An Interaction Mining Approach for Classifying User Intent on the Web

Loredana Caruccio, Vincenzo Deufemia, Giuseppe Polese
Department of Management and Information Technology
University of Salerno
84084 Fisciano (SA), ITALY
{lcaruccio, deufemia, gpolese}@unisa.it

Abstract

Predicting the goals of internet users can be extremely useful in e-commerce, online entertainment, and many other internet-based applications. One of the crucial steps to achieve this is to classify internet queries based on available features, such as contextual information, keywords and their semantic relationships. Beyond these methods, in this paper we propose to mine user interaction activities in order to predict the intent of the user during a navigation session. However, since in practice it is necessary to use a suitable mix of all such methods, it is important to exploit all the mentioned features in order to properly classify users based on their common intents. To this end, we have performed several experiments aiming to empirically derive a suitable classifier based on the mentioned features.

1 Introduction

During an Internet navigation session the user performs several actions that can provide hints on his/her future activities. Being able to capture and interpret the hidden goals behind such actions can provide organizations with a competitive advantage. For instance, e-commerce organizations might predict user needs, and advertise the products that users will most likely buy. Thus, multimedia catalogues, web and information retrieval systems need to embed search engines capable of capturing user intent, which is the focus of user intention understanding (UIU) research area [26].

Many approaches for user intent understanding are based on the analysis of search behaviors [4, 6, 7, 8, 10, 17], such as clicked URLs [31] and submitted queries. Most of them aim to capture semantic correlations among search behaviors of the same user, in order to let search engines produce customized results for each individual user.

Other studies analyzed user interactions with *Search Engine Result Pages* (SERPs) to infer their intent [2, 3, 14, 18, 22, 30]. However, by limiting the analysis to results contained in a SERP, such methods ignore many important interactions and contents visited from such results. For this reason, some approaches to user behavior analysis focus on user interactions with web pages to infer clues on their interest and satisfaction with respect to the visited contents [1, 16]. Following this trend, in this paper we define a new model for UIU analyzing both interactions with SERP results and those on the visited web pages. The interaction features considered in the proposed model are local page level statistics, that is, they are fine-grained and refer to portions rather than the whole web pages. This provides the basis for a more promising prediction of the user intent, since several experiments with eye-trackers revealed that users analyze web page contents by sections, overlooking those of low interest [27].

Other than interaction features, the proposed model considers additional features, such as query keywords and contextual information, all feeding a classification algorithm to understand user intent. The classification process uses a two-level taxonomy in which the first level defines *navigational*, *informational*, and *transational* types of queries[5], where the last two are further decomposed in the second level [28].

We also provide experimental results highlighting the efficiency of the proposed model for query classification. The proposed set of features has been evaluated with several classification algorithms. To this end, in order to more precisely compare the achieved results, and detect the most promising features, we have introduced a metric to evaluate the performances of the different classifiers.

The rest of this paper is organized as follows. In Section 2, we provide a review of related work. Then, we present the model exploiting interaction features for UIU in Section 3. Section 4 describes experimental results. Finally, conclusions and future work are given in Section 5.

DOI reference number: 10.18293/DMS2015-045

2 Related Work

As said above, many approaches for user intent understanding analyze search behaviors of users while they navigate and submit queries through the web [4, 6, 7, 8, 10, 17].

In the early 90s, a pioneer study on search behaviors focused highlighted three browsing strategies [9]: *scan browsing*, in which new information is scanned based on its relevance to changing tasks, representing transient browse goals; *review browsing*, in which, with respect to scan browsing, the scanned information is also reviewed and integrated; finally, *search-oriented browsing*, in which the new information is scanned, reviewed, and integrated based on its relevance to a fixed task.

Morrison *et al.* proposed three taxonomic classification schemes based on user responses to web activities that significantly impacted on their decisions and actions [24]. In particular, they formalized the main questions users ask themselves before starting a search session: *why*, *how*, and *what*, which represent the primary purpose of the search, the method used to find the information, and the content of the searched-for information, respectively, yielding three different taxonomies.

Sellen *et al.* extended previously defined taxonomies by extensively monitoring user search activities [29]. They ended up with a classification dividing web activities into six categories, in which two new types were introduced: *transacting* and *housekeeping*. The first concerns using the web to execute secure transactions targeted at products and services, such as ordering a product or filling out a questionnaire. The second concerns using the web to check or maintain the accuracy and functionality of web resources.

A taxonomy focusing on search queries has been defined by Broder [5], who identified the following three classes of queries based on user's intent: *navigational*, aiming to reach a particular web site, *informational*, aiming to collect information from one or more web pages, and *transactional*, aiming to perform some web-mediated activities, that is, to reach a web site where some service is offered, and from which further interactions are expected.

