International Journal of Pattern Recognition and Artificial Intelligence

PERSONALIZED SEARCH BASED ON A MEMORY RETRIEVAL THEORY

FABIO GASPARETTI

Department of Computer Science and Automation
Artificial Intelligence Laboratory
University of "Roma Tre"
Via della Vasca Navale, 79 00146 Rome, Italy
gaspare@dia.uniroma3.it

ALESSANDRO MICARELLI

Department of Computer Science and Automation
Artificial Intelligence Laboratory
University of "Roma Tre"
Via della Vasca Navale, 79 00146 Rome, Italy
micarel@dia.uniroma3.it

Personalization is the ability to retrieve information content related to users' profile and facilitate their information-seeking activities. Several environments, such as the Web, take advantage of personalization techniques because of the large amount of available information. For this reason, there is a growing interest in providing automated personalization processes during the human-computer interaction.

In this paper we introduce a new approach for user modeling, which grounds in the Search of Associative Memory (SAM) theory. By means of implicit feedback techniques, the approach is able to unobtrusively recognize user needs and monitor the user working context in order to provide important information useful to personalize traditional search tools and implement recommender systems. Experimental results based on precision and recall measures indicate improvements in comparison with traditional user models.

Keywords: User modeling; personalized search; conceptual search

1. Introduction

The instant availability of many interesting Web resources is a great opportunity that users can exploit for many of their every day tasks. Nevertheless, traditional search engines based on Information Retrieval (IR) approaches return long lists of ranked documents that users are forced to sift through to find relevant documents. Moreover, the same result list is returned for the same query, regardless of who submitted the query, despite users usually having different needs. For these reasons, the identification of the user's information needs and the personalization of the human-computer interaction are becoming fundamental research topics.

The acquisition of user knowledge and preferences is one of the most important problems to be tackled in order to provide effective personalized assistance. Learning techniques for user modeling can be partitioned by the type of input used to build the profile. Explicit feedback systems rely on direct user intervention, that usually suggests keywords or documents of interests, or answers to questions about his/her needs.

Even though explicit feedback techniques have been shown to improve retrieval performance, some studies have found that these techniques are not able to considerably improve the user model [Whiteet al.,2001], especially if no good interface is provided to manage the model [Wærn,2004]. Users are usually unwilling to spend extra effort to explicitly specify their needs, and are often not able to use those techniques effectively [Anick,2003,Teevanet al.,2005], or might find them confusing and unpredictable [KoenemannandBelkin,1996]. Moreover, research shows that users often start browsing from pages identified by less precise but more easily constructed queries, instead of spending time to fully specify their search goals [Teevanet al.,2004]. Aside from requiring additional time during the seeking processes, the burden on users is high and the benefits are not always clear, therefore the effectiveness of explicit techniques may be limited.

On the one hand, implicit feedback collects information about users while they perform their regular tasks. Basically, it unobtrusively draws usage data by tracking and monitoring user behavior, for example by means of server access logs or query and browsing histories [KellyandTeevan,2003,Claypoolet al.,2001,Chan,2000 Whiteet al.,2002,Teevanet al.,2005,RadlinskiandJoachims,2005]. Implicit and explicit feedback based on the same amount of information, namely snippets from result lists, are reasonably consistent in search engine domains. The fact that implicit feedback is readily available in large quantities overcomes possible bias in users' decisions, sometimes influenced by the trust they have in retrieval functions, and by the overall quality of result sets [Joachimset al.,2005].

Once enough knowledge has been collected, user profiles can be employed to provide personalized results. For instance, Google Personalized Search records a trail of all queries and Web sites the user has selected from the results, building a profile according to that information. During the search activity, the search engine adapts the results according to needs of each user, assigning higher scores to the resources related to what the user has seen in the past. Alternatively, user profiles can take part in a distinct re-ranking phase. Some systems implement this approach on the client-side, where the software connects to a search engine, retrieving query results that are then analyzed locally, e.g., [Pitkowet al.,2002,MicarelliandSciarrone,2004,SperettaandGauch,2005].

Just-in-Time IR (JITIR) [Rhodes,2000] is a further interesting approach for personalized search where the information system proactively suggests information based on a person's working context. Basically, the system continuously monitors the user's interaction with the software, such as typing in a word processor or browsing the Web, in a non-intrusive manner, automatically identifying their information needs and retrieving useful documents without requiring any action by users. The same approach could be employed to provide contextual advertisements to users.

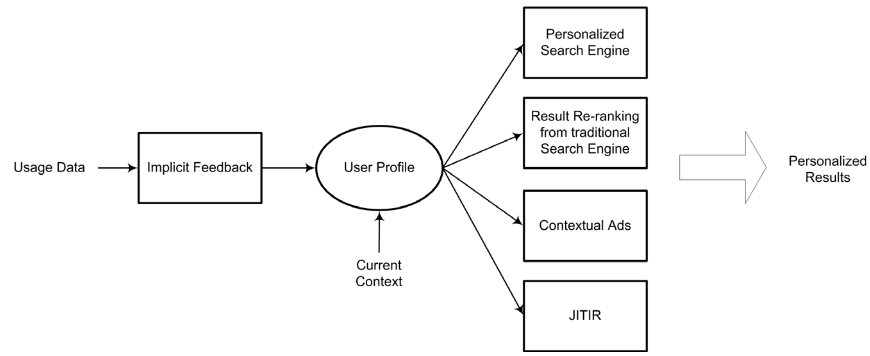

Fig. 1. A user profile based on implicit feedback techniques analyzes usage data, e.g., server access logs or query and browsing histories, in order to build representations of information needs of users. The profile can be employed in search engines, to directly provide personalized results or in distinct re-ranking activities. Contextual advertisements and JITIR are other tasks where user profiles can enhance the result precision. Monitoring the user working contexts allows to provide more updated recommendations related to the users' current activities.

