https://doi.org/10.48009/2_iis_2009_257-266

USER MULTI-INTEREST MODELING BASED ON SEMANTIC SIMILAR NETWORK IN PERSONALIZED INFORMATION RETRIEVAL

Zhiheng Qi, Central Michigan University, qi1z@cmich.edu
Nianbai Fan, Hunan University, China, nbfan6203@gmail.com
Zhenyu Huang, Central Michigan University, huang1z@cmich.edu

ABSTRACT

People spend far more time searching information over the Internet than using it, because the desired information is often buried within a long list of searched results. Personalized internet access is a feasible solution to solve this search vs. use dilemma, which helps identify the web documents users truly need. A user's interests are usually represented by a profile. In this research, an improved vector space model representation is proposed to improve the user interests management efficiency. Based on this, the research further proposes a method for user multi-interest modeling integrated with semantic similar network (SSN). It studies the feature selection in user modeling, and proposes a feature selection method combining TF and TF-IDF that is proved a better performance in the test. Finally a complete module design is presented, which provides a personalized recommendation system for practical applications.

Keywords: personalization, information retrieval, vector space model, semantic similar network

INTRODUCTION

Information retrieval tools, such as searching engines, have been widely used to reduce information overload (Kuropka, 2004). With these searching engines, information is retrieved by the matching of keywords entered by users. These conventional searching tools help, to some extent, relieve the difficulty in information searching, but the relevant results are still too much (Kim and Chan, 2003). For instance, if we search for NBA game information with keywords "NBA play-off 2008", search engines will render all results "matching" the keywords, even irrelevant texts in which these terms are just mentioned occasionally. In addition, since it is often hard to accurately prescribe the actual information need with a small list of keywords, users may choose inappropriate or inaccurate keywords as search terms, which then lead to unexpected retrieval results with little values to user information needs.

As information retrieval services are becoming more significant in business and daily use, both for organizations and individuals (Fisher and Everson, 2003), it is critical for them to move from being passive with little adaptation to their users, to being more proactive and personalized in offering and tailoring information for group and individual users (Liang, Lai and Ku, 2007). Personalization refers to adopting different service strategies and to providing different details, according to different users (French and Viles, 1999; Luce and Giacomo, 2004). In accordance to the characteristic and requirement of different users, personalization initiative services sort out and classify information resources, and further provide and recommend users information according to users demand and preference, which enable users to obtain better services with moderately small investment, as well as to extricate users from the information overload.

The foundation and premise of personalized services are to create a user profile, which is designed to reflect user actual requirements (Shepherd, Duffy, Watters and Gugle, 2001; Vallet, Fernandez and Castells, 2005). An accurate user profile is the key of user modeling. In recent years, the user modeling techniques are investigated in various aspects. Many approaches, in the representation of user model, for instance, the vector space model (Salton, Wong and Yang, 1975), ontology (Vallet, et al., 2005); techniques in the machine learning, Tf-idf (Salton and Buckley, 1988), Bayes classification (Mozina, et al., 2004); genetic algorithm (Mitchell, 1996), neural networks methods (Gardner and Derrida, 1988) in the model updating, have been proposed and developed.

In this research, a method for user multi-interest modeling based on SSN is proposed. The organization of the paper is as follows: a generic context and background are introduced in section 1. In section 2, concepts of model representation are introduced, and an improved representation of vector space model (VSM) is proposed. Section 3 introduces the modeling schemes, and proposes a model update approach based on the improved VSM representation in section 2. Conclusion and future outlook are in section 4.

INFORMATION RETRIEVAL MODEL

Information Retrieval (IR) systems

Information retrieval (IR) is the science of searching for information in documents, searching for

documents, searching for metadata which describe documents, or searching within databases, whether relational stand-alone databases or hypertextually-networked databases such as the World Wide Web. Information Retrieval systems are designed with the objective of providing, in response to a user query, references to documents that would contain the information desired by the user (Singhal, 2001).

For the information retrieval to be efficient, the documents are typically transformed into a suitable representation. An information retrieval model specifies representations used for documents and queries, and how they are compared (Turtle and Croft, 1992). Common representation models are categorized into two dimensions: the mathematical basis and the properties of the model (Kuropka, 2004), as shown in Figure 1.

