Graph-based Link Prediction in Cross-session Task Identification

Chao Xu[1], Mingzhu Zhu[1], Wei Xiong[2], and Yi-fang Wu[1]
[1]Information Systems, New Jersey Institute of Technology, Newark, NJ, USA
[2]Mathematics, Computer Science and Information Systems, Northwest Missouri State University, Maryville, MO, USA
Email: [1]{cx26, mz59, wu}@njit.edu, [2]xiong@nwmissouri.edu

Abstract **- The information needs of search engine users vary in complexity. Some simple needs can be satisfied by using a single query, while complicated ones require a series of queries spanning a long period of time. The search task, consisting of a sequence of search queries serving the same information need, can be treated as an atomic unit for modeling user search preferences and has been well applied in information retrieval to improve the accuracy of search results. Most existing studies have focused on over-session based task identification and heavily relied on human annotations for supervised classification model learning, which are not ideal in large, real time search applications where users have long-term interests spanning over multiple search sessions. In this study, a cross-session based method is proposed for discovering search tasks by modeling the latent structure of task information in the search log dataset, without needing human annotations. Experimental results show that the proposed cross-session based method contributes to an increased accuracy of task identification.**

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

General Terms

Algorithms, Performance, Experimentation

Keywords

Search Session, Search Task, Search Log

1. Introduction

The information needs of search engine users span a broad spectrum. Some simple information needs, such as finding a person's homepage or navigating a social networking site, can be accomplished in a single search session. Yet addressing complex information needs, such as planning a vacation, organizing a wedding, or repairing a laptop, requires a user to issue a series of queries, spanning a long period of time and over multiple search sessions. For example, if a user's laptop is broken and he wants to find the solution on the internet, usually, he will search a query first, such as "macbook pro broken", and then go through search results. If the user fails to find relevant information, he would most likely revise his query. Moreover, a user may open multiple web browsers and work on several search tasks at the same time. In this study, the user's search activity is examined at the task level based on the session information, where a search task is defined as a unit of representing one distinct information need.

In most of the existing studies [1, 2], a search task is defined as one or multiple sessions that correspond(s) to a distinct information need. The task is extracted based on the segmented session information, which is also used as the unit for extracting user's interests. These methods are referred to as over-session based task identification, because the task information is constructed over the session units. One obvious problem is that it oversimplifies the user's search activity by assuming that users only work on the same search task within a short period of time. Yet people might work on different search tasks at the same time. Thus, it is needed to examine the search task both within and cross session boundaries to improve the performance of task identification.

Recently, several studies have been conducted on identifying tasks within search sessions. For example, some studies [3, 4] adopt supervised methods to label search tasks using a pairwise classification methods. However, pairwise prediction might not be consistent. For example, two pairs: (query q_i and q_j), (query q_i and q_k) are predicted to be in the same task, while query q_j and q_k are not. Meanwhile, some studies [5, 6] use an external dataset such as the Open Directory Project or Wikipedia. Because the labels and categories of search tasks are generated from an external dataset, the total number of labels or categories is fixed rather than adaptive to the user's search activities. However, it is usually the case that most users have multiple information needs and they are dynamically changing [7]. To solve these problems, in this study, a cross-session based query analysis method with a best-link model is proposed to improve the performance of task identification. Specifically, search queries within a search session are segmented into sub-tasks by using the best-link model to learn query connections from users' search activities. And a graph-based representation method is utilized to calculate the contextual pairwise similarity of queries. Then, search tasks are identified by grouping similar sub-tasks from all search sessions together.

This paper makes the following contributions: 1) a cross-session based task identification method; 2) a best-link prediction method for identifying the structural dependencies of queries; and 3) a graph-based representation method for determining the link relation between a pair of queries.

The rest of the paper is organized as follows. Section 2 summarizes related studies. Section 3 presents the proposed cross-session based task identification method. Section 4 introduces the dataset, experimental design, evaluation methods, and performance comparison between the proposed method and baselines. Section 5 summarizes the main conclusions of this study.

