

PAPER • OPEN ACCESS

Identifying Behavioural Intents in a Complex Search Process Management System

To cite this article: Mengliang Zhu *et al* 2019 *J. Phys.: Conf. Ser.* **1302** 022076

View the [article online](#) for updates and enhancements.

You may also like

- [Design and Implementation of Oil and Gas Information on Intelligent Search Engine Based on Knowledge Graph](#)
Shujun Huang, Yuyan Wang and Xiang Yu
- [Space-fractional Fokker–Planck equation and optimization of random search processes in the presence of an external bias](#)
Vladimir V Palyulin, Aleksei V Checkin and Ralf Metzler
- [Double hashing technique in closed hashing search process](#)
Robbi Rahim, Iskandar Zulkarnain and Hendra Jaya



The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting

Oct 9 – 13, 2022 • Atlanta, GA, US

Early hotel & registration pricing ends September 12

Presenting more than 2,400 technical abstracts in 50 symposia

The meeting for industry & researchers in

BATTERIES
ENERGY TECHNOLOGY
SENSORS AND MORE!

 **ECS Plenary Lecture featuring M. Stanley Whittingham,**
Binghamton University
Nobel Laureate –
2019 Nobel Prize in Chemistry

 **Register now!**



Identifying Behavioural Intents in a Complex Search Process Management System

Mengliang Zhu, Ruikang Xu, Mingzhen Sun and Yin Zhang

School of Computer Science and Engineering, Northeastern University, Shenyang, China.

Email: zhangyin@cse.neu.edu.cn (Corresponding author: Yin Zhang)

Abstract. Understanding search intents is essential for improving search results and query recommendations. Currently, most search intent identification methods work on plain form search logs consisting of queries and clicks. Meanwhile, new search tools are being introduced by the CHI/HCI communities, allowing users to perform more types of behaviors than querying and clicking, and generating rich form search logs with more information than that in plain form search logs. In this paper, we seek to identify search intents by considering various types of behaviors performed in a complex search process management system. We use random forest to identify search intents based on the behaviors performed. Experimental results show that search intents could be identified by analyzing search behaviors, and certain types of behaviors can be good indicators of search intents.

1. Introduction

Understanding search intents is essential to choosing proper search result re-ranking or query recommendation algorithms [1]. Currently, most search intent identification methods follow a contextual approach to understand search intents, that is, to understand search intents by analyzing queries, clicks, and other contextual information [2].

A primary reason of adopting a contextual approach is that these methods focus on identifying search intents from plain form search logs generated by commercial search engines consisting of queries, clicks, timestamps, and user/session IDs. However, the CHI/HCI communities have been developing new tools to support complex search tasks [3]. These new tools allow users to perform more operations than simple querying and clicking and can generate rich form search logs with more information than that in plain form search logs [4]. Such a new trend inspires us to develop new methods of search intent identification by involving the rich behaviors performed in these new search tools.

This paper introduces out attempts to identifying behavioral intents in a complex search process management system. By identifying different types of behaviors and behavioral intents, we develop a random forest based method to identify behavioral intents based on the behaviors performed by searchers. Experimental results show that the proposed method can be used to identify behavioral intents, and certain features such as the number of clicked search results, and the time of reviews can be good indicators of different behavioral intents.

2. Behavioral Intents in a Complex Search Process Management System

In this section, we briefly introduce our complex search process management Time Tree, the behaviors a user could perform on Time Tree, and the corresponding behavioral intents of the behaviors.



2.1. The Complex Search Process Management System Time Tree

In our previous research, we proposed a complex search process management system Time Tree [5]. In Time Tree, each search task is represented as a tree, as shown in Figure 1. In Time Tree, queries are represented as circle nodes, while clicked search results are represented as square nodes. Detailed information about a node is provided in an infobox. Figure 2 shows an example of an infobox. A user could rate, comment on, or delete the node in the infobox. A higher rating will brighten up the node. Users could also reorganize nodes and arcs by drag-n-drop.