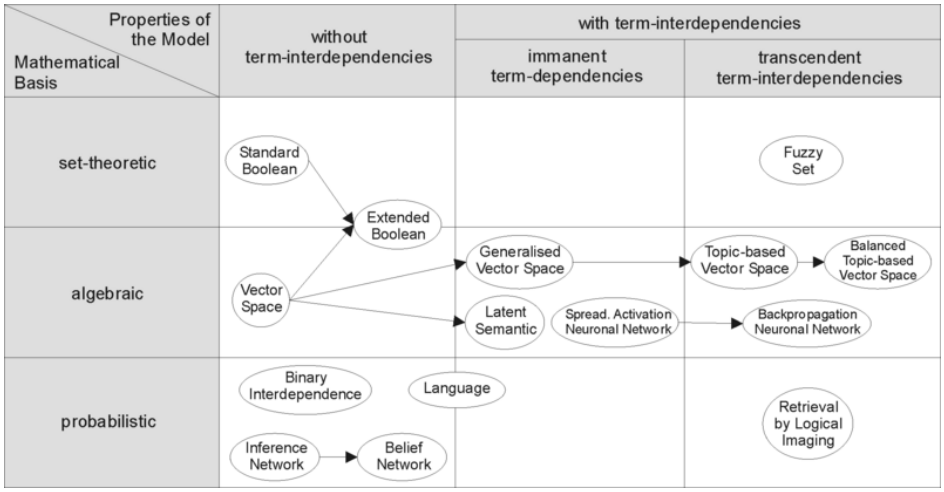

Figure 1. Categorization of information retrieval models

The Boolean, vector and probabilistic models are the three basic models used in document retrieval (Baeza-Yates and Ribeiro-Neto, 1999). These three types of model are included in mathematical basis, the first dimension. In the Boolean model, documents and queries are represented as set of index terms, and therefore this model is set theoretic. In vector space model (VSM), documents and queries are presented as vectors in multi-dimensional space, so the model is algebraic. Finally, in the probabilistic model, the framework for modeling document and query representations is based on probability theory, so the model is called probabilistic.

Property of the model is the second dimension. It reveals the interdependency of terms. Models without term-interdependencies treat different terms/words as independent, while models with term interdependencies allow a representation of interdependencies between terms. Models with immanent term interdependencies suggest that interdependency between two terms is defined by the model itself, but models with transcendent term interdependencies do not allege how the interdependency between two terms is defined (Kuropka, 2004). Based on three basic models, some modern IR models pay more attention to the term interdependency. The new models and the modeling techniques are rapidly developing, especially in recent years (Fisher and Everson, 2003; Wang, et al., 2008).

Traditional vector space model

From the three basic models described above, vector space model (VSM) is the most popular. It is simple, and can perform as good as the other two. For that reason, it has been widely used in the traditional IR field (Cavnar, 1995; Huang, et al., 2006). In the model both documents and queries are represented by an n-dimensional feature vector $\{(k_1, w_1), (k_2, w_2), \ldots, (k_n, w_n)\}$, with each single vector composed of keyword k_i and its weight w_i. Weight represents user's degree of interest to a concept or document, while keywords k_1, k_2, k_3,… can be all the terms in the document or some keywords obtained from feature selection (Salton, Wong and Yang, 1975; Raghavan and Wong, 1986). In general, in order to achieve an accurate searching in a high rate, user interest feature is usually represented by some certain keywords. Therefore, keyword selection is significant.

Representation of improved vector space model

User interests change all the time. It varies with time, so user profile ought to be able to catch the changes and update the user interests accordingly. In this paper, a feature term t is added in the traditional VSM. If user has n interest nodes, then user model is represented as follows:

$$Interest = \{f_1, f_2, \ldots, f_n\} = \{(k_1, w_1, t_1), (k_2, w_2, t_2), \ldots, (k_n, w_n, t_n)\}$$

f_i refers to the feature vector i of user interest; k refers to the key term, w refers to the weight of the term, and t refers to the latest update time of the term weight w.

Representation of user multi-interest profile modeling based on semantic-similar network

