

Probabilistic Matrix Factorization Recommendation Algorithm with User Trust Similarity

Yuxin Dong¹, Shuyun Fang¹, Kai Jiang¹, Fukun Chen¹ and Guisheng Yin¹

¹College of computer Science and Technology, Harbin Engineering University, Harbin, China

Abstract. In this paper, we describe the formatting guidelines for Conference Proceedings. Whether the user similarity calculation is reasonable in the traditional collaborative filtering recommendation algorithm directly affects the result of the collaborative filtering recommendation algorithm. This paper proposes a probabilistic matrix factorization recommendation algorithm with user trust similarity which combines improved similarity of users' trust and probability matrix factorization recommendation method. The results show that proposed algorithm could relieve user cold start issues and effectively reduce the error of recommendation.

1 Introduction

With the rapid growth of data, "information overload" has become the most important issue we need to solve at present. How to achieve accurate and personalized recommendations in massive data, this issue is particularly prominent in major e-commerce platforms, mobile applications and other fields. The traditional collaborative filtering recommendation algorithm is one of the most widely used recommendation methods and is also the most successful.[1-3] However, the CF needs to be based on user history information, such as the user's evaluation of the product.

The main drawback of this is that as the amount of data continue to increase, the number of points generated by users is limited. The data scarcity of the user history rating matrix severely restricts the calculation of similarity in the collaborative filtering method and reduces the recommendation accuracy of the algorithm.[4-5] Therefore, generating accurate recommendations for target users through user similarity calculations has become a problem to be solved at this stage.

Under these circumstances, some existing solutions focus on trust relationships between users.[6-7] The literature [8] proposed a PIP method, Consider the similarity between users from three aspects of Proximity and Impact and Popularity. The literature [9] proposed that the NHSM method achieved good results on the basis of the PIP method. Literature [10] use the matrix decomposition method which is based on ACOS and AMSD [11-12] to calculate the user similarity has got better application effect.

In order to solve the problems of the traditional recommendation algorithm and make full use of existing data information. This paper proposes a collaborative

filtering recommendation algorithm that integrates user implicit trust. The algorithm puts forward a new method by taking into account the differences in ratings between users' common scoring items and users and the differences in different user rating items. The results prove that the accuracy of the recommendation is improved.

2 User trust similarity

The traditional rating trust relationship is: If user A trusts user B, then user B also trusts user A. This is equivalent to the fact that the trust matrix is a symmetric matrix. But in real life, Trust relationships are often not equal. This relationship can be better reflected in social networks, so we consider that the real trust relationship matrix is asymmetrical. In other words, user A's trust in user B may not equal to user B's trust in user A. The main reason to consider the difference in trust is that the number of items in the two ratings is different. If two users have the same number of scoring items, they have the same degree of trust. If two users have different number of scoring items. In the case of a certain number of common scoring items between the two, users with more evaluation numbers have a lower degree of trust for users with a lower number of ratings.

For the two users A and B, considering the relative proportion of the ratings number, set the trust coefficient of A to B as formula (1):

$$T_{A, B} = \frac{|I_A \cap I_B|}{|I_A|} * I_B \frac{|I_A \cap I_B|}{|I_A \cup I_B|} \quad (1)$$

In the formula, I_A is the collection of items that user A rated excessive. Similarly, I_B is the collection of items that user B rated. In addition to the trust difference between users, the traditional algorithms only consider

the directional differences of different user rating vectors when calculating users' similarities, and don't take different users' specific scores into consideration. When the two vectors are in the same direction but the values are inconsistent, the traditional methods of calculation seem unreasonable. If the direction of the two users' interest vectors are the same as the direction of the third user, the smaller of the score difference with third user, the stronger of the trust relationship between them.

In response to the problem mentioned above, For User A and User B, consider the differences in the values of the ratings. The score difference coefficient between users is defined as formula (2):

$$D_{A,B} = \theta^{\sum_{j \in I} |R_{A,j} - R_{B,j}|} \quad (0 < \theta < 1) \quad (2)$$

In the formula, I is the collection of items for User A and User B rated. $R_{A,j}$ and $R_{B,j}$ are the rating of item j that user A and user B rated respectively. Max is the maximum score, Min is the minimum score, θ is a constant between 0 and 1.

