RESEARCH ARTICLE OPEN ACCESS

Solving the Accuracy Metrics and Diversity Measures For Personalised Recommendation

I. Jenitta Magdalene, Mrs.Rebecca Sandra
PG Scholar/Department of Computer Science, Karunya University, Coimbatore
Assistant Professor/Department of Computer Science, Karunya University, Coimbatore

ABSTRACT
In the vast amount of information in the internet, to give individual attention for each users, the personalised recommendation system is used, which uses the collaborative filtering method. By the result of the survey did with some papers, the main problems like the cold start and the sparsity which were found previously have been overcome. Filtering the users when the number is large is done by the nearest neighbour approach or by the filtration approach. 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. 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 and evaluate the method using two accuracy criteria and two diversity measures. Results show that the approach outperforms the ordinary user-based collaborative filtering method by not only enhancing the accuracy but also improving the diversity.

Keywords -Recommender system, Collaborative filtering, Personalised recommendation, User similarity network, Nearest neighbour

I. Introduction

The main objective of this method is to first construct a user similarity network for personalised 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 the objects are sorted in descending order to obtain the highest ranking of the objects to prepare the list of recommendation.

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.

Collaborative filtering is a technique which are widely used by some recommender systems. This filtering has two senses, one a narrow one and the other a more general one. Collaborative filtering is the process of filtering for information or patterns using the techniques which involves collaboration among multiple agents, viewpoints, etc Applications of the collaborative filtering typically involve large data sets.

III. Collaborative filtering methods

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.

3.1 Memory Based Collaborative Filtering

Memory-based collaborative filtering makes recommendations based on a collection of user preferences for items. The idea of this approach is that the interests of an active user will be more likely coincide with those of users who share similar preferences of the active user. Hence, the choice of a similarity measure between users is critical to rating items.[1]

A similarity update method that uses an iterative message passing procedure is proposed. Additionally, this work deals with a drawback of using the popular mean absolute error for performance evaluation, which ignores the distribution of ratings. A new modulation method and accuracy metric are presented in order to minimize the predictive accuracy error and to distribute predicted ratings over true rating scales. Results show that the proposed similarity update and prediction modulation techniques improve the rankings. However even some predictions may cluster around their significant values.

3.2 Content Based Collaborative Methods

Content-based collaborative filtering methods, where the systems will recommend an item to a target user, based upon a description of the object and a profile of the user's interests. Although the details of systems differ, this recommendation systems share in common a means for describing the items that may be recommended by the means for creating a profile of the user which describes the types of items which the user likes, and of comparing items to the user profile to determine what to recommend.

There are several limitations of this system, i) they require effort from the user and it is difficult to get many users to make this effort. ii) Customization systems do not provide a way to determine the order. However, when there are a small number of attributes, i) the performance, ii) simplicity and iii) understandability of decision trees for content-based models are all advantages of this particular method.[2]

3.3 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 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. [3]

The task of collaborative filtering is to predict the preferences of an active user. A novel regression-based approach is proposed that first learns a number of experts describing the relationships in ratings between pairs of items. Based on ratings provided by the user for the items, they are combined by the use of statistical methods to predict the user's preferences for the remaining items. [4]

The method was designed to address the problem of data sparsity and the prediction latency. The difference in the accuracy was more real when the number of ratings provided by an active user was small. Strong experimental evidence was obtained that the proposed approach can be applied to data with a large range of sparsity scenarios and is superior to non-personalised predictors even when ratings data are very sparse. The main advantage on this method are it is superior to non personalised predictors when data's are sparse, and the main drawback on this method which prevents the accurate evaluation of the recommendation is the prediction latency.

IV. Ordinary User Based Collaborative Filtering

In the ordinary used based collaborating filtering method only the individual users history will be analysed and the items which the user had preferred will be recommended in the future, but because of this the other user's similarity cannot be observed. Although user based collaborative filtering approaches have demonstrated remarkable successes in a variety of situations, the basic assumption that users sharing similar preferences in history would also have similar interests in the future may fail when some popular objects are present .

This is usually done by assigning a discriminant score to each candidate and then sorting the objects in non-ascending order according to their scores. In mathematics, given the collection of historical preferences of u users on o objects, represented as a matrix X = (xij)o×u denotes that the j-th user prefers the i-th object and zero otherwise , then the pair wise similarity scores for the users is calculated and obtain a user similarity matrix S = (sij)u×u. With this matrix, an ordinary user-based collaborative filtering method weights of users according to their similarity to the target user and then mix the preferences to obtain discriminant scores for candidate objects,

$$v_{ij} = \frac{\sum \mathbf{1} < k < u\, \boldsymbol{x_{ik} s_{kj}}}{\sum \mathbf{1} < k < u\, \boldsymbol{s_{kj}}}$$

Where vij is the discriminant score of the i-th object for the j-th target user, and skj the similarity score between the k-th and the j-th users.[6]

By this discriminant scores, the validation is done for about 20 times to obtain the accurate value. Sort the data's in the descending order and rank the scores, then the objects are recommended based on the rank they have got after this validation process. The main drawback of this method is the presence of the popular objects which reduces the accuracy measures and always goes in favour of the popular objects alone neglecting the accuracy detail of the whole process.

