



A Recommendation Algorithm Considering User Trust and Interest

Chuanmin Mi¹✉, Peng Peng¹, and Rafał Mierziak^{1,2}

¹ Nanjing University of Aeronautics and Astronautics,
Nanjing 210016, People's Republic of China
michuanmin@nuaa.edu.cn

² Poznan University of Technology, Poznan, Poland

Abstract. A traditional collaborative filtering recommendation algorithm has problems with data sparseness, a cold start and new users. With the rapid development of social network and e-commerce, building the trust between users and user interest tags to provide a personalized recommendation is becoming an important research issue. In this study, we propose a probability matrix factorization model (STUIPMF) by integrating social trust and user interest. First, we identified implicit trust relationship between users and potential interest label from the perspective of user rating. Then, we used a probability matrix factorization model to conduct matrix decomposition of user ratings information, user trust relationship, and user interest label information, and further determined the user characteristics to ease data sparseness. Finally, we used an experiment based on the Epinions website's dataset to verify our proposed method. The results show that the proposed method can improve the recommendation's accuracy to some extent, ease a cold start and solve new user problems. Meanwhile, the STUIPMF approach, we propose, also has a good scalability.

Keywords: Data mining · Recommender system
Collaborative filtering · Social trust · Interest tag
Probability matrix factorization

1 Introduction

With the expansion of network and information technology, the amount of data generated from human activities is rapidly growing. "Information Overload" problem is becoming serious [2]. Therefore, a recommender system might be an important tool to help users find the interested items and solve the problem of information overload. More and more e-commerce service providers, such as Amazon, Half. Com, CDNOW, Netflix, and Yahoo!, are using recommendation

This work was supported by the Project of National Social Science Foundation of China (17BGL055).

systems for their own customers with “tailored” buying advice [19]. A recommended algorithm as the most core and key part of these types of systems, determines their performance quality to a great extent [1]. Due to a simple operation, a reliable explanation, easiness of realization in the technical level, and collaborative filtering (CF), a recommendation algorithm is becoming one of the most widely used recommendation algorithm [4], which mainly makes uses of the ratings of the user to calculate the similarity to give recommendation. However, studies have shown that in big e-commerce systems, items rated by users generally will not exceed 1% in total number, so user rating inevitably has problems such as data sparseness, a cold start etc., which affect the precision and quality of the recommendation [21]. On the one hand, introducing a user trust relationship in a recommender system can solve a cold start problem. On the other hand, adding user interest can alleviate the problem of data sparseness.

In recent years, the number of Internet users has exponentially grown and as a result social network also has developed. The 38th China Internet development statistics report issued by China Internet Network Information Center (CNNIC) on August 3rd 2016 in Beijing shows that, the scale of Internet users in China has reached 710 million up to June 2016, whereas the Internet penetration rate reached 51.7%. Nielsen research agency examined factors that influence users’ trust in recommendations. The investigation showed that nearly ninety percent of users can believe the recommendation given by their friends [17].

According to the facts presented above, we established an implicit trust relationship between users and potential interest tags from the perspective of user rating. Next, we combined users trust relationship and user interest tag information into a probability matrix factorization (PMF) model. Finally, on the basis of PMF model, we proposed a probability matrix factorization model (STUIPMF) by integrating social trust and user interest.

2 Related Works

In a traditional CF recommendation algorithm there are problems with data sparseness and a cold start, which affect the precision and quality of a recommendation algorithm based on social trust [4, 6, 10, 11, 18]. In order to improve the performance of a recommender system, it ought to be roughly divided into two categories. One is a recommended approach, which examines a trust relationship based on a neighborhood model. Here we can distinguish a Mole Trust model, which utilizes a depth-first strategy to search users and predicts the trust value to target user B by considering the passing of the trust on the side of user A’s social network [18]. Similarly, Golbeck proposed a Tidal Trust model that improves a breadth-first strategy to forecast a user trust value [4], whereas Jamali proposed a Trust Walker model, which combines a recommender system based on an item with a recommender system based on trust [6]. However, these methods only consider a trust relationship between the neighboring users, but also neglect an implicit trust relationship between users, and the influence user ratings exerted to the result of recommendation.

The second category is a recommended approach, which fuses a trust relationship among users with rating data based on MF model. Take for example a recommendation method that adds a social regularization term to a loss function, which measures the difference between the latent feature vector of a user and those of their friends [16]. On the other hand, referring to Social MF model, it integrates all user trust information, which introduces a concept of trust propagation. Moreover, the model considers information about direct trustable users and “two-steps” users to generate recommendation. However, computing complexity is high, and it does not adopt different trust metrics [7]. Another example is a MF recommendation method that predicts the variation of ratings with time [13]. Next proposal is a stratified stochastic gradient descent (SSGD) algorithm to solve general MF problem. It provides sufficient conditions for convergence [3]. Finally, we can also find an incremental CF recommendation method based on regularized MF [14] and a Social Recommendation method, which connects users rating information and social information for research by sharing user implicit feature vector space [15]. All the methods described above focus on direct trust network, however they ignore mining implicit trust relationships between users.