2. Related Work

A search session, as defined by Boldi et al. [4], is a sequence of queries issued by a single user within a specific time limit. The related queries of the same session often refer to the same search goal or search activity. Based on this assumption, He et al. [5] propose to group queries into search sessions through detecting the topic shifts among queries. Hassan et al. [6] adopt topic models to extract session-level search goals. It is concluded that the method of examining user search activities through search sessions outperforms the traditional approaches that are based on only relevance feedbacks. Piwowarski et al. [7] model a hierarchy of users' search activities through a layered Bayesian network to identify distinct patterns of users' search behaviors. They use classification methods to learn the connection of latent states for a clicked document to the relevance assessment of that document without considering the document content. Mei et al. [8] propose a framework of studying the sequences of users' search activities, in which an algorithm is introduced to segment the query stream into goals and missions.

Recently, several studies have noticed the necessity of going beyond the session boundary and examining the user's information needs in a task. For example, Spink et al. [17] indicate that multi-tasking behavior occurs frequently in which users switch search tasks within a short period of time. Lucchese et al. [14] model task-based sessions to extract multiple tasks from the search session. Meanwhile, Hassan and White [9] indicate that a search task can be complex and span a number of search sessions. To tackle this, they propose a method to generate a task tour which comprises a set of related search tasks. Kotov et al. [11] explicitly define the cross-session task as the one extending over multiple sessions and corresponding to a certain high-level search intent. To extract cross-session tasks, Jones et al. [18] have built classifiers to identify task boundaries and pairs of queries belonging to the same task. Agichtein et al. [19] have examined the cross-session task identification by using a binary classification method and have found that different types of tasks have different life spans. Besides, a few studies [11, 20, 21, 22, 23, 24, 25] have proven the effectiveness of classifying queries and web pages into search tasks on improving the search performance. Although they prove that the search task information contributes to the improvement of search performance, all of them have two main issues. The first issue is that they define the search task manually. The fixed number of search tasks is not suited to predict the user's future search activities – since it will be an incomplete representation, if the number is too small; and noises will occur, if the number is too large. The second issue is that existing classification-based methods rely on human annotated dataset for training models, which is not applicable when only few manual annotations are available.

The main difference between this study and existing cross-session based task identification studies is that we model this problem as a link prediction problem rather than a binary classification problem. The advantage of this study is that the latent dependencies between queries within each task are modeled explicitly.

3. Methods

3.1 Task analysis

Search logs are proven as a valuable data resource for analyzing user's search activities and information needs. In this study, the AOL search log dataset is examined to extract users' search tasks. A search log is a dataset that records users' search activities, which can be denoted by the vector $< a_i, q_i, t_i, c_i, r_i >$, where a_i is the identifier of the user, q_i is the query submitted by the user a_i, t_i is the time of the user activity, c_i is the click on the relevant result returned for q_i, and r_i is the rank position of c_i [10].

The primary mechanisms for segmenting the logged query streams are session-based. A search session is usually considered as the basic unit of information in search log analysis [1]. In a search engine which works in the session mode, the user's search activities are recorded and earlier search data, i.e. queries and results clicked, in the same session is used to update user's current search actions. A search session is defined as a sequence of search activities $S = \{< a_j, q_j, t_j, c_j, r_j > \ldots < a_k, q_k, t_k, c_k, r_k >\}$ issued by a single user within a specific time limit.

Table 3.1 Sample of Session Segmentation

User_ID	Query	QueryTime	Clicked_URL	Rank
382351	apple warranty	2006-04-24 22:00:21	http://www.superwarehouse.com	6
382351	ipod questions	2006-04-24 22:17:42	http://www.maclink.co.uk	1
382351	dogwood festival	2006-04-29 21:46:30	http://www.fayettevilledogwoodfestival.com	5
382351	myrtle beach map	2006-05-29 22:58:09	http://travel.yahoo.com	3
382351	cherry grove south carolina	2006-05-29 23:03:03	http://www.tripadvisor.com	4
382351	cherry grove south carolina	2006-05-29 23:03:03	http://www.cherrygrovebeachhouses.com	9
382351	body kits for civic	2006-05-30 20:03:12	http://www.modacar.com	2
382351	motley crue jackets	2006-03-01 17:41:26	http://www.motley.com	9
382351	ticketmaster	2006-03-16 14:40:40	http://www.ticketmaster.com	1

