ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Improved Item Based Collaborative Filtering for Personalised Recommendation

I. Jenitta Magdalene, Rebecca Sandra
PG Scholar/Dept. of CSE, Assistant Professor/Dept. of CSE
Karunya University

***Abstract*- In the vast amount of information in the internet, to give individual attention for each users the personalized recommendation system is used, which uses the collaborative filtering method. Due to some popular objects the accuracy of the data's are lost. To remove this influence the method which is proposed here is a network based collaborative filtering which will create a user similarity network, where the users having similar interests of item or movies will be grouped together forming a network. Filtering the users when the number is large is done by the nearest neighbour approach or the filtration approach. Then we calculate discriminant scores for candidate objects. Validate the proposed approach by performing random sub-sampling experiments for about 20 times to get the accurate results. To improve and enhance the accuracy of the results the item based collaborative filtering is proposed. Results show that the approach out performs the network-based collaborative filtering method .**

Keywords: Recommender system, Collaborative filtering, Personalized recommendation, User similarity network, Nearest neighbour

I. INTRODUCTION

A series of survey is did regarding the issues that have been arising in the personalised recommendation systems, the old problems are the cold start and the sparsity where these problems arise in the case of a new website or for new item. After this some other problems occur like the improving a memory based collaborative filtering, the content based filtering and the item based collaborative filtering

II. PROBLEM STATEMENT

The problem statement in this approach is the presence of popular objects which adversely influence the correct estimation of similarities that have been obtained by the historical preference of the user as well as the pair wise user similarity between users and may further yield undesirable results of recommendation.

The main objective of this method is to first construct a user similarity network for personalized recommender systems using network based collaborative filtering, and to achieve a reasonable balance between accuracy and diversity measures which are obtained by removing the influence of the popular objects. This method starts with the construction of a user similarity network that are obtained from the historical data, and then by the pair wise similarity between the users for each and every objects. To remove the influence of the popular objects, they are being filtered by the nearest neighbour approach to filter out for each user a fraction of the weakest relationships between a user and other users. Alternatively, there is also a filtration approach that filters out weak relationships between users according to a pre-defined value and generates a filtration network. The discriminant scores for candidate objects is calculated by computing the historical preferences of the user and the pair wise similarity score for each objects and further

Vol. 2 Issue IV, April 2014
ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

the objects are sorted in descending order to obtain the highest ranking of the objects to prepare the list of recommendation.

III. NETWORK BASED COLLABORATIVE FILTERING

In order to avoid the previously explained drawback, a network based collaborative filtering method is used to remove the adverse influence which are done by the popular objects present there, by constructing a user similarity network with the use of historical data about preferences of users and the make recommendations based on this network.

It is possible that a tie occurs when two or more candidate objects are assigned equal discriminant scores. In such a situation, the tie is broken by putting objects with equal scores in random order. Alternatively, the average over ranks of objects is taken and assigns rank to the objects. The difference between these two strategies is negligible.

It is proposed to filter out the unreliable small user similarity scores according to the nearest neighbour strategy. Applying the filtering procedure to all users, the weight matrix W is obtained. First, given the collection of historical preferences of u users on o objects, represented as a matrix X = (xij), obtaining a pair wise user similarity matrix S = (sij). Applying the above filtering procedure to all users, the weight matrix is obtained. [10]

$$\text{Wij} = \begin{cases} s_{ij} & r_{ij} < \lambda \times \text{u}, \\ 0 & \text{otherwise} \end{cases}$$

An alternative approach for constructing a user similarity network is to define a threshold value, assign zeros to elements that are smaller than this cut off value, and then obtain the network corresponding to the resulting weight matrix. For this purpose, first map all users onto the constructed user similarity network and identify the set of neighbouring users that connect to the target user t. Then, weigh the preference of each of these users using the weight of the edge pointing from the user to the target t, and summate over all such neighbouring users and further perform a normalization to obtain the discriminant score for the target. It is to be ensured that the resulting discriminant score is in the range of [0, 1]. [10]

$$v_{ct} = \frac{\sum k \in u_t \, x_{ck} w_{kt}}{\sum k \in u_t \, w_{kt}}$$

IV.PROPOSED APPROACH

Although this method is proposed to target on improving user-based collaborative filtering, it is straightforward to incorporate the principles of the method into item-based collaborative filtering. This can be simply done by constructing an item similarity network by applying either the nearest neighbour strategy or the d-filtration scheme. The idea of constructing similarity networks can also be incorporated into content- based methods. In the proposed system the item based collaborative filtering is used. It will improve the performance of the system.

