International Journal of Knowledge and Language Processing *www.ijklp.org*
KLP International ©2013 ISSN 2191-2734
Volume 4, Number 2, **2013** pp.1–11

A CF Recommendation Method based on Domain Expert Trust

Zhao Tianzi and Jing Minchang

China University of Petroleum, 102249, Beijing, China
ztz@cup.edu.cn

Received April 2013; revised April 2013

ABSTRACT. *Traditional collaborative filtering recommendation algorithms mainly focus on the similarity of ratings made by users. The Follow-the-Leader model introduces the expert trust into the recommendation system, but the expert trustworthiness in this method is computed based on the overall rating scores without considering the domain attribute of the expert. Considering the expert domain attribute and real life experience, this paper improves the model by dividing experts into different domains according to different categories of rated items and defining the concept of the domain expert trustworthiness, and proposed a collaborative filtering algorithm based on prioritized domain expert trust in the recommendation process. The experimental results of the application on an open dataset named GroupLens show that the accuracy of the new algorithm in predicting user ratings is superior to those algorithms which do not consider the domain attribute of experts.*

Keywords: Collaborative Filtering; Follow-the-Leader model; trustworthiness; domain expert trust

1. **Introduction.** Since the 1990s, with the rapid development of internet, the information which people can obtain is growing of scale. When people enjoy the wealth of information, they are also facing the problem of information overload seriously at the same time. One of the methods to solve this problem is the personalized recommendation technology [1][2].

The personalized recommendation system relies on the powerful data mining technology to make itself has the ability of machine learning. In the continuous interaction with the user, it can understand the information of user's preferences, tastes, habits etc. more and more, and build the user knowledge. Then the user knowledge can be used in the recommendation information service process for users. Because having the large amount of information and knowledge, the recommendation engine will be "know you better than you".

Because of the initiative and personalized service characteristics, the recommendation system attracted the attention of researchers and the business communities. And it has been proved an effective method to solve the problem of information overload by practice. At the same time, it is a new information service method following the portals, search engines

[3]. Now it has been used widely in the field of e-commerce domain. It becomes an integral part of the basic functions of various information systems. Amazon is a typical case of application of recommended techniques to achieve e-commerce marketing. It is said, 35% of Amazon page sales credits to its recommendation engine. Alex Iskold [4] called Amazon "The King of recommendations" without overrated。

Among personalized recommendation systems, the collaborative filtering (CF) is one of the most widely used and successful recommendation algorithms. Most collaborative filtering recommendation system is based on the establishment of the user's close neighbors of the same preferences and uses this information for recommendation [5]. Essentially, the nearest-neighbor users can also be seen as a confidence-building mechanism, and the collaborative filtering recommendation process is a series of processes that build on trust among similar users, during which the target user within the system accepts the recommendation of the nearest neighbor with the trust on the neighbor [6]. Ziegler CN et al. consider that the higher the similarity of two users is, the more trust there are between them [7].

The introduction of the trust mechanism into the recommendation system is required by the rapid development of social networking sites such as MySpace, Facebook, which can improve the accuracy of the recommendation system and alleviate the problem of data sparsity and cold-start to some degree. A number of studies have been conducted in this area both at home and abroad [2-12]. These studies, however, only consider the direct or indirect trust between users, and leave out the discussion on the expert trust attributes of the users. Jebrin Al-Sharawneh et al. [13] believed that the expert trustworthiness is another dimension of the user credibility, and studied the influence of the expert trustworthiness of users on the recommendation result based on social network data. However, the expert trustworthiness in the previous study was evaluated using the overall score among the whole users, with no consideration of the differences between different expert domains.

For digital library users, information acquisition is their main demand. The user values more on creditable information; therefore, the expert trustworthiness, especially in specific domains, is of more importance for the digital library users than normal users using the personalized recommendation system in the domain of e-commerce. Seeking the nearest expert neighbors instead of looking for generally similar neighbors is more in line with the actual needs of the digital library.

Based on the studies of Jebrin Al-Sharawneh et al, we categorize experts in different domains using the project classification attributes, bring up the concept of domain expert trustworthiness, and propose a collaborative filtering recommendation algorithm which gives priority to the domain expert trustworthiness in the recommendation. This algorithm is validated by an application on an actual set of publicly available data, showing the effect of the domain expert on the recommendation.

