Personalized document ranking: Exploiting evidence from multiple user interests for profiling and retrieval

Lynda Tamine-Lechani, Mohand Boughanem, Nesrine Zemirli
Institut de recherche en informatique de Toulouse SIG-RI
118, route de Narbonne, 31062 Toulouse (France) CEDEX 09
lechani@irit.fr, bougha@irit.fr, nzemirli@irit.fr

ABSTRACT: The goal of personalization in information retrieval is to tailor the search engine results to the specific goals, preferences and general interests of the users. We propose a novel model for both user profiling and document ranking that consider the user interests as sources of evidence in order to tune the accuracy of the documents returned in response to the user query. User profiling is performed by managing the user search history using statistical based operators in order to highlight the user short-term interests seen as surrogates for building the long-term ones. The document ranking model’s foundation comes from influence diagrams which are extension of Bayesian graphs, dedicated to decision-making problems. Hence, query evaluation is carried out as an inference process that aims at computing an aggregated utility of a document by considering its relevance to the query but also the corresponding utility with regard to the user’s topics of interest. Experimental results using enhanced TREC collections indicate that our personalized retrieval model is effective.

Categories and Subject Descriptors
H.3.4 [Systems and software]: User profiles: I.7 [Document and text processing]: Document management

General terms: Document ranking, User Interests

Keywords: Spatial Consistency, Integrity Constraint Checking, Data Quality, Visual Language Parsing, Grammar Formalism, Topological Relationships.

1. Introduction

According to the cognitive view (Ingwersen, 1996), information retrieval (IR) implies a continuous process of interpretation and cognition in context, on both the system side and on the human actor side. This principle of cognitive influence (Ingwersen and Järvelin, 2005) is useful for enhancing the conception of several central concepts in IR: information need, relevance, knowledge acquisition, user interaction etc. Considering the amount of available information in several information environments (like Web, networks, digital libraries), we expect easily the advantages held by the introduction of the cognitive view in IR: (1) short queries are disambiguated (2) high retrieval precision is improved (3) social interaction is encouraged (4) domain-specific retrieval tasks are facilitated.

These are some of the reasons that explain the emergence, since 1990’s, of contextual IR research which claims that IR is a contextual task that aims to enhance the understanding of human needs, activities and intentions in order to deliver more accurate results in response to user queries. Context is determined by various features such as location, task, users’ goals, preferences and interests that have a huge impact on the user’s relevance assessment on the returned information in response to his query. There are two kinds of contexts; the short-term context, which is the surrounding information that emerges from the current user’s information need in a single session. The other kind of context is long-term one which refers generally to the user interests that have been inferred across the past user sessions.

The research presented in this paper focuses on the exploitation of the user interests as the main retrieval context feature to enhance the retrieval effectiveness. Considering the user interests, representing his profile (user’s model) during retrieval, leads more precisely to information personalization that has been investigated by numerous works this last decade (Micarelli et al., 2007; Crestani and Ruthven, 2007). Recent studies suggest furthermore that user’s searches may have multiple goals or topics of interest and occur within the broader context of their information-seeking behaviors (Spink et al., 2004). Research studies also indicate that IR researches often include such multiple topics, during a single session or multitasking search (Spink, 2004). They found that multitasking information seeking is a common behaviour as many users of IR system conduct information seeking on related or unrelated topics.

The objective of our contribution reported in this paper is to highlight the prevalence and the usefulness of the evidence extracted from multiple user interests in order to tune the accuracy of the results presented in response to the query. The main research questions addressed are: (1) how to gather user’s information interaction in order to outline his various topics of interest? (2) how to model a personalized document ranking within a broad variety of topics? In order to answer the first question, we leverage the relevance of information objects (either terms or documents) by the user’s familiarity about the topic addressed by the current topic; we investigate further the use of a statistical based method for tracking significant changes in his information need and adopt a strategy for tuning his profile. The latter is then exploited as a key component of the formal personalized document ranking model, which answers the second question. The underlying idea is that the relevance of a document could be measured by means of the utility of the decision related to present this document or not to the user according to several criteria: term and/or document relevance according to the query, user interest adequacy to the query topic. The theoretical retrieval model we attempt to specify is based on an influence diagram (ID) (Shachter, 1988) which is an extension of Bayesian networks to decision-making problems. To the best of our knowledge, no study has tried to turn contextual retrieval to a decision-making task, incorporating furthermore the user profile as an explicit part of the retrieval model.

The remainder of this paper is organized as follows. After reviewing related works in section 2, we present in section...
3, our approach for user profiling. Section 4 details our personalized document ranking model from both qualitative and quantitative view points. In section 5, we proceed to the validation of our proposed model according to a TREC evaluation protocol. The experimental set up is first described, followed by experimental results that show the effectiveness of our model. Section 6 concludes the paper and outlines future research.

2. An overview on personalized information retrieval

It is well known that a traditional IR process returns the same results for the same query regardless of who submitted the query. In contrast, personalized IR aims to customize search based on specific user interests, goals, preferences or every surrounding feature having an impact on the user’s relevance assessment. Therefore, as information personalization is intended for a wide variety of users with different profiles, it involves two potential challenges: accurately identifying the user profile and then exploiting it to improve document ranking in order to better fit the user’s expectations. The following paragraphs give a synthetic overview of works related to each of these two critical questions.

