International Journal of Hybrid Information Technology
Vol.8, No.12 (2015), pp. 171-180
http://dx.doi.org/10.14257/ijhit.2015.8.12.11

Research on a Collaborative Filtering Recommendation Algorithm

Qi Zheng

*School of Computer Science and Engineering,
Jilin Jianzhu University, Changchun 130000, China
181168571@qq.com*

Abstract

Aiming at the problem that the traditional Collaborative Filtering algorithm has low recommendation accuracy, in the paper, we propose a collaborative filtering recommendation algorithm based on the global trust degree integrating the direct trust information in the social networks. We first transform the local trust relationships to the global trust relationship by the rules in the trust network, and get the trust rank of all users in the trust networks; Then we use the global trust value to instead of the similarity information value as the weights of a predicted formula in the traditional collaborative recommendation algorithm, and integrate the weights to the matrix factorization-based recommendation model.

Keywords: *Recommendation, Trust of global, Matrix factorization*

1. Introduction

Related researches suggest that personalization recommendation systems can solve well the information overload problem mentioned before [1-2]. Of them, the most widely and successfully used recommendation system in e-commerce field is collaborative filtering technology, like Ebay and Amazon, which applied collaborative filtering recommendation technology [3-4]. The most important constitute in the recommendation system is recommendation algorithms. The performance and nature of recommendation systems depend largely on recommendation algorithms. However, contemporary scholars' recommendation systems are based on the target user's recommended item of the current user rated item, regarding user-item rating matrix as only information source [5]. But such matrix data are very sparse, ignoring the impact of the trust information between online social network users on user preferences, which largely causes low accuracy rate of recommendation results. Among plenty of factors impeding the development of e-commerce network, the absence of trust mechanisms is the major reason [6-7].

2. Matrix Decomposition CF (Collaborative Filtering) Algorithm based on Global Confidence Degree

2.1. Recommendation Model

The collaborative filtering recommendation algorithm fused with trust information belongs actually to user-based collaborative filtering recommendation technology. The scoring predicted value is calculated based on scoring of items by neighboring users. In the trust-based collaborative filtering recommendation algorithm, trust users constitute neighborhood user set, instead of being generated by user groups with similar interests like before. The Trust-Rank-MF model based on global trust degree is designed to

ISSN: 1738-9968 IJHIT

transform with PageRank algorithm the local credibility between users into global trust degree and then fuse into the recommendation process, using trust information to filter out any item without relation to user requirements and thus to make appropriate recommendations to target users.

The model has three parts: data input, recommendation processing, and recommendation service. The second part is core. The input of original data is also important. By integrating closely three parts, we can finally implement the recommendation oriented to trust information. The recommended model is as shown in Figure 1.

Data input refers to the entry of two rating matrixes, starting from the primitive data of recommendation systems to get two rating matrixes after quantification treatment: user-item rating matrix and user-user trust rating matrix.

The recommendation processing part: after recommendation processing, we can get user-item rating information and user-user trust rating information; then with the obtained trust information we can acquire one-to-one local trust degree of past recommendations offered by target users to adjacent users; further on, iterating with PageRank algorithm, we can get global credibility; next, by incorporating user's global trust degree into the recommendation algorithm based on matrix factorization, we have the collaborative filtering recommendation method based on global confidence degree.

In the recommendation service part, we can get predicted scores with the above method and calculate user's predictive marks for one item to eventually generate recommendation lists for target users.

