IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

SPECIAL SECTION ON RECENT COMPUTATIONAL METHODS IN KNOWLEDGE ENGINEERING AND INTELLIGENCE COMPUTATION

Received November 17, 2017, accepted December 22, 2017, date of publication December 28, 2017, date of current version September 7, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2787741

Field-Aware Matrix Factorization for Recommender Systems

ZHIYUAN ZHANG[1], YUN LIU[1], AND ZHENJIANG ZHANG[2]

[1]Key Laboratory of Communication and Information Systems, Department of Electronic and Information Engineering, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China

[2]School of Software Engineering, Beijing Jiaotong University, Beijing 100044, China

Corresponding author: Yun Liu (liuyun@bjtu.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 2018RC007, in part by the Fundamental Research Funds for the Central Universities under Grant W17JB00060, and in part by the National Natural Science Foundation of China under Grant 61271308.

ABSTRACT Predicting user response is one of the core machine learning tasks in recommender systems (RS). The matrix factorization (MF)-based model has been proved to be a useful tool to improve the performance of recommendation. Many existing matrix factorization-based models mainly rely on adding some side information into basic MF to enable the model to fully express the data. However, most of the side information is measured based on the statistics or empirical formula. Also, the latent features of side information cannot be deeply mined. In this paper, we focus on mining the influence of field information (useful side information) to improve the performance of prediction. Based on the MF framework, we propose a field-aware matrix factorization (FMF) model. In FMF, the interactions between user/item and field can be captured and learned in the latent vector spaces. We propose efficient implementations to train FMF. Then, we comprehensively analyze FMF and compare this model with the state-of-the-art models. The analysis of experiments on two large data sets demonstrates that our method is very useful in RS.

INDEX TERMS Recommender systems, matrix factorization, machine learning, field.

I. INTRODUCTION

Recommender Systems (RS) are computer applications and techniques for recommending specific items that may meet users' preference [1], [2]. Related recommendation techniques have been widely studied in researching communities of information retrieval [3], [4], machine learning [5]–[8] and data mining [9]–[11]. Among these approaches, the matrix factorization based models have attracted a lot of attention in the past decades. That is because upon most occasions, a user's preference (e.g., rating) about an item can be easily represented by the past user-item rating history matrix. Moreover, as a model-based approach, the matrix factorization based models can solve the limitations and disadvantages, such as data sparsity and cold start problem, to some extent by using the machine learning thought. In RS, ratings can be regarded as the most obvious explicit information, which can reflect users' interest in items. These ratings are filled into a big sparse matrix with one dimension representing users and one dimension representing items. Matrix factorization algorithm decompose the user-item matrix into a user latent factor matrix and an item latent factor matrix by using machine learning techniques. Finally, the learned two low-dimensional matrices will be utilized to predict the unknown ratings and to make recommendations.

Due to the data sparsity, dimension reduction technique in matrix factorization is able to improve model's ability of generalization. Therefore, it's easy to add various modeling elements and side information into the framework of MF. And these side information enables the model to fully express data. Over the past decades, plenty of previous research papers have been published to enhance the recommendation accuracy through incorporating some side information, e.g. social relations, into basic matrix factorization based models [1], [12]–[14]. However, there are some important limitations that should not be neglected, including: (1) Most of side information are measured based on statistics or empirical formula; (2) Trust/Distrust or social relations are mined according to similarity calculation, that is to say, these side information are directly defined before optimizing the objective function. We can get the conclusion that unstable data noise has existed before processing the optimization-based framework for MF. (3) field information is rarely considered,

TABLE 1. An artificial field-related user-movie data set.

field	user	relation	movie	field
Latent field 1	Tom		Roman Holiday	Romance
Latent field 2	Dustin		White Chicks	Comedy
Latent field 3	David		Final Destination	Horror
Latent field 2	Jack		Titanic	Romance
Latent field 4	Abby		Home Alone	Comedy

(4) all the side information are added into the regularization term to optimize the objective function, but the latent features of side information cannot be deeply mined. The machine learning method provides us a new point of view to solve the limitations of the method based on MF mentioned above. All the side information is not necessarily defined in a direct way. Learning the effect of side features seems to be crucial for MF based recommendations. Inspired by the machine learning method. We will pay attention to learn the influence of field information (one of side information) in this paper.

Considering the case of field features (most elements in recommender systems are either belonging to some fields or can be made to belong to some fields through discretization.) to answer the question that why mining the influence of field information is so important to improve recommendation. Here we give an illustration using the data set in Table 1. For most movie data sets like that in table 1, "movies" can be grouped in to "fields." In the example, five different "movies" belong to three "fields" and five different "users" belong to three "latent fields." For different users, speaking from experience, they may rate a same movie with different tastes. This assumption means different users' behaviors may be reflected by different latent factors or latent fields. More specifically, reconsidering the example in Table 1, users "Dustin" and "Jack" have watched and rated all the five movies. The phenomenon shows that "Dustin" and "Jack" belong to the same latent field, they love all kinds of movies and they have the same tastes. Different from "Dustin," user Tom has watched and rated two romance movies "Roman Holiday" and "Titanic," which means "Tom" belongs to another latent field, he has a passion for romantic movies. All the users and movies can be analyzed in a similar way. Field-aware MF proposed in this paper is a variant of MF that utilizes the field information. In traditional matrix factorization, the predicted rating R(Abby, Titanic)can be learned by dotting latent vector U(Abby) and V(Titanic). Both U(Abby) and V(Titanic) are single latent vectors related to user "Abby" and movie "Titanic." In this paper, considering the latent influence of fields, we seek to learn a set of latent vectors related to fields. Let's take the movie "Home Alone" for example, the learned latent vector will be V(Home Alone), (related latent field). For more details, please see section 4.

