DMSVIVA 2020

Proceedings of the 26th International DMS Conference on Visualization and Visual Languages

July 7 to 8, 2020
KSIR Virtual Conference Center
Pittsburgh, USA

DOI: 10.18293/DMSVIVA2020

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

PROCEEDINGS

DMSVIVA2020

The 26th International DMS Conference on Visualization and Visual Languages

Sponsored by

KSI Research Inc. and Knowledge Systems Institute, USA

Technical Program

July 7 to 8, 2020

KSI Research Virtual Conference Center, Pittsburgh, USA

Organized by

KSI Research Inc. and Knowledge Systems Institute, USA

ISBN: 1-891706-51-9
ISSN: 2326-3261 (print)
2326-3318 (online)
DOI: 10.18293/DMSVIVA2020

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 *26th International DMS Conference on Visualization and Visual Languages (DMSVIVA2020)*, 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 2020, the conference received 21 submissions. All the papers were rigorously reviewed by three members of the international Program Committee. Based on the review results, 9 papers have been accepted as regular papers with an acceptance rate of 43%. 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.

Qi Li, Shanghai University of Engineering Science, China; Program Co-Chair
Loredana Caruccio, Universiy of Salerno, Italy; Program Co-Chair

DMSVIVA2020

The 26th International DMS Conference on Visualization and Visual Languages

July 7 and 8, 2020

KSIR Virtual Conference Center, Pittsburgh, USA

Conference Organization

DMSVIVA2020 Conference Chair and Co-Chairs

Joseph J. Pfeiffer, Jr., New Mexico State University, USA; Conference Chair
Vincenzo Deufemia, University of Salerno, Italy; Conference Co-Chair
Andrew Fish, University of Brighton, UK; 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

Qi Li, Shanghai University of Engineering Science, China; Program Co-Chair
Loredana Caruccio, Universiy of Salerno, Italy; Program Co-Chair

DMSVIVA2020 Program Committee
Subcommittee on Visualization and Visual Languages

Danilo Avola, University of Rome, Italy
Rachel Blagojevic, Massey University, New Zealand
Paolo Bottoni, Universita Sapienza, Italy
Paolo Buono, University of Bari, Italy
Kendra Cooper, University of Texas at Dallas, USA
Gennaro Costagliola, University of Salerno, Italy
Vincenzo Deufemia, University of Salerno, Italy

Filomena Ferrucci, University of Salerno, Italy
Manuel J. Fonseca, University of Lisbon, Portugal
Jennifer Leopold, Missouri University of Science & Technology, USA
Weibin Liu, Beijing Jiao Tung University, China
Eloe Nathan, Northwest Missouri State University, USA
Kazuhiro Ogata, JAIST, Japan
Peter Rodgers, University of Kent, UK

Subcommittee on Sentient and Distributed Multimedia Systems

Andrew Blake, University of Brighton, UK
Loredana Caruccio, University of Salerno, Italy
William Cheng-Chung Chu, Tunghai University, Taiwan
F. Colace, University of Salerno, Italy
Gennaro Costagliola, Univ of Salerno, Italy
Andrea De Lucia, Univ. of Salerno, Italy
Martin Erwig, Oregon State University, USA
Larbi Esmahi, National Research Council of Canada, Canada
Daniela Fogli, Universita degli Studi di Brescia, Italy
Kaori Fujinami, Tokyo University of Agriculture and Technology, Japan
Ombretta Gaggi, Univ. of Padova, Italy
Angela Guercio, Kent State University, USA
Alan Liu, National Chung Cheng Univeristy, Taiwan
Max North, Southern Polytechnic State University, USA
Antonio Piccinno, Univ. of Bari, Italy
Giuseppe Polese, University of Salerno, Italy
Weiwei Xing, Beijing Jiao Tung University, China
Atsuo Yoshitaka, JAIST, Japan
Ing Tomas Zeman, Czech Technical University, Czech Republic

Subcommittee on Distance Education Technologies

Maiga Chang, Athabasca University, Canada
Yuan-Sun Chu, National Chung Cheng University, Taiwan
Mauro Coccoli, University of Genova, Italy
Rita Francese, University of Salerno, Italy
Angelo Gargantini, University of Bergamo, Italy
Pedro Isaias, University of Queensland, Australia
Paolo Maresca, University Federico II, Napoli, Italy
Andrea Molinari, University of Trento, Trento, Italy
Ignazio Passero, University of Salerno, Italy
Elvinia Riccobene, University of Milano, Italy
Michele Risi, University of Salerno, Italy
Michael Wybrow, Monash University, Australia

Publicity Chair

Jun Kong, North Dakota State University, USA; Publicity Chair

Table of Contents

Session VIVA1: Visualization and Visual Languages

Session VIVA2: Visualization and Visual Languages

Note: (S) denotes a short paper and (P) denotes a poster paper.

An Automated Visual Recognition System to Counteract Illegal Dumping in Smart Cities

Mauro Coccoli
DIBRIS
University of Genoa,
Genoa, Italy

Vincenzo De Francesco
DIETI
Federico II Naples University,
Naples, Italy

Antonio Fusco
Asia Napoli S.p.A.
Naples, Italy

Paolo Maresca
DIETI
Federico II Naples University,
Naples, Italy

Abstract—**In this paper we will describe the prototype form of an automated visual recognition system designed to mitigate illegal dumping, as the outcome of an experiential learning activity. The presented solution relies on the sensor networks of a smart city where the waste management system is supposed to be integrated with other municipality services for environment control and management. In particular, we want to take advantage of the pictures, frames and videos continuously recorded by cameras installed in the cities for traffic monitoring, for surveillance or any other reason. Such data are processed by means of cognitive computing techniques and a specific algorithm of image analysis has been trained to identify trash, especially bulky waste, where it should not be, and trigger an alarm to the municipality. Besides, an organization plan is also proposed for intelligent waste collection as well as some organizational ideas for scalability. The learning activity has been conducted within the program "Party Cloud Challenge per Genova" promoted by IBM in collaboration with the city municipality of Genoa, Italy.**

Keywords-smart city; cognitive computing; image recognition; waste management; environment

I. Introduction

Commonly, the idea of smart city is tied to using novel ICT solutions to develop mobile intelligent applications to provide effective services seamlessly integrated with the main city infrastructures and sensor networks for environment control and management. However, we must remember that this general view should include technological aspects as well as people and institutions. In fact, the adoption of sophisticated systems and advanced solutions are just enablers for the institutions to set active policies driven to enhance citizens' quality of life. Moving along these three main dimensions, the objectives of a smart city can be summarized as the following: *(i)* integration of infrastructures and technology-mediated services, *(ii)* social learning for strengthening human infrastructure, and *(iii)* governance for institutional improvement and citizen engagement [1]. In this respect, most projects for making a city smart, result in the definition of new models to mitigate current urban problems, to the aim of improving services to the citizens and, thus, making cities attractive places to live in. From a practical point of view, most smart cities implement a common reference model that addresses global sustainability challenges at a local level [2]. Regardless what are the implementation details, the available sensors, the installed instrumentations, the adopted interconnection techniques, the integration methodologies, the ways for exchanging data, possibly in open data format, and the number of possible applications, the main topics addressed in designing smart cities services are e-democracy, e-health, sustainable urban mobility, water, energy, environment, pollution. In this flavor, triggered by the initiative "IBM Party Cloud Challenge *per Genova*" –Italian for '*for Genoa*'– this paper focuses on the use of cognitive computing techniques to develop an automated alerting system based on visual recognition, to support the municipality of the city of Genoa, Italy, in counteracting illegal dumping of bulky waste in the streets. The developed system was used as a demonstrator in a Bachelor's Degree thesis in computer engineering, centered on the use of cognitive computing technologies and design methodologies, to show how they can accelerate and improve the learning processes in universities [3, 4].

Applying the smart city paradigm to the field of waste management is an important topic for the environment control and management in urban settings, and poses challenging issues. To find effective solutions, we explore the possibility of using novel cognitive computing technologies. Generally speaking, waste management refers to finding solutions and advanced techniques to manage the garbage lifecycle and disposal processes such as, e.g., using sensorized bins in an Internet of Things (IoT) framework [5] to optimize pickup operations and transports to decentralized dumps. Nevertheless, a smart waste management should embrace an integrated planning strategy tailored for resource recovery and efficiency within a circular economy framework [6]. The easiest and more economical solution is calling people to the action of reducing the volume of the rubbish they produce, also encouraging a careful separate collection of waste. Anyway, due to the possible presence of non-polite citizens, we also have to consider policies aimed to the reduction of illegal dumping of bulky waste. An alternative and sophisticated way to achieve this objective is using the large amount of data already available within a smart city environment owing to the installed sensor networks. Hence, we regard the smart city as a generator of images, coming from different sources, such as, e.g., the videos collected around by many cameras installed by the municipality and security agencies as well, for purposes other than dumping surveillance. Recalling the three-dimensional model cite above, our system implements

DOI reference number: 10.18293/DMSVIVA20-006

the following: *(i)* video surveillance system are integrated within the whole smart city framework and data circulate between applications; *(ii)* given the presence of an automated alerting system, citizens are encouraged to not-dumping illegally; *(iii)* the city government can take specific actions with dedicated prevention policies. Specifically, we developed an algorithm to generate alerts when a bulky waste is abandoned outside of authorized disposal points. In the following, we present the project submitted to the *Challenge per Genova*, which was awarded for the first prize (see requirements and evaluation criteria in Section 2).

The remainder of the paper is structured as follows: Section 2 describes the general framework of the Challenge *per Genova* initiative, while in Section 3 the proposed smart alerting system is presented. Following the implementation details, Section 4 discusses possible impact of the proposed solution in a smart-city framework. Finally, Conclusions and a glance on future works are reported in Section 5.

II. The "Challenge per Genova" Initiative

As anticipated, this paper reports the experience carried on during the so called "IBM Party Cloud Challenge per Genova". More precisely, the "IBM Party Cloud" is an online hackathon promoted by IBM to accelerate the technological innovation made in Italy at the service of sustainability. The "Challenge *per Genova*" track was launched on September 20, 2018 in Milan, lasting November 11, 2018, and it was open to developers, data scientists and IT people in general. This specific initiative was supported by IBM Italia, Dock Joined in Tech, Lifegate and Codemotion, under the patronage of the Genoa Municipality and its local inhouse garbage company. In fact, the objective of the hackathon was that of finding novel solutions to counteract illegal dumping more effectively, especially with reference to the disposal of bulky waste in unauthorized places in the city streets and in urban areas in general. Specifically, the statement of the problem posed by the challenge is that, presently, most critical situations are spontaneously reported by citizens, even if more than 300 surveillance cameras are installed all over the town. The video surveillance system requires a number of operators to watch videos to find possible abuses and this is not an efficient way of using human resources. Hence, the objective is developing a smart alerting solution able to find in real-time abuses and promptly trigger alarms to the operation center, which can manage timely counteractions to solve issues arisen. In more detail, from the technical point of view, the participants to the contest are called to develop an application prototype using the IBM Cloud platform. In such an integrated environment, designers and developers are provided with advanced image recognition tools and analytics services, which can be used to identify the abuses described above and, consequently, trigger alarms to signal them via SMS and/or other means. Moreover, any alert is required to include the following: *(i)* picture of the evidence, *(ii)* geo-localization coordinates, *(iii)* type of bulky waste and its spatial features to evaluate its possible dangerousness, *(iv)* possible information on "who" made the illegal dumping, e.g., a person as well as a vehicle and its plate number, *(v)* abnormalities, if any.

After the deadline for submitting the projects, the developed applications were evaluated based on the following criteria: *(i)* usefulness and value, *(ii)* bearing on the challenge issue, *(iii)* technical soundness and accuracy, including architectural blueprints, *(iv)* interface design and user experience, *(v)* creativity and novelty.

III. The Proposed Smart Alerting Sysytem

Given the strict system requirements specified within the framework of the challenge, the structure of the automated alerting system can be described with two functional blocks: *(a)* identify and classify a possible item of interest in a specific area under video surveillance, and *(b)* managing the alerting system. To achieve these results, we were provided with a selection of videos recorded by the local surveillance system of the Genoa Municipality, showing non-fictional scenes of illegal dumping actions of some bulky waste. By using consolidated techniques for comparing frames, it is easy to find suspect changes in the scene (see, e.g., [7] and references therein), i.e., the stable presence of a new object in the landscape. Such new and unknown object could be a car in a long-lasting stop as well as an object candidate to be considered a bulky waste, possibly abandoned by the suspect vehicle. When such conditions occur, an operator is called to perform a visual analysis of the video, which may lead to the identification of a possible abuse. We want to make this process unattended, implementing an automated visual recognition system able to identify the presence of bulky waste in a given scene and, consequently, trigger alerts.

Owing to the advanced cognitive computing capabilities offered by IBM Watson within the IBM Cloud platform, all the necessary tools for the development of a software prototype were available. Among these, we used: data analysis and visualization, data preparation and modeling, machine learning capabilities and visual recognition. The result is a system able to identify and classify in real time the bulky waste in unauthorized places and to manage alerts triggering. Specifically, for the development of the application prototype, we adopted Watson studio, which provides the environment and tools to accelerate infusion of artificial intelligence (AI) in applications. In more detail, we used the IBM Watson Visual Recognition service, which has suited algorithms to analyze images and to build a personalized model so to extract meaning from visual content. Figure 1 depicts the relevant reference architecture.

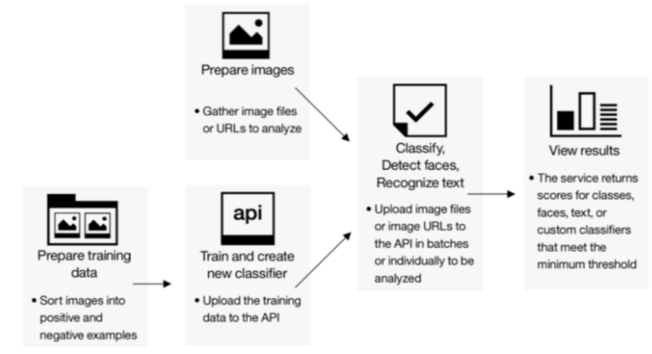

Prepare images
• Gather image files or URLs to analyze
Prepare training data
• Sort images into positive and negative examples
api
Train and create new classifier
• Upload the training data to the API
Classify, Detect faces, Recognize text
• Upload image files or image URLs to the API in batches or individually to be analyzed
View results
• The service returns scores for classes, faces, text, or custom classifiers that meet the minimum threshold

Figure 1. The architecture of Visual Recognition services within the IBM Watson Studio platform.

A. *Visual Recognition and Classification*

As a starting point, we designed our solution based on sampling videos and extracting frames, to the aim of creating a customized visual recognition model, able to identify and classify bulky waste in a given area. According to this model (Fig. 1), the supervised learning process for the recognition model is crucial and, depending on its accuracy, the application may fail or succeed. Hence, for the custom model to work effectively, an accurate training phase is needed for optimization and to guarantee good performances for the classifier, in terms of precision. Such a process is performed as described in Figure 2, showing the steps needed for creating and training a specialized visual recognition classifier. More precisely, the following *Steps 1-3* are repeated recursively:

Step 1 - prepare training images. Gather files to use as positive and negative data example;

Step 2 - create and train the classifier. Specify the location of the training images and call the Visual Recognition API to create the custom classifier;

Step 3 - test the custom classifier. Classify images with the new custom classifier and measure the classifier performance.

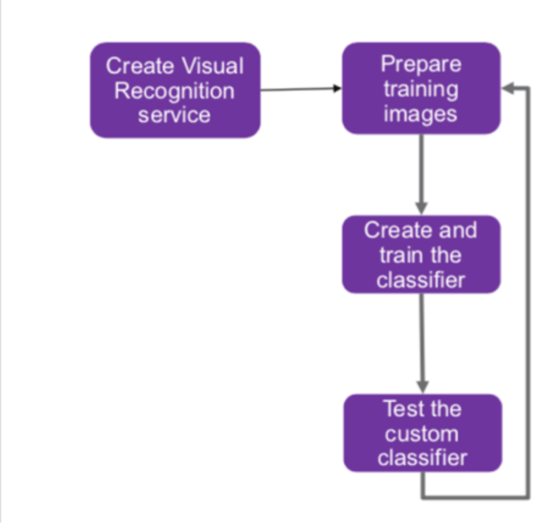
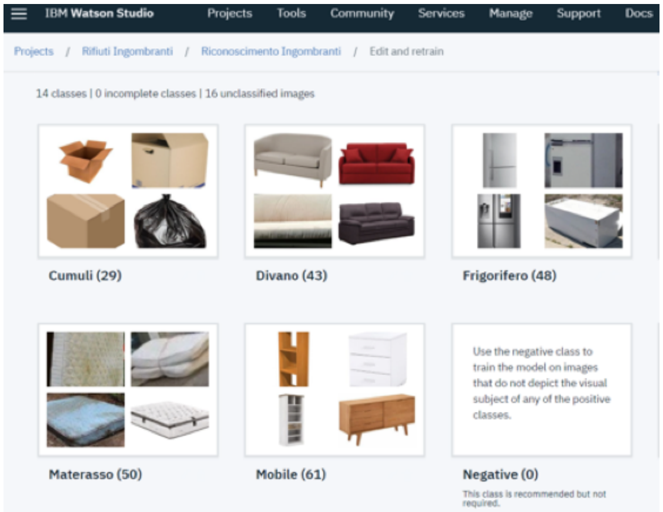

Create Visual Recognition service
Prepare training images
Create and train the classifier
Test the custom classifier

Figure 2. The recursive steps for training the visual recognition classifier.

To effectively perform this task, we have identified the types of waste that are dumped illegally more often, also matching all the ones included in the sample videos. This resulted in the definition of thirteen classes, i.e.: Refrigerator, Washing machine, Couch, Furniture, Freezer, Bicycle, Armchair, Mattress, Chair, Printer, Television, Bed base, and Heap. Then another class has to be added, that is the Negative one, for images that do not depict the visual subject of any of the positive classes, thus containing every non-classifiable item. For each of the above classes, many similar but heterogeneous pictures have been used to train the system, chosen e.g., with different values of brightness and contrast, taken from different angles and in varied light conditions.


IBM Watson Studio
Projects
Tools
Community
Services
Manage
Support
Docs
Projects / Rifiuti Ingombranti / Riconoscimento Ingombranti / Edit and retrain

14 classes | 0 incomplete classes | 16 unclassified images

Cumuli (29)

Divano (43)

Frigorifero (48)

Materasso (50)

Mobile (61)

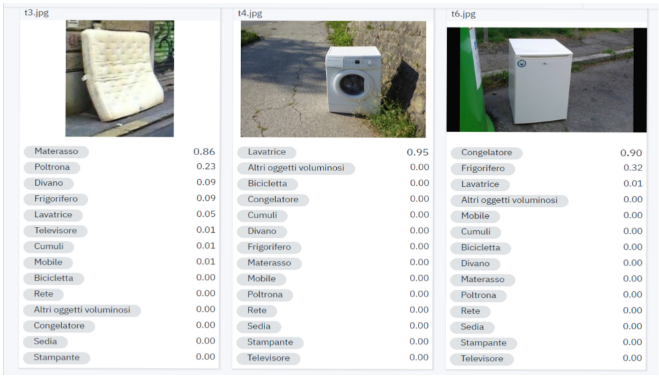
Use the negative class to train the model on images that do not depict the visual subject of any of the positive classes.

Negative (0)

This class is recommended but not required.

Figure 3. Results of the tests performed to evaluate the visual recognition system after its training phase.

Figure 3 shows a screenshot taken from the Watson Studio Visual Recognition working environment, presenting the results achieved for some of the considered classes, after a preliminary training phase. Please note that the system was programmed in Italian so, for a better understanding, consider the following Italian–English *legenda* when reading the picture: *Cumuli* – Heap; *Divano* – Couch, *Frigorifero* – Refrigerator, *Materasso* – Mattress, *Mobile* – Furniture. The numbers in parentheses specify the number of sample pictures associated to each class. Then, the real effectiveness of the training phase was tested against both the frames extracted from the original sample videos and generic pictures retrieved through common image search engines available over the Web such as, e.g., Google Images, always using the *abandoned* prefix in the search keys. For each of these pictures, the classifier assigned a matching score to the classes. Please note that such score is a number between 0 and 1, which estimates the model's confidence in the classification, based on the used training set, not an absolute percentage of accuracy.



t3.jpg

Materasso	0.86
Poltrona	0.23
Divano	0.09
Frigorifero	0.09
Lavatrice	0.05
Televisore	0.01
Cumuli	0.01
Mobile	0.01
Bicicletta	0.00
Rete	0.00
Altri oggetti voluminosi	0.00
Congelatore	0.00
Sedia	0.00
Stampante	0.00

t4.jpg

Lavatrice	0.95
Altri oggetti voluminosi	0.00
Bicicletta	0.00
Congelatore	0.00
Cumuli	0.00
Divano	0.00
Frigorifero	0.00
Materasso	0.00
Mobile	0.00
Poltrona	0.00
Rete	0.00
Sedia	0.00
Stampante	0.00
Televisore	0.00

t6.jpg

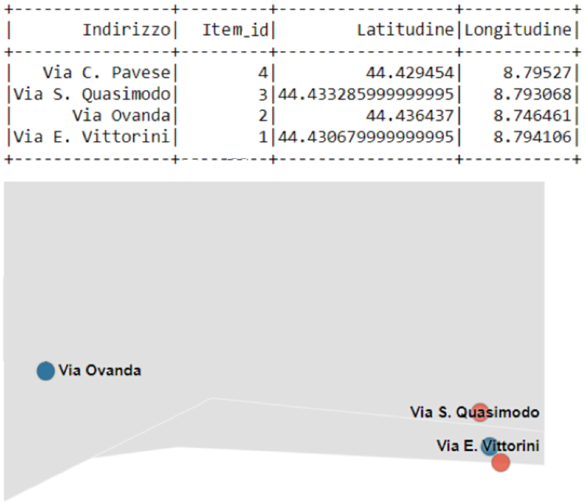
Congelatore	0.90
Frigorifero	0.32
Lavatrice	0.01
Altri oggetti voluminosi	0.00
Mobile	0.00
Cumuli	0.00
Bicicletta	0.00
Divano	0.00
Materasso	0.00
Poltrona	0.00
Rete	0.00
Sedia	0.00
Stampante	0.00
Televisore	0.00

Figure 4. Results of the test on the trained system.

According to the results shown in Figure 4, we conclude that the custom classifier should be considered reliable. For a better understanding, again, an Italian – English *legenda* is required, that is the following: *Materasso* – Mattress, *Lavatrice* – Washing machine, *Congelatore* – Freezer.

B. *The Alerting System*

When in working conditions, the visual recognition system described in the previous sub-section will be fed with frames extracted at a given sampling time from video surveillance cameras. When a suspect item is identified in at least two consecutive frames of the same scene, an alert is triggered, reporting the *class* for the suspect item and the relevant confidence *score* to evaluate the accuracy of the alert. Then, the alert message is completed with metadata attached to the video source, such as a unique *video identifier*, the timestamp through *date* and *time*, the geographical localization through *coordinates* and street *address*.



```
+----------------+---------+------------------+-----------+
|       Indirizzo|  Item_id|        Latitudine|Longitudine|
+----------------+---------+------------------+-----------+
|   Via C. Pavese|        4|         44.429454|    8.79527|
|Via S. Quasimodo|        3|44.433285999999995|   8.793068|
|      Via Ovanda|        2|         44.436437|   8.746461|
|Via E. Vittorini|        1|44.430679999999995|   8.794106|
+----------------+---------+------------------+-----------+
```

Figure 5. Graphical representation of alerts on a map of the city of Genoa.

More precisely, the alerts are stored in a database table whose fields are the following (names are reported in English and in the corresponding Italian, when used):

`Alert_id`

`Item_id`

`Class` – *`Classe`*

`Score`

`Video_id`

`Date` – *`Data`*

`Time` – *`Ora`*

`Latitude` – *`Latitudine`*

`Longitude` – *`Longitudine`*

`Street_address` – *`Indirizzo`*

When an alert is received the municipality can launch different procedures in reaction, such as, e.g., reporting to another competent authority (e.g., when a license plate is identified in the images and a vehicle has to be investigated) or collecting the waste.

In addition, the data collected within the analysis of the video frames also allow analyzing different aspects. For example, from an operational point of view, we can identify the position and dangerousness of the abandoned object. Other applications are possible, e.g., the analysis of social impact, the subsequent implementation of precise investigation criteria and, finally, in terms of implementing interventions addressed to prevent citizens to adopt such an undesired behavior. To provide an example, the data are processed with a Jupyter Notebook, one of the data science tools included in the Watson Studio, enabling interactive data analysis and visualization. A specific Python Notebook has been programmed to represent the distribution of alerts on a simplified map of Genoa, as shown in Figure 5.

IV. Possible Impact

According to the three dimensional smart-city model already mentioned in the introductory section [1], in the following we report some important considerations relevant to:

(a) prospective technology enhancements,

(b) social aspects,

(c) governance.

A. *Prospective Technology Enhancements*

What happens after the system triggers alerts with the relevant information is out of the scope of the developed prototype, which should be duly integrated at the business intelligence level of the existing legacy system through suited interfaces and specific middleware. Moreover, another block in cascade should be able to exploit profitably the information transferred by the classifier, possibly with semantic capabilities. For example, in the case the item identified in a bulky waste dumping event is a TV, it should be associated to a specific hazard class, i.e., the RAEE in Italy, equivalent to Electric and Electronic Equipment Waste, which is not considered in our model, due to limitations imposed by the strict timing and the original requirements of the challenge. This should classify the item on the basis of the subsequent pickup procedure and recovery. Furthermore, other data can drive the decisions. For example, the alert date and its persistency allow going back to the number of days of distress to the community due to the presence of the abandoned object. In this case, the collection system should keep into account a priority order in the organization of the service and logistics by choosing, e.g., the suited vehicle for both the characteristics of the item to be uploaded and the width of the street, that, in Genoa and elsewhere in Italy, can be very narrow. Furthermore, using the geographical coordinates can help calculating optimized routes of the collection paths on the road graph, to plan more efficiently the collection sequence and the overall execution times of the extraordinary withdrawal operations. Future developments should also consider security and privacy issues.

B. *Social Aspects*

From the point of view of the social impact, the spatial distribution and the identified commodity types allow to construct a dynamic map of the distribution and incidence of the abandonment phenomena, with special attention, e.g., to suburbs. Assuming a principle of proximity at the base of the phenomenon, timely communication and awareness campaigns for users can be started in order to discourage the recurrence of such phenomena. Furthermore, it is possible to or predict periods of the year in which such illegal dumping phenomena occur more often. This allows planning more frequent controls or other specific actions e.g., with more frequent collection services or making available new cumbersome stuff transfer facilities under a controlled regime, which may be both fixed or temporary, i.e., itinerant vehicles.

C. *Governance*

Both the above-described aspects are of paramount importance to reduce the impact and the incidence of the phenomena of bulky waste illegal dumping in cities. In fact, the economic and social advantages that may derive from the resolution of this phenomenon are far superior to the costs of activation of specific prevention programs and of a re-organization of existing services. The prototype presented in this paper, contributes first of all to enable a wider understanding of the phenomenon and its causes, on an analytical basis, which may be due to the lack of collaboration of the citizenry and/or the inadequacy of the relevant policies and planned services. In both cases, the availability of a database of reports on a geographical basis, with a precise classification of the abandoned waste, makes it possible to design and test targeted interventions to optimize the resources devoted to solve the problem, very felt by the community in terms of overall perception of the quality of the service provided by the civic operators.

V. Conclusions and Future Work

Following the *Challenge per Genova* initiative, we faced a specific issue of the waste management problem, which requires more complex solutions and the integration of different interoperable systems. Specifically, we focused our effort on developing and testing of an automated visual recognition system based on cognitive computing, and on collecting data and metadata to transmit with alerts, according to the given requirements. As we try expanding the boundaries of the problem analysis, several possibilities arise to extend the basic functionalities of the developed prototype. Depending on the observed filed of application, improvements can be done in different directions, to achieve different objectives. In the following we report three examples of possible enhancements:

(i) expanding the set of data collected, managed and processed, through a suited optimization of the presented visual recognition system. In fact, fully exploiting the cognitive computing capabilities available, we can estimate the overall geometrical characteristics of the detected items both in terms of dimensions (linear and volumetric) and in shape. Moreover, the simultaneous recognition of multiple objects in a single frame or at least the identification of the prevailing commodity fraction would be a desirable feature. With such information available, alerts would include the GPS coordinates of the interested place as well as the expected number of objects to be collected there, including their type, size, and possible dangerousness;

(ii) optimizing the collection path design, based on the information provided by the alerts for any point an abandoned item is detected in. A possible solution is performing a classification for the collection points, according to the gravity of the situation and the numerosity of waste and their dimensions and expected weight; as well as for the time spent from the alert. This can result in a priority-based list of sites to be visited and give hints on the best suited vehicle for safe collection and transport operations;

(iii) creating an interactive reporting service so that users, i.e., citizens, can signal abuses, taking pictures and uploading them with data such as, e.g., a description of the object, precise position with address, other useful information, including personal ones. The same interface can also be exploited to prevent illegal dumping by implementing a pick-up service on-demand, in the case a cumbersome object has to be trashed by any citizen.

As a final remark, we observe that this work is also relevant from the educational point of view, for it copes with emerging needs of both the modern society and the labor market, driving the evolution of the current higher education model. In fact, this activity can be regarded as an example of experiential learning, where knowledge is created through the transformation of experience [8]. Such a learning methodology requires students to actively acquire knowledge by doing and applying their own problem-solving abilities [9]. As students are actively involved in the learning process, their learning satisfaction increases as well as their understanding and retention of course material, and they voluntarily become continuous learners, also improving interpersonal communication skills, as well as developing analytical thinking, and critical thinking abilities [10]. In particular, for the specific case of software engineering classes, examples of laboratory experiences involving third parties can be found, e.g., in [11] and references therein.

Acknowledgment

The research has been conducted in the metropolitan area of the city of Genoa, Italy, within the program "Party Cloud Challenge *per Genova"* promoted by IBM Italia, Dock Joined in Tech, Lifegate and Codemotion, under the patronage of the Genoa Municipality and AMIU, the local inhouse garbage company.

References

[1] T. Nam and T.A. Pardo, "Conceptualizing smart city with dimensions of technology, people, and institutions," in Proceedings of the 12th Annual International Conference on Digital Government Research: Digital Government Innovation in Challenging Times, College Park, Maryland, USA, June 12-15, 2011, pp. 282-291.

[2] S. Zygiaris, "Smart city reference model: assisting planners to conceptualize the building of smart city innovation ecosystems," in Journal of the Knowledge Economy, no. 4, 2013, pp. 217-231.

[3] M. Coccoli, P. Maresca, and L.Stanganelli, “Cognitive computing in education,” Journal of e-Learning and Knowledge Society, vol. 12, no. 2, 2016, pp. 55-69.

[4] M. Coccoli, P. Maresca, and L.Stanganelli, “The role of big data and cognitive computing in the learning process,” Journal of Visual Languages and Computing, vol. 38, 2017, pp. 97-103.

[5] T. Anagnostopoulos, A. Zaslavsky, K. Kolomvatsos, A. Medvedev, P. Amirian, J. Morley and S. Hadjieftymiades, “Challenges and opportunities of waste management in IoT-enabled smart cities: A survey,” IEEE Transactions on Sustainable Computing, vol. 2, no. 3, 2017, pp. 275-289.

[6] A. Del Borghi, M. Gallo, C. Strazza, F. Magrassi and M. Castagna, “Waste management in smart cities: The application of circular economy in Genoa (Italy),” Impresa Progetto Electronic Journal of Management, vol. 4, 2014, pp. 1-13.

[7] H. F. Ng and C. Y. Chin, “Effective scene change detection in complex environments,” International Journal of Computational Vision and Robotics, vol.9, no. 3, 2019, pp. 310-328.

[8] D.A. Kolb, Experiential learning: Experience as the source of learning and development. Prentice-Hall, Inc. Englewood Cliffs, NJ, 1984.

[9] M. McCarthy, “Experiential learning theory: From theory to practice,” Journal of Business & Economics Research (JBER), vol. 14, no. 3, 2016, pp. 91-100.

[10] D. R. Brickner and E. R. Etter, “Strategies for promoting active learning in a principles of accounting course,” Academy of Education Leadership Journal, vol. 12, no. 2, 2008, pp. 87-93.

[11] M. Coccoli, P. Maresca, L. Stanganelli and A. Guercio, “An experience of collaboration using a PaaS for the smarter university model,” Journal of Visual Languages and Computing, vol. 31, 2015, pp. 275-282.

Better State Pictures Facilitating State Machine Characteristic Conjecture

Dang Duy Bui, Kazuhiro Ogata
School of Information Science, JAIST
1-1 Asahidai, Nomi, Ishikawa 923-1292, Japan
Email: {bddang,ogata}@jaist.ac.jp

***Abstract*—MCS shared-memory mutual exclusion protocol is used as an example to demonstrate that state picture designs should be better visualized. A case study has been conducted in which we demonstrate that better visualized state pictures make it possible to conjecture more non-trivial characteristics of state machines than the old state pictures. The lessons learned acquired through the case study are summarized as new tips on how to make state picture designs.**

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

I. Introduction

Our group has been developing a state machine graphical animation (SMGA) tool [1], [2]. Given a state picture design for a state machine, SMGA basically takes a sequence of states in text and plays a graphical animation for the state machine by regarding the sequence as a movie film based on the state picture design. SMGA has been used to make graphical animations of state machines formalizing Alternating Bit Protocol (ABP) [1], MCS shared-memory mutual exclusion protocol [3] and Suzuki-Kasami distributed mutual exclusion protocol [2]. One of the main goals to be achieved with SMGA is for human users to conjecture state machine characteristics or properties by visually/graphically observing graphical animations. This is because humans are good at visual perception [4].

We summarize some tips on how to design state pictures in our paper [2] published in 2019. The tips have been accumulated through the case studies for ABP [1], MCS protocol [3] and Suzuki-Kasami protocol [2]. They are still useful but we realized that there is some room to improve some state picture designs used for SMGA. Some part of each state of a state machine formalizing MCS protocol was not visualized sufficiently enough in that the part is almost the same as text representation. We have visualized the part, conducting a case study in which we have conjectured several non-trivial characteristics of (the state machine formalizing) MCS protocol. In this paper, we propose some more tips on how to make better state picture designs.

ShiViz [5] is a tool to visualize logs generated by distributed systems. The visualization used by ShiViz (still visualization) and the one (graphical animation) used by SMGA can be complementary. Magee, et al. [6] have proposed a way to visualize the behavior of a Labeled Transition System (LTS) described in FSP and developed a tool to support their proposed technique. One novelty of their approach to graphical animation of the LTS behavior is to use Timed Automata as formal semantics of animations. Their tool has been implemented with SceneBeans a library of JavaBeans. Their visualization is graphical animation like ours. SceneBeans could be used to implement a future version of SMGA.

We suppose that readers are falimilar with state machines and Maude [7] to some extent.

MCS protocol [8] is a shared-memory mutual exclusion protocol invented by Mellor-Crummey and Scott:

```
rs :  “Remainder Section”
l1 :  $next_p :=$ nop;
l2 :  $pred_p :=$ fetch&store($glock, p$);
l3 :  **if** $pred_p \neq$ nop {
l4 :      $lock_p :=$ true;
l5 :      $next_{pred_p} := p$;
l6 :      **repeat while** $lock_p$;  }
cs :  “Critical Section”
l7 :  **if** $next_p =$ nop {
l8 :      **if** comp&swap($glock, p,$ nop)
l9 :      **goto** rs;
l10 :     **repeat while** $next_p =$ nop;  }
l11 : $lock_{next_p} :=$ false;
l12 : **goto** rs;
```

It uses one global variable $glock$ and three local variables $next_p$, $pred_p$ and $lock_p$ for each process p. The protocol uses two atomic operations (or instructions): fetch&store and comp&swap.

II. Specification of MCS Protocol in Maude

We suppose that there are three processes that participate in MCS protocol. Let $\mathcal{M}_{\text{MCS}}$ formalize MCS protocol. Each state in $\mathcal{S}_{\text{MCS}}$ is expressed as follows:

```
{(glock: $bp$)
 (pc[p1]: $l_1$) (pred[p1]: $pp_1$)
            (lock[p1]: $b_1$) (next[p1]: $np_1$)
 (pc[p2]: $l_2$) (pred[p2]: $pp_2$)
            (lock[p2]: $b_2$) (next[p2]: $np_2$)
 (pc[p3]: $l_3$) (pred[p3]: $pp_3$)
            (lock[p3]: $b_3$) (next[p3]: $np_3$)}
```

where bp, pp_i and np_i for $i = 1, 2, 3$ are process IDs, l_i for $i = 1, 2, 3$ is a label, such as rs, l1 and cs, and b_i for

The work was partially supported by JSPS KAKENHI JP19H04082.
DOI reference number: 10.18293/DMSVIVA20-007

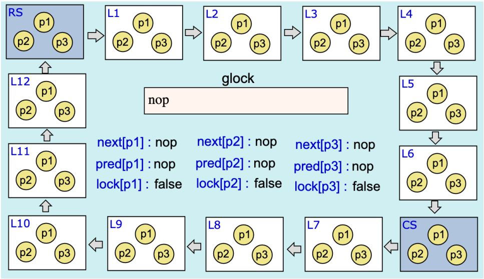
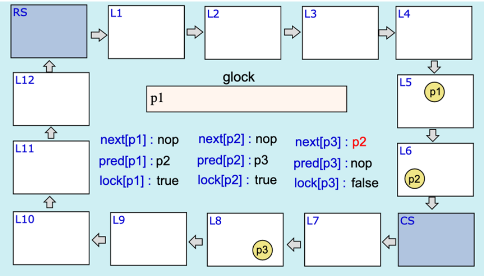


Figure 1. A state picture design for MCS protocol (1)



```
State 17 : (glock: p1 ) (pc[p1]: l5 ) (pc[p2]: l6 ) (pc[p3]: l8 ) (next[p1]: nop ) (next[p2]: nop ) (next[p3]: p2
) (lock[p1]: true ) (lock[p2]: true ) (lock[p3]: false ) (pred[p1]: p2 ) (pred[p2]: p3 ) (pred[p3]: nop )
```

Figure 2. A state picture for MCS protocol (1)

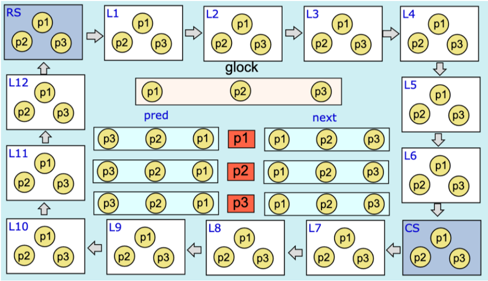
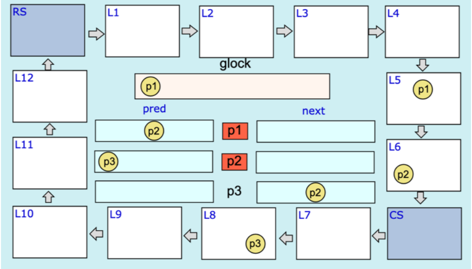


Figure 3. A state picture design for MCS protocol (2)



```
State 17 : (glock: p1 ) (pc[p1]: l5 ) (pc[p2]: l6 ) (pc[p3]: l8 ) (next[p1]: nop ) (next[p2]: nop ) (next[p3]: p2
) (lock[p1]: true ) (lock[p2]: true ) (lock[p3]: false ) (pred[p1]: p2 ) (pred[p2]: p3 ) (pred[p3]: nop )
```

Figure 4. A state picture for MCS protocol (2)

$i = 1, 2, 3$ is a Boolean value. Initially, bp, pp_i and np_i are nop, l_i is rs and b_i is false. $\mathcal{I}_{\mathrm{MCS}}$ consists of one state. Let `init` equal the initial state.

$\mathcal{T}_{\mathrm{MCS}}$ is described in terms of Maude rewrite rules. Among them are as follows:

```
rl [stprd] : {(glock: Q) (pc[P]: l2)
(pred[P]: Q1) OCs}
=> {(glock: P) (pc[P]: l3)
(pred[P]: Q) OCs} .
rl [chprd] : {(pc[P]: l3) (pred[P]: Q) OCs}
=> {(pc[P]: (if Q == nop then cs else l4 fi))
(pred[P]: Q) OCs} .
rl [stlck] : {(pc[P]: l4) (lock[P]: B) OCs}
=> {(pc[P]: l5) (lock[P]: true) OCs} .
```

where `OCs` is a Maude variable of observable component soups, `P`, `Q` and `Q1` are Maude variables of process IDs, and `B` is a Maude variable of Boolean values. `if` b `then` x `else` y `fi` equals x if b equals `true` and y if b equals `false`.

III. State Picture Design

States are expressed as braced soups of observable components, where observable components are name-value pairs. Therefore, it would be possible to automatically produce a straightforward state picture design. However, state pictures generated from the state picture design are almost the same as states in text. We summarize some tips on how to design state pictures for mutual exclusion protocols in our paper [2] published in 2019. Among the tips are as follows:

- *To recognize what sections there are at which each process or node is located, allocate the pane (or place) for each section such that the relations among the sections are visually perceived and display some diagram, such as a circle on which a process or node ID is written, on the designated pane;*
- *To recognize what pieces of information, such as the network for the Suzuki-Kasami protocol and variable* locked *for the test&set protocol, are shared by all processes or nodes, allocate the pane (or place) for each such piece of information such that we can visually perceive they are shared by all processes and nodes and display them on the designated panes adequately;*

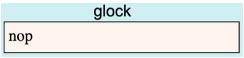
Nguyen and Ogata [3] made the state picture design shown in Fig. 1 for MCS protocol, which follows the tips. A state picture generated from the state picture design is shown in Fig. 2. The state picture allows us to immediately realize that processes p1, p2 and p3 are located at l5, l6 and l8, respectively. Nguyen and Ogata [3] conducted a case study in which several characteristics of MCS protocol can be discovered by observing graphical animations of MCS protocol such that each state picture used in the graphical animation is generated from the state picture design. For example, one of the characteristics found is as follows:

No state such that a process is at cs, l7, l8, l10, *or* l11 *and another process is at* cs, l7, l8, l10, *or* l11.

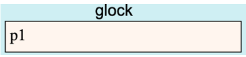
Let the characteristic be called Characteristic 0 in this paper.

Although the state picture shown in Fig. 2 also allows us to notice the values stored in the global variable $glock$ and the three local variables $next_p$, $pred_p$ and $lock_p$ for each process p, their representations on the state picture are almost the same as the text representation. Since SMGA requires and/or permits human users to make state picture designs, the representations must be able to be visually/graphically perceivable. To this end, we came up with the state picture design shown in Fig. 3. Fig. 4 shows a state picture generated from the state picture design.

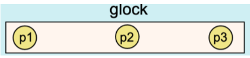
The design of the $glock$ representation used in Fig. 1 is as follows:



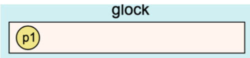
The value of $glock$ is nop, p1, p2 or p3. Regardless of the value, the value is displayed on the same place. For example, when the value is p1, it is displayed as follows:



The design of the $glock$ representation used in Fig. 3 is as follows:

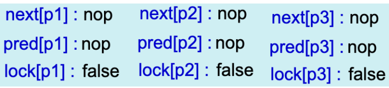


If the value is nop, nothing is displayed on the rectangle or pane for $glock$. If the value is p1, p1 is displayed at the left-most place of the rectangle for $glock$. If the value is p2, p2 is displayed at the middle place of the rectangle for $glock$. If the value is p3, p3 is displayed at the right-most place of the rectangle for $glock$. For example, when the value is p1, it is displayed as follows:



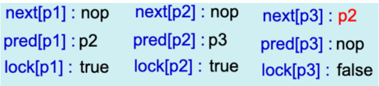
We can say that the $glock$ representation used in Fig. 3 and Fig. 4 helps human users more visually/graphically perceive its value than the one used in Fig. 1 and Fig. 2.

The design of the $next_p$, $pred_p$ and $lock_p$ representations for each process p used in Fig. 1 is as follows:



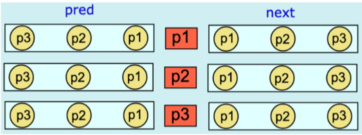
next[p1] : nop	next[p2] : nop	next[p3] : nop
pred[p1] : nop	pred[p2] : nop	pred[p3] : nop
lock[p1] : false	lock[p2] : false	lock[p3] : false

Regardless of the values of $next_p$, $pred_p$ and $lock_p$, their values are displayed on the same places. For example, when $next_{\mathrm{p1}}$ is nop, $pred_{\mathrm{p1}}$ is p2, $lock_{\mathrm{p1}}$ is true, $next_{\mathrm{p2}}$ is nop, $pred_{\mathrm{p2}}$ is p3, $lock_{\mathrm{p2}}$ is true, $next_{\mathrm{p3}}$ is p2, $pred_{\mathrm{p3}}$ is nop and $lock_{\mathrm{p3}}$ is false, those values are displayed as follows:

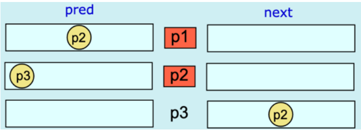


next[p1] : nop	next[p2] : nop	next[p3] : p2
pred[p1] : p2	pred[p2] : p3	pred[p3] : nop
lock[p1] : true	lock[p2] : true	lock[p3] : false

The design of the $next_p$, $pred_p$ and $lock_p$ representations for each process p used in Fig. 3 is as follows:



The $next_p$ representation appears at the right-most place, where there are three rectangles, the first, second and third ones of which from top are used for p1, p2 and p3, respectively. For example, if the value of $next_{\mathrm{p1}}$ is nop, nothing is shown on the first rectangle; if the value is pi, the circle on which pi is written is shown at the designated place on the first rectangle. The $lock_p$ representation appears at the middle place, which also indicates that the first, second and third rows are used for p1, p2 and p3, respectively. When $lock_{\mathrm{p}i}$ is true, the background color of pi is red; otherwise the color is non-red (or light blue). The $pred_p$ representation appears at the left-most place, where there are three rectangles, the first, second and third ones of which from top are used for p1, p2 and p3, respectively. For example, if the value of $pred_{\mathrm{p1}}$ is nop, nothing is shown on the first rectangle; if the value is pi, the circle on which pi is written is shown at the designated place on the first rectangle. For instance, when $next_{\mathrm{p1}}$ is nop, $pred_{\mathrm{p1}}$ is p2, $lock_{\mathrm{p1}}$ is true, $next_{\mathrm{p2}}$ is nop, $pred_{\mathrm{p2}}$ is p3, $lock_{\mathrm{p2}}$ is true, $next_{\mathrm{p3}}$ is p2, $pred_{\mathrm{p3}}$ is nop and $lock_{\mathrm{p3}}$ is false, those values are displayed as follows:

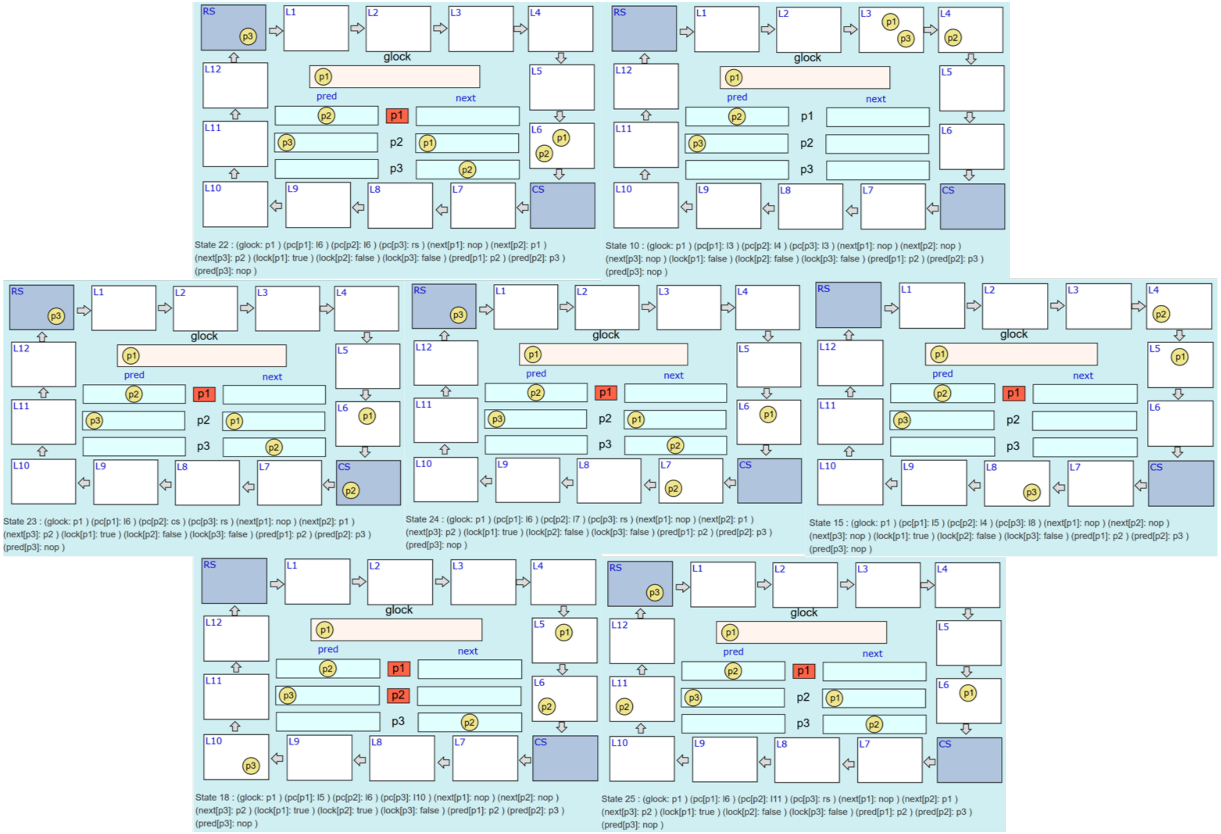


We can say that the $next_p$, $pred_p$ and $lock_p$ representations for each process p used in Fig. 3 and Fig. 4 help human users more visually/graphically perceive their values than the ones used in Fig. 1 and Fig. 2.

We advocate that state picture designs should be visualized as much as possible so that state pictures generated from the designs can help human users visually/graphically perceive the value of each observable component and some relations among multiple observable component values. In what follows, we demonstrate what characteristics of MCS protocol can be perceived when we use the state picture design shown in Fig. 3.

IV. Characteristic Discovery Based on New State Pictures

Carefully observing graphical animations for MCS protocol in which new state pictures, such as Fig. 2, were used, we realized that there is always at most one process at cs, l7, l8, l10 and l11 . This is exactly the same as Characteristic 0 discovered based on old state pictures, such as Fig. 4. We call these sections (cs, l7, l8, l10 and l11) CS region. We also noticed that there exists at most one process p such that p is located at l3 and $pred_p$ is nop and there exists at most



State 22 : (glock: p1) (pc[p1]: l6) (pc[p2]: l6) (pc[p3]: rs) (next[p1]: nop) (next[p2]: p1)
(next[p3]: p2) (lock[p1]: true) (lock[p2]: false) (lock[p3]: false) (pred[p1]: p2) (pred[p2]: p3)
(pred[p3]: nop)

State 10 : (glock: p1) (pc[p1]: l3) (pc[p2]: l4) (pc[p3]: l3) (next[p1]: nop) (next[p2]: nop)
(next[p3]: nop) (lock[p1]: false) (lock[p2]: false) (lock[p3]: false) (pred[p1]: p2) (pred[p2]: p3)
(pred[p3]: nop)

State 23 : (glock: p1) (pc[p1]: l6) (pc[p2]: cs) (pc[p3]: rs) (next[p1]: nop) (next[p2]: p1)
(next[p3]: p2) (lock[p1]: true) (lock[p2]: false) (lock[p3]: false) (pred[p1]: p2) (pred[p2]: p3)
(pred[p3]: nop)

State 24 : (glock: p1) (pc[p1]: l6) (pc[p2]: l7) (pc[p3]: rs) (next[p1]: nop) (next[p2]: p1)
(next[p3]: p2) (lock[p1]: true) (lock[p2]: false) (lock[p3]: false) (pred[p1]: p2) (pred[p2]: p3)
(pred[p3]: nop)

State 15 : (glock: p1) (pc[p1]: l5) (pc[p2]: l4) (pc[p3]: l8) (next[p1]: nop) (next[p2]: nop)
(next[p3]: nop) (lock[p1]: true) (lock[p2]: false) (lock[p3]: false) (pred[p1]: p2) (pred[p2]: p3)
(pred[p3]: nop)

State 18 : (glock: p1) (pc[p1]: l5) (pc[p2]: l6) (pc[p3]: l10) (next[p1]: nop) (next[p2]: nop)
(next[p3]: p2) (lock[p1]: true) (lock[p2]: true) (lock[p3]: false) (pred[p1]: p2) (pred[p2]: p3)
(pred[p3]: nop)

State 25 : (glock: p1) (pc[p1]: l6) (pc[p2]: l11) (pc[p3]: rs) (next[p1]: nop) (next[p2]: p1)
(next[p3]: p2) (lock[p1]: true) (lock[p2]: false) (lock[p3]: false) (pred[p1]: p2) (pred[p2]: p3)
(pred[p3]: nop)

Figure 5. Some pictures for extended CS region

one process p such that p is located at l6 and $lock_p$ is false. Moreover, if p is located at l3 and $pred_p$ is nop, there is no process in CS region and there is no process q at l6 such that $lock_q$ is false, and if p is located at l6 and $lock_p$ is false, there is no process in CS region and there is no q at l3 such that $pred_q$ is nop. l3 and l6 are only the sections from which processes enter CS region. We call CS region plus l3 and l6 extended CS region. The first characteristic that can be conjectured by carefully observing graphical animations for MCS protocol in which new state pictures, such as Fig. 2, are used is as follows:

- Characteristic 1: There exists at most one process except for processes p such that (1) p is located at l3 and $pred_p$ is not nop and (2) p is located at l6 and $lock_p$ is not false in extended CS region.

Extended CS region is one key concept that captures one important aspect of MCS protocol and Characteristic 1 is very crucial in that several other characteristics can be discovered based on the characteristic. Fig. 5 shows some state pictures that capture Characteristic 1. On the top left state picture, there are two processes p1 and p2 in extended CS region such that both are located at l6, $lock_{\mathrm{p1}}$ is true and $lock_{\mathrm{p2}}$ is false and therefore p2 is only the process in extended CS region in the sense of Characteristic 1 because p1 satisfies condition (2) in Characteristic 1. On the top right state picture, there are two processes p1 and p3 in extended CS region such that both are located at l3, $pred_{\mathrm{p1}}$ is p2 and $pred_{\mathrm{p3}}$ is nop and therefore p3 is only the process in extended CS region in the sense of Characteristic 1 because p1 satisfies condition (1) in Characteristic 1. Each of the other five state pictures shows that there exists one process at one section of CS region and for all processes q located at l3 and l6 if any, $pred_q$ is not nop and $lock_q$ is not false, respectively.

Focusing on a process in CS region or extended CS region, we can recognize the following characteristics:

- Characteristic 2.1: Whenever there is a process at l10, there is at least one process at l3, l4, l5 or l6;
- Characteristic 2.2: Whenever there is a process at l11, there is at least one process at l6.