In this work, we propose a user modeling approach based on a cognitive theory of retrieval processes in human memory. The theory allows us to define not only suitable structures to organize concepts, but also how the information is stored and kept updated during the information seeking process. The user model is able to unobtrusively learn user needs by means of implicit feedback techniques. Unlike many traditional user models, the proposed approach takes under consideration the current user context in order to provide personalization related to the user's current activity. Moreover, the user model's learning phase does not need any off-line phase, during which user provides data in order to train the model.

The paper is organized as follows: related work is discussed in Sect. 2, while Sect. 3.1 introduces SAM, the general theory on which the proposed user modeling approach is based on, which is discussed in Sect. 3.2. Section 3.3 presents an instance of the approach applied in the context of Internet browsing activities, while the related evaluation, with a comparison with two other user modeling approaches developed in our context, are discussed in Sect. 4. Section 5 closes the paper.

2. Related Work

Several statistical approaches for user profiling have been proposed in order to recognize user information needs while interacting with information sources. Nevertheless, a few of them are based on formal cognitive theories and implicit feedback techniques.

A number of approaches that address the personalization task are based on the traditional content-based IR approach and explicit feedback techniques, e.g., *Joachimset al.*,1997,*MoukasandMaes*,1998,*Liebermanet al.*,1998,*Montaneret al.*,2003. Some approaches reduce the burden on users to provide explicit feedback combining content and collaborative techniques [BalabanovicandShoham,1997]. Interesting approaches employ Naïve Bayes classifiers to build representations of the user needs, such as Syskill and Webert [Pazzaniet al.,1996]. Personal WebWatcher [Mladenic,2001] employs Bayes classifiers to represent the keywords contained in the hyperlinks selected by users, helping the user browse the Web highlighting interesting hyperlinks on visited Web pages. It is one of the first prototypes based on implicit feedback techniques.

Speretta and Gauch [SperettaandGauch,2005] analyze search histories in order to build user models based on ODP categories. Teevan *et al.* obtain interesting results performing a re-rank of search engine results according to user profiles built from the information created, copied, or viewed by a user, such as Web pages, emails, documents stored on the client PC, etc.. Watson [Budziket al.,2001] monitors the users actions and the files that he is currently working on to predict the users needs and offer them related resources. The TFxIDF technique [Saltonet al.,1975] is used to create the contextual query based on the currently active window that is submitted to the information sources, i.e., search engines. These kinds of user models have many similarities with ones we included in our evaluation.

Two systems use natural language processing and semantic or keyword networks in order to build long term user profiles and evaluate the relevance of text documents with respect to a profile. SiteIF project [MagniniandStrapparava,2004] uses semantic networks built from co-occurrence frequencies among keywords in News corpora, where each node represents the meaning of one keyword in a given news, identified by means of the WordNet database [MillerandFellbaum,1993]. ifWeb prototype [AsnicarandTasso,1997] makes use of a network of keywords in order to create a representation of the available topics in one domain. The explicit feedback updates the user model adding or removing subsets of topics, i.e., sub-networks, judged interesting for the network associated with the user.

Two works based on cognitive theories are focused on the prediction of user actions, SNIF-ACT model [PirolliandFu,2003] exploits ACT-R theory's concepts [AndersonandLebiere,2000] such as the declarative and procedural knowledge, trying to represent all the items a user can deal with during the search, e.g., links, browser buttons, etc., and simulates users' actions during information-seeking processes by means of a set of production rules. The action selection considers the mutual relevance between the user goals and the current Web contents. Cognitive Walkthrough for the Web (CWW) [Blackmonet al.,2003] looks at the degree of similarity between user goals and heading/link texts by means of Latent Semantic Analysis, a technique that estimates the semantic relatedness of texts, based on a statistical analysis of a large corpus [LandauerandDumais,1997]. Even though they have not been employed in the personalization domain, the underlying cognitive models and techniques are in part related to the proposed user modeling approach.

3. User Modeling Approach

We hereby provide a brief introduction of the SAM theory before investigating in depth the proposed user modeling approach.

3.1. *SAM: Search of Associative Memory*

In this section, we give a short description of the SAM theory. For a closer examination of this theory see for example [RaaijmakersandShiffrin,1981].

SAM is a general theory of retrieval from long-term memory that considers both the structure of the memory system and the processes operating within it. The structure refers to the items represented and their organization in the memory system, the processes refer to the main activities that occur within the memory, such as learning and recall of the stored information.

The memory is organized in two parts: Long-Term Store (LTS) and Short-Term Store (STS). The STS shows two key features: a limited capacity and a proneness to "forget" its content (if the buffer size is reached, an item at random will be replaced). It can be regarded as a temporarily activated subset of information enclosed in the permanent LTS storage, which contains all prior information plus new information transferred from STS. The role of STS corresponds to a working space for control processes, such as coding, rehearsal, decision-making, etc.. When a new external sensory input occurs, the related information is analyzed through the LTS structure, and data correlated with the input is activated and placed in the STS.

Both kinds of memories consist of unitized *images*, that is objects that may be learned and recalled (images also include temporal-contextual features that are not included in the discussion in order to make the description simple, see [RaaijmakersandShiffrin,1981] for details). The retrieval (or recall) process is based on the associative relationships between probe cues in STS, and LTS's memory images. In the SAM theory, probe cues are the pieces of information the subject has with regard to the current task to be accomplished, e.g., information from a question, information retrieved earlier in the search, etc.. The cue set activates a subset of LTS images. To what degree each of the images is activated is determined by a matrix that gives the strengths of relationship between each possible probe cue and each possible image (Fig. 3).

The recall process determines what images are sampled and made available to the user for evaluation and decision-making. It is convenient to organize the process in two phases:

- At each step of the process, STS cues are used as probes of LTS. The probability to *sample* a LTS's image is a function of the strength of association between the probe cues and the various images in LTS.
- The sampled LTS images are then accessed and evaluated by the so-called *recovery* process. This process depends on the strength between the selected images and the probe cues. The recovered images will be stored in the STS.

