International Journal of Security and Its Applications
Vol. 9, No. 4 (2015), pp. 81-90
http://dx.doi.org/10.14257/ijsia.2015.9.4.09

Fusion Trust Relation and Rating Data Algorithm

Xiaofeng Li

Department of Information Science,
Heilongjiang International University,
Harbin 150025, China
mberse@126.com

Abstract

A new algorithm FTRA has been proposed, which infuses users' trust network and rating data. The sparse problem of rating data will significantly reduce the accuracy of collaborative filtering recommendation. In addition to the users' ratings data on the Internet, other data sources which can be used in the process of recommend, and one of the more common is trust network data which describes the mutual relationship between users. To solve this problem, this paper will the data of trust network as an important supplement on the rating data, and bases on graph theory concepts or methods, the similarity method in the paper, and the Katz method which is used to calculate the similarity of link, proposes the FTRA algorithm which organic infuses this two data, and then better to solve the sparse problem of the rating data faced by collaborative filtering. The experimental results on the Epinions dataset show that the FTRA algorithm is superior to or significantly better than the comparison algorithms, which include the algorithms that only based on the rating data or the trust relationship, and the other algorithms infusing the two data sources.

Keywords*: Fusion trust network and rating data algorithm, Collaborative filtering recommendation algorithm, Similarity*

1. Introduction

User-based collaborative filtering method makes the hypothesis: recommending interesting contents to target user can be effected by finding out other users who have the similar preferences with the user and sending their interested contents to the user. So the main task for user-based collaborative filtering method is to accurately search other users who have alike interests with the target user [1]. Currently, it's common to use Pearson relativity and cosine similarity to calculate the preference similarity between different users. But, such methods are dependent on two users' co-rating items. Since the rating matrix between real user and item is too sparse that users who make co-ratings with target user is quite few. Besides, even if such users have co-ratings, the co-rated items are very few, as a result, the found users may not exactly have the same preference to target users'. Finally, the accuracy of recommendation or the precision of predicted scores is affected. In a word, the sparsity problem of rating data is one of the main challenges by the collaborative filtering method.

With advancement of web2.0, the information generated by users on websites is not only their scorings of items, and also social relationship information provided by them, like classmate relation, friend relation, trust relation etc. In recent years, some researchers introduced those social relation data to solve the inaccuracy of prediction caused by the sparseness of rating data [2-3], which, to a certain degree, alleviated the impacts of such sparseness [4-6]. But the existing techniques cannot settle the problem. After reviewing previous works, we propose a new method FTRA, i.e. fusion trust relation and rating data algorithm. It incorporates the trust relation between users and their rating data into a

ISSN: 1738-9976 IJSIA

graphical model. Then, on the basis of Katz similarity calculation method, it accurately measures the similarity between users to get neighbors with similar preference to the target users'. Further on, it makes prediction of rating values based on ratings by similar neighbors.

2. Related Work

Many methods [7-14] were developed to weaken the influences of data sparseness on recommendation results. Kautz and Granovetter [15] et al. found that people make use of social relation like friend relation, workmate relation and cooperative relation to accept and broadcast information. Resnick et al [16] pointed that in daily life, we get recommendation information by relying on others every day, such as word-of-mouth recommendation, book or movie review, restaurant recommendation. Either social relation or others' recommendation, it is a way to help users filter information. So it's important to fuse social relation to the recommendation method, but not simply base on mathematical equation like Pearson relativity cosine similarity to estimate the similarity between users, which might lead to better recommendation quality.

