Are Topics Interesting or Not? An LDA-based Topic-graph Probabilistic Model for Web Search Personalization

51

JIASHU ZHAO, Department of Physic and Computer Science, Wilfrid Laurier University, Canada
JIMMY XIANGJI HUANG, Information Retrieval and Knowledge Management Lab,
York University, Canada
HONGBO DENG, Alibaba Inc, China
YI CHANG, School of Artificial Intelligence, Jilin University, China
LONG XIA, Information Retrieval & Knowledge Management Research Lab, York University, Canada

In this article, we propose a Latent Dirichlet Allocation– (LDA) based topic-graph probabilistic personalization model for Web search. This model represents a user graph in a latent topic graph and simultaneously estimates the probabilities that the user is interested in the topics, as well as the probabilities that the user is not interested in the topics. For a given query issued by the user, the webpages that have higher relevancy to the interested topics are promoted, and the webpages more relevant to the non-interesting topics are penalized. In particular, we simulate a user's search intent by building two profiles: A positive user profile for the probabilities of the user is interested in the topics and a corresponding negative user profile for the probabilities of being not interested in the the topics. The profiles are estimated based on the user's search logs. A clicked webpage is assumed to include interesting topics. A skipped (viewed but not clicked) webpage is assumed to cover some non-interesting topics to the user. Such estimations are performed in the latent topic space generated by LDA. Moreover, a new approach is proposed to estimate the correlation between a given query and the user's search history so as to determine how much personalization should be considered for the query. We compare our proposed models with several strong baselines including state-of-the-art personalization approaches. Experiments conducted on a large-scale real user search log collection illustrate the effectiveness of the proposed models.

CCS Concepts: • **Information systems → Personalization;**

Additional Key Words and Phrases: Personalization, probabilistic model, Web search, Latent Dirichlet Allocation (LDA), topic-graph

This research is supported by the research grants (RGPIN-2020-05588 and RGPIN-2020-07157) from Natural Sciences and Engineering Research Council (NSERC) of Canada, York Research Chairs (YRC) program and Wilfrid Laurier University start-up program.
Authors' addresses: J. Zhao (corresponding author), Department of Physic and Computer Science, Wilfrid Laurier University, Waterloo, Canada; email: jzhao@wlu.ca; J. X. Huang (corresponding author) and L. Xia, Information Retrieval and Knowledge Management Lab, York University, Toronto, Canada; emails: jhuang@yorku.ca, long.phil.xia@gmail.com; H. Deng, Alibaba Inc, Hangzhou, China; email: hbdeng@acm.org; Y. Chang, School of Artificial Intelligence, Jilin University, Jilin, China; email: yichang@jlu.edu.cn.

1046-8188/2021/12-ART51 $15.00
https://doi.org/10.1145/3476106

ACM Reference format:
Jiashu Zhao, Jimmy Xiangji Huang, Hongbo Deng, Yi Chang, and Long Xia. 2021. Are Topics Interesting or Not? An LDA-based Topic-graph Probabilistic Model for Web Search Personalization. *ACM Trans. Inf. Syst.* 40, 3, Article 51 (December 2021), 24 pages.
https://doi.org/10.1145/3476106

1 INTRODUCTION AND MOTIVATION

With the rapid growth of Internet technologies, more and more people rely on the Internet for searching [36], networking [24], shopping [21, 45], and so on. According to the July 2020 Internet usage statistics,[1] there are over 4.57 billion Internet users all over the world. Meanwhile, these users have various information needs, which means different users require different information through the Internet. Therefore, there is a high demand for personalization that aims to provide satisfied results for individual users [3, 10]. In this article, we focus on the problem of personalization in **Information Retrieval (IR)**. Given an input query, a traditional search engine returns a ranked list of documents for all users. For example, given a query "java," a list of webpages returned by a search engine is shown in Table 1.

Different users expect to locate different documents. Some people may be looking for the programming language "java" and others may be interested in the "java" island. Providing the identical search results is not able to satisfy all the users. A document is relevant or non-relevant to a query should be dependent on the users. It could be relevant to a query for some users, but non-relevant to the same query for the other users. In Table 1's example, the users who are interested in the java island need to go through the top ranked but non-interesting documents before reaching the relevant document. The purpose of this article is to model the personalized relevancy between a document and a query for a given user.

Although the personalized approaches are known to be effective in many IR applications, how to represent a user's search history and how to integrate the user information into the query-document matching remain largely unexplored. In fact, how to properly represent a "user" is very critical, since it is impossible to track all users' real information and intentions at the time the users are inputing the queries. Query logs are usually used in analyzing the users for personalization [65, 72]. Query logs store the users' search histories, including the users' queries, the users' viewed and clicked behaviours. And the query logs are feasible to be collected by the search engine companies. In this article, we analyze the query logs and build user profiles based on the users' search histories.

Researchers have investigated matching a user (via query logs), a query issued by the user, and a given document with different approaches [35, 39, 47, 56]. One group of these approaches is to match term by term, where the user is represented as a list of the user's possibly interested terms [39]. However, the term space could be very sparse, which results in the overfitting issues for the personalization approaches. So some other techniques are proposed and try to deal with these issues. Another group of the approaches is to match the user, the query and the document through topics, where the users' interests are represented by topics. The ambiguous matching provides a more flexible way for search personalization. The document categories are commonly adopted as topics. Sontag et al. [56] and Liu et al. [35] used the pre-classified google news topic categories (e.g., "world," "sport," "entertainment"). Among these approaches, the users' interested topics are modelled and further matched with the queries and the documents' topics.

In this article, we model both the users' interested topics and their not-interested topics. The users and topics can be considered as a graph [7] as shown in Figure 1. A user is connected with

[1] https://www.statista.com/statistics/617136/digital-population-worldwide/.

Table 1. A Ranking List Example

Ranking	Webpage Content
No. 1	**Java** - Official Site: Get the latest **Java** Software and explore how **Java** technology provides a better ...
No. 2	Download free **java** software
No. 3	**java** SE Downloads | Oracle Technology Network | Oracle
No. 4	Welcome | **java**.net: **java**.net is the source for **Java** Technology Collaboration
No. 5	**Java** - Wikipedia, the free encyclopedia: **Java** (Indonesian: Jawa) is an island of Indonesia. ...
No. 6	**Java** (programming language) - Wikipedia, the free encyclopedia: **Java** is a computer programming language that is concurrent, class-based, object-oriented, and specifically designed to ...

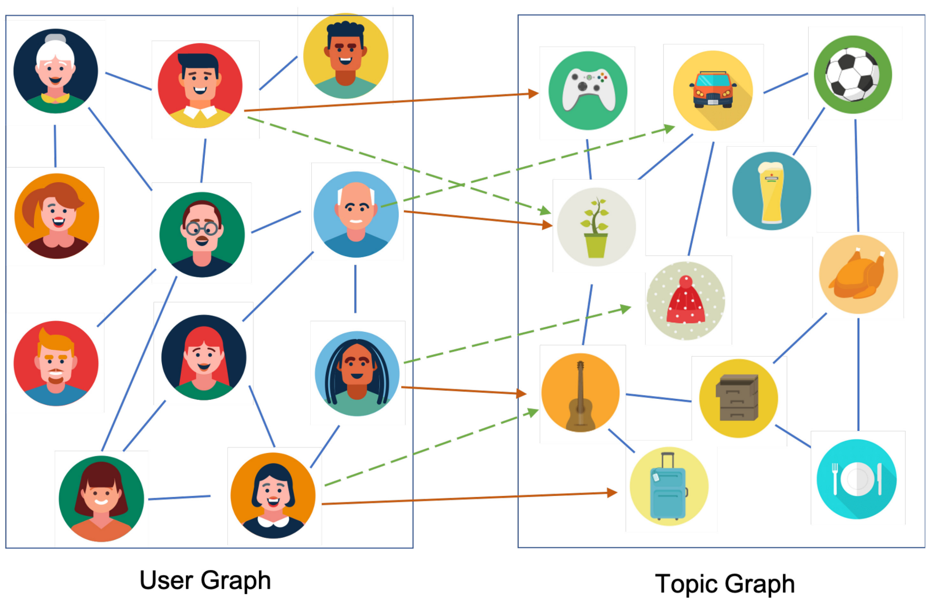


Fig. 1. The user and topic graph. The red solid lines represent the users' interested topics, and the green dashed lines represent the users' non-interested topics.

his/her similar users, who are more likely to share the same search interests. As the user connections are not explicit, we can implicitly represent such connections by the topics [66]. The users are all connected with their interested/non-interested topics. We aim to build a positive profile and a negative profile for a given user. Since the behaviours on the Internet only show a small portion of their real intentions. It would be more precise if we can use as much information as possible when modelling the user. In our earlier example shown in Table 1, a user might be looking for "java island" when the input query is "java". Such intentions would be expressed by the user's clicking on the fifth document and skipped the rest of the documents. In this case, the user is more likely to be interested in the topic "java island" and not interested in "java programming language".

In more details, we represent and match the users, queries, and documents in both the latent topic space and the word space. The topic probabilities provide an explicit representation of a document, where each document is associated with some latent topics and the topics are associated with some words. We further represent the users by the latent topics. In particular, we use **Latent Dirichlet Allocation (LDA)** [4], which is a generative model that extracts the hidden topic information from a document collection, and the topic distribution is assumed to have a Dirichlet prior.

The major contributions of this article are as follows:

(1) We propose a personalized probabilistic framework that explicitly study the relevancy of a document, a query and a user. It directly builds the users' positive and negative profiles based on the users' interested topics and the users' non-interesting topics from query logs.
   (a) The proposed model estimates the document-query matching for the user in the latent topic space based on LDA, and also considers how much we should trust the personalization process by modelling the relevancy of a query to a user in the word space. To the best of our knowledge, this is the first work that directly estimates the relevancy of a query to a user, and first query-document-user matching in both a latent topic space as well as a term space.
   (b) In addition, we further propose two strategies to build the users' negative profiles, the subtraction strategy and the orthogonal projection strategy. The intuition is that the user may skip a document if it shares some similar topics with the clicked documents. Topics in such skipped documents should not be considered in the user's negative profile.