The existing algorithm cannot perform recommendation very well. An improved collaborative filtering recommendation algorithm based on dynamic item clustering method was proposed. The new items can be added in the clusters dynamically without changing too much and less resource demanding. Item-based Collaborative filtering is a method of making automatic predictions about the interests of a user by collecting preferences or taste information from many users.

The underlying assumption of the approach is that those who agreed in the past tend to agree again in the future. By analyzing historical data, collaborative filtering technology generates result set which is the most similar with the current user's interests. There are three phases in the algorithm: data representation; finding the nearest neighbour and generating the recommended result sets.

A. Item-based collaborative filtering

The tremendous growth in the amount of information that are available in the internet and the number of visitors to Web

Vol. 2 Issue IV, April 2014

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

sites in recent years gives some key challenges for recommender systems. In traditional collaborative filtering systems the amount of work increases with the number of users in the system, this is considered to be a disadvantage. New recommender system are needed which will quickly produce high quality recommendations, even for large-scale problems.

To avoid these issues the item-based collaborative filtering techniques is proposed. Item-based techniques analyze the user-item values to identify relationships between different items, and then use these relationships to indirectly give recommendations for users. Even though these systems have been so successful in the past, their wide usage has exposed some of their drawbacks such as the i)the sparsity problems in the data set, ii) problems associated with high dimensionality regarding the datasets.

B. Improved item-based collaborative filtering approach

It is proposed to construct a user similarity network with the use of historical data about preferences of users, and then make recommendations based on this network. For this purpose, first map all users onto the constructed user similarity network and identify the set of neighbouring users that connect to the target user. Then, weight the preference of each of these users using the weight of the edge pointing from the user a to the target, say and summate over all such neighbouring users and further perform a normalization to obtain the discriminant score for the target.

It is possible that a tie occurs when two or more candidate objects are assigned equal discriminant scores. In such a situation, the tie is broken by putting objects with equal scores in random order. Alternatively, can average over ranks of objects in a tie and assign the mean rank to the objects. According to our experiences, the difference between these two strategies is often negligible. [5]

C. Data representation

A model is built to represent the rating items. The input data of the algorithm is presented by a matrix, m rows for m users. While n for n items being evaluated; items in the matrix represent evaluation of items by users. Different methods can be used to represent the evaluation. Discrete value (such as 1, 2, 3, 4, 5) can be used to represent preference degrees of the user to the items. We agree that the higher of the discrete value, the more prefer of the user on the item. [11]

D. Finding the nearest neighbour

At this stage, items which are most similar to the target user are found. First, Similitude between user and other users are acquired. A few of methods we can choose from to calculate similitude between two items, Such as Cosine-based Similitude, Adjusted Cosine Similitude and Pearson Correlation Coefficient. We use Pearson Correlation Coefficient in the paper. Suppose that i and j are two users in the user space, we use Pearson Correlation Coefficient to calculate the similitude of i and j sim(i, j).

$$sim(i,j) = \frac{\sum_{c \in I_{i,j}} (R_{i,c} * \overline{R_i})(R_{j,c} * \overline{R_j})}{\sqrt{\sum_{c \in I_{i,j}} (R_{i,c} - \overline{R_i})^2} \sqrt{\sum_{c \in I_{i,j}} (R_{j,c} - \overline{R_j})^2}}$$

In equation (1), sim(i, j) is the similitude of user i and user j. Ri,c is the evaluation value of item c by user i; Rj, c is the evaluation value of item c by user j; while $\overline{Ri}$ and $\overline{Rj}$ are the average evaluation value on the item by user i and user j; Ii,j for the item set that user i and user j have evaluated on in common. [11]

E. Generating recommended result sets

In this phase, recommended results set is generated by the algorithm. Suppose Iu is the set that user u had evaluated on, evaluation pu,i on item i by target user u can be got in this way:

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

$$p_{u,i} = \overline{R_u} + \frac{\sum_{v \in N_u} sim(u,v) * (R_{v,j} - \overline{R_v})}{\sum_{v \in N_u} sim(u,v)}$$

Items that are not evaluated before can be predicted by items that had been predicted. [11]

IV. VALIDATION METHODS

The random sub-sampling strategy is implemented to validate the proposed approach. In each validation run, for a target user, collect a set of test objects as those that link to the target in the test data and a set of control objects as those that neither link to the target in the training data nor in the test data is present.

