International Journal of u-and e-Service, Science and Technology
Vol.7, No.5 (2014), pp.359-368
http://dx.doi.org/10.14257/ijunnesst.2014.7.5.32

Research on Real-time Personalized Recommendation Algorithm

Liu Lijun

Department of Logistics Information, Hunan Vocational College of Modern Logistics,Changsha 410131,china
aoron@126.com

Abstract

Personalized recommendation algorithm is core to recommendation systems, which matters the quality of recommendations of such system. The paper proposed an improved Slope one recommendation algorithm, M-Slope one. Based on real-time user interest model, the new method can calculate similarities between users and establish neighboring user groups to narrow down search scope of related items and improve the average rating differential equation for items. The algorithm proves its effectiveness for improving the precision of recommendations.

Keywords: *User Interest Model;Slope one; Recommended Algorithm*

1. Introduction

Studies on personalized recommendations make use of data that provided by users or data mining technologies to collect users' preferences from historical records, aiding users in acquiring interesting information to solve the overload problem with the Internet information. Compared with information retrieval techniques like search engine, recommendation systems have the biggest advantage of their ability to offer users with individualized services. Search engines can only meet mainstream requirements and the process of getting information is in the lead of users. Search results are provided in accordance to users' key words [1-2]. Users need to find out what they really want from those results. For recommendation systems, the whole searching process is controlled bythe systems, which instruct users based on their personal information or browsing behaviors. For different users with unlike interests/favors, the feedback of the process will vary. If such systems can give high quality service of recommendations, there will be great effects on improving users' loyalty and preventing the loss of them. Individualized recommendation algorithm and user interest model are both focalon personalized recommendations, which is likewise significant to the recommendation quality of those systems [3-4]. The real-time concept for the real-time personalized recommend strategy has two parts:

a. Synchronization of user interest model, i.e. displaying users' interest and preference simultaneously and accurately and applying the real-time users' interest model to personalized recommendation algorithms [5];

b.Synchronization of feedback of recommendation results, returning accurate recommendation results to users in very short time. Firstly, through comparison of recommendation quality between the traditional collaborative filtering recommendation algorithm and the one based on real-time user interest model, the effect of such model will be verified on the improvement of recommendation quality; the model will be employed to upgrade the traditional Slope one recommendation algorithm and develop the new one, M-Slope one. The precision of the algorithm is improved through the test [6].

ISSN: 2005-4246 IJUNESST

2. Collaborative Filtering Recommendation Algorithm based on User Interest Model

2.1. Description of the Algorithm

Synergic filtering recommendation algorithm is the most useful and successful technique. The earliest use of the collaborative filtering technology was found in the system called Tapestry to overcome the problem of information overload at Xerox PARC Institution. Then, similar systems appeared successively such as GroupLens, Ringo and Jester, which covered plentiful resources e.g. news information, movies, music, and even jokes & humors [7-8]. In daily life,people used to ask for opinions from workmates and friends before making their own decisions. It is the same as the idea of collaborative filtering algorithm. The method relies on the proposition that similar users make similar scorings about the same item to make recommendations to target users. By describing users' interest and defining similarities between them based on their assessment of an item, the method finds out user groups which have identical interest and preferences as to generate recommendations after referring to those users' ratings to foresee the appraisal by target users about one item [9].

The traditional collaborative filtering recommendation algorithm has the following steps in the process of recommending: representation of user interest, i.e. user-item scoring matrix, selection of neighboring usersand generation of suggestions:

2.1.1. Representation of users' interest. Collaborative filtering algorithm consideredusers' interests and preferences from the point of their evaluation ofitems [10]. The method regards user-item scoring matrix as input data of it. The matrix is $m \times n$ dimensional, where, m and n means separately the number of users and items in the system (Table1); R_{ij} is $user_i$ rating about the $item_j$