2.1 User profiling

The related literature reveals that two key points characterize a user profiling approach: (1) information sources and acquisition techniques involved to build the user profile (2) representation model and profile updating algorithms. The acquisition of user knowledge and preferences is one of the most important problems to be tackled in order to provide effective personalized assistance (Micarelli et al., 2007). Numerous works exploit one of the following source of evidence to build the user profile:

1. **user behaviour**, seen as as set of implicit feedback indicators (Kelly, 2004) such as past click history, clickthrough data, browsing features (Teevan and Dumais, 2005; Shen et al., 2005; Gasparetti and Micarelli, 1998), eye-tracking (Joachims et al., 2007),
2. **bookmarks** as clues for predicting user preferences (Gowan, 2003),
3. **desktop informations** (Dumais et al., 2003) and contextual information sources like news sources (Reuters, New York Times), Blog sites and e-commerce sites (Budzik and Hammond, 2000) used for inferring the user interests.

The amount of information extracted from the user’s retrieval environment constitutes a rich repository managed by using data mining (Mobasher, 2007) or machine learning (Webb, 2001) strategies in order to build the user profiles seen as user’s background information like topics of interest, familiarity with the query topic, intent (achieved task), preferences etc. User interests, which is the main feature of the user’s profile, are generally expressed using flat term based vectors or vector classes (Gowan, 2003) or rich semantic structures enhanced with the use of ontologies (Liu and Yu, 2004; Speretta and Gauch, 2005; A.Sieg et al., 2007). Below descriptions of some of the related systems or profiling approaches.

WebPersonae (Gowan, 2003) is a browsing and searching assistant based on web usage mining. The different user interests are represented as clusters of weighted terms obtained by recording documents of interest to the user. (Speretta and Gauch, 2005) build a short term user contextual profile as a weighted ontology. The weight of a concept reflects the degree to which the concept represents the current user’s activities. This weight is computed using a classifier that maps a web page into the concepts of the ontology. The classification

is based on the computation of a similarity measure between web page’s vector visited by the user and each concept vector representation of the ontology. ARCH (Sieg et al., 2007) is a personalized IR system that uses both the user profile, which contains several topics of interest, and a concept hierarchy to enhance the user query. The system represents the long term user context as a set of pairs encapsulating the selected concepts and the deselected concepts that are relevant to the user’s information need across the search sessions. The short term context is pair of the selected and the deselected concepts in the current search session. When a long term user context exceeds a similarity threshold with the short term context, the system updates it by combining it with the short term context.

Comparatively to these previous works, our approach of user profiling has the following new features (Tamine et al., 2006):

1. related and unrelated user interests are dynamically inferred from the search history using a statistical rank-order correlation operator,
2. rather than using a basic Tf (Term frequency)-idf (Inverse document frequency) weighting scheme in the user profile representation, we propose a new measure to estimate the relevance of the words according to the user interests.

In section 4, we present the strategy of collecting and modeling the user’s search history and then we explain how they are used to update the user interests.

2.2 Personalizing document ranking

Document ranking personalization is achieved in practice by incorporating the user’s profile in one of the three main retrieval steps: (1) query refinement (2) document re-ranking (3) document retrieval. Most of the related works enhance the two first steps by exploiting evidence extracted from the user profile. According to the first approach, in (Sieg et al., 2004), personalization is achieved via query reformulation based on information issued from selected and unselected semantic categories. In (Shen et al., 2005), the authors propose context-sensitive retrieval algorithms based on statistical language models. More precisely, their contribution focusses on developing models for using implicit feedback information such as clickthrough history of the current search session to improve the retrieval accuracy. Firstly, they used the clickthrough history to update the query language model and then they compute the KL-divergence between the document language model and the updated query language model leading to the score of the document. Following the second direction, in (Spertett and Gauch, 2005) personalization is carried out by re-ranking the top documents returned to a query using a RSV\(^1\) function that combines both similarity document-query and document-user. In WebPersonae (Gowan, 2003), the relevance of a document is leveraged by its degree of closeness to each of the clusters representing the user interests. In (Liu and Yu, 2004) a user profile consists of a set of semantic categories. Retrieval effectiveness is improved using voting-based merging algorithms that aim to re-rank the documents according to their relatedness to the query categories.

Our approach for personalizing document ranking is different from those previously cited. Our approach attempts to exploit the user profile as an explicit part of the formal retrieval model and not as a source of evidence to re-rank the documents or adapt a basic relevance estimation function. For this aim, we explore the use of ID which are Bayesian probabilistic tools

---

\(^1\) Relevance Status Value
3. Exploiting evidence from multiple user interests for personalizing retrieval

Our research considers the following retrieval setting: a user \( U \) interacts with a document space \( D \) through retrieval sessions that express the successive following events: the user \( U \) submits at time \( s \) a query \( q \) to a search engine; the latter returns a ranked list of documents, then the user expresses his preferences on the documents of interest. We assume that a document returned by the search engine with respect to \( q \) is relevant if it generates some observable user behaviours (reading during a reasonable duration, saving, printing etc) or it is explicitly judged as relevant by the user. In this setting, users have a priori topics of interest \( s \), \( s \), \( s \), \( s \), \( s \) that we attempt first to infer and maintain through an appropriate updating algorithm and then exploit in the retrieval model for a personalization purpose. Following this general view illustrated in figure 1, we detail in the following, our approach for inferring the user interests and personalizing document ranking.

3.1 Inferring the user interests

Our approach uses the evidence collected across successive search sessions in order to track potential changes in the user interests. At time \( s \), the user profile is represented as \( I(s) \) = \( H(s) \cup I(s) \) where \( H(s) \) and \( I(s) \) represent respectively the search history and a set of interests of the user \( U \) at time \( s \). Our method runs into two main steps. The first step consists in representing the user search history by collecting information from the user feedback at each retrieval session, and then gathering this information w.r.t. the user profile. The second step consists in learning the user long-term interests by managing the short-term interests. A user’s short-term interest expresses the user’s information goals achieved within a limited number of search sessions. It is represented using a set of weighted dominant keywords discovered during the previous step. The updating profiling algorithm is based on a correlation measure used to estimate the level of changes in the user interests structure during a period of time.