The bottom left state picture of Fig. 5 is an example of Characteristic 2.1 and the bottom right state picture of Fig. 5 is an example of Characteristic 2.2.

Carefully observing the $next_p$ and $pred_p$ representations for each process p, nothing may be displayed and the circle on which q that is different from p is written may be displayed but the circle on which p is written is never displayed. Thus, we can realize the following characteristics:

- Characteristic 3.1: The value of $next_p$ for each process p is never p.
- Characteristic 3.2: The value of $pred_p$ for each process p is never p.

For example, taking a look at the $next_{\mathrm{p1}}$ representation or rectangle, it is visually/graphically observable that the circle on which p1 is written never comes into sight on the designated position. This is because $next_{\mathrm{p1}}$ is visually/graphically represented in the new state picture design.

Some relations between two observable components can be discovered by carefully observing graphical animations, from which some characteristics can be conjectured. A relation between the glock observable component and the $\mathrm{pc}[p]$ observable component can be perceived by graphical animations and allows us to conjecture the following characteristics:

- Characteristic 4.1: If $glock$ is nop, then there is no process at l3, l4, l5 or l6 or in CS region;
- Characteristic 4.2: If $glock$ is a process p, then p is located at l3, l4, l5 or l6 or in CS region;
- Characteristic 4.3: If $glock$ is not nop (or equivalently a process), then there exists a process in CS region.

A relation between the $\mathrm{pred}[p]$ observable component and the $\mathrm{pc}[p]$ observable component allows us to conjecture the following characteristic:

- Characteristic 5: If $pred_p$ for each process p is nop, then p is never located at l4, l5, l6 or l12.

A relation between the $\mathrm{pc}[p]$ observable component and the $\mathrm{next}[p]$ observable component allows us to conjecture the following characteristics:

- Characteristic 6.1: If each process p is located at l2 or l9, then $next_p$ is nop;
- Characteristic 6.2: If $next_p$ for each process p is nop, then p is not located at l11 or l12.

A relation between the $\mathrm{lock}[p]$ observable component and the $\mathrm{pc}[p]$ observable component allows us to conjecture the following characteristics:

- Characteristic 7.1: If $lock_p$ for each process p is true, then p is located at l5 or l6;
- Characteristic 7.2: If each process p is located at l5, then $lock_p$ is true.

Note that when a process p is located at l6, another process may set $lock_p$ to false.

It is necessary to fix the values of some observable components so as to discover some similarities of multiple states and/or some relations among observable components. It is not straightforward to do so by observing graphical animations because we need to remember states in which the former observable components have the fixed values. For example, it must not be reasonable to remember all states in a graphical animations such that $pred_{\mathrm{p1}}$ is p2 and p2 is located at rs, l1, l2, l9 or l12. One possible remedy for it is to use the Find Pattern functionality of SMGA [1], [3], [2]. Given a regular expression written by human users, the Find Pattern functionality finds all states in an input sequence of states such that they match the regular expression. When we would like to find all states in an input sequence of states such that $pred_{\mathrm{p1}}$ is p2 and p2 is located at rs, l1, l2, l9 or l12, it suffices to write the following regular expression:

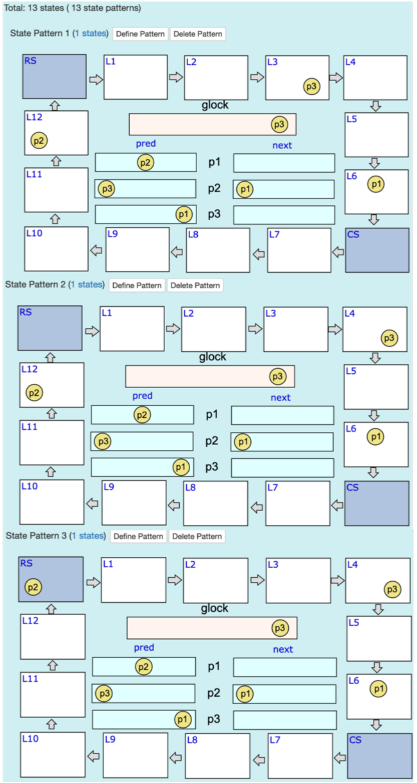

Total: 13 states (13 state patterns)
State Pattern 1 (1 states)
Define Pattern
Delete Pattern
State Pattern 2 (1 states)
State Pattern 3 (1 states)
RS
L1
L2
L3
L4
L5
L6
L7
L8
L9
L10
L11
L12
CS
glock
pred
next
p1
p2
p3

Figure 6. Three state pictures discovered by Find Pattern functionality.

```
state['pred[p1]'] == 'p2' &&
(state['pc[p2]'] == 'l1' ||
 state['pc[p2]'] == 'l2' ||
 state['pc[p2]'] == 'rs' ||
 state['pc[p2]'] == 'l9' ||
 state['pc[p2]'] == 'l12')
```

`state['pred[p1]'] == 'p2'` says that $pred_{\mathrm{p1}}$ equals p2. `&&` and `||` are logical conjunction and disjunction, respectively. Fig. 6 shows three state pictures among the states discovered by the Find Pattern functionality for the regular expression.

It does not suffice, however, to use the regular expression abovementioned so as to conjecture non-trivial characteristics. This is because the pair $(\mathrm{p1}, \mathrm{p2})$ is one possible combination and there are five more combinations to consider: $(\mathrm{p1}, \mathrm{p3})$, $(\mathrm{p2}, \mathrm{p1})$, $(\mathrm{p2}, \mathrm{p3})$, $(\mathrm{p3}, \mathrm{p1})$ and $(\mathrm{p3}, \mathrm{p2})$. Note that we do not need to take the three combinations $(\mathrm{p1}, \mathrm{p1})$, $(\mathrm{p2}, \mathrm{p2})$ and $(\mathrm{p3}, \mathrm{p3})$ into account because of

Characteristic 3.1. Let (p, q) be each of the six combinations to consider. Carefully observing all states discovered by the Find Pattern functionality for the six combinations to consider, we can conjecture some non-trivial characteristics:

- Characteristic 8.1: If $pred_p$ is q and q is located at rs, $\mathrm{l1}$, $\mathrm{l2}$, $\mathrm{l9}$, $\mathrm{l11}$ or $\mathrm{l12}$, then p is not located at $\mathrm{l3}$, $\mathrm{l4}$ or $\mathrm{l5}$;
- Characteristic 8.2: If $pred_p$ is q and $glock$ is q, then p is not located at $\mathrm{l3}$, $\mathrm{l4}$ or $\mathrm{l5}$;
- Characteristic 8.3: If $pred_p$ is q, $next_q$ is not nop and p is located at $\mathrm{l3}$, $\mathrm{l4}$, $\mathrm{l5}$, $\mathrm{l6}$, cs, $\mathrm{l7}$, $\mathrm{l8}$ or $\mathrm{l10}$, then q is not located at $\mathrm{l3}$, $\mathrm{l4}$ or $\mathrm{l5}$.

We noticed that there are both states in which $next_q$ is nop and those in which $next_q$ is not nop among the states found by the Find Pattern functionality for the regular expression of the condition of Characteristic 8.1, while there are only states in which $next_q$ is nop among the states found by the Find Pattern functionality for the regular expression of the condition of Characteristic 8.2. Then, finding the states with the Find Pattern functionality for the regular expression of the second sub-condition only of Characteristic 8.2 because the first sub-condition is shared by both characteristics, we realized that $next_q$ is nop in all of them. Therefore, we came up with the following characteristic:

- Characteristic 9: If $glock$ is a process p (or equivalently non-nop), $next_p$ is nop.

We have used the Maude reachability analyzer (or the search command) to model check that all characteristics conjectured in this section are invariant properties with respect to the state machine formalizing MCS protocol and have not found any counterexamples for each characteristic.

V. Some Lessons Learned

Let us summarize some lessons learned through the case study with SMGA and MCS protocol. We have already described one important lesson learned in Sect. III. Let us repeat it:

- State picture designs should be visualized as much as possible so that state pictures generated from the designs can help human users visually/graphically perceive the value of each observable component and some relations among multiple observable component values.

We have demonstrated that the new state picture design (more visualized) allows us to more visually/graphically perceive the values of $glock$ and $next_p$, $pred_p$ & $lock_p$ for each process p than the old state picture design (less visualized) and then conjecture more non-trivial characteristics of MCS protocol than the old ones in Sect. IV.

To make the lesson more practical, we describe some more concrete lessons or guides.

- When an observable component can have two different values, such as $lock_p$ for each process p, it should be visually/graphically represented as a light bulb. For example, if an observable component has one value, such as on, we should use a fancy or light color; if it is the other value, such as off, we should use a plain or dark color.
- When an observable component can have three or more (but moderate) different values, such as $glock$ and $next_p$ & $pred_p$ for each process p, we should prepare some designated area, such as a rectangle, and a specific position in the area for each value where some visual object, such as a circle on which the value is written, is displayed; if the observable component has a value, only the visual object for it should be displayed and the other visual objects for the other values should disappear; there may be some special value, such as nop, and if the observable component has such a value, nothing should be displayed.
- If there are some local variables, such as $next_p$, $pred_p$ and $lock_p$ for each process p, to processes or nodes, then we should design the layout of the visual representations for them so that we can visually/graphically identify what variables or observable components are local to what processes or nodes; for example, all local variables for each process should be aligned horizontally.

VI. Conclusion

We have used MCS protocol to demonstrate two things. One is that the old state picture design for the protocol can be more visualized. The other is that the new state picture design makes it possible for human users to more visually/graphically perceive interesting phenomena in graphical animations and then conjecture more non-trivial state machine characteristics. We have summarized some new tips on how to make state picture designs.

References

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

[2] 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.

[3] 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.

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

[5] I. Beschastnikh, et al., “Visualizing distributed system executions,” *ACM TOSEM*, vol. 29, no. 2, p. 38 pages, 2020.

[6] J. Magee, et al., “Graphical animation of behavior models,” in *22nd ICSE*, 2000, pp. 499–508.

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

[8] J. M. Mellor-Crummey and M. L. Scott, “Algorithms for scalable synchronization on shared-memory multiprocessors,” *ACM TOCS*, vol. 9, no. 1, pp. 21–65, 1991.

A General Parsing Algorithm with Context Matching for Context-Sensitive Graph Grammars

Yang Zou, Xiaoqin Zeng, Yun Zhu
Institute of Intelligence Science and Technology
School of Computer and Information, Hohai University
Nanjing, China
{yzou, xzeng}@hhu.edu.cn, zhuyunhhu@163.com

Tingting Sha
5G Innovation Center
China United Network Communications Group Co., Ltd.
Beijing, China
shatt@chinaunicom.cn

***Abstract*—Context-sensitive graph grammars have been intuitive and rigorous formalisms for specifying visual programming languages, as they are sufficient expressive and equipped with parsing mechanisms. Parsing has been a fundamental issue in the research of context-sensitive graph grammars. However, the existent parsing algorithms are either inefficient or confined to a minority of graph grammars. This paper proposes a general parsing algorithm with two embedded strategies, one is context matching, and the other is production set partitioning. The two strategies can greatly narrow down the search space of redexes and thus considerably improve the parsing performance, even though the worst-case time complexity is not theoretically reduced. Moreover, a case study along with detailed analysis is provided to demonstrate the paring process and parsing performance of the proposed algorithm.**

Keywords-visual languages; context-sensitive graph grammar; context matching; production set partitioning; parsing algorithm

I. Introduction

Visual Programming Languages (VPLs) have been widely and frequently adopted in many disciplines of computer science. Like the textual programming languages that are usually equipped with proper formal syntax definition and parser, VPLs also need the support from such mechanisms. Quite a few approaches have been proposed for the specification and parsing of VPLs [1-3]. As a natural extension of formal grammar theory, graph grammars offer the mechanisms that can formally specify and parse VPLs [4], just like formal grammars do for string languages. In contrast to formal grammars, graph grammars consist of a set of productions in form of a pair of graphs rather than strings. However, the generalization from formal grammars to graph grammars brings about a new problem, the embedding problem [5].

Various graph grammar formalisms have been proposed in the literature [6-9], most of which are context-free or context-sensitive. The expressive power of a graph grammar depends on the type it belongs to as well as the embedding mechanism being integrated into the productions [10]. Context-sensitive graph grammars tend to be more expressive than context-free ones, when identically confined to less complex embedding mechanisms and invariant embedding in particular. Since context-free graph grammars even have difficulty in specifying a large portion of graphical VPLs [8-9], recent research in this discipline pays much more attention to context-sensitive graph grammars [11-13].

The existent context-sensitive graph grammar formalisms include Layered Graph Grammar (LGG) [8], Reserved Graph Grammar (RGG) [9], Contextual Layered Graph Grammar (CLGG) [14], Spatial Graph Grammars (SGG) [15], Context-Attributed Graph grammar (CAGG) [16], Edge-based Graph Grammar (EGG) [17], Temporal Graph Grammar (TGG) [13], etc. LGG and RGG are representative formalisms among them. CLGG, SGG, and TGG are typical extensions of LGG, RGG, and EGG, respectively. CLGG supports three extra mechanisms in productions to define more complex VPLs, SGG extends RGG by adding to productions a kind of spatial specification mechanism, and TGG generalizes LGG by augmenting temporal specification ability to productions.

In context-sensitive graph grammars, contexts pertaining to a production generally refer to the subgraphs neighboring to the rewritten portion of its left graph in potential host graphs [5], which describe the situations under which the production can be applied. According to how the context portion of a production is expressed, these formalisms are divided into two categories: explicit and implicit [18]. LGG and RGG are the typical examples of the former and the latter, respectively. An inherent weakness of implicit formalisms is that they are not intuitive. This is due to the fact that the context portion of a production is even not the complete immediate contexts. For example, the context portion of RGG productions is a set of marked vertices whose exact meaning is undefined. To make implicit context in productions explicit, a formal definition of context is proposed and the properties are characterized in [18], which provide a theoretical foundation for the computation of context.

Parsing has been one of the fundamental issues in the study of graph grammars. The LGG formalism is equipped with a parsing algorithm that consists of a top-down phrase for searching potential redexes and a bottom-up phrase for derivation from the initial graph to a given host graph [8]. Obviously, the process is rather complicated. RGG provides a naive Selection-Free Parsing Algorithm with polynomial time complexity, provided that the graph grammars are selection-free [9]. The condition selection-free is similar to the concept of local confluence in the literature. Unfortunately, this condition is not frequently satisfied in applications [18]. To deal with this problem, a general parsing algorithm is described in [19].

DOI reference number: 10.18293/DMSVIVA20-008

Without a doubt, the mechanism of backtracking will inevitably appear in the parsing process, so as to traverse the search space that exponentially increases with the size of the host graph. Therefore, how to reduce backtracking becomes a critical challenge in parsing algorithms.

As to reducing backtracking, which is commonly caused by blind trials, i.e., reduction steps with unnecessary redexes, a feasible way is to try to avoid as many as possible irrelevant productions and unnecessary redexes in each step of reduction, namely to push less redexes into the stack in that they will probably be popped out for reduction in backtracking at a later time during the parsing process. Context can be employed to serve this purpose. When a redex is found in the host graph, the contexts of the corresponding production can be used to further check whether it is valid by matching the contexts with its circumstance in the host graph. If it is not matched, the redex can be directly discarded. Consequently, the potential backtracking is reduced. The approach to computation of context is proposed in [20], paving the way for context applications.

The technical contributions in this paper are as follows: Based on the RGG formalism, it proposes a general parsing algorithm with two essential strategies embedded, one is context matching, and the other is production set partitioning. The former can be utilized to reexamine the found redexes to exclude the unnecessary redexes, whereas the latter can be employed to precisely choose relevant instead of all the productions so as to narrow down the search for redexes. The two strategies will considerably improve the parsing performance of the proposed algorithm, even though the worst-case time complexity is not theoretically reduced. Moreover, a case study along with detailed analysis is provided to illustrate the paring process and parsing performance of the algorithm.

The remaining of the paper is organized as follows: Section 2 reviews the RGG formalism, and the notions of partial precedence graph and context. Section 3 presents a general parsing algorithm with two strategies established on the notions. Section 4 conducts an analysis of the algorithm from two aspects. Finally, Section 5 concludes the paper and proposes future research.

II. PRELIMINARIES

A. The RGG Formalism

A graph grammar consists of an initial graph and a set of productions (also called graph rewriting rules). Each production consists of two graphs called left graph and right graph respectively, and can be applied to another graph called host graph. Each node in a production is either a terminal or a non-terminal node. A graph grammar defines a graph language composed of those graphs that can be derived from the initial graph by repeated applications of the productions and whose nodes are all terminal ones.

In this paper, RGG is taken as the representative of implicit context-sensitive graph grammar formalism. RGG is a context-sensitive graph grammar formalism [9]. It introduces a node-edge format to represent graphs in which each node is organized as a two-level structure, where the large surrounding rectangle is the first level, called super vertex, and other embedded small rectangles are the second level, called vertices. Either a vertex or a super vertex can be the connecting point of an edge. In addition to the two-level node structure, the RGG also introduces a marking technique that divides vertices into two categories: marked and unmarked ones. Each marked vertex of a production is identified by an integer that is unique in the left or right graph where the vertex lies. A production is properly marked if each marked vertex in the left graph has a counterpart marked by the same integer in the right graph, and vice versa. In the process of a production application, when a redex is matched in a host graph, each vertex that corresponds to a marked vertex in the left or right graph preserves its associated edges connected to nodes outside of the redex, which avoids the appearance of dangling edges during the subsequent subgraph replacement provided that an additional embedding rule is also enforced. The embedding rule states that if a vertex in the right (or left) graph of a production is unmarked and has an isomorphic vertex in the redex of a host graph, then all the edges connected to the vertex should be completely inside the redex.

An RGG that specifies process flow diagrams is depicted in Figure 1, and a simple process flow diagram is shown in Figure 5(a).

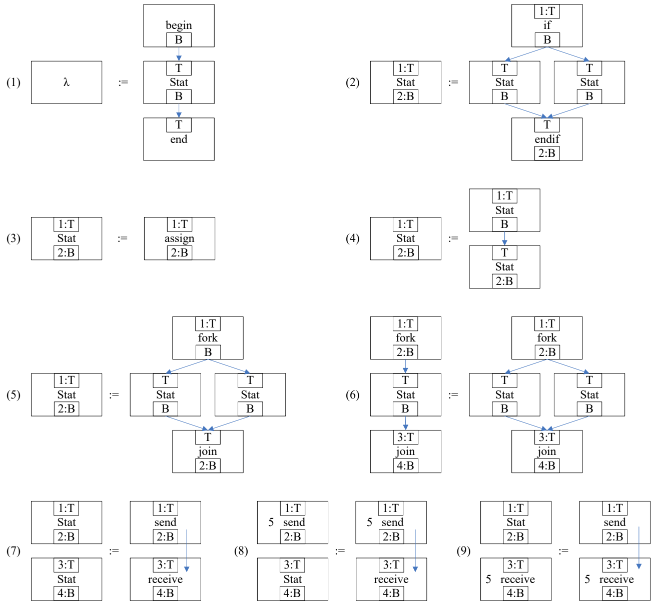


Figure 1. A graph grammar for process flow diagrams.

B. Basic Concepts and Notations

For the sake of clarity and simplicity, some basic concepts and notations are listed below. Note that graphs are directed ones in the node-edge format and only vertices in productions might be marked.

RGG: A reserved graph grammar is a triple (A, P, Ω), where A is an initial graph, P a set of graph grammar productions, Ω a finite label set consisting of two disjoint sets Ω^T and Ω^{NT}(called terminal label set and nonterminal label set, respectively). For any production $p := (L, R) \in P$, three conditions are satisfied: R is non-empty, L and R are both over Ω, and the size of R are not less than that of L.

$p \coloneqq (L, R)$: A production with a pair of marked graphs: the left graph L and right graph R. The notations $p.L$ and $p.R$ represent the left and right graph of a production p, respectively. For any graph G, $G.N$ and $G.E$ denote the set of nodes and edges, respectively; for any edge e, $s(e)$ and $t(e)$ represent the source and target vertex of e, respectively, and $l(e)$ is the label on e.

The function Mcc is a mapping from graphs to their sets of maximally connected subgraphs (components).

Redex: A redex is a subgraph in a host graph that is isomorphic to the left or right graph of a production.

A production's L-application to a host graph is to find in the host graph a redex of the left graph of the production and replace the redex in the host graph with the right graph of the production, while an R-application is a reverse process, i.e., to find a redex of the right graph of the production and replace the redex with the left graph.

Direct Partial Precedence: Let $gg \coloneqq (A, P, \Omega)$ be an RGG, and $p_1, p_2 \in P$ be two productions, p_1 directly partially precedes p_2, denoted as $p_1 \preccurlyeq_d p_2$, if $\exists S \subseteq Mcc(p_2.L)$ such that $S \sqsubseteq Mcc(p_1.R)$. The direct partial precedence relation between them is denoted by the pair $\langle p_1, p_2 \rangle$. The direct partial precedence relation on the set P of productions is defined as $\preccurlyeq_P = \{\langle p_1, p_2 \rangle | p_1, p_2 \in P \wedge p_1 \preccurlyeq_d p_2\}$.

The partial precedence relation is the closure of the direct partial precedence relation on a set P. Partial precedence is a kind of relation between a pair of components chosen from two distinct productions, whereas total precedence describes the same relation between two sets of components from a subset of productions and a single production respectively.

C. Partial Precedence Graph

The notion of partial precedence graph is proposed in [21], which is excerpted as follows:

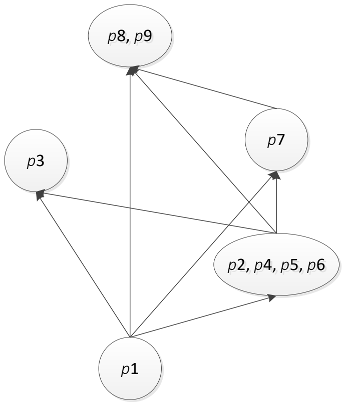


Figure 2. The partial precedence graph of the production set of an RGG.

Definition 1. Let P be the production set of an RGG and $\preccurlyeq_P$ the direct partial precedence relation on P. A subset $D \subseteq P$ is a precedence ring if $\forall p, p' \in D(p \neq p' \rightarrow p \preccurlyeq p' \wedge p' \preccurlyeq p)$. A precedence ring D is maximal if there does not exist another precedence ring $D' \subseteq P$ such that $D \subset D'$.

Definition 2. Let P be the production set of an RGG, $\preccurlyeq_P$ the direct partial precedence relation on P, and $W = \{D_1, \cdots, D_k\}$ the set of maximal precedence rings of P. The partial precedence graph of P is the pair (W, E), where $E = \{(D_i, D_j) | \exists p, p' \in P(p \in D_i \wedge p' \in D_j \wedge i \neq j \wedge p \preccurlyeq_d p'), 1 \leq i, j \leq k\}$ is the set of edges between the elements of Pr.

The partial precedence graph of the production set of the RGG in Figure 1 is depicted in Figure 2, where each node is a maximal precedence ring that is represented as a cycle or an eclipse, depending on the number of productions it contains. The node including only the first production $p1$ can be neglected. A conclusion with respect to the partial precedence graph is drawn in [21] that if a host graph can be parsed by a graph grammar, then a valid reduction that conforms to the partial precedence graph of the graph grammar exists.

D. Notion of Context

The sets of partial or total precedence relations with respect to a graph grammar establish an order of production applications, which can be exploited to discover potential situations in which any of the productions is applicable for derivation. The situations are referred to as contexts.

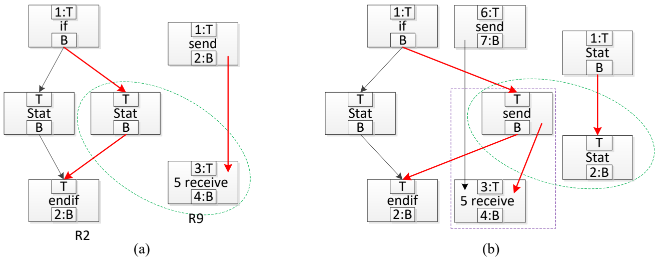


Figure 3. Contexts. (a) A level-1 context of $p9$. (b) A level-2 context of $p8$.

Given two productions p_1 and p_2, if p_1 directly partially precedes p_2, then $p_1.R$ contains a context of p_2 or merely a portion of a context, depending on whether $p_2.L$ consists of only one or at least two maximally connected components. As for the former case, $\{p_1\} \prec_d p_2$ readily holds and a context of p_2 immediately follows; whereas in the latter case, a subset of productions involving p_1 that directly totally precedes p_2 is pursued so as to form a complete context for p_2. As a third case, a total precedence relation can be sought to build a rather deeper complete context. A formal definition on context can be found in [20].

A context is of the form (U, Z), where U is the context subgraph, and Z is the connection between the context and the redex.

Two contexts at different levels of a production are shown in Figure 3. A level-1 context of $p9$ is shown in Figure 3(a), where the left component of the graph is $R2$ (the right graph of production 2), the right one is $R9$, the subgraph enclosed by the green dashed ellipse is the redex of $L9$ (the left graph of $p9$), and the context consists of two parts U and Z: U is the rest of the whole graph minus the redex and Z the set of thick red edges that connect U to the redex. Figure 3(b) shows a level-2 context of $p8$. In this graph, the left and right component is $R10$ (an extended production of $p9$ with context) and $R4$, respectively, the subgraph enclosed by the purple dashed rectangle is the isomorphic image of the underlying production of $p10$, and the one surrounded by the green dashed eclipse is the redex of the left graph of $p8$.

III. General Parsing Algorithm

In this section, two algorithms that correspond to two parsing strategies: production set partitioning and context matching, are presented, and then a general parsing algorithm with the two strategies embedded follows.

A. The Strategy of Context Matching

Apparently, backtracking is the main cause of high time complexity of a general parsing algorithm, due to its blind trial of reductions. During the process of parsing a host graph, some unnecessary redexes are found and the corresponding reduction steps are conducted, and consequently a final graph will be obtained that is not the initial graph of the involved graph grammar and to which no more reductions can be conducted. This gives rise to backtracking in the process of parsing. Therefore, identifying unnecessary redexes so as to avoid unnecessary reductions is a direct and effective way to improve parsing performance. In this paper, the unnecessary redexes are termed as false positive redexes.

```
Algorithm 1. ContextMatching.
input: A host graph $H$, a production $p$, a redex redex of $p$,
and a context set $p.C$ of $p$.
output: A Boolean value True or False.
begin
  $C' = p.C$;
  while $C' \neq \emptyset$ do
    Find an element $c = (U, Z) \in C'$;
    $q$ = OrderNodeSequence($U$);
    $D$ = FindNodeSequence($H$, $q$, redex, $Z$);
    for $\forall d \in D$ do
      if Match($H$, $d$, $U$) then
        return ("True");
    end for
    $C' = C' \backslash \{c\}$;
  end while
  return "False";
end
```

As for the RGG formalism, a subgraph will be found as a redex of a production's right graph in a host graph only when it is isomorphic to the right graph and the associated constraint on unmarked vertices is satisfied, regardless of the circumstance in which the subgraph lies. The circumstance, also called situation, refers to the rest of the host graph minus the written portion of the redex.

A strategy called context matching can be adopted to exclude false positive redexes by using the information of its situation. When a redex is found, an additional test is conducted to match the situation with contexts of the corresponding production.

Note that contexts of a production are potential situations in which it can be applied for derivation, and conversely, for reduction. Theoretically, the situation of a redex must match a context of any level (if applicable) for the corresponding production. The result of context matching is employed to determine whether a redex is false positive or true positive.

A redex is defined as true positive if the situation is matched with a level-1 context of the production to which it corresponds, or false positive, otherwise. If a redex is examined as false positive, it will be directly discarded, i.e., excluded from the set of candidates for reduction.

The procedure OrderNodeSequence(U) sequences the nodes of the graph U in terms of the alphabetic order of their labels; the procedure FindNodeSequenceSet(H, q, *redex*, Z) returns all the possible node sequences from the host graph H, each of which is the same as the given node sequence q such that the set Z exactly specifies the connection between the *redex* and the nodes in H that constitutes this node sequence; and the function Match(H, d, U) checks whether U is isomorphic to a subgraph in the host graph H whose node sequence is d, if so, "True" is returned.

B. The Strategy of Production Set Partitioning

Algorithm 2 computes the partition of a set of productions according to its partial precedence graph, and returns the result in the form of an array of sets.

```
Algorithm 2. SetPartitioning.
input: A set $P$ of productions.
output: An array of sets that is a partition of $P$.
begin
  Initialize $prt$[]; //An array of sets;
  Create the partial precedence graph $(V, E)$ of $P$;
  $i = 0$;
  while $V \neq \emptyset$ do
    Find an element $W \in V$ such that $\forall e \in E(s(e) \neq W)$;
    $prt[i] = W$;
    $i = i + 1$;
    $V = V \backslash \{W\}$;
    $E = E \backslash \{e | t(e) = W\}$;
  end while
  return $prt$;
end
```

The function map: $P \rightarrow prt$ is a surjective mapping from a production set to its partition set (a set of disjoint subsets of P whose union is P), where prt is the output of SetPartitioning(P). That is, if $p \in P$, then $map(p) = prt[i]$, for some constant i such that $p \in prt[i]$. The function $succ(Q)$ computes a successive subset in the partition, i.e., if $Q = prt[i]$, $succ(Q) = prt[i + 1]$, and if Q is the last element, then $succ(Q)$ is set to be an empty set.

C. The General Parsing Algorithm

Generally, parsing is such a process that attempts to perform a series of R-applications to reduce a given host graph to an initial graph with a designated graph grammar. It usually needs to deal with the following four interrelated steps:

- Search for the redexes of a production's right graph in a host graph;

- conduct an R-application with a certain redex to produce a new intermediate graph from the initial host graph or current intermediate graph;

- apply in turn the above two steps repeatedly, trace all the reduction paths, and backtrack if no more reductions can be made to the current intermediate graph that is not the initial graph;

- halt until a path leading to the initial graph is found or all possible paths have been exhausted.

Algorithm 3. Parsing.

```
input: A host graph $H$, a set $P$ of productions and their
contexts.
output: Success with a valid parsing path, or failure.
begin
  Initialize redexStack, hostStack, and traceStack; //Stacks
  $prt$ = SetPartitioning($P$);
  while $H \neq A$ do
    Push(redexStack, (DELIMITER, NIL));
    $Q = prt[0]$;
    tag = 0;
    while $tag = 0 \wedge Q \neq \emptyset$ do
      for $\forall p \in Q$ do
        redexset = FindRedex($H$, $p.R$);
        for ∀redex ∈ redexset do
          if ContextMatching($H$, $p$, redex, $p.C$) then
            Push(redexStack, (redex, $p$));
            tag = 1;
          end if
        end for
      end for
      $Q = succ(Q)$;
    end while
    (redex, $p$) = Pop(redexStack);
    while (redex = DELIMITER) do
      $H$ = Pop(hostStack);
      Pop(traceStack);
      (redex, $p$) = Pop(redexStack);
    end while
    if redex = NIL then
      return("Invalid");
    Push(hostStack, $H$);
    Push(traceStack, ($H$, redex, $p$));
    $H$ = RightApplication($H$, redex, $p$);
  end while
  return traceStack;
end
```

Algorithm 3 is a general parsing algorithm. It tries to trace all possible reduction paths starting from a given host graph to decide whether there exists one path that leads to the initial graph.

An assumption is that two basic procedures have already been available, one is searching for all the redexes of a production's right graph and the other is the accomplishment of an R-application. Further details on these two procedures is not discussed, since the procedures with similar functions can be found in the literature [1]. Three stacks are employed to record the redexes found, the intermediate host graph yielded, and the successful reduction path returned, respectively. As there is a need in the tracing to maintain a correspondence between a redex and its host graph for performing a reduction and this correspondence is usually many to one, it introduces a delimiter in the redex stack to delimit a group of redexes that corresponds to the same host graph. The delimiter makes the correspondence manageable by synchronizing the contents in the two stacks redexStack and hostStack. The algorithm takes a host graph and a set of productions and their contexts as input parameters and returns a definite answer to indicate whether the graph is valid or not. When the graph is valid, a successful reduction path is returned accordingly.

Two essential strategies are embedded in the general parsing process.

- Production set partitioning. The set of productions of a given graph grammar are partitioned into several disjoint subsets in terms of its partial precedence graph. Meanwhile, the subsets are ordered in accordance to the partial order suggested by the graph. When searching for redexes in an intermediate host graph, only a subset rather than the whole set of productions will be traversed. This operation is done by Algorithm 2.

- Context matching. Upon all the redexes of a production's right graph in an intermediate host graph are computed, the algorithm will not directly push them one by one into the stack *redexStack* in preparation for possible backtracking. Instead, it further decides whether each of them is an "appropriate" redex by matching the circumstance in the host graph where it is situated with its contexts. If a redex is verified in this way, it will then be pushed into the stack; otherwise, it will be discarded. This operation is done by Algorithm 1.

Some details of the algorithm are explained as follows. In the outmost while-loop, Q is an element from the partition set prt, denoting a subset of productions whose redexes will be searched in H. *tag* is Boolean value, indicating whether there is a redex of some production in Q that has been pushed into the *redexStack*. If so, *tag* is set to 1; otherwise, *tag* remains 0 and Q will be replaced by its successor. Noticeably, the first embedded while-loop collects merely the redexes of the productions in Q whose contexts have been matched, rather than all the redexes of P. The second while-loop conducts backtracking when no redexes are found in the intermediate host graph H during the first while-loop. *traceStack* is employed to record the parsing trace.

IV. Algorithm Analysis

In this section, we first analyze the time complexity of Algorithm 3, then present a case study to illustrate how the algorithm improves parsing performance, and finally summarize its characteristics and indicate the situations in which it can perform efficiently.

A. *Time Complexity*

Theorem 1. The worst-case time complexity of Algorithm 3 is $O((m/r!)^h(h!\,h^h)^r)$, where h is the number of nodes in the host graph to be parsed, m is the number of productions of the graph grammar, n is the maximal number of components in the left or right graphs of the productions, and r is the maximum number of nodes in the right graphs of the productions.

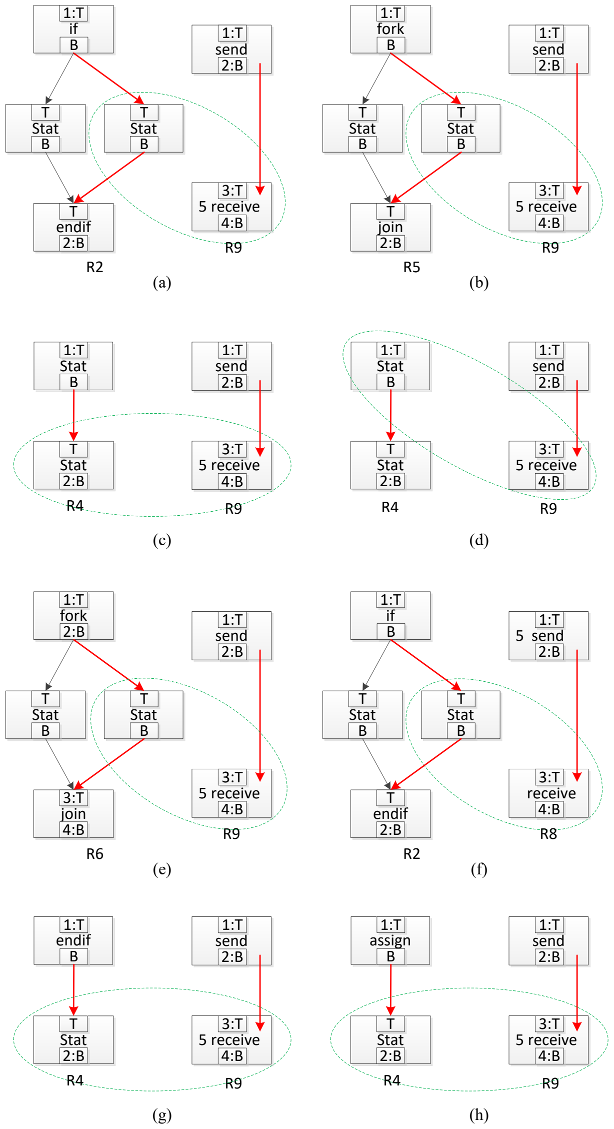


Figure 4. Some contexts of production $p9$.

Proof. An assumption is that the maximal degree of all the nodes in a host graph is a constant. In Algorithm 2, the creation of the partial precedence graph accounts for the time taken, that is $O(m^2(m+n^2r^2))$ [21], and so is the time complexity of the algorithm. But, from the perspective of Algorithm 3, Algorithm 2 takes $O(1)$, as m, n, and r are small constants for a graph grammar.

Similarly, the time complexity of a call for Algorithm 1 in Algorithm 3 is also $O(1)$, since the number of contexts for a production used for this purpose is a constant, and the time taken by the procedure FindNodeSequenceSet(H, q, *redex*, Z) is merely dependent on the degree of the nodes directly connected to the *redex* through Z, but has nothing to do with the cardinality of H.

The underlying structure of Algorithm 3 is similar to the algorithm described in [19]. In the worst case, calls for Algorithm 1 and Algorithm 2 may not essentially reduce the search space of productions and redexes. Hence, the time complexity of Algorithm 3 is also $O((m/r!)^h(h!\,h^h)^r)$. ■

B. *A Case Study*

we introduce a typical example to illustrate how the proposed algorithm reduces the search space during the parsing process.

Figure 4 lists some typical examples of the level-1contexts for the production $p9$. The left graph of $p9$ is composed of two components, one is the node "Stat", and the other is "receive". The former component appears 5 times in the right graphs of four productions, and the latter occurs 3 times in the right graphs of three productions. Therefore, the level-1 contexts of $p9$ totally is 15, i.e., the cardinality of the Cartesian product of the two constituents.

Figure 5 shows the deduction processes of a host graph parsed by the graph grammar depicted in Figure 1. The host graph is given in Figure 5(a). Graphs in (a)-(g) demonstrate a complete successful reduction, which can be generated by Algorithm 3. On the contrary, (h)-(j) present three intermediate host graphs that will lead to dead ends in next several deduction steps where backtracking have to be conducted.

In each graph of Figure 5, the subgraph surrounded by the dashed green rectangle (if exists) is a redex of some production's right graph, the bold red arcs connected to a redex are the contextual connection of the redex, and the subgraph surrounded by the dashed red rectangle (if exists) is the output of the R-application of a production to the redex in the anterior graph in a deduction.

Consider the case when a general parsing algorithm without the two strategies is adopted to parse the host graph in Figure 5(a).

Two redexes of $p3$ and one redex of $p9$, $p7$ and $p8$ can be found, which is shown in the dashed green rectangles. Ordinarily, five redexes will be pushed into *redexStack*, and then one will be popped out for deduction. We assume that the redex of $p8$ is popped out first. Note that this enclosed subgraph is also the redex of both $p9$ and $p7$ simultaneously. After an R-application, one obtains an intermediate host graph, shown in (h), where the subgraph enclosed in the dashed red rectangle is the output of the R-application of $p7$. In (h), only two redexes of $p3$ exist, and twice R-applications of $p3$ can be done accordingly, leading to a host graph that cannot be deduced any longer. As a matter of fact, other choices (except $p9$) lead to situations similar to (h), e.g., the graph (i). In these situations, backtracking is needed after a few reduction steps.

In contrast, Algorithm 1 tackles the host graph in Figure 5(a) in a quite different way.

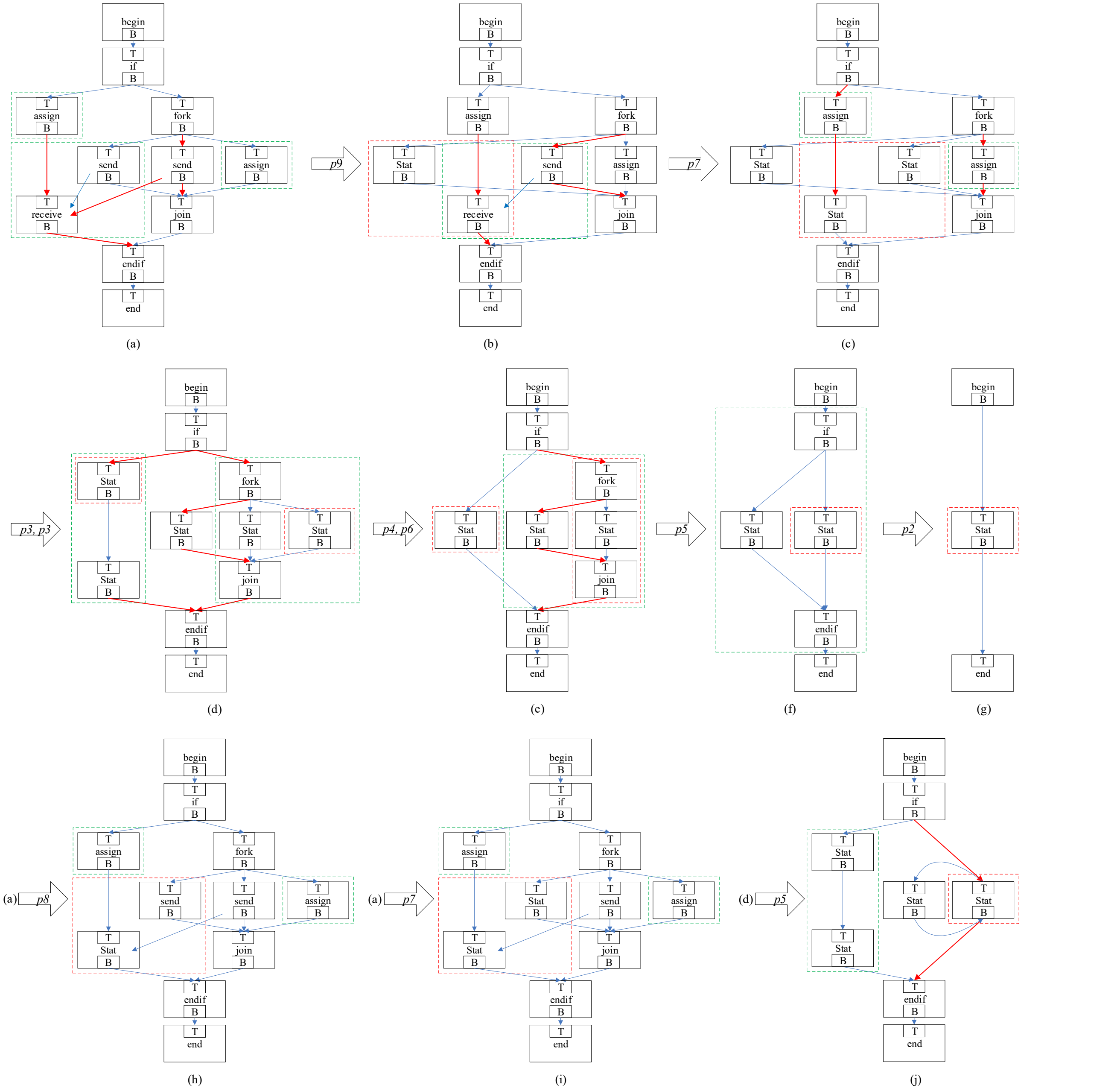


Figure 5. The parsing processes of a host graph. (a)-(g) A successful parsing path produced by Algorithm 3. (h)-(j) three intermediate host graphs where backtracking is necessary.

First, the set P of productions are partitioned into 4 subset: $prt[0] = \{p8, p9\}$, $prt[1] = \{p7\}$, $prt[2] = \{p3\}$, $prt[3] = \{p2, p4, p5, p6\}$, and these subsets are sequentially ordered, in terms of the partial precedence graph of P. The redexes of productions from them will be handled in priority according to the order.

Then, in the first iteration of the outmost while-loop, only the redexes of productions in $prt[0]$ are considered in this graph, that is, the redex of $p8$ or $p9$. Moreover, context matching can be utilized to further verify the redex found. Obviously, a context is matched for $p9$, whereas no context is matched for $p8$. A distinct character of the contexts for $p9$ is that a node labeled "send" is the source node of an edge directed to the node "receive" in the redex. After the application of $p9$ to the redex in (a), graph (b) is obtained.

Next, it enters the second iteration and repeats the above process. It searches graph (b) for redexes of the productions from $prt[0]$ and finds the subgraph enclosed in the green rectangle, which is a redex of both $p8$ and $p9$. However, no contexts can be matched for both productions. Therefore, it tries the second subset $prt[1]$ and obtains a redex of $p7$. Similarly, context matching is conducted to the redex enclosed in the green rectangle. Then the redex is verified and a corresponding deduction step outputs (c).

In the third iteration, as there is no more redexes of productions from both $prt[0]$ and $prt[1]$ in (c), the only production in $prt[2]$ is considered, which leads to (d).

The subgraph enclosed in the green dashed rectangle in (d) is a redex of both $p5$ and $p6$ from $prt[3]$. Once again, context matching is activated. A distinct character of the contexts for $p6$ rather than for $p5$ is that a node labeled “Stat” is simultaneously the target node of an edge originated from the node “fork” and the source node of an edge directed to the node “join” in the redex. Thus, $p5$ is excluded and $p6$ is selected for R-application, leading to (f). Finally, (g) follows. A valid deduction path is achieved even without backtracking. In the parsing process, only level-1 contexts are employed.

Otherwise, if $p5$ is selected for deduction in the intermediate host graph (d), it then produces (j). An R-application of $p4$ can be conducted to (j), and the resulting graph is irreducible. In that situation, backtracking is inevitable.

C. Discussion

As is known, backtracking is the main cause for high time complexity of a general parsing algorithm. An ideal way is to avoid as many as possible unnecessary productions and redexes in every reduction step. The introduction of the two strategies is exactly for this purpose.

In each reduction step, not the whole set of productions of the graph grammar will be chosen for searching for redexes in a host graph. Instead, only a subset of it is selected by using the strategy of production set partitioning. For example, in the iterations of the first inner while-loop, only 1 or 2 out of the total 9 productions are picked for searching in most cases. Thus, this strategy can exclude a majority of unnecessary productions in each reduction step. The effectiveness of production set partitioning originates from locality principle for programs, as a graph grammar is essentially a program. Obviously, the more subsets a production set can be partitioned into, the more productions will be excluded by the strategy in reduction steps.

After being found, a redex of a production will be verified by context matching to check whether it is a true positive one. This is a critical means to reduce backtracking. For instance, in Figure 5(a), there is a subgraph that is a redex of $p9$, $p7$ and $p8$; by means of context matching, the redex of $p9$ is verified whereas other two redexes are denied; otherwise, the execution of R-applications of $p8$ and $p7$ to the host graph will produce the graphs in (h) and (i), respectively, any of which will arrive at situations of backtracking in a few more reduction steps. A similar situation arises in (d) where the subgraph enclosed by the right dashed green rectangle is a redex of both $p6$ and $p5$, and the strategy excludes the latter and avoids the situation depicted in (j) accordingly. In this example, context matching removes all the unnecessary redexes so as to eliminate backtracking. Hence, the strategy can exclude a certain number of unnecessary redexes in each reduction step. Clearly, when the right graphs of two productions share a common subgraph, their contexts are supposed to be distinct from each other, which is a premise to ensure the effectiveness of the strategy.

In summary, we believe that the introduced strategies can considerably improve the parsing performance of a general parsing algorithm for context-sensitive graph grammars.

V. Conclusion and Future Work

Based on the RGG formalism, this paper proposes a general parsing algorithm for context-sensitive graph grammars. The algorithm is embedded with two essential strategies: one is context matching, and the other is production set partitioning. The former is utilized to reexamine the found redexes so as to exclude the unnecessary redexes, and the latter can be used to precisely choose the relevant productions to narrow down the searching space for redexes. Therefore, the two strategies can considerably improve the parsing performance, even though the worst-case time complexity is not theoretically reduced. Furthermore, a case study along with detailed analysis is provided to illustrate the paring process and parsing performance of the proposed algorithm.

Nevertheless, to what extent can the strategy of context matching reduce unnecessary redexes, and how is it related to the characteristics of contexts still need to be investigated. In our future work, feature analysis and integration for contexts will be conducted to facilitate the study on these issues.

Acknowledgment

This work is supported in part by the National Science Foundation of China under grant 61170089.

References

[1] F.Ferrucci, G. Pacini, G. Satta, et al., “Symbol-relation grammars: a formalism for graphical languages,” Information and Computation, 131(1), pp. 1–46, 1996.

[2] K. Marriott, “Constraint multiset grammars,” IEEE Symposium on Visual Languages, St. Louis, Missouri, pp. 118–125, 1994.

[3] G. Costagliola, V. Deufemia, and G. Polese, “Visual language implementation through standard compiler-compiler techniques,” Journal of Visual Languages and Computing, 18(2), pp. 165–226, 2007.

[4] G. Rozenberg (Ed.), Handbook on Graph Grammars and Computing by Graph Transformation, vol. 1: Foundations, World Scientific, 1997.

[5] J. L. Pfaltz, A. Rosefeld, “Web grammars,” International Joint Conference on Artificial Intelligence, pp. 609–619, 1969.

[6] G. Engels, H. J. Kreowski, and G. Rozenberg (Eds.), Handbook of Graph Grammars and Computing by Graph Transformation, vol. 2: Applications, Languages, and Tools, World Scientific, 1999.

[7] H. Ehrig, H. J. Kreowski, U. Montanari, and G. Rozenberg (Eds.), Handbook of Graph Grammars and Computing by Graph Transformation, vol. 3: Concurrency, Parallelism, and Distribution, World Scientific, 1999.

[8] J. Rekers and A. Schürr, “Defining and parsing visual languages with layered graph grammars,” Journal of Visual Languages and Computing, 8(1), pp. 27–55, 1997.

[9] D. Zhang, K. Zhang, and J. Cao, “A context-sensitive graph grammar formalism for the specification of visual languages,” The Computer Journal, 44(3), pp. 187–200, 2001.

[10] M. Nagl, “Set theoretic approaches to graph grammars,” International Workshop on Graph Grammars and Their Application to Computer Science, Lecture Notes in Computer Science, vol. 291, Springer Verlag, pp. 41–54, 1987.

[11] C. Zhao, J. Kong, and K. Zhang, Program behavior discovery and verification: a graph grammar approach, IEEE Transactions on Software Engineering, 36(3), pp. 431–448, 2010.

[12] Y. Liu, X. Zeng, Y. Zou, and K. Zhang, "A graph grammar-based approach for graph layout," Software: Practice and Experience, 49(8), pp. 1523–1535, 2018.

[13] Z. Shi, X. Zeng, Y. Zou, et al., A temporal graph grammar formalism. Journal of Visual Languages and Computing, vol. 47, pp. 62–76, 2018.

[14] P. Bottoni, G. Taentzer, and A. Schürr, "Efficient parsing of visual languages based on critical pair analysis and contextual layered graph transformation," IEEE Symposium on Visual Languages, pp. 59–60, 2000.

[15] J. Kong, K. Zhang, and X. Zeng, "Spatial graph grammars for graphical user interfaces," ACM Transactions on Computer-Human Interaction, 13(2), pp. 268–307, 2006.

[16] Y. Zou, X. Zeng, and X. Han, "Context-attributed graph grammar framework for specifying visual languages," Journal of Southeast University (English Edition), 24(4), pp. 455–461, 2008.

[17] Y. Liu, Z. Shi, and Y. Wang, "An edge-based graph grammar formalism and its support system," International DMS Conference on Visualization and Visual Languages, pp.101–108, 2018.

[18] Y. Zou, J. Lü, and X. Tao, "Research on context of implicit context-sensitive graph grammars," Journal of Computer Languages, vol. 51, pp. 241–260, 2019. https://doi.org/10.1016/j.cola.2019.01.002

[19] X. Zeng, K. Zhang, J. Kong, and G. Song, RGG+: An enhancement to the reserved graph grammar formalism. In Proc. IEEE Symposium on Visual Languages and Human-Centric Computing, pp. 272–274, 2005.

[20] Y. Zou, X. Zeng, Y. Liu, and H. Liu, "Context computation for context-sensitive graph grammars: Algorithms and Complexities," Journal of Visual Language and Computing, vol. 1, pp. 15–28, 2019.

[21] Y. Zou, X. Zeng, Y. Liu, and H. Liu, "Partial precedence of context-sensitive graph grammars," International Symposium on Visual Information Communication and Interaction, pp. 16–23, 2017.

A Chatbot for supporting users in Cultural Heritage contexts

Fabio Clarizia
DIIn
University of Salerno
Fisciano (Salerno), Italy
fcolarizia@unisa.it

Francesco Colace
DIIn
University of Salerno
Fisciano (Salerno), Italy
fcolace@unisa.it

Marco Lombardi
DIIn
University of Salerno
Fisciano (Salerno), Italy
malombardi@unisa.it

Domenico Santaniello
DIIn
University of Salerno
Fisciano (Salerno), Italy
dsantaniello@unisa.it

Abstract— **The distance education of today is performed through new technologies, which assist whit the aim to improving the educational process. New technologies brought e-learning environments more and more complete and able to supply learning experiences increasingly similar to traditional educational processes. The introduction of chatbots could represent an important aspect of the students' educational process. Chatbots are particular applications able to simulate typical human being conversations; through inference engines and modern Natural Language Processing techniques, a chatbot could be able to provide tailored educational paths. In this paper is presented a chatbot prototype in order to support students in the study of Cultural Heritage. This system takes advantage of inference techniques and approaches based on Natural Language Processing in order to provide a fluid conversation responding to the student's needs. The real experimentation was performed in the application context of the archaeological park of Paestum, where students were able to experience contextualized and personalized learning paths based on the recommendation capacity of the proposed system. In terms of student satisfaction, our proposed system shows very satisfactory results.**

Keywords—Chatbot, Technology-Mediated Learning, Cultural Heritage

I. Introduction

The world of distance education is enriching with new tools able to improve the training process. Modern e-Learning environments do not simply aim to reproduce the typical processes of the world of traditional education but seek to enrich them thanks to the support of new technologies. In this scenario, the introduction of the chatbots could be an effective enhancement for the distance learning process [1]. Chatbots, in fact, are becoming one of the most interesting approach in many fields: their design has become increasingly sophisticated and their use adopted commerce, entertainment, public sector, social networks and education [2] [4]. In fact, the introduction of chatbots comes with four main advantages. Firstly, the implementation of chatbots saves customer service costs by replacing nearly all human assistants. Secondly, chatbots increase user satisfaction by speeding up response times and being available twenty-four hours a day. Thirdly, chatbots can be proactive: they can anticipate user questions and needs. Fourthly, chatbots are also a powerful analysis tools, since conversations between users and chatbots can be analysed to better understand customer requirements [5] [6].

In distance learning field, chatbots promise real enhancements in the learning process and a real improvement in the field of Intelligent Tutoring Systems (ITS) [7] [8] [9].

The ITS are environments that incorporate artificial intelligence and support the students' learning process. These tools, generally, are limited to specific knowledge domains. Unlike ITS, chatbots focus on conversation and are able to analyse the student's learning context and needs. On the other hand, teaching is a relational act based on communication and interaction and chatbots have significant educational potential thanks to their communicative ability through natural language. This kind of approach is effective especially in large-scale learning scenario where the problem of individual student support is important. Another impressive feature of the chatbots is in their efficiency: they can operate every time supporting students everywhere and allow teachers to avoid answering repetitive questions [10].