Our work significantly departs from previous works on definition or consistency analysis of social relations or trust/distrust information, and aims to effectively learn field information in matrix factorization for effective recommendation. In this paper, we propose a matrix factorization based model to learn the latent connections between fields and users or items. Main work and results of this paper are the followings:

- The field-aware matrix factorization learns the latent connections between users/items and the related fields.
- The field-aware matrix factorization is the promotion of basic matrix factorization.
- The field-aware matrix factorization is used on preference prediction problem. We conduct experiments to seek out the difference in terms of recommendation accuracy.
- The field-aware matrix factorization is a machine learning model. We present methods for training the proposed model to solve the optimization problem.

The rest of the paper is organized as follows: Section 2 includes a brief description of previous related work. Section 3 describes the traditional matrix factorization model. Section 4 presents the proposed field-aware matrix factorization model, in which the influence of field information is considered. Section 5 describes the evaluation procedure, and provides encouraging results. Finally, Section 6 gives conclusions and outlook for further research in this area.

II. RELATED WORK AND CONTRIBUTIONS OF THIS WORK

In this section, we review several popular previous RS approaches, inspired by which we construct our model, including: (1) matrix factorization based methods and (2) field-aware factor models.

A. MATRIX FACTORIZATION BASED METHODS

Normally, the matrix factorization based methods always utilize the observed user-item rating information to train a predefined learning model. The user-item matrix will be factorized into two low-dimensional matrices, and the two low-dimensional matrices which have been learned can be employed to make further predictions. Due to its efficiency in handling very huge datasets, matrix factorization-based methods have become one of the most popular models among the model-based methods, for example, Singular Value Decomposition (SVD) model [15], [16], Matrix Factorization (MF) [17]–[19], Probabilistic Matrix Factorization (PMF) [7], [20]–[22] and Non-negative Matrix Factorization (NMF) [23], [24] are all very effective matrix factorization based methods. The SVD model is a powerful technical of dimensionality reduction. PMF is a probabilistic linear model with Gaussian observation noise, which models the rating matrix as a product of two low-rank matrices (users and items). Probabilistic Sparse Matrix Factorization (PSMF) [25], Bayesian Probabilistic Matrix Factorization (BPMF) [7], [26], and General Probabilistic Matrix Factorization (GPMF) [27] are all the effective probabilistic models. Non-negative Matrix Factorization [24] (NMF) is also called non-negative matrix approximation, which is greatly developed by Lee and Seung. In NMF model, the original rating matrix is factorized into two matrices, with the property that all values of the three matrices are no-negative.

Those methods assume that user preferences can be modeled by only a small number of latent factors. Specifically, in the basic form of matrix factorization, the latent features of user and item can be modeled by factorizing the initial rating matrix into two low-dimensional user-specific and item-specific matrices. The way to learn a proper lower dimensional feature space and to catch the latent factors are the key problems of those methods. However, the latent connections between different fields of different users or items is difficult for traditional matrix models to learn because each learned latent user-specific or item-specific factor is only concerned with user or item itself. Therefore, in this paper, we expand the range of latent factors to catch latent connections between fields and items/users.

Although particular matrix factorization based models are able to generate high-quality recommendations, these approaches also suffer from the data sparsity problem in real-world scenarios and fail to address users who rated only a few items. In order to overcome these limitations, many previous literatures try to compensate for the lack of information in the rating matrix with other sources of side information, such as trust and distrust relations [28], [29] or social regularization [13], [30]. More specifically, users generally tend to connect with other users due to some commonalities they share, often reflected in similar interests. Moreover, in many real-life applications it may be the case that only social information about certain users is available while interaction data between the items and those users has not yet been observed yet. We note that all these methods are MF-based methods which employ only heuristic algorithms to handle the side information. All the mentioned side information will be added into the regularization term to optimize the objective function. However, the inner relations between user-item matrix and side related information have not been studied systematically, especially in a machine learning method. Because these methods need to calculate the pairwise user similarities and pairwise user trust scores based on statistic data. It is worth noting that seldom model pay attention to catch the influence of field information (side information for the initial rating matrix). In this paper, we propose a novel field-aware model to catch the influence of field information. The model is a matrix factorization based method and the field information will reflect a latent connection between field and the related items or users.

B. FIELD-AWARE FACTOR MODEL

Factorization Machine (FM) [31], [32] is a generic method, which can mimic most of factorization models just using feature engineering. In FM, the interactions between categorical variables are modeled and learned by applying factorization approaches. All useful feature information can be directly incorporated into the model, FM learns the effect of feature conjunction by factorizing it into a product of two latent vectors, then FM can be regarded as an extension of MF in the point of feature engineering. As generic approaches, both MF and FM lack detailed discussion on catching field information. By considering field information, an effective solution called Pairwise Interaction Tensor Factorization (PITF) is proposed in [33]. In PITF model, they assume three available fields including User, Item, and Tag, and factorize (User, Item), (User, Tag), and (Item, Tag) in separate latent spaces. Because PITF is limited to three specific fields, [34] generalize PITF for more fields and effectively apply it on POI prediction. However, both [33] and [34] learn field feature in separate matrix space. The internal connections between fields cannot be effectively captured. By using the "field" concept for reference, Field-aware Factorization Machine (FFM) is proposed in [35], which has been used to win two click-through rate prediction competitions hosted by Criteo and Avazu. Essentially, FFM is a variant of FM that utilizes field information. In FFM, features of the property can be grouped into "fields." Ech feature has several latent vectors. Depending on the field of other features, one of them is used to do the inner product. FFM can be treated as an efficiently tool on CTR prediction problem.