3.1.1 Building the user profile using the search history

The user search session is represented by a Document-Term matrix \( S' (D' \times T') \) where \( T' \) is the set of terms indexing \( D' \) (\( T' \) is a part of all the representative terms of the previous relevant documents, denoted \( T(R) \)). Because the importance degree of a term within a document depends on the user’s background (a term that is important to one user is sometimes not important to others), we propose to weight terms in a document according to their relevancy to the user profile. More precisely, we address context-based term weighting, focusing on the statistical feature of term occurrence: if a user is not familiar with the topic, he may think general terms related to the topic are important. On the other hand, if a user is familiar with the topic, he may think more detailed terms are important. With this in mind, we improve the accuracy of document-term representation by introducing in the weighting scheme, a factor that reflects the user interest for specific terms. For this purpose, we use term dependencies as association rules checked among \( T' \) (Lin et al., 1998) in order to compute the personalized relevance measure defined as follows:

**Definition 1.** The personalized relevance measure of term \( T \) in document \( D \) at time \( s \) is defined as:

\[
PRM^S (T, D) = \frac{w_{mn}}{dl} \sum_{T' \in T, T' \cap (R')} \text{cooc}(T, T')
\]

where \( w_{mn} \) is the common Tf*Idf weight of term \( T \) in the document \( D \), \( dl \) is the length of document \( D \) (number of
distinct terms in the document), \( \text{cooc}(T,T') \) is the confidence value of the rule \( T \rightarrow T' \) computed using the EMIM measure as: \( \text{cooc}(T,T') = \frac{P(T,T') \log \frac{P(T,T')}{P(T)P(T')}}{P(T')}. \) \( P(T,T') \) is the proportion of documents among \( \mathcal{D}' \) containing both terms \( T \) and \( T' \). \( P(T') \) is the proportion of documents among \( \mathcal{D}' \) containing term \( T \). More a term is cooccurrent with user’s familiar terms, better is its relevance measured with \( \text{PRM}(T,D) \). In contrast, the related value is small if few or any user’s familiar terms are cooccurrent with the term under consideration. The session \( S'(D,T) \) is then determined as \( S' = (\text{PRM})' \).

The user search history is a \( R \times T(K) \) matrix, denoted \( H^r \), built dynamically by bringing document information from the matrix \( S' \) and using an aggregation operator defined below.

**Definition 2.** The operator \( \oplus \) is defined as follows:

\[
H^r(D,T) = S'(D,T)
\]

\[
H^r(D,T) \oplus S'(D,T) = \begin{cases} 
\alpha \cdot w_{ij} \cdot 0 \cdot \alpha \cdot w_{ij} \cdot S'(D,T) & \text{if } r \in R(K') \\
\alpha \cdot w_{ij} \cdot (1-\alpha) \cdot S'(D,T) & \text{if } r \in (K') \text{ and } d \in R(K') \\
0 & \text{ otherwise }
\end{cases}
\]

(2)

For each term, the operator \( \oplus \) combines its basic term weight in the document and its personalized relevance measure computed across the past search sessions as described above.

**Illustration.**

Consider the following document collection \( D = \{d_1,d_2,d_3,d_4,d_5\} \) and the related term index \( T = \{t_1,t_2,t_3,t_4,t_5\} \). Term frequency of each index term belonging to the documents are given below. \( d_1 = \{t_1,t_2,t_4,t_5\}, d_2 = \{t_2,t_3,t_5\}, d_3 = \{t_1,t_4,t_5\}, d_4 = \{t_2,t_3,t_5\}, d_5 = \{t_1\} \).

Table 1 gives the normalized TF*IDF term weights of each index term over the collection of documents. Table 2 gives the cooccurrence values between terms across the collection.

**1. Building \( S' \) and \( H^r \)**

Let’s assume that at the search session \( S'^r \), the user has judged \( \{d_1,d_2\} \) as relevant. Therefore, \( T(K') = \{t_1,t_2,t_4,t_5\} \). Considering the cooccurrence values given in table 2 and the formula (1), the \( \text{PRM}' \) values of terms in \( \{a,b,c\} \) are presented in table 3. Therefore, \( S' \) is represented by the following matrix:

\[
S'(D,T) = \frac{D}{D} \begin{bmatrix} T_1 & T_2 & T_3 & T_4 & T_5 & T_6 & T_7 \\
0.46 & 0.87 & 0 & 1.48 & 0.44 & 0 & 0.88 \end{bmatrix}
\]

At this initial stage \( H^r = S' \).

**2. Building \( S'^r \) and \( H^r \)**

Let’s assume that at the following search session \( S'^{r+1} \), the documents judged as relevant are \( \{d_1,d_2\} \), hence \( S'^{r+1}(D,T) \) is represented by:

\[
S'^{r+1}(D,T) = \frac{D_2}{D_1} \begin{bmatrix} T_1 & T_2 & T_3 & T_4 & T_5 & T_6 & T_7 \\
0.47 & 0.37 & 0.92 & 0 & 0.55 & 0.23 & 0 \end{bmatrix}
\]

Using the aggregation operator defined with the formula (2), we obtain:

\[
H^{r+1}(D,T) = \frac{D_2}{D_1} \begin{bmatrix} T_1 & T_2 & T_3 & T_4 & T_5 & T_6 & T_7 \\
0.46 & 0.87 & 1.48 & 0.44 & 0 & 0.88 \end{bmatrix}
\]

with \( \alpha = 0.5 \).

