Proceedings of the International MultiConference of Engineers and Computer Scientists 2008 Vol I
IMECS 2008, 19-21 March, 2008, Hong Kong

The Blog-article Recommendation System(BARS)

Li-Hua Li, Fu-Ming Lee, and Shang-Chi Chan, *Member, IAENG*

***Abstract*—According to Technorati's [1] 2007 report, the blog sites is up rising to 72 millions and the popularity of blogers has drawn many attention. This phenomenon has turned many web users into bloggers. The vast amount of blog information also brings the phenomenon of information overloading which is not handled by the blog function yet. In addition, the personalized recommendation service, which should be provided, is also not incorporated in the blog function now.**

To better service the bloggers and to overcome the above problems, this research proposes a Blog Article Recommendation System (BARS) which provides personalized article recommendation based on blogger's interests.

This research adapts the ontology technique in BARS to construct a personal preference tree for understanding blogger's interests. The ART (Adaptive Resonance Theory) network is also utilized to cluster the group with similar interests. In order to find the similar preference between target blogger and the corresponding neighbors, we apply the Collaborative Filtering (CF) technique to generate the recommendation. The "cold-start" problem, i.e. lacking of blogger's usage data at the very beginning, is handled by combining ontology and Content-Based (CB) filtering method to infer the potential preference in BARS. The purpose of this research is to achieve the followings.

(1) To solve the problem of information overloading.
(2) To propose the cold-start problem when making the recommendation.
(3) To build the BARS for personalized blog-article recommendation.

***Index Terms*—Blog, Recommendation System (RS), Adaptive Resonance Theory (ART), Ontology.**

I. Introduction

According to the Technorati 2007 reports [1], the history of blogs is less than five years, however, the blog culture has spread around the world beforehand. There are merely 8 millions blogs on March 2005, but in the end of March 2007, the blogs has been raising to 72 millions. This phenomenon has turned many web users into the faithful bloggers.

Due to the vast amount of information on a popular blog, it is often time consuming for reviewing and finding the blog-article to suit the bloger's mind. We often called this type of problems as information overloading and non-personalized service problem. To handle the information overloading problem, Belkin et al.[2] proposed a technique by using the information filtering and retrieval technique to solve the information overloading problem, however, their study did not incorporate the personalized service. The studies of [3][4][5] also proposed methods for providing personalized service and resolving the information overloading problem, however, these researchers did not consider the similar-preference neighbors to enhance efficacy of recommendation.

In order to meet the previous requirements, i.e., to provide the personalized service and to solve the information overloading problem, this research uses an academic blog as the research platform and proposes the BARS (Blog Article Recommendation System) for personalized blog-article recommendation. The detailed account of reasons of proposing BARS are given as below.

1) The categories of blog-article might represent diverse meanings on various blog-portal [6]. This usually creates the synonym problem and, therefore, hard to recommend by simply using the defined category.
2) The academic blogs have common domain categories. Articles highly recommended by the reviewers are often valuable to share; therefore, this research uses the academic blog as the platform for doing the blog-article recommendation research.

The proposed BARS utilizes two major technique, one is Collaborative Filtering (CF) technique, and the other is Content-based (CB) method. CF is a technique that relies on the group opinion or group preference to perform the recommendation. CF has become popular because it is highly used for providing personalized recommendation [7]. CB method is to recommend items based on user's profile or the requested content by measuring the similarity between items (contents).

In order to understand the blogger's interest, the user profiles are often acquired. The aim of using user profile is to understand the user preference and the historical browsing interests. To construct user profile, Middleton et al. [3] has proposed an ontology-based system, named Quickstep, to represent user interests with tree structure. Ontology is a technique that can help referring user's potential preference according to the user profile. It is, therefore, this research adopts the ontology technique to construct the personal preference tree to infer the user preference so that the recommendation can meet the personal interest.

The purpose of this research is to achieve the followings.