Despite all these advantages, the diffusion of chatbots in the education field is still low: a small number of researches have shown experiences in learning scenarios [11] [12]. A lot of them have been developed on IBM's Watson Platform [13] while other ones are created with specific tools. Chatbots in education can achieve two main aims [1]:

– To support students in tackling different topics and reflect amongst themselves on the basis of starting questions posed by the chatbot

– To infer information about the students' perceptions of a specific topic, interaction, situation or context.

Depending on the tasks carried out by chatbots in the education field, a taxonomy can be introduced [14]:

- Administrative and management tasks to foster personal productivity
- FAQs Management
- Student Mentoring
- Motivation
- Practice of specific skills and abilities
- Simulations
- Reflection and Metacognitive Strategies
- Student Learning Assessment

In this paper a chatbot for Cultural Heritage Learning, is introduced. The aim of this tool is supporting students in the learning of topics related to the Cultural Heritage field according an approach based on practice of specific skills and abilities, simulations and assessment [15][16][17]. In particular, the proposed approach adopts Natural Language Process methodologies and Context Aware Techniques for

DOI reference number: 10.18293/DMSVIVA20-009

inferring services and contents suitable for the students' needs. Thanks to probabilistic approach as the Latent Dirichlet Allocation [18] the proposed prototype can infer from the conversation The chatbot works at its best in operative scenarios as museums or archaeological parks: in these contexts, the proposed chatbot, starting from cultural artefact, can furnish adaptive contents and enrich the learning experience of the students [19]. The paper has the following organization: in the next section more details about chatbots will be furnished. The third section will show the methodology and the architecture of the chatbot. The description of the experimental results and the conclusions will end the paper.

II. Chatbots: Main Concepts

Chatbots are applications that can interact with people using language-based interfaces [20] [21]. Their purpose is to simulate a human conversation so that the speaker has an experience similar to a conversation via text or voice interactions with a real person [22] [23]. This definition includes all kinds of software enabling humans to have a conversation with a computer. In this way, virtual assistants as Amazon's Alexa, Google's Home or Apple's Siri can be considered chatbots. In literature, the first chatbot was Eliza: it analysed input sentences and created its response based on reassembly rules associated with a decomposition of the input [24]. Other interesting experiences related to the chatbots' development was Converse, Jabberwacky and Alice [25]. In particular, Alice adopts the AIML (Artificial Intelligence Mark-Up Language) allowing the introduction of Data Mining techniques. In the last period the introduction of more sophisticated Data Mining methodologies, based on Artificial Intelligence, of probabilistic methods and semantic tools has improved the skills of chatbots and their capacity to make decisions.

Generally, a chatbot has these main components:

- A Conversational Artificial Intelligence Module: this module is the Natural Language Processing engine that allows the building of the conversation. The first chatbots focused on the interpretation and recognition of patterns and rules. The modern chatbots adopt techniques based on Artificial Intelligence or deep learning approach for generating suitable responses.
- User Experience Module: this module allows a natural and coherent conversation among users and chatbot
- User Interface: this module defines the interface that user adopts for seeing or hearing the conversation with the chatbot
- Conversational design: this module allows the addition of the human logic to the artificial interaction human-machine.

The huge increase of chatbots in the last years defines three main dimensions to classify them. The first one can be defined as the "building approaches" that distinguishes between retrieval-based models and generative models. The chatbots based on retrieval-based models pick the appropriate response from a repository of predefined responses. Usually they adopt rule-based approach. Generative models generate responses out of the input with the help of machine learning techniques. This let the chatbot feel more human and capable of longer dialogs [26]. The second dimension is related to the input mode of chatbot. Users can interact with the chatbot in different ways depending on the communication interface. Three main approaches can be considered:

- Chatterboxes: users can interact with the chatbot through text or voice inputs and outputs.
- Embodied conversational agents: the interface is represented by an avatar which interacts with the user. This kind of approach involves audio, text, multimedia.
- Physical: in this case user interact with a chatbot that is embedded in a robot

The third dimension addresses the inclusion of contextual information such as linguistic and physical context in order to select the right responses.

The proposed Chatbot adopts a generative contextual model and adopts as the input mode text and voice. In the next section, more information about the chatbot architecture will be furnished.

III. Architecture and Approach

Before introducing the Architecture of the chatbot an example of its operation workflow. Imagine a student, which have to deepen the Greek culture in southern Italy and visit the Archaeological Park of Paestum. According to the context, the chatbot begins to provide generic content on the presence of the Greeks in southern Italy. Arriving near the temples it will begin to provide more details on the cults of the gods in Paestum. Depending on the student's requests, the chatbot will provide new contents. If for example the student will show interest in the cult of Athena, the chatbot will invite him to go to the museum in the park to view the weapons present at the time in Paestum. If the student was interested in the cult of the dead, he would be invited to see the famous Tomb of the Diver. Obviously, at regular intervals the chatbot, with some tests, will verify the student's learning. Furthermore, the chatbot is able to retrieve information [27], thanks to REST services [28], from environments such as Europeana [29] (https://pro.europeana.eu/) or DatabencArt (https://www.databencart.it/) that contain information relating to the world of cultural heritage.

The architecture of the proposed chatbot is based on some main modules, as shown in figure 1. The storytelling module is closely related to the bot ability to guide the user through the whole learning experience, making the way of proceeding while leaving the user free to express himself and immerse himself in the personalized story of a place [30].

Each user acts differently from the others, becoming part of a creative process and creating a unique and unrepeatable visit. In planning an itinerary and in designing the narration, therefore, not all progress must be defined but a plurality of scenarios must be prepared that the user can explore freely up to crucial moments, common to all, or almost all, of the scenarios. The Context-Aware Manager deals with representing all the possible contexts of use through the Context Dimension Tree and performing contextualized queries.

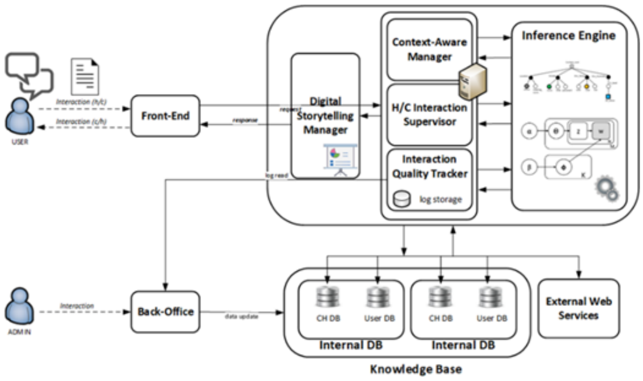


Fig. 1. The Proposed Architecture

In this way, it is possible to extract and provide personalized information by aggregating and custom-tailoring data and services extracted from different sources. Some resources are private (internal resources) or managed directly by the chatbot provider. These resources can be used, for example, to maintain the profiles of registered users, or the data of the museums reviewed. The chatbot can also interface with external services (external resources) [31].

Other modules, such as the Human/Computer Interaction Supervisor and the Interaction Quality Tracker, have the following objectives:

- supervise the dialogue, keeping track of timing of interaction, identifying ambiguous questions, or dialogue sessions that are not convergent or too long, and so on;
- carry out monitoring interactions between the user and the chatbot, producing synthetic quality indicators and highlighting critical aspects useful for the improvement of the system.

The core of the architecture is the Inference Engine, which includes text analysis and context extraction. It is assumed that the text generated by a chat is a mixture of contextual information and that the use of some words helps to define the different context elements useful in the search for the same context that can be identified through the Context Dimension Tree (CDT) [32].

Latent Dirichlet Allocation is a model suitable for this purpose as it can be used to explain the correlation between keywords and topics (in our case, context elements), as shown in the following figure 2. Through textual analysis it is possible to know more about the student, where he is or where he would like to be (museum, archaeological park, etc.), the learning purpose of his visit and what they need.

In practice, the interaction of the user with the chatbot is divided into shorter and simpler sentences (clusters), through appropriate Bayesian filters for keywords, assuming that each sentence is semantically related to the other. The proposed approach is therefore based on the following assumption: the probability that the word W belongs to the concept node N_c within the CDT is proportional to how much the argument (for example, the purpose of a tourist's visit to a city of art) has already been treated and the number of times that a word has been used for a specific topic. The application of this model provides a characterization of chats in an automatic way, without the needing of specify the semantic value of the words in the text.

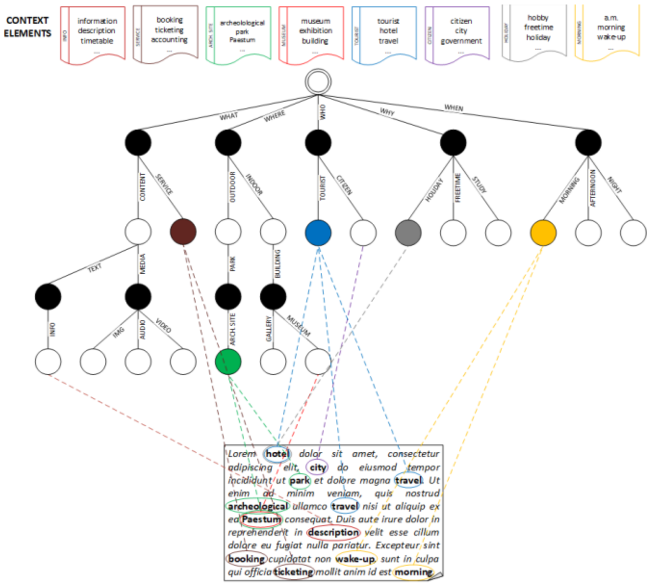

CONTEXT ELEMENTS
INFO
information
description
timetable
...
SERVICE
booking
ticketing
accounting
...
ARCH. SITE
archeolological
park
Paestum
...
MUSEUM
museum
exhibition
building
...
TOURIST
tourist
hotel
travel
...
CITIZEN
citizen
city
government
...
HOLIDAY
hobby
freetime
holiday
...
MORNING
a.m.
morning
wake-up
...
WHAT
WHERE
WHO
WHY
WHEN
CONTENT
SERVICE
OUTDOOR
INDOOR
TOURIST
CITIZEN
HOLIDAY
FREETIME
STUDY
MORNING
AFTERNOON
NIGHT
TEXT
MEDIA
PARK
BUILDING
INFO
IMG
AUDIO
VIDEO
ARCH. SITE
GALLERY
MUSEUM
Lorem hotel dolor sit amet, consectetur adipiscing elit, city do eiusmod tempor incididunt ut park et dolore magna travel. Ut enim ad minim veniam, quis nostrud archeological ullamco travel nisi ut aliquip ex ea Paestum consequat. Duis aute irure dolor in reprehenderit in description velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint booking cupidatat non wake-up, sunt in culpa qui officia ticketing mollit anim id est morning.

Fig. 2. Context Definition

Furthermore, using the LDA approach on a set of chats that belong to the same domain (in the case analysed, tourism), it is possible to automatically extract a Mixed Graph of Terms (mGT) that can be used both for the design of the tree of context and the constraints associated with it is to detect the context extracted in real time from the user's chat with the bot.

In particular, LDA was mainly used to generate topics within chats (text documents). These topics were processed by the system as contextual elements suitable during the use of the Context Dimension Tree. According to the LDA model, a distribution of terms for each topic i is represented as a multinomial distribution φ_i drawn from a symmetric Dirichlet distribution with parameter β:

$$p(\Phi_i|\beta) = \frac{\Gamma(W\beta)}{[\Gamma\beta]^W}\prod_{v=1}^{W}\phi_{iv}^{\beta-1}$$

The topic distribution for a document d is also represented as a multinomial distribution Θ_d drawn by a Dirichlet distribution with parameter α:

$$p(\theta_d|\alpha) = \frac{\Gamma(\sum_{i=1}^{K}\alpha_i)}{\prod_{i=1}^{K}\Gamma(\alpha_i)}\prod_{i=1}^{K}\phi_{di}^{\alpha_i-1}$$

In this way, the topic z_{dn} for each index n token can be chosen from the distribution of the document topics as:

$$p(z_{dn} = i|\theta_d) = \theta_{di}$$

Each token w is chosen from a multinomial distribution associated with the selected topic:

$$p(z_{dn} = i|\theta_d) = \theta_{di}$$

LDA aims to find patterns of co-occurrence terms in order to identify consistent topics. Through LDA it is possible to learn a topic i, if $p(w = v \mid z = i)$ is high for a certain term v, then every document d that contains the term v has a high probability for the topic i.

It is possible to state that all the terms co-occurring with the term v are more likely to have been generated by the topic i.

A complex structure like the mGT can allow to capture and represent the contextual information contained in a set of chats that belong to a specific domain (for example, tourism). This graph can be extracted automatically and used for the classification of the text, or to label the N_c concept nodes and know which of the nodes participate in the definition of the context at a given time. Formally, it can be defined as a graph $g \ = < N, E >$ where:

- $N \ = \{R, W\}$ is a finite set of nodes, those elements can be aggregates or aggregators.
- $E \ = \{E_{RR}, E_{RW}\}$ is a set of arcs that connect the aggregates and aggregators.

The proposed approach is essentially composed of two basic modules located within the Inference Engine: a module for the construction of the Mixed Graph of Terms and a module for extracting the context elements.

Mixed Graph of Terms building module: this module builds the mGT starting from a set of documents that belong to a specific domain (tourism) and that have been previously labeled in accordance with the contextual information contained. The mGT can also be used in the design phase of the Context Dimension Tree.

Context Mining Module: this module involves the extraction of the context, or rather of the different context elements, thanks to the use of mGT as a context filter. The input of this module consists of a generic chat, the mGT extracted and the CDT in relation to the specific domain. The output is the context related to the chat.

Each context element is associated with a dedicated section of the database, which contains relevant and specific data. The contextual query is performed automatically by defining a global view given by the composition of the associated partial views. In addition to simple data, the same mechanism can be used for the selection of useful services related to the context identified, as shown in figure 3.

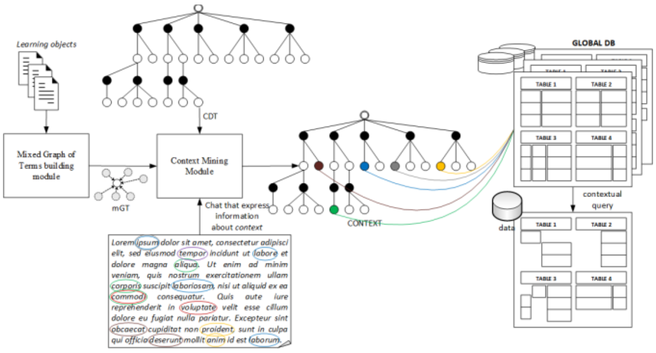
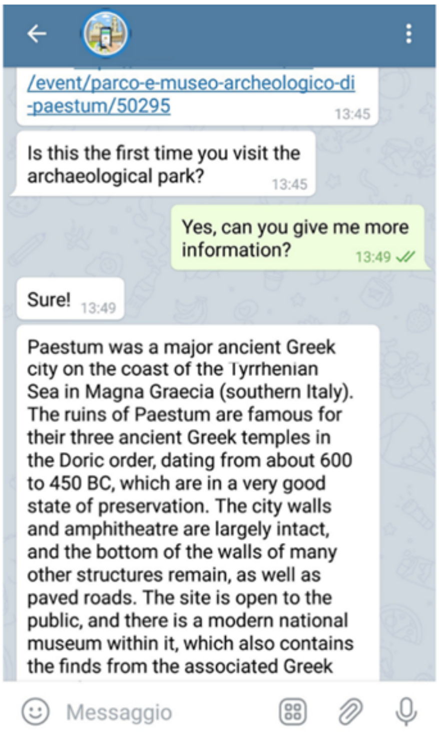

Learning objects
Mixed Graph of Terms building module
mGT
CDT
Context Mining Module
Chat that express information about context
Lorem ipsum dolor sit amet, consectetur adipisci elit, sed eiusmod tempor incidunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur. Quis aute iure reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint obcaecat cupiditat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.
CONTEXT
GLOBAL DB
TABLE 1
TABLE 2
TABLE 3
TABLE 4
contextual query
data

Fig. 3. Contextual queries

IV. Experimental Results

Based on the proposed architecture, an application prototype, implemented in Java, was developed: a chatbot, designed and implemented, along with a server-side component, as described above (figure 4). The chatbot was initially designed to support students visiting a the cultural site of the Campania region of Italy (the Archaeological parks of Paestum). In this experimental phase, the main services and contents potentially useful for students have been identified.


/event/parco-e-museo-archeologico-di-paestum/50295 13:45

Is this the first time you visit the archaeological park? 13:45

Yes, can you give me more information? 13:49

Sure! 13:49

Paestum was a major ancient Greek city on the coast of the Tyrrhenian Sea in Magna Graecia (southern Italy). The ruins of Paestum are famous for their three ancient Greek temples in the Doric order, dating from about 600 to 450 BC, which are in a very good state of preservation. The city walls and amphitheatre are largely intact, and the bottom of the walls of many other structures remain, as well as paved roads. The site is open to the public, and there is a modern national museum within it, which also contains the finds from the associated Greek

Fig. 4. The chatbot prototype at work

After the interaction with the chatbot, 73 students of the last year of the High School responded, according to the Likert scale, to a questionnaire comprising five sections. To each statement, present in a specific section, five possible answers were associated: "I totally disagree" - TD, "I disagree" - D, "Undecided" - U, "I agree" - A, "I totally agree" - TA.

Section A: recommendation

1. The proposed services and contents have satisfied the needs of the user, based on personal preferences and the current context.
2. The system has furnished effective contents for learning.

Section B: conversation

1. The dialogue with the chatbot took place smoothly and without unexpected jumps.
2. The system was able to correctly understand the intentions of the user.

Section C: presentation

1. The information has been presented appropriately.
2. The information provided was exhaustive.

Section D: usability

1. The chatbot interface is user-friendly.
2. Response times are adequate.

Section E: future developments

1. It would be useful to include in the chat other students
2. It would be interesting to apply the same approach in other scenarios.

According to Table 1, the positive responses (Agree and Totally Agree) reach a percentage always higher than 84%.

TABLE I. Analysis of Results

	Percentage		
Section	**Negative**	**Neutral**	**Positive**
A	7,52%	4,87%	87,61%
B	8,01%	6,11%	85,88%

C	8,93%	3,72%	87,35%
D	11,12%	3,95%	84,93%
E	7,21%	3,05%	89,74%

The lowest number of positive feedbacks is found in the conversation section: some users have found it difficult to converse fluently with the chatbot, encountering some problems related to understanding the requests made.

This can be largely due to grammatical errors or dialect words not yet learned by the system in this initial state. On the other hand, section A, relating to the context-based recommendation form, shows very satisfactory results: the services and content proposed have satisfied the needs of the user, dynamically adapting to the current context. The results therefore suggest. In particular, attention to machine learning in order to improve human-machine interaction.

V. Conclusions

In this paper, a chatbot for supporting students in the Cultural Heritage learning has been introduced. This chatbot adopts a natural language processing methodology based on Latent Dirichlet Allocation and Context Awareness. The proposed approach has been applied in a real scenario. The obtained results are promising. Future works aims to apply the proposed approach in other scenarios and to develop more sophisticated approach for making the conversation more "human oriented".

References

[1] R. Winkler and M. Soellner, "Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis," Acad. Manag. Proc., 2018 DOI:10.5465/ambpp.2018.15903abstract.

[2] N. Albayrak, A. Ozdemir, and E. Zeydan, "An overview of artificial intelligence based chatbots and an example chatbot application," in 26th IEEE Signal Processing and Communications Applications Conference, SIU 2018, 2018 DOI:10.1109/SIU.2018.8404430.

[3] F. Clarizia, F. Colace, M. Lombardi, F. Pascale, and D. Santaniello, "Chatbot: An education support system for student," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, vol. 11161 LNCS, pp. 291–302 DOI:10.1007/978-3-030-01689-0_23.

[4] A. Xu, Z. Liu, Y. Guo, V. Sinha, and R. Akkiraju, "A New Chatbot for Customer Service on Social Media," in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17, 2017, pp. 3506–3510 DOI:10.1145/3025453.3025496.

[5] M. Casillo, F. Clarizia, F. Colace, M. Lombardi, F. Pascale, and D. Santaniello, "An Approach for Recommending Contextualized Services in e-Tourism," Information, vol. 10, no. 5, p. 180, May 2019 DOI:10.3390/info10050180.

[6] M. Casillo, F. Clarizia, G. D'Aniello, M. De Santo, M. Lombardi, and D. Santaniello, "CHAT-Bot: a Cultural Heritage Aware Teller-Bot for supporting touristic experiences," Pattern Recognit. Lett., Jan. 2020 DOI:10.1016/j.patrec.2020.01.003.

[7] L. Benotti, M. C. Martínez, and F. Schapachnik, "A Tool for Introducing Computer Science with Automatic Formative Assessment," IEEE Trans. Learn. Technol., 2018 DOI:10.1109/TLT.2017.2682084.

[8] B. Liu, Z. Xu, C. Sun, B. Wang, X. Wang, D. F. Wong, and M. Zhang, "Content-Oriented User Modeling for Personalized Response Ranking in Chatbots," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 26, no. 1, pp. 122–133, Jan. 2018 DOI:10.1109/TASLP.2017.2763243.

[9] L. N. Paschoal, M. M. de Oliveira, and P. M. M. Chicon, "A Chatterbot Sensitive to Student's Context to Help on Software Engineering Education," in 2018 XLIV Latin American Computer Conference (CLEI), 2018, pp. 839–848 DOI:10.1109/CLEI.2018.00105.

[10] G. Molnar and Z. Szuts, "The Role of Chatbots in Formal Education," in 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), 2018, pp. 000197–000202 DOI:10.1109/SISY.2018.8524609.

[11] K. Stoeffler, Y. Rosen, M. Bolsinova, and A. A. von Davier, "Gamified assessment of collaborative skills with chatbots," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018 DOI:10.1007/978-3-319-93846-2_64.

[12] A. Kerly, P. Hall, and S. Bull, "Bringing chatbots into education: Towards natural language negotiation of open learner models," Knowledge-Based Syst., vol. 20, no. 2, pp. 177–185, Mar. 2007 DOI:10.1016/j.knosys.2006.11.014.

[13] A. Goel, B. Creeden, M. Kumble, S. Salunke, A. Shetty, and B. Wiltgen, "Using Watson for enhancing human-computer co-creativity," in AAAI Fall Symposium - Technical Report, 2015.

[14] G. Garcia Brustenga, M. Fuertes Alpiste, and N. Molas Castells, "Briefing Paper: Chatbots in Education," Barcelona, Sep. 2018 DOI:10.7238/elc.chatbots.2018.

[15] F. Amato, V. Moscato, A. Picariello, and F. Piccialli, "SOS: A multimedia recommender System for Online Social networks," Futur. Gener. Comput. Syst., vol. 93, pp. 914–923, Apr. 2019 DOI:10.1016/j.future.2017.04.028.

[16] F. Amato, A. Mazzeo, V. Moscato, and A. Picariello, "Semantic Management of Multimedia Documents for E-Government Activity," in 2009 International Conference on Complex, Intelligent and Software Intensive Systems, 2009, pp. 1193–1198 DOI:10.1109/CISIS.2009.195.

[17] G. D'aniello, M. Gaeta, and M. Z. Reformat, "Collective perception in smart tourism destinations with rough sets," in 2017 3rd IEEE International Conference on Cybernetics, CYBCONF 2017 - Proceedings, 2017, pp. 1–6 DOI:10.1109/CYBConf.2017.7985765.

[18] F. Clarizia, F. Colace, M. Lombardi, F. Pascale, and D. Santaniello, "Chatbot: An Education Support System for Student," Springer, Cham, 2018, pp. 291–302 DOI:10.1007/978-3-030-01689-0_23.

[19] F. Clarizia, S. Lemma, M. Lombardi, and F. Pascale, "A Mobile Context-Aware Information System to Support Tourism Events," 2017, pp. 553–566 DOI:10.1007/978-3-319-57186-7_40.

[20] M. Lombardi, F. Pascale, and D. Santaniello, "An application for Cultural Heritage using a Chatbot," in 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), 2019, pp. 1–5 DOI:10.1109/CAIS.2019.8769525.

[21] S. A. and D. John, "Survey on Chatbot Design Techniques in Speech Conversation Systems," Int. J. Adv. Comput. Sci. Appl., vol. 6, no. 7, 2015 DOI:10.14569/IJACSA.2015.060712.

[22] M. L. Mauldin, "ChatterBots, tinyMuds, and the turing test entering the loebner prize competition," in Proceedings of the National Conference on Artificial Intelligence, 1994.

[23] F. Amato, S. Marrone, V. Moscato, G. Piantadosi, A. Picariello, and C. Sansone, “Chatbots meet ehealth: Automatizing healthcare,” in CEUR Workshop Proceedings, 2017.

[24] J. Weizenbaum, “ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine,” Commun. ACM, 1983 DOI:10.1145/357980.357991.

[25] V. di Lecce, M. Calabrese, D. Soldo, and A. Giove, “Semantic management systems for the material support of e-learning,” J. E-Learning Knowl. Soc., 2010.

[26] C.-W. Liu, R. Lowe, I. Serban, M. Noseworthy, L. Charlin, and J. Pineau, “How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation,” in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 2122–2132 DOI:10.18653/v1/D16-1230.

[27] P. Bottoni and M. Ceriani, “SWOWS and dynamic queries to build browsing applications: On linked data,” J. Vis. Lang. Comput., 2014 DOI:10.1016/j.jvlc.2014.10.027.

[28] P. Bellini, P. Nesi, and A. Venturi, “Linked open graph: Browsing multiple SPARQL entry points to build your own LOD views,” J. Vis. Lang. Comput., vol. 25, no. 6, pp. 703–716, Dec. 2014 DOI:10.1016/j.jvlc.2014.10.003.

[29] A. Isaac, R. Clayphan, and B. Haslhofer, “Europeana: Moving to Linked Open Data,” Inf. Stand. Q., vol. 24, no. 2/3, p. 34, 2012 DOI:10.3789/isqv24n2-3.2012.06.

[30] F. Colace, M. De Santo, M. Lombardi, and D. Santaniello, “CHARS: a Cultural Heritage Adaptive Recommender System,” in Proceedings of the 1st ACM International Workshop on Technology Enablers and Innovative Applications for Smart Cities and Communities - TESCA’19, 2019, pp. 58–61 DOI:10.1145/3364544.3364830.

[31] F. Clarizia, F. Colace, M. De Santo, M. Lombardi, F. Pascale, and D. Santaniello, “A Context-Aware Chatbot for Tourist Destinations,” in 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2019, pp. 348–354 DOI:10.1109/SITIS.2019.00063.

[32] 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, 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.

Supporting Cyber Attack Detection via Non-Linear Analytic Prediction of IP Addresses: A Big Data Analytics Technique

Alfredo Cuzzocrea*

iDEA Lab, University of Calabria, Rende, Italy
and LORIA, Nancy, France
alfredo.cuzzocrea@unical.it

Enzo Mumolo

University of Trieste, Trieste, Italy
mumolo@units.it

Edoardo Fadda

Politecnico di Torino and ISIRES, Torino, Italy
edoardo.fadda@polito.it

Marco Tessarotto

University of Trieste, Trieste, Italy
marco.tessarotto@regione.fvg.it

Abstract

Computer network systems are often subject to several types of attacks. For example the distributed Denial of Service (DDoS) attack introduces an excessive traffic load to a web server to make it unusable. A popular method for detecting attacks is to use the sequence of source IP addresses to detect possible anomalies. With the aim of predicting the next IP address, the Probability Density Function of the IP address sequence is estimated. Prediction of source IP address in the future access to the server is meant to detect anomalous requests. In other words, during an access to the server, only predicted IP addresses are permitted and all others are blocked. The approaches used to estimate the Probability Density Function of IP addresses range from the sequence of IP addresses seen previously and stored in a database to address clustering, normally used by combining the K-Means algorithm. Instead, in this paper we consider the sequence of IP addresses as a numerical sequence and develop the nonlinear analysis of the numerical sequence. We used nonlinear analysis based on Volterra's kernels and Hammerstein's models.

DOI reference number: 10.18293/DMSVIVA2020-018

1 Introduction

User modeling is an important task for web applications dealing with large traffic flows. They can be used for a variety of applications such as to predict future situations or classify current states. Furthermore, user modeling can improve detection or mitigation of Distributed Denial of Service (DDoS) attack [11, 15, 13], improve the quality of service (QoS) [16, 10], individuate click fraud detection and optimize traffic management. In peer-to-peer (P2P) overlay networks, IP models can also be used for optimizing request routing [1]. Those techniques are used by severs for deciding how to manage the actual traffic. In this context, also outlier detection methods are often used if only one class is known. If, for example, an Intrusion Prevention System wants to mitigate DDoS attacks, it usually has only seen the normal traffic class before and it has to detect the outlier class by its different behaviour. In this paper we deal with the management of DDos because nowadays it has become a major threat in the internet. Those attacks are done by using a large scaled networks of infected PCs (bots or zombies) that combine their bandwidth and computational power in order to overload a publicly available service and denial it for legal users. Due to the open structure of the internet, all public servers are vulnerable to DDoS attacks. The bots are usually acquired automatically by hackers who use software tools to scan through the network, detecting vulnerabilities and exploiting the target machine. Furthermore, there is also a strong need to mitigate DDoS attacks near the target, which seems to be the only solution to the problem in the current internet infrastructure. The aim of such a protection system is to limit their destabilizing effect on the server through identifying malicious requests. There are multiple strategies with dealing with DDoS attacks. The most effective ones are the near-target filtering solutions. They estimates normal user behavior based on IP packet header information. Then, during an attack the access of outliers is denied. One parameter that all methods have in common is the source IP address of the users. It is the main discriminant for DDoS traffic classification. However, the methods of storing IP addresses and estimating their density in the huge IP address space, are different. In this paper, we present a novel approach based on system identification techniques and, in particular, on the Hammerstein models. A broader overview of state-of-the-art research on the available methods for DDoS traffic classification is given by [9]. The paper is organized as follows. In Sections 2 and 3 we present our proposed a technique based based on Hammerstein models and we recall some similar model. Although DDoS mitigation is the most important practical application for IP density estimation, we do not restrict the following work on this topic. Our generic view on IP density estimation may be valuable to other applications as well. One might think of preferring regular customers in overload situations (flash crowd events), identifying non-regular users on websites during high click rates on online advertisements (click fraud detection) or optimizing routing in peer-to-peer networks. Finally, in Section 4 we draw conclusions and indicate future work. The extended version of this paper appears in [7].

2 Analytic Prediction

Data driven identification of mathematical models of physical systems (i.e. nonlinear) starts with representing the systems as a black box. In other terms, while we may have access to the inputs and outputs, the internal mechanisms are totally unknown to us. Once a model type is chosen to represent the system, its parameters are estimated through an optimization algorithm so that eventually the model mimics at a certain level of fidelity the inner mechanism of the nonlinear system or process using its inputs and outputs. This approach is, for instance, widely used in the related *big data analytics* area (e.g., [6, 3, 5, 8])

In this work, we consider a particular sub-class of nonlinear predictors: the Linear-in-the-parameters (LIP) predictors. LIP predictors are characterized by a linear dependence of the predictor output on the predictor coefficients. Such predictors are inherently stable, and that they can converge to a globally minimum solution (in contrast to other types of nonlinear filters whose cost function may exhibit many local minima) avoiding the undesired possibility of getting stuck in a local minimum. Let us consider a causal, time-invariant, finite-memory,continuous nonlinear predictor as described in (1).

$$\hat{s}(n) = f[s(n-1), \ldots, s(n-N)] \tag{1}$$

where $f[\cdot]$ is a continuous function, $s(n)$ is the input signal and $\hat{s}(n)$ is the predicted sample. We can expand $f[\cdot]$ with a series of basis functions $f_i(n)$, as shown in (2).

$$\hat{s}(n) = \sum_{i=1}^{\infty} h(i) f_i[s(n-i)] \tag{2}$$

where $h(i)$ a re proper coefficients. To make (2) realizable we truncate the series to the first N terms, thus we obtain

$$\hat{s}(n) = \sum_{i=1}^{N} h(i) f_i[s(n-i)] \tag{3}$$

In the general case, a linear-in-the-parameters nonlinear predictor is described by the input-output relationship reported in (4).

$$\hat{s}(n) = \vec{H}^T \vec{X}(n) \tag{4}$$

where $\vec{H}^T$ is a row vector containing predictor coefficients and $\vec{X}(n)$ is the corresponding column vector whose elements are nonlinear combinations and/or expansions of the input samples.

2.1 Linear Predictor

Linear prediction is a well known technique with a long history [12]. Given a time series $\vec{X}$, linear prediction is the optimum approximation of sample $x(n)$ with a linear combination of the N most recent samples. That means that the linear predictor is described as eq. (5).

$$\hat{s}(n) = \sum_{i=1}^{N} h_1(i) s(n-i) \tag{5}$$

or in matrix form as

$$\hat{s}(n) = \vec{H}^T \vec{X}(n) \tag{6}$$

where the coefficient and input vectors are reported in (7) and (8).

$$\vec{H}^T = \begin{bmatrix} h_1(1) & h_1(2) & \dots & h_1(N) \end{bmatrix} \tag{7}$$

$$\vec{X}^T = \begin{bmatrix} s(n-1) & s(n-2) & \dots & s(n-N) \end{bmatrix} \tag{8}$$

2.2 Non-Linear Predictor based on Volterra Series

As well as Linear Prediction, Non Linear Prediction is the optimum approximation of sample $x(n)$ with a non linear combination of the N most recent samples. Popular nonlinear predictors are based on Volterra series [14]. A Volterra predictor based on a Volterra series truncated to the second term is reported in (9).

$$\hat{x}(n) = \sum_{i=1}^{N_1} h_1(i) x(n-i) + \sum_{i=1}^{N_2} \sum_{j=i}^{N_2} h_2(i,j) x(n-i) x(n-j) \tag{9}$$

where the symmetry of the Volterra kernel (the h coefficients) is considered. In matrix terms, the Volterra predictor is represented in (10).

$$\hat{s}(n) = \vec{H}^T \vec{X}(n) \tag{10}$$

where the coefficient and input vectors are reported in (12) and (12).

$$\vec{H}^T = \begin{bmatrix} h_1(1) & h_1(2) \dots h_1(N) \\ h_2(1,1) & h_2(1,2) \dots h_2(N_2, N_2) \end{bmatrix} \tag{11}$$

$$\vec{X}^T = \begin{bmatrix} s(n-1) & s(n-2) \dots s(n-N_1) \\ s^2(n-1) & s(n-1)s(n-2) \dots s^2(n-N_2) \end{bmatrix} \tag{12}$$

2.3 Non-Linear Predictor based on Functional Link Artificial Neural Networks (FLANN)

FLANN is a single layer neural network without hidden layer. The nonlinear relationships between input and output are captured through function expansion of the input signal exploiting suitable orthogonal polynomials. Many authors used for examples trigonometric, Legendre and Chebyshev polynomials. However, the most frequently used basis function used in FLANN for function expansion are trigonometric polynomials [17]. The FLANN predictor can be represented by eq.(13).

$$\hat{s}(n) = \sum_{i=1}^{N} h_1(i) s(n-i) + \sum_{i=1}^{N} \sum_{j=1}^{N} h_2(i,j) \cos[i\pi s(n-j)] + \sum_{i=1}^{N} \sum_{j=1}^{N} h_2(i,j) \sin[i\pi s(n-j)] \tag{13}$$

Also in this case the Flann predictor can be represented using the matrix form reported in (14).

$$\hat{s}(n) = \vec{H}^T \vec{X}(n) \tag{14}$$

where the coefficient and input vectors of FLANN predictors are reported in (15) and (16).

$$\vec{H}^T = \begin{bmatrix} h_1(1) & h_1(2) \dots h_1(N) \\ h_2(1,1) & h_2(1,2) \dots h_2(N_2, N_2) \\ h_3(1,1) & h_3(1,2) \dots h_3(N_2, N_2) \end{bmatrix} \tag{15}$$

$$\vec{X}^T = \begin{bmatrix} s(n-1) & \dots & s(n-N) & \\ \cos[\pi s(n-1)] & \dots & \dots & \cos[\pi s(n-N_2)] \\ \sin[\pi s(n-1)] & \dots & \dots & \sin[\pi x(s-N_2)] \end{bmatrix} \tag{16}$$

2.4 Non-Linear Predictors based on Hammerstein Models

Previous research [2] shown that many real nonlinear systems, spanning from electromechanical systems to audio systems, can be modeled using a static non-linearity. These terms capture the system nonlinearities, in series with a linear function, which capture the system dynamics as shown in Figure 1.

Indeed, the front-end of the so called Hammerstein Model is formed by a nonlinear function whose input is the system input. Of course the type of non-linearity depends on the actual physical system to be modeled.

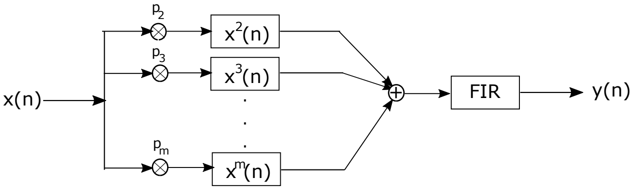


Figure 1. Representation of the Hammerstein Models

The output of the nonlinear function is hidden and is fed as input of the linear function. In the following, we assume that the non-linearity is a finite polynomial expansion, and the linear dynamic is realized with a Finite Impulse Response (FIR) filter. Furthermore, in contrast with [2], we assume a mean error analysis and we postpone the analysis in the robust framework in future work. In other word,

$$z(n) = p(2)x^2(n) + p(3)x^3(n) + \dots p(m)x^m(n) = \\ = \sum_{i=2}^{M} p(i)x^i(n) \quad (17)$$

On the other hand, the output of the FIR filter is:

$$y(n) = h_0(1)z(n-1) + \dots + h_0(N)z(n-N) = \\ = \sum_{j=1}^{N} h_0(j)z(n-j) \quad (18)$$

Substituting (17) in (18) we have:

$$y(n) = \\ \sum_{i=1}^{N} h_0(i)z(n-i) = \sum_{j=1}^{N} h_0(j) \sum_{i=2}^{M} p(i)x^i(n-j) = \\ \sum_{i=2}^{M}\sum_{j=1}^{N} h_0(j)p(i)x^i(n-j) \quad (19)$$

Setting $c(i,j) = h_0(j)p(i)$ we write

$$y(n) = \sum_{i=2}^{M}\sum_{j=1}^{N} c(i,j)x^i(n-j) \quad (20)$$

This equation can be written in matrix form as

$$\hat{s}(n) = \vec{H}^T \vec{X}(n) \quad (21)$$

where

$$\vec{H}^T = \begin{bmatrix} c(2,1) & c(2,2)\dots c(2,N_2) \\ c(3,1) & c(3,2)\dots c(3,N_2) \\ \dots c(M,1) & c(M,2)\dots c(M,N) \end{bmatrix} \quad (22)$$

$$\vec{X}^T = \\ \begin{bmatrix} s^2(n-2) & \dots & s^2(n-N) & \\ s^3(n-2) & \dots & s^3(n-N) & \\ s^M(n-2) & \dots & s^M(n-1) & \dots s^M(n-N) \end{bmatrix} \quad (23)$$

3 Estimation of Predictor Parameters

So far we saw that all the predictors can be expressed, at time instant n, as

$$\hat{s}(n) = \vec{H}^T \vec{X}(n) \quad (24)$$

with different definitions of the input, $\vec{X}(n)$, end parameters vectors $\vec{H}^T$. There are two well known possibilities for estimating the optimal parameter vector.

3.1 Block-based Approach

The Minimum Mean Square estimation is based on the minimization of the mathematical expectation of the squared prediction error $e(n) = s(n) - \hat{s}(n)$

$$E[e^2] = E[(s(n) - \hat{s}(n))^2] = E[(s(n) - \vec{H}^T\vec{X}(n))^2] \quad (25)$$

The minimization of (25) is obtain by setting to zero the Laplacian of the mathematical expectation of the squared prediction error:

$$\nabla_H E[e^2] = E[\nabla_H e^2] = E[2e(n)\nabla_H e] = 0 \quad (26)$$

which leads to the well known unique solution

$$\vec{H}_{opt} = \vec{R}_{xx}^{-1}\vec{R}_{sx} \quad (27)$$

where

$$\vec{R}_{xx}(n) = E[\vec{X}(n)\vec{X}^T(n)] \quad (28)$$

is the statistical auto-correlation matrix of the input vector $\vec{X}(n)$ and

$$\vec{R}_{sx}(n) = E[s(n)\vec{X}(n)] \quad (29)$$

is the statistical cross-correlation vector between the signal $s(n)$ and the input vector $\vec{X}(n)$. The mathematical expectations of the auto and cross correlation are estimated using

$$\vec{R}_{xx}(n) = \frac{\sum_{k=1}^{n} \vec{X}(n)\vec{X}^T(n)}{n} \quad (30)$$

is the statistical auto-correlation matrix of the input vector $\vec{X}(n)$ and

$$\vec{R}_{sx}(n) = \frac{\sum_{k=1}^{n} s(k)(n)\vec{X}(n)}{n} \quad (31)$$

3.2 Adaptive Approach

Let us consider a general second order terms of a Volterra predictor

$$y(n) = \sum_{k=0}^{N-1} \sum_{r=0}^{N-1} h_2(k,r) x(n-k) x(n-r) \tag{32}$$

It can be generalized for higher order term as

$$\sum_{k_1=1}^{N} \cdots \sum_{k_p=1}^{N} c_{k_1} \cdots c_{k_p} H_p \left\{ x_{k_1}(n), \cdots x_{k_p}(n) \right\} \tag{33}$$

where

$$\sum_{k=1}^{N} c_k x_k(n). \tag{34}$$

For the sake of simplicity and without loss of generality, we consider a Volterra predictor based on a Volterra series truncated to the second term

$$\hat{r}(n) = \sum_{i=1}^{N_1} h_1(i) r(n-i) + \sum_{i=1}^{N_2} \sum_{j=i}^{N_2} h_2(i,j) r(n-i) r(n-j) \tag{35}$$

By defining

$$H^T(n) = |h_1(1), \cdots, h_1(N_1), h_2(1,1), \cdots, h_2(N_2, N_2)| \tag{36}$$

and

$$\begin{aligned} X^T(n) = \big| r(n-1), \cdots, r(n-N_1), r^2(n-1) \\ r(n-1)r(n-2), \cdots, r^2(n-N_2) \,| \end{aligned} \tag{37}$$

Eq (35) can be rewritten as follows

$$\hat{r}(n) = H^T(n) X(n). \tag{38}$$

In order to estimate the best parameters H, we consider the following loss function

$$J_n(H) = \sum_{k=0}^{n} \lambda^{n-k} \left[\hat{r}(k) - H^T(n) X(k) \right]^2 \tag{39}$$

where λ^{n-k} weights the relative importance of each squared error. In order to find the H that minimizes the convex function (39) it is enough to impose its gradient to zero, i.e.,

$$\nabla_H J_n(H) = 0 \tag{40}$$

That is equivalent to

$$R_{XX}(n) H(n) = R_{rX}(n) \tag{41}$$

where

$$\begin{aligned} R_{XX}(n) &= \textstyle\sum_{k=0}^{n} \lambda^{n-k} X(k) X^T(k) \\ R_{rX}(n) &= \textstyle\sum_{k=0}^{n} \lambda^{n-k} r(k) X(k) \end{aligned} \tag{42}$$

It follows that the best H can be computed by

$$H(n) = R_{XX}^{-1}(n) R_{rX}(n) \tag{43}$$

Since

$$R_{XX}(n) = \lambda R_{XX}(n-1) + X(n) X^T(n) \tag{44}$$

it follows that

$$\begin{aligned} & R_{XX}^{-1}(n) = \\ & \quad \frac{1}{\lambda} \left[R_{XX}^{-1}(n-1) - k(n) X^T(n) R_{XX}^{-1}(n-1) \right] \end{aligned} \tag{45}$$

where $k(n)$ is equal to

$$k(n) = \frac{R_{XX}^{-1}(n-1) X(n)}{\lambda + X^T(n) R_{XX}^{-1}(n-1) X(n)} \tag{46}$$

Instead, matrix $R_{rX}(n)$ in (43) can be written as

$$R_{rX}(n) = \lambda R_{rX}(n-1) + r(n) X(n) \tag{47}$$

Thus, inserting Eq (47) and Eq (45) in Eq (43) and rearranging the terms, we obtain

$$H(n) = H(n-1) + R_{XX}^{-1}(n) X(n) \epsilon(n) \tag{48}$$

where

$$\epsilon = \hat{r}(n) - H^T(n-1) X(n) \tag{49}$$

By recalling Eq. (46), we can write Eq. (48) as

$$H(n) = H(n-1) + \epsilon(n) k(n) \tag{50}$$

By introducing, the new notation,

$$F(n) = S^T(n-1) X(n) \tag{51}$$

The previous equations can be resumed by the following system

$$\begin{cases} L(n) = S(n-1) F(n) \\ \beta(n) = \lambda + F^T(n) F(n) \\ \alpha(n) = \frac{1}{\beta(n) + \sqrt{\lambda \beta(n)}} \\ S(n) = \frac{1}{\sqrt{\lambda}} \left[S(n-1) - \alpha(n) L(n) F^T(n) \right] \\ \epsilon(n) = \hat{r}(n-1) - \alpha(n) L(n) F^T(n) \\ \epsilon(n) = H(n-1) + L(n) \frac{\epsilon(n)}{\beta(n)} \end{cases} \tag{52}$$

It should be noted that by using Eq (52) the estimation adapts in each step in order to decrease the error. Thus, the system structure is somehow similar to the Kalman filter.

Finally, we define the estimation error as

$$e(n) = r(n) - H^T(n) X(n) \tag{53}$$

It is worth noting that the computation of the predicted value from Eq. (38) requires $6N_{\text{tot}} + 2N_{\text{tot}}^2$ operations, where $N_{\text{tot}} = N_1 + N_2 (N_2 + 1)/2$.

4 Conclusions

In this paper, we presented a new way to deal with cyber attack by using Hammerstein models. Future work will have two objectives. First, we want to consider the problem in a stochastic optimization settings. Second, we want to test the approach on other case studies, by also exploiting *knowledge management methodologies* (e.g., [4]).

Acknowledgements

This research has been partially supported by the French PIA project "Lorraine Université d'Excellence", reference ANR-15-IDEX-04-LUE.

References

[1] A. Agrawal and H. Casanova. Clustering hosts in p2p and global computing platforms. pages 367– 373, 06 2003.

[2] V. Cerone, E. Fadda, and D. Regruto. A robust optimization approach to kernel-based nonparametric error-in-variables identification in the presence of bounded noise. In *2017 American Control Conference (ACC)*. IEEE, may 2017.

[3] G. Chatzimilioudis, A. Cuzzocrea, D. Gunopulos, and N. Mamoulis. A novel distributed framework for optimizing query routing trees in wireless sensor networks via optimal operator placement. *J. Comput. Syst. Sci.*, 79(3):349–368, 2013.

[4] A. Cuzzocrea. Combining multidimensional user models and knowledge representation and management techniques for making web services knowledge-aware. *Web Intelligence and Agent Systems*, 4(3):289–312, 2006.

[5] A. Cuzzocrea and E. Bertino. Privacy preserving OLAP over distributed XML data: A theoretically-sound secure-multiparty-computation approach. *J. Comput. Syst. Sci.*, 77(6):965–987, 2011.

[6] A. Cuzzocrea, R. Moussa, and G. Xu. Olap*: Effectively and efficiently supporting parallel OLAP over big data. In *Model and Data Engineering - Third International Conference, MEDI 2013, Amantea, Italy, September 25-27, 2013. Proceedings*, pages 38–49, 2013.

[7] A. Cuzzocrea, E. Mumolo, E. Fadda, and M. Tessarotto. A novel big data analytics approach for supporting cyber attack detection via non-linear analytic prediction of ip addresses. In *Computational Science and Its Applications - ICCSA 2020 - 20th International Conference, Cagliari, Italy, July 1-4, 2020, Proceedings*, 2020.

[8] A. Cuzzocrea and V. Russo. Privacy preserving OLAP and OLAP security. In *Encyclopedia of Data Warehousing and Mining, Second Edition (4 Volumes)*, pages 1575–1581. 2009.

[9] S. Dietrich, N. Long, and D. Dittrich. Analyzing distributed denial of service tools: The shaft case. pages 329–339, 12 2000.

[10] E. Fadda, P. Plebani, and M. Vitali. Optimizing monitorability of multi-cloud applications. pages 411–426, 06 2016.

[11] M. Goldstein, C. Lampert, M. Reif, A. Stahl, and T. Breuel. Bayes optimal ddos mitigation by adaptive history-based ip filtering. In *Seventh International Conference on Networking (icn 2008)*, pages 174–179, 2008.

[12] J. Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, 1975.

[13] G. Pack, J. Yoon, E. Collins, and C. Estan. On filtering of ddos attacks based on source address prefixes. pages 1–12, 08 2006.

[14] Z. Peng and C. Changming. Volterra series theory: A state-of-the-art review. *Chinese Science Bulletin (Chinese Version)*, 60:1874, 01 2015.

[15] H.-X. Tan and W. Seah. Framework for statistical filtering against ddos attacks in manets. pages 8 pp.–, 01 2006.

[16] Y. Yang and C.-H. Lung. The role of traffic forecasting in qos routing - a case study of time-dependent routing. pages 224 – 228 Vol. 1, 06 2005.

[17] H. Zhao and J. Zhang. Adaptively combined fir and functional link artificial neural network equalizer for nonlinear communication channel. *IEEE Transactions on Neural Networks*, 20(4):665–674, 2009.

Spatial reference frame based user interface design in the virtual reality game design

FU Yaqin
School of Art and Design
Shanghai University of Engineering Science
Shanghai 201620, China
fyq19946285149@163.com

LI Qi
School of Art and Design
Shanghai University of Engineering Science
Shanghai 201620, China
richaqli@yahoo.com

***Abstract*—Virtual reality (VR) creates a virtual environment for learning and education as its strong sense of presence and real-time interactivity. Much research focus on the issues of VR hardware and platform, however, little research involve the design of user interface in VR systems. Different from other interface, a VR interface is designed as a floating menu that virtually presented in the front of users. This paper aims to investigate the influence of user experience from the various spatial reference frames and how three spatial reference frames based user interface and the sense of presence impact user performance in the virtual environment. We developed a VR game for fire safety training and evaluated three types of VR user interface commonly used in VR systems. Sixty participants were involved in this evaluation based on the task completion times, completion rates and error rates. Our results show that the spatial reference frames can be critical important to the user experience when users interact with the game in the virtual environment. We found that the user interface of different spatial reference frames also has significant difference in the sense of presence and user performance of virtual environment.**

Keywords: User interface; VR; Virtual environment; Presence; spatial reference frame

I. INTRODUCTION

In the learning and education, virtual reality has been used to create the experience of immersion, interactivity and imagination [1]. Although VR systems have been developed for various purposes, designing user interface for a VR system still encounters many challengers, which includes the lack of hardware pointing tool and unestablished interactive modes [2]. It is partly because the limitation of VR hardware and design fails to meet human cognitive ability. User interface design is an important part of VR system, but the guidelines of user interface design relies on the traditional two-dimensional user interface paradigm on video game, which is difficult to meet the good uses' experience in three-dimensional immersive virtual environment [3].

The challenge is how users can easily find the information in a large free space [4]. Currently, the floating menus are used in the most of VR user interface design, which has no a physical screen similarly to that used in two-dimensional user interface device. The floating menus in virtual environment have a number of issues, including interface manipulation and unsatisfied user experience. For example, it is difficult for users to touch the floating menu with their fingers, which is different from the physical screen menu. When users encounter difficulty to manipulate the interface, they struggle to sort the problems that cause low sense of presence. In this paper, we aim to explore how user interface design influences the immersion and user performance in the virtual environments. Spatial reference frames are main design factors, which have important influence on the user experience in virtual environment. It lacks the sufficient experiment to prove how influence on the sense of presence and user satisfaction in virtual space [5].

II. THE SPATIAL REFERENCE FRAMES

There are three types of the space reference frames in 3D virtual environment, which are Head Reference Frame (HRF), Body Reference Frame (BRF) and Virtual World Reference Frame (VWRF), which are used in the design of user interface. The selection of reference frames may have different influence on the sense of presence and user performance of virtual environment. Some research demonstrates that menus displacement has impacted to users in virtual environment. For example, placing a menu above a user's head may cause the tiredness and using different reference frames affecting the user's performance with elements, such as buttons and slider bars on the menus [15]. Other research suggests that the displacement of a menu can also affect user's perception of distance when manipulating an element by hand, thus affecting accurate positioning [16]. It suggests that virtual environment emphasized spatial activities are more likely to depend on choosing reference frame, which directly leads to enhance performance.

Obviously, virtual reality has an advantage in the development of educational programs that focus on spatial activities, such as exploring the world, learning geography, and various safety training. The teaching tasks rely on the spatiality and immersion in virtual reality, such as learning the knowledge of galaxies or exploring spatial structure of a building from a virtual space. Therefore, we developed a VR game based on fire safety training to investigate the impact of interface design with three types of spatial reference frames. Sixty students were recruited to evaluate immersion and performance in the user interface based on HRF, TRF and VWRF as a set of variables. We collected the data of task completion times, completion rates and error rates [6]. We also collected the data from users' subjective experience via Semi-

DOI reference number: 10.18293/DMSVIVA20-020

structured interview and Presence questionnaire (PQ) to investigate participant’s sense of presence.

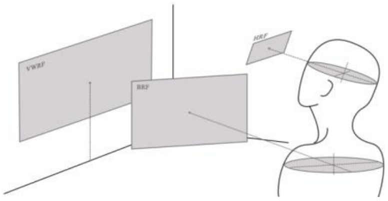


Figure 1. Three types of spatial reference frames, which are commonly used in virtual environment.

III. VR User Interface and Presence

User interface (UI) allows users to interact with a virtual environment [7], such as exploring the environment, switching scenes, adjusting proportions, settings and other tasks. A good user interface provides users instant information, and adapts quickly to the virtual environments and completes the tasks accurately; while a bad user interface can largely reduce the user's experience even cause them to give up. In the early research, Laurel and Mountford [8] highlight the UI design principles including user-oriented, consistent, immediate feedback and others. In the 1990s, research demonstrates that a virtual menu is an effective tool to interact with a virtual environment. To evaluate the menu presentation, Kim et al. [9] compared three positions, which include static with the scene, static with the user view and static with a scene object. However, Kim’s study failed to find out the impact of the menu on user’ performance. Dachselt and H übner [10] examined taxonomy of interfaces of controlled menus, in which most of the interfaces were considered as the “simple menu”. However, the choice of the division criteria has limited to a particular requirement, which may not require in a given approach [11].

Much research emphasize the relationship between virtual reality interface design and user experience [2] and how the UI design improve user experience and performance in virtual environment. Some studies argue that users' experience and performance are affected by the level of presence in the virtual environment. A higher level of presence will result in higher user experience and performance [12]. When a spatial task is completed through a monitor in the traditional computer, the information is a flat image in human brain [13] and users cannot perceive the depth. The notion of spatial depth is realized by binocular vision to detect the distance and stereoscopic object of three-dimensional space [14]. Therefore, the influencing factors of UI design in virtual environment not only include design factors, such as color and observation in two-dimensional environment, but also involve a three-dimensional space of x, y and z axis, which makes VR interface design more challenging.

