ATLANTIS PRESS
Advances in Computer Science Research, volume 71
4th International Conference on Machinery, Materials and Information Technology Applications (ICMMITA 2016)

Algorithm optimization of recommendation based on probabilistic matrix factorization

Qi He[1,a], Yan-fen Cheng[2,b]
[1]Wuhan University of Technology, school of Computer Science and Technology, WuHan
[2]Wuhan University of Technology, school of Computer Science and Technology, WuHan
[a]1145602766@qq.com, [b]995132428@qq.com

Keywords: Recommendation system；Probabilistic matrix factorization；User-trust network；Stochastic gradient descent.

Abstract. With the rapid development of the internet, the excessive information of user and item leads to user-item rating matrix becomes bigger and more sparse, also, accuracy of the traditional collaborative filtering recommendation algorithm gets lower. So in order to improve the precision of recommendation system, this paper considered user-trust network and user rating bias had influence on the accuracy of recommendation system, designed a probabilistic matrix factorization which integrates user-trust network and user rating bias. According to the results, the proposed algorithm is superior to probabilistic matrix factorization, and has a better prediction.

Introduction

The rapid development of Internet has changed the human's life-style, now people are used to listening to music, watching movies, shopping, reading news and so on with the Internet. At the same time, information is constantly updated and increased, which has made the problems such as the amount of information has been beyond the scope of the user processing capacity, users cannot find useful information effectively and so on. Currently the most widely and effective ways to solve these problems is to design a recommendation system. we can automatically recommend information to users through the recommendation system which save the user blind search time. Nowadays, the major electronic business sites, social networking sites, as well as music players have their own recommendation systems. They will be recommended according to user's browsing information after login and give users convenient and personalized experience.

In daily life, people tend to find a trusted friend to recommend an item, or put forward a proposal to help themselves. Therefore, the trust degree between friends has become a crucial factor in the recommendation system. At the same time, the user may be limited by interference of external factors when rating for shopping in the use of a web site so that the score is lower than the true value, which is the user rating bias.

With the popularity of the Internet, the large amount of users and project's information in the recommendation system has made user-item rating matrix increasingly large and sparse so that the recommendation quality of the recommendation system is getting lower and lower. Therefore, how to improve the recommendation quality of the recommendation system has become a widespread concern at the present stage. In this paper, the probability matrix factorization technique is applied to the user-item rating matrix. In view of data sparsity, this paper proposes a collaborative filtering algorithm based on probability matrix factorization. Combining user's interest and user's interest in social relationship, we can construct a probabilistic model to recommend and predict. Mapping the common interest of the user and trusted friends to the potential factor space, then analyzing the potential factors according to the design of the learning algorithm, this paper will realize the personalized recommendation based on the trust relationship of social network friends and join the mining user rating bias so as to improve the accuracy of the recommendation system.

Related Work

Collaborative filtering recommendation algorithm is currently the most popular recommendation algorithm. Different from the traditional recommendation algorithm, it is the first to analyze the user's interest to find the similar target user in the user group and then according to the evaluation of similar users to a project to decide whether to recommend to the target user. Collaborative filtering is mainly divided into memory-based filtering and model-based filtering. Memory-based collaborative filtering algorithm is divided into user-based collaborative filtering recommendation algorithm and item-based collaborative filtering recommendation algorithm, the former is recommended to the target users according to the preference information of similar users, the latter recommends several projects with high similarity to target users based on the similarity between the projects. In contrast to content-based recommendation algorithms, memory-based collaborative filtering can automatically filter information that machine are difficult to automatically filter, such as movie, goods. However, memory-based collaborative filtering is highly dependent on the similarity, which probably results in an inaccurate similarity when the score is very sparse. In the circumstances, the data sparsity problem is so serious while there is a growing problem of cold start and expansion with the increase of users and products. Therefore, the collaborative filtering algorithm based on model factorization is gradually being concerned in the field of recommendation.

Model-based collaborative filtering methods include matrix factorization[1], clustering algorithm[2], Bayesian network[3]and so on. Among them, the matrix factorization technique has recommended by Netflix recommended competition, because of its high recommendation accuracy and good scalability, it got more attention of domestic and foreign scholars in the recommended field. Matrix Factorization （MF） technology describes that user-item rating matrix R is decomposed into a form in which the user-factor matrix U and project-factor matrix V are multiplied, known as R=U*V. And then we can assume that the difference between the true rating and the prediction rating of the user to the items obeys Gauss distribution, which can derive the objective function .And the global optimal solution can be obtained by optimizing the objective function. A new recommendation algorithm based on singular value factorization has been proposed in the literature[4].The literature[5] that the probability distribution is added to the improved singular value factorization technique has presented Probabilistic Matrix Factorization（PMF）.Like other collaborative filtering algorithms, these models will have the data sparse problems resulting from a large amount of data. Sparse data can also arise the problem named cold start problem that new users can’t get recommended. Therefore, in the case of very sparse user-item rating matrix, how to improve the quality of recommendation is now the recommendation system should be solved. And now many studies have shown that social relationships among users can improve these problems. The literature[6] proposes to improve the recommendation quality of the recommendation system by combining the context information and the social network, The literature[7] introduces the concept of user similarity in social networks to improve the recommendation accuracy in social networks. But with the constant deepening of social network, researchers have turned their attention to the research of trust relationship in the recommendation system.