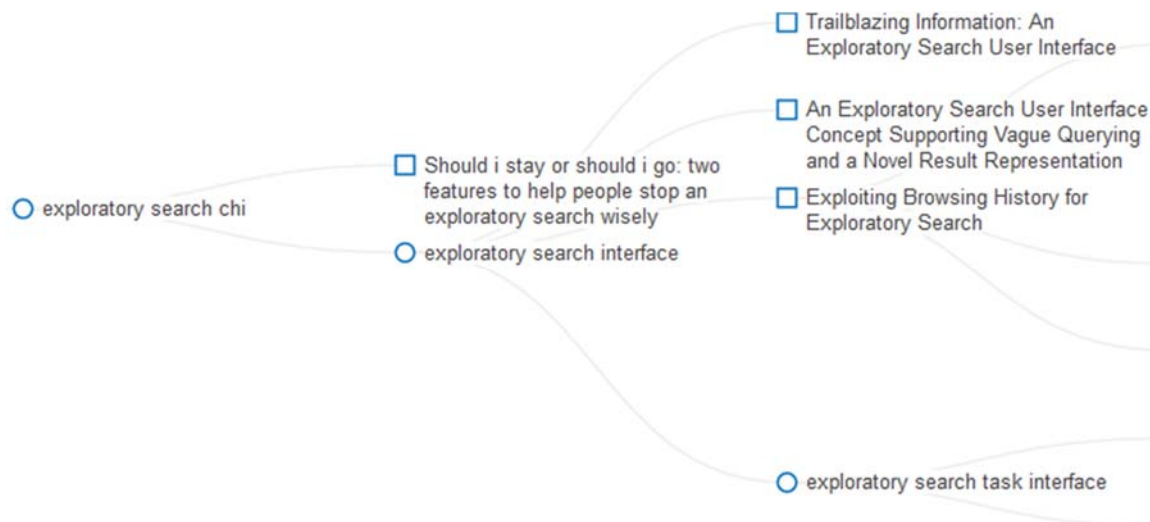


Figure 1. A fraction of a Time Tree for a complex search task. Queries are represented as circle nodes. Clicks are represented as square nodes.

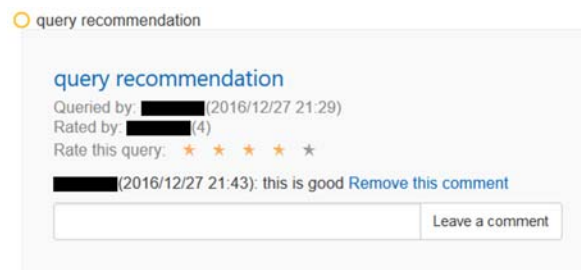


Figure 2. An example of a query node infobox.

2.2. Behaviors and Behavioral Intents in Time Tree

We can identify five types of behavior a user could perform on a Time Tree:

- Add comments: when a user learns something from a query or a click, the user could add comments to the query or the click.
- Delete comments: when a user thinks what he has learned could be wrong, the user could delete the corresponding comments.
- Change ratings: when a user evaluates the importance of a query or a click, the user could change the rating of the query or the click.
- Drag nodes: when a user figures out a better structure of the search process, the user could modify the current structure by dragging nodes.
- Review: a user could review tooltips to recall previous search processes.

We then extract behavior sequences the search behavior log. Given query Q_c , the father query of Q_c is referred as Q_p . Then "search behavior" is defined as the behavior sequence between Q_c and Q_p . There are two requirements for the behavior sequence between Q_c and Q_p :

- All the behaviors in the sequence have to be applied to the ancestors of Q_c ;

- All the behaviors in the sequence have to be reasonably related.
To extract reasonably related behaviors, Algorithm 1 is applied.

Algorithm 1. Algorithm of extracting all ancestor behaviors of Q_c

Input: All the behaviors A_{all} between Q_p and Q_c

Output: All the ancestor behaviors of Q_c

```

1. Begin
2.  Add all the ancestors of  $Q_c$  to set  $N_{ancestor}$ 
3.  For BEHAVIOR in  $A_{all}$ 
4.    If BEHAVIOR is applied to a node in  $N_{ancestor}$  Then
5.      Add BEHAVIOR to OUTPUT_SET
6.    End If
7.  End For
8. Endb

```

Based on Lau and Horvitz's work, we could identify five types of behavioral intents [6]:

- Initial query: the first query in a Time Tree. The initial query is the root node of a Time Tree.
- Specialization: a user is trying to dig into a topic.
- Generalization: a user is trying to get a more general view of a topic.
- Reformulation: a user is trying to use different keywords to describe the same topic.
- Changing directions: a user is changing the topic.

The behavioral intents used here coincide with the intents proposed by Lau and Horvitz. The differences are that, as Time Tree do not record duplicated and null queries, the "Request for additional results" and the "Blank queries" intents are removed.

3. Identifying Behavioural Intents Using Random Forest

In this paper, we use random forest to classify behavioral intents. The training process of classifying behavioral intents are:

- 1) Input tagged training data set X_{train} .
- 2) Randomly choose a training subset $X_{train-subset}$.
- 3) Randomly choose L features as F_L .
- 4) Build decision trees using $X_{train-subset}$ and F_L .
- 5) Repeat 2-4 to build m decision trees.

The process of building a decision tree is a recursive process. The key to the process is to choose a splitting feature. In this paper, we use the Gini impurity to choose splitting features. The definition of the Gini impurity is:

$$Gini(D) = 1 - \sum_{i=1}^m p_i^2 \quad (1)$$

In Equation (1), m gives the number of class C in dataset D . P_i gives the probability of an item in D belongs to class C_i . When using the Gini impurity to build binary decision trees, the new Gini impurity of splitting using feature R is:

$$Gini(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (2)$$

In Equation (2), D_1 and D_2 are nonempty sets and $D_1 + D_2 = D$. The minimum $Gini_R(d)$ is used as R 's Gini impurity. The feature with the maximum Gini increment is used to build the binary decision tree:

$$\Delta Gini(R) = Gini(D) - Gini_R(D)$$

4. Experiments

In this section, we use experiments to evaluate the proposed behavioral intention identification method. We designed complex search tasks and recruited volunteers to log their search behaviors. We then apply the proposed method to identify the behavioral intents and verify the results.

