CR-IIA:Collaborative Recommendation Algorithm Based on Implicit Item Associations

Pengfei He, Li Yu, Jingru Li and Chaozhun Wen
School of Electronic Information and Communications,
Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
Email: hepengfei@hust.edu.cn, hustlyu@mail.hust.edu.cn, li1jingru@163.com, U201313686@hust.edu.cn

***Abstract*—Recommendation system provides prediction of users' preference and guidance for users' purchase. Probabilistic matrix factorization (PMF) as a widely-used collaborative filtering approach so far, generally has excellent performance for personalised recommendation. Nevertheless, it still suffers from data sparsity problem. In this paper, we present a novel recommendation algorithm CR-IIA which combines matrix factorization and association rule mining. Our method employs *ItemCorrelation* to measure the latent correlation among items given by association rule mining. And *ItemCorrelations* are integrated into PMF model to optimize feature vectors of items. The factor of *ItemCorrelations* makes rating prediction more accurate and realistic. Compared with social trust model and context-based model, our approach improves the accuracy of ratings prediction without aggravating data sparsity problem caused by additional information. Moreover, our framework also allows additional dimensionality to be considered simultaneously, which benefits subsequent research in personalised recommendation based on matrix factorization. We conduct experiments on a real-world data set MovieLens, and the results show that our model outperforms the item-oriented approaches.**

I. Introduction

With the explosive development of electronic commerce businesses in the last decade, the items (movies, music, news, etc.) information flood has almost inundated users. Thus, recommendation system (RS), which is the mechanism that predicts users' tastes and offers items recommendation by the known preference of users, become more and more crucial. Some well-established recommender systems have been applied to business already, such as music recommendation at iTunes[1] and news recommendation at Yahoo!News[2], etc.

There mainly are two types of techniques for item recommendation: content-based methods and collaborative filtering (CF) methods. As to the content-based recommendation system, it utilizes the description of items and the profile of users to predict one's preference and meet his need. On the other hand, CF is a popular technique that predicts user's taste by leveraging a historical user-item ratings matrix. It can be further classified into two categories: memory-based methods and model-based methods. A typical memory-based model is K Nearest Neighbour model (KNN), which is the baseline method in RS. Among the model-based models, PMF is the most widely-used approach, and it maps both users and items to a joint latent factor space of dimensionality D. PMF was firstly proposed by Ruslan et al. [1], and it constructed a framework allowing the incorporation of additional information [2]. Although it has been proven that CF can yield good results in the industry, traditional CF still suffers from extreme sparsity of user-item matrix, and this problem leads to great difficulty in finding similar users.

For the weakness mentioned above, additional information, such as social trust information [3], [4], [5] and contextual information [6], [7], [8], are incorporated into PMF to enhance the recommendation accuracy. Social recommendation approaches are proposed based on the intuition that users usually have similar tastes to their trusted friends [9]. In [4], the linked users across social networking sites and e-commerce websites were taken into account for item recommendation. The authors seek to build a bridge to map users' social networking features to another feature representation. As regards studies on contextual information, time information was fused into similarity measurement and QoS prediction for web service recommendation in [8]. Despite phenomenal performance of the methods mentioned above, additional dimensionality not only aggravates data sparsity problem and exponential-increased computational complexity, but also confronts a tough challenge that the acquisition of additional information is not easy. Since there is always no social relations in majority of application areas and the amount of users in most e-commerce website is numerous [10].

In traditional methods, users are limited to see items similar to those already rated high scores, which is termed as lack of diversity [11]. Among previous researches, recommendation techniques based on association rule mining (ARM) have solved this problem efficiently through identifying transaction patterns in large data set. ARM is used to investigates latent relations among items rather than only comparing item similarities[12], [13]. However, recommendation methods based on ARM are usually incorporated with memory-based model, where ratings matrix is not fully utilized.

For the purpose of alleviating the problems mentioned above and facilitating the performance of recommendation system, we present a novel algorithm integrating ARM into PMF model. ARM is adopted to supplement and replace vacant and impractical relations among items where necessary. The relations among items in association rules are named as *ItemCorrelation*. It is obtained by means of applying hyperbolic tangent function. Finally, *ItemCorrelation* is added to objective

[1]http://www.apple.com/itunes/
[2]https://www.yahoo.com/news/

978-1-5090-5019-2/17/$31.00 ©2017 IEEE

function processed by stochastic gradient descent (SGD) to optimize feature vectors of users and items.