(2) We conduct experiments on the real personalized web search query-log dataset to show that our proposed framework consistently outperforms baseline methods.

The organization of this article is as follows. In Section 2, we discuss prior related work. In Section 3, we present a personalized probabilistic model that matches the user, a query and a document in both the latent topic and the word space. Two strategies are further proposed to process the users' profiles in Section 4. The experimental settings are introduced in Section 5. In Section 6, we conduct experiments to illustrate the proposed model on a commercial search engine corpus. In Section 6, we conclude our findings and discuss several possible future directions.

2 RELATED WORK

In Information Retrieval, probabilistic models [20, 25, 30, 31, 48] have been investigated for years. Though there are many variations of probabilistic information retrieval models, a basic assumption is that the terms distribute differently in the relevant documents and the irrelevant documents [73, 75]. The probabilistic models usually aim to rank documents in a way that the probability of the query terms' relevancy to a document is maximized and the probability of the query terms' irrelevancy to a document is minimized. In this article, we start from the basic ideas of the probabilistic models, and propose to adapt such ideas in the scope of personalization.

Personalized approaches are recognized to have improved retrieval performance in many IR applications [8, 13, 16, 32, 37, 38]. Most of these approaches tend to represent users with simplified user profiles (often based on historic interests), enabling the efficient calculation of personalized ranked lists [57]. Previous work includes the users editing their profiles manually, or to represent a user profile in terms of manual social tags [62]. However, different from personalization in other areas [74], using the users' real-world profile was found to be very challenging in IR [76]. Therefore the implicit user profiles [42, 61] are largely investigated, such as building a user ontology [60], a personalized weight assignment vector [28]. Some previous approaches consider users' short-term interests [11, 23, 70], while some focus long-term interests [59], or both long-term and short-term [9]. Our method will work on all of the available user behaviors, and thus focus on long-term interests. Five personalized search strategies (including two click-based and three profile-based ones) are evaluated in Reference [17], and the best strategy assumes the web pages frequently clicked by a user in the past are more relevant to the user than those seldom clicked. This article is inspired by the user historical behaviors, and aim to automatically extract the latent user profiles learned from the users' click graph.

In a click graph, clicked documents are widely utilized to characterize the users' search interests [2, 5]. Moreover, skipped documents can provide additional information to the clicked ones, and have also been studied in some web search tasks [18, 33, 53, 55]. However, it remains a challenge on how to better utilize such clicked and skipped information in personalized search. Previous work has incorporated a document's skip information into learning to interpret user behaviors as the document's features [1], or as the target (1 for click and −1 or skip) in learning to rank models [41, 68, 71]. In this article, we directly model both the clicked and skipped documents in the user profiles. We assume that for a given query the clicked documents reflect a user's interests and the skipped documents contain information that the user is not interested in. Then, we explicitly build probabilistic user positive profiles with the clicked documents, and negative profiles with the skipped documents.

To summarize the users' click graph into user profiles, users are often represented as categorized topics [56], or latent topics [54]. The limitation of using categorized topics for personalization is that many documents may not contain the topics covered in the ontology, and the human-generated topics require expensive manual effort to determine the correct categories for each document [64]. However, latent topics are learned from document contents, such as word embedding and topic modeling [12], providing a natural way to reduce the dimension of the user related documents into summarized user profiles. Recently, deep learning approaches are proposed to learn the vector representations of words and documents. For example, BERT [14] is pre-trained on predicting the masked words in the sentence, and then fine-tuned on a specific task such as query-answering. The item representation can be learned by a multi-layer RGCN with the encode-decode paradigm by minimizing the reconstruction loss [52]. In this article, the user profiles are built in an unsupervised manner. Deep generative models have been proposed based on the Variational Autoencoder [34, 46], Generative adversarial networks [22], Attention-based Aspect Extraction [26], and so on. LDA is a topic modeling approach with a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics [4, 6, 27, 50]. LDA has been employed in personalized search to extract latent topics for building user profiles [15, 44, 51, 64]. Although such approaches enable the efficient calculation of personalized ranked lists, they did not make full use of all the information. In this article, we incorporate LDA to generate latent topic representations, and investigate the effect of considering both the user interested and not-interested contents in IR personalization.

3 AN LDA-BASED PERSONALIZED MODEL

When a user issues a query to a search engine, (s)he may click some top ranked documents and skip the others. In this article, we assume that the clicked documents contain the interesting topics to the user's information need, and the non-clicked documents cover some non-interesting topics to the user's information need.

Given a user's query, we aim to locate the documents that have a higher probability of being relevant to the user's query, denoted as $P(R = 1|d, q, U)$, as well as a lower probability of being non-relevant to the user's query, $P(R = 0|d, q, U)$, where U is the user, q is the query, and d is a given document. A personalized ranking score can be defined as

$$\begin{aligned} Score(d|q,U) &= \frac{P(R=1|d,q,U)}{P(R=0|d,q,U)} \\ &= \frac{P(d|R=1,q,U)}{P(d|R=0,q,U)} \cdot \frac{P(R=1|q,U)}{P(R=0|q,U)} \\ &= f(d,q,U) \cdot g(q,U), \end{aligned} \tag{1}$$

where $P(d|R=1,q,U)$ and $P(d|R=0,q,U)$ are the probabilities of retrieving a relevant or non-relevant document d for U and q. $P(R=1|q,U)$ and $P(R=0|q,U)$ are the previous probabilities of retrieving a relevant or non-relevant document for U and q correspondingly. In the rest of the article, we denote $\frac{P(d|R=1,q,U)}{P(d|R=0,q,U)}$ as $f(d,q,U)$, and $\frac{P(R=1|q,U)}{P(R=0|q,U)}$ as $g(q,U)$ for simplicity. We can see that $f(d,q,U)$ matches a document and a user's query, and $g(q,U)$ indicates the relevancy between a query and a user. In Sections 3.1 and 3.2, we will propose approaches to estimate $f(d,q,U)$ and $g(q,U)$ correspondingly. In Section 3.3, the estimation of $Score(d|q,U)$ is combined with the non-personalized weighting model.

3.1 Document-Query Matching for the User

In this article, we match the document and the user's interests in the latent topic space to estimate $f(d,q,U)$. We first build LDA topics for all the documents, which generate the topic distributions on each document $P(T_k|d)$ and the word distributions for each topic $P(w|T_k)$, where T_k represent a latent topic. The details about the LDA model are described in Section 3.1.2.

The probability of retrieving a relevant document $P(d|R=1,q,U)$ can be estimated in the topic space by

$$\begin{aligned} P(d|R=1,q,U) &\propto \sum_k P(d|T_k)P(T_k|R=1,q,U) \\ &= \sum_k \frac{P(d)}{P(T_k)} P(T_k|d)P(T_k|R=1,q,U), \end{aligned}$$

where $P(d)$ is the probability of generating a document, $P(T_k)$ is the probability of generating a topic, and $P(T_k|R=1,q,U)$ is the probability of generating a topic from the contents that relevant to the user's query. We assume $P(d)$ and $P(T_k)$ are uniformly distributed in the collection. Thus $\frac{P(d)}{P(T_k)}$ is a constant, denoted as C_1. Then $P(d|R=1,q,U)$ could be obtained by

$$P(d|R=1,q,U) \propto C_1 \cdot \sum_k P(T_k|d)P(T_k|R=1,q,U).$$

Similarly, the probability to retrieve a non-relevant document is

$$P(d|R=0,q,U) \propto C_1 \cdot \sum_k P(T_k|d)P(T_k|R=0,q,U),$$

where $P(T_k|R=0,q,U)$ is the probability of generating a topic from the contents that are non-relevant to the user's query.

Therefore, $f(d,q,U)$ can be estimated as

$$f(d,q,U) = \frac{\sum_k P(T_k|d)P(T_k|R=1,q,U)}{\sum_k P(T_k|d)P(T_k|R=0,q,U)}. \tag{2}$$

$P(T_k|d)$ is obtained from the LDA model. Then we need to further analyze the probabilities $P(T_k|R=1,q,U)$ and $P(T_k|R=0,q,U)$. The probability could be estimated via bayesian approach as the following

$$\begin{aligned} P(T_k|R=1,q,U) &= \frac{P(T_k|R=1,U)P(q|T_k,R=1,U)}{\sum_j P(T_j|R=1,U)P(q|T_j,R=1,U)} \\ &= \frac{P(T_k|R=1,U)P(q|T_k)}{\sum_j P(T_j|R=1,U)P(q|T_j)}. \end{aligned} \tag{3}$$

Here it is assumed that the probability of generating a query from a topic is independent from the user, thus $P(q|T_k,R=1,U)=P(q|T_k,R=0,U)=P(q|T_k)$. $P(q|T_k)$ could be estimated by a unigram or n-gram language model from $P(w|T_k)$, where w is a word in query q. $P(T_k|R=1,U)$ is the probability of generating a topic from the contents relevant to the user. Similiarly, we can derive $P(T_k|R=0,q,U)$ as

$$P(T_k|R=0,q,U) = \frac{P(T_k|R=0,U)P(q|T_k)}{\sum_j P(T_j|R=0,U)P(q|T_j)}, \tag{4}$$

where $P(T_k|R=0,U)$ is the probability of generating a topic from the contents non-relevant to the user. Then we apply Equation (3) and (4) in Equation (2) for estimating $f(d,q,U)$. Both $P(T_k|R=1,U)$ and $P(T_k|R=0,U)$ depend on the user, and independent from the current query and the current document. Therefore, $P(T_k|R=1,U)$ and $P(T_k|R=0,U)$ are the user's positive and negative profiles that we intend to build from the user's previous query logs. In Section 3.1.1, we discuss how to build these user profiles from the user's search log. Then, in Section 3.1.2, we briefly introduce the core ideas of LDA.

