Using Temporal Bayesian Networks to Model user Profile Evolution

Farida Achemoukh[1], Rachid Ahmed-Ouamer[2]
[1] LARI Laboratory, Department of Computer Science, University of Mouloud Mammeri
15000 Tizi-Ouzou, Algeria
achemoukh.farida @ yahoo.fr
[2] LARI Laboratory, Department of Computer Science, University of Mouloud Mammeri
15000 Tizi-Ouzou, Algeria
ahm_r@yahoo.fr

Journal of Digital Information Management

ABSTRACT: *Modeling the user profile can be the first step towards personalization of information search. The user profile refers to his/her interests built across his/her interactions with the information retrieval system. It could be inferred from the recent search history limited to a single search session, during a short period of time to model short term user interests. On the other hand, from the whole search history, to model long term ones stable for a long time. In this paper, we present a personalized information retrieval approach for building and updating the user profile, based on Temporal Bayesian network. The theoretical framework provided by these networks allows better capturing and exploiting the change of user interests over time. Experiments carried out on TREC-1 ad hoc and TREC 2011 session Track collections show that our approach achieves significant improvements over a personalized search approach described in the state of the art and also to a baseline search information process that do not consider the user profile.*

Subject Categories and Descriptors
[H.3.1 Content Analysis and Indexing] [H.3.3 Information Search and Retrieval] [H.3.4 Systems and Software]; User profiles and alert services

General Terms: Bayesian Networks, Information Processing, User Profile

Keywords: Information Retrieval, Personalized Search, User modeling, User Interest, User Profile Evolution, Temporal Bayesian Network

Received: 17 July 2017, Revised 24 August 2017, Accepted 3 September 2017

1. Introduction

Nowadays, everyone searches, manipulates and exploits information. To find information adapted to their needs, users use web search. The web is considered as a privileged information source and in this space; the information in its various forms is growing continuously. Browsing this content becomes a difficult task given the presentation of data that does not meet user's aims and needs. This situation led to the emergence of Information Retrieval Systems (IRS) to improve continually the quality of information access services. Most IRS rely on the so called system-centered approach, behaves as a black box, which produces the same answer to the same search query, independently on the user's needs. But, different users may have different needs expressed via the same search query. For example, without considering the user, it is hard to know which sense "mouse" refers to in query [Shen et al., 2005], [Qiu et al., 2006]. A programmer uses this query to find information about computer devices, while a biologist uses the same query to get information about rodents.

To satisfy user needs, personalization is an appropriate solution to improve the IRS usability. Personalized Information Retrieval Systems (PIRS) rely on the so called user-centered approach, take a step further to better satisfy the user's specific information needs by providing search results that are not only of relevance to a query but are also of particular relevance to the user who submitted the query [Pitkow et al., 2002], [Gauch et al., 2003], [Liu et al., 2004], [Tan et al., 2006], [Teevan et al., 2010], [White et al., 2010].

Modeling the user profile can be the first step towards personalization of information search and retrieval. However, more personalized information retrieval approaches focused on the user profile construction in order to better identify his information needs.

User profile can be deduced explicitly by asking users questions [Ma et al., 2007] or implicitly, by observing their activities [Gauch et al., 2003] [Speretta et al., 2005], [Liu et al., 2010] , [Srinivasa et al.,2016], [Zhou et al.,2016]. It can be represented by a simple structure based on keywords [Shen et al., 2005], or by concept hierarchy issued from the user's documents of interests [Kim et al., 2003], [Speretta et al., 2005]. Or by using external domain ontology as an additional evidence to model the user profile as a set of concepts issued from predefined ontology [Gauch et al., 2003], [Daoud et al., 2009].

One of the main issues of user profile concerns its evolution. In most personalization search systems the profile evolution consists of adapting its content to the variations of the user information needs [Mathijs et al., 2011], [Teevan et al., 2011], [Belo 2016] [Sontag et al., 2012], [Kiseleva et al., 2013], [Febna e al., 2016]. User profile can be short-term or long-term types.

The short-term user profile is limited to a single search session, with consistent user interests for a short period of time. It is inferred from the recent search history [Shen et al., 2012]. Long term user profile holds persistent user interests generally stable for a long time. It is inferred from the whole user search history. Some personalized approaches investigate either short-term ones [Gauch et al., 2003], [Daoud et al., 2009], [Liao et al., 2013] or long term ones [Tan et al., 2006], [Shen et al., 2012] in a personalized document ranking.

In this paper, we present a personalized information retrieval relies on building and updating a user profile in search session, based on Temporal Bayesian network. Search session is defined as a sequence of user's interactions including queries that are related to the same user information need and occur within a unit of time. The notion of user profile considered here is modeled by his/her general interests represented as weighted vectors of terms and built across his/her interactions with the retrieval system. Therefore for each submitted query, we consider the relevant documents selected by the user at his/her interaction with the retrieval system as the data source involved to build his/her interest.

User profile building is intended to improve the relevance of search results that match the user's information needs. Therefore, we propose a variant of bayesian network approach for search personalization performed by integrating the user profile in the retrieval process. In particular, we extend the Bayesian belief network model proposed in [Ribeiro-Neto et al., 1996] to provide a structure for representing a user's interaction and we define the matching measure that integrates the user profile in the retrieval process by interpreting the query-document-user profile relevance as a belief in a document and in a user profile with respect to a query.

The user profile is updated by merging his/ her interests built across his/her interactions over same search session. We use a temporal sequence of Bayesian networks to observe user's interactions, named a Temporal Bayesian network [Dean et al., 1989]. In order to detect related interactions of the same search session, we define a session identification method based on user interest's change. We use the Pearson's correlation measure that tracks changes between user interests over the sequence of user's interactions.

Unlike previously cited work, our approach could be distinguished by several features. First, the user profile is modeled by his/her general interests represented as weighted vectors of terms. We consider the relevant documents selected by the user at his/her interactions with the retrieval system as the data source involved to build his/her interest. To estimate the relevance of document we use a bayesian approach for the matching measure by integrating the user profile as a separate component in the relevance retrieval function. The user profile is represented by a list of concepts issued from an external data source that is domain ontology in [Gauch et al., 2003] and [Daoud et al., 2009] then exploited in the ranking search results by combining the original score between the document and the query with the score between the document and the user profile [Daoud et al., 2009].

Second, we use temporal Bayesian networks to update the user profile and to identify a session by handling changes of user interests over time. Unlike [Murray et al., 2007], [Gayo- Avello 2009], [Gao et al., 2009], [Liao et al., 2013] works, where sessions are identified using time threshold.

The rest of this paper is organized as follows: section 2 gives an overview of related work. In Section 3 we describe our personalized search approach for user profile exploitation and its evolution based on Temporal Bayesian network. Section 4 presents the experimental evaluation and results. In the last section, we present our conclusion and the future work.

2. Related Work

Various approaches have been proposed to adapt and personalize the search results to a given user. Personalization consists of user modeling to build user profile and then its exploitation in the retrieval process. User profile refers to the user interests and can be either short-term, limited to a single search session, or long term, holds persistent user interests generally stable for a long time and may evolve over time.

In the next sections, we first present related work to the

personalization process, namely the user modeling and user profile exploitation in the retrieval process and user profile evolution with session identification.

2.1 User Modeling

A user model describes data that characterizes a user, such data related to user's preferences, goals and interests [Shen e al., 2005], [Sieg et al., 2007], [Micarelli et al., 2007], [Santos- Eugene et al., 2008], [Jenifer et al., 2015], [Srinivasa et al., 2016]. Most of user model approaches represent user profile as one or more vectors of terms [Gowan 2003], [Shen et al., 2005], [Tan et al., 2006]. Others organize user profile as hierarchical concepts structure representing the interest's domains [Gauch et al., 2003], [Kim et al., 2003], [Speretta et al., 2005] or with a structured model of predefined dimensions (personal data, interests, preferences… etc). Works presented in [Micarilli et al., 2007] describe the user profile with two dimensions represented by the interactions history with search system and the user information needs based on his/her interests. Other approaches use external domain ontology as an additional evidence to model user profile as a set of concepts issued from predefined ontology [Gauch et al., 2003], [Daoud et al., 2009].

