# Terms, Topics & Tasks: Enhanced User Modelling for Better Personalization

Rishabh Mehrotra
Dept of Computer Science
University College London
r.mehrotra@cs.ucl.ac.uk

Emine Yilmaz
Dept of Computer Science
University College London
emine.yilmaz@ucl.ac.uk

## ABSTRACT

Given the distinct preferences of different users while using search engines, search personalization has become an important problem in information retrieval. Most approaches to search personalization are based on identifying topics a user may be interested in and personalizing search results based on this information.

While topical interests information of users can be highly valuable in personalizing search results and improving user experience, it ignores the fact that two different users that have similar topical interests may still be interested in achieving very different tasks with respect to this topic (e.g. the type of tasks a broker is likely to perform related to finance is likely to be very different than that of a regular investor). Hence, considering user's topical interests jointly with the type of tasks they are likely to be interested in could result in better personalised experience for users.

We present an approach that uses search task information embedded in search logs to represent users by their actions over a task-space as well as over their topical-interest space. In particular, we describe a tensor based approach that represents each user in terms of (i) user's topical interests and (ii) user's search task behaviours in a coupled fashion and use these representations for personalization. Additionally, we also integrate user's historic search behavior in a coupled matrix-tensor factorization framework to learn user representations. Through extensive evaluation via query recommendations and user cohort analysis, we demonstrate the value of considering topic specific task information while developing user models.

## Categories and Subject Descriptors

H.3.3 [**Information Storage And Retrieval**]: Information Search and Retrieval—*User Modelling*

## Keywords

Search tasks; User modelling; Tensor Decomposition


*ICTIR'15,* September 27–30, Northampton, MA, USA.
© 2015 ACM. ISBN 978-1-4503-3833-2/15/09 ...$15.00.
DOI: http://dx.doi.org/10.1145/2808194.2809467.

## 1. INTRODUCTION

As consumers of the informational content, different users have distinct information seeking preferences; thus accurately understanding their respective information needs and decision preferences is crucial for providing effective support during search interactions. While user behaviours are largely determined by their own goals and preferences, the mined knowledge from log activity data reveals different user intentions and behaviour patterns, which provide unique signals for user centric optimization and personalization.

Web search personalization has recently received a lot of attention by the research community. Personalized search leverages information about an individual to identify the most relevant recommendations for that person. A challenge for personalization is in collecting user profiles that are sufficiently rich to be useful in settings such as result ranking and query recommendations.

Most previous work on personalization has focused on using long term search histories to provide better personalized results. In particular, most recent personalized search systems mainly focus on identifying topics a user might be interested in based on their search history and improving their search experience by identifying and using information from different topics [33, 4].

Even though using topical interest of users can be highly valuable in personalizing search results and improving user experience, it still ignores the fact that two different users that have similar topical interests may still be interested in achieving very different tasks with respect to this topic. For example, a stockbroker and a normal investor while being interested in the same topic (finance), perform quite different set of search tasks and as a result need different kinds and levels of support while tackling these tasks. More generally, while topical interests capture the heterogeneity among users stemming from varied topical interests, such task based approaches would assist in capturing the heterogeneity stemming from differences in user needs and behaviors. Hence, using task information together with topics could result in systems that can provide improved personalized search experience to users.

In this work, we focus on using *search task* information for user modeling, where a search task has been previously defined as an atomic information need that consists of a set of related (sub)tasks [11]. In a recent poster [22], we showed that search tasks can indeed be used for personalization. This work was based on replacing topic models with search tasks for personalization and building task based representations of users for topic modelling. Hence, this work ignores the fact that tasks users are interested in tend to be topic specific: people tend to be interested in achieving certain tasks only for certain topics. In this work, we investigate the idea of task based personalization in detail and develop a model that combines topic based user modelling with task based user models. Additionally, we look at the user's search history that provides information about user's term usage behavior. We integrate user's historical information to the task-topic tensor framework by proposing a coupled matrix-tensor factorization model which jointly learns user representations based on their search history, term usage behavior, topical interest profiles and search task behaviors.

In particular, we show that it is possible to represent the topic specific tasks users are interested in by representing users in terms of a 3-modal $< user - topic - task >$ tensor (multidimentional array). We show that tensor factorization can be used to learn *coupled task-topic based user representations* for each user, thereby incorporating tasks together with topics in representing the user population. The tensor based framework helps in encapsulating the complex interactions between topics and tasks across the entire user population and learns a low dimensional factor model wherein user's interests, preferences and behaviors are determined by an interplay between these latent factors. We further extend the tensor based framework to include user's search history information by proposing the use of coupled matrix-tensor factorization model [3] wherein the matrix captures user's topical interest and search task information while the matrix captures user's term usage behavior.

Finally, we show that the proposed methods result in better user profiles by evaluating the quality of our approach on a variety of tasks for personalisation including collaborative query recommendation, cluster based recommendation and user cohort analysis.

## 2. RELATED WORK

There is a growing interest in the information retrieval and machine learning communities in moving beyond context free search experiences, and towards examining how knowledge of a searcher's interests and search context can be used to improve various aspects of search (e.g., ranking, query recommendation, query classification). Even though there is significant amount of prior work on search task identification and personalisation based on topical representations of users, there is no prior work that uses topic based task representations for users. We first summarise the related work on representing users and personalization, and provide a summary of previous work on task identification from search engine query logs, which will be used by the models proposed in this work.

**User Representation & Personalization**

Irrespective of where the user's data comes from, a model must encode this data. A variety of such models have been used in the past including a vector of weighted terms (e.g. [20]), a set of concepts (e.g. [7]), using topic models (e.g. [8]) or a hierarchical category tree based on ODP and corresponding keywords (e.g. [4]).