At present, the behavior of some businesses to profit on e-commerce web sites ,such as making false evaluation, eliminating negative comments ,has made users in the online shopping no longer easily believe the commodity evaluation and more hope to seek a trusted friend's recommendation. So, trust as a key word in social network, was introduced into the recommendation system. Epinions[8] is the first the shopping site to provide a "trust mechanism" .Users make judgments by the merits of other users' product quality. If you trust someone, he / she will be included in the trust list , otherwise not included. The ski mountaineering community website Moleskiing[9] and movie recommendation website FilmTrust[10] are all recommendation algorithms based on the explicit trust. The paper[11] designs a set of transfer rules of trust relationship by using the transport properties of trust, and makes the similarity and trust combined to recommend. In the literature[12], the matrix factorization is combined with the social network, but the paper doesn't take into account the trust relationship among users. Although these methods improve the recommendation accuracy , they have not been studied

the relationship between user trust network and user-item rating matrix, and they need to calculate the similarity and trust value between any two users, which will lead to higher computational complexity and longer time in the use of large-scale data sets. At the same time, these methods do not take into account certain bias of the data set which will also affects the recommended quality.

Recommendation Algorithm based on the user trust network

Probabilistic matrix factorization model is a very successful recommendation model of all the matrix factorization models. The basic idea is to introduce the probability to the matrix factorization, which means probabilistic matrix factorization model R is decomposed into a form in which the user-factor matrix U and project-factor matrix V are multiplied, known as R=U*V, and then we can assume that the implicit feature vector matrix of users and projects are both consistent with Gauss's prior distribution[5],that is

$$p\left(U \mid \sigma_U^2\right)=\prod_{i=1}^{M} N(U_i \mid 0, \sigma_U^2 I) \tag{1}$$

$$p\left(V \mid \sigma_V^2\right)=\prod_{i=1}^{N} N(V_i \mid 0, \sigma_V^2 I) \tag{2}$$

There is that the Gauss distribution which expectation value is 0 and variance is σ.

Based on this assumption, the degree of user's preference for a project is a series of probability combination problems.

$$p(R \mid U, V, \sigma^2)=\prod_{i=1}^{M}\prod_{j=1}^{N}[N(R_{ij} \mid U_i^T V_j, \sigma^2)]^{I_{ij}} \tag{3}$$

Where $p(A \mid B)$ denotes the conditional probability of event A in case of occurring event B and I_{ij}^R denotes indicator function. If the user i choose the project j, and then I_{ij}^R equals to 1, otherwise equals to 0.

Bayesian inference is used to obtain the posterior probability of the implicit features of users and projects.

$$p\left(U, V \mid R, \sigma_R^2, \sigma_V^2, \sigma_U^2\right) \propto p\left(R \mid U, V, \sigma_R^2\right) \times p\left(U \mid \sigma_U^2\right) \times p(V \mid \sigma_V^2) \tag{4}$$

The Eq.4, as predictive equations, can be used to minimize the logarithm to the both sides of the equation so that the objective function of probabilistic matrix factorization can be obtained .But when the user-item rating matrix R is very sparse, there will be over fitting problems. The error of fitting performance on the training data set is very small while the error on the test data set increases. The solution to this problem is regularization, the objective function after adding the regularization items is as follows.

$$L(U, V)=\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{N} I_{ij}\left(R_{ij}-U_i^T V_j\right)^2+\frac{\lambda_1}{2}\sum_{i=1}^{M}\left\|U_i\right\|_{Fro}^2+\frac{\lambda_2}{2}\sum_{j=1}^{N}\left\|V_j\right\|_{Fro}^2 \tag{5}$$

Where λ_1 and λ_2 denotes the weight of regularization term and $\left\|\cdot\right\|_{Fro}^2$ means the square of the vector F's norm.

This paper considers that if user A trusts user B, the potential factor spaces of A and B should be similar and the degree of similarity depends on A's level of trust in B. In order to obtain the optimal solution, it is desired that the error between the users is minimized, that is

$$\min_U=\frac{1}{2}\sum_{i=1}^{M}\sum_{deTrust(i)} T_{ab}\left\|U_a-U_b\right\|_F^2 \tag{6}$$

Where Trust(i) denotes a collection of user A's trusted users. If the user A trust user B, the value of T_{ab} is 1, otherwise equals to 0.