The construction of the user profile consists of collecting information representing the user. It can be done in two ways; explicitly or implicitly [Micarelli et al., 2007], [Jenifer et al., 2015], [Zhou et al., 2016]. In the explicit approach the user is asked to be proactive and to directly communicate to the system his/her data and preferences [Ma et al., 2007]. However, an explicit request of information to the user implies to burden the user, and to rely on the user's willingness to specify the required information. To overcome this problem, several techniques have been proposed in the literature to automatically capture the user interests by implicit feedback techniques; this is done by monitoring the user's actions in the user system interaction, and by inferring from them the user's preferences. The proposed techniques range from clickthrough data analysis, query log analysis, desktop information analysis, document display time [Speretta et al., 2005], [Agichtein et al., 2006], [Srinivasa et al., 2016].

2.2 Personalization Process

User profile can be exploited before search to reformulate the query or after a search by re-rank the initial results [Micarelli et al., 2007].

Query reformulation consists of initial query expanding with the user profile terms [Koutrika et al., 2005], [Joachims et al., 2007] [Gan et al., 2008]. In [Qiu et al., 2006] user profile is incorporated in the query-document matching model. It consists of computing the document score by considering its relevance to the query and to the user profile.

Most of personalization approaches are based on initial results re-ranking by combining either original rank or score between the document and the query with the rank or score between the document and the user profile [Gowan 2003], [Liu and al., 2010], [Teevan and al., 2011], [Cai et al., 2017].

2.3 User Profile Evolution

In most personalization search systems the profile evolution consists of adapting its content to the variations of the user information needs. User profile can be long term or shortterm. A wide range of personalized IR approaches exploited long term user interests in the search personalization task [Tan et al., 2006],[Matthijs et al., 2011], [Teevan et al., 2011], [Sontag et al., 2012], [Raju et al., 2017]. Another range exploited short-term ones, which could be inferred either from the recent search activities that could hold multiple user information needs [Kiseleva et al., 2013], or from a single search session defined as sequence of related search activities belonging to the same user information need [Shen et al., 2012], [Liao et al., 2013]. Models of short-term user profile based on search queries and result clicks have been used to improve search quality [Daoud et al., 2009], [Xiang et al., 2010], [White et al, 2013], [Shen et al., 2012]. Although longterm models include short-term information, few explicitly study it separately [Bennett et al., 2012]. [Li et al., 2007] modeled short-term and long-term user activities but used different representations for each one. [Dou et al., 2009] learned both click -based and topic -based user models.

2.4 Session Identification

A user may have a single session or multiple sessions during a period of time. The term search session was never formally defined in the literature and its meaning differs in different works. [Sukanya et al., 2014], [Febna e al., 2016] defined a session as a series of queries by a single user made within a small range of time. A session is meant to capture a single user attempt to fill a single information need. The most commonly used session identification method is called timeout. A timeout is the time between two successive search activities, and it is used as a session boundary when it exceeds a certain threshold [He 2000]. This session identification method suffers from its difficulty to set the time threshold. Different users may have different query behaviors, and their time intervals between sessions may be significantly different. Even for the same user, intervals between sessions may vary. Often sessions are identified using a 30 minute timeout [Gao et al., 2009], [Liao et al., 2013]. Other time cutoffs have been proposed, from 5 to 120 minutes [Downey e al., 2007], to a per-user cutoff [Murray et al., 2007]. Nevertheless, the temporal-based approach has several disadvantages: It is unable to detect sessions which are split within a short time, and the long time between searches may not mean that the user's intent changed. The longer pause between queries may be caused by several factors, such as reading long article, breaks, or any other interruptions that are commonly encountered [Czerwinski et al 2004].

Another approach uses lexical distance of the queries.

This idea compares content of two queries to detect if the intent has changed [Jansen et al., 2007]. [Spink et al., 2006] defined a session as a sequence of queries from a web searcher, independent of the information need. [Gayo-Avello 2009] combines both temporal and lexical distances. Other approaches use retrieved documents. [Shen et al., 2005] use vectors of titles and snippets of 50 top-ranked documents for the query. The session split is determined by comparing cosine similarities of two successive queries. [Xiang et al., 2010], [Ustinovskiy et al., 2013] defined a search session as a series of intent-related user's requests to the search engines and used the cosine and jaccard distance between the search results and the query. Another approach uses document keywords. In [Sriram et al., 2004] a session is identified using a semantic similarity measure between successive queries using term relations, which are defined by the number of documents indexed by terms issued from both queries. Method described in [Daoud et al., 2009] extracts document's keywords using the TF.IDF and map them to ODP categories. The session is then split when the ODP category changes.

3. User profile Representation and Evolution Approach

In personalized search one of the main issues is how to infer user profile and how to allow its evolution over time. To address these issues, our general approach for search personalization relies on building and updating a user profile in search session. Then, this user profile will be used in retrieval process. First, the user profile is modeled by his/her general interests learned across his/her interactions with the retrieval system including queries. User interest is built from returned documents judged relevant by the user for a query. It is represented as a vector of weighted terms. The building user profile is used to improve relevant results that match the user information needs. We propose a variant of bayesian network approach for search personalization performed by integrating the user profile in retrieval process. In particular, we extend the Bayesian belief network model proposed in [Ribeiro-Neto et al., 1996] to provide a structure for representing a user interaction and interpreting the query-document-user profile relevance as a belief in a document and in a user profile with respect to a query.

The user profile is updated by merging the user interests built across his/her interactions of the same session. We use a temporal sequence of Bayesian networks to observe these interactions, named a Temporal Bayesian network [Dean et al., 1989]. In order to detect related interactions of the same search session, we define a session identification method based on user interest's change. We use the Pearson's correlation measure that tracks changes between user interests over the sequence of user's interactions.

We summarize below the terminology and notations used in our contribution, then we detail our approach.

3.1 Terminology and Notations

• User's Interaction

A user's interaction with the search system, noted i^n, includes a query submitted by the user, the returned documents and the subset documents judged relevant implicitly or explicitly by the user. It is also characterized by a duration that indicates the time that the user spent on this interaction.

• Search Session

A search session, noted s^m, is defined as a sequence of user's interactions $i^1 \ldots i^n$, related to the same user information need. It is characterized by a length and duration. The length represents the number of successive user's interactions including queries that are related to the same user information need. Duration represents time of all interactions.

• User Profile

User profile refers to the user interests learned across his/her interactions with the retrieval system related to the same search session. A user interest is issued from the relevant documents selected by the user at his/her interaction. It is also represented as a vector of weighted terms, noted $c_k = \{(t_1,w_{1k}), (t_2,w_{2k}), \ldots,(t_i,w_{ik})\}$, where w_{ik} denotes the weight of term t_i in user interest c_k. The weighting term value w_{ik} will be detailed below.

The user profile is initialized by the user interest inferred across the first interaction of the search session. Then it is updated by merging the user interests built across others interactions of the same search session. At time t, the user profile, noted Pr_t, is represented as a vector of weighted terms obtained by merging the previous one Pr_{t-1} and the user interest c_k inferred across the interaction at time t related to the same search session.

3.2 Generating the Short-term user Profile

The process of generating the short-term user profile in a particular search session is detailed in three main steps: (1) building the user interest learned across the first user's interaction at time t of the search session, (2) detecting a possible session when a new query is submitted in a new user's interaction at time $t+1$, (3) updating the user profile when there is no session identification encountered. We notice that to update the user profile in a particular search session, we use a linear function that combines user interests built in user's interactions of the same search session by updating the weights of common terms.

3.2.1 Building a Keyword user Interest

Building the user interest starts by collecting a set of relevant documents D_r returned with respect to a query q related to a user's interaction at time t. Each relevant document is represented as a vector of weighted terms, where the weight w_{ij} of term t_i in document d_j is computed using the TF-IDF weighting scheme:

$$w_{ij} = tf_{ij} \times \log \frac{N}{n_i} \quad (1)$$

Where tf_{ij} is the frequency of term t_i in document d_j, N is the total number of documents and n_i is the number of document that contain term t_i.

The user interest c_k is also represented as a weighted vector of the most relevant terms occurring in the relevant documents judged by the user. The weight w_{ik} of term t_i in user interest c_k is computed as follows:

$$w_{ik} = \frac{1}{|D_r|} \sum_{d_j \in D_r} w_{ij} . \log \frac{(r+0.5)(N-R-n+r+0.5)}{(n-r+0.5)(R-r+0.5)} \quad (2)$$

Where
N and R are respectively the total number of documents and the number of relevant documents to the query belonging to user interest c_k. r is the number of relevant documents that contain term t_i, n the number of documents that contain term t_i.