2. **Related research.**

2.1. **Follow-the-Leader model.** In reality, people often tend to consult an expert. Social psychology study also shows that people tend to be more willing to listen to the opinions of experts when making choices of recommended products [14]. To this end, Goldbaum D brought up the Follow-the-Leader model [15], in which people in the social network model are divided into two categories: followers and leaders. Leaders have a strong subjective consciousness and a high confidence, whereas followers are members more vulnerable to influences from those of high confidence in the social network.

This model divides the followers into three types based on the relationship between the followers and the leaders. Type 1 followers have fixed preferences, but do not have complete product information. In this case, the leaders can provide information or recommendations for the followers, so that they can use their information to the full effectiveness; Type 2 followers have some fixed preferences, but are susceptible to influences of the leaders or other people; Type 3 followers have no fixed preferences at all, and are affected completely by the opinions of the leaders.

2.2. **Expert-based Trust.** Trust, as a social attribute of individuals in the human society, has long been considered by researchers. Trustiness is usually used to indicate the degree of trust. The greater the trustiness is, the higher the level of trust is. The recommendation of users with higher trustiness in the recommendation system is more acknowledged by other users.

For example, for rating data, users with a higher trustiness may make ratings in a more reliable, honest, and objective manner and these ratings are therefore more useful in a recommendation algorithm; on the contrary, users with lower trustiness may make ratings in a less reliable manner, leading to recommendations of low quality if a recommendation algorithm uses these ratings. According to the interacting bodies of trust, trust can usually be divided into direct trust (DT) and indirect trust (IDT) [10].

Expert trust (ET) is another dimension of the credibility of users, which is different from the user's direct trust and indirect trust. Jebrin Al-Sharawneh et al [13], based on the user trust mechanisms, further enriched the Follow-the-Leader model, and applied this model in the recommendation system for the first time. They proposed the trust-aware Follow-the-Leader model, considering the credibility of users in two main aspects: Trustworthiness and Expertise. Meanwhile, they also decided that the credibility of users can be jointly determined by the direct trust, the indirect trust and the expertise.

Jebrin Al-Sharawneh's method is proved to have a successful application on the recommendation behavior in the social network; however, the applicability of this method is subject to certain restrictions if lack of a direct or indirect relationship of trust between users. In addition, the expert trustworthiness was calculated from the overall rating data, with no considerations of different domains of expertise, which is inconsistent with realistic experience. In reality, experts generally belong to one or a few domains, and no one can be experts in all the domains.

3. **CF algorithm based on prioritized domain expert trust.**

3.1. **Algorithm principals.** Domain expert trust recommendation model is proposed based on the following two considerations:

(a) In real life, users are generally familiar with only one or several areas, and very few people can be proficient in all areas. The so-called experts are only experts in their familiar domains, but not necessarily experts in other domains. Even if someone could become experts in many domains, his proficiency of knowledge "specialized" in different areas will be different. This is so-called "One masters specific knowledge". Therefore, experts may make unreliable ratings in areas other than theirs.

(b) In practical applications, users (or experts) make few ratings, which are usually concentrated in one or a few project categories according to their interest. Users (or experts) have more dense ratings in these categories. The calculation of trustworthiness is done over several item categories rather than the entire item space, so the amount of computation can get a great degree of reduction.

3.2. **Domain expert trustworthiness.** Based on the concept of the expert trust [16], we define the domain expert trust as the competing ability of users in the trust network to provide reliable information in a specific domain, showing the credibility of users possessing knowledge, ability, and skills in this domain.

Similar to the expert trust, the domain expert trust is also reflected on the number and quality of the ratings made on an item in collaborative filtering systems. Generally, it can be considered that users who make more ratings, especially more high-quality ratings, than others, have higher trustworthiness.