In the improved VSM model, keyword is extracted from the document. Due to the lexical synonymy and semantic divergency, and some word correlation factors, VSM sometimes cannot precisely and adequately represent the user interest features and therefore some queries may include inaccurate results. For instance, if a user intends to buy a "laptop", and then the user is potentially interested in the information "notebook computer"; if a user is a fan of "NBA", and then the user may also be likely to be willing to get the news of "Kobe Bryant" or "T-Mac". Supposing that the user model only contains the term "laptop" and "NBA", regarding to the articles describing the information of computers and basketball games, if those articles don't use the word "laptop" or "NBA", then they will not be searched out by retrieval system; it is very likely that other irrelevant articles, however, if mentioned those words incidentally, will be recognized under the searching result. As is often the cases, the user profile based on VSM describes user interest with over-precision, so that it cannot represent the semantic features, and thus restricts the fuzzy matching, of user interests (Lehmann, 1992). Aiming at the shortcomings of VSM representation discussed above, the paper proposes a user multi-interest model based on semantic similar network (SSN). By virtue of the existing linguistic knowledge system, through the statistical training in the large scale corpus, SSN can expand feature terms in the knowledge layer, associating the similar and relevant words, which improves the information retrieval from being based on the keywords to on the user actual interest, therefore improving the intelligent level in the personalized retrieval (Shackleford, 2005).

Besides, considering the diversity of user interests, a multi-interest model is advanced in this paper. In the model a user's interest is represented as a set of user interest categories, which reduces the interference among different interest features. If user has n preferences in different fields, then the user model is represented as follows:

$$Interest = \{(c_1, h_1, q_1, F_1), (c_2, h_2, q_2, F_2), \ldots, (c_n, h_n, q_n, F_n)\}$$

(c_i, h_i, q_i, F_i) refers to category i of user interest nodes; c refers to the name of the user interest category; h refers to the weight of the category, a ratio of the interest samples in category c to all the interest samples; q refers to the number of the interest samples in the category. F is the list of the interest nodes in category c, if c has m interest nodes, then:

$$F_i = \{(k_{i1}, w_{i1}, f_{i1}, t_{i1}), (k_{i2}, w_{i2}, f_{i2}, t_{i2}), \ldots, (k_{im}, w_{im}, f_{im}, t_{im})\}$$

$(k_{ij}, w_{ij}, f_{ij}, t_{ij})$ refers to the interest node j in category c; k refers to the key term, w refers to the weight of the term, and t refers to the latest update time of the term weight w. f is a mark, which has a value of 1 if the key term (k) is directly extracted from the document, and has a value of 0 if the key term (k) comes from the association of semantic proximity.

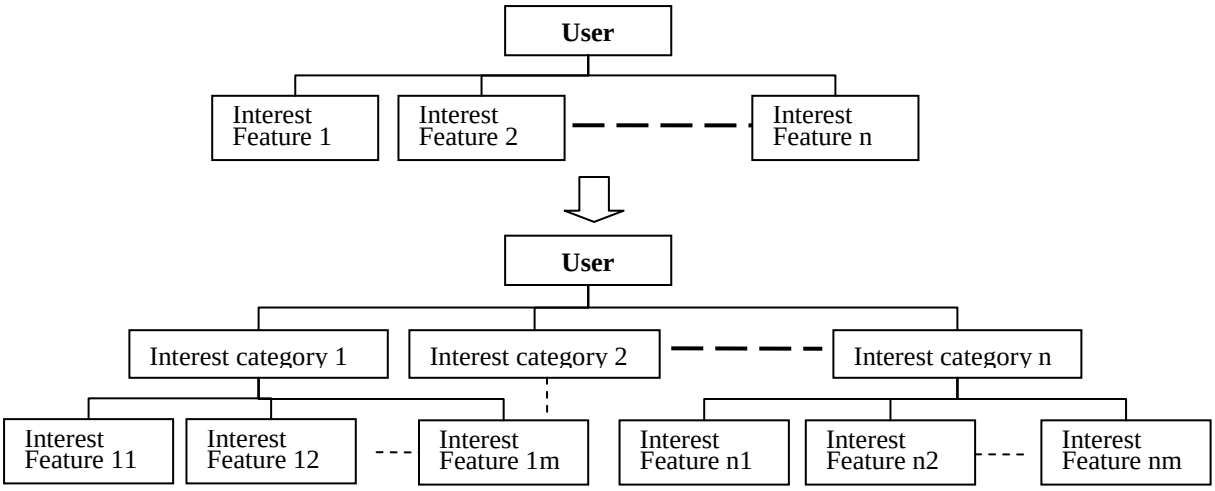

Figure 2. Structure comparison of user profile

USER MODEL CONSTRUCTION

User modeling scheme

User modeling is used to capture and record user needs and interests, by creating a user profile. A user profile records and manages user interests, and describes user's potential interest needs (Zhang, 2008). Before creating a user profile, information collection is required, for instance, information of user interest features, user activities and behaviors, and the acquisition of document set of user interests (Godoy and Amandi, 2005). In this research, the interest documents directly come from the user, that is, user actively provides the interest documents, based on which the user profile is then created through user model scheme.

