DMSVIVA 2021

Proceedings of the 27th International DMS Conference on Visualization and Visual Languages

June 29 to 30, 2021
KSIR Virtual Conference Center
Pittsburgh, USA

DOI: 10.18293/DMSVIVA2021

Proceedings preparation, editing and printing are sponsored by KSI Research Inc.

PROCEEDINGS

DMSVIVA2021

The 27th International DMS Conference on Visualization and Visual Languages

Sponsored by

KSI Research Inc. and Knowledge Systems Institute, USA

Technical Program

June 29 to 30, 2021

KSI Research Virtual Conference Center, Pittsburgh, USA

Organized by

KSI Research Inc. and Knowledge Systems Institute, USA

ISBN: 1-891706-53-5
ISSN: 2326-3261 (print)
2326-3318 (online)
DOI: 10.18293/DMSVIVA2021

Additional copies can be ordered from:
KSI Research Inc.
156 Park Square Lane
Pittsburgh, PA 15238 USA
Tel: +1-412-606-5022
Fax: +1-847-679-3166
Email: dms@ksiresearch.org
Web: http://ksiresearch.org/ seke/dmsviva20.html

Proceedings preparation, editing and printing are sponsored by KSI Research Inc. and Knowledge Systems Institute, USA.

Printed by KSI Research Inc. and Knowledge Systems Institute, USA.

FOREWORD

On behalf of the Program Committee of the *27th International DMS Conference on Visualization and Visual Languages (DMSVIVA2021)*, we would like to welcome you. This conference aimed at bringing together experts in visualization, visual languages and distributed multimedia computing and providing a forum for productive discussions about these topics.

It is our pleasure to announce that by the extended deadline of 20 April 2021, the conference received 20 submissions. All the papers were rigorously reviewed by three members of the international Program Committee. Based on the review results, 8 papers have been accepted as regular papers with an acceptance rate of 40%. We would like to thank all the authors for their contributions. We also would like to thank all the Program Committee members for their careful and prompt review of submitted papers
.

We would like to thank the Steering Committee Chair Professor Shi-Kuo Chang for his guidance and leadership throughout organization of this conference. The assistance of the staff at KSI Research and Knowledge Systems Institute is also greatly appreciated, which made the review process smooth and timely.

Stefano Cirillo, University of Salerno, Italy; Program Co-Chair
Yang Zou, Hohai University, China; Program Co-Chair

DMSVIVA2021

The 27th International DMS Conference on Visualization and Visual Languages

June 29 and 30, 2021

KSIR Virtual Conference Center, Pittsburgh, USA

Conference Organization

DMSVIVA2021 Conference Chair and Co-Chairs

Mark Minas, Universität der Bundeswehr München, Germany; Conference Chair
Loredana Caruccio, Universiy of Salerno, Italy; Conference Co-Chair

DMSVIVA2020 Steering Committee Chair

Shi-Kuo Chang, University of Pittsburgh, USA; Steering Committee Chair

DMSVIVA2020 Steering Committee

Paolo Nesi, University of Florence, Italy; Steering Committee Member
Kia Ng, University of Leeds, UK; Steering Committee Member

DMSVIVA2020 Program Chair and Co-Chair

Stefano Cirillo, University of Salerno, Italy; Program Co-Chair
Yang Zou, Hohai University, China; Program Co-Chair

DMSVIVA2020 Program Committee

Danilo Avola, University of Rome, Italy
Rachel Blagojevic, Massey University, New Zealand
Andrew Blake, University of Brighton, UK
Paolo Bottoni, Universita Sapienza, Italy
Bernardo Breve, University of Salerno, Italy
Loredana Caruccio, University of Salerno, Italy
Maiga Chang, Athabasca University, Canada
Yuan-Sun Chu, National Chung Cheng University, Taiwan
Stefano Cirillo, University of Salerno, Italy
Mauro Coccoli, University of Genova, Italy
Francesco Colace, University of Salerno, Italy

Kendra Cooper, University of Texas at Dallas, USA
Gennaro Costagliola, University of Salerno, Italy
Andrea DeLucia, Univ. of Salerno, Italy
Mattia DeRosa, University of Salerno, Italy
Vincenzo Deufemia, University of Salerno, Italy
Martin Erwig, Oregon State University, USA
Larbi Esmahi, Athabasca University, Canada
Edoardo Fadda, Politecnico di Torino, Italy
Daniela Fogli, Universita degli Studi di Brescia, Italy
Rita Francese, University of Salerno, Italy
Ombretta Gaggi, Univ. of Padova, Italy
Angelo Gargantini, University of Bergamo, Italy
Angela Guercio, Kent State University, USA
Pedro Isaias, University of Queensland, Australia
Manuel J. Fonseca, University of Lisbon, Portugal
Jennifer Leopold, Missouri University of Science & Technology, USA
Alan Liu, National Chung Cheng Univeristy, Taiwan
Paolo Maresca, University Federico II, Napoli, Italy
Mark Minas, Universität der Bundeswehr München, Germany
Andrea Molinari, University of Trento, Trento, Italy
Eloe Nathan, Northwest Missouri State University, USA
Paolo Nesi, University of Florence, Italy
mnorth@kennesaw.eduouthern Polytechnic State University, USA
Michela Paolucci, University of Florence, Italy
Ignazio Passero, University of Salerno, Italy
Giovanni Pilato, Italian National Research Council, Italy
Giuseppe Polese, University of Salerno, Italy
Elvinia Riccobene, University of Milano, Italy
Michele Risi, University of Salerno, Italy
Peter Rodgers, University of Kent, UK
Domenico Santaniello, University of Salerno, Italy
Michael Wybrow, Monash University, Australia
Atsuo Yoshitaka, JAIST, Japan
IngTomas Zeman, Czech Technical University, Czech Republic
Yang Zou, Hohai University, China

Publicity Chair and Co-Chair

Jun Kong, North Dakota State University, USA; Publicity Chair
Bernardo Breve, University of Salerno, Italy; Publicity Co-Chair

DMSVIVA2021

Table of Content

Note: (S) denotes a short paper.

Keynote I

Extreme Visual Languages

Shi-Kuo Chang
Professor
School of Computer and Information
University of Pittsburgh
USA

Abstract

Extreme Visual Languages are visual languages that can communicate an integrated/holistic meaning of an application, a philosophy or even the entire spatial/temporal dimensions of a universe in a single or a few visual expressions. An example is the aliens' visual language in the movie "The Arrival" where each visual sentence expresses the continuum of space and time. Another example is a visual diagram illustrating the spread of Covid-19 in the world. A third example is the Ying/Yang icon that appears in the national flag of the Republic of Korea (and many other places). A fourth example is the red flag. A natural extension to include sound and other media leads to *Extreme Multimedia Languages*. In this talk I will discuss extreme visual and multimedia languages, both abbreviated as *XVL*s, including empirical and comparative studies of existing XVLs and the investigation of syntax, semantics, dynamics and application integration of existing and/or new XVLs and related topics.

About the Speaker

Dr. Chang is the founder and President of KSI Research. Dr. Chang received the B.S.E.E. degree from National Taiwan University in 1965. He received the M.S. and Ph.D. degrees from the University of California, Berkeley, in 1967 and 1969, respectively. He was a research scientist at IBM Watson Research Center from 1969 to 1975. From 1986 to 1990, he was Professor and Chairman of the Department of Computer Science, University of Pittsburgh. He is currently Professor and Director of Center for Parallel and Distributed Systems, University of Pittsburgh. Dr. Chang is a Fellow of IEEE. He has been consultant for IBM, Laboratories, Standard Oil, Honeywell, and Naval Research Laboratory. His research interests include knowledge-based systems, pictorial information systems, visual languages and computer vision. Dr. Chang has published over two hundred papers, and written or edited eight books. His books, Principles of Pictorial Information Systems Design, (Prentice-Hall, 1989), and Principles of Visual Programming Systems, (Prentice-Hall, 1990), are pioneering advanced textbooks in these research areas.
Dr. Chang is the Editor of the Journal of Visual Language and Computing published by KSI Research, and the Editor of the International Journal of Software Engineering & Knowledge Engineering published by World Scientific Press.

Keynote II

Aesthetic Measurement of Paintings

Kang Zhang
Professor
United International College UIC
ZhuHai, China

Abstract

This talk will first introduce the recently emerging interdisciplinary research topics of computational aesthetics and aesthetic computing, and discuss their difference and complementary roles. We will then focus on computational aesthetics, in particular, on the aesthetic measurement of paintings. Based on the mathematician Birkhoff's 1933 seminal theory of aesthetic measurement, our first work was to measure the white space usage in Master Wu Guanzhong's paintings, and then the complexity of Chinese ink paintings. Our latest work is to compute the visual order of both Chinese ink paintings and Western styled oil painting and compare their commonality and differences, possibly influenced by different cultural backgrounds. If time permits, we will mention our recent efforts in generating abstract images of styles of Kandinsky, Miro, Pollock, and Picasso, and information visualization.

About the Speaker

Kang Zhang is a Professor at UIC and Professor Emeritus of Computer Science at the University of Texas at Dallas. He was a Fulbright Distinguished Chair, an ACM Distinguished Speaker, and held academic positions in the UK, Australia, and China. Zhang's research interests include computational aesthetics, visual languages, and software engineering; and has published 8 books, and over 100 journal papers. He is on the Editorial Boards of *Journal of Big Data, The Visual Computer, Journal of Visual Language and Computing*, *International Journal of Software Engineering and Knowledge Engineering*, *Visual Computing for Industry, Biomedicine and Art* and 《软件学报》.

Long-Term Prediction of Bikes Availability on Bike-Sharing Stations

Daniele Cenni, Enrico Collini, Paolo Nesi, Gianni Pantaleo, Irene Paoli
Distributed Systems and Internet Technologies Lab, Department of Information Engineering, University of Florence, Florence, Italy, Https://www.disit.org, Https://www.snap4city.org (<name>.<surname>@unifi.it)

Abstract**—** Bike-sharing systems have been adopted in many cities as a valid alternative to traditional public transports since they are eco-friendly, prevent traffic congestions, reduce the probability of social contacts which are probable in public means. On the other hand, they also bring some problems which include the irregular distribution of bikes on the stations/racks/areas and the difficulty of knowing in advance their status with a certain degree of confidence, whether there will be available bikes at a specific bike-station at certain time of the day, or a free slot for leaving the rented bike. Therefore, providing predictions can be useful for improving the quality of service. This paper presents a technique to predict the number of available bikes and free bike-slots in bike-sharing stations (which is still the best solution for e-bikes). To this end, a set features and predictive models have been compared to identify the best models and predictors for long-term predictions. The solution and its validation have been performed by using data collected in bike-stations in the cities of Siena and Pisa, in the context of Sii-Mobility National Research Project on Mobility and Transport and Snap4City Smart City IoT infrastructure. The Gradient Boosting Machine (GBM) offers a robust approach for the implementation of reliable and fast predictions of available bikes in terms of flexibility and robustness with respect to critical cases, producing long-terms predictions in critical conditions (when available bikes are few).

Keywords- available bikes prediction, bike-sharing, machine learning, prediction models, smart city.

I. Introduction

In recent decades, the city has become an increasingly large and complex body. The number of inhabitants living in urban areas is increasing. Today, about 55% of the world's population lives in urban areas, and the figure is expected to rise to 68% in 2050, according to the "World Urbanization Prospects 2018", published by the United Nations Department of Economics and Social Affairs [1]. Today, transportation is one of the most important causes of certain gas emissions and thus of air pollution. In this context, bike-sharing systems may represent a part of the solution. Bike-sharing systems are widely used in many cities, offering a more sustainable alternative to public transport and reducing congestion. The station can be smart when they are capable to detect the presence of the bike, their status, and can release the bike. The alternative could be floating bike-sharing in which the bikes are more intelligent, and capable to communicate with the central management servers their position, etc., such as Mobike solution. Floating solution are still not very effecting in the case of e-bike since the recharge can be easier on racks.

In the context of this article, the solution with simple bikes (even e-bike) and smarter stations is addressed. The bikes can be typically released at any station providing that a free slot is available, this may create discomfort to the users when the station is full, and the user has to move to next and return by walk. One of the problems of bike-sharing is related to the irregular distribution of bikes among the various stations and the impossibility to know with a certain confidence to find a at least a bike at a desired station in a precise time slot of the day, or just few minutes in advance. The same for the possibility to find a free slot to leave the bike. Therefore, predicting the availability of bikes (as well as free slots) per station over time can be useful for managing the demands for bikes per station and to perform the redistribution in advance [2].

In recent years, many researchers have studied urban bike-sharing systems, mainly on four main areas of interest.

The first area is the *design of Bike-Sharing Systems.* In [3], a mathematical model has been proposed to determine the number of docking stations needed, their locations and the possible structure of the cycle path network, as well as models to make predictions about possible routes taken by users between stations of origin and destination. The second area is related to the *analysis of the behaviour and dynamics of a Bike-Sharing system.* In [4] and [5], clustering and forecasting techniques have been used on the network of bike-sharing stations in Barcelona to obtain useful information to describe the city's mobility. In [6], the authors interpreted the system as a dynamic network by analysing how bicycle flows distribute spatially along the network. In [7], different bike-sharing services are analysed highlighting the differences in bike flows and routes.

The last area concerns the prediction of bikes availability [8]. In [4], four different predictive models to estimate the availability of bikes in stations have been compared. The authors used a Bayesian network to predict the status of a bike-station (full, almost empty or empty) using bike-station information and providing predictions at 2 hours, with an accuracy of 80%. In [5], ARMA (AutoRegressive Moving Average) models has been used to predict the number of vacancies one hour in advance, while in [1], the authors presented a model system for predicting bike traffic of a bike-sharing network in Lyon.

DOI Reference number: : 10.18293/DMSVIVA2021-001

A. Article Overview

The **main contribution of this paper** consists in presenting a solution for real-time prediction of the available bikes on bike-sharing stations, and thus of the number of free slots by knowing the size of the station and the number of broken bikes. To this aim, a model has been identified to predict the availability of bikes 24 hours in advance (long-term predictions) with a resolution of 15 minutes, and thus also the free slots in the stations. Prediction of available bikes is a non-linear process whose dynamic changes involve multiple kinds of factors, coming from the context. To this end, the solution has been obtained by taking into account different cities, and locations, and despite the changes in Siena and Pisa in both cases the identified features and model have been the same, thus demonstrating the validity of the derived results. The precision obtained for long terms prediction have been much better than those provided in the literature.

The solutions have been implemented in the context of Sii-Mobility project and infrastructure (national mobility and transport smart city project of Italian Ministry of Research for terrestrial mobility and transport, http://www.sii-mobility.org) solution based on Km4City model (https://www.km4city.org) and Snap4City tools [9], [10], [11]. Sii-Mobility project aimed at defining solutions for sustainable mobility, suggesting bikes availability status to users at least 15 minutes/1 hour in advance to allow them to take a conscious decision, and maybe change their own plan. As a result, the solution has been capable to produce reliable prediction even 24 hours in advance.

The paper is structured as follows. Section II provides a description of the bike-sharing data, and their characterization in terms of clustering in groups. In addition, the identification of several features at the basis of the predictive models is reported. In Section III, the machine learning approaches adopted to identify and validate the predictive models and framework are presented. Conclusions are drawn in Section IV.

II. DATA DESCRIPTION AND FEATURE IDENTIFICATION

As mentioned in the introduction, the main goal was to find a solution to predict the bikes availability in each bike-station (and by knowing the size of the bike-station and the number of broken bikes on rack, we can derive the number of free slots). Typically, the status of each station is checked and registered by the server every 15 minutes. The data refers to 15 stations located in the municipality of Siena and 24 located in Pisa. In order to understand the typical time trend H24 (multiple seasonality may be present, daily, weekly and seasons over year) of bikes availability per station located in Siena and Pisa cities. Since the service is evolving quite rapidly over time, the seasonal trends taken into account are those daily and weekly. We taken into account data from June 2019 to January 2020 for Siena stations, and from December 2019 to March 2020 for Pisa stations. A clustering approach has been applied in order to classify Pisa and Siena stations based on their trend of bikes availability, which is also correlated to the typical services in the neighbourhoods. In detail, the K-means clustering method has been applied to identify clusters. In K-means clustering, there is an ideal center point that represents a cluster. The clustering has been performed on the basis of the time trend H24, considering the normalized trend of bikes availability measures. The optimal number of clusters resulted to be equal to 3, and it has been identified by using the Elbow criteria [12]. In particular, each cluster represents a group of stations. The stations/racks belonging to **Cluster 1** are typically characterized by a decrement of bike availability at lunchtime, and are mainly located close to the railway stations, airport, etc. Bike racks belonging to **Cluster 2** are typically positioned in the central area of the cities and are characterized by an increment of the availability of bikes in the central part of the day (lunch hours, since most of the people are parking their bikes to get lunch). **Cluster 3** presents an almost uniform trend in the bike availability and bike racks are mainly positioned in the peripheral areas of the city.

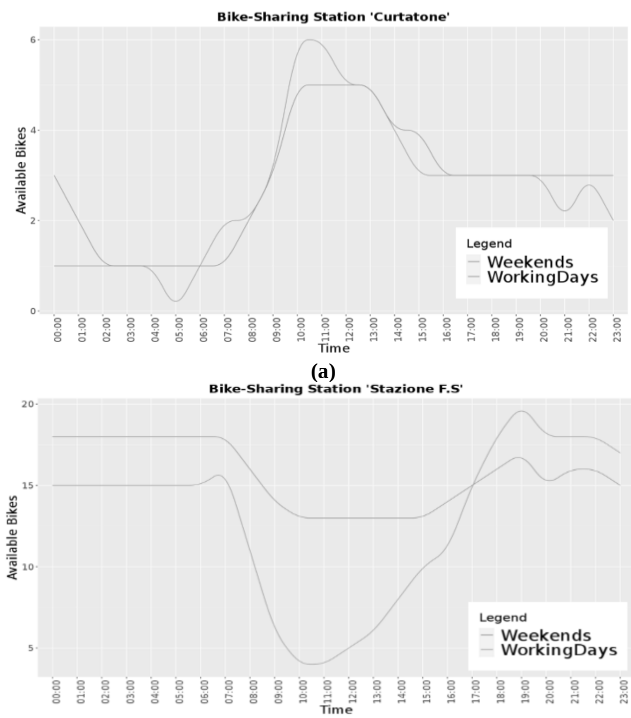


Figure 1. Working days/weekend trends of the (a) "Curtatone" bike-sharing stations in Siena and (b) "Stazione F.S" stations in Pisa municipality

For example, in Siena municipality "*Terminal Bus*" station that is a bike-sharing station positioned near by the train station in Siena, "*Ospedale*" station positioned near the Hospital, "*Due Ponti*" station positioned near the bus station/terminal, "*Curtatone*" station positioned near the stadium and "Napoli" positioned in residential areas. In Pisa municipality "*Comune Palazzo blu*" station is near the municipality building, "*Ospedale Cisanello*" station positioned near the Hospital, "*San Rossore F.S*" station

positioned near the train station San Rossore, "*Stazione F.S*" station positioned near the central train station and "*Marchesi*" station near the educational buildings.
Moreover, we have also detected some changes in the typical trends from working days and weekends as shown in **Figure** 1. **Figure** 1**(a)** reports the comparison between the trendsd for working days and weekends for "Curtatone" station in Siena, while **Figure** 1**(b)** shows the trends of working days/weekends for "Stazione F.S" in Pisa.

A. *Feature Identification*

With the aim of developing a prediction model, a set of features have been identified and tested. We analysed a large number of features for selecting the best, with the aim of conquering a higher precision with respect to the state-of-the-art solutions mentioned above. So that the hypothesis has been verified in the results reported in this paper for the case of bike-station status predictions. Features belonging to the ***Baseline* (time series)** category refer to aspects related to the direct observation of bike status over time as in [13]. Date and time when measures are taken, working day or not, number of bikes on racks, etc., belong to this category. Typically, the values are recorded every 15 minutes. Please note that the temporal window for the training is not based only on 15 minutes, but the measures over months are taken every 15 minutes. Features describing the **differences over time**. Usually, the trend of number of bikes is similar from one week to another for the same day (e.g., Monday to prev/next Monday), in the same month for example. Thus, two other features have been included in the model for capturing: (i) the difference between the number of bikes captured at the same time in the previous time slot of previous week (dPw); (ii) the difference between the number of bikes captured at the same time in the successive time slot of previous week (dSw). The value of the number of bikes related to the previous week respect to the observed one at the same time has been considered as additional feature (PwB).

Category	Feature
Baseline-Historical	Available Bikes in the past
	Time, month, day
	Day of the week
	Weekend, Holiday
	Previous week (PwB)
	Previous day (PdB)
Diff. from actual values and prev. observations	Previous observation's difference of the previous week (dPw)
	Subsequent observation's diff. of the previous week (dSw)
	Previous observation's difference of the previous day (dPd)
	Subsequent observation's difference of the previous day (dSd)
	Previous observation's difference between the previous week and two weeks earlier (dP2w)
	Previous observation's difference between the previous day and two days earlier (dP2d)
Real-time weather and weather forecast	Max Temperature
	Min Temperature
	Temperature
	Humidity
	Rain
	Pressure
	Wind Speed
	Cloud Cover Percentage
	Sunrise
	Sunset

Table 1. Overview of the feature used in the prediction models

Features belonging to the **real time weather and weather forecast** *collected* every 15 minutes (i.e., temperature, humidity and rainfall). Please note that, according to our analysis, the significant values for the weather are those related to the current time and the hour just before measured bike availability time. For example, in order to predict the number of available bikes at the rack at 3 pm, the weather features at 2 pm and at the current time are relevant. Thus, the weather conditions influence the decisions on using the bike or other transportation means. Similarly, the weather forecast influences the plan to get the bike.
The data collected from historical values of each bike rack are in practice all the data in the learning window (several weeks or months) of the past has described in Section II. For each time sample, the features of **Table** 1 are collected and when needed estimated and stored. When the long terms prediction is performed 24 hours in advance, the training/learning is performed once a day for each bike rack. To perform the training more often is not producing better results, and it is very computational expensive.

III. PREDICTION MODELS

In this section, number of machine learning techniques are considered and compared to identify the best solution to predict the bikes availability at bike-sharing stations/racks and to identify the features that could be the best predictors for the purpose. During our research study a number of techniques have been discharged since they did not produce satisfactory results -- e.g., Bayesian Regularized Neural Network that achieves an R-squared (https://en.wikipedia.org/wiki/Coefficient_of_determination) about 0.4 for each bike-sharing station. On the other hand, among the techniques we have presented here the comparison of the most effective solutions, which are **Random Forest** (RF) [14], **Gradient Boosting Machine** (GBM) [15] and the more traditional statistical approach such as **Auto-Regressive Integrated Moving Average** approach (e.g., ARIMA) [16]. The accuracy of each model has been evaluated in terms of R-squared, MASE (Mean Absolute Square Error), RMSE (Root Mean Square Error), and processing time considering the representative station per cluster. The RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(obs_i - pred_i)^2}{n}}$$

The MASE is calculated as follows:

$$MASE = mean\ (|q_t|), \qquad t = 1, \dots, n$$

and

$$q_t = \frac{obs_t - pred_t}{\frac{1}{n-1}\sum_{i=2}^{n}|obs_i - obs_{i-1}|}$$

where $obs_t = observation\ at\ time\ t$, $pred_t = prediction\ at\ time\ t$, n is the number of the values predicted over all test sets (96 daily observations per 7 days). The MAE (Mean Absolute Error) is estimated as follows:

$$MAE = \frac{\sum_{i=1}^{n}|obs_i - pred_i|^2}{n}$$

Note that, MASE is clearly independent on the scale of the data. When MASE is used to compare predictive models, the best model is the one presenting the smaller MASE.

A. Experimental Results

In the general framework, three different approaches were tested, i.e., RF, GBM, and ARIMA models applied on the features presented in **Table 1**. In detail, for GBM a regression tree with a maximum depth of 9 was used as a basic learner and the total number of trees was increased to 500 while the minimum number of observations in each leaf was increased to 5. The learning-rate has been set to 0.1. Note that, determining the optimal (hyperparameter) settings for the model is crucial for the bias-reduced assessment of a model's predictive power. The choice of GBM parameters has been obtained by a hyperparameter tuning implementation. Different combinations of parameter values have been tried on dataset (see **Table** 2).

Hyperparameter	Type	Start	End	Default
n.tree	Integer	100	10000	100
shrinkage	Numeric	0.01	0.3	0.1
interaction.depth	Integer	3	10	1
bag.fraction	Numeric	0.1	1	0.5

Table 2. Hyperparameter ranges and types for GBM model

The RF has been set with number of trees composing the forest equal to 500 and the candidate feature set equal to 1/3 of the number of the data set variables.

The ARIMA model has been executed as multi-step forward with updated iteration technique: the forecast was computed one hour in advance. Then, the training set is updated with the observations recorded in the predicted hour and a new forecast is executed for the next hour. The comparison of the needed processing time per each bike-sharing station, among the models considered above, is also relevant and it is reported in **Table 5.**

ARIMA Model Results		
Siena Bike-Sharing Stations	**MASE**	**RMSE**
Curtatone	1.23	1.58
Napoli	0.51	1.10
Terminal Bus	1.15	1.32
Due Ponti	0.52	1.15
Ospedale	0.23	1.39
Pisa Bike-Sharing Stations	**MASE**	**RMSE**
C.Marchesi	0.51	1.21
Comune Palazzo Blu	0.27	1.33
Ospedale Cisanello	0.86	1.13
San Rossore F.S.	1.01	1.22
Stazione F.S.	0.10	2.22

Table 3. ARIMA multi-step forward (short term online predictions) with updated iteration results in terms of MASE and RMSE per station in Siena

Table 3 shows the results for the ARIMA model for the main bike-sharing stations in Siena and Pisa. ARIMA model cannot be used for medium-long term forecasts due to the large errors produced. An approach to cope with this problem could be to apply the forecasting ARIMA technique as a multi-step forward to make 24-hour predictions (96 time slots). In other words, compute 24 forecasts (i.e., 1 hour in advance per 24 times): the real observations recorded in that hour (four slots of 15 minutes) are inserted into the training set, and the prediction for the next hour is computed with the new information. Therefore, the model needs to be trained every hour (see **Table 5**), so that 24 times per day per 15/20 bike-sharing stations per city, which is computationally more expensive than the others. For this reason, the solution has been discharged, despite to the fact that for the ARIMA, the obtained accuracy in terms of MASE is better than those obtained by machine learning techniques presented in **Tables 4** and **6**. Please remind that, the goal was to find a computationally viable solution to make satisfactory predictions in terms of precision for several different cases. As a further step, the comparison has been focused by considering RF and GBM on the whole set of bike-sharing stations in Siena (**Table 4**), exploiting all features presented in **Table 1**. The comparison of the predictive models has been estimated on a training period of 7 months. **Figure 4** reports the GBM predicted values vs real in a 96 time slots (24 hours) for "Curtatone" station in Siena city, which is a typical result.

"Curtatone" Station	**RF**	**GBM**
R2	**0.86**	0.78
MAE	2.42	**2.41**
MASE	0.82	**0.79**
RMSE	**2.90**	3.16
"Napoli" Station	**RF**	**GBM**
R2	**0.92**	0.81
MAE	2.22	**1.35**
MASE	1.10	**0.87**
RMSE	1.50	**1.45**
"Terminal Bus" Station	**RF**	**GBM**
R2	**0.91**	0.89
MAE	3.51	**3.37**
MASE	2.62	**2.52**
RMSE	2.2	**2**
"Due Ponti" Station	**RF**	**GBM**
R2	**0.96**	0.95
MAE	2.22	**1.85**
MASE	1.10	**0.92**
RMSE	2.60	**2.35**
"Ospedale" Station	**RF**	**GBM**
R2	**0.87**	0.79

MAE	**2.23**	2.35
MASE	**0.88**	0.92
RMSE	2.59	**2.35**

Table 4. Machine Learning Models results and comparison for different Siena stations

MASE and RMSE error measures have been estimated on a testing period of 1 week after the 7th January 2020. This comparison has highlighted that in Siena stations GBM approach achieved better results in terms of MASE, MAE and RMSE even if RF turned out to be the better ranked in terms of R-squared.

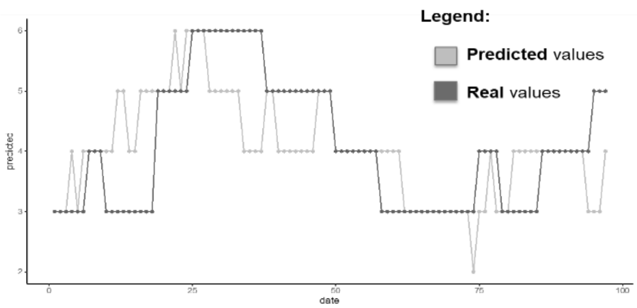


Figure 2. GBM predicted values vs real in a 96 time slots (24 hours) for "Curtatone" station in Siena

Table 5 shows that almost all the approaches may produce predictions every hour for the next hour in a reasonable estimation time. On one hand, in order to produce satisfactory predictions, the ARIMA approach needs to re-compute the training every hour (even if the online training can be seen as an alternative it is also a computational cost). This is a quite expensive cost of about 30s for each bike-sharing station, due to the fact that the charging stations can be hundreds. On the other hand, machine learning models (i.e., GBM and RF) provide predictive models with 96 values in advance with quite satisfactory results, they produce better results with less effort with respect to ARIMA. GBM processing time is quite low and results in terms of error measure are better respect the RF. GBM model can be considered the best solution for a real-time application.

Processing Time	ARIMA	RF	GBM
Average training time	30.9 sec	410.3 sec	**21.8 sec**
Training frequency	1 time per hour	1 time per day	1 time per day
Training period	1 months	7 months	7 months
Forecast window	1 hour	1 day	1 day

Table 5. Forecasting Models comparison in terms of processing time

Figure 3 shows the GBM model feature relevance [15] for "*Curtatone*" bike-sharing station in Siena (a similar figure could be presented for Pisa while it has been omitted for the lack of space). The most important features are those related to *Time*, *Day of the Week* and Weather category as *Air pressure, humidity and temperature.* The same features relative influence has been obtained for the other stations in Siena municipality. This result shows that exists a strong dependence between bicycle use and weather conditions. For this reason, the only use of data related to bike-sharing station are not enough and cannot produce satisfactory and flexible results. In addition, to confirm the strong dependence with weather features, a new model exploiting baseline feature only has been trained for "*Curtatone*" station in Siena. Results show that the R2 decreases to 0.48 while error measure RMSE increase to 2.90.

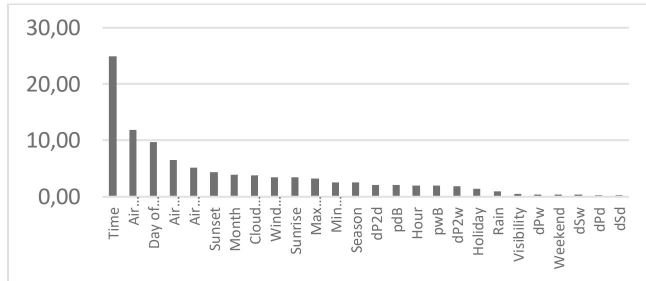


Figure 3. GBM model features relative influence for "*Curtatone*" station considering all features presented in Table 2.

The experiment above has been performed without applying feature selection. As additional analysis, a recursive feature elimination approach (RFE) based on RF was applied as a dimensionality reduction measure. The RFE technique implements a backward selection of the features by ranking their importance to an initial model using all predictors. The RFE selection method [17] is a recursive process that grades feature according to a certain degree of importance in order to filter unnecessary features and achieve a better performance of GBM model.

The RF-RFE optimization procedure has been applied to find the best performing subset of features, before applying the GBM model. The RF-RFE method identified a subset of 18 features (see **Figure** 4), in particular: *Time*, *Air Pressure*, *Day of the Week*, *Wind Speed*, *Cloud Cover Percentage*, *Hour*, *Air Temperature*, *Air Humidity*, *Sunset*, *Max Temperature*, *Min Temperature*, *Sunrise*, *Visibility*, *dP2d*, *dP2w*, *pwB*, *pdB.*

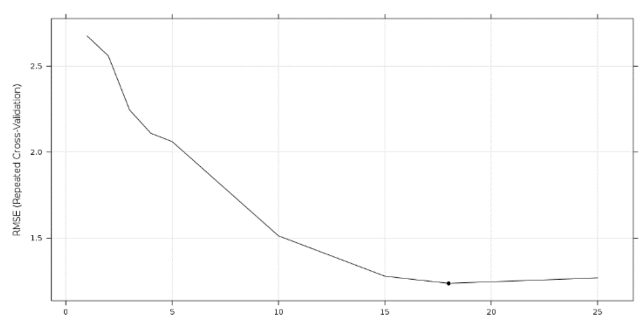


Figure 4. RF-RFE performance profile across different features subset sizes in terms of RMSE (the black point represents the best subset size of features, that is equal to 18)

Results from the GBM model trained on the identified subset of features have not shown a better accuracy with respect to those presented in **Table** 4. In conclusion, the features presented in **Table 1**, are those strictly necessary to obtain the best GBM performance in terms of R-square MASE and RMSE. In order to produce predictions, two GBM models have been trained to test the capabilities in predicting bike rack status in the next 15 and 30 minutes, respectively.

Predictions	**MASE**	**RMSE**
15 minutes	1.02	2.8
30 minutes	1.27	2.98

Table 6. GBM model predictions 15 and 30 minutes for "*Curtatone*" station, where the MASE and RMSE have been computed with respect to the true values by using 10 consecutive predictions.

Table 6 reports GBM models results for predictions showing that the RMSE does not improve much (passing from 2.9 to 2.8 for the 15 minutes), and the MASE seem to be worst with respect to the results presented in **Table 4** (the precision decreases with the distance from the last actual value, such as for ARIMA solutions). In these cases, the online learning has been performed at every time slot of 15 minutes, which is very expensive. It means that, in a perspective of online prediction and similar cost of a traditional method as the ARIMA, which can be preferable since the results in terms of error measures (in particular in terms of MASE) are better (see **Table** 3).

The same machine learning models presented above have been applied and compared on the bike-sharing stations of Pisa, considering all the features presented in **Table** 1. Note that, the amount of data available for the city of Pisa is lower than for Siena municipality. The comparison of the predictive models has been estimated on a training period of 3 months (from 1st December 2020 to 1st March 2020). **Table 7** reports the results of the comparisons of RF and GBM models for five representative stations presented in Section II. Contrary to the results achieved for stations in Siena, RF approach achieved slightly better results in terms of R-squared, MASE, MAE and RMSE in all five representative stations in Pisa, while the training period for the RF model remains significantly longer than for GBM. MASE and RMSE error measures have been estimated on a testing period of 1 week after the 1st March 2020.

"C. Marchesi" Station	**RF**	**GBM**
R2	**0.90**	0.84
MAE	**2.43**	2.77
MASE	**0.78**	0.89
RMSE	**2.91**	3.36
"Comune Palazzo Blu" Station	**RF**	**GBM**
R2	**0.94**	0.91
MAE	**1.91**	2.25
MASE	**0.95**	1.16
RMSE	**2.32**	3.01
"Ospedale Cisanello" Station	**RF**	**GBM**
R2	**0.92**	0.91
MAE	**1.18**	2.02
MASE	**0.90**	1.01
RMSE	**2.32**	2.60
"San Rossore" Station	**RF**	**GBM**
R2	**0.91**	0.86
MAE	**3.59**	4.01
MASE	**1.16**	1.25
RMSE	**4.08**	4.94
"Stazione F.S" Station	**RF**	**GBM**
R2	**0.94**	0.92
MAE	**5.11**	5.65
MASE	**0.73**	0.92
RMSE	**5.21**	6.15

Table 7. Machine Learning Models results and comparison for different Pisa stations

Also, the feature relative influence of RF model for Pisa shown that weather category variables are no longer the most influential in forecasting available bikes. The most important features are those related to the previous days/week information: in order of importance, are *PwB*, *PdB*, *Time*, *Day of the Week*, *Hour*, *DP2d* followed by *Air Pressure*, *Air Humidity* and *Air Temperature*. In the case of "Stazione F.S", the error trend was also evaluated based on the different times of the day. The results are shown in the **Table 8**.

"Stazione F.S" Station	*MAE*	*MASE*	*RMSE*
Night	5.54	1.94	6.58
Morning	**4.83**	**0.56**	**5.98**
Afternoon	4.93	0.98	6.05
Evening	5.52	0.85	6.31

Table 8. RF results for "Stazione F.S" station in Pisa per times of the day

As additional result, the same model has been trained for "Stazione F.S" bike-sharing station exploiting baseline feature only. The trained model seems not to be much worse in terms of prediction errors than the RF model results presented in **Table 8.** The RF predicted values vs real value in 24 hours for "Stazione F.S" station in Pisa, number of free bikes is reported in **Figure 5**. Results in terms of R2, MASE, MAE and RMSE are respectively 0.88, 0.76, 4.67 and 5.54.

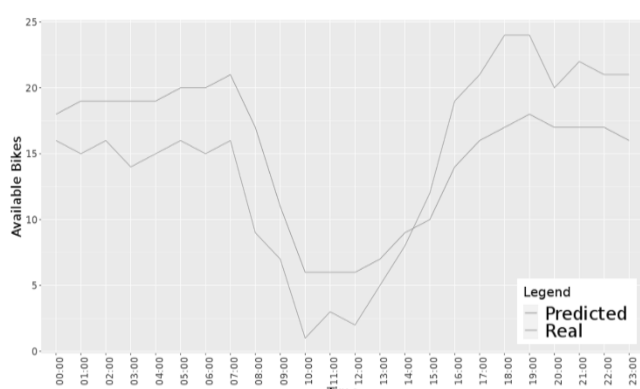


Figure 5. RF predicted values vs real in 24 hours for "Stazione F.S" station in Pisa, number of free bikes.

IV. Conclusions

In this paper, we proposed machine learning methods to predict bike availability for each station in bike-sharing systems. The proposed methods use a model which takes high dimensional time-series data from each station and uses real-time and forecast weather information as input to perform long term prediction the next 24 hours bikes availability for each bike-sharing station. The proposed solution demonstrated that in case of prediction (1 hour in advance), the ARIMA models may outperform in short time the predictions obtained using the RF and GBM algorithms. However, ARIMA model cannot be used for medium-long term forecasts because the iterative forecasting model should be trained at least 24 times per day per several bike-sharing stations per city. To this aim, RF and GBM algorithms have been considered as alternative finding a satisfactory computationally viable solutions to make medium-long term predictions that produce satisfactory results in terms of precision and able to suit for several cases.

In the models, we have considered several features, such as the *historical data, difference in days and weeks, and the weather conditions and forecast*. In almost all predictive models, the baseline/historical data and weather information have demonstrated high predictive capabilities in explaining the number of available bikes. The weather features have improved the accuracy of forecasting available bikes. Please note that, despite the changes in Siena and Pisa in both cases the identified features and model have been the same, thus demonstrating the validity of the derived results. The entire approach resulted to be very flexible and robust with respect of the sporadic lack of data samples. The predictive models can produce predictions 24 hours in advance, while they are provided on mobile applications, 30 minutes, 1 hour in advance directly, and if requested also a day in advance as possible general trend. The solution has been deployed as an additional feature on Smart City Apps in the Tuscany area to encourage sustainable mobility https://play.google.com/store/apps/details?id=org.disit.toscana .

Acknowledgment

The authors would like to thank the MIUR, the University of Florence and companies involved for co-founding Sii-Mobility national project on smart city mobility and transport. Km4City and Snap4City (https://www.snap4city.org) are open technologies and research of DISIT Lab. Sii-Mobility is grounded and has contributed to Km4City open solution. In addition, the authors would like to thank to Gabriele Bruni for his support in the early experiments of this research.

References

[1] Flandrin P. Robardet C. Rouquier J. Borgnat P., Abry P. and Fleury E. "Shared Bicycles in a City: a Signal Processing and Data Analysis Perspective," *Advances in Complex Systems,* vol.14, n.3, 2011, pp.415-438.

[2] Hulot, Pierre, Daniel Aloise, and Sanjay Dominik Jena. "Towards station-level demand prediction for effective rebalancing in bike-sharing systems." Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018

[3] Lin, Jenn-Rong, and Ta-Hui Yang. "Strategic design of public bicycle sharing systems with service level constraints," *Transportation research part E: logistics and transportation review,* vol.47, n.2, 2011, pp.284-294.

[4] Froehlich, Jon Edward, Joachim Neumann, and Nuria Oliver. "Sensing and predicting the pulse of the city through shared bicycling," *Twenty-First International Joint Conference on Artificial Intelligence,* 2009.

[5] Kaltenbrunner, Andreas, et al. "Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system," *Pervasive and Mobile Computing,* vol.6, n.4, 2010, pp.455-466.

[6] Flandrin P. Robardet C. Rouquier J. Borgnat P., Abry P. and Fleury E. "A dynamical network view of Lyon's Velo'v shared bicycle system," *Dynamics On and Of Complex Networks,* Volume 2. Birkhäuser, New York, NY, 2013, pp.267-284.

[7] Gupta S. Ma D. Bargar A., Gupta A. "Interactive visual analytics for multicity bikeshare data analysis," *The 3rd International Workshop on Urban Computing*, New York, USA, Vol. 45, 2014.

[8] Colace, Francesco, et al. "A multilevel graph approach for predicting bicycle usage in London area." Fourth International Congress on Information and Communication Technology. Springer, Singapore, 2020.

[9] C. Badii, P. Nesi, I. Paoli. "Predicting available parking slots on critical and regular services exploiting a range of open data, *IEEE Access*, 2018, https://ieeexplore.ieee.org/abstract/document/8430514/

[10] C. Badii, E. G. Belay, P. Bellini, D. Cenni, M. Marazzini, M. Mesiti, P. Nesi, G. Pantaleo, M. Paolucci, S. Valtolina, M. Soderi, I. Zaza. "Snap4City: A Scalable IOT/IOE Platform for Developing Smart City Applications," *Int. Conf. IEEE Smart City Innovation*, China 2018, IEEE Press. DOI: https://ieeexplore.ieee.org/document/8560331/

[11] C. Badii, P. Bellini, A. Difino, P. Nesi. "Smart City IoT Platform Respecting GDPR Privacy and Security Aspects," IEEE Access, 8 (2020): pp.23601-23623.

[12] Kodinariya, T. M., & Makwana, P. R. "Review on determining number of Cluster in K-Means Clustering," *International Journal*, vol.1, n.6, pp.90-95, 2013.

[13] Kim, Kyoungok. "Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations." Journal of transport geography 66 (2018): 309-320.

[14] Breiman, Leo. "Random forests," *Machine learning,* vol.45, n.1, 2001, pp.5-32.

[15] J. H. Friedman. "Greedy function approximation: A gradient boosting machine," *Annals of Statistics, vol.*29, n.5, pp.1189–1232, 2001.

[16] Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. "Time series analysis: forecasting and control," John Wiley & Sons, 2015.

[17] Kuhn I. Guyon, J. Weston, S. Barnhill and V. Vapnik, "Gene selection for cancer classification using support vector machines", Machine Learning, vol. 46, no. 1, pp. 389-422, 2002.

Regression-Based Model using Contextual and Personal information for filling Missing Values in Personal Informatics systems: A Case Study using Fitbit data-set

Nannan Wen
Computer Science Department
University of Pittsburgh
PA, USA
naw66@pitt.edu

ShiKuo Chang
Computer Science Department
University of Pittsburgh
PA, USA
schang@pitt.edu

***Abstract*—The potential of Personal Informatics data to support reflection is well explored within HCI. However, there is still work needed to understand better exactly how filling missing data support people in making sense of and gaining insight on their captured data. This paper aims to evaluate regression-based estimation models for filling missing data to study how it improves personal reflection in providing additional contextual information missed due to recharge and away from the device when an event occurs. Before we study the effect missing data has when users are trying to gain insight from the data they collected, we need to fill in the missing data used in the longitudinal study we plan to conduct. In this paper, we discuss the regression-based models for filling missing data in Personal Informatics during the data integration process and carefully review the characteristics of the data collected. Four regression models for predicting missing values combining context information and temporal and event pattern are being developed and evaluated through the data collected from a self-tracking device. A case study on a specific self-tracking device (Fitbit) is discussed as a representative example for the estimation model.**

***Index Terms*—PI systems, regression-based model, estimation, self-tracking, Fitbit**

I. Introduction

Over the years, people thrive to obtain self-knowledge. Recently, with the advancement in technologies, the Personal Informatics systems, defined as those that ”help people collect personally relevant information for self-reflection and gaining self-knowledge” [4] was widely adopted. The robustness of sensor-rich wearable devices and applications allows users to monitor users' behavior in a less obtrusive but more ubiquitous way. People employ tracking behavior for various reasons– for example, to set and achieve goals, document activities, or discover the correlated relationship between different factors.

Despite various intentions for the tracking behavior, challenges are plenty in making sense of the ever-increasing amounts of everyday self-tracking data retrieved across multiple domains due to limited support for fostering self-reflection and drawing personal insights. Studies show that most people, including Q-Selfers, fails to fully leverage personal data [8] [9] [22] [19] [23]. For example, common reasons for people to abandon their device include difficulty meeting their goals and disappointment with the level of information the device provides [22]. Moreover, it is challenging for nonprofessionals to interpret the visualizations, translate questions into data attributes [23]. Thus the data exploration and analytics capabilities for personal data analysis ”remains primitive,” leaving ”the heavy analytical lifting to the end-user” [24]

Although we can not expect the end-user to be an expert in data analytics, data exploration with visualization is a powerful way to facilitate the understanding of daily habitual patterns, and self-reflection [5]. Our closing goal is to shed some light on whether visualizing missing data would lead to a richer insight. As a first step, we evaluate the appropriate estimation methods for the missing data we collected. In the study, the missing data represented significant events that were not tracked due to low battery and away from the device when such an event occurs.

Regarding missing values, one of the approaches to deal with it is data imputation, which is a method that aims to replace the missing value in time series with some plausible values. Traditional machine learning-based imputation methods includes maximum likelihood Expectation-Maximization (EM) imputation [16], KNN based imputation [2], and Matrix Factorization (MF) [12]. Nevertheless, most of these statistical and traditional machine learning imputation methods underrate the time series's temporal information.

This paper primarily explores regression-based machine learning models to estimate the wearable-sensory time series proposed by [13] with activity patterns, contextual information, and personal characteristics as covariates. We present methods in replacing missing data, analyzing the role missing data plays in the reflection stage will be the subject of future work. The assumption we made here is that the sensor data is accurate; we will not consider the precision of the data collected by the wearable sensory; what we cared about is the missing data showed in the dashboard. To resolve the

DOI Reference number: : 10.18293/DMSVIVA2021-002

prediction of missing data, we aim at choosing the best practice estimation method for our evaluation. It is an essential step in our study as we study the effect missing data has in the reflection stage on PI systems. Different methods for data imputation are being considered, and a new methodology is being proposed and evaluated. In summary, the contributions of this paper are as follows:

- A novel approach combines contextual and personal information and wearable-sensory time series imputation with the regression-based model using limited data.
- To the best of our knowledge, no research has been done to discover the effect of missing data plays in personal reflection.

The remainder of this paper is organized as follows. We review related works in data imputation solutions in Section 2. In Section 3, we present the proposed data imputation method based on regression models. Section 4 describes the dataset we collected for evaluation and the comparison of imputation techniques under different conditions. Section 5 discusses the experimental result and based on our research questions. Section 6 describes the limitations we have and future work. Finally, Section 6 concludes the paper.

II. Related work

In this section, we review existing works for that researchers have been done on reflection and data imputation.

A. Data-Driven Self-Reflection

With the advancement of technology, the concept of reflection has earned more and more attention in the HCI community [25]. However, the term for reflection is often not clearly defined [25]. In the staged-based Personal Informatics model, reflection (or self-reflection) is described as a part of the personal data tracking process as an independent step [4]. However, we have little understanding of precisely what happened during this stage and what factors would facilitate reflection. The stage-based model proposed a step-by-step view of tracking, whereas Ploderer et al. describe two types of reflection: reflection through real-time feedback and reflection through aggregated feedback [21]. Taken all together, people can do different types of reflection based on which stage they are at, the time of receiving an intervention, and the type of feedback the system provides.

Reflection is often supported through personal visualization in PI systems, as interacting with visualization is an essential element of self-reflection with data [4]. Choe et al. describe self-reflection by including contextual information, confirmation, or contradiction of existing knowledge [10]. In our study, we consider missing data as part of contextual information and try to study its effect on self-reflection by visualizing it.

B. Imputation methods

a) Traditional Imputation: The most popular statistical imputation methods include mean substitution and median substitution, which require the signal to be close to static. Simultaneously, most wearable devices like heart rate or step count fluctuate over time. However, in PI systems, other data like weight, emotions do not have close relationships with fine-grained time series. Linear and stochastic regression is widely used in single imputation where it uses complete-data methods of analysis which single imputation allows [1]. In contrast, multiple imputations [3, 11, 16] is a predictive approach that aims to predict plausible imputations of the missing values and maintain the relationships between different variables simultaneously.

b) Time Series Imputation: Various deep learning techniques, such as recurrent neural networks (rNNs) [6, 13, 17], aims to predict fine-grained time series data, like heart rate and network flow imputation, from which patterns for the signal can be derived easily given the volume of input. Specifically, Lin et al. [13] studied the time series imputation using GAN with contextual information, which we would like to adopt, to predict missing values in heart rate signals. Wu et al. [17] developed a deep learning framework with transfer learning of the personalized factor when predicting heart rate sequence in Fitbit.

Motivated by these approaches, this work is built upon the regression-based multiple imputation model. Further, it incorporates the personalized temporal properties for each participant for data imputation for a longitudinal study in the reflection stage for PI systems.

III. Methods

A. Notation

Following the same definition as Wu et al. [17] and Lin et al. [13] for time series data, the data collected from wearables are mainly user-specific, here we use $U = \{u_1, ..., u_i, ..., u_I\}$ to denote the participant population in the data, where I is the number of participants. The collected dates are defined as a vector $t_i = \{t_i^1, ..., t_i^j, ..., t_i^{J_i}\}$ in a chronological order, where t_i^j denotes the date collected from participant u_i on $j-th$ day, and J_i is the total number of collected days. The time series collected on t_i^j is presented as $x_i^j = \{x_i^{(j,1)}, ..., x_i^{(j,\tau)}, ..., x_i^{(j,T)}\}(x_i^j \in \mathbb{R}^T)$. Each measurement $x_i^{(j,k)}$ is associated with a timestamp information.

B. Temporal Properties (TP)

For each instance collected in a wearable-sensory device (e.g. Fitbit) there will be a timestamp associate with it which allows us to order them chronologically, generating sequential properties (e.g. trend, frequencies) where adjacent timestamps usually have a higher similarity than distant ones. Here we define one of the sequential properties for user i as follows: $TP_i = \{(x_i^{j,T} - x_i^{(j,T-1)}), ..., (x_i^{j,T-k} - x_i^{(j,T-k-1)}), ..., (x_i^{j,T-\sigma} - x_i^{(j,T-\sigma-1)})\}$, where σ denotes the time gap used to calculate the difference between two consecutive days.

C. Contextual information (CI)

To exploit the contextual information around the missing intervals, we further incorporate the previous instance collected for the same participant in our model. Given the known data-set in previous dates for the target missing instance, we assume that it may involve extra features to assist the missing part imputation in the target instance. $CI_i = \{(x_i^{(j,T-1)}), ..., (x_i^{(j,T-k-1)}), ..., (x_i^{(j,T-\sigma-1)})\}$.

D. Linear & Polynomial Regression (LiR)

To construct the linear models, we applied ordinary least squares (OLS) regression analysis on the Fitbit dataset; the temporal and contextual properties extracted will be used as predictors for the missing values.

E. Logistic regression (LoR)

We implemented this model mainly because it can handle both dense and sparse input; the regression model uses L2, known as Ridge Regression, regularization. The loss function here is also known as the least-squares error (LSE).

$$Loss = LSE(y_i, \widehat{y_i}) + \lambda \sum_{i=1}^{N} w_i^2 \tag{1}$$

The model tries to minimize the error between the actual y and the predicted $\widehat{y}$ by adding the regularisation term as an optimization during the process.

F. Decision tree (DT)

The decision tree model builds the estimation model in the form of a tree structure. It first decomposes an extensive dataset into smaller and smaller subsets while, at the same time, an associated decision tree is developed incrementally. After the training process, there will be a tree with decision nodes and leaf nodes. A decision node has two or more branches, each representing values for the attribute tested. The leaf node represents an estimation on the numerical target—the topmost decision node in a tree that corresponds to the root node's best predictor. Decision trees can handle both categorical and numerical data. We use the mean squared error (RMSE) as a feature selection criterion for the supported criteria; we use the average of each terminal node's value to minimizes the L2 loss; we also applied different variances for turning the tree.

G. Random Forest (RF)

Random forest estimator consists of a massive number of individual decision trees on various sub-samples of the dataset. Each tree in the random forest gets perfectly trained on that particular sample data. In the case of the regression problem, the output utilizes multiple decision trees by averaging all the results to improve the predictive accuracy and control over-fitting. In short, A random forest is an ensemble technique with the use of multiple decision trees and a technique named Bootstrap and Aggregation, also known as bagging.

We implemented the regression models above and incorporated the TP and CI properties extracted as predictors when calculating the missing values.

We used a two-stage procedure to select the model for our research study. First, we use the original data set with missing data being deleted resulting in a complete-case data set denoted as $X_i^{'}$, then we incorporated the TP and CI properties extracted with the original dataset and denote it as $X_i^{''} = X_i^{'} * TP_i * CI_i$, we term the datasets as FD and FDTP respectively. Next, the regression-based prediction model is constructed on the two data sets by applying OLS regression analysis, and the prediction models are termed as ∗_FD, ∗_FDTP, and the ∗ can be replaced with LiR, LoR, DT and RF respectively.

IV. Experimental results

In this section, we illustrated the effectiveness of the proposed methods on the Fitbit dataset; we perform extensive experiments on the imputation task intending to answer the following research questions:

- **RQ1**: To what level does the proposed method gives a good estimation of our missing data?
- **RQ2**: How does incorporate TP and CI perform compared to without TP and CI as predictors?
- **RQ3**: How do the different tuning policies impact the model performance?
- **RQ4**: Which model is the best fit in visualizing the missing data used in the next step for our research?

A. Experiment Setting

a) Data: In our evaluation, we conducted experiments on Fitbit data. The details for the dataset are shown in Table. I. In the Fitbit data set, we have 34 unique participants with a period of 19 days. Thus it is a small data set, and there will be missing data in it. We would delete incomplete instances during the evaluation process of the models. The dataset will be divided into two subsets: a training set and a test set. We get the data set from a data dump of Fitbit, so it is a simulated dataset of our longitudinal study we would like to collect later on.

TABLE I
DESCRIPTIVE STATISTICS OF FITBIT DATA SET

variables	Statistics					
	N	***Mean***	***Median***	***StDev***	***Min***	***Max***
TotalSteps	396	7554	6895	5100	4	28497
TotalDistance	396	5.38	4.95	3.91	0	27.53
VActiveDistance	396	1.36	0.95	2.62	0	21.92
MActiveDistance	396	0.56	0.2	0.87	0	6.4
Calories	396	2284	2182	787	50	4526
TrackerDistance	396	5.31	4.94	3.91	0.0	27.53
LActiveDistance	396	0.19	3.41	0.88	0.0	12.51
SActiveDistance	396	0.02	0.0	0.009	0.0	0.1
VActiveMinutes	396	1.91	2.00	30	0.0	202
FActiveMinutes	396	1.34	6.00	21	0.0	141
LActiveMinutes	396	196	204	110	0.0	720
SMinutes	396	942	940	312	0.0	1440

b) Models: According to different regression-based models, we test our longitudinal data set on four models (i.e., LiR, LoR, DT, RF), Table. II shows the detail of these four models as well as the data properties we proposed.

TABLE II
DESCRIPTION ON MODELS

Models	Description
	Regression Models
LiR	OLS based linear regression model
LoR	Ridge Regression model
DT_i	Decision tree models, i stands for different tuning strategies
RF_i	Random Forest models, i stands for different tuning strategies
	Data Properties
FD	temporal closeness instance + contextual information
FD_{DT}	FD + daily trend sequence

B. Evaluation protocol

As evaluation metrics for the effort prediction model we applied the following methods. The overall goodness of fit was evaluated using Adjusted R^2 as:

$$R^2_{adj} = 1 - [\frac{(1-R^2)(n-1)}{n-k-1}] \quad (2)$$

Where n is the number of points in the data sample. k is the number of independent regressors. The accuracy of the models was evaluated using the Root Mean Square Error (RMSE) as:

$$RMSE = \sqrt{\frac{1}{z}\sum_i (v_i - \widehat{v}_i)^2} \quad (3)$$

Where $\widehat{v}$ and v are the predicted value and ground truth, respectively, z is the number of all predicted values.

V. DISCUSSION

A. Performance Comparison (RQ2, RQ4)

a) Overall Performance: In our experiments, we evaluate the performance of all compared methods over different settings of trend sequence and contextual information on the Fitbit dataset in predicting TotalSteps. Table. III shows the imputation average performance of all regression-based models using k-fold cross validation since we have limited data. From the RMSE score showed on the table, we have the following key observations: i.LiR model achieves the best performance over other methods in all cases. For example, the improvements of LiR over different conditions range from 11.6% to 90.7% for temporal closeness instance, from 19.3% to 93.2% for temporal closeness and trend sequence, and from 11.7% to 90.4% for temporal closeness and contextual information in terms of RMSE score. ii. Among all compared methods, the model that outperforms others is $LiR_{FD_{DT_1}}$ where temperal closeness and daily trend sequence are being applied to the Fitbit data set. iii. We also notice that the linear model always performs better than methods in all conditions, whereas the logistic model performs the worst in all the case, which means in later analysis, we can exclude variations of logistic models, and the liner model is the best fit for later research so far.

TABLE III
RMSE. THE SAMLLER, THE BETTER

Models	***RMSE***	**Adjusted** R^2
LiR_{FD}	69.68	0.969
LoR_{FD}	753.98	0.620
DT_{FD}	102.18	0.935
RF_{FD}	79.21	0.961
$LiR_{FD_{DT_1}}$	**66.27**	0.972
$LoR_{FD_{DT_1}}$	983.27	-1.675
$DT_{FD_{DT_1}}$	126.39	0.900
$RF_{FD_{DT_1}}$	82.13	0.957
$LiR_{FD_{DT_2}}$	75.52	0.965
$LoR_{FD_{DT_2}}$	790.93	0.91
$DT_{FD_{DT_2}}$	109.83	0.924
$RF_{FD_{DT_2}}$	85.58	0.958

b) Performance v.s. time gap: We performed experiments of all methods across different training and testing sets while fixing the corresponding gap size in calculating additional data properties. The gap size we used is daily based, that's based on the information showed on Fitbit dashbord as shown in Fig. 1. Given the gap size, from Table. III, we can observe that LiR model performs better than other compared methods in all the gap sizes, and we also note that the smaller the time gap we use when generating trend sequence the better the performance (e.g. FD_{DT_1} as 24 hours, FD_{DT_1} as 48 hours) for LiR and RF, however, this doesn't hold for LoR and DT models.

Fig. 1. Fitbit dashboard.

B. Model Analysis (RQ1, RQ3)

a) Parameter Study: We also conducted experiments to study the impact of different parameters, i.e. the depth of the decision tree. Except for the parameter being tested, all other parameters are set as the default. Fig. 2 shows the RMSE of the tuning parameter when others are controlled for. We notice

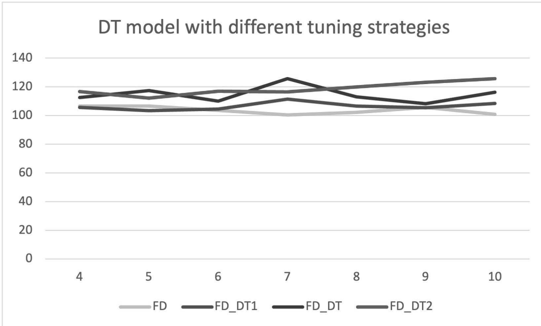


Fig. 2. Effects of tuning strategy on DT model.

that the tuning strategy have a relatively low influence on the performance of the DT model for our data set.

b) Model comparison: To better understand which model is the best fit for our later research, we evaluated the Goodness-of-Fit by reporting the adjusted R^2 value for each model. From the result showed in Table. III, we notice that the adjusted R^2 value is pretty similar across the models (LiR, DT, and RF), so theoretically all these three models can be used in later researches goals. However, we should exclude LoR model since it has a close to zero R^2 value.

VI. Limitations and Future work

However, our work has its limitations. First of all, the dataset we used are relatively small, and the dataset can only represent a simulated longitudinal study, we will solve this in our future works. Second, in this work, we talked only on filling missing data by data imputation, but in the next step of our research, we will visualize the estimated missing data and conduct user studies in understanding how it can affect users' self-reflection. Lastly, we didn't compare our methods with existing data imputation methods, for future work, we will compare our method with both traditional and top-notch methods in filling missing data.

VII. Conclusion

In this paper, we proposed a novel respect in understanding the factors might affect self-reflection. And presented our study on imputation methods in filling missing data in our Fitbit data-set intended for a longitudinal study which also has the temporal and contextual property. It addresses several limitations of using regression-based models to predict temporal data, including dealing with larger intervals of variable size. Our approach first introduced the historical time series, extracted personal features and contextual information as additional predictors. And then the model predicts the missing value based on the context information of the missing value. Extensive experiments demonstrate that our approach works best on the linear model, and the best fitting model for later research is $LiR_{DF_{DT_1}}$

References

[1] Patrician, Patricia A. "Multiple imputation for missing data." Research in nursing & health 25.1 (2002): 76-84.

[2] Shah, Jasmit S., et al. "Distribution based nearest neighbor imputation for truncated high dimensional data with applications to pre-clinical and clinical metabolomics studies." BMC bioinformatics 18.1 (2017): 1-13.

[3] Pampaka, Maria, Graeme Hutcheson, and Julian Williams. "Handling missing data: analysis of a challenging data set using multiple imputation." International Journal of Research & Method in Education 39.1 (2016): 19-37.

[4] Li, Ian, Anind Dey, and Jodi Forlizzi. "A stage-based model of personal informatics systems." Proceedings of the SIGCHI conference on human factors in computing systems. 2010.

[5] Prochoska, J. O., and W. F. Velicer. "The transtheoritical model of health behaviour change." Am J Health Promot 12 (1997): 38-48.

[6] Tlija, A., K. Wegrzyn-Wolska, and D. Istrate. "Missing-data imputation using wearable sensors in heart rate variability." Bulletin of the Polish Academy of Sciences. Technical Sciences 68.2 (2020).

[7] Li, Ian, Anind Dey, and Jodi Forlizzi. "A stage-based model of personal informatics systems." Proceedings of the SIGCHI conference on human factors in computing systems. 2010.

[8] Choe, Eun Kyoung, et al. "Understanding quantified-selfers' practices in collecting and exploring personal data." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2014.

[9] Epstein, Daniel A., et al. "A lived informatics model of personal informatics." Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015.

[10] Choe, E. K., & Lee, B. (2015). Characterizing visualization insights from quantified selfers' personal data presentations. IEEE computer graphics and applications, 35(4), 28-37.

[11] Huque, Md Hamidul, et al. "A comparison of multiple imputation methods for missing data in longitudinal studies." BMC medical research methodology 18.1 (2018): 1-16.

[12] O. G. Troyanskaya, M. Cantor, G. Sherlock, P. O. Brown, R. T. T. Hastie, D. Botstein, and R. Altman, "Missing value estimation methods for dna microarrays," Bioinformatics, vol. 17, pp. 520–525, July 2001.

[13] Lin, Suwen, et al. "Filling missing values on wearable-sensory time series data." Proceedings of the 2020 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2020.

[14] Krause, Robert W., et al. "Missing network data a comparison of different imputation methods." 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE, 2018.

[15] J. R. Quinlan, "Unknown attribute values in induction," in Proceedings of the Sixth International Workshop on Machine Learning, San Francisco, CA, USA, 1989, pp. 164–168.

[16] Shin, Tacksoo, Mark L. Davison, and Jeffrey D. Long. "Maximum likelihood versus multiple imputation for missing data in small longitudinal samples with nonnormality." Psychological methods 22.3 (2017): 426.

[17] Wu, Xian, et al. "Personalized Imputation on Wearable-Sensory Time Series via Knowledge Transfer." Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020.

[18] Q. Zhang and Z. Chen, "A weighted kernel possibilistic c-means algorithm based on cloud computing for clustering big data," International Journal of Communication Systems, vol. 27, no. 9, pp. 1378–1391, 2014.

[19] Li, I., Dey, A., & Forlizzi, J. (2010, April). A stage-based model of personal informatics systems. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 557-566). ACM.

[20] Cuttone, A., Petersen, M. K., & Larsen, J. E. (2014, June). Four data visualization heuristics to facilitate reflection in personal informatics. In International Conference on Universal Access in Human-Computer Interaction (pp. 541-552). Springer, Cham.

[21] Ploderer, Bernd, et al. "Social interaction and reflection for behaviour change." (2014): 1667-1676.

[22] Lazar, Amanda, et al. "Why we use and abandon smart devices." Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing. 2015.

[23] Rooksby, John, et al. "Personal tracking as lived informatics." Proceedings of the SIGCHI conference on human factors in computing systems. 2014.