3.1.1 Building the User Profiles. Our intuition is that a user only clicks on the documents relevant to the user and ignores the documents that are not relevant to the user, though these ignored documents could be ranked high. This is because the information needs for various users are different and the generalized search tends to find documents that can satisfy all users.

We build a positive user profile $\{P(T_k|R=1,U)\}_{k=1}^K$, which has positively effects toward the user's search interests, as well as a negative user profile $\{P(T_k|R=0,U)\}_{k=1}^K$, which reflects the topics that the user does not like to click. K is the dimension of the topic space. Both profiles are in the form of a vector of length K and the values are from the user's previous queries. We denote the user's previous queries as $\{r_1,\ldots,r_l,\ldots,r_L\}$. The positive user profile on the kth topic is

$$P(T_k|R=1,U) \propto \sum_l P(T_k|R=1,r_l,U)\cdot P(r_l|U), \tag{5}$$

where $P(T_k|R=1,r_l,U)$ is the probability of generating a topic from the contents that are relevant to a user's previous query r_l, and $P(r_l|U)$ is the probability of the user issues the query. We estimate $P(r_l|U)$ as the number of times that the user issues the query r_l divided by the total number of queries issued by the user,

$$P(r_l|U) = \frac{1}{L}\sum_{i=1}^{L} \mathbb{1}_{r_i=r_l}.$$

Correspondingly, the negative user profile is

$$P(T_k|R=0,U) \propto \sum_l P(T_k|R=0,r_l,U)\cdot P(r_l|U), \tag{6}$$

where $P(T_k|R=0,r_l,U)$ is the probability of generating a topic from the contents that are non-relevant to a user's previous query r_l.

The clicking event on documents for a previous query shows that the user has found the interested topics. We assume that the skipped documents contain some topics that the user is not

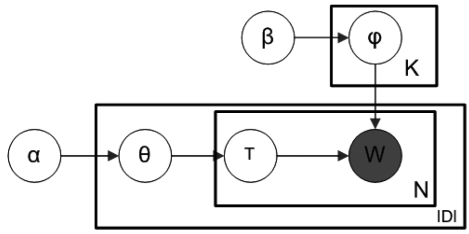


Fig. 2. Plate notation for the LDA model.

interested in. For a query r_l, a list of documents were given $\{ (d_{l,1}, c_{l,1}), \ldots, (d_{l,m_i}, c_{l,m_k}) \}$, where $c_{l,j} \in \{0,1\}$ represents whether U clicked $d_{l,k}$ ($c_{l,k} = 1$) or skipped $d_{l,k}$ ($c_{l,k} = 0$). We estimate $P(T_k|R = 1, r_l, U)$ in Equation (5) and $P(T_k|R = 0, r_l, U)$ in Equation (6) by the average of document topics over the clicked documents or the skipped documents, respectively,

$$P(T_j|R = 1, r_l, U) = \frac{\sum_k P(T_j|d_{l,k}) \cdot \mathbb{1}_{c_{l,k}=1}}{\sum_k \mathbb{1}_{c_{l,k}=1}}, \tag{7}$$

$$P(T_j|R = 0, r_l, U) = \frac{\sum_k P(T_j|d_{l,k}) \cdot \mathbb{1}_{c_{l,k}=0}}{\sum_k \mathbb{1}_{c_{l,k}=0}}. \tag{8}$$

$P(T_j|d_{l,k})$ can be obtained from the LDA model. By applying the above formula in Equations (5) and (6), we build the user profiles from the previous queries and the corresponding click events. For the same user, there might be non-zero probabilities for the same topic in both the positive profile and the negative profile. In this approach, we expect that there is a significant difference between these two profiles in those topics that the user has a strong preferences. If the user is very interested in a topic, then the probability of this topic in the positive profile will be much larger than the negative profile

3.1.2 Latent Dirichlet Allocation. In this article, the latent topics are built by the LDA model [4], which assumes the following generative process for each document d in a document collection D:

- Choose a multinomial distribution Φ_T for each topic T from a Dirichlet distribution with hyperparameter β. β is the parameter of the uniform Dirichlet prior on the per-topic word distribution.
- Choose a multinomial distribution θ_d for each topic d from a Dirichlet distribution with hyperparameter α. α is the parameter of the uniform Dirichlet prior on the per-document topic distributions.
- For each word w in document d, choose a topic $T_k \sim Multinomial(\theta_d)$ and choose the word w from the multinomial distribution of φ_{T_k}.

Thus, the probability of generating the collection D is given as following:

$$\begin{aligned} &P(d_1, \ldots, d_{|D|}|\alpha, \beta) \\ &= \iint \prod_{T=1}^{K} P(\varphi_T|\beta) \prod_{d=1}^{|D|} P(\theta_d|\alpha) \left(\prod_{i=1}^{N_d} \sum_{T_i=1}^{K} P(w_i|T, \varphi) \right) d\theta d\varphi. \end{aligned}$$

where $|D|$ is the number of documents in dataset D, N_d is the number of words in document d, K is the number of topics in the LDA model. Figure 2 depicts the plate notation for the LDA model, which can capture the dependencies among all the variables. α and β can be estimated via the Expectation maximization algorithm or Gibbs sampling [67]. In this article, we use online LDA with

Gibbs sampling for parameter estimation due to the large size of the dataset. The following probabilities can be generated from LDA: $P(T|d)$, $P(w|T)$, and $P(T)$. In particular, the corresponding probabilities are obtained using vowpal_wabbit[2] package.

3.2 Relevancy of a Query to the User

In a traditional binary model, the relevancy between a query and a user $g(q,U)$ is usually ignored, since it does not contain the document information. In this article, we argue that if a query is more relevant to the user, then we should emphasis more on personalization for matching this query with the document. However, if this query is less relevant to the user, then the user profile does not benefit much while ranking the documents. This component will judge whether a query should be return more personalized results or non-personalized results. To estimate $g(q,U)$, we match query q with the user's search logs. First, $g(q,U)$ can be written as

$$\begin{aligned} g(q,U) &= \frac{P(q|R=1,U)}{P(q|R=0,U)} \cdot \frac{P(R=1,U)}{P(R=0,U)} \\ &= C_U \cdot \frac{P(q|R=1,U)}{P(q|R=0,U)}, \end{aligned}$$

where $\frac{P(R=1,U)}{P(R=0,U)}$ is constant for a given user, denoted as C_U. $P(q|R=1,U)$ and $P(q|R=0,U)$ are the probabilities of generating the query from the contents that are relevant or non-relevant to the user, respectively. We build a unigram model of q on all the previous clicked documents to estimate $P(q|R=1,U)$ as follows:

$$\begin{aligned} P(q|R=1,U) &= \Pi_{w\in q} P(w|U\ clicked\ documents) \\ &= \Pi_{w\in q} \frac{\sum_{l,k} count(w,d_{l,k}) \cdot \mathbb{1}_{c_{l,k}=1}}{\sum_{l,k} |d_{l,k}| \cdot \mathbb{1}_{c_{l,k}=1}} \\ &\propto \frac{\sum_{l,k} count(w,d_{l,k}) \cdot \mathbb{1}_{c_{l,k}=1} + \mu \cdot P(w|Co)}{\sum_{l,k} |d_{l,k}| \cdot \mathbb{1}_{c_{l,k}=1} + \mu}, \end{aligned}$$

where $count(w,d_{l,k})$ is the number of w's occurrences in $d_{l,k}$ and $|d_{l,k}|$ is the length of document $d_{l,k}$. If one of the query terms w does not appear in the user's previous clicked documents, then $P(q|R=1,U)$ equals zero. Therefore, we use Dirichlet smoothing approach to assign a non-zero probability in the third step of the above formula, where $P(w|Co)$ is the probability of seeing term w over the whole collection and μ is the smoothing parameter in Dirichlet approach. We also build a smoothed unigram model of q on all the previous skipped documents to estimate $P(q|R=0,U)$,

$$P(q|R=0,U) \propto \frac{\sum_{l,k} count(w,d_{l,k}) \cdot \mathbb{1}_{c_{l,k}=0} + \mu \cdot P(w|Co)}{\sum_{l,k} |d_{l,k}| \cdot \mathbb{1}_{c_{l,k}=0} + \mu}.$$

Then we can get an estimation of $g(q,U)$ based on the user's searching history.

3.3 A Linearly Integrated Personalization Model with Non-linear Normalization

A linear combination is often applied to integrate the personalized score with the non-personalized score [40]. In this section, we linearly combine the personalized model $f(d,q,U) \cdot g(q,U) = \frac{P(R=1|d,q,U)}{P(R=0|d,q,U)}$ with the **original ranking score (ORS)**. The ORS is the non-personalized ranking score used in the search engine. In this article, we do not focus much on the non-personalized score. So we regard the non-personalized weighting as a black box that has already been optimized and

[2] https://github.com/JohnLangford/vowpal_wabbit/wiki/Latent-Dirichlet-Allocation.