Recently, there are more and more recommendation methods fusing social relation to collaborative filtering. In 2006, Golbeck et al [17] used social trust relation information. They estimated target user's evaluation of items according to the score made by the user who is trusted by the target. In 2007, Avesani et al [18] based on user's social trust relation to take the limited step-length trust propagation method. When target user's estimated trust value of other users was acquired, they got estimated marks on the basis of that. In 2009, Yuan et al. introduced users' friend relation, user group information and users' selected items information to construct a graph which contains three types of nodes. They used the graph-based random walk method to produce recommendation results. In 2010, Jebrin et al [19] employed users' trust relation and rating data of items to compute the global reputation value of every user. With the evaluation of items made by users with the highest "global reputation", they predicted the scorings of items by target users. All the above methods performed better than traditional collaborative filtering methods in terms of estimation accuracy and recommendation accuracy.

3. Idea of the Proposed Method

Sinha and Swearingen [20] stated that users prefer to accept recommendations by those who are known to or trusted by them. Ziegler and Lausen [21] concluded that users' trust relation is positively proportional to preference similarity through experiments. Thus, fusing trust relation information to the collaborative filtering would help accurately find users with similar interests to the target users'. More accurate prediction values will be obtained based on more accurate neighbors with similar preference.

Based on the above work, we present a new method FTRA to solve data sparseness. Its basis idea is: users may like items which are also loved by other users who have trust relation or preference similarities. Considering those two relations are internally associated, it's possible to get more dense user similarity relational graph by fusing the two relations. With similarity propagation based on dense graph, more accurate similar neighbors can be got, thus to acquire more accurate prediction values. FTRA has the following steps:

- Step 1 Calculate users' preference relation

According to user-item rating data, a similarity method based on co-rated items is used to get users' preference similarity relation;

- Step 2 Fuse two relations

Fuse the users' preference similarity relation to user trust relation, i.e. use the inferred similarity relation to replenish users' trust relation to get a new preference/trust relation between users;

- Step 3 Compute the similarity between users

With the new preference/trust relation after fusion, Katz similarity method is applied to accurately locate the similar neighbors of target user;

- Step 4 Evaluation prediction

Predict target user's scores of relative items based on the value by similar neighbors of one item. The algorithm is described as follows in Table 1.

Table 1. FTRA Algorithm Description

Input: User-item rating data training set Tr ,the test set Te , user trust matrix T; Output: rating prediction matrix P

1. FTRA_1[initialization]
2. FTRA_2[Calculation of the matrix T ']
FOR i=1 TO n DO
FOR j=1 TO n DO
IF T_{ij} ≠0 THEN
ELSE $\mathrm{T}_{ij}^{'}$ =sim(Tr, i, j)
3 FTRA_3[Calculation of limiting similarity matrix T"]
NORMALIZE (T').
4. FTRA_4[Calculation score prediction matrix P]
FOR i=1 TO n DO
FOR k=1 TO m DO
IF Te_{ik} ≠0 THEN P_{ik} =PREDICT$\left(\mathrm{Tr},\ \mathrm{T}^{''},\ \mathrm{i},\ \mathrm{k}\right)$

4 Experimental Analysis and Results

4.1Test Dataset and Evaluation Indicators

Epinions were used as the test dataset. The dataset is outlined in Table 3.1. In the experiment, we used MAE, RMSE and Recall to validate the performance of the proposed method. Next, we used cross validation method to divide the dataset to training set and test set. The rating data by each user was chosen at a proportion of 10% as the test set. The remaining was training set.

4.2 Comparison Algorithms

We selected four recommendation methods for comparison test, as seen in Table2.

The six comparing algorithms are classified to three types in terms of data they depend on, like: user-based recommendation method (such as user-based CF, item-based CF, userMeanR-based, itemMeanR-based); usertrust-based recommendation method (i.e. userTrust-based); the method fusing rating data and trust relation (i.e. userCredibility-based). Additionally, the paper proposed FTRA method, the transfer matrix P can be only structured by trust matrix T, i.e. $P_{ij} = T_{ij} \cdot (\sum_{k=1}^{n} T_{ik})^{-1}$, called TA method. On the regard, FTRA method depends only on trust relation to make evaluation prediction. They can be described as follows:

4.2.1 Average Score Algorithm based on User (userMeanR-based). According to the historical rating data of target user, can calculate user the average score on the item value estimate, as average score is target users on other item score value. Referred to as the recommendation algorithm of userMeanR-based, such as formula (1):

$$P_{io}^{UserMean} = \bar{r}_i \qquad (1)$$

4.2.2Average Score Algorithm based on Item (itemMeanR-based). Users were scored data According to the item, the average score is calculated all item value. The mean score of the target item is as the target user to estimates of its score value. Referred to as the recommendation algorithm of itemMeanR-based, such as formula (2):

$$P_{io}^{UserMean} = \bar{r}_o \qquad (2)$$

Table 2. Six Comparison of the Relevant Experimental Selection Algorithm

Algorithm Category	Algorithm name
Recommendation algorithm based on user item rating data	Collaborative filtering algorithm based user (user-based CF) Collaborative filtering algorithm based item (item-based CF) Average score algorithm based on user(userMeanR-based) Average score algorithm based on item (itemMeanR-based)
Recommendation algorithm based on user trust relationship	Based on trust relationship algorithm (userTrust-based)
Recommendation algorithm of data fusion and trust relationship	Based on user global reputation value algorithm (userCredibility-based)

4.2.3 Based on Trust Relationship Algorithm (userTrust-based). Golbeck is based on the target user direct trust user (step trust propagation), target item rating is to estimate the underlying target user item ratings. Referred to as the recommendation algorithm of userTrust-based, such as formula (3):

$$P_{io}^{Trust} = \bar{r}_i + \frac{\sum_{j \in Raters} FT_{ij} \cdot (r_{j,o} - \bar{r}_j)}{\sum_{j \in Raters} FT_{ij}} \qquad (3)$$

4.2.4 Based on User Global Reputation Value Algorithm (user Credibility-based). Jebrin and Williams according to the user is the direct trust and indirect trust user number is the number of users of the item, item score values is close to average score of degree three aspects to describe the user's global reputation value. Referred to as the recommendation algorithm of userCredibility-based, such as formula (4):

$$P_{io}^{Cr} = \bar{r}_i + \frac{\sum_{j \in Creditor} Cr(v_j) \cdot (r_{j,o} - \bar{r}_j)}{\sum_{j \in Creditor} Cr(v_j)} \qquad (4)$$

4.3 Results and Analysis

It showed the different results of MAE, RMSE and Recall of userTrust-based and userCredility-based methods changing with parameters, as follows:

4.3.1 From Fig1-2, whe it's μ =0.2, userTrust-based method reached the best value for both MAE and RMSE, respectively 0.8660 and 1.1385. They both improved in the two indicators, compared with the method considering only the direct trust relation (μ =1) or the second indirect trust relation (μ =0);

4.3.2 In terms of Recall, when $\mu \in (0,1)$, userTrust-based strategy considers both direct trust relation and the second indirect trust relation. But the two relations were of different concern, so the Recall is invariable, i.e. 0.8810, 89.18% more than the value 0.4657, which was obtained by the method depending on users' direct trust relation (μ =1);

4.3.3 UserCredibility-based method has several parameters, of which δ is chosen as per experience. In the work [2], when the parameter set (μ =0.1, μ =3/9, μ =1/9, μ =5/9) was given, the method's MAE, RMSE and Recall are separately 0.8747, 1.1464 and 94.98%. After several tests, the optimal parameter set (μ =1.0, μ =0.1, μ =0.1, μ =0.8) was reached for the dataset in the experiment. MAE, RMSE and Recall changed accordingly to 0.8251, 1.0752 and 99.97%.