1) To provide the personalized blog-article recommendation for bloggers and to resolve the information overloading problem.
2) To reduce the cold-start problem by using the ontology to perform the inference and using the CB technique to find the similar neighborhood for recommendation.
3) To express the blogger's profile in compliance with ontology personal preference tree (OPPT).
4) To discover the user potential preference according to

Li-Hua Li is with the Chaoyang University of Technology, Wu-Fong Township Taichung County, 41349 Taiwan (R.O.C.). (Phone: (04)2332-3000 ext. 4288; e-mail: lhli@cyut.edu.tw).

Fu-Ming Lee is with the Chaoyang University of Technology, Wu-Fong Township Taichung County, 41349 Taiwan (R.O.C.). (Phone: (04)2332-3000 ext. 4286; e-mail: fmlee@cyut.edu.tw).

Shang-Chi Chan is with the Chaoyang University of Technology, Wu-Fong Township Taichung County, 41349 Taiwan (R.O.C.). (Phone: (04)2332-3000 ext. 4168; e-mail: s9514627@cyut.edu.tw).

the clustering result generated by using Adaptive Resonance Theory (ART) network.

To verify the proposed method, this research has implemented the BARS as an academic blog for real-world experiment. Experiments are done by collecting real bloggers' usage and their feedback so that the recommendation correctness is verified. Two types of recommendation approach are proposed, i.e., the BBRL (User Based Recommendation List) and CBRL (Content Based Recommendation List) method. The experiment proves that BBRL produces better recommendation result than CBRL with 84% satisfaction by all the bloggers.

II. RELATED WORKS

A. Recommendation System

Recommendation system (RS) [8] is a typical personalized service mechanism which can find user's interest items and make recommendation according to user's preference. The other definition of RS by Rashid et al. [9] is to make decision for users in the complex information environment. In general, the mechanism of implementing RS can be divided into Content-based (CB) and Collaborative Filtering (CF) [10]. The detailed content of CB and CF is introduced as follows.

Content-based (CB) [10] is a method extended from Information Retrieval technique. CB analyzes the attributes and characteristics according to user's historical preference items and, then, matching the suitable ones for user's request. The similarity measure of items or attributes is used for finding the matched items. The measuring technique for CB is not suitable for recommending items such as music, art, movie, audio, photograph, video, etc. However, these items are frequently read in blog sites, hence, these types of article may not easily be analyzed for relevant attribute information [11].

Collaborative Filtering (CF) [8][12], on the other hand, is widely applied and used for article, movie, product, etc. CF recommends items based on the similar preference of a group, known as neighbor. The CF technique, therefore, requires methodology for clustering or finding the neighborhood. Although CF method can handle the wide variety of information, it may still suffer two common problems, i.e., sparsity and cold-start [10]. Sparsity means even there are many users, it could happen that the user accessing-matrix or rating-matrix is still sparse. This phenomenon will generate the low coefficients of similarity and, therefore, make the recommendation inaccurate. Cold-start is the phenomenon of few users at the beginning and, hence, not enough information can be calculated for recommendation.

To do the blog-article recommendation, this research incorporates both Collaborative Filtering (CF) and Content-based (CB) method for performing the personalized recommendation. The BARS (Blog Article Recommendation System) is designed to service the bloggers for finding academic articles and to overcome the cold-start problems.

B. Adaptive Resonance Theory network

Adaptive Resonance Theory (ART) network is originated from the Theory of Cognitive [13], and it was proposed by Grossberg [14] in 1976. ART network is a type of Artificial Neural Network (ANN), and it belongs to the unsupervised learning network. Since the ART network is similar to human neural operation, ART network also exhibits the features of both stability and plasticity. This is done by adjusting the vigilance value when forming the clusters.

The structure of ART network, as shown in Fig. 1, is described as follows.

1) *Input layer*: the input vector is the training data.
2) *Output layer*: the clustering results of the training data.
3) *Weight connections*: every connection between the input unit and output unit has both top-down (w_t) links and bottom-up (w_b) links.

