Journal of The Korea Society of Computer and Information
Vol. 20, No. 1, January 2015
www.ksci.re.kr
http://dx.doi.org/10.9708/jksci.2015.20.1.029

A Multi-Agent MicroBlog Behavior based User Preference Profile Construction Approach

Jee-Hyun Kim*, Young-Im Cho**

Abstract

Nowadays, the user-centric application based web 2.0 has replaced the web 1.0. The users gain and provide information by interactive network applications. As a result, traditional approaches that only extract and analyze users' local document operating behavior and network browsing behavior to build the users' preference profile cannot fully reflect their interests. Therefore this paper proposed a preference analysis and indicating approach based on the users' communication information from MicroBlog, such as reading, forwarding and @ behavior, and using the improved PersonalRank method to analyze the importance of a user to other users in the network and based on the users' communication behavior to update the weight of the items in the user preference. Simulation result shows that our proposed method outperforms the ontology model, TREC model, and the category model in terms of 11SPR value.

▸ Keywords : Interactive, Multi-agent, MicroBlog, PersonalRank, Preference, Ontology

•제1저자 : Jee Hyun Kim •교신저자 : Young Im Cho
•투고일 : 2015. 1. 13, 심사일 : 2015. 1. 20, 게재확정일 : 2015. 1. 27.
* 서일대학교 컴퓨터소프트웨어과(Department of Computer Software, Seoil University)
** 수원대학교 컴퓨터학과(Department of Computer Science, The University of Suwon)
※ The present research has been conducted by the Research Grant of Seoil University in 2014.

I. Introduction

With the rapid development of internet users and accessible web pages, the amount of information on the internet is growing exponentially. How to effectively get the required and the most relevant documents from the mass of information which is in all types of formats has become a serious problem. Today one of the most common effective ways to solve this problem is to get the end user's preference by recording the user's context and personalized applications, based on which to tailor the user′s needs according to their interests.

There are two kinds of behavior information that are meaningful to be used to analyze the users' interests: one is local user behavior which includes reading habit of local documents, reading duration of local document and information of application program windows switching operation; the other one is the users' network operating behavior information.

The pure ontology-based user preference profile technique is mostly limited to taxonomies of users' interests, while some other network operating behavior information analysis and mining based methods mainly abstract the users' interested web pages by considering the uses' browsing behavior, such as click frequency, reading duration, adding to favorites or not, sliding-time window, forgetting function and so on.

So many researchers devoted themselves to build a dynamic user preference profile by comprehensively considering the knowledge structure (such as Ontology and WordNet) and the users' browsing behaviors, based on which to retrieve more relevant document, insomuch as to improve the precision significantly.

These knowledge and user daily network operating behavior based user preference construction approaches achieved significant results in improving the precision and recall ratio in the information centered web 1.0 time, but with the web 2.0 time' arrival, only depending on the users' one way network operation, such as clicking, browsing cannot gain the users' overall personal preferences.

Web 2.0 is a user-centric application and the information is spread around the user. They are not only the gainer of the information, but also the provider of the user. Interactivity is the important characteristic of the web 2.0 based network time. So the users' operating behavior on the web 2.0 is based on network applications(such as MicroBlog, Social Networking Services and Instant Messaging) show the significance for the construction of dynamic user preference.

The first innovation of this paper is that, construction of a static personalized ontology user preference is based on the documents stored in the users' local devices and the user device active degree, which is an extension part of the domain ontology by comprehensively considering the knowledge structure and the users' local behavior.

The second innovation of this paper is that, we firstly extracted the information such as personal information, MicroBlog text, user relationship, user communication information, user impact from MicroBlog; then we use the improved PersonalRank method to analyze the importance of a user to other users in the network and update the weight of the items in the user preference based on the users' communication behavior such as forwarding behavior, and commented behavior.

The rest of our paper is organized as follows: section 2 gives a short review of the related work. Section 3 describes the framework and the realization of our system. Section 4 gives the simulation based on our proposed method and we arrive at the conclusion in Section 5.

II. Related Works

Most personalization systems are based on some type of user profile, and they are typically built from topics of interest to the user, and are generally represented as sets of weighted keywords, semantic networks, or weighted concepts, or association rules.