The sense of presence is considered as an important factor to create good user experience in virtual environment. Without an interactive environment conformed to the real human life, it is unlikely to have experience of presence. Therefore, the level of presence is also considered as important criteria to evaluate the user experience in the virtual environment.

IV. Research Methods

A. Participants

Sixty students were recruited for this study. They were thirty females and thirty males aged between 20 and 25 and the average age are 23. All participants have some knowledge of fire safety. None of the participants had any experience in VR fire safety training. All participants were randomly divided into three groups (A, B and C) with the same number of participants. The UI design in each group is based on the spatial reference frame factors and all other design is the same (Table 1). During the process in the virtual environment, all the participants' performances were recorded.

TABLE I. Design of experiment groups.

Group	NUMBER OF PEOPLE	SPATIAL REFERENCE SYSTEM TYPE
A	20	HRF
B	20	BRF
B	20	VWRF

B. Experimental design

HTC Vive device was used as the platform for the development of fire safety VR game. We tested three VR user interfaces under three different spatial reference frames, which were HRF interface layout, BRF interface layout and VWRF interface layout. For accuracy, other factors, such as color, font, layout of interface, icon were keep as the same. We explored the different level of presence and user performance with three different interfaces at three different spatial reference frames to each user. User performance includes task completion times, completion rates, and error rates. The task completion times (item/min) was defined as the task completion phase divided by the total manipulating time (in minutes). The completion rates was defined as the number of quests completed divided by the number of all quests in the game. Error rates were the number of mission errors recorded during the game. The level of presence was assessed by semi-structured interviews and PQ. After the semi-structured interview, we asked participants to complete a PQ questionnaire to assess the level of presence in the virtual environment.

C. Procedures

The experimental process is as follows: basic information was collected from the participants, including age, gaming experience, fire safety knowledge and experience (including field exercises and fire-related science videos). All participants were informed the VR game tasks, and received an adaptation period of half an hour to reduce the additional psychological load. We assigned sixty participants with three tasks of user interfaces: HRF-UI, BRF-UI and VWRF-UI.

Before the game began, participants need to familiarize themselves with the game environment. They were required to practice all game with HTC Vive device. Throughout the tests, the system automatically recorded each action and evaluated the user performance of the training, including the speed of manipulation, percentage of completion, and error rates. After completing the tasks, all participants were required to complete the semi-structured interview and a PQ questionnaire.

V. DISCUSSION

We analyzed the data collected from each participant. The results indicate that user interfaces based on different spatial reference influence the presence and user performance in the virtual environment. Group A that has 20 participants reported 35% prefer the HRF-UI manipulation. The distribution of group B reflected 40% likes and 60% dislikes. The distribution of group C was 55% liked and 45% dislikes. It is clearly that VWRF-UI is much more popular user interface than others. In addition, there is a significant correlation between the three groups' interface design patterns presence and user performance (Table 2). VWRF-UI is positively correlated with expressive presentation, which may be the fact that the user interface based on VWRF enables participants to accurately identify directions in the virtual space. However, when choosing HRF-UI to interact with virtual environment, the influence on user performance is not significant.

TABLE II. THE CORRELATION BETWEEN 3 TYPES OF UI AND USER PERFORMANCE.

spatial reference frame	**Indicators**	**Task completion times**	**Completion rates**	**Error rates**
HRF	Pearson correlation Sig.	0.341 0.197	-0.534* 0.033	-0.827* * 0.000
BRF	Pearson correlation Sig.	0.037 0.021	0.006 0.001	0.154 0.013
VWRF	Pearson correlation Sig.	-0.053 0.819	0.301 0.185	0.533* 0.013

* At the 0.05 level (1-tailed), the correlation is significant.

** At the 0.01 level (2-tailed), the correlation is significant.

VI Ê CONCLUSIONS

In this study, spatial reference frame, a major factor of interface design in VR games, is analyzed on the basis of experiments by the level of users' sense of presence and user performance. It explores how interface design in different spatial reference frames affect users' presence and user performance. The results show that participants could quickly interact with the virtual environment using HRF-UI, but the level of presence is relatively poor and the score of user performance is low. Using the BRF-UI, participants reported that the interface delayed as the body moved, but no delay in performing the manipulating task, while the level of presence and user performance were moderate. When using the VWRF-UI, most participants reported a high level of presence, and being able to accurately identify directions, complete individual game tasks, but the speed of the manipulation of interface was slow. Therefore, the design of user interface should help people to locate the spatial position and identify the spatial direction. Although the influence of spatial reference frame is important, other factors including the user's cognition, gender difference, and the level of fire safety knowledge will also affect the presence and user performance of users in the virtual environment. Therefore, other factors should also be considered in the design of virtual reality game.

ACKNOWLEDGMENT

The author wish to recognize the support of Shanghai Pujiang Program (18PJ068).

REFERENCES

[1] G. Yildirim, M. Elban, and S. Yildirim. Analysis of Use of Virtual Reality Technologies in History Education: A Case Study. Asian Journal of Education and Training, 4(2), 62-69. 2018.

[2] M. Bernatchez, and J. M. Robert. Impact of Spatial Reference Frames on Human Performance in Virtual Reality User Interfaces. Journal of Multimedia, 3(5). 2008

[3] C. Sun, W. Hu, and D Xu. Navigation modes, operation methods, observation scales and background options in UI design for high learning performance in VR-based architectural applications. Journal of Computational Design and Engineering, 2019,6(2).

[4] R. Li, Y. Chen, C. Sha, and Z. Lu. Effects of interface layout on the usability of In-Vehicle Information Systems and driving safety. Displays, 2017,49.

[5] S. Serino, F. Morganti, D. Colombo, E. Pedroli, P. Cipresso, and G. Riva. Disentangling the contribution of spatial reference frames to executive functioning in healthy and pathological aging: An experimental study with virtual reality. Sensors, 18(6), 1783. 2018

[6] D. Chen. Overview of interface design elements of VR pplication[J].Wireless interconnection technology, 2018,5(64-65)

[7] B. Shneiderman, C. Plaisant, M. Cohen, S. Jacobs, N. Elmqvist, and N. Diakopoulos. Designing the user interface: strategies for effective human-computer interaction. Pearson. 2016.

[8] R. Jacoby, and S. Ellis, Using Virtual Menus in a Virtual Environment Proceedings of SPIE: Visual Data Interpretation, vol. pp. 39-48, 1992.

[9] M. Mine, UNC Chapel Hill Computer Science, Technical Report TR96-029, 1996..

[10] N. Kim, and Kim, G. J., Park, C.-M., Lee, I., and Lim, S. H., Multimodal menu presentation and selection in immersive virtual environments Proceedings - Virtual Reality Annual International Symposium: IEEE Virtual Reality 2000, Mar 18 Mar 22 2000, vol. pp. 281 2000. 0-7695-0478-7. Pohang Univ of Science and Technology, South Korea.

[11] R. Dachselt, and A. Hübner, "A Survey and Taxonomy of 3D Menu Techniques," Proceedings of the 12th Eurographics Symposium on Virtual Environments (EGVE'06), Lisbon (Portugal), 2006.

[12] K.W. Su, S.C. Chen, P.H. Lin, and C.I. Hsieh. Evaluating the user interface and experience of VR in the electronic commerce environment: a hybrid approach. Virtual Reality, 2019.

[13] So-Yeon Yoon, Yun Jung Choi, and O. Hyunjoo. User attributes in processing 3D VR-enabled showroom: Gender, visual cognitive styles, and the sense of presence. International Journal of Human-Computer Studies,2015,82.

[14] D. Paes, E. Arantes, and J Irizarry. Immersive environment for improving the understanding of architectural 3D models: Comparing user spatial perception between immersive and traditional virtual reality systems. Automation in Construction, 2017,84.

[15] D. Larimer, and D. Bowman, "VEWL: A Framework for Building a Windowing Interface in a Virtual Environment.," Proceedings of INTERACT: IFIP TC13 International Conference on Human-Computer Interaction, pp. 809-812, 2003.

[16] Z, Xia, F, Hu, C, Cheng and M. Gu. Virtual reality space reconstruction based on visual space orientation theory[J].Chinese Journal of Liquid Crystals and Display.2019,34(02):215-219.

A Visualization Analysis of China's Leisure Sports Research

Sun Ming, Zeng Ji
School of Physical Education
Hubei University
Wuhan, China
Corresponding Author: Zeng Ji, 1337266477@qq.com

Hu Tian
Center for Studies of Education and Psychology of
Ethnic Minorities In Southwest China
Southwest University
Chongqing, China
454816870@qq.com

***Abstract*—The social economy's continual development has produced gradual increases in disposable income, which have enriched leisure activities. In order to engage progress in the research of leisure sports in China over the past 20 years and contribute a new angle and direction to the research of leisure sports in China, this paper summarizes the internal and external characteristics of the country's leisure sports research. The data visualization software CiteSpace5.0 is applied to provide a macro analysis of the development of China's leisure sports research over the past 20 years. This method has not been applied to research of this kind before, and this reiterates the innovative character of this paper. It achieves this by extracting the 1998-2017 leisure sports literature from the CSSCI database before applying bibliometric Visual Analysis Software CiteSpace 5.0 to this material. It finds that the momentum of the development of the county's leisure sports is positive, and observes the initial establishment of a cooperation network. However, it also highlights a number of problems, which include the insufficient depth of disciplinary research, the lack of interdisciplinary research results (as shown by the adoption of a single research paradigm) and the neglect of frontier issues.**

Keywords-Chinese Leisure Sports; Leisure Research Hotspots; Leisure Sports Research; Visualization Analysis

I. INTRODUCTION

China's leisure sports research mainly focuses on the history, development status and problems encountered by leisure sports research. Although this research had a late start (Kong Chuihui, 2009), its development has been rapid and it has provided enriching perspectives and generated methods that can be applied (Xiaoyu and Hai, 2016).

But there are a number of general problems in the study of China's leisure sports that need to be resolved. These are as follows: 1) The definition of the leisure sports concept is ambiguous and is not sufficiently comprehensive (Wu Jiangang, 2003; Li Xiangru, 2015); (2) Leisure sports research has limited depth and its results are not sufficiently influential (Xiaoyu and Hai, 2016); (3) Most of the research has an international focus, and does not sufficiently refer to China (Xiaoyu and Hai, 2016); (4) a single research method tends to be adopted and cooperative and interdisciplinary research is

DOI reference number: 10.18293/DMSVIVA20-005

limited; (ibid) (5) research groups are not sufficiently diverse and do not sufficiently acknowledge engagement in leisure sports activities across different social classes (Jiangang, 2003; Chuihui, 2009).

There are grounds for believing that the development of leisure sports will make an important contribution to the rapid transformation of Chinese society and its future modernization. China therefore provides an ideal opportunity to study the relationship between leisure sports research and the stage of national development. Here it should be remembered that most of the research samples of leisure sports have been taken from Western contexts. The study of the development of China's leisure sports research over the past 20 years will provide insight into its future direction.

This paper will explore dynamic changes in China's leisure sports hotspots during the period 1998-2018; it does so with the intention of identifying key characteristics of China's leisure sports research, predicting future research trends and highlighting problems that can be engaged by Chinese researchers. Although development trends in China's leisure sports research have been extensively engaged (Jiangang, 2003; Limin, 2007; Xiaoming, 2008; Chuihui, 2009; Xin, 2012; Xiaoyu, 2016; Xiuyu, 2016), there are still research shortcomings that need to be directly addressed. This paper seeks to contribute to this process by providing an overview of all literature related to leisure sports topics that were published by CSSCI journals and incorporated into the China Knowledge Network (CNKI) during the period 1998-2018. It will apply Citespace with the intention of gaining insight into the literature's features, strengths and deficiencies, along with its dynamic development.

II. Literature Review

China is the world's second largest economy. In 2018, its gross domestic product (GDP) reached $90.03 trillion (USD) and its population expanded to 1.395 billion, accounting for about one-fifth of the world's population. (World Bank, 2018) Since the 1980s, China's annual economic growth rate has, on average, exceeded 9 percent, which far surpasses the annual growth rate of 2.3 percent recorded for developed economies during the same period. (World Bank, 2011). Rapid economic development has greatly improved Chinese living standards (Li et al., 2012) and residents of mainland China now have the time and financial resources to participate in leisure activities. (Liang & Walker, 2011).

In recent times, China's booming economy and overall national strength has created huge dividends for its citizens. Personal income has also increased as a result of more efficient resource allocation and more incentives for private investment (Easterlin et al., 2012). During this transition period, both the leisure time (Yin, 2005) and leisure consumption of Chinese residents have increased significantly (Zhai & Xiao, 2004).

A. The conceptualization of 'leisure sports'

Chinese scholars have not yet reached an agreement on what 'leisure sports' are, and this has interrupted research into conceptualization. The most renowned experts in this area are Tian Hui, Ma Huizhen, Zhou Aiguang, Lu Feng and Xiao Huanqi. Ma Huizhen (2008) and Zhou Aiguang (2009). They define 'leisure sports' in similar terms, and observe they are a sports activity engaged in for leisure, which relieves stress, eliminates fatigue and benefits body and mind. Tian Hui (2006), Xiao Huanqi (2010) and Yu Kehong (2003) define them as sports activities that are engaged in during leisure time, which assist economic, environmental, personal and social development. Lu Feng (2004) suggests they are a collective term for various sports activities engaged under relatively free conditions.

The classification and definition of ‘leisure sports’ by scholars shows that China lacks metacognition of this concept . It is part of a foreign vocabulary, and has no Chinese counterpart. The word needs to be traced back to its root in order to explore its original meaning.

B. *Leisure Sports and individual and collective development*

A substantial amount of research has engaged with the developmental contribution of leisure sports. Since the beginning of the 21st century, population and economic aggregates have both increased and the individual and collective influence of leisure sports has attracted growing academic interest. Lu Feng (2004) suggests that the development of leisure sports has personal, developmental, social, social fashion (communication), social group (organization) and social symbolic functions. Zhang Rui (2014) observes that leisure sports do not just enhance physical health but also improve mental health, promote social interaction, create a relaxed and comfortable environment and contribute to an improved quality of life.

It is only by understanding and promoting leisure sports, in addition to grasping their essential purpose, that it will be possible to promote a healthier body and mind. Lu Gaofeng (2014) suggests that leisure sports provide a basis for various capital conversions. By engaging in leisure sports, higher social classes can accumulate cultural capital, expand social capital, and maintain or develop physical capital by investing in economic capital. Lower social classes can also transform physical capital into economic capital by engaging in leisure sports or can instead transform it into a form of cultural or social capital.

Qiu Yajun (2014) finds that women who do not participate in leisure sports are mainly held back by their own limitations and structural constraints. Female participants in recreational sports activities frequently encounter perceptual and experience limitations. Xiong Huan (2014) suggests that sport dissolves individual and micro-level limitations of women's leisure, and therefore promotes their freedom of choice and empowerment – this, however, is conditional on a more equal, free and reasonable social system. Only when the cultural environment is present will it be possible to effectively achieve this.

Jinyin Day (2015) finds that the leisure sports activities of Shanghai residents are characterized by activity space circle, activity demand differentiation, wide-area characteristics of travel space and the regularization of activity time. Wu Xiaoyang’s (2015) research of the literature demonstrates that leisure sports are one way through which rural citizens integrate into an urban environment. The number of people involved in leisure sports therefore provides insight into the extent of urbanization.

Ye Xin (2015) proposes that gender order is the most important part of women's leisure sports behavior, and suggests it establishes a basis for consciousness while putting in place a material foundation that underpins women's leisure sports behavior. Yan Ke (2016) suggests that the healthy development of leisure sports relies on a benign interaction between individual self-development and social norms. This entails that the psychology and behavior of social individuals will produce their active participation in leisure sports and will encourage their use of social laws, regulations, norms and ethics that control, regulate and stimulate society. Guo Xiujin (2016) observes that the best popular leisure sports involve nature and civilization, and she therefore presents leisure sports as a kind of 'green living style' that helps to make the world a more harmonious place. Yan Ke (2017) notes that individuals make the nation's leisure and sports life possible, and observes the development of national leisure sports promotes an accelerated individualization. Fashion and market tendencies are the social characteristics that correspond to the development of national leisure sports in an individualized era.

Chen Dexu (2017) notes how leisure sports have promoted material development by stimulating sports consumption and economic growth, strengthening social governance, enhancing the stability of political civilization, improving the quality of the population and promoting the healthy development of utility. Sun Fenglin (2018) applies the theory of the ecological food chain network to a case study, and observes that the park can meet the multi-level and multi-type leisure needs of the elderly. Wang Min (2018) suggests that leisure sports are the most active, effective and economical way to deal with an aging society. They can enhance the physical condition, mental health and life satisfaction of the elderly, while reducing their risk of illness. They can alleviate decreased participation in the labour market caused by an aging society, and can also reduce healthcare expenditures and the social burden.

Li Hui (2018) observes that women participate in leisure sport activities, and notes that this does not just challenge the traditional gender order, but also provides a new approach to the construction of this order in the new era. It also enables a healthy China to be constructed along both vertical and horizontal axes. In the first sense, the participation of women in leisure sports activities are based on a challenge to the traditional gender order, the realization of gender role self-identification, an empowered gender freedom, the expansion of gender space and the promotion of women's individuality on terms that encompass the whole life-cycle.

Chinese scholars have a lot of empirical research on the development of human and social development in leisure sports, and there are few theoretical innovations. The research results mainly reflect the subsidiary meaning of leisure sports, and the interpretation of its essential functions is less, and further research is needed.

C. Research into leisure sports majors

Since the start of the century, China's comprehensive national strength has continually grown, along with related sport activities. The concept of 'leisure sports' has become increasingly widely accepted. Colleges and universities have also established leisure sports-related majors that help to meet growing demand for professional leisure sports in China. Because the major is newly developing, it is not amateur enough.The construction of China's leisure sports major has gradually become a research 'hot spot'.

Peng Guoqiang (2014) compares the construction of leisure sports majors in American colleges and universities, and proposes that China's leisure sports majors should meet the needs of the social market, cultivate composite talents that combine sports and leisure, broaden the coverage of leisure sports courses and reflect the curriculum.

The dynamic development of the system, the reform and construction of the leisure sports curriculum and the professional setting reflect diversified social needs, the establishment of a third-party specialized agency and the need for a leisure sports professional audit system that is adapted to China. Wang Xiaoyun (2017) observes that current talent training in the construction of leisure sports is not closely integrated with social requirements, neglects the improvement of comprehensive quality, and also gives insufficient attention to professional practice and the low standardization of training programs. He suggests the country should be market-oriented, should analyze demand for leisure sports professional ability, should clarify the training specifications of leisure sports professionals and should build a competency-oriented and standard curriculum system that helps to achieve the overall optimization of leisure sports professionals. Xu Dapeng (2017) conducts a comparative study of leisure sports majors and social sports majors in the Capital Institute of Physical Education, which concludes that the professional setting of leisure sports should focus on innovation and entrepreneurship education, increase the specific sport choices that are available and also inspire students' innovative spirit.

Due to the lack of a correct understanding of the term "leisure sports", Chinese scholars have limited research on leisure sports majors, and the relevant results are not rich, and there is no research result that reflects the connotation of leisure sports.

D. The industrialization of leisure sports

The sports and leisure industry has become indispensable to China's national economic development and its construction of a harmonious society. The question of how to align China's leisure sports industry with national conditions has come to preoccupy many scholars. Luo Lin (2006) draws on the perspective of cultural chemistry to put forward three principles for the development of the leisure sports industry, specifically an orientation towards people, the accommodation of multicultural values and the development of a national traditional sports culture and valued leisure industry. Wang Xianliang (2015) finds that China's leisure sports industry is characterized by consumption and production simultaneity, industrial integration and sports characteristics. The development of the leisure sports industry should clarify the industrial value chain, optimize the value chain layout and give regional advantages in order to optimize regional layout. Ye Xiaoyu (2016) finds the research of China's leisure sports industry is developing in accordance with contemporary trends, and adopts a macro-level perspective of the layout of the leisure sports industry, development strategy and policy design. Yang Yukai (2017), after studying the development of the leisure sports industry in developed countries, concludes that China's leisure sports industry needs to acknowledge the leading role of leisure sports culture in the healthy development of leisure sports, grasp the brand effect and integrate the industrial chain; this, he observes, will promote its own development.

Zhao Lefa (2017) uses the ‘diamond model’ of strategic management to analyze factors that restrict the competitiveness of China's leisure sports industry. He cites the lag of the mass consumption concept, the small scale of the leisure sports industry, limited integration with other industries and a lack of professional talent. In response, he suggests it is necessary to deepen the health concept, strengthen macro guidance, integrate cross-border industry integration and broaden talent training channels. Yang Lei’s (2017) comparative research finds that, while the market of China's leisure sports industry has great potential, it has not yet shown short-term economic effects. Lei suggests that the sound development of China's leisure sports industry should be based on an understanding of its basic laws of development and distinctive characteristics. Both should be combined with innovation, coordinated development and scientific management.

The research on China's leisure sports industry mainly focuses on quantitative research. The results of qualitative research are limited. It is necessary to change the research paradigm and produce more theoretically in-depth research results.

When studying China’s leisure sports, it is important to explore their function and significance from different angles. The past 20 years have only revealed a few research hotspots and frontier changes in China's leisure sports research, and there is still an ongoing need for further research.

III. Methods

A. Study Setting and Data Collection

The Chinese Social Science Citation Index (CSSCI) database was selected and documents from the period 1998-2017 were searched by using the keywords "leisure sports" and "leisure". After screening, 371 documents that satisfied the requirements were selected as this study’s dataset.

B. Quantitative analysis

This study uses Visual analysis tools and CiteSpace 5.0 to study research hotspots and leisure sports trends in China that have emerged since the reform and ‘opening up’. Their application will avoid subjective judgement and will better demonstrate the process and distribution of leisure sports research. The Citespace 5.0 is installed and run on JAVA platform; it is then processed, before the document data is downloaded by CSSCI. The year slice is set to one year and the threshold value is set to 50. Keyword, author and dispatch agency are generated in the corresponding maps and are then used to analyze the corresponding high frequency vocabulary, core authors and research institutes.

Bibliographic information statistical analysis tools, including SATI3.0 and Microsoft Office Excel 2007, are applied. In addition, the number of papers, organizations, authors and other information in Citespace 5.0 software are summarized and counted for purposes of further analysis.

C. Qualitative analysis

The high citation-rate literature is cut out from a knowledge map, read in full text and then analyzed through a visual map. The high citation-rate literature cut out in the knowledge map is read in full text, and combined with the visualization map and related papers for analysis. Comment on the context, characteristics and future trends of leisure sports development in China, in order to provide reference for new researchers to discover new research points.

IV. Data Analysis

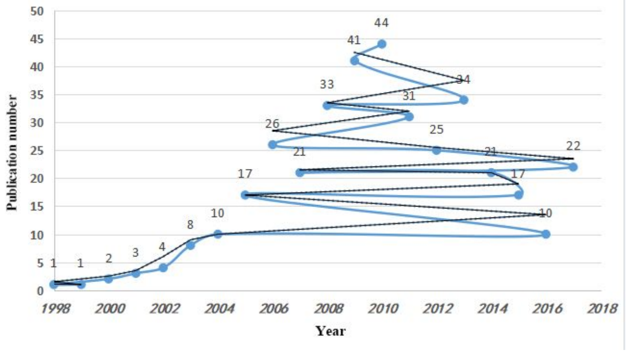


Figure 1. Distribution of leisure sports publications in China: 1998-2017

The time distribution of knowledge domain research can reflect the overall progress and development of this research field. The number of publications during the period 1998-2017 (see Fig. 1) show that, from the end of the 1990s onwards, Chinese scholars began to increasingly engage with the study of leisure sports. The concept of ‘leisure’ was introduced to China for the first time by the scholar Yu Guangyuan, who observes that ‘leisure is an important goal of productivity development. The length of leisure time is parallel to the development of humans’.

After it was gradually acknowledged that sports activities were part of leisure activities, research papers on leisure sports began to appear. Since the beginning of the 21st century, the amount of publications gradually increased, peaking in 2010, and then evidencing a spiral upward distribution trend. Although the volume of papers has declined, the annual number of published papers has remained above 10.

Changes in the number of articles published on leisure sports were closely related to the development of China's economy and society. The World Leisure Organization observes that per capita GDP of $2000 (USD) was the threshold for the rapid growth of leisure demand, and notes that leisure activities began to diversify beyond this point. (Li Xiangru, Ling Ping, Lu Feng,2011) When per capita GDP exceeds $3000 (USD), leisure demand generally arises. In 2006 and 2008, China's per capita GDP respectively exceeded $2000 and $3000 (both USD). Rapid increases in leisure demand led to the development of related research, and reiterated that leisure sports research is a contemporary concern.

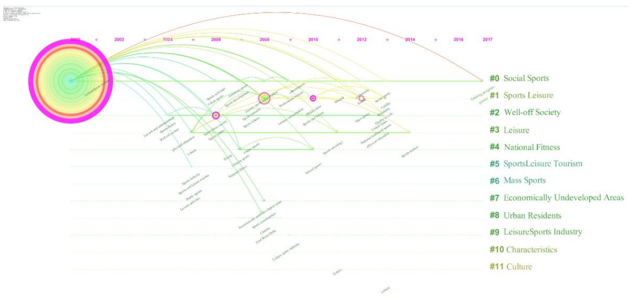
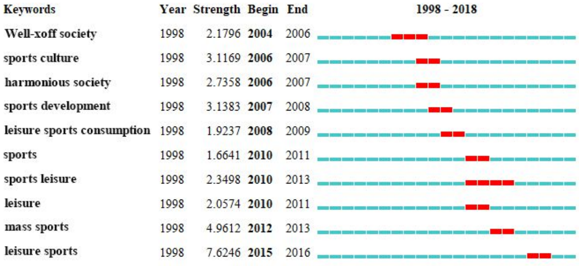


Figure 2. Timeline Analysis of Key Words



Keywords	Year	Strength	Begin	End	1998 - 2018
Well-xoff society	1998	2.1796	**2004**	2006	
sports culture	1998	3.1169	**2006**	2007	
harmonious society	1998	2.7358	**2006**	2007	
sports development	1998	3.1383	**2007**	2008	
leisure sports consumption	1998	1.9237	**2008**	2009	
sports	1998	1.6641	**2010**	2011	
sports leisure	1998	2.3498	**2010**	2013	
leisure	1998	2.0574	**2010**	2011	
mass sports	1998	4.9612	**2012**	2013	
leisure sports	1998	7.6246	**2015**	2016	

Figure 3. Emergence Information Analysis: Keywords

Leisure sports was a dynamic research field that was closely related to economic and social development in China. Material progress was a precondition for the development of leisure sports, which impacted related research fields. The Time Zone Distribution Chart (Fig. 2) and the keyword chart (Fig. 3) make it possible to identify dynamic changes in research hotspots that have occurred since reform and ‘opening up’. The combination of graphic illustrations and interviews with authoritative experts in the field makes it possible to divide research from the last 20 years into three periods:

A. Germination Period (1990s-2000)

Research of leisure sports in China began in the early 1990s, when scholars began to focus on sports and leisure. Some of the most influential contributions were *Leisure Sports Theory* (Cheng Zhili, 1990) and *Sports and the Leisure Life of Chinese Urban Residents* (Liu Depei, 1990). These papers provided theoretical and empirical insight, and highlighted the relationship between social development and the diversification of leisure lifestyle. Due to conceptual underdevelopment, leisure sports was only considered within ‘sports and health’ and ‘social sports’, and was even viewed as ‘tourism and entertainment’, which further underlined its struggle to become an independent discipline. In 1995, China issued an outline of the National Fitness Program, an outline of the Development of the Sports Industry and the Sports Law of the People's Republic of China.

Although research into leisure sports continued to develop, progress was hindered by the under-development of concepts and the inability of the field to distinguish itself from its predecessors. This meant that research results for leisure sports remained limited. At this time, China's per capita income was not sufficient to sustain leisure sports, and so related research remained under-developed.

B. Initial period (2001-2005)

As China became an increasingly developed country in the 21st century, the research on leisure sports began to progress. Some of the most well-known contributions were: *Cultural Significance of the Rise of Leisure Sports* (Chen Rong, 2002), *Investigation and Development Strategy of Urban Leisure Sports Consumption* (Hu Chunwang, 2003), and *Sports Leisure Science* (Lu Feng, 2005). These contributions considered the significance and value of leisure sports by drawing on cultural, economics, leisure and sports perspectives. At this stage, leisure sports mainly focused on "leisure and entertainment", "mass sports", "well-off society", "entertainment and leisure" and "sports and leisure tourism", in addition to other contents.

C. Development period (2006-2017)

During the period 2006-2017, the research hotspots of leisure sports mainly focused on "sports culture", "national fitness", "harmonious society", "American sports development" and "leisure sports consumption". The research results focused on "leisure sports industry" increased substantially. The content involved the basic theory, economic value, educational value, cultural connotation, planning and design of leisure sports. The most influential contributions included *Introduction to Leisure Sports* (Xu Zongxiang, 2007), *On Sports Leisure* (Hu Xiaoming, 2008), *Thoughts on the Cultivation of Leisure Sports Professionals* (Chen Qi, 2008), *Thoughts on the Construction of Leisure Sports Specialty in China from the Perspective of Leisure* (Li Xiangru, 2009), *Formation and Development of Leisure Sports Discipline* (Liang Limin, 2010). *Introduction to Leisure Sports* (Li Xiangru, 2011), *Perspective on China's Leisure Sports* (Li Xiangru, 2012), *Theory and Thinking of China's Leisure Sports Development* (Zhong Bingshu, 2015) and *Research on China's Leisure Sports Practice* (Li Xiangru, 2016).

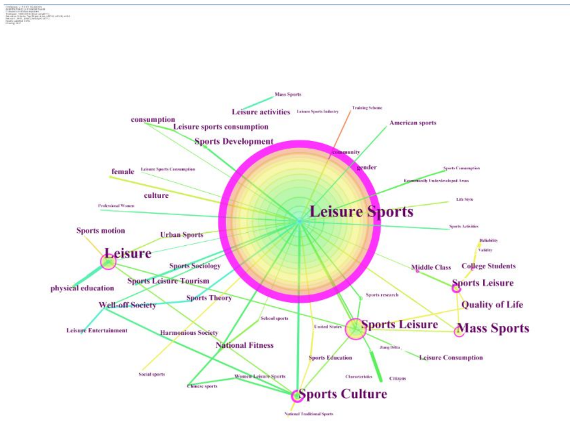


Figure 4. Cluster Co-occurrence Map: Keywords

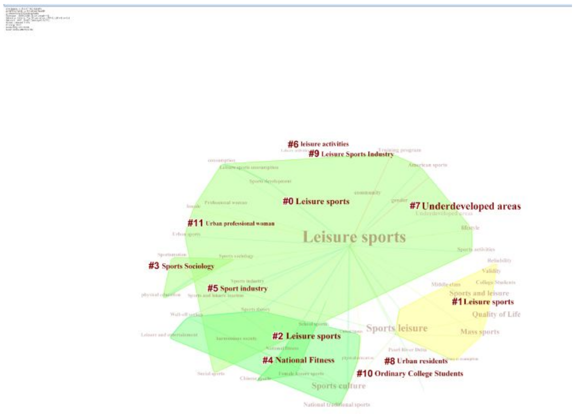


Figure 5. Clustering Common View: Keywords

The keywords in the literature are the core words extracted from the articles. They are core components of an article, and are referenced in the summary and conclusion. Keywords that occur frequency are used to identify hot issues in a research field. The most frequently occurring words collected from CSSCI data are 'leisure sports', 'sports leisure', 'leisure', 'sports culture' and 'sports leisure' (Fig. 4). This shows that, since the reform and 'opening up', scholars in the field of leisure sports have mainly focused on these contents, with the consequence that the research scope of leisure sports in China has generally progressed. Foreign experience, meanwhile, helped to stimulate domestic demand, increase employment and improve national happiness. But extended research into leisure sports in China was still insufficient. There were too many theoretical macro-studies and too few empirical micro-papers. Some leisure sports research directly copied from foreign research, and the results were divorced from China's national conditions.

Leisure sports was a compound subject produced by the combination of ‘Leisure Science’ and ‘Sports Science’. The current tendency to integrate disciplines made it clear that leisure sports cannot exist as an isolated island, and therefore needs to be combined with different disciplines to produce new academic breakthroughs. The research scope of leisure sports is currently mainly combined with the contents of ‘sports industry’, ‘national fitness’ and ‘gender’ (Fig. 5). However, there are too many quantitative studies, too few qualitative studies, insufficient research paradigms, no significant interdisciplinary research and insufficient integration with sociology, pedagogy, economics, anthropology and other disciplines.

For example, an exploration of the social significance of leisure sports can draw on sociology, while the behavioral motivations of leisure sports participants can be engaged from the perspective of Social Action Theory. The symbolic meaning of leisure sports can be explored from the perspective of Symbolic Interaction Theory, and the interactive and communicative value of leisure sports can be explored from the perspective of Daily Life Theory. Leisure sports can also be analyzed from the perspective of pedagogy. The interactive mode of leisure sports can be constructed with the theory of experiential learning, and the relationship between leisure sports and human status acquisition and identity can be analyzed by drawing on the theory of educational stratification and conflict. Leisure sports can also be analyzed by drawing on the theory of economics, while the economic value of leisure sports can be explored through theories of human capital and social reproduction. Anthropological perspectives can be drawn on to assess the cultural significance of leisure sports.

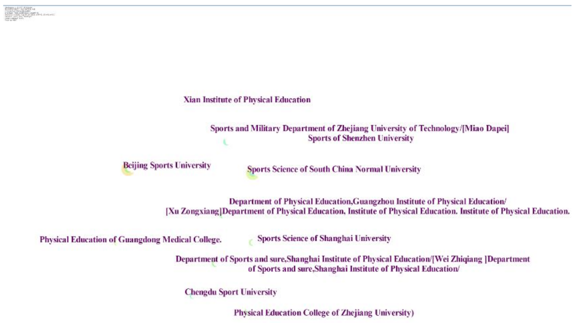

Xian Institute of Physical Education
Sports and Military Department of Zhejiang University of Technology/[Miao Dapei]
Sports of Shenzhen University
Beijing Sports University
Sports Science of South China Normal University
Department of Physical Education,Guangzhou Institute of Physical Education/
[Xu Zongxiang]Department of Physical Education, Institute of Physical Education. Institute of Physical Education.
Physical Education of Guangdong Medical College.
Sports Science of Shanghai University
Department of Sports and sure,Shanghai Institute of Physical Education/[Wei Zhiqiang]Department
of Sports and sure,Shanghai Institute of Physical Education/
Chengdu Sport University
Physical Education College of Zhejiang University)

Figure 6. Co-occurrence Map: Research Institutions

Leisure sports have emerged as a ‘hot spot’ of sports research in recent years. Analysis of the common view spectrum of leisure sports research institutions can provide insight into the core institutions that promoted leisure sports research, and this can help researchers to better understand related research fields from the perspective of research institutions.

Leisure sports research in China is concentrated in different universities (Fig. 6), of which sports and comprehensive universities account for the highest proportion (five and three respectively). Sports colleges and universities tend to focus more on leisure sports research. Comprehensive colleges and universities, meanwhile, engage on an interdisciplinary basis, adopt diverse research perspectives and have a clear advantage with regard to the quality of research. China’s leisure sports research institutions tend to be based in economically developed cities, and this reflects both economic thresholds and the objective law of disciplinary development.

While leisure sports research in China has given rise to different views, the total number of views has not been high; furthermore, the nodes of institutions have been relatively isolated, with no connection between them. Leisure sports research institutions have fewer cross-regional alliances, low efficiency of resource integration and greater disciplinary limitations, none of which are conducive to cross-disciplinary integration of leisure sports research. Research institutions should accordingly strengthen cooperation with the intention of achieving complementary research resources and should work to promote the development of leisure sports research.

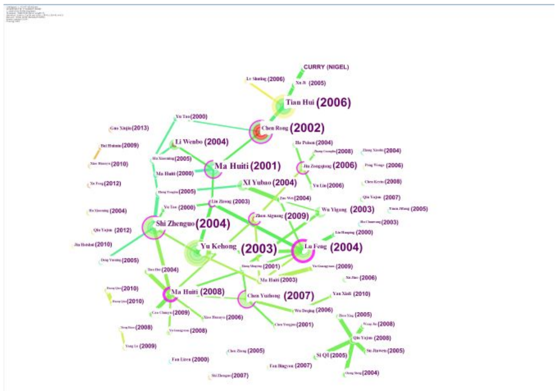
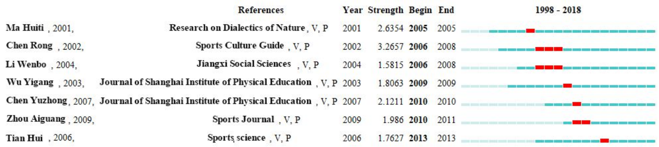


Figure 7. Co-occurrence map: highly-cited literature



References		Year	Strength	Begin	End	1998 - 2018
Ma Huiti , 2001,	**Research on Dialectics of Nature** , V, P	2001	2.6354	**2005**	2005	
Chen Rong , 2002,	**Sports Culture Guide** , V, P	2002	3.2657	**2006**	2008	
Li Wenbo , 2004,	**Jiangxi Social Sciences** , V, P	2004	1.5815	**2006**	2008	
Wu Yigang , 2003,	**Journal of Shanghai Institute of Physical Education** , V, P	2003	1.8063	**2009**	2009	
Chen Yuzhong , 2007,	**Journal of Shanghai Institute of Physical Education** , V, P	2007	2.1211	**2010**	2010	
Zhou Aiguang , 2009,	**Sports Journal** , V, P	2009	1.986	**2010**	2011	
Tian Hui , 2006,	**Sports, science** , V, P	2006	1.7627	**2013**	2013	

Figure 8. Co-occurrence information analysis

Frequently cited contributions to the literature can reveal the knowledge base of leisure sports research (Fig. 7). The most frequently cited paper is Yu Kehong's ‘On Leisure Sports from the Definition of Leisure’, which was published in *China Sports Science and Technology* in 2003. It mainly compares and summarizes the definitions of ‘leisure’ and ‘leisure sports’ put forward by Chinese and foreign scholars, with the intention of developing a more objective, comprehensive and clear definition. The second paper is Tian Hui's ‘Leisure, Leisure Sports and Its Development Trend’, which was published in *Sports Science* in 2006. It mainly analyzes and interprets the meaning of leisure and leisure sports by tracing the historical development of leisure sports and discusses its content. The third is ‘Cultural Significance of the Rise of Leisure Sports’ by Chen Rong, which was published in 2002 in *Sports Culture Guide*. This paper adopts a cultural perspective to discuss the cultural characteristics and significance of leisure sports. Shi Zhenguo published ‘Leisure, Leisure and Leisure Sports’ in *Sports Culture Guide* in 2004. It analyzes the historical origins of leisure and leisure sports, and also engages the concept of leisure sports activities. In 2004, Lu Feng’s ‘Discrimination of Leisure Sports Concepts’ was published in the *Journal of Chengdu Institute of Physical Education*. It defines the concept of leisure sports and incorporates the author’s personal opinions to establish three constructive dimensions of leisure sports.

Forward-looking research has also received extensive attention in the research community (Fig. 8). Ma Huidi’s ‘21st Century and Leisure Economy, Leisure Industry and Leisure Culture’ was published in *Dialectics of Nature* in 2001. It anticipates a stronger interrelation of the leisure industry and China’s culture and economy during the 21st century, and also calls for the strengthening of relevant academic research. It was widely cited in 2005. Li Wenbo’s ‘Leisure Sports Consumption Research: An Interpretation of Culture and Sociology’, which was published in *Jiangxi Social Sciences* in 2004, proposes that sports culture has a unique cultural symbolic significance. It was widely cited during the period 2006-2008.

Wu Yigang's ‘Current Situation and Problems of Leisure Sports Research at Home and Abroa’" was published in *Journal of Shanghai Institute of Physical Education* in 2003. It studies changes within leisure sports and their general development of leisure sports at both the domestic and international level, and also acknowledges limiting factors in each of these respects. It was quoted extensively in 2009. Tian Hui’s ‘Leisure, Leisure Sports and Its Development Trend in China’ was published in *Sports Science* in 2006. It traces the origin and evolution of leisure sports in China and also explains their significance and content. It was widely cited in 2013. Chen Yuzhong's ‘Future Trend of Leisure Sports Development in China’ was published in *Journal of Shanghai Institute of Physical Education* in 2007. It explores the conditions and historical stages of the rise of leisure sports in China, and provides insight into their future prospects. It was widely cited in 2010. These frontier contributions have made a vital contribution by establishing a basis for the future development of leisure sports research.

V. Conclusions and Future Research

A. Conclusions

The scientific knowledge map of core journal papers on leisure sports from the past 20 years is subject to analysis by Citespace 5.0 software, and this confirms that research on leisure sports spiraled upwards from the beginning of the 21st century, before peaking in 2010. Although a decline then followed, a relatively constant annual output was maintained. The research process of leisure sports was hierarchical and closely related to social development and policy dynamics.

Leisure sports research institutions mainly focused on professional sports colleges and comprehensive universities, while financial and economic colleges and normal universities paid less attention to leisure sports. There was less cooperation among scientific research institutions, as effective cooperation mechanisms were still absent. Researchers of leisure sports, including Chen Rong, Yu Kehong, Shi Zhenguo, Lu Feng, Tian Hui, Chen Yuzhong and Zhou Aiguang, contributed to the study of leisure sports by offering different perspectives and methods. But cooperation between different authors was still not sufficient. The hotspots and frontiers of leisure sports research were mainly based on policy, history and culture, and focused on industry, history and comparative sports. This research generally tended to be systematic, pluralistic and innovative.

Although substantial achievements have been made in the research of leisure sports in China, there are still some problems and shortcomings: The connotation of leisure sports research is still not sufficiently deep, and most contributions struggle to extend beyond a relatively superficial concept analysis or assessment of prospects. The results of interdisciplinary research have been insufficient to form horizontal and vertical cooperation mechanisms. In addition, there is also a gap between concrete developments and research. For example, as national fitness has improved, leisure sports items, including yachts, hot balloons, racing cars and RV camping, have also increased; this development has not, however, been reflected in domestic research. Although there have been individual breakthroughs in research paradigms, methods and perspectives, the overall impact has fallen short of what is required.

B. Future Research

In the future, the research on leisure sports can further improve the insufficiency of this research. The depth of connotation of leisure sports research is insufficient. It can be traced from the meta-research level, correctly define the meaning of leisure sports, and carry out theoretical innovation. Interdisciplinary research on leisure sports Too few results should be led by the government level, so that different disciplines can strengthen cooperation and form a sustainable cooperation mechanism; research on leisure sports should be paid attention to, and many emerging leisure sports should receive more attention, such as: Outdoor sports, yachts, motor homes, etc.; enrich the research paradigm of leisure sports, emphasizing the theoretical depth and practical value of leisure sports research through the perspective of qualitative research.

When compared against previous research, this research offers a significant methodological innovation. Its application of visualization software has shown how it is possible to provide an objective representation of the frontiers and hotspots of leisure sports development in China over the last 20 years. Visualization analysis saves a considerable amount of time in accessing the literature, and also makes it possible to extract research hotspots and deficiencies in a certain research field from a large number of papers. In future, it can be both applied to other disciplines and also used to open up new areas of research into leisure sports.

REFERENCES

[1] Z. The Central Committee of the Communist Party of China, State Council(2016). Outline of the "Healthy China 2030" Program. http://www.gov.cn/xinwen/2016-10/25/content_5124174.htm.

[2] J. Wang Liwei, Zhang Yongjun, Fan Suxiang(2007). Leisure Sports: A Historical Category of Interactive Economic Development. Journal of Chengdu Institute of Physical Education, 33 (5), 5-8.

[3] M. Yu Guangyuan(2005). On the Generally Leisure Society. Beijing: China Economic Publishing House, 12-16.

[4] M. Li Xiangru, Ling Ping, Lu Feng(2011). Introduction to Leisure Sports. Beijing: Higher Education Press, 11.

[5] Z. China's annual average GDP .https://www.kuaiyilicai.com/stats/global/yearly_per_country/g_gdp_per_capita/chn.html.

[6] J. Lu Feng, Liu Xishan, Wen Xiaoyuan(2006). Classification of Leisure Sports Activities. Journal of Wuhan Institute of Physical Education, 40 (12), 59.

[7] M. Li Xiangru, Ling Ping, Lu Feng(2011). Introduction to Leisure Sports. Beijing: Higher Education Press, 160.

[8] J. Yu Kehong, Liang Ruowen(2003). On Leisure Sports from the Definition of Leisure. China Sports Science and Technology, 39(1), 21-23.

[9] J. Tian Hui, Zhou Hong(2006). Leisure and leisure sports and their development trend in China. Sports Science, 26 (4), 67-70.

[10] J. Chen Rong(2002). Cultural Significance of the Rise of Leisure Sports.Sports Culture Guide, (2), 12-13.

[11] J. Shi Zhenguo, Tian Yupu(2004). Leisure, Leisure Sports. Sports Culture Guide, (8): 45-46.

[12] J. Lu Feng(2004). Discrimination of the Concept of Leisure Sports. Journal of Chengdu Institute of Physical Education, 30 (5), 32-34.

[13] J. Ma Huidi(2001). The 21st Century and Leisure Economy, Leisure Industry and Leisure Culture. Research on Dialectics of Nature, 17 (1), 48-52.

[14] J. Li Wenbo(2004). Research on Leisure Sports Consumption: An Interpretation of Culture and Sociology. Jiangxi Social Science, (9), 154-157.

[15] J. Wu Yigang(2003). Current situation and problems of leisure sports research at home and abroad. Journal of Shanghai Institute of Physical Education, 27 (3), 39-43.

[16] J. Tian Hui, Zhou Hong(2006). Leisure and leisure sports and their development trend in China. Sports Science, 26 (4), 67.

[17] J. Chen Yuzhong(2007). Future Trend of Leisure Sports Development in China. Journal of Shanghai Institute of Physical Education, 31 (1), 9.

Perceiving space through sound: mapping human movements into MIDI

Bernardo Breve, Stefano Cirillo, Domenico Desiato
Department of Computer Science
University of Salerno
84084 Fisciano (SA), Italy
{bbreve,scirillo,ddesiato}@unisa.it

Mariano Cuofano
MRes Architecture
Royal College of Art
South Kensington, London (UK)
mariano.cuofano@alumni.rca.ac.uk

Abstract

Gestural expressiveness plays a fundamental role in the interaction with people, environments, animals, things, and so on. Thus, several emerging application domains would exploit the interpretation of movements to support their critical designing processes. To this end, new forms to express the people's perceptions could help their interpretation, like in the case of music. This paper presents a novel algorithm for mapping human movements into MIDI music. The algorithm has been implemented in a system that integrates a module for real-time tracking of movements through a sample-based synthesizer using different types of filters to modulate frequencies. The system has been evaluated through a user study, in which several users have participated to a room experience, yielding significant results about their perceptions with respect to the environment they were immersed.

***Index terms*—** Movements Tracking, MIDI Sound, Synthesizer Sounds

1 Introduction

Gestures, movements, and body languages represent a common way through which it is possible to mark verbal communication. Indeed, they emphasise the language by adding significant characteristics useful for communication purposes. More specifically, it is possible to capture details that allow to associate a particular gesture with an emotion. In fact, by analysing the interaction between two persons, it is possible to understand the feelings produced during their communication through their gestures. In the state of the art, it is possible to find different tracking techniques for detecting human movements [30], and techniques to represent their semantics [1]. They have been mainly applied to scenarios in which it is necessary to support humans in the real-time interpretation of critical situations, such as in the context of video surveillance [8] and emergency management [28]. In the social context, an important role is played by the music. It is possible to define music as an art manifestation, since it consists of creating and producing sounds that are pleasant to the human ear. The music, in most cases, leads emotions in individuals who are providing and listening to it. Thus, also in this case, it is possible to associate a sentiment to a specific organisation of sounds. In fact, most of the time, people use music to regulate their emotions, this means that each individual through the listening of particular shades of sound modifies his/her emotional state or, more simply, a sound can create emotions such as happiness, sadness, gladness, and so on [11].

In this paper, we exploit the combination of human movements and sound synthesis techniques to associate sounds to movements in the space. More specifically, we propose a novel algorithm and a system for recognising human movements, and translating them into MIDI music, with a sample-based synthesizer, which uses different types of filters to modulate frequencies. The single frequency is used to associate a specific set of sounds to each movement captured in real-time. In particular, in order to quickly map spatial coordinates of people in sound, we defined a layer-based module, which is able to uniquely identify MIDI notes, and to manage their changes according to people's movements.

DOI reference number: 10.18293/DMSVIVA20-011

The paper is organised as follows. In Section 2, we describe recent works concerning movements tracking techniques and their application into several fields. In Section 3, we introduce some preliminary concepts useful to understand our proposal. In Section 4, we explain our methodology, whereas its validation is discussed in Section 5. Finally, conclusions and future research directions are provided in Section 6.

2 Related work

In this section, we describe approaches and methodologies defined in the literature concerning the mapping of gesture recognition [3, 12, 13]. In particular, we focus on different application domains in which the recognition of gestures plays an important role [21].

In the music field, interesting pioneering work is described in [20]. Authors present Conductor's Jacket, a wearable device interpreting physiological and gestural stimuli in order to apply them in a musical context. In fact, it uses sixteen sensors communicating with a musical software by collecting data over different reliable channels, also offering mcal-time graphical feedbacks to control the mapped gestures. Instead, in [27] authors present MATRIX (Multipurpose Array of Tactile Rods for Interactive eXpression), a musical interface for music amateurs and professionals. It permits the usage of hands to control music by exploiting a 3-dimensional interface allowing the manipulation of traditional musical instruments in conjunction with it. The MATRIX interface manipulates the parameters of a synthesis engine or effect algorithm in real-time, in response to the performer's expressive gestures. In [10], authors illustrate a real-time musical conducting gesture recognition system that supports music players in enhancing their performance. They used a single-depth camera to capture image inputs and to establish a real-time gesture recognition system. In the data mining field, one of the most relevant works is [26] in which the authors propose an innovative framework for progressive mining and querying of motion data, by also exploiting information extracted from data relationships [6].

Another application field concerns the application of gesture recognition applied to video surveillance. In [25], authors present a video surveillance framework for realtime multi-person tracking. It uses an adaptive background subtraction in order to identify foreground regions for catching users' movements. In [29], authors present an online approach to simultaneously detect 2D poses of multiple people in a video sequence. They exploit Part Affinity Field (PAF) representations designed for static images, and they propose an architecture that can encode Spatio-Temporal Affinity Fields (STAF) across a video sequence. In [32], authors present a novel multi-person tracking system for crowd counting and normal/abnormal events detection in indoor/outdoor surveillance environments. They use two challenging video surveillance datasets, such as PETS2009 and UMN crowd analysis datasets, to demonstrate the effectiveness of their proposed system, which achieved 88.7% and 95.5% of accuracy and detection rate, respectively. In [4] authors propose an algorithm for multi-person tracking in indoor surveillance systems based on a tracking-by-detection approach. They use Convolutional Neural Networks (CNNs) for detecting and tracking people. They also perform several experiments by tracking people in rapid-panic scenarios, achieving good performances in terms of classification accuracy. In [24], authors define a lightweight tracking algorithm named Kerman (Kernelized Kalman filter), which is a decision tree based hybrid Kernelized Correlation Filter (KCF) algorithm for human object tracking.

Finally, in the medical field, gesture recognition is applied to monitor diseases. Authors in [22] present an innovative approach to preterm-infants' limb pose estimation. They exploit spatio-temporal information to determine and track limb joint position from depth videos with high reliability. Instead, in [34], authors illustrate Ultigesture wristband, a hardware/software platform for gesture recognition and remote control. Ultigesture wristband offers full open API for third party research and application development.

3 Preliminaries

In this section, we provide preliminary notions to allow for a better understanding of the concepts underlying the proposed system.

The Musical Instrument Digital Interface (MIDI) is a standard de facto for enabling communication among digital musical instruments and processors of digital music, such as personal computers and sequencers [23]. A MIDI message carries data concerning the peculiarity of a certain sound, such as the vibrato, the tremolo, and so on. However, a MIDI file does not contain any actual waveform generated by the notes of a played musical instrument. Instead, it is a collection of data informing on how a type of sound can be simulated by the digital music processor, which is then responsible for playing the sound by retrieving the representation of the simulated music instrument assigned to those notes from its memory [14].

More specifically, a MIDI message is transmitted over 16 different channels in groups of 8 bits, each of which can be of two types called status byte and data byte, respectively. The latter defines the value associated with the message, whereas the former is used to specify the type of message sent. The bytes are distinguished through the first bit, that is, a status byte begins with the bit 1, whereas a data byte begins with the bit 0. A MIDI message is usually composed of a status byte, followed by one or two data bytes, and can belong to one of the two following categories: channel

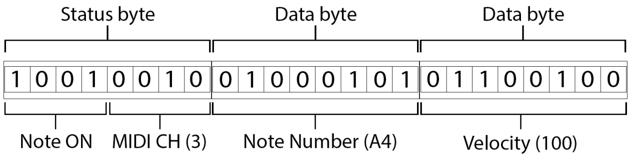


Figure 1: An example of MIDI message.

message or system message. Channel messages are sent to single channels and contain information about the musical performance; system messages are aimed at the MIDI system responsible for coordinating the succession of sounds.

Figure 1 shows an example of a MIDI message. In particular, it shows a channel message, composed of one status byte, which is mandatory for every type of message that has been sent and two data byte. The status byte contains information about the operation to be performed and the channel involved with that particular operation, which in this case of Figure 1 is ”play a note on the third channel”. The following two data bytes contain information about the note to be played and the velocity to be applied. The velocity is a value though which it is possible to emulate the amount of force exerted on the key. Alternatively, it can also describe the width of the output or the tone of the sound.

MIDI control changes, also known as or associated with MIDI Controllers or Control Changes, are MIDI messages conveying positional information related to performance control devices such as wheels, sliders, pedals, switches, and other control-oriented devices [31]. This type of information can be used to control a variety of functions, such as vibrato depth, brightness, and many other parameters. MIDI controller change messages are 128, and each of these has a control number and a control value parameter.

MIDI notes and control change are used in our system to map the spatial coordinates to sounds and to give a different effect to each sound. Details about the proposed system and on how it uses MIDI messages are provided in the next section.

4 A system for mapping human movements into sounds

We propose a new system named PIANO (maPpIng humAn movements iNto sOunds), which recognizes the movements of people in space and translates them into sounds.

PIANO is a modular system offering standalone modules, designed to be fast and easy to use, relying on non expensive hardware devices, available on the market. More specifically, PIANO consists of several modules, the first of which is a digital tracker relying on an infrared camera to distinguish the participants’ bodies and associate them to a virtual figure. When a person enters the room, movements are mapped into sounds. These are read by Cubase, a sequencer located on a platform and emitted as sound.

4.1 The Object/Human Detection Algorithm

Tracking movements of a person in a live video is not always a simple operation. Thus, several problems must be considered, such as the variability. In fact, a position detection algorithm must be able to trace the object considering the enormous variations in the appearance and position of the objects. Therefore, there are several variables that a tracking algorithm must take into consideration: the point of view, the position, the lighting conditions, the quality or the occlusions of the images. Furthermore, the large number of successive frames that may belong to a live or recorded video can make tracking activities particularly difficult. A tracking system of a moving object must be able not to lose the position of the object in subsequent frames. Thus, an algorithm must be able to consider the variation of all these variables and to perform real-time activities on any environment. These types of operations are even more complicated when needing to identify and/or track groups of objects/people.