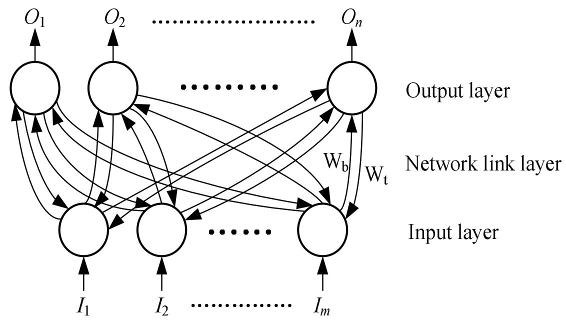

Fig. 1. The Structure of ART Network

In this research, we use ART network to cluster users with similar characteristics.

C. Ontology

Ontology is defined as a conceptualization of knowledge domain into a human-understandable in which machine-readable consists of entities, attributes, relationships, and axioms [15]. Researchers such as Hendler [16] has an explanation for ontology, i.e., ontology is composed of vocabularies, semantic interconnections, rules of inference, and logic for some particular topics. It may be viewed as the combination of knowledge terminologies.

Due to the advantages of ontology concept and the inference feature, Middleton et al. [4] has combined these two processes into RS (Recommendation System) so that the recommending performance can be improved. Their method is to use the hierarchical concepts of ontology to implement the user's profile, and the potential preference of users is predicted through the spreading and discovering of user's profile. Besides the mentioned methods, the Content-based method can also be improved by incorporating the ontology in the RS [17].

III. THE FRAMWORK OF BARS

This paper is proposing a Blog-article Recommendation System (BARS) and the aim of this research is to adopt the academic articles to perform the personalized bolg-article recommendation. The structure of the BARS system is shown in Fig. 2. The BARS includes two modules: Ontology Construction Module (OCM) and Article Recommender Module (ARM). The details of these two modules are discussed as below.

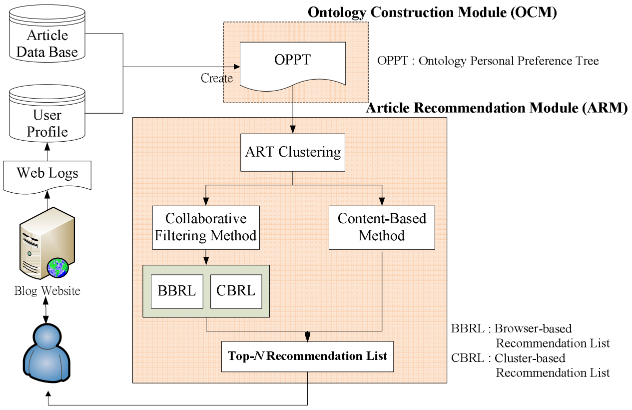

Fig. 2 The Structure of BARS

A. Ontology Construction Module (OCM)

The goal of OCM is to map between user's profile and article database. This part is implemented by using the Ontological Personal Preference Tree (OPPT) which representing the past reading records according to the ontological tree. The purpose of OPPT is to describe the read records of users by using predefined domain ontology. In addition, OPPT can also infer the potential concept of user depending on the hierarchical concepts of ontology. The construction procedures are described below.

i. Loading the user's profile and constructing OPPT.

ii. Initializing the u_{th} OPPT and performing a judgment whether the OPPT is null or not. If the result is *yes*, then we construct a new node on the root of the new OPPT. If the result is *no*, then we move to the next step.

iii. When user read the articles, adding these read articles into the OPPT. This is done by using the recursive method with the sequence of article type. During the recursive process, the judgment of whether the OPPT have the duplicate nodes of new affiliation is made. If the result is *yes*, then we return to the last step and add the reading records. If the result is *no*, then we add this node to OPPT and move to the next step.

iv. Finally, the process will check all the nodes whether these new nodes should be joined. If the result is *yes*, then we return to the first step and change to the next user for the $(u+1)_{th}$ OPPT construction. If the result is *no*, then we return to the last step.

B. Article Recommendation Module (ARM)

In Article Recommendation Module (ARM), there are three steps to be introduced. The ARM processes are shown in Fig. 3 and the details are explained as below.