The simplest user preference is the keyword preference profiles which have to capture and represent all (or most) words by interests which may be discussed in future documents. Amalthaea [1] weighted the keywords with the widely used tf × idf weighting scheme from information retrieval. Salton [2] employs a learning algorithm based on genetic algorithms to adapt and expand the user profiles. Some other researchers use Latent Semantic Indexing (LSI)[3] and Linear Least Squares Fit (LLSF) [4] to create the keyword-based feature vectors.

Semantic user profiles can explicitly model the relationship between particular words and higher-level concepts, which have an advantage over keyword-based profiles. The items in semantic user profile are more accurate because they represent concepts rather than individual words. Ontology based user preference profile has already been proposed in various applications, such as web search [5] and personal information management [6]. Gauch [7], Chaffee [8] and Trajkova [9] focus on creating user profiles based on ontology automatically. They encoded the information by giving a well-defined meaning with predefined ontology which consists of term descriptions and their interrelationships that allow making inferences and retrieving more relevant information than the keyword-based search. Susan Gauch [10] used classification techniques to create profiles for users on web sites automatically. The Semantic Web approaches only handle unlabelled links (assumed to represent parent-child relationships) between concepts rather than represent a wide variety of link types and link labels between concepts.

Actually, the amount of internet applications has been used in the construction of user preferences to help users solve the problem of information overload, and to provide the personalized recommendation service to different users. Shi Y.C. [11] proposed an adaptive context learning method for mobile user preferences to judge whether user's preferences are influenced by context introduced to the least squares support vector machine (SVM), and further to extract the user preferences. It is analyzed the cause of the change of user preference from the perspective of motivation, and in time discovered the change of the preference by analyzing the user behavior records and the users' feedback to make active response. Wang J.H [12], by using MicroBlog as platform, built the user preference profile based on cp-nets tools, and provided the users with forwarding decision metric by induced graph.

III. The Proposed Method

There are four agents in our entire system, with all agents communicating among each other, and each agent having its own responsibility, shown as Fig.1.

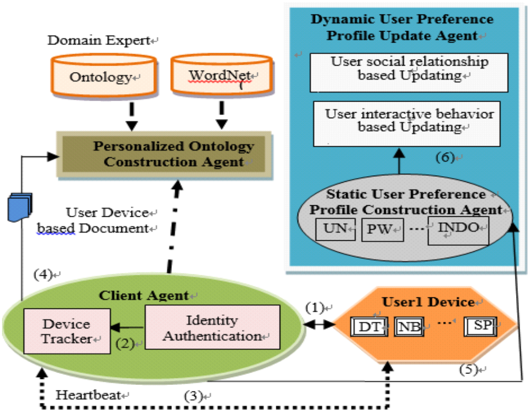

Fig. 1. Framework of Multi-Agent MicroBlog Behavior based User Preference Profile Construction Approach

1. Function of each agent in the proposed approach

First, we use Client Agent to verify user's identities by creating a union user account, and monitor whether a device is ready to operate a personalized ontology profile constructing task.

The Client Agent is responsible for verifying the user's identity (in the form of "UserID"), calculating the Active Degree (in the form of "UserID.ADj", where j is the index of the device) of different user devices, collecting the context information (such as metadata and users' daily working behavior) from all the smart devices possessed by a same user.

Second, we use a five-level Personalized Ontology Construction Agent to create a personalized ontology profile, which is an extension of domain ontology that considers both the Knowledge Structure and the users' local operation behavior.

The proposed personalized ontological profile is an extension of domain ontology. The ontology is extended with co-occurrence terms, semantically related concepts and highly related user device document terms. The terms are given weights to reflect the importance of the semantic relation between the concept and the terms

For the first level, a Domain Classification Algorithm (DCA) used to identify which domain the document stored in the user device belongs to. For the second level, a Concept Identification Algorithm (CIA) used to identify which concept in the WordNet is the corresponding one containing the words in the document. For the third level, we use the frequency of the terms spotted in the documents of the device alive to find the related terms for each concept, based on which it weighs each terms. For the forth level, we rely on the active degree to reweigh each term of the corresponding concept so as to indicate the activeness of different smart devices that belong to the same user. For the fifth level, we add the semantic related terms into the ontology profile according to the WordNet.