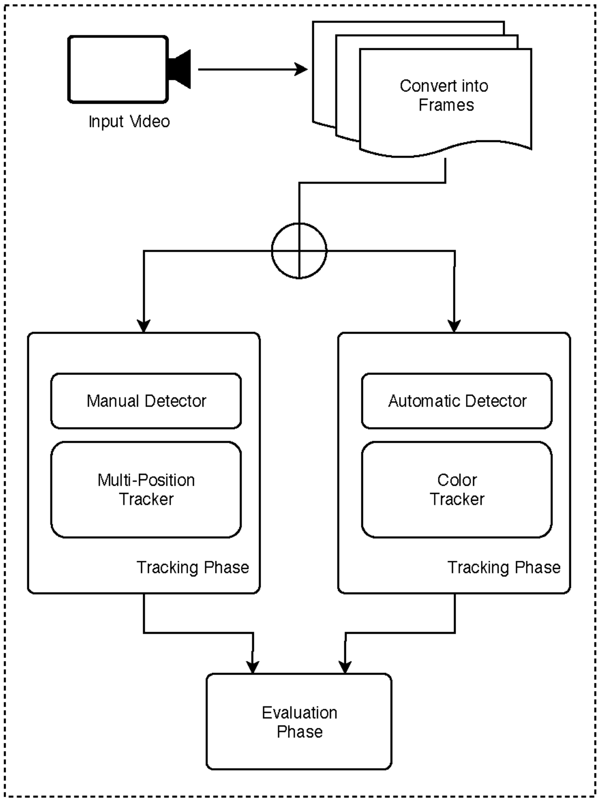


Figure 2: Flowchart of the object detection algorithm.

Within PIANO we designed a tracking module capable of using any camera for recognizing people. The proposed algorithm determines the position of one or many people. Starting from the video, it maps the position of each person in spatial coordinates.

Figure 2 shows the three phases of the PIANO tracking algorithm: detection of people, tracking of these subjects, and evaluation of the tracking results to describe semantic events and latent phenomena. In the first phase, the tracking algorithm reads the video as input and converts it to subsequent frames. In particular, the algorithm can work on static video sources or real-time input streams. This guarantees the high adaptability of PIANO to any type of situation and the repetition of itself several times, even after a live video recording. After selecting the input, PIANO allows choosing the tracking mode, i.e. manual or automatic.

Manual tracking enables a full control of the system and the arbitrary choice of whether or not to select a person to track. The user draws the selection rectangle of the objects or people s/he wants to monitor. In particular, each selection is a Region of Interest (ROI) to be tracked. Each selected ROI is associated with a person, and it will be independently recognised by the others. As shown in Figure 3(a), when a ROI is selected, the algorithm creates a virtual figure. Starting from this figure, the algorithm calculates the coordinates and defines the centre of the ROI. This is the reference point of the person's movement. Using different rectangles, it is possible to keep track of many people at the same time, so that PIANO can track separate music for each person. The algorithm can use different types of trackers: Boosting [15], Multiple Instance Learning (MIL) [2], Kernelized Correlation Filters (KCF) [16], TLD[18], Median-Flow [17], Minimum Output Sum of Squared Error (MOSSE) [5], and Discriminative Correlation Filter with Channel and Spatial Reliability (CSRT) [19]. Support for different types of trackers enables the algorithm to adapt itself to all tracking situations, regardless of hardware characteristics.

Automatic tracking has been created to track people in a dark room using an infrared camera. Unlike manual tracking, there is no need to select people or objects to track. The proposed tracker receives an input dictionary of colours, and after selecting the input, the individual frames are analysed. Each frame is resized, blurred, and converted to the HSV (Hue, Saturation, Value) colour space. Next, for each colour defined in the dictionary, the tracker checks objects in each frame. Then, it constructs a mask for the colour from the dictionary, and performs a series of implicit dilations and erosions in order to remove any small blobs left in the mask. Successively, it selects one of the colours in the dictionary and defines the contours of the figure, showing it within the frames. The shapes are defined as circles, and the centre is the reference point for the person's movements. Since this type of tracker has been created for shooting in the dark, two-colour scales have been defined in the algorithm, red and grey.

Algorithm 1 MAPPING_SPACE_MIDI

INPUT: The coordinates H_x, H_y of the person in the space extracted from the bounding box; The height F_h and the width F_w of the frame; The number of rectangles on X-axis of the frame S_w; The number of rectangles on Y-axis of the frame S_h

OUTPUT: The MIDI note value to be played

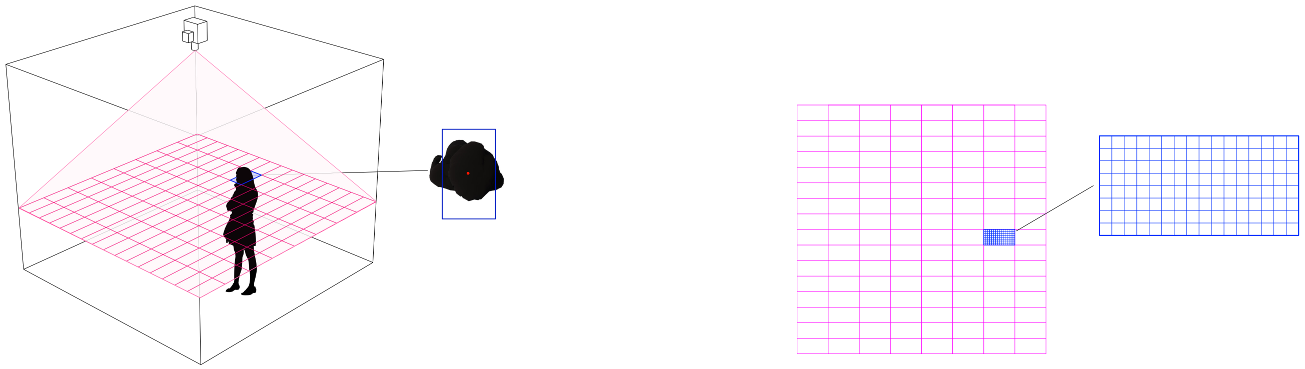
```
$R^1_{height} \leftarrow \frac{F_{height}}{S_h}$
$R^1_{width} \leftarrow \frac{F_{width}}{S_w}$
$G^1_x \leftarrow \frac{H_y}{R^1_{height}}$
$G^1_y \leftarrow \frac{H_x}{R^1_{width}}$
$MIDI_{note} \leftarrow S_h * G^1_x + G^1_y$
return $MIDI_{note}$
```

As shown in Figure 2, after the tracking phase, it is necessary to evaluate the output of each tracker. The evaluation phase defined in PIANO is described in the next section.

4.2 Mapping space to sound

PIANO uses two different functional systems: digital and analog. The analog system is based on a technology inspired by the "Theremin". It consists of an antenna capable of feeling the proximity of the electric voltage of the human body, and of translating it into an analog audio input, which is pre-amplified by an integrated circuit before issuing the sound through an active speaker. The three objects interacting with each other generate a magnetic field producing a white background noise. Bodies moving through this field alternate this with the magnetic status, and as a consequence, alter the white background noise and sound.

The digital system works on the actual relation of the bodies in the room, associating them to a sound only treated with their position in the room. In particular, we developed an algorithm to transform the movement of one or more people into sounds. This operation is particularly complex so that it is necessary to use a mapping methodology that is fast and precise. To this end, we have developed an approach to quickly map spatial coordinates to sound. After the video stream is read, the proposed methodology creates different virtual layers for each frame: the first layer defines the midi note that must be played, whereas the second layer represents the control change for each note. Figure 3 shows how the frame is divided. In particular, Figure 3(a) shows a representation of a frame from a video stream. We divided each frame into 128 different equal parts, corresponding to



(a)

(b)

Figure 3: Virtual grids in the environment.

the number of existing midi notes. To do this, PIANO automatically calculates the size of each section.

Formally, let F be the frame extracted from the video, F$_{width}$ and F$_{height}$ the height and the width of F, respectively, both expressed in pixel, S$_h$ and S$_w$ the maximum number of rectangles on F$_{width}$ and F$_{height}$, i.e. S$_w$ = 16 and S$_h$ = 8. Then, starting from the centre of the bounding box identified on the frame at the position (H$_x$, H$_y$), it is necessary to use the following formulas to define the midi note to play:

$$G_x^1 = \frac{H_x \cdot S_w}{F_{width}} \tag{1}$$

$$G_y^1 = \frac{H_y \cdot S_h}{F_{height}} \tag{2}$$

Moreover, Figure 3(b) shows the structure of the second layer. In particular, each rectangle of the first layer is in turn divided into 128 rectangles of equal size, which represent the control change. When the centre of the bounding box plays a midi note in the first layer, the control change is simultaneously set in the second layer. Starting from the formulas 1 and 2, we can define the type of control change when playing the midi note:

$$G_x^2 = \frac{(H_x - G_x^1 \cdot \frac{F_{width}}{S_w})}{2 \cdot S_w} \tag{3}$$

$$G_y^2 = \frac{(H_y - G_y^1 \cdot \frac{F_{height}}{S_h})}{2 \cdot S_h} \tag{4}$$

Using the formulas 1-4, we obtain a pair of coordinates (G_x^1, G_y^1) and (G_x^2, G_y^2) that allow us to identify the midi note and the control change in F to play.

Algorithm 1 shows the mapping function from spatial coordinates to MIDI notes. Given the size of each frame expressed in pixel, the number of rectangles in which the space has been divided, and the centre of the bounding box defined in the frame to track one person, the algorithm defines the height and the width of each rectangle in the first grid (lines 1-2), and then calculates the rectangle in which the person moves (lines 3-4). As said above, each rectangle on the grid corresponds to a different note, so that it is possible to define the MIDI note to be played by only considering the coordinates of the rectangle (line 5).

Algorithm 2 MAPPING_SPACE_CONTROLCHANGE

INPUT: The coordinates H$_x$, H$_y$ of the person in the space extracted from the bounding box; The height F$_h$ and the width F$_w$ of the frame; The number of rectangles on X-axis of the frame S$_w$; The number of rectangles on Y-axis of the frame S$_h$
OUTPUT: The control change value to be reproduced

```
1: $R^2_{height} \leftarrow \frac{R^1_{height}}{S_h}$
2: $R^2_{width} \leftarrow \frac{R^1_{width}}{S_w}$
3: $G^2_x \leftarrow \frac{(H_y - G^1_x \cdot R^1_{height})}{R^2_{height}}$
4: $G^2_y \leftarrow \frac{(H_x - G^1_y \cdot R^1_{width})}{R^2_{width}}$
5: $ControlChange \leftarrow S_h * G^2_x + G^2_y$
6: return $ControlChange$
```

Algorithm 2 provides the control change value to be played. Similarly to Algorithm 1, it starts by considering the dimensions of each rectangle in the first grid, aiming to define the height and the width of the rectangles in the second grid (lines 1-2). As said above, the control change can take 128 possible values, each defined from a single rectan-

Figure 4: Interaction Room.

gle in the second grid. Thus, the algorithm calculates the control change to be reproduced by considering the size of the frame and the position of the people (lines 3-5).

5 Interaction Room

In order to present the case study, it is necessary to provide some details about the perception of space aiming to understand how the user perceives space through unconventional instruments such as music.

Starting from the definition provided in [33], we consider the space as a room. The term *room* is derived from the archaic English *rum*, which is similar to the German world *Raum* (space). This, in turn, refers to the Latin derivation — *rus* — which can be translated as the act of making space. In this definition, the room is not described as a space associated to a specific role in the context where it is located, like for example, as a component of an apartment. Rather, room here is meant as a volume where phenomena take place. Our attempt is to understand the meaning of movements in the architecture domain and how to design the space based on them. The analysis starts by moving a critique to the existenzminimum, assuming it to be a course of predetermination paradigms of architectural design. The final outcomes of these models produce a monolithic city, able to grow and to allow for urban colonisation, triggering a wild world attitude to the gentrification of an urban environment. Starting from these ideas, we have been designed a specific environment called *Interaction Room*.

As spatial speculation, the Interaction Room represents a format of feedbacking architecture. It aims to offer an opportunity, a phenomenon, and a sound, to show how it is possible to generate sounds and emotions by using the space and the relative position each one takes within, coherently with the presence of other individuals.

Sounds are generated into two different modes. The objects mentioned before will have the analog system integrated into them, which creates an unpredictable sound relation with each possible room layout. However, digital sounds are generated by using the methodology defined in Section 4.2. The intention is to merge these two solutions to have more a homogeneous and deep sound.

The experience aimed to learn how to be in a space together, generating a sound that depends on how people perceive the relation with space. Through artistic and digital approaches, we have given people the opportunity to discuss their space recognition experience.

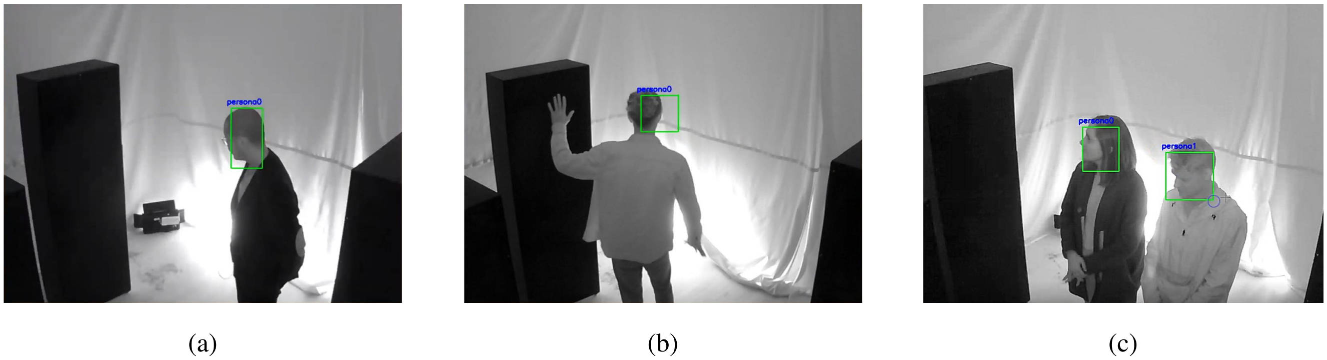
5.1 Experimental Design

The Interaction Room has been designed as cubic structure with dimensions of 3 meters (Figure 4). Space is interrupted by the presence of three objects, which simulate limitations, edges, and walls, that the participants could use to describe their own layout (Figure 4(b)). In particular, these structures are parallelepipedic objects having dimensions of 1.7 x 0.7 x 0.3 meters, and a weighing less than 10 kg.

PIANO has been executed on a computer with an Intel i9-9900k 3.6GHz 8-core CPU with 32 GB of RAM. The system interacted with an infrared camera with a resolution of 3 megapixels and an external sound card to reproduce the sounds. The latter has been connected to four 500-watt audio speakers located in the four corners of the room.

The evaluation session has been performed by involving people of different ages in supervised experience. In particular, people had different education levels, including high school diploma, master degree, and PhD. Moreover, some of them were architects and surveyors with considerable past experience.

The overall experience was based on the production of sounds. The idea was to link the sounds produced by the movements of participants to a specific position in the room, so as to relate it to a specific sound frequency (a note). The experience consists of accessing into the room, which ap-



(a)

(b)

(c)

Figure 5: An example of multi-position tracker.

pears as dark volume. In this room, people were able to face other participants only as silhouettes, and the obscure environment did not enable them to distinguish proper figures and images. Shadows are generated by a tiny light located on the edges of the room.

The aim of this experiment is to achieve pure perceptive feedbacks from participants, without triggering an "over-automatisation" process in adapting to space.

5.2 Experimental Results

As the most frequent answer, people that took part in this experience were describing the capability of manipulating the space, and interacting with the sound. Describing direct feedback connected with participants' emotions, the room was able to manifest an interactive space, reproducing any movements in the form of sound.

From another point of view, it could be considered as a proper act of space production. It is clear how the environment perceived was not related to the tectonic of the room itself and the volume was not related to predictable habits. It is therefore clear how the act of making space was strictly related to a phenomenon [33].

Most of the participants claimed to have the impression of manipulation on the density of the room. This was possible because of the background white noise produced by the interference of the three objects. There was clearly a relation with the quantity of background noise and the quality of the space. Passing through this imaginary plan inhabitants were altering a system with their presence. What the objects were able to transmit is a clear form of feedback. It is revealing that through a perceptive operation it was possible to generate presence in relation to the environment. This elementary, primitive, generation of space represents the opportunity for creating architecture based on dynamic feelings and perceptions. This architecture is nor a tailored made space, neither the application of an efficient program, based on common functions and habits, rather it is strictly connected to the individual in that given moment. In other words, it could be assumed that this architecture was just a representation of those individuals in the form of space.

6 Conclusion

We proposed an algorithm and a system to recognize human movements in the space, aiming to translate them into sounds. The system has been used in the architectural domain, aiming to provide useful insights for gathering people's perceptions with respect to the surrounding space.

In the future, we would like to use the system into different application domains, such as video surveillance, and evaluate its performances. This type of system could be applied to define customized alarms in order to simply recognize the intents of possible malicious actions through the sounds. To this end, we would like to define a novel translation model in order to enable the possibility to map and distinguish high-level critical scenarios. Moreover, we would like to further extend our technique to exploit multimedia dependencies [9], since these can be nowadays automatically extracted from multimedia databases [7].

References

[1] P. L. Albacete, S. K. Chang, and G. Polese. *Iconic language design for people with significant speech and multiple impairments*, pages 12–32. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[2] B. Babenko, M.-H. Yang, and S. Belongie. Robust object tracking with online multiple instance learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(8):1619–1632, 2011.

[3] A. D. Bagdanov, A. Del Bimbo, L. Seidenari, and L. Usai. Real-time hand status recognition from rgb-d imagery. In *Proc. of International Conference on Pattern Recognition (ICPR '12)*, page 2456–2459, 2012.

[4] R. Bohush and I. Zakharava. Robust person tracking algorithm based on convolutional neural network for indoor video surveillance systems. In *International Conference on Pattern Recognition and Information Processing*, pages 289–300. Springer, 2019.

[5] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui. Visual object tracking using adaptive correlation filters. In *2010 IEEE computer society conference on computer vision and pattern recognition*, pages 2544–2550. IEEE, 2010.

[6] L. Caruccio, S. Cirillo, V. Deufemia, and G. Polese. Incremental discovery of functional dependencies with a bitvector algorithm. In M. Mecella, G. Amato, and C. Gennaro, editors, *Proceedings of the 27th Italian Symposium on Advanced Database Systems*, volume 2400 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2019.

[7] L. Caruccio, V. Deufemia, and G. Polese. Evolutionary mining of relaxed dependencies from big data collections. In *Proceedings of the 7th International Conference on Web Intelligence, Mining and Semantics (WIMS)*, pages 5:1–5:10. ACM, 2017.

[8] 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.

[9] S. Chang, V. Deufemia, G. Polese, and M. Vacca. A normalization framework for multimedia databases. *IEEE Transactions on Knowledge and Data Engineering*, 19(12):1666–1679, 2007.

[10] F. Chin-Shyurng, S.-E. Lee, and M.-L. Wu. Real-time musical conducting gesture recognition based on a dynamic time warping classifier using a single-depth camera. *Applied Sciences*, 9(3):528, 2019.

[11] T. Cook, A. R. Roy, and K. M. Welker. Music as an emotion regulation strategy: An examination of genres of music and their roles in emotion regulation. *Psychology of Music*, 47(1):144–154, 2019.

[12] G. Costagliola, V. Deufemia, and M. Risi. A multi-layer parsing strategy for on-line recognition of hand-drawn diagrams. In *Proc. of IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 103–110, 2006.

[13] V. Deufemia, M. Risi, and G. Tortora. Sketched symbol recognition using latent-dynamic conditional random fields and distance-based clustering. *Pattern Recognit.*, 47(3):1159–1171, 2014.

[14] M. Flam. Musical instrument digital interface with speech capability, 2001. US Patent 6,191,349.

[15] H. Grabner and H. Bischof. On-line boosting and vision. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06)*, pages 260–267. IEEE, 2006.

[16] J. Henriques, R. Caseiro, P. Martins, and J. Batista. High-speed tracking with kernelized correlation filters. *IEEE transactions on pattern analysis and machine intelligence*, 37(3):583–596, 2014.

[17] Z. Kalal, K. Mikolajczyk, and J. Matas. Forward-backward error: Automatic detection of tracking failures. In *Proc. of International Conference on Pattern Recognition*, pages 2756–2759, 2010.

[18] Z. Kalal, K. Mikolajczyk, and J. Matas. Tracking-learning-detection. *IEEE Transactions on pattern analysis and machine intelligence*, 34(7):1409–1422, 2011.

[19] A. Lukezic, T. Vojir, L. Cehovin Zajc, J. Matas, and M. Kristan. Discriminative correlation filter tracker with channel and spatialreliability. *International Journal of Computer Vision*, 126(7):671–688, 2018.

[20] T. Marrin and R. Picard. The "conductor's jacket": A device for recording expressive musical gestures. In *Proceedings of International Computer Music Conference*, 1998.

[21] S. Mitra and T. Acharya. Gesture recognition: A survey. *IEEE Transactions on Systems man and Cybernetics C*, 37(3):311–324, 2007.

[22] S. Moccia, L. Migliorelli, V. Carnielli, and E. Frontoni. Preterm infants' pose estimation with spatio-temporal features. *IEEE Transactions on Biomedical Engineering*, 2019.

[23] F. R. Moore. The dysfunctions of midi. *Computer music journal*, 12(1):19–28, 1988.

[24] S. Y. Nikouei, Y. Chen, S. Song, and T. R. Faughnan. Kerman: A hybrid lightweight tracking algorithm to enable smart surveillance as an edge service. In *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 1–6. IEEE, 2019.

[25] W. Niu, J. Long, D. Han, and Y. Wang. Human activity detection and recognition for video surveillance. In *Proc. of International Conference on Multimedia and Expo*, pages 719–722, 2004.

[26] R. Ortale, E. Ritacco, N. Pelekis, R. Trasarti, G. Costa, F. Giannotti, G. Manco, C. Renso, and Y. Theodoridis. The daedalus framework: progressive querying and mining of movement data. In *Proc. of International Conf. on Advances in Geographic Information Systems*, pages 1–4, 2008.

[27] D. Overholt. The MATRIX: A novel controller for musical expression. In *New Interfaces for Musical Expression*, pages 38–41, 2001.

[28] L. Paolino, M. Romano, M. Sebillo, and G. Vitiello. Supporting the on-site emergency management through a visualisation technique for mobile devices. *Journal of Location Based Services*, 4(3-4):222–239, 2010.

[29] Y. Raaj, H. Idrees, G. Hidalgo, and Y. Sheikh. Efficient online multi-person 2D pose tracking with recurrent spatio-temporal affinity fields. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 4620–4628, 2019.

[30] F. Remondino. Tracking of human movements in image space. *Technical Report at IGP-ETH Zurich*, 2001.

[31] F. Rumsey. *MIDI systems and control*. Butterworth-Heinemann, 1994.

[32] A. Shehzed, A. Jalal, and K. Kim. Multi-person tracking in smart surveillance system for crowd counting and normal/abnormal events detection. In *2019 International Conference on Applied and Engineering Mathematics (ICAEM)*, pages 163–168. IEEE, 2019.

[33] M. Tattara and P. V. Aureli. *The Room of One's Own. The Architecture of the (Private) Room*. Black Square, 2017.

[34] H. Zhao, S. Wang, G. Zhou, and D. Zhang. Ultigesture: A wristband-based platform for continuous gesture control in healthcare. *Smart Health*, 11:45–65, 2019.

Visualizing Visual Parser Executions

Gennaro Costagliola, Mattia De Rosa
Dipartimento di Informatica, University of Salerno, Fisciano, SA, Italy
{gencos, matderosa}@unisa.it

Mark Minas
Universität der Bundeswehr München, Neubiberg, Germany
mark.minas@unibw.de

Abstract

Parsing of visual structures like diagrams and graphs is more complicated than parsing strings. This is so because visual structures are inherently more complex than strings, because visual grammars are more difficult to write than string grammars, and because the algorithms for parsing visual structures are usually more complicated than for parsing strings. The developer of a visual parser, therefore, needs more tool support than a developer of a string parser. In fact, developing and debugging a visual parser without proper visualization of the parsing process is very challenging.

This paper describes a visualization approach that arose from this need. Its main focus is on the interaction of the developer with the visualization tool in order to explore the execution process of the parser. It has evolved from experiences with developing and debugging parsers by applying different visual parsing methods. In order to better describe it we introduce a concrete example.

Keywords: *visual parsing, graph parsing, parser visualization.*

1. Introduction

Parsing strings with respect to a grammar is well-known and well-understood since some 50 years [18]. Every compiler of a (textual) programming language uses a string parser to analyze the syntactic structure of its input and to control its translation into other formats, in particular machine language or intermediate representations. Parser generators make building string parsers a simple task [26, 23, 14]. A significant part of the research on visual languages in the last 25 years [5] has focused on the study of their semantics and the possibility to specify them in a formal way, also for use in visual programming languages. Researchers have therefore tried to analyze them using approaches similar to those used for strings, but with less success than in the string domain [22]. One reason is the obvious fact that parsing strings is much easier than visual parsing. All established string parsing techniques take advantage of the linear structure of strings, in particular of substrings of the input. This is apparent for top-down and bottom-up parsers using LL and LR parsing, which process the input string from left to right, i.e., analyze prefixes of the input. Even table-based parsers like Cocke-Younger-Kasami parsers [33] depend on the strings' linear structure although they do not process input strings from left to right. Instead, they construct nonterminals for arbitrary substrings (and not just prefixes) of the input string, starting with substrings of length one and eventually for the entire input, if it is valid. Substrings are easily represented by just two numbers, e.g., start and length. Parsing errors can be easily communicated to the user that way. This task is more complicated when a visual parser fails; it must then visualize those parts of the input that have already been processed when the error occurred. The situation becomes even more complicated when an implementor develops a parser, even when using a parser generator. Understanding the flow of execution of a visual parser and its consequent validation without proper visualization of the parser's progress and its data structures is then tedious, other than very challenging.

The use of visual structures like graphs have recently gained some importance in the field of natural language processing (NLP) where the meaning of sentences is represented by graphs [11]. Their syntactic structure is defined by grammars, and (graph) parsers are used for analyzing them. Several approaches are used in this context, and some tools have emerged [10]. However, similar to the situation in earlier VL research, appropriate parser visualization techniques and tools are yet missing.

This paper extends [6] by describing the analysis-based approach that led to the development of prototypical tools

DOI reference number: 10.18293/DMSVIVA20-013

for properly visualizing the execution of visual parsers. We first describe the need for such tools which became apparent when the authors developed visual parsers. Based on these needs and experiences with first visualization prototypes, we reconsidered the problem and identified the primary use cases. We generalized these results by deriving requirements that parser visualization tools should fulfill in order to effectively and efficiently support the realization of visual parsers. A general parser visualization architecture has been developed from these requirements and realized in two independent prototypical tools.

The rest of the paper is structured as follows: we start by presenting our running example describing the basic concepts of Visual Generalized LR (VGLR) parsing and the semantic representation of natural language sentences in Section 2, then, in Section 3 we outline the primary use cases of a visualization tool for visual parsers and derive the visualization requirements in Section 4. In Section 5 we describe the proposed parser visualization architecture used to implement our prototypes and related work in Section 6. Section 7 concludes the paper.

2. The application example: VGLR parsing and NLP

The proposed visualization and exploration approach can be used with many different types of visual parsers, either top-down or bottom-up. In fact, it has already been used with two different VGLR parser approaches based on *[contextual] hyperedge replacement grammars* ([C]HRGs) [13, 9, 24] and *extended positional grammars* [7, 8].

In this paper, the running example is based on a VGLR parser built for a CHRG from the domain of natural language processing. In the following, the basic concepts of VGLR parsing are given, followed by a brief description of the natural language representations used as input sentences.

The Generalized LR (GLR) parsing algorithm [32, 28] extends the well-known LR parsing algorithm [18] to ambiguous string grammars and Visual GLR (VGLR) parsing algorithm extends GLR parsing to the case of graph languages.

We assume that readers are familiar with the standard LR parsing algorithm, which analyzes an input string from left to right, maintains a stack of states through the shift/reduce actions, and produces a single parse tree if the input string is valid. In order to handle nondeterminism, a generalized parser works on multiple stacks at the same time and produces multiple parse trees, one for each interpretation. A GLR parser is able to do this efficiently by storing the stacks in a so-called *graph-structured stack* (GSS; see Fig. 5 for an example) and packing the resulting parse trees in a *packed parse forest* (see Fig. 6 for an example). A GSS is a particular directed acyclic graph representing each individual stack as a path from some top-most state to the unique initial state. There are three main operations that can be performed on a GSS: splitting, combining and local ambiguity packing. Each time the parser faces two conflicting actions (shift/reduce or reduce/reduce) the current stack top is *split* to accommodate two new branches in the graph. Whenever a new stack top, resulting from a shift action, happens to be equal to an already existing stack top, they are *combined* into one node. A local ambiguity packing is the operation of merging two equal branches. This happens when the same fragment of the input can be reduced to the same nonterminal in different ways. The goal of these operations is to maximize the sharing of the common parts of the multiple stacks. In fact, working on the GSS instead of on a set of complete copies of different stacks does not only save space, but also time: instead of repeating the same operations on separated stacks that have common parts, the parser has to perform them only once.

By following the same idea, a packed parse forest stores the many parse trees produced by analyzing an ambiguous input by sharing all of their common subtrees. The relation between a GSS and its corresponding packed parse forest is given by the fact that each edge in the GSS corresponds to a vertex of the forest and the subtree rooted in it. In this way, a vertex (called packed vertex) in the forest may be root of distinct subtrees corresponding to the same shared GSS branch (due to local ambiguity packing).

When dealing with non-linear sequences such as graphs or other types of diagrams, a *visual GLR* (VGLR) parser must also deal with the fact that there is no a priori reading sequence of visual tokens. This may force a VGLR parser to pursue different reading sequences in parallel while it performs the search process. This has consequences as follows:

- Each parse stack corresponds to a specific subset of visual tokens that have been read already. Hence, the parser must store, for each stack separately, which visual tokens have been read.

  Sets of stacks are stored as a GSS like in GLR parsers. Each GSS node corresponds to a state. Additionally, each GSS node keeps track of the set of visual tokens that have been read so far. Note that GSS nodes may be shared only if both their states and their sets of visual tokens coincide.

- VGLR parsers cannot process their sets of stacks in rounds. When a stack is obtained by executing a *shift* action, the parser must not wait until the same visual token has been read in all the other stacks (as done for GLR parsing); they may read other tokens first. As a consequence, a VGLR parser needs different strategies from that adopted by GLR parsing to control the order in which stacks are processed. Strategies are beyond

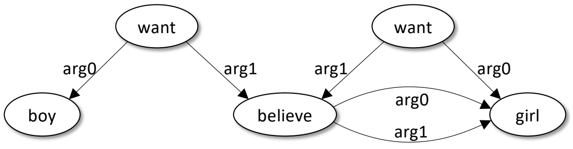


Figure 1. AMR graph for "*The boy wants the girl to believe in herself and this is what the girl wants, too.*"

the scope of this paper, but visualizations of parser executions must make them explicit to the user.

In order to describe our visualization approach we apply the VGLR parser to the graph language of *abstract meaning representations* (AMR). AMRs have been proposed as a semantic representation of English sentences. Each AMR is a directed graph with labelled nodes and edges, which represent concepts and their relations, respectively. For instance, the semantics of the sentence "*The boy wants the girl to believe in herself and this is what the girl wants, too*", taken from [11], can be represented by the AMR shown in Fig. 1. A description of AMRs is beyond the scope of this paper; details can be found in [1, 20], and a corpus of AMRs is described in [3].

3. Use Cases

The authors have realized several visual parsers and visual parser generator [7, 8, 9, 13, 24]. Developing VGLR parsers is challenging because one has to coordinate several non-trivial data structures like the input, the GSS, and the parse forest. It turned out that it is almost impossible to develop such parsers without proper visualization of all these data structures. In the following, we briefly describe the primary uses cases of visualization tools when developing VGLR parsers. We then elaborate on the requirements derived from these use cases in Section 4.

The main use case is inspecting the different data structures after a parsing error occurred. The error can either be due to an invalid input (syntax error) or to an error in the parser code or both. Only by inspecting the input, the GSS, and the parser forest at the time of the error and running the last steps that led to the error can help the parser implementor to detect the possible causes. Since the number of nodes in the GSS and the parse forest can be very high (sometimes more than 100) and many concurrent stack tops and tree roots may be present, it would be almost impossible to accomplish the task without proper visualization. Even though the number of nodes may make the data structure visualizations difficult to read, running back and forth the last actions preceding the error can help the implementor to individuate and zoom on specific nodes of the structures being modified.

Visualizing the input and those input tokens that have already been inspected by the parser is another use case. In contrast to textual input, visual input has in general no self-evident ordering of input tokens. The parser has, rather, to identify a correct parsing sequence, and it turned out that visualizing this information makes debugging VGLR parsers a lot easier.

Yet another use case is for the parser user to see all the possible syntactic interpretations of its input by looking at the parse forest visualization. Each of the contained parse trees corresponds to individual stacks within the GSS and possibly different parsing sequences through the input. It turned out to be manageable for the implementor to comprehend the correspondence of all these data structures with the help of proper visualizations.

4. Parser Visualization Requirements

The use cases outlined in the previous section allowed to derive requirements on a parser visualizer (in the following called just *visualizer*). It must essentially provide visual representations of the parser's data structures that change over time during its execution. In order to allow the implementor to analyze and validate these data structures, she must be able to stop the execution, to continue it, to watch it in single-step mode, to go backwards in time, i.e., to retrace the parser's steps, etc. In other words, the visualizer must provide control over the parser execution quite similar to a program debugger.

4.1. Granularity levels of execution control

A well known feature of program debuggers is that they allow to run a program step by step on quite different levels of granularity. At the lowest level, they allow one to stop after each statement or instruction. On a higher level, they can "step over" a procedure call, i.e., they stop automatically when the procedure call terminates. The user can choose the appropriate level of granularity freely. Similarly, a parser execution visualizer should be able to show parser executions at different levels of granularity and to allow one to easily pass from one level to the other. Furthermore, in order to better fit the mental model that a parser/language implementor has of the parser execution, each level should be programmable, in the sense that she should be allowed to define the operations included in a level. In the following, we use the action and step granularity levels.

In general, the action level visualizes the results of the execution of each parser action. As mentioned above, to better represent the user needs, a parser action may be split in more refined actions: as an example, the reduce action of a bottom-up parser may be split in "deletion" of the reduced path and "addition" of the new goto state. This gives a better understanding on how the reduction process is performed and on which states. As a further example, not represented here, in the case of visual parsers that use relations to navigate the input, the shift action may further be split in "move the cursor" to point to the next input symbol and "shift the pointed token".

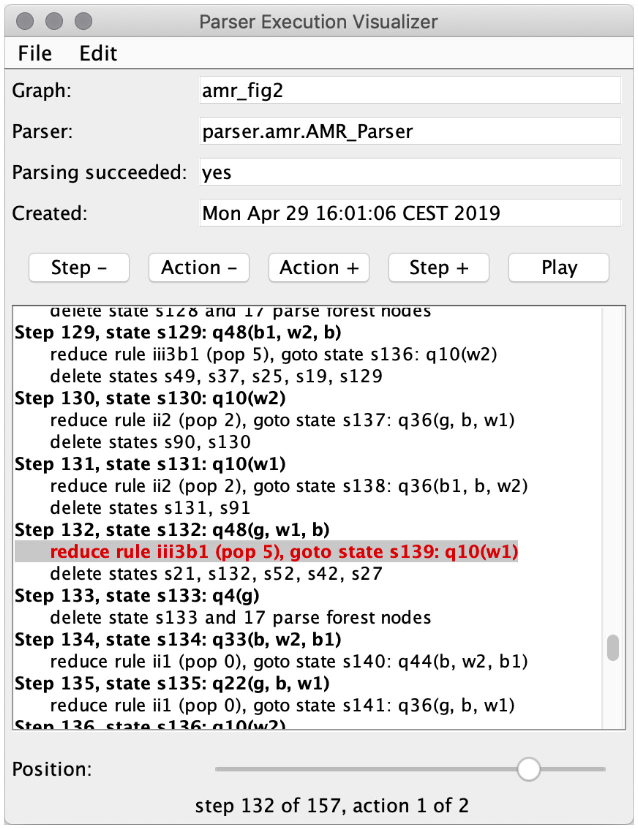
A step is usually the highest granularity level and it can vary depending on the particular strategy of execution adopted by the parser being implemented. It is usually used to synchronize the actions of the parser. As an example, in Tomita's parser the execution is synchronized by the input tokens. As a consequence, each step includes either all the shift actions to visit a new input token or all the possible reductions that can be applied on a token. In our case, a step includes the actions to be executed on a particular top state.

4.2. Execution control

A program debugger lets the user control the execution by showing her the program source code in which she can set breakpoints and in which the line of code is highlighted where the program has been stopped. The visualizer should offer a similar view that shows the sequence of parser operations on the selected level of granularity. Fig. 2 shows an example used in our prototype visualizing the execution of the AMR graph parser analyzing the graph shown in Fig. 1. The lines in bold-face are steps where the parser operates on a specific node of the GSS, which corresponds to a state, before it continues working on the next node of the GSS. The sequence of actions composing a step are shown in normal font below a bold-face step. The action where the parser execution has been stopped is highlighted in red. Here it is a reduce action that pops five states off the stack, starting at the current state $s_{132} : q_{48}(q, w_1, b)$ and which will then perform a goto to state $s_{139} : q_{10}(w_1)$. Again, details of the parser and its states are beyond the scope of the paper. Note that the action that follows below the highlighted line will then remove the states that have been popped off.

Note also that such a view is in fact different from the program source code of a program debugger in the sense that it does not show the parser program, but rather a trace of the parser execution on a specific level of granularity. In fact, it must be an a posteriori trace taken from a previous parser execution. Otherwise, the view could not show the steps and actions following the one where the execution has been stopped. We require (see Sect. 5) that the parser stores the trace in a log file, which is read later by the visualizer, a technique which is also well-known from the analysis of parallel programs [21].

The user of the visualizer must be able to run the parser, or rather retrace its steps from the log. She must be able to execute the parser actions step by step on different granularity levels, and must be able to go back and forth in time, realized by buttons in Fig. 2. Pushing it shall start a continuous animation visualizing the progress of the parser in the data views discussed in the next section. Furthermore, clicking on an action in the trace should fast-forward (or fast-backward, resp.) the visualization to this action.[1]



Graph: amr_fig2

Parser: parser.amr.AMR_Parser

Parsing succeeded: yes

Created: Mon Apr 29 16:01:06 CEST 2019

```
    delete state s128 and 17 parse forest nodes
Step 129, state s129: q48(b1, w2, b)
    reduce rule iii3b1 (pop 5), goto state s136: q10(w2)
    delete states s49, s37, s25, s19, s129
Step 130, state s130: q10(w2)
    reduce rule ii2 (pop 2), goto state s137: q36(g, b, w1)
    delete states s90, s130
Step 131, state s131: q10(w1)
    reduce rule ii2 (pop 2), goto state s138: q36(b1, b, w2)
    delete states s131, s91
Step 132, state s132: q48(g, w1, b)
    reduce rule iii3b1 (pop 5), goto state s139: q10(w1)
    delete states s21, s132, s52, s42, s27
Step 133, state s133: q4(g)
    delete state s133 and 17 parse forest nodes
Step 134, state s134: q33(b, w2, b1)
    reduce rule ii1 (pop 0), goto state s140: q44(b, w2, b1)
Step 135, state s135: q22(g, b, w1)
    reduce rule ii1 (pop 0), goto state s141: q36(g, b, w1)
Step 136, state s136: q10(w2)
```

Position:

step 132 of 157, action 1 of 2

Figure 2. Execution control of the parser execution visualizer prototype.

4.3. Data views

In order to allow the user to understand what is going on in a parser execution, different aspects of the parser must be visualized. Apparently, its main data structures must be shown in a way that match the user's mental model. For our running example using a VGLR parser, the visualizer must at least show the current GSS, the parse forest, and the parser input, as discussed in the following. These data structures are connected. For instance, leaves of the parse forest correspond to visual tokens of the parser input, and each edge in the GSS refers to a parse forest node. Visu-

[1] Two screencasts that demonstrate the user interactions with the prototype can be seen at [37].

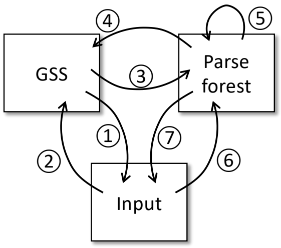


Figure 3. Different data views. Arrows indicate visual feedback triggered by user interaction described in the text.

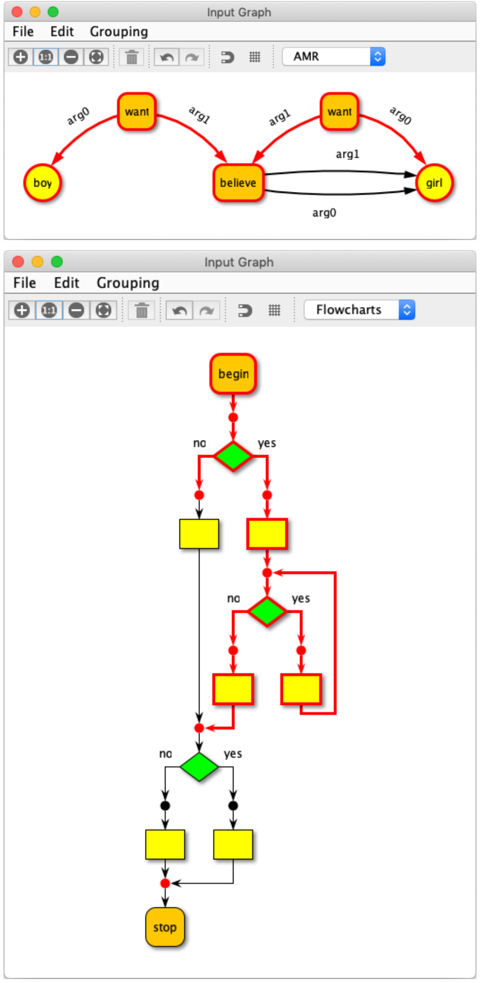


Figure 4. Input view of an AMR graph and a flowchart.

alizing all these aspects and their interconnections in a single view would produce just clutter. Instead, we suggest to provide separate views and to visualize interconnections between them by means of user interaction and visual feedback [30]. Fig. 3 symbolizes the three views that the visualizer at least should provide; arrows represent user interactions in one view and the corresponding visual feedback in a different or the same view. Arrow 7, for instance, includes highlighting a visual token in the input view when the corresponding parse forest leaf is clicked.

In the following, we first describe the requirements on the three different data views and afterwards their interconnection by means of visual feedback triggered by user interaction.

4.3.1. Input view

The parser must appropriately choose in which sequence it reads the visual tokens of the input, and it may be forced to choose different reading sequences in the same execution if the input is ambiguous, e.g., the AMR graph shown in Fig. 1. The visualizer, therefore, must provide an *input view* that shows the parser input and indicates which of its visual tokens have already been read in the current state of the execution, and which have not been read yet.

Visualizing the parser input is more complicated than showing the input of a string parser: Whereas a string can be simply shown as text, there is no uniform representation for all visual languages that can be analyzed by a visual parser. The input view, therefore, must be highly customizable, its visual representation should match the representation of the visual language. Fig. 4 shows two screenshots of our prototype. The upper screenshot depicts the AMR graph of Fig. 1, the lower one a flowchart in the process of being parsed. Visual tokens that have already been read by the parser are highlighted in red. More details are described below.

Of course, different parsers are used to analyze AMR graphs and flowcharts, but they use a common input format with less information than the concrete diagram. In general, the input format may in fact be just a kind of graph (as in our case), which does not contain any information on the layout of its visual tokens. The input view, therefore, must be customizable in the way how elements of the common input format shall be represented. Furthermore, it shall offer some automatic and manual layout functionality like in standard visual editors, but without the ability to modify the parser input semantically. Our prototype allows to switch between different customizations using the combo box in the window toolbar.

4.3.2. GSS view

The visualizer must show the main data structure of the parser. For Earley-parser or a Cocke-Younger-Kasami-based parser, it would be a table used for dynamic programming. In our VGLR example, however, this is the GSS, which shall be shown as a plain DAG as in Fig. 5.

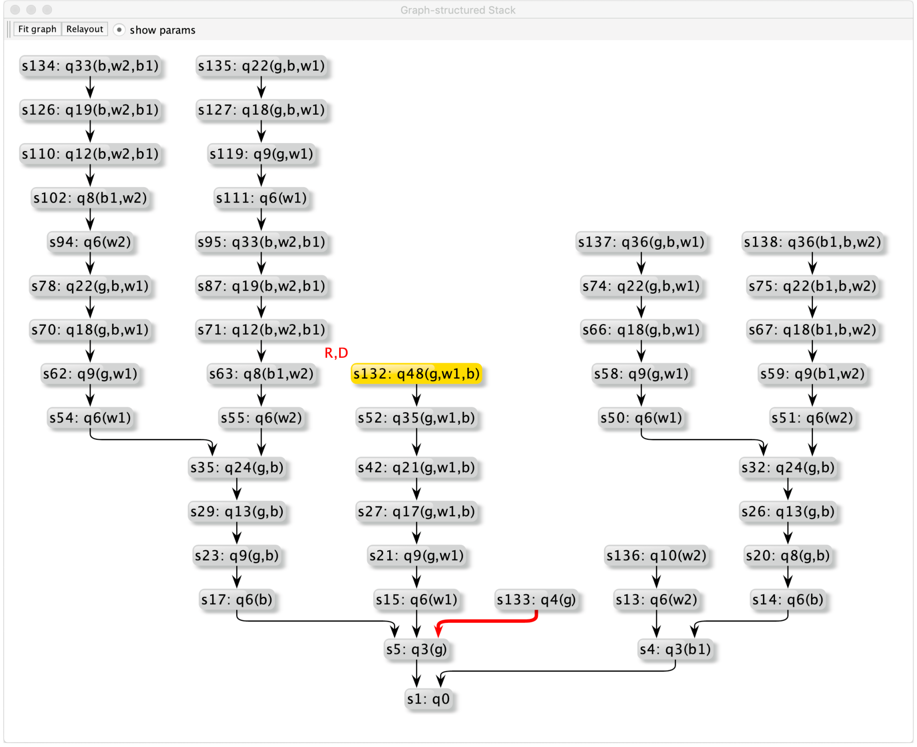


Figure 5. GSS view corresponding to Fig. 2.

The GSS changes with the parser execution, i.e., GSS nodes and edges are added and deleted. Moreover, GSSs can grow rather big as Fig. 5 shows, which can make it difficult for the user to follow these modifications. The GSS view shall provide some general features that allow the user to easily recognize any modification. First, the layout of the modified GSS should be computed from the old one incrementally in order to preserve the user's mental map [27]. Second, changes to the layout should not happen abruptly. Instead, they should be visualized in an animated way so that the user has time to see the changes happen. And finally, the view shall indicate the GSS node ("working node") that is currently processed by the parser, i.e., where changes happen, shown in orange in Fig. 5. Note that this GSS node is the same as the one indicated in the trace view of execution control (Fig. 2).

Moreover, the GSS view should also add further information about the current working node, which informs the user about the changes that will happen to this node. Possible changes are triggered by the actions within the step processing this node, i.e., shift, reduce, accept, and delete as described before. The initial letters of these actions are used here to mark the working node, here R and D, which corresponds to the actions of step 132 as shown in the log view (Fig. 2).

The highlighted edge in Fig. 5 is a visual feedback after selecting a parse forest node described in Sec. 4.3.4.

4.3.3. Parse forest view

The parse forest represents the syntactic structure of the parse input processed so far during the parser execution. It is the final parse forest, i.e., a packed form of the set of all parse trees of the input, if the parser terminates successfully, and it is empty if the parser fails. The parse forest is usually a DAG if sub-trees are shared in order to save space. And in general, it consists of several unconnected components as long as the input has not been analyzed completely yet. In fact, each edge of the GSS refers to a unique parse forest node. This interconnection of GSS and parse forest shall be communicated to the user by means of user interaction and visual feedback described in the next section.

Fig. 6 shows a part (note the scrollbar at the bottom) of the parse forest in the execution state shown in Fig. 2. Terminal parse forest nodes are drawn in yellow, nonterminal ones in green. Note also the nodes *b*, *g*, and *b*1 drawn in faded red; they are so-called contextual nodes which are a specific feature of contextual hyperedge replacement grammars used in our example (see Sec. 2 and [11]). Their incoming and outgoing arrows represent where these nodes have been created and where they are used as contextual nodes in the parse forest. Again, details are beyond the scope of this paper.

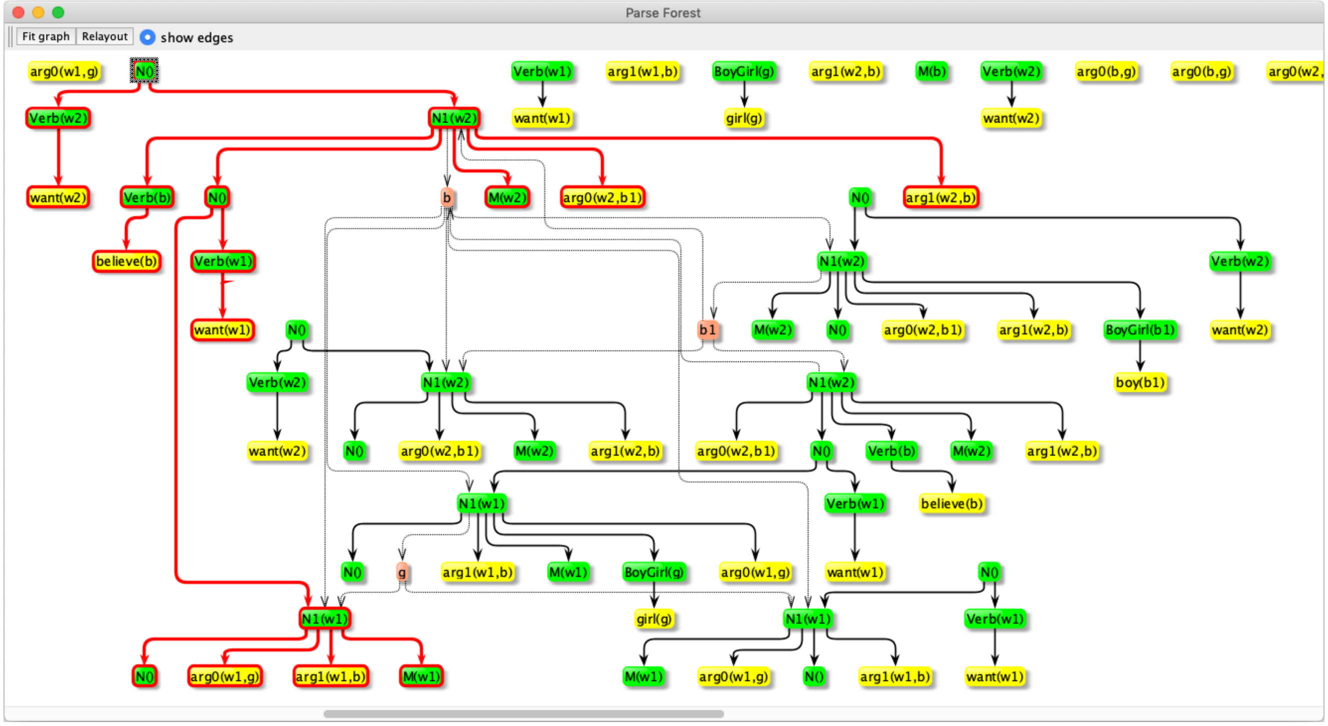


Figure 6. Parse forest created by the VGLR parser corresponding to Fig. 2.

Like the GSS, the parse forest changes with the parser execution. New parse forest nodes are added when edges are added to the GSS, possibly adding edges to child nodes in the parse forest (see the screencasts at [37]), and entire subgraphs of the parse forest may be deleted when a parse fails. The latter does not necessarily mean that the whole parser execution fails; it may be just one dead end in the search carried out by the parser. Like the GSS view and in order to prevent user confusion, the parse forest view must provide an automatic layout that allows to preserve the user's mental map.

Fig. 7 shows the complete parse forest after the parser eventually accepted the AMR graph of Fig. 1. It represents in fact two parse trees and uses local ambiguity packing (see Sec. 2) in order to save space: two different nonterminals $N_1(b)$ with different sub-DAGs are contained within a gray node, whose parent nonterminal $N(b)$ can select either of the two nodes $N_1(b)$ as a child, resulting in two different parse trees. One parse tree corresponds to the sentence "*The boy wants the girl to believe in herself and this is what the girl wants, too*", the other to the semantically equivalent "*The girl wants to believe in herself, and the boy wants the girl to believe in herself, too.*"

The nodes and edges of the parse forests of Figures 6 and 7 highlighted in red are visual feedbacks after selecting the top-most highlighted node $N()$ and $N_1(b)$, respectively, and is described in the next section.

4.3.4. User interaction and visual feedback

We are now going to describe the visual feedback triggered by selecting components in one of the views. The numbers of the following items correspond to the numbers used in Fig. 3.

① Whenever a GSS node becomes the current working node or if the user selects a GSS node in the GSS view, all visual tokens that have already been read in this parser state shall be highlighted. The nodes and edges of the AMR graph in Fig. 4 drawn in red have been read in the current state $s_{132} : q_{48}(q, w_1, b)$ in Fig. 5.

② When the user selects a visual token in the input view, all GSS nodes that have already read this token shall be highlighted in the GSS view.

③ Each edge of the GSS refers to a top-most node of the parse forest. If the user selects an edge in the GSS view, this node and its complete sub-DAG shall be highlighted in the parse forest view.

④ is in fact the opposite of ③: When a user selects a top-most node in the parse forest node, the edge of the GSS that refers to this parse forest node is selected in the GSS view. The edge in Fig. 5 highlighted in red is in fact the visual feedback for the selection of the top-most node $N()$ in the parse forest view (Fig. 6).

