Pattern Recognition 136 (2023) 109171

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Lower bound estimation of recommendation error through user uncertainty modeling

Heng-Ru Zhang [a,*], Ying Qiu [a], Ke-Lin Zhu [a], Fan Min [a,b]

[a] School of Computer Science, Southwest Petroleum University, Chengdu 610500, China
[b] Institute for Artificial Intelligence, Southwest Petroleum University, Chengdu 610500, China

ARTICLE INFO

Article history:
Received 8 March 2022
Revised 6 August 2022
Accepted 9 November 2022
Available online 13 November 2022

Keywords:
Magic barrier estimation
Mixture of exponential power
Recommender system
Uncertainty modeling

ABSTRACT

In machine learning, the Bayesian error is the lower bound of the prediction error induced by data distribution. In recommender systems, this is also known as the magic barrier (MGBR). MGBR estimation is an important issue because the recommended data frequently contain considerable uncertainties that are difficult to quantify. It is possible to determine the extent to which the recommendation algorithm can be optimized by obtaining the MGBR for a given dataset. MGBR estimation generally requires real user ratings that are not affected by external factors such as human emotions and living environment, which can be extremely difficult or even impossible to gather. Existing theoretical approaches based on simple models, such as Gaussian distributions, have limited estimation capabilities. In this paper, we propose a more sophisticated mixture of exponential power (MoEP) model, which enables adaptive parameter selection for intricate uncertainty. To fit the distribution of the real data, we constructed a flexible learning model that automatically adjusts super- or sub-Gaussian uncertainties using the MoEP components. To select parameters adaptively, we employed an expectation-maximization algorithm to infer the parameters of the components. To estimate the MGBR, we explored an approach for calculating the lower bound of the prediction error under the guidance of a probability model. Experiments on the four datasets validated the rationality of the proposed method. The results show that the MGBR estimated using the new model is marginally lower than the prediction error of state-of-the-art algorithms.

1. Introduction

Ratings are primarily used by recommender systems to convey user preferences for items proactively [1]. However, generally this approach has uncertainties owing to the influence of difficult-to-quantify factors [2,3]. Uncertainties are frequently associated with human behavioral patterns such as fallibility, reflexivity, and inconsistency [4]. Human fallibility originates from the limited ability of people to absorb, process, and communicate information [5]. Human reflexivity refers to the shifting, reorientation, and prioritization of personal priorities [6]. Human inconsistency is frequently related to changes in emotions and living circumstances [7].

Such uncertainty leads to a magic barrier (MGBR) [3,7–9] in recommender systems. The origins of this concept can be seen in generic machine-learning tasks [10] meant to derive models from data. Machine learning typically struggles with uncertainty and cannot guarantee perfect accuracy [11,12]. For classification tasks, the least amount of error caused by uncertainty is typically referred to as Bayesian error [13,14], commonly known as MGBR in recommendation tasks [7,8]. Although recommender systems have shifted to richer information, such as user sentiments [15] or timestamps [16], rating data remain at the core. The MGBR indicates the lower bound of the prediction error of a recommendation algorithm [3]. This estimation helps us evaluate the quality of data. Therefore, this is a fundamental research topic because the MGBR indicates the circumstances under which new recommendation algorithms are no longer required for the data at hand. For instance, Tang et al. [17] found that the mean absolute error (MAE) of the ML-1M dataset could be less than 0.6724 (on a five-point rating scale). Amatriain et al. [2] investigated the impact of uncertainty and found root mean squared errors (RMSEs) ranging from 0.5570 to 0.8156 for the Netflix dataset. Further, Nguyen et al. [18] concluded that the lower the uncertainty, the lower the prediction error.

MGBR estimation typically requires real user ratings that are unaffected by external factors such as human emotions and living environments. This is often extremely difficult or impossible to

* Corresponding author.
E-mail addresses: zhanghr@swpu.edu.cn (H.-R. Zhang), 202121000495@stu.swpu.edu.cn (Y. Qiu), 201922000379@stu.swpu.edu.cn (K.-L. Zhu), minfan@swpu.edu.cn (F. Min).

https://doi.org/10.1016/j.patcog.2022.109171
0031-3203/

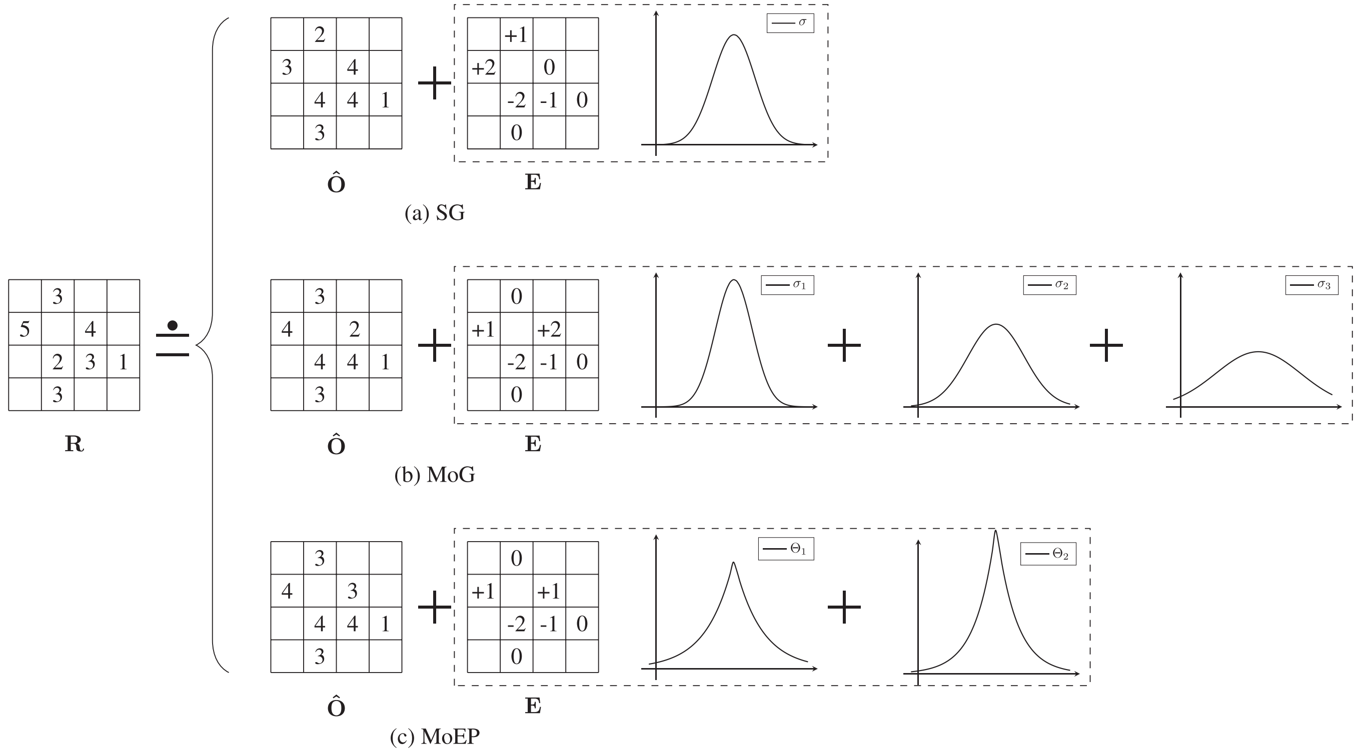


Fig. 1. The three uncertainty models. From top to bottom: (a) The uncertainty is assumed to follow an SG distribution with the standard deviation σ explicitly specified. (b) The uncertainty is assumed to follow an arbitrary distribution fitted by the three Gaussian components. (c) The uncertainty assumption is the same as in (b), and the two exponential power (EP) components are used to fit it.