The contributions of our work in this paper are as follows:

- We present a metric *ItemCorrelation* to measure the latent correlation among items given by ARM. The quantified correlation is obtained through a defined correlation formula via applying hyperbolic tangent function.
- We propose a mechanism to denote the feature vectors of items tuple, which means we can take relation of multiple items into consideration when applying SGD. It makes item feature vectors more consistent with actual situations.
- We introduce a novel method CR-IIA which incorporates ARM with PMF model to address the issue of data sparsity and achieve better recommendation accuracy than traditional methods and item-oriented approaches.

The remainder of this paper is organized as follows: In section II, we introduce basic techniques. And section III presents the proposed method. The experiment results and analysis are demonstrated in section IV, which is followed by the conclusions and future work in section V.

II. Preliminary

A. Probabilistic Matrix Factorization

Typically in a CF recommendation system, there is a set of users and items respectively and a user-item rating matrix $R = [r_{ui}]_{M \times N}$, where r_{ui} denotes the rating of user u for item i. Let $U \in R^{D \times M}$ and $V \in R^{D \times N}$ be latent user and item feature matrices, with column vectors U_u and V_i representing the D-dimensional latent vectors of user u and item i, respectively. The target of matrix factorization is to obtain latent feature vectors by means of performing SGD, and apply them to ratings prediction [2].

In PMF model, the researchers adopt a probabilistic linear model with Gaussian observation and define the conditional distribution over the observed ratings as follow:

$$p\left(R|U,V,\sigma^2\right) = \prod_{u=1}^{M}\prod_{i=1}^{N}\left[\mathcal{N}\left(r_{ui}|U_u^T V_i, \sigma_R^2\right)\right]^{I_{ui}}. \quad (1)$$

Where $\mathcal{N}\left(x|\mu,\sigma^2\right)$ is a Gaussian distribution representing the probability density function with mean μ and variance σ^2, I_{ui} is the indicator function that is equal to 1 if user u rated item i and equal to 0 otherwise. The zero-mean spherical Gaussian priors are placed on user and item feature vectors:

$$p\left(U|\sigma_U^2\right) = \prod_{u=1}^{M}\mathcal{N}\left(U_u|0,\sigma_U^2 I\right), \quad (2)$$

$$p\left(V|\sigma_V^2\right) = \prod_{i=1}^{N}\mathcal{N}\left(V_i|0,\sigma_V^2 I\right). \quad (3)$$

Next, the posterior probability of the latent variables U and V can be derived through a Bayesian inference.

$$\begin{aligned} &p\left(U,V|R,\sigma_R^2,\sigma_U^2,\sigma_V^2\right) \propto p\left(R|U,V,\sigma_R^2\right)p\left(U|\sigma_U^2\right)p\left(V|\sigma_V^2\right) \\ &= \prod_{u=1}^{M}\prod_{i=1}^{N}\left[\mathcal{N}\left(r_{ui}|U_u^T V_i,\sigma_R^2\right)\right]^{I_{ui}} \times \prod_{u=1}^{M}\mathcal{N}\left(U_u|0,\sigma_U^2 I\right) \\ &\times \prod_{i=1}^{N}\mathcal{N}\left(V_i|0,\sigma_V^2 I\right). \end{aligned} \quad (4)$$

We illustrate the corresponding graphical model in Fig.1(a). Using log function on Eq. (4), then we can infer that maximizing the posterior probability term equals to minimizing the objective function as follows:

$$\begin{aligned} \mathcal{E} = &\frac{1}{2}\sum_{u=1}^{M}\sum_{i=1}^{N} I_{ui}\left(r_{ui} - U_u^T V_i\right)^2 + \frac{\lambda_U}{2}\sum_{u=1}^{M}\|U_u\|^2 \\ &+ \frac{\lambda_V}{2}\sum_{i=1}^{N}\|V_i\|^2. \end{aligned} \quad (5)$$

Where $\lambda_U = \sigma^2/\sigma_U^2$, $\lambda_V = \sigma^2/\sigma_V^2$, and we can obtain a local minimum of the objective function from Eq. (5). Hence, the latent feature vectors U_u, V_i can be learnt, for any rating of user u on item i ,we can predict it by calculating $U_u^T V_i$.

