ISSN : 0974 - 7435 Volume 10 Issue 13

2014

BioTechnology

An Indian Journal

FULL PAPER

BTAIJ, 10(13), 2014 [7087-7092]

Recommendation algorithm based on collaborative filtering under social network environment

Caixia Lu

Huaian College Of Information Technology, Huai'an City Jiangsu Province 223003, (CHINA)

ABSTRACT

A collaborative recommendation algorithm based on community found is proposed in this paper. The idea of community is introduced in collaborative recommendation, at the same time the scoring pre-treatment is joined, in order to solve some problems in the traditional collaborative filtering recommendation. Results demonstrate that algorithm proposed in this paper is better than the algorithms based on the Pearson similarity and Cosine similarity.

KEYWORDS

Collaborative filtering; Social network environment; Recommendation algorithm.

© Trade Science Inc.

INTRODUCTION

With the wide popularity of Internet, traditional trade and business' activities have been revolutionarily changed by e-commerce, which also causes the reform from model of commodity-oriented to customer-oriented[1,2]. It is a trend for enterprises to provide personal service to meet the different needs for different people. E-commerce recommendation system appears in such a situation, which can effectively retain customers, increase enterprise sales, improve service quality and enhance the competitiveness of enterprises[3,4].

Recommendation system has a wide range of applications and good development prospect in the area of electronic commerce. It gradually becomes an important part of e-commerce and attracts a large number of researchers' attention[5]. Nowadays, there are three main recommendation technologies in the recommendation system, including content-based, collaborative filtering and hybrid recommendation technology[6]. Collaborative filtering is one of the most successful applications of recommendation technology. However, due to data sparseness and cold start issues of collaborative filtering and the growing of data scale in the E-commerce, e-commerce recommendation system faces many challenges[7].

In the next section, we introduce basic idea of collaborative filtering recommendation algorithm based on community discovery[8]. In Section 3, implementation of the proposed algorithm is given. In Section 4, the experiment data analysis and performance evaluation is given. In Section 5 we conclude the paper and give some remarks.

BASIC IDEA OF COLLABORATIVE FILTERING RECOMMENDATION ALGORITHM BASED ON COMMUNITY DISCOVERY

Collaborative filtering recommendation needs multiple users to score the project, and the system make recommendations based on the users and the score information. If it lacks of data support, collaborative filtering recommendation system becomes no root, it is difficult to recommend accurately. From the perspective of the realization of the recommendation system, if a large amount of information exists in the system, the processing speed of collaborative recommendation based on memory becomes slow, and it is difficult to realize real-time recommendation. Recommendation system based on model also exists question of overhead.

This paper proposes a collaborative recommendation algorithm based on community found to solve the problems existing in the traditional collaborative filtering recommendation system. Collaborative recommendation algorithm based on community found firstly establishes user network for offline data. Based on the user's network, the user is divided into several classes. Secondly choose a number of classes as the candidate set of users from known user categories. Thirdly, calculate all similarity of the users and the candidate user sets, and select the first k users as a neighbor. Finally recommendation is carried out based on the k number of neighbor users.

Collaborative recommendation algorithm based on community found is a combination of memory and model, and it reduces the busy time of the system by classifying the offline data. When there is a new user, a new project or new score entering into the system, the system will choose free time for reclassification to avoid occupying more resources at busy time in the system, which is advantageous for the system to make real-time recommendation. Next is the process of choosing neighborhood users for target users roughly. Calculate the similarity between target users and each class, and choose several classes nearest to the user as candidate set. And then select the k number of users nearest to the target user. After roughing selection, the system calculation time is reduced. Even if there is a lot of information, the recommender system also can make recommendation. In addition, when calculating the similarity of users, also the score pre-treatment is joined, which reduces the drawbacks of the recommendation algorithm brought by data sparseness.

IMPLEMENTATION OF THE PROPOSED ALGORITHM

Users-project relation is shown in Figure 1. The score is a integer belonging to[1,5]. Supposing there are m number of users and n number of projects. The relation matrix is $R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & & & \cdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}$.