[24] Fawcett, Tom. "Mining the quantified self: personal knowledge discovery as a challenge for data science." Big data 3.4 (2015): 249-

[25] Slovák, Petr, Christopher Frauenberger, and Geraldine Fitzpatrick. "Reflective practicum: A framework of sensitising concepts to design for transformative reflection." Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017.

Fashion Retail Recommendation System by Multiple Clustering

Pierfrancesco Bellini, Luciano Alessandro Ipsaro Palesi, Paolo Nesi, Gianni Pantaleo
University of Florence, DINFO dept, DISIT lab (https://www.disit.org)
emails: <name>.<surname>@unifi.it

Abstract— **Fashion retail has a large popularity and relevance, allowing customers to buy anytime finding the best offers and providing nice experiences in the shops. Consequently, Customer Relationship Management solutions have been enhanced by Information and Communication Technologies to better understand the behaviour and requirements of customers, engaging and influencing them to improve their buying experience, as well as increasing the retailers' profitability. Current solutions on marketing provide a too general approach, based on most popular or most purchased items, losing the focus on the customer centricity. In this paper, a recommendation system for fashion retail shops is proposed, based on a multi clustering approach of items and users' profiles in online and on physical stores. The proposed solution relies on association rules mining techniques, allowing to predict the purchase behavior of newly acquired customers, thus solving the cold start problems which is typical of current state of the art systems. The presented work has been developed in the context of the Feedback project founded by Regione Toscana, and it has been conducted on real retail company Tessilform, Patrizia Pepe mark. The recommendation system has been validated in store, as well as online.**

Keywords—recommendation systems, clustering, customer and items clustering composed.

I. Introduction

The competitiveness of retailers strongly depends on the conquered reputation, brand relevance and on the marketing activities they carry out. The latter aspect is exploited to increase the sales and thus a retailer, through marketing, should be able to encourage customers to buy more items or more valuable items. Today consumers tend to buy more on ecommerce and the COVID-19 situation also stressed this condition. Online shopping offers the possibility to buy at any time of the day; customers buy where they find the best offer, online as well as offline, and they are also influenced by an increasing amount of information from blogs, communities, and social networks. To retain a customer is therefore an extremely difficult achievement, and in some measure can get easily out of control.

Currently, ICT (Information and Communication Technologies) offers customer relationship management (CRM) solutions that enable the recording of user data profiles, from customer information to product details, to sales transactions. CRM comprises a set of processes and enables systems supporting a business strategy to build long term, profitable relationships with specific customers [36]. Customer data and information technology (IT) tools form the foundation upon which any successful CRM strategy is built. Swift [37] defined CRM as an "enterprise approach to understanding and influencing customer behavior through meaningful communications in order to improve customer acquisition, customer retention, customer loyalty, and customer profitability". However, CRM solutions on the market have a very traditional approach based on popular items or bundled offers, similar items or featured items and therefore often leave out the important customer centricity in any marketing strategy. In addition, there are IOT tools from the big vendors, which promise an evolved engagement on various levels, interacting with less queues, promotions, more involvement, assistance, but which are hardly triggered within companies, especially retail ones, which need more flexible solutions. To date, therefore, market solutions are unable to build and leverage user, historical, social, and behavioral profiles. Through transactions, retailers can generate knowledge about their consumer's behavior. One of the techniques receiving more attention from researchers to generate consumer knowledge, is machine learning, specifically clustering techniques. Clustering techniques are used to group customers by similarity. So, retailers can tailor marketing actions more effectively than general marketing actions. Understanding the reasons why consumers choose a specific item within the store is important to the retailer. In addition, knowing the consumer's needs through the factors that influence shopper's decision-making process is important for the store's own growth. This is what recommendation systems are all about. Recommendation systems are applications that assist users in finding items (products, services and information) that best suit their preferences [29]. The generated recommendations are personalized, in the sense that they have been generated for a user or a group of users, or not personalized (e.g. best-selling items, or selection of items). Non-personalized recommendations are typically not addressed by research.

State of the art recommendation systems today do not solve typical retail problems. Most of the retail companies today have both online and physical store customers assisted in purchasing by shop assistants. With the GDPR (General Data Protection Regulation) rules [22], often the customer demographics are differently collected in different areas and shops where different rules are adopted. Deep learning methods to improve accuracy are hard to be adopted for the scarcity of data. For example, in fashion retail shops most are anonymous transactions of one single item, and recurrent acquisition are performed every 8-12 months, for costs and seasonality aspects. Regarding classification methods, the multichannel nature of retailers provides data with different features and with many incomplete records, and such aspects don't allow to apply most of these methods. As for clustering methods, they use RFM (Recency – Frequency – Monetary Value) [5] LTM (Life-Time-Value) [6] and demographic values as input and do not take advantage of the intuition typical especially of deep learning of customer behavior towards items. Another problem related to the fashion retail industry is the seasonality of items. The life of most items is 6 months to 1 year.

In this paper, a recommendation solution in the context of fashion retail is proposed. In order to solve the above-mentioned problems of **cold start, computational complexity, low number of returns in the shops of fashion retails, more mediated interaction in the shop and more direct online, seasonality of products**, we realized a multi clustering approach by taking as input the RFM value of online and physical stores separately. To solve the problem of

DOI reference number: 10.18293/DMSVIVA2021-003

the seasonality of the products we have clustered the items. In addition, we enriched the input data with the customer behavior towards the items. To solve the cold starting problem of cluster-based recommendation systems, we have used the association rules mining technique that allows us to predict the purchase behavior of newly acquired customers. The work presented has been developed in the context of **Feedback** research and development project founded by Regione Toscana, Italy. Partners of the project are: VAR Group, University of Florence, TESSIFORM (*Patrizia Pepe* trademark), SICETELECOM, 3F CONSULTING and CONAD (External partner). The studies we are going to illustrate below were conducted on real retail company Tessilform: fashion retailer with online sales and different store in the world, mainly in Italy.

The paper is structured as follows. In Section II, related work on recommendation systems is presented. Section III described the system architecture adopted in Feedback solution. In Section IV, the proposed recommender systems based on multi clustering is presented using a number of subsections. The solution allowed to prepare the recommendations in advance and consume them in real time when the conditions occur, or for stimulating the customer to return in the shop. In Section V, the assessment and validation are reported. Conclusions are drawn in Section VI.

II. Related work

Recommendation techniques can be classified, according to the sources of knowledge they use, into six categories: [31], [23]. The **content-based approaches** recommend items by computing similarities among items and users through a set of features associated to them [39], [40]. For example, for a clothing item the considered features can be: the group (shirt, sweater, T-shirt, etc.), color, popularity, etc.; while for the users: demographic aspects, surveys answers, etc.

The **collaborative filtering-based** approach is based on the historical data of the user's interactions with the items, either explicit (e.g., user’s ratings) or implicit feedback (e.g., purchase, visit). The mathematical techniques used are the neighborhood method and the latent factor model [35]. The neighborhood method identifies relationships among elements or, alternatively, among users. The latent factor model sets a number of evaluation methods to characterize both items and users and it’s mainly based on the matrix factorization (for example the ratings-matrix). These approaches do not need a representation of the items, as they are based only on ratings, so they are the best recommendation systems in regard to scalability. The accuracy of recommendations increases as user interactions increase. They have cold start problems for both new items and new users. The **Demographic-based** approaches generate recommendations on the basis of the user's demographic profile (age, gender, education, etc.). They do not require a user ratings history, and they have cold start problems for new items. The **knowledge-based** approaches are based on the knowledge of item features which meet the users’ needs. They do not have cold start problems, require a broad knowledge of the domain and, in case of many items, they are very difficult to implement. The **community-based** systems make recommendation through the preferences of users' friends in contexts of social networks or communities. The basic concept is that a user tends to rely on recommendations from their friends instead of those of similar but anonymous users. This approach is very useful for cold-start recommendations. The **hybrid-based** recommender systems combine two or more recommendation approaches in different ways. Usually, considering two different approaches, the advantages of the former are used to mitigate the weakness of the latter.

The sources of knowledge are usually represented by three types of descriptors for: items, users and transactions (relations between user and item). Modern recommendation systems also use textual reviews, images and web page sequences, processed through data mining or deep learning methods, to generate recommendations.

The data mining methods for recommender systems can be summarized in three types of algorithms:

Classification. For example, the kNN classifier finds the closest k points (closest neighbors) from the training records. In [28] kNN has been implemented, with very good results in precise recommendations, to suggest short-term news to users. Decision Trees classifier works well when objects have few features, but in [27], [41] it has been shown that this technique can have low performance. In [26] it is used in the identification of target customers minimizing the recommendation errors, by selecting users to whom the recommendations should be addressed, according to which categories of purchases they have made in a selected period of time. In [28] a Naive Bayes classifier has been used to predict the user's long-term preferences in the news domain, with excellent results in accuracy. Support Vector Machines (SVM) classifier is used to find a linear hyperplane (decision boundary) that separates input data in such a way that the distance among data groups is maximized [30].

Cluster Analysis. It is the task of segmenting a heterogeneous population into a number of subgroups [1] [38]. Through Clustering it is possible to explore the data set and to organize the data for creating recommendations. Variables used in the clusters may be: Demographic [4], RFM [5], LTV [6], Demographic + RFM [7], Demographic + LTV [8], LTV+RFM [9]. The common clustering algorithms in use are: K-means (each cluster is represented by the center of the means of the data points belonging the cluster); K-Medoids (each cluster is represented by the most representative/central data point of the cluster); Clara (it is an extension to Partitioning Around Medoids, PAM, adapted to large data sets); Self-organizing map (SOM, based on artificial neurons clustering technique) [3], [33].

Association Rules aim at finding rules in the dataset that satisfy some minimum support and minimum confidence constraints. An association rule is an expression $X \Rightarrow Y$, where X and Y are item sets (e.g., Milk, Cookies ⇒ Sugar). Given a set of transactions T, and denoting MinSup and MinConf the minimum support and the minimum confidence constraint values, the goal of association rule mining is to find all rules having support greater than or equal to MinSup, and confidence greater than or equal to MinConf. The most common algorithms used for implementing association rule mining are Apriori [10], FP- Growth [11], SSFIM [12] and SETM [25].

In [14], a **hybrid recommendation** system combining content-based, collaborative filtering and data mining techniques has been proposed. The recommendation algorithm makes similar groups of customers using LTV value, for this the segmentation of customer based on costumer behavior through RFM attributes has been performed.

With the growing volume of data acquisition, the possibility of using **deep learning** in recommendation systems have been also considered, in order to overcome the obstacles of conventional models listed above, achieving a higher accuracy of recommendation. Through deep learning it is possible to detect non-linear and non-trivial relationships among users and items from contextual, textual and visual input [19]. The main limitations of deep learning based recommendation systems are that for content-based systems there are often privacy issues in the collection of information, while for collaborative filtering the acquisition of data from different sources often results in incomplete information that greatly affects the accuracy of recommendations. The main deep learning algorithms for recommender system are: (i) Multilayer Perceptron (MLP) which is a class of feedforward artificial neural network with multiple hidden layers between the input and the output layer. [18] uses a standard MLP approach to learn interaction among user and item latent features, providing the model with flexibility and non-linearity; (ii) Autoencoders (AE) represent an unsupervised model that generate an output by compressing the input in a space of latent variables. There are many variants of autoencoders; the most common are denoising autoencoder, marginalized denoising autoencoder, sparse autoencoder, contractive autoencoder and variational autoencoder [32]. (iii) Convolutional Neural Networks (CNN) are feedforward neural networks that use convolution in place of general matrix multiplication in at least one of their layers. They can capture the global and local features and improve the efficiency and accuracy [15]. They have been used in several implementations, such as AlexNet [16] and batch-normalized Inception [17]. (iv) Recurrent Neural Networks (RNN) are typically employed to trace dynamic temporal behavior, actually in this kind of neural network the connections among the nodes form a direct graph along a temporal sequence [21]. Other fields of research have achieved an improvement by exploiting Long-Short Term Memory networks (LSTMs) that minimize RNN problems regarding the gradient vanishing/exploding. Adversary Network (AN) is a generative neural network where two neural networks are trained simultaneously within a minimax game framework [24]. Deep reinforcement learning (DRL) combines deep learning and reinforcement learning that enables to learn the best possible actions to attain the expected goals [34].

Compared to the previously discussed data mining techniques, all deep learning algorithms have cold start problems and require a considerable amount of data to improve performance. Open problems in the literature for deeplearning based recommendation systems concern the scalability of recommendation systems and the explicability of generated recommendations. On the other hand, deep learning solutions are not applicable in this case, in which the number of acquisitions per user is low.

III. System Architecture

In the context of fashion retail, the shops are of small size, and the customers are directly followed step by step by the attendees which provide suggestion and are ready to support them on all aspects. A similar scenario may occur on the online shopping, in which an online assistant is ready to follow the customer. In both cases, the user profiles are improved with new data in the few occasions in which the customer interact, and thus the customer might be continuously engaged with suggestions. This is obviously not possible and not acceptable by the customers. So that, a moderated engagement tool has to be provided that may consume the possible recommendations by spending them toward the customers or the assistant a limited number of times per experience, and in specific conditions. The architecture is reported in **Figure 1**. In compliance with GDPR rules, the TOOL ADMIN stores the details of customers, items and transaction data of both stores and ecommerce website in a centralized database. Using sensors in the fitting rooms, totems in stores and RFID technology on items, customer interactions with products are stored. The recommender reads the information from the DB, generates the recommendations and stores them in the Recommendation table. The TOOL ENGAGER is responsible for sending recommendations to the customer or shop assistant within the store. After the recommendation is sent, the TOOL ENGAGER records the customer's interaction with the recommended product. The Recommender has to create a list of suggestions taking into account users' profiles and items descriptions. The recommendations have to be carefully suggested, since suggested item should not have been purchased by the customer recently, neither already proposed by the human Assistant. All the suggestions need to be generated on the basis of purchases made by the customer in the last few experiences and months, when possible. These last rules are applied directly at the final stage into the Tool Engager.

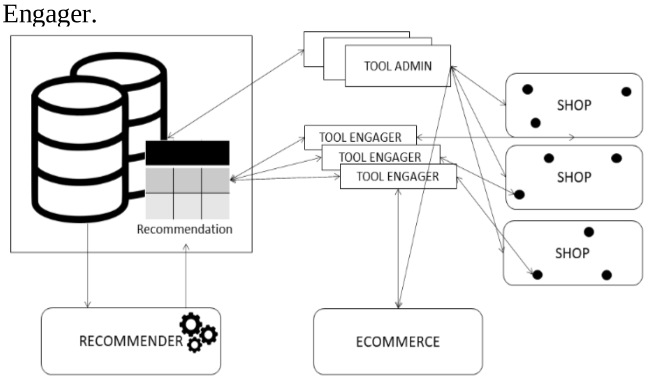


Figure 1 – General Architecture.

IV. Feedback Recommender System

The main goals consisted in increasing the customer recency, and thus to increase the number of times users contacts and sales may occur. For this purpose, the computational workflow reported in **Figure 2** has been adopted. The data are continuously collected by the administration (sales in shops: online and onsite); then a periodic clustering on items is performed. The results are taken into account in the computation of an integrated clustering driven by the user profiles and additional features to finally provide a set of suggestions of different kinds. The main steps of the workflow are described in the following subsections.

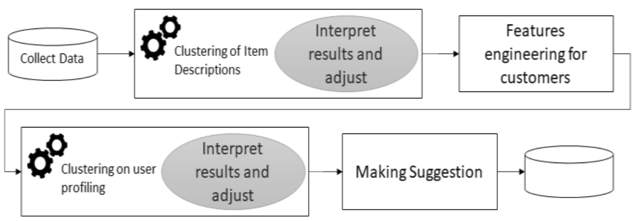


Figure 2 – General Data Computing workflow

The production of recommendations and their submission are asynchronous: (i) mediated by the assistant that may decide or not to accept and pass them to the customer, (ii) filtered by the Tool Engager according to the last actions performed by the customer, (iii) decided to be spent by sending them online via email when the time passed since the last contact with the users is greater than a reference value, etc. The produced pool of recommendations (for each potential returning user, and user kind) are generated on table, and the table can be refilled on demand or with a high rate. They include a programmed mix of *suggestions computed: by customer similarity, by items similarity, and by serendipity (randomly produced).*

A. *Clustering of Item Descriptions*

As described above, the first analysis has been performed to cluster the domain of item descriptions. This allows to reduce the space and weight the relevance of item categories. In the case of fashion retail, typically the number of products is not huge, differently to what you may have on supermarkets. In our test cases, the database contained about 50000 items with fields reported in Table I.

TABLE I. PRODUCT ITEM DESCRIPTIONS FIELDS

Field ID	*Item Description*	*Example*
TYPE	Type	"1A0145", "1A0333",…
CONFIGURATION	Configuration	"DRESS" ,"JACKET",…
PATTERN	Color	"White", "Red", "Navy blue", …
MODEL	Alphanumeric code model	"1A0145", "1A0333", …
PACKAGING_TYPE	Type packaging	"Packaging Basic PE", "Packaging Basic-Contin.", "Women's Packaging A/I",
PRODUCTION_CATEGORY	Production category	"Accessories", "Clothing", "Jeans", ……
MERCHANDISE_MCR_TYPE	Merchandise type	"Basic, Preview", "Women", "Main Women", ……
MERCHANDISE_TYPOLOGY	Merchandise tipology	"Preview Women SS", "Main Women AI", "Women PE", ……
MERCHANDISE_MCR_FAMILY	Merchandise family	"Coat", "Bag", "Dress", ……
MERCHANDISE_GROUP	Merchandise group	"Jewelry", "Dress", "Shirt", ……
GENDER	Gender	"Accessories Women", "Child", "Women", ……
BRAND	Brand	"VA", "GM", "PW", ……
STYLE_GROUP	Style	"P", "C", ……
BIRTH_SEASON	Season	"20201", "20062", "20071", ……
PERIODICITY	Periodicity	"C", "S", ……
IS_CLOTHING_ITEM	Marking if the item belongs to a clothing category	1,0 (yes/no)
NRM_CAT_LVL_1	Code normalized business classification level 1	"Accessories", "Clothing", "Jeans", ……
NRM_CAT_LVL_2	Code normalized business classification level 2	"Bag", "Clothing", "Coat", ……
NRM_CAT_LVL_3	Code normalized business classification level 3	"Shopping", "Dress", "Jacket", ……
NET_SOLD_PRICE	Price	1580.00
IN_STOCK	Whether an item is available or not	1,0 (yes/no)
132 X Hashtag tasche, abalze,…	Hashtag website	1,0 (yes/no)

Most of the fields are strings, only a few of them are numeric or Boolean. Clustering has been carried out by using *K*-medoids [20]. *K*-medoid is a classical clustering technique that partition a dataset of n objects into k a priori known clusters. A number of techniques can be used to identify the best compromise on the value of *K* [2]. To calculate the distance among items we used the Gower distance [13], which is computed as the average of partial dissimilarities across individuals. Each partial dissimilarity (and thus the Gower distance) ranges in [0,1].

$$d(i,j) = \frac{1}{p}\sum_{i=1}^{p} d_{ij}^{(f)}$$

Where: $d_{ij}^{(f)}$ is the partial dissimilarity computation which depends on the type of variable being evaluated. For a qualitative assessment the partial dissimilarity is 1 only if observations *xi* and *xj* have different values, and 0 otherwise. Through the silhouette method we determine the optimal number of clusters. The silhouette method calculates the average silhouette of observations for different values of *K* [2]. The optimal number of clusters *K* is the one that maximizes the silhouette over a range of possible values for *K*. In **Figure 3a**, the trend of silhouette index, as a function of *K*, is reported. **Figure 3b** shows the distribution of cluster size for *K=13*. From the trend, the value of *K=13* has been estimated as a good compromise. In **Table II,** the descriptions of the identified clusters, and the corresponding sales are reported. The main descriptions have been identified by a cluster analysis.

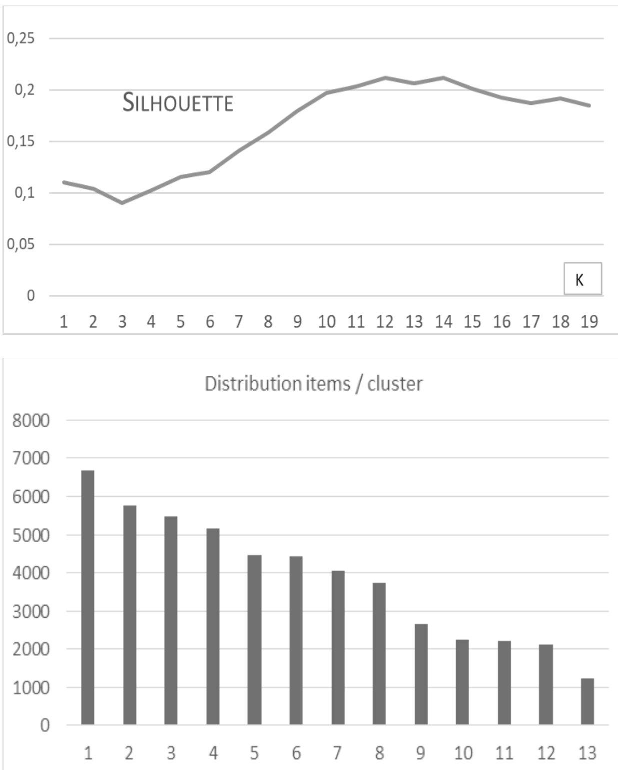


Figure 3 – (a) Trend of the Silhouette value as a function of the number cluster K for item dataset; (b) distribution of items/cluster when K=13.

TABLE II. MAIN DESCRIPTION OF CLUSTERS.

Cluster	*Derived descriptions of the item clusters*	*# sales*
1	BAG	969
2	DRESS	1171
3	TROUSERS	794
4	KNIT	678
5	T-SHIRT	674
6	ACCESSORIES (HAT - FOULARD - SCARF - NECKLACE - GLOVES - BRACELET)	596
7	SHIRT	838
8	COAT	388
9	SHOES	341
10	SKIRT	530
11	JACKET	292
12	BELT	237
13	CHILDREN'S CLOTHING	126

B. *Features engineering for customers*

The data collected by the administrations and the retail shops refer to the user behavior, which is associated with the user profile. The user profile has been enriched with information regarding customer behavior adding: (i) fields about the customer's maximum interest for an item within the cluster, such as: Interest (Yes/no), Observed (Totem, Online, etc.), Tried, Purchased item; (ii) fields describing the items purchased within the cluster.

In addition, a number of features (which in some sense are KPI, Key performance indicators) have been also computed, assessed, identified and discussed in this section by taking into account the experience of business developers. Among them: recency, frequency, and average spending. *Recency* is defined as the number of days passed since the last visit or access in a store or online; *Frequency* represents the frequency of purchase in number of days; *Average spending* is the average value of single ticket for the customer (estimated on the basis of the admin track record). In addition, in order to distinguish from online and in-store behavior, online and in-store frequency and recency have been computed.

C. *Second Level* ***Clustering on user profiling***

In this case, the number of user profiles has been 608447, of which 27346 have been acquired in the 2016-2019 temporal range. The user profile includes the features listed in Table III.

TABLE III. 35 USER CUSTOMER FEATURES (ALL NUMBERS).

Name profile feature	*Description*
RFM_TRN_DaysFrequency	Frequency transaction
RFM_TRN_DaysRecency	Recency transaction
RFM_TRN_AvgAmount	Average spending transaction
RFM_PRS_ONLINE_DaysFrequency	Frequency presence online
RFM_PRS_ONLINE_DaysRecency	Recency presence online
RFM_PRS_ONPREM_DaysFrequency	Frequency presence store
RFM_PRS_ONPREM_DaysRecency	Recency presence store
FidelityUsageRange	Fidelity card use
CUS_FIDELITY_CARD_LEVEL_CD	Fidelity card level
Cluster_k_Interest size[13]	Max interest for each cluster
Cluster_k_Purchased size[13]	Number of items purchased

On the basis of the user profile features, which include two vectors of preferences of the user for items clusters identified in the first phase, a clustering has been carried out through the K-means method. The Silhouette method has been used to determine the optimal number of clusters, in this case $k=14$ (see **Figure 4**). In Table IV, the derived descriptions of customers/user clusters and they corresponding size are reported.

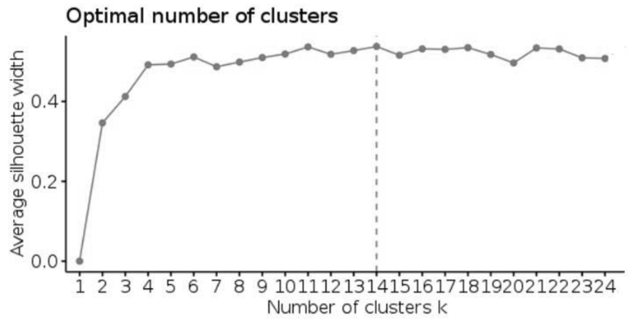


Figure 4 - Average silhouette width vs number of clusters

TABLE IV. DESCRIPTION OF USERS' CLUSTERS.

Cluster	*Derived Description from Customer cluster analysis*	*# total customer*
1	Customers with average spending amount not defined; the frequency is not defined neither in store neither online; day of the last purchase not defined	9195
2	Customers with low average spending amount, mainly online with undefined frequency and last purchase older than two years	3158
3	Customers with undefined average spending amount, mainly in store, with undefined frequency and last purchase older than two years mainly online	2433
4	Customers with low average spending amount, last purchase older than one year.	2302
5	Customers with low average spending amount in store, with frequency of about 4 months in store; last purchase has been made within one year. often using the fidelity card	2302
6	Customers with low average spending amount, more frequent in store with annual frequency; last purchase older than one year.	1657
7	Customer with low average spending amount, more frequent online, but also buyingin store with frequency of about 2 months online and about 6 months in store; last purchase older than one year, use fidelity card	1493
8	Customer with average spending amount not defined, mainly online; last purchase mid term days	1186
9	Customer with very high average spending amount in store	887
10	Customer with medium average spending amount more frequent in store but also buys in store with frequency about 230 days; last purchase about 262 days, use fidelity card	819
11	Customer with average spending medium amount in store; last purchase one year ago; frequency is not defined	797
12	Customer with average spending amount not defined, mainly online, with frequency of about 270 days; last purchase one year	717
13	Customer with medium average spending amount, mainly in store, with not defined frequency and last purchase older than one year	391
14	Online customers with annual frequency	9

According to the obtained results, cluster #1 was actually very large. For this reason, a second level clustering has been performed to split user cluster #1 in subclusters based on the same features. Also in this case, the silhouette method has been used to identify the best compromise in terms of the number of clusters. Analyzing the distributions of cluster size, we opted for $K=5$, with the aim of having maximum classifications and expression, as shown in Table V. The final distribution of clusters has been reported in Figure 5.

TABLE V. DESCRIPTION OF SECOND LEVEL CLUSTER OF CLUSTER #1.

Cluster	*Derived Description from Customer cluster analysis*	*# total customer*
1.1	Customers with average spending amount undefined; the frequency is undefined neither in store nor online; day of the last purchase undefined	5167
1.2	Customers with low average spending amount. They mainly buy in the product cluster #12	2411
1.3	Customers with very low average spending amount, mainly in the product clusters: #2, #10 and #12	1330
1.4	Customers with: recency of about 23 days, frequency of about 18 days	173
1.5	Customers with average spending amount of about 150 Euro; mainly buying in the product cluster #1	148

Figure 5 - Distribution of customers along the resulting 18 clusters

D. Final production of suggestions

As described above, the identified solution produces a number of recommendations for each user. Each possible suggestion is labeled with the kind, the date of emission, and a deadline. The Engager Tool also marks those that have been spent with: the date and time of emission, the channel adopted (shopID, mobileApp, website, shopID.totemID, etc.), the ID of the assistant, etc. This information is useful for the assessment of the level of acceptance, and thus for the validation, as described in the next Section. Therefore, the database with the suggestions is never discharged since the recommender has to take into account the already spent suggestions.

The recommendations are generated according to different kinds (as described in the following list) and they are consumed in different contexts by the Engager Tool. The **different kinds of recommendations** are by:

customer similarity: for each customer cluster the most representative items are built. They are identified among the most purchased items within the users' ones belonging to the same item cluster (they can be selected by using other criteria, for example: because they are the most frequently asked, or the company would like to push them, or they are closer to the cluster centroid, or to maximize the revenue or minimize the stock, etc.). In addition, the suggested item should have not been already purchased or proposed/suggested to same customer in the same season.

item similarity: considering the last items purchased by the customer according to the information contained into its profile, and randomly selecting items in the same item clusters, avoiding proposing items which have been already bought or proposed. Also in this case, the items can be filtered/selected by using additional criteria, for example: because they are the most frequently asked, or the company would like to push them, or they are closer to cluster centroid, or to maximize the revenue or minimize the stock, etc.

item complementary: considering items that may complement the last items that have been bought by the customer according to a table of complementary items; for example: a belt in combo with a bag. Please note that some of the item clusters are complementary each other, see the above descriptions – e.g., #1 and # 2 of Table II. To this end, through association rules using a priori algorithm [10] for each transaction in the dataset a set of metrics have been calculated; some examples are reported in Table VI, for the first 5 clusters. The used metrics are *support, confidence, lift, count*, and are defined as follows. Let *N* and *M* be two clusters. *Support({N}→{M})* is the ratio of the number of transactions/tickets including *N* and *M* wrt the total number of transactions. *Confidence({N}→{M})* is the ratio of the number of transactions containing *N* and *M* wrt the total number of transactions containing *N*. *Lift({N}→{M})* is the ratio of confidence of *N* wrt the total number of transactions containing *M*. *Lift({N}→{M})* is the number of transactions containing *N* or *M*. To generate the recommendations, we considered the top 5 clusters with highest *support* and suggested one of the best-selling items (*count*) within the cluster.

TABLE VI. EXAMPLE OF COMPLEMENTARY CLUSTERS ASSESSMENT BY USING METRICS: SUPPORT, CONFIDENCE, LIFT AND COUNT (PART)

Cluster	Complementary clusters				
	cluster	*support*	*confidence*	*lift*	*count*
1	2	0.26486066	0.6069351	1.106003	12935
	7	0.24864345	0.5697729	1.253423	12143
	3	0.24465057	0.5606231	1.213722	11948
	8	0.24336057	0.5576670	1.277549	11885
	4	0.22298667	0.5109797	1.282096	10890
2	3	0.34351004	0.6259701	1.355196	16776
	7	0.32391425	0.5902612	1.298495	15819
	8	0.31392182	0.5720522	1.310504	15331
	4	0.29840080	0.5437687	1.364367	214573
3	2	0.34351004	0.7436830	1.355196	16776
	7	0.30397035	0.6580814	1.447690	14845
	8	0.29868747	0.6466442	1.481385	14587
	4	0.27753548	0.6008511	1.507592	13554
	1	0.24465057	0.5296569	1.213722	11948
4	2	0.29840080	0.7487156	1.364367	214573
	3	0.27753548	0.6963625	1.507592	13554
	7	0.26578209	0.6668722	1.467029	12980
	8	0.27260069	0.6839807	1.566918	13313
	1	0.22298667	0.5594945	1.282096	10890
5	2	0.13366914	0.7559931	1.377628	6528
	8	0.12396339	0.7011002	1.606137	6054
	7	0.12224338	0.6913723	1.520926	5970
	3	0.12199767	0.6899826	1.493780	5958
2	4	0.12158814	0.6876665	1.725420	5938

item associated: in order to improve a customer's purchase frequency, we generated suggestions for customers who purchased an item in the last three months. For the generation we have proceeded as follows: through association rules using a priori algorithm [10] we have defined pairs of items *(i,j)* with *support >= 0.001* and *confidence >= 0.01*. If a customer buys item *i* then item *j* will be suggested. This is the typical suggestion which can be delivered for stimulating the return on the shop.

suggestions for serendipity: randomly selecting items to be suggested from the whole present collection, taking also into account what is available in the physical shop.

V. Assessment and Validation

The recommendation system has been validated in a store located in Florence and on the online store as follows. We have exploited the data collected since December 2019 to test and tune the solution, verifying if the suggestions produced were also provided by the Assistant in shops and finally acquired by the customers. The algorithm updates the clusters monthly and generates the new suggestions daily. With the suggestions generated, without stimulating customers, we verified in the period January - June 2020, through transactions and verifying the shop assistants (which are the reference experts), if there was a match between suggestions and items purchased by customers. This analysis showed that on about 400 customers who bought, about 10000 suggestions were generated. On suggestions generated, the 6.36% items were purchased. This was considered the minimum level of reaching with the efficiency since resulted to be possible without the tool. Then from July 2020 until December 2020, the recommendation system was tuned on operative to stimulate a certain class of users, entering in the store, using the totem in the store and by mail for ecommerce. This analysis with the stimulated customers showed that on 67 selected customers in the trial, 3050 suggestions have been generated, while only about the 20% has been actually sent to the customers (on shops and/or email). On the items suggested, the 9.84% of them were actually acquired or tested. Therefore, using the stimulus of the recommendation system, we have increased the customers' attention of the 3.48%. The period for the assessment and validation was also complicated by the COVID-19 pandemic which strongly limited the access to the stores, and the validation via the e-commerce without the effective verification of the shop assistant is not comparable with the conditions of the 2019.

VI. CONCLUSIONS

In this paper, a recommendation system in the context of fashion retail has been proposed and described, relying on a multi-level clustering approach of items and users' profiles in online and physical stores. The solution has been developed in the context of the Feedback project founded by Regione Toscana, and has been conducted on real retail company Tessilform, and it has been validated against real data from December 2019 to December 2020, showing that the use of the proposed recommendation tool generated stimulus to the customers which brought to an increase of buyers' attention and purchase increase of 3.48%. The solutions proposed has demonstrated to be functional also in the presence of low number of customers and items, and when suggestions are mediated by the assistants as happen in the fashion retail shops. Moreover, the proposed solution addresses and solved lacks and issues which are present in current state of the art tools, such as also the cold start problems in generating recommendations for newly acquired customers, since it relies on rules mining techniques, allowing to predict the purchase behavior of new users. Our solution is also GDPR compliant, addressing the current strict policies for users data privacy, solving one of the main issue for managing users' demographic details.

Acknowledgment

The authors would like to thank FEEDBACK project and partners for which we have developed a part of the solutions described in this paper, and Regione Toscana for the partial founding POR FESR 2020 Phase 2. https://www.vargroup.it/progetti-rd/

References

[1] Michael, J.A Berry A., Gordon S. Linoff. "Data Mining Techniques Second Edition", wiley publishigng,

[2] U. Maulik and S. Bandyopadhyay, "Performance evaluation of someclustering algorithms and validity indices,"IEEE Transactions onPattern Analysis and Machine Intelligence, vol. 24, no. 12, pp. 1650–1654, 2002.

[3] M. Namvar, M. R. Gholamian and S. KhakAbi, "A Two Phase Clustering Method for Intelligent Customer Segmentation," 2010 International Conference on Intelligent Systems, Modelling and Simulation, Liverpool, 2010, pp. 215-219.

[4] Michael Giering. 2008. Retail sales prediction and item recommendations using customer demographics at store level. SIGKDD Explor. Newsl. 10, 2 (December 2008), 84–89. DOI:https://doi.org/10.1145/1540276.154030

[5] Chen, D., Sain, S. & Guo, K. Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. J Database Mark Cust Strategy Manag 19, 197–208 (2012). https://doi.org/10.1057/dbm.2012.17.

[6] Su-Yeon Kim, Tae-Soo Jung, Eui-Ho Suh, Hyun-Seok Hwang, "Customer segmentation and strategy development based on customer lifetime value: A case study" Expert systems with applications, Vol. 31, pp. 101-107, 2006.

[7] John A. McCarty, Manoj Hastak, "Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression", Journal of Business Research, Vol. 60, pp. 656–662, 2007.

[8] Hyunseok Hwang, Taesoo Jung, Euiho Suh, "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry" Expert systems with applications, Vol. 26, pp. 181-188, 2004.

[9] Chu Chai Henry Chan,"Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer", Expert systems with applications, Vol. 34, pp. 2754-2762, 2008.

[10] Mining association rules between sets of items in large databases R. Agrawal, T. Imielinski and A. Swami ACM SIGMOD Rec., 22 (2) (1993), pp. 207-216

[11] Mining frequent patterns without candidate generation Han J., Pei J. and Yin Y. ACM SIGMOD Rec., 29 (2) (2000), pp. 1-12

[12] SS-FIM: single scan for frequent itemsets mining in transactional databases Y. Djenouri, M. Comuzzi and D. Djenouri Proceedings of the 21st Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer (2017), pp. 644-654

[13] J. C. Gower, "A general coefficient of similarity and some of itsproperties,"Biometrics, pp. 857–871, 1971.

[14] Fátima Rodrigues, Bruno Ferreira "Product Recommendation based on Shared Customer's Behaviour" https://doi.org/10.1016/j.procs.2016.09.133

[15] Tuinhof H., Pirker C., Haltmeier M. (2019) Image-Based Fashion Product Recommendation with Deep Learning. In: Nicosia G., Pardalos P., Giuffrida G., Umeton R., Sciacca V. (eds) Machine Learning, Optimization, and Data Science. LOD 2018. Lecture Notes in Computer Science, vol 11331. Springer, Cham

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, pages 1097–1105, 2012.

[17] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In ICML, pages 448–456, 2015.

[18] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In Proceedings of the 26th International Conference on World Wide Web (WWW '17). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 173–182. DOI:https://doi.org/10.1145/3038912.3052569

[19] R. Mu, "A Survey of Recommender Systems Based on Deep Learning," in IEEE Access, vol. 6, pp. 69009-69022, 2018, doi: 10.1109/ACCESS.2018.2880197

[20] Kaufman, L., & Rousseeuw, P. J. (1990).Finding groups in data: Anintroduction to cluster analysis. New York: Wiley.

[21] S. Wu, W. Ren, C. Yu, G. Chen, D. Zhang and J. Zhu, "Personal recommendation using deep recurrent neural networks in NetEase," 2016 IEEE 32nd International Conference on Data Engineering (ICDE), Helsinki, 2016, pp. 1218-1229.).

[22] C. Badii, P. Bellini, A. Difino, P. Nesi, "Smart City IoT Platform Respecting GDPR Privacy and Security Aspects", IEEE Access, 2020. 10.1109/ACCESS.2020.2968741 https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8966344

[23] Nageswara, R. K. , & Talwar, V. G. (2008). Application domain and functional classification of recommender systems a survey. DESIDOC Journal of Library and Information Technology, 28 (3), 17–35 .

[24] C. Wang, M. Niepert and H. Li, "RecSys-DAN: Discriminative Adversarial Networks for Cross-Domain Recommender Systems," in IEEE Transactions on Neural Networks and Learning Systems.

[25] M. Ho&ma and A. Swami. Set-oriented mining of association rules. Research Report RJ 9567, IBM Almaden Research Center, San Jose, Cali- fornia, October 1993.

[26] Cho Y.H., Kim J.K., Kim S.H. A personalized recommender system based on web usage mining and decision tree induction Expert Syst. Appl., 23 (2002), pp. 329-342 International Conference on Innovative Internet Community Systems (I2CS) – Jubilee Edition 2010 –. Bonn: Gesellschaft für Informatik e.V.. (S. 170-179).

[27] Pazzani, M., Billsus, D. Learning and Revising User Profiles: The Identification of Interesting Web Sites. Machine Learning 27, 313–331 (1997). https://doi.org/10.1023/A:1007369909943.

[28] Billsus, D., Pazzani, M.J. User Modeling for Adaptive News Access. User Modeling and User-Adapted Interaction 10, 147–180 (2000). https://doi.org/10.1023/A:1026501525781

[29] Tariq Mahmood and Francesco Ricci. 2009. Improving recommender systems with adaptive conversational strategies. In Proceedings of the 20th ACM conference on Hypertext and hypermedia (HT '09). Association for Computing Machinery, New York, NY, USA, 73–82. DOI:https://doi.org/10.1145/1557914.1557930.

[30] Min SH., Han I. (2005) Recommender Systems Using Support Vector Machines. In: Lowe D., Gaedke M. (eds) Web Engineering. ICWE 2005. Lecture Notes in Computer Science, vol 3579. Springer, Berlin, Heidelberg

[31] Ricci F., Rokach L., Shapira B. (2011) Introduction to Recommender Systems Handbook. In: Ricci F., Rokach L., Shapira B., Kantor P. (eds) Recommender Systems Handbook. Springer, Boston, MA.

[32] Suvash Sedhain, Aditya Krishna Menon, Scott Sanner, and Lexing Xie. 2015. AutoRec: Autoencoders Meet Collaborative Filtering. In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). Association for Computing Machinery, New York, NY, USA, 111–112. DOI:https://doi.org/10.1145/2740908.2742726

[33] S. Wei, N. Ye, S. Zhang, X. Huang and J. Zhu, "Collaborative Filtering Recommendation Algorithm Based on Item Clustering and Global Similarity," 2012 Fifth International Conference on Business Intelligence and Financial Engineering, Lanzhou, 2012, pp. 69-72

[34] Xiangyu Zhao, Liang Zhang, Zhuoye Ding, Dawei Yin, Yihong Zhao, and Jiliang Tang. 2017. Deep Reinforcement Learning for List-wise Recommendations. arXiv preprint arXiv:1801.00209 (2017)

[35] Y. Koren, R. Bell and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," in Computer, vol. 42, no. 8, pp. 30-37, Aug. 2009

[36] Ngai, E., Xiu, L., & Chau, D. (2009). Application of data mining techniques in cus- tomer relationship management: A literature review and classification. Expert Systems with Applications, 36 (2), 2592–2602. https://doi.org/10.1016/j.eswa.2008. 02.021 .

[37] Swift, R.S., 2001. Accelerating Customer Relationship using CRM and Relationship Technologies. Upper Saddle River: Prentice Hall PTR.

[38] P. Bellini, I. Bruno, D. Cenni, A. Fuzier, P. Nesi, M. Paolucci, "Mobile Medicine: Semantic Computing Management for Health Care Applications on Desktop and Mobile Devices", on Multimedia Tools and Applications, Springer. http://www.springerlink.com/content/q8512555u0j00584/ new May 2012, Volume 58, Issue 1, pp 41-79 (see more)

[39] P. Bellini, I. Bruno, P. Nesi, M. Paolucci, "A Static and Dynamic Recommendations System for Best Practice Networks", Proc. of the Human-Computer Interaction Conference, 21-26- July 2013, Nevada, USA, Press by Springer Verlag LNCS.new

[40] P. Bellini, D. Cenni, P. Nesi, "Optimization of Information Retrieval for Cross Media contents in a Best Practice Network", International Journal Multimedia Information Retrieval, 10-04-2014. http://link.springer.com/article/10.1007/s13735-014-0058-8?

[41] Cuzzocrea, A., Fadda, E., Data-Intensive Object-Oriented Adaptive Web Systems: Implementing and Experimenting the OO-XAHM Framework, Proceedings of the 12th International Conference on Management of Digital EcoSystems, MEDES 2020, 2020, pp. 115–123

Graphical Animations of the Lim-Jeong-Park-Lee Autonomous Vehicle Intersection Control Protocol

Win Hlaing Hlaing Myint, Dang Duy Bui, Duong Dinh Tran, Kazuhiro Ogata
School of Information Science
Japan Advanced Institute of Science and Technology (JAIST)
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
Email: {winhlainghlaingmyint,bddang,duongtd,ogata}@jaist.ac.jp

***Abstract*—The main goal of SMGA is to help human users be able to perceive (non-trivial) characteristics. The Lim-Jeong-Park-Lee autonomous vehicle intersection control protocol has been graphically animated with SMGA to demonstrate that SMGA can be applied to the wider class of systems. We have revised SMGA so as to handle composite data that are used in the protocol. We design a flexible state picture for the protocol so that it is possible to deal with different initial states when the number of vehicles is less than or equal to a given number. Some characteristics are guessed by observing graphical animations based on the state picture design, and the characteristics are confirmed with model checking. The paper also summarizes several lessons learned as tips on how to make a state picture design when composite data are used**

***Keywords*-graphical animation; LJPL protocol; SMGA; state machine; state picture design**

I. Introduction

SMGA [1] has been developed to visualize graphical animations of systems/protocols. The main purpose of SMGA is to help human users be able to perceive non-trivial characteristics of systems/protocols by observing its graphical animations because humans are good at visual perception [2]. Those characteristics could be used as lemmas to formally verify that the systems/protocols enjoy some desired properties. It implies the usefulness of the tool since lemma conjecture is a challenging problem in formal verification.

The Lim-Jeong-Park-Lee autonomous intersection control protocol (the LJPL protocol) has been graphically animated to demonstrate the potential of SMGA that can be applied to wider classes of systems/protocols. First, we need to carefully make a state picture design because it is a core task of the tool [3]. The specification of the protocol, however, contains some observable components whose values are composite (over one component value inside). One possible way to visualize such a composite value is to display each component value of the composite value. We, however, revise the tool so that users can design a state picture to be able to display the composite data explicitly. Because the LJPL protocol has multiple different initial states even if the number of vehicles is fixed, we make a flexiable state picture template that can be used for different initial states. Given a natural number n, we make a flexible state picture design so that any initial states in which the number of vehicles is up to n can be handled. After that, some characteristics are guessed by human users based on some graphical animations of the LJPL protocol and confirmed with Maude [4]. Some lessons learned are summarized as tips on how to make a good state picture design for state machine that use observable components whose values are composite

Bui and Ogata [5] have revised SMGA to visualize the network components. This work and ours can share working flow but we cannot apply those technique as they are to ours. One reason is that the behavior of a distributed mutual exclusion protocol and the LJPL protocol cannot share each other. María Alpuente, et al. [6] have proposed a methodology and implemented a prototype tool to check whether a Maude program is correct or not via logical assertions based on rewriting logic theories. The tool is to visualize the possible trace slides (as state sequences in our paper) to help users identify the cause of the error. In case that many possible rewriting rules are used, the visualization looks like a graph or a tree in which the states (displayed as text) are nodes. This visualization approach can be applied to our work, although its purpose is different than ours.

We suppose the readers are familiar with state machines and Maude [4] to some extent. The formal specification of the LJPL protocol used in the paper is followed by the work [8] in which they have proposed some modifications (i.e., two more statuses running and approaching are added as each vehicle's $status$). Please refer to the paper [8] in detail.

II. Specification of LJPL Protocol in Maude

In this paper, a state is expressed as a soup of observable components. Let b be a Boolean value, q a queue of vehicle IDs (i.e., a queue of natural numbers). Let vid, lid, t, lt be natural numbers, where vid and lid represent a vehicle ID and a lane ID, respectively, while t and lt represent the time.

This work was partially supported by FY2020 grant-in-aid for new technology research activities at universities (SHIBUYA SCIENCE CULTURE AND SPORTS FOUNDATION)

DOI reference number: 10.18293/DMSVIVA2021-004

To formalize the LJPL protocol as a state machine M_{LJPL}, we use the following observable components:

- (`clock` : t, b) - it says that the current time is t. `clock` represents the global clock shared by all vehicles. Initially, the first parameter of `clock` is set to 0, and increased by the time. However, if time is allowed to increase without any constraints, the reachable state space will quickly explode. That is the reason why we introduce the second counterpart b such that t only can increment when b is true. That is, whenever b is true, t can increment, b becomes false, and when a vehicle obtains the current time t (without changing t), b becomes true,
- (`v`[vid] : $lid, vstat, t, lt$) - it says that the vehicle vid is running on the lane lid, its current $status$ is $vstat$, it arrives at the intersection at the time t, and the lead vehicle of the lane lid reaches the intersection at the time lt,
- (`lane`[lid] : q) - it says that the queue of vehicles running on lane lid is q,
- (`gstat` : $gstat$) - where $gstat$ is either `fin` or `nFin`. When it is `fin`, all vehicles concerned have crossed the intersection.

Each state in S_{LJPL} is expressed as $\{obs\}$, where obs is a soup of those observable components. We suppose that five vehicles (from 0 to 4) participate in the LJPL protocol such that two vehicles are running on lane0, one vehicle is running on lane1, and two vehicles are running on lane5. The initial state of I_{LJPL} namely `init` is defined as follows:

```
{(gstat: nFin) (clock: 0,false) (lane[0]: oo)
(lane[1]: oo) (lane[2]: oo) (lane[3]: oo)
(lane[4]: oo) (lane[5]: oo) (lane[6]: oo)
(lane[7]: oo) (v[0]: 0,running,oo,oo)
(v[1]: 0,running,oo,oo) (v[2]: 1,running,oo,oo)
(v[3]: 5,running,oo,oo) (v[4]: 5,running,oo,oo)
(v[oo]: 0,stopped,oo,oo) (v[oo]: 1,stopped,oo,oo)
(v[oo]: 2,stopped,oo,oo) (v[oo]: 3,stopped,oo,oo)
(v[oo]: 4,stopped,oo,oo) (v[oo]: 5,stopped,oo,oo)
(v[oo]: 6,stopped,oo,oo) (v[oo]: 7,stopped,oo,oo)}
```

Initially, `gstat` is set to `nFin`, the value of the global clock is 0. Since the second value of the `clock` observable component is `false`, the abstract notion of the current time cannot increment. Each queue associated with each lane only consists of `oo` (denoting ∞), saying that there is no vehicle on the lane close enough to the intersection. `v[0]` & `v[1]` represent the two vehicles running on lane0, `v[2]` represents the vehicle running on lane1, and `v[3]` & `v[4]` represent the two vehicles running on lane5. There are eight `v[oo]` observable components that are used to represent dummy vehicles.

12 rewrite rules are used to specify T_{LJPL}. Let `OCs` and `OCs'` be Maude variables of observable component soups, `T`, `T'` and `T''` be Maude variables of natural numbers, and `B` is a Maude variable of Boolean values. When all vehicles have crossed the intersection, the state does not change anything, which is specified by the following two rewrite rules:

```
rl [stutter] : {(gstat: fin) OCs}
=> {(gstat: fin) OCs} .
```

```
crl [fin] : {(gstat: nFin) OCs}
=> {(gstat: fin) OCs} if fin?(OCs) .
```

where `fin?(OCs)` returns true iff all vehicles in `OCs` have crossed the intersection.

The rewrite rule `tick` is defined to specify the behavior of the global clock:

```
rl [tick] :
{(gstat: nFin) (clock: T,true) OCs} =>
{(gstat: nFin) (clock: (T + 1),false) OCs} .
```

The rewrite rule says that if the second value of the `clock` observable component is `true`, the abstract notion of the current time `T` increments and the second value becomes `false`.

Two rules are used to specify a set of transitions that change a vehicle status from running to approaching as follows:

```
rl [approach1] : {(gstat: nFin) (clock: T,B)
(lane[LI]: oo) (v[VI]: LI,running,oo,oo) OCs}
=> {(gstat: nFin) (clock: T,true)
(lane[LI]: VI) (v[VI]: LI,approaching,T,oo)
OCs} .
```

```
rl [approach2] : {(gstat: nFin) (clock: T,B)
(v[VI]: LI,running,oo,oo)
(lane[LI]: (VI' ; VS)) OCs}
=> {(gstat: nFin) (clock: T,true)
(lane[LI]: (VI' ; VS ; VI))
(v[VI]: LI,approaching,T,oo) OCs} .
```

where `LI`, `VI`, and `VI'` are Maude variables of natural numbers, `VS` is a Maude variable of queues of natural numbers and ∞, and `;` constructs the queue. The first rewrite rule specifies the case in which there is no vehicle close enough to the intersection on the lane where the vehicle is running, while the second one deals with the case in which there exists at least one vehicle close enough to the intersection on the lane.

Three rewrite rules are used to specify a set of transitions that change a vehicle status from approaching to stopped.

```
rl [check1] : {(v[VI]: LI,approaching,T,oo)
(gstat: nFin) (lane[LI]: (VI ; VS)) OCs}
=> {(gstat: nFin) (v[VI]: LI,stopped,T,T)
(lane[LI]: (VI ; VS)) OCs} .
```

```
rl [check2] : {(v[VI]: LI,approaching,T'',oo)
(gstat: nFin) (v[VI']: LI,stopped,T,T')
(lane[LI]: (VS' ; VI' ; VI ; VS)) OCs}
=> {(gstat: nFin) (v[VI]: LI,stopped,T'',T')
(v[VI']: LI,stopped,T,T')
(lane[LI]: (VS' ; VI' ; VI ; VS)) OCs} .
```

```
rl [check3] : {(v[VI]: LI,approaching,T'',oo)
(gstat: nFin) (v[VI']: LI,crossing,T,T')
(lane[LI]: (VS' ; VI' ; VI ; VS)) OCs}
=> {(gstat: nFin) (v[VI]: LI,stopped,T'',T'')
```

```
(v[VI']: LI,crossing,T,T')
(lane[LI]: (VS' ; VI' ; VI ; VS)) OCs} .
```

where `VS'` is a Maude variable of queues. The first rewrite rule specifies the case in which vehicle `VI` is the top of the queue (i.e., `VI` will be lead on the lane). The second one deals with the case in which there exists another vehicle `VI'` in front of the vehicle `VI` such that `VI'` is stopped (`VI` will be non-lead on the lane). The last one specifies the case in which there exists another vehicle `VI'` in front of the vehicle `VI` such that the *status* of `VI'` is crossing (`VI` will be lead on the lane).

Two rewrite rules `enter1` and `enter2` are used to specify a set of transitions that change a lead vehicle *status* from stopped to crossing. `enter1` deals with the case in which the ID of the lane on which the lead vehicle is located is even and `enter2` deals with the case in which it is odd. `enter1` is defined as follows:

```
crl [enter1] : {(v[VI]: LI,stopped,T,T)
(gstat: nFin) (lane[LI]: (VI ; VS)) OCs}
=> {(gstat: nFin) (lane[LI]: (VI ; VS))
 (v[VI]: LI,crossing,T,T) OCs'}
if isEven(LI) /\
 LI1 := (LI + 2) rem 8 /\
  (lane[LI1]: (VI1 ; VS1))
  (v[VI1]: LI1,VSt1,T11,T12) OCs1 := OCs /\
  VSt1 = stopped /\ T < T12 /\
 LI2 := (LI + 5) rem 8 /\
  (lane[LI2]: (VI2 ; VS2))
  (v[VI2]: LI2,VSt2,T21,T22) OCs2 := OCs /\
  VSt2 = stopped /\ T < T22 /\
 LI3 := (LI + 6) rem 8 /\
  (lane[LI3]: (VI3 ; VS3))
  (v[VI3]: LI3,VSt3,T31,T32) OCs3 := OCs /\
  VSt3 = stopped /\ T < T32 /\
 LI4 := (LI + 7) rem 8 /\
  (lane[LI4]: (VI4 ; VS4))
  (v[VI4]: LI4,VSt4,T41,T42) OCs4 := OCs /\
  VSt4 = stopped /\ T < T42 /\
 OCs' := letCross(VS,OCs) .
```

where `LI`i for $i = 1, \ldots, 4$ are Maude variables of natural numbers, `VI`i & `T`j for $i = 1, \ldots, 4$ & $j = 11, 12, \ldots, 41, 42$ are Maude variables of natural numbers & ∞, `VS`i for $i = 1, \ldots, 4$ are Maude variables of queues, `VSt`i for $i = 1, \ldots, 4$ are Maude variables of vehicle statuses, and `OCs`i for $i = 1, \ldots, 4$ are Maude variables of observable component soups. `isEven(LI)` holds if `LI` is even. The rewrite rule checks if all lead vehicles of the four conflict lanes (i.e., `LI1`, `LI2`, `LI3`, and `LI4`) are not crossing the intersection and the arrival time `T` of the vehicle `VI` is less than all arrival times of the lead vehicles on the conflict lanes. If the conditions are satisfied, the status of vehicle `VI` is changed to crossing from stopped and the statuses of all vehicles that follow `VI` and whose statuses are stopped also become crossing, which is done by `letCross(VS,OCs)`.

The rewrite rule `enter2` can be defined likewise. There are also two more rewrite rules `leave1` and `leave2` that are used to specify a set of transitions changing a vehicle *status* to crossed from crossing. All of them can be found from the webpage presented in Sect. I.

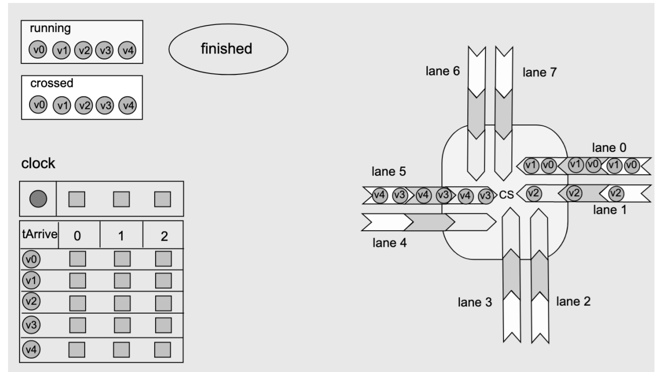


Figure 1. A state picture design for the LJPL protocol (1)

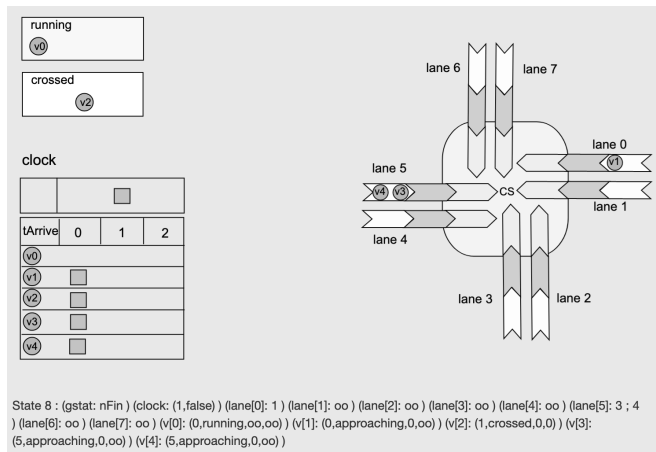


Figure 2. A state picture for the LJPL protocol (1)

III. Graphical Animation of LJPL Protocol

A. Idea

At the beginning of the work, the first author of the presented paper needs to deal with a problem of how to design a good state picture so that it can display well a composite value of some observable component. At that time, the version of SMGA could only visualize observable components whose value is text or designated place. The tool, however, cannot display specific component inside the composite values of the observable components. For example, considering the following observable component: ($time$: (YY, MM, DD)), SMGA cannot display the value "YY", "MM", or "DD". Therefore, the first author has modified the specification by adding some observable components that do not affect the behavior of the protocol. With the example above, three observable components are added: ($year$: YY) ($month$: MM) (day: DD) ($time$: (YY, MM, DD)). This

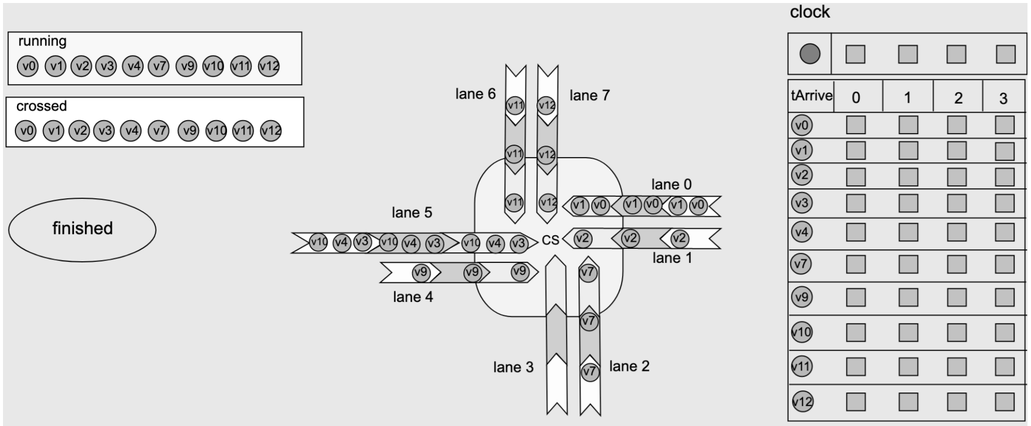


Figure 3. A state picture design for the LJPL protocol (2)

is the key idea to makes the tool be able to produce good graphical animations for the LJPL protocol in particular and display observable components whose values are composite in general.

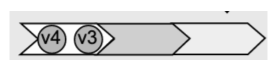
It is convenient for users if SMGA supports a functionality that can explicitly visualize specific component inside the composite values of the observable components without adding unneeded observable components. Therefore, the second author of the presented paper has revised the tool to support that functionality. The key idea is to add # followed by a natural number (start from 0) that represents a position inside the composite value of an observable component. For example, with the following composite value ($time$: $(YY, MM, DD(hh, mm, ss))$), we can extract the value mm by the notation $time\#2\#1$, where 2 denotes the third position of $time$'s value (i.e., $DD(hh, mm, ss)$), and 1 denotes the second position inside DD (i.e., mm). Therefore, users can display the component as a text or a designated place as SMGA provides.

B. State picture design

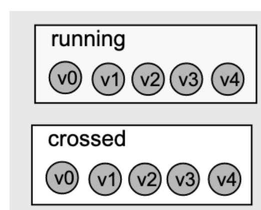
In SMGA, designing a good state picture is an important task because it can help humans better perceive the characteristics of the protocol [9]. In the LJPL specification, some observable components contain the same values of some information inside such as the lane ID (`laneID`), the $status$ (`vStat`) of a vehicle. By following the similarity principle of Gestalt [10], [11], they should be put together. Furthermore, the laneID of a vehicle cannot be changed, hence, we fix it as a constant text. After that, we come up with a state picture design for the initial state `init` mentioned in the previous section (shown in Fig. 1). A state picture generated from the state picture design is depicted in Fig. 2.

In Fig. 1, there are eight arrow shapes representing eight lanes. A lane representation designed is as follows:

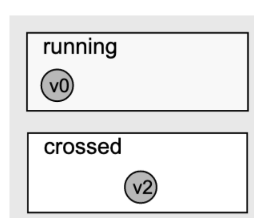
There are three colors: light green, pink, and light yellow that represent three statuses crossing, stopped, and approaching, respectively. For example, the status values of the fourth and fifth vehicles (i.e., v3 and v4) in the following figure are approaching:



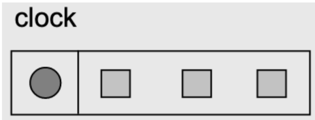
Two status values running and crossed representations used in Fig 1 are as follows:



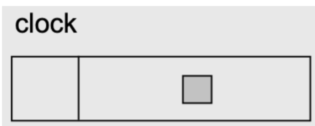
A rectangle whose color is light cyan represents the status running. A rectangle whose color is white represents the status crossed. For example, the status values of the first and third vehicles (i.e., v0 and v2) in the following figure are running and crossed, respectively:



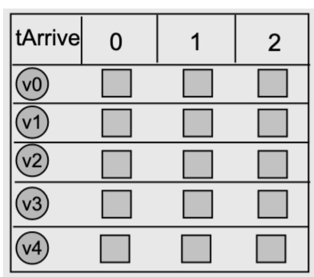
The design of the clock representation used in Fig. 1 is as follows:



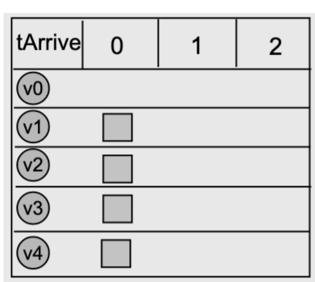
The value of the clock consists of two pieces of information: a natural number and a Boolean value (mentioned in Sect. II). Three blue squares represent the natural number from 0 to 2. If the value of the natural number is 0, the first blue square is displayed. A red circle represents the Boolean value. If the value is $true$, a red circle is displayed, otherwise, nothing is displayed. For example, when the value of the natural number is 1, and the Boolean value is $false$, those values are displayed as follows:



The design of the time arrivals of vehicles representation used in Fig 1 are as follows:



In each line, three blue squares represent the value of the time arrival (from 0 to 2). If nothing is displayed, the value is ∞. If the value is 0, the first blue square is displayed. For example, the figure below displays the case when the value of the first vehicle is ∞, the values of the four other vehicles are 0:



The design of the gstat representation used in Fig. 1 is as follows:

If the value of gstat is `fin`, the circle and text is displayed, otherwise, nothing is displayed.

Fig. 3 shows a state picture in which the initial state contains one vehicle in each lane1, lane2, lane4, lane6, and lane7, two vehicles in lane0, and three vehicles in lane5. It indicates that users need to redesign a new state picture since the initial state is changed. We design a flexible state picture such that it can be used when the number of vehicles participating in the protocol is small enough. Fig. 5 displays the flexible state picture design, in which each lane can contain up to four vehicles, the value of the natural number of clock, and the value of the time arrival are up to 6.

C. Graphical animation of LJPL protocol

Fig. 4 shows a state sequence for the LJPL protocol based on the state picture design depicted in Fig. 3. Six pictures correspond to six consecutive states from State 13 to State 18 in one state sequence randomly generated by Maude. Those pictures follow the rewrite rules mentioned in II. For example, State 17 is the successor of State 16 by the rewrite rule `leave1`. Taking look at the first picture (State 13) immediately makes us recognize that each of lane0 and lane5 contains two vehicles whose status are stopped, each of lane6 and lane7 contains one vehicle whose status is approaching, each of lane1 and lane2 contains one vehicle whose status is stopped, the values of time arrival of those vehicles are equal to 0 except two vehicles whose status are running have time arrival ∞. Taking look at State 13 and State 14 immediately makes us recognize that v12's status changes from approaching to stopped. Taking look at State 15 to State 17 immediately makes us recognize that v2's status changes from stopped to crossing and finally to crossed, and v4's status changes from running to approaching.