Kang *et al.* focused on analyzing two types of search activities [18]: *topic relevance*, that is, searching documents guided by a given topic, of informational type, and *homepage finding*, aiming to search main pages of several types of navigational web sites. Starting from common information used by Information Retrieval (IR) systems, such as web page content, hyperlinks, and URLs, the model proposes methods to classify queries based on the two categories mentioned above.

Agichtein *et al.* proposed a predictive model derived from real case studies, which is based on the analysis and the comprehension of user interactions during web navigation [2]. The model tries to elicit and understand user navigation behaviors by analyzing several activities, such as clicks, scrolls, and dwell times, aiming to predict user intention during web page navigation. Moreover, the study proposes to analyze features that are used to characterize the complex interactions following a click executed on a result page. Such interactions have been exploited also by Guo *et al.*, since they considered them useful to accurately infer two particular tightly correlated intents: search and purchase of products [14].

Lee *et al.* proposed a feature based model for the automatic identification of search goals, focusing on navigational and informational queries [22]. The model has been developed starting from experimental studies on real user navigation strategies, which have primarily revealed the possibility of effectively associating most queries to one of two categories defined within the taxonomy. They observed that queries not effectively associable to a category are usually related to few topics, such as proper nouns or names of software systems. More specifically, the model proposes two features: *past user-click behavior* to infer users intent from their past interactions with results, and *anchor-link distribution*, which uses possible targets of links sharing the same text with the query.

While the strategies described so far aim to classify search queries exclusively using features modeled to characterize search queries, Tamine *et al.* propose to analyze search activities previously performed in the same context [30]. To this end, the set of past queries represents the *query profile*, which helps deriving data useful for inferring the type of the current query.

3 A Model for User Intent Understanding

In this section we describe the model and the features used for the classification process. The model of this work is based on the model proposed in [12].

3.1 A two-level taxonomy for web queries

During a web search the user has a specific goal, generally described by a textual query, and classifiable in a taxonomy. In what follows, we introduce the two-level taxonomy that will be used in the proposed approach for classifying user queries, which is shown in Figure 1. It synthesizes concepts defined in the taxonomies proposed in [5, 28], which have been refined based on the analysis of the query set used in our experiments.

A brief description of the categories on both levels of the taxonomy follows:

- *Informational*: The aim of this kind of query is to learn something by reading or viewing web pages;

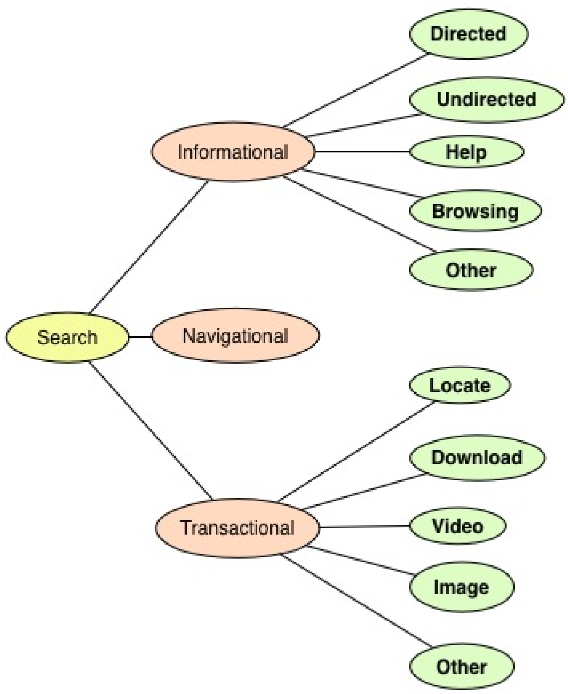

Figure 1: Two-level taxonomy.

- **Directed:** when searching something about a topic;
- **Undirected:** when the user wants to learn anything/everything about a topic;
- **Help:** when the user searches for advices, ideas, suggestions, or instructions;
- **Browsing:** when the user searches something like news, forums, or manuals;
- **Other:** when the informational query does not fall in any of the categories above.

- *Navigational*: The aim of this kind of query is to reach a known website. The only reason of this kind of search is that it is more convenient than typing the URL, or perhaps if its URL is not precisely known.

- *Transactional*: The aim of this kind of query is to retrieve a resource available on some web page.
  - **Download:** when the user aims to download a resource;
  - **Video:** when the user aims to watch a video;
  - **Image:** when the user aims to get an image;
  - **Locate:** when the user aims to verify whether or where some real world service or product is offered;
  - **Other:** when the query is transactional, but it does not fall in any of the categories above.

3.2 Search model: session, search, interaction

Several studies have proven the usefulness of user interactions to assess the relevance of web pages [1, 15, 16, 19], and to determine the intent of search sessions [14, 13]. However, there are additional interactions originating from SERP's contents, such as browsing, reading, and multimedia content fruition, which can potentially provide additional useful clues to UIU.