Teevan *et al.* [31] constructed user profiles from indexed desktop documents and showed that this information could be used to re-rank search results and improve relevance for individuals. Matthijs and Radlinski [20] constructed user profiles using users' browsing history, and evaluated their approach using an interleaving methodology. Their approach focused on using term based user profiles which often limit the scope of personalization as different users inherently follow different distributions over words. Dou *et al.* [7] investigated a number of heuristics for creating user profiles and generating personalized rankings. Bennett *et al.* [4] made use of hand picked Open Directory Project (ODP) topical categories to construct user profiles. While such topical categories are easily specified, much human effort is required in labelling queries for each topic. ODP categories based methods restricts topic coverage in a major way as search logs offer much richer content both in terms of the number of topics involved as well as the granularity level of each topic. Very recently, Wang *et al* [33] have proposed a generative model which models users as a mixture over latent user groups wherein each group shares a common distribution over queries and a common click preference pattern. Finally, Harvey *et al.* [8] use the topic model based approach to build user profiles from topics obtained and personalize search results based on the learnt user profiles.

Aiming for short-term personalization, Sriram et al. [29] describe a search engine that personalized based on the current user session. A longer term personalization click model can also be used, exploting clickthrough data collected over a long time period. For example, Speretta and Gauch [28] and Qiu and Cho [26] model users by classifying previously visited web pages into a topic hierarchy, using this model to re-rank future search results. Also, a particularly straightforward yet effective search interaction personalization approach is PClick, proposed by Dou et al. [7]. This method involves promoting URLs previously clicked on by the same user for the same query. The user representation model we present in this work could be easily used in any of these personalization techniques.

**Search Task Identification**

It was peviously shown that approximately 75% of user search sessions involve multi-tasking [18], which makes task identification an important step towards understanding user goals. People have been shown to pursue a wide range of different *search tasks* online [12, 19] and inferences about task behavior have been shown to have value in areas such as modeling search satisfaction [9]. Information about user's task involvements provides a completely different aspect of user's intents and provide strong contextual cues which could be leveraged by future recommendation and advertising engines to better serve user's needs.

Prior work on identifying search-tasks mainly explores task extraction from search sessions [18], wherein the objective is to segment a search session into disjoint sets of queries where each set represents a different task. Recent work on identifying cross-session tasks has targeted pairs of queries, and made predictions about whether they share the same goal or represent the same task [32, 14]. Unfortunately, pairwise predictions alone cannot generate the partition of tasks, and post-processing is needed to obtain the final task partitions [16].

In addition to extracting task clusters, recent efforts by Mehrotra *et al.* [21] have aimed at extracting hierarchies of search tasks and sub-tasks. Li *et al.* [15] model query temporal patterns using a special class of point process called Hawkes processes, and combine topic model with Hawkes processes for simultaneously identifying and labeling search tasks. While some recent research has considered supporting users in their pursuit of complex search tasks by recommending related tasks from a task graph [10], no explicit user models were proposed in their work.

White *et al.* [34] explore the idea of using task information to personalize search result ranking by finding other users performing similar tasks. In this work, we propose a similar yet different notion of using user's task information and propose techniques for learning user representations which could be used in rather diverse application settings.

To the best of our knowledge, this is the first work to consider user's task behavior information to couple varied user information like topical interests and search histories. We next describe our approach to jointly learn task based as well as topic based coupled user representations.

# 3. METHODOLOGY

We propose a new direction in learning user representations by modeling user's task behaviors. We posit that topics and tasks capture different set of insights about user's behavior and information needs and can be coupled with their term usage behavior to jointly learn richer user representations, which is the main goal of this work.

To this end, we intend to extract search tasks from a given search log and represent users in terms of these tasks. In the next sub-section, we describe the approach we use to extract search tasks. This is followed by briefly describing our initial efforts in modeling users based on tasks alone ignoring the topical information [22] in section 4. Finally, we present our approach of coupling task and topical information in Section 5 and extend it to include user's language model and term usage behavior in Section 6. We describe the experimental evaluation set up and results in Section 7, while section 8 concludes.

## 3.1 Notation & Background

We start with defining the notations used throughout the paper. Columns of a matrix are denoted by boldface lower letters with a subscript, e.g., $\mathbf{a_r}$ is the $r-th$ column of matrix $\mathbf{A}$. Entries of a matrix or a tensor are denoted by lowercase letters with subscripts, i.e., $i_1$ entry. Given two matrices $\mathbf{A} \in \Re^{I\times K}$ and $\mathbf{B} \in \Re^{J\times K}$, their Khatri-Rao product is denoted by $A \odot B$ and defined as column-wise Kronecker product. The result is a matrix of size $(IJ) \times K$ and defined by

$$A \odot B = [a_1 \otimes b_1 a_2 \otimes b_2 \dots a_K \otimes b_K] \tag{1}$$

where $\otimes$ denotes Kronecker product. For more details on properties of Kronecker and Khatri-Rao products, the reader is referred to *Kolda et al.* [13].

Table 1 shows a list of symbols used throughout the paper, together with their descriptions.

| Symbol | Description |
|---|---|
| ALS | Alternating Least Squares |
| CMTF | Coupled Matrix Tensor Factorization |
| $\mathbf{A} \odot \mathbf{B}$ | Khatri-Rao product |
| $a \circ b \circ c$ | $(a \circ b \circ c)(i,j,k) = a(i)b(j)c(k)$ |
| $\mathbf{A}_1^i$ | series of matrices or vectors, indexed by i |
| $\Vert A\Vert_F$ | Frobenius norm of $\mathbf{A}$ |
| $\mathbf{T}$ | User-Topic-Task tensor |
| $\mathbf{M}$ | User-Term matrix |
| $\mathbf{U}$ | User representation matrix |
| $\mathbf{S}$ | Search Task matrix |
| $\mathbf{L}$ | LDA topics matrix |
| $\mathbf{W}$ | User language model matrix |

Table 1: Table of symbols

## 3.2 Extracting Search Tasks

In order to build task based representaions of users, we first need to identify and extract search tasks users are likely to perform when they use a search engine. Here we describe our approach of extracting these tasks given a search log. Following the approach in Lucchese *et al.* [18], we employ a graph based query-clustering approach based on finding weighted connected components of a graph.