achieve. Therefore, it is necessary to model user uncertainty. Currently, there are two theoretical MGBR estimation models: Single Gaussian (SG) [8] and mixture of Gaussian (MoG) [9]. On the one hand, the SG model considers uncertainty as a normal distribution with an expert-determined standard deviation. On the other hand, the MoG model can approximate an arbitrary distribution when the number of components approaches infinity. However, because only a finite number of components can be specified in practice, super- or sub-Gaussian uncertainties are beyond the representative ability of the current MoG [19] model.

In this study, we propose a novel MGBR estimation approach to deal with arbitrarily distributed uncertainties. The key idea is to model the uncertainty as a mixture of exponential power (MoEP) [19,20] distribution, which can approximate any continuous density function with the fewest components. The MoG model has two advantages. First, the components of the MoEP are more sophisticated than those of MoG, allowing arbitrarily distributed uncertainty to be modeled with fewer components. Second, the MoEP model can automatically learn the number of distribution components from rating data without requiring expertise.

Fig. 1 shows examples of uncertainty modeling for the SG, MoG, and MoEP models. Let $\mathbf{R}$ denote the original rating matrix, $\hat{\mathbf{O}}$ the reconstructed rating matrix, and $\mathbf{E}$ the uncertainty matrix. σ denotes the standard deviation of the SG model. σ_1, σ_2, and σ_3 denote the three standard deviations of the MoG model, respectively. Θ_1 and Θ_2 denote the two component parameters of the MoEP, respectively.

Fig. 1 (a) assumes that the uncertainty $\mathbf{E}$ follows an SG distribution. However, the standard deviation σ must be specified according to expert experience. Fig. 1(b) assumes that the uncertainty $\mathbf{E}$ follows an arbitrary distribution. The uncertainty $\mathbf{E}$ is approximated using three Gaussian components. However, an unlimited number of Gaussian components are necessary to fit more complex distributions. Fig. 1(c) assumes the same assumption as in Fig. 1(b). Because EP components have more sophisticated representation capabilities than Gaussian components, fewer components can fit complex distributions.

To validate the newly proposed MGBR estimation model, we conducted extensive experiments on four publicly available datasets. The experimental results show that 1) our model can fit the intricate uncertainty more precisely with fewer components than the MoG model, 2) the number of EP components can be determined automatically, and 3) our estimated MGBR is close to or lower than that of state-of-the-art (SOTA) algorithms.

The main contributions of this study can be summarized as follows:

1) We propose the assumption that the MGBR is the lower bound of prediction error caused by user uncertainty (see Assumption 1). For a better realization of this assumption than in the SG [8] and MoG [9] models, we reinterpreted the fundamental meaning of the assumption.
2) We propose an MoEP model to estimate the MGBR of recommender systems. The MoEP model has an advantage over the MoG model [9] in that it enables the modeling of arbitrarily distributed uncertainty with fewer components.
3) The barriers are validated through comparison with SOTA recommendation algorithms.

The remainder of this paper is organized as follows: Section 2 reviews background information and summarizes the most relevant methods for MGBR estimation. Section 3 introduces the proposed algorithm in detail. Section 4 presents the outcomes of the experiments conducted on multiple datasets. Finally, Section 5 summarizes the contributions of this study and discusses future work. The implementation of the MoEP-based MGBR estimation for recommender systems is available at https://github.com/zhanghrswpu/LEUM.

Table 1
Notations.

Notation	Meaning
U	The set of all users
T	The set of all items
L	The rating levels
r_l	The lowest rating level
r_h	The highest rating level
r_s	The minimum interval between discrete ratings
n	The number of users
m	The number of items
Ω	$\Omega = \left\{(i, j) \mid r_{i,j} \neq 0, i \in [1, n], j \in [1, m]\right\}$
$\mathbf{R}$	The rating matrix
u_i	the ith user
t_j	the jth item
$\hat{\mathbf{R}}$	The predicted rating matrix
$\mathbf{O}$	The true rating matrix
$\hat{\mathbf{O}}$	The reconstructed rating matrix
$\mathbf{E}$	The uncertainty corresponding to $\mathbf{R}$
$o_{i,j}$	The true rating of u_i on t_j
$e_{i,j}$	The uncertainty in $r_{i,j}$
$r_{i,j}$	The rating of u_i on t_j
K	The number of mixture components
η_k	The precision parameter of the kth EP distribution
$\boldsymbol{\eta}$	$\boldsymbol{\eta} = [\eta_1, \eta_2, \ldots, \eta_K]$
p_k	The shape parameter of the kth EP distribution
$\mathbf{P}$	$\mathbf{P} = [p_1, p_2, \ldots, p_K]$
π_k	The mixing proportion of the kth EP distribution
$\boldsymbol{\pi}$	$\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_K]$
$\boldsymbol{\Theta}_k$	$\boldsymbol{\Theta}_k = \{\pi_k, \eta_k\}$
$\mathbb{P}$	The likelihood of $\mathbf{E}$
$\mathbb{P}_{\Theta_k}\left(r_{i,j} = y\right)$	The probability of $r_{i,j} = w$ for Θ_k
$\mathbb{P}_{\Theta_k}\left(o_{i,j} = y\right)$	The probability of $o_{i,j} = w$ for Θ_k
$\mathbb{P}_{\Theta_k}\left(r_{i,j} = y \mid o_{i,j} = w\right)$	The conditional probability of true ratings wrt. the rating for Θ_k
$\mathbb{P}_{\Theta_k}\left(o_{i,j} = w \mid r_{i,j} = y\right)$	The conditional probability of true ratings wrt. the true rating for Θ_k
$\frac{\parallel O - R \parallel_1}{N}$	The MGBR in terms of MAE
$\frac{\parallel O - R \parallel_F}{\sqrt{N}}$	The MGBR in terms of RMSE

Table 2
Rating matrix **R**.