When analyzing research referring to MF models, it is worth to notice that a recommendation algorithm based on MF uses latent factors. Thus, it is difficult to give an accurate and reasonable explanation to recommended results. Hence, Salakhutdinov described a matrix factorization problem from the perspective of probability, and put forward PMF model, which obtained the prior distribution of the user-recommended item’s characteristic matrix, and maximized the posterior probability of the forecast evaluation to make recommendations [20]. This model achieved very good prediction results on Netflix data sets. It is worth to mention that Koenigstein integrated some characteristic information of the items in the process of probability matrix factorization, and carried out experiments on the Xbox movie recommendation system, which verified the effectiveness of the proposed model [9].

Research also has shown that we take into account user interests, such as tags, categories and user profiles, there is a huge opportunity to improve the accuracy of recommendation. Considering the user interest model, it is conducive to make more accurate personalized recommendation. Lee combined user’s preference information with a trust propagation of social networking, and improved the quality of recommendation [5], while Tao proposed a CF algorithm based on user interest classification adapting to the user’s interests diversity. After that the improved fuzzy clustering algorithm was used to search the nearest neighbor [22]. What is more, Ji put forward a similarity measure method based on user interest degree. This way the combination of a degree of user interest in the different item category and user ratings were utilized to calculate the similarity between them [18]. However, most of these methods focus on the user’s rating value of the item. They do not consider user preferences and influence on the relationship between the user ratings and the item properties that affect the accuracy of recommendation. Furthermore, they also ignored the user trust relationships.

Therefore, in this paper, we have a comprehensive consideration of user rating and implicit trust relationship between users. Moreover, users trust relationship and user interest tag information on the basis of PMF model are introduced. And then we identify latent user characteristics hidden behind a trust relationship and user ratings. As a result, a STUIPMF model is proposed. According to the experimental results, this method comprehensively utilizes various information, which can enhance recommended accuracy.

3 Probability Matrix Factorization Recommendation Algorithm Combining User Trust and Interests

3.1 Probability Matrix Factorization Model (PMF)

The principle of PMF model is to predict user ratings for an item from the perspective of probability. To make the notation clearer, the symbols that we will use are shown in Table 1. The calculation process of PMF is as follows.

Assume that latent factors of users and items are subject to Gaussian prior distribution,

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

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

Table 1. Notation

Symbols	Descriptions
M, N, S	Number of users, number of items, number of interest labels respectively
K	Number of latent factors
$U_{M \times K}$	User latent factors
$V_{N \times K}$	Item latent factors
$F_{M \times K}$	Trust latent factors
$L_{S \times K}$	Interest tag latent factors
R_{ij}	Rating matrix
P_{ik}	Tagging times
T_{il}	Trust degree
\tilde{R}_{ij}	Predicted rating

Moreover, let's assume that conditional probability of user ratings data obtained are subject to Gaussian prior distribution,

$$P(R | U, V, \sigma_R^2) = \prod_{i=1}^M \prod_{j=1}^N [N(R_{ij} | g(U_i^T V_j), \sigma_R^2)]^{I_{ij}^R} \quad (3)$$

I_{ij}^R is an indicator function, if user U_i has rated V_j , $I_{ij}^R = 1$, otherwise 0. $g(x)$ maps the value of $U_i^T V_j$ to the interval, in this paper $g(x) = 1/(1 + e^{-x})$.

Through the Bayesian inference, we can gain posterior probability of users and items' implicit characteristics.

$$P(U, V | R, \sigma_R^2, \sigma_U^2, \sigma_V^2) \propto P(R | U, V, \sigma_R^2) \times P(U | \sigma_U^2) \times P(V | \sigma_V^2) \quad (4)$$

This way, we can learn about latent factors of users and items through rating matrix, and then get the most similar user rating by means of inner product formulated as follows:

$$\tilde{R} \approx U_i^T V_j \quad (5)$$

The corresponding probability graph model is presented in Fig. 1.

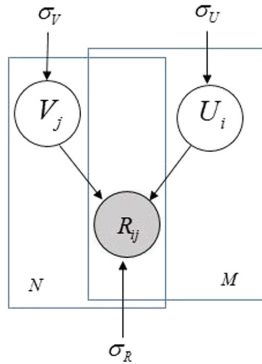


Fig. 1. Probability matrix factorization graph model

3.2 Mining User Implicit Trust Relationships

Most existing algorithms only consider a direct trust network, namely the dominant trust relationship between users [6,10,18]. They have less attention to mining user implicit trust relationship. Therefore, a user behavior coefficient and a user trust function are introduced to improve the measurement of user trust relationships.