Given a user session $\phi$, we build a complete graph $G_\phi = (V, E, w)$, whose nodes $V$ are the queries in $\phi$, and whose $E$ edges are weighted by the similarity of the corresponding nodes. The weighting function $w$ is a similarity function $w : E \to R \in [0, 1]$ that can be easily instantiated in terms of the distance functions $\mu$, which we describe a bit later. The graph $G_\phi$ describes the similarity between any pair of queries in the given session. For evaluating similarity between two queries, we make use of the following two similarity features:

- **Content-based**: Two queries that share some common terms are likely related. Sometimes, such terms may be very similar, but not identical, due to mispelling, or different prefixes/suffixes. To capture content distance between queries, following Lucchese *et al.* [18] we adopt a Jaccard index on tri-grams along with a normalized Levenstein distance which is widely accepted as the best edit-based feature for identifying goal boundaries [18].
- **Semantic-based**: Following Lucchese *et al.* [18], we assume that a Wikipedia article describes a certain concept and that the presence of a term in a given article is an evidence of the correlation between that term and that concept. We represent each term in a high-dimensional concept space, and sum over each query term to obtain a query's concept vectors. The cosine similarity between such concept-vectors of queries provides the semantic similarity between the two queries. The distance between two queries is defined as a (1-weighted average of the two similarities). For further details, users are referred to Lucchese *et al.* [18].

Based on the query pair distances obtained above, weak edges with low similarity are dropped, since the corresponding queries are not related, and clusters are built on the basis of the strong edges, i.e. with high similarity, which identify the related query pairs. The connected components of the pruned query-query graph identify the clusters of related queries and provides us with our set of search tasks. Lucchese *et al.* [18] provide further details on the above mentioned similarity features.

# 4. LEARNING TASK BASED USER REPRESENTATIONS

We postulate that in a web search setting, search logs contain information about various actions that users perform and profiling users based on search tasks would better capture the heterogeneity in user information and help us in modeling users. In a recent poster [22], we present some preliminary work which describes a purely search task based user representation system (ignoring topical information) as described in this section. We later propose a novel way of combining such task based representations with user's topical interest information to learn a coupled task-topical interest user profile and additionally incorporate user's term histories via a coupled matrix-tensor factorization framework described in Section 6.

**User-Task Association Matrix:** Based on the extracted search tasks, we construct a user-task association matrix which represents the search tasks users have been involved with. For each user $u_i$, we consider their search history and create a bag-of-queries representation from the list of queries issued by the user and compare each user with each of the search tasks $t_j$ obtained by the method described in section 3.2. For each user-task $<u_i, t_j>$ pair, we populate the corresponding value in the user-task association matrix ($R$) with the cosine similarity score ($r_{ij}$) we obtain for the pair. For tasks in which users do not have any matching queries, we assign a score of 0 to the corresponding pair. The overall motivation behind such a set-up is to capture information about whether or not users have performed such a search task before.

**Probabilistic Matrix Factorization for User Representations:** We wish to extract task-based user vector representations by jointly mapping users and tasks to a joint latent factor space. Following Salakhutdinov *et al.* [23], we model the user-task association in terms of probabilistic matrix factorization problem and learn latent vector representation for each user from the user-task association matrix by fitting a probabilistic model. Given the user-task association matrix $\mathbf{R}$, we find the user feature matrix $\mathbf{U} = [u_i]$ and task feature matrix $\mathbf{T} = [t_j]$. The conditional distribution over the observed user-task associations $R \in \Re^{N \times M}$ is given by:

$$P(R|U,T,\alpha) = \prod_{i=1}^{N} \prod_{j=1}^{M} \left[ \mathcal{N}(R_{ij}|U_i^T T_j, \sigma^2) \right]^{I_{ij}} \quad (2)$$

where $\mathcal{N}$ denotes the Gaussian distribution and $I_{ij}$ is the indicator function which is 1 if the user $i$ was involved in search task $j$. The latent vector representations for the users, the system minimizes the regularized error:

$$min_{u^*,t^* \in \kappa} \sum_{i,j} (r_{ij} - t_j^T u_i)^2 + \lambda(\|u_i\|^2 + \|t_j\|^2) \quad (3)$$

where $\kappa$ is the set of non-zero $r_{ij}$ values, $u_i$ represents the user and $t_j$ represents a task. The user matrix $\mathbf{U}$ obtained as a result, contains vector representations of each of the users which is used in further experiments.

So far, we have been able to extract collective search tasks from all users and learnt a user representation based on these search tasks. We show in Section 7 that task based user models indeed result in better performance than basic bag-of-term based or basic topical interest based representation which further motivates us to investigate combining the two different modalities of user information: topical interests and tasks associations. Indeed, the information carried by user's topical interest profiles and their task profile are different and it would make sense to couple both these informations to jointly learn user profiles. In the next section, we further augment our task based user profiles by incorporating user's topical interest profiles and describe our tensor based approach for the joint model.

# 5. COMBINING SEARCH TASKS WITH TOPICS

Our objective in this section is to build succinct user profiles from the search task information embedded in search logs while at the same time incorporating user's topical interest profiles. Building upon on prior work, we augment our task based user representation model with user's topical information by coupling the topical interest with task based information in the form of a tensor and learning user profiles based on the decomposition of the $< user, topic, task >$ tensor. We first describe the model we use for identifying topical interests of users and further show how we combine this model with task based representation.

## 5.1 Learning Topical Interest Profiles

Topical interests based methods are quite popular in learning user representations [4, 8]. Given user's history of search queries, we aim to develop a topic interest model which captures user's interest distribution over different topics. We make use of the Latent Dirichlet Allocation (LDA) model to learn the latent set of topics embedded in the search log [8]. It is to be noted that LDA topic model based approaches are standard methods to extract user's topical interest profiles and are widely used across user modelling applications.

We hypothesize that each search query is motivated by choosing a topic of interest first and subsequently a query is issued to describe that search need from the catalogue of words consistent with that particular topic. Based on this intuition, we learn an LDA based topic model and use the learnt model to do topical inference for each user to obtain a topic-distribution for the user over the set of learnt topics. We refer to this distribution as a user's ***topical profile***.

## 5.2 Coupling Topics & Tasks

Our main intuition behind leveraging both the topical profile as well as the search task profile of users is to better differentiate between users who share similar topical profiles. Topics and tasks capture different information: topical interest information help in capturing the user heterogeneity resulting from varied interests while task information helps in capturing user heterogeneity resulting from different information needs.

We formulate this intuition in our model by coupling task information with topical information on a per-user basis. We construct a 3-mode tensor $< user, topic, task >$ to jointly capture user's topical as well as search task based information. Next, we briefly describe the tensor formulation.