4.1. Data

Sixteen volunteers were recruited to complete two complex search tasks. The requirements of the two complex search tasks are shown as below:

- Write a report about drugs used in chemotherapy. Introduce common drugs used in chemotherapy, including their pharmacological actions, indications, administrations, and side effects. Introduce treatment strategies and common combination of chemotherapy regimens. Do not focus only on drugs that beat cancer. Consider also drugs that decrease the toxic effects of other drugs.
- Write a report about fine particles (PM 2.5) in China. Introduce the concept and sources of PM 2.5. Introduce how and why PM 2.5 affects human health. Introduce the top affected cities by PM 2.5 in China. Analyze the causes of PM 2.5 in the top affected cities. Introduce methods as well as the corresponding mechanisms and feasibilities to reduce PM 2.5. Analyze both the positive and the adverse effects of the introduced methods.

We observed 207 queries and 387 clicks in the experiment. On average, there was 1.86 clicks per query, 25 queries and 48 clicks per user. The number of behaviors observed is shown in Table 1.

Table 1. Behaviors observed in the experiment

	Total # of behaviors	Average # of behaviors per volunteer
Add comments	91	11.38
Delete comments	0	0
Change ratings	49	6.13
Drag nodes	62	7.75
Review	361	45.13

4.2. Results

4-fold cross validation is used in this paper to evaluate how could the proposed method identify behavioral intents. The number of decision trees used in the experiment is 50, and three features are used to build each decision tree. Precision, recall and the F1 measure are used to evaluate the results:

Precision, the proportion of true positives among all the positives

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positive}}$$

Recall, the proportion of true positives found.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

With precision and recall, F1 is defined as:

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The results of the precisions, recalls and the F1s are shown as Table 2.

Table 2. Precisions, recalls and F1s of different behavioral intents

	Precision	Recall	F1
Specialization	0.72	0.65	0.68
Generalization	0.33	0.11	0.17
Reformulation	0.31	0.5	0.38
Changing directions	0.42	0.57	0.48
Average	0.57	0.55	0.55

Table 2 shows that on average, the precision and recall are 57% and 55%. For different intents, the “Specialization” intent has the best precision and recall, following by the “Changing directions” intent. The “Generalization” intent has the worst precision and recall. By analyzing the features of different intents, we could see that the “Specialization” intent exposes the most obvious features. A possible reason for the low performance on the “Generalization” intent could be the limited size of training data for the intent.

The importance of different features is shown in Table 3. Table 3 shows that the number of clicks helps identify different intents the most, following by the time of reviews, and the number of reviews.

Table 3. The importance of different features

	Importance
# of drags	0.075
# of comments	0.11
# of ranks	0.125
# of reviews	0.195
Time of reviews	0.22
# of clicks	0.275

The results show that the proposed method can be used to identify behavioral intents on Time Tree.

5. Conclusion

The study of understanding behavioral intents in modern search tools is still not well developed. This paper introduced a method to identify behavioral intents of user behaviors in a complex search process management system. We introduced our complex search process management system TimeTree, and proposed a random forest based machine learning method to classify behavioral intents. The result has shown that behavioral intents can be identified by analyzing complex search behaviors, and certain behavioral features can be good indicators of behavioral intents. This paper can provide evidence to the future design of more sophisticated methods to understand user intent to provide better search results and query recommendations.

6. Acknowledgments

This research is supported by the National Natural Science Foundation Program of China (61603082, 61502089, 61572116, 61572117), the National Key Technology Support Program (2014BAI17B00), the National Key Technology R&D Program of the Ministry of Science and Technology (2015BAH09F02, 2015BAH47F03), the Special Fund for Fundamental Research of Central Universities of Northeastern University (N161604009), and the Technology Project of Hebei Province of China (16210342, 17210326).

7. References

- [1] Ruotsalo T, Jacucci G, Myllymäki, Petri and Kaski S 2014 Interactive intent modeling: information discovery beyond search *Commun. ACM* **58**(1) 86-92
- [2] Markus K, Petri L, Tuukka R and Mats S 2018 Proactive information retrieval by capturing search intent from primary task context *ACM Transactions on Interactive Intelligent Systems* **8**(3) 1-25
- [3] Zhao Y, Zhang Y, Zhang B and Gao K 2018 Recommending queries by extracting thematic experiences from complex search tasks *Entropy* **20**(6) 459
- [4] Nguyen P, Xu K, Bardill A, Salman B, Herd K and Wong B L 2016 Sensemap: supporting browser-based online sensemaking through analytic provenance *IEEE Conf. on Visual Analytics Science and Technology* (USA).
- [5] Zhang Y, Zhang B, Gao K, Li p, Zhao Y and Zhang C 2018 Towards a coherence-oriented complex search experience management method *Physica A: Statistical Mechanics and its Applications* **506** 844-56
- [6] Lau T 1999 Patterns of search: analyzing and modeling web query refinement *UM99 User Modeling* **407** 119-28