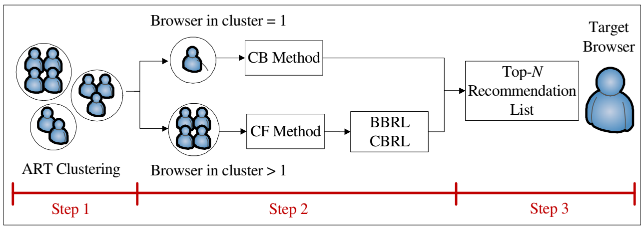

Fig. 3 The Processes of ARM

Step 1: Using ART for Clustering

Since the ART network adopts the binary codes, i.e., 1 and 0. We will assign *1* or *0* to the OPPT records where 1 means the users have read the target article and 0 means the target article is not read. Furthermore, to determine the preference level of article type read by the user, a set of condition bit is designed, that is, *f* columns are designed to determine the preference of user based on each article type. For example, if *f*=3, then these three columns could be defined as "2-4," "5-9," and "above 10," respectively, which tells how many articles of the target article type the user have read. For instance, if a user have read eight articles in a certain article type, we then put the bit *1* into the "2-4" and "5-9" at the same time and put the bit *0* into the column of "above 10." Therefore, the input vector for ART network for a certain article type can be expressed as (1, 1, 0). Hence, the input vector for the ART network for each user is, then, defined as the formula (1).

$$BRR = \{IC^l \mid 1 \le l \le L,\ L \in \mathbb{N}\}$$

$$IC^l = \{R_1^l, R_2^l, \ldots, R_f^l\},\ \ R_f^l \in [0,1],\ \ R_f^l = \begin{cases} 1 & \text{if } R_f^l \text{ is read} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

In formula (1), The User Reading Record (BBR) is primarily representing the record of each article type read by the user. The IC^l represents the l_{th} article type and R^l represents the sub-type of the l_{th} article type. The *f* index is representing the range of read frequency, i.e., "2-4," "5-9," and "above 10" of the sub-type of the l_{th} article type.

During the processes of ART network, the test of vigilance value ρ $(0 < \rho < 1)$ is examined by setting the ρ to 0.1, 0.2, …, 0.9 so that the appropriate results of cluster are determined. In formula (2), the cluster result is represented by the set of G^y where *y* represents the cluster number and B_x represents the user who falls into this cluster.

$$G^y = \{B_1^y, B_2^y, B_3^y, \ldots, B_x^y\},\ \ 1 \le x \le X. \tag{2}$$

Step 2: Collaborative Filtering and Content-Based

The two major modules of BARS are applying CF and CB methods and both are the important techniques for recommendation system.

CF is responsible for receiving the cluster results generated by the ART network and finds the cluster where the target user belongs to. The next step is to make the article recommendation based on the most similar preference around the vicinity of target user. To generate the proper recommendation list, this research proposes two list generation method, i.e., the User-based Recommendation List (BBRL) and Cluster-based Recommendation List (CBRL).

The BBRL is the recommendation list based on user's personal preference. The first step of BBRL is finding the highest preferred article type depending on the top-*N* preferred articles of the target user. The second step is finding the similar preferred article type among the neighborhood and target users. Finally, articles of highest click-rate and not read by the target user are generated for blog-article recommendation. The recommendation procedures are described as below.

Step 1: Calculating the threshold *t* of article type by using the formula (3) for each user. Finding and sorting the major preferred type of article for the user according to the setting of threshold. URC (User Reading Count) and URTC (User Reading Type Count) represent, respectively, the count of articles and sub-types that user had read. The variable *t*

ISBN: 978-988-98671-8-8

represents the threshold of article type for each user.

$$t = \frac{URC}{URTC} \tag{3}$$

Step 2: In formula (4), it decides how many articles can these different kinds of article type be recommended to the user? The Top-N_w represents the percentage of w article is recommendable as Top-N. The numerator and denominator represent all of the articles of certain type and all of the Top-N articles, respectively.