There is a certain user rating bias in the data set, which will also affect the recommendation

accuracy. In this paper, we join the user rating bias in the above-mentioned improvement, and propose a probabilistic matrix factorization algorithm based on user trust, which integrates the user rating bias. First, we need to define the user preferences for the project as μ, the initial value is set to the average of the training samples, and then a new objective function is obtained, finally, the objective function is optimized by stochastic gradient descent algorithm to obtain the optimal solution. If the user i exists empty set of trusted users, that is

$$U_i \leftarrow U_i - \alpha \times (r \times V_j + \lambda_1 \times U_i) \tag{7}$$

$$V_j \leftarrow V_j - \alpha \times (\beta \times d \times U_i + \lambda_2 \times V_j) \tag{8}$$

$$\mu \leftarrow \mu - \alpha \times r \tag{9}$$

Where α in the Eq.7 and Eq.8 refers to the step size, known as the learning rate, d is the difference between the predicted value and the true value, and β in the Eq.8 is the pre-defined normalization parameter.

If user i has non - empty set of trusted users, the Eq.7 changes to

$$U_i \leftarrow U_i - \alpha \times \left(r \times V_j + \lambda_1 \times U_i + t \times \beta \times U_i - \beta \times s\right) \tag{10}$$

Where user t refers to the rows of a set of trusted users, and s is the sum of the trusted users.

Experiment Analysis

Dataset Description

The dataset used in the experiment is Filmtrust, the dataset is selected because it is sparse enough, and also contains the score of user evaluation and trust. This dataset used in this experiment contains 1695 users who have rated 2056 films, a total of 653814 rating records, and 461528 trust records. Wherein the maximum value of the movie data that the user evaluates is 1024 and the minimum value is zero. The dataset of ratings range from 0.5 to 0.5, The higher the user's rating for the movie, the more likely the user will be.

Evaluation Criterion

In this paper, the root mean square error(RMSE) and mean absolute error(MAE) are used to verify the prediction accuracy of the proposed algorithm. The RMSE calculation method is as follows

$$RMSE = \sqrt{\frac{\sum_{u,i}(r_{ui} - \hat{r}_{ui})^2}{N}} \tag{11}$$

Where r_{ui} denotes the real rating, $\hat{r}_{ui}$ denotes the predictive rating, N denotes the number of ratings contained in the test set.

If the value of RMSE is smaller, the predictive value is closer to the true value, and the higher the recommendation accuracy is.

The MAE calculation method is as follows

$$MAE = \frac{\sum_{u,i}\left|r_{ui} - \hat{r}_{ui}\right|}{N} \tag{12}$$

Similarly, if the value of MAE is smaller, the predictive value is closer to the true value, and the higher the recommendation accuracy is.

This paper proposes a probabilistic matrix factorization which integrates user-trust network and user rating bias called BiasTrustPMF, and compared with the probability matrix factorization algorithm (PMF), then the experimental results are obtained.

Experimental Results

Randomly scramble the numbers 1 to n in the experimental data to obtain a random number sequence as an index, and then take the first 1000 index data as a test sample, the rest of the data as a

training sample. In this paper, a set of experiments are designed to compare the BiasTrustPMF and PMF algorithms.

In the Filmtrust dataset, BiasTrustPMF and PMF algorithms are calculated to obtain the predicted value, and then calculate the RMSE value (Figure 1) and MAE value (Figure 2). Where x coordinate is set to the iteration number of the stochastic gradient descent algorithm, The values are 10, 20, 30, ... , 100 times. The y coordinate of Fig. 1 is the value of RMSE, Fig. 2 is the value of MAE.