**Tensors: a primer**

A tensor is a multidimensional array. More formally, a N-way tensor or N-th order tensor is an element of the tensor product of N vector spaces each of which as its own coordinate system. A first-order tensor is a vector, a second-order tensor is a matrix, and tensors of order three or higher are called higher-order tensors. The *order* of a tensor is the number of dimensions, also known as *modes*. A third order tensor can be represented as $T \in \Re^{I_1 \times I_2 \times I_3}$ with each element of the tensor denoted as $t_{i,j,k}$ with $i \in (1, I_1)$,

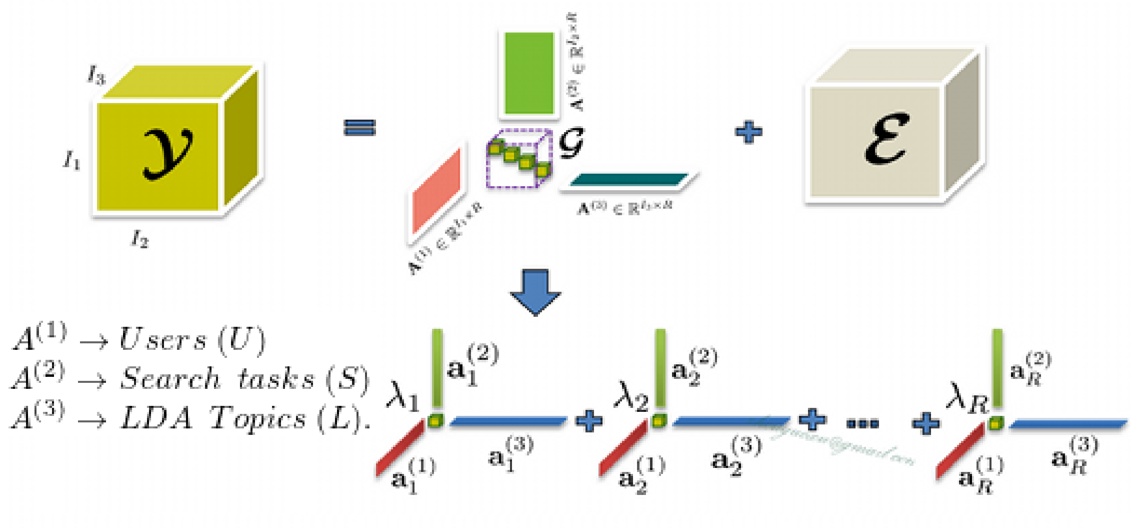


Figure 1: The overview of the user-topic-task tensor constructed by jointly considering user's topical interest profiles alongwith their search task interaction behavior. The tensor decompisiton breaks the tensor into latent factors which encode the complex interactions between the three different modes of the tensor.

$j \in (1, I_2)$ and $k \in (1, I_3)$. The symbol $\circ$ represents the vector outer product.

**Constructing $< user, topic, task >$ Affinity Tensor**
To jointly model the user's topical and task preferences, we construct a 3-mode tensor - users, topics and tasks. Each element of our tensor ($T \in \Re^{I_1 \times I_2 \times I_3}$), $t_{i,j,k}$ defines user $i$'s combined task based and topical preference - a user's participation in a certain task gets weighted by his topical affinity, thereby coupling his task based and topical affinity. More formally, we define each tensor-component value as follows:

$$t_{i,j,k} = U_{i_{topic_j}} \times U_{i_{task_k}} \tag{4}$$

where $U_{i_{topic_j}}$ is user $U_i$'s topical affinity for topic $j$ obtained from the LDA model learnt before while $U_{i_{task_k}}$ represents the task affinity for user $U_i$'s for search task $k$ obtained in earlier the user-task association phase (Section 4). To obtain user's topical affinity estimates ($U_i$), we train an LDA topic model on the entire query collection and use user's historical queries to create user's term profile which is then used for estimating the topic proportions using LDA inference techniques. $I_1, I_2, I_3$ are the different dimensions of the different modes of the tensor - in our case, these represent the number of users, number of topics and the number of search tasks extracted respectively. Thus, for each user we construct his coupled task-topic affinity value and populate the corresponding component in the tensor $T$.

**Tensor Decomposition**
Tensor decomposition methods are regarded as higher-order equivalents to matrix decompositions. The PARAFAC tensor decomposition [30] allows us to leverage connections between the different users across different topics and different search tasks. By PARAFAC, the input tensors are transformed into Kruskal tensors, a sum of rank-one-tensors. Formally, the tensor $T \in \Re^{I_1 \times I_2 \times I_3}$ is decomposed into component matrices $U \in \Re^{I_1 \times d}$, $T \in \Re^{I_2 \times d}$ and $S \in \Re^{I_3 \times d}$ and $d$ principal factors $\lambda_i$ in descending order. Via these, tensor $T$ can be written as a Kruskal tensor by:

$$T \approx \Sigma_{k=1}^{d} \lambda_k \cdot U^k \circ T^k \circ S^k \tag{5}$$

where $\lambda_k$ denotes the k-th principal factor. The goal is to compute a decomposition with $d$-components that best approximates our tensor $T$, i.e., to find

$$min_{\widetilde{T}} \|T - \widetilde{T}\| \tag{6}$$

such that

$$\widetilde{T} = \Sigma_{k=1}^{d} \lambda_k \cdot U^k \circ T^k \circ S^k \tag{7}$$

We make use of the Alternating Least Squares (ALS) approach [13] to solve the above objective - having fixed all but one matrix, the problem reduces to a linear least-squares problem.

Overall, the above formulation helps us to couple user's topical interests with their search task associations and learn a user representation based on this coupled tensor. This tensor decomposition based user modelling approach allows us to use multi-modal user information and leverage insights from each of them while learning user representations.

Similar to other works based on tensors, an important characteristic of the proposed user modelling approach is that this method is generic enough and allows us to plug-in other sources of user information - click models, data from advertisement responses, etc.