The proposed approach extends existing predictive models, by mining interactions between users and web pages during a search session. We believe that the actions performed on the visited pages, contrasted to the page format, provide a valuable source of knowledge to predict user intent. As an example, scrolling a web page containing flat text might imply a given user intent, which is different from the scrolling actions performed on framed web pages including both textual and multimedia contents.

In general, a web *session* can be seen as a sequence of search activities aimed at achieving a given goal. When the submitted query does not provide the desired results, the user tries to gradually approach the target, by refining or changing search terms and keywords. A *search* activity can be seen as the combination of the following user actions: submission of a query to a search engine, analysis of search results, and navigation through one or more hyperlinks inside them. The last two types of activities are accomplished by means of several types of *interactions*, which include mouse clicks, page scrolling, pointer movements, and text selection. If combined with features such as dwell time, reading rate, and scrolling rate, such interactions allow us to derive an implicit feedback of user experience with the web pages [12].

The proposed approach prescribes a fine-grained analysis of the traced interactions between users and web pages. Indeed, user interaction analysis is restricted to portions of web pages, e.g., blocks of text, images, multimedia content, which can have a variable length. The use of *subpage*-level analysis provides additional information in the assessment of the user interactions with respect to a global analysis of the entire page.

The data concerning user interactions during web navigation have been encoded into features, which are used by predictive models to characterize user behaviours. We organize the set of features into the following categories: *query*, *search*, *interaction*, and *context*.

Query. These features are derived from characteristics of a search query such as keywords, the number of keywords, the semantic relations between them, and other characteristics of a search or an interaction.

Search. These features act on the data from search activities such as: results, time spent on SERP, and number of results considered by the user. The *DwellTime* is measured from the start of the search session until the end of the last interaction originated by the same search session. The reaction time, *TimeToFirstInteraction,* is the time elapsed from the start of the search session and the complete loading of the first selected page. Other features dedicated to interactions with the results are *ClicksCount*, which is the number of visited results, and *FirstResultClickedRank*, determining the position of the first clicked result.

Interaction. These features act on the data collected from interactions with web pages and subpages, taking into account the absolute dwell time, the effective dwell time, all the scrolling activities, search and reading activities. The *DwellRate* measures the effectiveness of the permanence of a user on a web page, while the reading rate *ReadingRate*, measures the amount of reading of a web page [12]. Additional interactional features are: *ViewedWords*, the number of words considered during the browsing, *UrlContainsTransactionalTerms*, which verifies if the URL of the page contains transactional terms (download, software, video, watch, pics, images, audio, etc.), *AjaxRequestsCount*, which represents the number of AJAX requests originated during browsing.

Context. These features act on the relationship between the search activities performed in a session, such as the position of a query in the sequence of search requests for a session.

3.3 Logging Web Interaction Data

In the following we describe the module YAR we implemented for logging the user interaction actions, from which we derive the set of features contributing to the mining of user intent.

The architecture of the YAR system is depicted in Figure 2. It is based on a client/server model, where data concerning user interactions are collected on the client side by the *Logger*, and evaluated on the server side through the *Log Analyzer*. The *Logger* is responsible for "being aware" of the user's behavior while s/he browses web pages, and for sending information related to the captured events to the server-side module. The latter is responsible for analyzing the collected data and for applying metrics to derive the candidate taxonomy categories.

The *Logger* is based on the *AJAX* technology [25] to capture and log user's interactions with a web system through a pluggable mechanism, which can be installed on any web browser. Thus, it does not require modifications to the web sites, or any other legacy browser extensions.

4 Experiments

In this section we describe the dataset constructed for evaluating the proposed approach and the results achieved with different classification algorithms. In the following, we first provide an overview on the used evaluation metrics and the considered subsets of features, then experimental results are presented.

4.1 Experiment Setup

In order to build the dataset for evaluating the proposed model we recruited 31 participants, whose profiles are described in Table 1. For each participant the table shows the gender (18 males vs. 13 females), the age (ranging from 20 to 65 ages), and their experience in using the Web (ranging from 1 to 23 years). Since age, education, and Web experience might significantly influence the approach to Web search, we have tried to involve a balanced mix of profiles, in order to gain unbiased conclusions. Thus, we involved people with heterogeneous ages and web experience; similar considerations apply for education, even though the majority of them have a computer science or technical background (18 out of 31).