After a trust inference based on user rating and calculation of rating accuracy, a user behavior coefficient will be determined. Next, user implicit trust relationships will be established on the basis of user rating similarity. Accuracy of rating is denoted by the difference of the item rating between a target user and all the users. In general, whether user ratings are accurate or not will directly affect the degree of other users' trust. A user behavior coefficient is expressed with φ_u symbol and it is depended on accuracy of rating.

$$\varphi_u = \frac{1}{1 + \sum_{i=1}^N (R_{ui} - \bar{R}_i) \cdot I_{ui}} \quad (6)$$

R_{ui} expresses rating of user u to the item i . \bar{R}_i expresses an average rating of all users to the item i if user u has rated the item i , $I_{ui} = 1$, otherwise $I_{ui} = 0$.

Rating similarity $sim_{i,j}$ is measured by a popular Pearson correlation coefficient. The computational formula is as follows:

$$sim_{i,j} = \frac{\sum_{c \in U} (r_{i,c} - \bar{r}_i) \times (r_{j,c} - \bar{r}_j)}{\sqrt{\sum_{c \in U} (r_{i,c} - \bar{r}_i)^2 \times \sum_{c \in U} (r_{j,c} - \bar{r}_j)^2}} \quad (7)$$

$r_{i,c}$ and $r_{j,c}$ express ratings of user i and user j to the item c respectively, and \bar{r}_i and \bar{r}_j express the average.

User implicit trust relationships are denoted by TI , between user i and user j :

$$TI_{ij} = \varphi_i \cdot sim_{ij} \quad (8)$$

Use t_{ij} to denote the explicit trust relationships between user i and user j , when user i trusts user j , $t_{ij} = 1$, otherwise 0. Due to the asymmetry of trust, t_{ij} cannot reflect a dominant trust relationship between users accurately, which should be related to the number of users' trust and trustable users. For example, when user t_i trusts many users, a trust value t_{ij} between user t_i and user t_j will be reduced. On the contrary, when many users trust user t_i , the trust value between user t_i and user t_j should increase. Therefore, the dominant trust value between users is upgraded on the basis of user influence TE_{ij} , which expresses the improved dominant trust value.

$$TE_{ij} = \sqrt{\frac{d^-(u_i)}{d^+(u_j) + d^-(u_i)}} \cdot t_{ij} \quad (9)$$

$d^-(u_i)$ points out the number of users u_i by a user who is trusted, $d^+(u_j)$ is the number of users by user u_j trust.

A user trust function is denoted by T_{ij} , which is calculated after determining the weight coefficient of a dominant trust and implicit trust combined with a dominant trust relationship stated in trust network. α expresses a weight coefficient.

$$T_{ij} = \alpha \cdot TE_{ij} + (1 - \alpha) \cdot TI_{ij} \quad (10)$$

A user trust relationship matrix is denoted by T . T_{il} expresses the trust degree of user U_i and a friend F_l . A conditional probability distribution function of a user's trust is known as:

$$P(T | U, F, \sigma_T^2) = \prod_{i=1}^M \prod_{l=1}^M [N(T_{il} | g(U_i^T F_l), \sigma_T^2)]^{I_{il}^T} \quad (11)$$

I_{il}^T is an indicator function if user U_i and user F_l are friends, $I_{il}^T = 1$, otherwise 0.

The probability distribution of U and F is as follows:

$$P(U | \sigma_U^2) = \prod_{i=1}^M N(U_i | 0, \sigma_U^2 I), P(F | \sigma_F^2) = \prod_{l=1}^M N(F_l | 0, \sigma_F^2 I) \quad (12)$$

Through the Bayesian inference, we achieve:

$$P(U, F | T, \sigma_T^2, \sigma_U^2, \sigma_F^2) \propto P(T | U, F, \sigma_T^2) \times P(U | \sigma_U^2) \times P(F | \sigma_F^2) \quad (13)$$

The corresponding probability graph model based on a user trust relationship is demonstrated in Fig. 2.

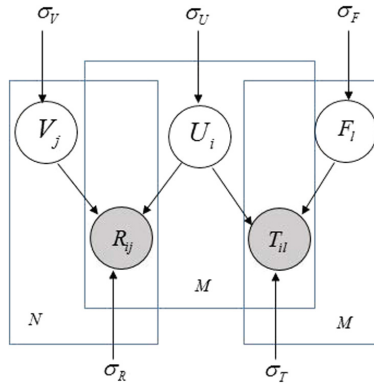


Fig. 2. Probability graph model based on user trust relationship

3.3 Mining the User Interest Similarity Relationship

A current recommendation algorithm based on user interest classification pays less attention to the influence, which user preference and the relationship between user ratings and item properties have, on recommended results [8, 9, 12, 23]. Thus, it is legitimated to combine the item information with user threshold based on the user-item rating matrix, and mining user implicit tag. As a result, a user-interest tag matrix is received and it is useful to fill user information and solve the problem of data sparseness.

