Item-Based Collaborative Filtering Recommendation using Self-Organizing Map

SongJie Gong[1], HongWu Ye[2], XiaoMing Zhu[1]

1. Zhejiang Business Technology Institute, Ningbo 315012, P. R. China
E-mail: zhuxiaomingzjbti@163.com

2. Zhejiang Textile & Fashion College, Ningbo 315211, P. R. China
E-mail: yehongwuzjbti@163.com

Abstract: Recommender systems can help people to find interesting things and they are widely used in Electronic Commerce. Collaborative filtering technique has been proved to be one of the most successful techniques in recommender systems. The main problems of collaborative filtering are about prediction accuracy, response time, data sparsity and scalability. To solve some of these problems, this paper presented an item-based collaborative filtering recommendation algorithm using self-organizing map. Firstly, it employs clustering function of self-organizing map to form nearest neighbors of the target item. Then, it produces prediction of the target user to the target item using item-based collaborative filtering. The item-based collaborative filtering recommendation algorithm using self-organizing map can efficiently improve the scalability and promise to make recommendations more accurately than conventional collaborative filtering.

Key Words: Collaborative Filtering, Recommender System, Self-organizing Map

1 INTRODUCTION

In many times in our everyday life, we become active seekers or passive receivers of information in order to make decisions. However, our experiences and knowledge often do not enough to deal with the vast amount of available information. Thus, methods to help find resources of interest have attracted much attention from both researchers and vendors. To deal with the problem, the personalized recommendation systems play a more important role and collaborative filtering has proved to be one of the most effective for its simplicity in both theory and implementation [1,2].

Collaborative filtering (CF) has been successfully used in various applications. The famous electronic commerce website Amazon and CD-Now have employed CF technique to recommend products to customers and it has improved quality and efficiency of their services. The CF assumes that a good way to find a certain user's interesting content is to find other people who have similar interests with him[3,4,5]. CF methods operate upon user ratings on observed items making predictions concerning users' interest on unobserved items. The sparsity of ratings problem is particularly important in domains with large or continuously updated list of items as well as a large number of users. The sparsity problem may occur when either none or few ratings are available for the target user, or for the target item that prediction refers to, or for the entire database in average [6,7]. Different treatments are required and different prediction techniques must be employed depending on the sparsity conditions, making the selection of an appropriate approach a cumbersome task. Current CF approaches are limited in the sense that they address specific aspects of the above problem.

To improve the recommendation performance, in this paper, we propose an item-based collaborative filtering algorithm using self-organizing map. Firstly, it employs clustering function of self-organizing map to form nearest neighbors of the target item. Then, it produces prediction of the target user to the target item using item-based collaborative filtering. Furthermore, the experimental results show that this method can increase the accuracy of the predicted values, resulting in improving recommendation quality of the collaborative filtering recommender system.

2 USING SELF-ORGANIZING MAP TO FORM NEAREST NEIGHBORS

2.1 Self-organizing map(SOM)

The self-organizing map network was first proposed in 1981 by Finland scholar Kohonen. As an unsupervised learning neural network model, SOM has been applied widely in many fields since it was brought forward. The network structure of SOM consists of two layers, the upper one is output layer and the down one is input layer. The number of neuron nods which are responsible for acquiring data in the input layer is consistent with the number of variables. A one dimensional or two dimensional network is formed in the output layer, and the network could ensure the domain relationships. The SOM network is a whole-connecting structure network, every neuron nod in the input layer connects with all the neuron nods in the output layer. When Euclidean Distance between the input vectors and certain weights, which link neuron nods in the input layer with ones in output layer, is minimum, the neuron nod corresponding with these weights is activated and acts as the output of the network. At the same time, the coefficients of connection weight are amended and become more contiguous with the input vectors, the output neuron is

A Project Supported by Scientific Research Fund of Zhejiang Provincial Education Department (Grant No. Y200806038).

also called winning neuron, and the corresponding coefficients of connecting weight are amended until the termination limitations are satisfied [8].

The Self-organizing map training algorithm proposed by Kohonen is summarized as follows [9].

Step 1. Initialization: Choose random values for the initial weights wj (0).

Step 2. Winner Finding: Find the winning neuron j* at time k, using the minimum-distance Euclidean criterion

$$j^* = \arg\min_j \| x(k) - w_j \|, j = 1, \ldots, N^2 \quad (1)$$

where x(k) = [x1(k), · · · , xn(k)] represents the kth input pattern, N^2 is the total number of neurons, and ‖ · ‖ indicates the Euclidean norm.