Gender	Age	Education	Web exp. (yrs)
M	36	Tech. High School Diploma	13
M	65	Tech. Professional Qualification	1
F	32	MSc in Graphics	11
F	59	Accountant Qualification	10
M	31	MSc in Computer Science	20
M	24	BSc in Computer Science	11
M	23	Undergrad. student in Biology	15
F	23	BSc in Biology	16
F	25	BSc in Computer Science	12
M	60	Tech. High School Diploma	23
M	25	BSc in Computer Science	13
M	27	BSc in Computer Science	10
M	24	Undergrad. student in Political Science	10
M	25	Undergrad. student in Computer Science	14
M	25	Undergrad. student in Computer Science	15
M	25	BSc in Computer Science	10
M	25	BSc in Computer Science	10
M	25	Grad. student in Computer Science	7
F	20	Undergrad. student in Linguistics	5
F	22	Undergrad. student in Civil Eng.	10
M	29	Grad. student in Computer Science	14
M	24	Grad. student in Computer Science	15
F	25	BSc in Computer Science	15
F	27	BSc in Education	11
M	25	BSc in Computer Science	9
F	33	MD specializing in Pediatrics	10
M	34	Grad. student in Microelectronics Eng.	20
F	25	Grad. student in Linguistics	8
F	25	Undergrad. student in Sociology	10
F	24	High School Diploma in Arts	8
F	27	BSc in Education	11

Table 1: Profiles of participants to the evaluation.

All participants were requested to perform ten search sessions organized as follows:

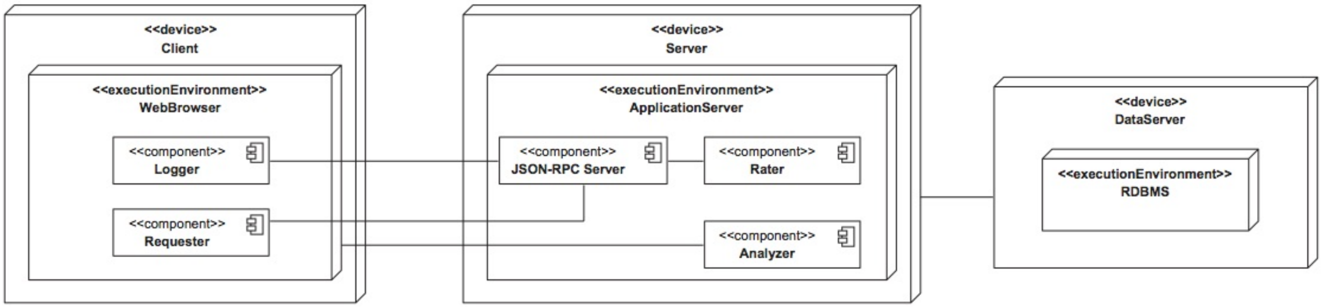

Figure 2: The YAR System Architecture.

- four guided search sessions;
- three search sessions in which the participants know the possible destination web sites;
- three free search sessions in which the participants do not know the destination web sites.

In the following, we list the goals of the guided search sessions:

- the London Metro map image;
- the official video of U2 song Vertigo;
- the e-mail address of an administrative office at the University of Salerno;
- the size of *Mona Lisa*, the famous painting of Leonardo.

This led to 129 sessions and 353 web searches, which were recorded and successively analyzed in order to manually classify the intent of the user according to the two-level taxonomy in Figure 1. Starting from web searches, 490 web pages and 2136 sub pages were visited. The interaction features were logged by the YAR plug-in for Google Chrome/Chromium [12].

4.1.1 Feature subsets

In order to analyze the effectiveness of the considered features, we have grouped them into several subsets:

- ***All***: subset of all the proposed features: *query*, *search*, *interaction*, and *context*;
- ***Query***: subset of all the features related to *queries*;
- ***Search***: subset of all the features related to *search* and *context*;
- ***Interaction***: subset of all the features related to *interactions*;
- ***Query+Search***: subset of the features derived as union from *Query* and *Search*. The goal is to evaluate the effectiveness of query classification by using the features considered in other studies [16, 2, 14];
- ***Transactional***: subset of all the features related to interactions over transactional queries *ViewWords*, *AjaxRequestsCount*, *ScrollingDistance*, *ScrollingCount*, and *UrlContainsTransactionalTerms*. The goal here is to evaluate the classification of transactional queries by adopting more specific features;
- ***Interaction−Transactional***: subset derived by the exclusion of the transactional features from the set *Interaction*. The goal here is to evaluate the effectiveness of the classification of transactional queries by comparing results achieved with interaction features to those achieved by excluding transactional features.
- ***All−Transactional***: subset derived by the exclusion of the transactional features from the set *All*. The goal here is to evaluate the effectiveness of the classification of transactional queries by comparing results achieved with all features to those achieved by excluding transactional features.

The set of features captured during the search sessions are available for download[1].

4.1.2 Classifiers

We considered three classifiers to evaluate the proposed feature model: SVM [11], CRF [20], and LDCRF [23].