B. Association Rule Mining

Association rule mining is investigated for discovering regularities between products in large-scale transactions recorded by supermarkets. Let $L = \{l_1, l_2, \cdots, l_n\}$ be a set of binary attributes called *items*. Let $DB = \{t_1, t_2, \cdots, t_m\}$ be a set of transactions called *database*. And each transaction t contains a subset of items in L where $t \subseteq L$. Each association rule is composed by two different sets of items (antecedent X and consequent Y), known as *itemsets*. If *itemset* Y has a large probability to show up when X presents, an association rule is defined as :$X \Rightarrow Y$, where $X, Y \subseteq L$.

The metrics of association rule are *support(sup)* and *confidence(conf)*, which reflect frequency and certainty of a rule respectively. The *support* of $X \Rightarrow Y$ refers to as the proportion of transactions t in the database DB which contain both X and Y.

$$support\left(X \Rightarrow Y\right) = \frac{|\{(X \cup Y) \subseteq t; t \in DB\}|}{|DB|}. \quad (6)$$

Confidence is an indication of how much it is possible for transactions involving Y under the circumstance of involving X.

$$confidence\left(X \Rightarrow Y\right) = \frac{support\left(X \Rightarrow Y\right)}{support\left(X\right)}. \quad (7)$$

III. CR-IIA Model

In this section, we firstly introduce item similarity regularization based on PMF. Then ARM is utilized to find strong association rules (SARs), where *ItemCorrelation* is computed in the way that wields hyperbolic tangent function. Finally *ItemCorrelation* is fused into a unified rating prediction model.

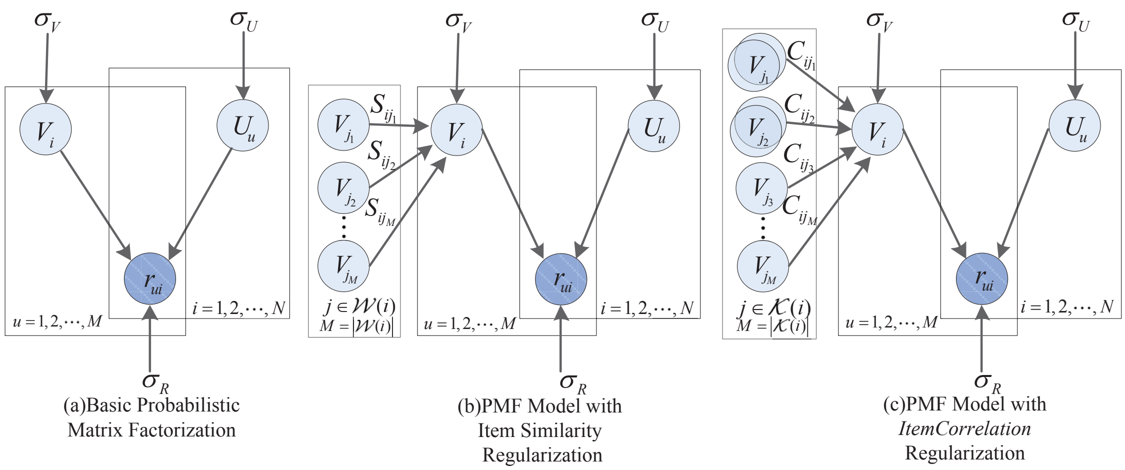


Fig. 1. Graphic model.

A. Item Similarity Regularization

In order to enhance recommendation accuracy, abundant researches pay attention to modelling social regularization. Motivated by user social regularization [9] and [14], it can be naturally extended to leverage implicit item similarities for regularization. The constraint term added into objection function is supposed to be effective and sensitive to both similar and dissimilar items. Thus the objective function of item similarity regularization is formulated as:

$$\mathcal{E} = \frac{1}{2}\sum_{u=1}^{M}\sum_{i=1}^{N} I_{ui}\left(r_{ui} - U_u^T V_i\right)^2 + \frac{\lambda_U}{2}\sum_{u=1}^{M}\|U_u\|^2 + \frac{\beta}{2}\sum_{i=1}^{N}\sum_{j\in\mathcal{W}(i)} S_{ij}\|V_i - V_j\|^2 + \frac{\lambda_V}{2}\sum_{i=1}^{N}\|V_i\|^2. \tag{8}$$