$$\text{Top-}N_w = \frac{\text{Number of articles of category}}{\text{All Top-}N_w} \times 100\% \tag{4}$$

Step 3: Finding the CPT^y, i.e., the Common Preference with Threshold, by using formula (5). CPT^y represents the set of the common preferred article type among the target user and neighborhood. NC^y is the set of the preferred article type of neighbor y. BC(t), on the other hand, represents the set of the preferred article type of the target user after the process of formula (3).

$$CPT^y = \{NC^y - (NC^y - BC(t))\} \tag{5}$$

Step 4: Finding the Article$_{HC}${MAX$_{top\text{-}N}$}, i.e., we retrieves the articles of highest click-rate and the article not read from target user before. The recommendation is made according to the set of the common preferred article type among the target user and the neighborhood.

The CBRL is another approach of generating recommendation list based on inter-cluster preference. Contrary to BBRL, CBRL, firstly, is to find the most similar Top-N preferred article type of neighborhood. Next, we sieve out the articles with the highest click-rate and not read from the target user before. The recommendation is made by generating these articles to the target user. The detailed process is described as below.

i. The step one of CBRL is finding the set of Top-N most common preferred article type among the target user and neighborhood by using the formula (6).

$$CP = \{NC^y - (NC^y - BC(I))\} \tag{6}$$

CP-Top-N is the set of Top-N reading article type

ii. The step 2 and step 3 are using the same process of step 2 and 4 of BBRL procedures.

To handle the cold-start problem, this research proposes the content-based recommendation using ontology for inferring the concept. The process steps are as follows.

i. Searching all of the reading records of article type depending on the user's OPPT and, then, computing the article type of higher reading rate by using formula (3).

ii. Recommending the articles with highest click-rate using the formula (4) and the article not reading by the target user.

C. *Evaluation of Effectiveness of Recommender*

To evaluate the performance of BARS, we collect the user feedback data by using the 5-point Likert scale from the user after reading and receiving the recommendation. Three major methods are used for evaluation, i.e., the quality of recommendation is examined using Precision, Recall, and F1 measure. The formulas of these three methods are defined as follows.

$$\text{Precision} = \frac{A_{\text{user's scores}}}{A_{\text{total scores of all recommended}}}$$

$$\text{Recall} = \frac{A_{\text{accept}}}{A_{\text{numbers of all recommended}}} \tag{7}$$

$$F1 = \frac{2 \times Precision \times Recall}{\text{Precision} + \text{Recall}}$$

A$_{\text{user's scores}}$: summation of total scores of Top-N articles, provided by user.
A$_{\text{total scores of all recommended}}$: summation of total default scores.
A$_{\text{accept}}$: If the given score is bigger than 3 by user, we use a unit of 1 to replace the original score and to sum the total scores.
A$_{\text{numbers of all recommended}}$: summation of recommended articles.

IV. Performance Evaluation

A. *Data Set*

We implemented a blog website (http://fblog.no-ip.org/) and we called it "Academic Research Blog." This blog website had 50 available sub-blogs, 660 articles, 25 article types, and 67 frequent users with information technology background. The portal site is as shown in Fig. 4.