Fig. 6 shows another state sequence. Three pictures correspond to three consecutive states from State 27 to State 29. Taking look at State 27 and State 28 makes us immediately recognize that three vehicles can change the status from stopped to crossing at the same time. It is interesting because crossing is regarded as the critical section such that at most one vehicle should be located in the critical section at the same time. Taking look at the State 28 makes us immediately recognize a case that exists two vehicles running on two different lanes, and their statuses are crossing. It can be explained that two vehicles running on two concurrent lanes (e.g., lane5 and lane2) are allowed to cross the intersection simultaneously.

IV. Confirmation of Guessed Characteristics with Maude

Observing the graphical animations, we first see that the status of a vehicle sometimes does not change when the value of clock changes (shown at State 15 in Fig. 4). Carefully focusing on the value of clock, we guess that when the Boolean value of clock is false, the time arrival value of any

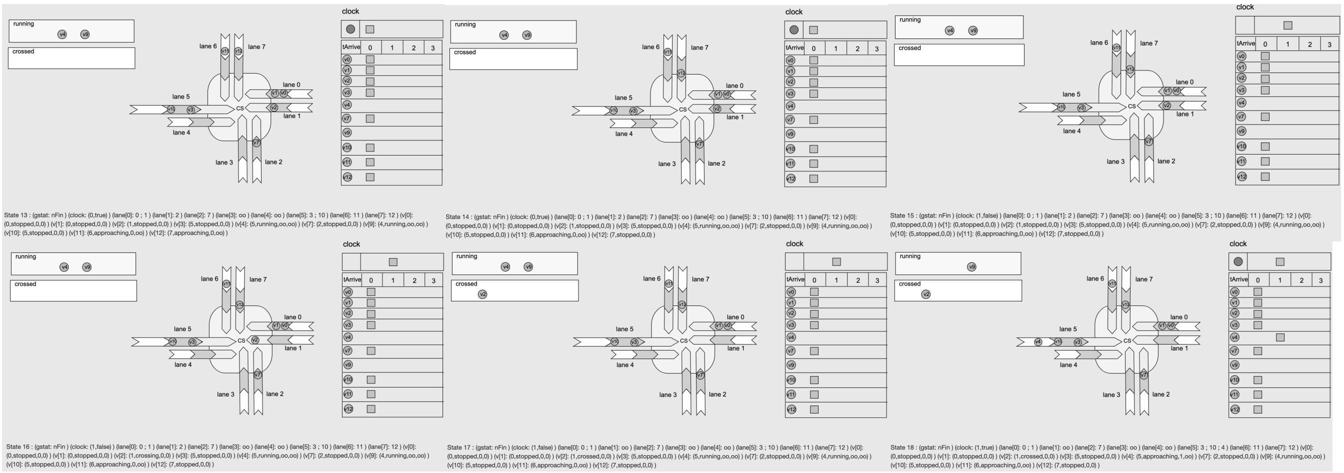


Figure 4. A state sequence for the LJPL protocol (1)

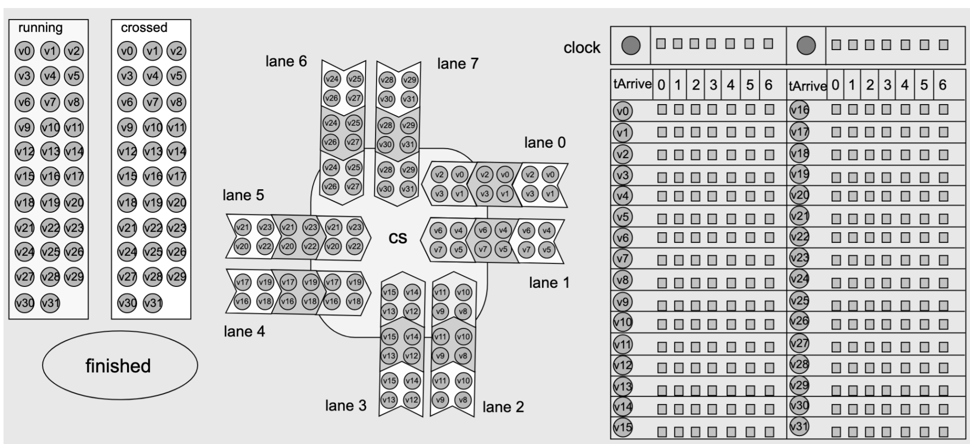


Figure 5. A flexible state picture design for the LJPL protocol (1)

vehicle cannot be greater than the first value of clock. The characteristic can be confirmed by Maude search command as follows:

```
search [1] in RIMUTEX :
init =>* {(clock: X:NatInf, false)
(v[i:NatInf]: j:NatInf, k:VStat,
Y:NatInf, t:NatInf) OCs}
such that
Y:NatInf >= X:NatInf and Y:NatInf =/= oo .
```

The search command above tries to find a reachable state in which the value of the time arrival of the vehicle `i` (i.e., `Y`) is greater than or equal to the first value of clock (i.e., `X`). Maude does not find any reachable state from the state `init` that satisfies the condition. Therefore, the guessed characteristic is confirmed with the initial state shown in Fig. 3.

Observing the graphical animations, we guess that if the first value of clock is equal to the time arrival of a vehicle, the status of such vehicle is not running. The characteristics can be confirmed by Maude search command as follows:

```
search [1] in RIMUTEX :
init =>* {(clock: X:NatInf, b:Bool)
(v[i:NatInf]: j:NatInf, K:VStat,
Y:NatInf, t:NatInf) OCs}
such that
Y:NatInf == X:NatInf and K:VStat == running .
```

The search command above tries to find a reachable state in which the value of the time arrival of a vehicle `i` is equal to the first value of clock, and the status of the vehicle `i` is running. Maude does not find any reachable state from the state `init` that satisfies the condition. Consequently, the guessed characteristic is confirmed with the initial state shown in Fig. 3.

V. Lesson learned

Through the case study with the LJPL protocol, we obtain several lessons on how to design a good state picture so that

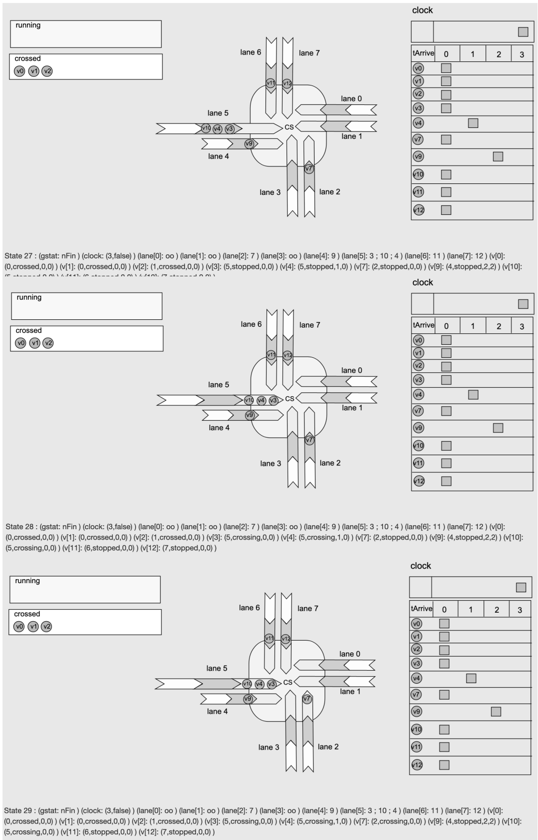


Figure 6. A state sequence for the LJPL protocol (2)

we can conjecture some non-trivial characteristics, especially when there are some observable components with composite values. Some of the lessons learned can be summarized as follows:

- When an observable component has a composite value, which consists of more than one component value inside, we need to carefully select which component values to visualize. For example, the second and the third component values (i.e., the $status$ and the time arrival) of the vehicle observable component are selected while the fourth component value (i.e., the time arrival of the lead) of the vehicle observable component is not used in our design.
- If a value of an observable component does not change, it should be expressed at a fixed label, such as laneID of each vehicle observable component.
- If there exist some observable components that have the same values, we should design and display their domain values together in a designated place.
- If there exist observable components that have a natural number as their values and the values are small enough, the values should be visually expressed nearby together so that we can see them simultaneously and compare them instantaneously. For example, the first value of clock (i.e., a natural number) and the time arrival of each vehicle have been visualized in our design.

VI. Conclusion

We have revised SMGA to be able to visualize composite data that are used in a state machine formalizing the LJPL protocol so that the LJPL protocol can be graphically animated. Some characteristics of the protocol have been conjectured by human users and confirmed with Maude based on our proposed state picture design. We have summarized our experiences as some tips on how to make a good state picture design for a state machine in which composite data are used. One future direction is to apply our work to other self-driving vehicle protocols, such as a merging protocol [12].

References

[1] T. T. T. Nguyen and K. Ogata, "Graphical animations of state machines," in *15th DASC*, 2017, pp. 604–611.

[2] K. W. Brodlie, et al., Ed., *Scientific Visualization: Techniques and Applications*. Springer, 1992.

[3] D. D. Bui and K. Ogata, "Better state pictures facilitating state machine characteristic conjecture," *MTAP*, 2021.

[4] M. Clavel, et al., Ed., *All About Maude*, ser. LNCS. Springer, 2007, vol. 4350.

[5] D. D. Bui and K. Ogata, "Graphical animations of the Suzuki-Kasami distributed mutual exclusion protocol," *JVLC*, vol. 2019, no. 2, pp. 105–115, 2019.

[6] M. Alpuente, et al., "Debugging Maude programs via runtime assertion checking and trace slicing," *J. Log. Algebraic Methods Program.*, vol. 85, no. 5, pp. 707–736, 2016.

[7] J. Lim, et al., "An efficient distributed mutual exclusion algorithm for intersection traffic control," *J. Supercomput.*, vol. 74, no. 3, pp. 1090–1107, 2018.

[8] M. N. Aung, Y. Phyo, and K. Ogata, "Formal specification and model checking of the Lim-Jeong-Park-Lee autonomous vehicle intersection control protocol (S)," in *SEKE 2019*, 2019, pp. 159–208.

[9] D. D. Bui and K. Ogata, "Better state pictures facilitating state machine characteristic conjecture," in *DMSVIVA 2020*, 2020, pp. 7–12.

[10] C. Ware, *Information Visualization: Perception for Design*. MKP Inc., 2004.

[11] D. Todorovic, "Gestalt principles," *Scholarpedia*, vol. 3, no. 12, p. 5345, 2008.

[12] S. Aoki and R. Rajkumar, "A merging protocol for self-driving vehicles," in *ICCPS 2017*, 2017, pp. 219–228.

Graphical Animations of the NSLPK Authentication Protocol

Thet Wai Mon, Dang Duy Bui, Duong Dinh Tran, Kazuhiro Ogata
School of Information Science
Japan Advanced Institute of Science and Technology (JAIST)
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
Email: {*thetwaimon,bddang,duongtd,ogata*}*@jaist.ac.jp*

***Abstract*—The behavior of the NSLPK authentication protocol is visualized using SMGA so that human users can visually perceive non-trivial characteristics of the protocol by observing graphical animations. These characteristics could be used as lemmas to formally verify that the protocol enjoys desired properties. We first carefully make a state picture design for the NSLPK protocol to produce good graphical animations with SMGA and then find out non-trivial characteristics of the protocol by observing its graphical animations. Finally, we also confirm the correctness of the guessed characteristics using model checking. The work demonstrates that SMGA can be applied to the wider class of systems/protocols, authentication protocols in particular.**

***Keywords*-graphical animation; SMGA; NSLPK protocol; state machine; state picture design**

I. Introduction

SMGA [1] has been developed to visualize graphical animations of protocols. The main purpose of SMGA is to help human users be able to visually perceive non-trivial characteristics of the protocols by observing its graphical animations because humans are good at visual perception [2]. Those characteristics can be used as lemmas to formally prove that systems/protocols enjoy desired properties. Several case studies have been conducted on some protocols with SMGA. Among the protocols are shared-memory mutual exclusion protocols [3], [4], [5], a distributed mutual exclusion protocol [6], and a communication protocol [1]. Any authentication protocols have not been yet tackled with SMGA. It is worth tackling authentication protocols with SMGA because such protocols, such as TLS, are infrastructure in our highly networked environment.

We aim at coming up with a brand-new way to visualize the behavior of an authentication protocol called NSLPK [7]. Since it is known that state picture designs affect how well human users can detect non-trivial characteristics of protocols [5], we carefully make a state picture design of the NSLPK protocol and produce graphical animations of NSLPK based on the state picture design. By observing the graphical animations, some non-trivial characteristics are guessed by human users and checked with Maude [8]. In the paper, we mainly focus on how to make the state picture design of the NSLPK protocol and how some characteristics could be found by observing graphical animations with detailed experiments.

This work was partially supported by JST SICORP Grant Number JPMJSC20C2, Japan and FY2020 grant-in-aid for new technology research activities at universities (SHIBUYA SCIENCE CULTURE AND SPORTS FOUNDATION).

DOI reference number: 10.18293/DMSVIVA2021-005

Bui and Ogata [6] have revised SMGA so as to visualize the network components in a distributed mutual exclusion protocol, which is applied partly to our work. VA4JVM [9] is a tool that can visualize outputs generated by Java Pathfinder (JPF). JPF outputs are often long and hard to read, especially when JPF finds something wrong, such as race-condition and deadlock. VA4JVM supports some functionalities, such as zooming, filtering, highlighting some specific parts of JPF outputs. Those functionalities can help human users observe some fragments that look interesting to be able to better comprehend JPF outputs. Counterexample generated by Maude LTL model checker can be graphically animated by SMGA [10]. Although Maude LTL model checker is a classical model checker and JPF is a software model checker, it would be worth considering some VA4JVM functionalities, such as zooming, filtering, and highlighting, to apply them to the future version of SMGA.

We assume that readers are familiar with state machines and Maude to some extent. NSLPK [7] is a modification of the NSPK authentication protocol [11]. The NSLPK protocol can be described as the following three message exchanges:

Init $p \to q : \varepsilon_q(n_p, p)$
Resp $q \to p : \varepsilon_p(n_p, n_q, q)$
Ack $p \to q : \varepsilon_q(n_q)$

Each principal such as p and q has a private/public key pair, and the public counterpart is shared with all principals but the private one is only available to its owner. $\varepsilon_p(m)$ denotes the ciphertext obtained by encrypting the message m with the principal p's public key. n_p is a nonce (a random number) generated by principal p. A nonce is a unique and non-guessable number that is used only once.

II. Formal Specification of NSLPK

We first introduce the following three operators to represent three kinds of ciphertexts used in the protocol:

```
op enc1 : Prin Nonce Prin -> Cipher1 .
op enc2 : Prin Nonce Nonce Prin -> Cipher2 .
op enc3 : Prin Nonce -> Cipher3 .
```

where `Prin` is the sort representing principals; `Nonce` is the sort denoting the nonce numbers; `Cipher1`, `Cipher2`, and `Cipher3` are the sorts denoting three kinds of ciphertexts contained in Init, Resp, and Ack messages, respectively. Given principals p, q and a nonce n_p term $\texttt{enc1}(q, n_p, p)$ denote the ciphertext $\varepsilon_q(n_p, p)$ obtained by encrypting n_p and p with the principal q's public key. `enc2` and `enc3` can be understood likewise. Hereinafter, let us use `Cipher1` (or `Cipher2`, or `Cipher3`) ciphertexts to refer to the ciphertexts contained in Init (or Resp, or Ack) messages. A `Nonce` is defined by the following operator:

```
op n : Prin Prin Rand -> Nonce .
```

where the third argument `Rand` is the sort denoting random numbers that makes the nonce globally unique and unguessable. Given principals p, q and random value r, term $\texttt{n}(p, q, r)$ denote a nonce created by principal p for authenticating p to principal q.

We specify three kinds of messages used in the NSLPK protocol as follows:

```
op m1 : Prin Prin Prin Cipher1 -> Msg .
op m2 : Prin Prin Prin Cipher2 -> Msg .
op m3 : Prin Prin Prin Cipher3 -> Msg .
```

where `Msg` is the sort denoting messages. `m1`, `m2`, and `m3` are the sorts denoting three kinds of messages Init, Resp, and Ack, respectively. The first, second, and third arguments of each of `m1`, `m2`, and `m3` are the actual creator, the seeming sender, and the receiver of the corresponding message. The first argument is meta-information that is only available to the outside observer and the principal that has sent the corresponding message, and that cannot be forged by the intruder; while the remaining arguments may be forged by the intruder.

The network is modeled as a multiset of messages, which the intruder can use as his/her storage. Any message that has been sent or put once into the network is supposed to be never deleted from the network because the intruder can replay the message repeatedly, although the intruder can not forge the first argument. Consequently, the empty network (i.e., the empty multiset) means that no messages have been sent.

In this paper, a state is expressed as a soup of observable components. Let ms, rs, ns, and ps be the collections of messages, random numbers, nonces, and principals, respectively. ps may contain an intruder. Let $c1s$, $c2s$, and $c3s$ be the collections of `Cipher1`, `Cipher2`, and `Cipher3` ciphertexts, respectively. To formalize the NSLPK protocol as a state machine M_{NSLPK}, we use the following observable components:

- (`nw` : ms) - it says that the network is constructed by ms,
- (`cenc1` : $c1s$) - it says that the collection of `Cipher1` ciphertexts gleaned by the intruder is $c1s$,
- (`cenc2` : $c2s$) - it says that the collection of `Cipher2` ciphertexts gleaned by the intruder is $c2s$,
- (`cenc3` : $c3s$) - it says that the collection of `Cipher3` ciphertexts gleaned by the intruder is $c3s$,
- (`nonces` : ns) - it says that the collection of $nonces$ gleaned by the intruder is ns,
- (`prins` : ps) - it says that the principals participating in the protocol are ps,
- (`rand` : rs) - it says that the available random numbers are rs. Every time a principal wants to send an Init or a Resp message, it needs to generate a random and globally unique number. To formalize that behavior, we provide a fixed collection of random numbers from the beginning, and every time a process needs to generate a random number, an element is extracted and used.

Each state in S_{NSLPK} is expressed as $\{obs\}$, where obs is a soup of those observable components. We suppose that two principals `p` & `q` together with an intruder participate in the NSLPK protocol, one initial state of I_{NSLPK} namely `init` is defined as follows:

```
{(nw: emp) (rand: (r1 r2)) (nonces: emp)
(cenc1: emp) (cenc2: emp) (cenc3: emp)
(prins: (p q intr))} .
```

where `intr` is a constant of `Prin` denoting the intruder, and `emp` denotes an empty collection.

Three rewrite rules `Challenge`, `Response`, and `Confirmation` formalize three actions when a principal sends an Init, a Resp, and an Ack message, respectively. Let `OCs` be a Maude variable of observable component soups; `P` & `Q` be Maude variables of principals; `Ps` be a Maude variable of collections of principals; `NW`, `R`, and `N` be Maude variables denoting a network, a random number, and a nonce, respectively; `Rs`, `CE1`, and `Ns` be Maude variables denoting a collection of random numbers, `Cipher1`, and nonces, respectively. The rewrite rule `Challenge` is defined as follows:

```
rl [Challenge] : {(nw: NW) (prins: (P Q Ps))
(rand: (R Rs)) (cenc1: CE1) (nonces: Ns) OCs }
=> {(nw: (m1(P,P,Q,enc1(Q,n(P,Q,R),P)) NW))
(cenc1: (if Q == intr then CE1 else
(enc1(Q,n(P,Q,R),P) CE1) fi)) (nonces:
(if Q == intr then (n(P,Q,R) Ns) else Ns fi))
(rand: Rs) (prins: (P Q Ps)) OCs} .
```

The rewrite rule says that when `R` is in `rand`, a new Init message is put into the network, intruder gleans the nonce and the ciphertext used in that message if that message sends to the intruder, and `R` is removed from `rand`.

In addition to the three rewrite rules that formalize sending messages exactly following the protocol mentioned above, we also introduce six more rewrite rules to formalize the intruder's faking messages:

- `fake12`, `fake22`, and `fake32`: a ciphertext `C` is available to the intruder, the intruder fakes and sends an

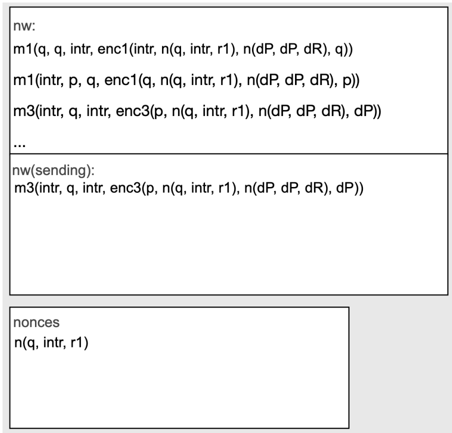


Figure 1. A simple state picture for the NSLPK protocol (1)

Init, or a Resp, or an Ack message using `C`, respectively.

- `fake11` and `fake31`: a nonce `N` is available to the intruder, the intruder fakes and sends an Init or an Ack message using `N`, respectively,
- `fake21`: two nonces `N1` and `N2` are available to the intruder, the intruder fakes and sends a Resp message using `N1` and `N2`.

The rewrite rule `fake11` is defined as follows:

```
rl [fake11] : {(nw: NW) (nonces: (N Ns))
(prins: (P Q Ps)) (cenc1: CE1) OCs} =>
{(nw: (m1(intr,P,Q,enc1(Q,N,P)) NW)) (cenc1:
(if Q == intr then CE1 else (enc1(Q,N,P) CE1)
fi)) (nonces: (N Ns)) (prins: (P Q Ps)) OCs} .
```

The rewrite rule says that when `N` is in `nonces`, a new intruder's faking Init message is put into the network, and the intruder gleans the ciphertext sent in that message.

The remaining rewrite rules can be defined likewise.

III. State Picture Design of NSLPK Protocol

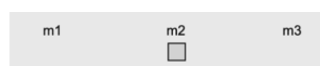
The network component, which consists of many messages, is the main part of the protocol that we should focus on. Initially, we try to make a design for the network in which Bui and Ogata [6] used, as shown in Fig. 1. The design, however, is hard to observe and/or analyze the messages in the network because there are many contents inside each message. As shown in Fig. 1, there are three rectangles in which the first rectangle represents a network that contains all messages, the second one displays the most recent message that has been put into the network, and the collection of *nonces* gleaned by the intruder is displayed in the last rectangle. "..." is displayed whenever the content of the network is overflowed. During making a better state picture design, by observing that the number of messages increases by one after each state, we come up with an idea that displays the contents of the most recent message that has been put into the network (hereinafter, let us call such a message as the latest message).

Although there are three kinds of ciphertexts (i.e., `enc1`, `enc2`, and `enc3`), in the state picture design, we use only one form to visualize ciphertexts. The form is as follows: enc$i$(public-key, nonce1, nonce2, cipher-creator), where public-key is a principal (possibly `intr`), nonce1 for m1, m2, and m3 is in the following form: nonce1(generator, random, forwhom); nonce2 is in the following form: nonce2(generator, random, forwhom). When the ciphertext is in the form of `enc3`, cipher-creator receives a dummy principal `dP` as its value. Similarly, when the ciphertext is in the form of `enc1` or `enc3`, nonce2 receives a dummy value denoted by nonce2(`dP`,`dR`,`dP`), where `dR` denotes a dummy random number.

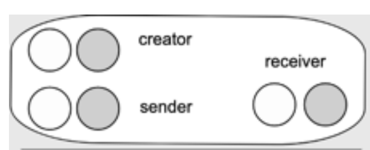
Fig. 2 depicts our state picture design. Some designs are used from state picture design tips of the work [5]. Fig. 3 displays a state picture. We first divide two roles that are creators and senders into two separate places. Then, observable components are put to the corresponding place in which their roles seem to belong. For example, public-key should be put to the receiver's side because the sender uses the public-key of the receiver for encrypting. Values are displayed with different colors and shapes. For example, pink and light yellow colors represent two different principals, blank represents `intr`, triangles represent the contents of the nonce.

We describe the details of the state picture design. The representation of the three types of messages designed in Fig.2 is as follows:

The type of the lastest message is represented by a small light gray square. For example, when the lastest message is a message m2, there is only one light gray square displayed under m2 as shown in the following picture:



The representations of the creator, sender, and receiver of the message used in Fig. 2 are as follows:



The creator of the message appears at the top-left place, pink and light yellow circles represent two different principals `q` and `p`. If the value is `intr`, nothing is displayed. The sender and receiver of the message appear at the bottom-left and bottom-right places, respectively. For example, when creator is `intr`, sender is `p`, receiver is `q`, it is displayed as follows:

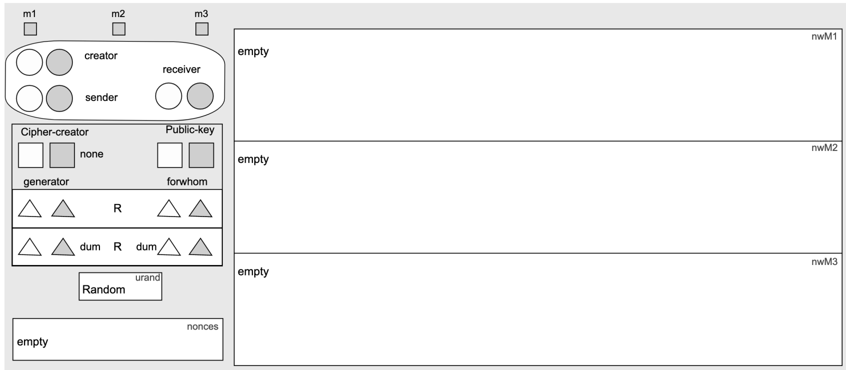


Figure 2. A state picture design for the NSLPK protocol (1)

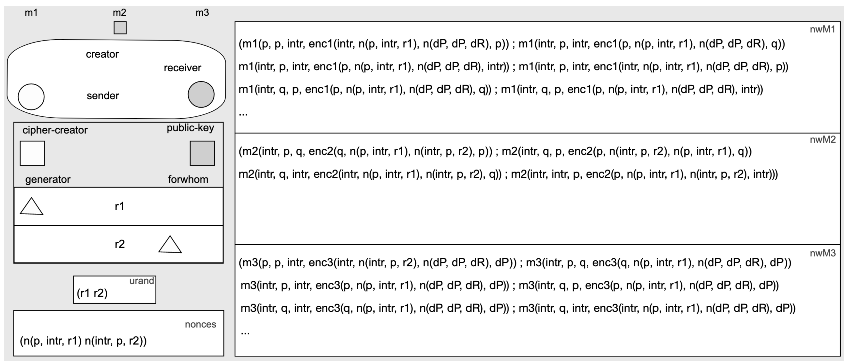
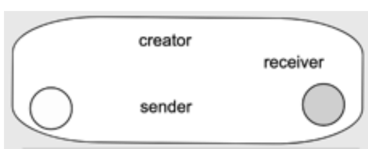
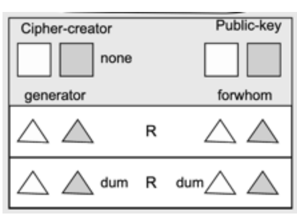


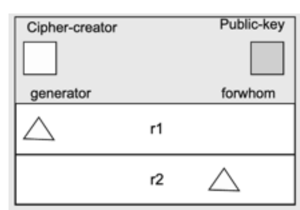
Figure 3. A state picture for the NSLPK protocol (1)



The representations of the contents of the ciphertext shown in Fig. 2 are as follows:



The cipher-creator of the ciphertext appears at the top-left place of the rectangle, pink and light yellow squares represent two principals `q` and `p`, respectively. If the value is `intr`, nothing is displayed. For the case the message is a message `m3`, the text "none" is displayed. The public-key of the ciphertext appears at the top-right place. If the value is `intr`, nothing is displayed. The two nonces of the ciphertext are shown with two rectangles inside the primary rectangle, where the upper rectangle visualizes the first nonce and the lower rectangle visualizes the second nonce. In the first nonce, the generator and forwhom representations appear at the left-hand side and right-hand side, respectively; pink and light yellow triangles are the principals `q` and `p`, respectively. If the value is `intr`, nothing is displayed. The random representation appears at the middle place in which the random number value used is displayed. The second nonce is represented likewise. If the message is a message `m3`, the text `dum` is displayed for the values of generator and forwhom, where `dum` denotes the dummy value `dP`. Considering the following example. cipher-creator is `p` and public-key is `q`. In the first nonce generator is `p`, random is `r1`, and forwhom is `intr`. In the second nonce, generator is `intr`, random is `r2`, and forwhom is `p`. Those values are displayed as follows:



In Fig. 2, the representations of urand and nonces are designed at the left-bottom corner. The values of both urand and nonces are displayed using two rectangles as follows:

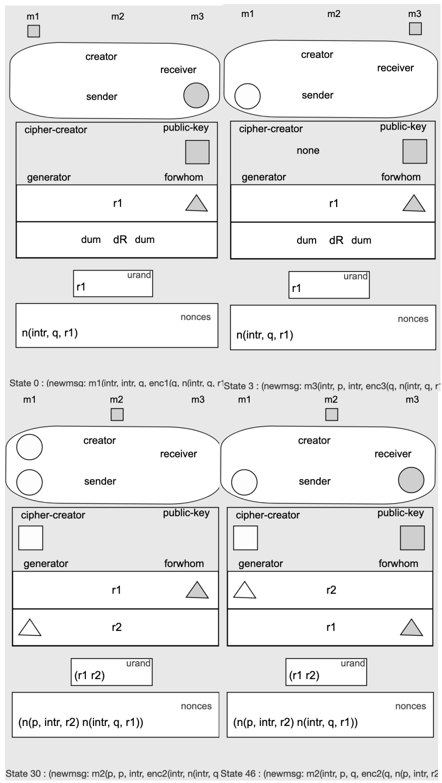
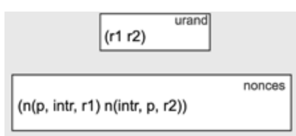
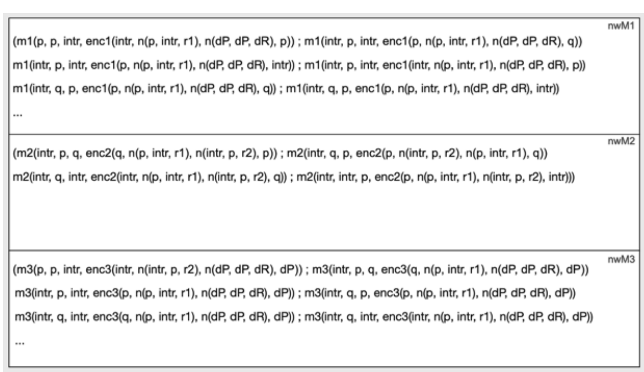


Figure 4. Some state pictures for the NSLPK protocol (1)



In Fig. 2, three types of network representations are designed on the right side. “...” is displayed whenever the messages are overflowed. This can be seen in the figure below:



IV. Characteristics Guessed Based on Our Design

We sometimes need to concentrate on some specific OCs when we observe the graphical animations. Most of the characteristics of the NSLPK protocol are straightforward to guess by observing graphical animations. However, some are not, precisely the characteristics that include two messages. We use Maude to generate a finite input sequence of states based on the Maude specification of the protocol, then feed it to SMGA which produces graphical animation of the input sequence of states.

Fig. 4 shows four pictures of states for M_{NSLPK}. Taking a look at the first picture (of State 0) and the second picture (of State 3) helps us recognize that there is `n(intr,q,r1)` in `nonces` when `generator` is `intr` and taking a look at the third picture (of State 30) and the fourth picture (of State 46) helps us recognize that there is `n(p,intr,r2)` in `nonces` when `forwhom` is `intr`. Any nonce gleaned by the intruder is stored in `nonces`. Hence, observing the graphical animation of these four pictures helps us guess the characteristic such that any nonce gleaned by the intruder has been generated by the intruder or a non-intruder principal that wanted to authenticate the intruder.

Taking a look at the second picture (of State 3) and the third picture (of State 30) allows us to guess another characteristic such that whenever `receiver` is `intr` (that displays blank in the state pictures) in the latest message, then the nonce of that message is in `nonces`. Carefully observing graphical animations helps us perceive one more characteristic. Taking a look at the four pictures of Fig. 4, we recognize the characteristic that when a nonce is in `nonces`, the random number used in the nonce is stored in the collection of used random numbers `urand`.

We prepare another input file that consists of a finite sequence of states so that we can guess more characteristics by observing the behavior of the protocol. To guess some non-trivial characteristics, we concentrate on the order in which messages have been sent. Carefully observing the order of messages, especially that a message `m2` should follow a message `m1`, as Fig. 5, we guess a characteristics that involves two messages. Taking a look at the first picture (of State 0), there exists a message `m1(p,p,q,enc1(q,n(p,q,r1),n(dP,dP,dR),p))` in `nwM1`. After some `m1` messages are faked by the intruder based on the gleaned information, there exists a message `m2(q,q,p,enc2(p,n(p,q,r1), n(q,p,r2),q))` in `nwM2` at the second picture (of State 5). Taking a look at the third picture (of State 19), we observe that the intruder creates many faked `m2` messages including `m2(intr,q,p,enc2(p,n(p,q,r1),n(q,p,r2), q))`. Observing the order of messages in the network allows us to conjecture the following characteristic:

- if there exists a message `m1` created by a non-intruder principal and sent to another non-intruder principal, and
- there exists a message `m2` (either created by the intruder or a non-intruder principal) that is sent to the sender of `m1`, then
- the message `m2` originates from a non-intruder principal who is the receiver of the `m1`.

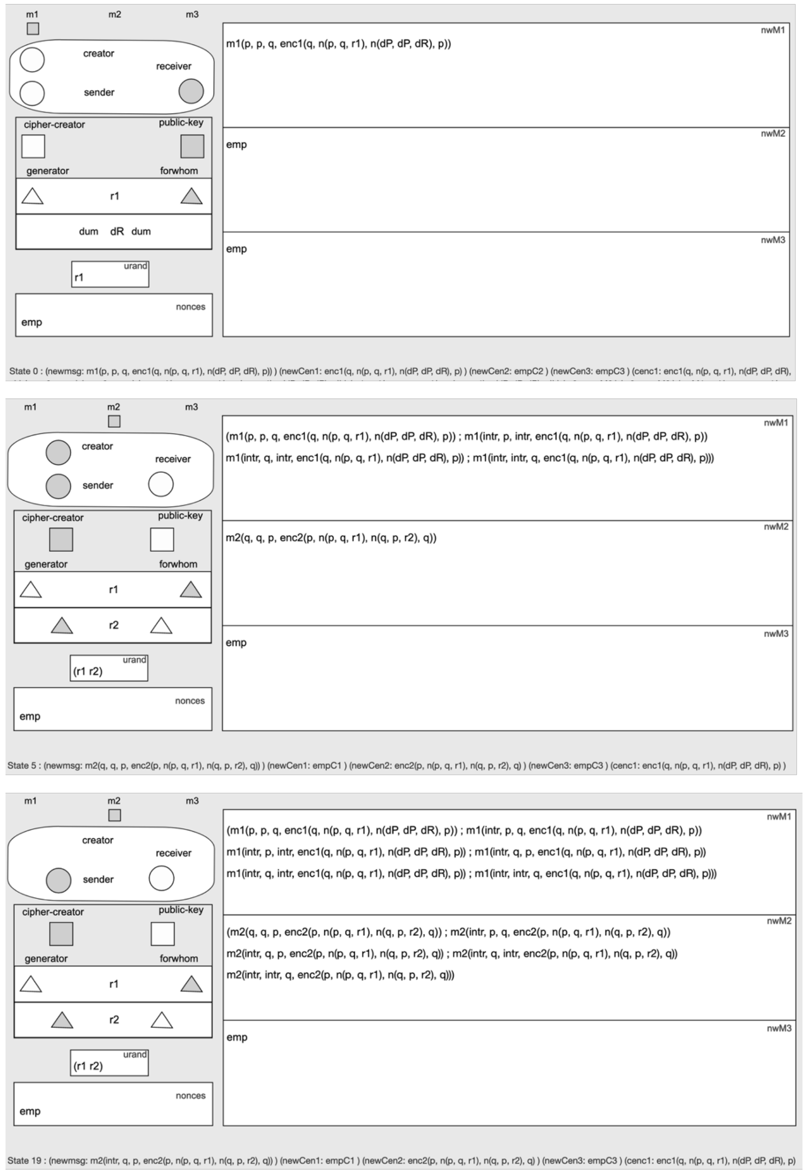


Figure 5. Some state pictures for the NSLPK protocol (2)

Similarly, we expect that a message `m3` should follow a message `m2`. There is a message `m2(q,q,p,enc2(p,n(p,q,r1),n(q,p,r2),q))` in `nwM2` at the second picture (of State 5). Taking a look at the first picture (of State 40) in Fig. 6, there exists a message `m3(p,p,q,enc3(q,n(q,p,r2),n(dP,dP,dR), dP))` in `nwM3`. At the second picture (of State 43), there exists a message `m3(intr,p,q, enc3(q,n(q,p,r2),n(dP,dP,dR),dP))` in `nwM3` which is created by `intr`. Carefully observing the order of the messages in the network, we also guess the following characteristic:

- if there exists a message m2 created by a non-intruder principal and sent to another non-intruder principal, and
- there exists a message m3 (either created by the intruder or a non-intruder principal) that is sent to the sender of the message m2, then
- the message m3 originates from the non-intruder principal who is the receiver of the message m2.

Maude search command can be used as an invariant model checker to check that the NSLPK protocol enjoys the guessed characteristics. The guessed characteristics are confirmed by the search command at a specific depth (depth 5) of the state space because the reachable state space (generated by Maude) of the protocol is too huge to be exhaustively traversed. The search command does not find any counterexample at depth

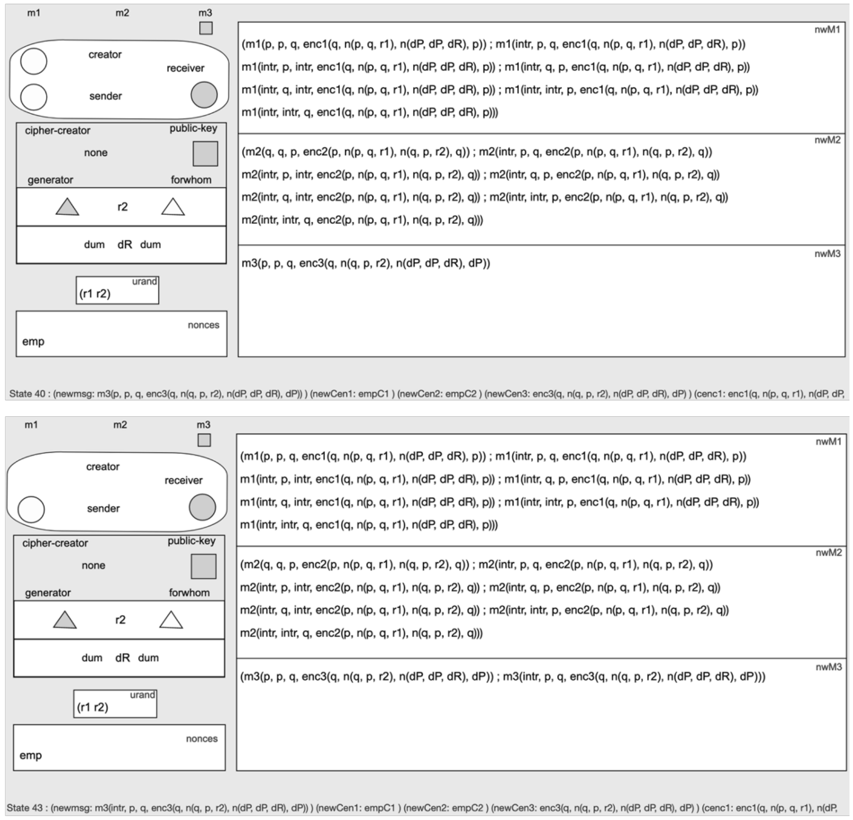


Figure 6. Some state pictures for the NSLPK protocol (3)

5. It means that the NSLPK protocol seems to enjoy the guessed characteristics.

V. Conclusion

We have graphically animated the NSLPK authentication protocol with SMGA. Observing the graphical animations based on our design allows us to guess some (non-trivial) characteristics of the state machine formalizing the NSLPK protocol. We have checked the characteristics by Maude search command. Although some model checking experiments were not completed because of the state space explosion problem, some characteristics of NSLPK have been proved [12], guaranteeing that the characteristics are invariant properties of NSLPK. One piece of our future work is to graphically animate state machines that formalize other authentication protocols, such as TLS [13], with SMGA.

References

[1] T. T. T. Nguyen and K. Ogata, "Graphical animations of state machines," in *15th DASC*, 2017, pp. 604–611.

[2] K. W. Brodlie, et al., Ed., *Scientific Visualization: Techniques and Applications*. Springer, 1992.

[3] M. T. Aung, T. T. T. Nguyen, and K. Ogata, "Guessing, model checking and theorem proving of state machine properties – a case study on Qlock," *IJSECS*, vol. 4, no. 2, pp. 1–18, 2018.

[4] T. T. T. Nguyen and K. Ogata, "Graphically perceiving characteristics of the MCS lock and model checking them," in *7th SOFL+MSVL*, 2017, pp. 3–23.

[5] D. D. Bui and K. Ogata, "Better state pictures facilitating state machine characteristic conjecture," *MTAP*, 2021.

[6] ——, "Graphical animations of the Suzuki-Kasami distributed mutual exclusion protocol," *JVLC*, vol. 2019, no. 2, pp. 105–115, 2019.

[7] G. Lowe, "An Attack on the Needham-Schroeder Public-Key Authentication Protocol," *Inf. Process. Lett.*, vol. 56, no. 3, pp. 131–133, 1995.

[8] M. Clavel, et al., Ed., *All About Maude*, ser. LNCS. Springer, 2007, vol. 4350.

[9] C. Artho, et al., "Visualization of concurrent program executions," in *31st COMPSAC*, 2007, pp. 541–546.

[10] T. T. T. Nguyen and K. Ogata, "A way to comprehend counterexamples generated by the Maude LTL model checker," in *SATE 2017*, 2017, pp. 53–62.

[11] R. M. Needham and M. D. Schroeder, "Using Encryption for Authentication in Large Networks of Computers," *Commun. ACM*, vol. 21, no. 12, pp. 993–999, 1978.

[12] K. Ogata and K. Futatsugi, "Rewriting-based verification of authentication protocols," *ENTCS*, vol. 71, pp. 208–222, 2002.

[13] T. Dierks and C. Allen, "The TLS protocol version 1.0," *RFC*, vol. 2246, pp. 1–80, 1999.

SENECA: An Attention Support Tool for Context-related Content Learning

Alessia Auriemma Citarella, Luigi Di Biasi, Stefano Piotto, Michele Risi, Genoveffa Tortora
Department of Computer Science
University of Salerno
84084 Fisciano (SA), Italy
{aauriemmacitarella, ldibiasi, piotto, mrisi, tortora}@unisa.it

Abstract

In this work, we proposed a tool named SENECA that aims to help the students who follow remote lessons to maintain/capture attention, allowing them to focus on learning led by the context. Among the disadvantages of distance education, especially for subjects who lack awareness, the greatest distractions at home are counted. These distractions cause a movement of the student's attention from the current lesson to disturbing events. For this reason, there is a need to experiment with new solutions also linked to Information Technology *(IT) to improve the focused learning during distance education. Our tool's technical idea is to create a real-time summary of the topic treated by the teacher. The system captures the text every five minutes, generates outlines, and scratches them and browses them to eliminate repetitive portions after each survey. On the general generated summary, Natural Language Processing techniques are applied to extract categories and keywords. The latter will show the highlights of the speech.*

Index terms— Natural Language Processing, Semantic Analysis, Distance Education, Learning, Attention, Content

1 Introduction

Into the current pandemic situation generated by the SARS-CoV-2 coronavirus, the educational environment worldwide has to face numerous challenges to continue teaching in schools and universities.

One of the most important aspects of cognitive function is the *"ability to keep"* relevant information in mind. *Working memory* is a system dedicated to the maintenance and temporary processing of information during cognitive processes. One of the components is represented by the *central executive* which carries out the coordination of subordinate systems, coordination of execution of tasks and recovery of strategies and attentional functions of both selection and inhibition [3]. The central executive controls the *phonological loop* which contains verbal and auditory information, the *visuo-spatial sketchpad* engaged in spatial representation and the *episodic buffer* which has a limited ability to link information from different sources with spatial and temporal parameters.

Specifically, each attentive act is divided into three phases: the orientation and perception towards the different stimuli; the processing phase that presents the function of selectivity and sustained attention overtime on a task or activity, the shift to move the focus quickly and the ability to pay attention to use the right cognitive resources in different situations; the specific response concerning the input stimuli [4].

Different studies have focused on the impact of technologies on cognitive functions in the present digitized era, both from the perspective of the benefits and disadvantages [38]. *Lodge and Harrison* [25] stressed as attention is subject to complex dynamics that impact learning, especially in educational contexts. The most important part of a sentence, oral or written, is the focus. Recent articles have demonstrated the importance of *marking elements* as a *guide* for better information exchange [24] between speakers and listeners. In particular, these studies argue that focus marking captures the listener's attention to what the speaker considers the most relevant part of the message. At the same time, this method helps the attention to be kept on the marked element allowing its representation [32].

Recent work has addressed the issue of distance education by administering questionnaires to both teachers and pupils. The most variable answers to the questions were also obtained on the degree of students' participation in distance lessons, emphasizing a wide range of behaviors. Furthermore, perception of difficulty during remote lessons was found to be linked to many factors: access to technology, motivation and support with a greater presence of negative experiences [26]. The new educational needs of online teaching and students' changing learning styles limit knowledge transfer comprehensively and effectively. During the detachment of presence foreseen by distance teach-

DOI reference number: 10.18293/DMSVIVA2021-006

Figure 1: A standard bidirectional media streaming.

ing, many dysfunctional behaviors can be generated, such as the loss of interest, attention and motivation for psycho-physical causes and non-adaptation to an abnormal situation. Compared to the standard educational environment, distance learning has a disadvantage in terms of distractions. It is an isolated experience in which there is no direct communication that makes participation much more active [14]. It would be useful to overcome the difficulties of maintaining a student's attention, regardless of the need or not of a situation that requires the use of tools for distance learning. Overcoming these issues would help refine each pupil's strategic learning styles and ensure a meta-cognitive self-assessment approach on one's limits and abilities, supported by technology. Recent meta-cognitive skills also include a student's ability to become aware of his or her ability to "*learning how to learn*". This ability means recognizing and then consciously applying appropriate behaviors and strategies useful for a more effective learning process [29].

This paper proposes a Distributed Multimedia System for support learning, designed to face the loss of attention during distance education. The purpose of the system is to be able to reawaken or maintain attention to the context (topic) that is being experienced during the activity in progress (in *real time*) to reduce the negative effect of distractions. The system also aims to provide the possibility of an in-depth analysis at the end of the lesson through auto-generated hyperlinks to lesson-related content. The architecture proposed rely on *Speech-to-text*, *Natural Language Processing* (NLP), Text Summarization [18] and *Semantic Analysis technology*.

This document is organized as follows. In Section 2, we described the most important related works about the technological systems that help in learning. In Section 3 we introduced the used methodologies. In Section 4 we have highlighted the working hypotheses on which we based our work. In Section 5 we presented the system architecture and in Section 6 we have detailed the performed experiments and the related results (Section 7). Finally, in Section 8 we will discuss research directions and future development of our work.

2 Related works

In this section, we present some related works about the use of semantic and NLP analysis technologies. Specifically, we will discuss different application contexts of these techniques. Most of a student's effort is to transfer information from working memory to long-term memory to acquire and memorize key concepts. Two strategies can be used: dual coding and chunking [10]. In cognitive psychology, a *chunck* is nothing more than a unit of information, and chunking is the operating mode in which this unit of data is recovered. When faced with new knowledge, the individual can grasp the relative chunk of information and bring it back to light later when recalling a similar situation or concept. Then, the initial piece can be expanded into more complex pieces following the management control and understanding of the flows of one's knowledge [35].

The standard structure of a *Distance Educational System* can be generalized as reported in Fig. 1. This kind of system relies on the classic bidirectional multimedia connection, like the common video-chat system based on SIP/VOIP system, such as Microsoft Teams [23]. However, a Distance Educational System supports multiple bidirectional connections between students and teachers (or teachers) and allows channel moderation.

Nowadays, the cloud's audio/video stream transfer services are implemented by the major world providers (*Amazon, Microsoft, Zoom*). A teacher can teach remotely by transmitting an audio/video stream from their home to one of these providers. Then these last provide a broadcast service to the students.

NLP techniques have been widely used in intelligent tutoring systems that helped acquire content knowledge [8]. For example, in Guzmán-García *et al.* [21], the analysis of the surgeons' speaking taken into the operating room through NLP techniques is proposed to obtain a deeper vision of intraoperative decision-making processes. This study aimed to develop a method of recognizing and evalu-

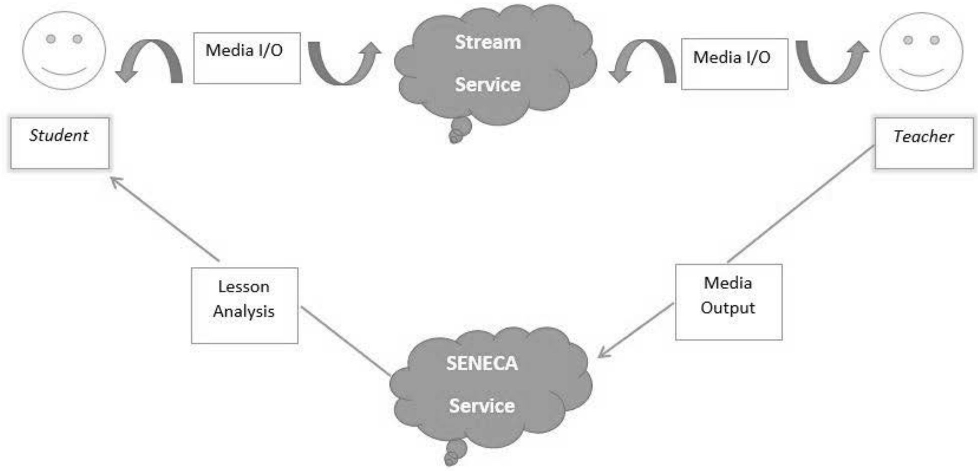


Figure 2: The additive layer for multimedia analysis.

ating the various surgical phases and developing a workflow comparable with the framework of the procedure to improve surgical learning in The Educational Operating Room.

Recent studies have highlighted how to identify the main contents to better understand the topic, especially in students with cognitive deficiency, attention, or memory. The ability to take advantage of text summarization techniques by explaining the main idea allows students to interface with the limits of their working memory and have a tool to overcome their difficulties [34].

Today, many educational and academic institutions benefit from the *Learning Management Systems* (LMS) to support and improve teaching processes [15][19]. Most LMS is software application systems that allow teachers to manage and deliver educational courses [2]. One of the requirements for the success of distance education is traceable in the self-management of learning which is the starting point for self-discipline in autonomous learning [36].

In 2019, Cobos *et al.* [12] have developed EdX-CAS, a content analyzer system for edX MOOCs, using NLP techniques for the Spanish language. The tool takes in input the video transcripts of the courses, with which users can interact specifically. It allows students to extract the text's main terms, the vector representation for each of the terms in the text, the linguistic diversity to understand how many different words are used, indications on the subjective opinion on the text and the representation with word clouds. The edX-CAS tool is oriented to Sentiment Analysis Opinion Mining for Detecting Subjectivity and Polarity Detection in Online Courses related to Madrid's Universidad Autónoma.

On the topic of the educational distance imposed by the Covid-19, to support students in self-training, a chatbot was proposed using NLP techniques [16]. The proposed solution involves sending a message to Moodle [1] by the student. An associated plugin tries to understand the text received and returns feedback. Based on the degree of assessment achieved by the student, the chatbot provides recommendations on the chapters for which the evaluation is not sufficient. The system presupposes the memorization of the evaluation outcomes of the student who are accessible to teachers. The chatbox, in this context, acts as a tutor and allows us to fill the gaps of the students.

3 Methods

We propose a new tool called SENECA (Support lEarning coNtEnt Context Attention), which involves using a new layer dedicated exclusively to analyzing the audio/video streams generated by the "teachers". Our purpose is to maintain the student focus or a rapid return to it by implementing this new layer. The proposed architecture is to be considered feasible for real-time distance lessons and not to the MOOC [5] or the on-demand recorded lessons. We did not consider the capabilities related to file upload, file sharing and homework as an added values.

The new layer is identified in the Fig. 2 as Seneca Service. In this proposal, we refer only to real-time audio/video streams, i.e., not recorded lessons held at a distance. The real-time component's presence allows us to immerse ourselves in a learning context susceptible to disturbances that distract the individual student.

SENECA's main goal is to help students avoid loss of concentration by providing multiple information that can allow students to maintain focus on the argument or bring attention back to the context in the event of distractions. Another important goal is to favor applying many text analysis techniques to improve the learning quality of the topic under study and integrate it, providing additional information that can be used at the end of the lesson to recover information further.

4 Working hypothesis

In this section, we aim to highlight some working hypotheses that allow us to have a more understandable general view of the SENECA tool's starting idea:

a) The real-time audio/video stream (from here called `STREAM`) generated by a remote lesson can be split into two unique sub-streams: `VIDEO` flow, containing the video frames and `AUDIO` flow containing audio buffer.

b) The `VIDEO` flow will contain information from slides or, in any case, projected material to provide a conceptual map to students.

c) The `AUDIO` flow will contain the lesson audio, and it is expected to add information on both the context under study and in-depth study (as well as student questions or others).

In this context, we assume that the `VIDEO` stream contains *information already summarized on the subject.* In our experiment, we considered the data extracted from `VIDEO` as already cleaned. On the other hand, the presence of heterogeneous data in `AUDIO` streams will require a more accurate analysis of the content.

The extracted data from `VIDEO` and `AUDIO` is referred following as `WORD STREAMS`.

5 System architecture

We designed a prototype architecture based on a pipeline approach, like Microsoft DirectShow [1] or ffmpeg [2]. In SENECA each computational block is called *Filter*.

An overview of a complete SENECA architecture is shown in Fig.3. For this proposal, we implemented only the following filters: SPLIT, OCR, STT, SUMMARY and SEMANTICS.

The filters are defined as following:

$$SPLIT(STREAM) \rightarrow \{AUDIO, VIDEO\}$$

$$OCR(VIDEO) \rightarrow \{WS\}$$

$$STT(AUDIO) \rightarrow \{WS\}$$

$$SUMMARY(WS, GLS) \rightarrow \{NEW\ GLS\}$$

$$SEMANTIC(GLS) \rightarrow SUGGESTIONS$$

The SPLIT filter takes as input the `STREAM` and splits it into two separate flows, called `AUDIO` and `VIDEO`.

The OCR technique allows the detection and extraction of text from images [28]. The SENECA OCR Filter takes as input a single frame video at a time. We used the videocr python module (v. 0.1.6)[3] for our experiment purpose. That module lies on Tesseract OCR 4.1.1. [4]. This filter analyzed each video frame from the pipeline and stored the detected text (handwritten and block letters) into a word stream (`WS`). Each word stream was enqueued into the next pipeline filter.

In SENECA, the STT filter performs a speech-to-text routine. Speech-to-text is a technique that allows the detection and the extraction of phrases from an audio flow `WS` [11]. Probably, the most commonly known example is Amazon Alexa or Google Assistant. Into our prototype, we used the Google Cloud Speech API https://cloud.google.com/speech-to-text/. For each audio frame extracted by the SPLIT, the STT filter generated a word-stream (`WS`) that was enqueued to the NPL filter.

One of the project goals is to provide a way to regain the attention on the topic focus after a distraction. In SENECA, one of the tips is to allow users to summarize the lesson in real-time. As shown in figure 3, the media flow (as an example, a streamed lesson) comes into the SPLIT filter that separates audio from video. For each video frame extracted, the OCR extracts the detected phrases, and STT makes the same action on the audio frame. In particular, these word streams came into the next filter into SUMMARY filter that is a delegate to create partial summaries using the word streams that came into the filter. Into our prototype, the SUMMARY filter computes a summary for the `WS`s using MEAD [18]. These multiple summaries are merged every 5 minutes into the Global Lesson Summary (`GLS`) that is processed again by MEAD. We have chosen the five minutes interval using the mean lesson length. Consequently, SENECA builds and refreshes a `GLS` by using SUMMARY filter output for each real-time lesson. Into our prototype, `GLS` is composed of phrases generated by applying MEDA text summarization algorithm on `WS`s.

The SEMANTIC filter extracts the `MEF` from the `GLS` every time a new `GLS` is deployed from the SUMMARY filter. We identify with the term `MEF` or *Most Expressed Features* of a string S, the dictionary D(S) of all possible *k-mers* extracted from S, using substrings length between m

[1] https://docs.microsoft.com/en-us/windows/win32/directshow/directshow
[2] https://ffmpeg.org/developer.html
[3] https://pypi.org/project/videocr/
[4] https://github.com/tesseract-ocr/tesseract

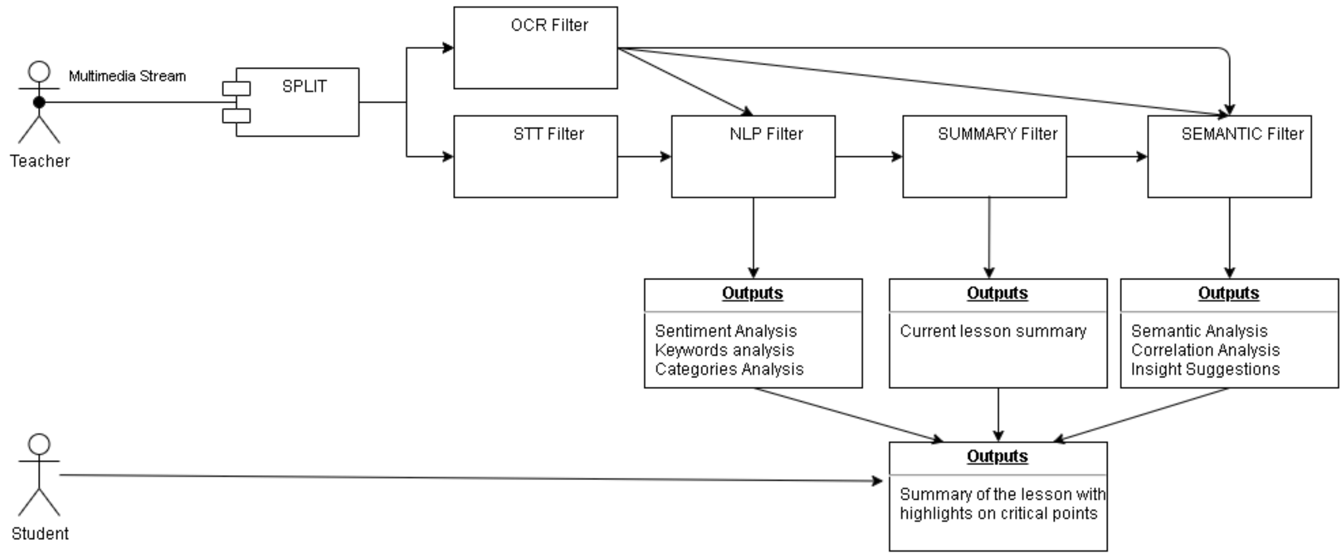


Figure 3: A basic SENECA module.

Table 1: Partial subset of lessons MEF.

Topics	**MEFs**
Cancer	Smoking, Colon Cancer, Surgery, Risk Factor
Diabetes	Beta Cell, Interleukin, Inflammatory, Physic
Evolution	Selection, Heritability, Billion Years, Coevolution
Terrorism	Terror, Povery, Success, Politican
Chemistry	Compound, Energy, Element, Electron

and M. The dictionary is ordered in a decreasing way, compared to the number of occurrences of each k-mers. The MEFs represent the object's functional parts, such as words or portions of sentences of a text repeated several times. SEMANTIC is designed to use the `MEF` for probing one or more scientific databases to make suggestions. In particular, it performs a combined NO-SQL Alignment-free search into the pre-processed PUBMED database (see next paragraph). For each GLS, SEMANTIC can extract the candidate literature papers indexed by MEF. It uses two metrics to compute (see experiment one) the suggested papers based on the semantic distance between GLS and candidate paper set.

For this prototype, we used the PUBMED database [9]. We execute the SEMANTIC filter on the entire PUBMED dataset, and we have extracted the `MEF`s, using substrings length interval between m=3 and M=15.

Due to the PUBMED dataset size, we used Amazon EC2 and Amazon RDS services [30] to distribute `MEF`s extraction and storage.

6 Experiment Execution

We simulated real-time lessons by using public videos from Coursera[5]. We selected five free courses, recovering from each it, the text subscription using STT. The main topics of the lessons are Cancer; Diabetes; Evolution; Terrorism, Chemistry.

We implemented two experiments. Our goal was to study the SUMMARY and SEMANTIC filter performances.

6.1 First Experiment

Into the first experiment, we sent all the entire lessons (5 merged videos per lesson) into the SENECA pipeline, sequentially, to compute separate `GLS` outputs for each entire lesson. Also, we sent each video (one at a time, not merged) into the pipeline to generate the GLS for each video.

We wanted to study if the SEMANTIC filter was able to discriminate between lesson topics. We applied the SE-

[5]https://www.coursera.org/

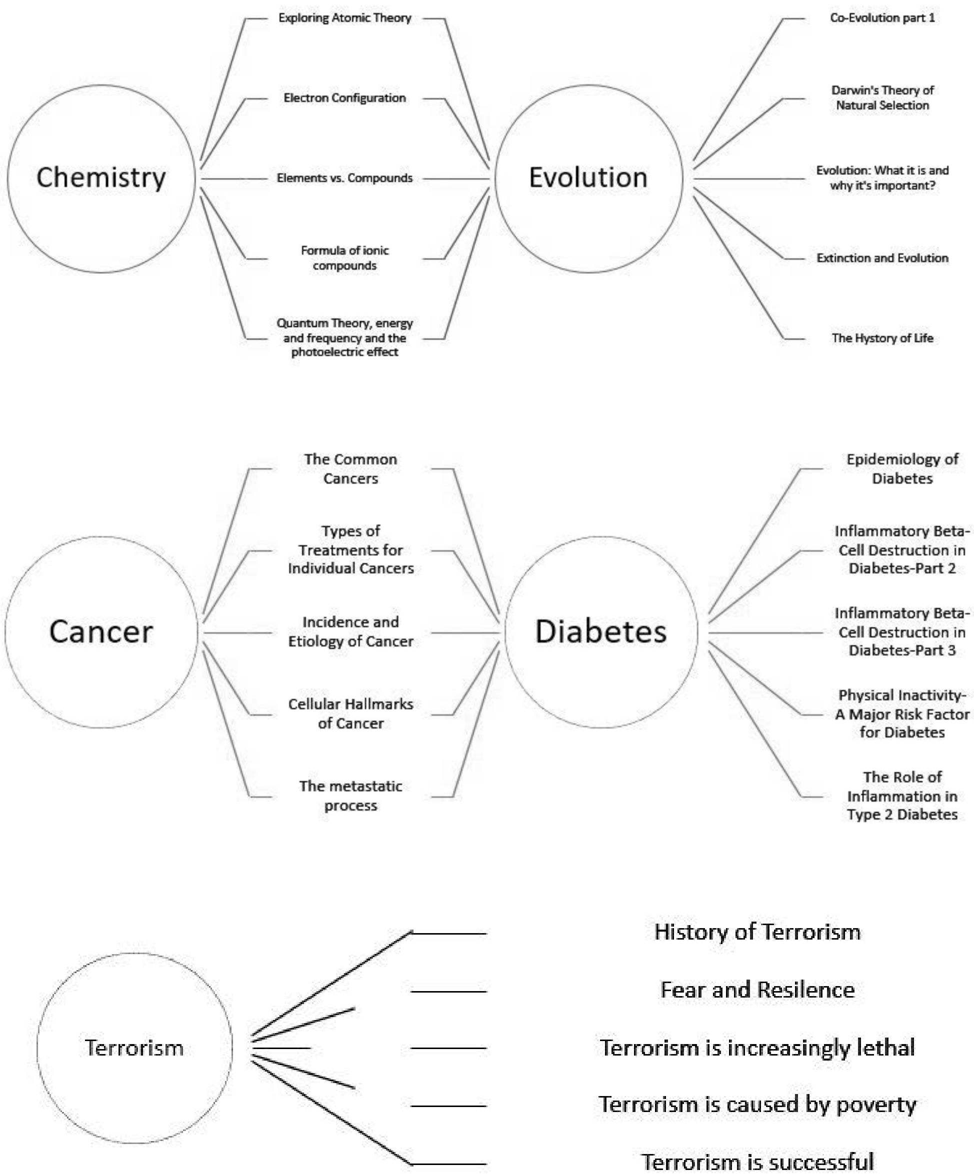


Figure 4: Lesson Topics clustering using Jaccard distance.