3.2.2 Session Detection

A search session is defined by a sequence of user's interactions related to the same information need. To determine the search session, we trace changes of user interests across successive user's interactions by using the Pearson's correlation measure to compute the differences between them. Therefore, we compare the user interest c_k learned in current user's interaction I_t at time t with the performed user profile Pr_{t-1} over the session at time t-1, in order to find relationship between them. Here, user interest c_k and user profile Pr_{t-1} are represented as weighted vectors of terms. Hence, the correlation degree, noted $P\_\text{corr}(c_k, Pr_{t-1})$, according to Pearson's correlation mesure is computed as follows:

$$P\_\text{corr}(c_k, Pr_{t-1}) = \frac{\sum_{ti} (w_{ik} - \overline{w_k})(w_{iPr} - \overline{w_{Pr}})}{\sqrt{\sum_{ti} (w_{ik} - \overline{w_k})^2} . \sqrt{\sum_{ti} (w_{iPr} - \overline{w_{Pr}})^2}} \quad (2)$$

Where
w_{ik} and w_{iPr} denote respectively the weight of term t_i in user interest c_k learned in the user's interaction I_t at time t and in the performed user profile Pr_{t-1} at time t -1.

$\overline{w_k} = \frac{1}{N} \sum_{ti} w_{ik}$ and $\overline{w_{Pr}} = \frac{1}{N} \sum_{ti} w_{iPr}$ represent respectively the arithmetic mean of the keyword user interest c_k and the keyword user profile Pr. N is the number of terms.

The correlation value $P\_\text{corr}(c_k, Pr_{t-1})$ is in the range [-1, 1]. Where a value closer to -1 or equal to 0 means that the user interest learned in current user's interaction I_t at time t is not correlated to the performed user profile Pr_{t-1} over the session at time t -1. Therefore the session will be changed. Otherwise a value closer to 1 means that the session is not changed and the short-term user profile will be updated.

3.2.3 Updating the user Profile

The short term user profile is updated based on user interests built across user's interactions of the same search session. The updating method is based on the following principles: (1) enhance the weight of possible common terms that can appear in user interest c_k inferred across the current user's interaction I_t and the performed user profile Pr_{t-1} for the user's interactions of the search session at time t-1; (2) alter the weight of noncommon terms using a decay factor $\alpha \in [0..1]$. The new weight of term t_i in the short term user profile is computed as follows:

$$w_{iPr} = \begin{cases} \alpha * w_{iPr} + (1-\alpha) * w_{ik} & \text{if } ti \in (c_k \text{ and } Pr_{t-1}) \\ \alpha * w_{ik} & \text{otherwise} \end{cases} \quad (4)$$

3.3 A Bayesian belief Network for Search Personalization

To improve relevant results that match the user information needs, we present a personalized information retrieval approach integrating the user profile in the retrieval process. Let us consider a submitted query q related to the user's interaction. Let $D = \{d_1, \ldots d_j, \ldots d_n\}$ the set of documents in the collection, $C\_I = \{c_1, \ldots c_k, \ldots c_m\}$ the set of user interests, and $T = \{t_1 \ldots, t_i, \ldots t_p\}$ the set of index terms used to index these documents and user interests. Furthermore, documents, user interest and query are modeled identically.

The relationship between user interests, documents and query q can be modeled as a Bayesian belief network that provide an effective and flexible framework for modeling distinct sources of evidence in support of a ranking. We propose to extend the Bayesian belief network model proposed in [Ribeiro-Neto et al., 1996] by integrating the user profile to provide a structure for representing a user's interaction and interpreting the query-document-user profile relevance as a belief in a document and in a user profile with respect to a query.

Bayesian belief network is represented by a directed acyclic graph $G(V, E)$, where nodes $V = T \cup D \cup C\_I \cup q$ correspond to the set of random variables and the set of arcs $A = V \times V$ represents conditional dependencies among them.

Figure (Figure 1) shows the topology of our belief network model for user's interaction where the terms nodes represent the network roots.

• Each term in the index terms, $t_i \in T$, is modeled by a random variable $t_i \in \{0, 1\}$. The event of "observing term t_i" is noted $t_i = 1$ or shortly t_i. The complement event that "term t_i is not observed", is noted $t_i = 0$ or shortly $\overline{t_i}$. Let p be the number of index terms present in the set of terms T. It exists 2^p possible term configurations represented by the set θ. A term configuration may represent a query, a document, or a user interest. It is represented by a vector of random variable $\vec{t} = (t_1, t_2, \ldots, t_p)$ where each variable indicates if the corresponding term is observed. For example, an index of 2 terms t_1 and t_2 presents $2^2 = 4$

possible term configurations represented by the set $\theta = \{(t_1, t_2), (t_1, \overline{t2}), (\overline{t1}, t_2), (\overline{t1}, \overline{t2})\}$. The event of observing a particular configuration $\vec{t}=\{t_1, t_2,\ldots,t_p\}$ is noted $\vec{t}$.

• Each document $d_j \in D$ is modeled by a random variable $d_j \in \{0, 1\}$ with two possible values 0 or 1. The event $d_j = 1$, simplified with d_j, denotes that the document d_j is observed. The event $d_j = 0$, simplified with $\overline{d_j}$, denotes that document d_j is not observed. A document d_j is represented as a term configuration $d_j = (t_1, t_2,\ldots,t_p)$ with t_i is a random variable indicating if either term t_i is present in the document or not. Obviously, observing a document in a retrieval process means that this document is relevant to the query.

• Each user interest $c_k \in C\_I$ is modeled by a random variable $c_k \in \{0, 1\}$. The event $c_k = 1$, simplified with c_k, denotes that the user interest c_k is observed. The complement event that "user interest c_k is not observed", is noted $c_k = 0$ or shortly c_k. A user interest $\overline{c_k}$ is represented as a term configuration $c_k = (t_1, t_2,\ldots,t_p)$ with t_i is a random variable indicating if either term t_i is present in the user interest or not. Obviously, observing a user interest means that this user interest is related to the query q.

• A user query q is represented by a random variable $q \in \{0, 1\}$. The two events of observing the query $(q = 1)$ or not observing the query $(q = 0)$ are noted q and $\overline{q}$, respectively. In our case, we interest only to a positive instantiation of q. In the same way as documents and user interests, a query is represented as a term configuration $q = (t_1, t_2,\ldots,t_p)$ with t_i is a random variable indicating if either term t_i is present in the query or not.

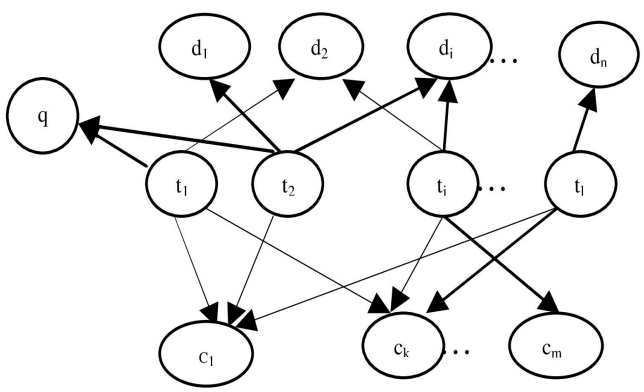


Figure 1. Belief network model for given user's interaction

To express conditional dependencies between random variables, three types of arcs are identified in the inference network model for search personalization: (1) Term to document: Arcs joining term node $t_i \in T$ to document node $d_i \in D$, (2) Term to user interest: Arcs joining term node $t_i \in T$ to user interest node $c_k \in C\_I$, (3) Term to query: Arcs joining term node $t_i \in T$ to user's query node q. Whenever term t_i belongs to document d_j, to user interest c_k and to a query q.

We detail in what follows the query evaluation process for the proposed belief network.