The corresponding median rating threshold set to user rating set of all the items is $A = \{A_1, A_2, \dots, A_m\}$, the attribute set of items is $L = \{L_1, L_2, \dots, L_k\}$. When $R_{ui} \geq A_i$, we regard that user u likes the item i . The attribute tag L_c of the item i is signed as an interest tag of the user u . We can extract the user's interest tag according to the item's attributes and the user's rating threshold. A user may be signed with the same interest tags repeatedly, and when the times are accumulated, we can get the user interest tag matrix $L_{me} = \{L_{uy}\}$. L_{uy} expresses the times of the interest tag user u signed to the item attributes L . Then, we make the rating which is below the user ratings threshold

0 to get a user-item median rating matrix. Combined with the item-attribute matrix, if the item belongs to some attribute, it is signed 1, otherwise 0. Therefore, when a link is established between the user and the item-attribute, we will get the user-interest tag matrix P .

The user-interest tag matrix is denoted by P , P_{ik} expresses the signed times of user U_i signed on the interest tag L_k . The probability distribution function of a user interest tag is known as follows:

$$P(P | U, L, \sigma_P^2) = \prod_{i=1}^M \prod_{k=1}^Q [N(P_{ik} | g(U_i^T L_k), \sigma_P^2)]^{I_{ik}^P} \quad (14)$$

I_{ik}^P expresses an indicator function, if the user U_i has signed on the interest tag L_k at least one time, otherwise 0.

The probability distribution of U_i and L is the following way:

$$P(U | \sigma_U^2) = \prod_{i=1}^M N(U_i | 0, \sigma_U^2 I), \quad P(L | \sigma_L^2) = \prod_{k=1}^Q N(L_k | 0, \sigma_L^2 I) \quad (15)$$

According to the Bayesian inference, we achieve:

$$P(U, L | P, \sigma_P^2, \sigma_U^2, \sigma_L^2) \propto P(P | U, L, \sigma_P^2) \times P(U | \sigma_U^2) \times P(L | \sigma_L^2) \quad (16)$$

The corresponding probability graph model based on a user interest tag is shown in Fig. 3.

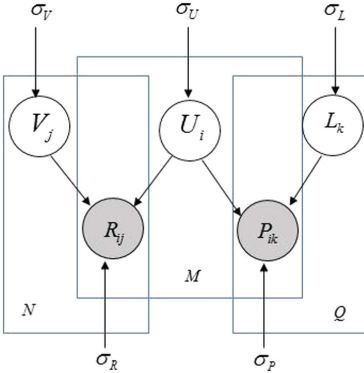


Fig. 3. Probability graph model based on user interest tag

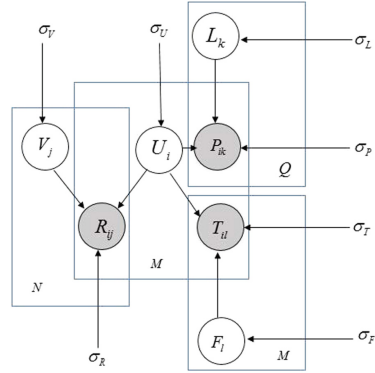


Fig. 4. STUIPMF probability graph model

4 STUIPMF Model Application

PMF algorithm is merely based on a user-item rating matrix and it studies the corresponding feature factor. However, it does not consider the trust relationship between the user and the user's interest on the result of recommendation. In order to reflect the effect, the model was improved by integrating the factorizations of three matrixes, which are a user trust relationship matrix, a user-interest tag matrix, and a user rating matrix respectively, and connected by a user latent feature factor matrix. Therefore, STUIPMF model is put forward, as it is demonstrated in Fig. 4.

The logarithm of posterior probability, after the conjunction, comes down to the Eq. (17).

In this research, a stochastic gradient descent method is used to study a corresponding latent feature factor matrix. Assuming that $\lambda_U = \lambda_V = \lambda_T = \lambda_L = \lambda$, a computational complexity is reduced. The values of λ_P and λ_F will be discussed in the latter part.