All the mentioned MF, PITF and FFM greatly motivates us to model the field feature relations into MF approach of RS. In our paper, the proposed FMF model is also the variant of MF that that utilizes field information, and it will be used to solve the rating prediction problem.

III. MATRIX FACTORIZATION FRAMEWORK

In this section, we will introduce the basic MF model and show how it works. What's more, we will clearly analyze why we can be inspired by the idea of basic MF.

Usually, the behaviors of users can be modeled in a big user-item matrix. Supposeing in a user-item rating matrix $R^{m\times n}$, $S = \{s_1, s_2, s_3, \ldots, s_n\}$ stands for a set of n users (n rows of matrix R) and $I = \{i_1, i_2, \ldots, i_m\}$ stands for a set of m items (m columns of matrix R). $R_{i,j}$ represent the rating value of user i for item j. The matrix was very sparse in most cases. So there are a mass of missing values in R. The problem we study in this paper is how to predict the missing values for the users effectively and efficiently by employing the initial user-item rating matrix. The method of MF to recommender system is to factorize the user-item matrix R to two low-dimensional matrices $U^{n\times f}$ and $V^{f\times m}$, Where $U^{n\times f}$ and $V^{f\times m}$ stand for user-latent and item-latent matrices,

TABLE 2. Toy example of initial user-movie-field data.

User	index	Field	index	Movie	index	Field	index
Tom	1	movie critic	1	Roman Holiday	1	romance	1
David	2	pastime	2	Home Alone	2	comedy	2
Jack	3	pastime	2	Titanic	3	romance	1
				Final Destination	4	horror	3

respectively, and f is f-dimensional specific potential feature of user and item, usually, $f << n, m$. We define the predicted matrix $\hat{R}$ can be computed by dotting the two learned latent matrices U and V, (i.e., $\hat{R}^{n\times m} = U^{n\times f}V^{f\times m}$, $\hat{R}^{n\times m} \approx R^{n\times m}$), then the missing values of R can be filled by $\hat{R}$ in every training iteration.

For user i and item j, the user-item interaction is modeled as an inner product in latent factor space, i.e., $\hat{R}_{ij} = U_i \cdot V_j$. Accordingly, each user i is associated with a latent vector U_i and each item j is associated with a latent vector V_j. The elements of U_i will control the extent of interest the user i has in items that are high on the corresponding factors, positively or negatively, the elements of V_j will control the extent to which the item j possesses those factors, again, positively or negatively. More specifically, for the given user i, his latent features can be learned and represented by a latent vector U_i with dimension f, for a given item j, its latent features can be learned and represented by a latent vector V_j, the dot product $U_i \cdot V_j$ captures the interaction between user i and item j – user i's interest for item j. All the mentioned phenomenon inspired us that if we expand the scope of the subjects being analyzed (i.e., more than just users and items), we may obtain more information in the latent space and capture deeper interaction between subjects.

In matrix factorization, the major challenge is to learn the latent matrix U and V, respectively. First of all, the factorization system minimizes the regularized squared error on the set of known ratings, the object function is given as:

$$\min\left[d = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(R_{ij} - \sum_{k=1}^{f} U_{ik}V_{kj})^2 + \frac{\lambda_U}{2}\|U\|_F + \frac{\lambda_V}{2}\|V\|_F\right], \tag{1}$$

where $\|\bullet\|$ is the Frobenius norm of a matrix, that is, $\|U\|_F = \sqrt{\sum_{a=1}^{n}\sum_{b=1}^{f}|U_{ab}|^2}$. $\frac{\lambda_U}{2}\|U\|_F$ and $\frac{\lambda_V}{2}\|V\|_F$ are regularization terms. The parameters λ_U and λ_V are regularizing coefficients for U and V respectively, which are used to prevent over-fitting.

Stochastic gradient decent is an effective method and then utilized to optimize Eq.(1). For this traditional MF problem, each desired element is obtained via the following training process:

$$\begin{cases} U_{ik} \overleftarrow{update} U_{ik} + \partial\left[(RV^T)_{ik} - (UVV^T)_{ik} - \lambda_U \cdot U_{ik}\right] \\ Q_{kj} \overleftarrow{update} Q_{kj} + \partial\left[(U^TR)_{kj} - (U^TUV)_{kj} - \lambda_V \cdot V_{kj}\right], \end{cases} \tag{2}$$

Where ∂ is the corresponding learning-rate. U_{ik} denotes the corresponding entry(rating) of U at row i and column k, V_{kj} denotes the corresponding rating of V at row k and column j. As mentioned before, these two parameters actually denote the kth latent feature of user s_i and the kth latent feature of item i_j, respectively.

IV. FIELD-AWARE MATRIX FACTORIZATION

A. MODEL

The idea of FMF originates from MF, FFM and PITF, which are described in section 2. In PITF, three efficient fields including user, item and tag are considered in the model. Then each pairwise matrix (user-item, user-tag, tem-tag) is factorized in different latent spaces. FFM have recently been established as a state-of-the-art method for CTR prediction by learning the multi-field features. However, PITF handle the field features in dispersed matrix spaces, FFM lack detailed discussion on MF framework. In this section we provide a more comprehensive study of FMF on learning the latent relations between missing values of sparse matrix.