3.3.1 Evaluation Process

Intuitively, we can express the personalization retrieval problem as follows: Given a query q, the search personalization consists in ranking documents according to the information need and user interest. In the network of (Figure 1), the ranking computation is based on interpreting the similarity between a document d_j, a user interest c_k and the query q as an intersection between dj, c_k and q. To quantify the degree of intersection of the document dj, the user interest c_k given the query q, we use the probability $(d_j, c_k \mid q)$. Thus, to compute a ranking, we use Bayes' law and the rule of total probabilities, as follows:

$$P(d_j, c_k \mid q) = \frac{P(d_j, c_k, q)}{P(q)} \qquad (5)$$

As the denominator $P(q)$ is a constant, we can use only the numerator in order to estimate the probability $P(d_j, c_k \mid q)$. Thus the formula (5) is computed as:

$$P(d_j, c_k, q) = \sum_{\vec{t} \in \Theta} P(d_j, c_k, q \mid \vec{t}) \times P(\vec{t}) \qquad (6)$$

The probability $P(\vec{t})$ corresponds to the likelihood of observing term configuration $\vec{t}$. We assume that all the configurations are independent and have an equal probability to be observed. Therefore, the probability $P(d_j, c_k, q)$ is then approximated with:

$$P(d_j, c_k, q) = \sum_{\vec{t} \in \Theta} P(d_j, c_k, q \mid \vec{t}) \times P(\vec{t}) \qquad (7)$$

In the network of (Figure 1), instantiation of the root nodes separates the document nodes, the user interest's nodes and the query node, making them mutually independent, which allows to write:

$$P(d_j, c_k, q \mid \vec{t}) = P(d_j \mid \vec{t}) P(c_k \mid \vec{t}) P(q \mid \vec{t}) \qquad (8)$$

By substituting $P(d_j, c_k, q \mid \vec{t})$ in formula 7, the probability $P(d_j, c_k, q)$ is estimated as:

$$P(d_j, c_k, q) = \sum_{\vec{t} \in \Theta} P(d_j \mid \vec{t}) P(c_k \mid \vec{t}) P(q \mid \vec{t}) \qquad (9)$$

Probabilistic inference in a Bayesian network is NP-Hard [Turtle et al., 1991]. To simplify the computation of probability $P(d_j, c_k, q)$, only instantiated terms in the query q are considered in term configuration $\vec{t}$ and other terms are assumed not effective for document and user interest relevance.

We detail in what follows the computation of the conditional probabilities in formula (9).

• **Probability** $P(q \mid \vec{t})$

$P(q \mid \vec{t})$ determines the probability of generating the query q from term configuration $\vec{t}$. As proposed in [Turtle et al., 1991] the probability $P(q \mid \vec{t})$ is computed using the And–combination:

$$P(q \mid \vec{t}) = \begin{cases} 1 & \text{if } \vec{t} = q \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

• **Probability** $P(d_j \mid \vec{t})$

The probability $P(d_j \mid \vec{t})$ that document d_j is generated by term configuration t is estimated as the similarity between the document d_j and term configuration $\vec{t}$. As described in [Ribeiro et al 1996], a Bayesian network can be used to represent the rankings generated by any of the classic models. For instance, a Bayesian network can be used to compute the vector space model ranking. So, the similarity between the document d_j and term configuration $\vec{t}$ is interpreted as an intersection between document d_j and terms configuration $\vec{t}$ .Then $P(d_j \mid \vec{t})$ is computed as follows:

$$P(d_j \mid \vec{t}) = \frac{\sum_{ti \in dj \wedge \vec{t}} w_{ij} \times w_{it}}{\sqrt{\sum_{ti \in dj} w^2_{ij}} \times \sqrt{\sum_{ti \in \vec{t}} w^2_{it}}} \quad (11)$$

w_{ij}, and w_{it} denote respectively, the weight of term t_i in document d_j and in term configuration $\vec{t}$.

• **Probability** $P(c_k \mid \vec{t})$

Analogously, the similarity between user interest c_k and term configuration $\vec{t}$ is interpreted as the similarity between the user interest c_k and term configuration $\vec{t}$. Then the probability $P(c_k \mid \vec{t})$ is computed as follows:

$$P(c_k \mid \vec{t}) = \frac{\sum_{ti \in ck \wedge \vec{t}} w_{ik} \times w_{it}}{\sqrt{\sum_{ti \in ck} w^2_{ik}} \times \sqrt{\sum_{ti \in \vec{t}} w^2_{it}}} \quad (12)$$

w_{ik} denotes the weight of term t_i in user interest c_k.

Given this latter probabilities, the formula (9) becomes:

$$P(d_j, c_k, q) = \frac{1}{\sum_{ti \in q} w_{iq}^2} \times \frac{\sum_{ti \in q \wedge ck} w_{ik} \times w_{iq} \cdot \sum_{ti \in q \wedge dj} w_{id} \times w_{iq}}{\sqrt{\sum_{ti \in ck} w_{ik}^2} \times \sqrt{\sum_{ti \in dj} w_{ij}^2}} \quad (13)$$

$\frac{1}{\sum_{ti \in q} w_{iq}^2}$ is a constant for a given document and user interest.

Ignoring it, formula (13) is rewritten as follows:

$$P(d_j, c_k, q) = \frac{\sum_{ti \in q \wedge ck} w_{ik} \times w_{iq} \times \sum_{ti \in q \wedge dj} w_{id} \times w_{iq}}{\sqrt{\sum_{ti \in ck} w_{ik}^2} \times \sqrt{\sum_{ti \in dj} w_{ij}^2}} \quad (14)$$

We can use an $m \times n$ matrix X, noted $X_{m,n}$ (m and n indicate respectively the number of user interests and documents) to represent resulting probabilities for each instantiations of document d_j and user interest c_k. It is defined as:

$$X_{m,n} = \begin{array}{c} \\ c_1 \\ c_2 \\ c_k \\ c_m \end{array} \begin{array}{c} \begin{array}{cccc} d_1 & d_2 & d_j & d_n \end{array} \\ \begin{pmatrix} P_{11} & P_{12} & \ldots P_{1j} & P_{1n} \\ P_{21} & P_{22} & \ldots P_{2j} & P_{2n} \\ P_{k1} & P_{k2} & \ldots P_{kj} & P_{kn} \\ P_{m1} & P_{m2} & \ldots P_{mj} & P_{mn} \end{pmatrix} \end{array}$$

Pk_j denotes the probability $P(d_j, c_k \mid q)$ of relevance of user interest c_k and document d_j for a given query q.

We consider that the most likely user interest, noted $\hat{c}$, given a query q, is selected as follows:

$$\hat{c} = \text{Arg max}_{\forall c_k \in C\_I} \frac{1}{n} \sum_{j=1}^{n} P_{kj} \quad (15)$$

Where $\frac{1}{n} \sum_{j=1}^{n} P_{kj}$ represents the arithmetic mean of the set $\{P_{k1}, \ldots, P_{kj}, \ldots P_{kn}\}$ for given user interest c_k.

Therefore, for a given query q and user interest $\hat{c}$, the probabilities P_{kj} presented in matrix $X_{m,n}$ are used to output a ranking list of documents.

3.4 User Profile Evolution

User profile refers to his/her interests and its evolution consists of adapting its content to the variations of the user information needs. It can be short-term or long-term types. A short-term user profile deduced from the user's interactions of a single search session and a long-term user profile reflecting permanents user interests. We propose to apply a temporal Bayesian network to Managing change in user interests over time based on continuous monitoring of interactions. A temporal Bayesian network extends the concept of a Bayesian network to incorporate temporal data. Just as with classic Bayesian network, a static causal model is created to represent a process at a single time slice; multiple copies of this model are then generated for each time slice and links between copies are inserted to capture temporal relations between random variables of different time slices [Friedman e al., 1998], [Kevin et al, 2002].

The proposed network formalism is depicted in figure (Figure 2). It is based on the previously defined Bayesian belief network (Fig.1) with extensions to represent the evolution of the following random variables at time slice t: S_t denotes the search session; I_t denotes the user's interaction and Dut denotes the stay duration of a user's interaction at time slice t. The time slices of the proposed

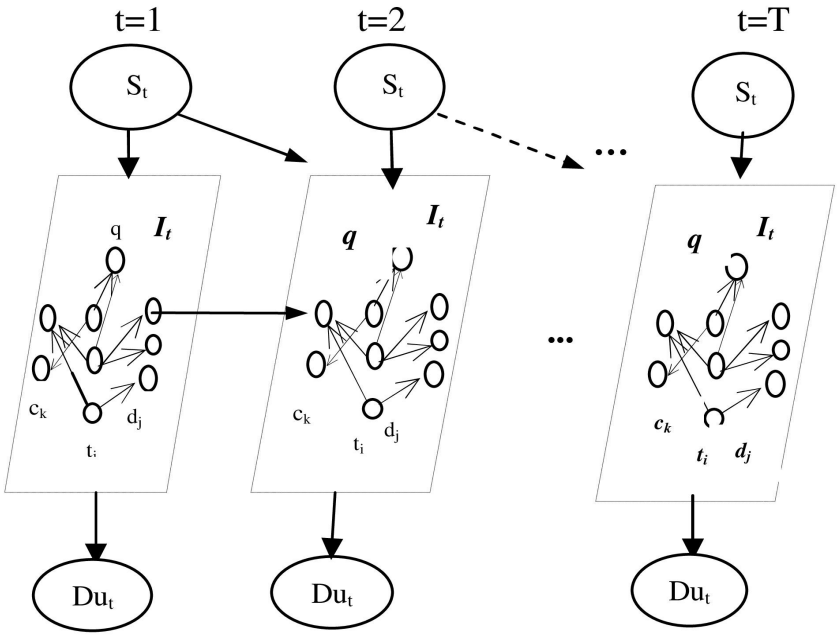


Figure 2. The proposed temporal bayesian network

temporal network are based on time points or periods. Each time point corresponds to user's interaction stay duration.