**Table 1.** TF-lOld term weights

**Table 2.** Term-term cooccurrence values

**Table 3.** PRM values of relevant terms

The above example highlights the following key points:

1. term importance in the user profile is reinforced for user’s familiar terms; the degree of reinforcement of a term importance depends on the average cooccurrence values with other familiar terms; as example term importance of \( T_1, T_2, T_7 \) have been improved respectively with 19.06%, 15.35% and 9.67% according to the
average cooccurrence values 2.65, 2.16 and 1.01.

2. term importance in the user profile is reduced for new terms (initial stage of learning); as example, the importance of terms belonging to $D_0$ are reduced according to their personalized relevance measure (PRM) assigned at the current session.

3.1.2 Updating the user interests

The process of updating the user interests (long-term interests) is based on a statistical method that tracks the changes in the short-term interests extracted periodically from the search history. A user short-term interest is defined as follows:

**Definition 3.** At learning time $s$, the ranked vector denoted $s \_ i^T$ reflecting a short-term interest, is built using the formula:

$$s \_ i^T(T) = \sum_{d \in R_s} H^s(D, T)$$

$s \_ c^T(T)$ is normalized as follows:

$$s \_ i^n_0(T) = \frac{s \_ i^T(T)}{\sum_{s \_ i^T(T)}}$$

In practice, $s \_ i^T$ is a vector of dominant and relevant terms that emerged at each learning period of time, from the user’s search history. To compute a new short-term interest, the average weight of each term across all the relevant documents (represented in the search history matrix) is computed. This gives a surrogate on the relative importance of the term during the past search sessions. The top ranked terms are then used to express the new short-term interest denoted by $s \_ i^T$.

In order to track the changes in the user interests, we compare the current user’s short-term interest $s \_ i^T$ (extracted at the current learning period) and the previous one $s \_ i^T$ (extracted at the precedent learning period) using the Kendall rank-order correlation operator $\rho(X, Y)$ computed as follows:

$$H^s = S^a \Delta l = \sum \frac{S_{pr}(s \_ i^T) + S_{pr}(s \_ i^T)}{\sqrt{\sum \frac{S_{pr}(s \_ i^T) \sum \frac{S_{pr}(s \_ i^T)}}{\sum \frac{S_{pr}(s \_ i^T) \sum \frac{S_{pr}(s \_ i^T)}}}}$$

where $S_{pr}(s \_ i^T) = \text{Signe}(s \_ i^T(T) - s \_ i^T(T)) = \frac{s \_ i^T(T) - s \_ i^T(T)}{s \_ i^T(T) - s \_ i^T(T)}$

$S_{pr}(s \_ i^T) = \text{Signe}(s \_ i^T(T) - s \_ i^T(T))$.

The coefficient value is in the range [-1, 1], where a value closer to -1 means the short-term interests are not similar and a value closer to 1 means that the short-term interests are highly related to each other. Furthermore, it is approximated by a normal law: $\Delta l \sim LG(0, \frac{2(n^2 + 5)}{3(n-1)}$, $n$ represents in our case $T(E)$.

Based on this coefficient value, we apply the following strategy in order to learn the user interests and so update the set $i^T$:

1. $\Delta l > \sigma$ ( $\sigma$ represents a threshold correlation value), No potential changes in the short-term interests, no information available to update $i^T$;

2. $\Delta l > \sigma$ There is a change in the short-term interests. In this case we gauge the level of change: does this reflect a refinement of a previously prior detected user interest or the occurrence of a new one? In order to answer this question we proceed as follows:

- select $C = \arg \max_{c \_ s^T} (C + s \_ c^T)$
- if $C \_ s \_ i^T > \sigma$ then refine the user interest $C$, update the matrix $H^s$ by dropping the rows representing the least recently documents updated, update consequently $R_s$;
- if $C \_ s \_ i^T < \sigma$ then add the new tracked interest in the library $i^T$, learn a period of time $C$: set $H^{s+1} = S^s$, $S = s$.

The user interests $i^T = \{s \_ C, s \_ C, \ldots, s \_ C\}$ learned in this first step are exploited as part of a personalized information access framework. The model that supports the framework is detailed in the following section.

3.2 The personalized document ranking model

3.2.1 Background

The inspiration and foundation of the present model comes from Bayesian theory. A Bayesian network is a graphical model that encodes probabilistic relationships among variables of interest (Jensen, 2001). A Bayesian network uses qualitative and quantitative components to model and manipulate n-dimensional probability distributions. The qualitative component is carried out through a Directed Acyclic Graph (DAG), $G = \langle V, E \rangle$ where each node in $X_i \in V$ encodes the random variable of interest and $E$ encodes the relationships among these variables. We note $Pa(X)$ the parent set of $X_i \in G$. The quantitative component outlines the estimation of the conditional dependencies among the variables. More precisely, for each variable $Xi \in V$, is attached conditional probability distributions $p(X_i / Pa(X))$ where $pa(X)$ represents any combination of the values of the variables in $Pa(X)$. The inference of new sources of evidence is done using the joint distribution law:

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^n p(X_i / pa(X_i))$$

An ID (Shatchter, 1998) is an extension of the Bayesian probabilistic model for solving decision-making problems. The basis of an ID are probabilities and utilities. Utilities are quantified measures for preference attached to each possible situation (scenario) concerned by the modeled decision-making problem. In practice, an ID is represented by an acyclic DAG containing qualitative and quantitative components.