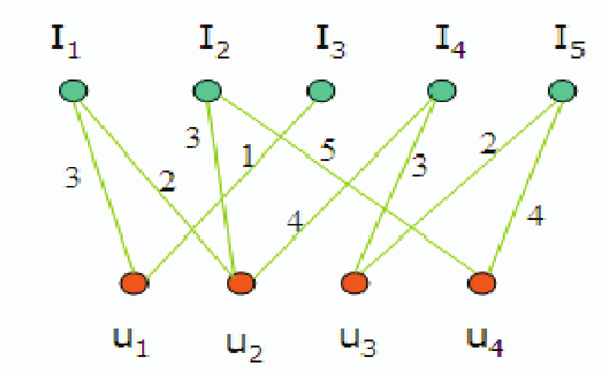

Figure 1 : Users-project

r_{ij} represents the score of user i to project j. If user does not score the project, the score is 0. The similarity between user x and user y is

$$S(x,y)=\frac{\sum_{s\in T_{xy}}(r_{x,s}-\overline{r_x})(r_{y,s}-\overline{r_y})}{\sqrt{\sum_{s\in T_{xy}}(r_{x,s}-\overline{r_x})^2\sum_{s\in T_{xy}}(r_{y,s}-\overline{r_y})^2}}.$$

The relation matrix of user-user is $S_u=\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1m} \\ S_{21} & S_{22} & \cdots & S_{2m} \\ \cdots & & & \cdots \\ S_{m1} & S_{m2} & \cdots & S_{mm} \end{bmatrix}$. The similarity of user i and user j is represented by S_{ij}.

After establishing the user's network, use the community discovery algorithm. Overlapping community discovery algorithm based on the center node is used to divide network. If there is a new user or project joining the system, using community discovery algorithm can realize the user reclassification. Then set up candidate user set.

In order to reduce the computational time of recommender system, this algorithm when choosing neighborhood users of target users, first has carried out the roughing process. Choose L number of classes, which is nearest to target user. All users of the L classes are taken as candidate user set U of the target user. Supposing there are t number of user class $\{C_1,C_2,\ldots,C_t\}$. Centroid vector of class C_i is $\overline{c_i}=(\overline{r_{c_i,1}},\overline{r_{c_i,2}},\ldots\overline{r_{c_i,n}})$. $\overline{r_{c_i,j}}$ represents average value of all users to project j in the class C_i. The similarity between target user u_a and class C_i is

$$S(u_a, C_i) = \frac{\sum_{s \in T_{u_a} \wedge T_{C_i}} (r_{u_a,s} - \overline{r_{u_a}})(\overline{r_{c_i,s}} - \overline{r_{c_i}})}{\sqrt{\sum_{s \in T_{u_a} \wedge T_{C_i}} (r_{u_a,s} - \overline{r_{u_a}})^2 (\overline{r_{c_i,s}} - \overline{r_{c_i}})^2}} .$$

T_{u_a} represents score item set of user u_a, T_{C_i} represents score item set of user in the class C_i. s represents the items that user u_a and class C_i score at the same time. $r_{u_a,s}$ represents score of user u_a to project s. $\overline{r_{u_a}}$ represents average value of scores of user u_a. $\overline{r_{c_i}}$ represents the average value of sum of each component for vector $\overline{c_i}$. In order to get the neighborhood user of user u_a, process of score pretreatment is added. In fact, user u can score for item i. The score can be expressed as

$$r_{u,i} = \begin{cases} r_{u,i}, & scored \\ \tilde{r}_{u,i}, & otherwise \end{cases}$$

$$\hat{r}_{u,i} = \bar{r}_u + \Delta r_{C_u}(i)$$

$$\Delta r_{C_u}(i) = \sum_{u' \in C_u(i)} (r_{u',i} - \overline{r_{u'}}) / \left| c_u(i) \right|$$

$\hat{r}_{u,i}$ is obtained by average score of u and score deviation of user to item i in the class C_u. Then calculate the similarity between target user u_a and user u.

$$S(u_a, u) = \frac{\sum_{s \in T(u_a)} (r_{u,s} - \overline{r_u})(r_{u_a,s} - \overline{r_{u_a}})}{\sqrt{\sum_{s \in T(u_a)} (r_{u,s} - \overline{r_u})^2 \sum_{s \in T(u_a)} (r_{u_a,s} - \overline{r_{u_a}})^2}}$$

$T(u_a)$ represents scored item set of user u_a. If user u does not score the item s, the pretreatment value is taken as scored value. The prediction score of item t is

$$R_{u_a}(t) = \overline{r_{u_a}} + \frac{\sum_{u \in C_i} S(u_a, u)(r_u(t) - \overline{r_u})}{\sum_{u \in C_i} S(u_a, u)}$$

C_i represents k number of neighborhood user set of user u_a.

THE EXPERIMENT DATA ANALYSIS AND PERFORMANCE EVALUATION

MovieLens data set is used to test the performance of collaborative recommendation algorithm based on community found in this paper, and compare with the traditional Pearson correlation and cosine similarity calculation algorithm. When K is 20, change the size of training set. Experiment result of different training set proportion is shown in TABLE 1. In TABLE 1, A represents the proposed algorithm, B represents collaborative filtering based on Pearson and C represents collaborative filtering based on cosin.