Table 1 The User Item Rating Matrix

	$item_1$	$item_2$	$item_3$	$\cdots$	$item_n$
$user_1$	R_{11}	R_{12}	R_{13}	$\cdots$	R_{1n}
$user_2$	R_{21}	R_{22}	R_{23}	$\cdots$	R_{2n}
$user_3$	R_{31}	R_{32}	R_{33}	$\cdots$	R_{3n}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$user_m$	R_{m1}	R_{m2}	R_{m3}	$\cdots$	R_{mn}

2.1.2. Selection of nearby users. Choosing neighbors is to establish neighboring user groups of target users. Users who have similar interest and favors as target users are put in the neighboring user groups. Suggestions are offered to target users based on neighbors' ratings about one item. In the process of selecting adjacent users, the user-item scoring matrix is used as data source, to figure out similarities between users and thus select K users who have higher similarities as neighbors. There are many similarity calculation between users,Methods commonly have cosine similarity, person correlation similarity, adjust cosine similarity.

2.1.3. Generation of recommendations. Based on the assumption that similar users rate identically the same item and with users' neighbors, the collaborative filtering recommendation method can predict target users' scores about items which are not evaluated by referring to other neighbors' marks. Under normal circumstances, there are two kinds of results recommended to target users: one is users' rating prediction of any item; the other is Top-N suggestion, i.e. computing users' predictable scorings of items which are not rated, with the first N highest scorings offered to target users. At present, there are three prediction techniques.

$$P_{ik} = \frac{\sum_{u_j \in NNU} R_{jk}}{K} \quad (1)$$

$$P_{ik} = \frac{\sum_{u_j \in NNU} sim(u_i, u_j) \times R_{jk}}{\sum_{u_j \in NNU} sim(u_i, u_j)} \quad (2)$$

$$P_{ik} = \bar{R}_i + \frac{\sum_{u_j \in NNU} sim(u_i, u_j) \times (R_{jk} - \bar{R}_j)}{\sum_{u_j \in NNU} sim(u_i, u_j)} \quad (3)$$

P_{ik} is u_i rating about the $item_k$, $sim(u_i, u_j)$ expressed u_i and u_j similarity of user. R_{jk} is u_j rating about the $item_k$,NNU(Nearest Neighbor User)indicates that nearest neighbor set of the user's u_i. K represents the number of neighbor users. $\bar{R}_i$ and $\bar{R}_j$ represent the user u_i and u_j item mean scores.

2.2. Collaborative Filtering Recommendation Algorithm based on Real-time User Interest Model

Here we present the collaborative filtering recommendation algorithm M-CF, which utilizes real-time user interest model to represent users' interests and preferences. The model can acquire users' preferences through virtue of their browsing interest, mitigating effectively the problem of cold start. Besides, the method does not require active offering of item's scoring data, reducing greatly the degree of coupling between the system and users' participation.

The process of the recommendation by the algorithm is described as follows:

Step1: Use users' labeling information to form mappings for standard labels through the process of standardization; displaying users' interest in the real-time users' interest model, user interest model is in short; standard labels are fewer and more stable than items. Matrix of user standard tag as shown in Table2.

Table2 Matrix of User Standard Tag

	Tag_1	Tag_2	Tag_3	$\cdots$	Tag_n
$user_1$	L_{11}	L_{12}	L_{13}	$\cdots$	L_{1n}
$user_2$	L_{21}	L_{22}	L_{23}	$\cdots$	L_{2n}
$user_3$	L_{31}	L_{32}	L_{33}	$\cdots$	L_{3n}
…	…	…	…	…	…
$user_m$	L_{m1}	L_{m2}	L_{m3}	$\cdots$	L_{mn}

Where, L_{ij} is interest of standard label Tag_j.

Step2: The measurement of users' interest degree is objective, different from users' scoring, which needs to consider the rating differentiation of different users.
Step3: Use users' interest model vectors to build linear regression equation to estimate users' interest degree of every single item, formula (4), then make recommendations to users by Top-N method.

