A Real-Time Eye Tracking Based Query Expansion Approach via Latent Topic Modeling

Yongqiang Chen[1], Peng Zhang[1], Dawei Song[1,2], Benyou Wang[1],
[1]Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, China
[2]Department of Computing and Communications, The Open University, United Kingdom
{cyq, pzhang, waby}@tju.edu.cn, dawei.song2010@gmail.com

ABSTRACT

Formulating and reformulating reliable textual queries have been recognized as a challenging task in Information Retrieval (IR), even for experienced users. Most existing query expansion methods, especially those based on implicit relevance feedback, utilize the user's historical interaction data, such as clicks, scrolling and viewing time on documents, to derive a refined query model. It is further expected that the user's search experience would be largely improved if we could dig out user's latent query intention, in real-time, by capturing the user's current interaction at the term level directly. In this paper, we propose a real-time eye tracking based query expansion method, which is able to: (1) automatically capture the terms that the user is viewing by utilizing eye tracking techniques; (2) derive the user's latent intent based on the eye tracking terms and by using the Latent Dirichlet Allocation (LDA) approach. A systematic user study has been carried out and the experimental results demonstrate the effectiveness of our proposed methods.

Category and Subject Descriptors: H.3.3 [Information Search and Retrieval]

Keywords: Eye Tracking, Query Expansion, Real Time, Implicit Relevance Feedback, LDA

1. INTRODUCTION

Query expansion based on relevance feedback has long been studied for its ability of finding out more relevant documents against ambiguous queries that users might type in. Compared with explicit relevance feedback that often causes the users extra cognitive overhead, implicit relevance feedback (IRF) has the advantage of obtaining the useful feedback information from the user interaction data to better infer users' search intention [6], yet without requiring the users explicit relevance judgments.

Traditional implicit feedback based query expansion methods usually return static results according to the searchers' historical log data, which cannot fully meet searchers' dynamic information needs. Recently various real-time IRF

approaches have been proposed [12] [7] [11], to better improve search performance and users' search experience. For example, Singh et al [11] argued that a query expansion framework should expand user's search query dynamically based on user's implicit feedback provided at the time of searching, in order to provide sufficient clues to reflect what the user wants.

Eye Tracking has been used in IR for its ability to record users' eye movement data, which can reveal the users' cognitive process when going through the retrieved documents in a natural way. Gwizdka *et al.* [5] have thoroughly examined the relevance of a document and cognitive effort a user may take using eye movement data. Ajanki *et al.* [1] designed an eye tracking experiment, in which participants were asked to search for relevant documents given a topic, and then the gaze locations were used to find relevant terms to reformulate the queries. However, they had to avoid scrolling the text due to the risk of missing some gaze location to word mappings. Furthermore, this work was not concerned about real-time query expansion.

Buscher *et al.* [3] used eye tracking to extract words based on optical character recognition (OCR) technology, and assign different words with different weights according to the words being read or skipped. They have proved that applying eye tracking as a new data source is feasible for implicit feedback. However, a limitation is that using OCR technique to extract words could not operate in real-time, i.e., the words can not be immediately grabbed when users are reading them. Furthermore, the number of words extracted by using eye trackers may be too small to fully express the users' search intention.

To tackle the problems described above, we propose a real time eye tracking based query expansion model via latent topic modeling. Different from Buscher *et al.* [3], we use a screen word-capturing technique to capture words that the searcher is reading in real-time, from which we could infer what the searcher is currently interested in. In our approach, we will have already refreshed the result list according to the captured words when the searcher clicks on the refresh or the next page button. The words captured by eye tracker are used to expanding the original query. According to [2], documents can be considered as being generated by different latent topics, each of which is a probability distribution of words. Based on this assumption, we apply LDA to further derive the searcher's latent information needs, which naturally relate to the words the user pays attention to. Our experiments show that combining the words captured by

eye-tracking and the related latent topics can improve the retrieval effectiveness.