UID / TID	t_1	t_2	t_3	t_4
u_1	0	3	0	0
u_2	5	0	4	0
u_3	0	3	2	2
u_4	4	0	0	0

2. Related works

This section presents relevant work, such as the definition of the rating system, a review of some representative collaborative filtering (CF) methods, and an analysis of existing MGBR estimation methods.

Table 1 lists the notations used in this paper.

2.1. Rating system

Let $U = \{u_1, u_2, \ldots, u_n\}$ denote the user set and $T = \{t_1, t_2, \ldots, t_m\}$ denote the item set. The rating mapping process [21] can be formulated as

$$\mathbf{R} : U \times T \to L, \tag{1}$$

where $\mathbf{R} = (r_{i,j})_{n\times m}$, and $L = \{r_l, r_l + r_s, r_l + 2r_s, \ldots, r_h\}$ denotes the rating levels. r_l, r_h, and r_s represent the lowest, highest, and minimum interval ratings, respectively.

Table 2 presents a simple example of a rating system, where $L = \{1, 2, 3, 4, 5\}$, $n = 4$, $m = 4$, $r_l = 1$, $r_h = 5$, and $r_s = 1$. The value 0 indicates that the item has not yet been rated by the user.

2.2. Problem statement of collaborative filtering

CF [22] frequently uses the rating system mentioned above to make recommendations. It recommends items from other users with preferences similar to those of active users [23]. Neighborhood-based CF [24,25] emphasizes the relationship between items or users. As an extension of CF recommendation, matrix factorization (MF)-based CF maps users and items to two latent subspaces, such as probabilistic MF [26] and sentiment-based MF [15].

The reduction of the prediction error is the main objective of CF algorithms. For CF algorithms, there are two different evaluation metrics: the error metric and the rank metric [27]. The first evaluates the predictive accuracy, and the second evaluates the quality of the recommendation list. In this study, we only covered the MGBRs associated with prediction accuracy.

The MAE and RMSE are two common error metrics [27]that determine the average deviation between predicted and actual values. Let $\hat{\mathbf{R}} = (\hat{r}_{i,j})_{n\times m}$ be a prediction matrix, where $\hat{r}_{i,j}$ is the prediction rating of u_i on t_j. Prediction error wrt. MAE and RMSE are defined as

$$\mathrm{MAE}(\hat{\mathbf{R}}, \mathbf{R}) = \frac{||\hat{\mathbf{R}} - \mathbf{R}||_1}{N} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}|\hat{r}_{i,j} - r_{i,j}|}{N} \tag{2}$$

and

$$\mathrm{RMSE}(\hat{\mathbf{R}}, \mathbf{R}) = \frac{||\hat{\mathbf{R}} - \mathbf{R}||_F}{\sqrt{N}} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}(\hat{r}_{i,j} - r_{i,j})^2}{N}}, \tag{3}$$

respectively, where $N = |\Omega|$, Ω is the index set of ratings in $\mathbf{R}$, and $||\cdot||_1$ and $||\cdot||_F$ represent the 1-norm and F-norm of a matrix, respectively. Evidently, the smaller the MAE or RMSE, the more accurate the prediction. MAE and RMSE are unlikely to be very close to the theoretical value of 0, given the uncertainty in the rating data.

2.3. Evaluating approaches for magic barrier

The MGBR suggests that the inherent nature of data induces a lower bound in terms of MAE and RMSE [8,9]. Existing MGBR estimation methods are mainly classified into three types [8,9]: empirical, semi-empirical, and theoretical methods.

Empirical methods [2,3,28–30] typically use several ratings for the same item by the same user to estimate MGBR. They hypothesize that such numerous ratings can eliminate the impact of uncertainty; however, such an operation may inevitably produce additional uncertainty. Furthermore, several ratings are inappropriate for large-scale data because they are tiresome and time consuming for users.

Semi-empirical methods [7,31] constitute simple statistical models based on the empirical methods. The crucial hyperparameters in these models require expertise. However, the inaccuracy of these artificially specified parameters generates additional uncertainty.

Theoretical methods constitute SG [8] or MoG [9] models for obtaining the rating data and followed by a probabilistic model to estimate the MGBR. However, real-world uncertainty comprises a mixture of super- and sub-Gaussian components. Consequently, both the SG and MoG models are limited in their ability to deal with such intricate uncertainties. Additionally, the required hyperparameters, such as the optimal deviation in the SG model and number of components in the MoG model, are difficult to specify.

3. Methodology

First, we introduce an MoEP-based uncertainty assumption. Second, we demonstrate how to acquire the MoEP parameters in detail. Finally, we elaborate on the MGBR calculation using the obtained MoEP parameters.

3.1. The estimation framework

First, we recall the assumption proposed by Zhang et al. [8,9].

Assumption 1. User Uncertainty MGBR Assumption.
The MGBR is the lower bound of the prediction error caused by user uncertainty.

$$\frac{||\mathbf{O}-\mathbf{R}||_1}{N} \leq \min \frac{||\hat{\mathbf{R}}-\mathbf{R}||_1}{N}, \tag{4}$$

and

$$\frac{||\mathbf{O}-\mathbf{R}||_F}{\sqrt{N}} \leq \min \frac{||\hat{\mathbf{R}}-\mathbf{R}||_F}{\sqrt{N}}, \tag{5}$$

where $\mathbf{O}$ is the true rating matrix, $\mathbf{R}$ is the rating matrix, $\hat{\mathbf{R}}$ is the prediction matrix, and $||\cdot||_1$ and $||\cdot||_F$ represent the 1-norm and F-norm, respectively. Furthermore, $\frac{||\mathbf{O}-\mathbf{R}||_1}{N}$ and $\frac{||\mathbf{O}-\mathbf{R}||_F}{\sqrt{N}}$ denote the mean error owing to the uncertainty wrt. the 1-norm and F-norm, respectively, and $\frac{||\hat{\mathbf{R}}-\mathbf{R}||_1}{N}$ and $\frac{||\hat{\mathbf{R}}-\mathbf{R}||_F}{\sqrt{N}}$ represent MAE and RMSE, respectively.

The "$\leq$" symbol in Assumption 1 can be understood as follows:

1) The left-hand sides of Eqs. (4) and (5) represent the uncertainty introduced by the data distribution.
2) The right-hand sides of Eqs. (4) and (5) contain both the uncertainty caused by the data distribution and the error introduced by the prediction algorithm.