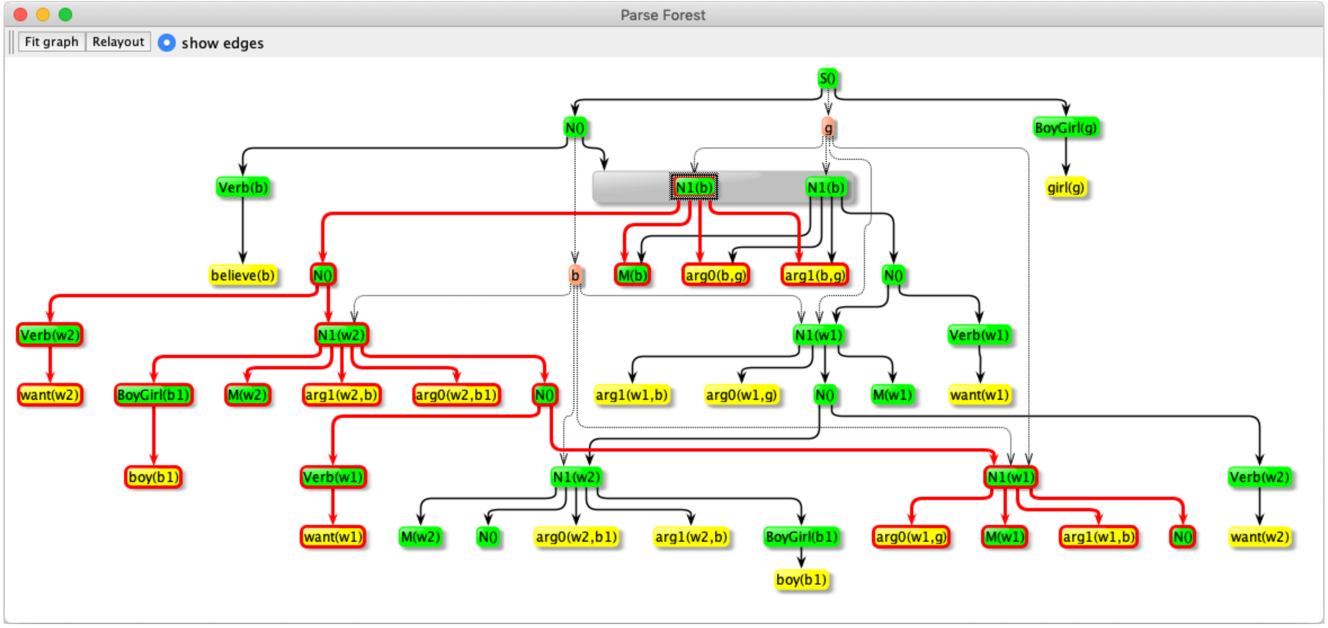


Figure 7. Final parse forest of the AMR graph shown in Fig. 1.

⑤ And whenever a node is selected in the parse forest node, all nodes and edges of its entire sub-DAG are highlighted, too. That is the reason for the other highlighted nodes and edges in Figures 6 and 7 selecting $N()$ or $N_1(b)$, respectively. This does in fact not visualize a connection between different data-views, but allows the user to recognize more easily which nodes of a parse forest belong to a sub-DAG. Otherwise, it would be rather tedious to see which parse forest nodes belong to either of the two nodes $N_1(b)$ within the gray packed node in Fig. 7.

⑥ When the user selects a visual token in the input view, the corresponding terminal parse forest nodes shall be highlighted in the parse forest view. Note that, according to ②, selecting a visual token also highlights all GSS nodes that have read the visual token.

⑦ When a parse forest node is selected, which highlights its entire sub-DAG according to ⑤, all visual tokens that correspond to terminal parse forest nodes in this sub-DAG shall be highlighted in the input view.

5. Parser Visualization Architecture

Fig. 8 shows the proposed architecture of the parser execution visualizer as it has also been realized in our prototype. Orange rectangles, yellow rounded rectangles, and green parallelograms represent data, processes, and UI components, respectively, arrows represent data flow.

We suggest that the visualizer does not visualize the parser during its execution. We rather suggest that the parser is instrumented so that it writes a trace of its actions into a log file, which is read later by the visualizer. This approach allows the visualizer to easily go back and forth in time. Moreover, the same parser execution can be visualized arbitrarily often, even if the parser runs nondeterministically [21]. We also assume that the log file contains the encoded parser input so that it can be visualized in the *input view*.

The *Log Reader* reads the file and internally stores the *parse trace* which consists of the *parser input* and the sequence of *parser events*, e.g., actions like shift and reduce. They are shown to the user in the *input view* (see Fig. 4) and the *log view* (see Fig. 2), respectively.

The *execution control process* reads the parse events forward and backward, controlled by the *execution control user interface* (see Fig. 2). This process keeps track of the current event, which is also highlighted in the *log view*. And if the user selects an action in the *log view*, execution control fast-forwards to the corresponding parser state. It maintains the parser state by means of the *parser data structures* based on the parsed events that have happened since the beginning of the parse trace. There are no uniform parser data structures, they rather depend on the specific parser type. In our example, they consist of the GSS and the parse forest. These data structures are visualized in the corresponding *data structure views* (here GSS view, Fig. 5, and parse forest view, Figs. 6 and 7) using some layouting facility. The *user interaction & feedback handler* reacts to pointing and

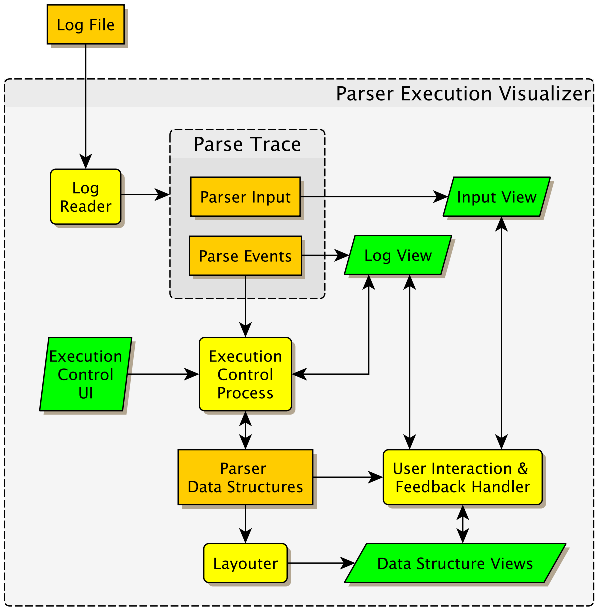


Figure 8. Parser execution visualizer architecture.

clicking in the different views and triggers visual feedback as described in Sec. 4.3.4.

6. Related Work

Data structure and algorithm visualizations have been studied for more than 35 years [31, 4, 17], and now many web resources exist implementing visualizations and animations of almost all the most common data structures and algorithms based on them, respectively [34, 35, 36]. However, lately, the research in this field has shrunk considerably. Most of the important papers are in the range from 1980 to 2000 and the applications have been basically two: visual debugging [25] and teaching and learning [29]. Among the still currently developed tools is JSAV [17], a JavaScript algorithm visualization library that is meant to support the development of general algorithm visualizations for online learning material.

Parser visualization tools either visualize the process of generating a parser from a grammar like LLparse and LRparse [2]. Or they visualize parser execution like PAT [12, 15] which has been used for the visualization and statistical comparison of various GLR parsers. Among other tools we can cite [16] for visualizing lexical generation processes and [19] that is an educational tool for visualizing compiling techniques based on deterministic parsers.

Our prototypes also visualize parser execution and, hence, are most closely related to the latter category. However, we are not aware of any tool that also allows to visualize the execution of visual parsers.

7. Conclusions

In this paper we have illustrated the requirements and the architecture of a parser visualizer system while using an example from the field of Natural Language Processing. In particular, we have discussed several visualization techniques that have proven useful in practice and then generalized the results by elicitation of visualization requirements that parser visualization tools should fulfill in order to effectively and efficiently support the realization of visual parsers.

A system based on the proposed specification is able to support a VGLR parser/language implementor at various levels of granularity and, as a side effect, it may also be used to help teachers to visualize the bottom-up parsing of a specific input when applied to simple grammars.

Two instances of the prototypical parser visualizer based on different VGLR parsing approaches exist following the guidelines and architecture presented in this paper. Even though the two instances have been specialized to the specific approach the needs to gain the maximum insight in the parser execution behavior resulted to be the same.[2] Because of the problem complexity, the use of a parser visualizer gave a huge contribution to the development of each phase of the two VGLRs: automatic generation of the VGLR parser from a grammar specification, execution of the generated parser and parsing of several languages including the example shown here. Another outcome of our parser visualizer architecture is that it allows for the visual comparison of the execution of different parsers obtained either by modified versions of the same approach or by different approaches, for the analysis of their differences and/or similarities at different level of granularity and for the discovery of specific parser behaviors.

References

[1] L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, K. Knight, P. Koehn, M. Palmer, and N. Schneider. Abstract meaning representation for sembanking. In *Proc. 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria, Aug. 2013. Assoc. for Computational Linguistics.

[2] S. A. Blythe, M. C. James, and S. H. Rodger. Llparse and lrparse: Visual and interactive tools for parsing. *SIGCSE Bull.*, 26(1):208–212, Mar. 1994.

[2] For completeness, some screenshots of the other instance not described here can be found at [37].

[3] F. Braune, D. Bauer, and K. Knight. Mapping between English strings and reentrant semantic graphs. In *Proc. of the Ninth Int. Conf. on Language Resources and Evaluation (LREC'14)*, pages 4493–4498, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA).

[4] M. H. Brown and R. Sedgewick. Techniques for algorithm animation. *IEEE Software*, 2(01):28–39, Jan. 1985.

[5] G. Costagliola, M. De Rosa, V. Fuccella, and S. Perna. Visual languages: A graphical review. *Information Visualization*, 17(4):335–350, 2018.

[6] G. Costagliola, M. De Rosa, and M. Minas. Visual parsing and parser visualization. In *2019 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 243 –247. IEEE Computer Society, Oct 2019.

[7] G. Costagliola, V. Deufemia, and G. Polese. A framework for modeling and implementing visual notations with applications to software engineering. *ACM Trans. Softw. Eng. Methodol.*, 13(4):431–487, Oct. 2004.

[8] G. Costagliola, V. Deufemia, G. Polese, and M. Risi. Building syntax-aware editors for visual languages. *J. Visual Lang. Comput.*, 16(6):508 – 540, 2005. Selected papers from Visual Languages and Formal Methods 2004 (VLFM '04).

[9] F. Drewes, B. Hoffmann, and M. Minas. Extending predictive shift-reduce parsing to contextual hyperedge replacement grammars. In E. Guerra and F. Orejas, editors, *Graph Transformation: 12th Int. Conf., ICGT 2019, Held as Part of STAF 2019, Proc.*, volume 11629 of *LNCS*, 2019.

[10] F. Drewes, B. Hoffmann, and M. Minas. Formalization and correctness of predictive shift-reduce parsers for graph grammars based on hyperedge replacement. *J. Log. Algebr. Methods*, 104:303–341, April 2019. Preprint available at arXiv:1812.11927 [cs.FL].

[11] F. Drewes and A. Jonsson. Contextual hyperedge replacement grammars for abstract meaning representations. In *13th Intl. Workshop on Tree-Adjoining Grammar and Related Formalisms (TAG+13)*, pages 102–111, 2017.

[12] G. R. Economopoulos. *Generalized LR parsing algorithms*. PhD thesis, Royal Holloway, Univ. of London, UK, 2006.

[13] B. Hoffmann and M. Minas. Generalized predictive shift-reduce parsing for hyperedge replacement graph grammars. In C. Martín-Vide, A. Okhotin, and D. Shapira, editors, *Language and Automata Theory and Applications, 13th Int. Conf., LATA 2019, Proc.*, volume 11417 of *LNCS*, pages 233–245, 2019.

[14] S. C. Johnson et al. *Yacc: Yet another compiler-compiler*, volume 32. Bell Laboratories Murray Hill, NJ, 1975.

[15] A. Johnstone, E. Scott, and G. Economopoulos. The grammar tool box: A case study comparing GLR parsing algorithms. *Electronic Notes in Theoretical Computer Science*, 110:97–113, 12 2004.

[16] A. Jorgensen, G. R. Economopoulos, and B. Fischer. VLex: visualizing a lexical analyzer generator - tool demonstration. In *Language Descriptions, Tools and Applications, LDTA 2011. Proc.*, page 12, 2011.

[17] V. Karavirta and C. A. Shaffer. JSAV: The JavaScript algorithm visualization library. In *Proc. of the 18th ACM Conf. on Innovation and Technology in Computer Science Education*, ITiCSE '13, page 159–164, 2013.

[18] D. E. Knuth. On the translation of languages from left to right. *Information and Control*, 8(6):607 – 639, 1965.

[19] N. Krebs and L. Schmitz. Jaccie: a Java-based compiler-compiler for generating, visualizing and debugging compiler components. *Sci. Comput. Program.*, 79:101–115, 2014.

[20] I. Langkilde and K. Knight. Generation that exploits corpus-based statistical knowledge. In *Proc. 36th Annual Meeting of the Association for Computational Linguistics and 17th Int. Conf. on Computational Linguistics, Volume 1*, pages 704–710. Assoc. for Computational Linguistics, Aug. 1998.

[21] T. J. J. LeBlanc and J. M. Mellor-Crummey. Debugging parallel programs with instant replay. *IEEE Trans. Comput.*, 36(4):471–482, Apr. 1987.

[22] K. Marriott and B. Meyer, editors. *Visual Language Theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1998.

[23] S. McPeak and G. C. Necula. Elkhound: A fast, practical GLR parser generator. In E. Duesterwald, editor, *Compiler Construction*, pages 73–88. Springer Berlin Heidelberg, 2004.

[24] M. Minas. Speeding up Generalized PSR parsers by memoization techniques. In R. Echahed and D. Plump, editors, *Proc. 10th Int. Workshop on Graph Computation Models (GCM 2019)*, volume 309 of *Electronic Proceedings in Theoretical Computer Science*, pages 71–86, 2019.

[25] S. Mukherjea and J. T. Stasko. Toward visual debugging: Integrating algorithm animation capabilities within a source level debugger. *ACM Trans. Comput.-Hum. Interact.*, 1:215–244, 1994.

[26] T. J. Parr and R. W. Quong. ANTLR: A predicated-LL(k) parser generator. *Software: Practice and Experience*, 25(7):789–810, 1995.

[27] H. C. Purchase and A. Samra. Extremes are better: Investigating mental map preservation in dynamic graphs. In G. Stapleton, J. Howse, and J. Lee, editors, *Diagrammatic Representation and Inference*, volume 5223 of *LNCS*, pages 60–73, Berlin, Heidelberg, 2008.

[28] E. Scott and A. Johnstone. Right nulled GLR parsers. *ACM Trans. Program. Lang. Syst.*, 28:577–618, 07 2006.

[29] C. A. Shaffer, M. L. Cooper, A. J. D. Alon, M. Akbar, M. Stewart, S. Ponce, and S. H. Edwards. Algorithm visualization: The state of the field. *Trans. Comput. Educ.*, 10(3):9:1–9:22, Aug. 2010.

[30] B. Shneiderman, C. Plaisant, M. Cohen, S. Jacobs, N. Elmqvist, and N. Diakopoulos. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson, 6th edition, 2016.

[31] J. T. Stasko. Simplifying algorithm animation with Tango. In *Proceedings of the 1990 IEEE Workshop on Visual Languages*, pages 1–6, Oct 1990.

[32] M. Tomita, editor. *Generalized LR Parsing*, volume 1. Springer US, 1991.

[33] D. H. Younger. Recognition and parsing of context-free languages in time n^3. *Information and Control*, 10(2):189 – 208, 1967.

[34] `http://www.algomation.com`.

[35] `https://visualgo.net`.

[36] `https://www.cs.usfca.edu/~galles/visualization/`.

[37] `http://cluelab.di.unisa.it/parser_execution_visualizer/`.

An Intrusion Detection Framework for Non-expert Users

Bernardo Breve, Stefano Cirillo, Vincenzo Deufemia
Department of Computer Science
University of Salerno
84084 Fisciano (SA), Italy
{bbreve,scirillo,deufemia}@unisa.it

Abstract

The wide spreading of the Internet and its integration in everyday objects lead to the born of a whole interconnected world, based on different devices communicating each other and performing operation remotely activated. Among all these devices, smart voice assistants are gaining particular attention thanks to their ease of use, allowing users to comfortably deploy commands for controlling other devices. The wide diffusion of these devices and their simplicity of use lead to that category of users with little or no knowledge, to interact with complex systems without being perfectly aware of the risks they are exposed to. For this reason, the common network defenses and monitoring systems are too complex to be used by non-expert users. This paper presents a framework for intrusion detection specifically designed to be configurable and usable by every category of users. The system will be based on specific automatic behaviors so as to minimize human intervention. Moreover, it will include several visual interfaces aiming to simplify the user interaction with the framework, allowing him/her to properly configure and run an Intrusion Detection System (IDS). The use of voice assistants as a communication channel between the user and the system will provide an additional contribution in order to improve the overall user experience.

Index terms— Intrusion detection system, Voice assistant, Computer networks, Human-computer interfaces

1 Introduction

The spread of the Internet in all socio-economic sectors has led to the need to educate people on the use of this tool. The main goal is to create a society able to exploit the power of the Internet with the aim of improving daily life. For this reason, the Internet of Things (IoT) has spread in the last few years. The IoT has become one of the most important technologies of this century, allowing us to connect each type of object, e.g. kitchen appliances, cars, thermostats, baby monitors, to the internet in order to establish continuous communication between people, processes, and things.

A large number of organizations benefit from the use of these types of devices in their business processes. In sectors as automotive [6, 22], public sector [33] and healthcare [3, 13], IoT has led to a real revolution. The use of intelligent systems that take advantage of IoT devices has improved safety in cars, has speeded up the time for the rescue of a person, or simply increased the productivity of the public administration. However, this has led to the birth of new security issues, since it is necessary to ensure that no one can interfere with their operations. Thus, the field of information security has become vitally important to the safety and economic well-being. The personal information of each person and what is connected to has enormous value and therefore must be preserved. To this end, new secure and safe information systems have been provided, by using firewalls, intrusion detection and prevention systems, authentication, and other hardware and software solutions.

The reasons that may induce an attack can been grouped into three categories: access information, alter information or render a system unusable [4]. These have led to the birth of intrusion detection systems (IDS), which provide tools for monitoring suspicious activities on the network. IDS can be defined as an alarm that monitors the network and reports intrusion to users. Over the years, a large number of IDSs [8] have been developed, which were later extended through the use of data mining tools [24], data relationships [10, 36], and machine learning approaches [16]. In general, we can consider several desirable characteristics for an IDS. In particular, an IDS should be run continuously without human supervision, and be fault-tolerant and survivable. Moreover, it should impose minimal overhead and be easily adapted to a specific network to observe the anomaly in network traffic.

Although there are a large number of IDSs, one of the main problems is to install and configure these systems to monitor a specific network. In fact, most of the existing

DOI reference number: 10.18293/DMSVIVA20-014

IDSs are used by domain experts who are able to carry out complex configurations and installations. However, most of the people subject to these types of attacks do not have skills for configuring these systems. Furthermore, being the IDSs similar to alarms, it is required to customize the devices and the notification methods of these systems. Although several visual languages and visualization techniques have been proposed to support the management of security issues in the context of Web applications [9, 11], it is necessary to use technologies that are familiar to a large part of users.

Voice assistants are increasingly popular and functional, and they have become a routine part of everyday life for many people. Initially, these assistants did not bring big news. But their developers knew they still had a bright future because, like any other technology, voice recognition needed some more time to evolve. In fact, over time, a large number of features have been developed that take advantage of artificial intelligence (AI) and machine learning for allowing users to use complex tools through their voice.

For these reasons, in this paper, we propose a new framework that allows non-expert users to install and configure an IDS on the network. In particular, we propose a new modular architecture with an easy-to-use user interface to customize an IDS. Moreover, we propose an innovative module for interacting with Alexa for executing and monitoring the status of an IDS via voice commands.

The paper is organised as follows. Section 2 describes recent work concerning IDSs and tools to support non-expert users in the use of systems that require a deep domain knowledge. Section 3 provides an overview of the different types of IDSs. Section 4 presents the architecture of the proposed framework by describing the underlying components. Section 5 briefly discusses the most crucial aspects for non-expert users interacting with an IDS and how we are focusing our efforts to meet their needs. Section 6 concludes the paper by presenting our conclusions and future directions.

2 Related Works

The goal of the proposed solution is to allow a large number of users to use IDS systems despite their inexperience. In fact, the recently proposed IDSs do not consider how they can be used by non-expert users.

In [18] the authors proposed a new IDS that takes advantage of a deep learning approach based on the self-taught learning technique (STL). The authors explicitly declare that this tool is targeted at network administrators, and not at common users.

One of the most relevant work has been presented in [20]. Here the authors proposed an innovative network IDS to combat increasingly sophisticated network attacks. It takes advantage of a Hidden Naïve Bayes multiclass classifier to create an effective IDS that outperforms one of the most used IDS based on SVM [2]. The goal of both researches were to create efficient tools without considering if non-expert users are capable to use them or not.

In this work we introduce an innovative framework to support non-expert users in the use of different types of IDSs. Through this framework we can increase the user's awareness of what is happening on their network. This topic has been widely discussed by researchers, who have created several tools and user interfaces to increase interaction between users and systems.

Recently, one of the studies that addressed the problem described above is [14]. The authors developed a visual interface for non-expert users, in order to increase awareness of what happens on the network during daily browsing sessions. Indeed, they have shown that most users are unaware of the type of information are exchanged during the browsing sessions and need specific tools to solve this problem. The proposal has been deeply evaluated and analyzed from the point of view of the user experience in [9].

In [7], authors have compared 13 different visualization tools for network analysis aiming to outline their pros and cons. They have used qualitative coding as part of their research design in order to select several metrics to evaluate the advantages and disadvantages of the analyzed tools. Their primary purpose is to increase the security analyst's situational awareness without considering the final users.

In literature, different tools have been proposed to facilitate the use of IDSs by non-expert users. In [28], authors have defined a simplified sound-assistant that mitigates the sound in order to uniquely notify network attacks. In particular, they exploit distinctive sounds for each attack scenario so that the users easily identify the type of attack. The proposed tool could be integrated within network IDSs.

Other research on human-computer interfaces for supporting IDS has focused on bimodal applications, visual and sound, to notify network intrusions. For example, in [25] the authors introduce immersive spatial audio representations of network events that exploit 3D visual representations for interactive auto-stereoscopic.

3 Overview of IDSs

In this section we provide a general overview of Intrusion Detection Systems (IDSs). The latter can be classified as Network-based IDSs (NIDSs) and Host-based IDSs (HIDSs) [35].

A NIDS is designed to observe the passing traffic on the entire subnet, detecting attacks that involve all the devices on the network [32]. A HIDS, instead, runs on an independent device of the network and monitor the incoming and outcoming packets from the device, looking for the presence of any malicious activity occurring to the system the HIDS is attached to. Alongside these two categories of IDSs, there also exist hybrid solutions, which combine the information provided by both the network and single devices' feedback to develop a complete view over the network system [34].

IDSs can also be classified based on the methodology used for the identification of intrusions. In this case, they can be classified in two main categories: Signature-based detection (SD), Anomaly-based detection (AD) [5, 19].

3.1 Signature-based detection

Signature-based detection systems relies on a set of specific patterns (or strings), called *signatures*, representing the network traffic trend during a certain type of known malicious attack. Following the analysis of the network traffic performed by the system, the extracted data is compared with the stored signatures. When a correspondence is found, this would immediately lead to a report of an attack in progress, and it would provide details about the type of attack and its characteristics. The comparison with the stored signatures can be performed with different techniques such as data mining [30], design patterns [15, 26], or involving both centralised and distributed components [17]. The main advantage of this intrusion detection methodology relies on the simplicity of the identification process, which mainly involves an extrapolation process followed by a comparison [23]. This method is ideal for identifying known attacks and obtaining simultaneously all the details about them. On the contrary, identifying an attack based on a restricted set of patterns limits the number of intrusions that could be recognized. Also, the knowledge base requires to be kept continuously updated, which is often a difficult and time-consuming process [31].

3.2 Anomaly-based detection

The anomaly-based detection methodology relies on the application of machine learning techniques for building a trustful activity model [29]. This methodology looks for any deviation from the known behavior derived from monitoring the system activities over a certain period. Indeed, after the system has been trained long enough to generate a model of what activities, hosts or even users affect the system, any incoming and outcoming anomaly traffic will be compared and declared dangerous if some of its characteristics cannot be found in the model.

The main advantages are the high dynamicity and extensibility of the model, since they capable of identifying new and unforeseen anomalies afflicting the system. On the contrary, its weak point relies on the low accuracy of acquiring attacks' information since the methodology does not use a proper knowledge base, as done by signature-based approaches, since it is strictly connected to the information acquired from the observed events. Moreover, the system cannot be operative straight away after its installation, but requires a certain amount of time for training the model, and adapt its analysis in response to the usual behavior of the network (or host) activities.

4 Framework

The proposed framework has been designed to improve the user' awareness of the network traffic and to simplify the IDS configuration process for receiving alerts when an attack is identified. The main idea is to create a modular framework that adapts to the different types of IDSs. This framework allows the user to manage their network through a visual interface and voice commands. The goal is to facilitate the installation and configuration of an IDS while ensuring the correct operations. The architecture of the proposed framework and its phases are described in the following sections.

4.1 System architecture

Normally, people use different types of devices without knowing what really happens during each usage and what the risks are. Therefore, it is difficult for them to configure network security tools. For these reasons, we have analyzed all the communication phases, starting from the interaction between users and devices, in order to design different components for the architecture of the proposed framework. The architecture involves components designed to ensure high modularity, adaptability, and ease of configuration.

The first challenge was to define an easily configurable and usable module by any type of device (Figure 1(a)). Thus, we have divided the architecture in three different layers that contain all the main components involving during network packets exchange. More specifically, our purpose is to monitor the outgoing and incoming traffic from the network during the connection between the devices and the local network controller. The first layer (Figure 1(b)) is divided into two distinct modules: devices and security configuration modules. The first is connected to Alexa so that it listens to voice messages and extracts the intents of the users. Through this module, the user defines the IDS configuration parameters. The security configuration module communicates with the third layer (Figure 1(c)). Using the parameters defined by the user, it automatically configures and executes the IDS manager on the network, in order to monitor web traffic. *IDS Manager* is a stand-alone component that can be easily replaced or updated as long as the framework configuration phase is repeated. Moreover, one of the main components of IDS is the *Notifier*. It is one of

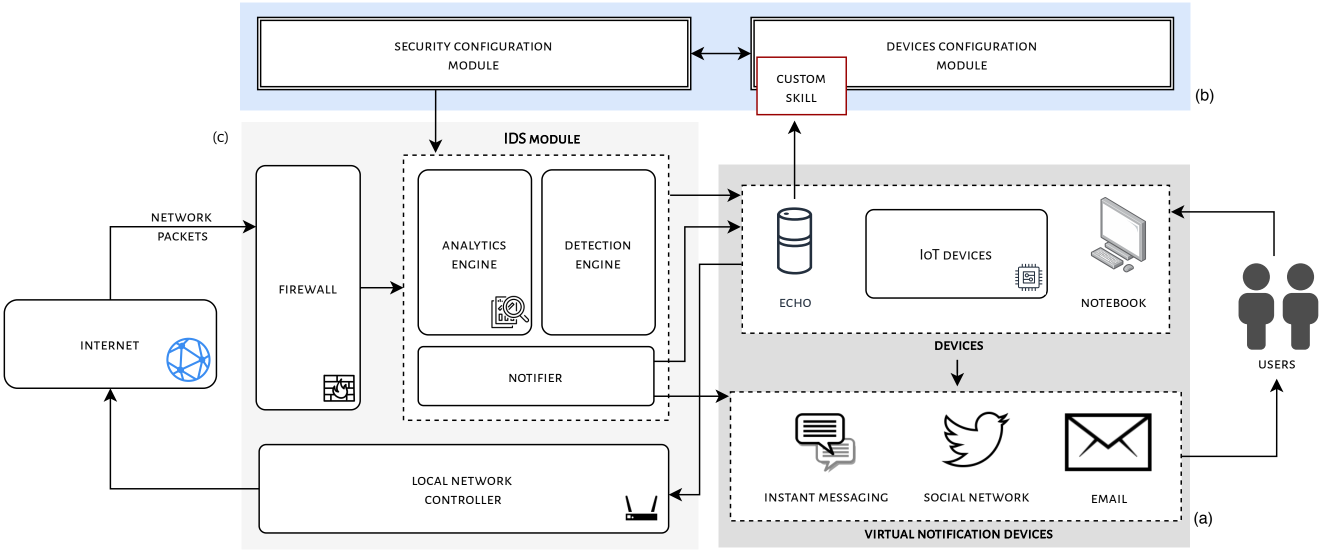


Figure 1: Architecture of the proposed framework.

the main elements of interaction with the user. In fact, the Notifier communicates directly with devices for sending reports of attacks to physical and virtual devices. The interaction modules with instant messaging apps and some of the most well-known social networks will be integrated into the framework. Users will be able to customize the notification devices by using the device configuration module.

The framework allows each user, with experience or not, to configure an IDS for their network and customize any device for receiving any alert.

4.2 Alexa custom skill architecture

One of the main proposals concerning this paper relies on the usage of voice assistant capabilities to ease the overall user experience with the system. Interaction with domestic voice assistants has been gaining prominence in the last period, becoming a very useful and simple communication channel [27]. Voice assistants, such as Google Home or Alexa, provide SDKs for the implementation of customized functionalities allowing for the definition of both the interaction model with the user, and the logic to deploy the commands on other devices. For this reason, we are planning the implementation of a customized functionality for the Amazon Alexa voice assistant, called ”skill” [1], through which the user can vocally interact with the IDS modules.

Figure 2 shows in detail how the vocal requests of the users are transformed into the corresponding commands that are deployed to the IDS modules. In particular, the user can launch the skill by pronouncing its name preceded by a keywords like: “Alexa run” or “Alexa start”. This starts the skill and enables the process of communication between the user and the framework through the voice assistant. Any pronounced command deployed by the user to the skill is received and passed through the API at the cloud orchestrator. It has the goal of communicating and synchronizing the actions of all the other modules in the Amazon cloud. The first involved module is the Speech Language Understanding (SLU) whose task is trying to match the specific request with the action. Indeed, all the actions a skill can execute, called *intents*, are associated with several utterances the user can pronounce to trigger that intent. When a match is found, the corresponding intent is passed back to the cloud orchestrator, which asks the Alexa Skill Service to perform that intent. The custom skill we are planning to design will at this point contact the IDS to fulfill the user’s request. After that, the system will return back to the skill with responses like the system status and notification about any intrusion occurring. In the last phase, the response received from the system are sent from the cloud orchestrator to the Text to Speech (TTS) module, which is responsible to translate the textual content into the voice that will be played by the Alexa device.

5 Discussion

In this section we will go through some of the most crucial and difficult aspects that a non-expert user needs to tackle down to correctly setup an IDS. We will also provide a general discussion about the solutions we are planning to implement for making all these steps possible.

We identified four main phases required for correctly using an IDS and we will walk through each of them describing our contribution plans.

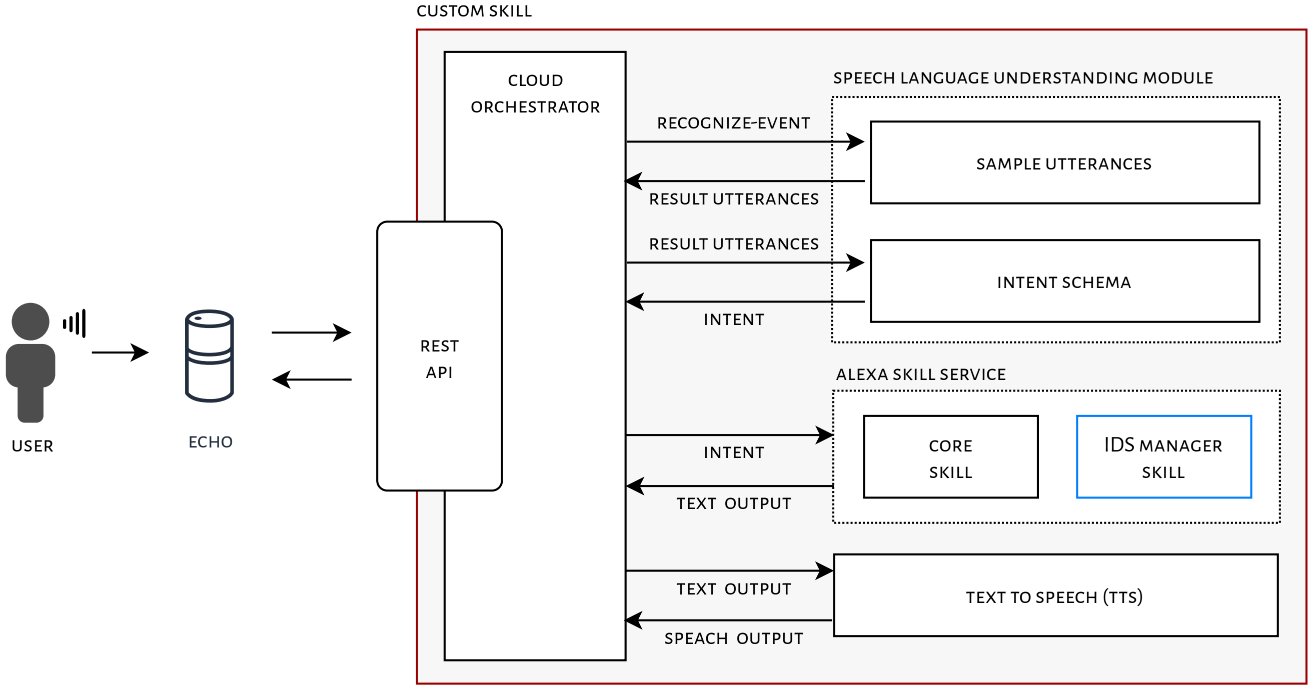


Figure 2: Architecture of the proposed skill.

5.1 Installation

The installation process is the first step a non-expert user needs to face when approaching an IDS. The main problem with this phase relies on the conspicuous amount of prerequisites the user have to deal with before actually proceed with the installation. Furthermore, most of the commands need to be deployed through a console command line, which represent a uncomfortable tool to interact with. To ease this issue, we are planning to include all the installation steps in an installer, a GUI-based software commonly found in the Microsoft Windows OS domain. An installer is composed of several windows describing the necessary steps to pursue with the installation of the software.

Hence, through a minimal interface user will have the possibility to specify the paths where all the required file will be saved and granting the mandatory authorizations for a correct execution of the IDS.

5.2 Configuration

After the installation phase, another crucial step is configuring the IDS. Indeed, it is necessary for the user to provide some essential parameters to obtain a correct network traffic monitoring together with the identification of intrusions. For example, it is fundamental for the user to provide the name of his/her network interface, i.e., the physical inbound and outbound connection port connecting the computer on which the IDS has been installed to the router and so the Internet.

To ease this type of process, we have planned to rely on a specifically designed visual interface, which will implement several visual metaphors designed to be suitable for the knowledge level of the non-expert users approaching it. The introduction of this new level of abstraction will help the users to complete a correct system configuration without getting lost into the details of technical terms.

5.3 Usage

Being a monitoring system whose main task is to silently monitor and evaluate in background the quality and type of network packets being exchanged, the active contribution by the non-expert users is reduced to the necessity of starting the IDS. However, this operation needs to be performed through a command launched from bash, which as mentioned above represents a particularly complicated step for inexperienced users. For this reason, we have planned the introduction of a series of automatism allowing the system to start without the user having to forcefully act on the system.

Alongside this choice, a useful alternative in this scenario would have been provided by the interaction capabilities offered by voice assistants. The Alexa custom skill we described earlier would allow the user to easily interact with the whole system, requesting to start the IDS and asking for information regarding its state of running.

5.4 Notification

Finally, the last step is the one that involves how to notify the user of the presence of an intrusion within the system. Even at this juncture, the use of a voice assistant providing immediate notification of the system's security status seems to be a suitable choice with respect to the knowledge level of the non-expert users. For this implementation phase, the challenge will be to program the type of message that the voice assistant will have to pronounce, avoiding phrases that can mislead the user and make the seriousness of the danger unclear. For example, the use of a phrase such as: "The system is under DDoS attack" is totally incomprehensible to a user who is unable to understand the seriousness of the danger of a DDoS attack which, in the IoT context, was the cause of one of the most devastating hacker attacks, the Mirai Botnet [21].

Therefore, it will be essential to find the right formulation to prevent the user from underestimating (or overestimating) the severity of the intrusion

5.5 User involvement

Being a framework specifically designed for end-users, the overall involvement of them in the realization and testing phases, plays a fundamental role in the achievement of a simple, functional, and effective system. For this reason, the development of the user interface will see the collaboration of some users, to whom we will submit some surveys to test their preferences. By doing so it is possible to direct the system towards the development of an interface more akin to user needs.

Another important phase will be the evaluation of the quality of the user experience. Thus, this type of evaluation, will be planned with the involvement of a large group of users, which we'll seek among who has little or none knowledge of computer technologies. Moreover, we will ask them to fill in different surveys in order to evaluate the usability and effectiveness of the framework.

6 Conclusion

Intrusion detection systems (IDS) have been defined as an essential security measure in any type of network. They are an important component that permits to identify network attacks by analyzing network traffic. This paper presents a framework that allows non-expert users to monitor their network and identify any attacks. In particular, we have defined two different modules connected to the main components of an IDS. Through this approach, it is possible to adapt our framework in different IDS systems. Moreover, an innovative skill for Alexa has been proposed, in order to allow users to run the IDS through voice commands.

In the future, we would plan to implement the framework, integrating it with different IDSs. Moreover, we intend to develop the skill for Alexa to test the framework with non-expert users in order to obtain constructive feedback and highlight their difficulties. Finally, we plan to extend the approaches proposed to capture the user navigation intents [12] for improving the intent understanding task.

References

[1] Amazon. Alexa Skill Kit. https://developer.amazon.com/it-IT/docs/alexa/sdk/alexa-skills-kit-sdks.html/, 2020. [Online; accessed 10-April-2020].

[2] T. Ambwani. Multi class support vector machine implementation to intrusion detection. In *Proceedings of the International Joint Conference on Neural Networks, 2003*, volume 3, pages 2300–2305. IEEE, 2003.

[3] S. Amendola, R. Lodato, S. Manzari, C. Occhiuzzi, and G. Marrocco. RFID technology for IoT-based personal healthcare in smart spaces. *IEEE Internet of things journal*, 1(2):144–152, 2014.

[4] J. P. Anderson. Computer security threat monitoring and surveillance. *Technical Report, James P. Anderson Company*, 1980.

[5] F. Anjum, D. Subhadrabandhu, and S. Sarkar. Signature based intrusion detection for wireless ad-hoc networks: A comparative study of various routing protocols. In *Proceedings of 2003 IEEE 58th Vehicular Technology Conference*, volume 3, pages 2152–2156. IEEE, 2003.

[6] I. B. Aris, R. K. Z. Sahbusdin, and A. F. M. Amin. Impacts of IoT and big data to automotive industry. In *Proceedings of 2015 10th Asian Control Conference (ASCC)*, pages 1–5. IEEE, 2015.

[7] A. E. Attipoe, J. Yan, C. Turner, and D. Richards. Visualization tools for network security. *Electronic Imaging*, 2016(1):1–8, 2016.

[8] S. Axelsson. Intrusion detection systems: A survey and taxonomy. Technical report, Chalmers University of Technology, 2000.

[9] B. Breve, L. Caruccio, S. Cirillo, D. Desiato, V. Deufemia, and G. Polese. Enhancing user awareness during internet browsing. In *Proceedings of the Fourth Italian Conference on Cyber Security*, pages 71–81. CEUR Workshop Proceedings 2597, 2020.

[10] L. Caruccio, S. Cirillo, V. Deufemia, and G. Polese. Incremental discovery of functional dependencies with a bitvector algorithm. In M. Mecella, G. Amato, and C. Gennaro, editors, *Proceedings of the 27th Italian Symposium on Advanced Database Systems*, volume 2400 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2019.

[11] L. Caruccio, V. Deufemia, C. D'Souza, A. Ginige, and G. Polese. A tool supporting end-user development of access

control in web applications. *International Journal of Software Engineering and Knowledge Engineering*, 25(2):307–331, 2015.

[12] L. Caruccio, V. Deufemia, and G. Polese. Understanding user intent on the web through interaction mining. *J. Vis. Lang. Comput.*, 31:230–236, 2015.

[13] L. Catarinucci, D. De Donno, L. Mainetti, L. Palano, L. Patrono, M. L. Stefanizzi, and L. Tarricone. An IoT-aware architecture for smart healthcare systems. *IEEE Internet of Things Journal*, 2(6):515–526, 2015.

[14] S. Cirillo, D. Desiato, and B. Breve. Chravat-chronology awareness visual analytic tool. In *Proceedings of 23rd International Conference Information Visualisation (IV)*, pages 255–260. IEEE, 2019.

[15] A. De Lucia, V. Deufemia, C. Gravino, and M. Risi. Detecting the behavior of design patterns through model checking and dynamic analysis. *ACM Trans. Softw. Eng. Methodol.*, 26(4):13:1–13:41, 2018.

[16] N. F. Haq, A. R. Onik, M. A. K. Hridoy, M. Rafni, F. M. Shah, and D. M. Farid. Application of machine learning approaches in intrusion detection system: a survey. *International Journal of Advanced Research in Artificial Intelligence*, 4(3):9–18, 2015.

[17] P. Ioulianou, V. Vasilakis, I. Moscholios, and M. Logothetis. A signature-based intrusion detection system for the internet of things. In *Proceedings of IEICE Information and Communication Technology Form*, 2018.

[18] A. Javaid, Q. Niyaz, W. Sun, and M. Alam. A deep learning approach for network intrusion detection system. In *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, pages 21–26, 2016.

[19] V. Jyothsna, V. R. Prasad, and K. M. Prasad. A review of anomaly based intrusion detection systems. *International Journal of Computer Applications*, 28(7):26–35, 2011.

[20] L. Koc, T. A. Mazzuchi, and S. Sarkani. A network intrusion detection system based on a hidden naïve bayes multiclass classifier. *Expert Systems with Applications*, 39(18):13492–13500, 2012.

[21] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas. DDoS in the IoT: Mirai and other botnets. *Computer*, 50(7):80–84, 2017.

[22] X. Krasniqi and E. Hajrizi. Use of IoT technology to drive the automotive industry from connected to full autonomous vehicles. *IFAC-PapersOnLine*, 49(29):269–274, 2016.

[23] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, and K.-Y. Tung. Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications*, 36(1):16–24, 2013.

[24] G. Nadiammai and M. Hemalatha. Effective approach toward intrusion detection system using data mining techniques. *Egyptian Informatics Journal*, 15(1):37–50, 2014.

[25] C. Papadopoulos, C. Kyriakakis, A. Sawchuk, and X. He. Cyberseer: 3D audio-visual immersion for network security and management. In *Proceedings of ACM Workshop on Visualization and Data Mining for Computer Security*, pages 90–98, 2004.

[26] R. M. Patil, M. R. Patil, K. V. Ramakrishnan, and T. Manjunath. Iddp: Novel development of an intrusion detection system through design patterns. *International Journal of Computer Applications*, 7(12):22–29, 2010.

[27] E. Polyakov, M. Mazhanov, A. Rolich, L. Voskov, M. Kachalova, and S. Polyakov. Investigation and development of the intelligent voice assistant for the internet of things using machine learning. In *Proceedings of Moscow Workshop on Electronic and Networking Technologies (MWENT)*, pages 1–5. IEEE, 2018.

[28] L. Qi, M. V. Martin, B. Kapralos, M. Green, and M. García-Ruiz. Toward sound-assisted intrusion detection systems. In *Proceedings of OTM Confederated International Conferences On the Move to Meaningful Internet Systems*, pages 1634–1645. Springer, 2007.

[29] P. Sangkatsanee, N. Wattanapongsakorn, and C. Charnsripinyo. Practical real-time intrusion detection using machine learning approaches. *Computer Communications*, 34(18):2227–2235, 2011.

[30] V. Singh and S. Puthran. Intrusion detection system using data mining a review. In *Proceedings of 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, pages 587–592, 2016.

[31] M. Uddin and A. A. Rahman. Dynamic multi layer signature based intrusion detection system using mobile agents. *International Journal of Network Security & Its Applications (IJNSA), Vol.2, No.4*, 2010.

[32] G. Vigna and R. A. Kemmerer. Netstat: A network-based intrusion detection system. *Journal of Computer Security*, 7(1):37–71, 1999.

[33] B. W. Wirtz, J. C. Weyerer, and F. T. Schichtel. An integrative public IoT framework for smart government. *Government Information Quarterly*, 36(2):333–345, 2019.

[34] K. Zaraska. Prelude IDS: current state and development perspectives. *URL http://www.prelude-ids.org/download/misc/pingwinaria/2003/paper.pdf*, 2003.

[35] B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga. A survey of intrusion detection in internet of things. *Journal of Network and Computer Applications*, 84:25–37, 2017.

[36] Y. Zuo and B. Panda. Fuzzy dependency and its applications in damage assessment and recovery. In *Proceedings from the Fifth Annual IEEE SMC Information Assurance Workshop, 2004.*, pages 350–357. IEEE, 2004.

Classification of Users' Transportation Modalities in Real Conditions

C. Badii, A.Difino, P. Nesi, I. Paoli, M. Paolucci
University of Florence, Department of Information Engineering, Distributed Systems and Internet Tech lab
DISIT Lab, https://www.disit.org , http://www.sii-mobility.org , https://www.km4city.org <name.surname>@unifi.it
Corresponding Author: Paolo Nesi, paolo.nesi@unifi.it

***Abstract* —** The modern mobile devices and the complete digitalization of the public and private transport networks have allowed to access useful information to understand the user's mean of transportation. This enables a plethora of old and new applications in the fields of sustainable mobility, smart transportation, assistance, and e-health. The precise understanding of the travel means is at the basis of the development of a large range of applications. In this paper, a number of metrics has been identified to understand whether an individual on the move is stationary, walking, on a motorized private or public transport, with the aim of delivering to city users personalized assistance messages for: sustainable mobility, health, and/or for a better and enjoyable life, etc. Differently from the state-of-the-art solutions, the proposed approach has been designed to provide results, and thus collect metrics, in *real operating conditions* (imposed on the mobile devices as: a range of different devices kinds, operating system constraints managing Applications, active battery consumption manager, etc.). The paper reports the whole experimentations and results. The solution has been developed in the context of Sii-Mobility Km4City Research Project infrastructure and tools, GDPR compliant. The same solution has been used in Snap4City mobile Apps with experiments performed in Antwerp and Helsinki.

Keywords—user behavior analysis, smart city, mobile phones, transportation modes, classification model, machine learning.

I. Introduction

With the complete digitalization of the public and private transportation networks, the capability of understanding the users' behavior and the mean of transportation have become important. The presence of GPS, accelerometers, sensors on mobile phones has made possible to create solutions exploiting the users' behavior and context. City users are from at the same time information providers and recipients of personalized suggestions and information [Lv et al., 2018]. The understanding of user behavior is the first step for providing suggestions and assistance to people on the move via mobile phones in smart city putting city users in the loop. For example, to allow the city user to receive suggestion to take more virtuous behavior, consume less energy, making more sustainable their transportation, having a healthier life walking more, saving money parking closer (and to monitor their reaction and acceptance level). The research addressed in this article aims to understand the users' mean of traveling taking into account contextual data and data coming from the phones. The correct classification of transportation means can be also used for providing suggestions in the context of public or private transportation. Thus, the above described problem is reconducted to the classification problem of the transportation modality/mean (car, bus, walk, bike, etc.), exploiting real time data coming from the devices and contextual information. Please note that, the contextual data are strongly different in different part of the city, and also change over time, for example busses have different timeline and paths: so that users are moving in the real space.

As described in the following section of related works, the problem of understanding the mean of traveling of users has been many times addressed, but not working in real operating conditions. Most of them, assume data collected from the mobile phones with high rates and high precision, identifying models only taking data in strongly controlled conditions: such as limited number of device type, limited number of users and directly engaged to keep the mobile app running in foreground, etc.

A. Related works

The problem of classifying users' mean of travelling has been addressed by a number of approaches in different research areas [Prelipcean et al., 2017]: *Location Based Services* (LBS), *Transportation Services* (TSc) and *Human Geography* (HG). The LBS solutions aim to understand the transportation meaning in real-time to provide useful information to the user whenever he/she asks. In TSc approaches, the correct segmentation of a trajectory is privileged with respect to velocity of response: [Biljecki et al., 2013]. The HG approaches focus on the segmentation of a trajectory into parts with domain-specific semantics: it is common to first split trajectories into segments where the object is stationary or moving. In LBS, the transportation means' classification is regarded as an online process: an algorithm that provides the current transportation mode of the user in real-time or quasi real-time. To this end, different types of data/sensors have been exploited: GPS, accelerometer and their combination. [Stenneth et al., 2011] compared five different models using data collected from GPS classifying the users' travelling means in six categories (walk, train, driving, stationary, bus, bike). [Hemminki et al., 2013] proposed a study that involves only accelerometer data. They have obtained an 80.1% accuracy and an 82.1% recall for seven transportation modes, by using both AdaBoost and Decision Tree (two-stages classification). [Yu

DOI reference number: 10.18293/DMSVIVA20-016

et al., 2014] have compared three different classifiers (Decision Tree obtaining an 84.81% average accuracy, AdaBoost with a 87.16% average accuracy, and SVMs with a 90.66% average accuracy). [Wang et al., 2010] have considered a small data-set of 12 hours (5544 samples of six transportation modes) from 7 different users, obtaining a 70% accuracy with a Decision Tree algorithm. [Reddy et al., 2010] demonstrated that, taking into account of both GPS and accelerometer the accuracy can be improved. They have achieved a 93.6% precision using a combination of Decision Tree and HiddenMarkov Model (two-stages classification), with both accelerometer and GPS features involved, using a sampling rate made a distinction among different type of non-motorized motion (walking, running, biking), vehicular and random movements, using the accelerometer sensors of mobile. [Manzoni et al., 2010] trained a Decision Tree classification model obtaining an 82.14% accuracy (with a gps and accelerometer sampling frequency rate of 1s and 0.04s respectively). [Ashqar et al., 2018], proposed a two-layer hierarchical classifier to predict five classes of transportation mode (car, bus, walk, run, bike), achieving a 97% accuracy. [Yanyun et al., 2017] presented a Convolutional Neural Networks (CNN) based method to automatically extracting features for the identification of transportation means, thus achieving a 98% accuracy to distinguish between train, bus, car, metro.

In **Table 1**, a summary of the state-of-the-art solutions for understanding the travel means is reported (the table report also additional experiments/papers with respect to those commented above). Almost all the state-of-the-art solutions adopted very **high rates for GPS data acquisition, with limited number of devices**. So that, those solution are almost unfeasible in *real operating conditions*. Mobile operating systems allow to keep the high rates (in the order of seconds) only when applications are running in foreground. In most cases, the precisions provided has been obtained with limited set of devices in unrealistic conditions.

Table 1. Related Work implementation overview.

Authors	**Classes**	**data exploited**	**Sampling**	**#users**	**#features**	**# of device types**	**Precision accuracy**
Wang et al.,2010	Stationary, Walk, Bike, Bus, Car, Metro	Accel	0.03s (accel)	7	23	**1**	70
Manzoni et al.,2010	Stationary, Walk, Bike, Motorcycle, Car, Bus, Metro, Train	Gps Accel	1s (gps) 0.04s (accel)	4	1	**1**	82.1
Reddy et al.,2010	Stationary, Walk, Run, Bike, Vehicle	Gps Accel	1s (gps) 0.03s (accel)	16	4	**1**	93.7
Stenneth et al.,2011	Stationary, Walk, Bike, Car, Bus, Train	Gps Gis	14s (gps)	6	7	**3**	92.8
Hemminki et al.,2013	Stationary, Walk, Car, Bus, Train, Metro, Tram	Accel	0.01s (accel)	16	27 +5	**3**	84.9
Prelipcean et al., 2017	Walk, Bike, Car, Bus, Metro, Train, Ferry	Gps Accel	50m (gps) 0.2s (accel)	9	11	-	90.8
Yu et al.,2014	Stationary, Walk, Run, Bike, Vehicle (Motorcycle, Car, Bus, Metro, Rail, Train)	Accel	0.03s (accel)	224	22 +8	**1**	91.5
Yanyun et al.,2017	Train, Metro, Bus, Car	Accel	0.01s (accel)	30	169	**1**	98
Ashqar et al.,2018	Car, Bus, Bike, Run, Walk	Gps Accel	0.04s (gps) 0.01s (accel	10	80	**2**	97

B. Research aims and Article Organization

The aim of our research has been to realize a solution overcoming the state-of-the-art solutions to classify the transportation modes to deliver personalized services for:

- sustainable mobility, to incentivize ecological transportation choices, suggesting alternative public mean of transport (bus/tram) instead of the private car/motorbike.
- healthy suggestions, better and enjoyable life, to stimulate users in dedicating a part of their time and moving needs to exercising their body.
- implementing city strategies to change city user attitudes [Badii et al., 2017b], [Badii et al., 2018].

With this pourpose, the real-time identification of a private transportation mode (car or motorbike) has a central role in assistance messages delivery. Therefore, according to the above described real operating conditions, the techniques have to produce high classifications accuracy to identify transportation modality of a user, in the presence of (i) large discontinuities samples of data (from sensors and sporadic communications to the central computation modules), (ii) relevant differences which may be due to the different kind of mobile phone features in terms of sensors and precision.

Therefore, the proposed solution overcomes the above-mentioned solutions at the state of the art, for the aspects focused on sensor energy consumption factors and real conditions. The solution has been tested on a real application (delivered to the users via official App stores such as Google Play Store, Apple App Store, and accepted by common users, see "*Tuscany where what….*" on the stores). As described in the following, it is capable to cope with the constrains introduced by terminal manufactures on battery usage for background and foreground services. Moreover, no restrictions on the modality of mobile device usage have been imposed, differently to what has been imposed in the state of the art experiments in which the devices have been asked to keep: (i) the application running in foreground to get more precise GPS data, (ii) the device in a proper position/ orientation during the usage; and/or to (iii) use specific devices.

The paper is focused on the classification of the transportation mode/means whether an individual is: (i) stationary, (ii) walking, (iii) on a motorized private transport (car or motorbike), or (iv) in a public transport (tram, bus or train). The classification model proposed has been produced by using open and real-time data of Sii-Mobility/Km4City project and infrastructure (which is national smart city project of Italian Ministry of Research for terrestrial mobility and transport, http://www.sii-mobility.org). Sii-Mobility is based on Km4City data aggregation and analytics infrastructure (https://www.km4city.org) in the Tuscany area, Italy, and its Smart City solutions. Research results have been produced with the aim of defining solutions for sustainable mobility, stimulating citizens towards virtuous behaviors, providing info mobility, etc. The research project conducted a large experimentation of the solution with the support of Public Transport Operators: BUSITALIA, CCTNORD.

This paper has been organized as described in the following. Section II reports the general architecture for data collection, from devices to server, and related data analytics. In Section III, a list of the identified metrics is reported, mainly related to: baseline, GPS, accelerometer, and historical data. Section IV proposes a comparison of predictive models exploiting the collected data from Km4City, to arrive at identifying the best resulting approach in terms of classification precision and recall. Conclusions are drawn in Section VI.