MANTIC filter on each `GLS` to extract the `MEF`s for each of them.

Using the `MEF`s, we were able to build two distance matrices. We used two different distance metrics. The distances were calculated using the *Jaccard Index* [31] (see Eq. 1), and by *Szymkiewicz–Simpson coefficient (SSC)* [39] (see Eq. 2).

Both metrics allow us to compute the similarity between pair of `MEF` dictionary. SSC is often identified as the *"overlay coefficient."*

The Jaccard index is defined as:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

where *A* and *B* are two different datasets, whilst the $|\bullet|$ operator computes the size of a set. In particular, the Jaccard index is represented as the size of the intersection divided by the size of the union of the datasets.

Given the two dictionaries *A* and *B*, the overlap coefficient is a measure that returns the overlap between them,

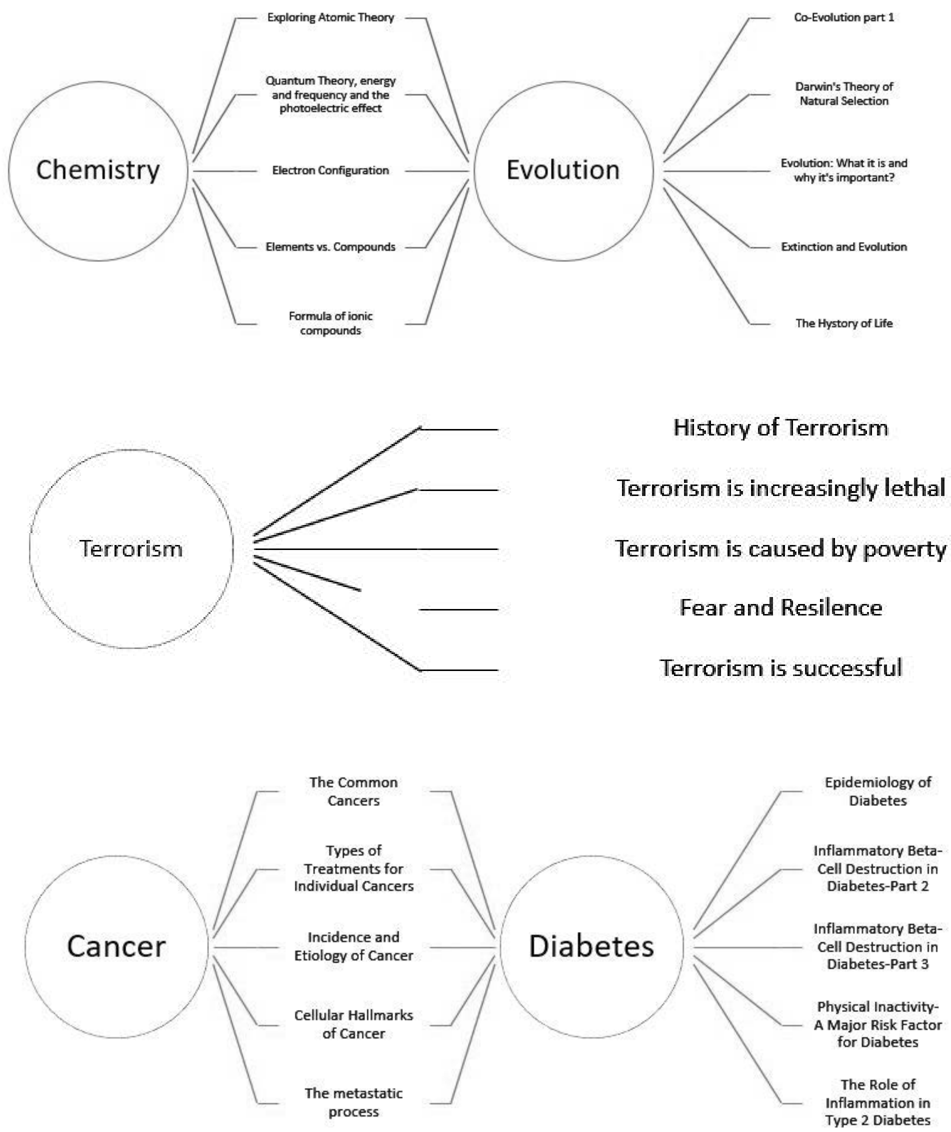


Figure 5: Lesson topics clustering using SSC.

and it is defined as the intersection divided by the smaller of the size of the two sets, as shown in Eq. 2.

$$SSC(A, B) = \frac{|A \cap B|}{min(|A|, |B|)} \quad (2)$$

We used the dictionaries and the distance matrices to extract the most expressed keywords and cluster the GLS.

6.2 Second Experiment

Into the second experiment, we sent each lesson, one at a time, to the pipeline to generate its GLS. For each GLS, we applied the SEMANTIC filter to extract the related MEFs to use them as probes for the subsequent insights.

We applied Jaccard and SSC metrics between each GLS and the PUBMED-MEF dataset.

We computed ten distance matrices from which we picked the first ten (see table 2) most similar results for each GLS. We were forced to used Apache Spark [37] to distribute this job across multiple slaves due to the heavy memory requirements of these tasks.

Table 2: Partial subset of suggested papers and their distance score.

Paper ID	Author	Value
P1	Belsky *et al.* [7]	0.91
P2	Huang *et al.* [22]	0.81
P3	Wu [41]	0.93
P4	Gaitanidis *et al.* [17]	0.83
P5	Bauer *et al.* [6]	0.88
P6	Wang *et al.* [40]	0.95
P7	Mays *et al.* [27]	0.94
P8	Schuck *et al.* [33]	0.93
P9	Guo *et al.* [20]	0.92
P10	Corsi *et al.* [13]	0.90

7 Experimental Results

The Table 1 shows the first four `MEF`s for each lesson. The SEMANTIC filter was able to detect keywords related to lesson context. Due to the GLS and MEF definitions, the MEF dictionaries contain up to 140K kmers for each GLS. We reported the most expressed that refer to complete word.

The results of topics clustering are available in Fig. 4 with the Jaccard index and Fig. 5 with the overlap coefficient.

We identified the topics which are part of the same branch of the tree with the same color and the subgroups of each topic using different colors. SEMANTIC separated the five treated topics well, with just little differences in lesson aggregation levels. Specifically, it is interesting to note how the system is able to cluster together the *Cancer* and *Diabetes* and *Chemistry* with *Evolution* topics.

In the second experiment, we used the extracted `MEF`s as if they represented ‘tags’ to recover suggested papers. For convenience, we have used only the `MEF`s of the lessons on the topics *Cancer* to show the results.

SEMANTIC identified 274 correlates documents recovered by `PUBMED-MEF` dataset. For example, in Table 2 we showed the first ten recovered papers with a score value greater than 0.80. In Fig. 6, we represented the position of suggested papers graphically compared to the target query, keeping in mind that the score of the distance from the target query is representative of the similarity between the set of `MEF`s of the *Cancer* topic and the individual retrieved papers.

8 Conclusions

The change to the basis of remote teaching is the transition from traditional education to smart education. The teacher is responsible for managing class to be student-centered, which involves greater responsibility and awareness of their limits and potential in self-learning behind one screen. It is not always possible. Disturbing factors related to the surrounding environment may disrupt the attention of the student. In an era in which new teaching tools are proposed, managing personal learning is also changed. One of the disadvantages of remote education is the different distractions that can intervene in maintaining the focus. In this work, we have proposed a tool that aims to maintain students’ attention on topics covered in the teacher’s lessons. The system works in real-time and allows us to generate further knowledge thanks to the possibility of an in-depth study by hyperlinks of treated topics. The preliminary experimentation suggests that the framework can provide insight and help regain attention on the lesson focus. Experimentation in real classes should be performed to understand the framework’s impact on student attention in remote lessons. Among the future directions’ objectives, we expect to explore techniques that improve text summarization and the extension of architecture for multilingual texts. An additional module for the real-time creation of concept maps may be provided.

References

[1] A. Al-Ajlan and H. Zedan. Why Moodle. In *Procs. of the 12th IEEE International Workshop on Future Trends of Distributed Computing Systems*, pages 58–64. IEEE, 2008.

[2] N. A. Alias and A. M. Zainuddin. Innovation for better teaching and learning: Adopting the learning management system. *Malaysian Online Journal of Instructional Technology*, 2(2):27–40, 2005.

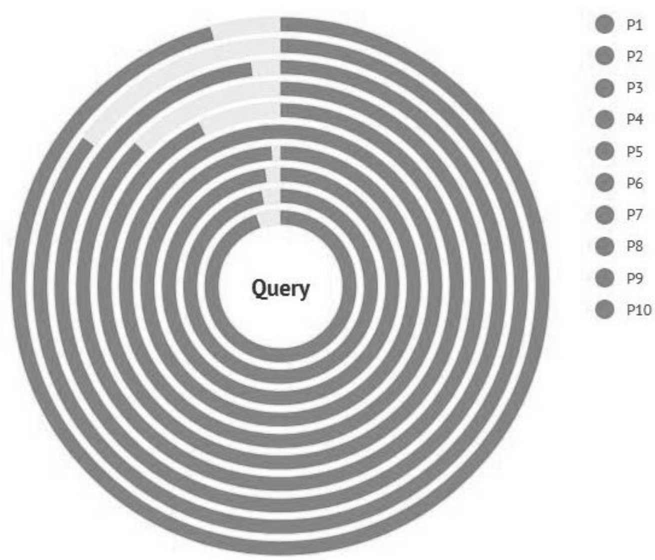


Figure 6: Position of suggested papers compared to the target query.

[3] A. Baddeley. Exploring the central executive. *The Quarterly Journal of Experimental Psychology - Section A*, 49(1):5–28, 1996.

[4] A. Baddeley. The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, 4(11):417–423, 2000.

[5] J. Baggaley. Mooc rampant. *Distance Education*, 34(3):368–378, 2013.

[6] M. Bauer, E. Morales-Orcajo, L. Klemm, R. Seydewitz, V. Fiebach, T. Siebert, and M. Böl. Biomechanical and microstructural characterisation of the porcine stomach wall: Location-and layer-dependent investigations. *Acta Biomaterialia*, 102:83–99, 2020.

[7] D. W. Belsky, T. E. Moffitt, T. B. Baker, A. K. Biddle, J. P. Evans, H. Harrington, R. Houts, M. Meier, K. Sugden, B. Williams, et al. Polygenic risk and the developmental progression to heavy, persistent smoking and nicotine dependence: Evidence from a 4-decade longitudinal study. *JAMA Psychiatry*, 70(5):534–542, 2013.

[8] J. Burstein. Opportunities for natural language processing research in education. In *Procs. of the International Conference on Intelligent Text Processing and Computational Linguistics*, pages 6–27. Springer, 2009.

[9] K. Canese and S. Weis. Pubmed: The bibliographic database. In *The NCBI Handbook [Internet]. 2nd edition*. National Center for Biotechnology Information (US), 2013.

[10] S. L. Chew and W. J. Cerbin. The cognitive challenges of effective teaching. *The Journal of Economic Education*, 52(1):17–40, 2021.

[11] C. C. Chiu, T. N. Sainath, Y. Wu, R. Prabhavalkar, P. Nguyen, Z. Chen, A. Kannan, R. J. Weiss, K. Rao, E. Gonina, N. Jaitly, B. Li, J. Chorowski, and M. Bacchiani. State-of-the-art speech recognition with sequence-to-sequence models. In *Procs. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4774–4778, 2018.

[12] R. Cobos, F. Jurado, and A. Blázquez-Herranz. A content analysis system that supports sentiment analysis for subjectivity and polarity detection in online courses. *IEEE Revista Iberoamericana de Tecnologías Del Aprendizaje*, 14(4):177–187, 2019.

[13] D. J. Corsi, C. K. Chow, S. A. Lear, S. Subramanian, K. K. Teo, and M. H. Boyle. Smoking in context: A multilevel analysis of 49,088 communities in Canada. *American Journal of Preventive Medicine*, 43(6):601–610, 2012.

[14] J. Cowan. The advantages and disadvantages of distance education. *Distance Education for Language Teachers: A UK Perspective*, pages 14–20, 1995.

[15] R. Francese, C. Gravino, M. Risi, G. Scanniello, and G. Tortora. Using project-based-learning in a mobile application development course-an experience report. *Journal of Visual Languages & Computing*, 31:196–205, 2015.

[16] K. Gaglo, B. M. Degboe, G. M. Kossingou, and S. Ouya. Proposal of conversational chatbots for educational remediation in the context of Covid-19. In *Procs. of the 23rd International Conference on Advanced Communication Technology (ICACT)*, pages 354–358. IEEE, 2021.

[17] A. Gaitanidis, N. Machairas, M. Alevizakos, C. Tsalikidis, A. Tsaroucha, and M. Pitiakoudis. Predictive nomograms for synchronous liver and lung metastasis in colon cancer. *Journal of Gastrointestinal Cancer*, pages 1–7, 2019.

[18] M. Gambhir and V. Gupta. Recent automatic text summarization techniques: A survey. *Artificial Intelligence Review*, 47(1):1–66, 2017.

[19] C. Gravino, M. Risi, G. Scanniello, G. Tortora, and R. Francese. Supporting mobile development project-based learning by software project and product measures. In *Procs. of the 22nd International Conference on Distributed Multimedia Systems (DMS)*, pages 41–47, 2016.

[20] Q. Guo, J. B. Unger, P. H. Palmer, C.-P. Chou, and C. A. Johnson. The role of cognitive attributions for smoking in subsequent smoking progression and regression among adolescents in China. *Addictive Behaviors*, 38(1):1493–1498, 2013.

[21] C. Guzmán-García, M. Gómez-Tome, P. Sánchez-González, I. Oropesa, and E. J. Gómez. Speech-based surgical phase recognition for non-intrusive surgical skills' assessment in educational contexts. *Sensors*, 21(4):1330, 2021.

[22] X. Huang, Y. Zou, L. Lian, X. Wu, X. He, X. He, X. Wu, Y. Huang, and P. Lan. Changes of T cells and cytokines TGF-β1 and IL-10 in mice during liver metastasis of colon carcinoma: Implications for liver anti-tumor immunity. *Journal of Gastrointestinal Surgery*, 17(7):1283–1291, 2013.

[23] M. Hubbard and M. J. Bailey. Mastering Microsoft Teams. *Mastering Microsoft Teams. https://doi. org/10.1007/978-1-4842-3670-3*, 2018.

[24] T. Káldi and A. Babarczy. Linguistic focus guides attention during the encoding and refreshing of working memory content. *Journal of Memory and Language*, 116:104187, 2021.

[25] J. M. Lodge and W. J. Harrison. Focus: Attention science: The role of attention in learning in the digital age. *The Yale Journal of Biology and Medicine*, 92(1):21, 2019.

[26] M. W. Marek, C. S. Chew, and W.-c. V. Wu. Teacher experiences in converting classes to distance learning in the Covid-19 pandemic. *International Journal of Distance Education Technologies (IJDET)*, 19(1):40–60, 2021.

[27] D. Mays, G. Luta, L. R. Walker, and K. P. Tercyak. Exposure to peers who smoke moderates the association between sports participation and cigarette smoking behavior among non-white adolescents. *Addictive Behaviors*, 37(10):1114–1121, 2012.

[28] J. Memon, M. Sami, R. A. Khan, and M. Uddin. Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR). *IEEE Access*, 8:142642–142668, 2020.

[29] C. M. Mowling and S. K. Sims. The metacognition journey: Strategies for teacher candidate exploration of self and student metacognition. *Strategies*, 34(2):13–23, 2021.

[30] T. Mufti, P. Mittal, and B. Gupta. A review on Amazon web service (AWS), Microsoft azure & Google cloud platform (GCP) services. *European Union Digital Library*, 2021.

[31] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu. Using of Jaccard coefficient for keywords similarity. In *Procs. of the International Multiconference of Engineers and Computer Scientists*, pages 380–384, 2013.

[32] A. J. Sanford, A. J. Sanford, J. Molle, and C. Emmott. Shallow processing and attention capture in written and spoken discourse. *Discourse Processes*, 42(2):109–130, 2006.

[33] K. Schuck, R. Otten, R. C. Engels, and M. Kleinjan. The role of environmental smoking in smoking-related cognitions and susceptibility to smoking in never-smoking 9–12 year-old children. *Addictive Behaviors*, 37(12):1400–1405, 2012.

[34] A. Shelton, C. J. Lemons, and J. Wexler. Supporting main idea identification and text summarization in middle school co-taught classes. *Intervention in School and Clinic*, 56(4):217–223, 2021.

[35] H. A. Simon. How big is a chunk?: By combining data from several experiments, a basic human memory unit can be identified and measured. *Science*, 183(4124):482–488, 1974.

[36] P. J. Smith, K. L. Murphy, and S. E. Mahoney. Towards identifying factors underlying readiness for online learning: An exploratory study. *Distance Education*, 24(1):57–67, 2003.

[37] A. Spark. Apache spark. *Retrieved January*, 17, 2018.

[38] A. Venditti, F. Fasano, M. Risi, and G. Tortora. The importance of interaction mechanisms in blended learning courses involving problem solving e-tivities. In *Procs. of the 13th International Conference on Digital Information Management (ICDIM)*, pages 124–129, 09 2018.

[39] M. Vijaymeena and K. Kavitha. A survey on similarity measures in text mining. *Machine Learning and Applications: An International Journal*, 3(2):19–28, 2016.

[40] M.-P. Wang, S.-Y. Ho, W.-S. Lo, and T.-H. Lam. Smoking family, secondhand smoke exposure at home, and nicotine addiction among adolescent smokers. *Addictive Behaviors*, 37(6):743–746, 2012.

[41] C.-Y. Wu. Initiatives for a healthy stomach. *Current Treatment Options in Gastroenterology*, 17(4):628–635, 2019.

Visual ECG Analysis in Real-world Scenarios

Bernardo Breve, Loredana Caruccio, Stefano Cirillo, Vincenzo Deufemia, Giuseppe Polese
Department of Computer Science
University of Salerno, Italy
{bbreve,lcaruccio,scirillo,deufemia,gpolese}@unisa.it

Abstract

Cardiac arrhythmia is an alteration of the heart rhythm, for which the heartbeat is irregular. Based on the severity of this condition, an arrhythmia could represent a serious danger for a patient. An ECG is a graphic representation of an heart rhythm, which provides an overview of heart's conditions over a specific time interval. ECG signal analysis is entrusted to trained clinicians, although complex and frantic environments, such as emergency settings, can make hard to delegate continuous monitoring to the medical personnel. In such scenarios, an automatic detection methodology could provide crucial support in promptly alerting clinicians towards a potential degeneration of a patient's conditions. To this end, we propose a heartbeat classification module capable of capturing the semantics of visual information of ECG signals provided by video frames. The module relies on feature extraction techniques derived from video projected images resulting in ECG data, which are then classified by means of deep-learning models. It can be used to support the early detection of some arrhythmia in critical contexts, such as emergency rooms. We show how the proposed module can be used to support clinicians in this context, and discuss an experimental evaluation performed over ground-truth datasets.

Index terms— ECG Analysis, Health scenarios, Arrhythmia, Multimedia

1 Introduction

Information and Communication Technology (ICT) as part of the health process is the inclusion of technology in managing and processing information for patients' direct or indirect benefit. In particular, it regards the use of digital devices and software products that allow users to create, store, exchange and use data, with the aim of providing a greater degree of efficiency in the management of health information and support professionals both in daily clinical activities in decision-making, such as those designed to identify possible therapies [11]. Particularly challenging are complex and frantic environments, such as emergency settings, where the translation of research evidence into clinical practice is particularly challenging.

Automatic monitoring systems can help professionals in scenarios that require continuous monitoring, and possibly alerting methods to catch the attention of clinicians when specific events occur [5, 8]. Such systems are becoming even more complex, and try to use different kinds of (smart) devices, which include modules typically focusing on specific tasks. Among these, Machine Learning (ML)-based modules can be used to classify monitored parameters and possibly to provide early detection of some pathologies.

Nevertheless, real-world contexts do not allow human pre-processing phases over raw data to be validated, and automatic support should also be devoted to simplifying the eliciting, processing, and visualizing of data [6, 10], trying to provide the best possible quality of data [7]. Thus, complete processes should be managed in real-world scenarios trying to not damage the accuracy that is fundamental in the health domain.

By focusing on the ECG Analysis task, in this paper we considered emergency rooms as basic scenario, where a patient can be monitored through several tools showing his/her vital signs (among which ECG signals can also be included), and several video recording devices are devoted to monitor patient's status and behavior [9]. Clinicians periodically check the patient's status by verifying his/her vital signs, and by interpreting audio/video ECG representations, and identifying the best evidence to guide clinical practice. At the same time, automatic support can be welcomed if it permits to help in solicitous monitoring and recording of clinical parameters and direct clinical inspection.

The proposal is focused on a new module that permits to classify heartbeats towards the early detection of arrhythmia by directly analyzing the visual representation of ECG signals. It aims to provide support in critical clinical sce-

DOI reference number: 10.18293/DMSVIVA2021-008

narios, where the continuous monitoring of patients cannot be accomplished constantly. In fact, by integrating such a module in a more complex monitoring system, it is possible to manage proper alarm triggering components to alert clinicians with the deterioration of ECG parameters related to the presence of some arrhythmia [22].

The module is able to analyze visual representation of ECGs, characterize proper features by considering the correspondent digital representation, and classify, through a ML model, the different types of signals according to the EC57 standard released by the Association for the Advancement of Medical Instrumentation (AAMI) [1]. The module is scalable towards different feature extraction techniques, and different ML models, and is able to provide labels on video frames in accordance with classification results.

The remainder of the paper is organized as follows: Section 2 presents the related work illustrating the main approaches of ECG analysis tasks provided in the literature. In Section 3, we introduce the proposed module by also describing the integrated models and techniques. Experimental results are discussed in Section 4. Finally, Section 5 provides the conclusion and future developments.

2 Related Work

Several studies have applied deep learning methodologies in ECG signal analysis. In [13], the authors provided a systematic literature review highlighting the main opportunities of ECG data analysis through shallow and deep-learning techniques, classifying the proposals based on the tasks they could perform:

- *Disease detection*. This task aims to detect anomalies in a patient's heartbeats, solutions for detecting arrhythmia, myocardial infarction, atrial fibrillation, and congestive heart failure, could represent important support to doctors for speeding up the diagnosis formulation;

- *QRS annotation*. The application of shallow and deep-learning techniques in the ECG analysis could be significant in annotating QRS waves that identify a single beat. Indeed, for factors depending both on the patient and the medical device, the conformation of the signal could present waves of different morphology, which could sometimes appear difficult to classify in a short amount of time;

- *Analysis of sleep patterns*. This application scenario could benefit from the application of ECG data analysis. In particular, the presence of sleeping syndromes such as obstructive sleep apnea could be classified through monitoring the ECG signals during sleep;

- *Biometric authentication*. With the advent of personal wearable devices capable of monitoring heartbeats and, for some recent devices, even provide an ECG analysis, some studies focused on trying to exploit this data as a biometric authentication means to improve the security of smart devices;

- *Denoising ECG signals*. Machine learning techniques for ECG data also served an important contribution in the steps of denoising signals as a result of an acquisition process.

In [4], authors presented a review of literature on heartbeat abnormality detection. In particular, Annam *et al.* propose a taxonomy of solutions of the arrhythmia distinguished in two main approaches: feature-based classification and timeseries-based classification. While with timeseries approaches the ECG are entirely and directly processed, in the feature-based classification approach, data are analyzed and decomposed to extract main features, which will then be employed for classifying ECG signals. For each major category, the authors also provided a further splitting between solutions that apply intra-patient or inter-patient categorizations. In the former, data concerning a single patient is partitioned between training and test sets. On the contrary, by applying an inter-patient paradigm, data from different patients are used to assess a strong generalization capability for unknown patient's data.

In [16], Kachuee *et al.* proposed a solution based on deep convolutional neural network (CNN) for the classification of the ECG heartbeat in different categories. Through a transferable representation, they designed a model for detecting arrhythmia. The authors then proved that this representation appears to be applicable in the prediction of myocardial infarction. The CNN defined obtained results comparable to the state-of-the-art in the classification of ECG signals, while it achieved better results in myocardial infarction classification than most of other proposed solutions.

In [25], the authors proposed a combination of wavelet transform and autoregressive models as a feature extraction method. Then, through the usage of SVM to classify the ECG heartbeat. The Wavelet transform is used for extracting the coefficients of the transform as characteristics of each ECG segment. The Autoregressive Model was also used for extracting coefficients, which combined with the wavelet coefficients, forming a vector of 32 dimensions for classification. The evaluation is then performed through an empirical study based on the MIT-BIH Arrhythmia Database.

In [23], a hierarchical ML method derived from the gradient boosting decision tree has been proposed. It extracts a large number of features from 6 categories (statistics, intervals, etc.) from pre-processed heartbeats via the Wavelet transform. The authors then used recursive feature elimination (RFE) to select the most relevant ones. The method has been applied to an inter-patient experiment conforming AAMI standard alongside several contrast experiments.

Instead, the method proposed in [3] by Alarsan *et al.* is based on Gradient Boosted Trees, Decision Tree and Random Forest, and uses the discrete Wavelet transform for features extraction. It has been evaluated and validated on the MIT-BIH arrhythmia dataset.

Finally, in [26] Zhou *et al.* propose an architecture composed of ResNet + Bi-LSTM, which combines the benefits of both to extract relevant characteristics and to carry out the classification of ECG signals.

All the works mentioned above are limited to the classification of individual ECG signals extracted from images. Differently to them, we propose a new ML-based module to directly classify in real-time multimedia representations of ECG signals, yielding to the early detection of arrhythmia.

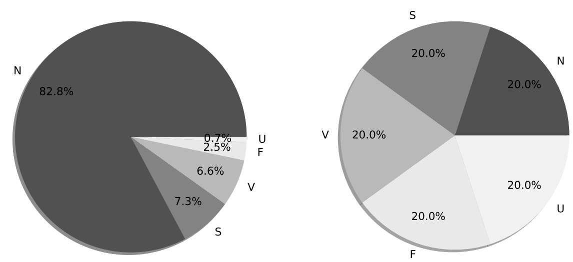
3 Monitoring ECG Heartbeats

The early detection of Arrhythmia or other pathologies through ML models is mainly accomplished by directly considering numerical ECG data. Nevertheless, in order to get classification processes applicable in different real-world scenarios, it is necessary to consider different types of raw data including images and/or video sequences. In what follows, we describe the proposed module by introducing different pre-processing steps based on the type of input data (Section 3.1), the CNN architectures we considered to perform some classification tasks (Section 3.2), and the real-world scenarios in which the proposed module can be applied (Section 3.3).

3.1 Preprocessing

The imbalance problem is one of the main issues that should be addressed in ML tasks during pre-processing steps. It entails adjusting class imbalance possibly characterizing the prior data distribution over classes. Sampling methods include undersampling andoversampling methods [2], yielding the possibility to remove samples, or add new samples even by creating copies of existing ones.

Figure 1 shows the application of resampling methods on the MIT-BIH dataset [20, 21], which represents a sample of the MIT-BIH arrhythmia database, including five classes of ECG signals (N, S, V, F, and U) in accordance with the EC57 AAMI standard [1]. In particular, an undersampling technique has been applied in order to randomly reduce the number of occurrences of signals labeled as "N". Moreover, oversampling techniques have been used to randomly increase the samples mapping the other classes of ECG signals (i.e., S, V, F, and U). Thus, the right plot in Figure 1 shows the balanced MIT-BIH dataset after the performed resampling process.



Before resampling. After resampling.

Figure 1: Distribution of classes for MIT-BIH dataset.

When the ECG data to analyze are images or video frames further pre-processing steps should be applied, with the aim to remove potential noise over multimedia data, by cleaning, resizing and recoloring images. To this end, we considered two main techniques based on Red Green Blue (RGB) and Hue Saturation Value (HSV) models, respectively. They allowed us to emphasize the ECG signal with respect to the image background.

The next step concerns the extraction of the feature vector, by firstly detecting peaks over the signal [15], then reconstructing QRS waves around the peaks, and then summarizing and normalizing them in the range $(0, 1)$.

3.2 Convolutional Neural Networks

CNNs represent one of the most popular classes of artificial neural networks capable of classifying images and/or video sources. This type of neural network has been first proposed in [18] and since then, it has achieved considerable success in several domains, and particularly in the field of health technology [14]. The architecture of a CNN is composed of one or more convolutional layers, and a fully connected layer at the top, containing weights and pooling layers.

In this paper, we adopt three different CNNs to identify arrhythmic diseases from real ECGs, composed of 14, 13, and 31 layers, respectively. Table 1 shows the architecture details of the proposed CNN. In particular, the Input Layer receives an image of dimensions 186x1, where 186 represents the features obtained after the preprocessing steps on the image. Then, the first block contains a Convolution1D characterized by 64 kernels larger than 6. It is important to notice that, this value is reduced in the following layers allowing CNN to further synthesize the images. In fact, due to the several convolutional layers, it is necessary to reduce the size of the kernels in order to avoid losing too much information. Next, a Batch Normalization Layer allows the CNN to normalize and scale the analyzed data. At the end of each

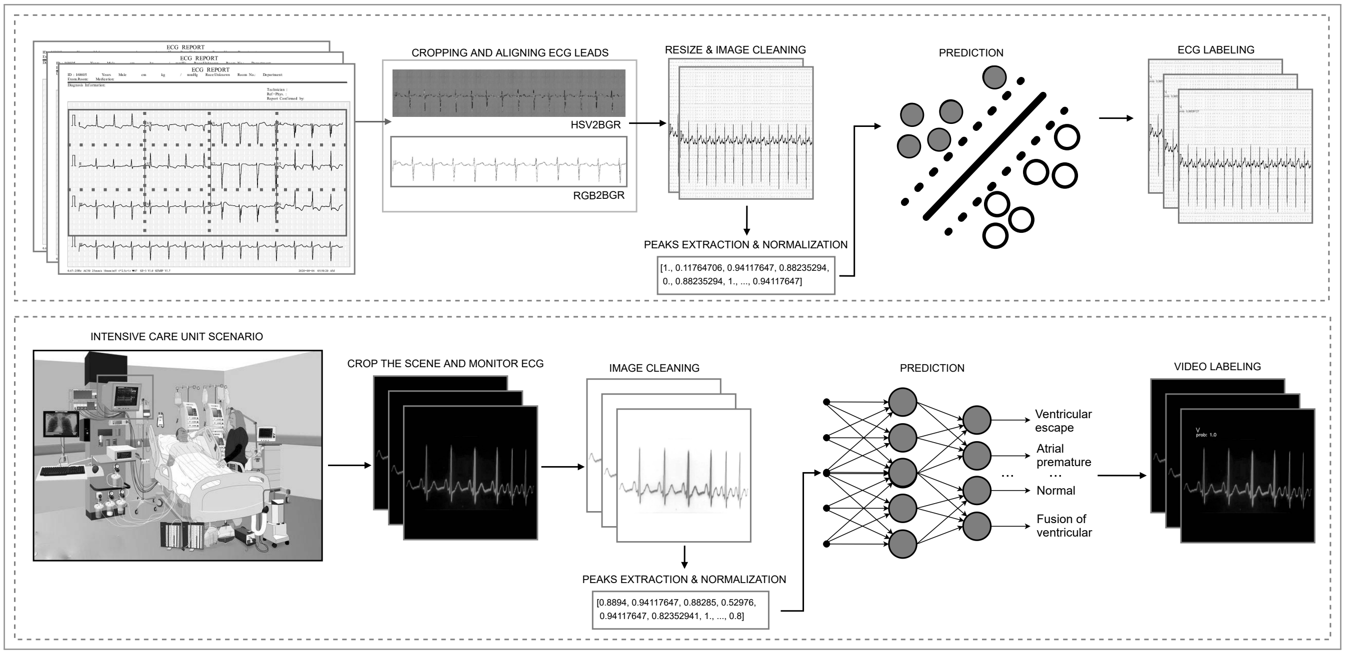


Figure 2: Classification of real ECG report and video frames.

Layers	Type	Output Shape	Kernel Size
0	Input Layer - ECG	186 x 1	-
1	Convolution1D	181 x 64	6
2	BatchNormalization	181 x 64	-
3	Maxpooling1D	91 x 64	3
4	Convolution1D	89 x 64	3
5	BatchNormalization	89 x 64	-
6	Maxpooling1D	45 x 64	3
7	Convolution1D	43 x 64	3
8	BatchNormalization	43 x 64	-
9	Maxpooling1D	22 x 64	3
10	Flatten	1408	-
11	Dense1	64	-
12	Dense2	32	-
13	Output Layer	5	-

Table 1: Architecture of the first CNN.

Layers	Type	Output Shape	Kernel Size
0	Input Layer - ECG	186 x 32	-
1	Convolution1D	185 x 32	3
2	Convolution1D	183 x 32	3
3	Maxpooling1D	91 x 64	3
4	Convolution1D	83 x 64	5
5	Convolution1D	79 x 64	5
6	Maxpooling1D	39 x 64	3
7	Convolution1D	31 x 128	5
8	Convolution1D	27 x 128	5
9	Maxpooling1D	13 x 128	3
10	Flatten	1664	-
11	Dense	64	-
12	Output Layer	5	-

Table 2: Architecture of the second CNN.

block, it has been introduced a Pooling Layer that exploits the Max-Pooling algorithm to perform a down-sampling operation to reduce the feature size.

As we can see, in this first CNN, we consider three similar blocks that differ from the others only in the size of the kernels. The Flatten Layer at the end of these blocks allows CNN to prepare the data to classify signals with a Neural Network (NN). In fact, two Fully Connected layhave been included in the architecture, with 64 and 32 neurons, respectively. Finally, the Output Layer has been composed of 5 neurons, where 5 represents the number of different arrhythmic diseases to be classified.

Table 2 details the architecture of the second CNN. Similar to the first model, this has been designed by considering an Input Layer and three similar blocks. However, this CNN considers two different convolutional layers in each block and a Pooling Layer that exploits the algorithm Max-Pooling. Unlike the first model, this CNN uses a dilated convolution with $k = 2$ that allows increasing the receptive vision (global vision) of the network and linear parameters. In particular, given a 2D image as input, setting the dilation rate $k = 1$ represents a normal convolution, $k = 2$ means skipping one pixel per input, and $k = 4$ means skipping 3 pixels. The resulting data from these layers are processed by a Fully Connected Layer with 64 neurons and an Output layer of 5 neurons.

The latest model has been introduced in [16], and represents one of the most recent proposals existing in the literature. In fact, experimental evaluations have shown that this model considerably improves the performance of many existing models.

In the next section, an experimental evaluation of CNNs will be made by applying a cross-validation technique to estimate the skill of each model. The evaluation of these three models allowed us to directly compare the performance of the proposed CNN with some of the most accurate models in the literature.

3.3 Applications in real-word scenarios

To highlight the effectiveness of the proposed module, we identified two real-world scenarios in which it could play a crucial role. In fact, as mentioned before, the frenetic pace at which the emergency rooms work does not always allow the specialized personnel to continuously monitor patients' ECGs. With the adoption of the proposed module, the healthcare personnel could be relieved of this task, being able to be called into question only if some abnormalities are detected by the model.

Figure 2 introduces the process of ECG signal classification in the aforementioned scenarios. They differ on the source from where an ECG signal is taken from. Indeed, data could be acquired both from an ECG report, printed out from an ECG machinery, and directly extrapolated from video frames, which could be recorded by video surveillance cameras.

3.3.1 Classification of ECG signal from report

We consider a scenario in which many patients have been performed an ECG, and the personnel has to end-up with many reports to analyze and process. To this end, in order to speed up the analysis process, the reports can be provided sequentially to the proposed module for the analysis.

The structure of an ECG report is composed of 12 leads, each of which represents the status of the heart read from different sensors positioned on the patient's body. For having a proper classification of the ECG, each lead needs to be taken into account, since they provide a different point of view of the heart conditions. Hence, the first step for this analysis is to isolate the 12 different leads that the ECG is composed of. After that, the leads follow the pre-processing, detailing, and cleaning activities detailed in Section 3.1.

The feature vector describing the QRS waves is then normalized and provided to the classifier. The output label of the classification process is added to the ECG report itself through post-processing image editing. The labeled ECG report is then printed for the personnel to evaluate the response.

Dataset	Data type	# Samples	Heartbeats classes
MIT-BIH	ECG data	87554	EC57 AAMI standard
PubMed	ECG images	1937	Binary

Table 3: Statistics on the considered datasets.

Dataset	Model	Times	
		Preprocessing (s)	Training (s)
MIT-BIH	Model (1)	0.93	1873.3
	Model (2)		1937.6
	Model (3)		2221.2
PubMed	SVM	1.2	1.4

Table 4: Time performances on the considered datasets.

3.3.2 Classification of ECG signal from video frames

The process of classifying and ECG signal from a video requires a different pre-processing approach. Indeed, in this scenario data are retrieved from video surveillance devices, which most of the time are not directly placed in front of the medical devices in charge of monitoring the patient. To extract ECG data the frame needs to be cropped, isolated, and occasionally transformed in order to flatten the detail for the ECG signal. The obtained single-lead ECG sequences are then processed as done in the previous approach.

Finally, after the prediction of the signal over a single frame, it is labeled and put on hold, waiting for more frames to be processed. After a certain amount of frames is processed, they are reassembled, producing a video sequence identical to the one in input, but with the label classifying the information provided by the signal.

4 Experimental evaluation

To evaluate the effectiveness of the proposed module and the considered CNNs, we performed two different experimental sessions. The first is mainly devoted to evaluate the accuracy of CNNs, whereas in the second session, we considered a novel ground-truth dataset, which allowed us to directly consider ECG images as input for both training and validation tasks.

4.1 Experimental settings

As mentioned above, we considered 2 different ground-truth datasets: MIT-BIH [20, 21] and PubMed [17], whose statistics are reported in Table 3.

MIT-BIH represents a sample of the MIT-BIH arrhythmia database, including five classes of ECG signals (N, S, V,

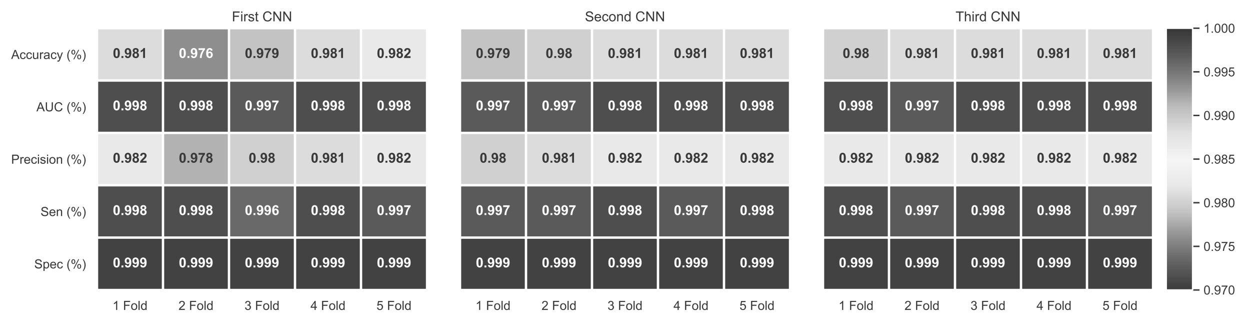


Figure 3: Cross-validation performance for MIT-BIH dataset.

F, and U) in accordance with the EC57 AAMI standard [1]. The first class represents samples with no anomaly detected, S and V classes categorize different types of heartbeat anomalies, class F includes ECGs with combined features from S and V classes, whereas U represents the class of unrecognized heartbeats. The MIT-BIH Arrhythmia Database is one of the most used datasets for evaluating automatic ECG Analysis techniques, and it contains ECG samples obtained from 47 subjects, and recorded in a 30 min duration and sampled at 360Hz. We used such dataset for evaluating the effectiveness of the two considered CNNs.

PubMed represents a novel dataset containing ECG images of Cardiac and COVID-19 patients collected in Cardiac Care and Isolation Units of different health care institutes across Pakistan. Image have been classified in five distinct categories (COVID-19, Abnormal Heartbeat, Myocardial Infarction (MI), Previous History of MI, and Normal Person). In our evaluation, we considered a sample of 1937 images including Abnormal Heartbeat and Normal Person images, yielding a binary classification task. We used such dataset for evaluating the effectiveness of the whole proposed methodology.

The time performances for the preprocessing steps are shown in Table 4. As we can see, the average preprocessing time for each image is less than 1 second for the MIT-BIH dataset, and 1.2 seconds for the PubMed dataset. In the latter case, the preprocessing step takes longer due to the identification of each lead, as described in Section 3.3.

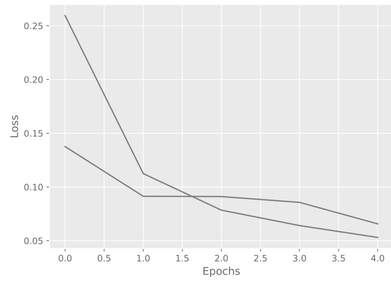
Moreover, the performance of the classification have been evaluated in each task by using several evaluation metrics. In particular, let TP, TN, FP, and FN be the values of True Positive, True Negative, False Positive, and False Negative, respectively, then we can define the evaluation metrics as follow:

- **Area Under the Curve** (AUC) represents the classifier's ability to distinguish classes;
- **Precision** (Prec) represents the number of positives that are correct, over all identified positives: $\mathrm{Prec} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$
- **Sensitivity** (Sen), also named **Recall** (Rec), represents the proportion of positives that are correctly identified: $\mathrm{Sen} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$
- **Specificity** (Spec) represents the proportion of negatives that are correctly identified: $\mathrm{Spec} = \frac{\mathrm{TN}}{\mathrm{TN+FP}}$
- **Accuracy** (Acc) represents the number of positives and negatives correctly identified over the over the total number of tested elements: $\mathrm{Acc} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}}$
- **F1-score** (F1) represents the armonic mean of precision and recall: $\mathrm{F1} = 2 \times \frac{\mathrm{Prec} \times \mathrm{Rec}}{\mathrm{Prec+Rec}}$

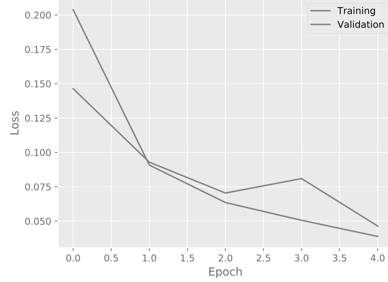
4.2 Evaluating CNN performances

To evaluate the accuracy and the effectiveness of predictive models, it is necessary to perform one or more evaluations of the errors obtained from the models. Generally, after the training phase, the model error is calculated in order to provide a numerical estimate considering the difference between the expected and original responses, called *training error*. However, this type of evaluation only provides a general idea of how the model works on the considered data. To this end, it is necessary to adopt cross-validation techniques that allow to accurately evaluate the performance of the models. These techniques do not use the entire dataset for the training phase. In fact, only a portion of data are used for training the model. The rest of data are used to evaluate the performance of the model.

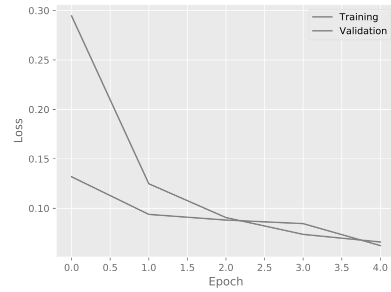
Among the most reliable cross-validation techniques, K-fold [24] is one of the most popular which provides for randomly dividing data into k groups of approximately equal size, called *folds*. The first fold is kept for testing, while the model is trained on the $k-1$ folds. Thus, this validation process is repeated k times considering different folds each time. Figure 3 shows the results after the K-fold validation of each considered CNN model over the 80% of data for training step, and the remaining 20% for testing with 5 different folds (i.e., $k = 5$).



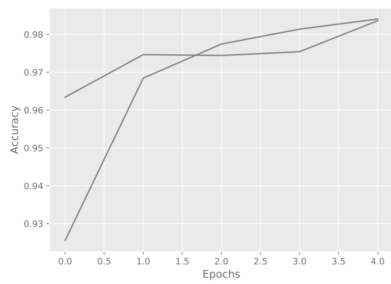
(a) Loss of the model (1).



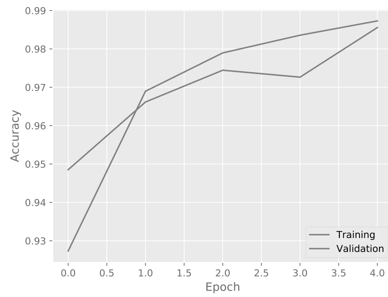
(b) Loss of the model (2).



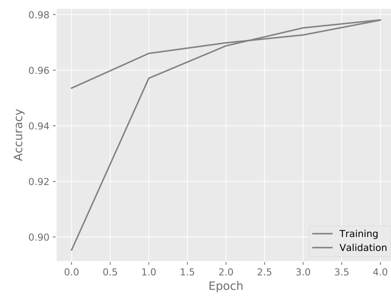
(c) Loss of the model (3).



(d) Accuracy of the model (1).



(e) Accuracy of the model (2).



(f) Accuracy of the model (3).

Figure 4: Learning curves of the considered CNN models.

Results over aforementioned metrics are detailed in Figure 3, which shows an accuracy value that reaches its maximum in the last fold for the first CNN (98.2%), whereas this value stabilizes already from the first fold for the other CNNs, which have an accuracy value of 98.1%. Moreover, the overall average values of the considered CNNs are 99.7%, 99.9%, 99.7%, and 98.2% for Sen, Spec, AUC, and Precision, respectively.

The learning curves, namely loss and accuracy, for all considered CNN models are shown in Figure 4. In general, learning curves plot performances on the training validation sets as a function of the training epochs, in terms of loss or accuracy. They can also identify under/over-fitting problems. By comparing plots of the considered models, they show very similar trend for the training line, whereas little differences in the trend monotonicity are shown for validation line. Nevertheless, none of the considered models seem to present the aforementioned problems.

Finally, classification results in terms of precision, recall, and F1-measure obtained over single classes or in average, are shown in Figure 5. We can notice that the second model obtains discordant results across different classes; whereas the other two models present less variability. In general, the fist model can be considered the best performing one.

4.3 Classification of ECG images

In the following section, we will discuss how we evaluated the proposed module according to the scenario presented in Section 3.3.1.

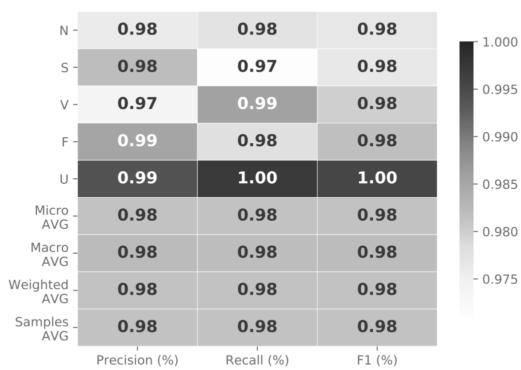
In details, we evaluated the module on the PubMed dataset described above. Starting from a total of 517 12-leads ECG reports divided between 284 ECGs showing normal heart activity, and 233 showing abnormal heartbeat conditions, we applied some pre-processing steps in order to extrapolate information about each ECG. In particular, we first split the ECG isolating the 12-leads it was composed of, and for each lead we extracted peaks with a vector of 186 numerical values. The leads have then been chained forming an array of 2232 integers. After labeling each array according to the type of ECG they were describing, we trained a Support Vector Machine (SVM) classifier, whose application in these scenarios has already been experimented [12].

Figure 6 reports the classification results obtained over a 5-fold cross validation process. The two datasets, representing normal and abnormal heartbeats, have been randomly shuffled in training and testing datasets, respectively, containing the 80% and 20% of all available reports. By repeating the same evaluation 5 times varying, as said, in a random way the ECGs applied in the training and testing phases.

This evaluation proved that the preprocessing approach

(a) Classification metrics for the first model.

(b) Classification metrics for the second model.



(c) Classification metrics for the third model.

Figure 5: Classification metrics of CNN models on MIT-BIH dataset.

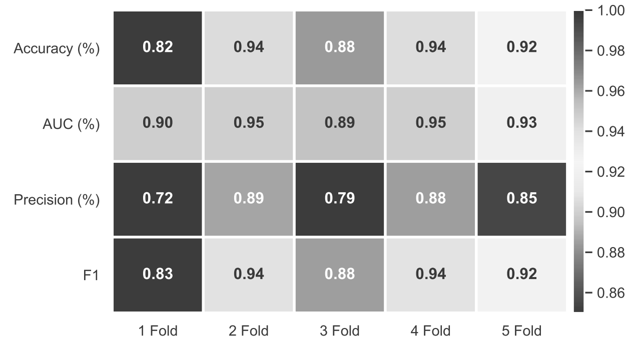


Figure 6: Cross-validation results for PubMed dataset.

has obtained an average accuracy value of 90%. The average values for AUC and Precision is 92.4% and 82.6%, respectively. Finally, the average F1-measure value emerged during the evaluation has been 0.92.

5 Conclusions

The automatic detection of arrhythmia from visual ECG data represents a particularly challenging task. Nevertheless, real-world scenarios require some support especially to empower the continuous monitoring of patients in frantic settings, such as emergency rooms. The proposed classification module tries to provide some findings in this respect. Although these results can be improved in the future, to the best of our knowledge, they represent the first-ever proposal applying these techniques to real-world scenarios. Modules proposed in literature perform their evaluation on numerical datasets, whereas data is collected and presented directly in their vector representation. Having to deal with images or video frames requires applying more accurate techniques to extract features since issues such as image noise or image distortion are quite common in this scenario. Considering that, the obtained results provide a good starting point for applying these procedures in real-world scenarios.

Future directions should consider such challenging scenarios, by providing further real-world datasets for heartbeat-related pathologies, by also considering data from several IoT sensors [19]. This could entail the analysis of the applicability of such kind of support in real-world settings. Further studies should also be considered in less expressive or distorted ECG visual representations.

References

[1] AAMI. AAMI EC57- Testing and reporting performance results of cardiac rhythm and st segment measurement algorithms, 2012.

[2] S. M. Abd Elrahman and A. Abraham. A review of class imbalance problem. *Journal of Network and Innovative Computing*, 1(2013):332–340, 2013.

[3] F. I. Alarsan and M. Younes. Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. *Journal of Big Data*, 6(1):1–15, 2019.

[4] J. R. Annam, S. Kalyanapu, S. Ch, J. Somala, and S. B. Raju. Classification of ECG heartbeat arrhythmia: A review. *Procedia Computer Science*, 171:679–688, 2020.

[5] M. Blount, M. R. Ebling, J. M. Eklund, A. G. James, C. McGregor, N. Percival, K. Smith, and D. Sow. Real-time analysis for intensive care: development and deployment of the

artemis analytic system. *IEEE Engineering in Medicine and Biology Magazine*, 29(2):110–118, 2010.
[6] B. Breve, L. Caruccio, S. Cirillo, V. Deufemia, and G. Polese. Visualizing dependencies during incremental discovery processes. In *EDBT/ICDT Workshops*, 2020.
[7] L. Caruccio and S. Cirillo. Incremental discovery of imprecise functional dependencies. *Journal of Data and Information Quality (JDIQ)*, 12(4):1–25, 2020.
[8] L. Caruccio, O. Piazza, G. Polese, and G. Tortora. Secure iot analytics for fast deterioration detection in emergency rooms. *IEEE Access*, 8:215343–215354, 2020.
[9] L. Caruccio, G. Polese, G. Tortora, and D. Iannone. Edcar: A knowledge representation framework to enhance automatic video surveillance. *Expert Systems with Applications*, 131:190–207, 2019.
[10] G. Costagliola, V. Fuccella, M. Giordano, and G. Polese. Monitoring online tests through data visualization. *IEEE Transactions on Knowledge and Data Engineering*, 21(6):773–784, 2008.
[11] G. Fitzpatrick and G. Ellingsen. A review of 25 years of CSCW research in healthcare: contributions, challenges and future agendas. *Computer Supported Cooperative Work (CSCW)*, 22(4-6):609–665, 2013.
[12] C. L. Herry, M. Frasch, A. J. Seely, and H.-t. Wu. Heart beat classification from single-lead ecg using the synchrosqueezing transform. *Physiological measurement*, 38(2):171, 2017.
[13] S. Hong, Y. Zhou, J. Shang, C. Xiao, and J. Sun. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in Biology and Medicine*, page 103801, 2020.
[14] W. N. Ismail, M. M. Hassan, H. A. Alsalamah, and G. Fortino. CNN-based health model for regular health factors analysis in internet-of-medical things environment. *IEEE Access*, 8:52541–52549, 2020.
[15] E. Jones, T. Oliphant, P. Peterson, et al. Scipy: Open source scientific tools for python. 2001.
[16] M. Kachuee, S. Fazeli, and M. Sarrafzadeh. ECG heartbeat classification: A deep transferable representation. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 443–444. IEEE, 2018.
[17] A. H. Khan, M. Hussain, and M. K. Malik. ECG images dataset of cardiac and COVID-19 patients. *Data in Brief*, 34:106762, 2021.
[18] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.
[19] P. Lo Giudice, A. Nocera, D. Ursino, and L. Virgili. Building topic-driven virtual iots in a multiple iots scenario. *Sensors*, 19(13):2956, 2019.
[20] R. Mark, P. Schluter, G. Moody, P. Devlin, and D. Chernoff. An annotated ECG database for evaluating arrhythmia detectors. In *IEEE Transactions on Biomedical Engineering*, volume 29, pages 600–600, 1982.
[21] G. B. Moody and R. G. Mark. The mit-bih arrhythmia database on cd-rom and software for use with it. In *[1990] Proceedings Computers in Cardiology*, pages 185–188. IEEE, 1990.
[22] E. Özcan, D. Birdja, L. Simonse, and A. Struijs. Alarm in the ICU! Envisioning patient monitoring and alarm management in future intensive care units. *Service Design and Service Thinking in Healthcare and Hospital Management*, pages 421–446, 2019.
[23] H. Shi, H. Wang, Y. Huang, L. Zhao, C. Qin, and C. Liu. A hierarchical method based on weighted extreme gradient boosting in ECG heartbeat classification. *Computer methods and programs in biomedicine*, 171:1–10, 2019.
[24] T.-T. Wong and P.-Y. Yeh. Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, 32(8):1586–1594, 2019.
[25] Q. Zhao and L. Zhang. ECG feature extraction and classification using wavelet transform and support vector machines. In *2005 International Conference on Neural Networks and Brain*, volume 2, pages 1089–1092. IEEE, 2005.
[26] Y. Zhou, H. Zhang, Y. Li, and G. Ning. ECG heartbeat classification based on ResNet and bi-LSTM. In *IOP Conference Series: Earth and Environmental Science*, volume 428, page 012014. IOP Publishing, 2020.

YouCare: a COVID-19 Telehealth App

Gennaro Costagliola, Mattia De Rosa, Vittorio Fuccella, Francesco Vitale
Dipartimento di Informatica, University of Salerno
Via Giovanni Paolo II, 84084 Fisciano (SA), Italy
{gencos, matderosa, vfuccella}@unisa.it

Abstract

The COVID-19 pandemic has caused disruption across the globe and put pressure on healthcare systems. In order to limit the use of hospital resources, the use of home care and telehealth has been very important to minimize direct human intervention in monitoring patients. The purpose of this work is to present YouCare: a multiplatform application that allows the collection of medical data on the health status of the user in order to allow physicians to efficiently monitor the status of the patient. As an important feature, it includes functions to monitor the general situation through statistics and interactions with the other users of the application. This might make the isolation period less stressful while exchanging current COVID experiences. The use of the application has been experimented with a usability test, obtaining positive feedback from the users. We also report other similar applications that have been developed and used in different parts of the world.

Keywords: *COVID-19, telehealth, app.*

1. Introduction

The COVID-19 pandemic, which spread in Wuhan (Hubei, China) in the early 2020s, caused a major change worldwide, strongly influenced people's lifestyle, and put a lot of pressure on national health care systems, in some cases leading them to collapse (with shortages of both hospital beds, doctors and medical supplies such as masks and oxygen [29]). In a substantial part of the world, therefore, great efforts have been made to improve the situation of healthcare systems, on the one hand, by expanding the capacity of hospital systems, although necessarily limited, and on the other, by expanding the use of home care and telehealth. In the most critical cases, the lack of healthcare personnel (or even just phone contact with a physician) has also negatively affected home care, so systems that can minimize the need for direct human intervention in monitoring patients at home can prove very valuable. However physicians' positive communication skills have a significant psychological effect on COVID-19 patients [2], so limiting direct interaction between patient and health professional may be negative from that perspective. Moreover, for such patients, in addition to the disease, typically a quarantine at home is imposed as well. Obviously, such a state of isolation and the consequent lack/reduction of social interactions and information about the disease can negatively affect mood.

The purpose of this work is to present YouCare: a multiplatform application (Android / iOS / Web, in order to ensure the widest possible availability), to allow the collection of medical data about the health status of home patients in order to efficiently allow physicians constant monitoring of the patient's status (also providing alerts in the case of critical values). The application also provides some features intended to make the isolation period less stressful.

In particular, the user can fill out a questionnaire in which he/she reports his/her clinical data (body temperature, sore throat, headache, muscle pain, nausea, cough, shortness of breath, bad mood, oxygen saturation level, breathing rate, heart rate, and blood pressure). Moreover the user can get in direct contact with other users through a Forum section, in this way other users (often in similar situations) can offer practical help and support through reply messages. In addition, the user can check statistics about the use of the app and aggregate information about other users' daily questionnaires (so he/she can monitor the condition of the majority of users, to have an idea of the global situation). The user may request a phone contact as well. The person's emotional state may also benefit from the opportunity to interact with health professionals and other people who are affected by COVID-19 and the ability to track overall statistics. The app also allows the user to "manage" another person, e.g. an elderly relative who does not have or is unable to use a smartphone or pc.

The paper is organized as follows: Section 2 describes previous work on COVID-19 apps; Section 3 describes the

DOI reference number: 10.18293/DMSVIVA2021-009

YouCare application, Section 4 shows its experimental evaluation, and Section 5 the evaluation results. Finally, Section 6 concludes the paper with a discussion on future work.

2. Related Work

Since the onset of the COVID-19 pandemic, numerous technologies and mobile applications have been proposed. Great interest has been devoted to contact tracing apps, with the study of different technologies and architectures, each with different privacy implications [1]. The use of this kind of technology has not been limited to contact tracing, but uses for telemedicine and remote diagnosis have also been explored, since in recent years, devices such as smartphones, smartwatches, and smart bands have seen an increase in usage in the medical and assistive technology fields [18, 10, 30, 4]. Some contact tracing apps have also included such features [3]. In this section, we will briefly focus on applications that offer these functionalities.

Among COVID-19 contact tracing apps, some include the ability for the user to report their health status (self-assessment). As an example, COVID Tracker is an app used by the Irish Health Service Executive for contact tracing on a national basis [13]. COVID Tracker monitors contacts between nearby devices via Bluetooth. Upon detection of contact with a positive, it alerts affected users via anonymous notifications. If the user expresses consent, instead of notification, they can receive a direct phone call from a health care provider. Daily, the user can fill in a short questionnaire in which he/she communicates his/her health situation based on a few parameters (fever, breathing problems, cough, taste, and smell problems). This application also has an informative side, in fact, on the homepage the citizen can read statistics on the progress of the infection in Ireland. Among the available statistics, there are: number of installations of the application, number of tests performed, number of symptomatic and asymptomatic infected people. Other examples of applications that include the facility for self-assessment used in different parts of the world include Mawid (Saudi Arabia) [5], Aarogya Setu (India) [17], NHS COVID-19 (UK) [22], PathCheck SafePlaces (USA) [23]. Another case is the Health Code system for contact tracing that has been integrated in China into two of the most popular apps used by the population, WeChat and Alipay [21]. According to the health code rules peoples are required to enter their personal information, including medical and health status information, into these apps. Based on contact tracing and entered symptom information the app assigns a health risk status to the person, which regulates his or her movement options and the possible need for medical/health intervention.

Other applications, on the other hand, are designed explicitly to allow the user to report their health status, both in the case of people without a COVID-19 diagnosis (preventive monitoring) and people with a COVID-19 diagnosis (home care). As an example, the Cvm-Health [26] web app, distributed by Sensyne Health in the UK and US, is a COVID-19 monitoring app. Users can sign up for the platform and daily record their vital signs and any COVID-19-related symptoms. The app creates a personalized digital medical record based on the recorded vital sign data and symptom information. This app also offers a social aspect: it allows users to help family and friends who are digitally disconnected by monitoring their health. The platform interacts directly with the NHS (UK National Health System) by sharing the data entered by users. Another example is the e-Covid SINFONIA app [27], used in Campania (Italy), which notifies the user of the outcome of COVID-19 tests performed in regional laboratories, and also allows the user to submit a questionnaire with the main infection indicators. These data are then made available to the general practitioners. Within the application, it is possible to register family members in order to view or receive notifications of their test results. Other examples include [8] in which teleconsultation is performed through a mobile application, tablet, or web browser, and [19] in which the approach of the Cleveland Clinic incorporates a self-monitoring app for patient engagement, monitors symptoms for early intervention.

Some applications, in addition to self-reporting capabilities, make use of wearable devices to detect the person's vital signs [7]. Examples include [15], where wearable devices were used to collect vital signs of people in quarantine, e.g., before or after a trip abroad. This data is monitored by algorithms and a medical team so that signs of deteriorating health can be detected and people can be transferred to a hospital if necessary. In [14], instead, wearable devices were used to monitor patients with chronic illnesses or who are recovering from a COVID-19, so that some patients who would have to be hospitalized could be allowed to stay at home instead. Another example is the ZCare Monitor [31], a home monitoring system for COVID-19 patients in the non-acute phase developed by the Zucchetti Group in collaboration with Doctors Without Borders for the Lodi Hospital, among the first to address the health emergency related to Coronavirus, and used by 11 different hospital facilities, for a total of more than 10,000 patients . The center contacts COVID-19 positive patients who are quarantined at home by phone. Patients, daily, enter their physiological data through the platform. In addition, a pulse oximeter, connected via Bluetooth, sends the following data to the platform twice a day: body temperature, blood saturation level, maximum blood pressure, pulse, and heart rate. This data is collected directly by the hospital platforms and monitored daily, also with the help of software for predicting clinical status (using machine-learning algorithms).

Other research has focused on the possibility of using mobile devices for diagnosis. As an example, in [20], the authors discuss the pre-symptomatic detection of COVID-19 using smartwatch data, also focusing on asymptomatic patients. Given that the asymptomatic status of the disease does not preclude the possibility of infection, it would be helpful to know about this condition. The authors asked study participants to record daily symptoms and share fitness tracker data. The types of data collected included heart rate, steps, and sleep over a period of several months. Two infection detection algorithms (RHR-Diff and HROS-AD) were developed, and based on these data, it was analyzed that 63% of COVID-19 cases could be detected prior to symptom onset via a two-level alert system based on the occurrence of extreme increases in resting heart rate relative to the individual baseline. Such an alarm would allow, even several days in advance, the recognition of a possible asymptomatic infection and the ability to proceed with standard tests. Other examples include *AI4COVID-19* [16], in which a preliminary COVID-19 diagnosis is performed from cough samples captured through the app, and [24], an app that collects self-reported symptoms, diagnostic testing results, and smartwatch and activity tracker data. This data, collected over time, is used to help identify subtle changes indicating an infection.

3. YouCare

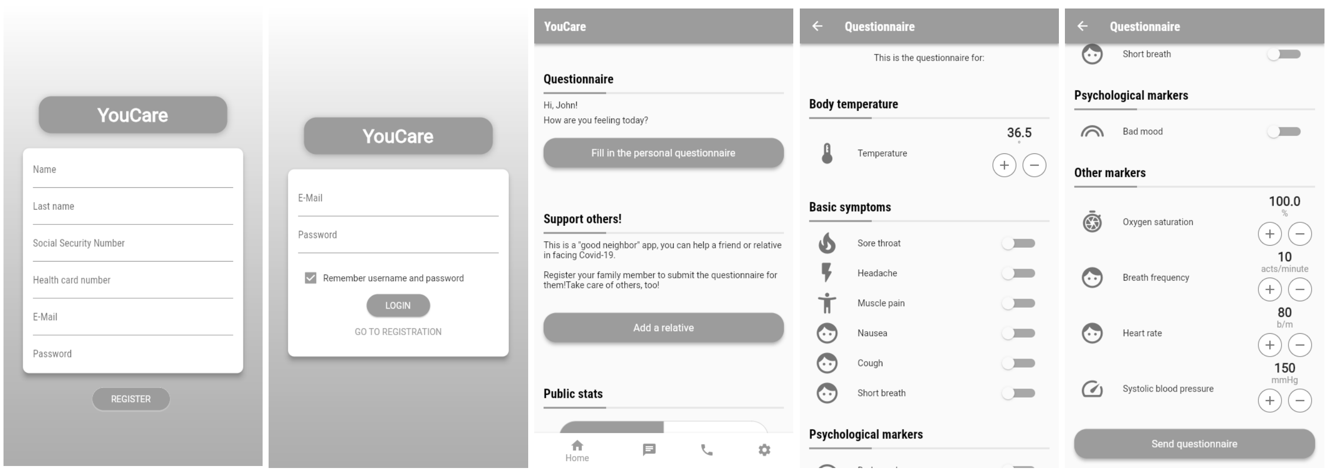
YouCare consists of three components: a cross-platform mobile application (Android/iOs/Web) used by the system's users, a server that receives/responds to requests from that application and stores user data, and a web platform that allows authorized physicians to access their patients' data. Based on what has been seen in the literature and from interviews with healthcare professionals and COVID-19 patients, the functional requirements of the application have been defined. Specifically, the application must allow:

- Registration and log in. The unauthenticated user must be able to login into the app. There must also be the ability to subscribe themselves. The app functionalities should only be available to people who have COVID-19 (or people with suspected COVID-19 and with a scheduled test or waiting for a test result).
- It should be possible to use the app's functionalities on behalf of another person. This is very important given the lower penetration of digital technologies among older people, who are also the category most at risk for COVID-19.
- Symptoms entry and statistics. The user must be able to enter his physiological data, useful for diagnosis, directly in his user area of the system on a daily basis. The application must allow the filling in of the questionnaire also for other users registered as relatives. The user shall also be able to consult global statistics about the users of the system.
- Forum. The user should have access to a Forum section in the application where he can read the topics of other users, reply to the topics and create his own. This should be possible in an anonymous way, possibly even restricting the geographic area in which messages can be viewed.
- Telephone support. The user should have the option of requesting telephone assistance. These requests will be handled by healthcare providers who, by accessing the request list, can make phone calls to the application users.
- View of users' status. The physician accessing the system with his credentials (provided by the administrator) can view the data collected with the questionnaires of his patients. Alerts and statistics about his patients should also be available.