Fig. 2 illustrates the framework for MGBR estimation. The uncertainty matrix $\mathbf{E}$ is fitted using two EP components, whose parameters are denoted by Θ_1 and Θ_2. From top to bottom, the MGBR estimation process includes the following steps:

1) Uncertainty estimation: For each rating $r_{i,j}$, the expectation maximization (EM) [32] method is employed to calculate the parameters Θ_1 and Θ_2, as well as the mixing ratio $\gamma = \{\gamma_{i,j,k} \mid (i,j) \in \Omega, k \in [1,K]\}$. The number of mixed components and the specific steps of EM estimation parameters are explained in detail in Section 3.3.
2) Conditional-probability computation: We can compute the conditional probability $\mathbb{P}_{\Theta_k}(r_{i,j} = y \mid o_{i,j} = w)$ using the parameters obtained in the previous step. We can calculate $\mathbb{P}_{\Theta_k}(o_{i,j} = w \mid r_{i,j} = y)$ using the Bayesian formula.
3) Magic barrier computation: Under the assumption of user uncertainty, the MGBR with respect to $\frac{||\mathbf{O}-\mathbf{R}||_1}{N}$ and $\frac{||\mathbf{O}-\mathbf{R}||_F}{\sqrt{N}}$ can be calculated using $\mathbb{P}_{\Theta_k}(o_{i,j} = w \mid r_{i,j} = y)$.

3.2. MoEP-based uncertainty assumption statement

In recommender systems, each rating $r_{i,j}$ in the rating data matrix $\mathbf{R}$ can be modeled as follows:

$$r_{i,j} = o_{i,j} + e_{i,j}, \tag{6}$$

where $o_{i,j}$ is the true rating of u_i on t_j, and $e_{i,j}$ is the uncertainty in $r_{i,j}$. It is almost impossible to obtain a user rating $o_{i,j}$ that is unaffected by external factors. Therefore, we focus on modeling the user uncertainty $e_{i,j}$.

We assume that the uncertainty $e_{i,j}$ follows an MoEP distribution [19]:

$$\mathbb{P}(e_{i,j}) = \sum_{k=1}^{K} \pi_k f_{p_k}\left(e_{i,j} \mid 0, \eta_k\right), \tag{7}$$

where K denotes the number of mixing components, π_k denotes the mixing ratio that satisfies $\sum_{k=1}^{K} \pi_k = 1$ and $\pi_k \geq 0$, and $f_{p_k}\left(e_{i,j} \mid 0, \eta_k\right)$ denotes the kth EP distribution with zero mean, precision parameter η_k, and shape parameter p_k.

As mentioned in Cao et al. [19], the density function of the EP distribution ($p_k > 0$) can be defined as:

$$f_{p_k}(e_{i,j} \mid 0, \eta_k) = \frac{p_k \eta_k^{\frac{1}{p_k}}}{2\Gamma\left(\frac{1}{p_k}\right)} \exp\left(-\eta_k \left|e_{i,j}\right|^{p_k}\right), \tag{8}$$

where $\Gamma(\cdot)$ is the gamma function. Different distributions can be obtained by adjusting the value of p_k. The Leptokurtic ($0 < p_k < 2$) and Platykurtic ($p_k > 2$) distributions are able to be obtained as long as changing p_k of EP components. In addition, EP component can describe the Laplace distribution with $p_k = 1$, the Gaussian distribution with $p_k = 2$ and the Uniform distribution with $p_k \to \infty$.

3.3. Uncertainty parameter estimation

There are three types of parameters to be solved: $\boldsymbol{\pi} = [\pi_1, \pi_2, \ldots, \pi_K]$, $\boldsymbol{\eta} = [\eta_1, \eta_2, \ldots, \eta_K]$ and $\mathbf{p} = [p_1, p_2, \ldots, p_K]$. For simplicity, we denote $\boldsymbol{\Theta} = \{\boldsymbol{\pi}, \boldsymbol{\eta}\}$. To estimate the value of $\boldsymbol{\Theta}$, we adopt the EM algorithm [32], where $\mathbf{p}$ can be varied.

We assume that each uncertainty $e_{i,j}$ has latent variable $\mathbf{z}_{i,j} = [z_{i,j,1}, z_{i,j,2}, \ldots, z_{i,j,K}]$.

$$z_{i,j,k} = \begin{cases} 1, & e_{i,j} \text{is derived from the } k\text{th EP distribution;} \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

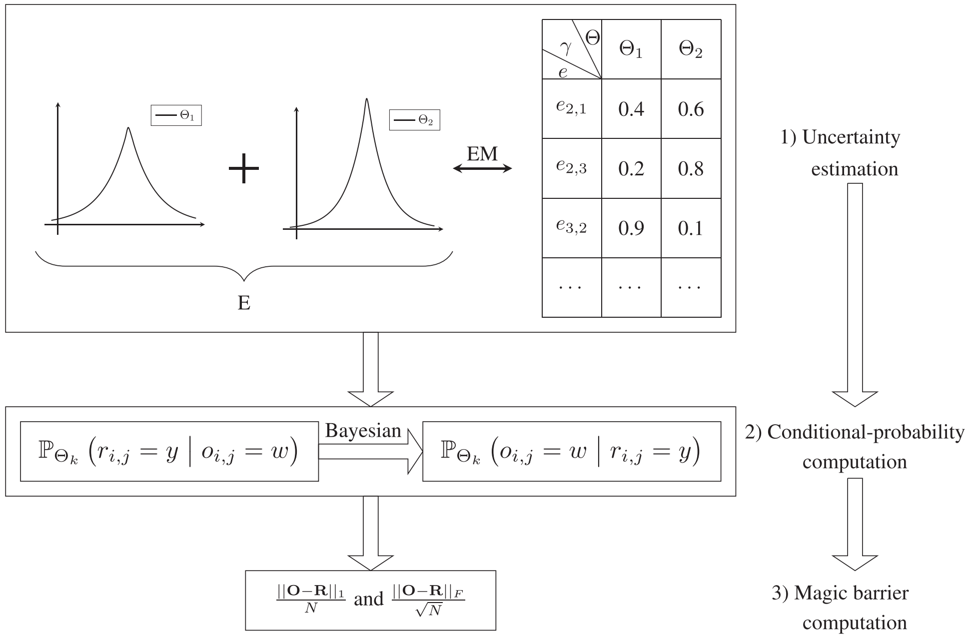


Fig. 2. Estimating framework. MGBR is determined in three steps: 1) Uncertainty estimation: The uncertainty is fitted using two EP components, and the parameters of these components are obtained using the EM algorithm; 2) Conditional-probability computation: The relevant parameters required to calculate the MGBR are obtained using the Bayesian formula; and 3) Magic barrier computation: Calculate the MGBR in terms of MAE and RMSE.