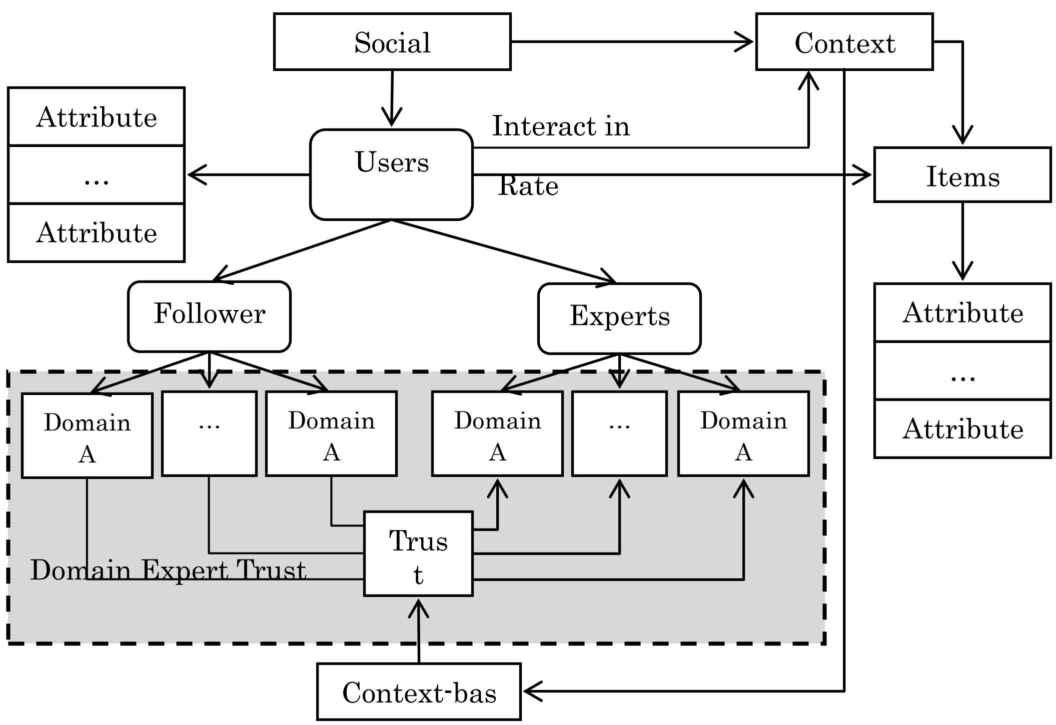

FIGURE 1. DOMAIN EXPERT TRUST RECOMMENDATION MODEL

Figure 1 is a domain expert trust model improved from the model brought up by Jebrin Al-Sharawneh et al [13]. The model areas typically can be divided according to the item category. In actual e-commerce sites or digital library systems, all products or documents are divided into a limited number of item classes. For example, the Chinese online bookstore Dangdang (http://www.dangdang.com) categorizes books into several classes such as literature, management, computer, and so on. Additionally, the Chinese Library Classification has even more detailed catalog classes.

Below gives a general definition of the domain expert trustworthiness: take $T_e(u,d)$ as the trustworthiness of user u in domain d, and use Jebrin Al-Sharawneh's method and reference [16]'s method, $T_e(u,d)$ can be expressed as follows:

$$T_e(u,d) = \frac{1}{M^d_{\max}} \times \sum_{i=1}^{M^d} T_e\left(r_u^i\right) = \frac{1}{M^d_{\max}} \times \sum_{i=1}^{M^d} \left\{1 - \frac{|\, r_u^i - R^i_{avg} \,|}{R^i_{\max}}\right\} \quad (1)$$

where d is the domain name (item category), $M\ d$ is the quality of the user's rating quality in domain d, and $M^d_{\max}$ is the maximum rating made by one single user.

We use the rating matrix data in the following table as an example to illustrate the calculation method of the domain expert trustworthiness. The items are divided into different categories in the table, as shown in Table 1.