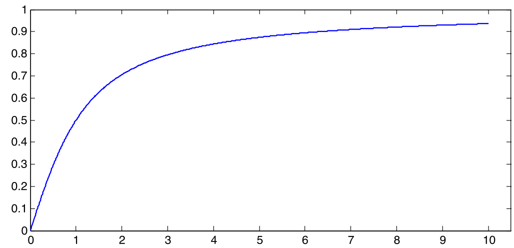


Fig. 3. The inverse trigonometric function for normalization.

use ORS directly in our model. A **linearly integrated personalization (LLP)** model is shown as follows,

$$LLP(d, q, U) = (1 - \lambda) \cdot h(ORS(d, q)) + \lambda \cdot h(f(d, q, U) \cdot g(q, U)), \tag{9}$$

where λ is the balancing parameter ranging from 0 to 1. To have the same schema, both $ORS(d, q)$ and $f(d, q, U) \cdot g(q, U)$ are normalized to be between 0 and 1 by a normalization function $h(\cdot)$. Since $g(q, U)$ is not related to the document d, a linear normalization approach would make $g(q, U)$ ineffective. Thereby, we adopt inverse trigonometric function, which has been applied as a normalization approach in Neural Network,

$$h(x) = arctan(x) \cdot \frac{2}{\pi},$$

where $x > 0$ is an input real number, $arctan(x)$ is the inverse trigonometric function, and $h(x)$ is our non-linear normalization function. Figure 3 shows the shape of $h(x)$. We can see that the domain of $h(x)$ is $[0, +\infty)$, and the corresponding range of $h(x)$ is $[0, 1)$.

This non-linear function $h(\cdot)$ balances the user's interestingness and non-interestingness. $f(d, q, U)$ and $g(q, U)$ are both ratios of probabilities, and therefore range in $(0, \infty)$. When both $f(d, q, U)$ and $g(q, U)$ equal 1, the probability o_ the document relevant to the user and query is the same as non-relevant. In this case, the personalization component should be neutral, and the corresponding normalized value should be 0.5. We can see that $h(1) = 0.5$, which is consistent with such heuristics. The more interesting documents have scores in the range of $(0.5, 1)$, and the less interesting documents have scores in the range of $(0, 0.5)$.

4 TWO STRATEGIES FOR BUILDING USER'S NEGATIVE PROFILE

Given a query and a document list presented to the user, there are various reasons that a user skips a document rather than the skipped document is not interesting to the user. For instance, a document d_s contains the topics that are interesting to the user. But these topics are covered by the another document d_c. Then the user might skip d_s and click on d_c. For another example, the document d_s outlines a few topics where one topic is interesting but the other topics are not so interesting. The user might skip d_s and click on a document that describes the interesting topic more precisely. Since we are studying the users in the latent topic space, we are able to distinguish these documents in terms of their topic distributions.

Therefore, we need to further process the user's negative profile, in case we broadly regard all topics in the skipped documents as "not-interesting." Intuitively, for a given query, if the skipped documents share some similar topics with the clicked documents, then these topics should not be considered in the user's negative topic profile. Given a query issued by the user, the interesting topics and not-interesting topics are represented as two vectors in the latent topic space, see Equations (7) and (8). These two vectors are manipulated by the strategies in this Section. For query r_l, we denote the interesting topic vector $P(T_j|R = 1, r_l, U)$ obtained from Equation (7) as $A = (a_1, a_2, \ldots, a_t)$, and denote the not-interesting topic vector $P(T_j|R = 0, r_l, U)$ from Equation (8) as $B = (b_1, b_2, \ldots, b_t)$, where t is the number of latent topics. t is determined by the *LDA* model parameter settings, which will be discussed in Section 5.3. Suppose the new non-interesting topic vector is $B' = (b'_1, b'_2, \ldots, b'_t)$. Here we present two strategies to eliminate the not-interesting topics that are covered by the interesting topics.

- Subtraction Strategy One strategy is to directly use the difference between vector A and B. Meanwhile, b'_i should be non-negative, since b'_i represents how much the user is not interested in the ith topic,
$$b'_i = max\{b_i - a_i, 0\}.$$
- Orthogonal Projection Strategy In this strategy, we use orthogonal decomposition to define the B'. We first find the hyperplane that is perpendicular to the vector A, and then project the vector B onto this hyperplane,
$$b'_i = max\left\{b_i - \frac{\sum_{i=1}^{t} a_i b_i}{\sum_{i=1}^{t} a_i^2} \cdot a_i, 0\right\}.$$

Then we use B' as the new non-interesting topic vector, and the interesting topic vector A remains the same. Each of these two strategies has its advantages. Subtraction strategy processes the profile vectors in a heuristic way. It focuses more on the values of the vectors. A negative profile with a small value would easily be fully covered by a positive profile with a large value. Under orthogonal projection strategy, the negative profile vector would have a non-zero projection as long as there is a non-zero angle between the positive and negative profile vectors. These two strategies are applied to the proposed LLP model. The corresponding new models are denoted as LLP-Subtraction and LLP-Projection.

5 IMPLEMENTATION AND EXPERIMENTS

Our experiments are conducted on a large-scale set of real user search logs from Yahoo! search. The corpus contains user query logs for a three months period, from August 1 to October 31, 2013. The users' search activities are collected, including the B-cookie,[3] query, timestamp, top 10 returned URLs and the corresponding user clicks. We focus on the users with a good number of queries for personalization. So we filter users by the number of issued queries with click event during these three months period. The queries are user issued full queries. It is recognized that too frequent B-cookies are more likely to be non-personal user, and personalized modeling on those users will not be reliable. Since the model requires user history behavior information to build profiles, we select users whose click frequency range between 300 and 5,000. Then we randomly select 100 frequent users, which include 76,640 queries with clicks. We extract all related URLs (top 10 returned URLs for these queries) as the document collection, which contains 551,358 records.

Since we aim to investigate the contents of the webpages, the URLs are crawled. We first extract the title, main text, and meta data from the webpages. Then in every field, each term is stemmed using Porter's English stemmer, and standard English stopwords are removed. An LDA model is built based on the processed documents.

To evaluate the generalized performance of our proposed models, the collection is splitted into a training set, a validation set and a testing set. The clicked documents are regard as ground truth. If more than 1 queries are clicked, then they are all regarded as ground truth. The statistics of the collection is shown in Table 2. The time format is Month/Date/Year. The training set contains two

[3]B-cookie is adopted as anonymous user id.

Table 2. Collection Statistics

	Duration	No. of queries
Training Set	08/01/2013–09/31/2013	53,923
Validation Set	10/01/2013–10/15/2013	12,023
Testing Set	10/16/2013–10/31/2013	10,694
Total	08/01/2013–10/31/2013	76,640

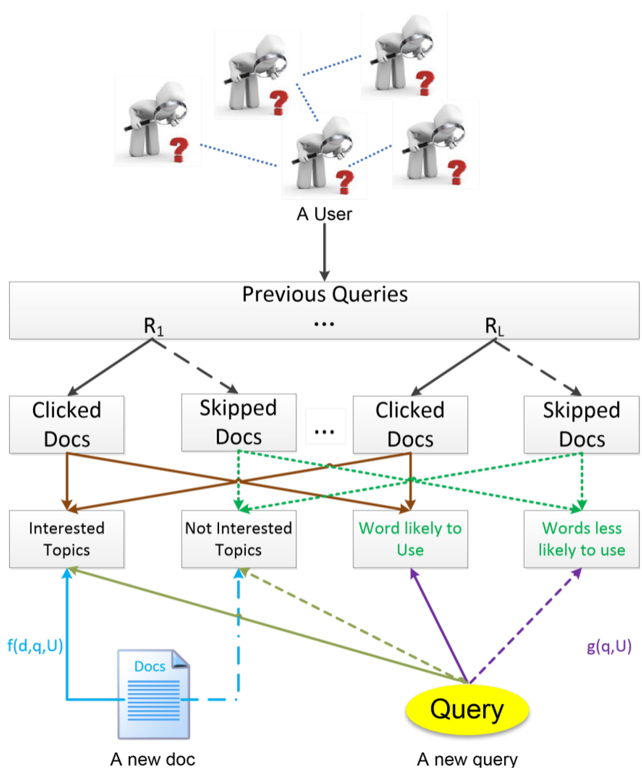


Fig. 4. The framework of the proposed model.

month query logs. Each of the validation and the testing set contains half of a month query logs of the same users. Personalizing search results based on long-term profiles is difficult, since users' interests can vary greatly over long time periods [6, 17]. The navigational queries (the goal is to go to specific known website URL) [49] are filtered on the validating set and testing set. Here, the navigational query list is maintained by the search engine company. We keep the navigational queries in the training set, because the topics related to the navigational queries are also interested to the user and therefore the interestingness can be in addition to the user training profiles. Significant tests use paired t-test with the significance levels of 0.05 and 0.01.

5.1 Models in Evaluation and Comparison

To evaluate whether we should consider the users' non-interesting topics in personalization, we compare our models with five well-performed approaches, including state-of-the-art personalization approaches in References [17, 56]. Five personalized strategies are investigated in

Table 3. Overview of the Improvement Rates over ORS

	MRR$^{\triangle}$	P@1$^{\triangle}$	P@3$^{\triangle}$	RScoring$^{\triangle}$
P-Click	+1.267%*	+2.525%*	+0.925%*	+0.425%*
Model1_LDA	−5.309%	−8.355%	−5.782%	−3.256%
Model2_LDA	−17.405%	−20.959%	−16.324%	−14.113%
Interesting Topics	+0.667%*	+1.778%*	+1.135%*	+0.305%*
LLP	+1.593%*†‡	+3.450%*†‡	+1.406%*†‡	+0.558%*†‡
LLP-Subtraction	+1.427%*†‡	+3.738%*†‡	**+2.008%***†‡	**+0.598%***†‡
LLP-Projection	**+1.878%***†‡	**+4.388%***†‡	+1.503%*†‡	+0.545%*†‡

*, †, and ‡ indicate significant improvement over ORS, P-Click, and Interesting Topics. correspondingly with the significance level of $p < 0.05$.