Simple user modeling scheme

According to the representation of improved vector space model, a simple user modeling scheme is proposed as shown in Figure 3. The flow of the modeling is quite simple. After removing interjections and particles from the interest document, key terms are selected and their weights are calculated by means of feature selection, to describe the user interests. At last, the vector consisted with term, its weight and update time is written into the user profile.

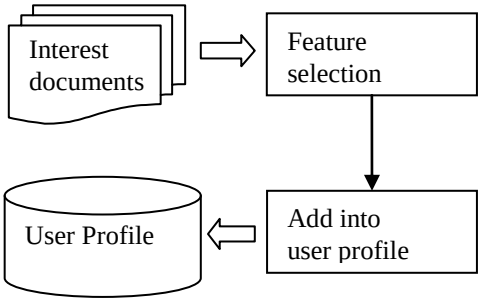

Figure 3. Simple user modeling scheme

User multi-interest modeling scheme based on semantic-similar network

The scheme of user multi-interest modeling based on SSN is shown as Figure 4. More complicated, the scheme is consisted with three modules: user modeling module, text classification training module and SSN training module.

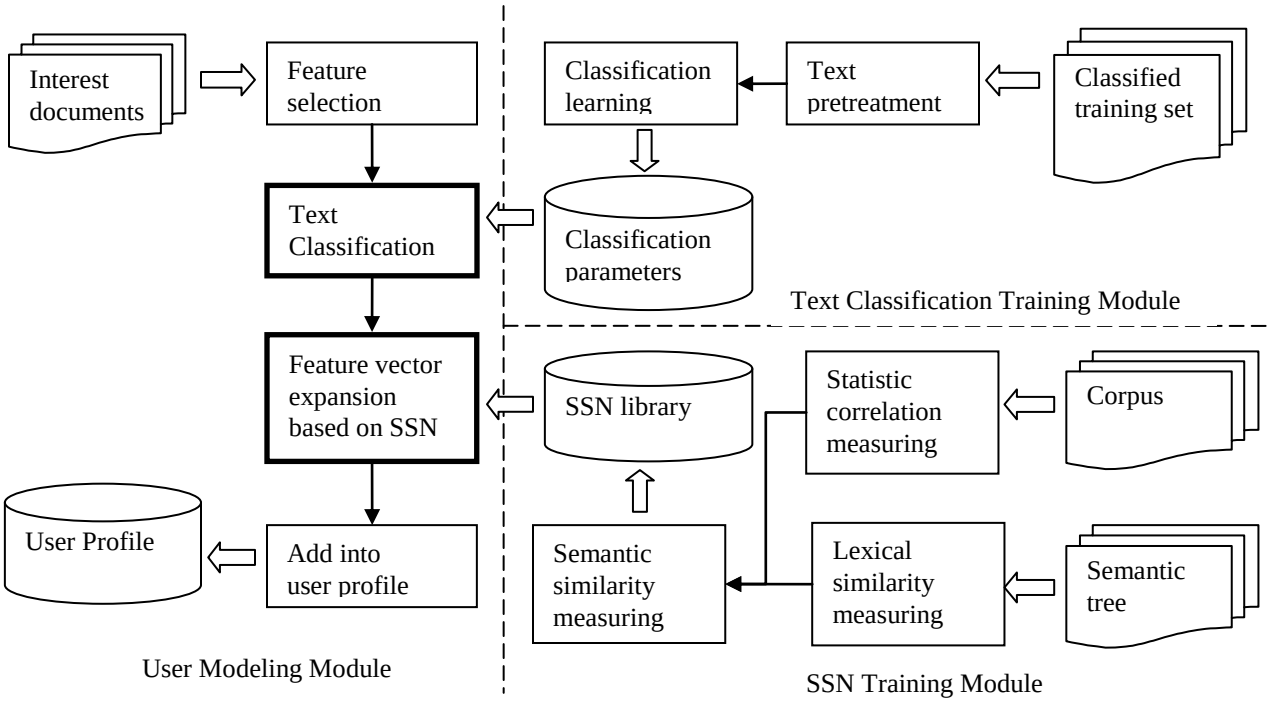

Figure4. User multi-interest modeling scheme based on semantic-similar network

The user modeling module is the main part of the modeling scheme. After feature vectors are obtained via feature selection, user interest features are classified into different categories. Feature vectors are expanded based on SSN next, through which similar words and relevant terms are associated. Finally the feature vectors, as well as the associated features, are added into user profile according to the information of interest classification. The text classification training module makes the preparation to the text classification. Through the classification learning on numerous training sets, the classification module offers experience parameter (Debole and Sebastiani, 2003). Similar to the text classification training module, SSN training module provides the basis for feature association, on the basis of the construction of corpus and artificial semantic tree (Lehmann, 1992; Sheth, et al., 2005).