Methods of extracting relevant sessions from search logs should examine all queries issued by a user. Short inactivity timeouts between user actions are applied as a means of demarcating session boundaries [4]. In the field of session segmentation, the relations between queries are categorized as Topic Continuation and Topic Shift. In Figure 3.1, query q_1 and q_2 are semantically related, so they should be grouped in the same session and the relation between them is Topic Continuation. On the contrary, q_2 and q_3 have no semantic relation, so the relation between them is Topic Shift, which generates a session boundary. In this study, user inactivity periods are adopted to segment the search session. The time interval within a search session should be less than a threshold σ (where σ is set at 25 minutes according to an empirical study). Table 3.1 shows a sample of segmented sessions.

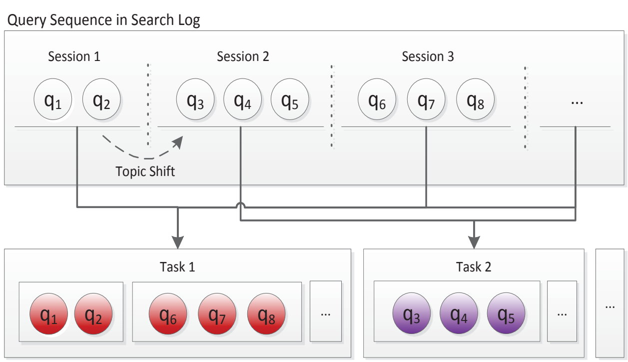

Figure 3.1 Task identification by grouping similar search sessions.

Meanwhile, search engine users have various search intentions. Addressing complex information needs usually requires a user to issue a series of queries, spanning across multiple search sessions. To tackle this problem, a fine-grained task identification method, which is also called the cross-session based task identification method, is proposed in this study. As shown in Figure 3.2, search queries within a search session are segmented into sets of queries which are formed to achieve specific search tasks. Each set of queries is called a sub-task. For example, in the first session, predicting q_2, q_4 and q_5 belonging to the same task would immediately lead to the conclusion that all these three queries are in the same task, even though q_2 and q_5 are not directly connected to each other. Then, after examining all search sessions of the user, search queries related to a particular search task are identified by grouping similar sub-tasks together.

To generate these sub-tasks for each search session, an unsupervised best-link model is proposed. The main idea is that the best-link defines a hierarchical tree structure of "strong" connections among the queries: rooted in the fake query q_0, and each sub-tree of q_0 corresponds to one specific search sub-task in a search session. For a new query, it can only belong to a previous search task or be the first query of a new task. Therefore, the temporal order provides a helpful signal to explore the dependency between queries.

Specifically, given a query sequence $Q = \{q_1, q_2, \ldots, q_m\}$ within a search session, f is introduced to refer the latent best-link structure. $f(q_i, q_j)$ indicates the existence of a link between q_i and q_j as following:

$$f(q_i, q_j) = \begin{cases} 1, & \varphi(q_i, q_j) > \gamma \\ 0, & \text{otherwise} \end{cases} \tag{3.1}$$

where $f(q_i, q_j) = 1$, if query q_i and q_j are directly connected; and otherwise, $f(q_i, q_j) = 0$. $\varphi(q_i, q_j)$ indicates the similarity between query q_i and q_j. To model the first query of a new search session, i.e., the query that does not have a strong connection with any previous queries, a fake query q_0 is added at the beginning of each search session. All the queries connecting to q_0 would be treated as the initial query of a new search sub-task. Besides, it is enforced so that a query can only link to another query in the past, or formally,

$$\sum_{i=0}^{j-1} f(q_i, q_j) = 1, \forall j \geq 1 \tag{3.2}$$

Note that the best-link method is conducted within each search session to generate a list of sub-tasks. Similar sub-tasks are grouped together as a search task using the hierarchical clustering [8].