Where $\mathcal{W}(i)$ is furtherly divided into $\mathcal{W}^+(i)$ and $\mathcal{W}^-(i)$. They represent the set of top-N similar and dissimilar items to item i respectively. S_{ij} denotes the similarity between item i and item j. The similarity of any two items is calculated by measuring co-rated items between them, otherwise set to 0 if there is none of co-rated items. In this paper, we employ Pearson Correlation Coefficient (PCC) to calculate items similarity. The PCC formula is as follows:

$$S_{ij} = \frac{\sum\limits_{q\in U(i)\cap U(j)} (r_{qi} - \bar{r}_i)\cdot(r_{qj} - \bar{r}_j)}{\sqrt{\sum\limits_{q\in U(i)\cap U(j)} (r_{qi} - \bar{r}_i)^2}\cdot\sqrt{\sum\limits_{q\in U(i)\cap U(j)} (r_{qj} - \bar{r}_j)^2}}. \tag{9}$$

Where $\bar{r}_i$ denotes the average rating of item i. Item similarity S_{ij} ranges from $[-1, 1]$, and the larger S_{ij}, the more similar for item i and j. $U(i)$ represents a set of users who has rated item i. Furthermore, we map the range of PCC similarities into $[0, 1]$ by mapping function $f(x) = (x+1)/2$. The graphical model of PMF with item similarity regularization is illustrated in Fig.1(b).

B. $ItemCorrelation$ Matrix

Based on the similarity obtained from PCC, we can take into account impact of similar and dissimilar items through adding item similarities regularization into object function. However, the essence of item relation is not only similarity between two items, but also includes latent correlation among multiple items. More specifically, if a user purchases bread and butter simultaneously, milk is more similar to the product tuple $\langle bread, butter\rangle$ than waffle and cheese, which are similar to bread and butter respectively. Therefore milk should be prior to be recommended. Motivated by this intuition, we adopt ARM to dig the correlation among items so that recommendation can be more in line with reality.

Definition 1. Let C_{xy} denotes quantified correlation value, where x and y represent a set of items respectively. we name C_{xy} as *ItemCorrelation*, it is used to indicate how much the quantified correlation is between x and y.

Aiming at calculating correlation among multiple items, Apriori algorithm is applied to yield SARs under the condition of empirical *minsup* and *minconf*. While obtained SARs always contain the rules like $\{i_1, i_2, i_3\} \Rightarrow \{i_4, i_5\}$, which is multiple items to multiple items and cannot be utilized directly. So obtained SARs should be split at first.

Assuming that *itemset* X and Y are denoted as $X = \{x_1, x_2, \cdots, x_d\}, d \in [1, N]$ and $Y = \{y_1, y_2, \cdots, y_e\}, e \in [1, N]$, thus splitting formula for SARs is:

$$X \Rightarrow Y = \begin{cases} x_1, x_2, \cdots, x_e \Rightarrow y_e \\ \qquad Y \in \mathrm{multi-element\ set, e} \in (1, N] \\ x_1, x_2, \cdots, x_d \Rightarrow Y \\ \qquad Y \in \mathrm{single-element\ set, d} = 1 \end{cases}. \tag{10}$$

Where $X \subseteq L, Y \subseteq L$, SARs actually are transformed to the rules whose form is $X \Rightarrow Y'$, where sconsequent Y' is totally split into single-element set.

When item similarity regularization applied into PMF model, we can derive an initial item similarity matrix (ISM) from employing PCC. As shown in Fig.2(a), each value S_{ij}

represents the similarity between item i and j. The values on the diagonal are set to 1.

$$\begin{array}{c} \\ \text{item 1} \\ \text{item 2} \\ \cdots \\ \text{item n} \end{array}\begin{array}{c}\begin{array}{cccc} \text{item 1} & \text{item 2} & \cdots & \text{item n} \end{array} \\ \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{bmatrix}\end{array} \Rightarrow \begin{array}{c}\begin{array}{ccccccc} \text{item 1} & \text{item 2} & \cdots & \text{item n} & \text{item n+1} & \cdots & \text{item n+i} \end{array} \\ \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} & C_{1n+1} & \cdots & C_{1n+i} \\ C_{21} & C_{22} & \cdots & C_{2n} & C_{2n+1} & \cdots & C_{2n+i} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} & C_{nn+1} & \cdots & C_{nn+i} \end{bmatrix}\end{array}$$

(a)ISM (b)IRM

Fig. 2. Transformation form ISM to IRM.