Fig. 4 The Portal Site of Blog

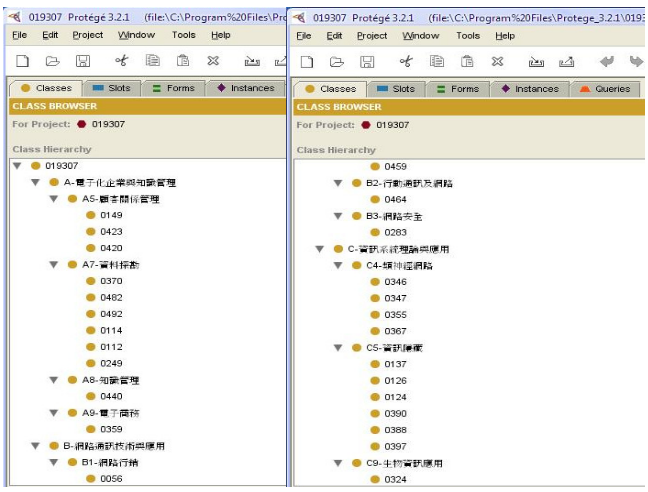

Fig. 5 A Segment of the OPPT

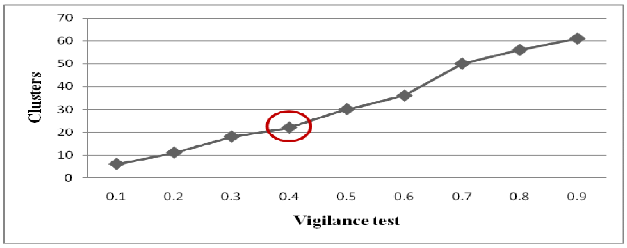

Fig. 6 Vigilance Test

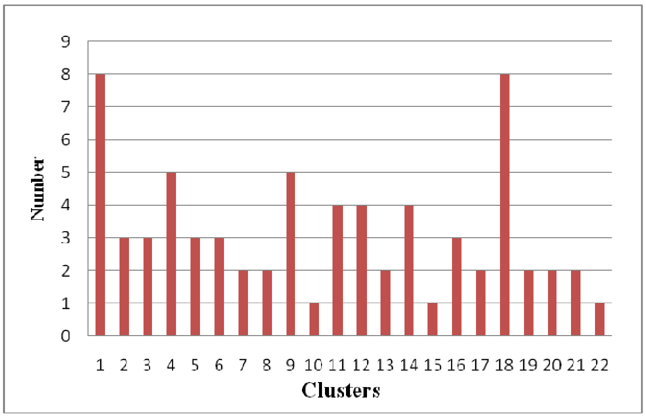

Fig. 7 The Cluster Results When Vigilance Value Set to 0.4

B. *Experimental Procedure*

The experiments were started by constructing the user profile according to Ontology Personal Preference Tree (OPPT). A segment of OPPT is shown in Fig. 5.

Next, we used Adaptive Resonance Theory (ART) network for clustering users with similar reading interests. When processing the ART network, the vigilance value is experimented and, finally, set to 0.4. Because the other vigilance values may generate clusters with only one user, which is not good for the similar preference finding and, hence, may not be able to generate the proper recommendation. For those clusters with fewer users, this research demonstrates the efficiency of CB method using the ontology concept to do the inference based on users unique reading preference and to solve the cold-start problem. The experiment data in Fig. 6 and Fig. 7 shows the clustering status of vigilance test and the clustering results when the vigilance value sets to 0.4.

Next, we will show the BARS's evaluation results and the performance for various recommendation list.

C. *Experimental Results*

The experiment was conducted to evaluated the performance of BBRL, CBRL, and Non-Personalized Recommendation (NPR) using the precision, recall, and F1.

In Fig. 8, Fig. 9, and Fig. 10, these results demonstrate the comparison results of BBRL, CBRL, and NPR using the precision, recall, and F1 measure, respectively.

In Fig. 8, it shows that the precision measure in BBRL has higher (75.82%) performance than CBRL (72.51%). And, the precision measure of NPR is showing the worst result, i.e., 64.06%. In Fig. 9, it shows that the highest recalling rate is achieved by using BBRL (94%), and the second is achieved by using CBRL (90%). The lowest recall rate is still the method of NPR with only 75%. In Fig. 10, it shows that the results of BBRL, CBRL, and NPR are 84%, 80%, and 69%, respectively. With the aspect of F1 measure in blog-article recommendation, the result varied with 4% between the BBRL and CBRL. Likewise, the comparison of BBRL, CBRL, and NPR shows the similar outcome, that is, the BBRL performs the best when using the F1 measure.

From these experiments results we can conclude that the BBRL and CBRL is suitable for blog-article recommendation when the personal preference is available or the similar preferences of neighbors are exist. The reason of NPR having the lowest precision, recall, and F1 measure is because only the population count of all users are calculated and, therefore, makes the NPR lacking of realistic understanding of user's preference. Therefore, our research shows that the performances of BBRL and CBRL are better than NPR for blog-article recommendation.