Finally, the improved trust similarity formula is defined as followed:

$$\text{simTrust}(A, B) = T_{A,B} * D_{A, B} * \text{sim}(A, B) \quad (3)$$

In formula (3), $\text{sim}(A, B)$ uses Pearson similarity as an improvement object in this paper.

3 TSPMF MODEL

3.1 User trust similarity probabilistic matrix factorization

User trust similarity matrix is decomposed into two low-dimensional implicit vector feature matrices. Where m indicates the number of users. The user feature matrix U is a $l * m$ dimension matrix. And the trust feature matrix Z is a $l * m$ dimensional matrix. It is necessary to get the most suitable U and Z to minimize the factorization error of the user trust similarity matrix, it is equivalent to let the inner product of U and Z satisfy the minimum error of the user trust similarity matrix C . The error formula defined as:

$$p(C|U, Z, \sigma_c^2) = \prod_{i=1}^m \prod_{k=1}^m [N(C_{ik}|U_i^T Z_k, \sigma_c^2)]^{I_{ik}^c} \quad (4)$$

In formula (4), the vectors U_i represents the characteristics of user i , and Z_k represents the trust characteristics of the group of k , $N(x|\mu, \sigma_c^2)$ represents the probability density function of a Gaussian distribution with mean μ and variance of σ_c^2 . When there is a trust relationship between user i and user k , Then $I_{ik}^c=1$, otherwise it is 0. The Gaussian distributions of user feature and user trust feature are defined as follows:

$$p(U|\sigma_U^2) = \prod_{i=1}^m N(U_i|0, \sigma_U^2 I)$$

$$p(Z|\sigma_Z^2) = \prod_{k=1}^m N(Z_k|0, \sigma_Z^2 I) \quad (5)$$

In formula (5), σ_U^2 is the variance of U distribution, σ_Z^2 is the variance of the Z distribution, I represents the identity matrix.

Mentioned in section 2.1, the similarity of trust takes into account between the different users number and scale of user preferences for projects. The similarity of trust between users takes the different number and scale of user preferences for items into account. where C_{ik} is the trust degree of user i to user k . It is shown as:

$$C_{ik} = \frac{|I_i \cap I_k|}{|I_i|} * \frac{|I_i \cap I_k|}{|I_i \cup I_k|} * \theta^{\sum_{j \in I} |R_{i,j} - R_{k,j}|} * \frac{\sum_{i \in I} (R_{i,j} - \bar{R}_i) * (R_{k,j} - \bar{R}_k)}{\sqrt{\sum_{i \in I} (R_{i,j} - \bar{R}_i)^2} \sqrt{\sum_{i \in I} (R_{k,j} - \bar{R}_k)^2}} \quad (6)$$

In formula (4), I_i is a set of items that rated by user i ; I_k is a set of items that rated by user k . $R_{i,j}$ is the user i 's rating to the item j ; $R_{k,j}$ is the user k 's rating to the item j . According to Bayesian theorem, the user trust similarity matrix is decomposed into user characteristic matrix and user trust matrix:

$$p(U, Z|C, \sigma_c^2, \sigma_U^2, \sigma_Z^2) \propto p(C|U, Z, \sigma_c^2) p(U|\sigma_U^2) p(Z|\sigma_Z^2) \quad (7)$$

3.2 User-item similarity probabilistic matrix factorization

Similarly, the user-item rating matrix is decomposed into two low-dimensional implicit vector feature matrices. Where n indicates the number of items. The user feature matrix U is a $l * m$ dimension matrix. And the item feature matrix V is a $l * n$ dimensional matrix. It is necessary to get the most suitable U and V to minimize the factorization error of the user trust similarity matrix, it is equivalent to let the inner product of U and V satisfy the minimum error of the user trust similarity matrix R . The error formula defined as:

$$p(R|U, V, \sigma_R^2) = \prod_{i=1}^m \prod_{j=1}^n [N(R_{ij}|U_i^T V_j, \sigma_R^2)]^{I_{ij}^R} \quad (8)$$