In the context of query classification, SVM assumes that the queries in a user session are independent, Conditional Random Field (CRF) considers the sequential information between queries, whereas Latent Dynamic Conditional Random Fields (LDCRF) models the sub-structure of user sessions by assigning a disjoint set of hidden state

[1] `https://goo.gl/ypH2ij`

variables to each class label. They have been configured as follows:

1. **SVM**. We used MSVMpack [21] as the SVM toolbox for model training and testing. The SVM model is trained using a linear kernel and the parameter C has been determined by cross-validation.

2. **CRF**. We used the HCRF library[2] as the tool to train and test the CRF model. For the experiments we used a single chain structured model and the regularization term for the CRF model was validated with values 10^k with $k = -1 \dots 3$.

3. **LDCRF**. We used the HCRF library for training and testing LDCRF model. In particular, the model was trained with 3 hidden states per label, and the regularization term was determined by cross-validation to achieve best performances.

4.1.3 Evaluation Metrics

In order to evaluate the effectiveness of the proposed model, we adopted the classical evaluation metrics of Information Retrieval: *accuracy*, *precision*, *recall*, and *F1-measure*, whose definition is given below:

$$\textbf{Accuracy} = \frac{\#TP + \#TN}{\#TP + \#TN + \#FP + \#FN}$$

$$\textbf{Precision} = \sum_{Category(i)} \Big(\frac{\#correctly\ classified\ queries}{\#classified\ queries} \times \frac{\#category\ queries}{\#total\ queries} \Big)$$

$$\textbf{Recall} = \frac{\#correctly\ classified\ queries}{\#total\ queries}$$

$$\textbf{F1} - \textbf{measure} = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

In addition, in order to simplify the comparison of performances for the different classifiers, in what follows, we apply more a suitable metrics. In fact, in order to evaluate the effectiveness of a classifier, the features need be grouped into several subsets, and executing each classifier by considering each subset of features once. Then, to contrast performances of classifiers we need to compare the results achieved on different pairs (*classifier*, *feature subset*). Thus, in our case, we need to compare 336 values since we have 3 classifiers, each executing on 8 feature subsets, for each of which we need to calculate 14 parameters.

[2] `http://sourceforge.net/projects/hcrf/`

The proposed metrics is based on the mean squared error (*MSE*), which is defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{x_i} - x_i)^2 \tag{1}$$

where $\hat{x_i}$ is the *i-th* predicted value, while x_i is the *i-th* correct value. For our purposes, we used MSE calculated on the *accuracy* measure. Thus, given the vector of *accuracy* values $\hat{a}$, the definition of the *Accuracy Mean Squared Error* (*AMSE* or MSE($\hat{a}$)) is

$$AMSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{a_i} - a_i)^2 \tag{2}$$

where a_i is equal to 1. We computed a relative AMSE value for each pair *Classifier-SubsetFeatures*.

AMSE is able to gain knowledge about the performance of the classifiers and the subsets of features, and how they influence each other.

Let

I = {*All, Query, Search, Transactional, Interaction, Query+Search, All-Transactional, Interaction-Transactional*}

J = {*CRF, LDCRF, MSVM*}

be the set of *SubsetFeatures* and the set of compared *Classifiers*, respectively. We designed four AMSE-based values for gaining knowledge about the classifier performances:

- **Global:** it returns the pair *Classifier-SubsetFeatures* with the minimum AMSE

  $$min\Big(AMSE_{i,j}\Big) \qquad \forall \;\; i \in I, j \in J \tag{3}$$

  It is useful to catch the best performance;

- **Subsets:** it predicts the classifier better performing on each subset of features.

  $$min_{j \in J}\Big(x_{i,j}\Big) \qquad \forall \;\; i \in I \tag{4}$$

  so that we can easily derive the best performing pairs *Classifiers-SubsetFeatures*;

- **FeaturesBehavior:** it computes the average behavior for each subset of features

  $$\frac{1}{|J|} \sum_{j \in J} AMSE_{i,j} \qquad \forall \;\; i \in I \tag{5}$$

  allowing us to gain knowledge about the subsets of features on which a classifier performs better;

- **ClassifiersBehavior:** it computes the average behavior for each *Classifier*

$$\frac{1}{|I|}\sum_{i\in I} AMSE_{i,j} \qquad \forall \quad j \in J \qquad (6)$$

allowing us to detect the best performing classifiers.

4.2 Results

In order to simulate an operating environment, 60% of user queries were used for training the classifiers, whereas the remaining 40% were used for testing them.

Figures 3-6 report the statistics based on the CRF classifier, which give an idea of how complex is the evaluation with conventional measures. On the other hand, Figures 7 and 8 provide a synthetic overview with all the used classifiers, which appears to be more effective. In particular, Fig. 7 highlights that MSVM achieves the best average performance, followed by CRF, which has almost the same MSVM value. Notice that, the lesser the AMSE value the better are the performances, since AMSE is an error measure.

