A Collaborative Filtering Recommendation Algorithm Based on SVD Smoothing

YiBo Ren
Zhejiang Business Technology Institute,
Ningbo 315012, P. R. China
e-mail: ryb@zjbti.net.cn

SongJie Gong
Zhejiang Business Technology Institute,
Ningbo 315012, P. R. China
e-mail: eigsj@sina.com

***Abstract*—Recommender system is one of the most important technologies in electronic commerce. And the collaborative filtering is almost the popular approach used in the recommender systems. With the development of electronic commerce systems, the magnitudes of users and items grow rapidly, resulted in the extreme sparsity of user rating data set. Traditional similarity measure methods work poor in this situation, make the quality of recommendation system decreased dramatically. Sparsity of users' ratings is the major reason causing the poor quality. To address this issue, a collaborative filtering recommendation algorithm based on singular value decomposition (SVD) smoothing is presented. This approach predicts item ratings that users have not rated by the employ of SVD technology, and then uses Pearson correlation similarity measurement to find the target users' neighbors, lastly produces the recommendations. The collaborative filtering recommendation algorithm based on SVD smoothing can alleviate the sparsity problems of the user item rating dataset, and can provide better recommendation than traditional collaborative filtering algorithms.**

Keywords-recommender system; collaborative filtering; singular value decomposition; sparsity

I. INTRODUCTION

While the rapid growth and wide application of the Internet and electronic commerce, information system has provided an unprecedented abundance of information resources, and it has also led to the problem of information overload. 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 [1,2].

Recommender system is one of the most important technologies in electronic commerce. And the collaborative filtering is almost the popular approach used in the recommender systems. With the development of electronic commerce systems, the magnitudes of users and items grow rapidly, resulted in the extreme sparsity of user rating data set. Traditional similarity measure methods work poor in this situation, make the quality of recommendation system decreased dramatically. Sparsity of users' ratings is the major reason causing the poor quality. The absence of a sufficient amount of available ratings significantly affects collaborative filtering methods reducing the accuracy of prediction. 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 [3,4]. 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.

To address the sparsity issue in the dataset, in this paper, a collaborative filtering recommendation algorithm based on singular value decomposition smoothing is presented. This approach predicts item ratings that users have not rated by the employ of SVD technology, and then uses Pearson correlation similarity measurement to find the target users' neighbors, lastly produces the recommendations. The collaborative filtering recommendation algorithm based on SVD smoothing can alleviate the sparsity problems of the user item rating dataset, and can provide better recommendation than traditional collaborative filtering.

II. COLLABORATIVE FILTERING RECOMMENDATION ALGORITHM BASED ON SVD SMOOTHING

A. Singular Value Decomposition (SVD)

Singular Value Decomposition (SVD), as part of Latent Semantic Indexing (LSI), was used widely in the area of Information Retrieval in order to solve the problems of synonymy and polysemy[5,6,7].

SVD is a matrix factorization technique which takes an $m \times n$ matrix M, with rank r, and decomposes it as follows:

$$SVD(M) = U \times S \times V^T$$

U and V are orthogonal matrices with dimensions $m \times m$ and $n \times n$, respectively. S, called the singular matrix, is an $m \times n$ diagonal matrix whose diagonal entries are non-negative real numbers.

The initial r diagonal entries of S(s1, s2, …, sr) have the property that si > 0 and s1 $\geqslant$ s2 $\geqslant$ … $\geqslant$ sr . Accordingly, the first r columns of U are eigenvectors of MM^T and represent the left singular vectors of M, spanning the column space. The first r columns of V are eigenvectors of M^TM and represent the right singular vectors of M, spanning the row space. If we focus only on these r nonzero singular values, the effective dimensions of the SVD matrices U, S and V will become $m \times r$, $r \times r$ and $r \times n$, respectively.

SVD can provide the best low-rank approximation of the original matrix, M, which is an attribute particularly useful in the case of Recommender Systems. By retaining the first

978-0-7695-3859-4/09 $26.00 © 2009 IEEE
DOI 10.1109/IITA.2009.491

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on September 24,2024 at 14:51:36 UTC from IEEE Xplore. Restrictions apply.

k ≤ r singular values of S and discarding the rest, which, based on the fact that the entries in S are sorted, can be translated as keeping the k largest singular values, we reduce the dimensionality of the data representation and hope to capture the important latent relations existing but not evident in the original representation of matrix M. The resulting diagonal matrix is termed S^k. Matrices U and V should be also reduced accordingly. U_k is produced by removing r − k columns from matrix U. V_k is produced by removing r − k rows from matrix V. Matrix M_k is defined as

$$M_k = U_k \times S_k \times V_k^T$$

Mk represents the closest linear approximation of the original matrix M with reduced rank k. Once this transformation is completed, users and items can be thought off as points in the k-dimensional space.

B. Employing the SVD to alleviate the sparsity

We apply the SVD method in our collaborative filtering recommendation system. After using SVD to the user-item matrix R, we get the matrices U, S, and V. US1/2 is a m×k matrix, which describes the relation of the users in the k dimensions. It can consider the user matrix and present the rating of the users. S1/2V is a n×k matrix and can consider the item matrix.

Then we transform the matrix using k-dimensional space and get the Uk, Sk, and Vk. Because the inter-multiples of UkSk and SkVk is the normalized rating of user to item, so we can convert-normalized multiples of UkSk and SkVk and then get the predictions as following.

$$P_{ui} = A_u + U_k\sqrt{S_k}(u) \times \sqrt{S_k}V_k^T(i)$$

Where Au is the average rating of the user u to the items, U, S, V are the matrices after using SVD to the rating matrix R, Uk, Sk, Vk are the reducing matrix of U, S, V, and k is the SVD reducing dimensional space.