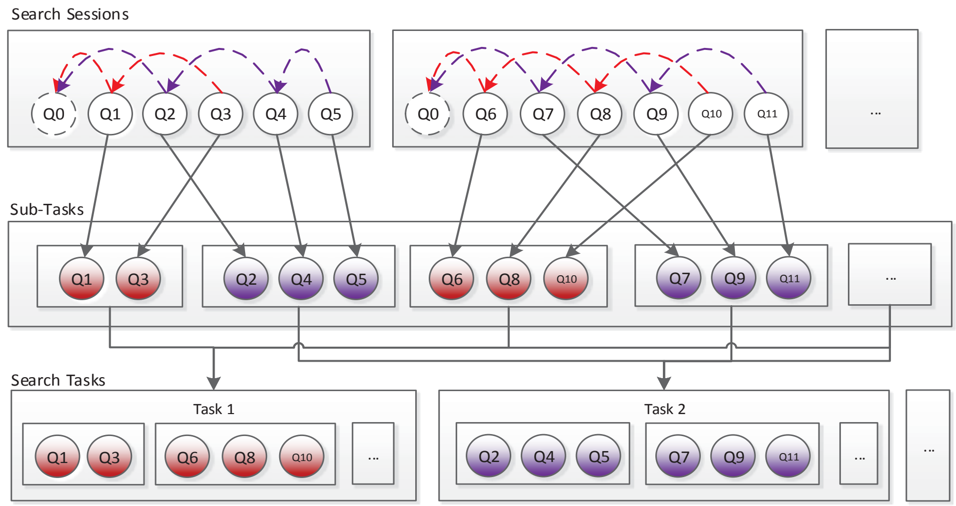

Figure 3.2 Task identification by grouping similar sub-tasks.

3.2 Graph-based Link Prediction

To achieve the latent structure $f(q_i, q_j)$ as defined in formula 3.1, $\varphi(q_i, q_j)$ should be determined first. As shown in Figure 3.3, the pairwise similarity between relevant feedback documents of q_i and q_j is adopted for determining the link relation between two queries. Specifically, the queries resulting in none click action are defined as invalid queries, such as q_3, q_4 and q_6. By contrast, the queries resulting in at least one clicked result are defined as valid queries, such as q_2 and q_5. All invalid queries are ignored in this study as did in one existing study [16]. For example, to determine if q_2 and q_5 belong to the same task, two similarities between the relevant feedback documents of these two queries are calculated, including $sim(d_{2,1}, d_{5,3})$ and $sim(d_{2,1}, d_{5,5})$, where $d_{2,1}$ denotes the first retrieved document of q_2, sim() represents the similarity between a pair of queries. Then q_2 and q_5 are segmented into the same task if $sim(d_{2,1}, d_{5,3})$ or/and $sim(d_{2,1}, d_{5,5})$ is/are bigger than the γ as indicated in formula 3.1.

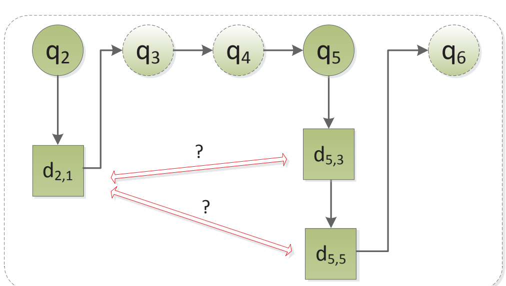

Figure 3.3 Example of the pairwise similarity.

However, there are two problems of calculating the above pairwise similarity using the original page contents, including data noise and data scarcity [12]. On one hand, many relevant documents contain other non-pertinent information such as advertisements and navigations, causing difficulty in summarizing their latent meanings. On the other hand, for a search log dataset, such as AOL, it does not contain snippets, but URLs that might not point to a live site anymore, or for which the content might have been changed after the dataset was created. To tackle these problems, a two-step graph-based representation method is proposed for predicting the pairwise similarity between the relevance feedback documents from two different search queries.

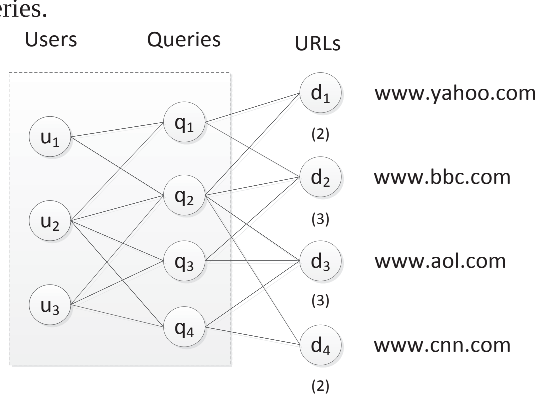

Figure 3.4 Example of a click graph.