To explain how FMF works, we consider the following new example. As stated in Table 2, we have three users and four movies. Each user or movie belongs to a related field. To make it easy to understand, each user/movie and field is marked with a number. (all the number are marked with different colors)

Let's take user 1,2 and item 3,4 as example, if user 2 has rated item 3, then we have $R_{2,3} = U_2V_3$, similarly, $R_{1,4}$ can be determined by U_1 and V_4. Then, we have learned the latent vector U_2 (related to user 2) and latent vector V_3 (related to movie 3). In MF, the preference of user 2 for movie 3 can be predicted by dotting U_2 and V_3. In this paper, if we consider the influence of field, the latent field-based vectors of user 2 and item 3 can be learned as $\mathrm{U}_{\mathbf{2-1}}$ and $\mathrm{V}_{\mathbf{3-2}}$, where the yellow color **1** represent item **3**'s related field number and the red color **2** is user **2**'s related field number.

From the above, the rating value R_{ij} that user i rated on item j can be learned by the following formula:

$$\hat{R}_{i,j} = U_i^{f_i}V_j^{f_j} + U_i^{f_j}V_j^{f_i}, \tag{3}$$

Where f_i and fj are the related field of i and j respectively, $U_i^{f_i}$ means the latent vector that user i in field f_i, $V_j^{f_j}$ means the latent vector that user j in field f_j, $U_i^{f_j}$ means the latent vector that user i in field f_j, $V_j^{f_i}$ means the latent vector that user j in field f_i. From a mathematical point of view, a big difference between traditional MF and FMF is: (1) In MF, for a target user i, if we want to predict his preference about item j in field 1 one and m in filed 2, we only need to learn a single latent vector U_i about user i. Note that MF neglect the influence of different filed. (2) However, for the same case, FMF will learn different latent vectors about user i according to different item's field. For a target item j, we can draw the same conclusion about the learned latent vector V_j. Figure 1 illustrates the above differences intuitively.

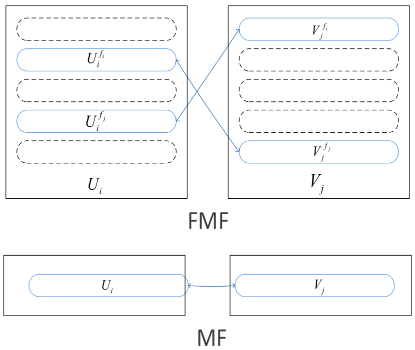

FIGURE 1. Difference between MF and FMF.

Then, we use the following formula as the loss function to learn the latent matrices U and V:

$$\ell=(R_{i,j}-\hat{R}_{i,j})^2+\lambda[(\left\|U_i^{f_i}\right\|^2+\left\|V_j^{f_i}\right\|^2)+(\left\|U_i^{f_j}\right\|^2+\left\|V_j^{f_j}\right\|^2)], \tag{4}$$

minimizing the loss function, which is min $\sum_{(i,j)\in\Omega}\ell(\bullet)$, will achieve the predicted missing ratings. The regularization term serves to reduce over-fitting. Finally, we can generate recommendations (top N or recommendation list) based on learning the missing ratings.

B. OPTIMIZATION

We use stochastic gradient descent to train model (3). For the target user i and item j, we set $e_{i,j} = R_{i,j} - \hat{R}_{i,j}$. The gradients of L($U_i^{f_i}$, $V_j^{f_j}$, $U_i^{f_j}$, $V_j^{f_i}$) with respect to $U_i^{f_i}$, $V_j^{f_j}$, $U_i^{f_j}$ and $V_j^{f_i}$ can be computed as:

$$\begin{aligned} g_i^{f_i} &= \partial\ell/\partial U_i^{f_i} = -e_{i,j}V_j^{f_j} + \lambda U_i^{f_i} \\ m_i^{f_j} &= \partial\ell/\partial U_i^{f_j} = -e_{i,j}V_j^{f_i} + \lambda U_i^{f_j} \\ p_j^{f_j} &= \partial\ell/\partial V_j^{f_j} = -e_{i,j}U_i^{f_i} + \lambda V_j^{f_j} \\ q_j^{f_i} &= \partial\ell/\partial V_j^{f_i} = -e_{i,j}U_i^{f_j} + \lambda V_j^{f_i}, \end{aligned} \tag{5}$$

Along the gradient direction, the model can be optimized as:

$$\begin{aligned} U_i^{f_i} &\leftarrow U_i^{f_i} - \eta g_i^{f_i} \\ U_i^{f_j} &\leftarrow U_i^{f_j} - \eta m_i^{f_j} \\ V_j^{f_j} &\leftarrow V_j^{f_j} - \eta p_j^{f_j} \\ V_j^{f_i} &\leftarrow V_j^{f_i} - \eta q_j^{f_i}, \end{aligned} \tag{6}$$

where λ is the regularization parameter, η is the learning rate.

In the training process, we set the initial rating matrix R as input. Latent matrix U of user and latent matrix V of item can be treated as output. Pseudocode of Table 3 illustrates the proposed FMF model in this paper.

TABLE 3. Pseudocode of FMF.