As we can see from ISM, if X is a single-element set, the correlation among X and Y' can be compared and updated directly in ISM by calculating *ItemCorrelation* formula (11), which will be put forward below. However, if X is a multi-element set, there is no column corresponding to the relation between tuple of items and single item. Thus, a modified item relation matrix (IRM) is proposed and demonstrated in Fig.2(b). The additional columns $n+i$ corresponds to tuples of original items. For example, I_{n+1} is made of I_2 and I_3, obviously the C_{1n+1} is denoted as the correlation between I_1 and the tuple $\langle I_2, I_3 \rangle$. In this paper we regard item tuples I_{n+i} as virtual combinations of items.

The *ItemCorrelation* for SAR is proposed to quantify how strong the association rule is. In order to calculate *ItemCorrelation* C_{xy}, hyperbolic tangent function is adopted to fuse sup and $conf$ to *ItemCorrelation* calculation. The *ItemCorrelation* calculation formula is as follow:

$$C_{XY'} = \frac{\sum\limits_{x_i \in X} S_{x_i Y'}}{|X|} \times \left(1 + \frac{e^t - e^{-t}}{e^t + e^{-t}}\right). \tag{11}$$

Where $X \Rightarrow Y', Y' \subseteq I$ is the association rule after split, x_i is the single element of antecedent, and consequent Y' is single-element set. t is denoted as the sum of sup and $conf$ for a specific SAR. Besides, $C_{XY'}$ is ranged from 0 to $C_{XY'}(t_{\max})$, we map $C_{XY'}$ into $[0,1]$ by mapping function $f(x) = x/C_{XY'}(t_{\max})$. Hence, IRM can be extended according Eq. (11) and lay the foundation for *ItemCorrelation* regularization.

C. A Unified Model

From subsection III.A to III.B, we describe traditional item similarity regularization and the approach of obtaining *ItemCorrelation* among items given by SARs. In this section, we propose our model that integrates ARM into PMF model based on item similarity regularization to enhance accuracy of prediction.

The fused objection function is demonstrated as follow:

$$\begin{aligned} \mathcal{E} = & \frac{1}{2}\sum_{u=1}^{M}\sum_{i=1}^{N} I_{ui}\left(r_{ui} - U_u^T V_i\right)^2 + \frac{\lambda_U}{2}\sum_{u=1}^{M}\|U_u\|^2 \\ & + \frac{\beta}{2}\sum_{i=1}^{N}\sum_{j\in\mathcal{K}(i)} C_{ij}\|V_i - V_j\|^2 + \frac{\lambda_V}{2}\sum_{i=1}^{N}\|V_i\|^2. \end{aligned} \tag{12}$$

Where C_{ij} represents the *ItemCorrelation* of item i and item j, or item tuple j. $\mathcal{K}(i)$ is the top-N similar and dissimilar items according to values in IRM. The corresponding graphical model for proposed model is illustrated in Fig.1(c).

By performing SGD, a local minimum of fused objection function given by Eq. (12) can be obtained as follow:

$$\frac{\partial\mathcal{E}}{\partial U_u} = \sum_{u=1}^{M} I_{ui}(U_u^T V_i - r_{ui})V_i + \lambda_U U_u, \tag{13}$$

$$\begin{aligned} \frac{\partial\mathcal{E}}{\partial V_i} = & \sum_{i=1}^{N} I_{ui}(U_u^T V_i - r_{ui})U_u + \lambda_V V_i \\ & + \beta_1 \sum_{i=1}^{N}\sum_{g\in\mathcal{K}^+(i)} C_{iq}\left(V_i - \frac{\sum\limits_{q\in f(g)} V_q}{|f(g)|}\right) \\ & + \beta_2 \sum_{i=1}^{N}\sum_{h\in\mathcal{K}^-(i)} C_{ip}\left(V_i - \frac{\sum\limits_{p\in f(h)} V_p}{|f(h)|}\right). \end{aligned} \tag{14}$$

Where $\mathcal{K}^+(i)$ and $\mathcal{K}^-(i)$ represent top-N similar and dissimilar items of item i, respectively. By the way, feature vector V_{n+i} of item tuple is denoted as the average of feature vectors that composes it. Thus, $f(g)$ represents the items that compose item tuple V_g, $|f(g)|$ is the amount of items in V_g. Equivalently for $f(h)$, we don't explain any more here. In order to reduce the model complexity, we set $\lambda_U = \lambda_V$ when we conduct experiment. In each iteration, U and V are updated based on the latent variables from the previous iteration.