User u_i is interest in the project p_j, calculated as:

$$I_{ij} = \sum_{t_k \in T} w_k \times p_k^j \tag{4}$$

Where, $p_k^j = \begin{cases} 0 & t_k \notin \text{text feature of item} \\ 1 & t_k \in \text{text feature of item} \end{cases}$ (5)

3. Improved Slope One Recommendation Algorithm

3.1. Slope One Algorithm

Slope one method considers other users' ratings about the predicted items and their evaluation of other items, but nothing about all items by all users in the whole system. The method for implementing Slope one algorithm is: firstly give a training set x, where x is the rating collection of users about items; then define user u rating about i and j, like respectively u_i and u_j. $S_{ij}(x)$ is a user set which rate together the item i and j; additionally, set $card(S)$ the number of elements in set S; finally we have the equation of average rating differential dev_{ij}, j between item i and j:

$$dev_{ij} = \sum_{u \in S_{ij}(x)} \frac{u_i - u_j}{card(S_{i,j}(x))} \tag{6}$$

3.2. M-Slope One Algorithm

Slope one algorithm owns brief and efficient pattern of operation. That is why it is often used to make real-time recommendations. However, its operational efficiency and precision are influenced because of the following shortcomings:

3.2.1. Number of related items to be rated.The rating prediction of item j is global. User u rating about item j was predicted on the basis of differences between other users' ratings about relevant items and j. Along with increasing number of related items to be rated, to compute scoring differences between item j and others will cost tremendously, which will not only affect the precision of recommendations but also limit the arithmetic speed of the algorithm.

3.2.2. Weak user relevance. User u predicted rating about item j involves all users who have rated item j. Some irrelevant users are filtered out by the way of computing here. But it is global for all users, which included users who meet requirements of rating items but have completely different interest, which will further cause bias to prediction results. As shown in

Table3, to predict $userC$'s rating about Item2, $userA$ and $userC$ hold the same preference, and their habit for rating items is similar; but $userB$ has totally different interest from $userA$ and $userC$, as well as the degree of love for items. According to the evaluation of the $userA$, $userC$ to predict j score is 5. According to the evaluation of $userA$ and $userB$ to predict j score is 1. Obviouslyit does not agree with our hypothesis. The outcome is not correct either.

Table3 The User Item Rating Scale

	$item_1$	$item_2$
$userA$	3	5
$userB$	5	1
$userC$	5	7

It is helpful to establish similar user groups to alleviate the problem mentioned above. Confidence level is enhanced through computing rating differences between items in users' neighboring group. Also the quantity of items to be calculated is considerably decreased to further improve the recommendation quality of the algorithm. The paper improves Slope one algorithm based on user interest model, which is called M-Slope one. With the use of user interest model to reckon similar user groups of target items, the method scales down data search scope for forecasting ratings of target items, develops the improved average differential equation for item scoring based on user similarities and enables similar users to contribute more to the computational work of average rating differences. The higher similarity to target users indicates the higher level of confidence by the user about scoring deviation. Otherwise, the level is lower. The algorithm demonstrates its ability to reduce the scope of items to be counted and improve computational reliability of related items and the recommendation quality of it.

.Improved average differential equation for item scoring based on users' similarities

$$dev_{ij} = \frac{\sum_{v \in S_{i,j}(x)} sim(u,v)(v_i - v_j)}{\sum_{v \in S_{i,j}(x)} sim(u,v)} \tag{7}$$

Where, $sim(u,v)$ is the similarity between user u and v.

.Recommendation process of M-Slope one algorithm:

Step1. Establish users' interest model and construct user-standard label matrix as seen in Table 2

Step2. Seek for users who have relevant items in similar user groups and utilize equation (7) to compute average different values of scoring.

Step3. Exploit equation (6) to calculate scores of predicted items and Top-N recommendation method to raise suggestions to users.