In our paper we define a personalized ontology profile like the following. This is a Concept Vector with each concept in the ontology with weighted related terms, i.e., for the same domain ontology, different users may have different ontology profiles including the same concepts but different weighted related terms. Suppose C is the concept set of the corresponding ontology. T is the set of n terms in the document collection used for construction of the personalized ontological profile. $t_i \in T$ denotes related weighted term i in the set of terms. Then the ontology profile for concept j is defined as the vector $C_j = [w_1, w_2, . . . , w_n]$ where each w_i denotes the semantic related weight for each term t_i with respect to concept C_j.

Third, we use Static User Preference Profile Construction Agent to create the initial static user preference profile according to his/her input information, shown as Tab.1.

Tab. 4. The example of user preference profile table

User preference with in a particular domain								
User name	password	Category 1 with its associated weight			…	Category n with its associated weight		
			…		…		…	

Finally, we use a two-level Dynamic User Preference Profile Update Agent to update the user's interest degree by comprehensively considering the users' interactive behavior and social relationship on web 2.0 application of MicroBlog. We split the users' preference into two types: one is short-term preference and the other one is long-term preference. Then we employ two different dynamic update strategies (Sliding-Time-Window Update Strategy for short-term preference, Time-Based-Forgetting Function Update Strategy for long-term preference) to update and readjust the weight of the short-term and long-term preference.

Each profile was initially built with concepts from the personalized ontology profile and then dynamically updated by the user's long term behavior and short term behavior.

Although multiple profiles were created, to lessen the burden on our subjects, we merged all concepts from all techniques into a single profile on which to collect judgments.

The profile was presented to the subject, and they were asked by the testers to identify the non-relevant concepts present in their profiles by crossing them out.

2. MicroBlog Content Based User Preference Updating Agent

User interests always constitute the most important part of the user profile in adaptive information retrieval system; meanwhile dynamic user profile is used to address frequently changed user interest.

2.1 User interactive behavior based analysis

Before Before the MicroBlog context based updating method, the stemmed words were extracted from the MicroBlog context by performing Part of Speech (POS) (SNLPG), Stop Words Removal, and Words Stemming. The feature words extraction approach is shown in Fig. 2.

Based on the feature words, we use browsing behaviors updating subagent to construct the user preference profile. If a user clicks a web page, it means that the user may have interest in this web page. However the degree of user interest in this web page depends on the series of behaviors after opening the web pages.

For each stemmed word we use the "Semantic Similarity Matching Algorithm" [12] to look up the most similar concept between the stemmed feature words set MBV {(mb1, w1),(mb2, w2), …, (mbn, wn)} and the static personalized ontology based user preference profile POPDi[13]. Here mbi is stemmed feature words in the content the MicroBlog, and wi is the weight of feature word mbi.

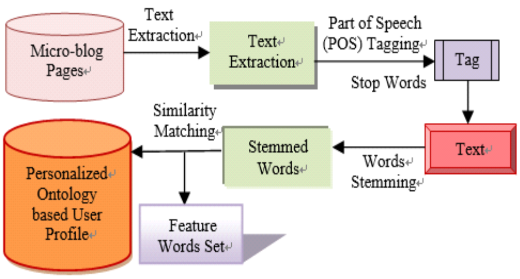

Fig. 2. Feature words extraction approach

For one stemmed word, if it has no similar concept in the personalized ontology based user preference profile, then discard this word from the feature word set MBV, otherwise use Eq. (1) to calculate the weight of the feature word mbi[14].

$$W_i = \beta \times tf_{fw,mb} \times idf_{fw,mb} \tag{1}$$

Here $tf_{fw,mbj}$ is the times of the feature word fw_i occurs in MicroBlog mb, $df_{fw,mb}$ is the number of the MicroBlog that was read by the user containing the feature word fw_{ij} at least once and the $idf_{fw,nb}$ is the inverse document frequency which can be calculated by Eq.(2) and β is a normalization constant.

$$idf_{fw,mb} = \log \frac{|mb|}{df_{fw,mb}} \tag{2}$$

Here |mb| is the number of the MicroBlogs the user has read.