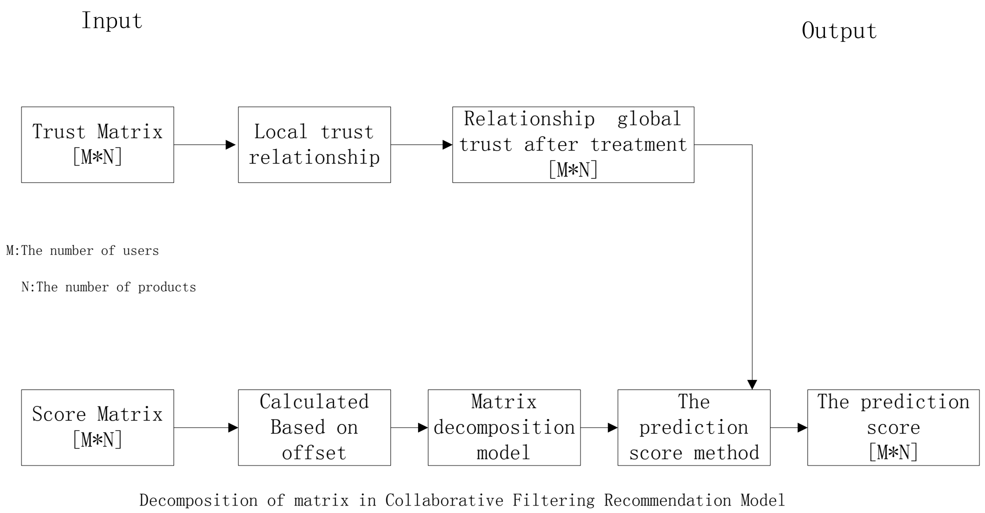

Figure 1. Recommendation Model based on Global Trust Degree

2.2. Basic Steps

2.2.1 Fetch user Trust Scores and Create Trust Rating Matrix. Convert users' trust information into numerical grades by some rules; then, establish user trust rating matrix in the matrix format. Here trust rating users are normal users left after some offensive ones are removed. The experiment here used Epinions dataset. The inter-user one-to-one local trust rating Figure is integer like 0 or 1. As shown in Table I, the rating is made with binary scoring mechanism as 1 or 0. Also from the table, we can know that direct trust ratings between users are very few, that is, direct trust matrix is quite sparse.

2.2.2. Measure of Global Trust Degree. The information of local trust between users which is acquired from trust rating matrix is what we said trust matrix, TM in short. Here we use well-known PageRank algorithm [8-9], which is renamed global trust iteration algorithm. It has the main idea: the global credibility between nodes is jointly determined by local opinions of other nodes with which it has interactive behaviors and the global confidence of those nodes themselves [10]. PageRank algorithm can evaluate objectively the importance of users. The formula is as follows:

$$PR(u) = (1-\alpha) + \alpha \sum_{v:(v,u)\in E} \frac{PR(v)}{O(v)} \tag{1}$$

By equation (1) we can get PR value of both u and v. In the beginning of iteration, PR value of each node is not known. So in the algorithm, we assume the initial PR value of each node in the whole directed graph is $1/(n+1)$, then by substituting it to equation (1), we can calculate $PR_1(u)$ and $PR_1(v)$; next, substitute the obtained result into equation (1) for the next iterative operation till $|PR_k(u) - PR_{k-1}(u)|$ is converged to a very little Figure ε, which is stop condition; return $PR_k(u)$ value, *i.e.*, PR value of node u. The algorithm is described as follows:

Algorithm 1: PageRank algorithm
Inputs: TM, u , v
Outputs: $PR(u)$
1 k=0
2 $PR_0(u) = 1/(n+1), PR_0(v) = 1/(n+1)$
3 For k=1, 2,…,n DO
4 $PR_k(u) = (1-\alpha) + \alpha \sum_{v:(v,u)\in E} \frac{PR_{K-1}(v)}{O(v)}$
5 Until $|PR_k(u) - PR_{k-1}(u)| < \varepsilon$
6 End For
7 Return $PR(u)$

Table 1. User-user Trust Rating Matrix

	$User_1$	$User_2$	$User_3$	$User_m$
$User_1$	0	1	0	0
$User_2$	0	0	1	1
$User_3$	1	0	0	1
$User_m$	1	1	0	0

2.2.3. Recommendation Generation of Matrix Factorization Collaborative Filtering Recommendation Model. We use u and v for users in the recommendation system; i and j for commodity items in the system; r_{ui} for user u's known evaluation marks for the commodity item i; $\hat{r}_{ui}$ for the predicted assessment point of user u for commodity item i; higher predicted scores suggest user's greater interest in the commodity. User rating information can be expressed as a rating matrix R with user and commodity item respectively as x and y axis. The two sets are set $U = \{User_1, User_2, ..., User_m\}$ which includes m users and set $I = \{Item_1, Item_2, ..., Item_n\}$ which has n commodity items. It shown in Table 2.