4. Experiment Design and Discussion

4.1. Experiment Data and Development Environment

For the experiment, we used MovieLens10M as test data and chose 1640 users with both labeling records and scoring records. All those users owned 50,000 tagged data and approximately 700,000 scoring data. The size of scoring data is between 0.5-5, at an interval of 0.5, 10 scores in all. To perform the validation test, the paper divided randomly those experimental data into 5 portions by user like Database1 to Database5 to verify reliability of the method. The experiment adopts Windows 7 operating system as the development environment and utilizes eclipse3.7 and SQL Server2008 to develop.

4.2. Method of Assessment

The evaluation of recommendation quality is an important part of the experiment. In the different applications of recommendation, the evaluating methods are varied.

4.2.1. Mean absolute error(MAE). Mean absolute error is used to measure the differences between predicted and actual scorings of the recommendation algorithm. The mean absolute error is smallest, meaning the highest accuracy of the method. It is the most commonly used method of rating by the existing recommendation systems. The computational formula is:

$$MAE = \frac{\sum_{i \in U, j \in I} | p_{ij} - r_{ij} |}{n} \qquad (8)$$

Where, p_{ij} is prediction score of the user i to j, r_{ij} is the actual score of user i to j. U is the set of users, I is collection for the project. n is the total score of the project.

4.2.2. Root Mean Squared Error(RMSE). Root Mean Squared Error is used to measure the size of mean errors. The smallest the value of such error is, the highest accuracy of the algorithm and recommendation have. The calculation formula is:

$$RMSE = \sqrt{\frac{\sum_{i \in U, j \in I} (p_{ij} - r_{ij})^2}{n}} \qquad (9)$$

4.2.3. Recall. Recall is the proportion of items which are accurately recommended to all testing sets. Set user i testing set T_i and the correctly recommended item set P_i. Recall ratio can be obtained through the expression:

$$\mathrm{Re}call = \frac{1}{n} \sum_{i=1}^{n} \frac{|T_i \cap P_i|}{|T_i|} \qquad (10)$$

4.2.4. Precision. Precision is used to estimate the percentage of right items recommended by users in TOP-N, which can be acquired through this:

$$\mathrm{Pr}ecision = \frac{1}{n} \sum_{i=1}^{n} \frac{|T_i \cap P_i|}{N} \qquad (11)$$

Where n is the total number of users, N is the number of recommended items.

4.2.5. F-measure .F-measure Recall and precision ratio is contradictory to some extent. Higher recall rate may suggest lower precision. To strike a balance between them, composite evaluation indicator F-measure is frequently adopted in the form of:

$$F_{measure} = \frac{2\,\mathrm{Pr}ecision * \mathrm{Re}call}{\mathrm{Pr}ecision + \mathrm{Re}call} = \frac{2}{1/\mathrm{Re}call + 1/\mathrm{Pr}ecision} \qquad (12)$$

4.3. Validation of Real-time User Interest Model and the Update Model

The part has two experiments, respectively validation test of real-time users' interest model and effectiveness test of the update model. We'll divide test data of both experiments into 8:2 as per the order of time where one item is rated, with the former 80% as training set and the remaining 20% as testing set.

4.3.1. Validation of real-time users' interest model. Through comparison of recall ratio by the traditional collaborative filtering recommendation method (CF) and the one based on real-time users' interest model (M-CF), the model's performance was verified. In the experiment, CF method used average ratings of items by users in the training set to fill up user-item scoring matrix, got similar degree between users through the modified cosine-similarity formula (1) and calculated scores of items to be predicted through formula (3). Finally it chose items of which the predicted scores were above 3 (recommended threshold) and recommended to users. M-CF algorithm utilized standard labeling information to which the training set corresponds to create user interest model, got similar degree between users through cosine-similarity formula (2) and obtained the interest degree of users towards items to be predicted through the formula (4). In the end, it recommended items of which the degree was over 0.5 (recommended threshold) to users. The experiment used recall ratio to make confirmation. In collaborative filtering recommendation algorithm, the number K of nearby users had effects on the precision of recommendations. If K is too small, item set cannot be adequately obtained; if K is too big, the searching cost will increase. For different K, the recall ratio is different, and also the accuracy of the algorithm. Experimental findings were listed in Table4.