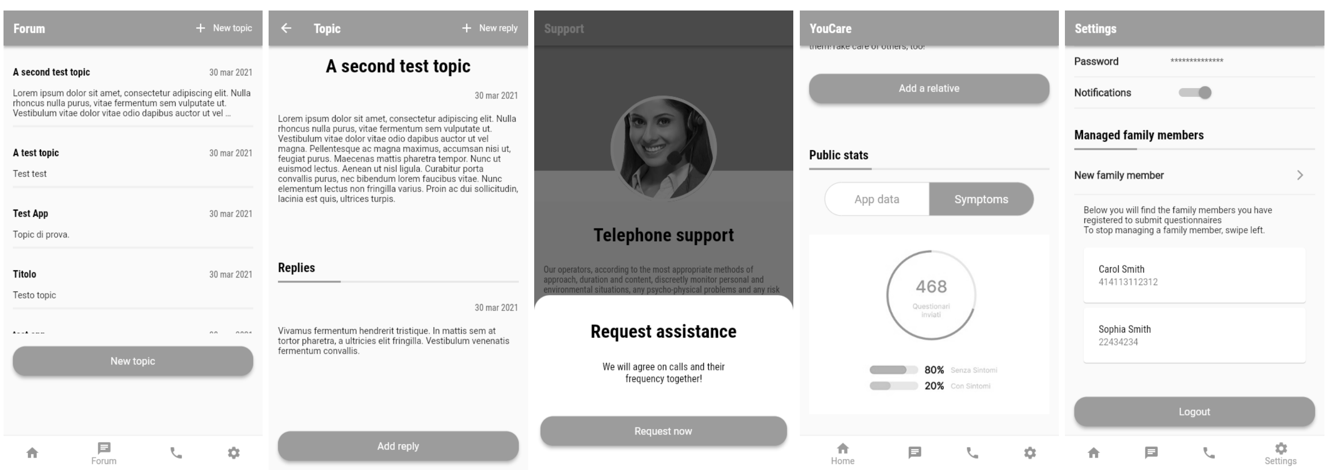
Non-functional requirements include:

- The registration process should be as fast as possible even in cases where the user is subscribing other people.
- The design of the system, especially with regard to the mobile application, must be as intuitive as possible due to the very diverse user base.
- The application must be available for as many devices as possible (both mobile and desktop).
- Access to the data and services offered by the system must meet the most common security standards. No user should be allowed to access the confidential data of other users to which he/she is not entitled. Access to the patient's data should be restricted to their physician.
- The system must be always online.

Based on the identified requirements we proceeded with the development of the app. The app development was performed using the Flutter framework, which allowed the development of a cross-platform Android, iOS, and Web application with a single codebase in Dart. As shown in Figure 1a, the user, in order to access the application must register to the platform by entering his personal data. In order to validate his identity, adapting the app to the Italian national health system, the number printed on the Italian health card is requested (a confidential number of which only the regional health authority should be aware; this validation should be adapted to the health systems of the various countries). This same check is used to allow one person to use the app on behalf of another. However, it is always possible for a person to register themselves, in which case they can decide whether to still allow the other person to use the app on their behalf or to remove that ability. Finally, the app's functionality will only be accessible to individuals for whom there is a positive (or scheduled/pending) COVID-19



(a) User registration (b) Log in (c) Home screen (d) Questionnaire (part 1) (e) Questionnaire (part 2)



(f) Forum (g) A topic (forum) (h) Telephone support (i) Public stats (j) Settings

Figure 1: Application screenshots.

test in the regional health system (again, this also needs to be adapted to the country's health systems).

After registration, the user can log in (see Figure 1b) and access the home screen of the application (see Figure 1c). From the home screen, the user can access the filling of the questionnaire (see Figure 1d and Figure 1e) that allows the user to enter his medical data daily and forward them to the server, thus allowing their vision to his physician. YouCare also allows the user to fill out questionnaires for other people (for example relatives who do not have a device with Internet access), in fact, the home screen presents a section for handling relatives. It is possible to add a relative by clicking the relative button (see Figure 1c) on the home screen (the number of the health card of the person to be added is required). After that, it is possible to fill in the daily questionnaire with the medical data. At any time, a person added by other users can register as a new user to the platform, thus taking possession of their user. The user, who previously managed the daily questionnaires, is notified of the registration by a notification on the device. YouCare offers features aimed at improving the psychological well-being of quarantined patients. These features include a forum where users can interact with other users (see Figures 1f and 1g), and a "Telephone support" section where users can request a telephone contact from their physicians or a health care professional (see Figure 1h). On the home screen the user can also view some global statistics about the application and the daily questionnaires of the users (see Figure 1i). Together these social and informational functions can be helpful to the person's emotive state. This information could be of moral support since it will show the run-time collective status of people in the same condition. Finally, there is a "Settings" section (see Figure 1j).

The application communicates with a server that provides a JSON API for the functionality needed by the app and takes care of storing the data entered by the users. The server uses PHP and MySQL technologies and uses the JSON WEB TOKEN standard (JWT - RFC 7519 [6] standard) for authentication management. The Google Firebase Cloud Messaging service [12] is also used to manage YouCare push notifications.

A web-based platform, shown in Figure 2, is also available for physicians to view patient data and medical information. The system administrator can add a physician to the system, and the association between a General Practitioner and his patients can be automatically performed thanks to the health card numbers (at least in countries that provide this association, such as Italy). The doctor can view the data entered by each patient through the daily questionnaires and statistics on the questionnaires. A MEWS scale [28] score is also provided for each patient in order to immediately alert the physician of a possible clinical instability of the patient. The score obtained from the scale ranges from a minimum of 0 to a maximum of 14. Above level 5 the patient is critical and unstable. For all other patients with normal values, however, the MEWS is an important tool for early detection of worsening clinical conditions.

	Task
1	Register a personal account from the app
2	Log in
3	Look at the public statistics from the app
4	Fill out a questionnaire for themselves
5	Add a first relative/family member
6	Add a second relative/family member
7	Remove one of the two relatives/family members
8	Fill out a questionnaire for a relative/family member
9	Add a new topic to the forum
10	Reply to one of the existing topics
11	Request phone support from the application

Table 1: Tasks performed by the experiment participants.

4. Evaluation

We carried out a user study aimed at evaluating the usability of YouCare. In the experiment, we asked participants to use the app's features and then evaluate the app by filling out a questionnaire.

4.1. Participants

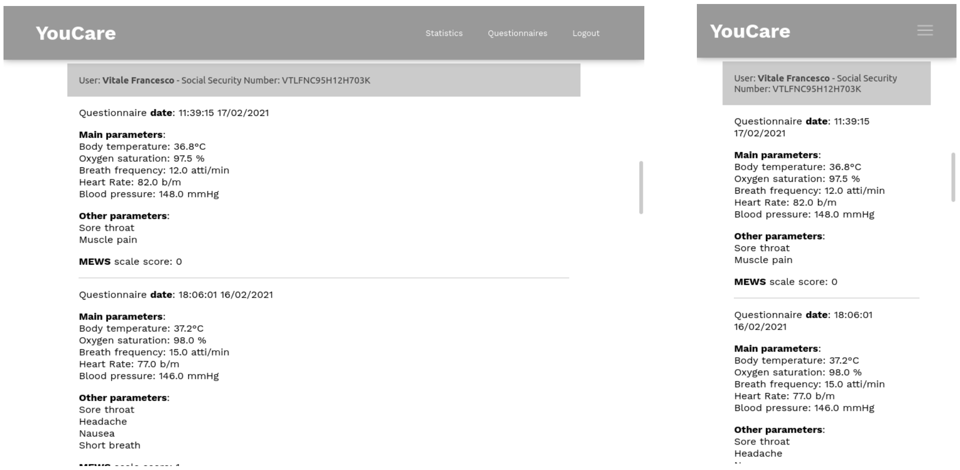
For the experiment, 15 participants (2 female) were recruited. They were all university students between 20 and 28 years old ($M = 22.3$, $SD = 2.6$) and chose to participate for free. All were usual users of computers and smartphones. Two of them had personally dealt with COVID-19.

4.2. Apparatus

The experiment was conducted on the individual participants' smartphones due to restrictions due to COVID-19 that prevented direct contact. All devices are recent Android smartphones from different manufacturers (Samsung, Xiaomi, etc.). The server was running an 8-core server with 32 GB of RAM running Ubuntu 20.04 and with a gigabit internet connection.

4.3. Procedure

Before starting the experiment, the participants filled out a questionnaire with the following information: personal data (age, gender), previous experiences with smartphones, personal experience with COVID-19.



(a) Desktop view (b) Mobile view

Figure 2: Web-based platform for physicians.

	Question
1	I think that I would like to use this system frequently
2	I found the system unnecessarily complex.
3	I thought the system was easy to use
4	I think that I would need the support of a technical person to be able to use this system
5	I found the various functions in this system were well-integrated
6	I thought there was too much inconsistency in this system
7	I would imagine that most people would learn to use this system very quickly
8	I found the system very cumbersome to use
9	I felt very confident using the system
10	I needed to learn a lot of things before I could get going with this system

Table 2: SUS questionnaire.

Participants were asked to use the app, performing a list of the most representative tasks (see Table 1). In order to assess the intuitiveness of the app, no explanations were given about it or on how to perform the tasks.

At the end of the experiment, the participants were asked to complete a System Usability Scale (SUS) [9] questionnaire. SUS includes ten statements (see Table 2), to which respondents had to specify their level of agreement using a five-point Likert scale. The questions alternate between positive and negative (since they are in a rather standard form we do not include them here). Each SUS questionnaire has a score between 0 and 100, of which we then calculated the averages on all participants.

In addition to SUS, we also collected further feedback through an open form and verbal interaction.

5. Results and discussion

All participants completed the experiment. For each participant, the experiment lasted about 20 minutes.

From the server logs, we have been able to verify that all participants have successfully completed all tasks. The average SUS score was 83.3 ($SD = 9.6$), which is a really good value [25]. Regarding free form comments and interviews, feedback was generally positive about both the usability of the app and its usefulness. The most common suggestion concerned the ability to view the status (confirmation) of the telephone support request. Other suggestions include adding reminders, integration with smart devices for health monitoring, keeping the name of sections always visible in the UI, adding more information about the general statistics. Some complaints were about the interaction required to remove a family member. Finally, a bug that occasionally occurred when adding/removing family members was reported.

Although the results obtained are good, it must be noted that the participants were all young people experienced in the use of smartphones, a less experienced user might have more difficulties. However, this is mitigated by the fact that one person can use the app in behalf of another person.

6. Conclusions and further works

In this paper, we presented YouCare, a COVID-19 patient telehealth application whose application design was influenced by the analysis of existing applications and interviews with COVID-19 patients.

Future work includes the possibility of directly interfacing the application with Bluetooth wearable devices in order to allow for the collection of some of the medical data without the need for the user to fill out the questionnaire in an ongoing and more reliable manner. It is also intended to integrate a video call feature to complement the phone contact and allow users to receive medication prescriptions from their doctor directly from the application.

7. Acknowledgment

This work was partially supported by the grant "*Fondo FSC 2014 2020 per il Piano Stralcio Ricerca ed Innovazione 2015-2017 - MIUR - progetto MEDIA*" (project code: PON03PE_00060_5/12 - D54G14000020005).

References

[1] N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, and S. K. Jha. A survey of covid-19 contact tracing apps. *IEEE Access*, 8:134577–134601, 2020. https://doi.org/10.1109/ACCESS.2020.3010226.

[2] W. Al-Zyoud, T. Oweis, H. Al-Thawabih, F. Al-Saqqar, A. Al-Kazwini, and F. Al-Hammouri. The psychological effects of physicians' communication skills on COVID-19 patients. *Patient Preference and Adherence*, Volume 15:677–690, Apr. 2021. https://doi.org/10.2147/ppa.s303869.

[3] T. Alanzi. A review of mobile applications available in the app and google play stores used during the COVID-19 outbreak. *Journal of Multidisciplinary Healthcare*, Volume 14:45–57, Jan. 2021. https://doi.org/10.2147/jmdh.s285014.

[4] F. Angellotti, G. Costagliola, M. De Rosa, and V. Fuccella. ifree: design and evaluation of a pointing method for disabled users on mobile devices. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, pages 1–6, 2020. https://doi.org/10.1109/ICHMS49158.2020.9209537.

[5] Arab News. Saudi arabia's mawid smartphone app offers coronavirus self-assessment, 2020. https://www.arabnews.com/node/1652171/saudi-arabia.

[6] Auth0. Json web tokens. https://jwt.io.

[7] J. Best. Wearable technology: covid-19 and the rise of remote clinical monitoring. *BMJ*, 372, 2021. https://doi.org/10.1136/bmj.n413.

[8] H. Bourdon, R. Jaillant, A. Ballino, P. El Kaim, L. Debillon, S. Bodin, and L. N'Kosi. Teleconsultation in primary ophthalmic emergencies during the covid-19 lockdown in paris: Experience with 500 patients in march and april 2020. *Journal Français d'Ophtalmologie*, 43(7):577–585, 2020. https://doi.org/10.1016/j.jfo.2020.05.005.

[9] J. Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.

[10] M. Cipriano, G. Costagliola, M. De Rosa, V. Fuccella, and S. Shevchenko. Recent advancements on smartwatches and smartbands in healthcare. In *Innovation in Medicine and Healthcare*. Springer Singapore, 2021. https://doi.org/10.1007/978-981-16-3013-2_10.

[11] G. Costagliola, M. De Rosa, and V. Fuccella. A technique for improving text editing on touchscreen devices. *Journal of Visual Languages & Computing*, 47:1–8, 2018. https://doi.org/10.1016/j.jvlc.2018.04.002.

[12] Google. Firebase cloud messaging. https://firebase.google.com/docs/cloud-messaging.

[13] Health Service Executive. COVID Tracker app - Ireland's coronavirus contact tracing app. https://covidtracker.gov.ie.

[14] Hospital Times. Northampton nhs at forefront of innovation for virtual ward tech, 2020. https://www.hospitaltimes.co.uk/northampton-nhs-at-forefront-of-innovation-for-virtual-tech/.

[15] Imperial College London. Wearable sensor trialled for remote covid-19 monitoring, 2020. https://www.imperial.ac.uk/news/196973/wearable-sensor-trialled-remote-covid-19-monitoring/.

[16] A. Imran, I. Posokhova, H. N. Qureshi, U. Masood, M. S. Riaz, K. Ali, C. N. John, M. I. Hussain, and M. Nabeel. Ai4covid-19: Ai enabled preliminary diagnosis for covid-19 from cough samples via an app. *Informatics in Medicine Unlocked*, 20:100378, 2020. https://doi.org/10.1016/j.imu.2020.100378.

[17] A. Jhunjhunwala. Role of telecom network to manage COVID-19 in india: Aarogya setu. *Transactions of the Indian National Academy of Engineering*, 5(2):157–161, June 2020. https://doi.org/10.1007/s41403-020-00109-7.

[18] J.-H. Lee. Future of the smartphone for patients and healthcare providers. *Healthc Inform Res*, 22(1):1–2, 2016. https://doi.org/10.4258/hir.2016.22.1.1.

[19] M. Medina, C. Babiuch, M. Card, R. Gavrilescu, W. Zafirau, E. Boose, K. Giuliano, A. Kim, R. Jones, and A. Boissy. Home monitoring for COVID-19. *Cleveland Clinic Journal of Medicine*, May 2020. https://doi.org/10.3949/ccjm.87a.ccc028.

[20] T. Mishra, M. Wang, A. Metwally, G. Bogu, A. Brooks, A. Bahmani, A. Alavi, A. Celli, E. Higgs, O. Dagan-Rosenfeld, B. Fay, S. Kirkpatrick, R. Kellogg, M. Gibson, T. Wang, E. Hunting, P. Mamic, A. Ganz, B. Rolnik, X. Li, and M. Snyder. Pre-symptomatic detection of covid-19 from smartwatch data. *Nature Biomedical Engineering*, 4(12), Dec. 2020. https://doi.org/10.1038/s41551-020-00640-6.

[21] P. Mozur, R. Zhong, and A. Krolik. In coronavirus fight, china gives citizens a color code, with red flags, 2020. https://www.nytimes.com/2020/03/01/business/china-coronavirus-surveillance.html.

[22] NHS UK. The NHS test and trace app support website. https://covid19.nhs.uk.

[23] Pathcheck Foundation. PathCheck foundation | COVID-19 technology & research. https://pathcheck.org.

[24] G. Quer, J. M. Radin, M. Gadaleta, K. Baca-Motes, L. Ariniello, E. Ramos, V. Kheterpal, E. J. Topol, and S. R. Steinhubl. Wearable sensor data and self-reported symptoms for COVID-19 detection. *Nature Medicine*, 27(1):73–77, Oct. 2020. https://doi.org/10.1038/s41591-020-1123-x.

[25] J. Sauro. 5 ways to interpret a sus score, 2018. https://measuringu.com/interpret-sus-score/.

[26] Sensyne Health. Cvm-health. https://www.cvm-health.com.

[27] So.Re.Sa. S.p.A. e-Covid SINFONIA app. https://www.soresa.it/Pagine/e-covid_sinfonia.aspx.

[28] C. Subbe, M. Kruger, P. Rutherford, and L. Gemmel. Validation of a modified Early Warning Score in medical admissions. *QJM: An International Journal of Medicine*, 94(10):521–526, 10 2001. https://doi.org/10.1093/qjmed/94.10.521.

[29] L. Taylor. Covid-19: Brazil's hospitals close to collapse as cases reach record high. *BMJ*, 372, 2021. https://doi.org/10.1136/bmj.n800.

[30] G. Vitiello, M. Sebillo, L. Fornaro, M. Di Gregorio, S. Cirillo, M. De Rosa, V. Fuccella, and G. Costagliola. Do you like my outfit? cromnia, a mobile assistant for blind users. Goodtechs '18, page 249–254, New York, NY, USA, 2018. Association for Computing Machinery. https://doi.org/10.1145/3284869.3284908.

[31] Zucchetti s.p.a. Zucchetti continues its solidarity commitment with zcare monitor, 2020. https://www.zucchetti.it/website/cms/comunicato-stampa-dettaglio/8661-zucchetti-continua-il-suo-impegno-solidale-con-zcare-monitor.html.

An Automatic Indoor Positioning Robot System Using Panorama Feature Matching

Jun-Wei Chang[1], Huan Chen[2], Krishna Kambhampaty[3], Bo-Chao Cheng[1]
[1]Department of Communications Engineering,
National Chung Cheng University,
Chia-I, Taiwan 621
[2]Department of Computer Science and Engineering
National Chung Hsing University,
Taichung, Taiwan 402
[3]Department of Computer Science
York College of Pennsylvania
York, PA 17403-3651

Abstract - **With Location-Based Service (LBS) support, a mobile user can enjoy specific information/services based on their location. The GPS-based positioning technique plays an essential role in outdoor LBS applications. Owing to physical barriers and potential interference inside a building, GPS-based positioning systems do not work well for commercial indoor LBS applications. Various wireless technology-based position systems have been proposed to solve indoor positioning problems (e.g., Wi-Fi, Bluetooth, infrared, and RFID), but they come with challenges involving accuracy, installation, interference, and power consumption. We propose a next generation indoor localization recognition system called IP-PFM (Indoor Position based on Panorama Feature Matching) to overcome the limitations and problems of wireless technology-based approaches. By applying panorama image recognition technology, IP-PFM accurately pinpoints the indoor position. Further, IP-PFM also includes an indoor position system chat bot feature to automate chat in real-time with users — using the Line Bot application programming interface (API)—which allows a mobile user to post a picture taken with their smartphone. Upon receiving the picture, IP-PFM determines the location and plots it on the floor plan. We have implemented a prototype of IP-PFM to demonstrate the end-to-end functionality of the indoor localization recognition chat bot system[1].**

***Keywords**: Location-based service; indoor positioning; panoramic; image recognition; Line Bot.*

I. INTRODUCTION

In recent years, positioning technology and location-based services (LBS) have become increasingly relevant to our lives, as they are used in navigation systems, such as Google Maps. Outdoor positioning technology is mostly ineffective for applications requiring indoor positioning systems (IPSs). However, IPSs may have high commercial value for use in department store merchandise shopping guidelines, home care service for seniors, airport station path guidance, and intelligent factory robot navigation are common applications [1]. As a result, Google Maps is moving toward in the direction of implementing IPS, and recently provided real-time pictures of 13 Tokyo Metro stations to reduce the number of passengers who lost their way in complex underground tunnels. If indoor positioning system can be used in conjunction with the Internet of Things (IoT), it can provide a broader range and ubiquitous services. Many IPSs based on wireless technology (such as Wi-Fi and Bluetooth) are usually used only in particular situations. For example, the user may have to be a staff member of the site to access the area's Wi-Fi system for positioning, meaning that visitors could not use the same approach for positioning. IPSs contain many variant wireless parameters, such as Angle of Arrival (AoA), Time of Flight (ToF), Return Time of Flight (RTOF), and Receive Signal Strength (RSS), but these technologies have bottlenecks and blind spots [2]. Wireless IPSs also have problems with accuracy, installation, interference, and power consumption [3].

Because of the accuracy and time delay in IPSs, researchers initially used image data to build a reference model (including architectural models, floor plans, windows, doors, walkways, or other building parts) to make improvement happen [4]. Some of them used the characteristics of the current image to match the architectural reference model or the objects and features in the floor plan, and finally, they extrapolated the location of the camera [5].

A system with machine learning in the back-end combined with reliable image recognition IPS could solve the complex problem of blurred images due to focal length or light. Therefore, some scholars developed a framework based on two digital cameras to solve the problem of the blurred image [6]. When we enter an unfamiliar building, even if we have the building directory, we cannot quickly determine our location. Therefore, in this paper, we demonstrate how to use feature recognition technology to achieve indoor positioning and propose an indoor positioning method based on panoramic image recognition combined with a Line Bot to create an automatic indoor positioning robot system. Owing to the pop-

[1] DOI reference number: 10.18293/DMSVIVA2021-010

ularity of mobile devices, we can take pictures anytime and anywhere, and users can shoot the image in front of them and upload it to a server instantly. After the images are processed by a server with panoramic photos in the database, we can obtain the user's location as well as the calculated distance and field of view to increase accuracy and user experience.

The structure of this paper is as follows. Section 2 lists some relevant researches that will be applied during the implementation stage. Section 3 introduces the detailed structure of the method proposed in this research and various calculation methods. Section 4 describes the experimental results. Section 5 shows the implementation and results of the relevant design methods. Section 6 summarizes and discusses future developments and applications of this paper.

II. RELATED WORK

This section elaborates on relevant technical applications and execution methods based on image feature matching. One approach is to apply Scale-Invariant Feature Transform (SIFT) [7] on a space vector to find the extreme point and extract its coordinates, image, and rotation invariant features through the algorithm to identify the object. The fault tolerance is quite high for light, noise, and error in the visual angle. In addition, the application of this algorithm is extensive. For these reasons, there are many improved feature recognition techniques that are designed based on SIFT. The feature recognition technology used in this article, Speeded Up Robust Features (SURF) [8], is also a SIFT-based scheme which is improved the performance with the use of the Hessian matrix to operate on the feature points. In terms of processing efficiency, SURF can achieve much higher speeds than SIFT, and it is more stable than SIFT in image transformation.

Salarian et al. [9] claimed that accurate information about the location and orientation of cameras in mobile devices is at the heart of the use of adaptive services, such as LBS. Most mobile devices rely on global positioning system (GPS) data, but the satellite's signal quality is not always good enough, which may lead to positioning errors. Therefore, they used SURF feature recognition technology on the image with the GPS tag image database to find the best match with the scene of the picture. However, this method compared the image with the general image data rather than the panoramic image; if there was no relevant information in the database, it was unable to make accurate corrections. As such, they constructed a location system by applying the technology of structure from motion (SfM) and a 3D camera position reconstruction model. In more than 170 sets of data experiments, the average error of their experimental results was less than five meters, which compares favorably with some other published methods. The disadvantage of this method is that its accuracy is related to the size of the database. If the information is incomplete, the error increases.

Xiao et al. [10] used the SUN360's panoramic image database and a Support Vector Machine (SVM) training model to enable the system to identify scenes and perspectives such as hospitals, restaurants, and theaters, as well as the orientations faced by observers. They trained two classifiers in the training model. The first was the scene classifier. For this classifier, the model trained m kinds of classifiers to train the experimental panoramic image database on m kinds of panoramic image categories. The second was a classifier of images with different angles of view. For this classifier, the same type of images with different angles of view were used for training. The number of different types of panoramic images was gradually increased to achieve effective classification results. The key limitation is that the author only classifies the panoramic image into four different types (symmetry, single axis symmetry, double axis symmetry, and full symmetry). Therefore, this approach to visual angle recognition has some problems. When encountering highly repeatable images in visual angle recognition and the angle of view is exactly on the axis of symmetry, it is easy to produce errors due to too high similarity, despite training through the SVM model.

Molnar and Kovacs [11] proposed a method for making panoramic images by using pixels that look for image boundaries and searching for overlapping areas to find the same boundaries for overlap. As their method does not need to grab feature points and feature comparisons, it can be run on low-resolution images. Because the red-green-blue (RGB) color in the pixel varies due to the brightness between different images, the RGB color was converted to grayscale before performing the calculations to prevent impacts on accuracy. They also use a graphics processing unit (GPU) acceleration to improve in the experiment significantly; however, the authors mentioned that if the range of the video arc is too large, there would be a reduction in accuracy.

Kawaji et al. [12] applied Locality Sensitive Hash (LSH) method to speed up the image matching process in their proposed indoor positioning system, and used the Railway Museum with 426 omnidirectional panoramas and 1067 supplementary images for the experimental study. LSH is a powerful approach to find the nearest neighbor technique, but it has a hash collision problem which is inevitable. Further, the proposed system does not consider images with different angles of view such that it needs many omnidirectional panoramas and supplementary images to construct a confidence database.

Huang et al. [13] proposed a Sorting Hat method to filter out unrelated element pairs to accommodate the mobile device environment where the image matching process needs to achieve a good design meeting the performance requirements. Sorting Hat starts the matching process from the innermost ring (containing the most critical features to determine similarity between two images) toward the outer ring. However, the proposed method was not practical and cost-effective because it requires extra sensors and computation power to calculate the user's current viewing direction.

III. IP-PFM

Figure 9 shows the process for the proposed IP-PFM which uses feature recognition technology, SURF, to calculate the characteristics of the user input image and find the features and the panoramic image in the database to perform a match calculation. Then, based on the matching results, the distance and where the user is facing will be estimated. Finally, the user's possible location will be computed and marked on the map. TABLE I shows the symbols related to IP-PFM algorithms and calculations.

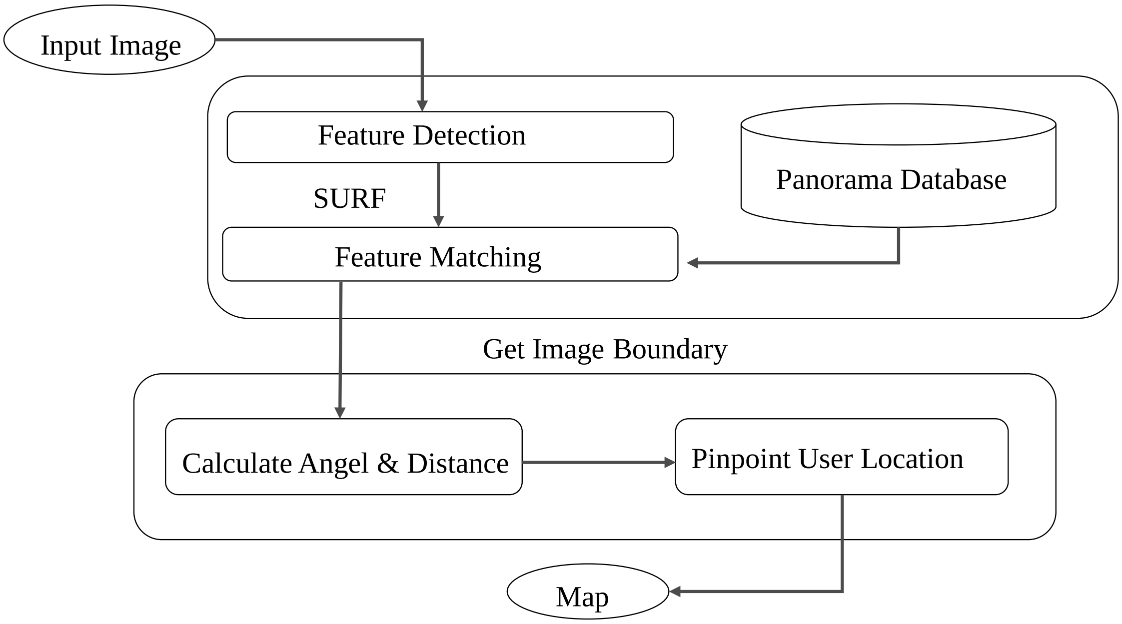


Figure 1. IP-IMP System architecture

TABLE I. Algorithmic symbols table

Symbol	Description
N	Number of panoramic images in the database
α	Match point number
α_{max}	Max match point number
M	Matched panorama
B_r	Right boundary
B_l	Left boundary
V_r	Matched range for four vertices
$[i]$	The panorama in the database
p	The image the user inputs
Cp	The coordinate information of the matched panorama on the map
Cu	The coordinate information of the user on the map
D	Prediction distance
θl	Angle of left boundary
θr	Angle of right boundary
θ	Angle of viewpoint
ω	Width of panorama
W_c	Width in the panorama for the picture taken at the panorama center
W_{60}	Width in the panorama for the picture taken at 60cm from the panorama center
θ_l	Angle of left boundary
θ_r	Angle of right boundary
θ	Angle of viewpoint

Now, we detail the IP-PFM implementation as following.

A. Pairs of Images and Panoramas

We use SURF to detect feature points of the photos inputted by users, and IP-IMP compares these with panoramic photos in the database (as shown in Algorithm 1: line 2~8). Subsequently, the input photos that correspond to the panoramic photo are determined. IP-PFM uses the best matching panoramic image to compute the left boundary (B_l) and right boundary (B_r) of the range of the user input picture in the panoramic image (as shown in Algorithm 1: line 9~11). In other words, the position information of the image input by the user is calculated to find where the user is situated in the panoramic image. As shown in IP-IMP finds the left and right boundaries of the range (green frame) of the user input image in the panoramic image.

Algorithm 1: Compare the image with a panorama

Input: $P[i]$, p

Output: B_l , B_r , α , α_{max}, M

```
01  αmax = 0;
02  For (i=1, i≦N, i++)
03  |   α= SURF(P[i], p);
04  |   If (αmax≦α)
05  |   |   αmax =α;
06  |   |   M= P [i];
07  |   End If
08  End For
09  [Va Vb Vc Vd] = getVertex [M];
10  Bl = (Va [0]+ Vd [0])/2;
11  Br = (Vb [0]+ Vc [0])/2;
```

Algorithm 1 Match user pictures and panoramic functions

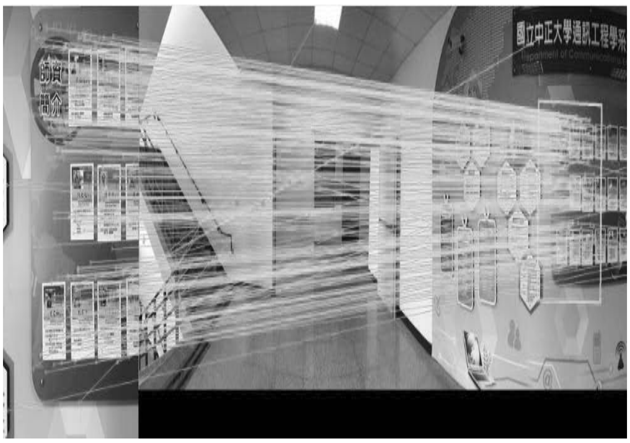


Figure 2 Photo by SURF and boundary calculation

B. Angle and distance calculations

Next step is to calculate the user's position information (distance from the center of the panoramic image and facing angle). The center of the panoramic image is the location where the panoramic image was taken, and the distance from the center of the panoramic image is the predicted value after calculations are performed by the system. IP-PFM determines how far the user must be from this position to take this picture. The facing angle means that the leftmost side of the panoramic image is set to 0 degrees. This means that IP-PFM needs to know which angle in the panoramic image where the user sees this scene.

Using the left and right boundaries obtained in Algorithm 1, the range of the user-shot image in the panoramic image is calculated using the left and right boundaries. Then, the distance between the user's shooting position and the matching panoramic image center and the orientation (expressed in degrees) are calculated. The angle can be obtained in Eq. (1) and illustrated the result of the inference in Figure 3.

$$\theta = \left\lfloor (B_l / \omega) \times 360 + (B_r / \omega) \times 360 \right\rfloor / 2 \qquad (1)$$

B_l and B_r are the left and right borders of the panoramic image, respectively, which are results calculated using Algorithm 1. When the leftmost side of the panoramic image is 0 degrees, the user's facing angle is obtained through Eq. (1). Example: If width of panorama image ($\boldsymbol{\omega}$) = 4,096, B_l = 2,000, and B_r = 2,096, as found in Algorithm 1, then the user's facing angle is calculated to be 180°.

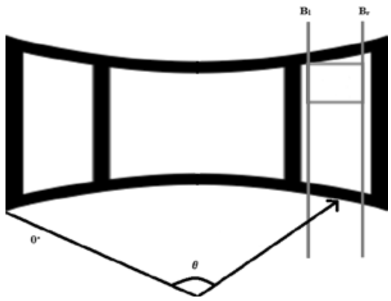


Figure 3 Angle calculation schematic

The distance is obtained by Eq. (2) , which is derived from Figure 4.

$$D = 60 \times \left[\frac{W_c}{(W_{60} - W_c)} \right] \times \frac{|B_l - B_r| - W_c}{W_c} \qquad (2)$$

W_c is the width of the picture taken by the user in the panoramic image; W_{60} is the width of the range of the photograph taken at 60 cm from the panoramic image center, and 60 is the calculated reference length.

With the shot remaining the same and without zooming, the field of view for taking the photo at the same location will be static. That is, W_c and W_{60} are constant values. In the camera lens used in this paper, these two values are approximately 1,600 and 1,200, respectively. As we have the B_l and B_r calculated by Algorithm 1, we can calculate that the distance through Eq. (2). In other words, the two above equations can thus obtain the face of the user's photograph and the distance between the location of the user's photo and the center of the matching panoramic image.

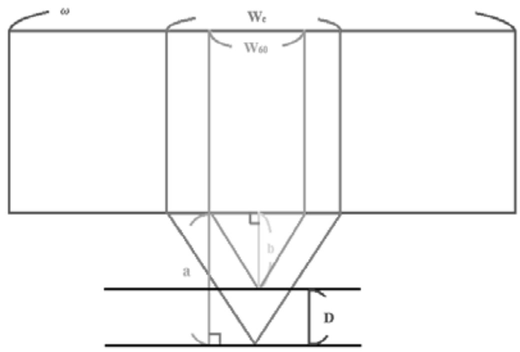


Figure 4 Diagram for the distance calculation

C. Output map and indicate user location

After Eq. (1) calculates the user's facing angle θ and Eq. (2) calculates the distance D between the location of the user's captured image and the center of the panoramic image, the user's location can be estimated based on these two pieces of information and marked on the map. As shown in Figure 5, C_p is the coordinate on the map with the center position of the matched panoramic image, and D is the estimated distance between the user and the center position of the panoramic image calculated by Equation (2). Here, θ is the estimated angle. Combining these two results and cooperating with the scale of the map, the coordinate position C_u of the user's image can be obtained through Eq. (3). Finally, we mark the location of C_u on the map.

$$C_u = C_p + [\cos\theta * D, \sin\theta * D] \qquad (3)$$

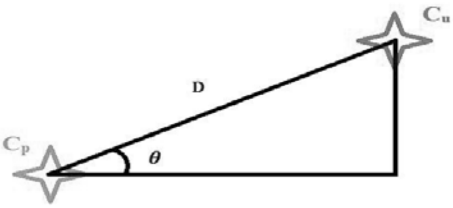


Figure 5 Panorama image center and user coordinate relationship

IV. EXPERIMENT

This experiment uses the feature recognition technology, SURF, which is a mature technology. It is used not only for image comparison but also in facial recognition technology-related research. As such, this paper adopts the current open source code of SURF in OpenCV 2.4.11. We choose a SURF identifier to provide a stable feature recognition function and then use the algorithm in IP-PFM to obtain the required information.

The client app developed and tested through Line Bot achieved the purpose of this paper. The image is sent to the server from the user, and the result is returned to the user through the calculation of the server. The indoor building directory is displayed in the app. The map also pinpoints the user's location. The experiment environment is shown in TABLE II.

TABLE II. EXPERIMENTAL ENVIRONMENT

SW/HW	Description
Processor	Intel Core i7-3770
RAM	16GB
Operating System	MS Windows 7 (64-bit)
Programming Language	Java
Camera	12 megapixels; Wide-angle: war/1.8 aperture

A. Distance from shooting location to panoramic center and error relationship

In a single scene, we shoot multiple images at different distances from the center of the panorama and use these images as system inputs. Then, we calculate the distance between the location where the image was shot and the center of the panorama. Figure 6 shows the results and the actual errors. Among them, -120, -60, 0, 60, and 120 are the actual distances from the center of the panorama when shooting the image. Negative values represent moving forward from the center (i.e., the field of view of the image is narrower and the coverage is smaller), and positive values represent moving backwards from the center (i.e., the field of view of the image becomes wider and covers a wider range). This also applies in terms of the distance estimation error.

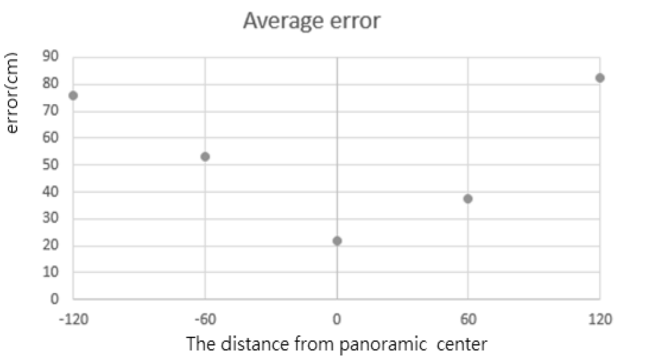


Figure 6 The shooting position and the panoramic center distance pair error graph

After the system calculations, an image taken farther from the center of the matching panoramic image will have a larger error. Conversely, if the scene image is taken closer to the center of the panoramic image, there will be a smaller error value after the calculation. Overall, the average error value was within one meter.

B. System prediction position and actual position error value

In this experimental simulation, 61 sets of panoramic images were set as benchmarks, and 70 scene images were samples. Regarding the estimation errors in this experiment after system calculations, Figure 7 shows the results obtained by using 10 images captured in the center of the panorama in different scenes. The green points represent the center of the panorama corresponding to the test image, which is where the image was taken. The red points mark the system's final estimation of the location where this image was taken. TABLE III shows the error values of these 10 different calculation results. The minimum value of the error value was 4 cm and the maximum value was 123 cm. This scale of error is considered acceptable for an indoor setting.

C. The relationship between the number of panoramic images and the error in the database

From Figure 8, we know that the more panoramic images there are in the database, the smaller the distance calculation error. This means that the more complete the panoramic image data in the database, the higher the accuracy of the system. This also solves the problem of increased error because the shooting position is further away from the panoramic center. As long as the number of panoramic images increases, the distance between the location of the captured image and the center of the corresponding panoramic image can be reduced, thereby decreasing the error value of the system judgment.

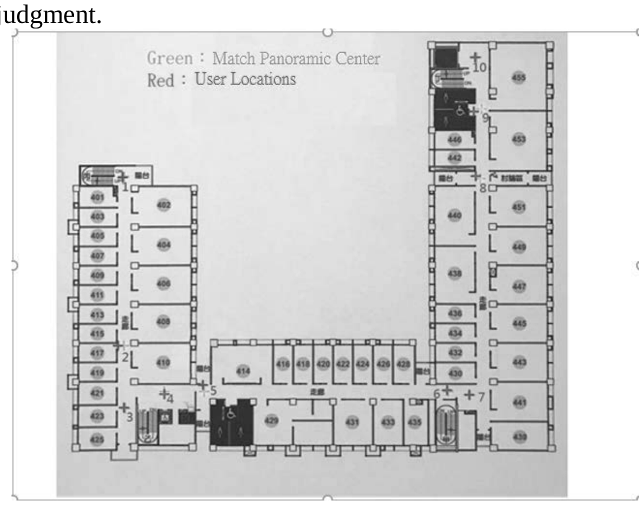


Figure 7 User location and panoramic map

TABLE III. ERROR VALUE TABLE

Number	Error (cm)	Number	Error (cm)
1	21	6	21
2	79	7	4
3	14	8	68
4	47	9	123
5	96	10	17

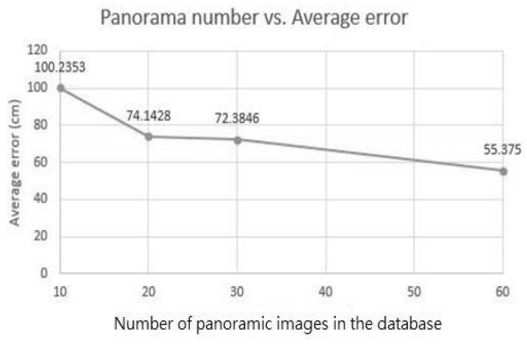


Figure 8 Panorama quantity vs. average errors

V. SYSTEM IMPLEMENTATION

The paper combines Line Bot with the method of image calculation and location estimation to make a usable indoor positioning robot system. Figure 9 shows the system with Line Bot as a bridge between the user and the IP-PFM. The first step is that the user sends a photo to Line Bot. Then the message is obtained by the server with Webhook, and the system can download the user transmitted information (photos) and use the method proposed in this paper to calculate the data. Finally, the system sends the map link to Line Bot, and the user receives the result.

Figure 10 is the conversation between the user and Line Bot implemented in the IP-PFM system. After the user sends a photo to Line Bot, the system receives the photo through Webhook and saves it. The bot sends a message to inform the user that the image has been successfully received and calculation is started. Upon finishing the calculation, the bot uploads the URL that contains the user's location picture. Finally, it sends the link URL message to the user. The user could use the URL to obtain the building directory. Figure 11 shows the map that users can see after connecting to the link.

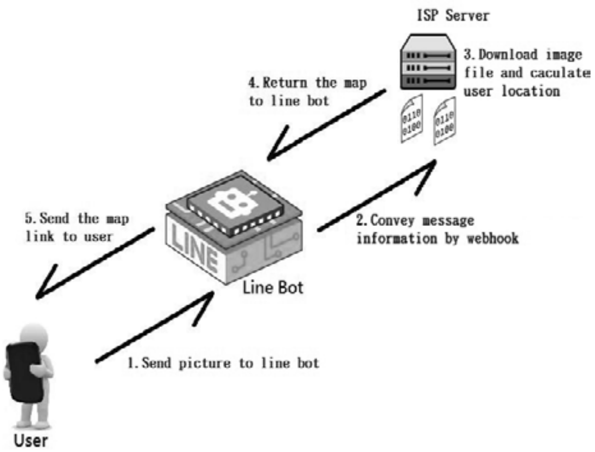


Figure 9. IP-PFM system architecture.

Figure 10 User and Line Bot conversation.

VI. CONCLUSION

This paper proposes a method, IP-PFM, for indoor positioning based on panoramic images and image recognition, which it combines with Line Bot. The complete procedure requires a user to input a photo of a scene, which is compared to the characteristics of panoramic images in the database. The system then estimates where the user has taken the photo and where the user is located by calculating the distance and direction between the user's input and the matching panoramic image. Finally, the system transmits the result back to the user by Line Bot.

This practical system can be customized according to local conditions. In the future, it may be possible to apply this method to robots to collect panoramic maps of buildings and 2D plane maps and to store these data directly in the computer system storage, which would be very useful for indoor positioning. In terms of computing speed, the experiment in this paper used only a central processing unit (CPU), and the processing was slow for image recognition. Adding a GPU may increase the processing speed so that users can access information faster.

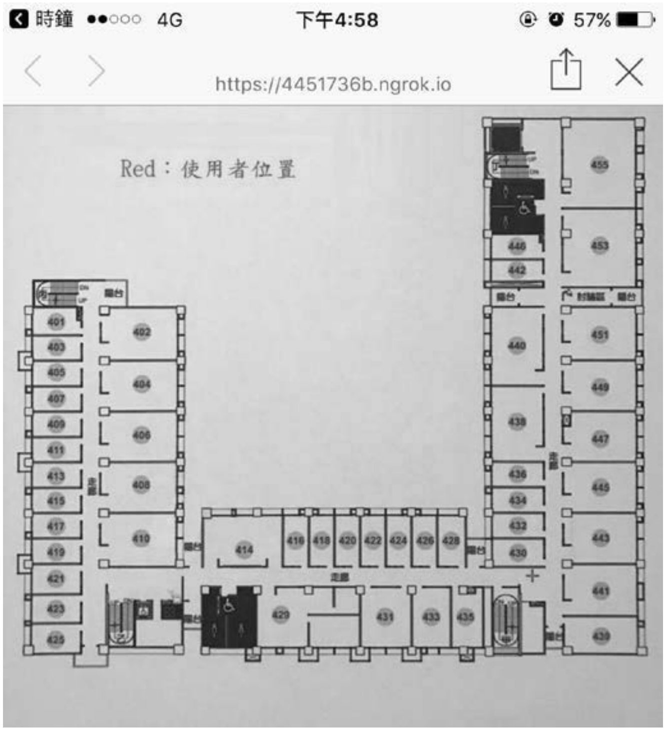


Figure 11 URL link image sent by the system.

ACKNOWLEDGEMENTS

This research was partially supported by Ministry of Science and Technology (MOST), Taiwan, R.O.C., under the grant number MOST 109-2221-E-005 -047 and 108-2221-E-194 -010 -MY2.

REFERENCES

[1] A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, C. Yuen, R. Raulefs, and E. Aboutanios, "Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications," IEEE Communications Surveys & Tutorials, vol. 19, no. 2, pp. 1327–1346, 2017.

[2] F. Zafari, A. Gkelias, and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2568–2599, 2019.

[3] F. Dwiyasa and M.-H. Lim, “A survey of problems and approaches in wireless-based indoor positioning,” 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Oct. 2016.
[4] H. An and N. Moon, “Image-based positioning system using LED Beacon based on IoT central management,” Multimedia Tools and Applications, Nov. 2020.
[5] M. Shahjalal, M. T. Hossan, M. K. Hasan, M. Z. Chowdhury, N. T. Le, and Y. M. Jang, “An Implementation Approach and Performance Analysis of Image Sensor Based Multilateral Indoor Localization and Navigation System,” Wireless Communications and Mobile Computing, vol. 2018, pp. 1–13, Oct. 2018.
[6] P. Pascacio, S. Casteleyn, J. Torres-Sospedra, E. S. Lohan, and J. Nurmi, “Collaborative Indoor Positioning Systems: A Systematic Review,” Sensors, vol. 21, no. 3, p. 1002, Feb. 2021.
[7] D. G. Lowe, “Object recognition from local scale-invariant features,” Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999.
[8] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-Up Robust Features (SURF),” Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346–359, Jun. 2008.
[9] M. Salarian, N. Ileiv, and R. Ansari, “Accurate Image Based Localization by Applying SFM and Coordinate System Registration,” 2016 IEEE International Symposium on Multimedia (ISM), Dec. 2016.
[10] Jianxiong Xiao, K. A. Ehinger, A. Oliva, and A. Torralba, “Recognizing scene viewpoint using panoramic place representation,” 2012 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2012.
[11] F. Molnar and A. Kovacs, “360-Degree image stitching with GPU support,” 2016 IEEE 20th Jubilee International Conference on Intelligent Engineering Systems (INES), Jun. 2016.
[12] H. Kawaji, K. Hatada, T. Yamasaki, and K. Aizawa, “Image-based indoor positioning system: fast image matching using omnidirectional panoramic images,” Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis - MPVA ’10, 2010.
[13] J.-Y. Huang, S.-H. Lee, and C.-H. Tsai, “A fast image matching technique for the panoramic-based localization,” 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), Jun. 2016.

UT-ATD: Universal Transformer for Anomalous Trajectory Detection by Embedding Trajectory Information

Yun Zhang*†, Nianwen Ning*†, Pengpeng Zhou*† and Bin Wu*†
*Beijing Key Laboratory of Intelligence Telecommunications Software and Multimedia, School of Computer Science
Beijing University of Posts and Telecommunications, Beijing, China
†{sigmarising, nianwenning, zhoupengpeng, wubin}@bupt.edu.cn

***Abstract*—Due to the development of the transportation industry, a large amount of trajectory data is pouring into the Internet all the time. Based on these trajectory data, anomalous trajectory detection technology provides great support for traffic safety assurance and traffic risk prediction. Most existing anomalous trajectory detection methods are based on trajectory's physical characteristics or representation learning, and they achieve good performance in a few scenarios. But they still face the following problems. (1) The imperfect utilization of trajectory points. (2) The sparsity of trajectory data, which leads to generalization issues. (3) Longer model training time consumed, which can't adapt to the large amount of trajectory data generated every day. To solve the above problems, we propose a novel anomalous trajectory detection model based on Universal Transformer, called UT-ATD. UT-ATD captures the information of trajectory positions by learning trajectory embedding for classification. UT-ATD has a faster training speed, relatively few model parameters, and sufficient portability, which are ideal for the realistic scene requirements. Our model achieves state-of-the-art performance in most aspects, and its effectiveness is verified by a series of experiments on the real-world taxi trajectory dataset.**

***Index Terms*—Anomalous Trajectory Detection, Trajectory Embedding, Universal Transformer**

I. Introduction

With the popularity of mobile smart terminals, numerous GPS trajectory information is generated in various scenarios. The trajectory information promotes the development of a large number of trajectory data mining studies, such as urban traffic navigation, urban construction planning, vehicle density monitoring, vehicle behavior prediction and other fields. Among them, anomalous trajectory detection becomes an increasingly important research direction.

Anomalous trajectory detection plays an essential role in two types of tasks. (1) To improve the service quality of taxis in the taxi field. Anomalous trajectory detection of taxis can prevent passengers from overpaying for deliberate bypasses by taxi drivers. It also allows taxi companies to respond in time and punish irresponsible drivers. (2) To detect the safety and unexpected status of transport vehicles in the field of transport tasks. For example, in remote prisoner escorts, anomalous trajectory detection can be used as an auxiliary detection method of the security system to assess the risk by determining whether the escort route meets the expectation. Anomalous trajectory detection can estimate whether the safety of valuables has been threatened in the transportation of extremely important items (such as museum relics).

DOI reference number: 10.18293/DMSVIVA2021-011

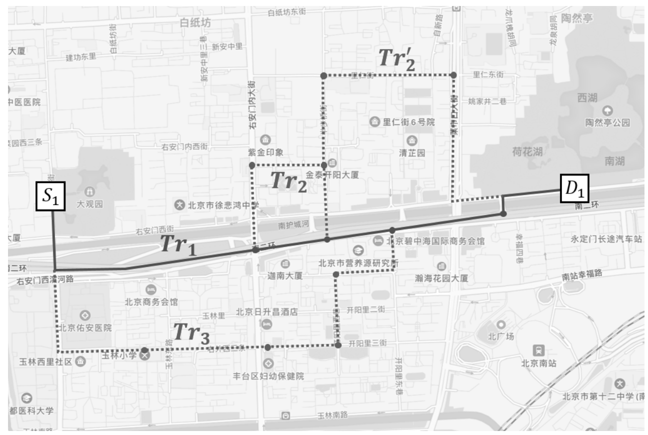


Fig. 1. The vital role of trajectory points information. Trajectory points play an important role in anomalous trajectory detection, as different anomalous trajectories can share the same anomalous trajectory points. Therefore, the trajectory points of one anomalous trajectory can make an important contribution to the detection of other anomalous trajectories.

In the real world, the anomalous trajectory detection task often needs to meet these requirements: (1) The model calculation speed is as fast as possible to fit the large amounts of trajectory data generated every moment. (2) The model should have good adaptation and generalization between different data sources. (3) The model is as portable as possible for broader deployment. In this paper, our target is to build an effective and flexible model to fit these requirements.

In the past years, there are many researches in anomalous trajectory detection, but they can't fit the real-world requirements well. The existing anomalous trajectory detection methods can be divided into two categories. The first category is based on trajectory's physical characteristics, such as density [1], direction [2], or isolation characteristics [3], [4] from the fragments of trajectories. However, these methods didn't take the importance of trajectory points into consideration. As it's shown in Fig. 1 that trajectories Tr_2 and Tr'_2 share the same red anomalous position.

The trajectory data sparsity is the main difficulty in anomalous trajectory detection. Because the number of trajectory points on the two-dimensional map are impossible to be summarized and recorded completely. With the increasing number of trajectory data, the methods based on trajectory's physical characteristics can cost too much in computing and have generalization issues. These methods based on trajectory's physical characteristics are seriously affected by the problem of trajectory data sparsity. The trajectory embedding by representation learning solve this problem to some extent.

The second category is based on trajectory embedding by representation learning. These methods [5], [6] map two-dimensional trajectories to one-dimensional sequences, and then use the models based on recurrent neural network(RNN) to encode the trajectory sequence into the low dimensional compact vectors. Compared to the first category methods, the process of trajectory embedding can make more use of the information of trajectory points. However, these methods stack to much RNN cells, which are time-consuming models during training. What's more, RNN can miss information when precessing long sequence [7]. In addition, these models usually have many parameters, which make them not portable enough and can not quickly adapt to the new trajectory data for self-training. For the above reasons, these methods can't apply well to the real world.

In this paper, we propose a novel model named **U**niversal **T**ransformer for **A**nomalous **T**rajectory **D**etection (**UT-ATD**), which is based on Universal Transformer. The model's self-attention mechanism makes the correlation degree of each trajectory point not decrease with the increase of sequence length. There are two entails substantial challenges in the real world anomalous trajectory detection. The first is the model needs to have enough effectiveness and probability in order to process the amount of trajectory data generated every day in different areas for self-training. The second is the model needs to have good generalization on the different data source to solve the trajectory data sparsity problem. To handle the above challenges, we utilize the self-circulation Universal Transformer, which uses the single Transformer block for trajectory embedding. We also change the trajectory points embedding layer with word2vec to make the model more scalable.

We carry out a series of experiments on a real-world taxi trajectory dataset, which has the highest sampling rate among all public datasets and better representation of real application scenarios. The experiments result showcase our model's better performance and generalization than other baselines.

The main contributions of this paper are listed as follows:

- We propose a novel anomalous trajectory detection model named UT-ATD, which utilizes the encoder of Universal Transformer as the core to learn the trajectory embedding.
- UT-ATD is effective and portable enough. It adapts to different data sources and has good generalization performance. The model's self-attention mechanism can encode trajectory points and trajectory sequence information better than existing methods. UT-ATD satisfies the real scenario requirements of anomalous trajectory detection task well.
- Experiments on the real-world taxi trajectory dataset verify the effectiveness of our proposed UT-ATD.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 defines the problem of anomalous trajectory detection. Section 4 introduces our model. Section 5 is the experiments and analysis. The summary and future work are shown in Section 6.

II. Related Work

In this section, we summary related work in two parts. Firstly, the research on anomalous trajectory detection based on trajectory's physical characteristics is summarized. Secondly, the techniques about trajectory embedding representation learning are introduced.

A. Anomalous Trajectory Detection Based on Trajectory's Physical Characteristics

The trajectory physical characteristics based detection methods mainly do calculation by classification, clustering or similarity [5], [8]. Zhu *et al.* [9], [10] mainly analyzed time-dependent outliers by the popular routes for each time interval. Lee *et al.* [1] divided the trajectories into sub-sequence and performed calculating based on distance and density. Knorr *et al.* [11] used large, multidimensional datasets based on distance to detect outliers. Li *et al.* [12] learned a model from multidimensional feature space-oriented on segmented trajectories. Kong *et al.* [13] did calculation based on the generated TS-segments. Zhang, Chen *et al.* [3], [4] used the isolation mechanism of anomalous trajectories. Zhang *et al.* [3] recorded a large number of historical trajectory and performed the adaptive iForest method for detection. Chen *et al.* [4] built an inverted index mechanism to get better retrieval performance. Zhang, Zhao *et al.* [14], [15] used the graph-based method for detection.

However, the detection methods based on distance, density, fragmentation didn't take the complete trajectory sequence information and positions context information into consideration. The calculation based on the historical database may suffer from a high time-consuming problem caused by trajectory data sparsity.

B. Trajectory Embedding By Representation Learning

Trajectory embedding is an extended field of word embedding [16]. Word Embedding is a good way to encode the information of words and sentences with a low dimension vector of fixed length. Zhao *et al.* [17] proposed a time-aware trajectory embedding model to deal with sequential information. Gao *et al.* [18] demonstrated that trajectory-user linking problem could be solved by trajectory embedding. Wu *et al.* [19] proposed a neural network algorithm based on spatial-temporal-semantic. In order to effectively capture the trajectory sequence information, many RNN based sequence models [5], [6], [20] are used for trajectory embedding training, and they achieved good results. However, for very

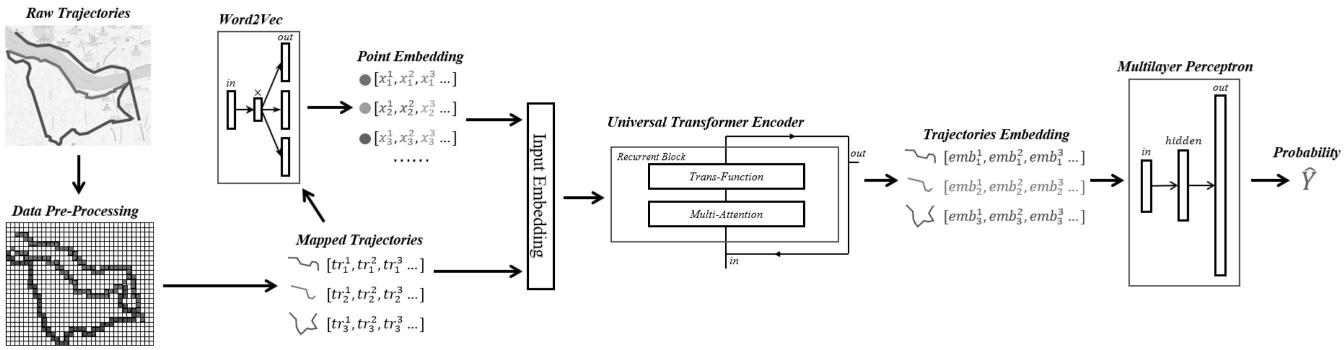


Fig. 2. The Architecture of UT-ATD. For all raw trajectories, get the mapped trajectories sequences by data pre-processing with f_{map} function. Then the word2vec(skip-gram) is used to pre-training the trajectory points embedding. Both mapped trajectories and points embedding are inputted into Universal Transformer Encoder by Input Embedding Layer. Then the output trajectory embedding is used for multilayer perceptron to calculate the probability.

long trajectory sequences, the RNN encoder cannot effectively capture all trajectory information [21]. And the training time-consuming of RNN is very high. Liu and Lane *et al.* [22] showed that the attention mechanism could improve the performance of RNN in sequence tasks. Vaswani *et al.* [7] proposed a network architecture based solely on attention mechanism named Transformer, which performed better and faster than RNN in sequence tasks. Dehghani *et al.* [23] upgraded the Transformer to Universal Transformer, which is Turing-complete and can be used in many fields such as video [24].

In order to obtain the complete trajectory information better and fully consider the role of positions in trajectory context, we use trajectory embedding by representation learning for anomalous detection. In this paper, the trajectory embedding is learned by Universal Transformer, which satisfies the real-world requirements of anomalous trajectory detection task.

III. Problem Definition

Definition 1. *Raw Trajectory Point.* The raw trajectory point tr^{pos} is the record information when GPS takes the sample, which is represented by

$$tr_i^{pos} = (id_i, time_i, lon_i, lat_i) \tag{1}$$

where (lon_i, lat_i) is the coordinate and $time_i$ is the timestamp. i denote the different raw trajectory point.

Definition 2. *Raw Trajectory.* Raw trajectory tr is a sequence of raw trajectory points collection.

$$tr_i = \left\{ tr_{i,1}^{pos} \rightarrow tr_{i,2}^{pos} \rightarrow ... \rightarrow tr_{i,len(tr_i)}^{pos} \right\} \tag{2}$$

where $tr_{i,1}^{pos}$ and $tr_{i,len(tr_i)}^{pos}$ is the source and destination of raw trajectory.

Definition 3. *Mapped Function.* For a given map, it can be divide equally into $m \times n$ grids. Every grid is represented by a unique $grid_{id}$. Mapped function f_{map} is used for convert raw trajectory point to $grid_{id}$ which it is located in.

$$f_{\text{map}}(map, m, n, tr^{pos}) = grid_{id} \tag{3}$$

where tr^{pos} is located in $grid_{id}$.

Definition 4. *Mapped Trajectory.* Mapped trajectory is a $grid_{id}$ sequence after performing mapped function on the specific raw trajectory. It's also the collection of *Mapped Trajectory point*. The mapped trajectory can be represented by equation

$$tr_i^{map} = \left\{ f_{\text{map}}(map, m, n, tr_{i,k}^{pos}) \middle| k = 1 \ldots len(tr_i) \right\} \tag{4}$$

Problem Statement. *Anomalous Trajectory Detection.* For a given collection of trajectories $A = \left\{ tr_1, tr_2, ...tr_{len(A)} \right\}$. Anomalous Trajectory Detection is to find out those trajectories B that are significantly different(according to the hierarchical clustering result analysis based on the similarity between trajectories) from the majority in historical datasets, where $B \subset A$.

IV. Methodology

A. Overview of UT-ATD

The architecture of UT-ATD can be found in Fig. 2. The workflow of UT-ATD is shown in Workflow 1.

The whole model can be divided into three parts: data pre-processing, trajectory embedding, and anomalous trajectory detection. In the process of data pre-processing, the two-dimensional trajectory points sequences (raw trajectory) are discrete to one-dimensional sequences (mapped trajectory). Then, we use word2vec [25] to pre-train the embedding of each mapped trajectory points. After that, the universal transformer encoder is applied to learn the embedding of mapped trajectories. Finally, a multilayer perceptron(MLP) is used to detect anomalous trajectories from trajectory embedding. In practice, the data pre-processing part is carried out separately. The Universal Transformer part and multilayer perceptron part execute joint training together.

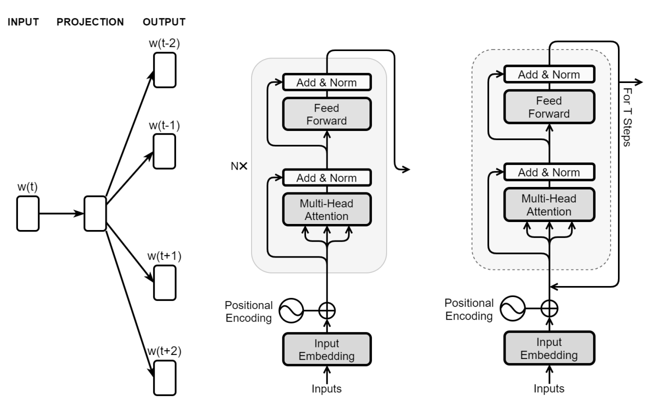
B. Data Pre-Processing

The original raw trajectory data are continuous numerical variables on the two-dimensional level. However, the number of points on the map is uncountable. The calculation cost and space cost brought by learning from raw trajectories are too expensive. The discrete step is needed to reduce the original data's dimension while retaining the complete serialization information. In detail, the map is divided into a separate grid of $m \times n$. Through the mapping function, all points in the same grid will be given the same $grid_{id}$. The missing points in the trajectory are padded to obtain a continuous trajectory sequence by using the pre-processing method provided in [3]. We find that $100m \times 100m$ is the optimal size of the grid through repeated experiments. Both the padding and the masking techniques are applied to mapped trajectories as well. In fact, anomalous trajectories rarely appear in the historical database. In order to avoid excessive influence on the optimization of model training, we add interference data(random select the grid in sequence, and replace it with its geographic neighbor grid on the map) to each anomalous trajectory to generate some negative samples for training.