Table 2. The User-Item Rating Matrix

	$Item_1$	$Item_2$	$Item_3$	$Item_n$
$User_1$	r_{11}	null	r_{13}	r_{1n}
$User_2$	r_{21}	r_{22}	null	r_{2n}
$User_m$	1	r_{m2}	null	r_{mn}

From Table 2, we learn that what's known is one user has rated one commodity item; what's lack is the user has not yet evaluated the commodity item. Traditional CF recommendation algorithm's task is to predict unknown scores in the matrix by referring to known ratings and other relevant information.

2.2.4. Collaborative Filtering Recommendation Algorithm based on Global Confidence Degree (TrustRank-MF). In the recommendation system, apart from user-item rating information, one user is allowed to add confidence and non-confidence information about other users. The local trust in trust networks is calculated with PageRank algorithm to get the global credibility of each user. We'll introduce the use of user's global trust degree weighting approach to compute the similarity weight between neighbors and the subject. The formula is as follows:

$$\hat{r}_{ui} = b_{ui} + \sum_{v \in R_k(u)} (r_{vi} - b_{vi}) \times w_u \tag{1}$$

2.3. The Algorithm Description

Algorithm 2: Collaborative filtering recommendation algorithm based on trust
Inputs: R,T,f,k
Outputs： $\hat{r}_{ui}$

(1)**For** u=1, 2,…, m **Do**
(2) $w_u \leftarrow PR(u)$ // Trust Rank Algorithm
(3) Initialize p_u, q_i=Random(-0.01,0.01)
(4) **For** *Iterations*=1 to15 **Do**
(5) **For** all $i \in S(u)$ **Do** //u have historical rating set $S(u)$
(6) $r_{ui} \leftarrow u + b_u + b_i + \sum_{v \in R_k(u)} (r_{vi} - b_{vi}) \times w_u + p_u \times q_i^{\mathrm{T}}$
(7) $e_{ui} \leftarrow r_{ui} - \hat{r}_{ui}$
(8) $b_u \leftarrow b_u + \gamma \cdot (e_{ui} - \lambda \cdot b_u)$
(9) $b_i \leftarrow b_i + \gamma \cdot (e_{ui} - \lambda \cdot b_i)$
(10) **For** $j = 0,1,\ldots,f$ **Do**
(11) $p_{uk} \leftarrow p_{uk} + \gamma(e_{ui} q_{ik} - \lambda p_{uk})$
(12) $q_{ik} \leftarrow q_{ik} + \gamma(e_{ui} p_{uk} - \lambda q_{ik})$
(13) **End For**
(14) **End For**
(15) **End For**
(16) $\hat{r}_{ui} \leftarrow \hat{r}_{ui} + (\mu + b_u + b_i + (p_u \cdot q_i^{\mathrm{T}}) \cdot w_u)$
(17)**End For**
(18)**Return** $\hat{r}_{ui}$

3. Experimental Analysis and Results

3.1. Experiment Data Set

The experiment adopts Epinions data set, collected from a famous e-commerce merchandise rating website. The data set statistics are listed in Table3, inclusive of two data sets frequently used by the recommendation system. Table3 shows the quantity of rated trust relationship among users. It can be deferred that the coverage of trust appraising is below 1%. The numerical score of user's rating about merchandise items is integral figure from 1 to 5. Such scores represent user's different fondness of items. Trust relationship among users can be rated 1 or 0. 1 means belief between two users, while 0 means no trust.