Table 4. Comparison of CF and M-CF Recall Rate

Number of neighbors	Database1		Database2		Database3		Database4		Database5	
	CF	M-CF	CF	M-CF	CF	M-CF	CF	M-CF	CF	M-CF
K=5	0.23	0.11	0.25	0.14	0.24	0.15	0.22	0.21	0.23	0.10
K=10	0.34	0.24	0.24	0.24	0.36	0.28	0.33	0.31	0.34	0.18
K=15	0.41	0.35	0.34	0.34	0.40	0.37	0.38	0.42	0.40	0.26
K=20	0.47	0.42	0.39	0.39	0.46	0.40	0.45	0.51	0.46	0.36
K=25	0.50	0.48	0.46	0.46	0.50	0.47	0.49	0.54	0.50	0.39
K=30	0.54	0.53	0.55	0.51	0.57	0.54	0.54	0.55	0.55	0.49
K=35	0.57	0.58	0.58	0.54	0.59	0.55	0.58	0.63	0.57	0.54
K=40	0.59	0.62	0.59	0.57	0.60	0.57	0.59	0.68	0.59	0.61
K=45	0.63	0.65	0.61	0.59	0.61	0.60	0.60	0.70	0.63	0.68
K=50	0.64	0.65	0.63	0.61	0.64	0.61	0.63	0.74	0.64	0.71
K=55	0.67	0.67	0.66	0.62	0.69	0.63	0.66	0.77	0.67	0.76
K=60	0.68	0.71	0.68	0.64	0.69	0.64	0.68	0.78	0.68	0.81
K=65	0.70	0.74	0.69	0.65	0.69	0.65	0.70	0.81	0.70	0.81
K=70	0.71	0.75	0.70	0.67	0.71	0.66	0.70	0.81	0.70	0.81
K=75	0.72	0.77	0.71	0.69	0.71	0.67	0.71	0.82	0.71	0.83
K=80	0.72	0.78	0.71	0.70	0.71	0.68	0.72	0.85	0.72	0.84
K=85	0.72	0.80	0.72	0.71	0.72	0.68	0.72	0.86	0.72	0.84
K=90	0.73	0.81	0.72	0.72	0.72	0.69	0.73	0.87	0.72	0.84
K=95	0.73	0.82	0.72	0.73	0.73	0.71	0.73	0.87	0.73	0.85
K=100	0.73	0.82	0.72	0.74	0.73	0.72	0.73	0.87	0.73	0.85

In the table4, the data of recall average represent a graph, as shown in figure1.