Reference [17], and we adopted P-Click as one of our baselines, since it has the best performance in Reference [17]. Some of the inferences we used are similar to the models in Reference [56] (e.g., Equation (5)), so we also adopted modified Model1 and Model2 in Reference [56] as baselines. We evaluate and compare our proposed models with the following models:

- The ORS without personalization: This is the score used in the commercial search engine, which is a strong baseline usually obtained from a mixture of several optimized approaches.
- The Personal-level Re-ranking (P-Click) model, which has the best performance in Reference [17]. The model assumes that the webpages that have more frequent clicks by a user in the past are more relevant to the user. The score is combined with ORS using Borda's ranking fusion method [19].
- Model1_LDA: This probabilistic personalization IR model is a direct revision of Model1 in [56]. Model 1 builds the user intent distribution over categorized topics (ODP categories). Model1 is shown to be effective on the search logs for the Bing search engine. Model1_LDAdirectly apply Model 1 to latent topics (continuous) instead of categorized topics (discrete).
- Model2_LDA: Similarly to Model1_LDA, this model directly applies Model2 in Reference [56] to latent topics. Model2 is a background model that reweights Model1 by the generic user distribution, which is computed by taking the weighted average of the topic distributions for each of the top-scoring search results.
- The model with only Interesting Latent Topics: Here we use our proposed estimation approach with only interesting topics, which is the numerator component in Equation (1).

We compare the above baselines with with our proposed models: LLP model in Equation (9), LLP model with subtraction strategy, and LLP model with orthogonal projection strategy. In this article, we focus on investigating the effect of non-interesting topics in personalization models. For the model with only Interesting Topics, we estimate all interesting topic probabilities as the methods in this article. This model share some similar ideas with Model1_LDA, and it is more comparable to the our proposed LLP models.

5.2 Evaluation Metrics

We evaluate the proposed personalized model by users' click-through data. The user-clicked webpages are regarded as relevant to the query. The re-ranked list from a personalized model is evaluated by comparing with the user-clicked webpage. Three evaluation metrics are applied: **Precision at ranking No. 1 (P@1)**, **Precision at ranking No. 3 (P@3)**, **Mean reciprocal rank (MRR)**, and Rank Scoring.

- The MRR [63] is the mean of the reciprocal of the rank at which the known item was found, averaged over all the queries $MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{rank_i}$.
- The Rank Scoring has been used to evaluate the personalized Web search accuracy [17, 58] . The expected utility of a ranked list of webpages is $R_s = \sum_i \frac{\delta(q,rank_i)}{2^{(rank_i-1)/(\alpha-1)}}$, where $\delta(q, rank_i)$ is 1 if the ith page is clicked and 0 otherwise, and α is set to 5 as in References [17, 58]. The Rank Scoring reflects the utilities of all test queries: $R = 100\frac{\sum_s R_s}{R_s^{max}}$, where R_s^{max} is the obtained maximum possible utility when all pages that have been clicked appear at the top of the ranked list.

5.3 Implementation and Parameter Settings

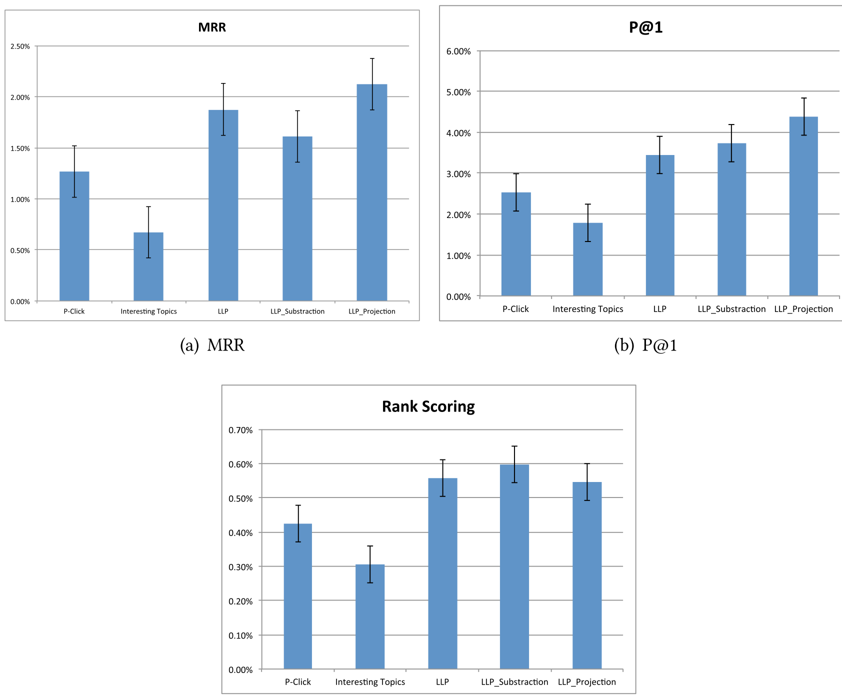
In Figure 4, we show an overview about how to implement our proposed models. A user is shown at top of the figure, which has a search history containing a series of training queries $\{r_1, r_2, \ldots, r_L\}$. Each of the query has 1 or more clicked webpages, as well as 0 or more skipped pages (0 when the first returned webpage is clicked). The interesting and non-interesting topics are generated from the clicked and skipped webpages correspondingly. They are the so-called users' positive and negative profiles. Also, we collect two bags of words extracted from the clicked and skipped webpages. These user-related variables are computed offline and stored after the training process. Given a new query q issued by the same user and a document d, we calculate $f(d, q, U)$ and $g(q, U)$ online with the variables obtained from the training data. $f(d, q, U)$, matches the topics in d, the distribution of q in the topics, and the users' positive and negative profiles. $g(q, U)$ matches the words in q with the words extracted from the clicked and skipped webpages. Finally, the personalization score is the product of $f(d, q, U)$ and $g(q, U)$, and then integrated with the original ranking score. In the search engine, this step serves the reranking purpose. The top ranked documents returned with the original ranking score will be reranked with the additional LLP approaches and presented to the users.

There are two parameters in the LLP models. λ is the linear combination parameter in LLP (see Equation (9)) and μ is the smoothing parameter. λ is set to be in $\{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.81, 0.9, 0.95, 1\}$, and μ is set to be in $\{100, 200, 500, 1000, 2000\}$. We train a user's profile on the training set, and validate the parameters λ an μ on the validation set for each user. Then the optimal parameters are applied on the testing set. For each user's query, the top 10 documents returned by the search engine are reranked and evaluated.

We build an LDA model on all the crawled webpages. Since the data collection is relatively large and the topics over the internet are very diverse, we choose 1,000 as the number of topics, which has been adopted in References [29, 43]. Common settings of parameters are $\beta = 0.01$ and $\alpha = 20/(\#\text{ of topics})$, or $\alpha = 50/(\#\text{ of topics})$ [69]. The smaller alpha indicates the more sparse the distribution. We use Dirichlet priors in the LDA estimation with $\alpha = 0.01$ and $\beta = 0.01$.

5.4 Overall Performance

The overall performance is shown in Table 3, where each row represents a model. MRR$^\triangle$, P@1$^\triangle$, P@3$^\triangle$, and Rank Scoring$^\triangle$ denote the improvement rates over ORS. ORS is obtained directly from the search engine. The rest of the models are personalized models, and their performance is compared with ORS. The results are averaged over the users. We compare with several state-of-the-art personalized models. First, the commercial search engine's ORS already has good performance. P-Click personalized ranking model boosts the performance of ORS. Second, we can see that Model1_LDA and Model2_LDA do not outperform ORS. The results indicate that Model1 and Model2 are not suitable to be directly applied on the continuous latent topic distributions or not



(a) MRR

(b) P@1

(c) Rank Scoring

Fig. 5. The overall improvement rates with error bars.

suitable for our dataset, although they have very good performance on discrete (categorized) topic distributions in Reference [56].

In the rest of this article, we will not discuss these two models in detail. Instead, we focus on the model considering only the interesting topics, which is a different implementation (using the estimations proposed in this article) of Model 1 in Reference [56]. From Table 3, we can see that only considering the interesting topics in personalization can improve the performance to some extent. The proposed LLP model outperforms both ORS and the model considering only the interesting topics model. This indicates that both the interesting or not interesting topics could characterize the users' searching intent. We also propose the subtraction strategy and the orthogonal projection strategy on the LLP model, which are denoted as LLP-Subtraction and LLP-Projection in Table 3. We can see that all personalization approaches have significant improvement over ORS. Adding negative profiles (e.g., LLP, LLP-Subtraction, and LLP-Projection) provides additional performance improvement compared to using only the interesting topics. This indicates the usefulness of the negative profile. The LLP-based models have the best performance, and the LLP-Subtraction and LLP-Projection can further improve the LLP model in most of the cases (3 out of 4 metrics). Especially for the top ranked positions, LLP-Subtraction and LLP-Projection can both outperform LLP in terms of P@1 and P@3. The significant tests with p-value set as <0.05 show that the P-Click, Interesting Topics, LLP, LLP-Subtraction, and LLP-Projection can improve the baseline ORS significantly, and LLP models further significantly improves the model with only interesting topics. Therefore, our proposed model are at least comparable to, if not better than, the existing state-of-the-art personalized models.

In addition, we plot the improvement rates with error bars in Figure 5. We can observe that the error bar of Interesting Topics does not overlap with error bars of LLP models. It shows that the

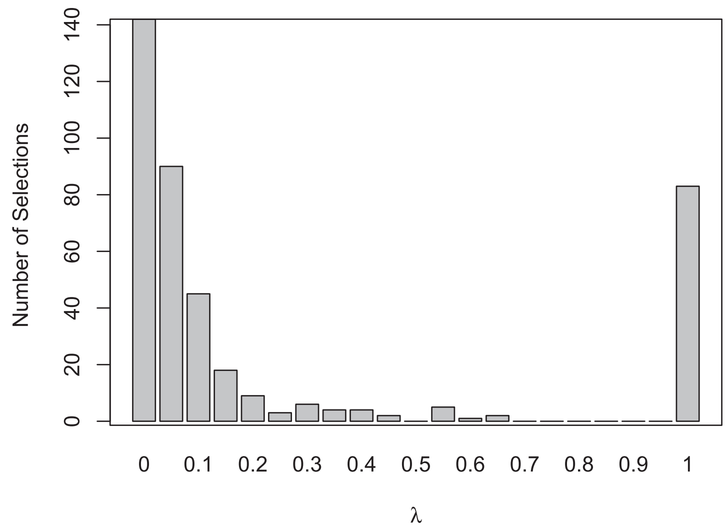


Fig. 6. Parameter Sensitivity with λ for *LLP*.