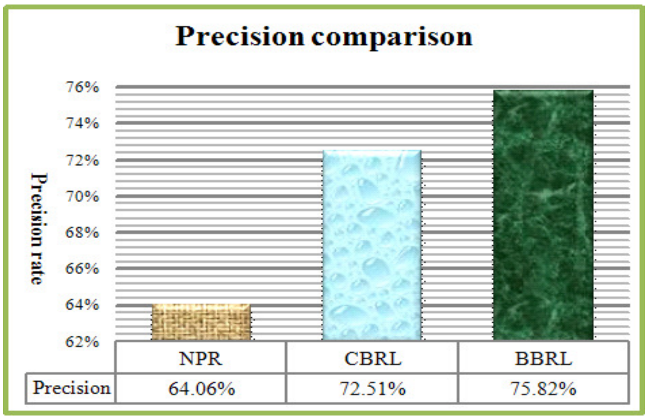

Fig. 8 The comparison results of precision measure

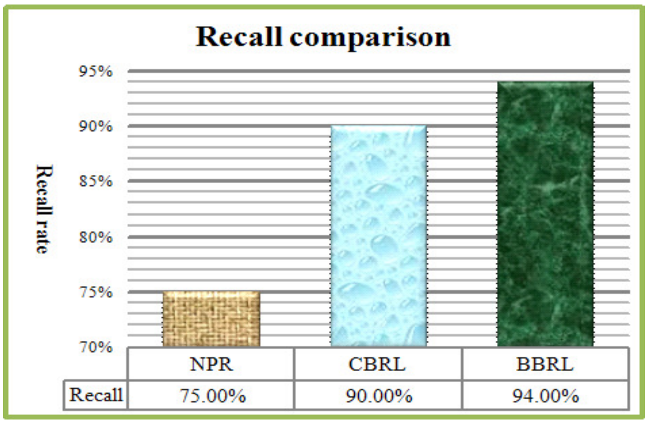

Fig. 9 The comparison results of recall measure

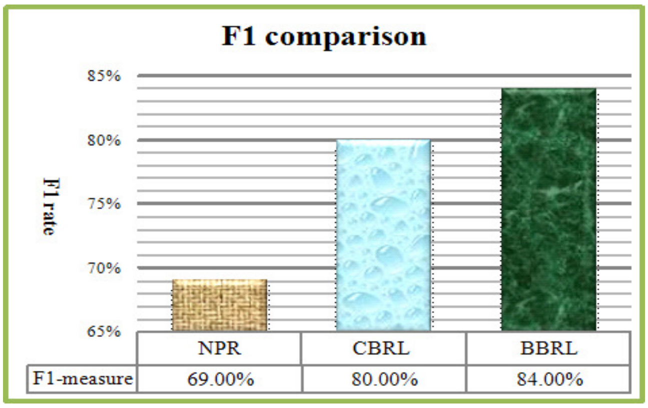

Fig. 10 The comparison results of F1 measure

V. Conclusion

This paper presents a Blog-article Recommendation System (BARS) which provides personalized recommendation for blog-article. BARS incorporates ontology to build the personal preference tree (OPPT) to understand the blogger's interest and applies ART network to do the clustering so that the potential preferences of neighbors with similar interests are utilized.

Furthermore, this paper also proposed a Content-based recommendation process by incorporating the ontology inference concept. This idea helps the research to meet our goals, i.e., to improve the effectiveness of recommendation and to handle the cold-start (i.e., lacking of new-user profile) problem.

Based on the experiments carried out above, this study has proved to obtain the following research results, i.e.,

1) reducing the user cold-start problem,
2) enhancing the performance and quality of personalized article recommendation, and
3) enabling the blog-article recommendation.