IV. Experiment and Analysis

In this section, the description of dataset MovieLens is given at first. Then we introduce some preparatory work for our experiments. Finally, we report our experimental results which compare with several relevant methods and the impact of some parameters is presented behind.

A. Dataset Description

Our experiments are conducted on a real-world dataset MovieLens[1]. The experimental dataset is a widely-used dataset of ratings to movies, which includes 943 users who rated a total of 1682 different movies. There are 100,000 ratings contained in the dataset MovieLens (ranged from 1 to 5), a larger value of rating indicates users show more interest on the movies and vice versa. What's more, the experimental dataset has been cleaned up for practical research. Users who had less than 20 ratings were removed from this dataset. Sparsity of experimental dataset is more than 93.69%.

More statistics of the experimental dataset is shown in Table I. The average number of ratings given by per user is 106 and the average number of ratings issued on per item is 59.

[1]https://grouplens.org/datasets/movielens/100k/

TABLE I
STATISTICS OF DATASET MOVIELENS.

Statistics	Users	Items
Max. num. of ratings	737	583
Min. num. of ratings	20	1
Avg. num. of ratings	106	59

B. Experiment Setup

Five-fold cross validation is performed in our experiments. In each fold, 80% of data is applied as training set and the remaining 20% as test set. We use mean absolute error (MAE) and root mean square error (RMSE) as evaluation metrics for measuring the prediction accuracy.

$$MAE = \frac{\sum |r_{ui} - \hat{r}_{ui}|}{|N_{test}|}, \tag{15}$$

$$RSME = \sqrt{\frac{\sum (r_{ui} - \hat{r}_{ui})^2}{|N_{test}|}}. \tag{16}$$

Where r_{ui} denotes the real rating of user u on item i, $\hat{r}_{ui}$ represents the corresponding pair (u, i) predicted rating. N_{test} is the amount of ratings in test dataset. Obviously from the definitions, the smaller MAE or RMSE is, the better performance recommendation system has.

In order to apply Apriori algorithm, transactions need to be extracted from ratings matrix. As to a specific user, we collect his/her ratings record and convert the value to 1 if the rating is greater than his/her average rating, or 0 otherwise. Then we regard users' converted ratings record vectors as the sets of transaction, and search for association rules that meet threshold *minsup* and *minconf* among them.

When selecting top-N similar and top-N dissimilar items, aiming at reducing the noises we adopt rules as follows: the similarity between two top-N similar items should be greater than 0.75 and the similarity between top-N dissimilar items should be less than 0.25.

In our experiments, we compare the results with some general recommendation methods, including baseline methods: user-oriented KNN (UKNN), item-oriented KNN (IKNN) and several relevant methods: PMF and PMF model with item similarity regularization (Item-PMF), which are introduced detailedly in [1] and [14].

C. Experimental Result

We set $\lambda_U = \lambda_V = 0.1$, $\beta_1 = 0.04$, $\beta_2 = 0.02$ and the learning rate $\alpha = 0.01$. Top-N similar and dissimilar items number N (neighbor numbers) is set to 5, 10 respectively. Dimensionality D is conducted in 10 and 30. Association rule mining parameters *sup* and *conf* are 0.13 and 0.6.