$\lambda_P = \sigma_R^2/\sigma_P^2$, $\lambda_T = \sigma_R^2/\sigma_T^2$, $\lambda_U = \sigma_R^2/\sigma_U^2$, $\lambda_V = \sigma_R^2/\sigma_V^2$, $\lambda_L = \sigma_R^2/\sigma_L^2$, $\lambda_F = \sigma_R^2/\sigma_F^2$ are all fixed regularization parameters, $\|\cdot\|_F$ expresses the Frobenius of the matrix.

$$\begin{aligned}
& \ln P(U, V, L, F \mid R, P, T, \sigma_R^2, \sigma_P^2, \sigma_T^2, \sigma_U^2, \sigma_V^2, \sigma_L^2, \sigma_F^2) \\
&= -\frac{1}{2\sigma_R^2} \sum_{i=1}^M \sum_{j=1}^N I_{ij}^R (R_{ij} - g(U_i^T V_j))^2 - \frac{1}{2\sigma_P^2} \sum_{i=1}^M \sum_{k=1}^Q I_{ik}^P (P_{ik} - g(U_i^T L_k))^2 \\
&\quad - \frac{1}{2\sigma_T^2} \sum_{i=1}^M \sum_{l=1}^M I_{il}^T (T_{il} - g(U_i^T F_l))^2 - \frac{1}{2\sigma_U^2} \sum_{i=1}^M U_i^T U_i - \frac{1}{2\sigma_V^2} \sum_{j=1}^N V_j^T V_j \\
&\quad - \frac{1}{2\sigma_L^2} \sum_{k=1}^Q L_k^T L_k - \frac{1}{2\sigma_F^2} \sum_{l=1}^M F_l^T F_l - \frac{1}{2} \left(\sum_{i=1}^M \sum_{j=1}^N I_{ij}^R \right) \ln \sigma_R^2 \\
&\quad - \frac{1}{2} \left(\left(\sum_{i=1}^M \sum_{k=1}^Q I_{ik}^P \right) \ln \sigma_P^2 + \left(\sum_{i=1}^M \sum_{l=1}^M I_{il}^T \right) \ln \sigma_T^2 \right) - \frac{1}{2} ((M \times K) \ln \sigma_U^2 \\
&\quad + (N \times K) \ln \sigma_V^2) - \frac{1}{2} ((M \times K) \ln \sigma_F^2 + (Q \times K) \ln \sigma_L^2) + C \tag{17}
\end{aligned}$$

$$S(U, V, L, F, R, P, T)$$

$$\begin{aligned}
&= \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N I_{ij}^R (R_{ij} - g(U_i^T V_j))^2 + \frac{\lambda_P}{2} \sum_{i=1}^M \sum_{k=1}^Q I_{ik}^P (P_{ik} - g(U_i^T L_k))^2 + \\
&\quad \frac{\lambda_T}{2} \sum_{i=1}^M \sum_{l=1}^M I_{il}^T (T_{il} - g(U_i^T F_l))^2 + \frac{\lambda_U}{2} \sum_{i=1}^M \|U\|_F^2 + \frac{\lambda_V}{2} \sum_{j=1}^N \|V\|_F^2 \\
&\quad + \frac{\lambda_L}{2} \sum_{k=1}^Q \|L\|_F^2 + \frac{\lambda_F}{2} \sum_{l=1}^M \|F\|_F^2 \tag{18}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial S}{\partial U_i} &= \sum_{j=1}^N I_{ij}^P g' (U_i^T V_j) (g (U_i^T V_j) - R_{ij}) V_j \\
&\quad + \lambda_P \sum_{k=1}^Q I_{ik}^P g' (U_i^T L_k) (g (U_i^T L_k) - P_{ik}) L_k \\
&\quad + \lambda_T \sum_{l=1}^M I_{il}^T g' (U_i^T F_l) (g (U_i^T F_l) - T_{il}) F_l + \lambda_U U_i
\end{aligned} \tag{19}$$

$$\frac{\partial S}{\partial V_j} = \sum_{i=1}^M I_{ij}^R g' (U_i^T V_j) (g (U_i^T V_j) - R_{ij}) U_i + \lambda_V V_j \tag{20}$$

$$\frac{\partial S}{\partial L_k} = \lambda_P \sum_{i=1}^Q I_{ik}^P g' (U_i^T L_k) (g (U_i^T L_k) - P_{ik}) U_i + \lambda_L L_k \tag{21}$$

$$\frac{\partial S}{\partial F_l} = \lambda_T \sum_{i=1}^M I_{il}^T g' (U_i^T F_l) (g (U_i^T F_l) - T_{il}) U_i + \lambda_F F_l \tag{22}$$

U_i, V_j, L_k, F_l are adjusted in each iteration as follows: $U_i \leftarrow U_i - \gamma \cdot \frac{\partial S}{\partial U_i}$, $V_j \leftarrow V_j - \gamma \cdot \frac{\partial S}{\partial V_j}$, $L_k \leftarrow L_k - \gamma \cdot \frac{\partial S}{\partial L_k}$, $F_l \leftarrow F_l - \gamma \cdot \frac{\partial S}{\partial F_l}$. γ is a predefined step length.