## 6. INCORPORATING HISTORICAL BEHAVIOR

One widely used aspect of user behavior that provides especially strong signals for delivering better personalized services is an individual's history of queries and clicked documents. To construct the profiles necessary for personalization, evidence of a user's interests can be mined from observed past behaviors which can be sourced from their short-term (e.g., the current search session) or the long-term (e.g., across many previous sessions) search histories

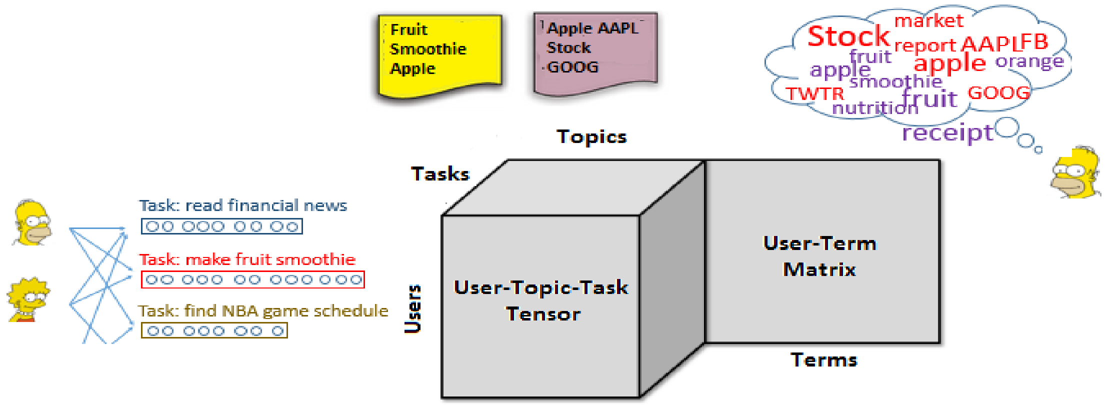


Figure 2: The coupled matrix-tensor obtained by coupling user's term usage behavior matrix with the user-topic-task tensor. The matrix and the tensor share a common mode of 'users'. On the left, we highlight some task related activity of the users and the associated topics obtained and the terms used on the top and right parts of the figure respectively.

[4]. User's term history comprises of the set of terms users used to compose search queries. The tensor based approach described in the previous section looks at utilizing user's topical interest profile along with user's task association information. We hypothesize that additional signals about user's profile could be obtained by jointly modeling user's term usage behavior together with their task and topical interests information.

Overall, our motivation is to combine user's historic term usage behavior with their topical and task based information to learn user representations. We construct a user's term usage behavior over a set of combined vocabulary space. Combining the different users term histories together provides us with a user-term matrix ($W$), which we intend to jointly factorize while performing tensor factorization of the user-topic-task tensor ($T$). The idea behind the coupled matrix-tensor decomposition is that we seek to jointly analyze T and M, decomposing them to latent factors who are coupled in the shared user dimension. More specifically, the first mode of T shares the same low rank column subspace as M; this is expressed through the latent factor matrix U which jointly provides a basis for that subspace.

## 6.1 Coupled Matrix-Tensor Factorization

In the topic-task tensor we described earlier, we have a user by topic by task tensor which encodes user's topical interest profiles and task activities. We also have a semantic matrix which provides additional information for the same sets of users - the user by term matrix. In such cases, we may say that the tensor and the matrix are *coupled* in the *user* mode. Following Acar *et al.* [3], we next describe the joint analysis of a matrix ($M$) and a 3th-order tensor ($T$) with one mode in common, where the tensor is factorized using the CP model and the matrix is factorized by extracting latent factors using matrix factorization.

Let $T \in \Re^{I_1 \times I_2 \times I_3}$ and $M \in \Re^{I_1 \times I_4}$ have the first mode (user) in common; the objective function for coupled analysis is defined by [3]

$$f(U,S,L,W) = \frac{1}{2}\|T - [U,L,S]\|_F^2 + \frac{1}{2}\|M - UW^T\|_F^2 \quad (8)$$

Our goal is to find the matrices U, L, S, W that minimize this objective. In order to solve this optimization problem, we can compute the gradient and then use any first-order optimization algorithm [24]. Rewriting the equation,

$$f(U,S,L,W) = f_1 + f_2 \quad (9)$$

where $f_1 = \|T - [U,L,S]\|_F^2$ and $f_2 = \|M - UW^T\|_F^2$. The partial derivative of $f_1$ with respect to the different matrices has been derived in [2] so we just present the results here. Let $Z = [U,L,S]$, then

$$\frac{\partial f_1}{\partial U} = (Z_i - T_i)U^{(-i)} \quad (10)$$

where $U^{(-i)} = U^{(I_1)} \odot \ldots U^{(i+1)} \odot U^{(i-1)} \odot \ldots \odot U^{(1)}$. Similar computations can be made for the other matrices components $L$ and $S$. The partial derivatives of the second component, $f_2$, with respect to $U, L, S$ and $W$ can be computed as

$$\begin{aligned} \frac{\partial f_2}{\partial U} &= -MW + UW^TW \\ \frac{\partial f_2}{\partial U} &= -W^TU + WU^TU \end{aligned} \quad (11)$$

Combining the above results, the partial derivative of f with respect to factor matrix can be computed as

$$\begin{aligned} \frac{\partial f}{\partial U} &= \frac{\partial f_1}{\partial U} + \frac{\partial f_2}{\partial U} \\ \frac{\partial f}{\partial W} &= \frac{\partial f_2}{\partial W} \end{aligned} \quad (12)$$

Similar computations can be made for the $S$ and $L$ components. With these gradients, the aforementioned coupled matrix-tensor optimization problem can then be solved using any first-order optimization algorithm [3, 24].

On solving the coupled factorization objective[1], we obtain latent factor matrices which could be used as latent representations. More specifically, by making use of the latent factor matrix $U$ we're able to learn user representations that jointly express user's topical, task and term profile information.

[1]We make use of the CMTF toolbox provided by [3]: http://www.models.life.ku.dk/joda/CMTF_Toolbox

| User Profile Information | TermSim | LDA | Task | TT-Tensor | CMTF |
|---|---|---|---|---|---|
| Term History | ✓ | | | | ✓ |
| Topical Interests | | ✓ | | ✓ | ✓ |
| Search Task information | | | ✓ | ✓ | ✓ |

Table 2: User profile information encapsulated in each of the compared approaches. We notice that the proposed TT-tensor and CMFT based methods maximally incorporate the different user profile information available.