where $\sum_{k=1}^{K} z_{i,j,k} = 1$. $\mathbf{z}_{i,j}$ is subject to a multinomial distribution $\mathbf{z}_{i,j} \sim \mathbb{P}_K(K : \pi_1, \pi_2, \ldots, \pi_K)$. Then, we have:

$$\mathbb{P}(e_{i,j} \mid \mathbf{z}_{i,j}) = \prod_{k=1}^{K} f_{p_k}(e_{i,j} \mid 0, \eta_k)^{z_{i,j,k}}, \tag{10}$$

$$\mathbb{P}(\mathbf{z}_{i,j} \mid \boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{z_{i,j,k}}. \tag{11}$$

By denoting $\mathbf{E} = (e_{i,j})_{n\times m}, \mathbf{Z} = (\mathbf{z}_{i,j})_{n\times m}$, and $\Omega = \{(i,j) \mid r_{i,j} \neq 0, i \in [1,n], j \in [1,m]\}$, the complete likelihood function can be expressed as:

$$\mathbb{P}(\mathbf{E}, \mathbf{Z} \mid \boldsymbol{\Theta}) = \prod_{(i,j)\in\Omega} \prod_{k=1}^{K} \left[\pi_k f_{p_k}\left(e_{i,j} \mid 0, \eta_k\right)\right]^{z_{i,j,k}}. \tag{12}$$

To simplify the calculations, the goal was to maximize the log-likelihood function for the proposed model parameters. Thus, the **target function** can be written as:

$$\begin{aligned}\max_{\boldsymbol{\Theta}} L(\boldsymbol{\Theta}) &= \ln \mathbb{P}(\mathbf{E} \mid \boldsymbol{\Theta}) \\ &= \ln \sum_{\mathbf{Z}} \mathbb{P}(\mathbf{E}, \mathbf{Z} \mid \boldsymbol{\Theta}).\end{aligned} \tag{13}$$

Next, we can obtain

$$\max_{\boldsymbol{\Theta}} L(\boldsymbol{\Theta}) = \sum_{(i,j)\in\Omega} \sum_{k=1}^{K} z_{i,j,k}\left[\ln \pi_k + \ln f_{p_k}(e_{i,j} \mid 0, \eta_k)\right]. \tag{14}$$

An advantage of the proposed model is that it can automatically determine the number of components. In addition, we employed an efficient trick proposed by Huang et al. [33] to determine the EP mixture number. Hence we have

$$\max_{\boldsymbol{\Theta}} L(\boldsymbol{\Theta}) = \sum_{(i,j)\in\Omega} \sum_{k=1}^{K} z_{i,j,k}\left[\ln \pi_k + \ln f_{p_k}(e_{i,j} \mid 0, \eta_k)\right] - P(\boldsymbol{\pi}; \lambda), \tag{15}$$

here,

$$P(\boldsymbol{\pi}; \lambda) = n\lambda \sum_{k=1}^{K} D_k \log \frac{\epsilon + \pi_k}{\epsilon}, \tag{16}$$

where λ is a fine-tunable penalty factor greater than zero, ϵ is a very small positive value, and D_k is a constant that has no influence on $\boldsymbol{\Theta}$. The EM algorithm was then used to iteratively estimate the model parameters, that is, to solve Eq. (15). $\boldsymbol{\Theta}^{(t)} = \{\boldsymbol{\pi}^{(t)}, \boldsymbol{\eta}^{(t)}\}$ represent the estimated parameters in iteration tth. The EM algorithm is divided into two steps: E-step and M-step.

In the E-step, the reconstructed rating matrix $\hat{\mathbf{O}}$ replaces the true rating matrix $\mathbf{O}$ via low-rank matrix factorization (MF) because $\mathbf{O}$ is difficult to obtain in practice. Let $\hat{\mathbf{O}} = \mathbf{U}\mathbf{V}^{\mathrm{T}}$, where $\mathbf{U} \in \mathbb{R}^{n\times r}$ and $\mathbf{V} \in \mathbb{R}^{m\times r}$ $(r \ll \min(n,m))$. The conditional expectation of $z_{i,j,k}$ given $e_{i,j}$ is denoted by $\gamma_{i,j,k}$. Hence, $\gamma_{i,j,k}$ can be calculated using the Bayesian rule

$$\gamma_{i,j,k}^{(t+1)} = \frac{\pi_k^{(t)} f_{p_k}\left(r_{i,j} - \mathbf{u}_i^{(t)}\left(\mathbf{v}_j^{(t)}\right)^T\right) \mid 0, \eta_k^{(t)}\right)}{\sum_{l=1}^{K} \pi_l^{(t)} f_{p_l}\left(r_{i,j} - \mathbf{u}_i^{(t)}\left(\mathbf{v}_j^{(t)}\right)^T\right) \mid 0, \eta_l^{(t)}\right)\right)}, \tag{17}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ represent the ith rows of $\mathbf{U}$ and $\mathbf{V}$, respectively.

Then, the Q function can be written as

$$\begin{aligned}Q\left(\boldsymbol{\Theta}, \boldsymbol{\Theta}^{(\mathbf{t})}\right) = &\sum_{(i,j)\in\Omega} \sum_{k=1}^{K} \gamma_{i,j,k}^{(t+1)}\left[\log f_{p_k}\left(e_{i,j} \mid 0, \eta_k\right) + \log \pi_k\right] \\ &-n\lambda \sum_{k=1}^{K} D_k \log \frac{\epsilon + \pi_k}{\epsilon}.\end{aligned} \tag{18}$$

In the M-step, we maximize Eq. (18) to update $\boldsymbol{\Theta}$. Because $\boldsymbol{\pi}$ is constrained using $\sum_{k=1}^{K} \pi_k = 1$, it is natural to utilize the Lagrange multiplier method to address such optimal problems. The objective is to maximize the following Lagrange function:

$$\begin{aligned}&\sum_{(i,j)\in\Omega} \sum_{k=1}^{K} \gamma_{i,j,k}^{(t+1)}\left[\log f_{p_k}\left(e_{i,j} \mid 0, \eta_k\right) + \log \pi_k\right] \\ &-n\lambda \sum_{k=1}^{K} D_k \log \frac{\epsilon + \pi_k}{\epsilon} + \tau\left(\sum_{k=1}^{K} \pi_k - 1\right),\end{aligned} \tag{19}$$

where τ denotes the Lagrange multiplier. For simplicity, by removing the part of Eq. (19) that is not related with the first derivative of π_k, we obtain the following function

$$\sum_{(i,j)\in\Omega}\sum_{k=1}^{K}\gamma_{i,j,k}^{(t+1)}\log\pi_k - n\lambda\sum_{k=1}^{K}D_k\log\frac{\epsilon+\pi_k}{\epsilon}+\tau\left(\sum_{k=1}^{K}\pi_k-1\right). \tag{20}$$

When the first derivative of Eq. (20) wrt. π_k to zero, we can obtain

$$\pi_k^{(t+1)}=\max\left\{0,\frac{1}{1-\lambda\hat{D}}\left[\frac{\sum_{(i,j)\in\Omega}\gamma_{i,j,k}^{(t+1)}}{|\Omega|}-\lambda D_k\right]\right\}, \tag{21}$$

where $\hat{D}=\sum_{k=1}^{K}D_k=2K$. Note that D_k is a constant with no effect on $\mathbf{\Theta}$. Similarly, by setting the first derivative of the function $Q\left(\mathbf{\Theta},\mathbf{\Theta}^{(\mathbf{t})}\right)$ wrt. $\boldsymbol{\eta}$ to zero, we obtain:

$$\eta_k^{(t+1)}=\frac{\sum_{(i,j)\in\Omega}\gamma_{i,j,k}^{(t+1)}}{p_k\sum_{(i,j)\in\Omega}\gamma_{i,j,k}^{(t+1)}\left|r_{i,j}-\mathbf{u}_i^{(t)}\left(\mathbf{v}_j^{(t)}\right)^T\right|^{p_k}}. \tag{22}$$

The above formula provides us with the process of calculating the parameters required for the MoEP.