#Algorithm: Stochastic gradient methods for field-aware matrix factorization.

input matrix *R-training, R-test*
output $U,V,$
initialization $U_i^{f_i}$, $V_j^{f_j}$, $U_i^{f_j}$, $V_j^{f_i}$, learning rate, d
for z=1:Z **do**
  for v=1:$|\Omega|$ in R **do**
    sample $R_{i,j}$ from *R-training*
    calculate gradient for each latent vector:

$$g_i^{f_i}=\frac{d\ell}{dU_i^{f_i}}=-e_{i,j}V_j^{f_j}+\lambda U_i^{f_i},\quad m_i^{f_j}=\frac{d\ell}{dU_i^{f_j}}=-e_{i,j}V_j^{f_i}+\lambda U_i^{f_j}$$

$$p_j^{f_j}=\frac{\partial\ell}{\partial V_j^{f_j}}=-e_{i,j}U_i^{f_i}+\lambda V_j^{f_j},\quad q_j^{f_i}=\frac{\partial\ell}{\partial V_j^{f_i}}=-e_{i,j}U_i^{f_j}+\lambda V_j^{f_i}$$

    learn and update each latent vector:

$$U_i^{f_i}\leftarrow U_i^{f_i}-\eta g_i^{f_i},\quad U_i^{f_j}\leftarrow U_i^{f_j}-\eta m_i^{f_j}$$

$$V_j^{f_j}\leftarrow V_j^{f_j}-\eta p_j^{f_j},\quad V_j^{f_i}\leftarrow V_j^{f_i}-\eta q_j^{f_i}$$

    calculate $\hat{R}_{i,j}$ and update *R:*

$$\hat{R}_{i,j}=U_i^{f_i}V_j^{f_j}+U_i^{f_j}V_j^{f_i}$$

      if *R-training(i,j)* = 0
       R-training(i,j) = $\hat{R}_{i,j}$
  end
end

C. DISCUSSION ABOUT EXTENSION OF THE PROPOSED METHOD

In recent years, to alleviate the data sparsity problem and the cold-start problem, researchers propose some matrix factorization based recommendation methods fusing side information as extra regularization term, such as social relations among users and tag information, with rating data. The model function is given as:

$$E = \min(\ell_{MF} + \textit{regularization side information}), \tag{7}$$

Where ℓ is the traditional matrix factorization function with rating data. Regularization side information are fused social relations or some other side information. Formula (7) can improve the performance of the recommender systems further. In this paper, we propose a novel model to enhance the performance of ℓ (traditional MF) by learning field influence. However, as a basic and unified recommendation model, side information such as social relations also can be fused into FMF model. We briefly introduce how to extend FMF by fusing side information. The extended objective function is given as:

$$E = \min(\ell_{FMF} + \textit{regularization side information}), \tag{8}$$

Choosing side information needs data support and is out of the scope of this paper, we remain it as a future research.

Note that object function (8) can also be minimized by gradient descent.

V. EXPERIMENTS AND RESULTS

A. EXPERIMENTAL SETUP

Our experiments are conducted on the MovieLens 1M dataset, which consists of 1,000,209 ratings (scale 1–5) assigned by 6040 users to a collection of 3900 items. Moreover, the users file contains demographic information about each user, for example, gender is denoted by a "M" for male and "F" for female, age is chosen from 7 ranges. Occupation can be chosen from 20 choices. The movies file contains 18 genres information about each movie. We select the top 5 occupation and genres from the initial datasets in order to ensure we can evaluate the model in different field with a big distinction. The statistics of data source is summarized in Table 3.

A cross-validation technique will be used in the paper. We randomly select 90% of 1M as the training-sets to validate the performance of remaining (10% as test-sets). The random selection was carried out 5 times independently. Noting that we use latent feature dimension of 5, learning rate 0.01 and regularization parameter 0.005 in our implementation for FMF in all the experiments.

All tested models are implemented in MATLAB R2012a, and are tested on a PC Server with a 2.2 GHz CPU and 8 GB Memory.

B. MODELS FOR COMPARISON

In order to test the effectiveness of the proposed method, the following models will be used as the comparisons in the experiments.

MF: this model is proposed in [19], which is the baseline model in this paper. All elements in MF are controlled by an additive updating rule. Field influence is not considered in this model.

NMF: this model is proposed in [24], all the predicted ratings are considered as an non-negative value. Different from the MF, this model has a multiplicative updating rule. Field influence is not considered in this model.

FMF-1D: this model neglect the influence of field itself, then the predicted rating R_{ij} will be learnt by:

$$\hat{R}_{i,j} = U_i^{f_j} V_j^{f_i}, \tag{9}$$

FMF: a field aware model proposed in this paper.

C. METRICS

Generally, MF-based techniques learn latent features of users and items from the observed ratings in the user-item matrix, which are further used to predict unobserved ratings. The final purpose of predicting unobserved ratings of initial matrix is to generate a proper recommendation result according to the rating values. Traditional measures RMSE and MAE are employed to compute the rating prediction accuracy. However, high accuracy of rating prediction does not mean that the recommended result(top N recommendation or recommendation list) is accurate. Therefore, We choose the evaluation metrics as Hit Ratio and Mean reciprocal rank(MRR), which are more sensitive to the result of ranked items.

Hit Ratio and MRR can be computed as follows:

$$HitRatio = \frac{\sum_u I(T(u) \in R(u))}{|u|}, \tag{10}$$

$$MRR = \frac{1}{|U|} \sum_{(i)\in U} \frac{1}{toprank(i)}. \tag{11}$$

In Hit Ratio, T means a recommendation list generated by the proposed model form training sets, R means the real recommendation list in test sets. I $*$ is a judgement function. If an item of T belongs to R, a hit happens. If we consider the order of recommendation list, Hit Ratio will be replaced by :

$$HitRatio = \frac{\sum_u I(T(u) \cap R(u))}{|u|}, \tag{12}$$

if an item of T belongs to R in the same position, a hit happens. Higher Hit Ratio value means higher prediction accuracy.

MRR focus on the performance of the top one recommendation. In formula (11), $toprank(i)$ denotes the ranking position of the top one recommendation of user i's recommended list in the test sets. For example, if user i's top one recommendation is in the first position of test list, $toprank()$ value will be 1, if user i's top one recommendation is in the nth position of test list, $toprank()$ value will be n. Higher MRR value means higher prediction accuracy. In the experiment, the recommendation list size is set as 10.