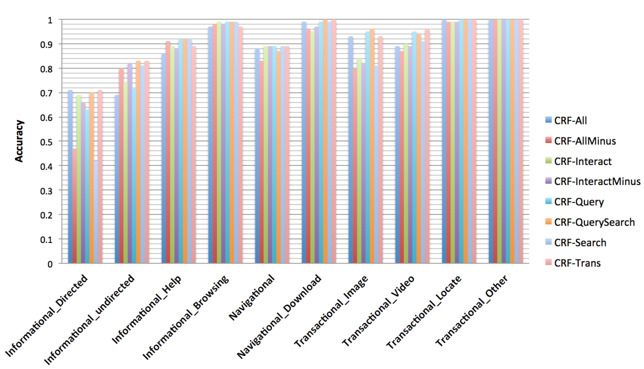

Figure 3: Accuracy obtained with the CRF model.

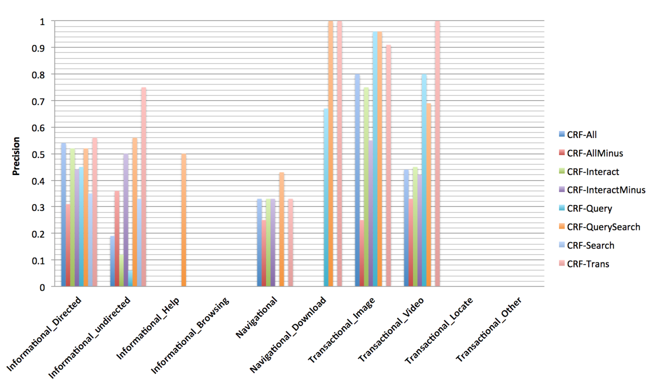

Figure 4: Precision obtained with the CRF model.

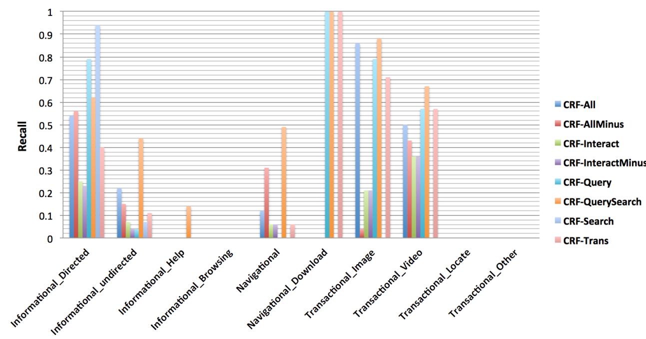

Figure 5: Recall obtained with the CRF model.

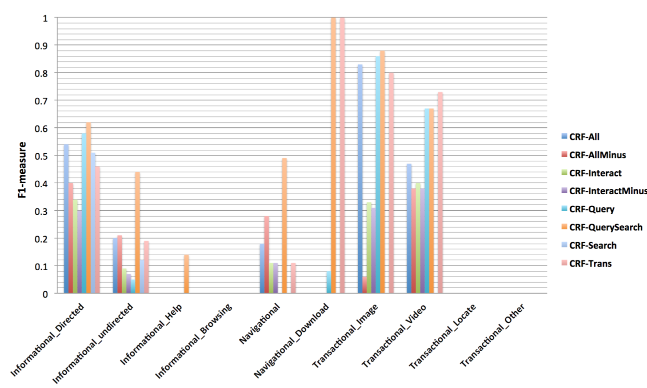

Figure 6: F1-measure obtained with the CRF model.

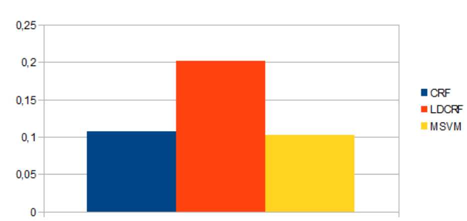

Figure 7: *AMSE ClassifiersBehavior* values of classifiers.

Fig. 8 shows the AMSE values obtained for the different subsets of features. In particular, *Transactional* achieves the best performances, followed by *Query*, and *Query+Search*. Instead, the worst performances are given by the subset *All*, since it yields the maximum value for AMSE.

The *AMSE global* values in Fig. 9 highlight that the best pair (classifier, features) has been *CRF-Transactional*. The *Transactional* features have shown a good discriminative power, since there are two classifiers achieving the best AMSE based on them. Conversely, the *AMSE Subsets* values shown in Fig. 10 highlight that the *CRF* classifier is the one showing best performances for most subsets of features, since it outperforms the other classifiers on 4 out of 8 subsets of features).