TABLE 1. USER RATING DISTRIBUTION AMONG DIFFERENT ITEM CLASSES

	Literature		***Management***		***Computer***		
	I_1	I_2	I_3	I_4	I_5	I_6	I_7
U_1	5	3	2.5				
U2	2	2.5	5	2			
U_3	2.5			4	4.5		5
U_4	5		3	4.5		4	
U_5	4	3	2	4	3.5	4	
R^i_{avg}	*3.7*	*2.83*	*3.13*	*3.63*	*4.0*	*4.0*	*5*

According to Formula (1), we calculated the expert trustworthiness of the five users at different domains, and the results are listed below:

(a) User U_1

$T_e(U_1, literature) = \frac{1}{2} * (0.81 + 0.94) = 0.88$;

$T_e(U_1, management) = \frac{1}{2} * (0.87) = 0.44$;

$T_e(U_1, computer)$ Not applicable ;

(b) User U_2

$T_e(U_2, literature) = \frac{1}{2} * (0.66 + 0.89) = 0.78$;

$T_e(U_2, management) = \frac{1}{2} * (0.62 + 0.64) = 0.63$;

$T_e(U_2, computer)$ Not applicable ;

(d) User U_3

$$T_e(U_3, literature) = \frac{1}{2} * (0.76) = 0.38$$;

$$T_e(U_3, management) = \frac{1}{2} * (0.92) = 0.46$$

$$T_e(U_3, computer) = \frac{1}{2} * (0.89 + 1) = 0.95$$;

(e) User U_4

$$T_e(U_4, literature) = \frac{1}{2} * (0.74) = 0.37$$;

$$T_e(U_4, management) = \frac{1}{2} * (0.97 + 0.81) = 0.89$$

$$T_e(U_4, computer) = \frac{1}{2} * (1) = 0.5$$;

(f) User U_5

$$T_e(U_5, literature) = \frac{1}{2} * (0.94 + 0.94) = 0.94$$;

$$T_e(U_5, management) = \frac{1}{2} * (0.77 + 0.92) = 0.85$$

$$T_e(U_5, computer) = \frac{1}{2} * (0.89 + 1) = 0.95$$

From the results shown above, the expert trustworthiness of users in various domains and among the whole ratings is different as shown in Table 2; therefore, in predicting the rating of a particular item, one should choose to trust experts with high trustworthiness in that particular domain.

TABLE 2. COMPARISON OF THE DOMAIN EXPERT TRUSTWORTHINESS

Domains	Comparison of users' expert trustworthiness
All domains	$U_5 > U_3 > U_4 > U_2 > U_1$
Literature	$U_5 > U_1 > U_2 > U_3 > U_4$
Management	$U_4 > U_5 > U_2 > U_3 > U_1$
Computer	$U_5 = U_3 > U_4$; U_1、U_2 not applicable

3.3. **Recommendation algorithm.** Traditional collaborative filtering algorithms determine the nearest neighbor of the user by considering only similarity of user ratings, not the identity attributes of the nearest neighbor. However, studies show that it can indeed improve the accuracy of recommendations and the success rate of the rating prediction by taking into account the expert attributes of users [13] [16].

In this paper, based on the study of the domain expert trustworthiness shown above, we propose that priority should be given to users with high domain expert trustworthiness when determining close neighbors for each user, and that the domain expert rating opinions should be taken into account for recommendations. We call this process "recommendation algorithm based on the prioritized domain expert trustworthiness” (EPT-D) ".

$$p(a,i) = R_a^{avg} + \frac{\sum_{u=1}^{m} T_e(u,d) \times (r_u^i - R_u^{avg})}{\sum_{u=1}^{m} T_e(u,d)} \tag{2}$$

where $p(a,i)$ is the predicted rating of the user a on the item i, m is the number of experts making ratings on i, d is the category of i, r_u^i is the actual rating of the domain expert *u* on *i*, R_u^{avg} is the average rating of the expert *u*, R_a^{avg} is the average rating of the item i. Different from the method used by Jebrin Al-Sharawneh[9], this formula takes into account the trustworthiness of the domain experts, instead of the factors of the direct or indirect trust, and the choice of user neighbors focuses more on the expert users with expertise in specific domains.

4. **Experiments and analysis.**

4.1. **Data source and evaluation standard.** In order to verify the effectiveness of this method, we make tests with the MovieLens ml-100k dataset provided by the Grouplens work group [17]. The MovieLens dataset includes 100 000 ratings ranging from 1 to 5, with a total of 943 users and 1682 films. Each user evaluates at least 20 films. The sparsity of the dataset is 0.943.