# 7. EXPERIMENTAL EVALUATION

In order to evaluate the performance of the proposed user modelling techniques, we use three techniques of evaluation based on collaborative query recommendation, query recommendation based on user groups and user cohort analysis.

## 7.1 Compared Approaches

We consider the following baselines to evaluate the performance of the proposed tensor based method:

- ***TermSim*** (TermSim) is a method that only uses bag-of-words based representation for each user where the terms are extracted from user queries and similar users found using cosine similarity between each user's bag-of-word based representations[20].
- ***LDA Topic Based*** (LDA) is a method of representing users in terms of their topical interests where the topics are extracted via a common Latent Dirichlet Allocation setup [8]. It s important to note that topic based representations are one of the most commonly used representations for personalization.
- ***Task Based (section 4):*** (Task) The first step towards coupling tasks with topics is representing users just in terms of search tasks. We use the user representations obtained in Section 3 as a result of matrix factorization as another baseline to compare the gain in performance obtained as a result of adding the topical aspect on top of user's search task information [22].
- ***TT-Tensor(section 5):*** (TT) Topic-Task Tensor (TT-Tensor) based user representation is the proposed technique which combines user's task information with their topical interests.
- ***CMTF(section 6):*** Coupled Matrix Tensor Factorization (CMTF) [3] based user representation is our second novel contribution which takes into account the user histories in addition to their topical and task based profiles.

Each of the compared approaches work with different user information. In Table 2 we summarize the different modalities of user information used by the different approaches.

## 7.2 Dataset

We make use of the AOL log dataset which consists of ∼20M web queries collected over three months and use data for a subset of ∼1200 users who have issued more than certain threshold (550) number of queries. We run our Task Discovery algorithm on the set of queries for each of these users which results in a total of ∼0.12M tasks which we cluster to obtain a set of 1521 search tasks. Such a setting for task extraction is in line with the original proposed research by Lucchese *et al.* [18]. These tasks are then used to create the user-task association matrix, as described in Section 4 and for constructing the coupled matrix-tensor, as described in Section 6. To make fair comparisons between the topical and task based user profiles, we keep the number of latent factors for tasks same as the number of latent topics.

## 7.3 Collaborative Query Recommendation

A good user profile for query recommendation should capture a user's specific interests and informational needs. Based on this intuition, we evaluate performance of the proposed approach on ***Collaborative Query Recommendation*** [33] where the goal is to recommend queries to a user based on queries issued by similar users. For each user we select the n-most similar users where the similarity is calculated by a cosine similarity score using the user representations learnt. We calculate the weighted frequency of a candidate query for most similar users of the target user $u$, and select the top-k queries as recommendation.

To evaluate the performance of the above mentioned techniques, we consider the test-set of queries in the target user as relevant, and computed average number of relevant queries matched in the recommendation query set as the performance metric. The training/test set per user is populated based on a 20% split across all user queries. We use the training set for populating the matrix/tensor while the test set of queries per user for evaluating the quality of the recommended queries. We plot precision@10 and precision@20 values based on the average number of query matches between the recommended set of queries (top-10 (left) and top-20 (right)) and user's own test set of (unseen) queries

**Discussion**

Our results (Figure 3) show that the proposed Topic-Task Tensor based user modelling approach(*TT-Tensor*) and the coupled matrix factorization method (CMTF) performs better than *TermSim* as well as *TaskBased* which demonstrates that combining search task information with user's topical interests thus help us better capture different aspects of user profiles and can serve as potent user modelling tools. Since *TermSim* relies strictly on term matching for measuring user similarities, its coverage is limited: it might not capture insights for the users with too few queries or those who shared the same search interest but issued different queries or performed different tasks. Task based user modelling can help in better differentiating between users which have similar topical interests but perform different tasks.

The proposed tensor based approach combines the best of both the worlds and hence was able to leverage the topical user profile information with the task aspect. Additionally, the CMTF model combines information from all available data modalities and learns a joint user representation. We see that the CMTF model outperforms the other methods which highlights the importance of jointly considering user's term, topic and task information. On analysis of the dataset, we figured out that the overall lower average query recall values can be attributed to the less query overlap between users, i.e., the upper limit of common query among users is indeed low on average.

## 7.4 Cohort based Query Recommendation

It is well-known that preferences across a user population often decompose into a smaller number of communities of

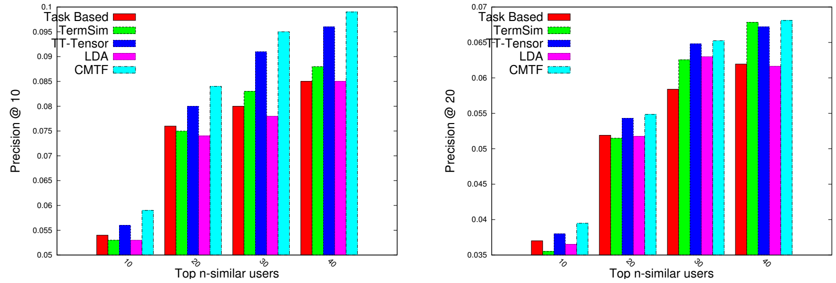


Figure 3: Performance on Collaborative Query Recommendation (left figure: Precision@10 & right figure: Precision@20). Based on the average number of query matches between the recommended set of queries and user's own test set of (unseen) queries, the precision at 10 and precision at 20 values are plotted against the number of similar users considered ($n$). The results obtained at n=10, 20, 30 (left) and n=10, 20 (right) were statistically significant ($p<0.05$) based on pairwise tests between the proposed method and the best performing baseline.

| | **DB Index** | | | | | **SI Index** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **nClusters** | **TermSim** | **LDA** | **Task** | **TT** | **CMTF** | **TermSim** | **LDA** | **Task** | **TT** | **CMTF** |
| 10 | 1.61 | 1.55 | 1.98 | 1.52 | **1.46** | 0.19 | 0.20 | 0.16 | 0.43 | **0.48** |
| 30 | 1.69 | 1.66 | 1.83 | **1.48** | 1.47 | 0.23 | 0.26 | 0.24 | 0.36 | **0.41** |
| 50 | 1.58 | 1.65 | 1.84 | 1.52 | **1.50** | 0.27 | **0.28** | **0.28** | 0.27 | 0.27 |
| 80 | 1.71 | 1.67 | 1.80 | 1.58 | **1.57** | 0.29 | 0.35 | 0.28 | 0.47 | **0.51** |
| 100 | 1.75 | 1.65 | 1.76 | 1.63 | **1.59** | 0.31 | 0.57 | 0.32 | 0.58 | **0.62** |