Next, the word2vec method is used to get the pre-trained embedding of each trajectory point, in order that the Universal Transformer can have a good learning effect. In practice, we choose skip-gram model for pre-training step. The model of skip-gram is shown in Fig. 3(a). The loss function is:

$$J(\theta) = -\frac{1}{T}\sum^{T}_{c=1,-m\leq j\leq m, j\neq 0}\sum \log p(w_{c+j} \big| w_c;\theta) \tag{5}$$

where θ is the set of all parameters, T is the number of words in the entire corpus, m is the radius of the context window and w is the word.



(a) Skip-Gram (b) Transformer Encoder (c) Encoder of Universal Transformer

Fig. 3. Model of skip-gram and Encoder of Transformer

C. Trajectory Embedding

The encoder part of Universal Transformer is used to learn the embedding of mapped trajectory. Trajectory embedding can be learned by providing the complete mapped trajectory sequence and pre-embedding of mapped trajectory points to the encoder. The information of the trajectory points will be fully utilized in the process of calculating the trajectory embedding.

1) Transformer: Transformer is inspired by the attention mechanism. It completely replaces the RNN structure with multi-head attention and feed forward networks, which are faster and more effective than RNN based approach [7]. Transformer consists of an encoder part and a decoder part. In this paper, the encoder part is used to learn the mapped trajectory sequences and key position in trajectory context. The structure and calculation process of Transformer's encoder is shown in Fig. 3(b). Each part of this encoder will be explained below.

Scaled dot product attention. This is the main calculation logic in multi-head attention. The input consists of queries and keys of dimension d_k, and values of dimension d_v. The queries, keys, and values are packed separately as matrices Q, K and V. The calculation equation is shown below:

$$\mathrm{Attn}(Q,K,V) = \mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{6}$$

Multi-head attention. This mechanism helps model learn sequence information better. And the calculation is as follows.

$$\begin{gathered}\mathrm{MultiHead}(Q,K,V) = \mathrm{Concat}(head_1,...,head_h)W^O \\ \text{where } head_i = \mathrm{Attn}(QW_i^Q, KW_i^k, VW_i^V)\end{gathered} \tag{7}$$

where h is the number of multi-head, d_{model} is the outputs dimension. The projections are parameter matrices $W_i^Q \in \mathbb{R}^{d_{\text{model}}\times d_k}$, $W_i^K \in \mathbb{R}^{d_{\text{model}}\times d_k}$, $W_i^V \in \mathbb{R}^{d_{\text{model}}\times d_v}$ and $W^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$.

Feed-forward network. The representation is:

$$\mathrm{FFN}(x) = W_2 \cdot \mathrm{ReLU}(W_1 \cdot x + b_1) + b_2 \tag{8}$$

where W_1, W_2, b_1, b_2 are parameters of this feed-forward network. And x is the input of the network.

Positional encoding. Transformer adds the positional encoding function to get the position information of tokens.

$$\mathrm{PE}(pos, 2i) = \sin(pos/10000^{2i/d_{model}}) \tag{9}$$

$$\mathrm{PE}(pos, 2i+1) = \cos(pos/10000^{2i/d_{model}}) \tag{10}$$

where pos is the position of the current word in the sentence, i is the index of each value in the vector. Sine encoding is used in the even position, and cosine encoding is used in the odd position.

Add&Norm. The residual connection is used to ameliorate the problem of gradient disappearance in the model [26]. The output of input x after normalization is $\mathrm{LayerNorm}(x + \mathrm{SubLayer}(x))$, where $\mathrm{SubLayer}(x)$ is the function implemented by sub-layer.

2) Universal Transformer: Universal Transformer is the updated version of Transformer. It uses recursive function to allow the number of the layers of Transformer can vary at will. Compared with vanilla Transformer, Universal Transformer is based on the self-circulation mechanism, which has lower model complexity, better generalization ability, lower cost of parameters, more portability and mobility. Its encoder is shown in Fig. 3(c). An Adaptive Computation Time (ACT) [27] based dynamic halting algorithm is used to optimize computing speed.

After the calculation of Universal Transformer, the output is stacked to form the mapped trajectory embedding (refer to *trajectory embedding* in following paper). All points in trajectory are involved in the calculation of trajectory embedding.

$$Emb_i = \text{Concat}(Y_{i,1}, Y_{i,2}, ...Y_{i,MTL}) \tag{11}$$

$$Y_{i,j} = [y_{i,j}^1, y_{i,j}^2, ..., y_{i,j}^{d_{model}}] \tag{12}$$

where Emb_i is the embedding of ith mapped trajectory, and MTL is the maximum trajectory length in dataset. $Y_{i,j}$ is the output of the jth point in ith mapped trajectory in Universal Transformer.

D. Anomalous Detection

In order to detect anomalous trajectory, a MLP for classification is designed.

$$\begin{aligned} y_i =& \sigma(W_2 \cdot \sigma(W_1 \cdot Emb_i + b_1) + b_2) \\ J_{BCE} =& -\frac{1}{n}\sum_{i}^{n}(y_i^{'} \log y_i + (1 - y_i^{'}) \log(1 - y_i)) \end{aligned} \tag{13}$$

where W_1, W_2, b_1, b_2 is parameters of MLP, n is the number of samples, y is the output of MLP, and $y^{'}$ is the ground truth. For a given dataset $A = \{tr_1, tr_2, ...tr_{len(A)}\}$, the Binary Cross Entropy(BCE) loss function can be minimized to conduct joint training for the model. Dropout techniques [28] is used here to avoid over fitting problem.

E. The Workflow of UT-ATD

Workflow 1 Our UT-ATD

Input: Raw trajectory collection A

Output: The trained model UT-ATD

1: For each raw trajectory $tr_i \in A$, get the mapped trajectory set $A^{map} = \{tr_i^{map} | i \in [0..\text{len}(A)]\}$ after the calculation of f_{map}. Then pad the missing points for each tr_i^{map}.

2: Get the pre-embedding matrix Emb_{pre}^{point} of each mapped trajectory point by training word2vec on A^{map}.

3: Get the maximum length K of mapped trajectories in A^{map}. Then append the 0 mark to each tr_i^{map} until its length is K. And the same length mask sequence $Mask_i$ is used to record the 0 mark information of tr_i^{map}. $Mask = \{Mask_i...Mask_{\text{len}(A)}\}$.

4: Input Emb_{pre}^{point}, A^{map}, and $Mask$ to Universal Transformer Encoder. For each tr_i^{map}, the output is trajectory embedding Emb_i.

5: The MLP obtain Emb_i as input. The output is the probability of whether tr_i is anomalous.
Note. The Universal Transformer Encoder and MLP execute joint training.

V. Experiments

A. Evaluation

To evaluate the performance of each model, Accuracy(Acc), F_1 are used as performance evaluation criteria. The evaluations are all carried out on trajectories level.

B. Dataset

We carried out experiments on the read-world taxi trajectory dataset, which is collected from 442 taxis in Porto, Portugal from Jan. 07, 2013 to Jun. 30, 2014. The average sampling rate of GPS is $15s/point$, which is the highest sampling rate among all public datasets. This dataset can better reflect the real scene of anomalous trajectory detection. We extract five source-destination pairs (refer to *sd-pair*, *sdPair* or *sdp*) with sufficient historical trajectories data for training and testing.

Existing works tend to label the outlier manually, but the solution from [29] provided a good way to do this automatically. This methods adopts a complete-linkage clustering algorithm to hierarchically cluster the trajectories, as for outliers are "few" and "different". This method can effectively annotate the dataset. But this automatic annotating method takes a week long to extract 5 sd-pairs, which is time-consumed expensive and unbearable in real-world application as for large amount of new trajectory data is generated every day. Therefore, we only use it as a method to annotate data. The information of dataset is shown in Table I.

TABLE I
THE INFORMATION OF DATASET

dataset	sdp1	sdp2	sdp3	sdp4	sdp5
trajectories	1233	765	617	1379	4973
anomalous	54	28	37	44	270
avg trajectory len	32	31	61	51	67
max trajectory len	95	74	187	256	321
min trajectory len	17	13	42	35	40
trainset	1150	693	537	1247	4565
trainset anomalous	29	16	21	22	132
testset	83	72	80	132	408
testset anomalous	25	12	16	22	138

C. Baselines

We compare our method with the below baselines.

LCS [30]: The Longest Common Sub-sequence mechanism is a widely used method for measuring trajectory similarity. We implement LCS by comparing all trajectories in training set for every given testing set.

XGBoost [31]: This is an efficient, flexible and portable model based on gradient enhancement decision tree.

TOP-EYE [2]: TOP-EYE uses a decay function to mitigate the influences of historical trajectories, which is based on density and moving direction. We conduct this method by counting the density of each grid and compute abnormal score for test data.

LoTAD [13]: LoTAD consists of TS-segments creation and anomaly index computation. The anomaly index is computed through the density of each trajectory points.

iBOAT [4]: Anomalous trajectories will be isolated from the majority of historical dataset. iBOAT uses the inverted index mechanism to fast retrieve the relevant trajectories.

ATD-RNN [5]: This is our main compared target. In practice, it has ATD-LSTM and ATD-GRU. We test both of them in single sd-pair test. In other tests, we use ATD-LSTM as its default.

D. Results and Analysis

1) Single Source-destination Pair: The proposed model is implemented by pytorch, and set transformer's d_{model} to 64, number of multi-head-attention to 4, transformer layers to 8. The MLP's dimension is 128 and the dropout probability is 0.5. AdamW [32] is used to optimize the model.

We run our model and above baselines on different sd-pairs, the results are shown in Table II.

TABLE II
THE RESULT ON SINGLE SD-PAIR

dataset		sdPair1	sdPair2	sdPair3	sdPair4	sdPair5
LCS	Acc	0.8434	0.9444	0.9625	0.9242	0.7819
	F_1	0.7797	0.8182	0.9032	0.7059	0.5340
XGB	Acc	0.8795	0.8810	0.8375	0.9470	0.7574
	F_1	0.7619	0.4444	0.4348	0.8205	0.4469
TOPEYE	Acc	0.9629	0.9444	0.9625	0.9470	0.8431
	F_1	0.9230	0.8571	0.9032	0.8444	0.8118
LoTAD	Acc	0.6747	0.7639	0.8125	0.8333	0.6593
	F_1	0.8029	0.8661	0.8951	0.9091	0.7935
iBOAT	Acc	0.9506	**0.9583**	0.9625	0.9394	0.8750
	F_1	0.9167	0.8696	0.8276	0.7778	0.8198
ATD-LSTM	Acc	0.9518	**0.9583**	**0.9875**	0.9848	0.9020
	F1	0.9167	0.8696	0.9697	0.9545	0.8374
ATD-GRU	Acc	0.9638	**0.9583**	0.9750	**0.9924**	0.9167
	F1	0.9412	0.8800	0.9412	0.9767	0.8722
UT-ATD	Acc	**0.9880**	**0.9583**	0.9750	**0.9924**	**0.9191**
	F1	**0.9915**	**0.9752**	**0.9841**	**0.9954**	**0.9430**

The relatively low Acc scores of LCS and XGBoost may due to the fact that these methods only consider the shape information of the trajectory and ignore the historical sequences information of trajectories. The LoTAD's result low may because it lacks the information of complete sequence. TOP-EYE and iBOAT achieve considerable performance may because they use historical trajectories data for calculation. iBOAT out-performs TOP-EYE in most circumstances may be that iBOAT not only takes the similarity into consideration but also makes use of the local sequential information.

It can be seen that UT-ATD achieves the highest accuracy on most sd-pairs and gets the best F_1 on all sd-pairs. It proves that UT-ATD is superior to ATD-RNN in its ability to obtain information from trajectory. It also shows that the information of different locations in trajectory context plays a very important role in trajectory embedding. One of the reasons why on sd-pair3 have the minimum amount of data, and UT-ATD is not trained enough. But in contrast, UT-ATD achieves a better F_1 on sd-pair3 than ATD-RNN.

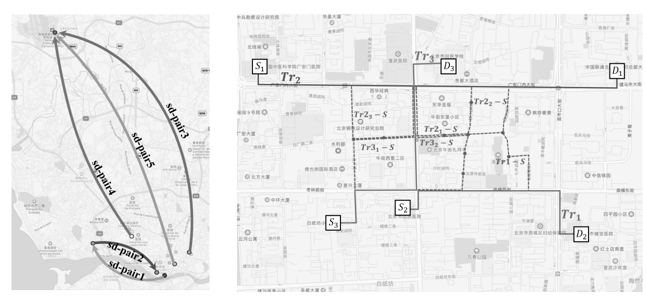
TABLE III
THE RESULT ON MULTI-SD-PAIR

dataset	ATD-RNN		UT-ATD	
	Acc	F1	Acc	F1
sdPair3+4	0.9340	0.9611	**0.9528**	**0.9714**
sdPair3+5	0.8996	0.9299	**0.9262**	**0.9459**
sdPair4+5	0.9148	0.9415	**0.9315**	**0.9520**
sdPair3+4+5	0.9145	0.9422	**0.9339**	**0.9550**
sdPair1+2+3+4+5	0.9123	0.9413	**0.9406**	**0.9598**

2) Multi-source-destination Pair: Through trajectory embedding, the extensibility of calculation among different sd-pairs is possible. Because the trajectory sequences between different sd-pairs can complement each other, the pressure of data sparsity can be alleviated to some extent [5]. We pack different sd-pairs together to make up multi-sd-pair. Then we compare UT-ATD with ATD-RNN on different multi-sd-pairs.

The geographic relationship of five sd-pairs can be seen in Fig. 4(a). What stands out in this figure is that sd-pair1 and sd-pair2 are close to each other. And sd-pair3, sd-pair4, and sd-pair5 are close to each other. The test results on combination of different sd-pairs are shown in Table III.

It can find out that, for geographical closed or not closed sd-pairs combination, UT-ATD's generalization ability is better than ATD-RNN. The probable reason is that the key positions in one sd-pair offer information for other sd-pairs. As shown in Fig. 4(b), the key positions(red) in $Tr3$-S will offer information for detecting $Tr1$-S and $Tr2$-S. For lots of very long trajectories, the information key positions offered will be more. Another possible reason is the RNN's poor ability to encode long trajectories, as for the trajectories in sd-pair3, sd-pair-4 and sd-pair5 are longer than others.



(a) Relationship of five SD-Pairs (b) Key Position Impact for Multi-Trajectories

Fig. 4. Validity of Word2vec and Universal Transformer

3) Training Speed Comparison: To compare the training speed of ATD-RNN and UT-ATD, we train both models with epoch(40), batch-size(32), mapped trajectory points pre-embedding dimension(64), and layers of neural network(6) on Linux server(Intel Xeon Gold 5118, 250G RAM, NVIDIA RTX 2080 Ti) and take average time for multiple runs.

The result can be seen from Fig. 5(a) that when training on sd-pair1 to sd-pair5, the training efficiency of UT-ATD is significantly faster than ATD-RNN. This is mainly because ATD-RNN must follow the sequence order as trajectory embedding is computed. The next state can only be computed after the calculation of the previous state. This doesn't constrain UT-ATD, and it can compute all information in parallel. So UT-ATD naturally has a faster speed than RNN based method in training.

(a) Training Time Comparison (b) Number of d_{model}

(c) Number of Multi-head-attention (d) Number of layers

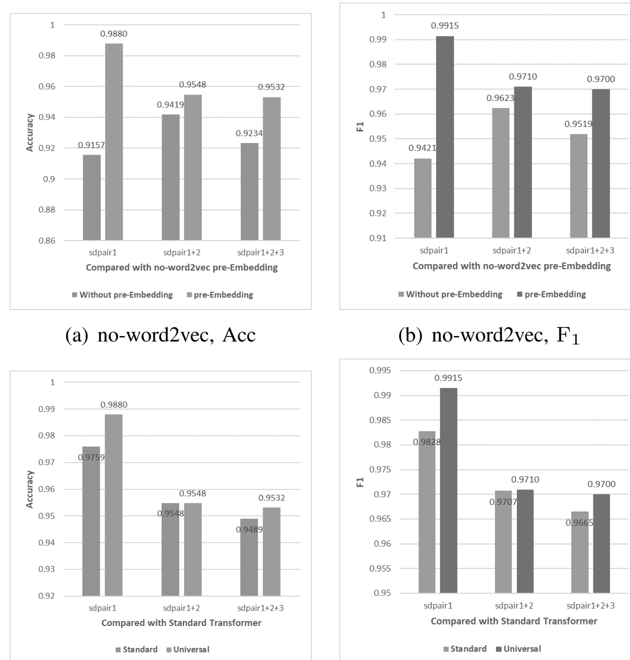
Fig. 5. Training Time and Parameter Influence

4) Parameter Experiments: To explore the influence of different parameters, we test the influence of d_{model}, number of multi-head-attention, and layers for UT-ATD on sd-pair1.

It can be seen from Fig. 5(b) that, with the increase of d_{model}, the accuracy and F_1 will increase and tend to be stable. The reason is that with the increase of trajectory embedding dimension, the represented information will increase. When $d_{model} = 64$, the information is saturated and there is no need to increase d_{model}. The Fig. 5(c) shows that accuracy and F_1 will fluctuates as num of multi-head-attention grows. One possible reason is that the length of trajectories in sd-pair1 is short and the effect of multi-head-attention to neutralize effective resolution is not obvious. Fig. 5(d) shows that the more number of layers, the lower performance will get. The reason is that the increase of layers will make model more complex, which may lead to over-fitting problem.

5) Ablation Experiments: We test the UT-ATD without word2vec pre-training (the pre-embedding is randomly initialized and adjusted during training) and compare UT-ATD to standard transformer on different scale sd-pairs.

As it's shown in Fig. 6(a) and Fig. 6(b) that the pre-training of word2vec do help the model to perform better. Fig. 6(c) and Fig. 6(d) show that Universal Transformer performs better than standard transformer, which may because the Universal Transformer is Turing-complete and more portability.



(a) no-word2vec, Acc (b) no-word2vec, F_1

(c) stand-transformer, Acc (d) stand-transformer, F_1

Fig. 6. Validity of Word2vec and Universal Transformer

VI. Conclusions and Future Work

In this paper, we propose an anomalous trajectory detection method **U**niversal **T**ransformer for **A**nomalous **T**rajectory **D**etection (**UT-ATD**). UT-ATD uses the Universal Transformer encoder to learn the embeddings of the trajectory effectively. Compared with traditional methods and RNN-based model, UT-ATD not only captures the information of complete trajectory sequence better but also considers the importance of positions in trajectory context. UT-ATD is not limited by the given source-destination pairs. It performs well in multiple combined source-destination pairs. The experiments on the real-world taxi trajectories dataset demonstrate the effectiveness of UT-ATD.

In the future, we will extend the UT-ATD for online detection and study the influence of few-shot learning on our model. Whether multi-modal based method has a good effect on anomalous trajectory detection is also worth discussing.

Acknowledgment

This work is supported by National Key Research and Development Program of China (2018YFC0831500), National Natural Science Foundation of China under Grant No.61972047, and NSFC-General Technology Basic Research Joint Funds under Grant U1936220.

References

[1] J. Lee, J. Han, and X. Li, "Trajectory outlier detection: A partition-and-detect framework," in *Proceedings of the 24th International Conference on Data Engineering, ICDE 2008, April 7-12, 2008, Cancún, Mexico*. IEEE Computer Society, 2008, pp. 140–149.

[2] Y. Ge, H. Xiong, Z. Zhou, H. T. Ozdemir, J. Yu, and K. C. Lee, "Top-eye: top-k evolving trajectory outlier detection," in *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM 2010, Toronto, Ontario, Canada, October 26-30, 2010*. ACM, 2010, pp. 1733–1736.

[3] D. Zhang, N. Li, Z. Zhou, C. Chen, L. Sun, and S. Li, "ibat: detecting anomalous taxi trajectories from GPS traces," in *UbiComp 2011: Ubiquitous Computing, 13th International Conference, UbiComp 2011, Beijing, China, September 17-21, 2011, Proceedings*. ACM, 2011, pp. 99–108.

[4] C. Chen, D. Zhang, P. S. Castro, N. Li, L. Sun, S. Li, and Z. Wang, "iboat: Isolation-based online anomalous trajectory detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 806–818, 2013.

[5] L. Song, R. Wang, D. Xiao, X. Han, Y. Cai, and C. Shi, "Anomalous trajectory detection using recurrent neural network," in *Advanced Data Mining and Applications - 14th International Conference, ADMA 2018, Nanjing, China, November 16-18, 2018, Proceedings*, ser. Lecture Notes in Computer Science, vol. 11323. Springer, 2018, pp. 263–277.

[6] Y. Cheng, B. Wu, L. Song, and C. Shi, "Spatial-temporal recurrent neural network for anomalous trajectories detection," in *Advanced Data Mining and Applications - 15th International Conference, ADMA 2019, Dalian, China, November 21-23, 2019, Proceedings*, ser. Lecture Notes in Computer Science, vol. 11888. Springer, 2019, pp. 565–578.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, 2017, pp. 5998–6008.

[8] A. Belhadi, Y. Djenouri, and J. C. Lin, "Comparative study on trajectory outlier detection algorithms," in *2019 International Conference on Data Mining Workshops, ICDM Workshops 2019, Beijing, China, November 8-11, 2019*. IEEE, 2019, pp. 415–423.

[9] J. Zhu, W. Jiang, A. Liu, G. Liu, and L. Zhao, "Time-dependent popular routes based trajectory outlier detection," in *Web Information Systems Engineering - WISE 2015 - 16th International Conference, Miami, FL, USA, November 1-3, 2015, Proceedings, Part I*, ser. Lecture Notes in Computer Science, vol. 9418. Springer, 2015, pp. 16–30.

[10] ——, "Effective and efficient trajectory outlier detection based on time-dependent popular route," *World Wide Web*, vol. 20, no. 1, pp. 111–134, 2017.

[11] E. M. Knorr, R. T. Ng, and V. Tucakov, "Distance-based outliers: Algorithms and applications," *VLDB J.*, vol. 8, no. 3-4, pp. 237–253, 2000.

[12] X. Li, J. Han, S. Kim, and H. Gonzalez, "ROAM: rule- and motif-based anomaly detection in massive moving object data sets," in *Proceedings of the Seventh SIAM International Conference on Data Mining, April 26-28, 2007, Minneapolis, Minnesota, USA*. SIAM, 2007, pp. 273–284.

[13] X. Kong, X. Song, F. Xia, H. Guo, J. Wang, and A. Tolba, "Lotad: long-term traffic anomaly detection based on crowdsourced bus trajectory data," *World Wide Web*, vol. 21, no. 3, pp. 825–847, 2018.

[14] J. Zhang, "Smarter outlier detection and deeper understanding of large-scale taxi trip records: a case study of NYC," in *Proceedings of the ACM SIGKDD International Workshop on Urban Computing, UrbComp@KDD 2012, Beijing, China, August 12, 2012*. ACM, 2012, pp. 157–162.

[15] X. Zhao, Y. Rao, J. Cai, and W. Ma, "Abnormal trajectory detection based on a sparse subgraph," *IEEE Access*, vol. 8, pp. 29 987–30 000, 2020.

[16] Q. V. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, ser. JMLR Workshop and Conference Proceedings, vol. 32. JMLR.org, 2014, pp. 1188–1196.

[17] W. X. Zhao, N. Zhou, A. Sun, J. Wen, J. Han, and E. Y. Chang, "A time-aware trajectory embedding model for next-location recommendation," *Knowl. Inf. Syst.*, vol. 56, no. 3, pp. 559–579, 2018.

[18] Q. Gao, F. Zhou, K. Zhang, G. Trajcevski, X. Luo, and F. Zhang, "Identifying human mobility via trajectory embeddings," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*. ijcai.org, 2017, pp. 1689–1695.

[19] F. Wu, K. Fu, Y. Wang, Z. Xiao, and X. Fu, "A spatial-temporal-semantic neural network algorithm for location prediction on moving objects," *Algorithms*, vol. 10, no. 2, p. 37, 2017.

[20] Y. Liu, K. Zhao, G. Cong, and Z. Bao, "Online anomalous trajectory detection with deep generative sequence modeling," in *36th IEEE International Conference on Data Engineering, ICDE 2020, Dallas, TX, USA, April 20-24, 2020*. IEEE, 2020, pp. 949–960.

[21] G. Tang, M. Müller, A. Rios, and R. Sennrich, "Why self-attention? A targeted evaluation of neural machine translation architectures," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*. Association for Computational Linguistics, 2018, pp. 4263–4272.

[22] B. Liu and I. Lane, "Attention-based recurrent neural network models for joint intent detection and slot filling," in *Interspeech 2016, 17th Annual Conference of the International Speech Communication Association, San Francisco, CA, USA, September 8-12, 2016*. ISCA, 2016, pp. 685–689.

[23] M. Dehghani, S. Gouws, O. Vinyals, J. Uszkoreit, and L. Kaiser, "Universal transformers," in *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net, 2019.

[24] M. Bilkhu, S. Wang, and T. Dobhal, "Attention is all you need for videos: Self-attention based video summarization using universal transformers," *CoRR*, vol. abs/1906.02792, 2019.

[25] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," in *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, 2013.

[26] L. J. Ba, J. R. Kiros, and G. E. Hinton, "Layer normalization," *CoRR*, vol. abs/1607.06450, 2016.

[27] D. Fojo, V. Campos, and X. Giró-i-Nieto, "Comparing fixed and adaptive computation time for recurrent neural networks," in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Workshop Track Proceedings*. OpenReview.net, 2018.

[28] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[29] H. Wu, W. Sun, and B. Zheng, "A fast trajectory outlier detection approach via driving behavior modeling," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*. ACM, 2017, pp. 837–846.

[30] J. J. Ying, W. Lee, T. Weng, and V. S. Tseng, "Semantic trajectory mining for location prediction," in *19th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems, ACM-GIS 2011, November 1-4, 2011, Chicago, IL, USA, Proceedings*. ACM, 2011, pp. 34–43.

[31] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*. ACM, 2016, pp. 785–794.

[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

Integration Strategies for the Personas technique within Agile Process

Patricia Losana
Depto. Ing. Informática, Escuela Politécnica Superior
Universidad Autónoma de Madrid
Madrid, Spain
patricia.losana@estudiante.uam.es

John W. Castro*
Depto. Ing. Informática y Ciencias de la Computación
Universidad de Atacama
Copiapó, Chile
john.castro@uda.cl

Silvia T. Acuña
Depto. Ing. Informática, Escuela Politécnica Superior
Universidad Autónoma de Madrid
Madrid, Spain
silvia.acunna@uam.es

***Abstract*—Agile development processes are increasing their consideration over usability, integrating different user-centered design techniques throughout development. One such technique is Personas, which proposes the design of fictitious users with real preferences to drive application design. As applying this technique conflicts with the time constraints of agile development, Personas has been adapted over the years. Our objective is to determine the adoption level and type of integration, as well as to propose improvements to the Personas technique for agile development. A systematic mapping study was performed, retrieving 28 articles which we grouped by agile methodology type. We found some common integration issues irrespective of the type of agile process, such as the life cycle stage or the target user analysis, and also some frequent problems, mainly related to project timing and representing the Persona context. Based on these difficulties, we propose possible improvements, such as using templates for a preliminary Persona. This study analyzed the different integration strategies for the Personas technique within agile processes, evaluating the difficulties found and proposing solutions. The number of publications is increasing, which reveals a growing interest in the adoption of this technique to develop usable agile software.**

Keywords—Personas; User Profiling; Human-Computer Interaction; User-Centered Design; Agile Methodology; Software Engineering; Systematic Mapping Study

I. INTRODUCTION

Usability is a software quality characteristic used in most of the classifications [1][2], which must be addressed throughout the entire interactive software development process [3][4]. Usability and User-Centered Design (UCD) techniques have been studied with increasing interest by development teams using agile software development processes (ASDP), as it has been shown that using only agile methods does not necessarily guarantee product usability [5]. To develop a usable software system, it is required to know how are the users to whom the system is destined [6][7]. There are several techniques within the Human-Computer Interaction (HCI) discipline that study and model the person that will use the software system, among which can be found the Personas technique [8], which has achieved promising results in software development [9]. ASDP have adopted 52% of the usability techniques related to requirements engineering. Of this group, Personas is the most used technique [10]. This user analysis technique is based on designing a user model from data obtained through interviews with real users, guiding the application design with the users' preferences and avoiding that the developers create a design based on their own inclinations.

The Personas technique has been systematized at the same level as the software engineering (SE) techniques through the works of [11] and [12]. Later, in the study performed by [13], the technique has been adapted to be integrated within an agile development process and has later been evaluated through a case study by [14], which allowed to test the viability and impact of applying the Personas technique adapted within a real agile project.

The next step in this line of research corresponds to analyze the state of the art of the incorporation of the Personas technique within agile processes, in order to establish how this technique is being used within agile projects and to identify potential improvements for it. For this purpose, the research work aims to identify the different integration approaches of the Personas technique through a literature review carried out by a Systematic Mapping Study (SMS), using as a reference the guidelines described in the study of [15] and [16] due to its great representativeness within the studies of HCI and SE. The result of the SMS is reported in the present research work.

Paper organization. In Sec. 2, we describe the Personas Technique. In Sec. 3, we present related work. In Sec. 4, we describe the research method of the SMS. In Sec. 5 we discuss the results of the SMS. Sec. 6 presents possible threats to validity, and finally, the conclusions are presented in Sec. 7.

II. THE PERSONAS TECHNIQUE

The Personas technique, attributed to Alan Cooper [17], is a UCD tool that seeks to conceptualize the behavior of real users within user models, with the objective of improving the usability of the design. In this way, although a persona is fictitious, the objectives it aims to cover are real, since they are synthesized from observations of final users. This allows the design and development team to empathize more easily with user preferences [18]. The following is a description of the steps that make up the Personas technique of Cooper et al. [9]:

- Step 1: *Identify Behavioural Variables*. In this step the different behavioral aspects observed are listed. This list is known as the set of behavioral variables.

* Corresponding Author.

DOI reference number: 10.18293/DMSVIVA2021-012

- Step 2: *Map Interview Subjects to Behavioural Variables*. Once the list of behavioral variables obtained from the subject interviews has been identified, the next step is to map each interviewed subject to each behavioral variable. These behavioral variables represent either a continuous range of behaviors or multiple discrete selections.
- Step 3: *Identify Significant Behavioural Patterns*. In this step, groups of subjects that fall into multiple ranges or variables are observed. A group of subjects grouping six to eight different variables could represent a significant pattern of behavior, which will form the basis of a persona.
- Step 4: *Synthesize Characteristics and Relevant Goals*. In this step, the details of each of the significant behavior patterns identified in the previous step are synthesized. This synthesis should describe the potential use environment, a typical workday, and relevant relationships with others. Brief statements describing the characteristics of the behaviors are sufficient for this synthesis.
- Step 5: *Check for Completeness and Redundancy*. In this step, the mappings, the characteristics of the personas and their objectives are checked to determine if there are any gaps that need to be filled. It is important to make sure that the set of personas is complete, and that each persona is significantly different from the others.
- Step 6: *Expand the Description of Attributes and Behaviours*. Third-person narrative is very useful for conveying the attitudes, needs, and problems of the persona to other team members. A typical description should be a synthesis of the most important details observed during the investigation that are relevant to the persona.
- Step 7: *Designate Persona Types*. Development requires a target. Typically, the most specific goal is the best. The goal is to find a single persona whose needs and objectives can be completely satisfied by a single interface without disappointing any of the other personas. This is accomplished through a process of designating persona types. There are six types of personas: (i) primary, (ii) secondary, (iii) supplemental, (iv) customer, (v) served, and (vi) negative.

It is important to mention that the application of the Personas Technique impacts the use cases, since a persona model is attached to each actor. The annotated use case diagrams and the use case specification provide project stakeholders with a model providing a common understanding for deciding what the software system should do according to each persona.

III. Related Work

Even though the Personas technique belongs to the HCI branch and not to the agile methodology, Personas has been sought to be used in agile processes to help development teams to have a better design [19]. However, the original idea of Personas may conflict with the agile philosophy in the process of obtaining details from end users regarding the system [20].

Although there are different agile methodologies, all of them are characterized by being iterative, promoting developer-client collaboration, and receiving feedback from the client throughout the development life cycle. The most relevant methodologies are: Dynamic Systems Development (DSDM) [21], Functionality Driven Development (FDD) [22], Lean Software Development [23], Scrum [24] and eXtreme Programming (XP) [25]. The agile philosophy is characterized by evaluating the functionality of prototypes with users over short iterations, which tends to have a direct impact on the usability of the design. Therefore, in order to develop usable software and prevent disuse, the integration of UCD techniques within agile methodologies has increased in the recent years [26][27]. Within these integrations, an agile version of the Personas technique stands out, which consists of a partial application of the technique at the beginning of the development and its refinement and completeness throughout the iterations. This *agilized* version of the Personas technique helps to overcome the time constraints that exist in the agile development process [28][29].

Different examples have been found in the agile literature proving that the Personas technique helps both to improve the usability of interfaces and to meet user requirements during the agile lifecycle. This makes the technique useful for mediating communication between developers and designers, measuring design effectiveness, and determining how a product should behave [30][31][32]. Therefore, the next step would be to know how such technique is being integrated in the different types of agile development processes in order to achieve an effective usability result in the software product, and to learn from the experiences of the different researches to unify the integration guidelines, thus improving its application in future work.

IV. Research Method

The information extracted from the primary studies aims to answer the following research questions: **(RQ1)** What is the state of the art of the Personas technique adoption in agile processes? **(RQ2)** What are the main ways of integrating the Personas technique in agile software development according to the primary studies? **(RQ3)** What are the main difficulties of integrating the Personas technique in agile software development and what improvements can be made?

A. Define the Search Strategy

The SMS begins with the identification of the keywords, which are those that appear most frequently in the Control Group (CG) articles: a reduced set of 13 papers directly related to the research area: [19][20][26]-[31][33]-[37]. Thus, to assess the validity of the search strings formed, we checked the number of CG articles retrieved within the Scopus database, since, being the largest database, it is where the highest number of CG articles are most likely to be found. To obtain the keywords, a table was generated with the frequency of all the words and combinations of words that appear in the CG articles, with the help of the Atlas.ti software. Only those words directly related to the research questions and that were present in a significant percentage of the GC articles were selected.

Once the keywords were identified, several search strings were constructed. For their construction, the words were grouped into synonyms of different components: words related to (i) the Personas technique, (ii) usability, and (iii) agile processes. The logical operator AND was used to join each of these components, while the logical operator OR was used to include synonyms of words from the same component. The different strings generated were tested within the Scopus database, and finally we selected the one that retrieved the bigger amount of CG articles: *Personas AND (usability OR user OR ucd OR "user-centered design" OR ux OR "user experience" OR hci OR "interface design" OR "interaction design") AND (agile OR "agile development" OR "extreme programming" OR scrum OR sprint OR "user stories" OR "agile method" OR "agile software development" OR "agile process")*. The criteria used to select the primary studies are summarized below.

a) Inclusion criteria: the paper is directly related to the use of the Personas technique in agile software development OR the paper describes the application of the Personas technique in agile software development OR the paper integrates the Personas technique in agile software development.

b) Exclusion criteria: the paper is a secondary study OR the paper is a primary study, but the topic is not directly related to integration or the use of the Personas technique in agile software development OR the paper is not written in English.

B. Select the Studies

Although the search string tests were performed in Scopus, the largest database of peer-reviewed literature [38], the searches were also performed in ACM Digital Library and IEEE Xplorer in order to cover more results. The databases were analyzed sequentially: first Scopus (*"Title OR Abstract OR Keywords"*), then ACM (*"Abstract OR Title")* and finally IEEE Xplorer (*"Abstract OR Title"*). If a duplicate appeared, the first result was kept.

A total of 104 papers were found in the different databases, among which 9 GC articles were retrieved. After excluding the articles that had appeared duplicated, the number was reduced to 78. Next, a peer review was carried out on these articles, applying the selection criteria to the title and abstract. The selected articles were validated upon a consensus meeting, analyzing together the abstracts of the conflicting articles, thus reducing the total to 38 pre-selected articles. After the meeting, the selection criteria were again applied to the remaining articles, but this time to the complete text, obtaining a final selection of 28 primary studies [27]-[30][34][35][39]-[60].

V. RESULTS AND DISCUSSION

A. State of the Art of the Personas technique adoption

To assess the state of the art of the adoption of the Personas technique in agile processes, each of the 28 selected studies was classified according to the type of agile process on which it was integrated. Fig. 1 synthesizes the results using two bubble scatter plots. The upper graph represents the number of articles published per year according to the type of publication (conference, journal or book chapter). Similarly, the lower graph plots the type of publication against the agile methodology on which it has been integrated. Thus, the bubbles are located at the intersections between the two axes and their size is proportional to the number of publications for each combination of values.

Although there have been studies integrating the Personas technique in agile processes since 2003, the interest in its integration in agile developments is increasing since 2016. This increase in the application of the technique may be due to the increasing use of agile methodologies because of their benefits. As the interest grows, the need to know the users arises, and therefore the use of the Personas technique is increasing. In addition, most primary studies have focused on Scrum and XP agile processes, and have been published in specialized conferences and journals, suggesting that the interest of the scientific community in integrating this technique in agile processes is increasing.

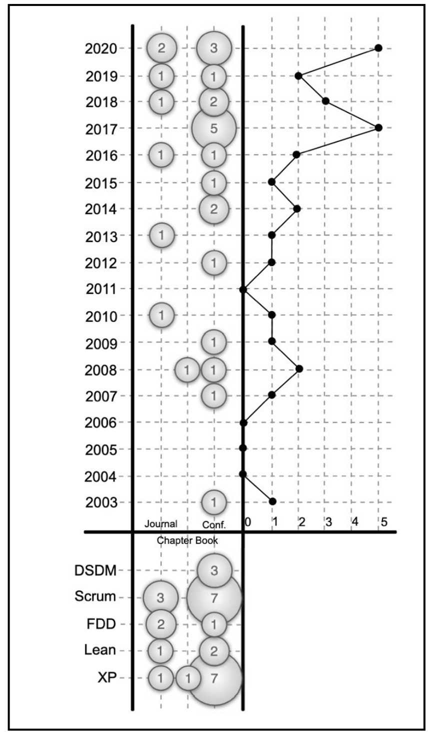


Figure 1. Mapping for the primary study distribution.

B. Main ways of integrating the Personas technique

We identified and extracted the main forms of integration of the Personas technique in agile software development for the selected articles, the description of how this integration was carried out and the life cycle activity where it was integrated. A synthesis of these results is shown on a mind map in Fig. 2.

1) DSDM: The studies conducted on DSDM agile processes create the personas models from both an interview and an analysis of the user stories. In the case of [39], instead of a narrative, they create drawn sketches of the personas based on the information obtained within the interviews. In [40], a previous design thinking session is carried out in which user stories are analyzed among all team members. In both cases the technique is integrated during the elucidation and requirements analysis activity, and additionally in [41] it is also integrated during the planning and design activity. In all three studies they validate the assignment of personas to user stories with the end users before starting the design. Moreover, in all of them, they validate each of the solutions after elaborating their designs.

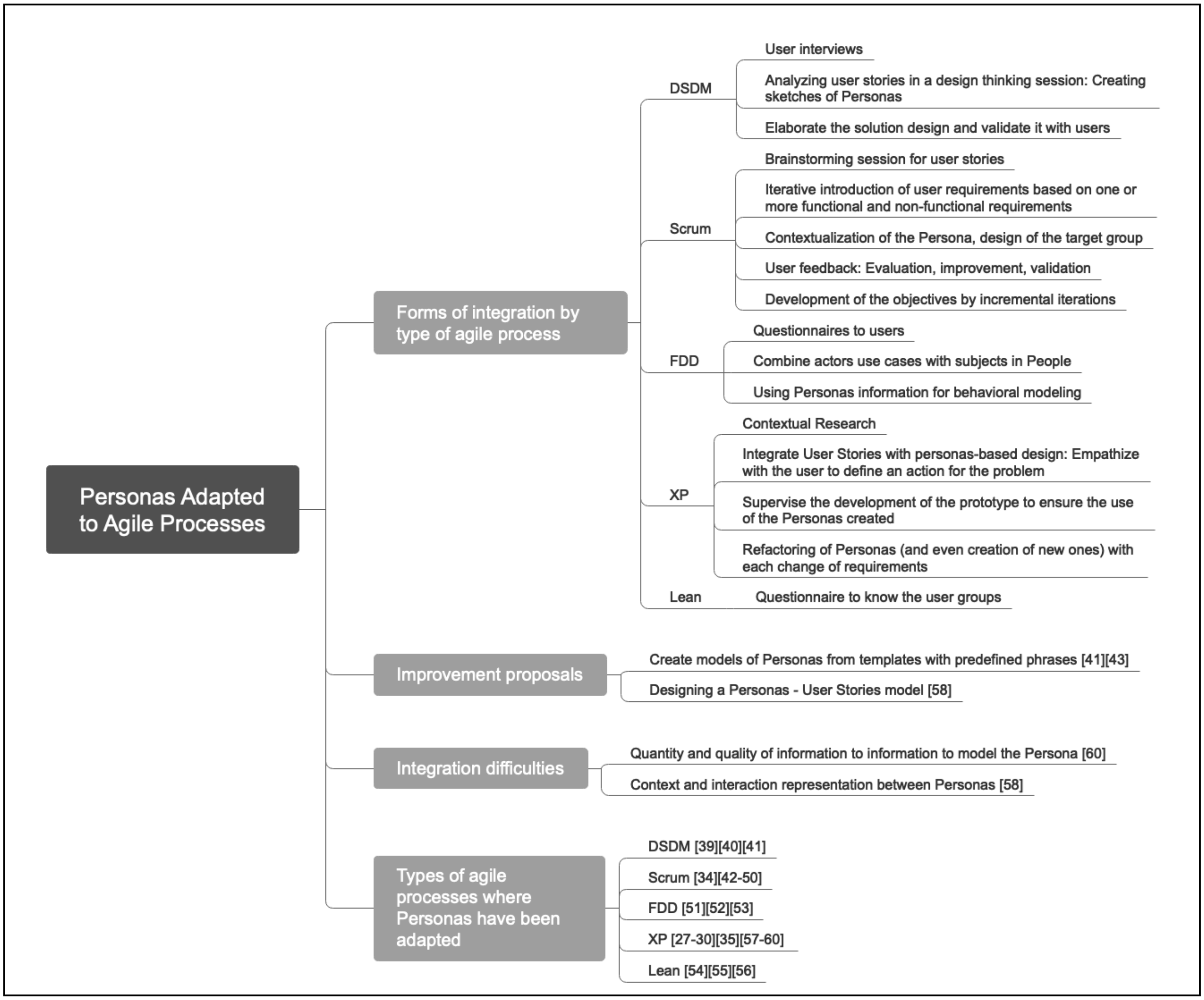


Figure 2. Main forms of integrations of the Personas technique in agile software development, main difficulties and proposed improvements.

2) Scrum: The studies that have integrated Personas into Scrum propose creative team sessions prior to the start of development to complete the Personas narratives. There are several studies in which they conduct a brainstorming session with students [42][43][44], where they complete the Personas narratives with previously generated sentences, and later associate them to the most convenient User Stories [45]. In the study by [46], it is proposed to use mind maps to connect the different Personas. The studies [34][47][48] associate each persona with a specific context, with a short description with preferences and a motivation that makes it easier for developers to empathize with end users during development. They all address user goals in incremental iterations, validating the functionality of the goals with users after each iteration. Moreover, in [49][50] they include non-functional requirements as goals as well, in order to obtain high-fidelity prototypes.

3) FDD: As for studies that have integrated Personas in FDD, they seek to analyze the interaction of people to establish behavioral patterns. In the studies of [51] and [52], they abstract patterns from user stories and assign them to specific subjects, and in [53] they further conduct interviews involving emotional analysis experts in order to more easily identify end-user personalities.

4) Lean: The integration of Personas in the studies analyzed on Lean start by knowing the user groups targeted by the development, either through questionnaires [54] or contextual investigations [55]. In [56] they group the results into clusters of users based on the preferences and behaviors found, customizing subsequent designs for the patterns found in each cluster.

5) XP: Finally, the studies carried out on the integration of Personas in XP interview and investigate the context of users in order to empathize more easily with them, and thus orient the development to their preferences [32][33][34][39][57][58][59]. Furthermore, [27] and [30] propose to collect information asynchronously to the project development as the team receives new information from users, refactoring and even creating new personas if they fit better the new user requirements. As in [46], in the study by [60] it is also proposed to design a mind map to connect what the persona wants and how they want it, using colors to highlight what is most relevant.

C. Main difficulties of integrating the Personas technique

Throughout the literature review, two main difficulties have been found to integrate the Personas technique in agile software development, generalized throughout the different studies. The main difficulty is to find the amount of necessary and sufficient information that should appear in the initial description of the persona. It should be detailed enough for the development team to empathize with the user's needs, but not as detailed as to conflict with the time restrictions within an agile process [41] [60]. An interesting solution could be to create the initial models of personas from templates with predefined phrases, as was proposed in [41] and [43]. Although the personas created by self-reported information during interviews may not be reliable [60], after the analysis performed over all the primary studies, we consider that this could be standardized within the Personas technique integrated to agile methodologies. This way, the first persona model would be created with a much lower temporal impact over the project. The first persona sketch would be simple, but it would be refined over iterations, as applied in studies [28][52][51][55][59].

The other difficulty shared by the different studies is to represent the context where a persona wants to perform an action, and the possible interaction with other personas within one same requirement [58]. Personas are created independently of each other, with the purpose of solving specific use cases [39] [41][47]. In the study of [58] they propose to design an Entity-Relationship model to allow differentiating the relationships between the different personas and their user stories.

Within the model there would be three entities: *User Story*, *Persona* and *Navigation Relationship*. The *User Story* entity would have a user value attribute, with the objective of prioritizing the list of requirements. On the other hand, the *Persona* entity would contain the attributes related to context of use, so that it would be possible to differentiate between different types of requirements according to the user. Finally, the *Navigation Relationship* entity would include attributes representing the interactions between *Persona* and *User Stories*, thus allowing different contexts of use between different *Personas* for the same *User Story* and, therefore, representation of more complex usage scenarios.

VI. Validity Threads

Throughout this study, certain aspects that could jeopardize the validity of the study have been assessed. The main threat to its validity is related to the possibility of bias in the selection process of the primary studies. To reduce this bias, we followed the guidelines proposed by the authors [15][16]. In addition, this SMS was carried out using three databases (Scopus, ACM, and IEEE Xplore), since they were considered the most relevant for the purpose of this search. However, if additional databases had been included, new results and complementary information would have been obtained. Numerous tests were also carried out to ensure the adequacy of the search string, checking that the maximum possible number of papers belonging to the CG were returned by the final search string. Another threat to validity is related to the application of the selection criteria and the analysis of the abstracts of the articles found. In order to minimize subjectivity, the selection process was carried out in parallel by two members of the research team, and the selected articles were subsequently agreed upon in a group meeting.

VII. Conclusions

Throughout this work, we have presented a secondary study on the integration of the Personas technique over different types of agile processes, with the objective of understanding the current state of the art of integration and to establish a knowledge base that would allow proposing future improvements to the technique. The study started by identifying the keywords in a set of articles called Control Group. These keywords were combined to formulate a search string that allowed us to carry out an in-depth analysis of all primary studies related to the integration of both concepts (the Personas technique on agile methodologies) over different databases. Subsequently, we applied a selection criterion to exclude those publications that did not contain the information to answer the research questions. From these 28 studies analyzed in depth, it has been possible to see that the integration of the Personas technique in agile developments has been increasing since 2016, which reflects a growing interest of the scientific community in the field, especially within the agile processes Scrum and XP.

After the synthesis of the results of the different publications, we have observed that, regardless of the type of agile process in which the Personas technique was integrated, there were some common aspects among them. On the one hand, integration always takes place at least during the activity of elucidation and requirements analysis, although it may also involve other activities of the software development process. In addition, regarding the integration steps, the first step always consists of performing an analysis of the target users, either by questionnaires, interviews, or brainstorming. This step allows to obtain a first persona model that can be refined or adapted according to new user requests that come with each iteration. On the other hand, the main difficulties in integrating the technique within the agile methodology are related to the difference in paradigm between User-Centered Design (where usability and detailed knowledge of the end user is a priority), and agile development (where the objective is to cover functionalities from early iterations with value for the client, affecting the time dedicated to the design).

In conclusion, this secondary study has analyzed numerous articles focused on the integration of the Personas technique within agile processes, in which different integration approaches, difficulties and proposed solutions have been found. This comprehensive analysis of articles from reliable sources provides support for the development and application of this technology on the future, aiming to obtain agile development processes with increasingly user-centered results. Future work will attempt to propose a more agile technique to validate it in a case study.

ACKNOWLEDGMENT

This research was funded by Spanish Ministry of Science, Innovation and Universities research grant PGC2018-097265-B-I00 and MASSIVE project (RTI2018-095255-B- I00). This research was also supported by the Madrid Region R&D programme (project FORTE, P2018/TCS-4314).

REFERENCES

[1] ISO/IEC Std. 25010. (2011). Systems and Software Engineering - Systems and Software Quality Requirements and Evaluation (SQuaRE) - System and Software Quality Models. ISO/IEC.

[2] ISO Std. 9241-11. (2018). Ergonomics of Human-System Interaction - Part 11: Usability: Definitions and Concepts. ISO.

[3] Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., and Diakopoulos, N. (2016). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson.

[4] Ferreira, J. M., Acuña, S. T., Dieste, O., Vegas, S., Santos, A., Rodríguez, F., and Juristo, N. (2020). Impact of usability mechanisms: An experiment on efficiency, effectiveness and user satisfaction. *Information and Software Technology*, 117, 1-17.

[5] Fox, D., Sillito, J., & Maurer, F. (2008). Agile methods and user-centered design: How these two methodologies are being successfully integrated in industry. In *Proceedings of the Agile 2008 Conference*. Toronto, ON, Canada, pp. 63-72.

[6] Nielsen, J., Berger, J. M., Gilutz, S., and Whitenton, K. (2003). Return on Investment (ROI) for Usability, 4th Edition, Nielsen Norman Group. [Online] https://n9.cl/9cwd. Last Access on December, 2020

[7] Rodríguez, F. D., Acuña, S. T., and Juristo, N. (2015). Design and programming patterns for implementing usability functionalities in web applications. *Journal of Systems and Software*, *105*, 107-124.

[8] Cooper, A., Reimann, R., and Cronin, D. (2007). *About Face 3.0: The Essentials of Interaction Design*. Wiley Publishing, Indianápolis: Wiley Publishing.

[9] Cooper, A., Reimann, R., Cronin, D. (2007). *About Face 3.0: The Essentials of Interaction Design*. Wiley Publishing, Indianápolis.

[10] Magües, D. A., Castro, J., & Acuña, S. T. (2016). Requirements engineering related usability techniques adopted in agile development processes. In *Proceedings of the Twenty-Eighth International Conference on Software Engineering and Knowledge Engineering (SEKE'16)*. San Francisco Bay, California, USA, pp. 537-542.

[11] Castro, J. W. (2009). Extensión de la Técnica Personas para su Incorporación en la Actividad de Análisis del Proceso de Ingeniería del Software. Trabajo de Fin de Máster. Directora: Acuña, S.T. Universidad Autónoma de Madrid.

[12] Acuña, S. T., Castro, J. W., Juristo, N. (2012). A HCI Technique for Improving Requirements Elicitation. *Information and Software Technology*, 54(12): 1357-1375.

[13] Magües, D. A. (2016). Transformación de Técnicas de Usabilidad relacionadas con las Actividades de la Ingeniería de Requisitos para su Incorporación en los Procesos de Desarrollo Ágil. *Trabajo Fin de Máster, Escuela Politécnica Superior de Informática, Universidad Autónoma de Madrid*. Directores: Acuña, S. T., y Castro, J. W.

[14] Losana, P. (2019). Aplicación de la Técnica Personas en un Proyecto de Desarrollo Ágil con Scrum. *Trabajo Fin de Grado, Escuela Politécnica Superior de Informática, Universidad Autónoma de Madrid*. Directora: Acuña, S. T.

[15] Kitchenham, B., and Charters, S. (2007). *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Tech. rep., Keele University and Department of Computer Science University of Durham.

[16] Kitchenham, B. A., Budgen, D., & Brereton, O. P. (2011). Using mapping studies as the basis for further research–a participant-observer case study. *Information and Software Technology*, *53*(6): 638-651.

[17] Cooper, A. (2004). *The Inmates are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity*. Sams - Pearson Education, Indianápolis.

[18] Caballero, L., Moreno, A. M., & Seffah, A. (2014, September). Persona as a tool to involving human in agile methods: Contributions from HCI and marketing. In *Proceedings of the Intern. Conf. on Human-Centred Software Engineering (HCSE'14).* Paderborn, Germany, pp. 283-290.

[19] Anwar, S., Motla, Y. H., Siddiq, Y., Asghar, S., Hassan, M. S., & Khan, Z. I. (2014, December). User-centered design practices in scrum development process: A distinctive advantage?. In *Proceedings of the 17th IEEE Intern. Multi Topic Conference 2014*. Karachi, Pakistan, pp. 161-166.

[20] Chamberlain, S., Sharp, H., & Maiden, N. (2006). Towards a Framework for Integrating Agile Development and User-centred Design. In: Abrahamsson, P., Marchesi, M., and Succi, G. (eds.) Extreme Programming and Agile Processes in Soft. Eng., pp. 143-153.

[21] Stapleton, J. (Ed.). (2003). *DSDM: Business Focused Development*. Pearson Education.

[22] Palmer, S. R., and Felsing, M. (2001). *A Practical Guide to Feature-Driven Development*. Prentice Hall.

[23] Poppendieck, M., and Poppendieck, T. (2003). *Lean Software Development: An Agile Toolkit*. Addison-Wesley Professional.

[24] Schwaber, K., and Beedle, M. (2002). *Agile Software Development with Scrum*, 1st Edition. Pearson.

[25] Beck, K., & Gamma, E. (2000). *Extreme programming explained: embrace change*. Addison-Wesley Professional.

[26] Sy, D. (2007). Adapting usability investigations for agile user-centered design. *Journal of Usability Studies*, 2(3): 112-132.

[27] Wolkerstorfer, P., Tscheligi, M., Sefelin, R., Milchrahm, H., Hussain, Z., Lechner, M., and Shahzad, S. (2008). Probing an agile usability process. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI'08).* Florence, Italy, pp. 2151-2158.

[28] Haikara, J. (2007). Usability in agile software development: extending the interaction design process with personas approach. In *Proceedings of the International Conference on Extreme Programming and Agile Processes in Software Engineering (XP'07).* Como, Italy, pp. 153-156.

[29] Rahim, W. A., Isa, W. M., Lokman, A. M., Taharim, N. F., and Wahid, N. D. (2014). Engineering m-learning using agile user-centered design. In *Proceedings of the 2014 Eighth Intern. Conf. on Next Generation Mobile Apps, Services and Technologies.* Oxford, UK, pp. 60-65.

[30] Broschinsky, D., & Baker, L. (2008). Using persona with XP at LANDesk Software, an Avocent company. In *Proceedings of the Agile 2008 Conference*. Toronto, ON, Canada , pp. 543-548.

[31] Hussain, Z., Slany, W., and Holzinger, A. (2009b). Current state of agile user-centered design: A survey. In *Proceedings of the Symposium of the Austrian HCI and Usability Engineering Group (USAB'09).* Linz, Austria, pp. 416-427.

[32] Øvad, T., and Larsen, L. B. (2015). The prevalence of UX design in agile development processes in industry. In *Proceedings of the 2015 Agile Conference*. Washington, DC, USA, pp. 40-49.

[33] Hudson, W. (2003). Adopting user-centered design within an agile process: A conversation. *Cutter IT Journal*, *16*(10): 5-12.

[34] Hudson, W. (2013). User stories don't help users: Introducing Persona stories. *Interactions*, 20(6): 50-53.

[35] Hussain, Z., Milchrahm, H., Shahzad, S., Slany, W., Tscheligi, M., and Wolkerstorfer, P. (2009). Integration of eXtreme Programming and user-centered design: Lessons learned. In *Proceedings of the Intern. Conf. on Agile Processes and Extreme Programming in Software Engineering (XP'09).* Sardinia, Italy, pp. 174-179.

[36] Najafi, M., and Toyoshiba, L. (2008). Two case studies of user experience design and agile development. In *Proceedings of the Agile 2008 Conference.* Toronto, ON, Canada, pp. 531-536.

[37] Nakao, Y., Moriguchi, M., and Noda, H. (2014). Using agile software development methods to support human-centered design. *NEC Technical Journal*, 8(3): 37-40.

[38] *Scopus. An Eye on Global Research: 5,000 Publishers. Over 71 M Records and 23,700 Titles*, Amsterdam, The Netherlands, Elsevier, 2018.

[Online]. Available: https://www.elsevier.com/__data/assets/pdf_file/0017/114533/SC_FS_overview_WEB.pdf

[39] Vilchez-Sandoval, J., Vasquez-Paragulla, J., and Llulluy-Nuñez, D. (2020). On the of use agile methodologies to re-design a Networks and Data Communications course. In *Proceedings of the IEEE World Conf. on Engineering Education (EDUNINE'20).* Bogotá, Colombia, pp. 1-5.

[40] Parizi, R., Moreira da Silva, M., Couto, I., Trindade, K., Plautz, M., Marczak, S., Conte, T., and Candello, H. (2020). Design thinking in software requirements: What techniques to use? A proposal for a recommendation tool. In *Proceedings of the 23rd Ibero-American Conf. on Software Engineering (CIbSE'20),* Curitiba, Brazil, pp. 1-14.

[41] Cleland-Huang, J., Babar, M. A., & Mirakhorli, M. (2014, May). An inverted classroom experience: engaging students in architectural thinking for agile projects. In *Companion Proceedings of the 36th International Conference on Software Engineering (ICSE Companion'14)*. Hyderabad, India, pp. 364-371.

[42] Quade, S., and Schlüter, O. (2020). Ideation: Generating as many ideas as possible. *Logos*, 31(1): 48-53.

[43] Gaikwad, V., Joeg, P., and Joshi, S. (2017). AgileRE: Agile Requirements management tool. In *Proceedings of the Computational Methods in Systems and Software (CoMeSySo'17).* Szczecin, Poland, pp. 236-249.

[44] Hakim, J., Spitzer, T., and Armitage, J. (2003). Sprint: Agile specifications in shockwave and flash. In *Proceedings of the 2003 Conference on Designing for User Experiences (DUX'03).* San Francisco, CA, USA, pp. 1-14.

[45] Choma, J., Zaina, L. A., & Beraldo, D. (2015, August). Communication of design decisions and usability issues: a protocol based on Personas and Nielsen's heuristics. In *International Conference on Human-Computer Interaction*, pp. 163-174.

[46] Sekar, B. (2017). Enterprise software experience design: Journey and lessons. In *Proceedings of the IFIP Conference on Human-Computer Interaction (INTERACT'17).* Bombay, India, pp. 356-359.

[47] Dirks, S. (2020). Persona design in participatory agile software development. In *Proceedings of the 22nd Intern. Conference on Human-Computer Interaction (HCI'20).* Copenhagen. Denmark, pp. 52-64.

[48] Choma, J., Zaina, L. A., & Beraldo, D. (2016, July). UserX story: incorporating UX aspects into user stories elaboration. In *Intern. Conf. on Human-Computer Interaction*, pp. 131-140.

[49] Paiva, S. C., and Carvalho, D. B. F. (2018). Software creation workshop: A capstone course for business-oriented software engineering teaching. In *Proceedings of the XXXII Brazilian Symposium on Software Engineering (SBES'18).* Sao Carlos, Brazil, pp. 280-288.

[50] Losada, B. (2018). Flexible requirement development through user objectives in an agile-UCD hybrid approach. In *Proceedings of the XIX International Conference on Human Computer Interaction (Interacción'18).* Palma, Spain, pp. 1-8.

[51] Forbrig, P., and Dittmar, A. (2019). Applying agile methods and personas to S-BPM. In *Proceedings of the 11th International Conference on Subject-Oriented Business Process Management (S-BPM ONE'19).* Sevilla, Spain, pp. 1-10.

[52] de Carvalho, C. R. M. (2010). MEX experience boards: A set of agile tools for user experience design. In *Proceedings of the IX Symposium on Human Factors in Computing Systems (IHC'10*). Belo Horizonte Minas Gerais, Brazil, pp. 213-216.

[53] Abdullah, N. N. B., Grundy, J., McIntosh, J., How, Y. C., Saharuddin, S., Tat, K. K., Ye, E. S., Rastom, A. J. A., & Othman, N. L. (2020, August). Using Work System Design, User Stories and Emotional Goal Modeling for an mHealth System. In *2020 IEEE First Intern. Work. on Requirements Engineering for Well-Being, Aging, and Health (REWBAH).* Zurich, Switzerland, pp. 1-10.

[54] Perdana, R. A., Suzianti, A., and Ardi, R. (2017). Crowdfunding website design with lean product process framework. In *Proceedings of the 3rd International Conference on Communication and Information Processing (ICCIP'17).* Tokyo, Japan, pp. 369-374.

[55] Bhattarai, R., Joyce, G., & Dutta, S. (2016, July). Information security application design: understanding your users. In *International Conference on Human Aspects of Information Security, Privacy, and Trust*, pp. 103-113.

[56] Mereu, S., Newman, M., Peterson, M., Taylor, E., White-Sustaita, J., and Yeats, D. (2017). Top-down vs bottom-up approaches to user segmentation: The best of both worlds. In *Proceedings of the Human Factors and Ergonomics Society 2017 Annual Meeting,* 61(1): 515-519.

[57] Sohaib, O., Solanki, H., Dhaliwa, N., Hussain, W., and Asif, M. (2019). Integrating design thinking into eXtreme Programming. *Journal of Ambient Intelligence and Humanized Computing*, 10, 2485-2492.

[58] Sedeño, J., Schön, E. M., Torrecilla, C. J., Thomaschewski, J., Escalona, M. J., and Mejías, M. (2017). Modelling agile requirements using context-based Persona stories. In *Proceedings of the 13th International Conference on Web Information Systems and Technologies (WEBIST'17),* Porto, Portugal, pp. 196-203.

[59] Hussain, Z., Lechner, M., Milchrahm, H., Shahzad, S., Slany, W., Umgeher, M., Vlk, T., Köffel, C., Tscheligi, M., and Wolkerstorfer, P. (2012). Practical usability in XP software development processes. In *Proceedings of the Fifth International Conference on Advances in Computer Human Interactions (ACHI'12),* Valencia, Spain, pp. 208-217.

[60] Khanh, N. T., Daengdej, J., and Arifin, H. H. (2017). Human stories: A new written technique in agile software requirements. In *Proceedings of the 6th International Conference on Software and Computer Applications (ICSCA'17).* Bangkok, Thailand, pp. 15-22.

Gene Ontology Terms Visualization with Dynamic Distance-Graph and Similarity Measures

Alessia Auriemma Citarella, Fabiola De Marco, Luigi Di Biasi, Michele Risi, Genoveffa Tortora
Department of Computer Science
University of Salerno
84084 Fisciano (SA), Italy
{aauriemmacitarella, fdemarco, ldibiasi, mrisi, tortora}@unisa.it

Abstract

In the biological field, having a visual and interactive representation of data is useful, particularly when there is a need to investigate a large amount of multilevel data. It is advantageous to communicate this knowledge intuitively because it helps the users to see the dynamic structure in which the correct connections are interacting and extrapolated. In this work, we propose a human-interaction system to view similarity data based on the functions of the Gene Ontology *(Cellular Component, Molecular Function, and Biological Process) for Alzheimer's and Parkinson's disease proteins/genes. The similarity data was built with the Lin and Wang measures for all three areas of gene ontology. We clustered data with the K-means algorithm and then we have suggested a dynamic and interactive view based on SigmaJS with the aim of allowing customization in the interactive mode of the analysis workflow by users. In this way we have obtained a more immediate visualization to capture the most relevant information within the three vocabularies of Gene Ontology. This facilitates to obtain an omic view and the possibility of carrying out a multilevel analysis with more details which is much more useful in order to better understand the knowledge of the end user.*

Index terms— Protein Visualization, Gene Ontology, Clustering

1 Introduction

The importance of having an omic vision is becoming increasingly important to define biological systems at an increasingly detailed level. Omic sciences aim to produce useful knowledge that can be used to characterize and interpret a biological system [17]. For omic sciences we refer to the wide range of biomolecular disciplines characterized by the suffix -omics including genomics, transcriptomics, proteomics, and metabolomics. In this sense, technological innovation aids the growth of complex system biology by allowing researchers to investigate various intrinsic and extrinsic influences and events at the base of life. Biological data is multidimensional and highly interdependent. The current challenge is to gain a more detailed integrative view of the dynamics of cellular processes in a cell or an organism rich in biological and spatial-temporal information [18]. Clear visualization methods can provide more immediate access to their content information.