The SAM Simulation (SAMS) approach is employed to build and keep the LTS structure updated,. This approach consists of a buffer rehearsal process [AtkinsonandShiffrin,1968] that updates the image-image strength as a function of the total amount of time the pair of objects are simultaneously present in the STS. For example, let t_{ij} be the time that objects i and j are together in the STS simultaneously, then the strength is $S_T(i, j) = a\ t_{ij}$ where a is a parameter. The strength between a probe set and an image also increases when a successful recovery occurs.

3.2. *The Proposed User Modeling Approach*

During each information-seeking session, when users are looking for documents that satisfy their information needs, they have to assemble and deal with sets of concepts related to these needs. The identification of these concepts allows us to personalize the human-computer interaction, e.g., improving the ranking of resources retrieved by traditional tools such as search engines or recommending new resources to the user. For this reason, we have decided to ground our user modeling approach in the cognitive theory of human memory developed by Raaijmakers and Shiffrin [RaaijmakersandShiffrin,1981]. It suggests important characteristics that have to be implemented in structures used to store memory concepts, and gives important recommendations about the algorithms to store and retrieve information from these structures.

Figure 2 shows the user model's internal organization and how it interacts from the outside. Probe cues are drawn from usage data (1), which is information related to users' behavior while interacting with the system. Examples of usage data are past queries submitted to search engines or contents of the browsed pages. This information is transferred in the temporary STS store (2) and is used by the sampling and recovering processes to probe the LTS (3), the permanent store containing all prior information transferred from STS. The probability to activate and transfer information from LTS to STS is a function of the strength of association between the STS probe cues and the information stored in LTS. These strength values saved in the LTS structure are assumed to be proportional to the total amount of time a given information unit remains in STS. The successfully recovered LTS units are transferred to STS and correspond with the information related to the user's current working context, e.g., a browsed page or an e-mail the user is currently writing. This information is useful to personalize the interaction with the system (4), for example generating queries to several sources of information and presents them in a separate window.

The SAM theory does not require particular memory representations for LTS and STS buffers, therefore we have to define the information encoding, e.g., the kind of images stored in the LTS and the STS cues used for probing. In our approach, both kinds of memories consist of words, therefore they correspond with the unitized information that may be learned and recalled. This choice is justified by its simplicity and adaptation to the Web personalization context used in the

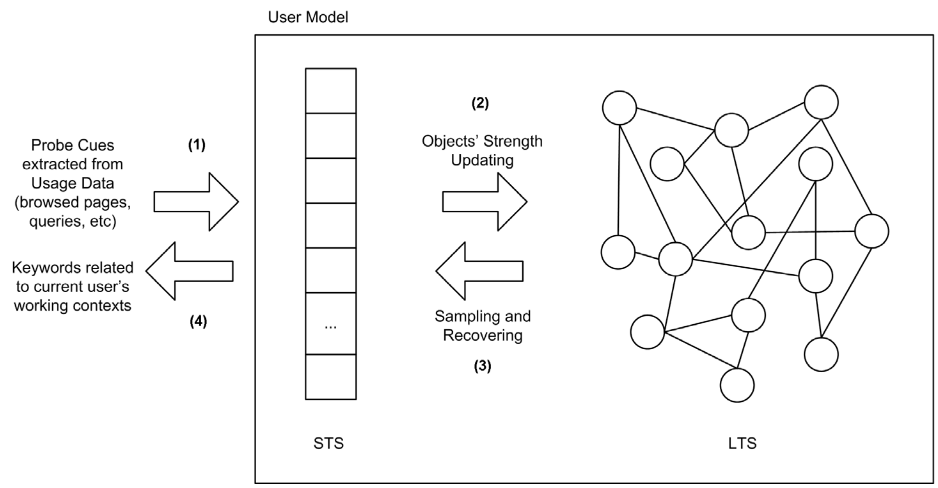

Fig. 2. The user profile is organized in two memory structures: STS and LTS. The former is a working buffer that contains probe cues extracted from usage data and subsets of LTS images (or words). The latter is the permanent store, containing all prior information. By means of learning and recall processes, both STS and LTS are kept updated according to the input of the user profile. The STS is also used to provide suggestion about the current user activities, in the form of sets of keywords.

evaluation. We have considered an environment where documents are represented by text sentences, therefore words seem the obvious elements that could be stored and analyzed.

No temporal or contextual information has been stored along with each word in this prototype. Nevertheless, Natural Language Processing (NLP) techniques can be successfully employed in order to assign a unique semantic meaning to each word extracted from usage data, for example analyzing the sentence in which the word occurs and using this co-occurrence information to query the *WordNet* database [MillerandFellbaum,1993] (see [MagniniandStrapparava,2004] for details).

LTS stores associative relationships between words. The amount of word-word information stored is assumed to be proportional to the total amount of time those two words are simultaneously present in the rehearsal buffer STS. The associations are stored in a *strength matrix* (see Fig. 3). Given a cue Q and a word I, the matrix stores the strength between these two objects $S_T(Q, I)$.

The STS size is a parameter of the system. Miller claimed that about seven chunks could be held into this kind of short-memory [Miller,1956], where a *chunk* stands for an integrated unit of information. Instead of remember a single stimulus, humans usually group input events, apply a new name to the group, and then remember the new name rather than the original input events. In this way, since the memory span is a fixed number of chunks, it is possible to increase the number of bits of information that it contains simply by building larger and larger chunks

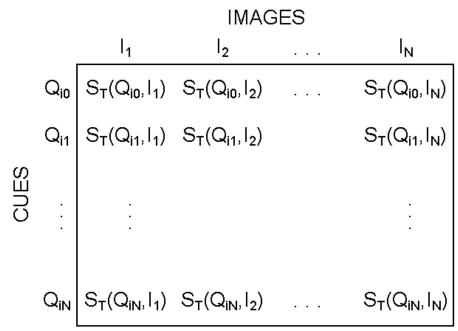

Fig. 3. The *strength matrix* used in the sampling and recovery processes to determine the probability of selection of words. It represents the information contained in LTS. The generic entry in the matrix corresponds to the strength between cue Q and word I (I stands for image in the SAM theory).

by means of a *recoding* process. Nevertheless, this kind of process has not been included in the user modeling approach.