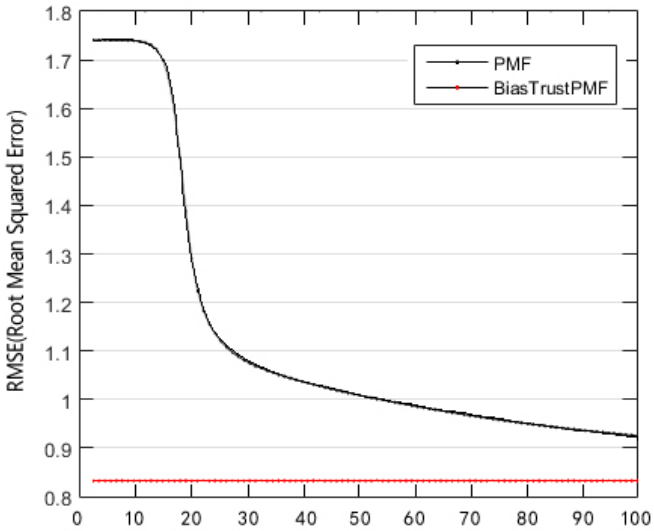

Fig. 1 Comparison of RMSE value of each algorithm on Filmtrust data set

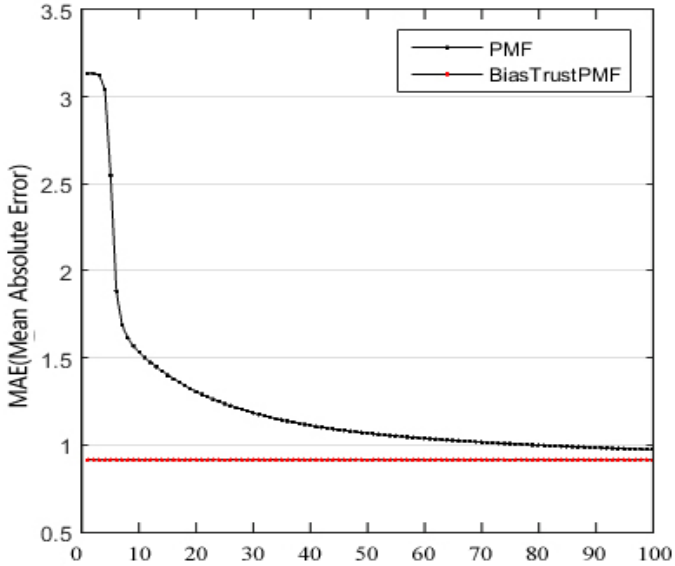

Fig. 2 Comparison of MAE value of each algorithm on Filmtrust data set

Comparing the data in figures 1 and figure 2, with the increase of the number of iterations, the RMSE and MAE of the algorithm proposed in this paper are smaller, while the number of iterations has little effect on the RMSE and MAE values. Therefore, the algorithm proposed in this paper is more accurate and the precision is higher.

Conclusion and Future work

In this paper, through the research of collaborative filtering algorithm and probability matrix factorization algorithm, taking into account the user's trust and user rating bias in the big data

environment, probability matrix factorization algorithm is optimized and improved, and proposes a probability matrix factorization algorithm which combines user trust and user rating bias. Finally, it reduces the sparsity of the recommendation system in the case of large amount of data and improves the recommendation accuracy. The next research direction is to improve the iteration speed of the algorithm proposed in this paper under the big data platform, at the same time, the performance of the BiasTrustPMF algorithm will be more fully tested.

References

[1] Li J. Research on Collaborative Filtering Recommendation Algorithm Based on Matrix Decomposition Method[C]// International Conference on Mechatronics, Materials, Chemistry and Computer Engineering. 2015.

[2] Herlocker J. Clustering Items for Collaborative Filtering[J]. Proceedings of the Acm Sigir Workshop on Recommender Systems, 1999.

[3] Ono C, Motomura Y, Asoh H. Recommendation Algorithm based on Bayesian Network[J]. Ieice Technical Report Neurocomputing, 2004, 104:55-60.

[4] Sharifi Z, Rezghi M, Nasiri M. New algorithm for recommender systems based on singular value decomposition method[C]// International Econference on Computer and Knowledge Engineering. 2013:86-91.

[5] Mnih A, Salakhutdinov R. Probabilistic matrix factorization[C]// International Conference on Machine Learning. 2012:880-887.

[6] Panniello U, Tuzhilin A, Gorgoglione M. Comparing context-aware recommender systems in terms of accuracy and diversity[J]. User Modeling and User-Adapted Interaction, 2014, 24(1):35-65.

[7] Rong H G, Huo S X, Hu C H, et al. User similarity-based collaborative filtering recommendation algorithm[J]. Journal on Communications, 2014.

[8] Zhang Y,Wu Z H,Chen H J,et al.Mining Target Marketing Groups from Users'Web of Trust on Epinions[C]//Proceedings of the 2008 AAAI Spring Symposium.2008:116-121.

[9] Avesani P, Massa P, Tiella R. A trust-enhanced recommender system application: Moleskiing[C]// ACM Symposium on Applied Computing. 2010:1589-1593.

[10] Golbeck J. Generating Predictive Movie Recommendations from Trust in Social Networks[C]// Trust Management, International Conference, Itrust 2006, Pisa, Italy, May 16-19, 2006, Proceedings. 2006:93-104.

[11] Gong S. A Collaborative Filtering Recommendation Algorithm Based on Trust Network and Trust Factor[J]. Journal of Convergence Information Technology, 2013, 8(5):1111-1118.

[12] Liu J, Wu C, Liu W. Bayesian Probabilistic Matrix Factorization with Social Relations and Item Contents for recommendation[J]. Decision Support Systems, 2013, 55(3):838-850.