First, a click graph is constructed for generating the pseudo-document of each clicked URL. An example of a click graph with four queries and four URLs is shown in Figure 3.4. The edges of the graph capture the relationships between the queries and the URLs. Based on the observation that different users may use different queries to describe their latent topics of interests within a particular web page, it is proposed to generate a pseudo-document for each URL by combining all its connected queries in this graph. For example, two different queries (q_1 and q_2) from different users (u_1, u_2 and u_3) are connected to the same URL, “www.yahoo.com”. The queries (q_1 and q_2) are then combined to represent the pseudo-content of "www.yahoo.com".

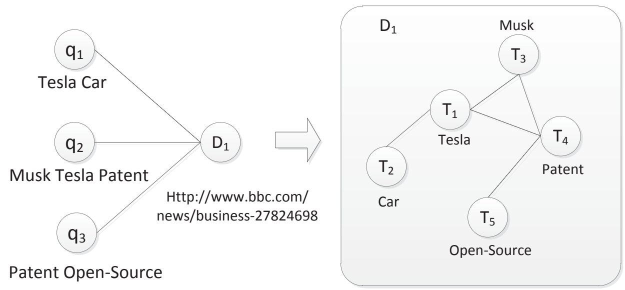

Figure 3.5 Graph-based representation of a relevance feedback document.

Second, simply adopting a bag-of-word to represent the content of a document will lose the structural semantic information. To tackle it, a graph-based representation of the pseudo-document is proposed. Specifically, the unique terms, denoted as $\{T_i\}$, are extracted from the pseudo-document. For example, as shown in Figure 3.5, there are five unique terms within the pseudo-content of D_1, including T_1: "Tesla", T_2: "Car", T_3: "Musk", T_4: "Patent", and T_5: "Open-Source". Afterwards, a pair-wise examination is automatically conducted within each query string to determine the existence of a binary non-directional edge between two terms. For example, T_1 and T_2 are connected with an edge because they are in the same query q_1; T_2 and T_3 are not connected because no query in D_1 contains both of them. Then each pseudo-document is represented as a graph G = (N, E), where N denotes the nodes (unique terms) and E denotes the edges. Finally, given two semantic graphs $G_1 = (N_1, E_1)$ and $G_2 = (N_2, E_2)$ constructed for two relevance feedback documents, a graph similarity measure is adopted to estimate their semantic relatedness. Specifically, the metric called "p-homomorphism" [13] is adopted as the underlying graph matching method, because the p-homomorphism concept extends the traditional graph homomorphism and sub-graph isomorphism concepts by additionally mapping edges from one graph to their corresponding edge paths in another graph.

4. Experimental Design

4.1 Data Sets and Evaluation Methods

Lucchese et al. [14] develop a Web application that helps human assessors manually identify the optimal set of user tasks from the AOL query log. They produce a ground truth that can be used for evaluating any automatic user task discovery method, which is also publically available at "http://miles.isti.cnr.it/~tolomei/downloads/aol-task-ground-truth.tar.gz". It contains a total of 554 search tasks with average 2.57 queries per task. And 143 cross-session tasks are contained in this dataset. In this experiment, this dataset was adopted as the ground truth for comparing the performance of the proposed task identification method and the baselines.

To evaluate the performance of the proposed task identification method, it is necessary to measure the degree of consistency between manually-extracted user tasks of the ground truth and search tasks generated by our algorithms. Specifically, both classification- and similarity-oriented measures [14] were adopted in this experiment. Predicted task indicates the user task where a query is assigned by a specific algorithm, while true task indicates the user task where the same query is in the ground truth.