A repeated training process, after each iteration, calculates and validates an average absolute error. When the change of the objective function S value is smaller than a predefined small constant iterative process is terminated. After obtaining a terminated iteration U_i, V_j, L_k, F_l , we can predict the user U_i unknown rating to the item V_j . To each target user, a proposed commodity is sorted from high to low according to a calculated predicting rating, and then Top-N recommended list is produced.

5 Experiment and the Analysis of the Results

Dataset in this research is provided from the studies conducted by Massa and Avesani [18] and “Epinions.com” website, since it is among the most often-used datasets for evaluating trust inference performance. Due to the fact that a trust system was built, it expresses the trust relationship between the users and helps the users determine whether to trust the comments of the item [5, 10, 11]. Statistics concerning this dataset is presented in Table 2.

Commonly used evaluation indexes, namely MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error) were adopted to evaluate the accuracy of the prediction, and then compare the effect of our proposed algorithm with models proposed in literature i.e. PMF model [20], SocialMF model [7], and SoReg [15].

The assignments of λ_P, λ_F are crucial in the proposed method, which plays the role of balance. When we assign $\lambda_P = 0$, the system only considers the user rating matrix and an implicit interest tag. When it recommends something, it

Table 2. Characteristics of dataset

Dataset	Epinions
Number of user	49290
Number of item	139738
Number of rating	664813
Number of trust relationship	487181
Number of interest tag	154

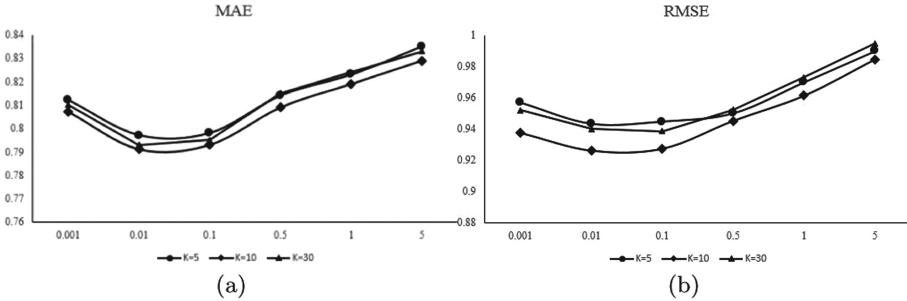


Fig. 5. The influence of parameter λ_P on MAE and RMSE

does not consider the trust relationship between users. If we assign high values to λ_P , the system only recognizes the trust relationships between users, but when recommending, it does not analyze other factors. Similarly, when $\lambda_F = 0$, the system only examines the user rating matrix and the trust relationship between users, however, when recommending, it does not deal with an implicit interest tag of users. When λ_F is enormous, the system only studies the implicit interest tag of users when recommending, not considering other factors.

Figure 5 shows the influence of a parameter λ_P on MAE and RMSE when the number of latent factors are 5, 10 and 30, and other parameters set constant. With the increase of λ_P , MAE and RMSE decrease, namely the accuracy of the prediction is improved. When λ_P reaches a certain threshold with the increase of λ_P , MAE and RMSE increase, namely the accuracy of the prediction is reduced. In conclusion, when $\lambda_P \in [0.01, 0.1]$, the accuracy of recommendation is higher. In latter experiments, we adopt the interval average $\lambda_P = \lambda_F = 0.005$ as the approximate optimal value to conduct an experiment. Figure 6 shows that the influence of a parameter λ_F has similarly more details.

In order to verify the experimental effect, we choose 80% of the whole data as the training set and the remaining 20% of the data constitutes the test set. Recommendations are generated on the basis of known information in the training set. Subsequently, the test set is used to evaluate the performance of recommendation algorithms [1, 5, 12]. Respectively, 90% of the whole data is the training set, and 10% of the remaining data is the test set for experiments.