Feature selection testing

Many different measures for evaluating the performance of information retrieval systems have been proposed. The measures require a collection of documents and a query. All common measures described here assume a ground truth notion of relevancy: every document is known to be either relevant or non-relevant to a particular query. In practice queries may be ill-posed and there may be different shades of relevancy.

Precision and **Recall** are two widely used measures for evaluating the quality of results in domains such as Information Retrieval and statistical classification. Precision can be seen as a measure of exactness or fidelity, whereas Recall is a measure of completeness. In an Information Retrieval scenario, Precision is defined as the number of relevant documents retrieved by a search divided by the total number of documents retrieved by that search, and Recall is defined as the number of relevant documents retrieved by a search divided by the total number of existing relevant documents (which should have been retrieved).

In the testing, Precision and Recall are defined in terms of a set of standard keywords (*{Keywords}*) and a set of relevant terms retrieved by feature selection (*{Selected}*).

$$Precision = \frac{\left| Keywords \cap Selected \right|}{\left| Selected \right|}$$

$$Recall = \frac{\left| Keywords \cap Selected \right|}{\left| Keywords \right|}$$

The testing data are 120 texts downloaded from Yahoo! news, in which a set of standard keywords are artificially selected. Different feature selection

methods are measured during the testing, and the result is shown in Table 1.

Table 1. Precision and Recall comparison with different feature selection methods

Feature selection methods	**Precision average**	**Recall average**
Tf (Term frequency)	0.502	0.311
Tf-Idf (Inverse document frequency)	0.579	0.376
Tf (non-noun) + Tf (noun)	0.662	0.381
Tf (non-noun) + Tf-Idf (noun)	0.701	0.670
Tf-Idf (non-noun) + Tf-Idf (noun)	0.714	0.688

Term frequency (Tf) and inverse document frequency (Idf) are used to represent the importance of terms in classifying a text document. A term's weight can be represented by the product of Tf and Idf:

$W = Tf * Idf$ (Tf-Idf method) or sometime is represented by Tf only (Tf method) for simple calculation.

$Idf = \ln(N/n)$, N refers to the number of documents in topic set, n refers to the number of documents that the term appears in.

In the first two methods during the test, all terms are retrieved without lexical distinguishing. While in the last three methods, noun and non-noun terms are separated in the process of feature selection. In general, concerning the result, the precision and recall average of last three methods are higher than the first two, and it proofs that the method splitting noun and non-noun terms can improve the performance of feature selection.

In the first two methods, the precision of Tf-Idf is higher than Tf, and in all five methods, using Tf-Idf on noun terms and non-noun terms respectively performs the most efficiency. Since the Tf-Idf considers both term frequency (Tf) and inverse document frequency (Idf), therefore its performance of feature selection is better than Tf. However, the Tf-Idf method is time-consuming. During the test, when scoring a 1000-word document for 1000 times, Tf finished in 11.05ms, while Tf-Idf consumed 3396.41ms, about 300 times the time Tf used. From the table above, it can be seen that the combination of Tf (non-noun) and Tf-Idf (noun) performs nearly as well as the last method does, but saves a lot of time. In the real application, when processing large amount of textual data, balancing the time and efficiency is necessary. Therefore, in this research, the method of combination of Tf (non-noun) and Tf-Idf (noun) is used for feature selection, in the personalized recommendation system.

Module design

In personalized information retrieval system, there are two main tasks of the modules of user models. The first is, according to the text collections that users supplied, to create the user profile through learning algorithm. The other is to supply retrieval basis, the information of user interest, to personalized information retrieval system.

In the user multi-interest modeling proposed in this research, the task of modeling is achieved through user interest management module, which mainly contains two parts, the user interest manager and user database manager. User interest manager creates user profile and gathers information of user interests, while user database manager is designed to read and write the user profile. Other assisted class, such as SSN manager and match engine (Turney, et al., 2006), are also necessary in the module design, as shown in Figure 5.