Step 3. Weights Updating: Adjust the weights of the winner and its neighbors, using the following rule:

$$w_j(k+1) = w_j(k) + p(k) N_{j^*}(k)(x(k) - w_j(k)) \quad (2)$$

where p(k) is a positive constant and $N_{j^*}(k)$ is the topological neighborhood function of the winner neuron at time k. It should be emphasized that the success of the map formation is critically dependent on how the values of the main parameters (i.e., p(k) and $N_{j^*}(k)$), initial values of weight vectors, and the number of iterations are prespecified.

2.2 Using SOM to cluster items

For the item-based collaborative filtering recommender system model, we focus on the clustering capability of SOM. For the collaborative filtering algorithm, the forming of neighbors is an important step. We consider the excellence of the clustering function of the self-organizing map, and firstly form the target item nearest neighbors using this method, as the figure 1 shows.

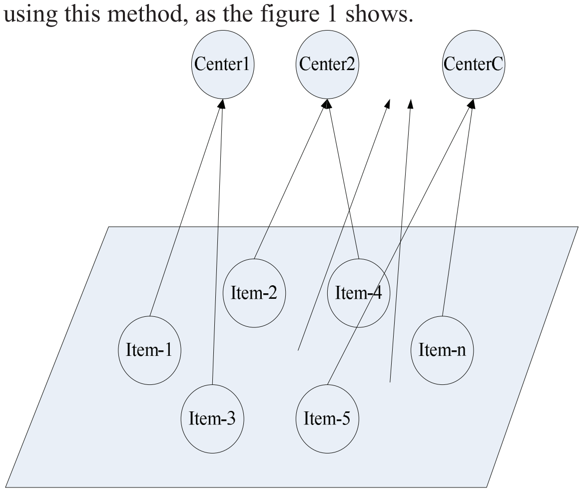

Fig 1. Employing SOM to cluster items

3 PRODUCING THE PREDICTION

3.1 Measuring the item rating similarity

There are several similarity algorithms that have been used [10,11]: Pearson correlation, cosine vector similarity, adjusted cosine vector similarity, mean-squared difference and Spearman correlation.

Pearson's correlation, as following formula, measures the linear correlation between two vectors of ratings as the target item t and the remaining item r.

$$sim(t,r) = \frac{\sum_{i=1}^{m} (R_{it} - A_t)(R_{ir} - A_r)}{\sqrt{\sum_{i=1}^{m} (R_{it} - A_t)^2 \sum_{i=1}^{m} (R_{ir} - A_r)^2}} \quad (3)$$

Where R_{it} is the rating of the target item t by user i, R_{ir} is the rating of the remaining item r by user i, A_t is the average rating of the target item t for all the co-rated users, A_r is the average rating of the remaining item r for all the co-rated users, and m is the number of all rating users to the item t and item r.

The cosine measure, as following formula, looks at the angle between two vectors of ratings as the target item t and the remaining item r.

$$sim(t,r) = \frac{\sum_{i=1}^{m} R_{it} R_{ir}}{\sqrt{\sum_{i=1}^{m} R_{it}^2 \sum_{i=1}^{m} R_{ir}^2}} \quad (4)$$

Where R_{it} is the rating of the target item t by user i, R_{ir} is the rating of the remaining item r by user i, and m is the number of all rating users to the item t and item r.

The adjusted cosine, as following formula, is used for similarity among items where the difference in each user's use of the rating scale is taken into account.

$$sim(t,r) = \frac{\sum_{i=1}^{m} (R_{it} - A_i)(R_{ir} - A_i)}{\sqrt{\sum_{i=1}^{m} (R_{it} - A_i)^2 \sum_{i=1}^{m} (R_{ir} - A_i)^2}} \quad (5)$$

Where R_{it} is the rating of the target item t by user i, R_{ir} is the rating of the remaining item r by user i, A_i is the average rating of user i for all the co-rated items, and m is the number of all rating users to the item t and item r.

3.2 Prediction using item-based CF

Since we have got the membership of item, we can calculate the weighted average of neighbors' ratings, weighted by their similarity to the target item.

The rating of the target user u to the target item t is as following:

$$P_{ut} = \frac{\sum_{i=1}^{c} R_{ui} \times sim(t,i)}{\sum_{i=1}^{c} sim(t,i)} \quad (6)$$

Where Rui is the rating of the target user u to the neighbour item i, sim(t, i) is the similarity of the target item t and the neighbour item i, and c is the number of the neighbours.