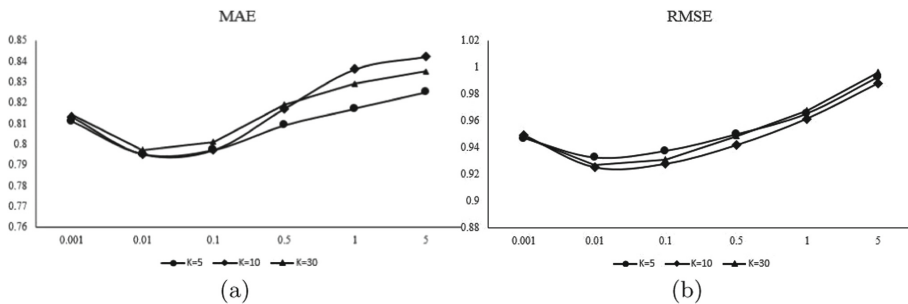


Fig. 6. The influence of parameter λ_F on MAE and RMSE

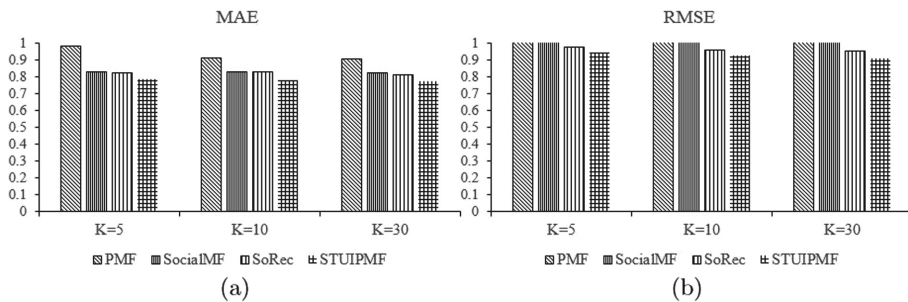


Fig. 7. Comparison of STUIPMF method and other methods under 80% training set

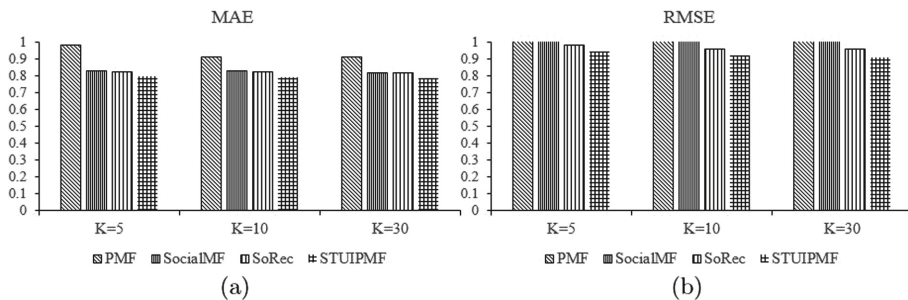


Fig. 8. Comparison of STUIPMF method and other methods under 90% training set

In the experimental process, relevant parameters are selected mainly according to the experimental results for the optimal choice. The parameters' settings in STUIPMF are as follows: $\lambda_U = \lambda_V = \lambda_T = \lambda_L = \lambda = 0.001$, $\lambda_P = \lambda_F = 0.005$. The numbers of latent factors are 5, 10 and 30 respectively.

The parameters' settings in other methods are as follows: in PMF model $\lambda_U = \lambda_V = 0.001$, in Social MF model $\lambda_U = \lambda_V = 0.001$, $\lambda_T = 0.5$, in SoReg model $\lambda_U = \lambda_V = 0.001\alpha = 0.1$. The comparison of experimental results of a STUIPMF method with other methods is presented in Figs. 7 and 8.

According to Figs. 7 and 8, we can come up with the following conclusions, namely

- (1) STUIPMF model, we proposed, comprehensively considers the user rating information, user's trust and interest in the case of all experimental parameters chosen optimally. When 80% is the training set, 20% is the test set, then compared with PMF, Social MF, and SoReg, MAE has reduced 17%, 5.8%, 5.3% respectively and RMSE has reduced 21%, 13%, 4% respectively. When 90% is the training set, 10% is the test set, then compared with PMF, Social MF, and SoReg, MAE has reduced 16.2%, 4.1%, 3.7% respectively and RMSE has reduced 20.8%, 13.5%, 4.1% respectively. Therefore, taking into account the analyzed data, the proposed method has improved the recommendation accuracy.
- (2) With the increase of latent factors' dimensions, the accuracy of recommendation has improved, but on the other hand, there may be fitting problems. Moreover, computational complexity has increased.
- (3) The probability matrix factorization of a user's trust relationship matrix and an interest tag matrix can increase the prior information of user characteristics, so as to solve the problems of a cold start and new users in recommender systems to a large extent.

6 Conclusions and Further Works

With the status and importance of the personalized service in modern economics and social life, it is increasingly prominent to accurately grasp the user's real interests and requirements through user's behavior. What is more, providing high quality personalized recommendation has become the current necessity. Taking into consideration a cold start and data sparseness problems in traditional CF method, we proposed STUIPMF model by integrating a social trust and user interest. We studied an implicit trust relationship between users and potential interest tags from the perspective of user rating. Next, we used PMF model to conduct MF of user ratings information, users trust relationship, and user interest tag information. In result, we analyzed the user characteristics to use data and generate more accurate recommendations. Our proposed method was verified with the use of an experiment based on representative data. The results showed that STUIPMF can improve the recommendation accuracy, make a cold-start easier and solve new user problems to some extent. Meanwhile, it occurred that the STUIPMF approach also has good scalability.