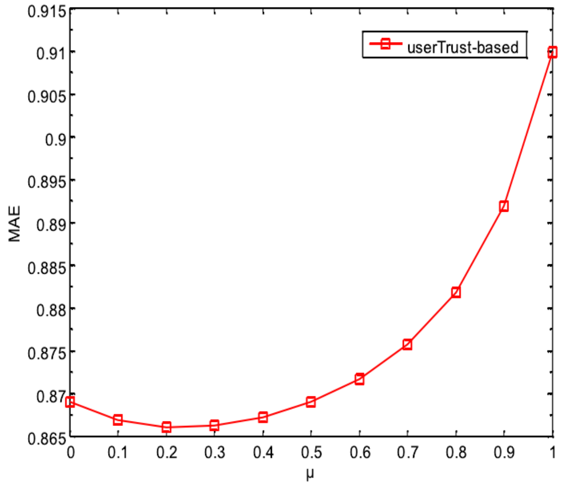

Figure 1. UserTrust-based Algorithm in the MAE Index with μ the Value Change Curve

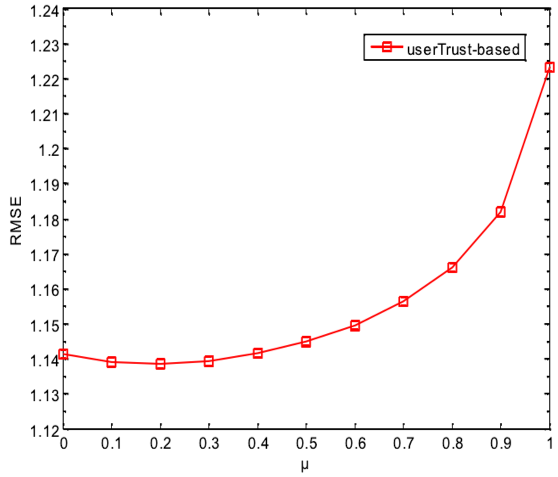

Figure 2. UserTrust-based Algorithm in the RMSE Index with μ the Value Change Curve

It provided curves of test results of FTRA and TA methods with parameter t, R and ρ. In FTRA and TA algorithms, step t, parameter R, decay factor ρ are parameters. Through iterative tests, we found when R=2.0, FTRA reached the best. So set R=2.0. We'll discuss when step t is infinite, i.e. (t=1, 2, 3), how FTRA and TA methods react to different ρ. The process is described in the following:

4.3.4 Fig3 shows the change of two methods' MAE values with different ρ when t=1, 2, 3. When t is bigger, on the whole, the MAE value diminishes and FTRA is always superior to TA. When t=1, the MAE value doesn't change with different ρ because at this point, trust was transferred only one step. The decay factor ρ does not affect the relative value of trust resources between terminal nodes. When t=2 or 3, the MAE value declines with ρ growing up. When t=3 and ρ=1.0, FTRA and TA reached optimal values, respectively 0.8143 and 0.8706.

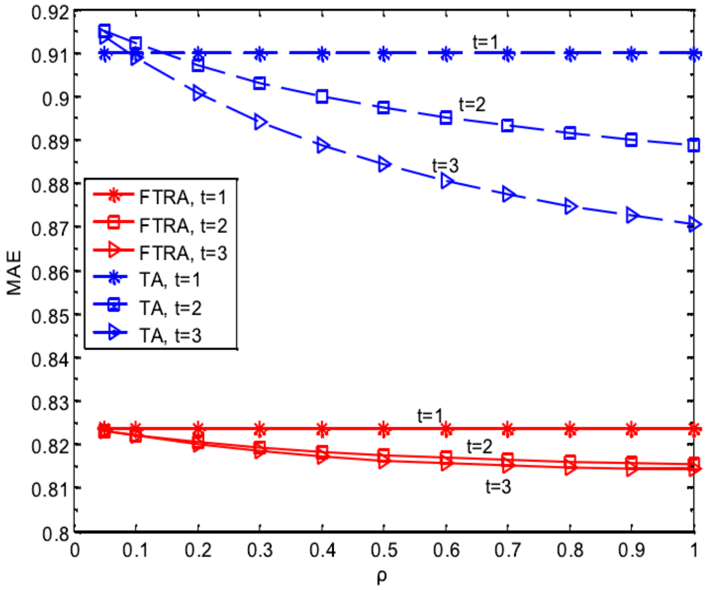

Figure 3. FTRA Algorithm and TA Algorithm in the MAE Index with ρ the Value Change Curve at Step t=1, 2, 3

4.3.5 Fig4 is the change of two methods' RMSE values with different ρ when t=1, 2, 3. Clearly, the RMSE changed similar to MAE with different ρ. When t=3, ρ=1.0, FTRA and TA performed the best, respectively 1.0635 and 1.1476 for RMSE.