- **Qualitative component:** there are three types of nodes (chance, decision and utility nodes) and two types of arcs (influence and informative arcs). Chance nodes, usually drawn as circles, represent random variables that are relevant to the decision problem and cannot be controlled. Decision nodes, usually drawn as rectangles, represent variables that the decision maker controls directly. Utility nodes, usually drawn as diamonds, express the preference degree of benefit attached to the consequences derived from the decision process. They are quantified by the utility of each possible combination of their parent nodes.

The influence arcs join chance nodes and express probabilistic dependencies as in Bayesian networks. The influence arcs join chance nodes to decision nodes and express the fact that the benefit depends on the value that these chance nodes take.

- **Quantitative component:** the dependencies between chance nodes, representing random variables, are carried out using classical Bayesian probability distributions. For each utility node $V$, related to a decision node, is attached
a real-valued function $\mu$ over $pa(U)$ specifying for each combination of values for the parents of $V$, a measure expressing the benefit attached to this combination for the decision maker.

Given a particular situation $S$, the diagram evaluation is carried out using an evidence propagation algorithm which aims to determine the decision alternative that will lead to the optimal utility called expected utility, denoted $EU(S)$.

### 3.2.2 Problem specification

Intuitively, the problem of personalizing IR may be expressed basically as follows:

Given a query $Q$, the IR system’s problem is to identify those documents $D$ that are relevant to the information need of the user $U$. From the probabilistic point of view, the IR system’s goal is to find the a posteriori most likely documents for which the probability of relevance of the document $D$, considering the query $Q$ and the user $U$, noted $p(d_i|q,u)$, is highest. By Bayes’ law,

$$p(d_i|q,u) = \frac{p(q|d_i,u)p(d_i|u)}{p(q|u)}$$  \hspace{1cm} (6)

Where $d_i$, $q$ and $u$ are the random variables associated to respectively $D$, $Q$ and $U$.

As the denominator $p(q|U)$ is a constant for a given query and user, we can use only the numerator in order to rank the documents. Thus we define the $RSV$ of a document as:

$$RSV_u(Q,D_i) = p(d_i|q,u)p(d_i|u)$$  \hspace{1cm} (7)

The first term of equation (7) is query dependent reflecting the closeness of the document $D_i$ and the query $Q$ according to the user $U$. The second term, in contrast, is query independent, highlighting the usefulness of the document to the user. This may express the suitability of the document to the whole interests of the user when seeking information. In the case that we state that the user is modelled using a set of topics $\{C_1,C_2,...,C_n\}$, the formula (7) gives:

$$RSV_u(Q,D_i) = p(q|d_i,c_1,c_2,...,c_n)p(d_i|c_1,c_2,...,c_n)$$

where $ck$ refers to a random variable associated to the user’s interest $C_k$.

This formula highlights that:

1. two key conditions are prevalent when computing the relevance of documents : (1) relevance condition that ensures that the selected documents are close to the query, (2) the usefulness condition that ensures that the selected documents are consistent with the user’s topics of interest,

2. maximum likelihood of a document is achieved when maximizing the coverage of the information according to the different topics. The user may choose the degree of relevance to integrate either all or a sublist of topics of interest during the personalization process.

By considering this manner of addressing the information personalization problem in the context of multi-user interests, we are hence attracted by formulating it in a mathematical model based on a utility theory supported by ID wich are extension of Bayesian models. The problem is globally expressed through $ID(D,C_i)$:

- user’s interests variable set $C=\{c_1,c_2,...,c_n\}$ where $u$ is the $u$th topic of interest,
- utility set $\mu=\{\mu_1,\mu_2,...,\mu_v\}$ where $\psi$ is an appropriate aggregation operator that combines evidence values from $(C_1,C_2,...,C_j)$. With respect to the probabilistic view illustrated above, the problem takes form of:

$$RSV_u(Q,D_i) = \psi_{k+\alpha}(\mu(d_i,c_k)p(q|d_i,c_k))$$  \hspace{1cm} (8)

### 3.2.3 Model topology

The proposed network architecture appears in figure 2. From a qualitative point of view, nodes in the graphical component represent different kinds of information expressed by three types of nodes: the chance nodes, the utility nodes and the decision nodes.

The **chance nodes** represent the whole of binary random variables used in our model expressed by the set $V=Q \cup D \cup C \cup U \cup T$ where the set $D=\{D_1,D_2,...,D_n\}$ represent the set of documents in the collection, $C=\{C_1,C_2,...,C_j\}$ represent the set of specific user’s long-term interests, $T=\{T_1,T_2,...,T_n\}$ represent the set of terms used to index these documents and user’s interests and $Q$ represents the user’s query. Each chance node $X$ in the set $V$ takes values in a binary domain $\{0,1\}$ which indicates the positive and the negative instantiation respectively. More precisely, for each document node in $D$, $d_i$ expresses, as in the Turtle model (Turtle and Croft, 1990), that the document $D_i$ has been observed and so introduces evidence in the diagram, all the remaining documents nodes are set to $\bot$ alternatively to compute the posterior relevance. Similarly, $c_k$ and $\top$ express that the context $C_k$ is observed or not observed respectively. For each term node, $t_j$ expresses that term $T_j$ is relevant to a given query, and $\bot$, that term $T_j$ is not relevant to a given query.

The relevance of a term means its closeness to the semantic content of a document. In the domain value of the query $q$ means that the query is satisfied and $\overline{q}$ that it is not satisfied. As only the positive query instantiation is of interest, we only consider $Q=q$.

A **utility node** is attached to each decision node related to present the document by taking into account the user interest. So, we associate for each document $D_i$ and each user interest $C_k$ one utility node. All the values given by the pair $(D_i,C_k)$ are used by a specific utility node in order to compute the global utility attached to the decision to return this document $D_i$ according to the whole user interests.