However, our research has revealed many challenges for further study. Take for example, the value λ we used in the model is the approximate optimal value, thus we will determine the optimal value λ and dynamic value changes to improve accuracy of recommendation. In the further research, we are going to verify the effects of the proposed algorithm for new users and for new items in detail. In addition, we will consider adding more information into the proposed model, e.g. text information, location information, time, etc., and pay more attention to the update of the user trust and interest. What is more, we will recognize a conjunction of the distrust relationship between users into the proposed model.

References

1. Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A.: Recommender systems survey. *Knowl.-Based Syst.* **46**, 109–132 (2013)
2. Borchers, A., Herlocker, J., Konstan, J., Reidl, J.: Ganging up on information overload. *Computer* **31**(4), 106–108 (1998)
3. Gemulla, R., Nijkamp, E., Haas, P.J., Sismanis, Y.: Large-scale matrix factorization with distributed stochastic gradient descent. In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 69–77. ACM (2011)
4. Golbeck, J.: Personalizing applications through integration of inferred trust values in semantic web-based social networks. In: *2005 Proceedings on Semantic Network Analysis Workshop*, Galway, Ireland (2005)
5. Guo, G., Zhang, J., Zhu, F., Wang, X.: Factored similarity models with social trust for top-N item recommendation. *Knowl.-Based Syst.* **122**, 17–25 (2017)
6. Jamali, M., Ester, M.: Trustwalker: a random walk model for combining trust-based and item-based recommendation. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 397–406. ACM (2009)
7. Jamali, M., Ester, M.: A matrix factorization technique with trust propagation for recommendation in social networks. In: *Proceedings of the Fourth ACM Conference on Recommender Systems*, pp. 135–142. ACM (2010)
8. Kim, H., Kim, H.-J.: A framework for tag-aware recommender systems. *Expert Syst. Appl.* **41**(8), 4000–4009 (2014)
9. Koenigstein, N., Paquet, U.: Xbox movies recommendations: variational Bayes matrix factorization with embedded feature selection. In: *Proceedings of the 7th ACM Conference on Recommender Systems*, pp. 129–136. ACM (2013)
10. Lee, W.P., Ma, C.Y.: Enhancing collaborative recommendation performance by combining user preference and trust-distrust propagation in social networks. *Knowl.-Based Syst.* **106**, 125–134 (2016)
11. Li, J., Chen, C., Chen, H., Tong, C.: Towards context-aware social recommendation via individual trust. *Knowl.-Based Syst.* **127**, 58–66 (2017)
12. Lim, H., Kim, H.-J.: Item recommendation using tag emotion in social cataloging services. *Expert Syst. Appl.* **89**, 179–187 (2017)
13. Lu, Z., Agarwal, D., Dhillon, I.S.: A spatio-temporal approach to collaborative filtering. In: *Proceedings of the Third ACM Conference on Recommender Systems*, pp. 13–20. ACM (2009)
14. Luo, X., Xia, Y., Zhu, Q.: Incremental collaborative filtering recommender based on regularized matrix factorization. *Knowl.-Based Syst.* **27**, 271–280 (2012)
15. Ma, H., Yang, H., Lyu, M.R., King, I.: SoRec: social recommendation using probabilistic matrix factorization. In: *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pp. 931–940. ACM (2008)
16. Ma, H., Zhou, D., Liu, C., Lyu, M.R., King, I.: Recommender systems with social regularization. In: *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, pp. 287–296. ACM (2011)
17. Ma, H., Zhou, T.C., Lyu, M.R., King, I.: Improving recommender systems by incorporating social contextual information. *ACM Trans. Inf. Syst. (TOIS)* **29**(2), 9 (2011)
18. Massa, P., Avesani, P.: Trust-aware recommender systems. In: *Proceedings of the 2007 ACM Conference on Recommender Systems*, pp. 17–24. ACM (2007)

19. Mi, C., Shan, X., Qiang, Y., Stephanie, Y., Chen, Y.: A new method for evaluating tour online review based on grey 2-tuple linguistic. *Kybernetes* **43**(3/4), 601–613 (2014)
20. Mnih, A., Salakhutdinov, R.R.: Probabilistic matrix factorization. In: *Advances in Neural Information Processing Systems*, pp. 1257–1264 (2008)
21. Sun, X., Kong, F., Ye, S.: A comparison of several algorithms for collaborative filtering in startup stage. In: *2005 IEEE Proceedings of Networking, Sensing and Control*, pp. 25–28. IEEE (2005)
22. Tao, J., Zhang, N.: Similarity measurement method based on user's interesting-ness in collaborative filtering. *Comput. Syst. Appl.* **20**(5), 55–59 (2011)
23. Zuo, Y., Zeng, J., Gong, M., Jiao, L.: Tag-aware recommender systems based on deep neural networks. *Neurocomputing* **204**, 51–60 (2016)