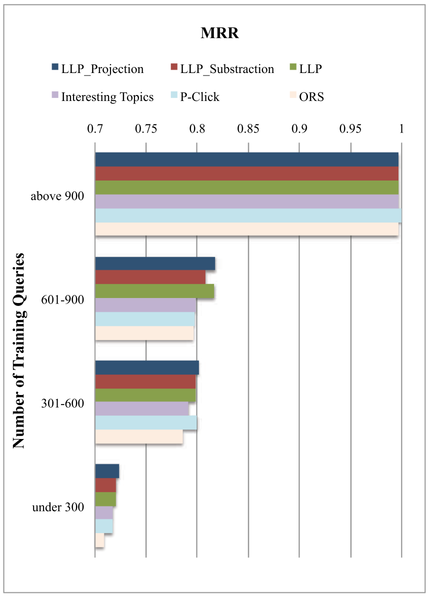
LLP models improve Interesting Topics significantly. Incorporating non-interesting topics can be useful addition to character users' search intents.

5.5 Parameter Sensitivity

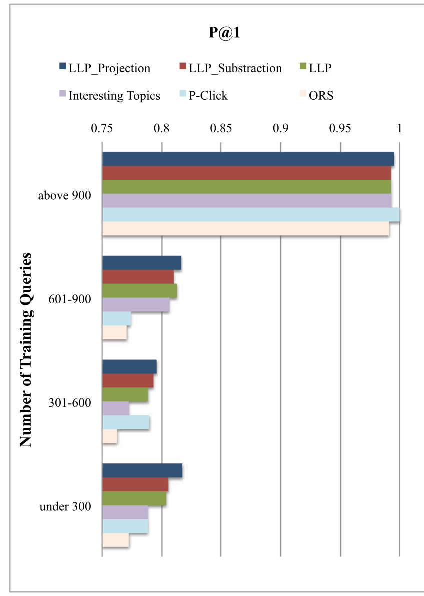
The balancing parameter λ represents how much personalization affects the overall weighting, raining from 0 to 1. λ is trained for each user on the validation set. Figure 6 shows how many times a λ value is selected as optimal for a user for an evaluation metric. Here we look at model *LLP*, and the other proposed models have very similar tendencies. We can see that the result are very extreme that personalization works very well on some user ($\lambda = 1$). For some other users, personalization does not benefit retrieval ($\lambda = 0$). Many users benefit both from the personalized and non-personalized ranking. Based on the fact that the ranking results in ORS are already optimized in Yahoo! Search, the proposed models can promote the retrieval performance for a large portion of real users. For the smoothing parameter μ for $g(q, U)$, the selection of its value does not affect the ranking performance very much, and we will not discuss the sensitivity for μ in this article.

5.6 Performance with the User Frequencies

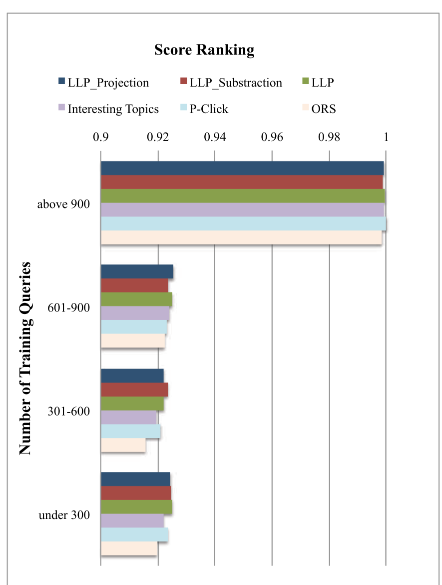
Here we group the users by their training query sizes, and study how the user frequency can affect personalization performance. The groups are user training set with the following queries: {under 300, 301-600, 601-900, above 900}. The results are shown in Figure 7. Due to privacy issues, the values are normalized where all results are divided by the highest value. We can see from the figure that all personalized approaches have better performance when the user searches more frequently. It is because more information is available for users with larger training query sets. Frequent users are more likely to click on the popular websites. However, personalization is difficult for the less frequent users, whom we know less information for new queries. For those groups {under 300, 301-600, 601-900}, LLP models generally have more significant improvements. In most of the cases, LLP-Projection has the best performance, especially on the difficult cases with less information. It indicates that LLP-Projection is at least comparable to, if not better than, the existing state-of-the-art personalized models, for different groups of users.



(a) MRR



(b) P@1



(c) Rank Scoring

Fig. 7. Normalized personalization performance with the users' training set size.

5.7 Performance with the Query Length

In this section, we analyze how the query length can affect personalization performance. Here the query length refers to the number of words in a query after stopwords are removed. The query length statistics are shown in shown in Table 4, and the according normalized results are in Figure 8. The values in the figure are normalized by being divided by the max value. We can

Table 4. Query Statistics

QueryLength	Percentage in the Testing Set
1	8.35%
2	29.70%
3	22.64%
4	14.65%
>5	24.67%

Table 5. Comparison between F_dqu (LLP without the Relevancy of a Query to a User) and LLP

	MRR$^{\triangle}$	P@1$^{\triangle}$	P@3$^{\triangle}$	RScoring$^{\triangle}$
F_dqu	+1.392%	+3.058%	+1.244%	+0.481%
LLP	+1.593%	+3.450%	+1.406%	+0.558%

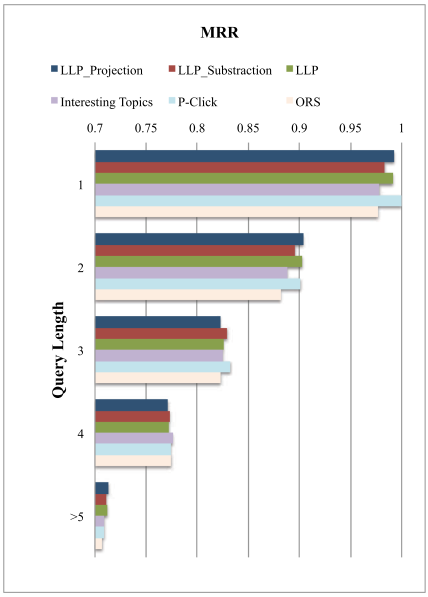
see that the personalization models have the best performance when the query length is 1. When the query length increases, all models have lower accuracy. The possible reason could be that the longer queries often too specific and occur less frequently. Therefore, the longer queries are often more difficult. The personalization approaches improve ORS more on the long queries than on the short queries. LLP-Projection and LLP-Subtraction have higher performance, especially when the query longer, for example in Figure 8(b), they outperform LLP and other baselines when queries are longer (include 3, 4, or greater than 5 terms). P-Click is a frequency-based personalization approach. So it favours more on the queries with length 1. In general, compared to P-Click, the topic-based approaches can match user, document, and long query by topics rather than exact match.

5.8 Effect of the Relevancy of a Query to a User

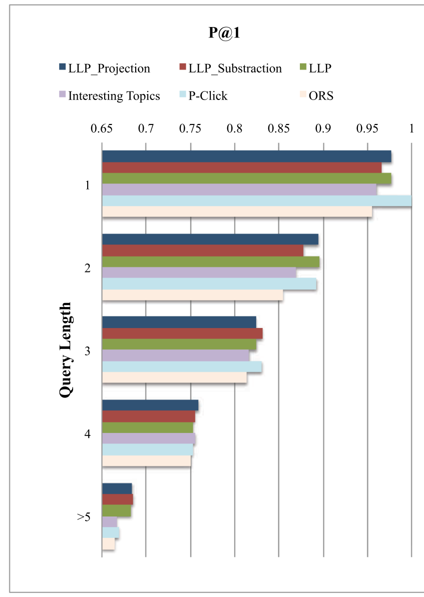
We further study the effect of the relevancy of a query to a user in the model, $g(q, U)$ in Equation (1). $g(q, U)$ is a user-specific parameter controlling how much personalization should be considered for a query. We consider the model with only the document-query matching for the user component F_dqu: Here we use our proposed estimation approach with only the $f(d, q, U)$ component in Equation (1). The results are shown in Table 5. LLP has better performance than F_dqu, which indicates effectiveness of additionally considering the relevancy of a user to a query. Please note that $g(q, U)$ will be different across various users, and even different for the same user across various queries. However, the global parameter λ in Equation (8) indicates how much overall personalization will be considered for all users and queries. Therefore, $g(q, U)$ can bring additional effect to adjust personalization of the proposed model.

5.9 Discussion and Analysis

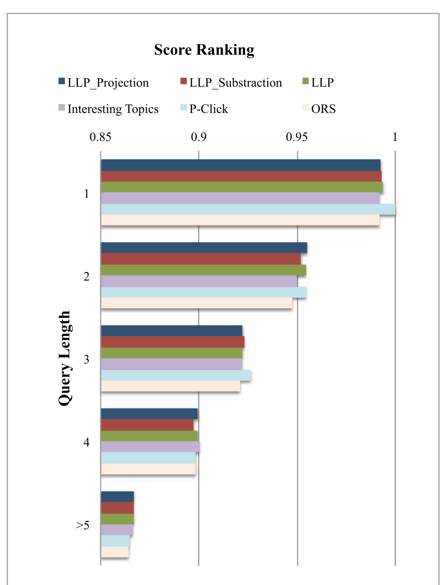
In this article, our experiments are conducted on a large-scale real user search log dataset from Yahoo! search engine. The experimental results show that considering the non-interesting topics is an useful addition to the interesting topics in personalizing the users' search intent. The positive profile aggregates interesting topics. Using the negative profile characterizes the boundary between the interesting and not-interesting topics. These two user profiles describe the users' history more accurately. The processing strategies for negative profiles are proposed to remove the misleading information in the user profiles and therefore further boost the retrieval performance. LLP-Projection usually has the best performance among all the models. We analyze the performance of the models with respect to the user search frequency and query length. The proposed



(a) MRR



(b) P@1



(c) Rank Scoring

Fig. 8. Normalized personalization performance with query length.

models have more performance improvements on the difficult cases when there are insufficient number of training queries and/or longer queries. The findings would be beneficial for further applications.