Epinions data set used in the experiment is rather sparse in terms of either user-item rating data or user-user trust relationship rating data. According to statistics, 48.4% of users in the set have less than five rating records, the density of rating matrix below 0.015%; 52.2% of user trust relationship rating records is below 5, the coverage rate of trust rating less than 1%. From Table3 it's noted that Movielens and Eachmovie are two most well-known data sets for the collaborative filtering recommendation system. Also based on scores of items, we can get scoring density of the two sets: 4.25% and 2.29%. But the experiment used Epinions data set instead of Movielens or Eachmoive data set because the latter two don't have user-user trust rating information.

Table 3. Data Set Description

Data Sets	Users	Items	ItemsRatings	TrustRatings
Epinions	40100	149856	663478	377191
Movielens	5060	3800	1111309	-
Eachmovie	68834	1578	3811789	-

3.2. Test Environment Configuration

This experiment mainly using Java language in this paper. In Eclipse platform to achieve. The experimental configuration specific includes two parts of hardware configuration and software configuration.

(1) Hardware configuration: Intel Core Duo processor, 4G memory, 500G hard disk.

(2) Software configuration: the development tools of Eclipse7.0, the compiler environment is Jdk1.6.0, the operating system is win7.

3.3. Evaluation Index in the Experiment

To assess the quality of the recommendation algorithm, we introduce two key indicators:

3.3.1. Accuracy Rate. Root mean square error (RMSE) is the most common indictor for assessing accuracy rate of the recommendation method. It's used to measure the square deviations between the assumed rating value and user's actual rating value. If one user's comments are known about some items, the standard is adopted. Here we use it to evaluate the precision rate of the recommendation method. By calculating the difference between user's predicted score and the actual one, RMSE examines the accuracy rate of such prediction. The smaller the RMSE is obtained, the higher the accuracy rate implies and the better quality the recommended service proves.

RMSE Formula is defined:

$$RMSE = \sqrt{\frac{\sum_{(u,i)|R_{test}} (r_{ui} - \hat{r}_{ui})^2}{|R_{test}|}} \quad (2)$$

3.3.2. Coverage Rate. The coverage rate is the percentage of the number of successful predicted scores by the recommendation system against that of all ratings in the whole test set. If the coverage rate is lower, the quantity of items recommended by the system is fewer and that the user gets worse recommendation results. Otherwise, the user gets better recommendation results because of higher coverage rate and more items recommended by the system. The formula is as follows

$$COV_R = \frac{N_d}{N} \quad (3)$$

3.4. Validation of Collaborative Filtering Recommendation Algorithm based on Global Confidence Degree

To verify the effectiveness of evaluation method, we compare the proposed algorithm and the following method. Bias-MF [11] is the most basic recommendation approach based on matrix factorization, to which both user and item's benchmark offsets are introduced, without consideration of trust relationship.

3.4.1. Determination of Experimental Parameters. The experiment divided randomly Epinions data set to 80% training set and 20% testing set. In testing 80% training set, parameter step factor γ and penalty factor λ are reached by cross validation; hidden dimension and iterative times are decided by measuring accuracy rate and required time. In all experiments here, make $\gamma = 0.01$; $\lambda = 0.003$; f is hidden dimension, f=100; iterative times is 15; then do experiment on 20% testing set; this will repeat five times. The final experimental results are arithmetic mean of values got from five iterative experiments.

The proposed Trust Rank-MF algorithm includes several parameters. The size of their values affects often the performance of recommendation methods. So in this part we discuss about how to determine those parameters. Such parameters include penalty factor λ (*i.e.,* regular parameter), step length factor γ, hidden dimension as well as iterative times. In the experiment, we find step length factor γ affects greatly RMSE results of the algorithm, especially affecting the convergence speed of the gradient descent algorithm. If the optimal step length γ is too small, the gradient descent's converging speed will become slow; but if the optimal step length γ is too big, in the optimization process, it won't converge instead of moving round the extreme point, causing the divergence. In order to ensure rapid convergence of target function rather than divergence, it's required firstly to select appropriate value of step length factor γ. As seen from experimental results, the influence of penalty factor λ is not quite noticeable; but different values of λ have certain impacts on the variations of RMSE results.