3.4. Magic barrier estimation

There are four steps in magic barrier estimation. The first step is to compute the conditional probability $\mathbb{P}_{\Theta_k}\left(r_{i,j}=y \mid o_{i,j}=w\right)$ using the obtained parameters η_k and p_k. Theoretically, the independent variable range of the EP distribution has a series of continuous values ranging from negative to positive infinity. However, in practice, the values of $o_{i,j}$ and $r_{i,j}$ are discrete and finite integers. According to a rating system with rating domain $L=[r_l,r_h]$, we have

$$\mathbb{P}_{\Theta_k}\left(r_{i,j}=y \mid o_{i,j}=w\right)=\int_{y-\frac{r_s}{2}}^{y+\frac{r_s}{2}}\frac{p_k\eta^{\frac{1}{p_k}}}{2\Gamma\left(\frac{1}{p_k}\right)}\exp\left(-\eta_k|x|^{p_k}\right)\mathrm{d}x, \tag{23}$$

where $y\in L$, $w\in L$, and $k\in[1,K]$.

The second step is to calculate the probability $P_{\Theta_k}(r_{i,j}=y)$. For each Θ_k, we have

$$\mathbb{P}_{\Theta_k}(r_{i,j}=y)=\frac{\sum_{r_{a,b}=y}\gamma_{a,b,k}}{\sum_{(a,b)\in\Omega}\gamma_{a,b,k}}. \tag{24}$$

Meanwhile, using the total probability formula, we obtain

$$\mathbb{P}_{\Theta_k}\left(r_{i,j}=y\right)=\sum_{w=r_l}^{r_h}\mathbb{P}_{\Theta_k}\left(o_{i,j}=w\right)\mathbb{P}_{\Theta_k}\left(r_{i,j}=y \mid o_{i,j}=w\right). \tag{25}$$

We employ the Gauss–Seidel iteration to calculate $|L|\times K$ unknown $\mathbb{P}_{\Theta_k}\left(o_{i,j}=w\right)$.

The Bayesian formula is used in the third step to obtain:

$$\mathbb{P}_{\Theta_k}\left(o_{i,j}=w \mid r_{i,j}=y\right)=\frac{\mathbb{P}_{\Theta_k}\left(o_{i,j}=w\right)\mathbb{P}_{\Theta_k}\left(r_{i,j}=y \mid o_{i,j}=w\right)}{\mathbb{P}_{\Theta_k}\left(r_{i,j}=y\right)}. \tag{26}$$

There are $L^2\times K$ unknown $\mathbb{P}_{\Theta_k}\left(o_{i,j}=w \mid r_{i,j}=y\right)$ to be determined.

In the last step, the MGBR in terms of MAE and RMSE can be respectively calculated as follows:

$$\frac{\|\mathbf{O}-\mathbf{R}\|_1}{N}=\frac{\sum_{y=r_l}^{r_h}\sum_{w=r_l}^{r_h}\sum_{(i,j)\in\Omega}\sum_{k=1}^{K}\mathbb{P}_{\Theta_k}\left(o_{i,j}=w \mid r_{i,j}=y\right)\pi_k\cdot|w-y|}{N}, \tag{27}$$

and

$$\frac{\|\mathbf{O}-\mathbf{R}\|_F}{\sqrt{N}}=\sqrt{\frac{\sum_{y=r_l}^{r_h}\sum_{w=r_l}^{r_h}\sum_{(i,j)\in\Omega}\sum_{k=1}^{K}\mathbb{P}_{\Theta_k}\left(o_{i,j}=w \mid r_{i,j}=y\right)\cdot\pi_k\cdot(w-y)^2}{N}}. \tag{28}$$

In this particular case, only one EP component exists, that is, $K=1$, $\pi_1=1$. Algorithm 1 outlines the MoEP-based MGBR estimation.

Algorithm 1 MGBR estimation using MoEP model ([34], this reference is included in Algorithm).

Input: Rating matrix $\mathbf{R}\in\mathbb{R}^{n\times m}$;Initialized number of components K_{start} and parameters $\mathbf{\Theta}$, preset candidates $\mathbf{p}=\left[p_1,p_2,\dots,p_{K_{start}}\right]$;
Output: MGBR estimation;
1: **while** not *converge* **do**
2: **Estep**: Update the posterior probability $\gamma^{(t)}$ by Eq. (17);
3: *Mstep*1: Update the proportional parameter $\pi_k^{(t)}$ by Eq. (21);
4: **if** EP component with $\pi_k^{(t)}=0$ **then**
5: Remove this component;
6: **end if**
7: *Mstep*2: Update the parameters $\eta^{(t)}$ by Eq. (22);
8: *Mstep*3: Update Matrices **U** and **V** by Augmented Lagrange Multiper (ALM) [34];
9: **end while**
10: Calculate $\mathbb{P}_{\Theta_k}\left(r_{i,j}=y \mid o_{i,j}=w\right)$ with above obtained parameters by Eq. (23);
11: Calculate $\mathbb{P}_{\Theta_k}\left(o_{i,j}=w\right)$ by Eqs. (24) and (25);
12: Calculate $\mathbb{P}_{\Theta_k}\left(o_{i,j}=w \mid r_{i,j}=y\right)$ by Eq. (26);
13: Calculate $\frac{\|\mathbf{O}-\mathbf{R}\|_1}{N}$ and $\frac{\|\mathbf{O}-\mathbf{R}\|_F}{\sqrt{N}}$ by Eqs. (27) and (28), respectively.

3.5. Time complexity analysis

We elaborate on the time complexity of Algorithm 1 as follows.