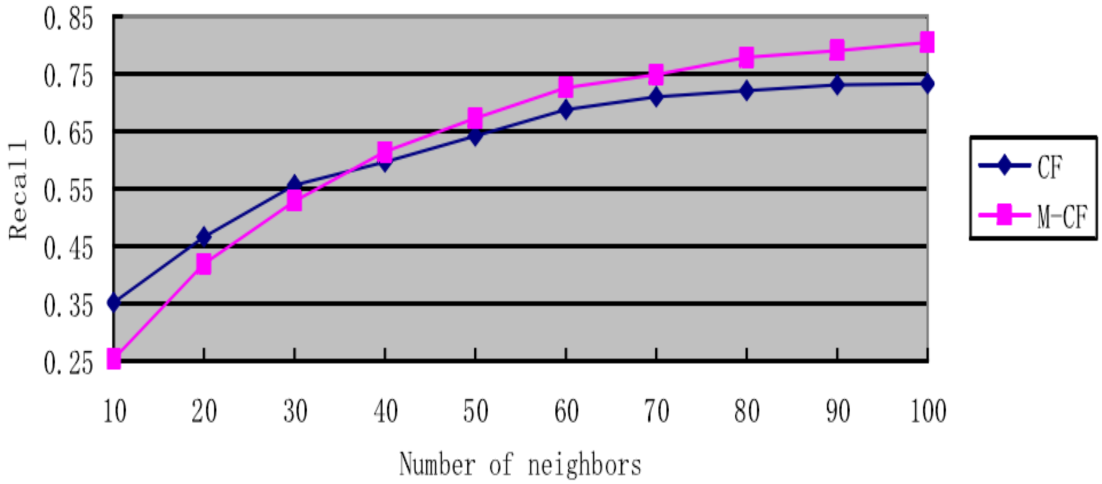

Figure 1. Recall Effect of Different user Interest Model

4.3.2. Validation of the Update Model. In this part we validated the update users' interest model, which combines with the collaborative filtering recommendation algorithm to form UM-CF. The experiment was designed to ensure that after comparison of recall ratio with M-CF algorithm, the update user interest model expressed users' interest degree more accurately. The experiment selected users whose labeled items were more than 10 in the training set as test objects for division of update data set. The same experiment was performed on M-CF method. For UM-CF, the former 90% data in the training set was regarded as training set of the model according to the labeled time of each item, while the rest 10% as update data set of it. We did something like that in order to avoid too big update data set and the loss of items in the model. The way of modeling was not different from M-CF. The time of last data in the model's training set was used as the point for creating the model and the time of last data in the update training set as the current point of time. Both M-CF and UM-CFadoptedto Top-N recommendation and used precision ratio to do tests. Along with the change of N, the precision varied. See Figure 2 for testing rests.

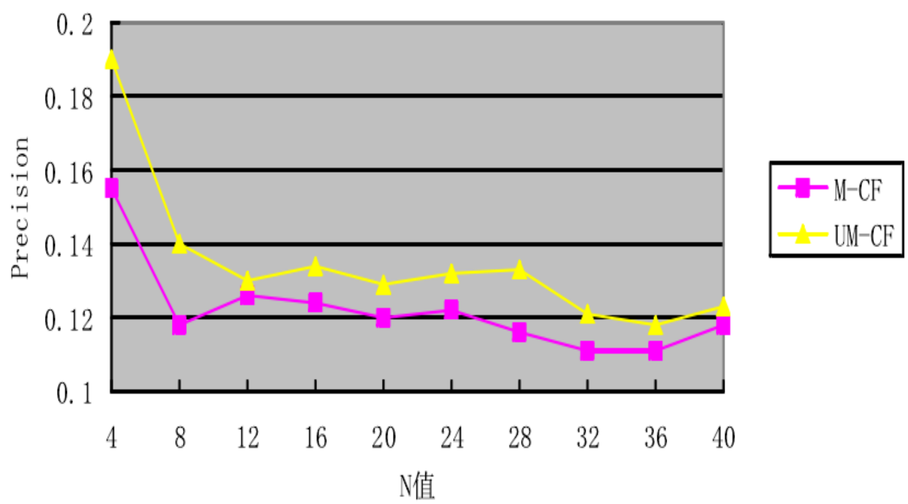

Figure 2. Precision Effectof Update User Interest Model

4.4. Experimental Validation of M-Slope One Algorithm

The precision and real-time capability of M-Slope one algorithm was confirmed in the part. Users'labeling information was used to create users' interest model and produce nearby user groups. Similarities between users were employed to improve average scoring differential equation for items. Finally suggestions were made. Performance of the algorithm was evaluated through MAE. After comparing the precision of traditional Slope one recommendation method with M-Slope one, we compared their time/space complexity and analyzed the running efficiency.

For testing data, we still divided it at the same way by 4:3, first 80% as training set and the rest 20% as testing set. As the classic Slope one algorithm didnot require choosing close users, the number of those users didn't affect the accuracy of the algorithm. The average value of MAE for five data sets is 0.753. For the improved Slope one recommendation algorithm based on users' interest model, the average value of items' MAE was varied from different adjacent user groups, as Figure 3.