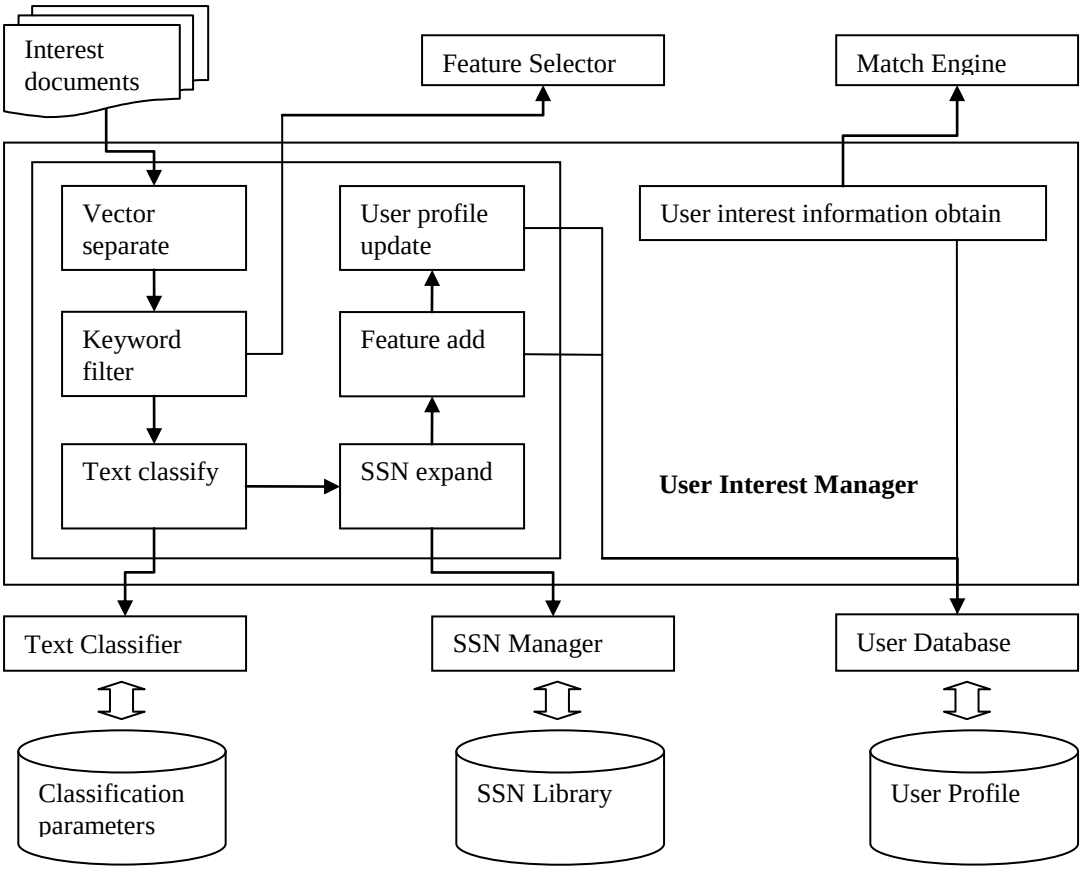

Figure 5. Module structure of user modeling based on semantic similar network

As shown in Figure 5, the flow of user interest manager processing is as follows: At first, the terms in the interest documents are separated by the method of Vector Separate, into noun vectors and non-noun vectors; then two different feature selection methods are used during Keyword Filter, though Feature Selector. For noun vectors, Tf-Idf (Inverse document frequency) is adopted while Tf (Term frequency) is used on non-noun vectors; after that a set of keywords are obtained, then these keywords are classified into categories in the class Text Classify using Text Classifier; next some of the keywords are expanded through SSN Manager, in the process SSN Expand; then according to the information of term categories, feature vectors and expanded feature vectors are added into user profile. In the process of Feature Add, a mark value 1 is used if a term is directly retrieved from the documents, and a mark value 0 is specified for expanded terms; at last, in the method User Profile Update, features with relatively low weights in the user profile are eliminated, to avoid the volume excess.

In addition to the construction of user profile, user interest management module also provides the user interest information that it created before, to the match engine in the personalized system. In this research, it is implemented by the method of User Interest Information Obtain (UIIO). UIIO accesses user profile though the interface of user database, and then extracts user interest information according to the request of match engine. If an accurate query is requested, then only the feature vectors with SSN expansion mark value 1 are selected; if the match engine requires a fuzzy match using all relevant interest information, then all feature vectors, with both SSN expansion mark value 1 and 0, will be read for the query.