A **decision node** $R_i$ is associated to each document $D_i$ in the collection. It represents the decision to state that the document $D_i$ is relevant. The node $R_i$ represents a binary random variable taking values in a domain $\text{dom}(R_i) = \{\top,\bot\}$.

**Informative arcs** join each term node $T_j$ to each document node $D_i \in D$ and each user interest node $C_k \in C_i$, whenever $T_j$ belongs to $D_i$ and $C_k$. This simply reflects the dependence between the relevance values of both document, user interests and terms used to index them. There are also arcs which connect each term node to the query node. We note $Pa(.)$ the parent sets for each node in the network:

$$\forall T_j \in T, Pa(T_j) = \{D_i\} \cup \tau(C_k) \land \forall D_i \in D, Pa(D_i) = \emptyset, \forall C_i \in C, Pa(C_i) = \emptyset$$

where $\tau(.)$ and $\tau(.)$ represent the index terms.
Influence arcs specify the influence degree of the variables associated with a decision. More precisely, they join in our model, the decision nodes, the user interest nodes and the document nodes by using an aggregation operator specified below.

\[ p(t_i/d,c_i) = \frac{\text{wtd}(j,i) \cdot \delta_j}{\text{wtd}(l,i) + \sum_{\text{wtd}(l,k)} \delta_k} \quad \text{if } T_j \in \pi(C_i) \]

where \( \text{wtd}(j,i) \) and \( \text{wtd}(l,k) \) are respectively the weights of term \( T_j \) in document \( D_l \) and user interest \( C_k \), \( \delta_j \) and \( \delta_k \) constant values \((0 \leq \delta_j, \delta_k \leq 1)\) expressing the default probability value. We consider in this paper \( \delta_j=0.5 \) and \( \delta_k=0.5 \). We compute

\[ p(t_j/d_j) = 1 - p(t_j/d_j), \quad p(t_i/c_i) = 1 - p(t_i/c_i) \]

3.2.5 Utility value

A utility value expresses the degree of the closeness between the document \( D_l \) and the user interest \( C_k \). We propose the following formula to compute \( \mu(t_i/d_i, c_j) \):

\[ \mu(t_i/d_i, c_j) = \frac{1 + \sum_{T_j \in \pi(C)} \text{nidf}(T_j)}{1 + \sum_{T_j \in \pi(C)} \text{nidf}(T_j)} \]

If there are no terms shared between document \( D_l \) and user interest \( C_k \) then \( \mu(t_i/d_i, c_j) = 1 \) which is the lowest value.

\[ \mu(t_i/d_i, c_j) = \frac{1}{\mu(t_i/d_i, c_j)} \]

3.2.6 Relevance scoring

Following the decision theoretical support of our approach, we propose the following mapping function which ranks the documents according to the ratio between the expected utility of retrieving them and the expected utility of not retrieving them, computed as:

\[ RSV_U \begin{align*}
R & \rightarrow R \\
RSV_U(Q, D) & \rightarrow EU(r_i / d_i)
\end{align*} \]

where \( EU(r_i/d_i) \) is the expected utility of the decision "D is relevant, to be presented" (resp. "D is irrelevant, not to be presented")

\[ EU(r_i/d_i) = \psi \cdot \mu(t_i/d_i, c_j) \cdot p(q/d_i, c_j) \]

By applying the joint law and assuming that terms are independent, \( EU(r_i/d_i) \) is computed as:

\[ EU(r_i/d_i) = \psi \cdot \sum_{\text{q/d}_i} p(q/\theta^q) \prod_{t_i \in \text{q/d}_i} p(\theta^q/t_i) \]

\[ EU(\bar{r}_i/d_i) = \psi \cdot \sum_{\text{q/d}_i} p(q/\theta^q) \prod_{t_i \in \text{q/d}_i} p(\theta^q/t_i) \]

where \( \theta \) represents the whole possible configurations of terms in \( q/d_i \), \( \theta^q \) the s order configuration, \( \theta^q \) the s order configuration for the term \( T_j \) in \( q/d_i \), \( T_{q,c} \) represent the query terms belonging to \( D_l \) or \( C_k \), or both; \( \psi \) an appropriate aggregation operator specified below.

For instance if \( Q \) node is related to nodes \( \{n_1, n_2, n_3, n_4, n_5\} \), the instance \( \theta^1 \) of term \( T_j \) in the first configuration

\[ \theta = \{n_1 n_2\}, \quad \theta^1 = \theta \]

Assuming that terms are independent,

\[ p(\theta^1/d_i, c_j) = p(\theta^1/d_i) \cdot p(\theta^1/c_i) \]
3.2.7 Relevance aggregation

The problem addressed at this level concerns the joint utility estimation of a document according to the whole user interests representing his profile. Assuming that the query may cover one major topic or various sub-topics, we shall specify the aggregation operator \( \psi \) on the basis of the relatedness of the user interests.

- **Hypothesis 1:** User interests are unrelated

In this case, the rank of a document should be high according to the suitable user’s interest and low according to the others. A Possible formulation of the aggregation operator is:

\[
\psi(z_1, z_2, \ldots, z_n) = \text{Max}(z_1, z_2, \ldots, z_n)
\]

where \( z_k = u(r_j / d_i, c_j) p(q / d_i, c_j) \) according to the formula (14), \( u \) is the number of user interests.

- **Hypothesis 2:** User interests are related

The relatedness of user interests implies a possible reinforcement of the information relevance according to the query. This could express in some cases as the presence of subtopics of a general topic as in hierarchical representations. A Possible formulation of the aggregation operator is:

\[
\psi(z_1, z_2, \ldots, z_n) = \sum (z_1, z_2, \ldots, z_n)
\]

Previous experiments (Tamine-Lechani et al., 2007) showed that the maximum operator outperformed the sum operator in the case of related or unrelated domains. However, we outlined that such relatedness has been assumed considering intuitively the main topics addressed by the query domains. We believe that further experimentation using surrogates from data distribution is necessary in order to achieve more reliable conclusions.