In the dataset, the films are divided into 19 different categories (i.e. domains as defined herein), and these 19 classes can be regarded as different domains. The trustworthiness of the user ratings on specific items can be regarded as the trustworthiness of the expert trustworthiness in this domain. As an item can belong to more than one domain, we categorize the trustworthiness into a number of domains when calculating the trustworthiness of the experts who make ratings.

Recommendation rating prediction has a variety of evaluation criteria, and we use the mean absolute error (MAE) to measure the prediction accuracy. The MAE is a commonly used method for measurement errors, which calculate the accuracy of the prediction by measuring the deviation between the computed predicted user ratings and the actual user ratings. The smaller the MAE value, the greater the recommendation accuracy. The calculating formula is as follows:

$$MAE = \frac{\sum_{i \in N} | p_i - r_i |}{N} \tag{3}$$

where *N* is the number of all the rating items in the tested dataset, p_i is the predicted user rating, and r_i is the actual user rating.

4.2. **Experiment method and result analysis.** Reference [16] considered that priority should be given to the expert trust factor in the information recommendation process and they proposed a CF filtering algorithm based on the prioritized expert trust (EPT), which is proved to be superior to the traditional nearest neighbor algorithm (KNN). This paper compares the EPT-D, EPT and KNN methods in predicting the rating accuracy and the success rate of the predication.

First, we trained the algorithm on the training data set (ua.base), and then used the test data set (ua.test) for final result verification. Expert users are considerably different in the 19 domains; therefore, the recommendation results are different when using different nearest neighbors of the users in the recommendation process.

In the experiments, we set different expert rates (2% to 20%) among the users for comparison. The results show that both algorithms give better recommendations with the increase of the number of experts among the users (Figure 2). However, with the same number of experts, particularly when the expert rate is less than 10% among users, the prediction accuracy of the EPT-D algorithm is obviously superior to the EPT algorithm. As the number of experts increases (greater than 15%), both recommendation accuracy tends to be the same. Nevertheless, considering that the proportion of experts are generally no more than 15% in real situations [13], the EPT-D algorithm proposed here is more applicable and useful.

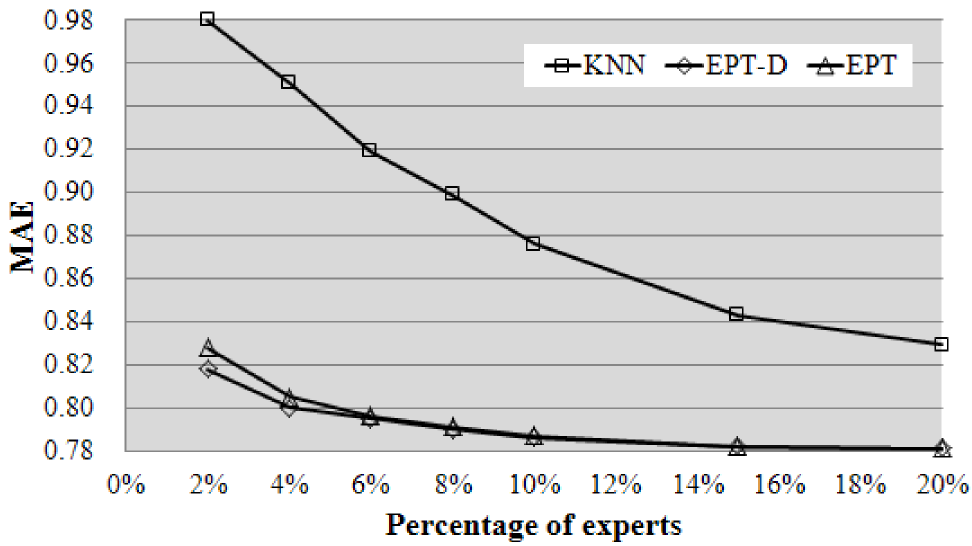

FIGURE 2. COMPARISON OF MAE WITH EPT-D, EPT AND KNN ALGORITHMS

In the prediction of the success rate, both methods produce higher success rates with increasing proportion of experts among users (Figure 3). With the same number of experts, the EPT-D algorithm produces a lower success rate than the EPT method, nevertheless exceeding 98% in the minimum, which is to say that most ratings can match the predicted values. The traditional KNN algorithm can not achieve the same effect when the proportion of experts is same (Figure 4).