In formula (8), the vectors V_j represent the trust characteristics of item j , $N(x|\mu, \sigma_R^2)$ represents the probability density function of a Gaussian distribution with mean μ and variance of σ_R^2 . If item j have been rated by user i , then $I_{ij}^R=1$, otherwise it is 0. The Gaussian distributions of user feature and item feature are defined as follows:

$$p(U|\sigma_U^2) = \prod_{i=1}^m N(U_i|0, \sigma_U^2 I)$$

$$p(V|\sigma_V^2) = \prod_{j=1}^n N(V_j|0, \sigma_V^2 I) \quad (9)$$

In formula (9), σ_U^2 is the variance of U distribution, σ_V^2 is the variance of the V distribution, I represents the identity matrix. According to Bayesian theorem, the user-item matrix is decomposed into user characteristic matrix and item characteristic matrix:

$$p(U, V|R, \sigma_R^2, \sigma_U^2, \sigma_V^2) \propto p(R|U, V, \sigma_R^2) p(U|\sigma_U^2) p(V|\sigma_V^2) \quad (10)$$

3.3 Joint probabilistic matrix factorization

The joint probabilistic matrix factorization can be used to obtain a user characteristic matrix that satisfies both the user trust relationship and the user-item matrix. Furthermore, the missing values in the user-item scoring matrix are obtained from the inner product of the user feature matrix and the item feature matrix under the constraint conditions. The logarithmic value of posterior of joint probability matrix is defined as:

$$\begin{aligned} \ln p(U, V, Z|C, R, \sigma_C^2, \sigma_R^2, \sigma_U^2, \sigma_V^2, \sigma_Z^2) \\ = -\frac{1}{2\sigma_C^2} \sum_{i=1}^m \sum_{k=1}^m I_{ik}^C (C_{ik} - U_i^T Z_k)^2 \\ -\frac{1}{2\sigma_R^2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R (R_{ij} - U_i^T V_j)^2 \\ -\frac{1}{2\sigma_U^2} \sum_{i=1}^m U_i^T U_i - \frac{1}{2\sigma_V^2} \sum_{j=1}^n V_j^T V_j - \frac{1}{2\sigma_Z^2} \sum_{k=1}^m Z_k^T Z_k \\ -\frac{1}{2} \left(\sum_{i=1}^m \sum_{k=1}^m I_{ik}^C \right) \ln \sigma_C^2 + \left(\sum_{i=1}^m \sum_{j=1}^n I_{ij}^R \right) \ln \sigma_R^2 \\ -\frac{1}{2} (m \ln \sigma_U^2 + n \ln \sigma_V^2 + m \ln \sigma_Z^2) + C \end{aligned} \quad (11)$$

In formula (11), C is a constant. In order to obtain the minimum error factorization of the user trust similarity matrix and the user-item rating matrix, which equivalent for formula (11) to take the maximum value. Formula (11) is simplified to formula (12), it's said that the maximum value of formula (11) corresponds to the minimum value of formula (12).

$$\begin{aligned} E = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij}^R (R_{ij} - U_i^T V_j)^2 \\ + \frac{\lambda_C}{2} \sum_{i=1}^m \sum_{k=1}^m I_{ik}^C (C_{ik} - U_i^T Z_k)^2 \\ + \frac{\lambda_U}{2} \|U\|_{Fro}^2 + \frac{\lambda_V}{2} \|V\|_{Fro}^2 + \frac{\lambda_Z}{2} \|Z\|_{Fro}^2 \end{aligned} \quad (12)$$

In formula (12), $\lambda_C = \sigma_R^2 / \sigma_C^2$, $\lambda_U = \sigma_R^2 / \sigma_U^2$, $\lambda_V = \sigma_R^2 / \sigma_V^2$, $\lambda_Z = \sigma_R^2 / \sigma_Z^2$, $\|\cdot\|_{Fro}^2$ represents the Frobenius norm.

4 EXPERIMENT

The experiment mainly verified the following questions:

- The influence of the choice of θ value in the user trust similarity formula on the recommendation effect, and the θ value of the best recommended effect;
- After θ is determined, test the λ_C value of the user trust similarity matrix through experiments;
- Finally evaluate the recommended effect of the TSPMF algorithm.

4.1 Joint probabilistic matrix factorization

The data set selected movielens-100k as the experimental data set, 80% of the data set are training sets, the remaining 20% is test set. The rating range is 1-5. 1 is not liked, 5 is liked.