The choice to use words as the unitized information that may be learned and recalled from the memory, and the absence of the recoding process lead us to reconsider the initial STS size. During empirical evaluations, users were asked to remember as many words as they could after having read sequences of documents. Being able to organize many words into single chunks, they could recall many more words than expected. Following these preliminary evaluations, we decided to set the STS size in the 50-100 unit range.

As explained in the previous section, the retrieval process consists of two phases: the sampling of words in the LTS, and their evaluation in a recovery process. Both steps exploit the strength matrix, which corresponds to the information stored in the LTS.

Given the current cues Qs, the sampling phase draws the probabilities for each word I_i in LTS as a function of the strength of association between them:

$$P_S(I_i|Q_1, Q_2 \dots Q_M) = \frac{\prod_{j=1}^{M} S_T(Q_j, I_i)^{W_j}}{\sum_{k=1}^{N} \prod_{j=1}^{M} S_T(Q_j, I_k)^{W_j}} \qquad (1)$$

where M is the current number of cues in STS, and N is the matrix dimension.

Equation 1 assigns high probabilities to the words with the highest product of strengths, hence, those that tend to be greatly associated with all the current cues. W_j represents weights used to give different significance to cues. For instance, in the Web domain, Inverse Document Frequency (IDF) values [Jones,2004] can decrease the weight for words that frequently occur in a given corpus, therefore very common and hardly relevant. In our evaluation W_j takes on the expression:

$$W_j = \left(\frac{idf_j}{maxIdf} \right)^3 \qquad (2)$$

where $idf_j = \log \frac{|Docs|}{df_j}$, df_j is the Document Frequency of the j word, $maxIdf = \log |Docs|$ and $Docs$ is a given input set of documents.

Once a word has been sampled, the recovery process takes place. The probability to successfully retrieve the sampled word I_i corresponds to:

$$P_R(I_i|Q_1, Q_2 \dots Q_M) = \frac{\sum_{j=1}^{M} S_T(Q_j, I_i)^{W_j}}{\max_{k\in[1,N]} \sum_{j=1}^{M} S_T(Q_j, I_k)^{W_j}} \tag{3}$$

This expression differs considerably from others in the literature, e.g., $_{Raaijmakers and Shiffrin, 1981}$:

$$P_R(I_i|Q_1, Q_2 \dots Q_M) = 1 - e^{[-\sum_{j=1}^{M} W_j S_T(Q_j, I_i)]} \tag{4}$$

The reason concerns the size of the rehearsal buffer, which is larger than the size used in other systems. If there are many cues, the exponent in Eq. 4 gets very low values and the probability tends to 1. In other words, the cue effect is so marginal that all the sampled images will be successfully recovered. The normalization in Eq. 3 prevents this effect. As in the original formulation, large cue weights affect the recovery positively and probabilities get high values even if one strength is high.

The strength matrix is kept updated according to the total time a word or a pair of words are stored together in the STS buffer. Given t_i the time spent in the buffer by the word I_i, and t_{ij} the time images I_i and I_j occur together in the buffer, we have:

$$S_T(I_i, I_j) = S_T(I_j, I_i) = b\ t_{ij}, \quad t_{ij} \neq 0$$

$$S_T(I_i, I_i) = c\ t_i$$

If a word pair has never appeared together in the buffer, they assume a non negligible residual strength d. The values b, c and d are parameters of the model.

The strengths are also increased whenever a successful recovery occurs. In this case, the strength between the cues Q_is and the word I_j, and the self-association strength $S_T(I_j, I_j)$ is incremented:

$$S'_T(Q_i, I_j) = S_T(Q_i, I_j) + f, \quad t_{ij} \neq 0$$

$$S'_T(I_j, I_j) = S_T(I_j, I_j) + g$$

where the S are the strengths before incrementing, and f and g are further parameters of the model.

The currently activated words in STS and their relationships with the words in LTS are used to build new LTS relationships. An important feature of the model is that units of information in the STS tend to be stored jointly. For example, when the user is browsing a given page and some content is extracted and given as input to the model, the words are stored in LTS with relationships that connect the words one another. When some of these words are recovered, the relationships help to retrieve the correlated information.

In natural language domains, such as the Web, this kind of implicit context helps to disambiguate the meaning of words. The relationships among them connect each word with the implicit context present at the time of learning. It may happen that users analyze the same word again, but in different contexts. In this case, different sets of connections are built between the word and the new context. Probing the LTS with some words related to the right context help recognize the right meaning of the word.

Moreover, the word significance depends on the user's current working context. Traditional approaches assign weights to words according to explicit and/or implicit feedbacks without taking under consideration the user's current activity. In our approach, a word is judged interesting if it is connected to the context, that is, the information stored in the control and decision-making buffer STS.

After having explained the sampling, recovering and learning processes, we move on to the general retrieval process that occurs each time the user interacts with information sources (see Fig.4). The process can be broken down to a inner retrieval process that is based on past SAM cognitive models. In the first cycle, there are K_{MAX} attempts to sample and recover words stored in LTS by means of the current probe cues (the context stored in STS). A failure in the retrieval process occurs when the sampling and recovering probabilities do not assume values over a given threshold, or if the sampled words have not been successfully sampled in previous cycles and STS has not been altered since then.

If a word I_i has been successfully sampled, the strengths $S_T(Q_j, I_i)$ with the current STS's words Q_js, and the self-association $S_T(I_i, I_i)$ are increased, and the word is included in the STS. The updated buffer will be used to sample other correlated words. The second cycle in the inner retrieval process corresponds to a *rechecking* phase. It ensures that all associative retrieval routes starting from the STS's words have been checked thoroughly, at least L_{MAX} times.