D. RESULTS

Hit Ratio: In this section, we will evaluate the prediction accuracy of each model under the following three conditions: 1. considering the order of recommendation list (evaluated by formula (10)); 2. the order of recommendation list is not considered (evaluated by formula (12)); 3. recommendation based on classification, each genres of movie will be recommended according to the rating values, then the recommendation list size is 5(there are five type of movies in the datasets)(evaluated by formula (12)). 1. Note that field aware model outperformances the traditional MF and NMF model. FMF achieves the best performance, because FMF learns all the related relations between fields (FMF-1D neglects the influence of field itself). As shown in Table 4, FMF achieves a performance improvement of ca. 13.6% over MF, ca. 16.1% over NMF and ca. 9% over FMF-1D. Especially, FMF achieves performance improvement of ca. 64% over MF and ca.73% over NMF under the experimental conditions with that recommendation list size is 30.

Table 5 states the comparison results of Hit Ratio in condition 2. Different from the result shown in Table 4, if the order of recommendation list is not considered in the experiment. The hit ratio will increase as the recommendation list increases. FMF achieves the best performance compared with MF, NMF and FMF-1D, because field aware

TABLE 4. Experimental data statistics.

user	number/classification	field
totoal	4926	
gender	2	F、M
age	3	(-24)、(24--44)、(45+)
occupation	5	student、artist、lawyer、writer、farmer

movie	number/classification	field
total	2651	
genres	5	Action、Romance、Horror、Sci-Fi、Drama

TABLE 5. Evaluation of hit ratio (consider the order of recommendation list).

List size	1	2	3	4	5	10	15	20	25	30
MF	0.5634	0.4237	0.3566	0.2968	0.2460	0.1515	0.1206	0.0863	0.0822	0.0584
NMF	0.5587	0.4301	0.3578	0.2945	0.2422	0.1563	0.1235	0.0895	0.0811	0.0543
FMF-1D	0.5758	0.4413	0.3647	0.3001	0.2579	**0.1693**	0.1392	**0.0898**	0.1011	0.0677
FMF	**0.5904**	**0.4717**	**0.3769**	**0.3095**	**0.2640**	**0.1687**	**0.1480**	**0.0808**	**0.1089**	**0.0963**

TABLE 6. Evaluation of hit ratio (the order of recommendation is not considered list).

List size	1	2	3	4	5	10	15	20	25	30
MF	0.5634	0.6671	0.7150	0.7345	0.7535	0.8325	0.8817	0.8964	0.9216	0.9417
NMF	0.5587	0.6623	0.7132	0.7321	0.7603	0.8311	0.8821	0.8978	0.9187	0.9424
FMF-1D	0.5758	0.6704	0.7179	0.7415	0.7613	0.8386	0.8785	0.8981	**0.9278**	0.9498
FMF	**0.5904**	**0.6895**	**0.7293**	**0.7547**	**0.7756**	**0.8426**	**0.8854**	**0.9069**	0.9244	**0.9519**

model learns all the related relations between fields (FMF-1D neglects the influence of field itself). As shown in Table 5, FMF achieves a performance improvement of ca. 2% over MF, ca. 2.2% over NMF and ca.1.3% over FMF-1D.

Table 6 states the comparison results of Hit Ratio in condition 3. In this condition, each type of movie is recommended one of the best, then the recommendation list size is 5. As shown in Table 6, FMF achieves a performance improvement of ca. 9% over MF, ca. 8.9% over NMF and ca.8% over FMF-1D.

In conclusion, by learning the latent relations between fields, the proposed FMF model's improvement is significant for all the conditions.

TABLE 7. Evaluation of hitratio (recommendation based on classfication).

model	HitRatio
MF	0.0088
NMF	0.0089
FMF-1D	0.0092
FMF	**0.0096**

MRR: Table 7 states the comparison results of MRR. Note that FMF achieve the best performance compared with MF, NMF and FMF-1D. MRR value of FMF is 0.7039, which leads the performance improvement of ca. 1.6% over MF, ca. 2.3% over NMF and ca.0.6% over FMF-1D. We can get the same conclusion that by learning the latent relations between fields, the proposed FMF model is able to enhance the performance of preference prediction, especially when it is the top one recommendation.

TABLE 8. Evaluation of MRR

model	MRR
MF	0.6925
NMF	0.6879
FMF-1D	0.6992
FMF	**0.7039**

Stability of model learning: Figure 2 reports the stability of model learning on Hit Ratio and MRR. The comparative performance experiments of MF and FMF are carried out 10 times. Recommendation list size is set as 5 in this section. As shown in Figure 2, the graphics lines of FMF are more smooth and steady than those of MF. Therefore, we can get the conclusion that, FMF has strong stability of model learning. By mining the latent relations between field, FMF achieves stable and correct prediction results.