In the experiment, we considered the time required for each iteration and the results of the algorithm's performance indicator RMSE. Parameters are defined by firstly choosing 15 as iterative times; hidden dimension of data set for the validation test is f=100.

To begin, we make cross validation of step factor γ and then verify penalty factor λ. Based on the previous experience, we choose well the value range of parameters. Those selected values will enable the algorithm to reach better accuracy rate. Next, from those parameter values, we define one fixed value for testing. Finally after repeated tests, we make the value $\gamma = 0.01$, $\lambda = 0.003$ of two parameters which are supposed to make the algorithm get satisfactory experimental results.

Figure 2 shows how the optimal step length factor γ affects RMSE when its value is different.

In the experiment, nine different values of γ is selected from $0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.02$. In Figure 1, when penalty factor $\lambda = 0.003$, with the aggrandizement of step factor γ, RMSE value are waning; when step factor $\gamma = 0.01$, RMSE value is minimal; when step factor $\gamma > 0.01$, no specific RMSE value is acquired, suggesting that when step factor $\gamma > 0.01$, the experimental result overflows. Hence for better results in the experiment, we need to select properly the value of $\gamma = 0.01$.

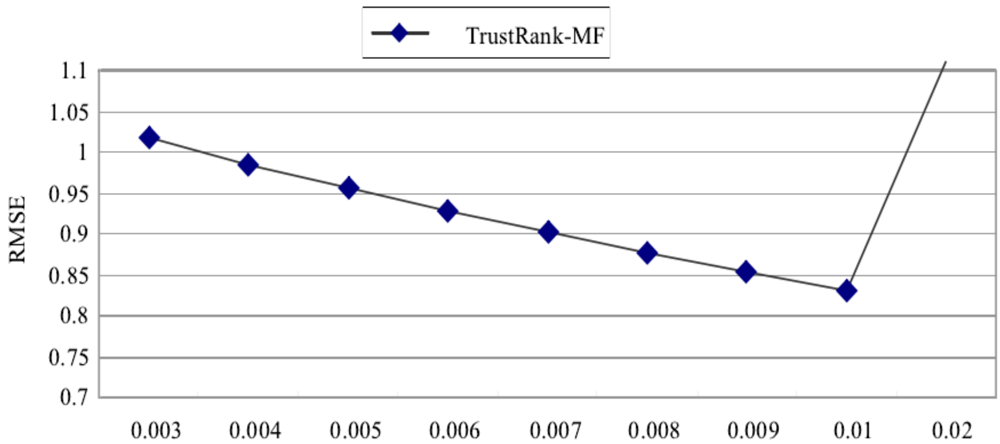

Figure 2. Effect of γ Value on RMSE

Table 4 shows how the penalty factor λ affects RMSE when its value is varied. The value of penalty factor λ has some effects on the recommendation results.

In the experiment, six different values of penalty factor λ is selected from $0.006,\ 0.005,\ 0.004,\ 0.003,\ 0.002,\ 0.001$. In the table, looking at λ 's value column from top to bottom, we learn that with decreasing of λ 's value, RMSE results tend to decrease progressively for both Bias-MF and TrustRank-MF, although the reduction is not great. When λ 's value is one of them, the two methods' RMSE results are both minimal. Then with continuous reduction of λ 's value, RMSE results are increasing. It indicates that when step factor $\gamma = 0.01$ and penalty factor $\lambda = 0.003$, the two methods' RMSE results can reach the least values. So in the experiment we choose penalty factor $\lambda = 0.003$.