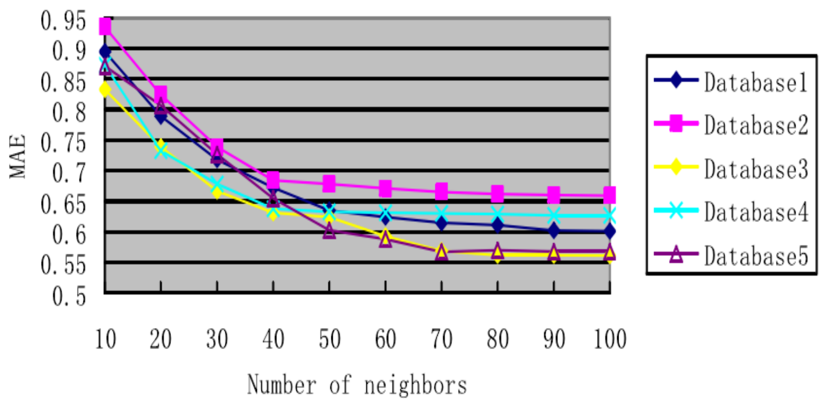

Figure 3. MAE value of MAE-Slope one Recommendation Algorithm

5. Conclusion

Abandon real-time of the user- item rating matrix collaborative filtering algorithm based on users' interest model, using which that based on label real-time users' interest model to calculate the similarity between users. And using user interest model linear regression equation to calculate interest degree of users' target project, Recommended by Top-N method which make recommendations for the user. The purpose of this algorithm is proposed reflect the role of improving the quality of the algorithm recommended which based on label users' interest model. Experiments show that users' interest model based on standard label can improve the recommendation quality of the algorithms.

References

[1] Z. Liang, H. Naijing and Z. Shouzhi,"Research and development of personalized recommendation algorithm design", computer, vol. 8, **(2002)**, pp. 986-991.
[2] S. Ying, L. Jianguo, G. Qiang and T. Dagang,"Considering the recommendation algorithm", operations research and the management of negative evaluation of personalized, vol. 6, **(2012)**, pp. 17-22.
[3] L. Hui, H. Yun, L. Cunhua and W. Xia,"Computer engineering and Application Research on the algorithm",nearest neighbor relationship of personalized recommendation, vol. 36, **(2012)**, pp. 205-209.
[4] X.Juqin, J. Yan and S. Shaolong,"Computer engineering and application integrated user preference model and BP neural network for personalized recommendation", vol. 2, **(2013)**, pp. 57-60.

[5] C. Airu and T. Xiaochun,"Recommendation algorithm computer and modern Web people collaboration platform based personalization", vol. 3, **(2013)**, pp. 50-53.
[6] G.Qingju, Z. Rangming and M. Juntao,"The recommendation algorithm research", software learning interest based personalized, vol. 9, **(2013)**, pp. 51-53.
[7] L.Jianguo, Z. Tao and W.Binghong,"Research progress: progress in natural science", personalized recommendation system, vol. 1, **(2009)**, pp. 1-15.
[8] Y.Huaizhen, C. Xiaoqi and L. Meilian,"Research on Recommendation Algorithm", computer engineering and science time weighted based personalization, vol. 6,**(2009)**, pp. 126-128.
[9] G.Yanhong and D. Guishi,"Algorithm. Application Research of computers to recommend a personalized collaborative filtering", **(2008)**, vol. 1, pp. 39-41.
[10] W. Jie and T.Xiaochun,"Algorithm Research", Application Research of computers community network content based personalized recommendation, vol. 4, **(2011)**, pp. 1248-1250.

Author

Liu Lijun is a lecturer at Department of Logistics Information, Hunan Vocational College of Modern Logistics. He is in the research of computer network, Internet of Things technology and application.