Proposition 1. *The time complexity of the proposed algorithm is $\mathcal{O}(TKnm)$.*

Proof. Suppose our algorithm iterates T times and the number of ratings is d. If the rating data is compressed, then $d\ll(n\times m)$. Line 1 initializes Θ in $\mathcal{O}(K)$. Line 2 updates the posterior probability $\gamma^{(t)}$ in $\mathcal{O}(K(n+m))$. Line 3 updates the proportional parameter $\pi_k^{(t)}$ in $\mathcal{O}(K^2(n+m))$. Lines 4 through 6 determine whether the component can be removed in $\mathcal{O}(1)$. Line 7 updates the parameters $\eta^{(t)}$ in $\mathcal{O}(K^2(n+m))$. Line 8 updates the matrices **U** and **V** with $\mathcal{O}(Knm)$. Line 10 calculates $\mathbb{P}_{\Theta_k}\left(r_{i,j}=y \mid o_{i,j}=w\right)$ in $\mathcal{O}(KL^2)$. Line 11 calculates $\mathbb{P}_{\Theta_k}\left(o_{i,j}=w\right)$ in $\mathcal{O}(dL^2)$. Line 12 calculates $\mathbb{P}_{\Theta_k}\left(o_{i,j}=w \mid r_{i,j}=y\right)$ in $\mathcal{O}(dL^2)$. Line 13 calculates MGBRs in $\mathcal{O}(dKL^2)$.

Because d, K, and L are much smaller than $n\times m$, the total time complexity is $\mathcal{O}(K)+\mathcal{O}(T)*(\mathcal{O}(K(n+m))+\mathcal{O}(K^2(n+m))+\mathcal{O}(1)+\mathcal{O}(K^2(n+m))+\mathcal{O}(Knm)+\mathcal{O}(KL^2)+\mathcal{O}(dL^2)+\mathcal{O}(dKL)+\mathcal{O}(dKL^2))=\mathcal{O}(TKnm)$.

This completes the proof. Table 3 describes the complexity of Algorithm 1 briefly. □

4. Experiments

In this section, we present an experimental scheme to answer the questions related to estimating MGBRs using the MoEP model.

Table 3
Time complexity of Algorithm 1.

Lines	Complexity
Line 1	$\mathcal{O}(K)$
Line 2	$\mathcal{O}(K(n+m))$
Line 3	$\mathcal{O}(K^2(n+m))$
Lines 4–6	$\mathcal{O}(1)$
Line 7	$\mathcal{O}(K^2(n+m))$
Line 8	$\mathcal{O}(Knm)$
Line 10	$\mathcal{O}(KL^2)$
Line 11	$\mathcal{O}(dL^2)$
Line 12	$\mathcal{O}(dL^2)$
Line 13	$\mathcal{O}(dKL)$
Total	$\mathcal{O}(TKnm)$

Table 4
Datasets.

Dataset	n	m	N	Density
FilmTrust	1508	2071	35,497	1.14%
MovieLens-100K (ML-100K)	943	1682	100,000	6.30%
MovieLens-1M (ML-1M)	6040	3952	1,000,209	4.19%
MovieLens-10M (ML-10M)	71,567	10,681	10,000,054	4.60%

1) To estimate the uncertainty in a dataset, how many EP components are required?
2) Are the MGBRs calculated by our model similar to the results of SOTA algorithms?

The answers to these questions help assess the rationality of the proposed model.

The calculation was repeated 10 times with varied initial values for the parameters $\mathbf{\Theta} = \{\Theta_k \mid k \in [1, K]\}$. The final result was the average of 10 computations.

4.1. Datasets

Experiments were conducted empirically using four popular movie datasets. The sizes of these datasets range from 10 thousand to 10 million, covering several movie recommendation scenarios. Table 4 summarizes the statistics for these datasets.

FilmTrust [35]. This is a small dataset containing only user, item, and rating information crawled from the entire FilmTrust website in 2011. With 1508 users and 2071 movies, this dataset provides 35,497 movie ratings ranging from 0.5 to 4. The FilmTrust dataset is sparser than three MovieLens datasets.

MovieLens [36,37]. Many of these researchers used datasets from GroupLens. The first includes 100,000 movie ratings from 943 users for 1682 movies. The second includes 1 million movie ratings from 6040 users for 3952 movies. The last includes 10 million movie ratings from 71,567 users for 10,681 movies.

Fig. 3 shows the rating distributions for the various datasets. Fig. 3(a) shows that the rating distribution tends to conform to the following law: the higher the rating level, the greater the proportion. The **FilmTrust** dataset does not follow a normal distribution. Fig. 3(b)–(d) show that the highest proportion of ratings fall within level-4, followed by level-3, level-5, level-2, and level-1. The **ML-100K** and **ML-1M** datasets basically follow a normal distribution, whereas **ML-10M** dataset does not.

4.2. Baselines

We conducted comparative experiments from two perspectives to validate the reliability of the proposed model. On the one hand, the proposed MoEP model was compared with three MGBR methods: Said [3], SG [8], and MoG [9] to examine the estimation ability. On the other hand, our proposed method was compared with SOTA methods to verify its reliability in terms of MAE and RMSE.

- **Said** [3]: A semi-empirical MGBR evaluation method that requires human involvement in the setting of key parameters.
- **SG** [8]: A theoretical MGBR evaluation method. The uncertainty is assumed to follow an SG distribution, and the key parameter σ must be specified manually.
- **MoG** [9]: A theoretical MGBR evaluation method. The uncertainty is assumed to follow an arbitrary distribution fitted by an MoG distribution. Similarly, the key parameter K should be specified manually.
- **SOTA:** Multiple SOTA recommendation algorithms.

4.3. Parameter learning

The number of MoEP components K is the most important parameter for selection. Once K has been determined, the other parameters $\mathbf{P} = [p_1, \dots, p_K]$, $\boldsymbol{\eta} = [\eta_1, \dots, \eta_K]$ and $\boldsymbol{\pi} = [\pi_1, \dots, \pi_K]$ can be determined using the iterative process of the EM algorithm.

Fig. 4 shows the variation in K with respect to the number of iterations. Following convergence, the number of MoEP components on all datasets was two, that is, $K = 2$. For the Filmtrust dataset, there were four iterations until convergence. The ML-100K dataset required 12 iterations to achieve convergence. The ML-1M dataset required 24 iterations to achieve convergence. For the ML-10M dataset, the number of convergence iterations was 89. The dataset requires more iterations as the number of interactions between users and items increases.

The learning process for other parameters is described below. We only considered the ML-100K dataset as an example because the learning process for the other parameters is the same across all datasets. Table 5 lists the parameter values for the rounds corresponding to the beginning and end of K change. In round 0, we set K to six, the shape parameter $\mathbf{P}$ to a known distribution, and $\boldsymbol{\eta}$ and $\boldsymbol{\pi}$ to random values. In round 2, the proportions of the components $p_2 = 0.2$ and $p_6 = 2$ gradually tended to zero and were removed with the iteration of the EM. In the 11th round, three components remained. In the 12th round, two components remained and tended to converge. $\mathbf{P}$, $\boldsymbol{\eta}$, and $\boldsymbol{\pi}$ tended to stabilize in the 100th round.