The proposed network shows that each session displays one or more user's interactions. Links from sessions to user's interactions show the chance of observing a particular user's interaction in the session. The user profile during a search session is updated based on user interests built across user's interactions.

The lower level of the network represents the transitions between user's interactions in one hand and the links between user's interactions and their respective stay durations on the other hand. Furthermore, the user's interaction at time slice t depends on the stay duration at time slice t and the user's interaction at the previous time slice t -1.

We detail in what follows the conditional probabilities distributions of the proposed temporal bayesian network**.** We need to specify the belief of a next user's interaction belonging to a session in order to updating user profile.

Formally, we denote the set of user's interactions as $I = \{i^1, i^2, \ldots i^n\}$ (assuming that there are $\boldsymbol{n}$ user's interactions and each user's interaction contains query q). The set of sessions is denoted as $S=\{s^1, s^2, .., s^m\}$ (assuming there are $\boldsymbol{m}$ sessions) and the set of stay durations denoted as $Du = \{du^1, du^2, \ldots, du^d\}$ (assuming there are $\boldsymbol{d}$ stay durations).

3.4.1 Probabilistic Description

This section describes the session probability $P(S_t)$ at time slice t, for each $\{s^1, s^2, .., s^m\}$, the initial user's interaction probability $P(I_1 \mid S_1)$ at time slice t =1, for each $\{i^1, i^2, \ldots i^n\}$, the user's interaction transition probability $P(I_t \mid I_{t-1}, S_t)$ and the stay duration probability $P\ (Du_t | I_{t,,})$, for each $\{du^1, du^2, \ldots, du^d\}$. Moreover each conditional probability distribution can be represented as a stochastic matrix (i.e. A matrix is called stochastic if each of its row summing to 1). This has the advantage of simplifying the writing of inference calculations.

3.4.1.1 Session Conditional Probability Distributions

$S=\{s^1, s^2, .., s^m\}$ is the set of the $\boldsymbol{m}$ search sessions. The following session distribution is defined for each $s^m \in S$ by:

$$P(S_t=s^m)=v_{t,m} \qquad (16)$$

Where $v_{t,m}$ is the probability that the session s^m is at time slice t. Conditional probability distribution v_t is then a sum-to-one vector of $\boldsymbol{m}$ elements.

3.4.1.2 Initial user's Interaction Conditional Probability Distributions

$I = \{i^1, i^2, \ldots i^n\}$ is the set of the $\boldsymbol{n}$ user's interactions. The initial user's interaction conditional probability distribution allows to describe the initial user's interaction characteristics according to the initial search session, $P(I_1 = i^n \mid S_1 = s^m)$. So for each $i^n \in I$ and each $s^m \in S$:

$$P\ (I_1=i^n|S_1=s^m) = B_{1,m,n} \qquad (17)$$

Where $B_{1,m,n}$ gives the probability to start with user's interaction i^n given the session s^m**.** Conditional probability distribution B_1 is then a matrix consisting of $\boldsymbol{m}$ rows and $\boldsymbol{n}$ columns.

3.4.1.3 User's Interaction Transition Conditional Probability Distributions

The user's interaction transition conditional probability distribution aims to describe how going from one user's interaction to another given the session s^m at time slice t,

$P(I_t=i^n|I_{t-1}=i^{n'}, S_t=s^m)$. It is writing for each $i^n \in I$, $i^n \in I$ and each $s^m \in S$ as:

$$P(I_t=i^n|I_{t-1}=i^{n'}, S_t=s^m) = Q_{n',m,n} \quad (18)$$

Where $Q_{n',m,n}$ gives the probability of transitioning from one user's interaction $i^{n'}$ at time slice t-1 to another user's interaction i^n given the session s^m at time slice t. Thus Q is represented as a stochastic matrix consisting of ***mn*** rows and ***n*** columns.

3.4.1.4 Stay Duration Conditional Probability Distributions

$Du = \{du^1, du^2, ..., du^d\}$ is the set of the ***d*** stay durations. The stay duration conditional probability distribution gives the time spent in a given user's interaction,

$P(Du_t=du^d| I_t=i^n)$. It is writing for each $i^n \in I$ and each $du^d \in Du$ as:

$$P(Du_t=du^d|I_t=i^1) = F_{n,d} \quad (19)$$

Where $F_{n,d}$ is a matrix consisting of ***n*** rows and ***d*** columns. It gives the probability to spend dud time units in the user's interaction i^n.

3.4.2 Probabilistic Inference

The purpose of this section is to evaluate the probability of any sequence of sessions and user's interactions of length T. This in order to handle the user interest's changes over time referred to as search session detection.

Remind that S_t, I_t and Du_t denote respectively the random variables associated to the session, the user's interaction and the stay duration at time slice t, taking their values in domains $S=\{s^1, s^2, .., s^m\}$, $I=\{i^1, i^2, \ldots i^n\}$ and $Du = \{du^1, du^2, ..., du^d\}$, respectively.

Each sequence of length T is defined as subsets of sessions, user's interactions and stay duration, denoted $(\bar{S}, \bar{I}, \overline{Du})_{1\le t\le T}$ where $\bar{S} \subseteq S, \bar{I} \subseteq I$ and $\overline{Du} \subseteq Du$, respectively. Let $\beta_t = (\bar{S}_t, \bar{I}_t, \overline{Du}_t)$ and $\beta = (\beta_t)_{1\le t\le T}$ is a sequence. The probability of an observation sequence β of length T, $P(\boldsymbol{\beta}) \in [0,1]$, is computed as follows:

$$P(\beta) = P(S_1 \in \bar{S}_1, I_1 \in \bar{I}_1, Du_1 \in \overline{Du_1}, ..., S_T \in \bar{S}_T, I_T \in \bar{I}_T, Du_T \in \overline{Du_T}) \quad (20)$$

To simplify the expression (20) we can write:

$$P(\beta) = P(\beta_1, \beta_2, \ldots, \beta_T) \quad (21)$$

The probability $P(\beta)$ allows to extract various information related to the user interest's inferred in user's interactions and their changes over time referred to as search session detection. According to the choice of $\beta=(\beta_t)_{1\le t\le T}$, the probability of a particular sequence $P(\beta)$ is recursively expressed for each $i^n \in I$, each $s^m \in S$ and each $du^d \in Du$ by:

$$P(\beta_t) = \begin{cases} \sum_{s^m \in S} v_{1,m} B_{1,m,n} F_{n,d} & \text{if } t=1 \\ \sum_{i^{n'} \in \beta_{t-1}} P(\beta_{t-1}) \sum_{s^m \in S} v_{t,m} Q_{n',m,n} F_{n,d} & \text{if } t>1 \end{cases} \quad (22)$$

4. Experimental Evaluation

Our experiments have two main objectives. The first one is to evaluate the effectiveness of our personalized search approach. The second one is to evaluate the impact of user profile evolution on the search results.

4.1 Evaluating the Effectiveness of our Personalized Search Approach

Our purpose is to compare the performance of our search personalization approach to the approach proposed by Daoud et al.,(2009). We recall that in our approach, the user profile is integrated in the retrieval process by interpreting the query document-user profile relevance as a belief in a document and in a user profile with respect to a query. In [Daoud et al. 2009] approach, personalization consists of re-ranking the search results by combining query-document score and profile document score.

The experiments have been handled in TREC data set from disk 1& 2 of the TREC ad hoc collections AP88 (Associated Press News, 1988) and WSJ90-92(Wall Street Journal, 1990- 92). Collections contain 741670 documents, queries and relevant judgments. We particularly tested the queries among $q51$ - $q100$.

The choice of this test collection is due to the availability of a manually annotated domain for each query. This allows us, to simulate user interests changing over different domains of TREC. We used the same domain categorization than [Daoud et al 2009] approach.

Table 1 shows six domains of TREC including 25 queries provided by TREC collection.