C. Smoothing

One of the challenges of the collaborative filtering is the data sparsity problem. To prediction the vacant values in user-item rating dataset, we make explicit use of SVD as prediction mechanisms.

Based on the SVD results, we apply the prediction strategies to the vacant rating data as follows:

$$R_{ij} = \begin{cases} R_{ij} & if\ user\ i\ rate\ the\ item\ j \\ \mathrm{P}_{\text{ij-SVD}} & else \end{cases}$$

Where Pij-svd denotes the prediction value for user i rating towards an item j, and the value has calculated in above specific algorithm.

D. Calculating similarity

An important step of user based collaborative filtering algorithm is to calculate the similarity of users. In this paper, we use Pearson's correlation, as following formula, to measure the linear correlation between two vectors of ratings.

$$sim(i,j) = \frac{\sum_{c\in I_{ij}} w(R_{i,c} - A_i)(R_{j,c} - A_j)}{\sqrt{\sum_{c\in I_{ij}} (R_{i,c} - A_i)^2 \sum_{c\in I_{ij}} (R_{j,c} - A_j)^2}}$$

Where w is the parameter weight if it is a actual value or the value after the SVD method, Ri,c is the rating of the item c by user i, Ai is the average rating of user i for all the co-rated items, and Iij is the items set both rating by user i and user j.

E. Selecting neighbors

After we calculate the similarity of users, we can select the neighbors who will serve as recommenders. Two techniques have been employed in recommender systems:

(a) Threshold-based selection, according to which users whose similarity exceeds a certain threshold value are considered as neighbors of the target user.

(b) The top-n technique in which a predefined number of n-best neighbors is selected.

F. Producing recommendations

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

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

$$P_{ut} = A_u + \frac{\sum_{i=1}^{c} (R_{it} - A_i) * sim(u,i)}{\sum_{i=1}^{c} sim(u,i)}$$

Where A_u is the average rating of the target user u to the items, Rit is the rating of the neighbour user i to the target item t, Am is the average rating of the neighbour user i to the items, sim(u, i) is the similarity of the target user u and the neighbour user i, and c is the number of the neighbours.

III. DATASET AND MEASUREMENT

A. Data set

We use MovieLens collaborative filtering data set [8,9]. 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. 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.

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on September 24,2024 at 14:51:36 UTC from IEEE Xplore. Restrictions apply.

B. Performance measurement

The metrics for evaluating the accuracy of a prediction algorithm can be divided into two main categories [9,10]: statistical accuracy metrics and decision-support metrics. Statistical accuracy metrics evaluate the accuracy of a predictor by comparing predicted values with user provided values. Decision-support accuracy measures how well predictions help user select high-quality items. In this paper, we use decision-support accuracy measures.

Decision support accuracy metrics evaluate how effective a prediction engine is at helping a user select high-quality items from the set of all items. The receiver operating characteristic (ROC) sensitivity is an example of the decision support accuracy metric. The metric indicates how effectively the system can steer users towards highly-rated items and away from low-rated ones. We use ROC-4 measure as the evaluation metric. 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 ROC-4 definition as following:

$$ROC\text{-}4 = \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} v_i}$$

$$u_i = \begin{cases} 1, & p_i \geq 4 \ \ and \ \ q_i \geq 4 \\ 0, & otherwise \end{cases}$$

$$v_i = \begin{cases} 1, & p_i \geq 4 \\ 0, & otherwise \end{cases}$$

The larger the ROC-4, the more accurate the predictions would be, allowing for better recommendations to be formulated.

IV. CONCLUSIONS

Personalized recommender system is one of the most important technologies and the collaborative filtering is almost the popular approach used in the recommender systems. With the development of electronic commerce systems, the magnitudes of users and items grow rapidly, resulted in the extreme sparsity of user rating data set. Traditional similarity measure methods work poor in this situation, make the quality of recommendation system decreased dramatically. Sparsity of users' ratings is the major reason causing the poor quality. To address this issue, in this paper, a collaborative filtering recommendation algorithm based on singular value decomposition smoothing is presented. This approach predicts item ratings that users have not rated by the employ of SVD technology, and then uses Pearson correlation similarity measurement to find the target users' neighbors, lastly produces the recommendations. The collaborative filtering algorithm based on SVD smoothing can alleviate the sparsity problems of the user item rating dataset, and can provide better recommendation than traditional collaborative filtering algorithms.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

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

[5] M.G. Vozalis, K.G. Margaritis, Using SVD and demographic data for the enhancement of generalized Collaborative Filtering, Information Sciences 177 (2007) 3017–3037.

[6] Zhao Liang, Hu NaiJing, and Zhang ShouZhi algorithm design for personlalized recommendation systems, journal of computer research and development, Vol 39,No 18 ,2002.

[7] Gong SongJie, Ye HongWu, Dai YaE, Combining Singular Value Decomposition and Item-based Recommender in Collaborative Filtering, Second International Workshop on Knowledge Discovery and Data Mining, 2009. WKDD 2009. Page(s):769 – 772.

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

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

[10] Gao Fengrong, Xing Chunxiao, Du Xiaoyong, Wang Shan, Personalized Service System Based on Hybrid Filtering for Digital Library, Tsinghua Science and Technology, Volume 12, Number 1, February 2007,1-8.

Authorized licensed use limited to: Università di Bari Aldo Moro. Downloaded on September 24,2024 at 14:51:36 UTC from IEEE Xplore. Restrictions apply.