The contributions of this paper are proposing two recommendation processes, i.e., the User-based Recommendation List (BBRL) and Cluster-based Recommendation List (CBRL), which utilize the CF and CB method, respectively. The results of both methods using the F1 measure for examining the recommendation performance are as follows.

(1)BBRL has higher F1 (84%) then the CBRL (80%).

(2) Both BBRL and CBRL methods are showing the better results then the NPR method (69%).

Acknowledgement

This study was partially supported by the National Science Council, Taiwan, ROC, under grant number NSC-95-2416-H-324-011 and NSC-96-2416-H-324-006.

References

[1] Technorati (2007). Available at: http://www.technorati.com/

[2] N. J. Belkin and W. B. Croft, "Information Filtering and Information Retrieval: Two Sides of the Same Coin?," *Communications of the ACM*, vol. 35, 1992, pp. 29-38.

[3] H. J. Lee and S. J. Park, "MONERS: A News Recommender for The Mobile Web," *Expert Systems with Applications*, vol. 32, 2007, pp. 143-150.

[4] S. E. Middleton, N. R. Shadbolt, and D. C. De Roure, "Ontological User Profiling in Recommender System," *ACM Transactions on Information Systems (TOIS)*, vol. 22, 2004, pp. 54-88.

[5] S. S. Weng and H. L. Chang, "Using Ontology Network Analysis for Research Document Recommendation," *Expert Systems with Applications*, 2007, In press.

[6] P. L. Hsu, P. C. Liu, and Y. S. Chen, "Using Ontology to Map Categories in Blog," *Proceedings of the International Workshop on Integrating AI and Data Mining(AIDM'06)*, 2006, pp. 65-72.

[7] Y. Li, L. Lu and L. Xuefeng, "A Hybrid Collaborative Filtering Method for Multiple-interests and Multiple-content Recommendation in E-Commerce," *Expert Systems with Applications*, vol. 28, 2005, pp. 67-77.

[8] B. Sarwar, Karypis, G. Konstan, and J. Riedl, "Analysis of Recommendation Algorithms for E-commerce," *Proceedings of ACM E-commerce 2000 Conference*, 2000, pp. 158-167.

[9] A. M. Rashid, I. Albert, D. Cosley, S. k. Lam, S. M. McNee, J. A. Konstan, and J. Riedl, "Getting to Know you: Learning New User Preferences in Recommender Systems," *Proceedings of the 7th International Conference on Intelligent User Interfaces*, 2002, pp. 127-134.

[10] F. Chesani, "Recommendation Systems," *Corso di laurea in Ingegneria Informatica*, 2002, pp. 1-32.

[11] Y. S. Kim, B. J. Yum, J. S. Su, and S. M. Kim, "Development of A Recommender System Based on Navigational and Behavioral Patterns of Customers in E-commerce sites," *Expert Systems with Applications*. vol. 28, 2005, pp. 381-393.

[12] S. Upendra and M. Pattie, "Social Information Filtering: Algorithms for Automating "Work of Mouth"," *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1995, pp. 210-217.

[13] R. C. Chen, J. Y. Liang, and R. H. Pan, "Using Recursive ART Network to Construction Domain Ontology Based on Term Frequency and Inverse Document Frequency," *Expert Systems with Applications*, vol. 34, 2008, pp. 488-501.

[14] S. Grossberg, "Adaptive Pattern Classification and Universal Recoding: I. Parallel Development and Coding of Neural Feature Detectors," *Biological Cybernetics*, vol. 23, 1976, pp. 121-134.

[15] N. Guarino, C. Masolo, and G. Vetere, "Ontologies and Knowledge Bases: Towards a Terminological Clarification," *In: Mars, N. (Ed.), Proceedings, Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing.* 1995, pp. 25-32.

[16] J. A. Hendler, "Agents and The Semantic Web," *IEEE Intelligent Systems*, vol. 16, 2001, pp. 30-37.

[17] N. Guarino, C. Masolo, and G. Vetere, "OntoSeek : Content-Based Access to the Web," *IEEE Intelligent Systems*, vol. 14, 1999, pp. 70-80.