Table 3: Cluster Analysis of User Representations - cluster evaluation metrics performance for the different approaches are shown. $BoT$ represents the simple Bag-of-Terms baseline, $LDA$ represents the topic model based user representations, *Task* represents user representations learnt via PMF by using task information while $TT$ represents the proposed Task-Topic Tensor based user representations.

commonly shared preferences [1, 25]. In this study, we investigate the performance by means of *groupization*: a variant of personalization whereby other users' profiles can be used to personalize current user's experience. As opposed to finding similar users from the entire user population for collaborative query recommendation, we explore the use of user-cohorts obtained above and leverage information from users belonging to the same cluster to aid in query recommendation. A good cluster should contain better similar users - users who are indeed more representative of the current user. Based on this, we evaluate the performance of the proposed approach on Cohort based Query Recommendation where the goal is to recommend queries to a user based on queries issued by users in the same cluster. Following similar set up as before, we present cohort-based query recommendation results (clustering performed with 10 clusters) in Fig. 4.

**Discussion**

The proposed approach of encapsulating user's historic term usage behavior with their topical and task oriented interests consistently performs better than our baselines in terms of recommending queries from users from the cluster. As can be seen in Fig. 4, the CMTF and coupled task-topic representation performs significantly better right at the start with the difference between the approaches slimming down as we go towards more query recommendations. This is indeed expected since we are measuring precision of queries and eventually not-so-efficient methods will eventually be able to recommend better queries as we increase the number of queries suggested.

Recent research on groupization has focussed on developing different ways of building user cohorts based on topical interests, location, etc [34]. In the present study, we used simple clustering on user features for building cohorts; in future study we intend to compare cohorts of varying sizes and variants of cohort construction techniques to obtain detailed insights on user cohort behaviors.

In addition to performing cohort based query recommendation, we also investigate the *goodness* of the user cohorts we obtain, which were used for query recommendation as described above. We next describe the experimental set-up to analyze the performance of the compared approaches on the task of user cohort formation.

#### *7.4.1 User Cohort Analysis*

We believe that incorporating task behavior of users while learning user representations enables us to better *decompose* users into user cohorts or clusters. In this study, we test the hypothesis that a good user modeling scheme would allow for good cluster formation based on the learnt user representations. We evaluate the user representations learnt in terms of the quality of user clusters formed. Unlike external cluster validation measures, which use external information ("true" cluster membership) not present in our data, internal cluster validation measures only rely on information in the data [17]. In Table 3 and Table 4, we present the cluster validation results on a variety of different metrics, which, to the best of our knowledge, represent a good coverage of the validation measures available in different fields, such as data mining, information retrieval, and machine learning.

The different measures used capture different *goodness* measures of clusters based on inter-cluster and intra-cluster similarities. The Davies-Bouldin index (DB) [6] is calculated as follows. For each cluster C, the similarities between C and all other clusters are computed, and the highest value is assigned to C as its cluster similarity. Then the DB in-

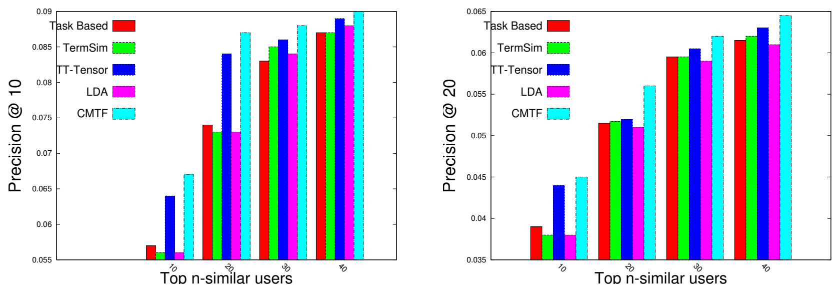


Figure 4: Performance on Cohort Query Recommendation (left figure: Precision@10 & right figure: Precision@20). Based on the average number of query matches between the recommended set of queries and user's own test set of (unseen) queries, the precision at 10 and precision at 20 values are plotted against the number of similar users from user's cluster considered ($n$). The results obtained between the CMTF and the best performing baseline at n=10, 20 (left) and n=10, 20, 30 (right) were statistically significant (p<0.05) based on pairwise tests between the proposed method and the best performing baseline.

| | CH Index | | | | |
|---|---|---|---|---|---|
| **nClusters** | **TermSim** | **LDA** | **Task** | **TT** | **CMTF** |
| 10 | 453 | 643 | 352 | 534 | **658** |
| 30 | 297 | 353 | 203 | 377 | **411** |
| 50 | 213 | 258 | 151 | 285 | **299** |
| 80 | 178 | 192 | 116 | 212 | **234** |
| 100 | 96 | 165 | 99 | 182 | **194** |

Table 4: Cluster Analysis of User Representations - internal cluster evaluation metric (CH Index) performance for the different approaches are shown. *BoT* represents the simple Bag-of-Terms baseline, *LDA* represents the topic model based user representations, *Task* represents user representations learnt vi PMF by using task information while *TT* represents the proposed Task-Topic Tensor based user representations.

dex can be obtained by averaging all the cluster similarities. The smaller the index is, the better the clustering result is.

The Silhouette index (SI) [27] validates the clustering performance based on the pairwise difference of between and within-cluster distances. The Calinski-Harabasz index (CH) [5] evaluates the cluster validity based on the average between and within cluster sum of squares.

**Discussion**

As can be seen in Table 3 and table 4, the user clusters obtained from via using topic-task coupled representations indeed perform better than the clusters obtained via just Bag-of-Terms or task baselines. This is in line with our hypothesis that capturing task behaviors across user populations indeed helps us in forming *well-knit* user clusters and thus could help us perform better in "*groupization*". Having good clusters could be useful for many applications, one of them being collaborative query recommendation, as shown above.