Domains	queries
Environment	59 77 78 83
International Politics	61 74 80 93 99
International relations	64 67 69 79 100
Law and government	70 76 85 87
Military	62 71 91 92
US Economics	57 72 84

Table 1. TREC domains used for simulating user interests

4.1.1 Experimental Design and Results

The evaluation is based by simulating user interest's process based on N-fold cross validation strategy [Mitchell 1997] explained as follows:

• For each TREC domain, divide the query set into N subsets. We repeat experiments N times, each time using

a different subset as the test set and the remaining N-1 subsets as the training set.

• For each query in the training set, the 1000 top documents are first returned by BM25 Model provided by terrier-3.5 platform then an automatic process uses the returned top documents which are listed in the assessment File (qrels) provided by TREC collections , to generate the user interest vector of weighted terms, using formula (2).

• Then for each query in the test set, an automatic evaluation process (cf. section 3.3.1) generates the matrix given the relevance scores of documents and user interests.

Table 2 shows the percentage of improvement of our approach compared to [Daoud et al., 2009] approach computed at P5, P10 and MAP (Mean Average Precision) and averaged over the queries belonging to the same domain.

We notice that our approach gives higher performance than Daoud et al. (2009) approach for most of the queries in the all domains at P5, P10 and mean average precision (MAP) .

The percentage of improvement (P-Imp) in average precisions between the two approaches presented in table 2 shows that the maximum improvements achieved by our approach are 100% at mean average precision and 125% at P10 and 150% at P5. Based on the overall evaluation results, the conclusion we can made is that the integration of user profile in the matching model of retrieval process as computing the query document- user profile relevance can better improve the search that the re-ranking of search results for a given query using the user profile as done in [Daoud et al 2009].

4.2 Evaluating the Impact of user Profile Evolution on the Search Results

The goal of this experiment is to evaluate the system performance by introducing the user profile over an entire session. It consists of evaluating whether we are able to improve performance for given query by using previous queries of the same session. We compare our approach to the baseline BM25 Model [Robertson et al 1998] provided by terrier-3.5 platform, using only the query ignoring any user profile.

We use a TREC 2011 session Track collection. It consists of clueweb09_English1 collection of documents and includes relevance judgments, 75 query sessions for 61

Domain	Daoud et al. (2009) approach	Our approach	Our approach % Daoud et al. (2009)	Daoud et al. (2009) approach	Our approach	Our approach % Daoud et al. (2009)	Daoud et al. (2009) approach	Our approach	Our approach % Daoud et al. (2009)
	P5	*P5*	*P_Imp*	*P10*	*P10*	*P_Imp*	*MAP*	*MAP*	*P_Imp*
Environment	0,35	0,80	128,5%	0,37	0,70	89,9%	0,19	0,19	00%
Inter. Politics	0,20	0,40	100%	0,16	0,36	**125%**	0,07	0,10	42,8%
Inter. Relations	0,16	0,40	**150%**	0,16	0,32	100%	0,02	0,04	**100%**
Law and Gov.	0,20	0,20	00%	0,10	0,20	100%	0,14	0,19	35,7%
Military	0,35	0,45	28,5%	0,32	0,40	25%	0,07	0,12	71,4%
US Economics	0,33	0,33	00%	0,36	0,40	11,1%	0,10	0,17	70%

Table 2. Performance comparison of the two approaches

```
<session num="1" >
<topic>    <title>peacecorp</title>
<desc>…</desc>        <narr>…</narr>
</topic>
<interaction num="1">   <query>peace corp</query>
</interaction>
<interaction num="2">   <query>peace corp apply</query>
</interaction>
<currentquery>    <query>peace corp application</query>
</currentquery>
</session>
```

Figure 3. Session num="1" on TREC 2011 session Track collection

main queries (topics) (some topics had more than one session corresponding to them). Each topic has a number of subtopics distributed as follows: 202 interactions queries and 75 currents queries. Interactions queries and current query are a sequence of reformulations of the main query.

Table 3 shows the statistics data characteristics of the test collection.

Number of documents	about 50.000.000 documents
Number of Sessions	75 sessions
Number of Main queries (topics)	61 queries
Number of Interactions queries	202 queries
Number of Currents queries	75 queries
Total queries	338 queries

Table 3. Statistics data of the test collection

4.2.1 Experimental Design and Results

The evaluation scenario we adopted is the following:

• **Initial user Profile of a Search Session:** A session is composed of a main query (topic), interactions queries and a current query. For each main query, the 1000 top documents are first returned by BM25 Model provided by terrier-3.5 platform then an automatic process uses the returned top documents which are listed in the assessment File (qrels) provided by TREC to generate the user interest vector of weighted terms, using formula (2). The vector represents the user interest interpreted as the initial user profile of the session.

• **Short-term user Profile Building:** Short- term user profile is built for a given session. Session is characterized by one or more of interaction queries. Each interaction query is interpreted as a user's interaction. An automatic evaluation process generates the short-term user profile evolution for each session by using the formulas (3) and (4).

• **Personalization Process:** It consists of ranking the search results of a current query by using the short-term user profile.

We ignored session where their current queries achieve zero mean average precision (MAP).We present in figure (Figure 4) the precision improvement graph obtained by our approach introduced the user profile of session compared to the baseline BM25 Model [Robertson et al 1998] using only the query ignoring any user profile, at P5, P10, P20 and MAP averaged over the current queries belonging to the different sessions.

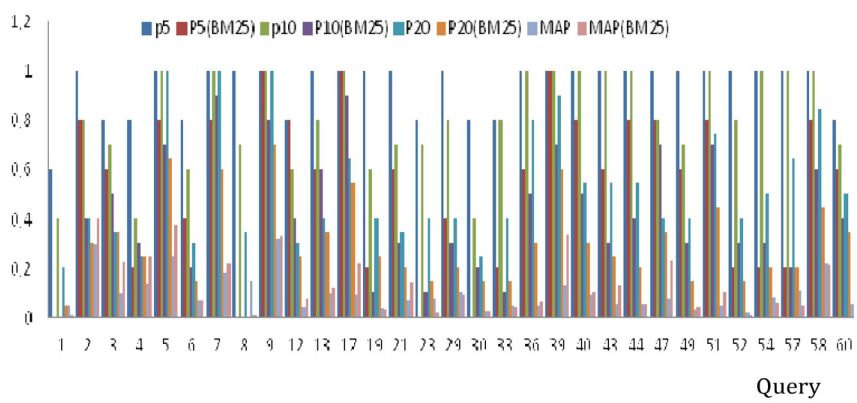
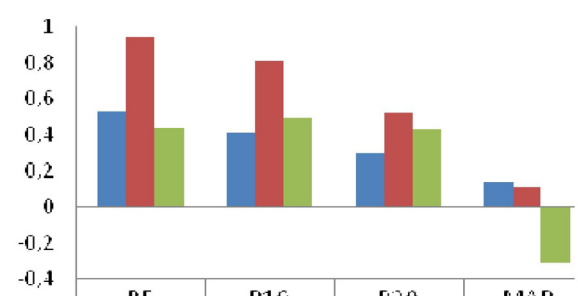


Figure 4. Query by query comparison results between the baseline BM25 and our approach



	P5	P10	P20	MAP
BM25 Model	0,529	0.4096	0,2983	0,135
Our approach	0,9354	0.8064	0,5225	0,103
Our approach % BM25	43,44%	49,21%	42,91%	-31,06%