The outer flowchart's part of the retrieval process concerns the human-computer interaction with information sources. In our domain, we analyze the user's current activity and extract content that is given to the model as probe cues. Examples of probe cues are: queries, document's snippets, categories selected by the user. The available information is included in the current context stored in STS, and the retrieval process takes place to recover and suggest correlated information and, at the same time, update the LTS. The process periodically checks if the current user activity is changed, e.g., the user closed the browser window and opened a word processor. In this case, the temporary information stored during the session is wiped out and the STS is cleared.

Each time the recovery process is completed, the STS words correspond to the information related to the current information needs of the user. As described in the evaluation they can be used for query expansion, enhancing the query using words or phrases related to the set of documents seen by the user and the current user activity.

Traditional interactive systems ask users to mark documents as relevant or non-

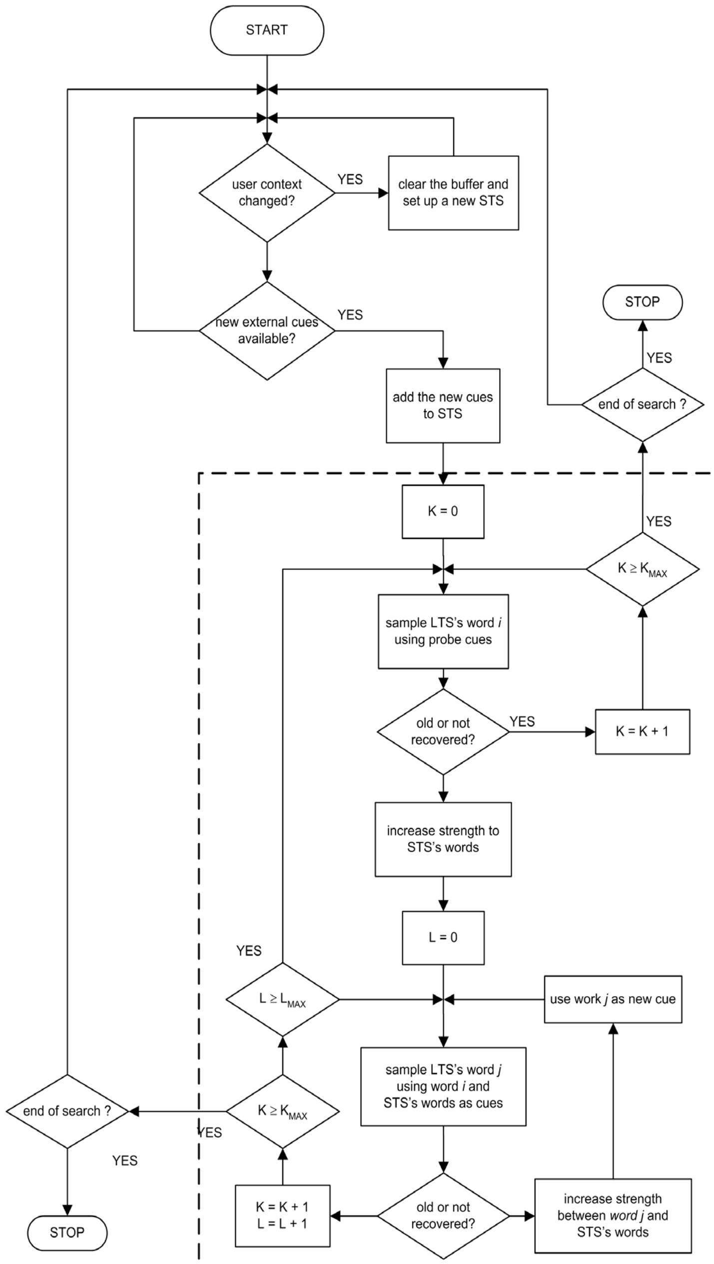

Fig. 4. An extension of the retrieval process flowchart in Raaijmakers and Shiffrin 1981 to consider the external cues originated by the interaction of users with information sources and their effects on the process. The broken box margins the original chart.

relevant. These documents are used as training data for a relevance feedback query expansion approach. On the contrary, the proposed user modeling approach is based on implicit feedback techniques, where usage data are analyzed in order to draw information used to build the model and keep it updated

The *renting* technique used in order to remove concepts from user models that are no longer judged interesting for the user has not been considered. There are two reasons to justify its presence. During information-seeking tasks, users analyze new concepts affecting their knowledge and beliefs, and therefore their current information needs. User models should be able to recognize these alterations, as well as alterations on users' goals. The second reason regards the user modeling accuracy. Several approaches use content-based techniques to extract content from Web documents used to build user models. Nevertheless, Web documents often cover different topics, even though users are usually interested in some of them. If user models are updated with irrelevant information, *forgetting* techniques able to remove concepts no longer taken into account by users is needed.

The described user modeling is based only on an additive learning process. At the same time, it is always possible to ignore concepts judged not interesting. In fact, the learning process tends to increase the relationship strengths between concepts that frequently occurred during the interaction with information sources. Wrong concepts may be included in the model, but their strengths with other words will not be increased if they no longer appear in the probe cues. Therefore, the probabilities to recover these concepts decrease as new concepts are included in the model. In other words, 'forgetting' follows the failure in the attempt to retrieve concepts.

3.3. *An User Model Instance for Browsing Activities*

This section presents an instance of the proposed user modeling approach in the Web domain, while the related evaluation is shown in Sect. 4. The amount of the available information that can be exploited during any information-seeking task makes the Web a fundamental environment where user modeling approaches can be evaluated. Recognizing what users are looking for during browsing sessions, that is, their information needs, is the first step towards efficient personalization techniques.

In order to employ our user modeling approach on the Web, we must resort to a methodology to identify the probe cues that are used in learning and recall processes, as discussed in the previous sections. The notion of *information scent* [Chiet al.,2000,Chiet al.,2001] developed in the context of Information foraging theory [PirolliandCard,1999], already evaluated in different tasks [Chi,2004], is a valid approach in recognizing these cues by means of the anchor text associated to each link. While browsing, users use these proximal cues to decide whether to access the distal content, that is, the page pointed by the link. Formally, the information scent is the imperfect, subjective perception of the value or cost of the information sources obtained from proximal cues.