4.2 Experimental evaluation index

Mean absolute error (MAE) and root mean squared error (RMSE) are used as experimental evaluation indicators. The formulas are as follows:

$$\begin{aligned} MAE &= \frac{1}{T} \sum_{i,j} |R_{i,j} - \bar{R}_{i,j}| \\ RMSE &= \sqrt{\frac{1}{T} \sum_{i,j} (R_{i,j} - \bar{R}_{i,j})^2} \end{aligned}$$

In the formula, T is the record numbers of items predicted. $R_{i,j}$ is the score predicted by TSPMF algorithm. $\bar{R}_{i,j}$ is the score that user i rated before.

4.3 Experimental analysis

In the experiment, the result of Fig. 1 is the MAE value of θ when different K-nearest-neighbors are selected.

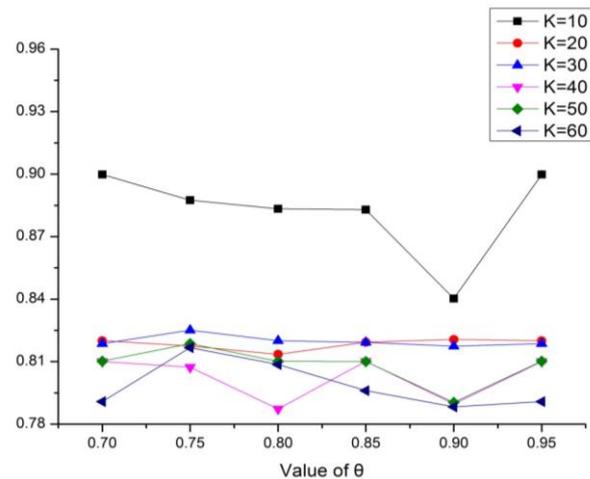


Fig 1. MAE under different K

In Fig.2, the mean value of the MAE is calculated to confirm that the value of θ in the user trusts similarity formula, when MAE takes the minimum value, $\theta = 0.9$.

After the θ value is selected, the parameter λ_C in the joint probabilistic matrix factorization is determined by calculation. When it is determined that the factorization dimension is $d=10$, Different MAE values corresponding to λ_C in Fig.3, when MAE takes the minimum value, $\lambda_C = 10$.

In Fig.2, the mean value of the MAE is calculated to confirm that the value of θ in the user trusts similarity formula, when MAE takes the minimum value, $\theta = 0.9$.

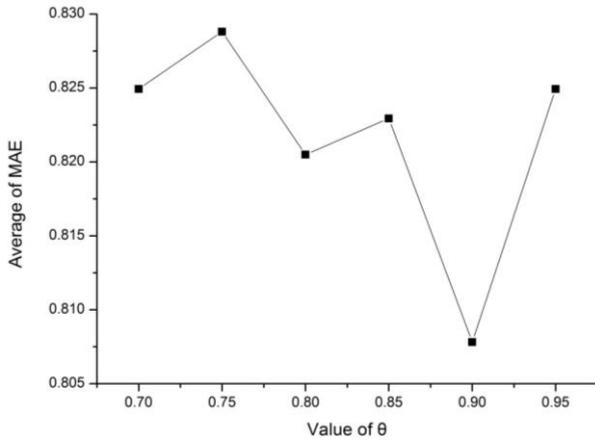


Fig. 2. Average of MAE

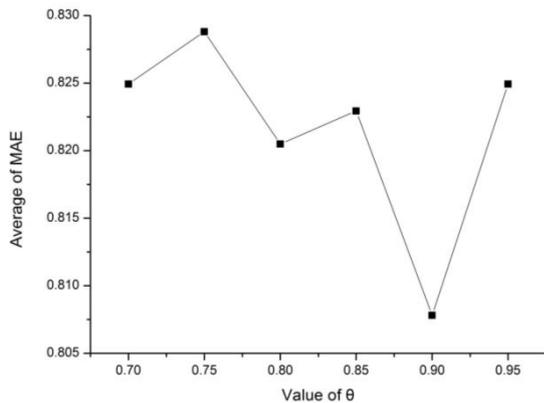


Fig. 3. Value of MAE at different λ_c .

After the θ value is selected, the parameter λ_c in the joint probabilistic matrix factorization is determined by calculation. When it is determined that the factorization dimension is $d=10$, Different MAE values corresponding to λ_c in Fig.3, when MAE takes the minimum value, $\lambda_c = 10$.