The visualization of biological data has become increasingly relevant in Biosciences, as O'Donoghue *et al.* [13] point out because it helps researchers to interpret heterogeneous data more quickly. One of the most current issues in omic data analysis is the inability to investigate relationships between multi-omic states to incorporate them and combine higher-level expertise [22].

In this paper, we report the preliminary results we obtained regards visualization of the similarity of the proteins based on the protein annotations. Protein similarity visualization not based on sequence alignment can be tricky due to inter-class dissimilarities and inter-class similarity [1]. Clustering and Machine Learning algorithms could fail to good abstract interdependencies between the objects [7]. This fact often does not allow to generates a clear visual representation of the information.

Our idea is to show how a dynamic graph generation aided by a human can help abstract functional relationships between proteins to generate a clear data visualization where a standard clustering algorithm fails. For this contribution, we focused on two diseases: *Alzheimer* and *Parkinson*.

Alzheimer's disease (AD) is a form of degenerative dementia that occurs after 65 years. In this pathology, there is a deposition of an Aβ peptide B with the formation of senile plaques and the intracellular aggregation of *tau* protein [4]. Parkinson's disease (PD) is the second most com-

DOI reference number: 10.18293/DMSVIVA2021-013

mon neurodegenerative disorder in the senile age in which neuronal loss is found in the substance nigra and formation of α-synuclein aggregates that are neuropathological [14]. These pathologies show similar neurodegeneration mechanisms supported by scientific evidence with genetic, biochemical, and molecular studies. Pathological pathways involving α-synuclein and *tau* proteins, oxidative stress, mitochondrial dysfunction, iron pathway, and *locus coeruleus* are among these findings [21]. They were chosen as an example for our search workflow because of this overlapping between their pathological mechanisms. This aspect introduces intra-class, and extra-class overlaps able to fool standard clustering methods.

The paper is structured as follows. In Section 2 we describe the most important related works in the examined field. In Section 3 we explain the conducted experiments and in Section 4 we discuss respectively the dataset, methodology, and performance measures which we have used in our research. Finally, we expose the visual results in Section 5 and the conclusions with future work in Section 6.

2 Related work

In the literature, several web interfaces can query the terms of the Gene Ontology. The *Gene Ontology* (GO) is a bioinformatics project that supports the standardization of biological information about attributes of genes and gene products through the use of ontology. It is structured as an acyclic oriented graph where each GO-term is identified by a word or strings and a unique alphanumeric code [6].

QuickGO allows us to find and display GO terms and generate a list of correspondence results based on the user's question. This tool returns a directed acyclic graph (DAG) containing a single GO term and its associated terms and annotations. It is designed with JavaScript, Ajax, and HTML. Statistics with interactive graphs and views of term location tables are available on the fly, indicating which words are frequently noted simultaneously. The user can create a subset of annotations based on different parameters (Specific protein, Evidence Codes, Qualifier Data, Taxonomic Data, Go Terms) and download them [2].

GOrilla[1] identifies enriched GO terms in ordered lists of genes using simple, intuitive, and informative graphics, without explicitly requiring the user to provide targets or background sets. It is a GO analysis tool that employs a statistical approach with flexible thresholds to identify GO terms significantly enriched at the top of a classified gene (very useful when genomic data can be represented as a classified list of genes). The analysis's results are presented in the form of a hierarchical structure that allows for a clear view of the GO terms [5].

[1] Gorilla: http://cbl-gorilla.cs.technion.ac.il

Blast2GO (B2G)[2] is an interactive platform that supports non-model species functional genomic research. It is a data sequence-based tool that combines high-performance analysis techniques and evaluation statistics with a high degree of user interaction. Similarity searches produce results on direct acyclic graphs [3].

NaviGO[3], in order to measure the similarity or relation between the terms of the GO, use six different scores: Resnik, Lin, and the relevant semantic Similarity score for semantic similarity, and *Co-occurrence Association Score* (CAS), *PubMed Association Score* (PAS), and *Interaction Association Score* (IAS) for GO associations. A `Funsim` score for functional similarity is also introduced [20].

More recently, the open-source software *AEGIS* allows us to visually explore the GO data in real-time, taking into input the entire dataset GO. Any Go terms can be chosen as the anchor and have a root, leaf, or waypoint, represented with a DAG. Each source can include all the descendants of the anchor term, the leaves will only include the ancestors, and the Waypoint anchors will constitute a DAG consisted of both ancestors and descendants [24].

3 Experimental setup

We explored two ways to calculate semantic similarity. We calculated the similarity for all three ontology gene domains, both for Alzheimer's and Parkinson's proteins, separately. For this first experiment, we considered both Lin's similarities and Wang's method. For simplicity, in this work we only show the results concerning the similarity of Lin while the future tool will allow user the setting of both measures. Subsequently, we clustered the data obtained for both similarity measures in about BPs, CCs, and MFs domains for AD and PD with the K-means algorithm, trying with n=3 and n=5 clusters.

4 Methods

In this work, we have used the R environment[4], a free software environment for statistical computing and graphics, and SigmaJS, a JavaScript library dedicated to graph drawing[5]. We used the standard SigmaJS renderer to show the graph view.

4.1 Datasets

Protein datasets for AD and PD were downloaded from UNIPROT [16]. Data cleaning has been carried out, removing all duplicates. Furthermore, for each UNIPROT protein

[2] Blast2GO: https://www.biobam.com
[3] NaviGO: https://kiharalab.org/web/navigo/views/goset.php
[4] R: https://www.r-project.org
[5] SigmaJS: https://sigmajs.org

ID, the reference gene has been obtained and linked to the STRING, removing all the proteins which were not mapped in this database. STRING database allows us to consider any protein-protein interactions (PPI) based on a score calculated on experimental evidences [15]. We have recovered a total of 216 genes for AD and 137 genes for PD.

4.2 Gene Ontology

The Gene Ontology is based on two types of relationships between objects: *instances* and *part of*. Three considered all the organisms share biological domains and that constitute structured and controlled vocabularies:

- *Biological Process*: refers to all those events that take place within an organism resulting from an orderly set of molecular functions;
- *Cellular Component*: concerns the location of the entity in question at the level of cellular and/or subcellular structures;
- *Molecular Function*: describes the processes that occur at the molecular level.

We have identified these domains with the following acronyms: biological process (BP), cellular component (CC), and molecular function (MF). We have recovered from UNIPROT[6] all the GO terms belonging to these three fields both for Alzheimer's and Parkinson's diseases with `UniProt` package in R.

4.3 Distance Metrics

We used two types of metric to calculate pairwise semantic similarities: *Lin* and *Wang* similarities with the `GOSemSim` package in R [23].

4.3.1 Lin's measure

Lin measure is based on *information content* (IC). The negative log of a concept's probability is formally known as information content (IC). This method computes the ratio between the amount of "common information" and the amount of "total information" in the descriptions regards an object pair. This ratio corresponds to the similarity between two objects [10].

In this case, this approach can measure the similarity of the knowledge content of the GO terms for each protein dataset referring to the two diseases. The frequency of two GO words involved and their closest common ancestor in a particular corpus of GO annotations are used in the estimation. The most basic definition two concepts share as an ancestor is suggested by the term *Least Common Subsammer* (LCS) [12]. So, we can consider the following Equation 1:

$$sim_{lin} = \frac{2 * IC(lcs(c_1, c_2))}{IC(c_1) + IC(c_2)} \quad (1)$$

where c_1 *and* c_2 are two concepts, *IC* is the information content and *lcs* is the function that computes the least common subsammer. In our experiment, c_1 and c_2 reflect the concepts represented by the GO terms referring to the BP, CC, and MF domains. The similarity is measured across all proteins in the pathological reference dataset, both for AD and PD.

4.3.2 Wang measure

The Wang method is based on a *graph-based* semantic similarity. The GO terms are converted into a numeric value by aggregating the terms of their ancestors in a GO graph [19].

Given two GO terms, *A* and *B*, we can represent DAG_A = (A, T_A, E_A) and DAG_B = (B, T_B, E_B), where T_n is the set of GO terms including the term *n* and all of its ancestor terms in the GO graph while E_n are the semantic relations represented as edges between the GO terms. The semantic similarity between these two terms are calculated as in Equation 2:

$$S_{GO}(A, B) = \frac{\sum_{t \in T_A \cap T_B} S_A(t) + S_B(t)}{SV(A) + SV(B)} \quad (2)$$

where $S_A(t)$ and $S_B(t)$ denote the S-value of a GO term *t* related to term *A* and term *B*.

Wang measures the semantic meaning of GO term *n*, SV(n), after obtaining the S-values for all terms in DAG_n with the Equation 3 below:

$$SV(n) = \sum_{t \in T_n} S_n(t) \quad (3)$$

4.4 K-means

K-means is one of the most common and widely used partitioning clustering algorithms because it divides a set of objects into K clusters based on their attributes [11]. A cluster is simply an aggregation of data based on similarities. The division into K clusters is done *a priori*, based on the goal to be achieved or using heuristic techniques, and the clusters represent the number of centroids required by the dataset. As the name implies, a centroid is a real or imaginary point that represents the cluster's center and is updated with each algorithm iteration.

The procedure is composed by four steps:

- *Step 1:* determine the value of K;

[6]UniProt: https://www.uniprot.org

- *Step 2:* randomly select K points as initial centers of the clusters;
- *Step 3:* assign each new point to the cluster with the closest Euclidean distance to its center. Formally, if c_i is a centroid of the set of centroids C then each point x will be assigned to a cluster based on:

$$\arg\min_{c_i \in C} dist(c_i, x)^2 \tag{4}$$

where dist (.) represents the Euclidean distance;

- *Step 4:* recalculate the updated cluster centers by averaging the points associated with each cluster:

$$c_i = \frac{1}{|S_i|} \sum_{x_i \in S_i} x_i \tag{5}$$

where Si is the cluster's set of points.

The procedure repeats steps 3 and 4 until a convergence is achieved.

The algorithm ensures speed of execution while leaving the data free to group and move away. Due to the goal of this research, we limited the max number of clusters to five. No PCA techniques were used. Figures 2 and 3 report how the GO objects are partitioned regarding the BP features for both diseases. The axis reports the distance between each item to its centroid. We used `cluster` and `factoextra` packages in R to perform clusterization. It is hard to read this kind of visualization due to arbitrary parameter values that the final user can assign to the number of the cluster.

4.5 Dynamic Distance-Graph

We propose a *dynamic build cyclic distance graph* (DCDG) to visualize and transfer knowledge regarding the GO terms. Our goal is to provide a clearer visualization of the GO interconnections than other visualization methods like clustering or partitioning. We used a web-based workspace built with Javascript and SigmaJS to allow the user to explore this interconnection. Workspace is designed to be as clean as possible. It starts as an empty web app with a single callable overlay menu on the upper left corner, allowing users to search the entry point protein into datasets.

The BP, CC, and MF distance matrices, calculated before the execution of the *k*-means algorithm, were used as datasets. When selected, the entry protein becomes the root of the graph. Users can click on each graph node to show a context menu (as depicted in Fig. 1) in which it is possible to choose extension (explosion) operation for the node itself.

We defined three kinds of extensions for this contribution, each of them related to one dataset: BP, CC and MF,

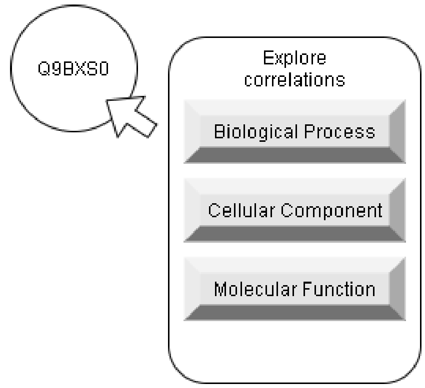


Figure 1: The contextual menu is available for each node.

whose definitions are those intended by the three vocabularies of the GO. The distance between each node pairs is written on the arcs between them. Also, the distance value is used to separate nodes into spaces.

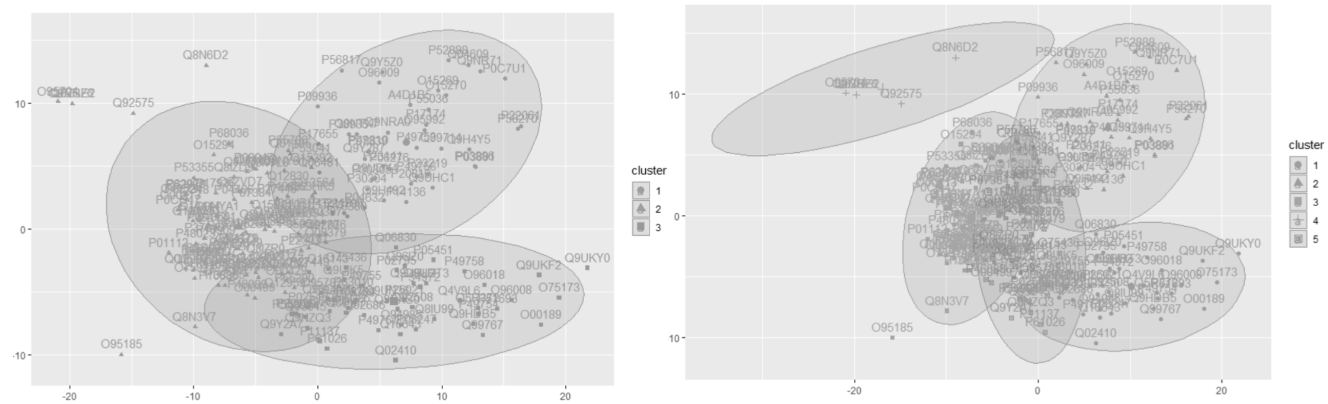
The ForceAtlas2 algorithm is used to avoid overlapping between near nodes. In particular, we used ForceAtlas2 embedded into SimgaJS [9]. ForceAtlas2 is a layout algorithm for force-directed graphs. This algorithm allows us to position each node depending on the other nodes using the distances between them as edge weights. Just because of this condition, the position of a node must always be confronted with the other nodes. The fundamental advantage of using ForceAtlas2 for the representation of protein graphs is to have an easier view of the structure because the structural proximity present in the original datasets is converted to visual proximity.

In order to better empathize the functionality distance between GO, we defined a spatial distance SD with the following equation. Given two nodes, A and B and their own distance d:

$$SD = log_e(d) \tag{6}$$

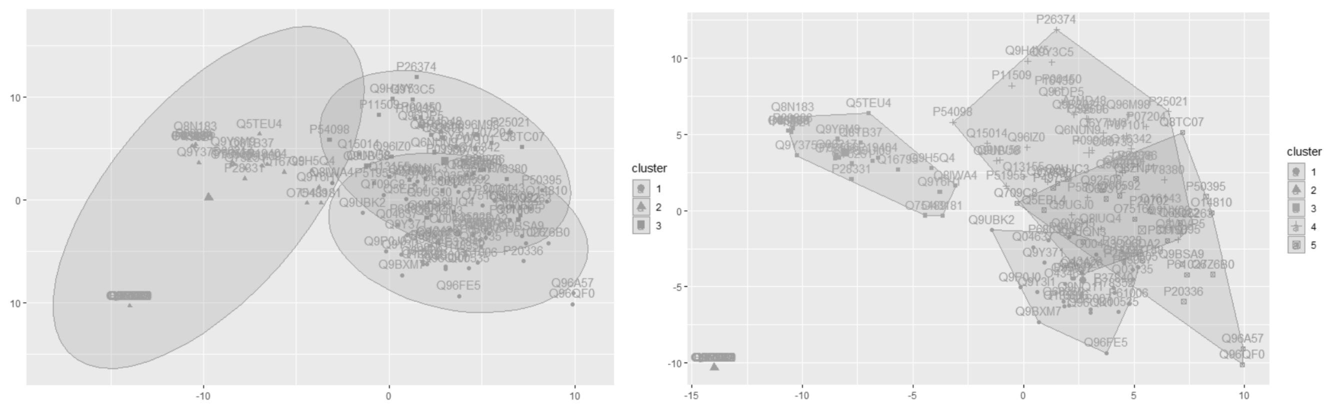
where *d* is the distance and the log_e is the natural logarithm with base the number of Nepero.

Note that SD is used only for graphical purposes in the rendering routines. Figure 5 shows no linear proportionality into edge lengths: see the distance between (Q8IZY2, Q9BS0) and (Q93045, Q9BS0). Still, for graphical purposes, we defined a threshold th_i as the mean of all the distances into the dataset i used for node expansion. As an example, given the node Q9BXS0 (see figure 5), the threshold th_Q9BXS0 is the mean of the edge's weight between Q9BXS0 and the related nodes. When the distance SD between two node A and B is greater than th_i, then node A and B are considered belonging to a different cluster. A dotted line renders each class separation.



(a) K = 3 **(b)** K = 5

Figure 2: K-means for BP for Alzheimer's disease with Lin's measure.



(a) K = 3 **(b)** K = 5

Figure 3: K-means for BP for Parkinson's disease with Lin's measure.

5 Results

5.1 K-means visualization

We represented the images related to Alzheimer's and Parkinson's diseases and calculated with Lin's measure for convenience. We have found that clustering with the K-means algorithm produces visually misleading and uninformative overlaps. In Figures 2 and 3 is shown what happens when the data of Alzheimer ans Parkinson diseases for only BP component is clustered with *n*=3 and *n*=5.

5.2 DCDG visualization

For our test, we considered the G9BXS0 protein from the similarity matrices obtained and we identified the proteins of his neighbor to build our view of node expansion. Before testing DCDG view, we carried out a simple statistic of the common GO terms, even in this case for the only BP component, between this *root* protein and its neighbors. We represented them with a Venn diagram [8] (see Fig. 4), on the basis of GO Lin's similarity matrix.

This view allows us to evaluate which elements are common among the different sets of the terms GO for all the selected proteins. It is clear that a simple statistics of the terms does not make useful information beyond the simple observation that there are terms, even if minimal, common to all five sets of the Terms Go of each protein. Instead, introduce similarity based on the *information content* of the

Figure 4: Venn Diagram for G9BXS0.

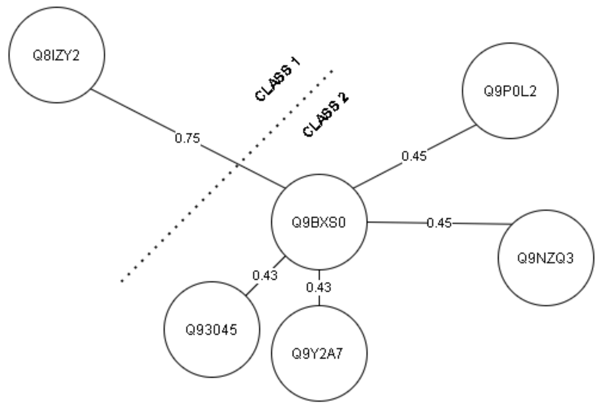


Figure 5: The result of Q9BX80 expansion by BP dataset.

GO terms is useful for expanding knowledge regarding biological aspects that would be omitted by a simple statistical analysis.

Figure 5 shows the BP expansion related to node G9BXS0, a protein produced by *COL25A1* gene for *Homo Sapiens* organism with the DCDG view. This protein inhibits the fibrillization of β-amyloid peptide which constitutes amyloid plaques present in Alzheimer's disease. It also assembles the amyloid fibrils in aggregates which are resistant to the demerger mechanisms.

The DCDG view allows the user to see and understand immediately the proteins belonging to the two distinct BP classes: **CLASS 1**, related to many biological processes such as signaling pathway and positive and negative regulation of cellular and chemical complexes and **CLASS 2**, concerning the organization of fibrils, microtubules, and structures of the cytoskeleton.

Figure 6 highlights the successive expansion of Q8IZY2 and Q9P0L2 proteins. Due to distances, a new class was identified by the system (**CLASS 3**). The visualization clearly states, from a point of view of biological meaning, that the added third class emphasizes further involvement of proteins indicated in different biological processes compared to previous classes. In particular, this class intervenes in broader biological regulation processes involving energy homeostasis and cell cycle regulation systems.

6 Conclusion

In this paper, we explored an alternative way to graphically view the relationships between the GO terms based on their information content. In particular, we have proposed a *human interaction*-based viewing system that allows the users to have a complete omic vision of data. In particular, by ensuring the direct representation of the inter-class and intra-class correlations between involved proteins. The strategy proposes an instrument to investigate the GO with a customizable and flexible approach providing information to a more general or selective level.

We presented a distance cyclic distance graph (DCDG) as a GO terms visualization approach to immediately represent interconnection between elements. The prototype was written as a web app by using the SigmaJS framework.

We used two similarity methods on the three GO vocabularies (*Biological Process, Cellular Component, and Molecular Function*) for two neurodegenerative diseases, Alzheimer's and Parkinson's diseases: Lin's and Wang's Methods. Thanks to these metrics, we built three different distance matrices (BP, CC, and MF) for each condition.

We explored the differences between the standard cluster view and the proposed DCDG view. The datasets were clustered using the K-means algorithm to show a classic clustering plot. Also, we use the proposed DCDG method to plot the same information into a graph view.

By applying a classic display of clustering, visually was not possible to recover the information immediately, also due to the problem of overlapping of some clusters elements. On the other hand, the display with DCDG allows a more immediate understanding of the interactions present between the proteins based on the similarity representative of the three vocabularies of the GO. The existence of well-outed protein clusters in a system is one of the purposes of our work as it represents a fundamental topological characteristic to understand the entire network of connections. This subdivision makes it possible to view the existing relationships between proteins and provides a tool which meets the need to identify and understand why some structural elements are grouped at different levels (cellular, biological and molecular) of in-depth.

As future work, we plan to improve the web-based tool prototype into a web app for exploring protein data based on the proposed assumptions in this research study, guaranteeing user-target customization of the tools available.

References

[1] M. Arif. Similarity-dissimilarity plot for visualization of high dimensional data in biomedical pattern classification. *Journal of Medical Systems*, 36(3):1173–1181, 2012.

[2] D. Binns, E. Dimmer, R. Huntley, D. Barrell, C. O'donovan, and R. Apweiler. QuickGO: A web-based tool for gene ontology searching. *Bioinformatics*, 25(22):3045–3046, 2009.

[3] A. Conesa, S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18):3674–3676, 2005.

[4] C. Duyckaerts, B. Delatour, and M.-C. Potier. Classification and basic pathology of Alzheimer disease. *Acta Neuropathologica*, 118(1):5–36, 2009.

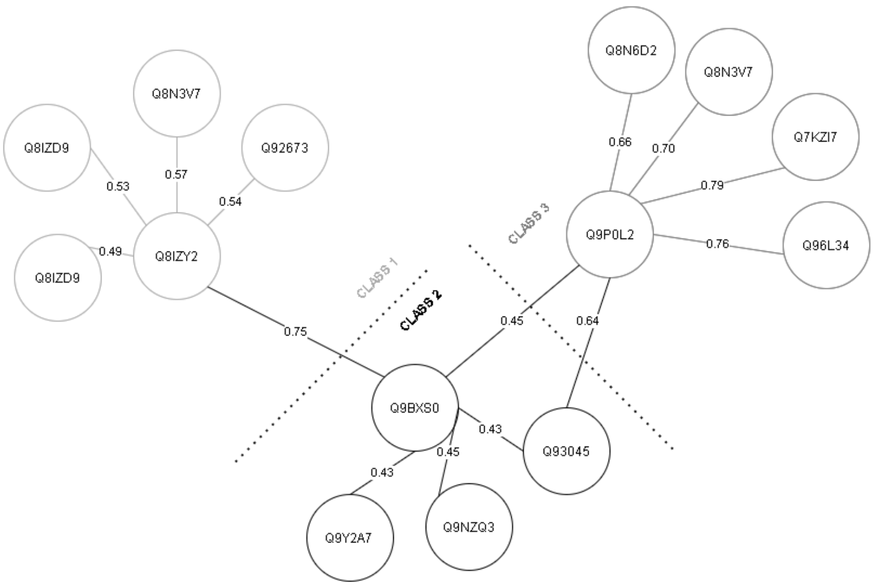


Figure 6: The result of Q8IZY2 and Q9P0L2 expansion by BP dataset.

[5] E. Eden, R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini. GOrilla: A tool for discovery and visualization of enriched go terms in ranked gene lists. *BMC Bioinformatics*, 10(1):1–7, 2009.

[6] Gene Ontology Consortium. The gene ontology project. *Nucleic Acids Research*, 36(suppl_1):D440–D444, 2008.

[7] M. Goyal, T. Knackstedt, S. Yan, and S. Hassanpour. Artificial intelligence-based image classification for diagnosis of skin cancer: Challenges and opportunities. *Computers in Biology and Medicine*, page 104065, 2020.

[8] D. W. Henderson. Venn diagrams for more than four classes. *The American Mathematical Monthly*, 70(4):424–426, 1963.

[9] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one*, 9(6):e98679, 2014.

[10] D. Lin. Extracting collocations from text corpora. In *Proceedings of the First Workshop on Computational Terminology*, pages 57–63, 1998.

[11] J. MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297, 1967.

[12] B. T. McInnes and T. Pedersen. Evaluating measures of semantic similarity and relatedness to disambiguate terms in biomedical text. *Journal of Biomedical Informatics*, 46(6):1116–1124, 2013.

[13] S. I. O'Donoghue, A.-C. Gavin, N. Gehlenborg, D. S. Goodsell, J.-K. Hériché, C. B. Nielsen, C. North, A. J. Olson, J. B. Procter, D. W. Shattuck, et al. Visualizing biological data—now and in the future. *Nature Methods*, 7(3):S2–S4, 2010.

[14] W. Poewe, K. Seppi, C. M. Tanner, G. M. Halliday, P. Brundin, J. Volkmann, A.-E. Schrag, and A. E. Lang. Parkinson disease. *Nature Reviews Disease Primers*, 3(1):1–21, 2017.

[15] D. Szklarczyk, J. H. Morris, H. Cook, M. Kuhn, S. Wyder, M. Simonovic, A. Santos, N. T. Doncheva, A. Roth, P. Bork, et al. The STRING database in 2017: Quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research*, page gkw937, 2016.

[16] UniProt Consortium. UniProt: A hub for protein information. *Nucleic Acids Research*, 43(D1):D204–D212, 2015.

[17] M. Vailati-Riboni, V. Palombo, and J. J. Loor. What are omics sciences? In *Periparturient Diseases of Dairy Cows*, pages 1–7. Springer, 2017.

[18] T. D. Veenstra. Omics in systems biology: Current progress and future outlook. *Proteomics*, 21(3-4):2000235, 2021.

[19] J. Z. Wang, Z. Du, R. Payattakool, P. S. Yu, and C.-F. Chen. A new method to measure the semantic similarity of GO terms. *Bioinformatics*, 23(10):1274–1281, 2007.

[20] Q. Wei, I. K. Khan, Z. Ding, S. Yerneni, and D. Kihara. NaviGO: Interactive tool for visualization and functional similarity and coherence analysis with gene ontology. *Bmc Bioinformatics*, 18(1):1–13, 2017.

[21] A. Xie, J. Gao, L. Xu, and D. Meng. Shared mechanisms of neurodegeneration in Alzheimer's disease and Parkinson's disease. *BioMed Research International*, 2014.

[22] J. Yan, S. L. Risacher, L. Shen, and A. J. Saykin. Network approaches to systems biology analysis of complex disease: Integrative methods for multi-omics data. *Briefings in Bioinformatics*, 19(6):1370–1381, 2018.

[23] G. Yu, F. Li, Y. Qin, X. Bo, Y. Wu, and S. Wang. GOSemSim: An R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*, 26(7):976–978, 2010.

[24] J. Zhu, Q. Zhao, E. Katsevich, and C. Sabatti. Exploratory gene ontology analysis with interactive visualization. *Scientific Reports*, 9(1):1–9, 2019.

Towards A Deep-Learning-Based Methodology for Supporting Satire Detection

Alfredo Cuzzocrea*
iDEA Lab, University of Calabria, Rende, Italy & LORIA, Nancy, France
alfredo.cuzzocrea@unical.it

Giosué Lo Bosco
Dept of Computer Science, University of Palermo & IEMEST, Palermo, Italy
giosue.lobosco@unipa.it

Mariano Maiorana
Dept of Computer Science, University of Palermo, Palermo, Italy
mariano.maiorana@community.unipa.it

Giovanni Pilato
ICAR-CNR, Palermo, Italy
giovanni.pilato@cnr.it

Daniele Schicchi
Istituto Euro-Mediterraneo di Scienza e Tecnologia, Palermo, Italy
danieleschicchi@iemest.eu

Abstract

This paper describes an approach for supporting automatic satire detection through effective deep learning (DL) architecture that has been shown to be useful for addressing sarcasm/irony detection problems. We both trained and tested the system exploiting articles derived from two important satiric blogs, Lercio *and* IlFattoQuotidiano*, and significant Italian newspapers.*

1 Introduction

Satire is a way of criticizing people (or ideas) by ridiculing them on political, social, and morals topics (e.g., [18]. Most of the time, such a language form is utilized to influence people's opinions. It is a figurative form of language that leverages comedic devices such as *parody* (i.e to imitate techniques and style of some person, place or thing), *exaggeration* (i.e to represent something beyond normality make it ridiculous), *incongruity* (i.e to present things that are absurd concerning the context), *reversal* (i.e to present the opposite of normal order), *irony/sarcasm* (i.e to say something that is the opposite of what a person mean). Moreover, satire masks emotions like irritation and disappointment by using ironic content.

The easy way of denouncing political and societal problems exploiting humor has brought consensus to satire that has been widely accepted. It leads people to constructive social criticism, to participate actively in the socio-political life, representing a sign of democracy. Unfortunately, the ironic nature of satire tends to mislead subjects that can believe the humorous news as they were real; therefore, satirical news can be deceptive and harmful.

Detecting satire is one of the most challenging computational linguistics tasks, natural language processing, and social multimedia sentiment analysis. It differs from irony detection since satire *mocks* something or someone, while irony is intended to be a way for causing laughter. Tackling such a task means both to pinpoint linguistic entities that characterize satire and look at how they are used to express a more complex meaning.

As satirical texts include figurative communication for expressing ideas/opinions concerning people, sentiment analysis systems may be negatively affected. In this case, satire should be adequately addressed to avoid performances degradation of such systems, mainly if sar-

*This research has been made in the context of the Excellence Chair in Computer Engineering — Big Data Management and Analytics at LORIA, Nancy, France

DOI reference number: 10.18293/DMSVIVA2021-016

casm/irony is used [1]. Moreover, reliably detecting satire can benefit many other research areas where figurative language usage can be a problem, such as *Affective Computing* [15]. An autonomous way of detecting satire might help computers interpret human interaction and notice its emotional state, improving the human-computer experience.

In this paper, we tackle automatic satire detection through effective deep learning (DL) architecture that has been shown to be effective for addressing the sarcasm/irony detection problem. The Neural Network (NN) exploits articles derived from two important satiric blogs, *Lercio* and *IlFattoQuotidiano*, and major Italian newspapers. The dataset has been specifically created for the task, and it includes news concerning similar topics. Experiments show an optimal performance achieved by the network that is capable of performing well on satire recognition. The network demonstrates the ability to detect satire in a context where it is not marked as in IlFattoQuotidiano. In fact, in this special case, news are so realistic that they seem to be true. [15]. An autonomous way of detecting satire might help computers interpret human interaction and notice its emotional state, improving the human-computer experience.

On the other hand, studying these techniques as combined with the emerging *big data trend* (e.g., [11, 10, 6, 8, 9]) is an interesting challenge.

The extended version of this paper appears in [7].

2 The Overall Proposed Methodology

Recognizing *satire* can be modeled as a classification task subdividing *satiric* and *non-satiric* articles in two different classes. Such a task has been widely tackled by using machine learning algorithms, and it has been shown that it is important to consider various aspects related to the application domain. For what concerns the subject problem, many factors should be taken into account: the way the text is represented and how it is structured (sec. 2.1), the model's architecture for tackling the task and its tuning (sec 2.2 and 2.3). Le Hoang Son et al. [13] have introduced a deep learning model that promises optimal performances for detecting sarcasm/irony. We believe that such a network can also help recognizing the main aspects of the satire; a detailed description is given in sec. 2.2.

2.1 Preprocessing

The preprocessing phase deals with the input arrangement to make it analyzable to the model as best as possible. Most of the time, the text is changed by removing punctuation marks, stop-words, etc. In this case, since the articles have been harvested from online resources we focused on the removal of the *author's name*, *HTML tags*, *hyperlinks*, and *hashtags*. Subsequently, the input text is split into tokens (i.e., words and punctuation marks) using NLTK [1]. To level out the lengths of the articles, we have analyzed the cumulative frequency of the length of the texts, and then we have selected a value $L = 4500$ words such that we considered 95% of the entire set of articles. Finally, each token is mapped to a 300-dimensional space by a pre-trained embedding tool that relies on FastText [3, 12]. Therefore, each article is represented by a matrix of real values of size $(L, 300)$. We crop texts longer than L, and we pad with 0s texts that are shorter.

2.2 Architecture

The network's architecture is inspired from the one presented by Le Hoang Son et al [13], that exploits *Bidirectional Long Short Term Memory* (BiLSTM), *Soft Attention Mechanism*, *Convolutional* NNs, and *Fully Connected* NNs. Moreover, such a model consider five different auxiliary characteristics that have been shown to be relevant to sarcasm/irony detection: number of exclamation marks (!), number of question marks (?), number of periods (.), number of capital letters, number of uses of *or*. A complete model representation is given in figure 1.

2.2.1 Input Layer

The first network's layer is the *Input* layer which manage the pre-processed text in order to allow the analysis by the BiLSTM.

2.2.2 BiLSTM Layer

BiLSTM is composed of two LSTM layers which examine respectively the input sequence in *forward* (from the first token x_0 to the last one x_T) and *backward* (from the last token x_T to the first one x_0) ways. LSTM *cell*, is a neural unit created specifically for overcoming the vanish/exploding gradient problem [2] that affects the training phase by using the backpropagation through time algorithm. The *cell* is composed of a set of *gates* (i.e input, forget, and output gate) which control the flow of information. The *forget* gate deals with choosing the information part should be kept and what should be gotten rid, the *input* gate proposes new information that is worth to be considered, and the *output* gate mix the contributes given by both the *input* and *forget* gates for creating the final cell's output. LSTM cell leverages two *feedback* loops (i.e internal and external) which allow to track the sequence of elements the cell has already analyzed through a sequence of internal states $h_1, \ldots, h_T$. The final output of the LSTM cell is its final internal state that is strictly dependent of the previous ones. The formulation of

[1] www.nltk.org

a LSTM unit, named *memory unit*, is described in by the following equations [14]:

$$\begin{aligned} f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\ i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\ c_t &= tanh(W_c x_t + U_c h_{t-1} + b_c) \\ s_t &= f_t \odot s_{t-1} + i_t \odot c_t \\ h_t &= tanh(s_t) \odot o_t \end{aligned}$$

where f_t, i_t, o_t are respectively the input, forget and output gates, the $\odot$ is the element-wise multiplication, the b_f, b_i, b_o, b_c are bias vectors, while tanh is the hyperbolic tangent and sigma is the sigmoid function.

The analysis of the input text in these two opposite directions create two representation of the input sequence: straight and reversed. BiLSTM layer merges the output of the two LSTM layers into a single output by concatenating them. The final vector, if examined through the soft attention, allow the network to capture the salient words considering the input text totally.

2.2.3 Soft Attention Layer

The Soft Attention is a mechanism that weight the input sequence elements on the basis of their relevance for the classification task, suggesting on what elements leverage for classifying the input correctly. It exploits the sequence of LSTM states during the examination of the input sequence.

The attention layer's output is the *context-vector*. It is computed as the weighted sum of the *attention weights* α_t and the LSTM's states $h_0, \ldots, h_T$. The approach is described by the following formulas, considering w_α the weights matrix:

$$\begin{aligned} z_t &= h_t w_\alpha \\ \alpha_t &= \frac{e^{z_t}}{\sum_{i=1}^{T} e^{z_i}} \\ c &= \sum_{i=1}^{T} \alpha_i h_i \end{aligned}$$

In this case, the context-vector c is extended by concatenating the auxiliary features. Finally, one-dimensional vector C which contains the analysis of the BiLSTM layer and the Pragmatic features becomes the input of the next convolutional layer.

2.2.4 Convolutional Layer

We stacked three convolutional layers for the feature learning. Each convolving filter of size s slides over the input vector to compute a localized feature vector v_j for each possible word through a nonlinear activation function. For each filter, a transition matrix T is generated. Such a matrix is iteratively applied to a part of the input vector to compute the features as following :

$$v_j = f(\langle T, F_{j:j+s-1} \rangle + b_a)$$

where $\langle \cdot, \cdot \rangle$ is the inner product, $F_{g,l}$ is the part of the input vector which includes elements from position g to position l, b_a is a bias related to the specific filter, and f is a non linear function.

The output of the convolutional layers is a vector of features $v = v_1, v_2, \ldots, v_{n-s+1}$ where n is the length of the input vector.

A max-pooling layer then processes the convolutional layer's output. Such a layer extracts the largest computed feature for each filter, considering only the most relevant ones. The output layer then analyzes the output vector that included the selected features.

2.2.5 Output Layer

The output layer is a Fully Connected NN activated by Softmax. Such a layer takes as input the features extracted by the max-pooling layer. Employing the Softmax activation function computes the probability that the input text belongs to the either *satiric* or *non-satiric* class.

Table 1. List of the model's hyperparameters.

Embedding size	300
LSTM neurons	500
Batch size	10
Convolutional layers	3
Kernel size	3
Convolutional activation function	ReLU
Dropout BiLSTM	0.2
Dropout ConvNet	0.4
Optimization algorithm	Adam
Learning rate	0.0001
Dense layer neurons	350

2.3 Parameters

Hyperparameters have been chosen empirically and taking inspiration from [13, 1]. Different tries have shown that taking a small learning rate and using a small minibatch coupled with Dropout regularization factors helps the network improve its performance by diminishing the loss. A complete list of them can be found in table 1.

3 Conclusion and Future Works

Satire aims at criticizing either something or someone leveraging on comedic devices. Its automatically detec-

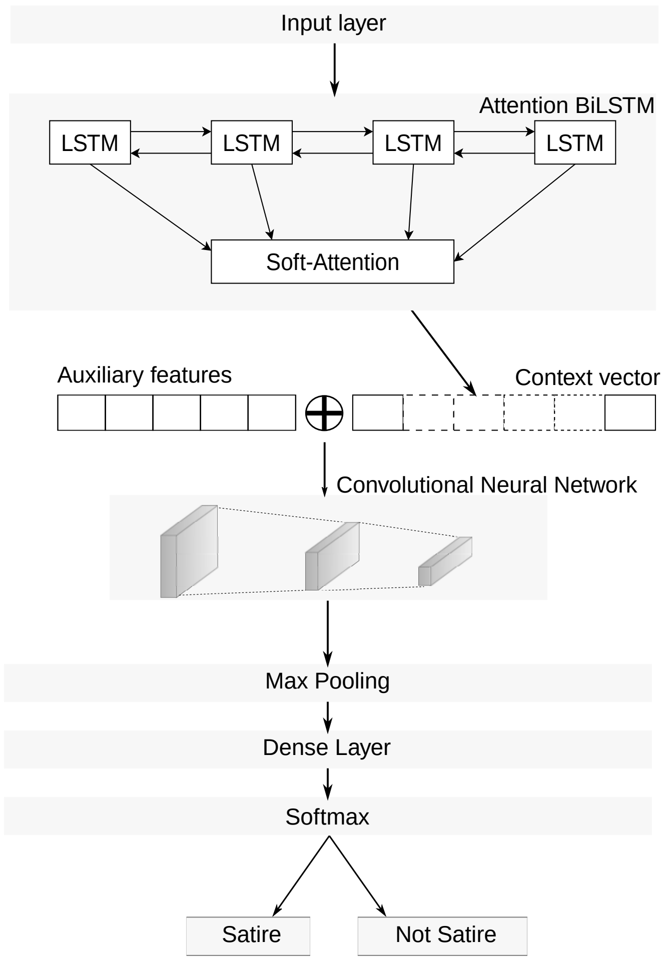


Figure 1. The representation of the Neural Network's architecture. The first layer manages the input in order to make it available for analysis. BiLSTM layer analyses the input in the forward and backward way to give a complete representation of the text. The attention mechanism is exploited for detecting the most relevant words for accomplishing the classification task. Its output is concatenated to the auxiliary features and then it is given as input to the convolutional layer. Such a layer extract prominent features, which are processed by a fully connected layer activated by softmax.

tion is a non-trivial task that have to consider the components it is composed such as *parody*, *exaggeration*, *reversal*, *irony/sarcasm* which often are related to stand-alone research topics.

In this paper, we have introduced a powerful DL model that tackles the satire detection problem by examining lexical, syntactical, and auxiliary features. To support the analysis by the system, we exploited an effective pre-trained embedding tool based on FastText.

Future work will further analyze the network's behavior by exploiting incremental data [5] and clustering [4]. Moreover, we are going to study how satire might affect the text comprehension [16] and if it might be reproduced through automatic creative processes [17].

References

[1] Teresa Alcamo, Alfredo Cuzzocrea, Giosue Lo Bosco, Giovanni Pilato, and Daniele Schicchi. Analysis and comparison of deep learning networks for supporting sentiment mining in text corpora. In *22th International Conference on Information Integration and Web-based Applications and Services (iiWAS2020)*, 2020.

[2] Y. Bengio, P. Simard, and P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, March 1994.

[3] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *CoRR*, abs/1607.04606, 2016.

[4] G. Casalino, G. Castellano, and C. Mencar. Incremental adaptive semi-supervised fuzzy clustering for data stream classification. In *2018 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, pages 1–7, 2018.

[5] Gabriella Casalino, Ciro Castiello, Nicoletta Del Buono, and Corrado Mencar. A framework for intelligent twitter data analysis with non-negative matrix factorization. *International Journal of Web Information Systems*, 2018.

[6] Alfredo Cuzzocrea. Improving range-sum query evaluation on data cubes via polynomial approximation. *Data & Knowledge Engineering*, 56(2):85–121, 2006.

[7] Alfredo Cuzzocrea, Giosue Lo Bosco, Mariano Maiorana, Giovanni Pilato, and Daniele Schicchi. A novel approach for supporting italian satire detection through deep learning. In *Flexible Query Answering Systems - 134th International Conference,*

FQAS 2021, Bratislava, Slovakia, September 19–24, 2021, Proceedings, Lecture Notes in Computer Science. Springer, 2021.

[8] Alfredo Cuzzocrea and Ugo Matrangolo. Analytical synopses for approximate query answering in olap environments. In *International Conference on Database and Expert Systems Applications*, pages 359–370. Springer, 2004.

[9] Alfredo Cuzzocrea, Rim Moussa, and Guandong Xu. Olap*: effectively and efficiently supporting parallel olap over big data. In *International Conference on Model and Data Engineering*, pages 38–49. Springer, 2013.

[10] Alfredo Cuzzocrea, Domenico Saccà, and Paolo Serafino. A hierarchy-driven compression technique for advanced olap visualization of multidimensional data cubes. In *International Conference on Data Warehousing and Knowledge Discovery*, pages 106–119. Springer, 2006.

[11] Alfredo Cuzzocrea and Paolo Serafino. Lcs-hist: taming massive high-dimensional data cube compression. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, pages 768–779, 2009.

[12] Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. Learning word vectors for 157 languages. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[13] Le Hoang Son, Akshi Kumar, Sangwan Raj Saurabh, Anshika Arora, Anand Nayyar, and Mohamed Abdel-Basset. Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network. *IEEE Access*, 7:23319–23328, 2019.

[14] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[15] Rosalind W Picard. *Affective computing*. MIT press, 2000.

[16] Daniele Schicchi, Giosué Lo Bosco, and Giovanni Pilato. Machine learning models for measuring syntax complexity of english text. In *Biologically Inspired Cognitive Architectures Meeting*, pages 449–454. Springer, 2019.

[17] Daniele Schicchi and Giovanni Pilato. Wordy: a semi-automatic methodology aimed at the creation of neologisms based on a semantic network and blending devices. In *Conference on Complex, Intelligent, and Software Intensive Systems*, pages 236–248. Springer, 2017.

[18] Aman Sinha, Parth Patekar, and Radhika Mamidi. Unsupervised approach for monitoring satire on social media. In Prasenjit Majumder, Mandar Mitra, Surupendu Gangopadhyay, and Parth Mehta, editors, *FIRE '19: Forum for Information Retrieval Evaluation, Kolkata, India, December, 2019*, pages 36–41. ACM, 2019.

A Novel Framework for Supporting Mobile Object Self-Localization via Emerging Artificial Intelligence Tools

Alfredo Cuzzocrea*
iDEA Lab, University of Calabria, Rende, Italy & LORIA, Nancy, France
alfredo.cuzzocrea@unical.it

Luca Camilotti
University of Trieste, Trieste, Italy
camilotti@units.it

Enzo Mumolo
University of Trieste, Trieste, Italy
mumolo@units.it

Abstract

This paper focuses the attention on the problem of supporting mobile object self-localization. The proposed algorithm estimates the distance from the camera lens to the center of the landmarks using only ceiling vision. Localization can be easily obtain from such distance estimations. Projections are geometrically described and the distance estimation is based on the pixels mapping information obtained by a two-dimensional dynamic programming algorithm (2D-DPA).

1 Introduction

Self-localization of mobile objects is a fundamental requirement for autonomy (e.g., [25]). Mobile objects can be for example a mobile service robot, a motorized wheelchair, a mobile cart for transporting tasks or similar. Self-localization represents as well a necessary feature to develop systems able to perform autonomous movements such as navigation tasks. Self-localization is based upon reliable information coming from sensor devices situated on the mobile objects. There are many sensors available for that purpose. The early devices for positioning are rotary encoders. If the encoders are connected to wheels or legs movement actuators, relative movements of the mobile object during its path [1] can be measured. Then, mobile object positioning can be obtained with dead-reckoning approaches. Dead reckoning [1] is still widely used for mobile robot positioning estimation. It is also true that dead-reckoning is quite unreliable for long navigation tasks, because of accumulated error problems. Other popular sensor devices for self-localization are laser or sonar based range finder devices and inertial measurement devices. In outside scenarios the most popular approaches are based on Global Positioning System (GPS). Due to the importance of self-localization, many other solutions for indoor environment have been proposed so far with different cost and accuracy characteristics. For example the Ultra Wide Band radio signal indoor localization systems [2], or the Bluetooth-based angle of arrival radio devices [3], or a combination of them. However these systems have serious limitations in cost and reliability, respectively. Another important type of sensors which may be used for cost effective self-localization are the CCD cameras, which require computer vision algorithms for localization such as for example visual odometry, [14]. Mobile objects vision self-localization is currently an open research field [5] and an increasing number of new methods are continuously proposed. As a matter of fact we have to consider that self-localization of mobile objects requires centimeter-level accuracy and Computer Vision is one of the most cost-effective techniques able to reach that accuracy. Consequently, some surveys of Computer Vision based self-localization techniques appeared recently in the literature, [6].

In this paper we describes a novel Computer Vision algorithm for estimating the distance from the camera lens to the center of ceiling landmarks with circular shape using a monocular low cost webcam. From the distance, mobile object localization approaches can be easily developed and a simple example is provided in this paper. The images of the ceiling landmarks are projected on the image plane of the camera. The projection is analytical described, but the projections distortions, which may arise especially when low cost devices are used, may affect the results. To take into account the projection distortions in order to obtain a better precision of the results, we use an approximation of

*This research has been made in the context of the Excellence Chair in Computer Engineering – Big Data Management and Analytics at LORIA, Nancy, France

DOI reference number: 10.18293/DMSVIVA2021-019

the two-dimensional dynamic programming (2D-DPA) algorithm [4] which finds a sub-optimal mapping between the image pixels of the ceiling landmarks and the image plane pixels of the projected landmarks. Since optimum 2D-DPA is NP-complete, in fact, many approximations have been developed. For example, the 2D-DPA technique described by Levin and Pieraccini in [16] has an exponential complexity in the image size, while Uchida and Sakoe describe in [15] a Dynamic Planar Warping technique with a complexity equal to $O(N^3 9^N)$. Lei and Govindaraju propose in [17] a Dynamic Planar Warping approximation with a complexity of $O(N^6)$. However each approximation has some limitation in terms of continuity of the mapping. In this paper we use a approximation of the optimum 2D-DPA with a complexity of $O(N^4)$ [13] which is implemented on a GPU to obtain real-time performance. When the landmark is far from the camera or if the environments has low lighting conditions, an high quantization noise may arise in acquired images. However the algorithm we describe in this paper is particularly robust against noise due especially to the use of two-dimension DPA.

Another interesting aspect of our research consists in considering the emerging integration of these topics with the innovative *big data trend* (e.g., [7, 8, 9]). Here, the main research perspective is that to take into account the well-known *3V model* of big data, including *volume*, *velocity* and *variety*.

The extended version of this paper appears in [22].

2 Problem Formulation

We show in Figure 1 a mobile object in an indoor environment. The movable object is equipped with a camera set tilted towards the ceiling at an angle φ. We call h the distance between the camera and the ceiling. Moreover in Figure 2 the horizontal and vertical angles of view of the camera, called θ_x, and θ_y respectively, are highlighted.
The direction towards which the camera is oriented is shown with the 'Camera Direction' arrow. The ceiling landmark is shown in Figure 1 with a segment with a greater thickness and the image plane of the camera is shown with a segment orthogonal to the camera direction. The ceiling landmark is projected to the landmark on the image plane. The visual landmarks positioned on the ceiling used in this approach are the lighting holders shown as that shown in Figure 3. We choose landmarks with isotropic shapes on the plane because in this way the distortion components due to image rotation can be eliminated. The simpler isotropic shape is the circle. As shown in Figure 3, the lines of pixels on the image plane are all parallel to the reference abscissa on the ceiling plane regardless of the angle of the camera with respect to the landmark. It is important to remark that each landmark must be distinguishable from the others and its co-

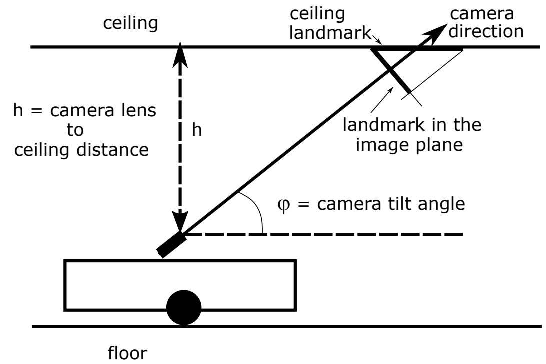


Figure 1. A mobile object with a camera on it, tilted toward the ceiling.

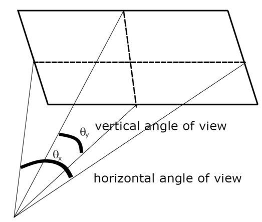


Figure 2. The horizontal and vertical angles of view of the camera.

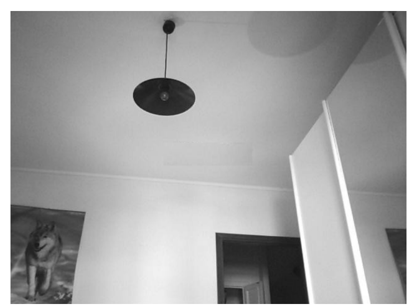


Figure 3. An example of the circular lamp holder used in this paper

ordinate in the global reference system must be known. A schematic representation of a mobile object and some landmarks with the orthonormal reference system centered on the camera lens is shown in Figure 4. The reference abscissa changes dynamically in relation to the direction of the focal axis. The reference abscissa, in fact, is always normal to the focal axis and at the same time it is parallel to the horizon.

The landmarks must be distinguishable from each other. There are many possible solutions for making the landmarks distinct. A simple possibility is to paint each holder with a different color. More recently, the characteristic fre-

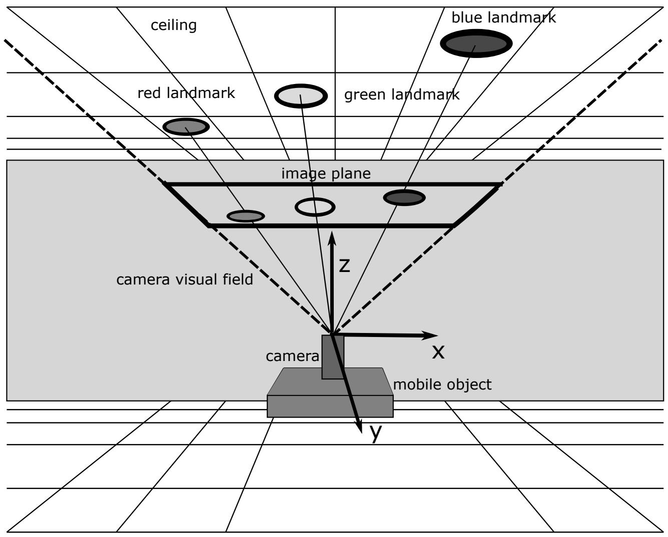


Figure 4. Schematic representation of orthonormal reference system, landmarks and image plane.

quency of fluorescent lights has been used, for instance in [11]. In this paper we used the simplest solution, namely we painted adjacent lamp holders with different colors. For this reason the landmarks in Figure 4 are represented with different colors, where for simplicity the three circular landmarks positioned on the ceiling are colored in red, blue and green. Figure 4 shows that the landmarks which fall within the visual field of the camera are projected onto the image plane of the camera. Of course we know in advance the physical position of each landmark in the global reference system. On the other hand the landmark colours can be detected using well known computer vision techniques.

3 Intelligent Transformations

The projective transformation is the linear transformation of coordinates reported in (1).

$$p' = Tp \tag{1}$$

where p represents a generic point in space expressed in homogeneous coordinates, relative to the orthonormal reference system S described by the quadruple $(O, \hat{\imath}, \hat{\jmath}, \hat{k})$. The projected point p' is expressed in coordinates relative to the reference system S' described by the quadruple $(O', \hat{\imath}', \hat{\jmath}', \hat{k}')$, where $\hat{\imath}' = \hat{\imath}$, $\hat{\jmath}'$ has the direction of the segment $\overline{MQ}$ and $\hat{k}'$ has the direction of the normal to the segment $\overline{MQ}$.

Since p is expressed with the three components (x_p, y_p, z_p) and p' has the three components $(x_{p'}, y_{p'}, z_{p'})$, eq. (1) can be also written as follows

$$\begin{pmatrix} x_{p'} \\ y_{p'} \\ z_{p'} \end{pmatrix} = T \begin{pmatrix} x_p \\ y_p \\ z_p \end{pmatrix} \tag{2}$$

Such a transformation maintains the properties of collinearity, that is, the points which in S belong to a line, are aligned in a line also in S'. However, projective transformation may not be defined for every point of S, in the sense that some points could be mapped in S' at infinity.

Let us view Figure 4 from the left side, that is the $y - z$ plane of the orthonormal reference system which has its origin coinciding with the center of the camera lens. This plane is highlighted in Figure 5, where the ceiling is at $z = h$, and the field of view of the camera is shown with points M and E. Let us assume that a landmark falls within the vertical angle of view. Then, the center of the landmark is the point C. On the other hand, if we view Figure 4 from the front side, that is the $x - z$ plane, we obtain the system shown in Figure 6. Of course the camera image plane, which is the plane normal to the focal axis in Figure 4, is shown with the segment $M - Q$ in Figure 5 and segment $G - I$ in Figure 6.

Suppose we fix a point P on the ceiling. If the point falls within the field of view of the camera it is shown as P in Figure 5. Let (p_x, p_y, p_z), with $p_z = h$, be the coordinates of P. The point P is projected to the image plane of the camera to the point P', which has coordinates $(x_{p'}, y_{p'}, z_{p'})$. Also the center of the landmark in Figure 5 is projected to the point C' and the segment $M - E$ is projected to the segment $M - Q$ in the image plane. In

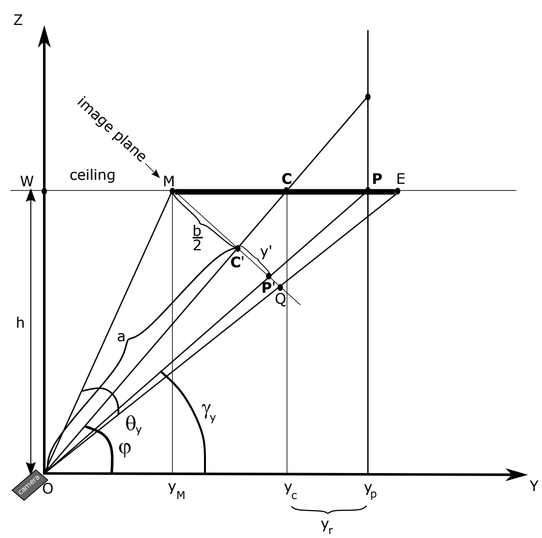


Figure 5. Plane $y - z$ in orthonormal reference system

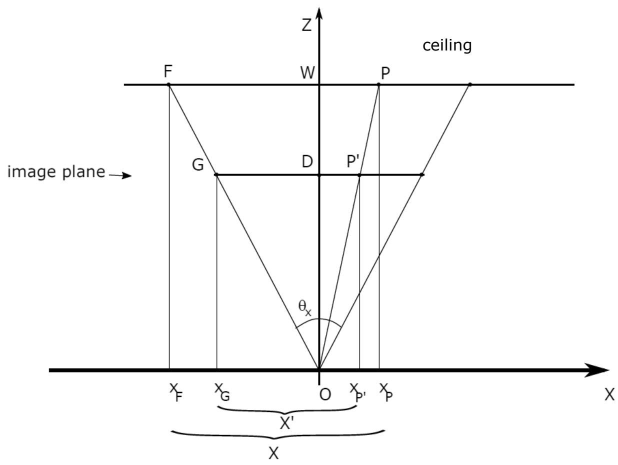


Figure 6. Plane $x - z$ in orthonormal reference system

this model, the focal distance of the device or other characteristic parameters are not taken into account. It is in fact a purely ideal model, which has the only purpose of deriving the relations that define the projective transformation from the orthonormal system whose origin coincides with the center of the camera lens to the image plane system. The latter is chosen independently of the characteristics of the camera. With reference to the Figures 5 and 6, we introduce the following geometric variables characteristic of the problem.

- $$\Phi = \varphi + \frac{\theta_y}{2} - \frac{\pi}{2} \quad (3)$$
- The distance a from the origin to the barycenter of the landmark projected on the image plane:
$$a = \overline{OC'} = \frac{h}{\sin(\varphi)} - h(\tan(\varphi) + \frac{1}{\tan(\varphi)})\cos(\varphi) \quad (4)$$
- The abscissa of the point P' on the image plane:
$$\frac{b}{2} = \overline{MC'} = \overline{C'Q} = h(\tan(\varphi) + \frac{1}{\tan(\varphi)}\sin(\varphi)) \quad (5)$$

Moreover, we define the following two variables:

$$G = -h\,(\tan\Phi\tan\varphi + 1) \quad (6)$$

and

$$F = h\,(\tan\varphi - h\tan\Phi) \quad (7)$$

We remark that the following considerations are based on three coordinate systems, namely an orthonormal reference system centered on the camera lens, shown in Figure 4, an orthonormal reference system on the image plane and a system on the ceiling plane which is simply translated by h with respect to that centered on the camera lens. In general, points on the systems centered on the camera lens and on the ceiling are denoted with a capital letter, such as **P**, while that on the image plane of the camera are denoted with a capital letter plus an apex such as **P'**. In this case, **P'** is the **P** point projected on the image plane. If we look at the landmark seen from the orthonormal reference system centered on the camera lens, its barycenter is located at (x_c, y_c). A generic point on the ceiling has coordinate (x, y) and the same point projected on the image plane is (x', y'). The coordinates of a generic point on the landmark is given relative to its barycenter: $(x = x_c + x_r)$ and $(y = y_c + y_r)$. According to Figures 6 and 5 the offsets x_r, y_r are projected to the image plane in x', y'.

Assume now we have an optimum mapping between images. In other words, assume that, having two images A and B, $A = \{a(i,j)|i,j = 1,...,N\}$ and $B = \{b(u,v)|u,v = 1,...,M\}$, we can estimate the mapping function

$$F(i,j) = \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x(i,j) \\ y(i,j) \end{bmatrix} \quad (8)$$

which maps each pixel (i,j) of one image to the pixel (u,v) of the other image such that the difference between the two images is minimized, as shown in (9).

$$min \sum\sum \|a(i,i) - b(u,v)\| \quad (9)$$

where $u = x(i,j)$ and $v = y(i,j)$. Such mapping is performed through a two dimensional Dynamic Programming operation [15]. 2D-DPA is the base of image matching algorithms called Elastic Image Matching. Unfortunately, the Elastic Image Matching operation is NP-complete [19]. For this reason we devise an approximation which reduces the 2D-DPA operation complexity to $O(N^4)$, as described below.

The barycenter of the landmarks, (x_c, y_c), are estimated using the following Proposition. By measuring the abscissa and ordinate (x', y') of a generic point on the landmark projected on the image plane we can estimate the coordinate (x_c, y_c) of the ceiling landmark using the following equations:

$$x_c = \frac{h\cos(\varphi - \gamma_y)(x' - g)}{a\sin(\gamma_y)} + g - x_r \quad (10)$$

$$y_c = \frac{aG + ay_r\tan(\varphi) - (y' - \frac{b}{2})(y_r + F)}{y' - \frac{b}{2} - a\tan(\varphi)} \quad (11)$$

A different estimation of the coordinates of the landmark barycenter is obtained for all the points $\boldsymbol{P}$ inside the landmarks. A sequence of barycenter coordinates $\boldsymbol{x_c}$, $\boldsymbol{y_c}$ are thus obtained, of which we compute the expected value. The algorithm is thus sequentially divided into two parts: estimation of $\mathbf{E}(x_c)$ and $\mathbf{E}(y_c)$ by measuring the dimension y' and x' of the distorted image on the image plane.

The distance from the camera lens and the landmark in the ceiling reference system is thus the following:

$$d = \sqrt{\mathbf{E}(x_c)^2 + \mathbf{E}(y_c)^2)} \quad (12)$$

with reference to Figures 6 and 5, where $C = (x_c, y_c, z_c)$ is the barycenter of the landmark in the reference system (O, i, j, k). We obtain he sub-optimal correspondence, pixel by pixel, between a reference image and a distorted image by means of approximated two dimensional dynamic programming, . Our algorithm therefore uses the deformation of the image to derive the distance of the landmark, i.e. it is intended to determine how the perspective has distorted the image.

The coordinates of the barycenter of the ceiling landmarks are obtained using the coordinate x' measured on the image plane and x_r using the mapping function, and in terms of y' and y_r. Clearly (x_r, y_r) and (x', y') are both known because they are derived from the coordinates of the pixels in the pattern and in the test images respectively. What associates the two pixels is the mapping relationship described in (8) obtained by 2D-DPA.

The characteristic that differentiates the algorithms present in the literature from the one developed in this paper is the statistical character of the obtained estimate. The algorithm based on dynamic programming is able to calculate a position estimate for each single pair of associated pixels from the mapping. The advantage is that a large number of points are used, which contribute to the calculation of the average distance value. This makes the estimate more truthful, especially when the landmark is very distant, which results in a smaller image and a greater quantization error.

4 2D-DPA-Based Image Mapping Technique

For the sake of coherence with what we write below, we repeat now the mapping considerations summarized above about images A and B using instead images X and Y. Given the two images, $X = \{x(i,j)\}$ and $Y = \{y(u,v)\}$, the mapping of one image to the other is represented by the operation

$$D(X,Y) = \min \sum_{i=1}^{N} \sum_{j=1}^{N} \|x(i,j) - y(u,v)\|$$

where $u = x(i,j)$, $v = y(i,j)$ is the mapping function between the pixels of X and Y. The quantity $D(X,Y)$ gives a distance between the image X and the optimally deformed Y, the optimal warping function $x(i,j), y(i,j)$ gives an interpretation of the image X according to the generation model Y.

Given the $i-$th row of the X image and the $j-$th row of the Y images, namely $Y_j = (y_{j,1}, y_{j,2}, \ldots, y_{j,N})$, $X_i = (x_{i,1}, x_{i,2}, \ldots, x_{i,N})$ respectively, the distance between the two rows is obtained by applying a 1D-DPA [18] for finding a warping among the two rows as described in (13). Here the map M' is, say, over (n, m) coordinates, so that $M'_l = ((i_l, n_l), (j_l, m_l))$.

$$d(X_i, Y_j) = \frac{\min\limits_{M'} \sum_{l=1}^{M'} d(M'_l)}{M'} = \frac{\min\limits_{M'} \sum_{l=1}^{M'} \|x_{i_l,n_l} - y_{j_l,m_l}\|}{2N} \quad (13)$$

Finally, the distance between the two images is obtained by (14). In this case the map $\overline{M'}$ is between all the rows of X and Y. As before, $|\overline{M'}|$ is the length of the path.

$$D(X,Y) = \frac{\min\limits_{\overline{M'}} \sum_k d(\overline{M'}_k)}{|\overline{M'}|} =$$

$$= \frac{\min\limits_{\overline{M'}} \sum_k d(X_{i_k}, Y j_k)}{|\overline{M'}|} = \frac{\min\limits_{\overline{M'}} \sum_k \frac{\min\limits_{M'} \sum_{l=1}^{|M'|} d(M'_l)}{2N}}{2N} =$$

$$= \frac{\min\limits_{\overline{M'}} \{\sum_k \min\limits_{M'} \sum_{l=1}^{|M'|} \|x_{i,n_l} - y_{j,m_l}\|\}}{4N^2} \quad (14)$$

Let us assume that the images are of equal size, that is $N \times N$ pixels. Then the length of the optimum path between the two images is equal to $2N$. The local distances in each point of this path is obtained with other 1D-DPA with paths of length $2N$. The total length is the sum of $2N$ along the $2N$ long path, giving $4N^2$ at the denominator. The complexity of the described operation is $O(N^2N^2) = O(N^4$ where N is the image dimension.