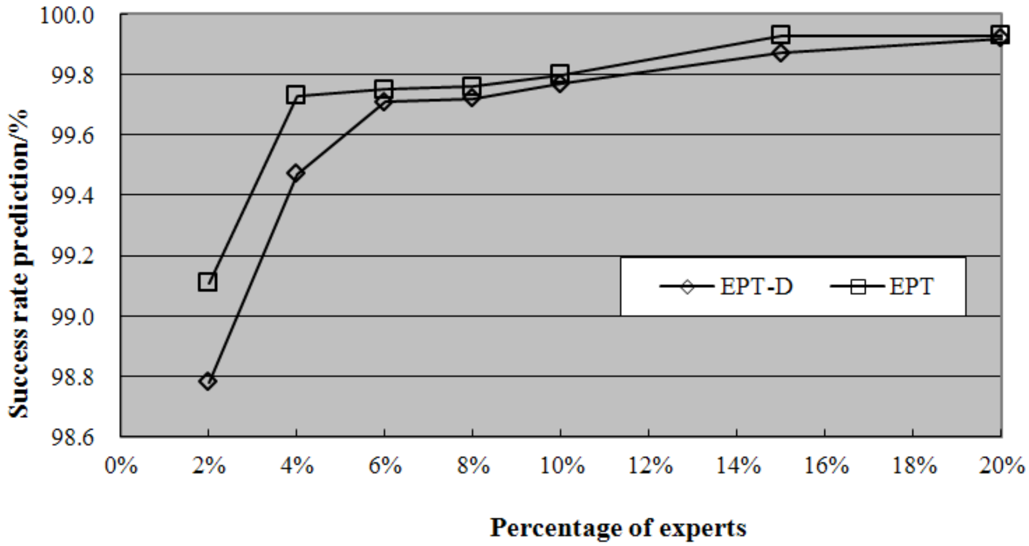

FIGURE 3. COMPARISON OF SUCCESS RATE BETWEEN EPT-D AND EPT ALGORITHMS

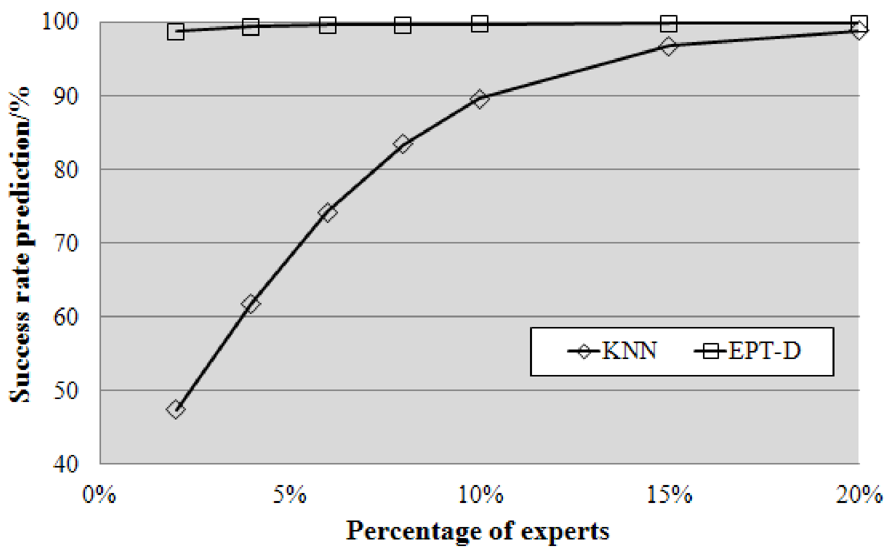

FIGURE 4. COMPARISON OF SUCCESS RATE BETWEEN EPT-D AND KNN ALGORITHMS

5. **Conclusions.** This paper introduces the application of the user trust in the collaborative filtering recommendation and its deficiency, and brings up the concept of the expert trust. We propose a collaborative filtering recommendation method based on prioritized domain expert trustworthiness (EPT-D), considering the factor of the expert trust primarily, and the domain attribute of the expert in the meantime, in the filtering collaborative recommendation process. We categorize the experts into different domains in accordance with the classification attributes of rating items. Experts in different domains have different expert trustworthiness, and the recommendation priority is different for the experts

specialized at different domains. As the digital library recommendation system needs to recommend authoritative and reliable literature information to library users, the role of the expert or the domain expert is particularly important; therefore, this method produces a better result in terms of the recommendation effect, compared with other methods without considering the domain expert trust. Moreover, the computation of the expert trustworthiness is narrowed down to a few items classes, rather than the entire item space; therefore, the amount of computation can also get a great degree of reduction.