Table 4. Effects of Penalty Factor λ on RMSE

		Bias-MF	TrustRank-MF
$\gamma = 0.01$	$\lambda = 0.006$	0.929	0.831
	$\lambda = 0.005$	0.928	0.830
	$\lambda = 0.004$	0.927	0.831
	$\lambda = 0.003$	0.925	0.829
	$\lambda = 0.002$	0.925	0.830
	$\lambda = 0.001$	0.924	0.830

Besides, in the experiment we analyzed when other parameters are fixed, how different values of the hidden dimension f affects RMSE results. It shown in Figure 3.

From Figure 3, when the hidden dimension f is growing, the proposed algorithm's RMSE results are becoming small. When f varies in the interval [50, 200], the proposed algorithm's RMSE results are decreasing rapidly. When f>200, RMSE results are becoming stable, *i.e.*, no any change of RMSE results. In this part, we consider both the required time for iteration and accuracy rate of the algorithm. Let hidden dimension f=100.

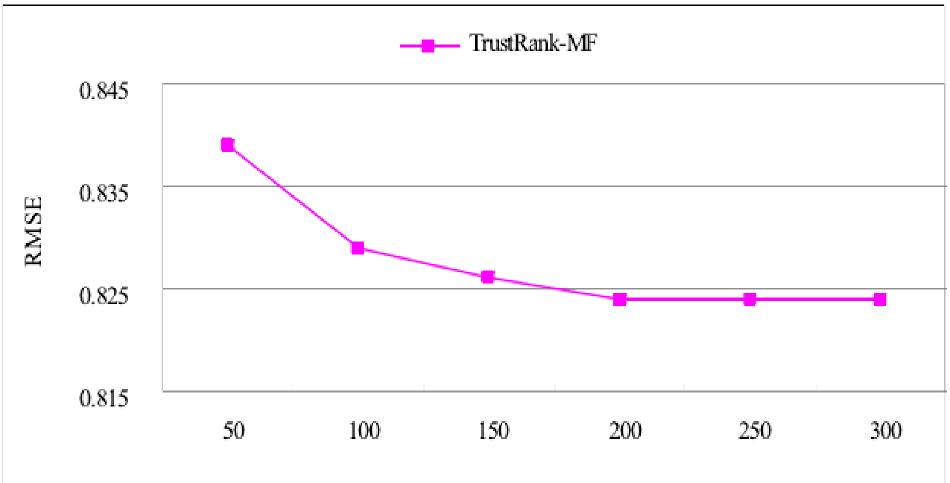

Figure 3. Effects of f Value on the RMSE

3.4.2. The Analysis of Experimental Results. We tested Bias-MF collaborative filtering recommendation algorithm with Epinions dataset for the validation. When parameters are fixed and the condition is same, the optimal result is 0.925. After validating Bias-MF collaborative recommendation algorithm and the improved recommendation strategy proposed here, we compared their RMSE results. The proposed method improved clearly RMSE results than Bias-MF algorithm. See Table 3 for details.