Classification-oriented approaches measure how closely predicted tasks match true tasks. F1 is one of the most popular measures in this category, as it combines both precision and recall. In this study, precision measures the fraction of queries that were assigned to a user task and that were actually part of that user task. Instead, recall measures how many queries were assigned to a user task among all the queries that were really contained in that user task. Globally, F1 evaluates the extent to which a user task contains only the queries that were actually part of it. Two notations, $p_{i,j}$ and $r_{i,j}$, are introduced to represent the precision and recall of predicted task i with respect to true task j, then F1 corresponds to the following weighted harmonic mean of $p_{i,j}$ and $r_{i,j}$.

$$F1 = 2 \times p_{i,j} \times r_{i,j} / (p_{i,j} + r_{i,j}) \qquad (4.1)$$

Similarity-oriented measures consider pairs of objects instead of single objects. Let T be the sets of predicted tasks, four values were computed, including: 1) t_n - number of query pairs that are in different true tasks and in different predicted tasks (true negatives); 2) t_p - number of query pairs that are in the same true task and in the same predicted tasks (true positives); 3) f_n - number of query pairs that are in the same true task but in different predicted tasks (false negatives); 4) f_p - number of query pairs that are in different true tasks but in the same predicted task (false positives). Then, two different measures were adopted as following:

Rand index: $$R(T) = (t_n + t_p) / (t_n + f_p + f_n + t_p) \qquad (4.2)$$

Jaccard index: $$J(T) = t_p / (f_p + f_n + t_p) \qquad (4.3)$$

4.2 Experimental Setup and Results

The experiment analyzed the contributions of the proposed cross-session based task identification methods including best-link method (BL) and best-link with graph-based representation method (BL-G). The difference is that BL adopts the bag-of-word method for representing the features of the pseudo-document while BL-G uses the proposed graph-based representation method for modeling rich semantic features.

Three baselines were adopted in this experiment, including one over-session based method and two cross-session based methods. The best performing over-session based method (OS) is proposed by Luxenburger et al. [15] who adopt a hierarchical clustering method in which the atomic units to be clustered are past sessions. The two best performing cross-session based methods, QC_wcc and QC_htc, are proposed by Lucchese et al. [14]. Specifically, QC_wcc performs clustering by dropping “weak edges” among queries and extracting the connected components as tasks. QC_htc assumes that a cluster of queries can be well represented by only the chronologically first and last queries in the cluster; therefore only the similarity among the first and last queries of two clusters is considered in the agglomerative clustering.

The annotated log dataset was randomly split into a training set with 270 annotated search tasks, and a test set with the other 270 annotated tasks. The parameters in each model were tuned by a 5-fold cross-validation on the training set. All baselines and our methods were trained on the same training set.

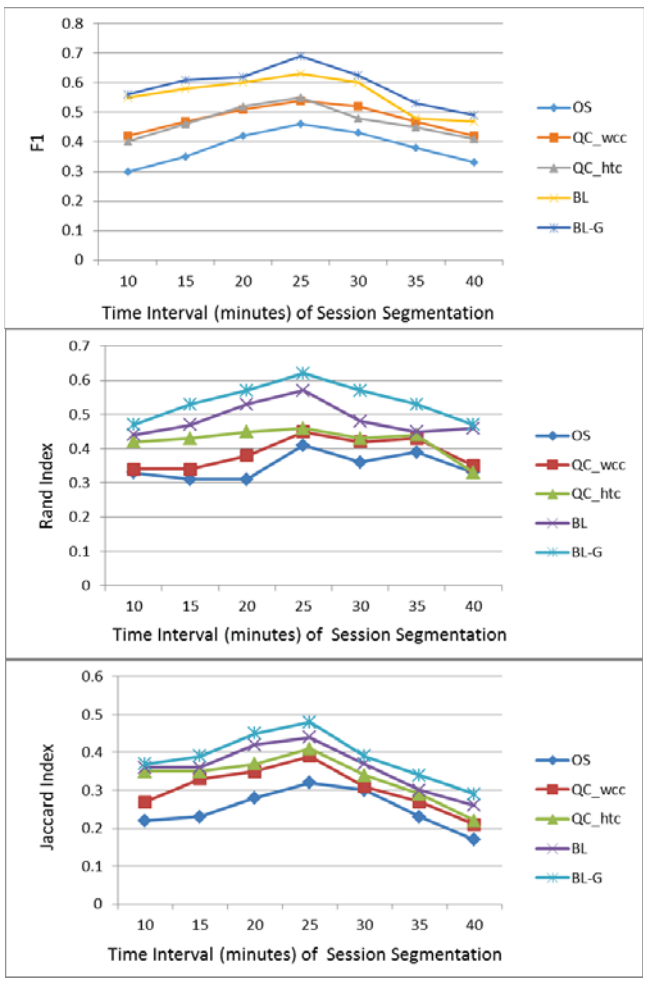

Figure 4.1 Performance comparisons of proposed methods with baselines.