Each time the user selects a link and visits the corresponding page, the link's anchor text can be extracted and used in learning and recall processes. However, preliminary evaluations show how sometimes this information is not enough to recognize valuable probe cues, especially if the text consists only of a few common words, e.g., "full story", "page two", "link", "rights reserved", etc..

For this reason, we have developed an algorithm to collect information related to a given link selection: the anchor text is combined with the title of related page visited by the user. Afterwards, by means of the page's Document Object Model[a], the page is divided into units whose boundaries are defined by a subset of HTML tags, e.g., `TR`, `P`, `UL`, etc., and the text of the deepest unit that contains the link is retrieved. Finally, the retrieved text plus anchor and title are compared to the other units of the pointed page, in order to find further related text. The text comparison is based on the IR similarity function used in the TextTiling algorithm [Hearst,1997].

As previously mentioned, no temporal or contextual information has been stored along with each word, nor have NLP techniques to analyze semantic meanings been included in this first prototype.

4. Evaluation

Evaluating user models based on implicit feedback techniques that draw usage data from browsing histories with standard measures is not an easy task, because test beds that include browsing sessions and descriptions of the related information needs are not available. Moreover, there are a few prototypes that employ implicit techniques to build user profiles for Web personalization.

Instead of making ad-hoc corpus of documents extracted from browsing histories, and manually recognizing the information needs that led users through specific paths, we have decided to consider some categories in a directory service Web site in order to train and test the user modeling approach. The paramount advantage is that measures of search effectiveness, e.g.., precision and recall, can be adopted without considering time consuming and costly experiments with human subjects.

The input of the user models is the information collected by the technique described in the previous section based on the anchor texts, the titles and the correlated text in a Web page. The images stored in the user model correspond to the word extracted from these data.

After having randomly chosen 10 categories from the Open Directory Project[b] (ODP), a subset of urls (25%) in each category is selected for the training phase. The 10 categories correspond to the different information needs of a single user. After a preliminary evaluation, we have chosen the following values for the system parameters: $a = 0.1$, $b = 0.1$, $c = 0.1$, $d = 0.2$, $e = 0.7$, $f = 0.7$ and $g = 0.7$. The size of the STS is 100 words.

[a] http://www.w3.org/DOM/
[b] http://www.dmoz.org

	ODP Category
1	Business/Mining and Drilling/Consulting
2	Sports/Cycling/Human Powered Vehicles
3	Computers/Home Automation/Products and Manufacturers
4	Games/Roleplaying/Developers and Publishers
5	Business/Agriculture and Forestry/Fencing

Table 1. The ODP categories randomly chosen for the evaluation.

The evaluation is based on the user model's ability to suggest further Web sites related to the user needs, namely, the remaining Web sites in the 10 categories that are not used for training. In order to provide the input probe cues, we employed the methodology presented in Sect. 3.3 based on the anchor text and related text retrieved from the linked pages.

After the training phase, the name of the category is given as probe cues and the keywords suggested by the user model are used for query expansion. A standard IR search engine indexes the Web sites contained in the first three layers of the ODP hierarchy. The following cosine rule is used to match the expanded query Q with the ODP sites Ds:

$$M(Q,D) = \sum_{t \in Q} \frac{tf_{Q,t} \cdot idf_t}{\sqrt{\sum_{t \in Q}(tf_{Q,t} \cdot idf_t)^2}} \cdot \frac{tf_{D,t} \cdot idf_t}{\sqrt{L_D}} boost_t \tag{5}$$

where t is a term of the query Q, L_D is the number of terms in the document to analyze D, $t_{x,t}$ is the term frequency of t in the query or document x, and finally idf_t is the inverse document frequency of t. The expanded queries are composed of the category name, and the suggested keywords weighted 0.5 by the $boost_t$ factor.

Two traditional user modeling approaches provide a benchmark to compare our evaluation results. The first one is based on a content-based technique where documents are represented through the Vector Space model (VSM). A Relevance Feedback (RF) technique updates the model according to the content from the training pages. The top 10 keywords with highest rank, measured by means of the TFxIDF measure, are used in the query expansion. The second is a Naïve Bayes classifier trained on the same textual input data [Mitchell,1997]. It is used for filtering the ODP corpus of Web sites according to the different user needs. Since it is not possible for this user model to submit a query and obtain keywords for query expansion, we have decided to collect the 5000 urls with the highest rank assigned by the classifier. This set is than compared with the testing set of the ODP input categories.

The different evaluation phases are summarized as follow:

- Random selection of 10 ODP categories, 5 of them corresponding to stable information needs

ODP	Naïve Bayes classifier		VSM-based Approach		Proposed Approach	
Category	Precision	Recall	Precision	Recall	Precision	Recall
1	0.08	0.05	0.18	0.11	0.36	0.23
2	0	0	0.08	0.25	0.06	0.19
3	0	0	0.06	0.11	0.08	0.14
4	0.02	0.05	0.12	0.3	0.16	0.4
5	0.06	0.03	0.22	0.10	0.6	0.30

Table 2. Evaluation results for the five ODP categories (see Table 1). Precision and recall values at 50 of the proposed SAM-based user modeling approach are to be seen in the sixth and seventh column. The same measures related to the benchmark user model based on VSM and relevance feedback are shown in the fourth and fifth column, while the results of the Naïve Bayes classifier analyzing the top 5000 ranked results are on the other two columns.

- Random selection of 25% urls in each category for the training phase
- Extraction of probe cues by means of the algorithm described in Sect. 3.3
- UM training through learning and recovery processes (Sect. 3.2)
- The category names are given as probe cues to the UM in order to retrieve related sets of keywords
- Query expansion with the retrieved keywords
- Precision and recall of the top 50 documents returned by a search engine which indexed the ODP

Table 1 shows 5 out of 10 categories chosen in the evaluation. The other 5 categories represent different topics that we have included in order to cause noise during the training phase. As in real scenarios, some of the browsing sessions are related to relatively stable information needs, others concern transient concepts that less frequently re-occur in the future, e.g., weather in a holiday resort, price of a LCD, etc..