2. METHODOLOGY

In this section, we first introduce how we capture a set of terms (or words), denoted as W^{Eye}, from the screen using an eye tracker when the user is viewing the returned documents. The words in W^{Eye} can be considered as words of interest (WoI) and should be properly weighted. We then use LDA to model user's latent topical preferences by extracting another set of terms W^{LDA} from the topics that are ranked according to W^{Eye}. Finally, we combine the weights of the each word in W^{Eye} and W^{LDA} by different strategies to expand the original queries.

2.1 Capturing Words of Interest (WoI) through Eye Tracking

Gwizdka *et al.* [5] have revealed that a user might gaze for a longer time on relevant words than non-relevant ones, and the user's pupil diameter will expand when seeing interesting words during reading. Accordingly, we develop a WoI capture method to capture the words a user is interested in, based on the user's real-time eye-tracking activity.

To begin with, the high accuracy of modern eye trackers allow us to use suitable APIs to obtain the exact set of raw coordinates $\{X_i, Y_i\}$ on the screen in the millimeter level, for the regions where a user gazes on. In addition, there are existing screen word-capturing technologies for capturing words based on mouse cursors, as long as the mouse cursor stay steady over a word for a certain period. Therefore, we could replace the mouse cursor positions with the coordinates set $\{X_i, Y_i\}$ provided by eye tracking. However, the eye tracker's sampling frequency is 300 HZ. Thus the coordinates in $\{X_i, Y_i\}$ may change too fast for the mouse-based screen word-capturing system to capture. To solve the problem, we adopt a real-time fixation filtering method [9] to cluster the raw coordinates in order to generate more stabilized fixation coordinates. Considering the mean minimum time to acquire the full meaning of a word is 151ms [10], we filter out those clustered fixation coordinates that last less than 151 ms. After we get the stabilized fixation coordinates, we can make use of the screen word-capturing technique to capture words with eye tracking in real-time. The captured words (after removing the stopwords but without doing stemming) $W^{Eye} = \{w_1^{Eye}, w_2^{Eye}, \cdots w_s^{Eye}\}$, where s denotes the number of words in W^{Eye}, are written into a temporary file for each user in each search session, which is archived or deleted after the session is completed.

Different from the traditional off-line extracted eye movement features such as regression ratio, which can be only calculated after collecting all eye movement data, we simply record 4 real-time computable features for each captured word w, namely fixation times FT_w, fixation duration FD_w, left eye pupil diameter LEP_w and right eye pupil diameter REP_w.

2.2 Weighting Scheme for WoI

Our weighting scheme for WoIs takes into account the traditional term frequency and the eye tracking features reflecting user's interest.

For each term t in W^{Eye}, we first normalize its term frequency as follows:

$$\mathcal{L}_t = \frac{wf_t^{Eye}}{wf_{max}^{Eye} + \mathcal{C}} \tag{1}$$

where $\mathcal{L}_t$ represents the normalized score for term t. wf_t^{Eye} represents the frequency of each term t and wf_{max}^{Eye} denotes the max frequency of the terms in W^{Eye}. $\mathcal{C}$ is a positive integer constant to guarantee $\mathcal{L}_t$ falls into the range from 0 to 1.

Furthermore, we use eye-tracking information to measure the searcher's degree of interest in the term t, denoted as $\mathcal{I}_t$:

$$\mathcal{I}_t = \mu FT_t + \nu FD_t + (1 - \mu - \nu)\mathcal{G}_t \tag{2}$$

where FT_t and FD_t represent the fixation times and fixation duration for t, respectively. The parameters μ and ν are coefficients for combining different features.

The function $\mathcal{G}_t$ denotes the changing rate of pupil diameters of searcher's both eyes, and it can be calculated by:

$$\mathcal{G}_t = \frac{1}{2}(LEP_t + REP_t) - \frac{1}{s}\sum_{i=1}^{s}(LEP_i + REP_i) \tag{3}$$

where LEP_t and REP_t denote the pupil diameters of the user's left eye and right eye, respectively, and s is the number of words in W^{Eye}.