V. Directed Random Walks

Random walks have been successfully used to measure user and the object similarities in collaborative filtering recommender systems, which has high accuracy but low diversity. A key challenge of a CF system is that the reliably accurate results are obtained with the help of peer's recommendation, but the most useful individual recommendations are hard to be found among diverse niche objects Without relying on any context-specific information, they are able to obtain accurate and diverse recommendations, which outperform the state-of-the-art CF methods. This work suggests that the random-walk direction is an important factor to improve the personalised recommendation performance. The directed random walk process indeed has been defined as a local index of similarity in link prediction, community detection and so on.[9]

Meanwhile, similarities based on the global structural information, have been used for information filtering, such as the transferring similarity and the Page Rank index, communicability and so on. Although the calculation of such measures is of high complexity, it's very important to the effects of directed random walks on these measures. Finally it is found that the direction of random walks is very important for information filtering, which may be helpful for deeply understanding of the applicability of directed similarity.

VI. 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.

Wij = {$s_{ij}\, r_{ij}$<λ×u,
{0 otherwise

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 to 1.[6]

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

VII. 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.

VIII. Evaluation Criteria

There are two criteria to evaluate the accuracy of the proposed method in recommending user preferred objects and two criteria to measure the diversity of recommendations for different users which has the influence of the popular objects. With the accurate values that are obtained by the large scale random sub sampling strategy, the evaluation is done.

8.1 Accuracy Metrics

Accuracy metrics is the values that are obtained from the regular method which gets the values from the historical method that the system has already stored in the database, and based upon the rank the objects had achieved, it will recommend the items. To improve these processes of evaluation with the historical data's of the user these two metrics are implemented.

8.1.1 Mean Rank Ratio

The first criterion for evaluating the accuracy is called the Mean rank ratio (MRR). Sort the objects in non-ascending order according to their discriminant scores and obtain the rank for the test object. In the situation that multiple objects have equal discriminant scores, the tie is broken by putting these objects in random order. Then further divide the rank of the test object with the total number of objects in the sorting process to obtain the rank ratio for the test object. When the rank ratios for all objects in the test set are averaged, then the criterion of the mean rank ratio is obtained.

8.1.2 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.

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.[10]

8.2 Diversity Measures

These measures are the ones which arise due to the inclusion of the popular objects, since it will give out the results which are not assumed or expected by the user. Many regular users like to have their own regular items and objects to be recommended for them whenever they visit the website. In order to improve the accuracy metrics, one should not completely ignore the popular objects, so to maintain a balance between the accuracy and diversity measures, two metrics are proposed under the diversity measures, they are the mean personality and mean novelty.

8.2.1 Mean Personality

It is used T = 20 in the calculation of this criterion. It is also obvious that a method of higher recommendation accuracy will have a larger recall enhancement. A criterion called the Mean personality (MP) is adopted to quantify the diversity of recommendations made by the proposed method for different users. Given the discriminant scores calculated for a list of objects, sort the objects in non-ascending order according to their scores.

MP (T)=1 $-\frac{1}{T}\frac{2}{u(u-1)}\sum|\Omega j(T) \cap \Omega k(T)|$ [10]

8.2.2 Mean Novelty

Use T = 20 in the calculation of this criterion. It is also evident that a method of higher recommendation diversity will have a larger mean personality. A criterion called the Mean novelty (MN) is improved to quantify the novelty of recommendations made by the proposed method. For an object, calculate the fraction of users that are relevant to the object in history.

MN(T)=$\frac{1}{u}\sum \log_2 fi$ [10]

IX. Conclusion

In this approach, the proposed network-based collaborative filtering approach will achieve personalised 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”.

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

[3] Sarwar, B., Karypis, G., J. (2001). “Itembased collaborative filtering recommendation algorithms”.

[4] Slobodan Vucetic., (2004). Collaborative Filtering Using a Regression-Based Approach

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

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

[7] Van Rijsbergen, C. J. (1979). “Information retrieval”. London Boston: Adaptive personalised 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] Mingxin Gan a, Ruiiang .,(2013) ”Constructing the user similarity network to remove the adverse influence of popular objects for personalised recommendation”