4. Empirical validation

Generally, the effectiveness of a traditional IR model is evaluated using the well known recall and precision metrics that allow measuring its ability to select relevant documents at the top. We outline, however, that they focus on topical relevance which is user independent. According to this view, a laboratory evaluation model has been proposed through TREC\(^*\) that provides data collections (document, queries and judgements) allowing comparative evaluation of algorithms, models and techniques in IR. The emergence of personalized IR challenged the effectiveness evaluation laboratory model towards user dependent relevance (Kekalainen, 2005) such as situational and cognitive relevance (Borlund, 2003). In the literature, the challenge has been assumed by two main approaches: user study based evaluation and TREC framework adaptation for the purpose of evaluating a personalized retrieval model. We adopted this latter approach for evaluating our model. We present below our framework evaluation and then describe our experiments for evaluating the query tests according to our model.

4.1. Experimental setup

4.1.1 Data sets

We used a TREC data set from disk 1 and disk 2 of the ad hoc task containing 741670 documents issued from journals\(^2\) Text REtrieval Conference (http://trec.nist.gov/)

like Associate Press (AP) and Wall street journal (WJS). We particularly used the topics among \( q_{100} - q_{100} \) because they are enhanced by the domain meta data that gives the query domain of interest. The collection contains queries addressing 12 domains of interest. We choose randomly six among them: Environment, Law & Government, International Relations, Military, International Economics and U.S. Economics. For each query, we used both title and description fields (td). Below, an example of a topic related to the Military domain:

\(<\text{top}>
<\text{head}> Tipster Topic Description
<\text{num}> Number: 062
<\text{dom}> Domain: Military
<\text{title}> Topic: Military Coups D’etat
<\text{desc}> Description: Document will report a military coup d’etat, either attempted or successful, in any country.
<\text{smry}> Summary: Document will report a military coup d’etat, either attempted or successful, in any country.
</top>

We exploited the domain meta data in order to enhance the TREC data test collection with user interests. In order to map the query domains to realistic and dynamic user interests, we applied the OKAPI algorithm that allows us to build a user interest vector according to the formula:

\[
wtc(i, k) = \log \frac{(r + 0.5) / (R - r + 0.5)}{(n - r + 0.5) / (N - n - R - r + 0.5)}
\]

where \( R \) is the number of relevant documents to the queries belonging to \( C_r \), \( n \) the number of relevant documents containing term \( T_r \), \( N \) is the total number of documents in the collection. For each specific domain tested addressed with \( n \) queries, we built \( n \) different user interests exploited for evaluating the query tests according to our model.

4.1.2 Baseline models

We compared our influence-based personalized retrieval model to a naive Bayesian model (NB) (Turtle and Croft, 1990), and a re-ranking based personalized retrieval model published in (Speretta and Gauch, 2005).

- In (Turtle and Croft, 1990), the relevance of a document according to the query is computed as:

\[
p(q / d) = \sum \sum \sum p(q / \theta_i) \prod_{j \in \theta_i} p(\theta_j / d)
\]

- In (Speretta and Gauch, 2005) the authors proposed to re-rank the documents by their conceptual similarity to produce their conceptual rank. The final rank of a document is obtained using the formula:

\[
\text{FinalRank} = \alpha \cdot \text{ConceptualRank} + (1 - \alpha) \cdot \text{InitialRank}
\]

where ConceptualRank is the document rank obtained by computing the similarity between the document profile and the user profile using the similarity function:

\[
sim(user, d) = \sum wt(l, k) + wt(l, i) \cdot \alpha\text{ is a constant having a value between 0 and InitialRank is the document rank given by the search engine.}
We highlight that according to our framework evaluation, the user profile still be a set of distinct basic terms (not concepts) built via a simulation based on the Okapi algorithm.

4.1.3 Evaluation metrics

We evaluated our personalized retrieval model using two main metrics, namely precision at X documents denoted P@X and Mean Average Precision denoted MAP. Each metric focuses on a particular aspect of retrieval, performed as described below.

- **P@X**: reports the proportion of relevant documents ranked in the top X results. In other words, this metric expresses user satisfaction over the X top documents and this is the main reason why it is the most useful metric to measure the performances of a personalized retrieval model. We used in this paper two variants of this metric P@5 and P@10.
- **MAP**: reports the mean of the precisions computed at each retrieved relevant document. This metric is the most commonly used to express the average performance of an information retrieval system over a set of queries by means of both recall and precision.

For each query, 1000 top documents are retrieved; we give the results evaluation by means of P@5, P@10 and MAP for each query and then averaged over all the test queries belonging to each domain. We outline that the improvement performance over the baseline model is significant if the corresponding value is greater than 5%. Analogously, performance decreasing is significant if the corresponding value is lower than -5%.

4.2 Experimental results

Table 4 shows the retrieval performances obtained with both NB model and our model. We can notice that our personalized IR model is effective and achieve significant performance improvements over the NB model for 5 domains among 6 especially according to the high precision oriented measures P@5 and P@10. We notice also that the MAP is improved significantly for two domains (Law& Government, Military) but decreases significantly for two others (Environment, USEconomics). This is an acceptable result as focus is on high precision retrieval performances namely P@5 and P@10.