In general, the performance cost for the proposed personalized approach is not high. The process for generating latent topics from the collection is conducted offline. During retrieval, we are reranking 10 top documents, and the personalized scores are calculated generatively without iterations. It makes our proposed models scalable and relatively fast even on large dataset.

6 CONCLUSIONS AND FUTURE WORK

We first propose a novel model LLP that integrates the users' personalization information into the Web search process. In this model, we analyze the users' search logs and aggregate both the interesting topic and non-interesting topics for each user over the training queries. For a given query, the user clicked documents are regarded as the context containing the interesting topics, and the skipped documents have a higher probability of including some topics that are not interesting to the user. A positive user profile corresponds to a vector where each element is the probability of that the user is interested in a topic. A negative user profile contains the probabilities of that the user is not interested in the topics. Both the positive and negative profiles are then matched with the new query and the documents for a new personalized ranking list. The parameters used in the model are obtained by training on the training set and validating on the validation set.

We further argue one of the reasons a user skips a document might be that the clicked documents include similar topics as in the skipped document. That means the topics in the skipped document are covered by the clicked documents so that the user does not have to click a similar document. We propose two strategies to process the document topic representations so that we can eliminate the interesting topics from the skipped topics for a given query. The experimental results show the effectiveness of our proposed models and strategies. Our proposed models can make significant improvements compared with existing approaches, in terms of MRR, P@1, P@3, and Score Ranking. In addition, our proposed models are robust and can be easily reproduced on other datasets. This is because the proposed models are not domain specific, and the latent topics can be generated directly from any given dataset.

There are several future directions of this work. One direction is to analyze non-interesting topics for several groups of users and utilize the group negative profile to boost the personalization on the individual users. Another direction is to analyze the information of the queries and summarize their behaviours in terms of personalization. Since the proposed model requires enough user behaviors to build the model, we are currently focusing on the long-term personalization. The proposed models could be extended to deal with more sparse user data and applied in short-term (sessional) personalization tasks in the future. Further, the personalized scores can be added to learning to rank approaches as one of the features. The presentation biases, which might influence the user behaviors other than user's interest, can also be studied. Furthermore, the models proposed in this article can be applied on other datasets can compared with more baselines. We will also propose novel approaches on LDA to generate latent representations that can better fit the personalized framework.

ACKNOWLEDGMENTS

We greatly appreciate the anonymous reviewers and associate editor for their insightful comments.

REFERENCES

[1] Eugene Agichtein, Eric Brill, Susan Dumais, and Robert Ragno. 2006. Learning user interaction models for predicting web search result preferences. In *Proceedings of the 29th International ACM SIGIR Conference on Research and Development in Information Retrieval.* 3–10.

[2] Wasi Uddin Ahmad, Kai-Wei Chang, and Hongning Wang. 2018. Intent-aware query obfuscation for privacy protection in personalized web search. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval.* 285–294.

[3] Paul N. Bennett, Ryen W. White, Wei Chu, Susan T. Dumais, Peter Bailey, Fedor Borisyuk, and Xiaoyuan Cui. 2012. Modeling the impact of short-and long-term behavior on search personalization. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval.* 185–194.

[4] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3 (2003), 993–1022.
[5] Fei Cai, Shuaiqiang Wang, and Maarten de Rijke. 2017. Behavior-based personalization in web search. *J. Assoc. Inf. Sci. Technol.* 68, 4 (2017), 855–868.
[6] Mark J. Carman, Fabio Crestani, Morgan Harvey, and Mark Baillie. 2010. Towards query log based personalization using topic models. In *Proceedings of the 19th Proceedings of the 29th ACM International Conference on Information and Knowledge Management.* ACM, 1849–1852.
[7] Xu Chen, Kun Xiong, Yongfeng Zhang, Long Xia, Dawei Yin, and Jimmy Xiangji Huang. 2020. Neural feature-aware recommendation with signed hypergraph convolutional network. *ACM Trans. Inf. Syst.* 39, 1 (2020), 1–22.
[8] Haibin Cheng and Erick Cantú-Paz. 2010. Personalized click prediction in sponsored search. In *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining.* ACM, 351–360.
[9] Qiannan Cheng, Zhaochun Ren, Yujie Lin, Pengjie Ren, Zhumin Chen, Xiangyuan Liu, and Maarten de de Rijke. 2021. Long short-term session search: Joint personalized reranking and next query prediction. In *Proceedings of the International Conference on the World Wide Web.* 239–248.
[10] Zhicong Cheng, Bin Gao, and Tie-Yan Liu. 2010. Actively predicting diverse search intent from user browsing behaviors. In *Proceedings of the 19th International Conference on the World Wide Web.* ACM, 221–230.
[11] Mariam Daoud, Lynda Tamine-Lechani, and Mohand Boughanem. 2008. Learning user interests for a session-based personalized search. In *Proceedings of the 2nd International Symposium on Information Interaction in Context.* ACM, 57–64.
[12] Abhinandan S. Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google news personalization: Scalable online collaborative filtering. In *Proceedings of the 16th International Conference on the World Wide Web.* ACM, 271–280.
[13] Luis M. de Campos, Juan M. Fernández-Luna, Juan F. Huete, and Eduardo Vicente-Lopez. 2014. Using personalization to improve XML retrieval. *IEEE Trans. Knowl. Data Eng.* 26, 5 (2014), 1280–1292.
[14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT'19)*, 4171–4186.
[15] Sahraoui Dhelim, Nyothiri Aung, and Huansheng Ning. 2020. Mining user interest based on personality-aware hybrid filtering in social networks. *Knowl.-Bas. Syst.* 206 (2020), 106227. https://doi.org/10.1016/j.knosys.2020.106227
[16] Mengting Diao, Zhongbao Zhang, Sen Su, Shuai Gao, and Huafeng Cao. 2020. UPON: User profile transferring across networks. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM'20).* Association for Computing Machinery, New York, NY, 265–274. https://doi.org/10.1145/3340531.3411964
[17] Zhicheng Dou, Ruihua Song, and Ji-Rong Wen. 2007. A large-scale evaluation and analysis of personalized search strategies. In *Proceedings of the 16th International Conference on World Wide Web.* ACM, 581–590.
[18] Zhicheng Dou, Ruihua Song, Xiaojie Yuan, and Ji-Rong Wen. 2008. Are click-through data adequate for learning web search rankings? In *Proceedings of the 17th ACM Proceedings of the 29th ACM International Conference on Information and Knowledge Management.* ACM, 73–82.
[19] Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. 2001. Rank aggregation methods for the web. In *Proceedings of the 10th International Conference on World Wide Web.* ACM, 613–622.
[20] Norbert Fuhr. 1992. Probabilistic models in information retrieval. *Comput. J.* 35, 3 (1992), 243–255.
[21] Jingyue Gao, Yang Lin, Yasha Wang, Xiting Wang, Zhao Yang, Yuanduo He, and Xu Chu. 2020. Set-Sequence-Graph: A multi-view approach towards exploiting reviews for recommendation. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM'20).* ACM, New York, NY, 395–404. https://doi.org/10.1145/3340531.3411939
[22] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. *Commun. ACM* 63, 11 (2020), 139–144.
[23] Dongyi Guan, Sicong Zhang, and Hui Yang. 2013. Utilizing query change for session search. In *Proceedings of the 36th International ACM SIGIR Conference on Research & Development in Information Retrieval.* ACM, 453–462.
[24] Bahareh Harandizadeh and Sameer Singh. 2020. Tweeki: Linking named entities on Twitter to a knowledge graph. In *Proceedings of the 6th Workshop on Noisy User-generated Text (W-NUT 2020).* Association for Computational Linguistics, 222–231.
[25] Ben He, Jimmy Xiangji Huang, and Xiaofeng Zhou. 2011. Modeling term proximity for probabilistic information retrieval models. *Inf. Sci.* 181, 14 (2011), 3017–3031. https://doi.org/10.1016/j.ins.2011.03.007
[26] Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2017. An unsupervised neural attention model for aspect extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).* 388–397.

[27] Matthew Hoffman, Francis R. Bach, and David M. Blei. 2010. Online learning for latent dirichlet allocation. In *Advances in Neural Information Processing Systems*. 856–864.

[28] Vagelis Hristidis, Yao Wu, and Louiqa Raschid. 2014. Efficient ranking on entity graphs with personalized relationships. *IEEE Trans. Knowl. Data Eng.* 26, 4 (2014), 850–863.

[29] Chi-I Hsu and Chaochang Chiu. 2017. A hybrid Latent Dirichlet Allocation approach for topic classification. In *Proceedings of the IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA'17)*. IEEE, 312–315.

[30] Xiangji Huang, Fuchun Peng, Dale Schuurmans, Nick Cercone, and Stephen E. Robertson. 2003. Applying machine learning to text segmentation for information retrieval. *Inf. Retr.* 6, 3–4 (2003), 333–362. https://doi.org/10.1023/A:1026028229881

[31] Xiangji Huang, Ming Zhong, and Luo Si. 2005. York University at TREC 2005: Genomics track. In *Proceedings of the 14th Text Retrieval Conference (TREC'05)*, Ellen M. Voorhees and Lori P. Buckland (Eds.). National Institute of Standards and Technology.