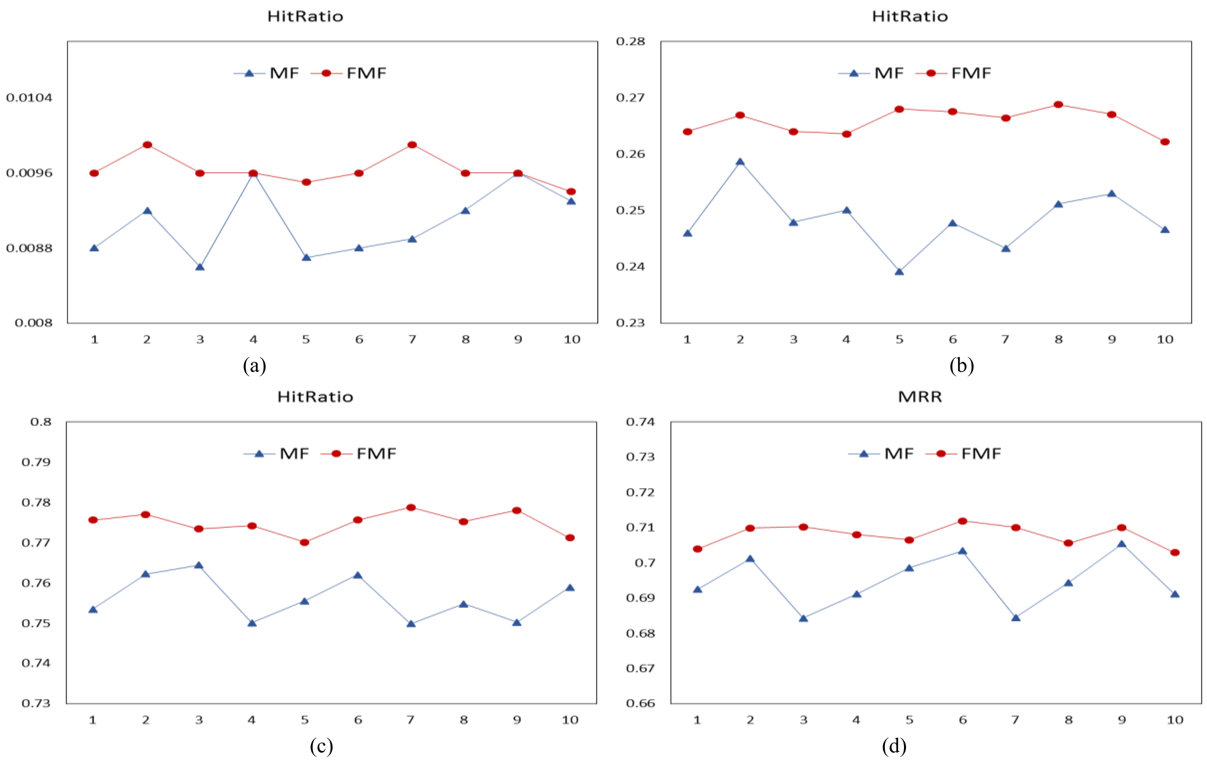

FIGURE 2. Evaluation of stability of models. (a: recommendation based on classification; b: consider the order of recommendation list c: the order of recommendation list was not considered; d: MRR)

VI. CONCLUSION AND FUTURE WORK

In this paper, we have introduced the FMF, an MF-based approach that exploits a field aware technique in order to realize improvement over the performance of the state-of-the art matrix factorization techniques for the task of recommendation. In FMF, the interactions between user/item and field can be captured and learned in the latent vector space. We also provide optimization method to train the model. The experiments demonstrate that FMF outperforms traditional MF, significantly in most cases.We also analyze the stability of model by learning of FMF and find its achievements of stable and correct prediction results.

Future work involves the following direction. The proposed FMF approach, like most of the current CF recommendation algorithms, could be regarded as a vibrational recommendation approach. where the common evaluation metrics, root mean square error is not directly related to the final recommendation but related to rating prediction. Ranking related object function could be further exploited for improving recommendation performance by directly optimizing learn to rank evaluation metrics.

REFERENCES

[1] Z. J. Zhang and H. Liu, "Social recommendation model combining trust propagation and sequential behaviors," *Appl. Intell.*, vol. 43, no. 3, pp. 695–706, Oct. 2015.

[2] L. O. Colombo-Mendoza, R. Valencia-García, A. Rodríguez-González, G. Alor-Hernández, and J. J. Samper-Zapater, "RecomMetz: A context-aware knowledge-based mobile recommender system for movie showtimes," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1202–1222, Feb. 2015.

[3] Y. Zhang and J. Koren, "Efficient Bayesian hierarchical user modeling for recommendation system," in *Proc. SIGIR*, Amsterdam, The Netherlands, 2007, pp. 47–54.

[4] K. Yu, S. Zhu, J. Lafferty, and Y. Gong, "Fast nonparametric matrix factorization for large-scale collaborative filtering," in *Proc. SIGIR*, Boston, MA, USA, 2009, pp. 211–218.

[5] M. A. Ghazanfar, S. Szedmak, and A. Prugel-Bennett, "Incremental kernel mapping algorithms for scalable recommender systems," in *Proc. 23rd IEEE Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2011, pp. 1077–1084.

[6] J. D. M. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *Proc. ICML*, 2005, pp. 713–719.

[7] R. Salakhutdinov and A. Mnih, "Bayesian probabilistic matrix factorization using Markov chain Monte Carlo," in *Proc. ICML*, 2008, pp. 880–887.

[8] S. A. Khan, E. Leppäaho, and S. Kaski, "Bayesian multi-tensor factorization," *Mach. Learn.*, vol. 105, no. 2, pp. 233–253, Nov. 2016.

[9] D. H. Alahmadi and X. J. Zeng, "ISTS: Implicit social trust and sentiment based approach to recommender systems," *Expert Syst. Appl.*, vol. 42, no. 22, pp. 8840–8849, Dec. 2015.

[10] F. Li, G. Xu, and L. Cao, "Two-level matrix factorization for recommender systems," *Neural Comput. Appl.*, vol. 27, no. 8, pp. 2267–2278, Dec. 2016.

[11] H. Steck, "Gaussian ranking by matrix factorization," in *Proc. ACM Conf. Recommender Syst.*, 2015, pp. 115–122.

[12] F.-X. Hong, X.-L. Zheng, and C.-C. Chen, "Latent space regularization for recommender systems," *Inf. Sci.*, vol. 360, pp. 202–216, Sep. 2016.