4 DATASET AND MEASUREMENT

4.1 Data set

We use MovieLens collaborative filtering data set to evaluate the performance of proposed algorithm. MovieLens data sets were collected by the GroupLens

Research Project at the University of Minnesota and MovieLens is a web-based research recommender system that debuted in Fall 1997. Each week hundreds of users visit MovieLens to rate and receive recommendations for movies [5]. The site now has over 45000 users who have expressed opinions on 6600 different movies. We randomly selected enough users to obtain 100, 000 ratings from 1000 users on 1680 movies with every user having at least 20 ratings and simple demographic information for the users is included. The ratings are on a numeric five-point scale with 1 and 2 representing negative ratings, 4 and 5 representing positive ratings, and 3 indicating ambivalence.

4.2 Performance measurement

Several metrics have been proposed for assessing the accuracy of collaborative filtering methods. They are divided into two main categories: statistical accuracy metrics and decision-support accuracy metrics. In this paper, we use the statistical accuracy metrics [12].

Statistical accuracy metrics evaluate the accuracy of a prediction algorithm by comparing the numerical deviation of the predicted ratings from the respective actual user ratings. Some of them frequently used are mean absolute error (MAE), root mean squared error (RMSE) and correlation between ratings and predictions. All of the above metrics were computed on result data and generally provided the same conclusions. As statistical accuracy measure, mean absolute error (MAE) is employed.

Formally, if n is the number of actual ratings in an item set, then MAE is defined as the average absolute difference between the n pairs. Assume that p1, p2, p3, ..., pn is the prediction of users' ratings, and the corresponding real ratings data set of users is q1, q2, q3, ..., qn. See the MAE definition as following:

$$MAE = \frac{\sum_{i=1}^{n} | p_i - q_i |}{n} \tag{7}$$

The lower the MAE, the more accurate the predictions would be, allowing for better recommendations to be formulated. MAE has been computed for different prediction algorithms and for different levels of sparsity.

5 CONCLUSIONS

Recommender systems can help people to find interesting things and they are widely used in Electronic Commerce. Collaborative filtering technique has been proved to be one of the most successful techniques in recommender systems. The main problems of collaborative filtering are about prediction accuracy, response time, data sparsity and scalability. To solve some of these problems, this paper presented an item-based collaborative filtering recommendation algorithm using self-organizing map. Firstly, it employs clustering function of self-organizing map to form nearest neighbors of the target item. Then, it produces prediction of the target user to the target item using item-based collaborative filtering. The presented algorithm can efficiently improve the scalability and promise to make recommendations more accurately than conventional collaborative filtering.

REFERENCES

[1] Jong-Seok Lee, Chi-Hyuck Jun, Jaewook Lee, Sooyoung Kim, Classification-based collaborative filtering using market basket data, Expert Systems with Applications 29 (2005) 700–704.

[2] Hyung Jun Ahn, A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem, Information Sciences 178 (2008) 37-51.

[3] George Lekakos, George M. Giaglis, Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors, Interacting with Computers 18 (2006) 410–431.

[4] Manos Papagelis, Dimitris Plexousakis, Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents, Engineering Application of Artificial Intelligence 18 (2005) 781-789.

[5] Sarwar B, Karypis G, Konstan J, Riedl J. Item-Based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International World Wide Web Conference. 2001. 285-295.

[6] Chong-Ben Huang, Song-Jie Gong, Employing rough set theory to alleviate the sparsity issue in recommender system, In: Proceeding of the Seventh International Conference on Machine Learning and Cybernetics (ICMLC2008), IEEE Press, 2008, pp.1610-1614.

[7] Yu Li, Liu Lu, Li Xuefeng, A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce, Expert Systems with Applications 28 (2005) 67–77.

[8] ZHOU ShaoHua, FU Lue, LIANG BaoLiu, Clustering analysis of ancient celadon based on SOM neural network, Science in China Series E: Technological Sciences, 2008, 51(7):999-1007.

[9] WANG Ling, MU Zhi-Chun,GUO Hui, Combining Self-organizing Feature Map with Support Vector Regression Based on Expert System,ACTA AUTOMATICA SINICA, 2005, 31(4):612-619.

[10] Songjie Gong, Chongben Huang, Employing Fuzzy Clustering to Alleviate the Sparsity Issue in Collaborative Filtering Recommendation Algorithms, In: Proceeding of 2008 International Pre-Olympic Congress on Computer Science, World Academic Press, 2008, pp.449-454.

[11] SongJie Gong, The Collaborative Filtering Recommendation Based on Similar-Priority and Fuzzy Clustering, In: Proceeding of 2008 Workshop on Power Electronics and Intelligent Transportation System (PEITS2008), IEEE Computer Society Press, 2008, pp. 248-251.

[12] Huang qin-hua, Ouyang wei-min, Fuzzy collaborative filtering with multiple agents, Journal of Shanghai University (English Edition), 2007,11(3):290-295.