Figure 4.1 shows the performance comparisons between proposed methods and baselines. It was first observed that the session boundary does impact the performance of all compared task identification methods. Most of them achieve the highest performance on these three evaluation metrics when the time interval is set at 25 minutes, which is consistent with existing studies [7, 15]. The proposed methods BL and BL-G outperformed QC_wcc and QC_htc significantly in all three metrics. The reason is that both QC_wcc and QC_htc target on predicting whether two queries represent the same task. However, the pairwise prediction cannot directly generate the task information and post-processing is required to obtain the tasks. Such a post-processing is independent from the classifier training therefore is not necessarily optimal.

Also, the OS baseline, as the over-session based method, performed much worse than the others especially on Rand Index and Jaccard Index metrics. The possible reason is that it assumes that users work on the same task within each period of a search session which results in a high f_p value. Finally, BL-G performs better than BL, because BL-G utilizes the proposed graph-based representation while BL adopts the bag-of-word representation in which the semantic structure is lost.

Table 4.1 Performance Comparisons between Session-based and Non-session based Task Identification Methods

Task Identification Methods		Evaluation Metrics		
		F1	Rand Index	Jaccard Index
Non-session based	BL-NoSS	0.560	0.478	0.422
	BL-G-NoSS	0.603	0.539	0.439
Session-based	BL	0.628	0.571	0.446
	BL-G	**0.695**	**0.619**	**0.483**

So far, the proposed best-link model for task identification is conducted within a session scope. One interesting question is whether the session information is contributive in the proposed best-link method. Table 4.1 illustrates the performance comparisons between the best-link methods using the search session and the ones without using the session data (denoted as BL-NoSS and BL-G-NoSS respectively). Note that both BL and BL-G were optimized by setting session interval at 25 minutes. It was observed that the proposed methods, BL and BL-G, using session information performed much better than the ones without using the session data, i.e., BL-NoSS and BL-G-NoSS. For example, the F1 scores of BL and BL-G were 0.628 and 0.695, whereas those of BL-NoSS and BL-G-NoSS were 0.560 and 0.603. The major reason for these performance differences is that the session plays the role of setting a temporal boundary for identifying the latent link structure of queries from the same search task. And this boundary prevents the incorrectly predicted link information from spanning so that the prediction error made in previous session will not affect the prediction accuracy in the current session. Furthermore, the fact that BL-G and BL-G-NoSS outperformed BL and BL-NoSS respectively, indicates that the proposed graph-based representation for query similarity computation is more effective.

5. Conclusions

Users switch search tasks frequently during their search activities, thus developing methods to extract these tasks from historical data is central to understanding longitudinal search behaviors and developing search systems to support users' long running tasks. In this study, a new cross-session based method is presented for extracting search tasks from users' historic search activities. Specifically, a best-link model is introduced which is capable of learning query connections from users' searching activities. Then a graph-based representation method is utilized to estimate the contextual pairwise similarity of queries. Finally, an experiment using a publically available annotated dataset from AOL log is conducted to demonstrate the superior performance of our method in identifying search tasks versus a number of state-of-the-art algorithms.

6. References

[1] B. Tan, X. Shen, and C. Zhai, "Mining long-term search history to improve search accuracy," Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining, Philadelphia, PA, USA, 2006.

[2] M. Ji, J. Yan, S. Gu, J. Han, X. He, W. V. Zhang, "Learning search tasks in queries and web pages via graph regularization," Proceedings of the 34th international ACM SIGIR conference on research and development in Information Retrieval, Beijing, China, 2011.

[3] D. H. Widyantoro, T. R. Ioerger, and J. Yen, "An adaptive algorithm for learning changes in user interests," Proceedings of the eighth international conference on information and knowledge management, Kansas City, Missouri, United States, 1999.