Figure 5. Performance Comparison between BM25 and our approach

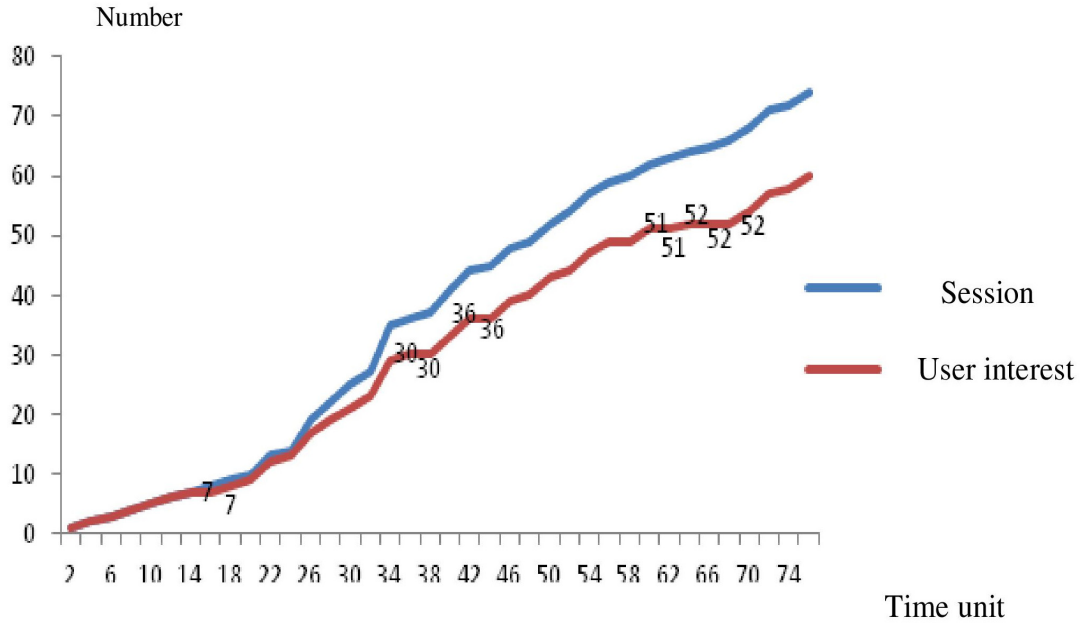


Figure 6. User Interests inferred throughout search sessions

We present in figure (Figure 5) the percentage of improvement of our approach comparatively to baseline BM 25 Model [Robertson et al 1998] computed at P5, P10, P20 and MAP and averaged over the currents queries.

We notice that our approach gives higher performance than BM25 Model at P5, P10 and P20. More particularly, our approach brings an improvement of 43.44% in P5, 49.21% in P10 and 42.91% in P20, but there is a decrease in the mean average precision (MAP). However, these results are acceptable given the values of P5 P10 and P20.

Figure (Figure 6) shows the changes of user interests inferred on several search sessions, defined on successive time slices.

We see several persistent user interest as 7 th, 30 th, 36 th, 51 th and 52 th ones, stable for a long time slices, which correspond respectively to subsets of sessions with the same user information needs. They defined subset of sessions with the same main query.

These results show the ability of our approach to detect persistent user interests on several search sessions about spontaneous user interests.

5. Discussion

Our research work consists first of all on the construction of user profile and its adaptation in a particular search session, and in the second times its integration into the retrieval process. Our intuition was based on the assumption that the search system provides the probability that a document is relevant to a user query, the goal is to estimate this probability by taking into account the user profile. For this purpose, our user profile modeling relies on building and updating a user profile across related user's interactions in the same search session so as to improve search results. Following this general view, our approach could be distinguished by several features in the personalized search community. The first one concerns the user profile integration in the search process and the second one concerns the user profile evolution over a search session.

The first feature of our approach concerns the user profile construction, and its integration in the retrieval process. The user profile is modeled by his/her interests represented as weighted vectors of terms. We consider the relevant documents selected by the user at his/her interactions with the retrieval system as the data source involved to build his/her interest. Then to estimate the relevance of document we use a bayesian approach for the matching measure by integrating the user profile as a separate component in the relevance retrieval function. While in [Gauch et al., 2003] and [Daoud et al., 2009], a user profile is represented by a list of concepts issued from an external data source that is domain ontology and original score between the document and the query with the score between the document and the user profile [Daoud et al., 2009]. The main assumption behind this representation is that we aim at representing the user profile as weighted vector of terms and incorporate it in the query-document matching model both represented as vector of terms using probabilistic approach.

The second feature of our approach concerns the strategy adopted for updating the user profile. We use of temporal Bayesian networks to handle the user interest's changes over time referred to as search session identification from an ongoing tracking of user's interactions. Unlike [Murray et al., 2007], [Gayo-Avello 2009], [Gao et al., 2009], [Liao et al., 2013] works, where sessions are identified using

time threshold, in our approach we used the Pearson's correlation measure based on tracking changes of user interests over the sequence of user's interactions to identify a session.

Therefore the user interest is used to initialize and update the user profile in the search session. Thus, the user profile is updated by merging the user interests built across the user's interactions of the same session. Comparatively to other approaches [Sieg and al., 2007, Rabeb 2016] the user profile evolution is presented by an aggregation and updating of the queries contexts in the search sessions. In [Tamine et al., 2008] the profile evolution consists of an interest dimension evolution based on a rank correlation measure which evaluates the change degree between uses contexts associated with successive periods.

6. Conclusion

In this paper, we have explored our approach for the user profile representation and its evolution in personalized search. It consists of three basic steps: (1) inferring user interest at user's interaction (2) updating the user profile based on user interests related to the user's interactions of the same session (3) incorporating the user profile in the matching model of retrieval process. The user profile refers to the user interests built across his/her user's interactions of a particular session. To handle the user interest changes over time referred to as search session detection, we use a temporal sequence of Bayesian networks representing the sequence of user's interactions.

To evaluate the performance of our approach, we have conducted two experiments, based on using standard test collections in order to allow accurate comparative evaluation. First, to evaluate the effectiveness of our personalized search approach, we use TREC ad hoc collections. We compared our approach to Daoud et al., (2009) approach. In our approach we integrate the user profile in the matching model by interpreting the query-document-user profile relevance as a belief in a document and in a user profile with respect to a query. In Daoud et al., (2009) approach, personalization consists of re-ranking the search results of a given query using the user profile. Moreover, our experimental evaluation shows an improvement of personalized retrieval effectiveness compared to Daoud et al., (2009) approach. Second, to evaluate the user profile evolution impact on the search results, we use clueweb09_English1 test collection and we compared our approach to baseline BM25 Model of the Terrier-3.5 platform, using only the query ignoring any user profile. The obtained results show that our approach gives higher performance than BM25 Model.

As future work, we plan to use long-term user profile in later search sessions to improve the system performance for a recurring query and then undergo experiments in order to evaluate the impact of introducing the user profile in personalizing search results by comparing our approach to another personalized approach.

Acknowledgments

I would like to express my deep gratitude to Pr Mohand Boughanem, Professor at Paul Sabatier University, for the valuable suggestions and comments he has provided in this work.

References

[1] Agichtein, E., Brill, E., Dumais, S., Ragno, R. (2006). Learning user interaction models for predicting Web search preferences. *In*: Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 3–10.

[2] Bennett, P. N., White, R. W., Chu, W., Dumais, S. T., Bailey, P., Borisyuk, F., Cui. X. (2012). Modeling the impact of short-and longterm behavior on search personalization. *In*: Proceedings of the 35th international ACM SIGIR, 185-194.

[3] Cai, F.,Shuaiqiang, W., Maarten, R. (2017). Behavior-Based Personalization in Web Search. *Journal of The Association For Information Science And Technology*, 855–868.

[4] Czerwinski, M., Horvitz, E., Wilhite, S. (2004). A diary study of task switching and interruptions. *In*: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM ,CHI '04, New York, NY, USA, 175-182.

[5] Daoud, M. Tamine-Lechani, L., Boughanem, M., Chebaro, B. (2009). A session based personalized search using an ontological user profile. *In*: SAC '09: Proceedings of the ACM symposium on Applied Computing, New York, 1732–1736.

[6] Dean, T., Kanazawa, K. (1989). A model for reasoning about persistence and causation. *Computational Intelligence*, 142–150.

[7] Dou, Z., Song, R., Wen, J.-R., Yuan, X. (2009). Evaluating the effectiveness of personalized web search. *Knowledge and Data Engineering*, IEEE Transactions on, 21(8), 1178-1190.

[8] Downey, D., Dumais, S. (2007). Models of searching and browsing: *Languages, Studies and Applications*. *In Proc. IJCAI,* 1465-1472, 2007.

[9] Febna V., Abraham A. (2016). User search goals evaluation with feedback sessions. ICETEST, 1256-1262

[10] Friedman, N., Murphy, K., Russell, S. (1998). Learning the structure of dynamic probabilistic net-works». *In*: Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence, 139–147.

[11] Gan.I, Wang, S., Wang, M., Xie, Z., Zhang, L., Shun Z. (2008). Query Expansion based on Concept Clique for Markov Network Information Retrieval Model. Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery, 29-33.

[12] Gao, J., Yuan, W., Deng, X. Li, Nie. J.Y. (2009). Smoothing clickthrough data for web search ranking. In

SIGIR ACM, 355– 362.

[13] Gauch, S., Chaffee, J., Pretschner, A. (2003). Ontology-based personalized search and browsing. *Web Intelligence and Agent Systems*, 219– 234.