Table 3. RMSE Value of Two Algorithms

Algorithm	RMSE
Bias-MF	0.925
TrustRank-MF	0.831

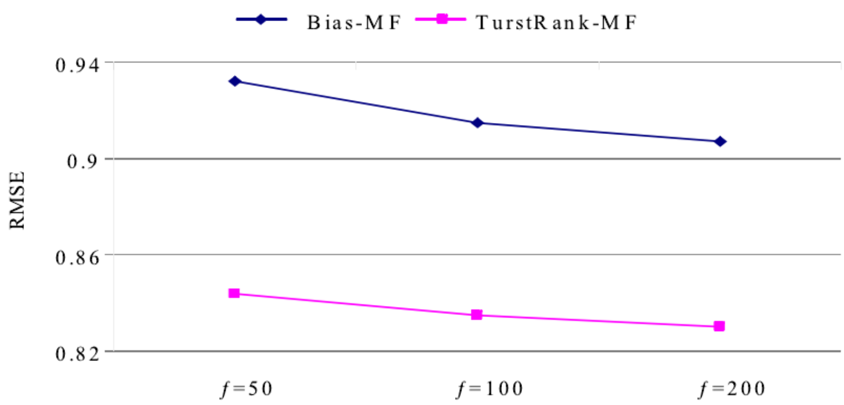

Figure 4. The Two Algorithms in Different Dimensions f on RMSE

As learnt from the above Table 3, the proposed algorithm improved RMSE than Bias-MF collaborative filtering recommendation method by approximately 0.1. Both strategies applied the currently popular matrix factorization technology, but they were improved from different aspects. The main difference is TrustRank-MF algorithm introduced the method involved with trust degree in sociology. It considered that confidence degree will make recommendation methods more comprehensive and thus recommendation results accord better to actual situations.

In Figure 4, with increasing hidden dimension f, two methods' RMSE results tend to decrease gradually and precision rates become better and better; however, the recommendation precision RMSE values are not enlarged greatly; and time for each iteration becomes longer and longer along with growing dimension f. For the trade-off between recommendation precision and time required for iterations, we make f=100, $\gamma = 0.01$, $\lambda = 0.003$ as benchmark in the experiment.

4. Conclusion

This paper proposes a collaborative filtering recommendation algorithm based on global trust. The local trust users of the Webpage ranking PageRank algorithm transformed into the global trust value, instead of the traditional collaborative filtering recommendation prediction formula of similarity weights. Proposes an improved collaborative filtering recommendation algorithm based on trust. This paper describes the concrete realization of the algorithm, effectively improve the accuracy of recommendation system.

References

[1] X. Xiaowei, Research and application of collaborative filtering recommendation algorithm based on trust recommendation system in e-commerce based on Donghua University, **(2013)**.

[2] L. Xudong, C. De and W. Huimin, "An improved collaborative filtering algorithm", Journal of Wuhan University of Technology (Information & Management Engineering Edition), vol. 04, **(2010)**, pp. 550-553.

[3] J. Yaya and Y. C. Mu, "An improved collaborative filtering recommendation algorithm based on trust degree", The modern technology of Library and information science, vol. 10, **(2010)**, pp. 49-53.

[4] C. Hao, "Research on Collaborative Filtering Recommendation Algorithm Based on Web usage mining", Zhejiang Sci-Tech University, **(2010)**.

[5] X. Xiaowu, "Research on Collaborative Filtering Recommendation Algorithm Based on trust model", University of Science & Technology China, **(2011)**.

[6] X. Zhe, "The beam Jingfan, Zhu Qing. Collaborative filtering recommendation algorithm of adaptive multi dimension", Mini micro systems, vol. 11, **(2011)**, pp. 2210-2216.

[7] W. Hui, B. Yijie, Z. Zhe and M. Ruimin, "A collaborative filtering algorithm based on trust", The application of the computer system, vol. 07, **(2014)**, pp. 131-135.

[8] Z. Min and L. Feng, "Study on the algorithm of PageRank", Micro computer information, vol. 27, no. 8, **(2011)**, pp. 143-144.

[9] M. Weiyu, "Evaluation method of Web information credibility based on improved PageRank", Hebei University, **(2011)**, pp. 5-44.

[10] M. Gori and A. Pucci, "Item rank: a random-walk based scoring algorithm for recommender engines", IJCAI, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India. USA: Morgan Kaufmann Publishers Incorporated, **(2007)** January 6-12, pp. 778-781.

[11] Y. Koren, R. Bell and C. Volinsky, "Matrix factorization techniques for recommender systems", IEEE Computer, vol. 42, no. 8, **(2009)**, pp. 30–37.

Author

Qi Zheng received her M.S degree from Tianjin Polytechnic University. She is a Lecturer in Jilin Jianzhu University. Her research interests include computer application technology.