Then, calculate discriminant scores for both test and control objects, and rank each test object against all control objects. Repeating the ranking procedure for all users, obtain a set of ranking lists and further calculate four criteria to evaluate the performance of the proposed method. To account for uncertainties in the data splitting process, further repeat the above validation run 20 times and summarize over all repeats to obtain means and standard errors of the criteria.

V. EVALUATION CRITERIA

With the accurate values that are obtained by the large scale random sub sampling strategy, the evaluation is done.

A. Recall enhancement

The second criterion for evaluating the accuracy is called Recall enhancement (RE). Given a threshold T, one can claim a test object as successfully recommended if the object has been ranked among top T in the ranking list. For a user who has collected a number of objects in the test data, then count the number of successful recommendations among these objects and calculate the fraction of successfully recommended objects to obtain the recall for the user.

$$\text{RE(T)}=\frac{R(T)}{R^{(rand)}(T)} = \frac{O}{T} \times R(T)$$

Finally, averaging over recalls for all users who have collected at least one object, obtain the recall under the threshold. Although the recall itself can be used as a criterion to evaluate the accuracy of a method, more careful reasoning suggests the comparison against random guesses, yielding a criterion called recall enhancement. [12]

VI. CONCLUSION

In this approach, the proposed network-based collaborative filtering approach will achieve personalized recommendation by filtering out low similarities between users. It is found to have outperformed the ordinary user-based collaborative filtering and also the previous methods like the item and content based collaborative filtering methods and enhance not only the accuracy but also the diversity of recommendation results. Such relationships, mainly resulting from the share of popular objects between users, adversely affect the correct calculation of discriminant scores for candidate objects in the ordinary collaborative filtering approach.. As a result, this method achieves significant improvements in both the accuracy and the diversity of the resulting recommendations thus creating a balance between these two metrics.

REFERENCES

[1] Jeong, B., Lee, J., & Cho, H. (2010). “Improving memory-based collaborative filtering via similarity updating and prediction modulation”. 180,602–612.

[2] Kazama, K., Imada, M., & Kashiwagi, K. (2012). “Characteristics of information diffusion in blogs, in relation to information source type”. Neurocomputing, 76,84–92

[3] Liu, J. G., Shi, K. R., & Guo, Q. (2012). “Solving the accuracy–diversity dilemma via directed random walks”. Physical Review E, 85.

[4] Pazzani, M., & Billsus, D. (2007). “Content-based recommendation systems”. In The adaptive web, lecture notes in computer science.

[5] Rennie, J. D. M., & Srebro, N. (2005). "Fast maximum margin matrix factorization for collaborative prediction". Machine learning .Germany

[6] Sarwar, B., Karypis, G., J. (2001). "Item - based collaborative filtering recommendation algorithms".

[7] Van Rijsbergen, C. J. (1979). "Information retrieval". London Boston: Adaptive personalized recommendation based on adaptive learning. Neurocomputing, 74.

[8] Wilson, D. C., Smyth, B., & O'Sullivan, D. (2003). "Sparsity reduction in collaborative recommendation" International Journal of Pattern Recognition, 17, 863–884.

[9] Zhou, T., Kuscsik, Z., Liu, J. G., Medo, M., Wakeling, J. R., (2010). "Solving the apparent diversity–accuracy dilemma of recommender system

[10] I.Jenitta Magdalene,Rebecca Sandra (2014), "Solving the accuracy metrics and diversity measures for personalised recommendation".IJERA.

[11] Badrul Sarwar, George Karypis ," ItemBased Collaborative Filtering Recommendation Algorithms".

[12] Mingxin Gan a, Rui Jiang b, " Constructing a user similarity network to remove adverse influence of popular objects for personalized recommendation"