Figure 5 presents the evaluation results (see also Tab. 2) for each ODP category. In particular, the precision and recall at 50 (only the top 50 search engine results were analyzed) for the proposed approach and the traditional VSM user model based on relevance feedback. As for the Naïve Bayes classifier, the precision and recall values are drawn considering the first 5000 results filtered by this model. We obtained better results, both in terms of precision and recall, in four categories, while one category shows slightly worse results.

In the Web domain, the proposed user model outperforms both the traditional VSM-based and the approaches based on Naïve Bayes classifiers, being able to recognize important keywords according to the user's working context, i.e., the probe cues. By means of query expansion, the keywords re-ranked the results of a traditional search engine suggesting better personalized documents. The evaluation also shows how the Naïve Bayes classifier already employed in information filtering with good results, is basically non-effective in the considered personalization task,

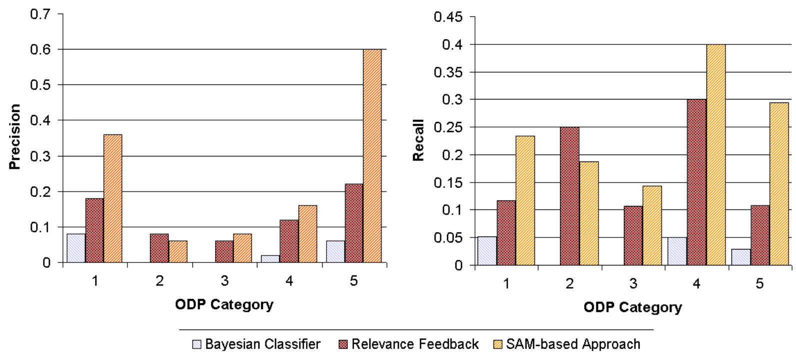

Fig. 5. A chart of the precision and recall values obtained in the evaluation.

whereas the VSM-based approach shows satisfactory performance.

5. Conclusions

We have presented a user modeling approach based on a cognitive theory able to recognize information needs during the user interaction with information sources. An instance of the approach for the Web domain has been discussed. The evaluation results outperform traditional approaches based on VSM or Naïve Bayes Classifier, proving the ability to recognize important keywords according to user working contexts. These keywords can be considered in query expansion techniques in order to provide personalization of search engine results or recommendation related to the current user activities.

Potential enhancements concern the inclusion of contextual information in each word stored in STS and LTS. This kind of information is available through common categorization techniques that assign a category to each analyzed document. NLP techniques able to assign unique semantic meanings to each word can increase the performance of the approach further.

Finally, from our own standpoint, user modeling should be organized in different layers. The lower layers concern the basic human memory processes of learning and retrieval, as the goal of the proposed approach. The higher layers deal with the information-seeking strategies and plans users undertake when a particular task ought to be accomplished. The latter layers are the subject of our future work.

References

Anderson and Lebiere, 2000. Anderson, J. R. and Lebiere, C. (2000). *The atomic components of thought*. Lawrence Erlbaum Associates, Mahwah, NJ, USA.

Anick, 2003. Anick, P. (2003). Using terminological feedback for web search refinement: a log-based study. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 88–95, New York, NY, USA. ACM Press.

Asnicar and Tasso, 1997. Asnicar, F. A. and Tasso, C. (1997). ifweb: a prototype of user model-based intelligent agent for document filtering and navigation in the world wide web. In *Proceedings of Workshop Adaptive Systems and User Modeling on the World Wide Web (UM97)*, pages 3–12, Sardinia, Italy.

Atkinson and Shiffrin, 1968. Atkinson, R. C. and Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In Spence, K. and Spence, J., editors, *The psychology of learning and motivation: Advances in research and theory*, volume 2, pages 89–195. Academic Press, New York, USA.

Balabanovic and Shoham, 1997. Balabanovic, M. and Shoham, Y. (1997). Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66–72.

Blackmon et al., 2003. Blackmon, M. H., Polson, P. G., Kitajima, M., and Lewis, C. (2003). Cognitive walkthrough for the web. In *Proceedings of the ACM conference on human factors in computing systems (CHI2002)*, pages 463–470, Minneapolis, Minnesota, USA.

Budzik et al., 2001. Budzik, J., Hammond, K. J., and Birnbaum, L. (2001). Information access in context. *Knowledge-Based Systems*, 14(1-2):37–53.

Chan, 2000. Chan, P. K. (2000). Constructing web user profiles: A non-invasive learning approach. In *WEBKDD '99: Revised Papers from the International Workshop on Web Usage Analysis and User Profiling*, pages 39–55, London, UK. Springer-Verlag.

Chi, 2004. Chi, E. H. (2004). Transient user profiling. In *Proceedings of the Workshop on User Profiling (CHI2004)*, pages 521–523, Vienna, Austria.

Chi et al., 2001. Chi, E. H., Pirolli, P., Chen, K., and Pitkow, J. (2001). Using information scent to model user information needs and actions on the web. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI2001)*, pages 490–497, Seattle, WA, USA.

Chi et al., 2000. Chi, E. H., Pirolli, P., and Pitkow, J. (2000). The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a web site. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI2000)*, pages 161–168, Hague, Netherlands.

Claypool et al., 2001. Claypool, M., Le, P., Wased, M., and Brown, D. (2001). Implicit interest indicators. In *IUI '01: Proceedings of the 6th international conference on Intelligent user interfaces*, pages 33–40, New York, NY, USA. ACM Press.

Hearst, 1997. Hearst, M. A. (1997). Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64.

Joachims et al., 1997. Joachims, T., Freitag, D., and Mitchell, T. M. (1997). Webwatcher: A tour guide for the world wide web. In *Proceedings of the 15h International Conference on Artificial Intelligence (IJCAI1997)*, pages 770–777.

Joachims et al., 2005. Joachims, T., Granka, L., Pan, B., Hembrooke, H., and Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 154–161, New York, NY, USA. ACM Press.