4.3.6 When t=1, 2, 3, FTRA got the recall respectively 98.82%, 99.97% and 99.97%; TA got respectively 46.57%, 88.10% and 95.48%.

With the use of Katz indicator, we analyzed the change of the two methods with different ρ, when t becomes infinitely bigger, as follows:

4.3.7 From Fig5, in terms of MAE, FTRA performed generally better then TA. When ρ is 0.8 and 0.9, FTRA and TA reached the optimum, i.e. 0.8141 (better than 0.8143 when t=3, ρ=1) and 0.8560 (better than 0.8706 when t=3, ρ=1). FTRA improved MAE by 4.89% than TA.

4.3.8 In Fig6, when ρ is 0.8 and 0.9, FTRA and TA got respectively best values of RMSE, i.e. 1.0632 (better than 1.0635 when t=3, ρ=1) and 1.1218 (better than 1.1476 when t=3, ρ=1), with RMSE improved by 5.22%.

4.3.9 The two methods' recall didn't change with ρ because ρ only affects the attenuation of the trust propagation process. FTRA's recall is 99.97% and TA's 96.05%.

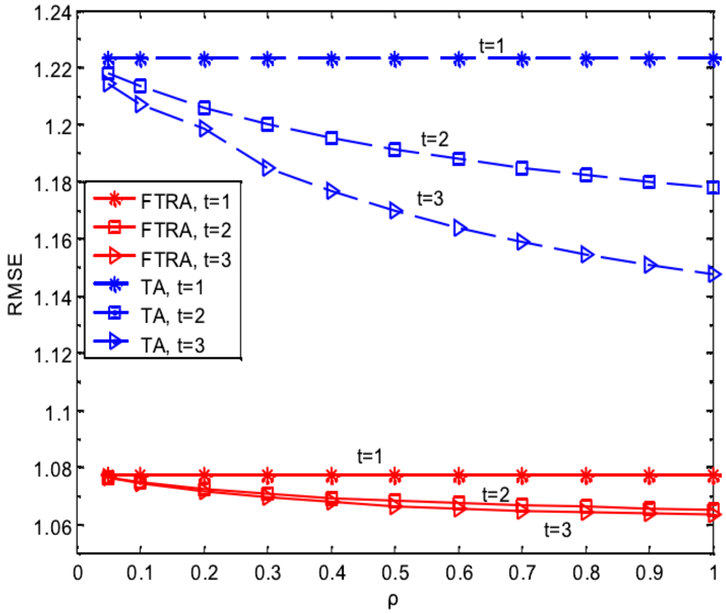

Figure 4. FTRA Algorithm and TA Algorithm in the RMSE Index with ρ the Value Change Curve at Step t=1, 2, 3