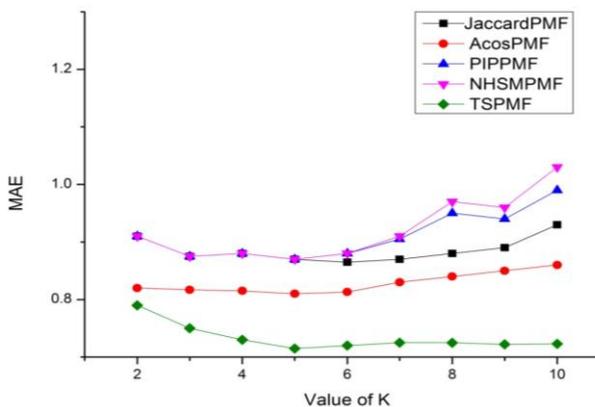


Fig. 4. MAE under different methods.

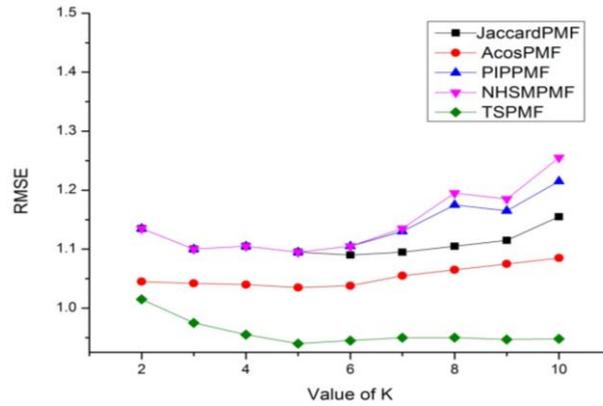


Fig. 5. RMSE under different methods

Experimental results show, the TSPMF algorithm proposed in this paper is obviously better than the other four algorithms of the same type. Prove that the method of integrating user trust to make recommendations is effective. It improves the consequent of the recommendation.

5 Conclusion

In this paper, a method of user trust probability matrix factorization (TSPMF) is proposed. To integrate the trust relationship between users into the traditional matrix factorization. Not only did it get accurate user similarity, but it further enhanced the effectiveness of the recommendations through trust relationships between users. The comparison experiment results show that the trust relationship between users plays an important role in the user's recommendation. The recommendation effect of TSPMF method is more stable and better than the others.

References

1. Bobadilla J, Ortega F, Hernando A. Recommender systems survey[J]. Knowledge-Based Systems, **46**(1):109-132(2013).
2. Sarwar, Badrul Munir, et al. "Item-Based Collaborative Filtering Recommendation Algorithms." Proceedings of the 10th International Conference on World Wide Web, pp. 285–295(2001)
3. Wang H, Wang N, Yeung D Y. Collaborative Deep Learning for Recommender Systems[J].1235-1244(2014)
4. Patra B K, Launonen R, Ollikainen V, et al. Exploiting Bhattacharyya Similarity Measure to Diminish User Cold-Start Problem in Sparse Data[C]// International Conference on Discovery Science. Springer, Cham, 252-263(2014)
5. Patra B K, Launonen R, Ollikainen V, et al. A new similarity measure using Bhattacharyya coefficient

- for collaborative filtering in sparse data[J]. Knowledge-Based Systems, **82**(C):163-177(2015)
6. Berkovsky S, Taib R, Dan C. How to Recommend?: User Trust Factors in Movie Recommender Systems[C]// International Conference on Intelligent User Interfaces. ACM, 287-300(2017)
 7. YIN Gui-sheng, ZHANG Ya-nan, DONG Yu-xin, HAN Qi-long, A Constrained Trust Recommendation Using Probabilistic Matrix Factorization, [J], Acta Electronica Sinica, 42(5):904-911.(2014)
 8. Ahn H J. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem[J]. Information Sciences, 178(1):37-51(2008)
 9. Liu H, Hu Z, Mian A, et al. A new user similarity model to improve the accuracy of collaborative filtering [J]. Knowledge-Based Systems, **56**(3):156-166(2014)
 10. Pirasteh P, Hwang D, Jung J J. Exploiting matrix factorization to asymmetric user similarities in recommendation systems[J]. Knowledge-Based Systems, 83(1):51-57(2015)
 11. Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks[C]// ACM Conference on Recommender Systems. ACM, 135-142(2010)
 12. Salakhutdinov R, Mnih A. Probabilistic Matrix Factorization [C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 1257-1264(2007)