II. ARCHITECTURE AND DATA COLLECTION OVERVIEW

The proposed solution relies on a client-server architecture, where the mobile application can be installed on different operating systems with different versions [Badii et al., 2017a]. The sensors' values collected on the mobile device (client-side) are sent to the server that enriches them with additional context information derived data (GIS, geographical information system and knowledge), etc., as described in the sequel. At the same time, the server executes the real time classification algorithm to compute the transportation mean classification for each user. The information is stored on server as a report on the preferred user's travel mean. On this basis, strategies triggered when the user behavior reaches certain specific conditions has been activated – i.e., to assist and/or engage the users in their daily activities (even rewarding them, in the cases of virtuous behavior). For example, a strategy for stimulating the city users may be based on a firing condition which sends a suggestion to all city users that take their private car to perform the same trip path at least 3 times per week, and at the same time the trip could be easily performed by using public transportation. Thus, the system may inform those city users of the possible alternative, and some of them may follow the suggestion. As a result, by exploiting the user behavior analysis, the solution may detect the acceptance of the suggestion by detecting of change of behavior and may automatically reward the user with a bonus or discount, and deliver congratulations. See [Badii et al., 2017b], on rules and strategies. Figure 1 provides a high-level overview of the software architecture and its main components.

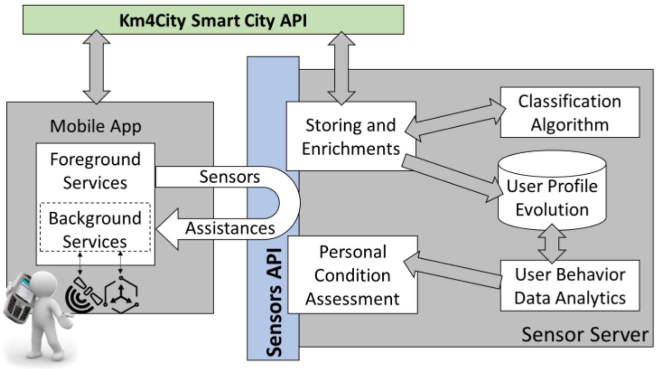


Figure 1. System architecture.

III. DATA AND FEATURES DEFINITIONS

The features considered by the classification algorithm have been selected from a larger set considered during the preliminary analysis and experiments. The process of features reduction has been performed by assessing their relevance in the classification. The aim was to identify the smallest subset of features without reducing significantly the precision of the travel mean's classifications. As a result, **Table 2** includes the selected metrics and the features, classified in 4 categories, collected from the mobile as **Sensor Data Package**, with those computed from the server-side to be used by the classification algorithm. Some of these features can be used for both users' traveling mean classification, and for creating firing conditions for implementing strategies. In **Table 2** "Where" can be: "D" when the measure is produced on the Device, and "S" when is computed on server-side. Each measure is collected/referred at a given **Day and Time**, and from this value can be easily derived from the device or from server if the day is a working day or not (**Non-Working Day**). The same approach can be followed to detecting the **Time Slot** in which the measure has been collected. The Time Slot strongly influences the attitude of the city users to move by using different means.

Table 2. Overview of Sensor Data Package feature measured at a given time from the mobile or computed on server-side.

categories	Metrics	Description of metric variable	Where
Day/Time Baseline and GPS	Day and Time	Day and Time of the sample package	D
	Non-Work-ing day	1 if weekend or vacation, 0 if it is a working day	D/S
	Time Slot	Slot of the day (morning, afternoon, evening, night)	S
	GPS latitude and longitude	Position of the device in GPS coordinates	D
	Accuracy	GPS Sensor's Accuracy from the mobile device	D
	Location Measure kind	Types of Location measure: GPS, Network, Mixed/Fused	D
	Speed	Speed as provided by the GPS driver of the mobile (as m/s)	D/S
	Average Speed	Average speed of the measures collected in the last two minutes	D/S
	Phone Year	Year/age of the terminal	D
	BDS	Availability of a BDS compliant	D

		GPS Sensor	
	User Type	User Type: commuter, citizen, students, tourist, etc.	D/S
Accelerometer	Average linear magnitude of acceleration	Average of the acceleration magnitude calculate on five measurements	D
	Linear acceleration of X-axis	Acceleration of the device along the X-axis, purged by Earth gravity	D
	Linear acceleration of Y-axis	Acceleration of the terminal along the Y-axis, purged by Earth gravity	D
	Linear acceleration of Z-axis	Acceleration of the terminal along the Z-axis, purged by Earth gravity	D
Proximity	Rail Line	Bool indicating if the device is in proximity of a rail line	S
	Sport Facilities	Bool indicating if the device is in proximity of a sport facilities	S
	Tourist Trail	Bool indicating if the device is in proximity of a tourist trail	S
	Green Areas	Bool indicating if the device is in proximity of a green areas	S
	Bus/ Light-rail Line	Bool indicating if the device is in proximity of a bus line or a light-trail line	S
	Cycle Paths	Bool indicating if the device is in proximity of a cycle path	S
Temporal window	Previous speed	Speed of the device of the previous 12 minutes	S
	Previous average speed	Average speed on the measures collected in a 12 minutes time slot	S
	Previous median speed	Median speed on the measures collected in a 12 minutes time slot	S
	Speed distance	Speed (m/s) calculated on the distance between two consecutive coordinates and the time passed between the observations	S

As described in Section II, the information about the user's movements is collected from the device sensors. If the user has the mobile application in foreground, the data are sent to the server every 1 minute and 30 seconds (sending interval). This interval can be reduced by the user (via the setting of the App) to an update up to 30 seconds, to have a more accurate assistance. If the App is not used, the data collection is performed in background modality, thus the measures and sending rates may become up to 3/5 minutes, forced by the operating system/device, which in some cases can hibernate the App. Therefore, in order to make the solution viable in real conditions (differently from the state-of-the-art solutions), a set of strategies and robust classification algorithms have been put in place. Among them, solutions for filtering noise and GPS errors, and for smoothing the sequence of the user locations (user trajectory) have been used.

A **Sensor Data Package** l_i represents the user context at a specific time t_i and is composed by the **GPS latitude and longitude** (according to a Location Measure kind), speed, and accuracy of the measure plus a list of N additional features (*feat-1…feat-n*):

$$l_i = \{latitude_i, longitude_i, speed_i, accuracy_i, feat\text{-}1_i, \dots, feat\text{-}n_i\}$$

A user trajectory t_{ir} is a sequence of l_i that describes the movements of a user to move from l_i to l_r:

$$t_{ir} = \{l_i, \dots, l_r\}$$

A segment s_{uv} is a trajectory t_{uv} in $t_{ir}$where a user keeps the same mobility mean:

$$l_i \rightarrow mobility\text{-}A \rightarrow \underbrace{l_u \rightarrow mobility\text{-}B \rightarrow l_v}_{t_{uv}} \rightarrow mobility\text{-}C \rightarrow l_r$$

The distance between l_u and l_v can be approximated by using flat-surface formulae between the two coordinates $<latitude_u, longitude_u>$ and $<latitude_v, longitude_v>$.

A measure of the terminal **Speed** can be directly retrieved from the GPS sensor (for example, every 30 seconds or at the rate imposed by the device). On the other hand, the above-mentioned Average Speed of **Table 2** is calculated over the sequence of l_i in the same sending slot from the mobile device, to cut out eventual errors coming from GPS sensor. If the mobile App is in foreground the **Average Speed** is computed every 2 minutes (4 measures of 30s, if any). If the mobile App service for collecting data is in background, and 2 minutes passed before a measure is available (probably the operating system put the application in hibernate mode). The service tries to wake up whenever it is possible (if the operating system on the device allows us to wake the service up), to retrieve a bounce of new l_i to calculate a more precise Average Speed. **Figure 2** overviews the scenario.

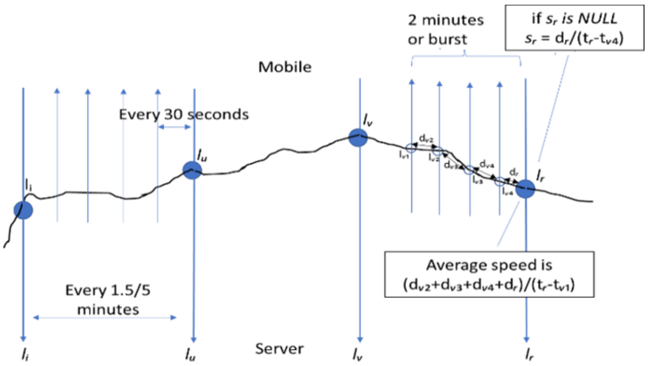


Figure 2. Speed and average speed.

The **Location Measure kind** is an important feature to understand the location measures reliability. Usually, the measures obtained and marked as "GPS" by the mobile device are quite accurate, even if they suffer time by time of well-known problem of shading (e.g., urban canyoning) or blocked (under the bridge) [Misra et al., 2006]. The location measures, labeled as "Network", resume the position from the location of the available Wi-Fi hotspots or GSM/4G/5G in the mobile connection; while those marked as "Mixed" modality is obtained by the operating system by merging the previous strategies according to different algorithms that may depend on the operating system kind, sensor kind, etc. The Location Measure kind strongly depends on the factory settings of the device, that make very difficult to force a pre-determinate modality from the App. The **Accuracy** of the GPS measure is reported in meters from the device and can be used from the classificatory algorithm to eventually discharge entries. Terminal model and its characteristics are also tracked and passed to the classification algorithm. Thus, the **Phone Year** of production of the device and the characteristics of the GPS sensors strongly influence the reliability of measure and thus have and have been considered as variable, differently from the state-of-the-art solutions. Old terminals usually support just A-GPS modality, meanwhile new ones' support also GLONASS and BDS standards.

A. ACCELEROMETER FEATURES

Values from the **Accelerometers** of the terminal/device are always available and are sampled. Using the linear acceleration of the device avoids taking measures influenced by device orientation (horizontal or vertical, in the hand or in the pocket). Not all the mobile devices provide this information (some of them just return the non-linear values, influenced by the gravitational acceleration, and orientation, thus needing a de-rotation). On the other hand, almost all the relatively new devices already have this aggregated measurement available. **Phone Year** variable allows us to take this into account. Thus, the three measures of linear acceleration on three axes have been considered aggregating five consecutive acceleration measures for computing an average magnitude as:

$$\textit{Average Linear Magnitude of Acc} = \sum_{k=1}^{5} \frac{\sqrt{acc_{x_k}^2 + acc_{y_k}^2 + acc_{z_k}^2}}{5}$$

B. DISTANCE FEATURES

On the server-side, the **Sensor Data Package** collected from the devices via the App are enriched by computing and, in most cases, exploiting the Km4City knowledge base of the City via Smart City API. This allows to retrieve contextual information about the closeness of the device/user with respect to: Railway Line, Sport Facilities, Tourist Trail, Green Areas, Bus/Light-rail Line, and Cycle Paths. The closeness features are binary values that specify if the location is closer to those structures, in the range of 30mt. This derived information is very valuable for understanding some transportation means. For example, to be close to a Rail and/or Bus/Light-rail line for a number of points of a trip permits to infer bus/train modality (train, bus and light-rail run just in their closeness) with a very high probability. On the other hand, the closeness to a cycle path cannot directly infer that a user is using a bike because the user can be in its proximity by a car and with similar speed or a bike can run also away for the cycling path (see **Figure 3**).

Besides having instantaneous measurements about device/user's mobility, the speed values in the last 12 minutes time-frame is also computer on server-side, as well as the average and mean value between these measurements. This allows to reduce the noise overcoming disruptive mobility conditions mainly related to traffic congestion or temporary signal absence. This is also due to the fact that the service for collecting data on the mobile device runs on a real application (foreground/background) conforming to the policy of "energy saving" of the user to have shortage of data for up to 3/5 minutes. So that, in real conditions, it is very important to avoid battery drainage warning, that may stimulate the user to un-install the App from the device. In order to perform an addition refinement on speed measures, mean and median speed and distance between GPS coordinates are also computed.

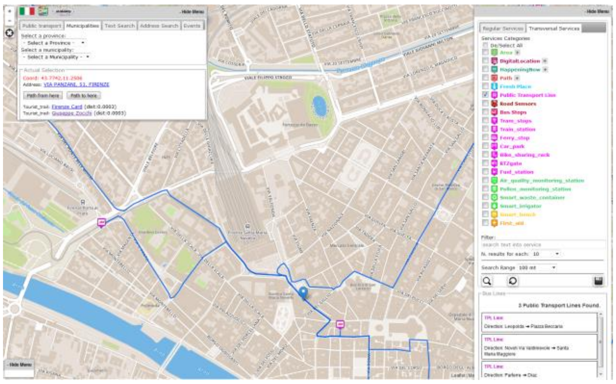


Figure 3. Bus-line in proximity computation

C. TEMPORAL WINDOW FEATURES

The **User Type** specified by the user in the App during installation or setup permits contribute to the classifications and to the strategies. The User Types are: citizen, commuter, student, tourist, etc. We noticed that different profiles present a different approach in everyday mobility and, so on the transportation mode they normally use.

IV. RESULTS FROM CLASSIFICATION/PREDICTION MODELS

According to the above described data, the challenge was to predict the transportation mode, whether an individual is stationary, or is walking, or moving on a motorized private transport (car or motorbike) or using a public transport (tram, bus or train). The experiment has been conducted on about 30.000 observations, collected from April to August on **38 different users and 30 different kinds of devices.** Note that, each user can use the mean of transport they want. When the mode of transport is changed, the user was asked to notify the change to the App for creating the learning set and for validation. As mentioned above, no restriction was imposed on how the phone should be held during movement (foreground/background, on hand or bag, etc.). Unlike the experiments reported in the literature, most of the data was collected in the background because the phones were kept in pocket or bag, in fact there is a non-conformity in the frequency distribution of the collected data. In details, the frequency average is equal to 180 seconds and the variance is equal to 13240 seconds. The frequency distribution of the sampling period is reported in **Figure 4**.

The training set has been created by randomly selecting the 80% of the collected data, while the test set was the remaining 20%.

In the general framework, three different approaches were more successfully considered -- i.e., Random Forest (RF), Extremely Randomized Trees (Extra-Trees), and the

Extreme Gradient Boosting procedure (XGBoost). Those approaches have been tested by using the above presented features/metrics (see **Table 2**), classified by categories as: *baseline* and *GPS* features, *accelerometer* features, *distance* features and *temporal window* features. The comparison among those models has been reported in **Table 3**, in terms of resulting data. From the comparison, it is evident that all the approaches are capable to produce satisfactory predictions (the accuracy for each model exceeds 90%) for the identification of the transportation means.

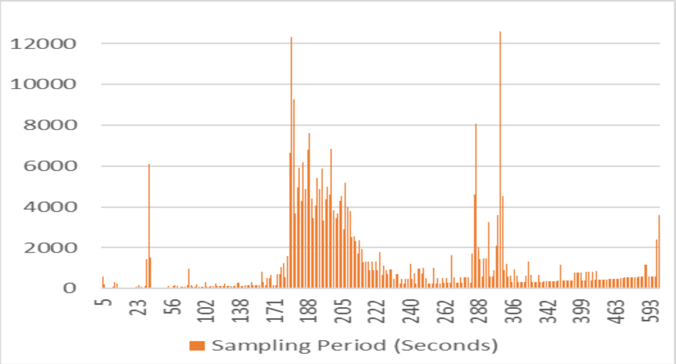


Figure 4. Frequency Distribution of Sampling Period.

According to the data results of **Table 3**, the differences among the different approaches provide the evidence that the **Extra-Trees** resulted to be the better-ranked approach in terms of accuracy and F_1 score. In **Table 3**, the F_1 score is reported: F_1 score has been used to measure the models' performances. This is a measure to evaluate the robustness of a model for making predictions, as a compromise between precision and recall:

$$F_1\ score = 2 \times \frac{Precision\ \times\ Recall}{Precision\ +\ Recall}$$

$$Precision = \frac{\#\ \text{correctly classified istances into class}\ i}{\#\ \text{istances classified as class}\ i},$$

$$Recall = \frac{\#\ \text{correctly classified istances into class}\ i}{\#\ \text{istances belonging to the class}\ i},$$

Table 3. Classification Models Comparison on four classes of transport mode: stationary, non-motorized, private transport, public transport.

Classifier Models	Accuracy	Precision	Recall	F_1 score
Extreme Gradient Boosting	0.947	0.773	0.828	0.800
Random Forest	0.942	0.774	0.869	0.819
Extra-Trees	**0.953**	**0.827**	**0.869**	**0.847**

According to this our first result, the Extra-Trees algorithm achieves an accuracy of 0.953, and a precision of 0.827. It should be remarked that, these results have been obtained and can be produced by observing data coming from a large range of devices and a variable sampling rate (up to 5 minutes). The model produce allows to understand if a user is moving with a public or private transport.

On the contrary, in [Reddy et al., 2010], a precision of 0.937 has been obtained by using a single device, Nokia n95, and a constant sampling rate of 60s, which is not realistic with present mobile operating systems. With Reddy's classification was only possible to know if a user is moving with a motorized vehicle. The same considerations apply to: [Stenneth et al., 2011] where data come from three different devices and they are taken with a constant rate of 15s achieving a precision of 93.7%; and to [Yu et al., 2014] achieving a precision of 91% with accelerometer sensor data only, without distinguishing the type of motorized transport.

Moreover, Table 4 reports the assessment of the results performed for each traveling mean classification for the Extra Tree procedure according to our first result. The traveling mean class with lower accuracy is Walk. This is probably due to the fact that, it is not easily to understand if a user is walking or not, since the GPS sensors accuracy is very noisy in indoor scenarios, with frequent jumps passing from the different modalities: wifi- mixed, etc.

Table 4. Extra-Trees Prediction Model: Statistic by class.

Extra Trees Model	Stay	Walk	Private Transport	Public Transport
Sensitivity	0.978	0.731	0.869	0.917
Specificity	0.901	0.988	0.987	0.996
Pos Pred Value	0.977	0.770	0.827	0.936
Neg Pred Value	0.904	0.985	0.990	0.994
Balanced Accuracy	0.940	0.859	0.928	0.956

We also tested the effect of combining the solution with a SuperLearner approach without obtaining better results.

A. Assessing the Influence of Features

A comparison in terms of accuracy, precision and recall of the Extra-Trees multi-class approach has been computed considering four combinations of the different categories of data (as reported in **Table 2**):

- baseline features and distance feature;
- baseline, distance feature and accelerometer features;
- baseline, distance feature and temporal window features;
- baseline, distance, accelerometer, temporal features together. (**Full Model**)

This set of combinations of feature categories permits to assess the flexibility of our approach in real operative conditions, where a variety of devices have to be supported, since not all devices support the full combination of categories. The results obtained by using different subsects of feature categories are reported in **Table 5.** Please note that the differences among the different cases for feature categories are substantial. The results suggest that the best choice in terms of precision is still the usage of model exploiting all the categories together, thus demonstrating that the model is flexible and resilient with respect to the device kind. Please note that the Boolean value detecting a close transportation line (i.e., proximity feature in the table) improves the classification effectiveness: the accuracy passed from 0.91 to 0.92 and higher.

Table 5. Extra Tree Model results on four classes of transport modality (stationary, non-motorized, private transport, public transport) considering four combinations of the different features.

Model features categories	Extra Tree Model results			
	Accuracy	Precision	Recall	F_1 Score
Baseline and GPS	0.910	0.682	0.751	0.714
Baseline and GPS + proxi mity	0.924	0.739	0.691	0.715
Baseline and GPS + proxi mity + Accelerometer	0.926	0.814	0.744	0.777
Baseline and GPS + proxi mity + Temp window	0.949	0.805	0.787	0.787
Baseline and GPS + proxi mity + Accelerometer + Temporal window	**0.953**	**0.827**	**0.869**	**0.847**

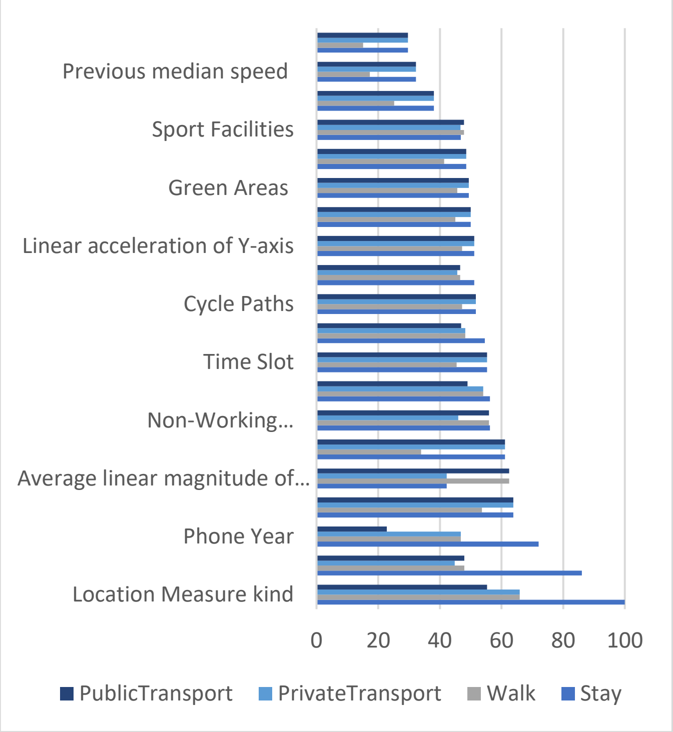


Figure 5. Variables Importance across the classes of the Extra-Trees full model.

In **Figure 5,** the features listed in **Table 2** are reported in order of importance across the classes for the prediction of the Extra-Trees Full Model, (the model with all the categories of covariates).

B. REAL CONDITION SCENARIO VS HIERARCHICAL APPROACH LIMITATIONS

Several considerations have been already presented about the critical aspect of working on real operating conditions. Battery drainage and the opportunity to support a contextual service for the users, even with the application in background mode, drove our research, despite little decrease of accuracy and precision. We decided to design a client-server architecture to support a finer classification, using GIS data easier available on the server side (avoiding user terminal network bandwidth usage to eventually download from remote) and to support technologies to aggregate information cross-terminal and user agnostic. Implementing a central server-side classification algorithm leaves open also the chance to auto-update scenario with feedbacks provided directly by the user. However, a real condition scenario can be affected by some limitations that cannot be solved either if a hierarchical approach is applied. This is due to the fact that the phone/user characteristics can be manifold, e.g., the presence of accelerometer information, the different type/generation of gps sensor, the presence of information related to the temporal window, etc. For this reason, the classification model has to be flexible and the training dataset has to be as much as possible various (e.g., any kind of generations, manufactures, years, characteristics, etc.) without any restriction. The application of a two-steps approach, may lead to a loss of accuracy due to a loss of information and can be more time consuming in terms of execution time and number of different training models. In detail, during the second step, six different training models have to be executed, one for each combination between pairs of the transportation modes (selected during the step-one), considering that the classes of transportation means are four. In addition, a specific model has to be created depending on the characteristics of the device and of the users, considering four combinations of the different categories of data (reported in Table 2).

A two-steps hierarchical approach has been proposed. In the first step a multi-class classifier algorithm has been adopted to classify the transportation modalities. After the first classification, the classes with a probability lower than a threshold of 0.90 (*prob* < 0.90) have been re-classified in the second step, while the classes that have a probability higher than 0.90 are considered as correct and excluded from the re-classification test set. During the second step six different binary classification model have been trained, one for each pair of transportation modality.

On the other hand, a single step classification model has been presented and different models have been compared. The Extra-Tree algorithm can be considered as the best and final solution: it was found to produce the best performance in terms of average accuracy (**0.953**) and time consuming. In detail, four different models have been trained to make the approach as flexible as possible. The necessity of this flexibility is because the solution has to be applied in a real condition scenario, for different phone/user characteristics, in any pseudo real-time context.

V. CONCLUSIONS

This research has been focused on presenting a solution to create a classification system that uses mobile devices' sensor values and GIS data (user contextual information) to identify the transportation mean of users: stationary, walking, on a motorized private transport (car or motorbike) or in a public transport (tram, bus or train). The goal has been to define a solution for sustainable mobility, delivering to the user useful personalized assistance messages. A number of metrics and features have been chosen as the *baseline and GPS*, the *distance*, the *accelerometer* data and the *temporal windows* data. The research documented in this paper demonstrated that a one-step multi-class classifier solution was found to produce the best performance in terms of average accuracy and time consuming if compared to a hierarchical approach. In detail, the Extremely Randomized Trees exploiting all the discussed above data can be a robust approach for reliable, precise and fast estimation of transportation means. The proposed solution overcome those of the literature since it presents a solution that is capable to produce reliable results in real conditions (i.e., real-time applications and background modality of operations) with a real set of devices and in particular: (i) addressing a large number of devices providing different features, different GPS sensors, different accelerometer sensors, etc., (ii) working with time variable samples of the data that may be due to the different operating systems, energy saving setting, etc., which are not under control of the App and thus are a strong constraint to realize real applications, background/foreground modality of operation; (iii) exploiting a number of different features and obtaining results with higher precision and accuracy. For these reasons, features related to the type of phone, e.g., the presence of accelerometer, phone year, location provider etc., have been considered in the prediction model, contributing to perform corrections in the model. The prediction model proposed has been created by exploiting open and real-time data of the Sii-Mobility (national smart city project of Italian Ministry of Research for terrestrial mobility and transport, http://www.sii-mobility.org). Sii-Mobility is un turn based on Km4City infrastructure http://www.km4city.org active in the Florence area, Italy since 2015. The solution presented has been deployed as an additional feature on Smart City Apps in the Tuscany and Florence areas for sustainable mobility, which is now in place for stimulating the private mover toward a more sustainable mobility with the collaboration of three major public transportation operators: ATAF, BUSITALIA and CTTNORD. Most of the computations were conducted in R Statistical Environment (https://www.R-project.org/), and then implemented in real time. In addition, the same solution has been used in Snap4City mobile Apps with experiments performed in Antwerp and Helsinki on Android mobile Apps that are on Google Play. In those cases, the collection of data from the mobile have been authorized thanks to the signed consent according to the GDPR of Snap4City [Badii, et al., 2018b].

VI. Acknowledgements

The authors would like to thank the MIUR Smart City national founding SCN_00112 and to the University of Florence and companies involved for co-founding. Snap4City has been founded by Select4Cities PCP of the European Commission. Snap4City and Km4City are 100% open technologies and research of DISIT Lab. Thanks to the collaboration of three major public transportation operators in Italy for their support in putting the solution in operative conditions: ATAF, BUSITALIA and CTTNORD.

VII. References

[Ashqar et al., 2018] Ashqar, H. I., Almannaa, M. H., Elhenawy, M., Rakha, H. A., & House, L. (2018). Smartphone Transportation Mode Recognition Using a Hierarchical Machine Learning Classifier and Pooled Features From Time and Frequency Domains. IEEE Trans. on Intelligent Transportation Systems.

[Badii et al., 2017b] C. Badii, P. Bellini, D. Cenni, A. Difino, P. Nesi, M. Paolucci, "User Engagement Engine for Smart City Strategies", 3rd IEEE Int. Conf. on Smart Computing, 2017.

[Badii et al., 2018] C. Badii, P. Nesi, I. Paoli, "Predicting available parking slots on critical and regular services exploiting a range of open data", IEEE Access, 2018

[Badii, et al., 2018b] C. Badii, et al., "Snap4City: A Scalable IOT/IOE Platform for Developing Smart City Applications", Int. Conf. IEEE Smart City Innovation, Cina 2018, IEEE Press. https://ieeexplore.ieee.org/document/8560331/

[Biancat et al., 2014] Biancat, Jacopo, Chiara Brighenti, and Attilio Brighenti. "Review of Transportation Mode Detection techniques." ICST Trans. Ambient Systems 1, no. 4 (2014): e7.

[Biljecki et al., 2013] Biljecki, Filip, Hugo Ledoux, and Peter Van Oosterom. "Transportation mode-based segmentation and classification of movement trajectories." International Journal of Geographical Information Science 27.2 (2013): 385-407.

[Hemminki et al., 2013] Hemminki, S., Nurmi, P., & Tarkoma, S. (2013, November). Accelerometer-based transportation mode detection on smartphones. In Proc. of the 11th ACM Conf. on Embedded Networked Sensor Systems (p. 13). ACM.

[Lv et al., 2018] Lv, Zhihan, et al. "Government affairs service platform for smart city." Future Generation Computer Systems 81 (2018): 443-451.

[Manzoni et al., 2010] Manzoni, Vincenzo, et al. "Transportation mode identification and real-time CO2 emission estimation using smartphones." SENSEable City Lab, Massachusetts Institute of Technology, nd (2010).

[Misra et al., 2006] Misra, Pratap, and Per Enge. "Global Positioning System: signals, measurements and performance second edition." Massachusetts: Ganga-Jamuna Press (2006).

[Prelipcean et al., 2017] Prelipcean, Adrian C., Gyözö Gidófalvi, and Yusak O. Susilo. "Transportation mode detection–an in-depth review of applicability and reliability." Transport Reviews 37.4 (2017): 442-464.

[Reddy et al., 2010] Reddy, Sasank, et al. "Using mobile phones to determine transportation modes." ACM Transactions on Sensor Networks (TOSN) 6.2 (2010): 13.

[Stenneth et al., 2011] Stenneth, Leon, et al. "Transportation mode detection using mobile phones and GIS information." Proc. of the 19th ACM SIGSPATIAL International Conf. on Advances in Geographic Information Systems. ACM, 2011.

[Wang et al., 2010] Wang, Shuangquan, Canfeng Chen, and Jian Ma. "Accelerometer based transportation mode recognition on mobile phones." Wearable Computing Systems (APWCS), 2010 Asia-Pacific Conference on. IEEE, 2010.

[Yanyun et al., 2017] Yanyun, G., Fang, Z., Shaomeng, C., & Haiyong, L. (2017, September). A convolutional neural networks based transportation mode identification algorithm. In Indoor Positioning and Indoor Navigation (IPIN), 2017 International Conference on (pp. 1-7). IEEE.

[Yu et al., 2014] Yu, Meng-Chieh, et al. "Big data small footprint: the design of a low-power classifier for detecting transportation modes." Proc. of the VLDB Endowment 7.13 (2014): 1429-1440.

Towards an Intelligent System for Supporting Gesture Acquisition and Reproduction in Humanoid Robots

Agnese Augello
ICAR-CNR, Palermo, Italy
agnese.augello@cnr.it

Angelo Ciulla
ICAR-CNR, Palermo, Italy
angelo.ciulla@cnr.it

Alfredo Cuzzocrea
iDEA Lab, University of Calabria, Rende, Italy
alfredo.cuzzocrea@unical.it

Salvatore Gaglio
University of Palermo and ICAR-CNR, Palermo, Italy
salvatore.gaglio@unipa.it

Giovanni Pilato
ICAR-CNR, Palermo, Italy
giovanni.pilato@cnr.it

Filippo Vella
ICAR-CNR, Palermo, Italy
filippo.vella@cnr.it

Abstract

In this paper, an intelligent system for supporting gesture acquisition and reproduction in humanoid robots, which is based on the well-known Microsoft Kinect framework, is introduced and discussed in this paper. The idea that has inspired the paper is represented by endowing an humanoid robot with the capability to mimic the motion of a human user in real time. As a further extension, the latter amenity may serve as a basis for further gesture based human-robot interactions.

1 Introduction

Nowadays, the interaction between human beings and robots has become a very relevant issue in a wide range of applications (e.g., [15, 21, 19]). It is commonly agreed that communication between humans is based on both verbal and not verbal cues. A humanoid robot capable of interacting with people combining speech and gestures would dramatically increase the naturalness of social interactions. On the other hand, other studies like [18, 14, 8] consider *knowledge management techniques* (e.g., [8]) to improve this phase.

Furthermore, the Microsoft Kinect is a popular choice for any research that involves body motion capture. It is an affordable and low-cost device that can can be used for non invasive, marker-less tracking of body gestures. As an example, Baron et al. [5] controlled a Mindstorm NXT artificial arm with sensor Kinect, employing gesture recognition to regulate arm movement. Chang et al. [6] developed a Kinect-based gesture command control method for driving a humanoid robot to learn human actions, using a Kinect sensor and three different recognition mechanisms: dynamic time wrapping (DTW), hidden Markov model (HMM) and principal component analysis (PCA).

Meanwhile, Sylvain Filiatrault and Ana-Maria Cretu [12] used sensor Kinect to mimic the motion of a human arm to an NAO humanoid robot. In their case, the software architecture is based on three modules: Kinect Manager, Interaction Manager, and NAO manager. The Kinect Manager deals with the events and data captured by the Kinect. The class Kinect Transformer is used to get the Euler angles of the desired joints. The Interaction Manager is the intermediary between the Kinect and the robot and contains the repository for the joints used by the other two modules. The use of a joint repository of all articulations allows reducing the data to be processed as some joints are not needed. Finally, the NAO manager contains the static and dynamic constraints to apply to each one of the articulations, as well as the methods that allow the control of the robot movements.

To be sure that the robot has enough time to execute the gesture, a delay of 200 ms between one cycle and the next has been introduced. Itauma et al. [13] used a Kinect to

DOI reference number: 10.18293/DMSVIVA2020-017

teach an NAO robot some basic Sign Language gestures. The aim was teaching Sign Language to impaired children by employing different machine learning techniques in the process. Shohin et al. [16] used three different methods to make a robot NAO imitate human motion: direct angle mapping, inverse kinematics using fuzzy logic and iterative Jacobian.

In some cases, neural networks were used: Miguel et al. [17] used a Kinect sensor and a Convolutional Neural Network (CNN) trained with the MSRC-12 dataset [1] to capture and classify gestures of a user and send related commands to a mobile robot. The used dataset was created by Microsft and had 6244 gesture instances of 12 actions. To have gestures of the same length, without losing relevant information, the system used a Fast Dynamic Time Warping algorithm (FastDTW) to find the optimal match between sequences by non linearly warping them along the time axis. This resulted in all gestures normalized to sequences of 667 frames, with each frame having 80 variables, corresponding to the x,y,z values for each of the 20 joints, plus a separation value for each joint. The resulting 667x80 matrix is used as the input of the CNN, which classifies it in one of the 12 possible gestures. The CNN was trained using two strategies, combined training consisting of a single CNN to recognize all 12 gestures and individual training with 12 different CNN, each capable of recognizing only one gesture. The accuracy rates were 72.08% for combined training and 81.25% for the individual training.

Moreover, Unai et al. [20] developed a natural talking gesture generation behavior for $Pepper$ by feeding a Generative Adversarial Network (GAN) with human talking gestures recorded by a Kinect. Their approach in mapping the movements detected by Kinect on the robot is very similar to what we used, but while they feed the resulting values to a neural network (a GAN), we use the (filtered) values directly.

This paper reports the implementation of a system able to acquire and reproduce the gestures performed by a human during an interactive session. In our approach, we exploited a $MicrosoftKinect$ sensor to capture the motion data from a user and then, we have defined a mapping algorithm to allow a *SoftBank Pepper* robot to reproduce the tracked gestures as close as possible to the original ones.
In particular, we used the *OpenNi* driver for the *Kinect*, the NiTE 2.2 libraries for detecting the user skeleton, and the Kinetic version of ROS with the module pepper_dcm to provide package exchange and bridging between the computer and the robot and Ubuntu 16.04. We focused on the movements of the arms and the head, laying the basis for the extension of the same approach to the remaining parts. The extended version of this paper appears in [4].

2 The Proposed Solution

The developed system is structured in a set of modules, to increase versatility for future projects and to simplify possible extensions to the current project. Besides the Kinect itself, the first module is named *Viewer*, which extracts data frames (consisting of nine float values: three values for a joint position in 3D space, four values for quaternion from the origin and reliability values for both) from the Kinect and sends them in a pipe. The module also provides the feed of the Kinect camera with the overlay of the tracked user's skeleton. The pipe, long 8640 chars (64 chars for each joint value, 9 values for each joint, 15 joints total), is read by the second module, Gesture_Brain.

The *Gesture_Brain* module works both as a gateway for the ROS system [2] and as the module where actual data processing takes place. The gathered data cannot be used directly: a mapping is required to correctly associate each joint user position to the equivalent one in the Pepper robot. For this reason, the data is parsed and structured in a 15×9 float matrix, which is then separated into three matrices: one for coordinates, one for quaternions, and one for reliability values. In our algorithm, we decided to use only the first matrix for simplicity reasons, neglecting the quaternion matrix, and not performing any reliability check on the joints. We assume that the joint data is accurate enough for our purpose, as the Kinect already discards joints whose reliability values are too low. The joint position data is used to estimate Pepper joint angles, specifically shoulder pitch, shoulder roll, elbow roll and elbow yaw for both arms and head yaw for the head (there are three more joint angles that could be estimated, left and right wrist yaw and head pitch, but the Kinect is too imprecise to allow a good estimate, so they have been fixed to a value of 0). The details about this estimation are discussed in the next section.

After all required values are collected, we can use the ROS threads provided by the bridge pepper_dcm to send the joint angles to the robot. These threads consist of multiple joint angles divided into groups, each group representing a body part. As we are interested only in the movement of arms and head, we use three: head, left arm, right arm. The bridge reads the sent values and the time between each capture to dynamically calculate the gesture trajectory in real-time. This means that to allow the system to be as accurate as possible, the gesture should be executed quite slowly. The bridge itself was modified to activate the in-built Self Collision Avoidance (part of the NaoQi library) and to deactivate wait and breathe animations, as they interfere with the commands sent by the pepper_dcm.

```
[ Coordinates ] [ Quaternion ] [ C_conf ] [ Q_conf ]
```

Figure 1. Structure of a single row of the array sent by the Viewer module

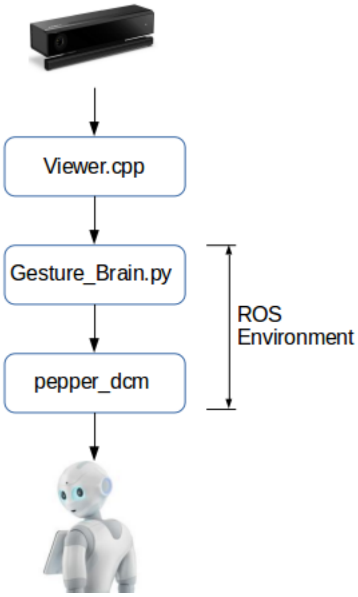


Figure 2. Structure of the algorithm

3 Mapping between User and Pepper

The *Pepper* robot has five Degree of Freedom for each arm (each one associated with a joint), unlike human beings who have seven. A mapping is thus required. From the *Kinect* the Cartesian coordinates for each joint, the quaternion for each segment (both referenced globally), and a reliability value for both are extracted. The bridge *pepper_dcm* uses Euler angles to communicate to the robot the new position of its joint angles. 3D space coordinates are thus used since quaternions have proven unsuitable. This is because the quaternions extracted do not represent the rotation from the previous frame, but rather the rotation from a reference quaternion. This leads to excessive inaccuracies once converted in Euler angles.

Let $\overline{x}$, $\overline{y}$ and $\overline{z}$ be the unit vectors for each axis, that is:

$$\overline{x} = (1,0,0)$$
$$\overline{y} = (0,1,0)$$
$$\overline{z} = (0,0,1)$$

Let S_L , E_L and W_L be the coordinates of the shoulder, the elbow and the wrist of the left arm respectively, $\overline{S_LE_L}$ and $\overline{E_LW_L}$ are defined as:

$$\overline{S_LE_L} = E_L - S_L$$

$$\overline{E_LW_L} = W_L - E_L$$

$SR_L$is the supplementary to the angle between $\overline{S_LE_L}$ and $-x$ axis:

$$SR_L = \frac{\pi}{2} - arcos(\overline{S_LE_L} \cdot -x) \quad (1)$$

$SP_L$is the angle between the projection of $\overline{S_LE_L}$ on zy plane and z axis, shifted in range to avoid the jump discontinuity at 180 and -180:

$$SP_L = \pi - mod_{2\pi}(\frac{3}{2}\pi + arctan(\overline{S_LE_L}_z, \overline{S_LE_L}_y) \quad (2)$$

For values of SR_L close to $\frac{\pi}{2}$, SP_L become unstable. As such, in the algorithm is assigned a value of 0 for $SR_L >$ 1.3.

$ER_L$is the angle between $\overline{E_LW_L}$ and $\overline{S_LE_L}$, shifted by $\frac{\pi}{2}$:

$$ER_L = \frac{\pi}{2} + arcos(\overline{E_LW_L} \cdot \overline{S_LE_L}) \quad (3)$$

$EY_L$is the angle between the projection of $\overline{E_LW_L}$ on zy plane and z axis, shifted in range for stability reasons, plus $-SP_L$:

$$EY_L = \pi - mod_{2\pi}(\frac{3}{2}\pi + arctan(\overline{E_LW_L}_z, \overline{E_LW_L}_y) - SP_L \quad (4)$$

The right arm is almost the same as the left arm, the only difference is that some angles have the opposite sign. Let $\overline{HN}$ be the difference between the coordinates of the joints H (head) and N:

$$\overline{HN} = H - N$$

The head yaw HY is the angle between the projection of $\overline{HN}$ on the xz plane and the z axis:

$$HY = -arctan(\overline{HN}_z, \overline{HN}_y) - \frac{\pi}{2} \quad (5)$$

3.1 Line of Best Fit

$Kinect$ joint detection is based on the shape of the user, which is redrawn at every frame. While calibrating the sensor helps to reduce the resulting jerkiness, there is still a significant amount of noise left. This noise can be approximately classified in two categories: a constant Gaussian noise caused by small alteration on the shape detected and large ”spikes” when the $Kinect$ fail to guess the position of one or more joints (especially common when part of the limb is outside of the frame or when two or more joints overlap). A simple way to compensate part of this noise is to use a line of best fit.

Given k points in (x, y) coordinates system, we must find the values c_0 and c_1 in the equation:

$$p(x) = c_0x + c_1$$

that define the straight line minimizing the squared error:

$$E = \sum_{j=0}^{k} |p(x_j) - y_j|^2$$

in the equations:

$$\begin{gathered} x_0c_0 + c_1 = y_0 \\ x_1c_0 + c_1 = y_1 \\ \dots \\ x_kc_0 + c_1 = y_k \end{gathered}$$

The result is a smoother movement, especially when Kinect is not able to detect the precise coordinates of a given joint. This is because, given a disturbing signal, the line of best fit can be seen as an approximation of the tangent that the signal would have at that point if the noise were removed. This is not always true, especially when the signal changes rapidly, but it's close enough in most cases to give a generally cleaner movement.

3.2 Modes of Operation

Besides mimicking the user movement, the Gesture_Brain module also has some additional features implemented to increase the breadth of experiments that can be performed with the system or to help with future projects. The behavior of the program is managed by the input arguments. These are, in order: mode, mirror_flag, json_file_name, LAjpos, RAjpos, Hjpos. The first one determines which of the three different modes of operation will be used (default 0), the second one determines if the mirror mode is activated or not (default false), the third defines the name of the text file used to record (in mode 0 and 1) or read (mode 2) the gestures (the default value is NULL, that is no recording) and set a flag (record_flag) to 1, the fourth, fifth and sixth ones are used to determine the pose to use in mode 1 (as default, the robot will spread its arms parallel to the ground, in a pose that in animation is known as ”T-pose”). More in details, the modes of operation of the main program are:

- Mode 0 or ”Mimic Mode”, is the default mode and makes the robot mimic the movements of the user. The record flag makes it, so the output is not just sent to the ROS publishers, but recorded in a JSON(JavaScript Object Notation) file, to be reproduced later. If the mirror flag is active, every movement is mirrored. In case both the record and mirror flags are active, the mirrored movement will be recorded and saved in the specified txt file.

- Mode 1, or ”Pose Mode”, make the robot execute a pose (defined at the beginning by the value of the given arguments) that the user must try to emulate. A distance algorithm calculates how close is the user pose to that of the robot, evaluated separately for the head, right upper arm, left upper arm, right forearm, and left forearm. If the user pose keeps all body parts below their respective thresholds (defined separately for each boy part), the program will communicate the success and shut down. The record flag makes it, so the distance values returned are written in a file, while the mirror flag makes it so the user must try to mirror the pose shown.

- Mode 2, or ”Playback Mode”, consists of reproducing a previously recorded gesture. The mirror flag, even if selected, doesn't have any effect on the algorithm. The name necessary to activate the record flag is used as the name of the file with the gesture to execute.

As an example, an experiment that was conceptualized consisted in using the Pepper robot to show a specific pose that the user must replicate as closely as possible. The experiment envisaged the use of both the normal mode and the mirror mode.

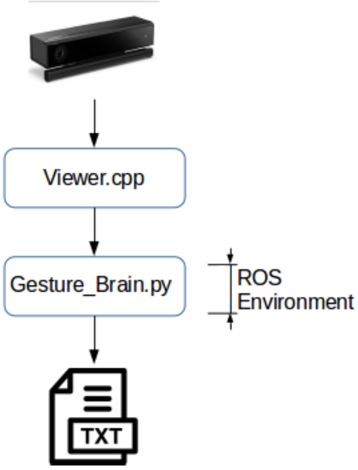


Figure 3. Structure of the algorithm when recording (the text file can be either the recorded gesture in mode 0 or the record of distance values in mode 1)

4 Conclusions and Developments

The system illustrated in this paper is capable of detecting the user poses with the Kinect with sufficient accuracy. The first experiments show that the reproduced movements are precise and smooth; the mirroring is accurate; the pose

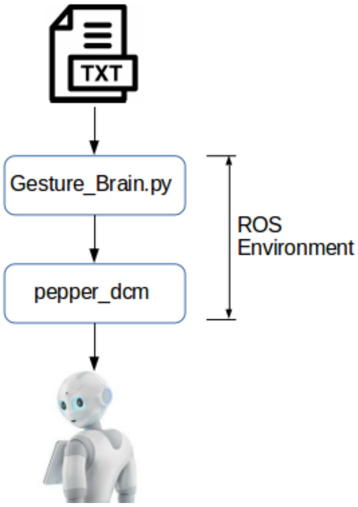


Figure 4. Structure of the algorithm in mode 2 (the txt file in this case is the coding of a previously recorded gesture)

evaluation is coherent; Furthermore, the recording and execution of the gestures are very close to the real-time movements. However, sometimes, certain positions cannot be reliably detected, due to imprecise behavior of the Kinect output when joints overlap each other, and to excessive reliance on the silhouette to detect the human body and the lack of joints in key points of the detected skeleton (like the hands). There is also an environmental factor, like lightning and positioning, that can make accurate user detection problematic. Currently, we are setting up two experiments: the first one is to make the robot autonomously capable of acting both as an instructor and a learner of the Semaphore Flag Signalling System [3], exploiting the gesture mirroring features; the second one is to make the robot capable of both encoding and decoding simple sentences from natural language to the flag semaphore system and vice-versa.

In future works, we plan to extend our framework as to deal with novel and emerging *big data trends* including performance (e.g., [10, 7]), and privacy and security (e.g., [9, 11]).

References

[1] Msrc-12 dataset, `https://www.microsoft.com/en-us/download/details.aspx?id=52283`.

[2] Ros kinetic, `http://wiki.ros.org/kinetic`.

[3] Semaphore flag signalling system, `https://en.wikipedia.org/wiki/Flag\_semaphore`.

[4] A. Augello, A. Ciulla, A. Cuzzocrea, S. Gaglio, G. Pilato, and F. Vella. A kinect-based gesture acquisition and reproduction system for humanoid robots. In *Computational Science and Its Applications - ICCSA 2020 - 20th International Conference, Cagliari, Italy, July 1-4, 2020, Proceedings*, 2020.

[5] G. Baron, P. Czekalski, D. Malicki, and K. Tokarz. Remote control of the artificial arm model using 3d hand tracking. In *2013 International Symposium on Electrodynamic and Mechatronic Systems (SELM)*, pages 9–10. IEEE, 2013.

[6] C.-w. Chang, C.-j. He, et al. A kinect-based gesture command control method for human action imitations of humanoid robots. In *2014 International Conference on Fuzzy Theory and Its Applications (iFUZZY2014)*, pages 208–211. IEEE, 2014.

[7] G. Chatzimilioudis, A. Cuzzocrea, D. Gunopulos, and N. Mamoulis. A novel distributed framework for optimizing query routing trees in wireless sensor networks via optimal operator placement. *J. Comput. Syst. Sci.*, 79(3):349–368, 2013.

[8] A. Cuzzocrea. Combining multidimensional user models and knowledge representation and management techniques for making web services knowledge-aware. *Web Intelligence and Agent Systems*, 4(3):289–312, 2006.

[9] A. Cuzzocrea and E. Bertino. Privacy preserving OLAP over distributed XML data: A theoretically-sound secure-multiparty-computation approach. *J. Comput. Syst. Sci.*, 77(6):965–987, 2011.

[10] A. Cuzzocrea, R. Moussa, and G. Xu. Olap*: Effectively and efficiently supporting parallel OLAP over big data. In *Model and Data Engineering - Third International Conference, MEDI 2013, Amantea, Italy, September 25-27, 2013. Proceedings*, pages 38–49, 2013.

[11] A. Cuzzocrea and V. Russo. Privacy preserving OLAP and OLAP security. In *Encyclopedia of Data Warehousing and Mining, Second Edition (4 Volumes)*, pages 1575–1581. 2009.

[12] S. Filiatrault and A.-M. Cretu. Human arm motion imitation by a humanoid robot. In *2014 IEEE International Symposium on Robotic and Sensors Environments (ROSE) Proceedings*, pages 31–36. IEEE, 2014.

[13] I. I. Itauma, H. Kivrak, and H. Kose. Gesture imitation using machine learning techniques. In *2012 20th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4. IEEE, 2012.

[14] M. C. Lau, J. Anderson, and J. Baltes. A sketch drawing humanoid robot using image-based visual servoing. *Knowledge Eng. Review*, 34:e18, 2019.

[15] C. A. Monje and S. M. de la Casa Díaz. Modeling and control of humanoid robots. *Int. J. Humanoid Robotics*, 16(6):1902003:1–1902003:3, 2019.

[16] S. Mukherjee, D. Paramkusam, and S. K. Dwivedy. Inverse kinematics of a nao humanoid robot using kinect to track and imitate human motion. In *2015 International Conference on Robotics, Automation, Control and Embedded Systems (RACE)*, pages 1–7. IEEE, 2015.

[17] M. Pfitscher, D. Welfer, M. A. d. S. L. Cuadros, and D. F. T. Gamarra. Activity gesture recognition on kinect sensor using convolutional neural networks and fastdtw for the msrc-12 dataset. In *International Conference on Intelligent Systems Design and Applications*, pages 230–239. Springer, 2018.

[18] P. Regier, A. Milioto, P. Karkowski, C. Stachniss, and M. Bennewitz. Classifying obstacles and exploiting knowledge about classes for efficient humanoid navigation. In *18th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2018, Beijing, China, November 6-9, 2018*, pages 820–826, 2018.

[19] S. Saeedvand, H. S. Aghdasi, and J. Baltes. Robust multi-objective multi-humanoid robots task allocation based on novel hybrid metaheuristic algorithm. *Appl. Intell.*, 49(12):4097–4127, 2019.

[20] U. Zabala, I. Rodriguez, J. M. Martínez-Otzeta, and E. Lazkano. Learning to gesticulate by observation using a deep generative approach. *arXiv preprint arXiv:1909.01768*, 2019.

[21] A. Zhang, I. G. Ramirez-Alpizar, K. Giraud-Esclasse, O. Stasse, and K. Harada. Humanoid walking pattern generation based on model predictive control approximated with basis functions. *Adv. Robotics*, 33(9):454–468.

On the impact of lightweight ciphers in automotive networks

Arcangelo Castiglione*, Francesco Palmieri†
Department of Computer Science
University of Salerno
arcastiglione@unisa.it*, fpalmieri@unisa.it†

Francesco Colace§‖, Marco Lombardi ¶,
Domenico Santaniello**
Department of Industrial Engineering
University of Salerno
fcolace@unisa.it§, malombardi@unisa.it¶,
dsantaniello@unisa.it**

Abstract—**The ever more exposure of modern vehicles to computer networks has led, in recent years, to increase risks due to cyberattacks. The internal computer network of vehicles, used to connect several electronic components present on cars, is even more threatened as it is potentially exposed to external attacks. Securing the Controller Area Network (CAN) protocol, used to govern those networks, is becoming increasingly important to ensure a safe driving experience.**

CAN is an ISO standard that dates back to 1983, over the years it has undergone very few changes coming to be outdated. It has been designed to minimize latency and data transmission errors through two essential features: small frames and unencrypted information transfer. The latter feature, in particular, appears to be the weak point of this protocol. Securing the communication channel is needed, but it must be done by preserving all the main features that ensure the performance of this protocol, in particular the low latency. Furthermore, the modification cannot introduce low-level alterations.

CAN security can be improved by acting at a higher level. In this work we investigate the feasibility of using symmetric encryption algorithms for securing messages exchanged on the CAN-bus. In particular, this paper evaluates the effectiveness of using lightweight ciphers, designed with the aim of introducing encryption also on devices, which have limited hardware and software resources, such as microcontrollers.

Keywords*—*Automotive security; Controller Area Network; Lightweight Cryptography; Cybersecurity; Encryption.

I. INTRODUCTION

We live in the Internet of Things (IoT) era, in which innumerable devices are interconnected and connected to the Internet, constantly exchanging information for several purposes [1]–[6]. In particular, IoT indicates a set of technologies that allow any type of device to be connected to the Internet and interact. The purpose of this type of solutions is varied [7]–[13] and could concerns monitoring, controlling and transferring information to perform consequent actions. The automotive field, not least, has progressively invested in the IoT paradigm. In fact, modern vehicles are equipped with many sensors that enhance their features and services. It is estimated that by 2022 new vehicles will be able to communicate with each other [14]. Modern vehicles are able to exchange information over the network using wireless technologies and cloud services. However, these communication channels expose the vehicle to potential vulnerabilities that may prejudice their functionality by attacking the internal network [15], [16].

Internet access exposes vehicles to potential attacks that can be different and varied. An attacker may be interested in spying on users to find personal information or even remotely control theirs vehicle. The Controller Area Network (CAN) represents the internal network of the vehicle which, in modern vehicles, consists of 70 nodes, which are the Electric Control Units (ECUs). The ECUs are responsible for controlling each component of the vehicle, from the ventilation system up to the braking system and steering [17], [18]. The communication channel, called CAN-bus, collects the exchange broadcast information between the various ECUs. Therefore, access to the CAN-bus would compromise the entire safety of the vehicle [19], [20].

The CAN-bus security issue has been addressed in the literature in several ways. One of the approaches is to redesign the CAN standard from a hardware perspective. However, this type of solution, with the exception of a substantial hardware upgrade, brings the disadvantage of not ensuring security to all vehicles already on the market [21]. Different approaches aim to equip the internal vehicle network with an Intrusion Detection System (IDS), which however, in many cases exceed the computational capacity of the microcontrollers [22], [23]. In attempt to preserve the exchange of information unchanged and considering the insufficient computational capability of microcontrollers, also possible solutions based on the Message Authentication Code (MAC) are to be excluded [24].

The aim of this work is to envision a solution for securing messages exchanged on the CAN-bus throught symmetric encryption algorithms. In particular, a solution using lightweight ciphers is proposed and evaluated. Lightweight ciphers provide encryption facilities on devices which have limited hardware and software resources, such as microcontrollers.

With the aim of extending our solution to all vehicles already on the market, the selected lightweight ciphers are PRESENT [25], SIMON and SPECK [26], which should produce safety improvement with little delay in communications. With the goal of evaluating the proposed approach, a prototype using Automotive Grade Linux (AGL) [27] was developed to evaluate the impact of the proposed solution on system performance.

This paper is organized as follows. In Section II we provide the background underlying the proposed approach. In Section III we provide our solutions in order to protect modern vehicles. In Section IV we describe a proof of concept of our proposed solution. In Section V we discuss a preliminary test activity.