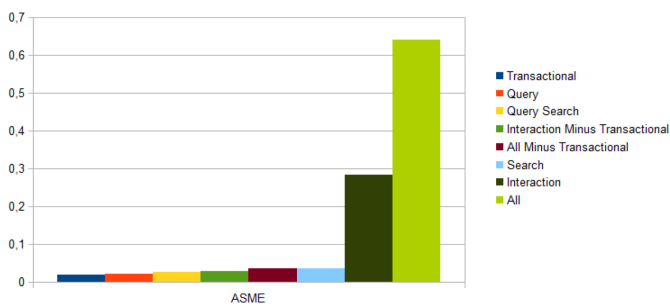

Figure 8: *AMSE Subset* values for each subset of features.

Global		
Classifiers	**SubsetFeatures**	**AMSE**
CRF	Transactional	0,01446
LDCRF	Transactional	0,02086
MSVM	Query	0,02105

Figure 9: *AMSE Global* values for each classifier.

Subsets		
SubsetFeatures	**Classifiers**	**AMSE**
All	MSVM	0,58953
All Minus Transactional	LDCRF	0,02903
Interaction	CRF	0,02262
Interaction Minus Transactional	CRF	0,02204
Query	LDCRF	0,02102
Query Search	CRF	0,01475
Search	MSVM	0,02391
Transactional	CRF	0,01446

Figure 10: Lower *AMSE FeaturesBehavior* values for the considered classifiers.

4.3 Discussion

From the experimental results we can conclude that the use of interaction features to mine the intent of the user during search sessions is a promising approach. In fact, we have observed best classification performances when using the transactional features, which embed a considerable amount of interaction actions. However, we have observed best performances when interaction features are analyzed in a specific context (transactional), rather than in generic contexts. This is due to the fact that is easier to mine the intent when the interaction is performed on a specific type of web page. Vice versa, when no specific assumption can be made on the structure of the web page, each interaction action can convey many different meanings. For instance, a scrolling action yields different interpretations if it is performed on a plain text web page with respect to framed pages like those of online magazines.

5 Conclusions and Future Work

We have proposed a model for UIU focusing on both interactions with SERP results and on the visited web pages. The model predicts user intent by exploiting local page level statistics, and additional features, such as query keywords and contextual information, all feeding a classification algorithm. The latter uses a two-level taxonomy, defining *navigational*, *informational*, and *transational* query types at first level [5], furtherly decomposing the last two types at the second level [28].

We have also empirically compared the performances of main classifiers, and have devised a suitable metrics to detect the best classifier and the best subset of features. In particular, the experiments highlighted that the MSVM classifier achieves the best average performances, followed by the CRF classifier, whereas the *Transactional* features outperformed the others, followed by *Query* and *Query + Search* feature.

In the future, other than investigating the possibility of monitoring additional features, we would like to investigate machine learning approaches for infering a suitable predictive model from a larger set of training data. Moreover, we need to perform a precise classification of web site types, in order to customize the interpretations of user interactions on the specific type of web page. We also need to perform similar investigations in order to tailor the interpretation of the user interaction actions to the type of client device. In fact, smartphones and tablets use different interaction paradigms for which we need further experiments to understand how to mine user interactions for intent understanding purposes.

References

[1] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *Proc. of the International Conference on Research and Development in Information Retrieval*, SIGIR'06, pages 19–26. ACM, 2006.

[2] E. Agichtein, E. Brill, S. Dumais, and R. Ragno. Learning user interaction models for predicting web search result preferences. In *Proc. of the International Conference on Research and Development in Information Retrieval*, SIGIR'06, pages 3–10. ACM, 2006.

[3] A. S. B. J. Jansen, D. L. Booth. Determining the user intent of web search engine queries. In *Proceedings of the International Conference on World Wide Web*, WWW'07, pages 1149–1150. ACM, 2007.

[4] S. Beitzel, E. Jensen, O. Frieder, D. Lewis, A. Chowdhury, and A. Kolcz. Improving Automatic Query Classification via Semi-Supervised Learning. In *Proceedings of the fifth IEEE International Conference on Data Mining*, pages 42–49, 2005.

[5] A. Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002.

[6] H. Cao, D. H. Hu, D. Shen, D. Jiang, J.-T. Sun, E. Chen, and Q. Yang. Context-aware query classification. In *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '09, pages 3–10, New York, NY, USA, 2009. ACM.

[7] H. Cao, D. Jiang, J. Pei, E. Chen, and H. Li. Towards context-aware search by learning a very large variable length hidden markov model from search logs. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 191–200, New York, NY, USA, 2009. ACM.

[8] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li. Context-aware query suggestion by mining clickthrough and session data. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, pages 875–883, New York, NY, USA, 2008. ACM.