MicroBlog provides a communication platform for different users that can interact with any users that they want to communicate, and then produce large amounts of information interaction behavior, such as users forwarding behavior, comment behavior and @ behavior. Statistics found that for the same MicroBlog content, if it was forwarded by the person with close ties with the target user or by the person with high acceptance, then the higher probability the MicroBlog would be forwarded or commented by the target user. That is to say, the feature words that occur in the above MicroBlogs have a more important impact on users' preference than other feature words.

We use vector $FW(u)=(f_0,f_1,f_2,\cdots,f_n)$ to indicate the number of the MicroBlog which were created by user $i(i=0,1,2,\cdots,n)$ and were forwarded by user u. $CT(u)=(C_0,C_1,C_2,\cdots,C_n)$ to indicate the number of the MicroBlog which were created by user $i(i=0,1,2,\cdots,n)$ and were commented by user u. $A(u)=(A_0,A_1,A_2,\cdots,A_n)$ to indicate the number of the MicroBlog which were created by user $i(i=0,1,2,\cdots,n)$ and were @ by user u. Then we calculate the importance of the MicroBlog for user u which was created by user i based on the interactive information, as shown in Eq.(3).

$$IT_u(IT_0,IT_1,IT_2,\ldots,IT_n)=\frac{FW(u)\times K1+CT(u)\times K2+A(u)\times K3}{FW(u)+CT(u)+A(u)} \quad (3)$$

Here, K_1, K_2, K_3 ($K_1 > K_2 > K_3$, and $K_1+K_2+K_3=1$) is the weighting parameters to indicate the importance of the interactive behavior.

2.2 User social relationship based analysis

In reality, users' social relationship plays an important role in the analysis of users' interest. Two users with close relationship tend do have similar preference. Here we used improved PersonalRank algorithm to analyze the importance of a user to other users in the network. It supposes that the users browse from one MicroBlog page to another page randomly and continuously jump through hyperlinks, so as to form a random walk process. After plenty of random walk, the probability of each page was accessed by a certain user tend to reach a stable value, thus this value can represent the importance of the page in the network. We use iterative Eq.(4) to calculate the importance of the MicroBlog pages I.

$$IMB_i=\begin{cases}(1-d_j)+d_j\sum\limits_{t\in in(i)}\frac{IMB_t}{|out(t)|} & if \quad i==start \quad point\\ d_j+\sum\limits_{t\in in(i)}\frac{IMB_t}{|out(t)|} & if \quad i\neq start \quad point\end{cases} \quad (4)$$

Here IMBi is the probability of MicroBlog page i that would be accessed by a user, dj is the probability of a user that continue random walks when he reached page j. Thus when the probability of a page that would be accessed tend to be stable, the value of IMBi will be the similarity degree of page i with start page. The higher the value of IMBi, the greater the page i can reflect the users' preference.

Finally, we update the weight of the user preference profile based on Eq(3) and (4), as shown in Eq.(5)

$$W_i=\ln\left(\frac{average\left(\sum\limits_{fw\in mb_t and \in pages\ created\ by\ user_j}W_i \times IT_u(j)\times IBM_t\right)}{N(fw)}\right)\times PV(fw) \quad (5)$$

N(fw) is the number of the MicroBlog that contains feature word fw, PV(fw) is the weight of the item fw in the static user preference profile which was calculated by personalized ontology based user preference profile according to users' local behavior.

IV. Simulation

In this simulation, we choose a branching factor of double with a depth of three levels in the hierarchy of the Open Directory which contained 153 concepts in the hierarchy, a total of 1837 documents indexed and 300 MicroBlog pages under various concepts.

The experimental data set were pre-processed and divided into two separate sets including a training set, and a test set. The training set consist of 895 documents and 126 MicroBlog pages used for the learning of the personalized ontology based user preference profile, meanwhile the concept terms and corresponding term weights were computed using the equation described in section III. A total of 942 documents and 174 MicroBlog pages were included in the test set

The performance of the proposed approach was measured by the precision averages at 11 Standard Recall Levels (11SPR). The 11SPR value is computed by summing the interpolated precisions at the specified recall cutoff, and then divided by the number of topics, as shown in Eq. (6)

$$\frac{\sum_{i=1}^{N} precision_{\lambda}}{N}; \qquad \lambda = \{0.0, 0.1, 0.2, 0.3, \ldots, 1.0\} \tag{6}$$

Here N denotes the number of topics, and λ indicates the cutoff points where the precisions are interpolated. We compared our system with ontology model, TREC model, and the category model, the simulation result on 11SPR is shown as Fig. 3.