The results also show that the introduction of the domain attribute of experts into the collaborative filtering algorithm not only is feasible, but also produces better recommendations to some degree. Although the prediction success rate is slightly lower than the algorithm which does not consider the user's domain expert attribute, the rate can still reach more than 98%, which can fully meet the functional requirements of the recommendation system.

REFERENCES

[1] Liu Jianguo, Zhou Tao, Wang Binghong. Research progress of Personalized recommendation System[J]. Progress in Natural Science, 2009, 19(1):1-15.

[2] Xu Hailing, Wu Xiao, Li Xiaodong,etc. Comparison Study of Internet Recommendation System[J]. Journal of software, 2009, 20(2) : 350-362.

[3] Wu Jinlong. Collaborative Filtering algorithms in Netflix Prize[D]. Beijing : Peking University,2010.

[4] Alex Iskold. The Art, Science and Business of Recommendation Engines [EB/OL]. http://www.readwriteweb.com/archives/recommendation_engines.php,2011-5-5.

[5] Koren Y and Bell R. Advances in Collaborative Filtering [EB/OL]. [2012-9-26]. http://research.yahoo.com/pub/3503.

[6] Guo Yanhong, Deng Guishi, Luo Chunyu. Collaborative Filtering Recommendation Algorithm Based on Factor of Trust[J]. Computer Engineering, 2008, 34(20):1-3.

[7] Ziegler C N, Golbeck J. Investigating interactions of trust and interest similarity [J]. Decision Support Systems. 2007, 43(2):460-475.

[8] Donovan J O and Smyth B. Trust in recommender system [C]. Proceedings of the 10th international conference on intelligent user interfaces. ACM New York, NY, USA. 2005:167-174.

[9] Massa P, Avesani P. Trustaware Collaborative Filtering for Recommender Systems[C]. Proceeding s of International Conference on Cooperative Information Systems. Berlin: Springer, 2004: 492 -508.

[10] Zhang Fuguo. Research on the Key Issues of Trust-based Personalized Recommender Systems in E-commerce[D]. Jiangxi: Jiangxi University of Finance & economics, 2009.

[11] Jin Yaya, Mou Yuanchao. Collaborative Filtering Recommendation Algorithm Based on Improved Trustworthiness[J].New Technology of Library and Information Service,2010,198(10):49-53.

[12] Lu Zhubing, Tang Yan. A Trust Network-based Collaborative Filtering Recommendation Strategy [J]. Journal of Southwest China Normal University (Natural Science Edition), 2008, 33 (2) : 123-126.

[13] Jebrin Al-Sharawneh and Mary-Anne Williams. Credibility-aware Web-based Social Network Recommender: Follow the Leader[C]. Proceedings of the 2nd ACM RecSys'10 Workshop on

Recommender Systems and the Social Web. Barcelona, Spain, 2010.9.
[14] Guanfeng Liu, Yan Wang and Orgun M. Trust Inference in Complex Trust-Oriented Social Networks[C]. International Conference on Computational Science and Engineering, CSE'09, 2009: 996-1001.
[15] Goldbaum D. Follow the Leader: Simulations on a Dynamic Social Network [EB/OL].[2012-9-26]. http://www.business.uts.edu.au/finance/research/wpapers/wp155.pdf.
[16] Jing Minchang, Tang Diguan, Yu Yinghui. A Recommending Method Based on Expert Prior Trust in Collaborative Filtering[J]. Library and Information Service, 2012, 56(11): 105-108.
[17] GroupLens Research. MovieLens Data Sets [EB/OL]. [2012-9-26]. http://www.grouplens.org/node/73.