[‖] Corresponding author: Francesco Colace, Department of Industrial Engineering, University of Salerno, fcolace@unisa.it, Via Giovanni Paolo II, 132 I-84084 Fisciano (SA), ITALY. Phone: +39089964256, Fax: +39089964218

DOI reference number: 10.18293/DMSVIVA20-010

Finally, in Section VI we draw conclusions and future research directions.

II. Background

A. The Controller Area Network Standard

The Controller Area Network bus, also known as CAN-bus, is a robust bus standard used to allow the interconnection of Electronic Control Units (ECUs), which controls all electronic vehicle systems [28].

The CAN-bus is used manly in automotive sector, although it is adopted also in other applications, such as industrial automation or healthcare industry.

The CAN protocol was designed around the 1980s at the Bosh laboratories. In the following years, Bosh released a new version of the protocol that was published as the ISO 11898 standard [29]. The protocol was designed to interconnect ECUs components, one by one in an economic and efficient way. The CAN is based on a broadcast network through which all the ECUs listen and communicate the new messages. In this way, the CAN protocol preserves the efficency needed in real-time applications such as automotive sector.

1) CAN Architecture: CAN is designed to minimize the number of messages reducing overhead and lost messages amount, thus ensuring as efficient as possible communication. The protocol provides a hierarchy among the messages exchanged on the channel, through priorities avoiding, of critical messages during simultaneous transmission of other less important signals. Frames (also called messages) perform the actual data transmission. There are four types of frames that can be transmitted on the CAN-bus: data, remote, error and overload.

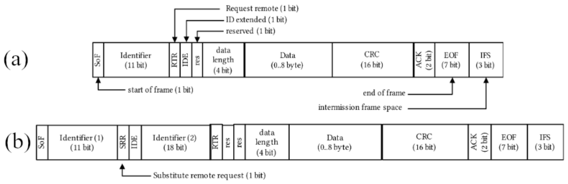


Fig. 1: Data Frames CAN 2.0: (a) Standard, (b) Extended.

The messages can be in two formats: basic frame format (shown in Fig. 1a) with 11 bits and extended frame format (shown in Fig. 1b) with 29 identification bits. The data frame is responsible for transmitting data and no more than 64 bits of information can be transferred without encryption.

In the design phase, the strong presence of electromagnetic interference in the operating environment and the presence of priority relationships between the messages were not neglected. In fact, there is a very low probability that the messages exchanged on the CAN-bus suffer alterations due to the city electromagnetic fields. It is also possible to define priorities, so as to prevent critical signals from being lost due to the simultaneous transmission of other less important messages. The interconnection is made physically through the use of a twinned pair to which all nodes involved are connected. The latter consists of a CPU, capable of decoding incoming messages and encoding those in output, by a CAN controller, which deals with the recomposition of the input message and the serialization of the output bits towards the CAN-bus (often this module is integrated in the CPU), and by a transceiver, which carries out a translation of the data stream in a format understandable by the CAN controller.

B. Automotive Grade Linux

The variety of software and hardware of the vehicles on the market is vast. Each manufacturer tends to propose different solutions both for mechanical and electronic system, which govern the vehicles. This generates fragmentation by not allowing the use of unique software. Automotive Grade Linux arises against the software segmentation present in the automotive sector, offering an open source solution based on Linux [30]. AGL promises to improve safety and entertainment of vehicles through APIs present at all levels. The software started in 2012 by the Linux Foundation with the support of distinguished companies of automotive sector. Today AGL counts on more than hundred participating companies. Toyota was the first, in 2018, to produce Carmy, the first vehicle with AGL on board, which represents one of very successful sedans in the world.

1) Message Exchange: Automotive Grade Linux (AGL) offers an API service, which allows managing the CAN-bus interaction as shown in Fig. 2. This service is divided as follow: CAN High Level Binding(s), which deals with the applications user layer that offers logic and advanced functions; CAN Low Level Binding(s), which interacts with low-level communication, such as encoding and decoding of messages, security checks, transactions, etc. AGL, which presents an architecture similar to the one defined in the OpenXC project [31] designed by Ford, offers convenient management of interactions with the CAN bus. The API for the exchange of messages is divided

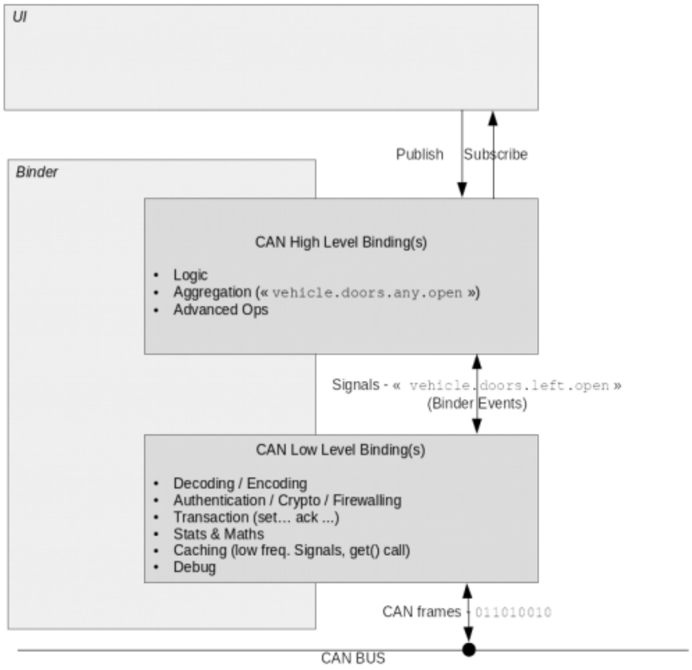

UI
Publish
Subscribe
Binder
CAN High Level Binding(s)
• Logic
• Aggregation (« vehicle.doors.any.open »)
• Advanced Ops
Signals - « vehicle.doors.left.open »
(Binder Events)
CAN Low Level Binding(s)
• Decoding / Encoding
• Authentication / Crypto / Firewalling
• Transaction (set... ack ...)
• Stats & Maths
• Caching (low freq. Signals, get() call)
• Debug
CAN frames - 011010010
CAN BUS

Fig. 2: Exchanging Message Architecture [27].

into two levels, that is high and low. The former deals with the logic and advanced functions that are offered to user applications, whereas the latter carries out security checks, transactions, caching, encoding and decoding of messages. The system is in fact designed to extract the various messages that can be admitted starting from a JSON file. Therefore, the applications require the transmission of a signal simply by using the name assigned to them during compilation. The lowest level of the CAN API then deals with the translation of this string into a frame of data and vice versa. The services are defined in files that are used by the afb-daemon binder and the latter can be queried via HTTP or WS specifying the name of the API to be used followed by the desired method. The actual writing and reading on/from the communication channel is done using the native Linux drivers.

III. Proposed Solution

This section deals with the proposed solution, which aims to protect connected vehicles from possible attacks. In fact, an attacker, through the use of malware, could attacks and controls an entire fleets of vehicles. Furthermore, it is necessary pay careful to make changes that do not compromise the performance of the protocol. In other words, security must be increased without compromising the efficient execution of message exchange. [32]

Four types of messages can be exchanged on the CAN bus, but the only one on which encryption could be applied is the data frame, which is involved in the exchange of information. However, the length of the data field is often of variable length and does not exploit all transmissible bytes. In this regard, to improve the resistance of the system to brute force attacks, it is necessary to use all the 64 bits available, using dummy data in order to introduce noise.

However, in the use of cryptographic algorithms, which would improve vehicle safety, there are very important limitations. Microcontrollers, i.e., hardware with limited memory capacity, the need to preserve the performance of the CAN protocol and the limited amount of data to be encrypted are the main limitations to consider when choosing ciphers. Lightweight ciphers are the best solution in the conditions described above. The main feature of lightweight encryption is to find the best compromise between security and lightness, reducing the resources needed to execute the algorithm, both in terms of space and time. In our case, due to the type of application, it is necessary to use ciphers capable of working on low performance hardware and with a particularly high encryption and decryption speed. These main features are peculiar to be implemented in the automotive sector. Some of the most well-known lightweight ciphers are PRESENT [25], SIMON and SPECK [26].

A. Protection of CAN Data Frames

The field to consider, to prevent that an attacker takes control of the vehicle, is the data frames field, which contains information exchanged to all nodes on the internal network. Therefore, encrypting this field could be crucial for security improvements. We remark that for performance reasons, the protocol tends to transmit the minimum number of bits needed for each message. In fact, most of the messages exchanged on the bus do not use all the 64 bits available. However, for reasons related to security and correct use of ciphers, it is necessary to include dummy data to fill the entire field. It must not be forgotten that this operation could compromise the performance of the CAN protocol.

The proposed solution includes the using of a fixed key [33], [34], usable for the entire life of the vehicle or to be changed at regular intervals, to encrypt the messages exchanged on the CAN-bus. In fact, this solution does not imply hardware changes or synchronization problems between the different components responsible for communication. This key must be generated during production, remaining secret for the entire life of the vehicle.

However, if an attacker could discover the key, the safety of the entire vehicle would be compromised [35]. In this regard, the use of a fixed key is not very recommendable and this key should be changed at regular intervals.

The proposed solution involves the addition of a small operation during the production phase of the vehicle. Using a fixed key implies a little change in the production chain between the vehicle manufacturer and electronic component suppliers. More precisely, the latter will have to provide a software for flashing the firmware of the individual ECU, allowing the inclusion of the key during the assembly phase. Finally, we point out that although the key must be kept secret, however, it must be available for technical interventions such as replacing an ECU.

IV. Experimental Phase

To implement the proposed solution it is necessary to modify the modules dedicated to writing and reading messages on the CAN-bus, present in Automotive Grade Linux. It was therefore necessary to modify these modules by inserting an encryption phase after writing and a decryption phase before the reading. AGL is an open source environment, so it allows us to clone the source of the modules by modifying them as necessary.

In AGL, the *low CAN service* module is responsible for communication with the CAN-bus, through which it is possible to read, write and subscribe signals via HTTP or WS interface. This module could be modified at various points, to make the safety improvements proposed. The appropriate solution concerns to intervene at the binding definition level, making changes to the functions responsible of sending and receiving a message, without changing the embedded libraries.

A. Message Encryption

The binding is defined in the file **low-can-binding/binding/low-can-cb.cpp** and is designed in such a way that the writing of messages is managed by two different functions, depending on the mode used. In particular, the function **write_signal()** is invoked when we specify only the name and value of the signal to be sent. The writing of a raw frame instead passes through the function **write_raw_frame()**. Both of the functions mentioned above invoke the function **send_frame()**, which represents an excellent point to enter and carry out the encryption before the frame is passed to the sending module [36].

Listing 1: Encryption in writing phase

```
static int send_frame(struct canfd_frame&
    cfd, const std::string& bus_name)
```

```
{
  if(bus_name.empty()) {
    return -1;
        }

  std::map<std::string, std::shared_ptr<
      low_can_subscription_t> >& cd =
      application_t::instance().
      get_can_devices();

  if( cd.count(bus_name) == 0)
    {cd[bus_name] = std::make_shared<
        low_can_subscription_t>
    (low_can_subscription_t());}

  //Start Encryption

  encrypt(cfd);

  //End Encryption

  return cd[bus_name]->tx_send(cfd,
      bus_name);
}
```

The modification consists simply in accomplishing, before the end of the function **send_frame()**, the encryption of the field **data** and eventually the modification of the value of the field **len**, keeping it consistent with the number of bytes actually used in **data**. In fact, the structure received as a parameter by the function **send_frame()** has several fields, among which there is **data** containing the raw information of the message and **len** which indicates the length of the message. In the code shown in Listing 1, the structure is passed as a parameter to the function **encrypt()** that inside it must provide for the implementation of the chosen cipher scheme and, if necessary, for the insertion of *dummy data* and for the modification of the **len** field.

B. Message Decryption

Intervention in reading requires a modification to the class **can_message_t** in order to make the vector editable **data_**. In fact, in this field the raw information is processed at a higher level. The modification concerns both the header in which the class is defined, and the file in which the behavior of the various methods is defined. The files in question are, respectively, **low-can-binding/can/can-message.hpp** and **low-can-binding/can/can-message.cpp** [36].

Listing 2: Adding set in prototype in **can-message.hpp**

```
void set_data_vector(std::vector<uint8_t
    >&);
```

Listing 3: Implementation of the method in **can-message.cpp**

```
void can_message_t::set_data_vector(std::
    vector<uint8_t>& data)
{
  data_ = data;
}
```

Once the problem of updating the data field has been solved, it is possible to intervene again in the definition of the binding, so as to define the decryption operation. In practice, the point in which to make changes is within the function **read_message()**.

Listing 4: Decryption in reading phase

```
int read_message(sd_event_source *
    event_source, int fd, uint32_t revents
    , void *userdata)
{
  low_can_subscription_t* can_subscription
      = (low_can_subscription_t*)userdata
      ;
  if ((revents & EPOLLIN) != 0)
  {
    can_message_t cm;
    utils::socketcan_bcm_t& s =
        can_subscription->get_socket();
    s >> cm;

    // Sure we got a valid CAN message ?
    if(! cm.get_id() == 0 && ! cm.
        get_length() == 0)
      {
        //Start Encryption
        std::vector<uint8_t> data (
            CANFD_MAX_DLEN);

        data = decrypt(cm.get_data_vector
            ());

       cm.set_data_vector(data);
        //End Encryption

        push_n_notify(cm);
      }
  }

  // check if error or hangup
  if ((revents & (EPOLLERR|EPOLLRDHUP|
      EPOLLHUP)) != 0)
  {
    sd_event_source_unref(event_source);
    can_subscription->get_socket().close()
        ;
  }

  return 0;
}
```

It remains to define the encryption mode, according to those chosen for the encryption process.

V. Experimental Results

The function **system_time_us()** is used to obtain the delay between the reading and writing phases of a message exchanged on the CAN-bus. This function is used before the encryption and after decryption phases, providing the time spent during the entire process. The time obtained represents the delay brought by the use of the proposed methodology.

The average of time obtained is between 300 μs and 1ms and, according to [37], it is more efficient to use hardware-oriented ciphers than the software one. To give an idea of the two effects to the delay generated by our solution, we can refer to a fundamental phase of driving a vehicle, which is the braking phase. The proposed solution generates delays in communication, which we can influence the various phases of driving the vehicle. Even if these delays are very low, to appreciate their significance we consider a crucial phase, which is the braking phase, transforming the delay in communication in terms of distance accumulated. However, even if, in the worst case, 1ms represents a very low value, it causes an extra distance to travel of about 40 mm, in terms of break reaction if a driver was travelling at 150 Km/h of speed. The delays obtained in terms of distance, in the best (300 μs) and in the worst (1 ms) case, are reported in Table I.

TABLE I: Excess distance due to the delay introduced by the encryption

Speed [Km/h]	Distance (300 μs) [mm]	Distance (1ms) [mm]
50	4.2	13.9
100	8.3	27.8
150	12.5	41.7
200	16.7	55.5
250	20.8	69.4
300	25.0	83.3

According to the Table I, the results are encouraging. There are two main causes of the accumulated delay during the encryption and decryption phases, that are the algorithm used and the hardware on which the system is installed. The realized prototype has been installed on a Raspberry Pi 3 to simulate a generic ECU. However, on the vehicles present in today's market, microcontrollers are installed that could be even less efficient.

VI. Conclusions

The study conducted in this paper deals with a cybersecurity problem in the automotive sector. In particular, was analysed the problem of the exchange of unencrypted messages on the CAN-bus and the feasibility of encryption that at least manages to prevent attacks on vehicles. According to the literature, an attacker who remotely penetrates into a vehicle's communication channel has the same control level of the driver. In this paper was analysed how important it can be to secure the CAN-bus, by obfuscating messages without altering their structure. This choice arises from the desire to keep the proposed solutions compatible with all the devices on the market, opening, in fact, to the possibility of correcting the security problem even on vehicles that already populate our roads. It has therefore been seen how message encryption is possible for all transmitted identifiers, as well as for sections containing data. We remark that due to the application field, the choice of the cipher is rather important, ensuring the least possible delay. Therefore, the use of lightweight ciphers has been evaluated. The study also focused on the implementation of a prototype through Automotive Grade Linux operating system and the interactions between the latter and the CAN-bus, achieving reasonable results. In fact, the maximum delay, introduced of our solution, is 1 ms and if the vehicle travels at 150 Km/h of speed, the extra distance generated by the delay is about 40 mm, which is a reasonable value. Further considerations on error handling in the decryption phase are demanded to future steps.

Acknowledgment

The authors would like to thank Aniello Carotenuto, who developed a preliminary version of our proposal.

References

[1] K. Ashton *et al.*, "That internet of things thing," *RFID journal*, vol. 22, no. 7, pp. 97–114, 2009.

[2] F. Colace, M. Lombardi, F. Pascale, D. Santaniello, A. Tucker, and P. Villani, "Mug: A multilevel graph representation for big data interpretation," in *2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 2018, pp. 1408–1413.

[3] Q. Zhang, L. T. Yang, A. Castiglione, Z. Chen, and P. Li, "Secure weighted possibilistic c-means algorithm on cloud for clustering big data," *Information Sciences*, vol. 479, pp. 515–525, 2019.

[4] G. DAngelo, R. Pilla, J. B. Dean, and S. Rampone, "Toward a soft computing-based correlation between oxygen toxicity seizures and hyperoxic hyperpnea," *Soft Computing*, vol. 22, no. 7, pp. 2421–2427, 2018.

[5] M. Carratù, M. Ferro, A. Pietrosanto, and P. Sommella, "Wireless sensor network for low-cost air quality measurement," in *Journal of Physics: Conference Series*, vol. 1065, no. 19. IOP Publishing, 2018, p. 192004.

[6] F. Clarizia, F. Colace, M. Lombardi, F. Pascale, and D. Santaniello, "A multilevel graph approach for road accidents data interpretation," in *International Symposium on Cyberspace Safety and Security*. Springer, 2018, pp. 303–316.

[7] F. Amato, N. Mazzocca, F. Moscato, and E. Vivenzio, "Multilayer perceptron: An intelligent model for classification and intrusion detection," in *2017 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA)*. IEEE, 2017, pp. 686–691.

[8] G. D'Aniello, M. de Falco, and M. Sergio, "Analysis of the collective perception using granular computing and virtual sensors," in *2018 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS)*. IEEE, 2018, pp. 1–6.

[9] M. Carratù, M. Ferro, A. Pietrosanto, P. Sommella, and V. Paciello, "A smart wireless sensor network for pm10 measurement," in *2019 IEEE International Symposium on Measurements & Networking (M&N)*. IEEE, 2019, pp. 1–6.

[10] B. Carpentieri, A. Castiglione, A. De Santis, F. Palmieri, and R. Pizzolante, "One-pass lossless data hiding and compression of remote sensing data," *Future Generation Computer Systems*, vol. 90, pp. 222–239, 2019.

[11] 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*. Springer, 2020, pp. 353–362.

[12] F. Colace, M. Lombardi, F. Pascale, and D. Santaniello, "A multi-level approach for forecasting critical events in smart cities," in *The 24th International DMS Conference on Visualization and Visual Languages*, 2018, pp. 45–46.

[13] 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)*, 2018, pp. 40–47.

[14] S. Kulandaivel, T. Goyal, A. K. Agrawal, and V. Sekar, "Canvas: Fast and inexpensive automotive network mapping," in *28th {USENIX} Security Symposium ({USENIX} Security 19)*, 2019, pp. 389–405.

[15] C.-W. Lin and A. Sangiovanni-Vincentelli, "Cyber-security for the controller area network (can) communication protocol," in *2012 International Conference on Cyber Security*. IEEE, 2012, pp. 1–7.

[16] D. S. Fowler, M. Cheah, S. A. Shaikh, and J. Bryans, "Towards a testbed for automotive cybersecurity," in *2017 IEEE International Conference on Software Testing, Verification and Validation (ICST)*. IEEE, 2017, pp. 540–541.

[17] T. Hoppe, S. Kiltz, and J. Dittmann, "Security threats to automotive can networkspractical examples and selected short-term countermeasures," *Reliability Engineering & System Safety*, vol. 96, no. 1, pp. 11–25, 2011.

[18] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham *et al.*, "Experimental security analysis of a modern automobile," in *2010 IEEE Symposium on Security and Privacy*. IEEE, 2010, pp. 447–462.

[19] H. Onishi, "Paradigm change of vehicle cyber security," in *2012 4th International Conference on Cyber Conflict (CYCON 2012)*. IEEE, 2012, pp. 1–11.

[20] J. Reilly, S. Martin, M. Payer, and A. Bayen, "On cybersecurity of freeway control systems: Analysis of coordinated ramp metering attacks," *Transportation Research, Part B*, 2014.

[21] R. Li, C. Liu, and F. Luo, "A design for automotive can bus monitoring system," in *2008 IEEE Vehicle Power and Propulsion Conference*. IEEE, 2008, pp. 1–5.

[22] H. M. Song, H. R. Kim, and H. K. Kim, "Intrusion detection system based on the analysis of time intervals of can messages for in-vehicle network," in *2016 international conference on information networking (ICOIN)*. IEEE, 2016, pp. 63–68.

[23] M. Casillo, S. Coppola, M. De Santo, F. Pascale, and E. Santonicola, "Embedded intrusion detection system for detecting attacks over canbus," in *2019 4th International Conference on System Reliability and Safety*. IEEE, 2019, pp. 136–141.

[24] R. Zalman and A. Mayer, "A secure but still safe and low cost automotive communication technique," in *Proceedings of the 51st Annual Design Automation Conference*. ACM, 2014, pp. 1–5.

[25] A. Bogdanov, L. R. Knudsen, G. Leander, C. Paar, A. Poschmann, M. J. Robshaw, Y. Seurin, and C. Vikkelsoe, "Present: An ultra-lightweight block cipher," in *International workshop on cryptographic hardware and embedded systems*. Springer, 2007, pp. 450–466.

[26] R. Beaulieu, S. Treatman-Clark, D. Shors, B. Weeks, J. Smith, and L. Wingers, "The simon and speck lightweight block ciphers," in *2015 52nd ACM/EDAC/IEEE Design Automation Conference (DAC)*. IEEE, 2015, pp. 1–6.

[27] "Automotive grade linux," https://www.automotivelinux.org/.

[28] R. I. Davis, A. Burns, R. J. Bril, and J. J. Lukkien, "Controller area network (can) schedulability analysis: Refuted, revisited and revised," *Real-Time Systems*, vol. 35, no. 3, pp. 239–272, 2007.

[29] R. Bosch *et al.*, "Can specification version 2.0," *Rober Bousch GmbH, Postfach*, vol. 300240, p. 72, 1991.

[30] M. Amiri-Kordestani and H. Bourdoucen, "A survey on embedded open source system software for the internet of things," in *Free and Open Source Software Conference*, vol. 2017, 2017.

[31] M. J. Cronin, "Ford finds its connection," in *Top Down Innovation*. Springer, 2014, pp. 13–24.

[32] A. Castiglione, F. Palmieri, F. Colace, M. Lombardi, D. Santaniello, and G. DAniello, "Securing the internet of vehicles through lightweight block ciphers," *Pattern Recognition Letters*, 2020.

[33] L. Zhou, Q. Wang, X. Sun, P. Kulicki, and A. Castiglione, "Quantum technique for access control in cloud computing ii: Encryption and key distribution," *Journal of Network and Computer Applications*, vol. 103, pp. 178–184, 2018.

[34] A. Castiglione, A. De Santis, A. Castiglione, and F. Palmieri, "An efficient and transparent one-time authentication protocol with non-interactive key scheduling and update," in *2014 IEEE 28th International Conference on Advanced Information Networking and Applications*. IEEE, 2014, pp. 351–358.

[35] A. Castiglione, A. De Santis, B. Masucci, F. Palmieri, and A. Castiglione, "On the relations between security notions in hierarchical key assignment schemes for dynamic structures," in *Australasian Conference on Information Security and Privacy*. Springer, 2016, pp. 37–54.

[36] A. Castiglione, F. Colace, M. Lombardi, F. Palmieri, and D. Santaniello, "Lightweight ciphers in automotive networks: A preliminary approach," in *2019 4th International Conference on System Reliability and Safety*. IEEE, 2019, pp. 142–147.

[37] P. Mundhenk, A. Paverd, A. Mrowca, S. Steinhorst, M. Lukasiewycz, S. A. Fahmy, and S. Chakraborty, "Security in automotive networks: Lightweight authentication and authorization," *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, vol. 22, no. 2, p. 25, 2017.

Self-training algorithm combining density peak and cut edge weight

Yang Liu
School of Statistics
Chongqing University
Chongqing 610031, China
statsyangliu@163.com

Abstract： In view of the influence of mislabeled samples on the performance of self-training algorithm in the process of iteration, a self-training algorithm based on density peak and cut edge weight is proposed. Firstly, the representative unlabeled samples are selected for labels prediction by space structure, which is discovered by clustering method based on density of data. Secondly, cut edge weight is used as statistics to make hypothesis testing. This technique is for identifying whether samples are labeled correctly. And then the set of labeled data is gradually enlarged until all unlabeled samples are labeled. The proposed method not only makes full use of space structure in formation, but also solves the problem that some data may be classified incorrectly. Thus, the classification accuracy of algorithm is improved in a great measure. Extensive experiments on real datasets clearly illustrate the effectiveness of proposed method.

Key words: self-training; density; cut edge weight; hypothesis testing

I. Introduction

Data classification is a very active research direction in the field of machine learning. In order to train an effective classifier, traditional supervised classification methods often require a large number of labeled samples. However, in practical applications, the acquisition of labeled samples requires a large price and is not easy to obtain, and the acquisition of unlabeled samples is relatively easy. Therefore, when the number of labeled samples is small, supervised classification methods are difficult to train an effective classifier. (Dong et al., 2016; Zhu, 2017) In this case, the semi-supervised classification method, which requires only a small number of labeled samples and makes full use of a large number of unlabeled samples, has attracted more and more attention. (Liu et al., 2019; Tanha et al., 2017) Self-training is one of the commonly used methods in semi-supervised classification. First, an initial classifier is trained with a small number of labeled samples, and the unlabeled samples are classified. Then, select unlabeled samples with higher confidence and their predicted labels, expand the labeled sample set, and update the classifier. These two processes continue to iterate until the algorithm converges.

(Pavlinek & Podgorelec, 2017; Vijayan et al., 2016; Xu et al., 2017)Self-training methods do not require any specific assumptions, are simple and effective, and have been widely used in many fields such as text classification, face recognition, biomedicine, and so on. But self-training classification algorithms also have some drawbacks, such as the classification performance is Based on the ST-DP algorithm, this paper proposes a Self-training method based on density peak and cut edge weight (ST-DP-CEW). This method not only selects unlabeled samples, uses the density clustering-based method to discover the underlying spatial structure of the data set, and selects representative samples for label prediction. Further, the correctness of the predicted labels can be identified by using the statistical method of cutting edge weights. Cutting edge weights and density peak clustering make full use of the sample spatial structure and unlabeled sample information, solve the problem of some samples being labeled incorrectly, reduce the accumulation of errors during iteration, and can effectively improve the performance of the classifier.

II.Algorithm construction

In this paper, we improve the classification accuracy of the self-trained semi-supervised classification algorithm by starting with the wrongly labeled samples during the self-training process. Based on ST-DP, the ST-DP-CEW algorithm is proposed. First, the spatial structure of the data set is discovered by density clustering method, and labeled.

DOI reference number: 10.18293/DMSVIVA20-004

1.Spatial structure of data

In this paper, let $L=\{(x_i,y_i)\}$ be the labeled sample set, where x_i is the training sample, and y_i is its label. $y_{i1}\in\{\omega_1,\omega_2,\cdots,\omega_s\}$, $i=1,2,\cdots,m$. S is the number of categories. $U=\{x_{m+1},x_{m+2},\cdots,x_n)$ is the unlabeled sample set. The local density of sample x_i is defined as follows:

$$\rho_i=\sum\chi\left(d_{ij}-d_c\right)$$

Among them:

$$\chi(x)=\begin{cases}1, & x<0\\ 0, & x\geq 0\end{cases}$$

d_{ij} is the Euclidean distance between samples x_i and x_i , and d_c is called the truncation distance. It is a constant that has no fixed value and is related to the data set itself(Wang & Xu, 2017). After calculating the ρ_i value of each sample x_i , find the sample x_{j} that is closest to sample x_i and has a greater local density, point x_i to x_{j} , and find the spatial structure of the data set.

2. Statistical method of cutting edge weights

(Triguero et al., 2014)Trim weighting is a method to identify and process mislabeled samples. First, in order to illustrate the similarity of the samples, a relative adjacency graph is established on the data set. The two samples x_i and x_{j} are connected side by side, if the following conditions are met: $d(x_i,x_j)\leq\max\left(d(x_i,x_m),d(x_j,x_m)\right),\forall m\neq i,j$,Where $d(x_i,x_j)$ is the distance between samples x_i and x_{j}. In an adjacency graph, if two samples with edges connected by different labels, this edge is called a cut edge. In an adjacency graph, if two samples with edges connected by different labels, this edge is called a cut edge. If x_i has many cut edges, that is, most of the samples in the neighborhood have labels that are different from those of x_i , it is considered that it may be labeled incorrectly. Therefore, cut edges play an important role in identifying mislabeled samples. For different samples, they may have the same number of cutting edges, but the importance of each cutting edge is different, so each edge in the adjacent graph is given a weight. Let W_{ij} be the weight of the edges connecting samples x_i and x_{j} .

. Finally, the hypothesis test was used to identify whether sample x_i was labeled incorrectly. The sum of the trimming weights J_i of sample x_i is defined as follows:

$$J_i=\sum_{j=1}^{n_l}w_{ij}I_i(j)$$

Among them,

$$I_i(j)=\begin{cases}1, & y_i\neq y_j\\ 0, & y_i=y_j\end{cases}$$

$\mathbf{n}_i$ is the number of samples with edges connected to sample x_i , and y_i is the label of sample x_i . If the J_i value of the sample x_i to be tested is large, it is considered that the sample may be labeled incorrectly. For hypothesis testing, the null hypothesis is defined as follows:

H_0 : All samples in the adjacent graph are labeled independently of each other according to the same probability distribution pro_y. pro_y represents the probability that the sample label is y .In order to do a bilateral test, you must first analyze the distribution of J_i under H_0 . Under the null hypothesis, $I_i(j)$ is an independent identically distributed random variable subject to a Boolean parameter of $1-pro_{y_i}$. So the expected μ_0 and variance σ^2 of J_i under H_0 are:

$$\mu_0=\left(1-\mathrm{pro}_{y_i}\right)\sum_{j=1}^{n_i}w_{ij}$$

$$\sigma^2=\mathrm{pro}_{y_i}\left(1-\mathrm{pro}_{y_i}\right)\sum_{j=1}^{n_i}w_{ij}^2$$

J_i follows the normal distribution $J_i\sim N\left(\mu_0,\sigma^2\right)$ under the original hypothesis H_0 , so the selected test statistic is

$$u=\frac{J_i-\mu_0}{\sigma}$$

Given a significance level of α , the rejection domain is:

$$W=\{|\,u\,|\geq u_{1-\alpha/2}\}$$

The rejection domain that gets the sum of the trimming weights is

$$W=\left[-\infty,\mu_0-\sigma\cdot u_{1-\alpha/2}\right]\cup\left[\mu_0+\sigma\cdot u_{1-\alpha/2},+\infty\right]$$

The main steps of the algorithm for identifying wrongly labeled samples using the edge-cut weights statistical method are as follows:

Step1. Establish a relative adjacency graph for the sample set, and initialize the labeled sample set correctly.

Step2. Assign weights to each edge in the adjacency graph.

Step3. Given the significance level, calculate the rejection domain.

Step4. If the value 1 falls into the rejection domain, the tag is correct and the correct tag set is updated; if it is not in the rejection domain, the wrong tag set is updated.

Step5. Repeat the above steps until all samples are tested.

3. Weight selection

The weight of each edge plays an important role in the statistical method of the edge weight. In this paper, the weight is first used to normalize the other nearest neighbor distances in the neighborhood by using the maximum nearest neighbor distance of each sample. Then calculate the probability that the sample has the same label as each neighboring sample, which is the weight of the edge.

Use the k -th nearest neighbor sample distance of x_i to normalize the distance from the first $k-1$ adjacent samples to x_i , then the normalized distance is:

$$D\left(x_{i,j},x_i\right)=\frac{d\left(x_{i,j},x_i\right)}{d\left(x_{i,k},x_i\right)},\quad j=1,2,\cdots,k$$

The weight of each edge in the adjacency graph is:

$$w_{ij}=P\left(x_{i,j}\mid x_i\right)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{D\left(x_{i,j},x_i\right)}{2}\right)$$

4. Self-training algorithm based on density and trimming weights

Classifier-based methods have extremely high requirements for the partitioning of sample sets and the selection of learning algorithms. The selection of distance metrics and values based on the nearest neighbor method need to be set in advance. If it is not selected properly in advance, it will cause a judgment error and affect the final classification effect. In addition, neither of these two methods uses a lot of valuable information carried by unlabeled samples in the recognition process, which reduces the accuracy of recognition. The method of cutting edge weight statistics to identify wrongly labeled samples does not need to set any parameters in advance, and it can also make full use of the information of unlabeled samples. Therefore, in order to improve the classification accuracy of the self-training algorithm, this paper incorporates the method of cutting edge weights to statistically identify the wrong label samples into the ST-DP algorithm, and proposes the ST-DP-CEW algorithm. The algorithm first uses the density clustering method to discover the spatial structure of the data set, and uses the spatial result information to preferentially select representative unlabeled samples for label prediction during the iteration process, which improves the accuracy of predicting labels. Then use the method of cutting edge weight statistics to judge whether the prediction label is correct. Use the correctly labeled samples for the next training. The specific steps of the algorithm are described as follows:

Step1. Use the density clustering method to find the true space structure of the entire data set.

Step2. (a) Use KNN or SVM as the base classifier, and train an initial classifier with the initial labeled sample set;

(b) label prediction on the "next" unlabeled sample of all samples in;

(c) identify whether the "next" sample is correctly labeled by using the method of trimming edge weights to obtain a correctly labeled sample;

(d) Repeat (a) through (c) until all "next" samples of have been marked.

Step3. (a) Perform label prediction on the "previous" unlabeled samples of all the updated samples;

(b) Identify the "previous" sample using the edge-cut weighting statistical method to obtain the correct labeled sample, and then update the classifier;

(c) Repeat (a) and (b) until all "previous" samples of have been marked.

II. Experimental results and analysis

In order to illustrate the effectiveness of the algorithm, the proposed algorithm is compared with existing self-training algorithms on 8 real data sets. The datasets are derived from the KEEL database. Samples with missing values are deleted from the Cleveland and Dermatology datasets, and the rest of the datasets are not processed. The comparison algorithms used are: traditional self-training algorithms using KNN and SVM as classifiers, self-training classification algorithms based on fuzzy c-means clustering (ST-FCM), density-based self-training classification algorithms (ST-DP), and Self-training classification algorithm (ST-DE) based on differential evolution.

1. Implementation of the experiment

A ten-fold cross-validation strategy was used to perform experiments on the dataset using KNN and SVM as base classifiers. Take one fold as the test set and the remaining nine fold as the training set. In each experiment, 10% of the samples in the training set are randomly selected as the initial labeled sample set, and the rest are unlabeled sets. In order to ensure the accuracy of the experiment, the ten-fold cross-validation experiment was repeated ten times, and the average value of the ten experiments was finally selected as the final experimental result. Accuracy rate (AR), Mean accuracy rate (MAR), and SD-AR are used as comparison criteria for the classification performance of the algorithm. Calculated as follows:

$$AR = \frac{1}{N_{T_s}} \sum_{i=1}^{N_{T_S}} \psi\left(\omega, f\left(x_i\right)\right)$$

$$MAR = \frac{1}{n} \sum_{k=1}^{n} AR_k$$

$$SD - AR = \sqrt{\frac{1}{n} \sum_{k=1}^{n} \left(AR_k - MAR\right)^2}$$

MAR represents the classification performance of the algorithm, and SD-AR represents the robustness of the algorithm. MAR ± SD-AR is selected as the basis for judging the performance of the algorithm.

Tables 1 and 2 show the experimental results of the data set with KNN and SVM as the base classifier, respectively. The bold data indicates that the algorithm performs better in classification. As shown in Tables 1 and 2, when the initial labeled sample is 10%, the average classification accuracy of ST-DP-CEW on multiple data sets is significantly better than other comparison algorithms. However, when the algorithm is based on the SVM classifier, the classification accuracy of ST-DP-CEW on the dataset Cleveland has basically not improved. This is mainly because the values of most attributes in the dataset are close to 0. For the same attribute, The differences between the samples are small, resulting in a small difference between the samples as a whole, and the discrimination of each category is reduced, which affects the final classification effect.

Table 1 Experimental results when the base classifier is KNN (MAR ± SD-AR, %)

data	Classifier: KNN				
set	KNN	ST-F	ST-D	ST-D	ST-DP
Bupa	54.48	56.91	58.88	59.13	**62.27**
Clevel	46.79	46.47	48.16	49.15	**52.17**
Derma	53.60	56.18	70.94	73.98	**78.19**
Glass	50.54	5L58	55.26	57.40	**61.65**
Haber	67.59	67.92	69.31	68.91	**72.19**
Ionosp	74.35	72.35	80.61	81.20	**83.45**
pi ma	67.72	64.98	66.40	66.93	**70.05**
yeast	45.96	48.32	49.19	50.74	**53.10**

Table 2 Experimental results when the base classifier is SVM (MAR ± SD-AR, %)

data	Classifier: SVM				
set	KNN	ST-F	ST-D	ST-D	ST-DP
Bupa	60.86	62.57	65.50	65.80	**67.01**
Clevel	**53.84**	**53.84**	53.82	53.82	**53.84**
Derma	56.41	57.28	68.14	72.36	**78.25**
Glass	44.81	46.34	49.46	51.36	**54.72**
Haber	70.59	71.61	71.85	72.24	**74.62**
Ionosp	78.33	79.75	80.92	82.34	**84.92**
pi ma	71.75	72.53	75.12	75.78	**77.23**
yeast	31.54	30.76	31.21	32.43	**35.81**

III. Conclusion

In this paper, based on the ST-DP algorithm, a self-training algorithm based on density peaks and edge trimming weights is proposed based on the samples that may be mislabeled during the self-training iteration process. That is, the method of statistically identifying cut-off weights to identify incorrectly labeled samples is integrated into the ST-DP algorithm. It not only considers the spatial structure of the data set, but also solves the problem that the samples are incorrectly labeled. In addition, the calculation of the weights in the adjacency graph also makes better use of the spatial structure of the data set and the information carried by the unlabeled samples. The effectiveness of the ST-DP-CEW algorithm is fully analyzed on the real data set. Especially when the proportion of initially labeled samples is low, the proposed algorithm has greatly improved performance compared to existing algorithms. In the subsequent work, we will discuss how to better construct the adjacency graph, and introduce a function that measures the probability of label error in the recognition process to make label recognition more accurate.

References

Dong, A., Chung, F., & Shitong, W. (2016). Semi-supervised classification method through oversampling and common hidden space. *Information Sciences*, *349*. https://doi.org/10.1016/j.ins.2016.02.042

Gan, H., Sang, N., Huang, R., Tong, X., & Dan, Z. (2013). Using clustering analysis to improve semi-supervised classification. *Neurocomputing*, *101*, 290–298. https://doi.org/https://doi.org/10.1016/j.neucom.2012.08.020

Liu, J., Gong, M., & He, H. (2019). Deep associative neural network for associative memory based on unsupervised representation learning. *Neural Networks : The Official Journal of the International Neural Network Society*, *113*, 41–53. https://doi.org/10.1016/j.neunet.2019.01.004

Pavlinek, M., & Podgorelec, V. (2017). Text classification method based on self-training and LDA topic models. *Expert Systems with Applications*, *80*, 83–93. https://doi.org/https://doi.org/10.1016/j.eswa.2017.03.020

Rodriguez, A., & Laio, A. (2014). Clustering by fast search and find of density peaks. *Science*, *344*(6191), 1492–1496. https://doi.org/10.1126/science.1242072

Tanha, J., van Someren, M., & Afsarmanesh, H. (2017). Semi-supervised self-training for decision tree classifiers. *International Journal of Machine Learning and Cybernetics*, *8*(1), 355–370. https://doi.org/10.1007/s13042-015-0328-7

Triguero, I., Sáez, J. A., Luengo, J., García, S., & Herrera, F. (2014). On the characterization of noise filters for self-training semi-supervised in nearest neighbor classification. *Neurocomputing*, *132*, 30–41. https://doi.org/https://doi.org/10.1016/j.neucom.2013.05.055

Vijayan, A., Kareem, S., & Kizhakkethottam, J. J. (2016). Face Recognition Across Gender Transformation Using SVM Classifier. *Procedia Technology*, *24*, 1366–1373. https://doi.org/https://doi.org/10.1016/j.protcy.2016.05.150

Wang, X.-F., & Xu, Y. (2017). Fast clustering using adaptive density peak detection. *Statistical Methods in Medical Research*, *26*(6), 2800–2811. https://doi.org/10.1177/0962280215609948

Wu, D, Luo, X., Wang, G., Shang, M., Yuan, Y., & Yan, H. (2018). A Highly Accurate Framework for Self-Labeled Semisupervised Classification in Industrial Applications. *IEEE Transactions on Industrial Informatics*, *14*(3), 909–920. https://doi.org/10.1109/TII.2017.2737827

Wu, Di, Shang, M., Luo, X., Xu, J., Yan, H., Deng, W., & Wang, G. (2018). Self-training semi-supervised classification based on density peaks of data. *Neurocomputing*, *275*, 180–191. https://doi.org/https://doi.org/10.1016/j.neucom.2017.05.072

Wu, Di, Shang, M. S., Wang, G., & Li, L. (2018). *A self-training semi-supervised classification algorithm based on density peaks of data and differential evolution*. 1–6. https://doi.org/10.1109/ICNSC.2018.8361359

An investigation on the influence of interactive aesthetics in virtual industrial design

Gao Zhu
School of Art and Design
Shanghai University of Engineering Science
Shanghai, China
Gao1965@163.com

Li Qi
School of Art and Design
Shanghai University of Engineering Science
Shanghai,China
qili@sues.edu.cn

Abstract—**Virtual reality (VR) is an emerging technology that creates a three-dimensional virtual environment. It emphasizes user's experience on immersion, imagination and interaction. Some research demonstrates that aesthetic used in VR design could enhance interactivity between virtual body and environment. Aesthetic design on industrial design emphasizes the product appearance rather than function or usability. However, there is little research in aesthetic contribution to industrial design in virtual reality. From the perspective of aesthetics, the paper outlines the aesthetic studies to a virtual reality and product design. Based on the user experience of perception, it analyses a model of aesthetic perception of industrial design in virtual environment, which provides a general guidance for virtual industrial product design.**

Keywords- aesthetics, virtual reality, interactivity, industrial design

I. Introduction

In the past decades, virtual reality technology has been used in the field of industrial product design. Virtual reality (VR) uses head-mounted displays (HMDs) to create a potentially different way of displaying virtual environment. As a popular visualizing design tool, VR technology has been used to explore design prototyping, visualization and communication in industrial product design. It has being employed to decision making in design, evaluation, and training processes across multiple disciplines [1]. Both industrial and academic communities have contributed to a large knowledge base on numerous virtual reality topics, which involve technical advance and aesthetics. Technical advances have enabled designers to explore and interact with data in a natural way; while aesthetic contribution to product appearance has been demonstrated to impact on product performance or price attributes [2]. Yamamoto and Lambert [2] argue that visual aesthetic can be reflected in many consumer experiences, such as fashion and arts, however, it also has significant in industrial products such as automobiles, home appliances and transportations. The aesthetics in industrial design emphasizes a sense of beauty and users' psychological and emotional needs for products. Aesthetic principles have been used in virtual industrial design, which should follow the beauty of product.

Aesthetic approach to industrial design involves two aspects: functional beauty and the beauty of form. In the conventional design, function plays an important role, which allows users to use the product effectively and efficiency. It meets the actual needs of users in daily life. Functional beauty is an important factor in industrial design, as its well-designed function is considered as a source of positive (or negative) aesthetic value, in which aesthetic theory includes art, the everyday, animals and organic nature and environments and artifacts [3]. The functional beauty has a long tradition within aesthetics, from classical ancient philosophy to contemporary aesthetic theory [4]. The outcome is both impressive and insightful in the way it brings together considerations from aesthetics and philosophy of science to a comprehensive and unified account of aesthetic experience. The beauty of form is subjective perception, which can design industrial products with combination of various methods. It provides users a sense of beauty and enhances user's satisfaction with the products, which includes consistency, balance, scales, contrast and rhythm.

Aesthetic of virtual reality emphasizes an immersion into a world of numbers that appears in forms, images, and sensations, which involve a virtual body and interactions with avatar [5]. Different from other types of digital technologies, VR has a special kind of interactivity, in which virtual body is an entity that is exceptionalized through interaction. Carroll (2008) investigates the photorealistic virtual reality to create "engaging" experience rather than traditional human-computer interaction approaches that "primarily focused on the performance and efficiency issue of the technology" [7]. Carroll's study highlights a concept of "aesthetic interaction", which examines the role of narrative in VR and the visual-narrative employed the use of aesthetics to "engage" the spectator in its process of storytelling.

However, there is little research related to aesthetic experience of industrial product design in virtual reality. This paper will firstly address the virtual reality used in industrial design, the concept of functional beauty and then aesthetic perception for product design. It aims to investigate a potential on aesthetic model influencing virtual industrial design.

II. Virtural reality and industrial design

Virtual reality (VR) is a new technology emerging at the end of the 20th century, which integrates digital image processing, computer graphics, multimedia technology, sensor technology and other information technologies. Research suggests that the advantage of VR is able to create the users

DOI reference number: 10.18293/DMSVIVA2020-021.

experience of immersion, interaction and imagination. Where human-computer interaction in virtual reality is considered as a natural interaction, users can interact with the computer through head-mounted helmets and gloves. The virtual environment created by the computer can make the users' experience a sense of presence and becomes a part of the virtual environment. In the last two decades, VR technology has been extendedly used as a tool in the field of industrial design and industrial training and education. For example, Wald et. al [8] proposed a ray tracing based virtual reality framework that supports direct ray tracing of trimmed freeform surfaces. Many car manufactures adopt VR technology to design and test vehicle system in real-time performance parameters.

Research on VR applied to industrial design focuses on the following aspects: Firstly, virtual reality in industrial design emphasizes an aesthetic convergence, which involves design methodology based on a multidisciplinary approach using VR as a tool [9].

Secondly, one of the primary goals of VR technology applied in industrial design highlights an alternative aesthetic and interactive sensibility for real-time, interactive, 3D computer graphics [10]. It is important to design an immersive virtual space to allow users to shed their ways to observe the world.

Thirdly, virtual industrial design is a new design method of product development based on computer using virtual reality technology. The effective application of industrial design can establish a virtual realistic world and users can conduct relevant research and better understand the relationship of products. For the industrial products, VR can offer the ideal unconstrained interface for free artistic practice and bridge the gap between creative experimentation and precise manufacturing-oriented modelling [11].

Lastly, VR can be used as a general design tool focusing on the decision making process in product development projects. Berg and Vance's research demonstrates that immersive VR technology can effectively assistant designers and engineers to identify design issues and potential solutions, which cannot be sorted out using traditional computer tools [12]. Their study shows that participants are being encouraged an increased sense of team engagement through virtual environment.

III. AESTHETIC AND VIRTUAL REALITY

Aesthetic examines subjective and sensori-emotional values, or sometimes called judgments of sentiment and taste [13]. It is the study of beauty and taste. In virtual environment, aesthetics is considered as user perception for immersion, interaction and imagination. In this sense, aesthetic approach to virtual reality has value to positive side of user experience. This aesthetic experience is defined as the "lively integration of means and ends, meaning and movement, involving all our sensory and intellectual faculties is emotionally satisfying and fulfilling. Each act relates meaningfully to the total action and is felt by the experiencer to have a unity or a wholeness that is fulfilling" [14]. This statement highlights felt life is critical important in aesthetic approach to product design. This is defined as pragmatist aesthetics by Petersen et al. However, virtual industrial design emphasizes the interaction in user perception. Petersen et al. argue that there are two aesthetic perspectives. The first focuses on creating involvement, user experience, surprise and serendipity in the interactive; the second focuses on promoting bodily experiences and symbolic representations interacting with the system [15].

These are a growing interest in the aesthetic of interactivity in industrial product design. It suggests a response of need for alternative frames of reference in interactive systems design and alternative ways of understanding the relationships [16]. Wright, Wallace and McCarthy (2008) argues that aesthetic of interactive design in VR involves three themes: (1) a holistic approach in which a user with emotions and thoughts focus on design; (2) a constructivist stance wherein user's self is seen as continuously engaged and constituted in making sense of experience and (3) a dialogical ontology that self, others, and technology are constructed as multiple centers of value. Based on three themes, researchers critic into the aesthetic of interaction and suggest a concept of "sensibilities" for designing aesthetic interaction.

Virtual reality has aesthetic characteristics of technology. Technology provides usability for the realization of virtual industrial design, through the computer system, can create a set of dynamic, audio and video functions in one of the three-dimensional space environment, and give a sense of immersive, fully mobilize users to participate in the interaction. The technical aesthetic embodied in the industrial design is mainly manifested in the aspects of modeling, vision and hearing, which not only makes effective use of science and technology, but also enhances its artistry and good users' experience. The technical aesthetic of industrial design can effectively enhance the aesthetic value of industrial product design and play a positive role in increasing economic benefits. In industrial design, the full integration of aesthetic theories, such as technical aesthetic is conducive to the optimization of labor productivity and the promotion of the economic development and scientific and technological advancement.

Although the virtual reality design takes the technology as the important support and shows the strong technical aesthetic characteristics, from the perspective of the form of expression, the design has the artistic characteristics of modeling. According to art theory, design is a special form of artistic activities. In industrial design, relevant personnel must carry out visual design on layout, structure, material, texture and other aspects, giving users a strong visual impact. Thus, industrial design contains aesthetic characteristics of artistic beauty. In the industrial design, we strive to set up an immersive virtual reality world for users, with three-dimensional visual effect, artistic expression, appeal and artistic beauty.

IV. AESTHETIC AND VIRTUAL INDUSTRIAL DESIGN

In Figure 1, a conceptual framework displays the relationship between the virtual reality, industrial design, aesthetic experience, and virtual aesthetic approach to industrial design. In this model, it is clearly shown that virtual aesthetics approach to industrial design draws from virtual reality technology, industrial product design and aesthetics. The characteristics of three fields are combined together to construct the essential properties in aesthetic approach to virtual industrial design. The immersion, interaction and imagination are fundamental to virtual reality; while usability, effectiveness and efficiency are basic to industrial product design. These are also linked to judgment of aesthetic and user experience. Thus, from this conceptual model, aesthetic approach to virtual industrial design not only presents aesthetic values in virtual environment, user interaction and imagination but also usability and effectiveness.

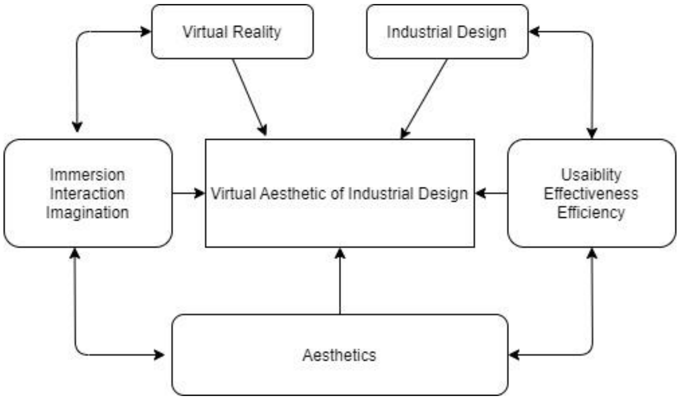


Figure 1. The model of virtual aesthetic of industrial design

This model can be used to inform the design principles for industrial design in virtual reality. It could be a valuable design framework for design of industrial products with characteristics of immersion, interactivity and imagination.

V. CONCLUSION

This paper presents an overview of interactive aesthetic that explores user's perception and experience in the virtual environment for the field of industrial design. Traditional industrial design emphasizes function and usability. Using virtual reality to industrial design emphasizes the immersion, interactivity and imagination. Aesthetic plays an important role in the design of virtual industrial products for its positive values. Functional beauty is a kind of aesthetic experience for the appearance of products. The beauty of form emphasizes the subjective perception, which evokes a sense of beauty and enhances user's satisfaction with the products. At the last, it concludes a conceptual model of virtual aesthetic in industrial design, which integrates the characteristics of VR, industrial design and aesthetics.

ACKNOWLEDGMENT

The project was supported by Shanghai Pujiang Program.

REFERENCES

[1] Berg, L. P., & Vance, J. M. (2017). Industry use of virtual reality in product design and manufacturing: a survey. Virtual reality, 21(1), 1-17.

[2] Yamamoto, M., & Lambert, D. R. (1994). The impact of product aesthetics on the evaluation of industrial products. Journal of Product Innovation Management, 11(4), 309-324.

[3] Davies, S. (2010). Functional beauty examined. Canadian Journal of Philosophy, 40(2), 315-332.

[4] Parsons, G., & Carlson, A. (2008). Functional beauty. Oxford University Press.

[5] Diodato, R. (2012). Aesthetics of the Virtual. SUNY Press.

[6] Carroll, F. (2008). Engaging photorealistic VR: An aesthetic process of interaction (Doctoral dissertation, Napier University, Edinburgh, Scotland).

[7] Wald, I., Dietrich, A., Benthin, C., Efremov, A., Dahmen, T., Gunther, J., ... & Slusallek, P. (2006). Applying ray tracing for virtual reality and industrial design. In 2006 IEEE Symposium on Interactive Ray Tracing (pp. 177-185). IEEE.

[8] Guerlesquin, G., Mahdjoub, M., Bazzaro, F., & Sagot, J. C. (2012). Virtual reality as a multidisciplinary convergence tool in the product design process. Journal of Systemics, Cybernetics & Informatics, 10(1), p51.

[9] Davies, C., & Harrison, J. (1996). Osmose: towards broadening the aesthetics of virtual reality. ACM SIGGRAPH Computer Graphics, 30(4), 25-28.

[10] Fiorentino, M., De Amicis, R., Stork, A., & Monno, G. (2002). Surface design in virtual reality as industrial application. In DS 30: Proceedings of DESIGN 2002, the 7th International Design Conference, Dubrovnik (pp. 477-482).

[11] Carroll, F. (2008). Engaging photorealistic VR: An aesthetic process of interaction (Doctoral dissertation, Napier University, Edinburgh, Scotland).

[12] Zangwill, N. (2003) "Aesthetic Judgment", Stanford Encyclopedia of Philosophy. Retrieved on May 02,2020.

[13] Wright, P., McCarthy, J., & Meekison, L. (2003). Making sense of experience. In Funology (pp. 43-53). Springer, Dordrecht.

[14] Petersen, M. G., Iversen, O. S., Krogh, P. G., & Ludvigsen, M. (2004, August). Aesthetic interaction: a pragmatist's aesthetics of interactive systems. In Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques (pp. 269-276).

[15] Wright, P., Wallace, J., & McCarthy, J. (2008). Aesthetics and experience-centered design. ACM Transactions on Computer-Human Interaction (TOCHI), 15(4), 1-21.