Jones, 2004. Jones, K. S. (2004). Idf term weighting and ir research lessons. *Journal of Documentation*, 60(5):521–523.

Kelly and Teevan, 2003. Kelly, D. and Teevan, J. (2003). Implicit feedback for inferring user preference: a bibliography. *SIGIR Forum*, 37(2):18–28.

Koenemann and Belkin, 1996. Koenemann, J. and Belkin, N. J. (1996). A case for interaction: a study of interactive information retrieval behavior and effectiveness. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 205–212, New York, NY, USA. ACM Press.

Landauer and Dumais, 1997. Landauer, T. K. and Dumais, S. T. (1997). A solution to platos problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104:211–240.

Lieberman et al., 1998. Lieberman, H., Dyke, N. W. V., and Vivacqua, A. S. (1998). Let's browse: a collaborative web browsing agent. In *Proceedings of the 4th International Conference on Intelligent User Interfaces (IUI'99)*, pages 65–68, Los Angeles, CA, USA. ACM Press.

Magnini and Strapparava, 2004. Magnini, B. and Strapparava, C. (2004). User modelling for news web sites with word sense based techniques. *User Modeling and User-Adapted Interaction*, 14(2):239–257.

Micarelli and Sciarrone, 2004. Micarelli, A. and Sciarrone, F. (2004). Anatomy and empirical evaluation of an adaptive web-based information filtering system. *User Modeling and User-Adapted Interaction*, 14(2-3):159–200.

Miller, 1956. Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63:81–97.

Miller and Fellbaum, 1993. Miller, G. A. and Fellbaum, C. (1993). Lexical and conceptual semantics. In Levin, B. and Pinker, S., editors, *Advances in Neural Information Processing Systems*, pages 197–229. Blackwell, Cambridge and Oxford, England.

Mitchell, 1997. Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.

Mladenic, 2001. Mladenic, D. (2001). Using text learning to help web browsing. In *Proceedings of the Ninth International Conference on Human-Computer Interaction*.

Montaner et al., 2003. Montaner, M., López, B., and Rosa, J. L. D. L. (2003). A taxonomy of recommender agents on the internet. *Artificial Intelligence Review*, 19(4):285–330.

Moukas and Maes, 1998. Moukas, A. and Maes, P. (1998). Amalthaea: An evolving multi-agent information filtering and discovery system for the WWW. *Autonomous Agents and Multi-Agent Systems*, 1(1):59–88.

Pazzani et al., 1996. Pazzani, M. J., Muramatsu, J., and Billsus, D. (1996). Syskill webert: Identifying interesting web sites. In *Proceedings of the National Conference on Artificial Intelligence (AAAI-96)*, pages 54–61, Portland, OR, USA. AAAI Press.

Pirolli and Card, 1999. Pirolli, P. and Card, S. K. (1999). Information foraging. *Psychological Review*, 106:643–675.

Pirolli and Fu, 2003. Pirolli, P. and Fu, W.-T. (2003). Snif-act: A model of information foraging on the world wide web. In *Proceedings of the 9th International Conference on User Modeling*, Johnstown, PA, USA.

Pitkow et al., 2002. Pitkow, J., Schütze, H., Cass, T., Cooley, R., Turnbull, D., Edmonds, A., Adar, E., and Breuel, T. (2002). Personalized search. *Commun. ACM*, 45(9):50–55.

Raaijmakers and Shiffrin, 1981. Raaijmakers, J. G. and Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88(2):93–134.

Radlinski and Joachims, 2005. Radlinski, F. and Joachims, T. (2005). Query chains: learning to rank from implicit feedback. In *KDD '05: Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 239–248, New York, NY, USA. ACM Press.

Rhodes, 2000. Rhodes, B. J. (2000). *Just-In-Time Information Retrieval*. PhD thesis, MIT Media Laboratory, Cambridge, MA.

Salton et al., 1975. Salton, G., Wong, A., and Yang, C. (1975). A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620.

Speretta and Gauch, 2005. Speretta, M. and Gauch, S. (2005). Personalized search based on user search histories. In *Web Intelligence (WI2005)*, France. IEEE Computer Society.

Teevan et al., 2004. Teevan, J., Alvarado, C., Ackerman, M. S., and Karger, D. R. (2004). The perfect search engine is not enough: a study of orienteering behavior in directed search. In *CHI '04: Proceedings of the SIGCHI conference on Human factors in com-*

puting systems, pages 415–422, New York, NY, USA. ACM Press.

Teevan et al., 2005. Teevan, J., Dumais, S. T., and Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 449–456, New York, NY, USA. ACM Press.

Wærn, 2004. Wærn, A. (2004). User involvement in automatic filtering: An experimental study. *User Modeling and User-Adapted Interaction*, 14(2-3):201–237.

White et al., 2001. White, R., Jose, J. M., and Ruthven, I. (2001). Comparing explicit and implicit feedback techniques for web retrieval: Trec-10 interactive track report. In *TREC*.

White et al., 2002. White, R., Ruthven, I., and Jose, J. M. (2002). The use of implicit evidence for relevance feedback in web retrieval. In Crestani, F., Girolami, M., and van Rijsbergen, C. J., editors, *Advances in Information Retrieval, 24th BCS-IRSG European Colloquium on IR Research Glasgow, UK, March 25-27, 2002 Proceedings*, volume 2291 of *Lecture Notes in Computer Science*, pages 93–109. Springer.

Biographical Sketch and Photo

Fabio Gasparetti is a Postdoctoral Fellow at the Artificial Intelligence Lab of the Department of Computer Science and Automation, University of "Roma Tre". His research interests include: Personalized Search, User Modeling and Focused Crawling.

Alessandro Micarelli is a Full Professor of Artificial Intelligence at the University of "Roma Tre", where he is in charge of the Artificial Intelligence Laboratory at the Department of Computer Science and Automation. His research interests include: Adaptive Web-Based Systems, Personalized Search, User Modeling, Artificial Intelligence in Education.