[32] Thorsten Joachims. 2002. Optimizing search engines using clickthrough data. In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 133–142.

[33] Jin Young Kim, Kevyn Collins-Thompson, Paul N. Bennett, and Susan T. Dumais. 2012. Characterizing web content, user interests, and search behavior by reading level and topic. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*. ACM, 213–222.

[34] Diederik P. Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *Proceedings of the International Conference on Learning Representations*.

[35] Jiahui Liu, Peter Dolan, and Elin Rønby Pedersen. 2010. Personalized news recommendation based on click behavior. In *Proceedings of the 15th ACM Annual Conference on Intelligent User Interfaces (IUI'10)*. ACM, 31–40.

[36] Shang Liu, Wanli Gu, Gao Cong, and Fuzheng Zhang. 2020. Structural relationship representation learning with graph embedding for personalized product search. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM (CIKM'20)*. ACM, New York, NY, 915–924. https://doi.org/10.1145/3340531.3411936

[37] Weiwen Liu, Qing Liu, Ruiming Tang, Junyang Chen, Xiuqiang He, and Pheng Ann Heng. 2020. Personalized re-ranking with item relationships for e-commerce. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM'20)*. Association for Computing Machinery, New York, NY, 925–934. https://doi.org/10.1145/3340531.3412332

[38] Peter Lofgren, Siddhartha Banerjee, and Ashish Goel. 2016. Personalized pagerank estimation and search: A bidirectional approach. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining*. ACM, 163–172.

[39] Nicolaas Matthijs and Filip Radlinski. 2011. Personalizing web search using long term browsing history. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*. ACM, 25–34.

[40] Ph Mylonas, David Vallet, Pablo Castells, Miriam Fernández, and Yannis Avrithis. 2008. Personalized information retrieval based on context and ontological knowledge. *Knowl. Eng. Rev.* 23, 01 (2008), 73–100.

[41] Eugene Opoku-Mensah, Fengli Zhang, Fan Zhou, and Philemon Kibiwott Kittur. 2017. Understanding user situational relevance in ranking web search results. In *Proceedings of the 8th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON'17)*. IEEE, 405–410.

[42] Athanasios Papagelis and Christos Zaroliagis. 2012. A collaborative decentralized approach to web search. *IEEE Trans. Syst. Man. Cybernet. Part A: Syst. Hum.* 42, 5 (2012), 1271–1290.

[43] Ian Porteous, David Newman, Alexander Ihler, Arthur Asuncion, Padhraic Smyth, and Max Welling. 2008. Fast collapsed gibbs sampling for latent dirichlet allocation. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 569–577.

[44] Shinjee Pyo, Eunhui Kim, et al. 2015. LDA-Based unified topic modeling for similar TV user grouping and TV program recommendation. *IEEE Trans. Cybernet.* 45, 8 (2015), 1476–1490.

[45] Xiaoru Qu, Zhao Li, Jialin Wang, Zhipeng Zhang, Pengcheng Zou, Junxiao Jiang, Jiaming Huang, Rong Xiao, Ji Zhang, and Jun Gao. 2020. Category-Aware graph neural networks for improving e-commerce review helpfulness prediction. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM'20)*. ACM, New York, NY, 2693–2700. https://doi.org/10.1145/3340531.3412691

[46] Alec Radford, Luke Metz, and Soumith Chintala. 2016. Unsupervised representation learning with deep convolutional generative adversarial networks. In *Proceedings of the 4th International Conference on Learning Representations (ICLR'16)* (2016).

[47] Mohammad Mustaneer Rahman and Nor Aniza Abdullah. 2018. A personalized group-based recommendation approach for Web search in E-learning. *IEEE Access* 6 (2018), 34166–34178.

[48] Stephen E. Robertson and Steve Walker. 1994. Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval. In *Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, W. Bruce Croft and C. J. van Rijsbergen (Eds.). ACM/Springer, 232–241. https://doi.org/10.1007/978-1-4471-2099-5_24

[49] Daniel E. Rose and Danny Levinson. 2004. Understanding user goals in web search. In *Proceedings of the 13th International Conference on the World Wide Web*. ACM, 13–19.

[50] Michal Rosen-Zvi, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. 2004. The author-topic model for authors and documents. In *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI'04)*. AUAI Press, 487–494.

[51] Jitao Sang, Changsheng Xu, and Dongyuan Lu. 2012. Learn to personalized image search from the photo sharing websites. *IEEE Trans. Multimedia* 14, 4 (2012), 963–974.

[52] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *Proceedings of the European Semantic Web Conference*. Springer, 593–607.

[53] Xuehua Shen, Bin Tan, and ChengXiang Zhai. 2005. Implicit user modeling for personalized search. In *Proceedings of the 14th ACM Proceedings of the 29th ACM International Conference on Information and Knowledge Management*. ACM, 824–831.

[54] Wei Song, Yu Zhang, Ting Liu, and Sheng Li. 2010. Bridging topic modeling and personalized search. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING'10)*. ACL, 1167–1175.

[55] Yang Song and Li-wei He. 2010. Optimal rare query suggestion with implicit user feedback. In *Proceedings of the 19th International Conference on the World Wide Web*. ACM, 901–910.

[56] David Sontag, Kevyn Collins-Thompson, Paul N. Bennett, Ryen W. White, Susan Dumais, and Bodo Billerbeck. 2012. Probabilistic models for personalizing web search. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*. ACM, 433–442.

[57] Ben Steichen, Helen Ashman, and Vincent Wade. 2012. A comparative survey of Personalised Information Retrieval and Adaptive Hypermedia techniques. *Inf. Process. Manage.* 48, 4 (2012), 698–724.

[58] Jian-Tao Sun, Hua-Jun Zeng, Huan Liu, Yuchang Lu, and Zheng Chen. 2005. Cubesvd: A novel approach to personalized web search. In *Proceedings of the 14th International Conference on World Wide Web*. ACM, 382–390.

[59] Bin Tan, Xuehua Shen, and ChengXiang Zhai. 2006. Mining long-term search history to improve search accuracy. In *Proceedings of the 12th ACM Conference of the Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD'06)*. ACM, 718–723.

[60] Xiaohui Tao, Yuefeng Li, and Ning Zhong. 2011. A personalized ontology model for web information gathering. *IEEE Trans. Knowl. Data Eng.* 23, 4 (2011), 496–511.

[61] Jaime Teevan, Susan T. Dumais, and Eric Horvitz. 2005. Personalizing search via automated analysis of interests and activities. In *Proceedings of the 28th International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 449–456.

[62] David Vallet, Iván Cantador, and Joemon M. Jose. 2010. Personalizing web search with folksonomy-based user and document profiles. In *Advances in Information Retrieval*. Springer, 420–431.

[63] Ellen M. Voorhees and D. M. Tice. 2000. Overview of the TREC-9 question answering track. In *Proceedings of the Text Retrieval Conference (TREC'00)*.

[64] Thanh Vu, Alistair Willis, Son N. Tran, and Dawei Song. 2015. Temporal latent topic user profiles for search personalisation. In *Advances in Information Retrieval*. Springer, 605–616.

[65] Hongning Wang, Xiaodong He, Ming-Wei Chang, Yang Song, Ryen W. White, and Wei Chu. 2013. Personalized ranking model adaptation for web search. In *Proceedings of the 36th International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 323–332.

[66] Hongwei Wang, Fuzheng Zhang, Mengdi Zhang, Jure Leskovec, Miao Zhao, Wenjie Li, and Zhongyuan Wang. 2019. Knowledge-Aware graph neural networks with label smoothness regularization for recommender systems. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19)*. ACM, New York, NY, 968–977. https://doi.org/10.1145/3292500.3330836

[67] Wei Wang, Payam Barnaghi, and Andrzej Bargiela. 2010. Probabilistic topic models for learning terminological ontologies. *IEEE Trans. Knowl. Data Eng.* 22, 7 (2010), 1028–1040.

[68] Yue Wang, Dawei Yin, Luo Jie, Pengyuan Wang, Makoto Yamada, Yi Chang, and Qiaozhu Mei. 2016. Beyond ranking: Optimizing whole-page presentation. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining*. 103–112.

[69] Xing Wei and W. Bruce Croft. 2006. LDA-based document models for ad-hoc retrieval. In *Proceedings of the 29th International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 178–185.

[70] Ryen W. White, Wei Chu, Ahmed Hassan, Xiaodong He, Yang Song, and Hongning Wang. 2013. Enhancing personalized search by mining and modeling task behavior. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 1411–1420.

[71] Jing Yao, Zhicheng Dou, and Ji-Rong Wen. 2020. Employing personal word embeddings for personalized search. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1359–1368.

[72] Puxuan Yu, Wasi Uddin Ahmad, and Hongning Wang. 2018. Hide-n-Seek: An intent-aware privacy protection plugin for personalized web search. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 1333–1336.

[73] Jiashu Zhao, Jimmy Xiangji Huang, and Ben He. 2011. CRTER: Using cross terms to enhance probabilistic information retrieval. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 155–164.

[74] Jiashu Zhao, Jimmy X. Huang, and Xiaohua Hu. 2013. BPLT+: A Bayesian-based personalized recommendation model for health care. *BMC Genom.* 14, Suppl 4 (2013), S6.

[75] Jiashu Zhao, Jimmy Xiangji Huang, and Zheng Ye. 2014. Modeling term associations for probabilistic information retrieval. *ACM Trans. Inf. Syst.* 32, 2 (2014), 7.

[76] Xujuan Zhou, Sheng-Tang Wu, Yuefeng Li, Yue Xu, Raymond Y. K. Lau, and Peter D. Bruza. 2006. Utilizing search intent in topic ontology-based user profile for web mining. In *Proceedings of the 2006 Web Intelligence*. IEEE, 558–564.

Received November 2020; revised July 2021; accepted July 2021