[4] P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna, "The query-flow graph: model and applications," Proceedings of the 17th ACM conference on information and knowledge management, Napa Valley, California, USA, 2008.

[5] D. He, A. Göker, and D. J. Harper, "Combining evidence for automatic web session identification," information process of management , 2002.

[6] A. Hassan, R. Jones, and K. L. Klinkner, "Beyond DCG: user behavior as a predictor of a successful search," Proceedings of the third ACM international conference on Web search and data mining, New York, New York, USA, 2010.

[7] B. Piwowarski, G. Dupret, and R. Jones, "Mining user web search activity with layered bayesian networks or how to capture a click in its context," Proceedings of the second ACM international conference on Web search and data mining, Barcelona, Spain, 2009.

[8] Q. Mei, K. Klinkner, R. Kumar, and A. Tomkins, "An analysis framework for search sequences," Proceedings of the 18th ACM conference on information and knowledge management, Hong Kong, China, 2009.

[9] A. Hassan and R. W. White, "Task tours: helping users tackle complex search tasks," Proceedings of the 21st ACM international conference on information and knowledge management, Maui, Hawaii, USA, 2012.

[10] B. Zhou, D. Jiang, J. Pei, and H. Li, "OLAP on search logs: an infrastructure supporting data-driven applications in search engines," Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining, Paris, France, 2009.

[11] A. Kotov, P. N. Bennett, R. W. White, S. T. Dumais, and J. Teevan, "Modeling and analysis of cross-session search tasks," Proceedings of the 34th international ACM SIGIR conference on research and development in Information Retrieval, Beijing, China, 2011.

[12] H. C. Wu, R. W. P. Luk, K. F. Wong, and K. L. Kwok, "Probabilistic document-context based relevance feedback with limited relevance judgments," Proceedings of the 15th ACM international conference on information and knowledge management, Arlington, Virginia, USA, 2006.

[13] W. Fan, J. Li, S. Ma, H, Wang., and Y. Wu, "Graph homomorphism revisited for graph matching," Proceeding of the VLDB endowment, 2010.

[14] C. Lucchese, S. Orlando, R. Perego, F. Silvestri, and G. Tolomei, "Identifying task-based sessions in search engine query logs," Proceedings of the fourth ACM international conference on Web search and data mining, 2011.

[15] J. Luxenburger, S. Elbassuoni, and G. Weikum, "Matching task profiles and user needs in personalized web search," Proceedings of the 17th ACM conference on information and knowledge management, California, USA, 2008

[16] L. Bing, W. Lam, and T.-L. Wong, "Using query log and social tagging to refine queries based on latent topics," Proceedings of the 20th ACM international conference on information and knowledge management, Glasgow, Scotland, UK, 2011.

[17] A. Spink, M. Park, B. J. Jansen, and J. Pedersen, "Multitasking during web search sessions," Information processing and management, 2006.

[18] R. Jones and K. L. Klinkner, "Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs," Proceedings of the 17th ACM conference on information and knowledge management, 2008

[19] E. Agichtein, R. W. White, S. T. Dumais, and P. N. Bennet, "Search, interrupted: understanding and predicting search task continuation," Proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval, 2012

[20] C. Xu, M. Zhu, Y. Liu, and Y. B. Wu. "User Profiling for Query Refinement", 20th Americas Conference on Information Systems, 2014

[21] D. Shen, J. Sun, Q. Yang, and Z. Chen. "Building bridges for web query classification," Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval, 2006

[22] Z. Liao, Y. Song, L. He, and Y. Huang, "Evaluating the effectiveness of search task trails," Proceedings of the 21st international conference on World Wide Web, 2012

[23] C. Xu, M. Zhu, Y. Liu, Y. B. Wu, "Personalizing Query Refinement Based on Latent Tasks", 2014 International Conference on Data Mining, 2014

[24] M. Zhu, C. Xu and Y. B. Wu, "Topic Model based Query Intent Prediction for Search by Multiple Examples". 2014 International Conference on Artificial Intelligence, 2014

[25] P. S. Dhillon, S. Sellamanickam, and S. K. Selvaraj, "Semi-supervised multi-task learning of structured prediction models for web information extraction," Proceedings of the 20th ACM international conference on information and knowledge management, 2011