Finally, the overall weight of each term t in W^{Eye} is computed as:

$$\Omega_t^{Eye}(t|t \in W^{Eye}) = \mathcal{L}_t * \mathcal{I}_t \tag{4}$$

2.3 Deriving and Incorporating Latent Topics

A document may consist of several latent topics, yet users may be just interested in one or two of them. Furthermore, users often tend to check a small number of documents in the search engine result pages (SERP) to find out the information they need. Thus it is likely that the number of captured words of interest W^{Eye} may remain too small to fully model the user's latent intention. To tackle this problem, we adopt LDA, a probabilistic laten topic model to further locate users' latent search intention. LDA assumes that documents are generated by different latent topics which are probabilistic distributions over different words. Practically, we train a topic model by LDA using all the documents in the test collection, and use Gibbs sampling to conduct approximate inference of all the parameters in LDA.

Given a query $Q = \{q_1, q_2, \cdots q_l\}$ (where l denotes the number of terms in the query) and a WoI set W^{Eye}, we first rank the topics according to the probabilities that they generate Q and W^{Eye}, and choose the top $\mathcal{M}$ words ranked in each topic distribution to expand the original query. The formula for ranking topics is given as follows:

$$P(z_n) = \gamma \sum_{j=1}^{l} f_j^Q P(q_j|z_n, \alpha, \beta) + \\ (1 - \gamma) \sum_{j=1}^{s} f_j^{W^{Eye}} P(w_j^{Eye}|z_n, \alpha, \beta) \tag{5}$$

where $P(z_n)$ is the probability for a topic z_n generating Q and W^{Eye}; f_j^Q and $f_j^{W^{Eye}}$ denote the frequency of term q_j in the query Q and the frequency of w_j^{Eye} in W^{Eye} respectively. γ is a constant parameter to balance the combination

between Q and W^{Eye}. In addition, α and β are the hyper-parameters used in LDA.

We then rank all the topics in descending order according to Formula (5). Then we choose top $\mathcal{M}$ words from each of the top ranked $\mathcal{N}$ topics for each query Q. Consequently, a total of $\mathcal{M}*\mathcal{N}$ words are selected, forming a new set of WoIs, denoted as W^{LDA}, which can be used for query expansion.

For each term t in W^{LDA}, the weight Ω_t^{LDA} is calculated as below:

$$\Omega_t^{LDA}(t|t \in W^{LDA}) = \sum_{i=1}^{\mathcal{N}} P(t|z_i) \qquad (6)$$

where z_i is one of the top ranked $\mathcal{N}$ topics and $P(t|z_i)$ is the probability that topic z_i generates the term t.

Now we get two sets of WoIs W^{Eye} and W^{LDA}, and all of them contain useful information for inferring users' search intention. Therefore, it is reasonable to integrate them in order to achieve better search performance. We use $W^{Eye+LDA}$ to denote the union of W^{Eye} and W^{LDA}. If a term appears in both W^{Eye} and W^{LDA}, we sum up its weights from the two sets and then normalize the summed weight into the range between 0 and 1.

2.4 Generating Expanded Queries

After we have calculated the weight of each term, we use Lemur toolkit [1] and Indri [2] search engine to implement our algorithm and to generate the expanded queries from W^{Eye} and $W^{Eye+LDA}$:

$$\#weight((1-\lambda)Q + \lambda W^{Eye}) \qquad (7)$$

$$\#weight((1-\lambda)Q + \lambda W^{Eye+LDA}) \qquad (8)$$

where Q represents the original query and λ is the combination coefficient.

3. EXPERIMENTS

To verify whether the proposed eye tracking based query expansion can improve retrieval performance, we have conducted an empirical study consisting of a series of user task-based experiments.

We implemented two proposed approaches for eye-tracking based query expansion. The first method we take only uses the eye-tracking WoIs W^{Eye}, denoted as Eye-QE. We use Eye-LDA-QE to denote the another one, which combines eye-tracking and latent topic modeling, i.e., $W^{Eye+LDA}$, for query expansion. For comparison, we take two traditional methods as baselines. The first one is a classical language model LM [13], which performs well in most cases without query expansion. The second baseline, $RM3$ [8], is a widely accepted query expansion approach for its power in boosting the search performance.