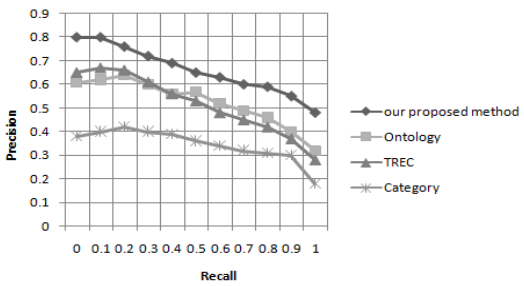

Fig.3. The11SPR results of the proposed method versus the other 3 approaches

Where the 11SPR curves show that our personalized ontology model is better than the other 3 approaches. The simulation result shows that our personalized search provides the user with results that more accurately satisfy their specific goal and intent for the search. The queries used in our experiments are intentionally designed to be short in order to demonstrate the effectiveness of our personalized information retrieval approach, especially in the typical case of web users who tend to use very short queries.

From the figure we can see that the ontology model discovered user background knowledge from user local instance repositories, rather than documents read and judged by users. Thus, the ontology user profiles were not as precise as the TREC user profiles.

For the TREC model, every document in the training sets was read and judged by the users, which ensured the accuracy of the judgments. The topic coverage of TREC profiles was limited. However, the ontology profiles had a broad topic coverage. The substantial coverage of possibly-related topics was obtained from the use of the world knowledge base which is a taxonomy constructed as a directed acyclic graph, and the large number of training documents.

As a result, when considering precision results, the TREC model's 11SPR performance was better than that of the ontology model. However, when taking into account recall results compared with the TREC model, the ontology model had better recall but relatively weaker precision performance.

Based on the user preference profile, the proposed personalized information retrieval has a broad topic coverage, takes into account the user's reading behavior of local documents and website browsing, and also makes up the shortfall of the ontology model. Therefore, the proposed method can achieve better precision and recall.

Because the proposed ontology and MicroBlog reading and forwarding behavior based personalized information retrieval has broader topic coverage, and comprehensively considers the user's reading behavior of local documents and MicroBlog browsing behavior, so it can make up the shortfall of the ontology model. Therefore, the proposed method can achieve better precision and recall than the other 3 models.

In order to evaluate the performance of long term preference and short term preference, we also compare our system results with another two methods: one is the pure sliding window method which can grasp the user's short-term performance more accurately based on the recent data as reference, and the other one is the pure forgetting strategy method which mainly considers the user's previous interest factor. Fig.4. shows the precision ratio at the recall ratio of 0.5.

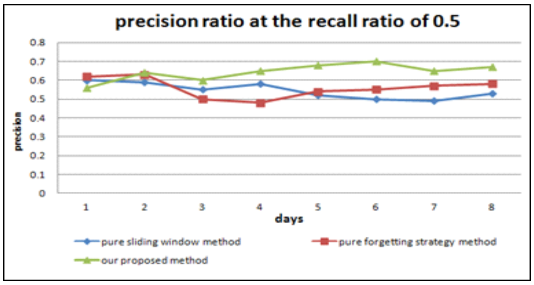

Fig.4. The precision ratio of our proposed method at the recall ratio of 0.5

The proposed our method just considers the ontology structure and the constantly changing document information stored in the user's local smart device, and does not consider the user's browsing behavior, clicked behavior and the user's feedback, therefore, the personalized ontology profile itself cannot reflect the user's preference very accurately.

The user's preference can be divided into long term preference and short term preference, which are not considered by personalized ontology profile, so the precision averages at 11 standard recall levels based only on personalized ontology profile based information retrieval are worse than those based on the TREC when the recall values are not very high. However, with the increase of the recall values, the TREC model's performance drops and the precision is close to the proposed personalized ontology profile based method.