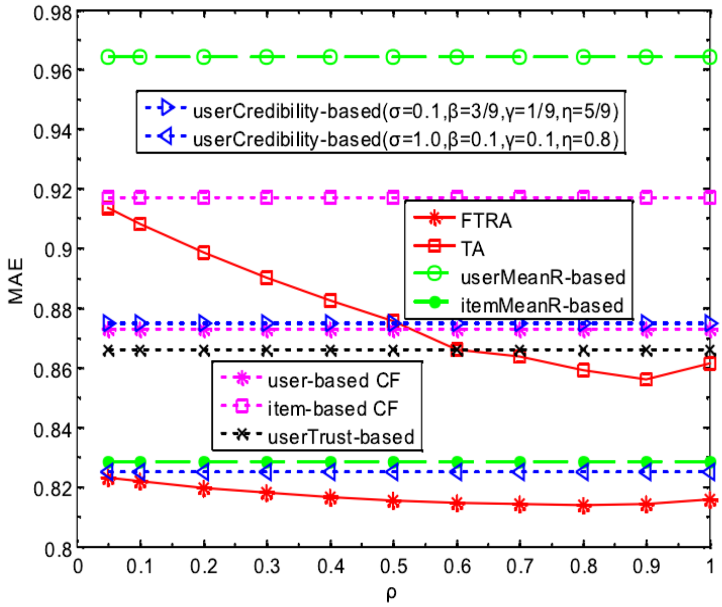

Figure 5. All Algorithm in the MAE Index with ρ Change Curve

The results of MAE and RMSE of the above methods with different ρ are presented in Fig. 5-6. Except FTRA and TA, other methods don't have ρ. That's why it's a straight line. From Fig. 5, when step t becomes infinitely bigger and ρ=0.8, FTRA performed better than others for MAE. Also in Fig. 6, FTRA's RMSE is still better than others in the same condition.

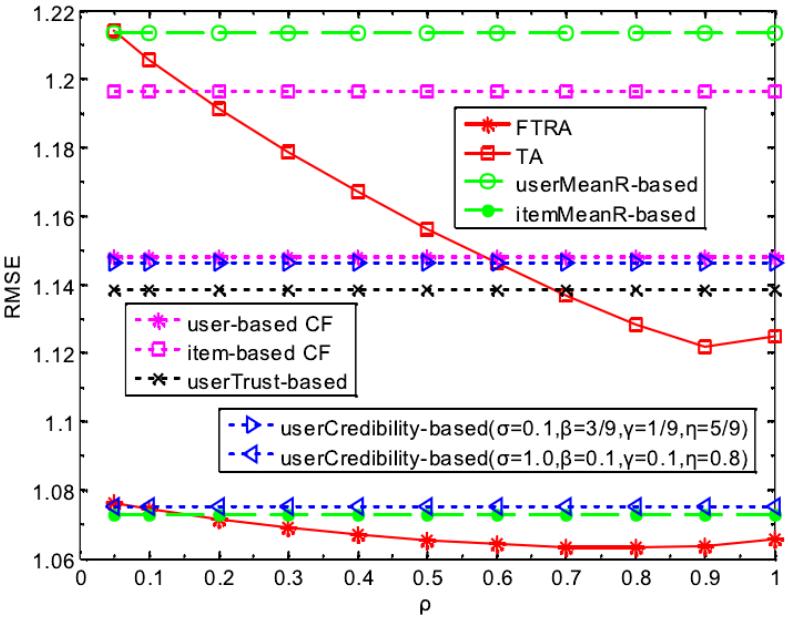

Figure 6. All Algorithm in the RMSE Index with ρ Change Curve

5 Conclusion

Traditional collaborative filtering algorithms lead to inaccurate results because of rating data sparsity when the similarity calculation method was applied to find similar neighbors. That eventually affected the accuracy of evaluation prediction. To solve the problem, on account of the internal association between user similarity and trust relation, the paper presented a new approach FTRA. The method firstly fused user trust relation and rating data information to construct a graphical model. Then based on Katz similarity and with full consideration of indirect relationship between users, it found more accurate similar neighbors for the target user, thus to generate more accurate predictive evaluation values.

References

[1] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: Asurvey of thestate-of-the-art and possible extensions”, IEEE Transaction on Knowledge and Data Engineering, vol. 17, no. 6, **(2005)**, pp. 734-749.

[2] Q. Yuan, S. W. Zhao and L. Chen, “Augmenting collaborative recommender by fusing explicit social relationships”, ACM Conference on Recommender System workshop on Recommender Systems and the Social Web, **(2009)**; New York.