[9] E. Carmel, S. Crawford, and H. Chen. Browsing in hypertext: a cognitive study. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 22:865–884, 1992.

[10] Z. Cheng, B. Gao, and T.-Y. Liu. Actively predicting diverse search intent from user browsing behaviors. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 221–230, New York, NY, USA, 2010. ACM.

[11] C. Cortes and V. Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, Sept. 1995.

[12] V. Deufemia, M. Giordano, G. Polese, and G. Tortora. Inferring web page relevance from human-computer interaction logging. In *Proc. of the International Conference on Web Information Systems and Technologies*, WEBIST'12, pages 653–662, 2012.

[13] Q. Guo and E. Agichtein. Exploring mouse movements for inferring query intent. In *Proceedings of the International Conference on Research and Development in Information Retrieval*, pages 707–708. ACM, 2008.

[14] Q. Guo and E. Agichtein. Ready to buy or just browsing? detecting web searcher goals from interaction data. In *Proceedings of the International Conference on Research and Development in Information Retrieval*, SIGIR'10, pages 130–137. ACM, 2010.

[15] Q. Guo and E. Agichtein. Towards predicting web searcher gaze position from mouse movements. In *Proceedings of the International Conference on Human Factors in Computing Systems*, CHI EA'10, pages 3601–3606. ACM, 2010.

[16] Q. Guo and E. Agichtein. Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior. In *Proceedings of the International Conference on World Wide Web*, WWW'12, pages 569–578. ACM, 2012.

[17] Q. He, D. Jiang, Z. Liao, S. C. H. Hoi, K. Chang, E.-P. Lim, and H. Li. Web query recommendation via sequential query prediction. In *Proceedings of the 2009 IEEE International Conference on Data Engineering*, ICDE '09, pages 1443–1454, Washington, DC, USA, 2009. IEEE Computer Society.

[18] I. Kang and G. Kim. Query type classification for web document retrieval. In *Proceedings of the Conference on Research and Development in Informaion Retrieval*, SIGIR'03, pages 64–71. ACM, 2003.

[19] D. Kelly and J. Teevan. Implicit feedback for inferring user preference: a bibliography. *SIGIR Forum*, 37(2):18–28, 2003.

[20] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the International Conference on Machine Learning*, ICML'01, pages 282–289, 2001.

[21] F. Lauer and Y. Guermeur. MSVMpack: a multi-class support vector machine package. *Journal of Machine Learning Research*, 12:2269–2272, 2011.

[22] U. Lee, Z. Liu, and J. Cho. Automatic identification of user goals in web search. In *Proceedings of the International Conference on World Wide Web*, WWW'05, pages 391–400. ACM, 2005.

[23] L.-P. Morency, A. Quattoni, and T. Darrell. Latent-dynamic discriminative models for continuous gesture recognition. In *Proceedings of IEEE Conference Computer Vision and Pattern Recognition*, CVPR'07, pages 1–8, 2007.

[24] J. Morrison, P. Pirolli, and S. C. SK. A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. In *Extended Abstracts on Human Factors in Computing Systems*, CHI EA'01, pages 163–164. ACM, 2001.

[25] G. Murray. Asynchronous javascript technology and XML (ajax) with the java platform. `http://java.sun.com/developer/technicalArticles/J2EE/AJAX/`, 2006.

[26] H. Nguyen. Capturing user intent for information retrieval. In D. L. McGuinness and G. Ferguson, editors, *Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications*

of Artificial Intelligence, July 25-29, 2004, San Jose, California, USA, pages 997–998. AAAI Press / The MIT Press, 2004.

[27] J. Nielsen. F-shaped pattern for reading web content, April 2006. `http://www.useit.com/articles/f-shaped-pattern-reading-web-content/`.

[28] D. Rose and D. Levinson. Understanding user goals in web search. In *Proceedings of the International Conference on World Wide Web*, WWW'04, pages 13–19. ACM, 2004.

[29] A. Sellen, R. Murphy, and K. Shaw. How knowledge workers use the web. In *Extended Abstracts on Human Factors in Computing Systems*, CHI'02, pages 227–234. ACM, 2002.

[30] L. Tamine, M. Daoud, B. Dinh, and M. Boughanem. Contextual query classification in web search. In *Proceedings of International Workshop on Information Retrieval Learning, Knowledge and Adaptability*, LWA'08, pages pp. 65–68, 2008.

[31] J. Wang, A. P. de Vries, and M. J. T. Reinders. A user-item relevance model for log-based collaborative filtering. In M. Lalmas, A. MacFarlane, S. M. Rger, A. Tombros, T. Tsikrika, and A. Yavlinsky, editors, *ECIR*, volume 3936 of *Lecture Notes in Computer Science*, pages 37–48. Springer, 2006.