We notice that the performances decrease significantly for the domain International Relations. A possible explanation is that this domain covers various topics expressed using a wide vocabulary and so the user interest is not able to suggest useful terms to enhance the query; in contrast, it provides noisy terms. We also outline that the degree of improvement varies from a query to another. This is probably depending, in one hand, on the relatedness between the simulated user interests and the query domain (expressed in our model using a utility measure) and in the other hand, on the performance level of the baseline and the query length.

In order to explore the latter aspect, we classified the test queries according to their length (expressed by their number of distinct terms) and retained only the sets containing at least two queries of the same length in order to achieve reliable conclusions. For those queries, we plotted then the curves presented in figure 3 representing the P@5, P@10 and MAP improvements over the baseline.

The graphs show that the improvements levels are positive and correlated within the different query lengths with a significant degree for P@5 comparatively to P@10 and MAP.

For all the curves, the best performances are achieved for queries having an average length (between 6 and 7,5) and the improvements are more significant for short queries than for longer ones. A possible explanation is that first, our model is able to clarify short queries; as example we achieve an improvement between 20% to 140% for the queries q_6_3, q_6_4 and q_6_5 containing respectively 4, 4 and 5 terms; such queries are probably ambiguous without exploiting evidence from the user’s background. Secondly, the analysis of the results obtained with long queries, having between 8 and 12 distinct terms, suggest that the improvements are less significant especially according to P@10 and MAP because of the degree of query difficulty. A query could be long but difficult like queries q_8_9_ and q_8_10 having worse results at the baseline and slight better ones using our model. In contrast, a long query like q_8_10 could be sufficiently clear and consequently, the baseline results are relevant and no much improvement is achieved using extra knowledge from the user.

We further the validation of our model by comparing its performance to Speretta’s model. Before discussing the results, we outline that the results of Speretta’s model, presented in this paper, are based on a re-ranking similarity measure based on the term level (user interests still be a set of terms) rather than the concept level; furthermore we choose alpha=0.5 rather than alpha=1 as claimed by the authors in (Speretta and Gauch, 2005) because it is the value giving the best results considering various parameters: our inverted file built with our indexing method, our strategy for building the user interests.

<table>
<thead>
<tr>
<th>Environment</th>
<th>Naive Bayes</th>
<th>Our model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P5</td>
<td>P10</td>
</tr>
<tr>
<td>Average Improvement</td>
<td>0,55</td>
<td>0,55</td>
</tr>
<tr>
<td></td>
<td>+45%</td>
<td>+36%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>International-Relations</th>
<th>Naive Bayes</th>
<th>Our model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P5</td>
<td>P10</td>
</tr>
<tr>
<td>Average Improvement</td>
<td>0,10</td>
<td>0,12</td>
</tr>
<tr>
<td></td>
<td>-37%</td>
<td>-7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Law-Government</th>
<th>Naive Bayes</th>
<th>Our model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P5</td>
<td>P10</td>
</tr>
<tr>
<td>Average Improvement</td>
<td>0,50</td>
<td>0,57</td>
</tr>
<tr>
<td></td>
<td>+60%</td>
<td>+13%</td>
</tr>
</tbody>
</table>
Table 4. Retrieval performances per domain

<table>
<thead>
<tr>
<th>Domain</th>
<th>P5</th>
<th>P10</th>
<th>MAP</th>
<th>P5</th>
<th>P10</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Military</td>
<td>62</td>
<td>0.2</td>
<td>0.40</td>
<td>0.33</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>71</td>
<td>1</td>
<td>1.00</td>
<td>0.80</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>91</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>92</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.80</td>
<td>0.60</td>
</tr>
<tr>
<td>Average Improvement</td>
<td>0.30</td>
<td>0.35</td>
<td>0.28</td>
<td>0.50</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+66%</td>
</tr>
<tr>
<td>US. Economics</td>
<td>57</td>
<td>0.40</td>
<td>0.60</td>
<td>0.33</td>
<td>0.60</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>84</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.40</td>
<td>0.20</td>
</tr>
<tr>
<td>Average Improvement</td>
<td>0.20</td>
<td>0.30</td>
<td>0.17</td>
<td>0.50</td>
<td>0.30</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+83% +68% -37%</td>
</tr>
</tbody>
</table>

Figure 4 and figure 5 show the comparison between the Map and P@10 values obtained using the two personalized retrieval models averaged over all the queries belonging to each domain. We observe that our model performs better for five domains among six that what confirms our intuition behind the benefit of integrating the user interests in the retrieval step rather than in the post-retrieval one.
5. Conclusion and further work

In this article, we have presented our general approach for exploiting evidence from multi-user interests to personalizing IR. Two main critical questions have been addressed: user profiling and document ranking. Our user profiling approach is based on statistical methods to gather the search history and track changes in the user interests. Unlike most previous related work, we focus on the updating of the search history representation using the user’s relevance point of view on familiar words, in order to build and update his profile. The user profile represents a key component of our document ranking model. The related Bayesian theoretical support offers a solid foundation for representation of uncertainty about various kinds of information (documents, terms, query, user interests) and dealing accurately, when seeking information, with preferences embedded in the user interests. This model has been endowed with an inference propagation process that allows performing a personalized query evaluation. The experimental evaluation results using an enhanced TREC data test collection show that our model is successful at selecting more relevant documents according to the user interests comparatively to other ranking models.

As future work, we plan to involve a user study to better measuring the impact of the user profiles in real life applications according to our document ranking model. Furthermore, we plan to study new probability estimations and other appropriate aggregation operators in order to combine more accurately the evidence issued from the broad variety of user interests.

6. Acknowledgments

This research was partially supported by the French Ministry of Research and New Technologies under the ACI program devoted to Data Masses (ACI-MD), project MD-33.

REFERENCES