As the comparison shown in Table II, we can summarise following conclusions:

- For those methods which leveraging neighborhood information, such as IKNN, UKNN, Item-PMF, CR-IIA, the prediction quality of recommendation system is influenced significantly by the amount of neighbors. Observed from results above, the performance improves as the neighbor quantity increases from 5 to 10. On the other hand, if the numbers of neighbors surpass a certain value, it will cover more global efforts and cause high computational complexity when more information of neighbors is considered.
- Generally, the dimensionality D controls the flexibility of items and users in latent space, a larger value of D will make items and users feature vector more concrete and consistent with the behaviors in reality. For matrix factorization models like PMF, Item-PMF and CR-IIA, no matter what neighbors quantity is, the prediction error diminishes with the increase of dimensionality from 10 to 30. However, the increase of D leads to the fact that learned hypothesis fits training data set so well that fails to generalize to test data set, which is called overfitting.
- Our proposed model obtains higher accuracy on prediction of users' preference than the corresponding recommendation system models, which verifies our initial intuition that applying association rule mining to measure implicit items correlation could improve RS quality. Besides, it can be obviously seen that the model-based approaches outperform the memory-based ones under all conditions, which is in line with previous researches.

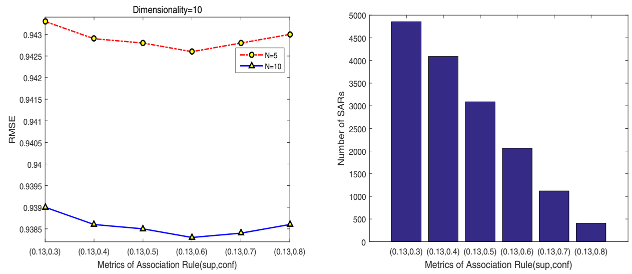
TABLE II
PERFORMANCE COMPARISON.

Dimensionality	Neighborhood	Metric	UKNN	IKNN	PMF	Item-PMF	CR-IIA
D=10	N=5	MAE	0.8373	0.8309	0.7489	0.7426	**0.7415**
		Improve[1]	11.44%	10.76%	0.99%	0.15%	
		RMSE	1.0639	1.0638	0.9547	0.9437	**0.9426**
		Improve	11.40%	11.39%	1.27%	0.12%	
	N=10	MAE	0.8011	0.7981	0.7489	0.7411	**0.7403**
		Improve	7.59%	7.24%	1.15%	0.11%	
		RMSE	1.0171	1.0202	0.9547	0.9392	**0.9383**
		Improve	7.75%	8.03%	1.72%	0.10%	
D=30	N=5	MAE	0.8373	0.8309	0.7513	0.7416	**0.7405**
		Improve	11.56%	10.88%	1.44%	0.15%	
		RMSE	1.0639	1.0638	0.9532	0.9387	**0.9375**
		Improve	11.88%	11.87%	1.65%	0.13%	
	N=10	MAE	0.8011	0.7981	0.7513	0.7389	**0.7382**
		Improve	7.85%	7.51%	1.74%	0.09%	
		RMSE	1.0171	1.0202	0.9532	0.9337	**0.9329**
		Improve	8.28%	8.56%	2.13%	0.09%	

[1]Improve indicates that the increased percentage of CR-IIA outperforms other methods.

D. Impact of Parameter t

The parameter t denotes the sum of *sup* and *conf*, it controls the numbers of SARs and value of *ItemCorrelation*, thereby affecting prediction accuracy. If the value of t is set to relatively smaller, the relation between association rules is weak and there will be tens of thousands meaningless rules obtained, which not only reduces the efficiency of processing but can't reflect real impact of latent item relations. Whereas, if we employ a relatively larger value of t, the quantities of SARs meeting the larger *sup* and *conf* are negligible. Majority of unprevailing items are unable to be excavated out.



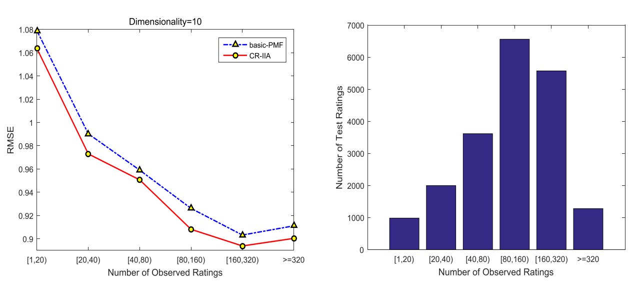
(a) RMSE Comparisons (b) Distribution of SARs

Fig. 3. Performance comparisons on different confidence

Fig.3 illustrates the RMSE of our model in different range of values t. We set *conf* scaled from 0.3 to 0.8 with *sup* fixed, the curve in left panel reveals the results with D=10 and N=5 and 10 respectively. From distribution of SARs, we can observe the quantities of SARs gradually cut down in according with our intuition.