CONCLUSION

The explosive growth of information has severely restricted users' efficient use of resources. In response to the more and more serious situation of "information overload" and "information confusion" on the internet, intelligent information retrieval and personalized information service currently become the focus of information service studies. However, the base of the intelligent and personalized

information retrieval system cannot get away with the "personalized user profile", and therefore the creation and application of user profile has become an active and broad research field in personalized information services.

With the development of concepts and techniques in artificial intelligence and machine learning fields, a variety of information retrieval systems have come forth continuously in recent years. At the same time, techniques related to information retrieval are rapidly developing, such as the web mining techniques, user activities research methods and different neural network algorithms, which all provide supports to the research of user interest profile. How to capture user actual and continuously changing requirements, how to describe and represent user requirements in the semantic knowledge level, and how to accurately create and real-timely update user interest profile, are all difficulties needed to be solved in the research. Not reflecting the diversity and implementing the adaptive update of the user interest, at present, are disadvantages of the traditional vector space model. Aiming at the shortcomings of VSM representation, this paper improves the traditional vector space model, and proposes a multi-interest model based on SSN. In this research, information resource is in text type, and therefore the modeling techniques is suitable to other applications in the text information retrieve field, for instance, the intelligent web searching engine, personalized services in digital library, recommendation services of personalized news, etc. Thereby, the investigation in this paper contains theoretical and practical value.

Improvement of modeling techniques has exploited new research direction for personalized services (Shahabi and Chen, 2003; Hofmann, 2004), for instance the research oriented to groups or communities. Single user model predominates the present user models; however, it shows more significance modeling with a group of users in some circumstance. Digital TV program recommendation becomes more and more popular, and usually the using object is a family. How to coordinate the preference in family members with different ages, different occupations, and different knowledge contexts, as well as the priority of selecting programs, is the primary problem in the modeling oriented to group users.

REFERENCE

1. Badi, R., Bae, S., Moore, J.M., Meintanis, K., Zacchi, A., Hsieh, H., Shipman F. and Marshall, C.C. (2006). Recognizing user interest and document value from reading and organizing activities in document triage. *Proceedings of the 11th international conference on Intelligent user interfaces*, pages 218-225.
2. Baeza-Yates, R. and Ribeiro-Neto, B. (1999). Modern Information Retrieval. New York: Addison-Wesley-Longman.
3. Cavnar, W.B. (1995). Using an N-gram-based document representation with a vector processing retrieval model. *Proceedings of the Third Text Retrieval*, pages 269-277.
4. Debole, F. and Sebastiani, F. (2003). Supervised Term Weighting for Automated Text Categorization. *Proceedings of the 18th ACM Symposium on Applied Computing*, pages 784-788.
5. Fisher, M. and Everson, R. (2003). Representing interests as a hyperlinked document collection. *Proceedings of the twelfth international conference on Information and knowledge management*, pages 378-385.
6. French, J.C. and Viles, C.L. (1999). Personalized information environments: An architecture for customizeable access to distributed digital libraries. *D-Lib Magazine*, *5*, No. 6.
7. Friedman, N., Geiger D. and Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, *29*, 131-163.
8. Gardner, E.J. and Derrida, B. (1988). Optimal storage properties of neural network models. *Journal of Physics A 21*, 271-284.
9. Godoy, D. and Amandi, A. (2005). User profiling for web page filtering. *IEEE Internet Computing*, *9*, 4, 56-64.
10. Hadrich, T. and Priebe, T. (2005). A Context-Based Approach for Supporting Knowledge Work with Semantic Portals, *International Journal on Semantic Web and Information*, *1*, 3, 64-88.
11. Hliaoutakis, A., Varelas, G., Voutsakis, E., Petrakis, E.G.M. and Milios, E. (2006). Information Retrieval by Semantic Similarity, *International Journal on Semantic Web and Information Systems*, *2*, 3, 55-73.
12. Hofmann, T. (2004). Latent Semantic Models for Collaborative Filtering. *ACM Transactions on Information Systems*, *22*, 1, 89-115.
13. Huang, C., Tseng, T. and Liang, H. (2006). Rough-set-based approach to manufacturing

process document retrieval. *International Journal of Production Research*. *44*, 14, 2889-2911.