[13] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King, "Recommender systems with social regularization," in *Proc. 4th ACM Int. Conf. Web Search Data Mining*, 2011, pp. 287–296.

[14] H. Ma, "An experimental study on implicit social recommendation," in *Proc. SIGIR*, 2013, pp. 73–82.

[15] Y. Koren, "Factorization meets the neighborhood: A multifaceted collaborative filtering model," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2008, pp. 426–434.

[16] Z. Sharifi, M. Rezghi, and M. Nasiri, "New algorithm for recommender systems based on singular value decomposition method," in *Proc. 3rd Int. Conf. Comput. Knowl. Eng. (ICCKE)*, 2013, pp. 86–91.

[17] S. Anbumalar, R. A. Natarajan, and P. Rameshbabu, "Non-negative matrix factorization algorithm for the deconvolution of one dimensional chromatograms," *Appl. Math. Comput.*, vol. 241, pp. 242–258, Aug. 2014.
[18] I. Christensen and S. Schiaffino, "Matrix factorization in social group recommender systems," in *Proc. 12th Mexican Int. Conf. Artif. Intell. (MICAI)*, 2013, pp. 10–16.
[19] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *IEEE Comput.*, vol. 42, no. 8, pp. 30–37, Aug. 2009.
[20] B. Bayar, N. Bouaynaya, and R. Shterenberg, "Probabilistic non-negative matrix factorization: Theory and application to microarray data analysis," *J. Bioinformat. Comput. Biol.*, vol. 12, no. 1, p. 1450001, Feb. 2014.
[21] M. C. Cobanoglu, M. Can, C. Liu, F. Hu, Z. N. Oltvai, and I. Bahar, "Predicting drug–target interactions using probabilistic matrix factorization," *J. Chem. Inf. Model.*, vol. 53, no. 12, pp. 3399–3409, Dec. 2013.
[22] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 880–887.
[23] N. Del Buono and G. Pio, "Non-negative matrix tri-factorization for co-clustering: An analysis of the block matrix," *Inf. Sci.*, vol. 301, pp. 13–26, Apr. 2015.
[24] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS*, vol. 13, 2001, pp. 556–562.
[25] D. Dueck, Q. D. Morris, and B. J. Frey, "Multi-way clustering of microarray data using probabilistic sparse matrix factorization," *Bioinformatics*, vol. 21, pp. I144–I151, Jun. 2005.
[26] J. Liu, C. Wu, and W. Liu, "Bayesian probabilistic matrix factorization with social relations and item contents for recommendation," *Decis. Support Syst.*, vol. 55, no. 3, pp. 838–850, Jun. 2013.
[27] B. Liu, H. Xiong, S. Papadimitriou, Y. Fu, and Z. Yao, "A general geographical probabilistic factor model for point of interest recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 5, pp. 1167–1179, May 2015.
[28] P. Victor, C. Cornelis, M. De Cock, and P. P. Da Silva, "Gradual trust and distrust in recommender systems," *Fuzzy Sets Syst.*, vol. 160, no. 10, pp. 1367–1382, May 2009.
[29] C.-N. Ziegler and G. Lausen, "Propagation models for trust and distrust in social networks," *Inf. Syst. Frontiers*, vol. 7, nos. 4–5, pp. 337–358, Dec. 2005.
[30] Z. B. Sun *et al.*, "Recommender systems based on social networks," *J. Syst. Softw.*, vol. 99, pp. 109–119, Jan. 2015.
[31] I. Bayer, "fastFM: A library for factorization machines," *J. Mach. Learn. Res.*, vol. 17, no. 184, pp. 1–5, 2016.
[32] S. Rendle, "Factorization machines with libFM," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 3, 2012, Art. no. 57.
[33] S. Rendle and L. Schmidt-Thieme, "Pairwise interaction tensor factorization for personalized tag recommendation," in *Proc. ACM Int. Conf. Web Search Data Mining*, 2010, pp. 81–90.
[34] X. Q. Zhao *et al.*, "Crafting a time-aware point-of-interest recommendation via pairwise interaction tensor factorization," in *Knowledge Science, Engineering and Management. KSEM* (Lecture Notes in Computer Science), vol. 9403. Cham, Switzerland: Springer, 2015, pp. 458–470.
[35] Y. Juan, Y. Zhuang, W.-C. Chin, and C.-J. Lin, "Field-aware factorization machines for CTR prediction," in *Proc. ACM Conf. Recommender Syst.*, 2016, pp. 43–50.

ZHIYUAN ZHANG received the bachelor's degree from the University of International Relations, China, in 2012. He is currently pursuing the Ph.D. degree with the School of Electronics and Information Engineering, Beijing Jiaotong University. His current research fields include recommender system, machine learning, data mining, and artificial intelligence.

YUN LIU received the Ph.D. degree in electronic engineering from Beijing Jiaotong University, China. She is currently a Professor and the Director of the Department of Communication and Information System. She is also the Director of the Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University. She has edited many books and published over 200 papers and book chapters, and also participated in many international academic activities, including the organization of several international conferences. Her current research interests include telecommunication, computer networks, network security, intelligent transportation system, and social dynamics. She is an IET Fellow.

ZHENJIANG ZHANG received the Ph.D. degree in communication and information system from Beijing Jiaotong University, Beijing, China, in 2008. Since 2008, he has been an Assistant Professor with the Department of Electronic and Information Engineering, Beijing Jiaotong University. He has published about 60 professional research papers. His research interests are wireless sensors networks techniques, including multi-source data fusion, security and privacy, routing, and energy management.