3.1 Eye Tracker Setup and Participants' Tasks

In our experiment, participants are required to complete a number of ad-hoc retrieval tasks, with queries and documents selected from the TREC AP8890 collection (to be detailed in the next subsection). We recruited 20 postgraduate students from three different departments to diversify

the users' background. They are all seated in front of a 23" LCD monitor at a screen resolution of 1920*1080 pixels. The eye tracker we employ is Tobii TX300, which has a sample frequency of 300 HZ and the accuracy is $0.3°$ of visual angle. Before conducting the experiment, each participant has to make a calibration to ensure the eye movement data captured at high precision. The participants are then instructed to complete several queries in front of the eye tracker, and for each query they have to find at least 5 relevant documents based on the description and narrative of the query.

3.2 Data Selection

Traditional query expansion methods can perform badly upon the so-called "hard" queries. In our experiment, the hard queries are determined by the low average precision of the retrieval results using the RM3 model. We expect our models would perform better especially when dealing with those queries. Considering that reading too many documents may increase the participants' cognitive burden and may even exhaust them, we select 6 queries for all participants, including 4 hard queries and 2 normal queries, from AP8890 data set. During the process of searching for relevant documents, the average number of documents the participants examined is about 10. After all participants have finished all the queries we collected almost 20*6*10=1200 documents in total and each participant was paid 10 \$ for about 1 hour of work.

3.3 Results and Discussions

In this section, we present our experiment results of the 4 comparative methods we have described.

We use the Mean Average Precision (MAP) as the main performance indicator, as it has been widely used and have a good discrimination power. We first compare the MAP scores of our proposed methods with the traditional models. For each query, we first calculate the Average Precision (AP) score by averaging all the 20 participants' AP scores. Then, we average all the queries' AP scores to obtain MAP for our proposed model. The results are shown in Table 1, under the manually selected optimal λ. The statistical significance is indicated by * ($p < 0.05$, one-way ANOVA) and ** ($p < 0.01$). The results demonstrate that using the words

Methods	MAP	%↑ over LM	%↑ over RM3
LM	0.1734	-	-
RM3	0.2773	59.91%	-
Eye-QE	0.3411	96.71%**	23.01%*
Eye-LDA-QE	0.3508	102.3%**	26.51%*

Table 1: MAP of different methods

of interest captured by eye trackers is feasible and improves search performance dramatically. Further improvement is achieved by combining eye tracking and LDA model, thus verifying that employing LDA model to enrich WoIs can help dig out user's latent search intention to a large extent.

To investigate the impact of WoIs captured by eye-tracking on query expansion, we test different settings of the parameter λ. The dynamic MAP scores along with changing λ are shown in Figure 1. We can see that, when λ is small (meaning more weight is given to the original query), $RM3$ is better than our proposed model. The possible reason is that smaller λ limits our captured WoIs to take an effect.
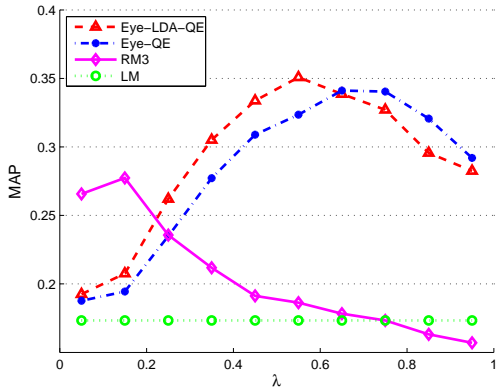
Figure 1: MAP scores along with changing λ

However, with the increase of λ, our models *Eye-QE* and *Eye-LDA-QE* become to outperform *LM* and *RM3*. Both our proposed models achieve the best performance when λ is larger, implying that the captured WoIs contain more resultful information than the original query.