TABLE 1 : Experiment result of different training set proportion

Proportion of test set	Mean absolute error			Root mean square error		
	A	B	C	A	B	C
20%	0.8034	0.8477	0.8417	1.0038	1.034	1.0312
30%	0.8020	0.8426	0.8273	0.9712	1.012	0.9902
40%	0.7955	0.8376	0.8170	0.9546	0.9942	0.9846
50%	0.7812	0.8375	0.8025	0.9438	0.9648	0.9644
60%	0.7713	0.8350	0.7989	0.8923	0.9441	0.9268
70%	0.7703	0.8301	0.7882	0.8345	0.9021	0.8722
80%	0.7698	0.8234	0.7773	0.7906	0.8454	0.8223

Comparison of mean absolute error of three different algorithms is shown in Figure 2. The blue line represents the proposed algorithm, the red line represents collaborative filtering based on cosin and the green line represents collaborative filtering based on Pearson. The mean absolute error of proposed algorithm is less than the other two algorithms.

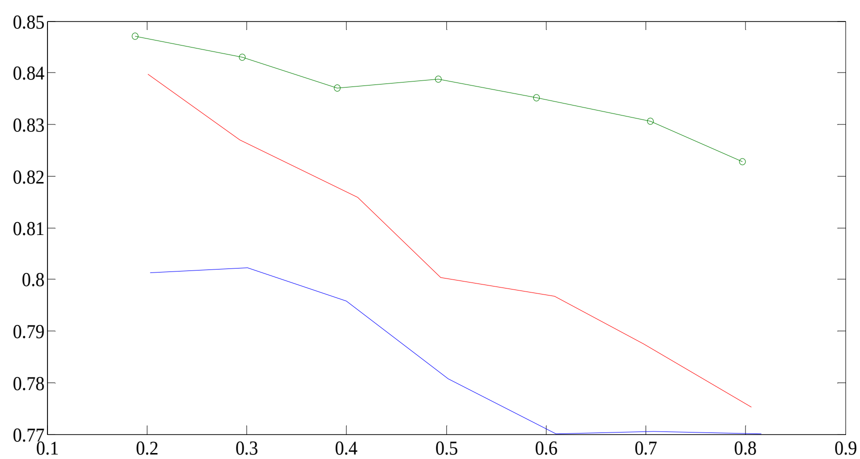

Figure 2 : Comparison of mean absolute error

CONCLUSION

With the full understanding of the principles and problems of the collaborative filtering recommendation, further useful exploration and research are made in this paper. We propose a collaborative filtering recommendation method based on community detection, which brings methods of community detecting into collaborative filtering. We select a part of user communities from the user network projected by user-item network as the candidate neighboring user set for the target user, thereby reducing calculation time and increasing recommendation speed of recommendation system. Finally, in order to make up for defect of few rating information, we add pre-rating mechanism into collaborative filtering to solve the problems arising from data sparseness, rasing precision of recommendation system. This paper has a perfect combination of social network technology and collaborative filtering technology, which can greatly increase recommendation system performance.

We carry out experiments by MovieLens data set to test two performance indexes which including MAE and RMSE. Results demonstrate that algorithm proposed in this paper is better than the algorithms based on the Pearson similarity and Cosine similarity.

REFERENCES

[1] G.Karypis; Evaluation of Item-based top-n Recommendation Algorithms, the Tenth International Conference on information and Knowledge Management (CIKM), **(2001)**.

[2] M.E.J.Newman, M.Girvan, Finding and Evaluating Community Structure in Networks [J], Phys Rev E, **69(2)**, 026113 **(2004)**.

[3] N.Manouselis, C.Costopoulou; Experimental analysis of design choices in multiat tribute utility collaborative filtering, International Journal of Pattern Recognition and Artificial Intelligence, **21(2)**, 311-331 **(2007)**.

[4] Hihua Zhang, RuiSheng Wang, XiangSun Zhang; Identification of overlapping community structure in complex networks using fuzzy c-means clustering. Physica A, **374**, 483-490 **(2007)**.

[5] DuanBing Chen, YanFu, MingSheng Shang; An efficient algorithm for overlapping community detection in complex. GlobalCongress on Intelligent Systems, 244-247 **(2009)**.

[6] M.S.Shang, C.H.Jin, T.Zhou, Y.C.Zhang; Collaborative filtering based on multi-channe difusion[J], Physica A, **388(23)**, 4867-4871 **(2009)**.

[7] G.Adomavicius, A.Tuzhilin; Toward the Next Generation of Recommender Systems:A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering, **17(6)**, 734-749 **(2005)**.

[8] F.Wu, B.A.Huberman; Finding communities in linear time: a physics approach. European Physical Journal B, **38(2)**, 331-338 **(2004)**.