[14] Gayo-Avello, D. (2009). A survey on session detection methods in query logs and a proposal for future evaluation. *Inf. Sci.,* 1822-1843 (May).

[15] He. D. (2000). Detecting session boundaries from web user logs. *In*: Proceedings of the BCS-IRSG 22nd Annual Colloquium on Information Retrieval Research, 57–66.

[16] Jansen, B. J., Spink, A, Blakely, C., Koshman, S. (2007). Defining a session on web search engines. *J. Am. Soc. Inf. Sci. Technol*., 862-871.

[17] Jenifer, K., Prabhu, Yogesh., M., Gunasekaran, N. (2015). A survey on web personalization web approaches for efficient information retrieval on user interests. *J. Recent Res. Eng. Technol*, 2349–2260.

[18] Joachims, T., Granka, L., Hembrooke, H., Radlinski F., Gay, G. (2007). Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search, *ACM Transactions on Information systems,* 25 (2) 7.

[19] Kevin, P., Murphy. (2002). Dynamic Bayesian Networks: Representation, Inference and Learning, PhD thesis, U.C. Berkeley.

[20] Kim, H., Chan, P.K. (2003). Learning implicit user interest hierarchy for context in personalization. *In*: IUI '03: Proceedings of the 8th international conference on Intelligent user interfaces, New York, 101–108.

[21] Kiseleva, J., Lam, Thanh, H., Pechenizkiy, M., Calders, T. (2013). Discovering temporal hidden contexts in web sessions for user trail prediction. *In*: Proceedings of the 22nd international conference on World Wide Web companion, 1067-1074.

[22] Koutrika, G., Loannidis, Y. (2005). A unified user profile framework for query disambiguation and personalization. *In*: Proceedings of Workshop on New Technologies for Personalized Information Access, 44-53, (July).

[23] Li, L., Yang, Z., Wang, B., Kitsuregawa, M. (2007). Dynamic adaptation strategies for long term and short-term user profile to personalize search, *In*: Proc. APWeb /WAIM, 228-240.

[24] Liao, Z., Jiang, D., Pei, J., Huang, Y., Chen, E., Cao, H., Li, H. (2013). An approach to context-aware search. *ACM Transactions on the Web* (TWEB).

[25] Liu, J., Belkin, N. J. (2010). Personalizing information retrieval for multi-session tasks: the roles of task stage and task type. SIGIR, 26–33.

[26] Liu, F., Yu, C., Meng, W. (2004). Personalized web search for improving retrieval effectiveness.*IEEE Transactions on Knowledge and Data Engineering,* 16 (1) 28–40.

[27] Ma, Z., Pant, G., Sheng, O. (2007). Interest-based personalized search. *ACM TOIS*, 25 (1).

[28] Mathijs, N., Radlinski, F. (2011). Personalizing web search using long term browsing history, *In:* Proc. WSDM, 25-34.

[29] Micarelli, A., Gaspaetti, F., Sciarrone, F., Gauch, S. (2007). Personalized Search on the World Wide Web. Lecture Notes in Computer Science, 195-230.

[30] Mitchell, T. (1987). Machine Learning. McGraw-Hill Higher Education. ACM, (October).

[31] Murray, C. G., Lin, J., Chowdhury, A. (2006). Identification of user sessions with hierarchical agglomerative clustering, *In:* Proceedings of the American Society for Information Science and Technology, 43 (1) 1-5.

[32] Pitkow, J., Schütze, H., Cass, T., Cooley, R., Turnbull D., Edmonds A., Adar, E., Breuel, T. (2002). A contextual computing approach may prove a breakthrough in personalized search efficiency. ACM p.50-55, (September).

[33] Qiu, F., Cho, J. (2006). Automatic identification of user interest for personalized search. *In:* Proc. of WWW, 727-736 (May).

[34] Raju, Y., Babu, Suresh., D., Anuradha, K. (2017). An Efficient Web Personalization Approach based on Periodic Accessibility and Web Usage Mining, *In:* Advances in Computational Sciences and Technology, 2289-2308.

[35] Ribeiro-Neto, B., Muntz, R. (1996). A belief network model for IR. *In*: Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 253-260.

[36] Robertson, S., Walker, M., Jones, Sparck. (1995). Okapi at trec-3. 28–40,*In*: Second Text Retrieval Conf (TREC-3), 109-126.

[37] Robertson, S. E., Jones, Sparck., K. (1998). Relevance weighting of search terms. 143-160.

[38] Robertson, S. E., Walker, S., Beaulieu, M. (2000). Experimentation as a way of life: Okapi at TREC. *Information Processing & Management*, 36 (1) 95-108.

[39] Shen, S., Hu, B., Chen W., Yang, Q. (2012). Personalized click model through collaborative filtering. WSDM, 323–332.

[40] Shen, X.,Tan, B., Zhai, C. (2005). Implicit user modeling for personalized searchz. *In:* Proc. of CIKM, 43–50.

[41] Sieg, B., Mobasher, B., Burke, R. (2007). Web search personalization with ontological user profiles, *In:* Proceedings of the 16th ACM conference on Conference on information and knowledge management, New York, NY, USA, ACM, 525–534.

[42] Sontag, D., Thompson, Collins., K., Bennet, P. N., White R. W., Dumais, S.T., Billerbeck, B. (2012).

Probabilistic models for personalizing web search, *In*: Proc WSDM, 433-442.

[43] Speretta, M., Gauch, S. (2005). Personalized search based on user search histories. *In*: Proc. of Int Conf on Web Intelligence, p. 622- 628.

[44] Spink, A., Park, M., Jansen, B.,J., Pedersen, J. (2006). Multitasking during web search sessions. *Process. Manage.*, 42 (1) 264– 275.

[45] Srinivasa, R. K., Kalyani, Renuka., A., Krishnamurthy M. (2016). A New Approach of Inferring User Interest for User Search Goals with Web Log Sessions. *Journal of Web Development and Web Designing*, 1-15.

[46] Sriram, S., Shen, X., Zhai, C. (2004). A session-based search engine, *In:* Proceedings of the International ACM SIGIR Conference, 492- 493.

[47] Sukanya, S., Gawade, Gyankamal, J. (2014). Chhajed Using Feedback Sessions for Inferring User Search Goals. *International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE),* 7015-7019, (June).

[48] Tamine, L., Boughanem, M., Zemirli, W. N. (2008). Personalized document ranking: Exploiting evidence from multiple user interests for profiling and retrieval. *Journal of Digital Information Management* 6 (5) 354–365.

[49] Tan, B., Shen, X., Ch., Zhai. (2006). Mining long-term search history to improve search accuracy. KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA. ACM, 718–723.

[50] Teevan, J., Liebling, D., Geetha, G. R. (2011). Understanding and prediction personal navigation, *In*: Proc.WSDM, 85-94.

[51] Teevan, J., Dumais, S. T., Horvitz, E. (2010). Potential for personalization. *ACM Transactions on Computer-Human Interaction* (TOCHI), 17 (1) 4.

[52] Turtle, H. R., Croft, W. B. (1991). Evaluation of an inference networkbased retrieval model, *ACM Transaction on Information Systems* 9 (3) 187–222.

[53] Ustinovskiy, Y., Serdyukov, P. (2013). Personalization of web-search using short-term browsing context. *In:* Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management, 1979-1988.

[54] White, R. W., Bennett, P. N., Dumais, S. T. (2010). Precting short-term interests using activity based search context. *In:* Proc .CIKM, 1009- 1018.

[55] White, R. W., Chu, W., Hassan, A., He, X., Song, Y., Wang, H. (2013). Enhancing personalized search by mining and modeling task behavior. *In*: Proceedings of the 22nd international conference on World Wide Web, 1411-1420.

[56] Xiang, B., Jiang, D., Pei, J., Sun, X., Chen, E., Li, H. (2010). Context aware ranking in web search, *In*: Proc . SIGIR, 451- 458.

[57] Zhou, D., Lawless, S., Xuan, W., Zhao, W., Liu, J. (2016). A study of user profile representation for personalized cross-language information retrieval, 448-477.

[58] Mbarek, Rabeb., Tmar, Mohamed., Boughanem, Mohand (2016). On Improving Pseudo-Relevance Feedback using an Absorbing Document, *International Journal Web Applications*, 8 (1) 8-15.

[59] Belo, Orlando., Rodrigues, Paulo., Barros, Rui., Correia, Helena (2016). A Personalization System for Data Visualization Platforms, *International Journal of Computational Linguistics Research* , 7 (4) 116.