As Granka et al. [4] pointed out, the majority of searchers usually examine the documents that are ranked higher, especially the top two documents. It is essential to rank relevant documents higher in order to improve the users' search experience. Therefore, in addition to MAP, we also averaged the 20 participants' *Precision at K* scores to as another performance indicator. The results (the best performing runs under different λ) are shown in Figure 2. We can find that our proposed models, especially the *Eye-LDA-QE* model, perform much better than the other ones when K is smaller. By taking advantage of latent topic derivation through LDA, we can more precisely locate user's hidden search intention and improve search experience.
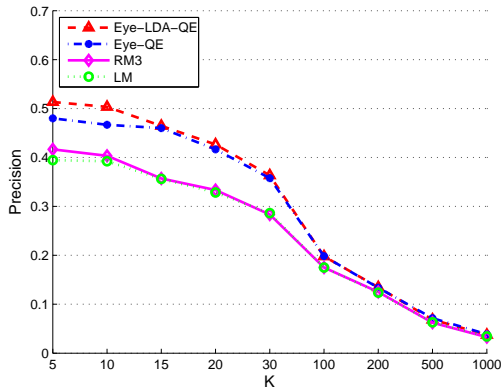

Figure 2: Precision at K for all methods

4. CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed to expand the original query using the words of interest captured by eye tracking in real time. Moreover, considering the words captured by eye tracker are limited due to the insufficient feedback data, the user's latent search intention may not be fully ex-

pressed. Therefore, we also train a latent topic model using LDA and then rank the topics, based on which the generated words can be used to further enrich the expansion model. The experimental results have demonstrated that capturing the words of interest based on eye-tracking can considerably improve the retrieval performance. Further, combining the eye-tracked words and the words in topic models can generate further improvement.

In the future, we will test our models on a larger-scale by recruiting more participants on larger datasets. We can further take into account the user's reading sequence or the reading pattern in a feedback-based query expansion model. Moreover, given that the number of mobile searchers is growing rapidly, it will be interesting to explore if and how we could use cameras on mobile devices to track user's gaze to better capture user's preferences.

5. ACKNOWLEDGEMENT

6. REFERENCES

[1] A. Ajanki, D. R. Hardoon, S. Kaski, K. Puolamäki, and J. Shawe-Taylor. Can eyes reveal interest? Implicit queries from gaze patterns. *User Modeling and User-Adapted Interaction*, 19(4):307–339, 2009.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *JMLR*, 3:993–1022, 2003.

[3] G. Buscher, A. Dengel, and L. van Elst. Query expansion using gaze-based feedback on the subdocument level. In *ACM SIGIR*, pages 387–394. ACM, 2008.

[4] L. A. Granka, T. Joachims, and G. Gay. Eye-tracking analysis of user behavior in www search. In *ACM SIGIR*, pages 478–479. ACM, 2004.

[5] J. Gwizdka. Characterizing relevance with eye-tracking measures. In *Proceedings of the 5th Information Interaction in Context Symposium*, pages 58–67. ACM, 2014.

[6] D. Kelly and J. Teevan. Implicit feedback for inferring user preference: a bibliography. In *ACM SIGIR Forum*, volume 37, pages 18–28. ACM, 2003.

[7] J. Y. Kim, M. Cramer, J. Teevan, and D. Lagun. Understanding how people interact with web search results that change in real-time using implicit feedback. In *ACM CIKM*, pages 2321–2326. ACM, 2013.

[8] V. Lavrenko and W. B. Croft. Relevance based language models. In *ACM SIGIR*, pages 120–127. ACM, 2001.

[9] P. Olsson. Real-time and offline filters for eye tracking. 2007.

[10] E. D. Reichle, A. Pollatsek, and K. Rayner. E–z reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, 7(1):4–22, 2006.

[11] S. R. Singh, H. A. Murthy, and T. A. Gonsalves. Inference based query expansion using user's real time implicit feedback. In *Knowledge Discovery, Knowledge Engineering and Knowledge Management*, pages 158–172. Springer, 2013.

[12] R. W. White and G. Marchionini. Examining the effectiveness of real-time query expansion. *Information Processing & Management*, 43(3):685–704, 2007.

[13] C. Zhai and J. Lafferty. Model-based feedback in the language modeling approach to information retrieval. In *ACM CIKM*, pages 403–410. ACM, 2001.