[3] H. Kautz, B. Selman and M. Shah, “Combining social networks and collaborative filtering”, Communicatins of the ACM, vol. 40, no. 3, **(1997)**, pp. 63-65.

[4] P. Avesani, P. Massa and R. Tiella, “A trust-enhanced recommender system application: Moleskinga”, Proceedings of the ACM symposium on applied computing, Santa, **(2005)**.

[5] P. Massa and P. Avesani, “Trust-aware recommender systems”, Proceedings of the ACM conference on recommender systems, **(2007)**; Minneapolis.

[6] A. S. Jebrin and M. A, Williams, “Credibility-aware Web-based Social Network Recommender: Follow the Leader”, Proceedings of the 2nd ACM RecSys Workshop on Recommender System and the Social Web, **(2010)**; Barcelona.

[7] J. S. Breese, D. Heckerman and C. Kadie, “Empirical analysis of predictive algorithms for collaborative filtering”, Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, **(1998)**; San Francisco.

[8] D. Billsus and M. Pazzani, “Learning collaborative information filters”, Proceedings of International Conference on Machine Learning, **(1998)**; San Francisco.

[9] J. Canny, “Collaborative filtering with privacy via factor analysis”, Proceedings of the 25th International ACM SIGIR Conference on Research and Development in Information Retrieval, **(2002)**; New York.

[10] S. Deerwester, S. Dumais, G. Furnas, T. Landauer and R. Harshman, "Indexing by Latent Semantic Analysis, The Journal of the American Society for Information Science, vol. 40, no. 6, **(1990)**, pp. 391-407.
[11] T. Hofmann, "Latent semantic models for collaborative filtering", ACM Transaction Information Systems, vol. 22, no. 1, **(2004)**, pp. 89-115.
[12] T. Hofmann, "Unsupervised learning by probalilistic latent semantic analysis", Machine Learning, vol. 42, **(2001)**, pp. 177-196.
[13] C. Zeng, C. X. Xingand L. Z. Zhou, "Similarity Measure and Instance Selection for Collaborative Filtering', Proceedings of the 12th Internationa Conference on World Wide Web, Budapest, **(2003)**.
[14] Z. G. Wei, L. D. Yi and L. Peng, "Collaborative filtering recommendation algorithm based on cloud model", Journal of software, vol. 18, no. 10, **(2007)**, pp. 2403-2411.
[15] M. S. Granovetter, "The strength of weak ties. American Journal of Sociology, vol. 78, no. 6, **(1973)**, pp. 1360-1380.
[16] P. Resnick and H. R. Varian, "Recommender system", Communications of the ACM, vol. 40, no. 3, **(1997)**, pp. 56-58.
[17] J. Golbeck, "Generating predictive movie recommendations from trust in social networks", Paper presented at the iTrust, vol. 39, no. 86, **(2006)**, pp. 93-104.
[18] P. Avesani, P. Massa and R. Tiella, "A trust-enhanced recommender system application: Molesking", Proceedings of the ACM symposium on applied computing, Santa, **(2005)**.
[19] A. S. Jebrin and M. A. Williams, "Credibility-aware Web-based Social Network Recommender: Follow the Leader", Proceedings of the 2nd ACM RecSys Workshop on Recommender System and the Social Web, **(2010)**; Barcelona.
[20] R. Sinha and K. Swearingen, "Comparing recommendations made by online systems and friends", Paper presented at the second DELOS network of excellence workshop on personalization and recommender systems in digital libraries, **(2001)**, pp.106-110.
[21] C. N. Ziegler and G. Lausen, "Analyzing correlation between trust and user similarity in online communities", Paper presented at the second international conference on trust management, Oxford, vol. 29, no. 95, **(2004)**, pp. 251-265.

Author

Li Xiaofeng, he is an advanced member of china computer federation and he is an associate professor in Heilongjiang International University. His research interest includes data mining and intelligent algorithm.