The algorithm described in this paper is summarized in the following Algorithm. The inputs of the algorithm are the two gray-scale images img_A and img_B which are the landmark on the image plane and on the ceiling respectively. We perform the 2D-DPA algorithm on these two images to obtain the mapping function as result. the mapping function is represented with a linked list where each node is the map related to the two pixels. The function $get()$ give as result the value of the pixel on the image indicated as input and is used to get the values of the two pixels linked by the map on the two landmark images. To decide if the pixel is a landmark pixel or not, we consider their gray levels. The landmarks have a lower values with respect to the environment and thus if the pixel values is less then a thereshold, the pixel is a landmark pixel.

Input: img_A, img_B
Output: $distance$
img=Detect(img_A); ▷ get the landmark in the image plane
id=identify(img); ▷ identify the landmark
head=2D-DPA(img_A, img_B);
ptr=head; ▷ head is the list of mapping function
repeat
  $pixA = get(img_A, ptr)$; ▷ pixel of img_A
  $pixB = get(img_B, ptr)$; ▷ pixel of img_B
  if $(pixA \leq L)\&\&(pixB \leq L)$ **then** ▷ the pixels are in the landmark
    Compute x_c, y_c with (10) and (11)
    $sum_y + = y_c$;
    $sum_x + = x_c$;
    counter++;
  $ptr = ptr \rightarrow next$;
until $ptr == NULL$
$y_c = y_s um/counter$;
$x_c = x_s um/counter$;
$distance = \sqrt{x_c^2 + y_c^2}$;
return $distance$

5 Conclusions and Future Work

In this paper we present an algorithm to measure the distance of a mobile object to the lightning lamps used as ceiling landmarks in indoor environment. The algorithm has many attractive features, mainly the accuracy, which is better than many other visual-based algorithms. Also, the distance measurements algorithm is robust against noise. Quantization noise can be high in low lighting condition of the environment and if the distance from landmarks and camera is high.

This paper naturally opens to the development of localization algorithms based on our distance estimation algorithm. The global localization is in fact under development. Another open important issue is the landmark placement (e.g., [23, 24]). Finally, the estimation of the orientation of the mobile object is another fundamental problem not addressed in this paper. The use of the characteristic frequencies of fluorescent lamp is an interesting method to identify the landmarks. In this case, adaptive and artificial intelligence metaphors, perhaps inherited by different scientific context (e.g., [20, 21]), may be considered. Future works will be focused on these open points.

References

[1] Sean Campbell, Niall O' Mahony, Anderson Carvalho, Lenka Krpalkova, Daniel Riordan and Joseph Walsh, Where am I? Localization techniques for Mobile Robots A Review, 6th ICMRE International Conference, Barcelona, Spain, February 12-15, 2020, pp.43–47

[2] Chengyang He, Yinqiu Xia, Chengpu Yu and Chaoyang Jiang, A multi-hop distributed indoor localization algorithm for ultra-wide-band sensor network, 16th ICARCV International Conference, Shenzhen, China, December 13-15, 2020, pp.1335–1340

[3] Gaurav Kumar, Vrinda Gupta and Rahul Tank, Phase-based Angle estimation approach in Indoor Localization system using Bluetooth Low Energy, 2020 ICOSEC International Conference, Trichy, India, September 10-12, 2020, pp.904–912

[4] Glasbey, C.A., Two-dimensional generalisations of dynamic programming for image analysis, Stat Comput Vol. 19, No. 49, 2009, pp. 49–56

[5] Wei A. Shang, Survey of Nobile Robot Vision Self-localization, Journal of Automation and Control Engineering, Vol.7, No.2, 2019, pp.98–101

[6] Anca Morar, Alin Moldoveanu, Irina Mocanu, Florica Moldoveanu, Ion Emilian Radoi, Victor Asavei, Alexandru Gradinaru and Alexandru Butean, A Comprehensive Survey of Indoor Localization Methods Based on Computer Vision, Sensors, Vol.20, No.9, 2020, pp.1–36

[7] G. Chatzimilioudis, A. Cuzzocrea, D. Gunopulos, N. Mamoulis, A novel distributed framework for optimizing query routing trees in wireless sensor networks via optimal operator placement, Journal of Computer and System Sciences 79 (3), 2013

[8] A. Cuzzocrea, S. Mansmann, OLAP visualization: models, issues, and techniques, Encyclopedia of Data Warehousing and Mining, Second Edition, 2009

[9] A. Cuzzocrea, C. De Maio, G. Fenza, V. Loia, M. Parente, OLAP analysis of multidimensional tweet streams for supporting advanced analytics, Proceedings of the 31st ACM SAC Symposium, 2016

[10] Marios Avgeris , Dimitrios Spatharakis , Nikolaos Athanasopoulos , Dimitrios Dechouniotis and Symeon Papavassiliou, Single Vision-Based Self-Localization for Autonomous Robotic Agents, 7th FiCloud International Conference, pp.123–129, 26-28 Aug., Istanbul, Turkey, 2019

[11] Zhang, C.; Zhang, X., Visible Light Localization Using Conventional Light Fixtures and Smartphones, IEEE Trans. Mob. Comput. 18, 2018, 2968–2983

[12] Yasuo Ogawa and Joo-Ho Lee and Syunji Mori and Akira Takagi and Chie Kasuga and Hideki Hashimoto, The positioning system using the digital mark pattern-the method of measurement of a horizontal distance, 1999 IEEE SMC International Conference, Tokyo, Japan, 1999, pp.731–741

[13] Alfredo Cuzzocrea and Enzo Mumolo and Daniel Pirro and Gianni Vercelli, An efficient CUDA-based approximate two-dimensional dynamic programming algorithm for advanced computer vision applications, 2016 IEEE SMC International Conference, Budapest, Hungary, October 9-12, 2016, pp.4251–4258

[14] Ming He and Chaozheng Zhu and Qian Huang and Baosen Ren and Jintao Liu, A review of monocular visual odometry, Vis. Comput., 36, 5, 2020, pp. 1053–1065

[15] Seiichi Uchida and Hiroaki Sakoe, A monotonic and continuous two-dimensional warping based on dynamic programming, Fourteenth ICPR International Conference, Australia, 16-20 August, 1998, pp.521–524

[16] Esther Levin and Roberto Pieraccini, Dynamic planar warping for optical character recognition, 1992 IEEE ICASSP International Conference, San Francisco, California, USA, March 23-26, 1992, pp.149–152

[17] Hansheng Lei and Venu Govindaraju, Direct Image Matching by Dynamic Warping, 2004 IEEE CVPR International Conference, Washington, DC, USA, June 27 - July 2, 2004

[18] K.Vinotha, Bellman Equation in Dynamic Programming, International Journal of Computing Algorithm, 05, 1, 2016, pp = 35–37

[19] Daniel Keysers and Walter Unger, Elastic image matching is NP-complete, Pattern Recognit. Lett., 24, 1-3, 2003, pp. 445–453

[20] M. Cannataro, A. Cuzzocrea, A. Pugliese, XAHM: an adaptive hypermedia model based on XML, Proceedings of the 14th SEKE International Conference, 2002

[21] A. Cuzzocrea, Combining multidimensional user models and knowledge representation and management techniques for making web services knowledge-aware, Web Intelligence and Agent Systems: An international journal 4(3), 2006

[22] Alfredo Cuzzocrea and Luca Camilotti and Enzo Mumolo, An Innovative Monocular Mobile Object Self-localization Approach Based on Ceiling Vision, AINA 2021 International Conference, Toronto, ON, Canada, 12-14 May, 2021, pp. 713–730

[23] Jerry Wang and Shiaofen Fang and Meie Fang and Jeremy Wilson and Noah Herrick and Susan Walsh, Automatic Landmark Placement for Large 3D Facial Image Dataset, 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, December 9-12, 2019, pp. 5088–5093

[24] Javier Perez-Ramirez and Deva K. Borah and David G. Voelz, Optimal 3-D Landmark Placement for Vehicle Localization Using Heterogeneous Sensors, IEEE Trans. Veh. Technol., vol. 62, no. 7, 2013, pp. 2987–2999

[25] Jason A. Janét and Ren C. Luo and Caglan M. Aras and Michael G. Kay, Sonar windows and geometrically represented objects for mobile robot self-referencing, Proceedings of 1993 IEEE/RSJ IROS International Conference, Tokyo, Japan, July 26 - 30, 1993, pp. 1324–1331

Towards a Project Management-enabled learning environmentS: a case study

Andrea Molinari
Univ. of Trento, Dept. Of Industrial Engineering (ITALY)
Lappeenranta University of Technology,
(Finland) andrea.molinari@unitn.it

***Abstract*—The paper addresses the issue of managing activities within virtual learning communities, a topic virtually ignored by LMS platforms and systems that offer collaboration services, not least social networks. The question is not surprising in itself, as the issue of managing time, costs and resources, linked to the discipline of Project Management, is historically a difficult issue for the IT world, from a cultural, technical and organizational point of view. The issue becomes even more complex if applied to the management of learning processes, where in any case we have times, deadlines, associated costs, resources that work there etc. In this work we will present a component for the management of projects, activities and resources within a collaborative platform used mainly in e-learning contexts, based on the metaphor of virtual learning communities. The introduction of this new component within the system wants to respond to the need to manage collaborative activities between learners, providing a tool for managing and controlling the progress of the activities assigned to the various members of the community**

Keywords-component: Project Management, virtual communities, social network

I. Introduction

In this paper, we present some services provided inside a virtual community platform as extensions towards managing time of the users, specifically teachers, students and administrative personnel involved in educational tasks. These services are rarely available (if any) inside Learning Management Systems (LMS), or have the form of simple tasks lists, to-do list or educational tasks scheduled in some workflow systems. Functionalities that provide support to (educational) projects are something significantly different. Project Management is a well-established discipline, where we can find different standards, guidelines and certification processes provided by world-wide organizations. The advent of web platforms that are claiming to provide Project Management services has changed the scenario, with a lot of competition in the market of Project Management services. If we aim to provide a Project Management approach to the conduction of educational activities, we find different solutions: a) “traditional” Project Management platforms, fully equipped with standardized services for scope, time, cost and resource management, but with a lack of support to collaboration and social processes inside project teams b) a lot of web platforms that support collaboration processes [1] and claim to support Project Management processes but that are not very efficient in this sense, lacking basic mechanisms like critical path method (CPM), timesheet cost and resource management, milestones, constraints, deadlines etc. [2]

This paper explores what should be needed, in the opinion of the author, inside a Learning Management System and inside educational contexts to support the many activities that fall into the definition of “project” [3]. We designed and realized a set of services specifically oriented towards the Project Management concepts and methodological tool and added them to a virtual community platform (with all basic collaboration, social media-oriented services). The platform, from providing traditional collaboration, education-oriented services, has been extended with a fully-fledged set of tools compatible with the most used Project Management standards, including task planning, costs, and resource management functionalities.

Since the advent of open source LMSs, Technology-Enhanced Learning (TEL) is a consolidated research topic, and a lot of tools and techniques are available for creating, delivering and managing online educational paths with plenty of solutions for every educational institution. What is less available inside the technological solutions that support our daily educational tasks is the integration with Project Management tools and techniques to manage tasks, resources, and costs within educational settings.

We can find a plethora of platforms and cloud services available today, even for free, that support at different level activities related to a project, but in order to implement Project Management discipline suggestions inside educational processes, a teacher (or a student) is forced to exit from his learning environment (for example, Moodle, Forma.LMS or Sakai LMSs), and use an external platform to manage project. This, in turn, could be problematic as most of the items that are the subject of the educational project are inside the LMS, so the user is forced to duplicate the material. Then, if we consider not only the trivial aforementioned problem of document sharing, but we extend the analysis to other typical Project Management tasks (like for example checking deadlines), the situation is again complex: the educational platform and the Project Management platform are completely disconnected.

What we can mostly find inside LMS under the label “project” or “Project Management” is a sort of task-list or to-do list service, that must be managed directly by the user. In few other cases, the idea of managing projects is simply implemented with a set of steps of a predefined workflow, in some way linked each other, but this is clearly very limited even respect to elementary Project Management tools.

If we consider the world of education in its full complexity and what educators do every day, we can find plenty of activities that can be fully defined as a “project”, but strangely software

DOI reference number: 10.18293/DMSVIVA2021-020

platforms that are so useful for educational purposes are not providing adequate support to this.

II. PROJECTS IN EDUCATIONAL CONTEXTS

The de-facto standard in the Project Management discipline [3] clearly identifies what a "project" is: "a temporary endeavour undertaken to create a unique product, service, or result" [3], The discipline of Project Management [4] is the application of a collection of tools and techniques (such as the CPM and matrix organization) to direct the use of diverse resources toward the accomplishment of a unique, complex, one-time task within time, cost and quality constraints.

After many years of mostly being ignored by the software industry (with some notable exceptions), nowadays we can find on the market, many different tools that project managers can choose to plan and manage their projects. Nevertheless, the application of tools and techniques derived from the standards is rare inside these platforms, where the focus is mainly on extended functionalities for Agile and in general collaboration services, rather than focusing first on core services like critical path method (CPM), Critical Chain Method (CCM), PERT, etc.

On one side, the perspective of using a collaborative platform as a Project Management tool is very interesting and fascinating, but this should happen without compromises respect to what Project Management methods provided to millions of skilled project managers around the globe, thus limiting the power of Project Management concepts. Deadlines, cost management, task duration, resources, these and many others are problems that everyone involved in any kind of project-related activities are forced to face.

The next consideration relates to the presence of these tools inside another kind of collaboration software, i.e., LMS. If we look at educational contexts, there are plenty of educational tasks that are part of a "projects" in the plain sense of the definition, enriched with many other aspects that are typically managed in project contexts. The many relations among students, between students and teachers, between students and the educational institution or external organizations (like stages) can be very easily and profitably managed as projects. Moreover, using Project Management concepts and tools for students of any course would be a fundamental growth in any direction their future professions will lead them.

Everybody talks about team working, tasks, milestones, deliverables, scope, risks, i.e., uses the typical jargon of Project Management, but the application of this complex discipline has not been so widespread, and surely not applied extensively in education. According to several studies [5], even today that we have a wide range of tools and techniques available, projects are frequently out of time and/or out of budget [6]. The poor results shown by these (and other) researches in terms of successful performance are subject to different interpretations, the most common being "Project Management continues to fail because included in the definition are a limited set of criteria for measuring success, cost, time and quality, which even if these criteria are achieved simply demonstrate the chance of matching two best guesses and a phenomenon correctly" [6].

Project Managers, today, can choose among many techniques and software to plan and manage their projects. The widespread usage of network approaches, like Gantt charts, critical path method (CPM), etc. have simplified the planning and controlling steps, while Project Management software has reached a solid maturity level. In order to define the scope of the project, the Work Breakdown Structure has been adopted as a (graphical) tool to delimit the things that should be done in the project, separating them by unwanted and/or unpaid requests. WBS is the hierarchical decomposition of the work to be executed by the project team to fulfill the objectives of the project and make deliverables. It organizes and evaluates the overall scope of the project. Information for a WBS is basically taken from project objective statement, historical files of previous projects and project performance reports. An appropriate WBS encourages a systematic planning process, reduces the possibility of omission of key project elements and simplifies the project by dividing into manageable units.

Another area heavily interlaced with Project Management and relevant for our argumentation is collaboration. Collaborative spaces are available within the project team to contribute to the success of the project's objectives. New generation tools of Project Management enable this functionality. Project is led and developed by the whole team, and each member has complete information about the project, with all the related documents. Project's progress is visible to everyone anytime, according to permissions granted to the subject.

When the project manager is free from the routine tasks; s/he can put more efforts into project vision and choose the direction for the project development. The authors of [7] discuss methods and tools for collaborative Project Management; if we cope this element with the widely recognized collaborative nature of educational processes, we should expect a convergence of these two disciplines and the relative tools.

On the contrary, the only field where we have found the application of Project Management tools and techniques inside educational contexts is the production of learning objects [8]. Here the concept of Project Management is not focused on providing tools inside the LMS for the management of activities as a project, but rather on managing the creation of learning objects with the typical five phases of the lifecycle of a project (initiation, planning, executing, monitoring and controlling, closing). This means treating the production of learning objects using the project's lifecycle as stated in [3], but the LMS remains in the background with the traditional set of functionalities not equipped with Project Management functionalities.

In our opinion, educational processes in general (and not only the production of educational material) can be profitably managed using the pillars of the discipline. Following this idea, we have integrated in our self-made LMS an entire set of Project Management functionalities. The management of tasks within an educational environment shows a series of constraints and issues that need to be managed with appropriate tools, like those supplied by Project Management. The pandemic we are living demonstrated that it is very important to provide a precise workplan for students, that are not allowed to follow physical lectures. Our "Virtual communities" platform provides affiliated users a set of features strictly related with Project Management tools and techniques: a) define and manage projects and their scope b) prepare a fully-functional Work Breakdown Structure with predecessors and constraints c) implement the Critical Path Method in the calculation of start/finished dates and free/total slack d) assign resources to tasks and check their allocation e) assign and control costs associated to a project, with a combination of role-permission to a level that can properly administer the security, confidentiality and privacy of the activities.

The integration of these features inside a Learning Management system guarantees the increase of application fields for these platforms, allowing them to be used not only for traditional educational activities but also for more collaboration and cooperation-oriented tasks. The problem we see in today's collaboration platforms that claim to be Project Management-enabled platform is exactly the approximation and imprecision of implementation of Project Management services.

In our experience, this incompleteness causes users to start using the service and then abandoning it (and the platform consequently), or on the contrary, considering Project Management as the discipline of the coloured sticky notes attached on a Kanban board. Being the world of education very poor in terms of Project Management tools and techniques, this could be a very good way to improve also the awareness of educational actors about how to manage their interaction with the institution. Books and manuals are full of words like "educational projects", "educational tasks", "learning milestone", "educational deliverable", but even inside modern LMSs there's no real, native, theoretically-grounded services that could support educational actors in exerting their tasks under the umbrella of Project Management theories and tools.

3. PROJECT MANAGEMENT FUNCTIONALITIES IN LMS

The typical collaboration services available in enterprise platforms (like wikis, blogs, and collaborative planning tools. All these tools, together with file sharing (documents, reports, agendas, comments…) represent a clear stimulus to consider also an e-learning platform as a possible provider of support and services to PM. These services, like many others, are very common both in collaborative environments and in e-learning platforms, as presented in [9][10]. Likewise, many different situations in educational settings could take advantage of PM services. Some examples can be the following:

- the management of a thesis assigned to a student is a real project with tasks, milestones, deliverables, and costs (even if not directly sustained by somebody);
- a research project led by a teacher or researcher is, by definition, a project, involving again different resources, costs, deliverables, and milestones;
- an educational path and all the tasks that any participant has to manage is another example of a "project"
- a complex training path providing professionals with a certification at the end of the activities, with respective assistance of external resources and tutors, is a project from the perspective of the organizing institution. Here we have a typical mix of educational needs (the LMS's most traditional services) and PM tasks
- a massive open online course (MOOC) initiative, with all the tasks related to the various phases of creation, marketing deployment, execution, support, and final certification is a project, both for the institution that delivers the MOOC and for the participant that has to perform tasks, to pay attention to milestones, to respect deadlines.

The possibilities of using these tools and services inside LMSs is even bigger, if we imagine to use a Virtual Communities system not only for managing "communities" devoted to educational purposes but also in larger contexts typical of collaboration, like a research group, a recreation organization, a secretariat, a board of directors, a club, a sport team, etc. All these "communities" need services that are available inside LMSs (like document sharing, forums, wikis, FAQ, sync, and async communication, etc.).

E-learning became so popular thanks to many factors, like network availability, multimedia, increased power of client workstations, flexibility, low costs, etc., but the role of software platforms like Moodle™, Docebo™, Dokeus™, Sakay™, Webct™ is clearly a central role. These platforms have proven to be effective in contexts not necessarily connected to academic education, therefore posing the issue of the evolution of software platforms towards services that are not necessarily related to traditional academic tasks. Last but not the least, the integration of e-learning (or collaborative) software platforms with the rest of the information system of the hosting organization represents clear evidence of the role of software platforms today in education.

From a meta-architectural point of view, e-learning platforms have based their pillars on the idea of "course", or "class", meaning that the basic container for relationships among users of the platform is a virtual place that resembles in some way what happens in any educational organization: collecting people in a (virtual) classroom.

What clearly emerged in past studies [9] and from our preliminary experiments is a need of a different funding paradigm for software platform: the "community", or "virtual community". The virtual community is a container ready for didactic processes, but not only: research teams, recreation groups, friends, secretariats, the board of directors, colleagues, anything that could be an aggregation of people around scope using virtual spaces on the web. The core of the application is composed of some abstract entities, i.e., virtual communities as an aggregation of people to which some communication services are available in order to obtain certain objectives. "Online Communities" [10] is a space on the web devoted to a collaboration objective, populated by people who communicate with each other, using a series of communication systems. With this approach, it could be possible to represent all the hierarchical relationships between different types of communities (such as faculties, didactic paths, master degrees, courses, etc.). The main characteristics of a virtual community could be summed up as follows:

- a community is a composition of services for a virtual space of interaction involving end-users for that community

- the services are general applications that enable the users to communicate in a synchronous and asynchronous way, to publish contents, to exchange files, to coordinate events, etc.
- the potential services of a community are activated by a manager of the community according to the needs, and the users of a community can use them with different rights and duties
- the communities can be aggregated into larger communities with hierarchic mechanisms and infinite nesting levels
- the communities can be aggregated in an arbitrary way into larger communities disregarding the possible position of a hierarchical structure
- all users are recognized.

The addition of Project Management services inside e-learning came mainly from the experience of the team in the techniques of Project Management, on the one hand, but also from everyday tasks: consider, for example, as part of learning community college, the need for a teacher to coordinate a number of undergraduates involved in the long task of drawing up their thesis. The activities of the individual, the professor, or those that are shared between them have often intertwined/ associates and impose the need to manage time, deadlines, relationships, and mutual dependencies.

More complicated is the situation on the teacher's side, where s/he could have more thesis to follow, so more projects of this type to manage. We, therefore, believe that the lack of a tool of this kind can be solved naturally with valuable tools for planning and managing existing projects, but these: a) do not integrate platforms b) on average, they are complex c) they are much more appropriate for people with specific expertise in the complex and multifaceted discipline of Project Management.

4. IMPLEMENTING PROJECT MANAGEMENT SERVICES INSIDE AN LMS

The idea of implementing Project Management services inside learning contexts benefitted a lot of the availability of "Online Communities", the virtual spaces dedicated to each community (what we call the virtual community). This collaboration space is equipped with services provided to the users, so it was simply a matter of creating new services related with the Project Management (PM) discipline but aware of all the rest the platform provides, and integrate the PM services with this. Another fundamental factor from the virtual community that we used in assembling the new services was the concepts of "roles", "rights", and "permissions" that are assigned to each user for each separated community. This allows a fine-grained, sophisticated way of managing and controlling who is doing what on a certain task.

The Project Management services contain some sophisticated features that are typical of top software solutions, strictly related with the critical path method (CPM) and its calculation mechanisms. These options have been added in the recent version as a substantial improvement, because some users were complaining about the lack of the following options. These options (fig.1) allow to obtain the same results we can obtain, in terms of calculation of time, start, finish etc. with Project Management professional tools. These options are:

- defining a project calendar with working and non-working days
- defining the start date of the project from which the CPM will start to calculate all start/finish dates according to the predecessors
- default resources: this allows to set the resources you want to automatically assign when a new task is created
- Milestone: allows the project to use "Milestone", i.e., tasks with a duration equal to zero days

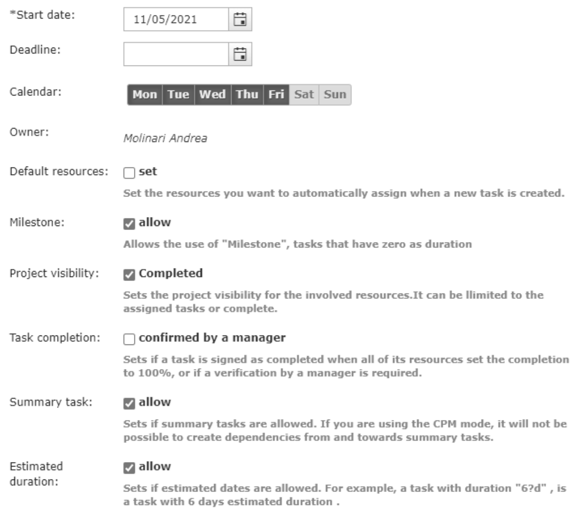


Fig.1 – Options for advanced settings of an educational project

- Project visibility: this option sets the project visibility for the involved resources. It can be limited only to the assigned tasks to that resources, or only when tasks are completed.
- Task completion: this sets the possibility for the created project to have task completion confirmed by a manager. This implies that a task is signed as completed when all of its resources set the completion to 100%, or if a verification by a manager is required.
- Summary task: allows the project to have the powerful feature typical of the WBS, i.e., tasks that summarized all data (start, finish, costs etc.) of the subtasks. If you are using the project in the CPM mode, it will not be possible to create dependencies from and towards summary tasks. This is partly a simplification respect to the full-fledged software where the CPM is implemented also for the summary tasks, but it's clearly more complicated in the calculation engine, and we have decided to skip this part in this version.
- Estimated duration: this is instead a very interesting niche feature available only in the most advanced software, but we consider it very important. This option sets the possibility of setting the duration of tasks in estimated days, i.e., days of duration considering a linear calendar of 365/366 days per year without non-working days. This feature dates are

allowed. For example, a task with duration "6?d" , is a task with 6 days estimated duration.

Tasks and users can also be shared among different communities, with the same inheritance mechanism. Users have on the one hand an institutional role inside the organization, and one or more functional roles in each community to which they participate. Examples of institutional roles are those of the classic academic institution (student teacher etc.) As examples of functional roles, we have the administrator of a community, participant, moderator, blogger, secretary, member, dean, writer, etc. Roles can be freely created, assigned with respective permissions for each service available in the platform, PM services included.

We therefore decided to add an extra feature that takes advantage of the possibilities of creation of a project inside a virtual community: member of the virtual community can be assigned as resources of the project (fig.2). This is very frequent (and logic, so to say) in educational communities, where the community itself exists because of the need of managing a project. A workgroup, for example, created with some students that have to perform a common educational task, with milestones, deadlines and detailed WBS, perfectly fits as an application scenario of the following feature. We can also add external users, so avoiding to limit the management of the project to people enrolled in the community.

This feature, natural as it may seem, has been very complicated to implement because of the need of validation of the users' actions inside the different part of the portal. In this perspective, external users have very limited actions to perform normally inside the platform, while in the case of educational projects, they can act as an important stakeholder thus needed different permissions on the project. For example, imagine an enterprise tutor following the workgroup to support with her ideas the students.

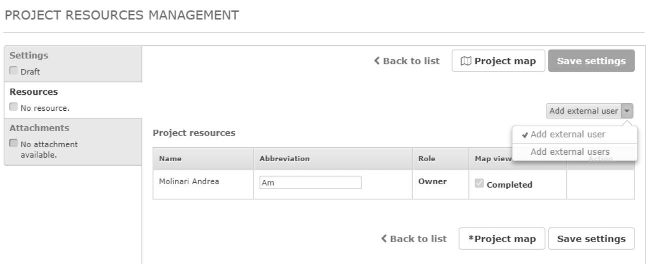


Fig.2 Managing resources and roles inside the project

In order to have a consistent approach to both the creation of complex projects or simple reminder/to-do lists, and to follow standard planning procedures in commonly available scheduling software, the Project Management service provides the user a general activity for the project at level zero, indicating the root of the project, distinguishing from the other only by the absence of a parent. From this point, the rest of the WBS is managed, together with the rest of inputs (duration and predecessors) by the CPM engine that we have implemented, this means that starting from a) the definition of a calendar b) the start date of the project c) the tasks and their dependencies, our system is able to produce as output a) the start and finish date of all the tasks b) the critical path, i.e., the tasks that have a total float equal to zero c) the finish date of the project.

What is normal in a Project Management context, it's not so "normal" in educational contexts. For example, forcing thesis students to thing about their thesis as a "project" (what else?) with tasks, deadlines, constraints, etc. has been a great improvement in both students and supervisor daily work.

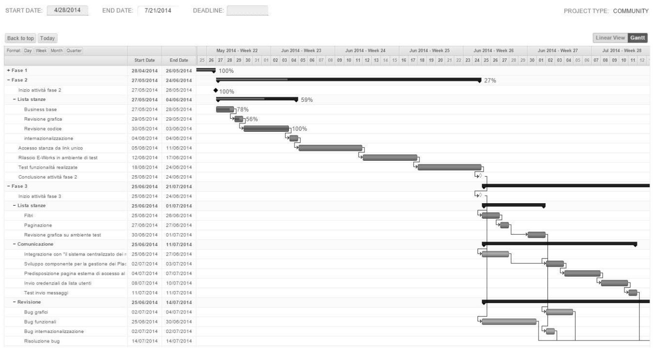


Fig.3: the WBS with critical path, predecessors and completion percentage

As further implementation, we have improved the previous version of our services, adding for each task or project the possibility of adding constraints and deadlines. A task has several other features, like a status (indicates at what stage of development is the project), priority, temporal constraints and the creation of milestones (used to indicate the achievement of the objectives set at the design stage).

On the side of security and protection, the community where the project is created/managed guarantees a sort of "sandbox" for the permissions management. A user holding the appropriate permissions is allowed to create a project with an arbitrary number of sub-tasks, to which different resources can be assigned. The roles that we have decided to support are:

- Owner of the project: role assigned when creating the project. Owner is the user that has total control over the project and has no limit in respect of assignment of roles, cancellation of tasks, attachment, etc.

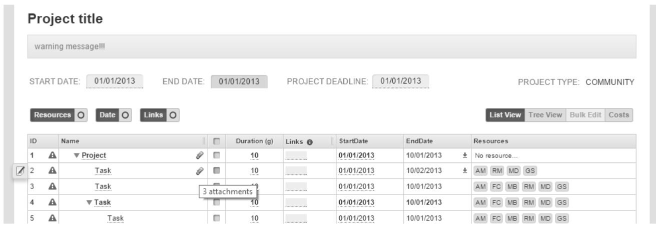


Fig.4: WBS with attachments and assignments of resources taken from the community's users

- Manager: this role will have the same potential of the project owner, with some restrictions on the tasks created by other managers. A manager may appoint other managers or simple resources. This can be done only on the task of which the manager is the owner (creator). Same goes for the cancellation and modification of activities. Note that the role of the owner of the project is distinguished from all others because of its total control over every single part of the task list regardless of the assignments. Task manager and the owner will also be asked to indicate the status of a task or project, thus introducing the control over the work of other users.

- Resources or executors of the project: these users will have a limited subset of actions since its main purpose is to perform the task and inform the manager through a report.
- Guest: this role is meant for those users who want to enable you to view a project, without, however, afford to interact with it in any way.

In the platform, you can create three different types of projects or task lists depending on confidentiality and context required. You can create personal and public projects within a community or personal projects at portal level, i.e., outside any specific community. A task list is visible only to the creator of the same and to the assigned people.

A public project, however, provides the necessary permissions for users with admin rights within the community to view and interact with all users involved, inviting external people from other communities or even not enrolled in the platform. We get more flexibility in the case of a portal task list, conceptually associated with a super-community, where all subscribers to the platform (here we are at the highest level of the communities' hierarchy) are considered within the same context. This allows us to engage in a project potentially all people registered to the platform, regardless of the inclusion in any community.

5. CONCLUSIONS

In this paper, we present a specific part of our LMS called "Online Communities", a software platform providing collaborative services among members of a (virtual) class, that in the end, we consider as a particular type of a virtual community. In this platform, we have added some Project Management tools that are following the tools and techniques provided by the Project Management discipline. These tools and techniques have been made available for any user inside the platform, in order to implement the idea of project integrated with a collaborative and educational software platform. This approach revealed two positive aspects: a) the appropriateness of Project Management concepts inside educational contexts, because many of the activities we perform during educational tasks can be seen as part of a project b) the advantage of implementing these services inside a virtual communities' environments, that provides a natural and fertile ground for the development of these services, and their availability to community members. Next steps of this evolution are the completion of project management services respect to some specific features (overallocation management, budgeting work, cost and labour resources, charting and reporting) and become a serious competitor on one side for project management tools, and on the other side for educational environment where these tools and techniques are not implemented, or implemented in very primitive stages.

REFERENCES

[1] Levitt R. (2011) Towards Project Management 2.0, Engineering Project Organization Journal, 1:3, 197-210

[2] Molinari A. (2015), Collaboration Services As A New Perspective For E-Learning Systems, Multidisciplinary Academic Conference on Education, Teaching and Learning (MAC-ETL 2015), Prague, 6-8 2015

[3] Project Management Institute. A Guide To The Project Management Body Of Knowledge (PMBOK Guides), 5th Edition. Project Management Institute, 2013

[4] Atkinson R (1999) Project Management: cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteriam International Journal of Project Management Vol. 17, No. 6, pp. 337±342, 1999 # 1999 Elsevier Science Ltd and IPMA.

[5] Meredith, J. R., & Mantel Jr, S. J. (2011). *Project Management: a managerial approach*. John Wiley & Sons.

[6] The Standish Group (2013), Chaos Manifesto 2013: Think Big, Act Small, retrieved from http:// www.versionone.com/assets/img/files/CHAOSManifesto2013.pdf

[7] Mohamed, B., & Koehler, T. (2012). The Effect of Project Based Web 2.0-Learning on Students' Outcomes. In *Towards Learning and Instruction in Web 3.0* (pp. 51-70). Springer New York.

[8] Nahla El Zant El Kadhi, Hanaa Al-Sharrah (2011) New Methodology for Developing Digital Curricula, ICCGI 2011 : The Sixth International Multi-Conference on Computing in the Global Information Technology, pp 148-153, ISBN: 978-1-61208-139-7

[9] Jackson, D. W. (2010). Tool, Collaboration vs Communication: Selecting the Appropriate. Law Library Journal , 102 (2), 315-324.

[10] Colazzo, L.; Molinari, A.; Villa, N. (2009). Collaboration vs. Participation: The Role of Virtual Communities in a Web 2.0 World, Education Technology and Computer, 2009. ICETC '09. IEEE International Conference, 321-325, 17-20 April 2009 Singapore

Generative Design for project optimization

Andrea Di Filippo
Department of
Civil Engineer
University of Salerno
Fisciano (SA), Italy
anddifilippo1@unisa.it

Marco Lombardi
Department of
Industrial Engineering
University of Salerno
Fisciano (SA), Italy
malombardi@unisa.it

Francesco Marongiu
Department of
Industrial Engineering
University of Salerno
Fisciano (SA), Italy
f.marongiu@studenti.unisa.it

Angelo Lorusso
Department of
Industrial Engineering
University of Salerno
Fisciano (SA), Italy
alorusso@unisa.it

Domenico Santaniello
Department of
Industrial Engineering
University of Salerno
Fisciano (SA), Italy
dsantaniello@unisa.it

***Abstract*—The present work concerns the applications of Generative Design in the Architectural sector: this represents an iterative process that involves an algorithm for the generation of a few outputs to satisfy certain constraints. Generative Design can offer advantages to traditional building design and the urban planning processes, given its capability to manage complexity by optimizing specifically preselected criteria while still at the drafting stage. The case study is a building currently under construction at the Valladolid Campus, and the objective is to analyse whether or not its design process has been exhaustive. The first step was to define inputs and variables depending on the ultimate goals. Successively it was time to set the boundaries for the outputs in terms of the different displayable and measurable solutions. Then the last step was to select the optimal solution. The tools employed are Formit for basic modelling, Revit for BIM modelling, Dynamo for visual programming, using graphs and nodes customizable through Python scripts. Specifically, for the Generative Design part and results displaying, Refinery used an Autodesk beta version software. Refinery implements a genetic algorithm with an NSGA-II optimization, a meta-heuristic algorithm for multi-objective optimization. Once the optimal solution has been determined, it is possible to import the results back into a Dynamo for Revit environment to complete the BIM modeling according to the Level of Geometry (LOG) and Information (LOI) required.**

Generative Design; optimization; BIM; Parametric Design; Genetic Algorithm.

I. INTRODUCTION

Nowadays, in addition to the interest in an aesthetic architecture aimed at satisfying the tastes of their clients, designers are paying more attention and emphasis to the performance of the building (structure, environment, building, socio-economic and cultural, etc.) [1], [2]. This change in design approach leads designers to adopt new technologies to support innovation, changing the traditional 2D design paradigm. Thus, civil engineers and architects have started a process of contamination towards experiences from other fields by adopting and learning technology from industrial design and mechanical engineering, for example, in which performance plays a crucial role. This process leads to adopting new computational design methods through computer science methodologies, such as generative and parametric approaches or isomorphic surfaces. In addition, many changes have been resulted from the advent and spread of low-cost and non-commercial electronic devices. Such devices are introduced to improve the quality of life through services [3],[4],[5] that aim to increase human comfort and security [6],[7],[8]. Despite some issues regarding security, such devices bring considerable benefit to the well-being of human life [9],[10],[11] by giving the possibility to build a digital twin of real environments [12].

Given these changes and the new technological perspective, this study explores the potential of a BIM environment integrated with new computational design methods in order to maximize design proposal opportunities. The work leverages generative design to identify alternatives in the early stages of design, and parametric algorithm and existing BIM tools for modification of chosen alternatives and change management during the more advanced stages of design through construction. The following sections will explain each feature in more detail.

Among the emerging systems for automating basic ideas, generative systems are helping designers to have different design solutions explored quickly, saving time and work, and evaluating the most significant number of alternatives to the chosen performance requirements. While generative tools help AEC designers immerse themselves in their designs, they otherwise fail to meet the basic principles of information modelling and desired data management. To address this problem, the answer has been Generative Design, as the development of space - generative - BIM environment that allows for creativity, fluidity, and flexibility in design, while making minimal changes to the standard design process. In fact, using such an integrated platform, information relevant to design requirements can be the system input, while design algorithms can generate the design output. Therefore, this platform integration can help designers to solve complex multi-criteria design problems. The enhancement of creativity characterizes this approach by allowing designers to use the idea seeding technique [13], supporting the design process by generating problem-free alternative designs [14], through altering various design parameters, observing (and reflecting) on the results in real-time [15].

The present work has as its object the study of a new design model, through the aid of Generative Design, with a first theoretical contextualization and a subsequent practical application in order to describe the entire design process, from concept to model.

The study, in fact, has been approached from an engineering point of view through the development of an application case referred to a new university building being designed at the Valladolid Campus to analyse the completeness of the analysis modelling processes. This synthesis presents a study conducted to explore different specifications of starting requirements and various possibilities of integrating generative design algorithms with existing BIM platforms, considering the complete process, from input setup, process description to the result and related outputs.

The main objective of this paper is to develop a conceptual method of the Generative BIM platform, developed with Autodesk's Refinery program in beta version, to maximize the efficiency of design processes and outline a methodology for supporting BIM applications: thus, from the very early basic idea stages to the final detailed design stages.

II. The Generative Design Based Approach

This section will explore the essential concepts that underlie the ability to understand how Generative Design works as applied to the Revit and Dynamo programs through Refinery.

A. Background

In the late 1970s, William J. Mitchell referred to Generative Design as a system able to generate potential solutions to a given problem. In the early 21st century, Thomas Fischer, and Christiane Herr [16] then defined it as a design approach in which "during the design process the designer does not interact with materials and products directly but through a generative system of some model." For Christiane Herr [17], a generative system refers to computer-aided generative systems that architects or engineers generally develop, thus reflecting the uniqueness of architectural design problems. John Frazer [18] described Generative Design as the use of "virtual computer space in a manner analogous to evolutionary processes in nature"; similarly, Jeffrey Krause [19], [20] defined it as the development of "systems that can enhance, evolve, or design architectural structures, objects, or spaces more or less autonomously." For Andre Chaszar and Sam Conrad Joyce [21], Generative Design overcomes the shortcomings of traditional manual design methods by "leveraging computational power and addressing issues of speed and accuracy, as well as complexity" and increased inventiveness as it increases "the number of design variations "and" the range of variations, "which include "happy accidents," i.e., unexpected outcomes that positively affect the design process. Fakhri Bukhari [22] more recently limited the scope of Generative Design to the use of "algorithms to generate a set of different solutions from a given set of design goals and constraints."
complexity" and increased inventiveness as it increases "the number of design variations "and" the range of variations, "which include "happy accidents," i.e., unexpected outcomes that positively affect the design process. Fakhri Bukhari [22] more recently limited the scope of Generative Design to the use of "algorithms to generate a set of different solutions from a given set of design goals and constraints."

B. The Generative Design Definition

A possible shared definition based on state of the art, then, can be to consider the generative approach as the "ability to produce or create something." Some authors define Generative Design as a design process that refers primarily to the evolution of techniques in creating and producing design solutions [23]. In contrast, others do not limit Generative Design to evolutionary processes, considering it a design approach based on algorithmic or rule-based processes that generate multiple and, possibly, complex solutions [24]. Considering these two perspectives, it is trivial to observe how the association of Generative Design to evolutionary processes is restricted. It excludes other methods that generate design and differentiate it from other terms, such as Parametric Design. Therefore, it can be understood as a design paradigm that uses more autonomous algorithmic descriptions than Parametric Design. After initiating the generative process, the Generative Design system executes the coded instructions iteratively to meet the chosen criteria. As a result, complex outputs can be generated even from simple algorithmic descriptions. In many cases, the algorithm is difficult to correlate with the generated product, making it difficult to predict the outcome from simply reading the algorithmic description; it is therefore supported by visual design programs to support design choices.

Generative Design can refer to any design process in which the designer uses a system, most likely through one or more software programs, to solve the design problem with some level of automation. Briefly, one can refer to Generative Design as a collaborative design process between humans and computers. During this process, the designer defines design parameters. The computer produces design studies (alternatives). These are only evaluated based on quantifiable objectives established by the designer, improving the results based on studies previously carried out and feedback from the designer himself. Finally, outputs are classified according to how closely they approximate the designer's original objectives.

Generative Design represents a design framework that combines computational computing and human creativity to achieve results that would otherwise be unseen and/or unimaginable. It involves integrating a geometric system, a set of measurable goals, and a system to automatically generate, evaluate, and evolve a diverse number of design options.

This approach can offer advantages for design, such as managing complexity and optimizing for specific criteria; this concept of complexity includes a great deal of input from experience, current demands, and specific needs.

Thus, there is the possibility of combining solutions based on real data, structuring discussion among stakeholders on design features and project goals, providing transparency on design assumptions, and offering a preliminary "live model" to build a complete design.

The framework consists of three main components:

- the generation of a broad design space of possible solutions through a custom design based on a geometric system;
- the evaluation of each solution through measurable objectives;
- the evolution of design generations through evolutionary computation.

Generative Design is, therefore, a framework that can be used on different scales and is flexible: in fact, its field of application varies in a wide range of design problems and scales, from industrial components, where it found its first applications, to buildings and cities.

C. *Used Tools*

The tools used were: Formit for basic design; Revit for modelling in a BIM environment; Dynamo for visual spreadsheet processing with nodes, graphs, and customization in Python. For the more specific part of Generative Design and visualization of the results was used Refinery, a beta version of Autodesk. The latter, through a genetic algorithm, implements NSGA-II optimization, a metaheuristic algorithm for multi-objective optimization.

D. *The Proposed Approach*

In Generative Design, the workflow involves the following consequential steps:

- *Creative*: design options are generated by the system, using algorithms and parameters specified by the designer.
- *Analysis*: the designs generated in the previous step are now measured and/or analysed according to the degree to which they meet the goals defined by the designer.
- *Ranking*: design options are sorted or ranked based on the results of the analysis.
- *Evolution*: the process of ranking design options to identify in which direction the design should be developed or evolved.
- *Exploration*: the designer compares and explores the generated designs, inspecting both the geometry and the evaluation results.
- *Integration*: a preferred design option is chosen and integrated into the larger project or design work.

Each of these phases can be further broken down into definition, execution, and results phases, where the definition phase is the designer's responsibility. In contrast, the execution phase and results are computer developed. In the definition phase, the designer must first establish the generation algorithm: this is the logic that defines how designs are generated, including constraints and rules. This definition phase is vital to all phases of the Generative Design process, as the validity of the results relies on the quality of the designer's input at this stage. Once the algorithm and its parameters have been defined, the computer begins to process, which means it starts generating different design options. The solutions generated

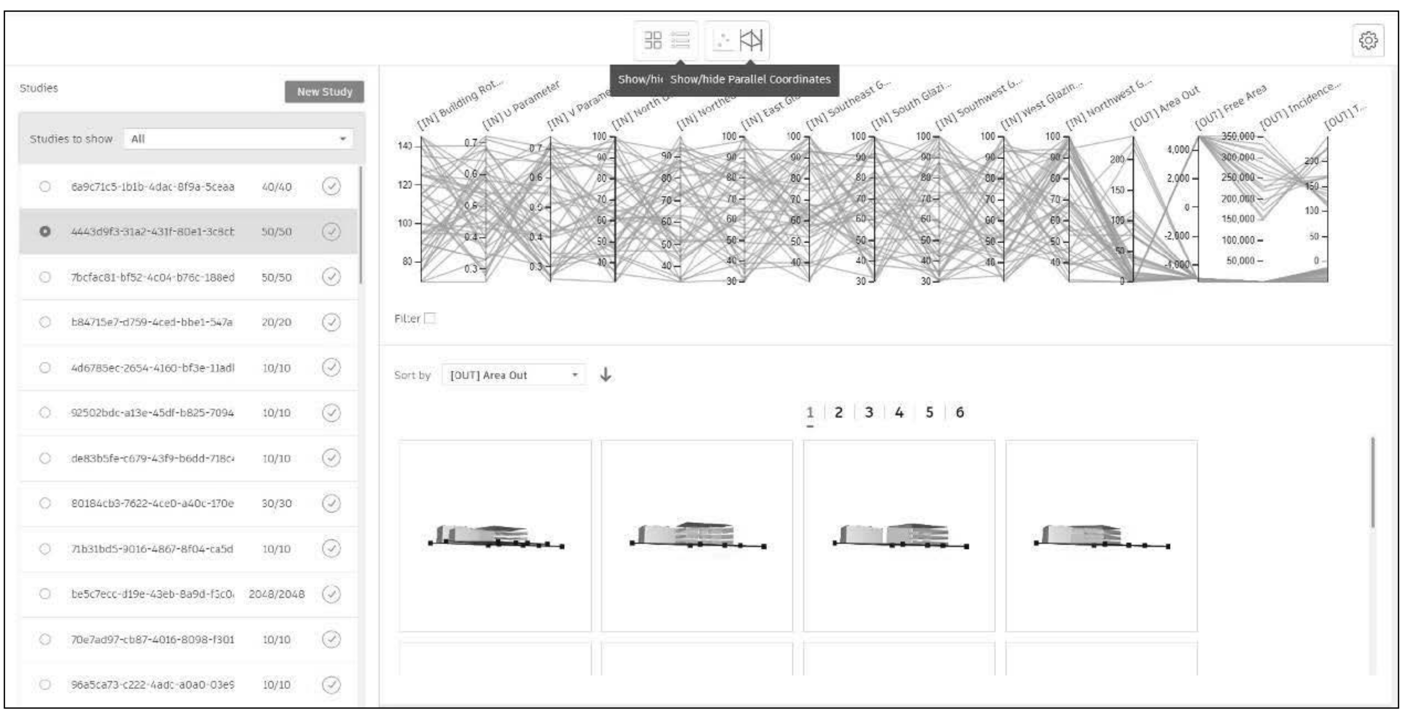


Fig. 1 System Output

during the execution phase are the final outputs of each phase; used, then, as inputs or parameters in subsequent phases. For example, the designs created in the generation phase will be used as one of the input parameters in the analysis phase. (Fig.1)

So, the designer, in the problem definition phase, will need to start with the following basic data: work to be designed, the design parameters, conditions and needs to be met, what is not desired in the final design and what aspects are desired to maximize or minimize, etc. This phase will help the designer break down the problem into less complex sub-problems that will be the subject of analysis at each stage of the generative design process.

Underlying the entire Generative Design process is a multi-objective genetic algorithm. Generative algorithms are the logical paths that create new potential solutions in this approach. In other words, they are the engine of the algorithm: they give the rest of the program something to evaluate. Evaluator algorithms receive potential solutions from the generator and evaluate how good those options are. The evaluators must be mathematically specified in the design because they will have to provide a value that the algorithm must use to discriminate between solutions. These evaluators allow the user, or the program, to query each design option and choose the best one based on the aspects they wish to include. Solver algorithms can automatically execute a script multiple times, which may contain both generators and evaluators. Solvers typically require precise inputs. Optimize is the method of running an optimization function with Generative Design. During an optimization sequence, the design will be developed based on the results of the evaluator. The optimization process works by defining multiple "generations" (or iterations) of a design. Each iteration will use the input configurations from the previous generation to optimize the new design options. [25] Other types of solvers are: Randomize, Like This, and Cross-Product.

III. CASE OF STUDY

The case analysed is the design of the new structure of the Faculty of Architecture within the Valladolid Campus.

The study consists of two phases: the first involves designing the building volume in principle, the second focuses on the positioning and consequent study of radiation in the predetermined site. The study, therefore, begins with the design from the volumetric point of view of the building in question; in principle, it was thought to separate it into three buildings connected to each other, diversifying the functions into classrooms, laboratories, and offices, considering the needs of the users of the structure. In the Dynamo environment, is first set the basis of the generative process by identifying the inputs and constraints. In this case, the inputs go to define the dimensional base of the volumes. Instead, the starting constraints delimit the inter-floor heights of each building.

Once defined the starting points of the process, it was 'build' the connections of Dynamo to obtain the desired volumes and then set the outputs necessary for our results.

Defined the part of the graph inherent in the design of the volume, we start with the second phase of problem definition, inherent in positioning the volume and the radiation of the same. For this purpose, the position of a selected mass within the site boundary will be moved and rotated. Therefore, the solar incidence expressed concerning the area can be maximized or minimized.

This latter workflow relies heavily on Dynamo's "Solar Analysis" node, making external hyperlinks to a web service to

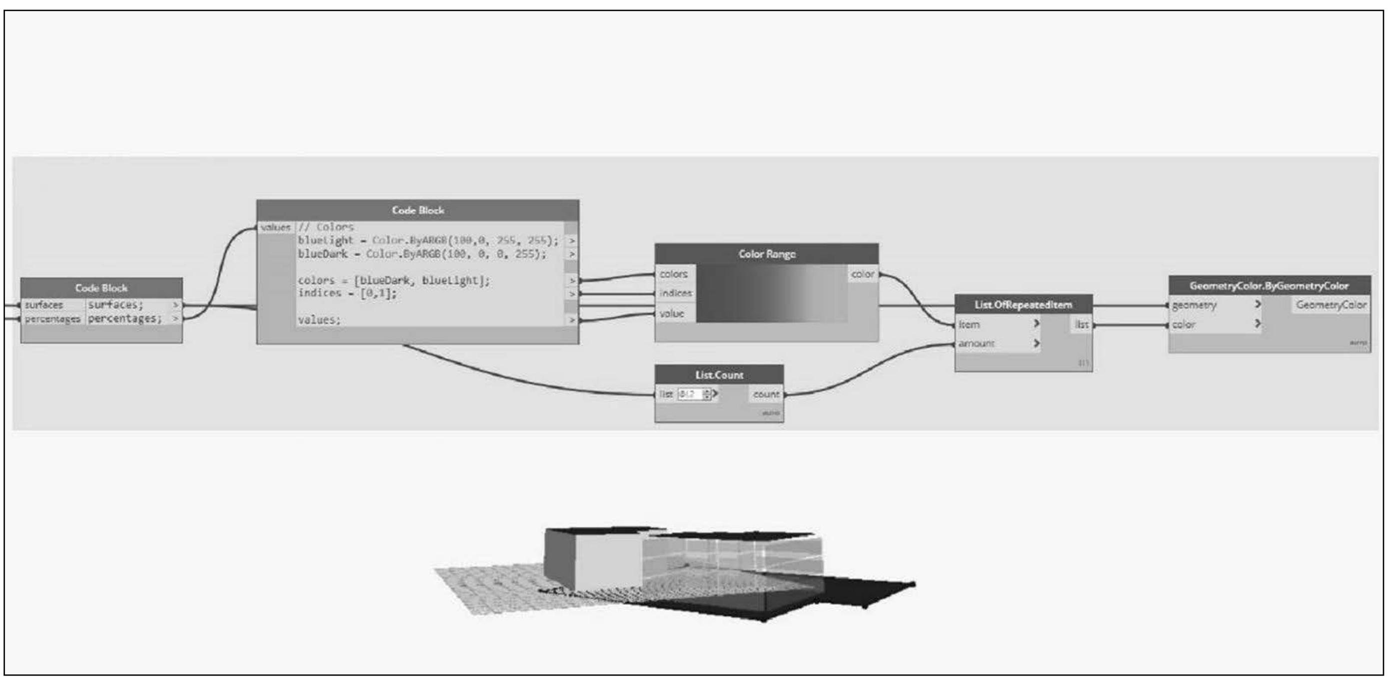


Fig. 2 Result of glazing surface.

collect the information needed for analysis. Consequently, each iteration may take some time to be performed; moreover, often, the analysis may not be performed correctly and/or with inefficient rototranslations that would lead to place the building itself beyond the boundaries of the site under study.

Trivially, in this case, the following additional inputs and constraints are defined in order to obtain new and more optimal outputs:

- Site boundary, the site boundary lines from the Revit model (model curves) should be selected;
- Surrounding buildings, select the surrounding context that will influence the solar analysis;
- Volume, the mass (building), defined by the previous graphs (which will be repositioned);
- Site offset, a value that defines the offset from the site boundary;
- Average interstorey height, used for "Solar Analysis";
- Location coordinates, these are the coordinates used for Solar Analysis;
- Building rotation (in degrees), the angle between 0° and 360° by which the building will rotate from its initial position;
- U value, the parameter of displacement of the building within site along a given axis;
- V value, parameter of displacement of the building inside the site along the axis perpendicular to U;
- Glazed surface, value in percentage of glazed surface, depending on orientation. (Fig. 3)

Once all these values are set, we continue with the generation of functions using Dynamo nodes. The script comprises a series of functions divided into groups within the graph; each group will have a name and a brief description. The name will indicate the type of function being executed, and the description will explain the related process in more detail.

The graph uses the Revit mass (or building) and extracts the geometry into Dynamo. The generator in this script provides a new position (based on the U and V values), along with a new rotation. The building is then moved to the new position point and rotated to fit the new angle. Once the building is in its new position relative to the site boundary, the solar analysis takes place, reviewing all exterior vertical surfaces of the building and calculating their solar incidence. To complete the Dynamo file with the contextual information, Formit was used, which was helpful to take advantage of the speed of mass modelling and the quick link with Bing satellite maps.

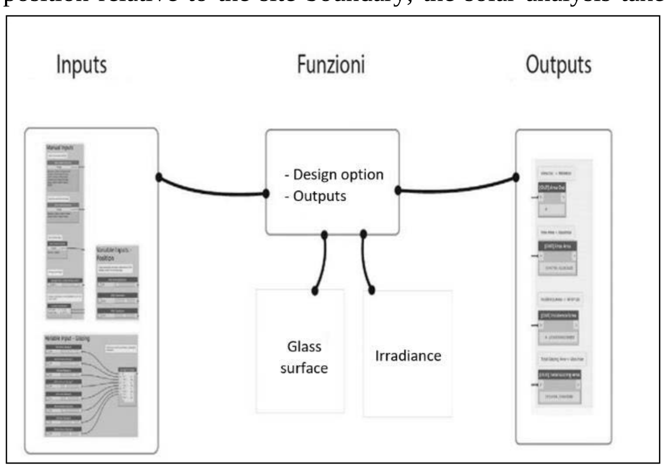


Fig. 3 Workflow.

So, after having modelled in masses the volumes of neighbouring structures, the file is saved in Formit format and exported to Revit (to take advantage of the potential of Generative Design, it is necessary to work in Dynamo, which is integrated with Revit since version 2020). Once the file has been exported with the volume in Revit, the latter recognizes the Formit masses in families in Revit automatically. At this stage, the lot used as the "Site Offset" constraint is delimited.

Having defined the geometry in Dynamo, it is often necessary to use other auxiliary geometries to facilitate the overall process. Note that all unnecessary geometry is disabled in Dynamo, to ensure that the displayed geometry shows only the final geometry output. Thus, all nodes with the preview disabled will not make the output visible in Generative Design. In this case, only the main building and the resulting solar analysis will be visible.

The solar analysis is represented on the exterior surfaces of the building as coloured grid points. These vary in colour from yellow to red, where yellow indicates low incidence and red indicates high incidence. Similarly, to display the corresponding glazed area, a scale of blue was set with the following parameters:

- Out area (m^2), area of the building that will be outside the site boundary;
- Free area (m^2), area of the inner boundary of the site that is not occupied by the base of the building;
- Average incidence (m^2), the average incidence of the exterior walls of the building. (Fig. 2)

If "area out" is greater than 0, the options are invalid, and all outputs are penalized according to whether they should be enlarged or minimized, and likewise, all null results will not be displayed and therefore not used.

So, having set the inputs and constraints, created the Dynamo graph, and chosen the outputs needed to discretize the results of the design problem, it is possible to move on to the development phase in Generative Design.

To export the work from Dynamo to Refinery, it was enough to save the project and export it to Refinery. In this phase, according to the choice and the need of the designer, the Generative Design development is set: Randomize, Optimize, Cross Product, and Like this. Once the study is complete, the results can be explored through the tables and charts in Refinery's "Explore Results" dialog box. Finally, in the Results View phase, it is possible to choose from the results obtained and then refine the optimal result with the settings chosen in Dynamo to continue the design in the BIM environment and complete the project in the defined Level of Development.

Generalizing and without automating the process of coding design options, that is, not running these scripts in Dynamo, the user would have to intervene 'manually' on the building until he finally managed to find the desired solutions. This process, besides being more time consuming, is in any case less constrained by verifications.

Since the goal in this example was to define the best position and rotation for the minimum or maximum solar incidence, it was possible to use the "Optimize" method: larger site offset values would limit the space the building can move and thus also reduce the risk of it falling outside the site boundaries.

IV. CONCLUSION

This paper proposed implementing a generative design workflow to analyse, in the first instance, the impact of radiation on a building, based on the placement of a building on the site and its surroundings. The course shows how the Generative Design process can result in good design strategies while revealing higher-level insights into the potential conflicts and trade-offs between design goals.

It was possible to highlight how at the base of Generative Design, there is always the choice of inputs and constraints that do not take responsibility away from the engineer and/or architect, but that can be gradually refined by further improving the generated strategies and leading to a more informed design. The Generative Design could also extend its potential to meet stakeholder needs for living comfort (temperature, air quality, acoustics, etc.), energy comfort, and general well-being. This can be achieved by extending the process to the evaluative component of generative design for architectural space planning and describing a set of new metrics for the automatic assessment of end-user satisfaction within spaces defined [26]. Machine learning can thus be a great tool to complement Generative Design when large datasets are available. Technically, we could try to leverage Machine Learning at any stage of the generative design process.

REFERENCES

[1] S. Abrishami, J. S. Goulding, F. P. Rahimian, and A. Ganah, “Integration of BIM and generative design to exploit AEC conceptual design innovation,” *Journal of Information Technology in Construction*, vol. 19, 2014.

[2] I. Caetano and A. Leitão, “Integration of an algorithmic BIM approach in a traditional architecture studio,” *Journal of Computational Design and Engineering*, vol. 6, no. 3, 2019, doi: 10.1016/j.jcde.2018.11.004.

[3] G. D’Aniello, M. Gaeta, F. Orciuoli, G. Sansonetti, and F. Sorgente, “Knowledge-Based Smart City Service System,” *Electronics*, vol. 9, no. 6, Jun. 2020, doi: 10.3390/electronics9060965.

[4] F. Colace, M. de Santo, M. Lombardi, F. Pascale, D. Santaniello, and A. Tucker, “A Multilevel Graph Approach for Predicting Bicycle Usage in London Area,” in *Fourth International Congress on Information and Communication Technology. Advances in Intelligent Systems and Computing, vol 1027*, vol. 1027, Y. XS., S. S., D. N., and J. A, Eds. Springer, Singapore, 2020, pp. 353–362. doi: 10.1007/978-981-32-9343-4_28.

[5] F. Colace, D. Santaniello, M. Casillo, and F. Clarizia, “BeCAMS: A behaviour context aware monitoring system,” 2017. doi: 10.1109/IWMN.2017.8078374.

[6] M. Carratu, M. Ferro, A. Pietrosanto, P. Sommella, and V. Paciello, “A Smart Wireless Sensor Network for PM10 Measurement,” in *2019 IEEE International Symposium on Measurements and Networking, M and N 2019 - Proceedings*, Jul. 2019, pp. 1–6. doi: 10.1109/IWMN.2019.8805015.

[7] R. Guida, M. C. de Simone, P. Dašić, and D. Guida, “Modeling techniques for kinematic analysis of a six-axis robotic arm,” *IOP Conference Series: Materials Science and Engineering*, vol. 568, Sep. 2019, doi: 10.1088/1757-899X/568/1/012115.

[8] F. Colace, M. Lombardi, F. Pascale, and D. Santaniello, “A Multilevel Graph Representation for Big Data Interpretation in Real Scenarios,” in *2018 3rd International Conference on System Reliability and Safety (ICSRS)*, Nov. 2018, pp. 40–47. doi: 10.1109/ICSRS.2018.8688834.

[9] C. Esposito, A. Castiglione, F. Palmieri, and A. de Santis, “Integrity for an Event Notification Within the Industrial Internet of Things by Using Group Signatures,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, Aug. 2018, doi: 10.1109/TII.2018.2791956.

[10] F. Colace, M. Lombardi, F. Pascale, and D. Santaniello, “A multi-level approach for forecasting critical events in smart cities,” 2018. doi: 10.18293/DMSVIVA2018-002.

[11] A. Castiglione *et al.*, “Securing the internet of vehicles through lightweight block ciphers,” *Pattern Recognition Letters*, vol. 135, pp. 264–270, Jul. 2020, doi: 10.1016/j.patrec.2020.04.038.

[12] F. Tao and M. Zhang, “Digital Twin Shop-Floor: A New Shop-Floor Paradigm Towards Smart Manufacturing,” *IEEE Access*, vol. 5, 2017, doi: 10.1109/ACCESS.2017.2756069.

[13] J. O. Blosiu, “Use of synectics as an idea seeding technique to enhance design creativity,” in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 1999, vol. 3. doi: 10.1109/icsmc.1999.823365.

[14] Y. S. Kim, “Personal characteristics and design-related performances in a creative engineering design course,” 2003.

[15] RIBA, *BIM Overlay to the RIBA Outline Plan of Work*. 2012.

[16] T. Fischer and C. Herr, “Teaching generative design,” *... of the 4th Conference on Generative Art*, 2001.

[17] C. M. Herr, “Generative architectural design and complexity theory,” 2002.

[18] J. Frazer, J. Frazer, L. I. U. Xiyu, T. Mingxi, and P. Janssen, “Generative and Evolutionary Techniques for Building Envelope Design,” *5th International Generative Art Conference GA2002*, vol. 2002, no. December, 2002.

[19] J. Krause, “Reflections: The Creative Process of Generative Design in Architecture,” *Generative Arts Conference*, 2003.

[20] S. Abrishami, J. Goulding, F. Pour Rahimian, and A. Ganah, “Virtual generative BIM workspace for maximising AEC conceptual design innovation: A paradigm of future opportunities,” *Construction Innovation*, vol. 15, no. 1, 2015, doi: 10.1108/CI-07-2014-0036.

[21] A. Chaszar and S. C. Joyce, “Generating freedom: Questions of flexibility in digital design and architectural computation,” *International Journal of Architectural Computing*, vol. 14, no. 2, 2016, doi: 10.1177/1478077116638945.

[22] F. a Bukhari, “A Hierarchical Evolutionary Algorithmic Design (HEAD) System for Generating and Evolving Building Design Models,” 2011.

[23] P. Zhang and W. Xu, “Quasicrystal structure inspired spatial tessellation in generative design,” in *CAADRIA 2018 - 23rd International Conference on Computer-Aided Architectural Design Research in Asia: Learning, Prototyping and Adapting*, 2018, vol. 1.

[24] “Autodesk Refinery program.” Accessed: May 31, 2021. [Online]. Available: www.autodesk.com/solutions/generative-design/architecture-engineering-construction

[25] “Generative Design and Dynamo Primer.” Accessed: May 31, 2021. [Online]. Available: https://www.generativedesign.org/

[26] L. Villaggi, J. Stoddart, D. Nagy, and D. Benjamin, “Survey-Based Simulation of User Satisfaction for Generative Design in Architecture,” in *Humanizing Digital Reality*, 2018. doi: 10.1007/978-981-10-6611-5_36.

Authors

Keywords