Ⅴ. Conclusion

Interactivity is very common between different users, therefore in this paper, we proposed a preference analysis and indicating approach based on the users' MicroBlog reading, forwarding and @ behavior, so as to dynamically update the ontology based user preference profile based on both the users' local behavior and web 2.0 network application behavior.

Firstly, we create a personalized ontology profile which is an extension of domain ontology by comprehensively considering the Knowledge Structure and the users' local operation behavior. Then we extracted the information such as personal information, MicroBlog text, user relationship, user communication information from MicroBlog and used the improved PersonalRank method to analyze the importance of a user to other users in the network and update the weight of the items in the user preference based on the users' communication behavior.

Simulation result shows that our proposed method outperforms the ontology model, TREC model, and the category model in terms of 11SPR value. In the future, we will research more refined platform-based personalized information retrieval system.

참고문헌

[1] Moukas A., "Information Discovery and Filtering Using a Multiagent Evolving Ecosystem," *Applied Artificial Intelligence, An International Journal*, Vol.11. pp.437-457, 1997.

[2] Salton G. and McGill M., "Introduction to Modern Information Retrieval," McGraw-Hill,

Inc. New York, NY, USA., 1983.

〔3〕 Deerwester S., Dumais S., Furnas G., Landauer T.K. and Harshman R., "Indexing by Latent Semantic Analysis," *Journal of the American Society for Information Science*, Vol.41. No.6. pp.391-407, 1990.

〔4〕 Liu F., Yu C. and Meng W., "Personalized web search by mapping user queries to categories," *In Proceedings the eleventh international conference on Information and knowledge management*, pp.558-565, 2002.

〔5〕 Trajkova J. and Gauch S., "Improving Ontology-based User *Profiles," Proc. of RIAO 2004, University of Avignon (Vaucluse)*, pp.380-389, 2004.

〔6〕 Katifori V., Poggi A., Scannapieco M., Catarci T. and Y. Ioannidis., "OntoPIM: how to rely on a personal ontology for Personal Information Management," Semantic Desktop Workshop of CEUR Workshop Proceedings, Vol.175, 2005.

〔7〕 Gauch S., Chaffee J. and Pretschner A., "Ontology Based Personalized Search," In Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), pp.391-398, 1999.

〔8〕 Chaffee J. and Gauch S., "Personal Ontologies For Web Navigation," In Proceedings of the 9th International Conference On Information Knowledge Management (CIKM), pp.227-234, 2000.

〔9〕 Trajkova J. and Gauch S., "Improving Ontology-based User Profiles," *Proc. of RIAO 2004, University of Avignon (Vaucluse)* , pp.380-389, 2004.

〔10〕 Gauch S., Speretta M. and Pretschner A., "Ontology-Based User Profiles for Personalized Search," Integrated Series in Information Systems, Vol.14. pp.665-694, 2007.

〔11〕 Shi Y.C, Meng X.W, Zhang Y.J, "Adaptive learning approach of contextual mobile user preferences," Journal of software, Vol. 23, No.10. pp. 2533-2549, 2012.

〔12〕 Wang J.H, He L.J, Zhang W, "Research of user preference based information dissemination in microblog," Computer engineer and science, Vol.35, No.3, pp. 140-144, 2013.

〔13〕 Youn-Hee Kim, Jee-Hyun Kim, "The Scheme for Path-based Query Processing on the Semantic Data," Journal of the Korea Society of Computer and Information, Vol. 14, No.10, pp.31-41, 2009.

〔14〕 Youn-Hee Kim, Jee-Hyun Kim, "A Query Processing Method for Hierarchical Structured e-Learning System," Journal of the Korea Society of Computer and Information, Vol.16, No.3, pp.189-201, 2011.

저 자 소 개

Jee Hyun Kim

1978: B.S at Ewha Womans Univ.

1994: MBA from the Business Admin. Dankook Univ.

2005: Ph. D from the Computer Science, Dankook Univ.

Now : Professor at Seoil Univ.

Interesting area: SQM, DB model

Email : jhkim@seoil.ac.kr

Young Im Cho

1988: B.Sc. at Korea University

1990: Master from Dept of Computer Science, Korea University

2004: Ph.D. from Dept of Computer Science, Korea University

Now: Professor at Suwon Univ.

Interesting area : AI, Big data

Email : cyi@suwon.ac.kr