## 8. CONCLUSION

We presented a novel approach to couple user's topical interest information with their search task information and their term usage behavior to learn a joint user representation technique. We demonstrated that coupling user's task information with their topical interests indeed helps us build better user models. We show through extensive experimentation that our task based method outperforms existing query term based and topical interest based user representation methods. This clearly demonstrates the value of considering *search tasks* rather than just query terms or topics during personalization. Future work involves the development of more sophisticated and generalizable models of task behavior that can model task-relevant activity beyond search engine interactions. The flexibility of the tensor based framework makes our proposed method generic enough to add more data sources and modalities. The user representations learnt can be used for various different applications, something we intend to explore as future work.

## 9. REFERENCES

[1] E. Abbasnejad, S. Sanner, E. V. Bonilla, and P. Poupart. Learning community-based preferences via dirichlet process mixtures of gaussian processes. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 1213–1219. AAAI Press, 2013.

[2] E. Acar, D. M. Dunlavy, and T. G. Kolda. A scalable optimization approach for fitting canonical tensor decompositions. *Journal of Chemometrics*, 25(2):67–86, 2011.

[3] E. Acar, T. G. Kolda, and D. M. Dunlavy. All-at-once optimization for coupled matrix and tensor factorizations. *arXiv preprint arXiv:1105.3422*, 2011.

[4] P. N. Bennett, R. W. White, W. Chu, S. T. Dumais, P. Bailey, F. Borisyuk, and X. Cui. Modeling the impact of short-and long-term behavior on search personalization. In *ACM SIGIR*, 2012.

[5] T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

[6] D. L. Davies and D. W. Bouldin. A cluster separation measure. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (2):224–227, 1979.

[7] Z. Dou, R. Song, and J.-R. Wen. A large-scale evaluation and analysis of personalized search strategies. In *Proceedings of the 16th international conference on World Wide Web*, pages 581–590. ACM, 2007.

[8] M. Harvey, F. Crestani, and M. J. Carman. Building user profiles from topic models for personalised search.

In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 2309–2314. ACM, 2013.

[9] A. Hassan, Y. Song, and L.-w. He. A task level metric for measuring web search satisfaction and its application on improving relevance estimation. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 125–134. ACM, 2011.

[10] A. Hassan Awadallah, R. W. White, P. Pantel, S. T. Dumais, and Y.-M. Wang. Supporting complex search tasks. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 829–838. ACM, 2014.

[11] R. Jones and K. L. Klinkner. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, Napa Valley, California, USA, October 26-30, 2008*, pages 699–708. ACM, 2008.

[12] M. Kellar, C. Watters, and M. Shepherd. A field study characterizing web-based information-seeking tasks. *Journal of the American Society for Information Science and Technology*, 58(7):999–1018, 2007.

[13] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[14] A. Kotov, P. N. Bennett, R. W. White, S. T. Dumais, and J. Teevan. Modeling and analysis of cross-session search tasks. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 5–14. ACM, 2011.

[15] L. Li, H. Deng, A. Dong, Y. Chang, and H. Zha. Identifying and labeling search tasks via query-based hawkes processes. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 731–740. ACM, 2014.

[16] Z. Liao, Y. Song, L.-w. He, and Y. Huang. Evaluating the effectiveness of search task trails. In *Proceedings of the 21st international conference on World Wide Web*, pages 489–498. ACM, 2012.

[17] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu. Understanding of internal clustering validation measures. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 911–916. IEEE, 2010.

[18] C. Lucchese, S. Orlando, R. Perego, F. Silvestri, and G. Tolomei. Identifying task-based sessions in search engine query logs. In *WSDM*, 2011.

[19] B. Ma Kay and C. Watters. Exploring multi-session web tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1187–1196. ACM, 2008.

[20] N. Matthijs and F. Radlinski. Personalizing web search using long term browsing history. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 25–34. ACM, 2011.

[21] R. Mehrotra and E. Yilmaz. Towards hierarchies of search tasks & subtasks. In *Proceedings of the 24th International Conference on World Wide Web Companion*, pages 73–74. International World Wide Web Conferences Steering Committee, 2015.

[22] R. Mehrotra, E. Yilmaz, and M. Verma. Task-based user modelling for personalization via probabilistic matrix factorization. In *Proceedings of the 8th ACM Conference on Recommender Systems (RecSys)*, volume 20, 2014.

[23] A. Mnih and R. Salakhutdinov. Probabilistic matrix factorization. In *NIPS*, 2007.

[24] J. Nocedal and S. J. Wright. *Least-Squares Problems.* Springer, 2006.

[25] A. Postlewaite. Social norms and social assets. *Annu. Rev. Econ.*, 3(1):239–259, 2011.

[26] F. Qiu and J. Cho. Automatic identification of user interest for personalized search. In *Proceedings of the 15th international conference on World Wide Web*, pages 727–736. ACM, 2006.

[27] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[28] M. Speretta and S. Gauch. Personalized search based on user search histories. In *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*, pages 622–628. IEEE, 2005.

[29] S. Sriram, X. Shen, and C. Zhai. A session-based search engine. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 492–493. ACM, 2004.

[30] A. Stegeman and N. D. Sidiropoulos. On kruskalâĂŹs uniqueness condition for the candecomp/parafac decomposition. *Linear Algebra and its applications*, 420(2):540–552, 2007.

[31] J. Teevan, S. T. Dumais, and E. Horvitz. Personalizing search via automated analysis of interests and activities. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 449–456. ACM, 2005.

[32] H. Wang, Y. Song, M.-W. Chang, X. He, R. W. White, and W. Chu. Learning to extract cross-session search tasks. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1353–1364. International World Wide Web Conferences Steering Committee, 2013.

[33] H. Wang, C. Zhai, F. Liang, A. Dong, and Y. Chang. User modeling in search logs via a nonparametric bayesian approach. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 203–212. ACM, 2014.

[34] R. W. White, W. Chu, A. Hassan, X. He, Y. Song, and H. Wang. Enhancing personalized search by mining and modeling task behavior. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1411–1420. International World Wide Web Conferences Steering Committee, 2013.