14. Java, A., Nirenburg, S., McShane, M., Finin, T.W., English, J. and Joshi, A. (2007). Using a Natural Language Understanding System to Generate Semantic Web Content, *International Journal on Semantic Web and Information Systems*, *3*, 4, 50-74.
15. Kalyanpur, A., Parsia, B. and Hendler, J.A. (2005). A Tool for Working with Web Ontologies, *International Journal on Semantic Web and Information*, *1*, 1, 16-49.
16. Kim, H.R. and Chan, P.K. (2003). Learning implicit user interest hierarchy for context in personalization. *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 101-108.
17. Korfhage, R.R. (1997). Information Storage and Retrieval. Wiley, 368 pages.
18. Kuropka, D. (2004). Modelle zur Repräsentation natürlichsprachlicher Dokumente. Ontologie-basiertes Information-Filtering und -Retrieval mit relationalen Datenbanken. Advances in Information Systems and Management Science, bd.10, 264 pages.
19. Lan, M., Tan, C., Low, H. and Sung, S. (2005). A comprehensive comparative study on term weighting schemes for text categorization with support vector machines. *Proceedings of the 14th international conference on World Wide Web*, pages 1032-1033.
20. Lehmann, F. (Ed.) (1992). Semantic Networks in Artificial Intelligence, Pergamon Press, Oxford.
21. Liang, T., Lai, H. and Ku, Y. (2007). Personalized content recommendation and user satisfaction: Theoretical synthesis and empirical findings. *Journal of Management Information Systems*, *23*, 3, 45-70.
22. Luce, R. and Giacomo, M.D. (2004). Personalized and collaborative digital library capabilities: responding to the changing nature of scientific research. *Science & Technology Libraries*, *24*, No. 1/2, 135-152.
23. Middleton, S.E., Shadbolt, N.R. and De Roure, D.C. (2004). Ontological User Profiling in Recommender Systems. *ACM Transactions on Information Systems*, *22*, 1, 54-88.
24. Mitchell, M. (1996). An Introduction to Genetic Algorithms, MIT Press, Cambridge, MA.
25. Mozina, M., Demsar, J., Kattan, M. and Zupan, B. (2004). Nomograms for Visualization of Naive Bayesian Classifier. *Proceedings of PKDD-2004*, pages 337-348.
26. Nasukawa, T. and Nagano, T. (2001). Text analysis and knowledge mining system. *IBM Systems Journal*, *40*, 4, 967-984.
27. Raghavan, V.V. and Wong, S.K.M. (1986). A critical analysis of vector space model for information retrieval. *Journal of the American Society for Information Science*, *37*, 5, 279-287.
28. Rish, I. (2001). An empirical study of the naive Bayes classifier. *Proceedings of IJCAI-01 Workshop on Empirical Methods on Empirical Methods in Artificial Intelligence*.
29. Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, *24*, 5, 513-523.
30. Salton, G., Wong, A. and Yang, C.S. (1975). A Vector Space Model for Automatic Indexing. *Communications of the ACM*, *18*, 11, 613-620.
31. Shackleford, T.C. (2005). Using data mining techniques to develop measures of document relevance. George Mason University.
32. Shahabi, C. and Chen, Y. (2003). An Adaptive Recommendation System without Explicit Acquisition of User Relevance Feedback. *Distributed and Parallel Databases*, *14*, 2, 173-192.
33. Shepherd, M., Duffy, J.F., Watters, C. and Gugle, N. (2001). The Role of User Profiles for News Filtering. *Journal of the American Society for Information Science and Technology*, *52*, 2, 149-160.
34. Sheth, A.P., Ramakrishnan, C. and Thomas, C. (2005). Semantics for the Semantic Web: The Implicit, the Formal and the Powerful, *International Journal on Semantic Web and Information Systems*, *1*, 1, 1-18.
35. Singh, R., Iyer L.S. and Salam, A.F. (2005). Semantic eBusiness, *International Journal on Semantic Web and Information*, *1*, 1, 19-35.
36. Singhal, A. (2001). Modern Information Retrieval: A Brief Overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, *24*, 4, 35-43.
37. Turney, P.D. (2006). Similarity of Semantic Relations. *Computational Linguistics*, *32*, 3, 379-416.
38. Turtle, H.R. and Croft, W.B. (1992). A comparison of text retrieval models. *The Computer Journal*, *35*, 3, 279-290.
39. Vallet, D., Fernandez, M. and Castells, P. (2005). An Ontology-Based Information Retrieval Model. *Proceedings of the Second European Semantic Web Conference, ESWC 2005*, pages 455-470.

40. Wang, X., Ju, S. and Wu, S. (2008). Challenges in Chinese Text Similarity Research. *2008 International Symposiums Information Processing*, pages 297-302.
41. Zhang, Y. (2008). Complex adaptive filtering user profile using graphical models. *Information Processing and Management: an International Journal, 44*, 6 1886-1900.