E. Performance on Different Users

In this subsection, we conduct the experiments on different user groups classified according the number of ratings observed in training set. The users are grouped into 6 classes based on how many ratings they have rated in training set, and evaluated prediction errors of different classes of users. The 6 classes include: [1,20), [20,40), [40,80), [80,160), [160,320) and over 320. For the purpose of interpreting trend clearly, the method for comparison contains the basic PMF model, where there is none of additional constraint term added into objective function. The other parameters setting of this experiment are as follows: D=10, N=10, *sup*=0.13 and *conf*=0.6, and the parameters for SGD remain the same with former setting.



(a) RMSE Comparisons (b) Distribution of Test Data

Fig. 4. Performance comparisons on different users

From Fig.4, we can observe that as the number of ratings increases, prediction accuracy improves sharply for both models. Meanwhile, our model in different user groups outperforms PMF model, which validates the effectiveness of CR-IIA in different classes. Since our model simultaneously considers complementary roles of the *ItemCorrelations* among multiple items and the similarities between any two items, this makes our model captures more items relation information and achieve better prediction of ratings.

V. Conclusion

In this paper, we propose a novel approach integrating association rule mining into matrix factorization model. *ItemCorrelation* is defined to indicate the latent correlation among items given by ARM. On the basis of *ItemCorrelation*, we put forward a method to represent the feature vector of items tuple and consider the relation of multiple items simultaneously when latent matrix U and V iterated. And finally, a unified recommendation model fused with *ItemCorrelation* is proposed to predict the ratings of users' preference more realistic and accurate. Experiments conducted on a real-world dataset validate the effetiveness of our approach. Our future work intends to concentrate on digging implicit users' relationship by incorporating context information to furtherly enhance the accuracy of prediction.

References

[1] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *International Conference on Machine Learning*, Corvallis, USA, 2007, pp. 880–887.
[2] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
[3] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized recommendation combining user interest and social circle," *Knowledge and Data Engineering IEEE Transactions on*, vol. 26, no. 7, pp. 1763–1777, 2014.
[4] W. X. Zhao, S. Li, Y. He, and E. Y. Chang, "Connecting social media to e-commerce: Cold-start product recommendation using microblogging information," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 5, pp. 1147–1159, 2016.
[5] L. Sun, X. Wang, Z. Wang, H. Zhao, and W. Zhu, "Social-aware video recommendation for online social groups," *IEEE Transactions on Multimedia*, vol. 19, no. 3, pp. 609–618, March 2017.
[6] K. Niemann and M. Wolpers, "Creating usage context-based object similarities to boost recommender systems in technology enhanced learning," *IEEE Transactions on Learning Technologies*, vol. 8, no. 3, pp. 274–285, July 2015.
[7] Z. Yu, R. Wong, and C. H. Chi, "Efficient role mining for context-aware service recommendation using a high-performance cluster," *IEEE Transactions on Services Computing*, vol. PP, no. 99, pp. 1–1, 2015.
[8] Y. Hu, Q. Peng, X. Hu, and R. Yang, "Time aware and data sparsity tolerant web service recommendation based on improved collaborative filtering," *IEEE Transactions on Services Computing*, vol. 8, no. 5, pp. 782–794, Sept 2015.
[9] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Forth International Conference on Web Search and Web Data Mining*, Hong Kong, China, 2011, pp. 287–296.
[10] Q. Liu, C. Wang, and C. Xu, "A modified pmf model incorporating implicit item associations," in *IEEE International Conference on Tools with Artificial Intelligence*, Athens, Greece, 2012, pp. 1041–1046.
[11] C. H. Lee, Y. H. Kim, and P. K. Rhee, "Web personalization expert with combining collaborative filtering and association rule mining technique," *Expert Systems with Applications*, vol. 21, no. 3, pp. 131–137, 2001.
[12] U. Wanaskar, S. Vij, and D. Mukhopadhyay, "A hybrid web recommendation system based on the improved association rule mining algorithm," *Journal of Software Engineering and Applications*, vol. 6, no. 8, pp. 32–36, 2014.
[13] S. K. Yong and B. J. Yum, "Recommender system based on click stream data using association rule mining," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13 320–13 327, 2011.
[14] H. Ma, "An experimental study on implicit social recommendation," in *International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2013, pp. 73–82.