Table 6 presents the optimal MoEP parameters. Regarding the FilmTrust, ML-100K, and ML-1M datasets, we obtained the same shape parameter distributions $\mathbf{P} = [0.1, 0.5]$, with a similar precision parameter $\boldsymbol{\eta}$ and component proportion $\boldsymbol{\pi}$. For the ML-10M dataset, the shape parameter distribution was $\mathbf{P} = [0.1, 1]$. The two components are applicable to the four datasets. This is because the EP components have more expressive power than the pure Gaussian components.

4.4. Comparison with other MGBR models

We conducted comparative experiments from two perspectives to validate the reliability of the proposed model. On the one hand, the MoEP estimation model was evaluated in comparison to other MGBR models. On the other hand, we compared the MGBR of the proposed model to the MAE and RMSE of SOTA algorithms. This section presents a comparison of MoEP with other models. Table 7 lists the comparison between the MGBRs of MoEP and other models in terms of MAE and RMSE.

We present the experimental results of the Said model [7]. This model computes the MGBR using a single item attribute to generate multiple ratings for each user who owns the same item genre. In the FilmTrust dataset, there were no item attributes. Consequently, the FilmTrust dataset was unsuitable for the Said model. The model can estimate the MGBR in terms of RMSE for datasets with item attributes. The model's experimental results for RMSE on datasets ML-100K, ML-1M, and ML-10M were 0.8554, 0.8786, and 0.8547, respectively. The calculated MGBRs of the RMSE for the

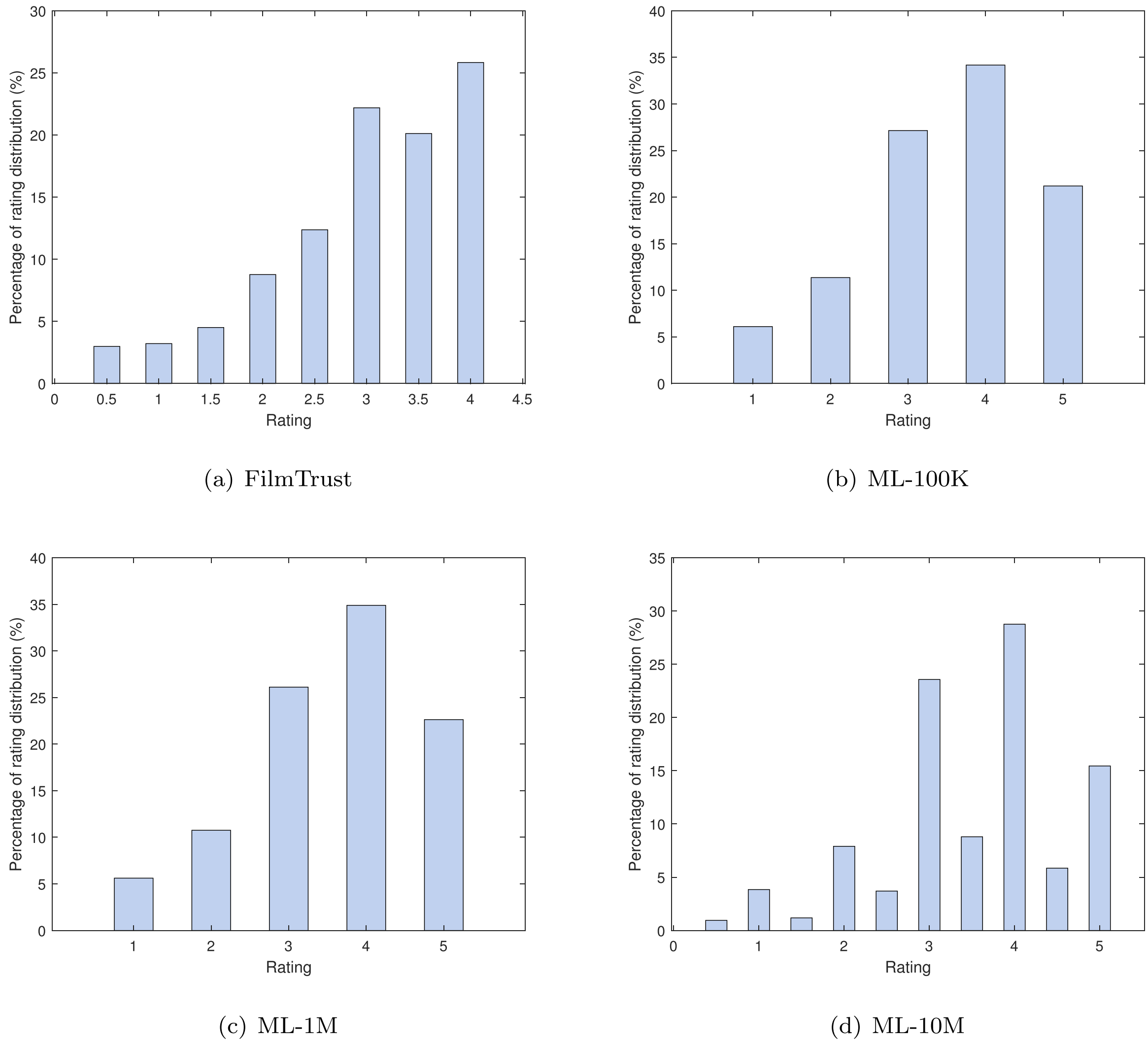


Fig. 3. datasets statistic information: (a) FilmTrust; (b) ML-100K; (c) ML-1M; (d) ML-10M.

Table 5
The learning process for other parameters using the ML-100K dataset as an example.

Iteration	K	$\mathbf{p}$	$\boldsymbol{\eta}$	$\boldsymbol{\pi}$
0	6	[0.1,0.2,0.5,1,1.8,2]	[2.2, 5.39, 8.14, 6.29, 1.06, 0.29]	[0.32, 0.03, 0.18, 0.24, 0.06, 0.17]
2	4	[0.1, 0.5, 1, 1.8]	[8.79, 4.93, 3.44, 0.45]	[0.30, 0.25, 0.22, 0.24]
10	4	[0.1, 0.5, 1, 1.8]	[14.41, 16.44, 0.99, 0.34]	[0.21, 0.29, 0.21, 0.29]
11	3	[0.1, 0.5, 1]	[14.83, 18.47, 0.93]	[0.39, 0.21, 0.40]
12	2	[0.1, 0.5]	[14.24, 19.10]	[0.35, 0.65]
100	2	[0.1, 0.5]	[10.77, 14.20]	[0.12, 0.88]

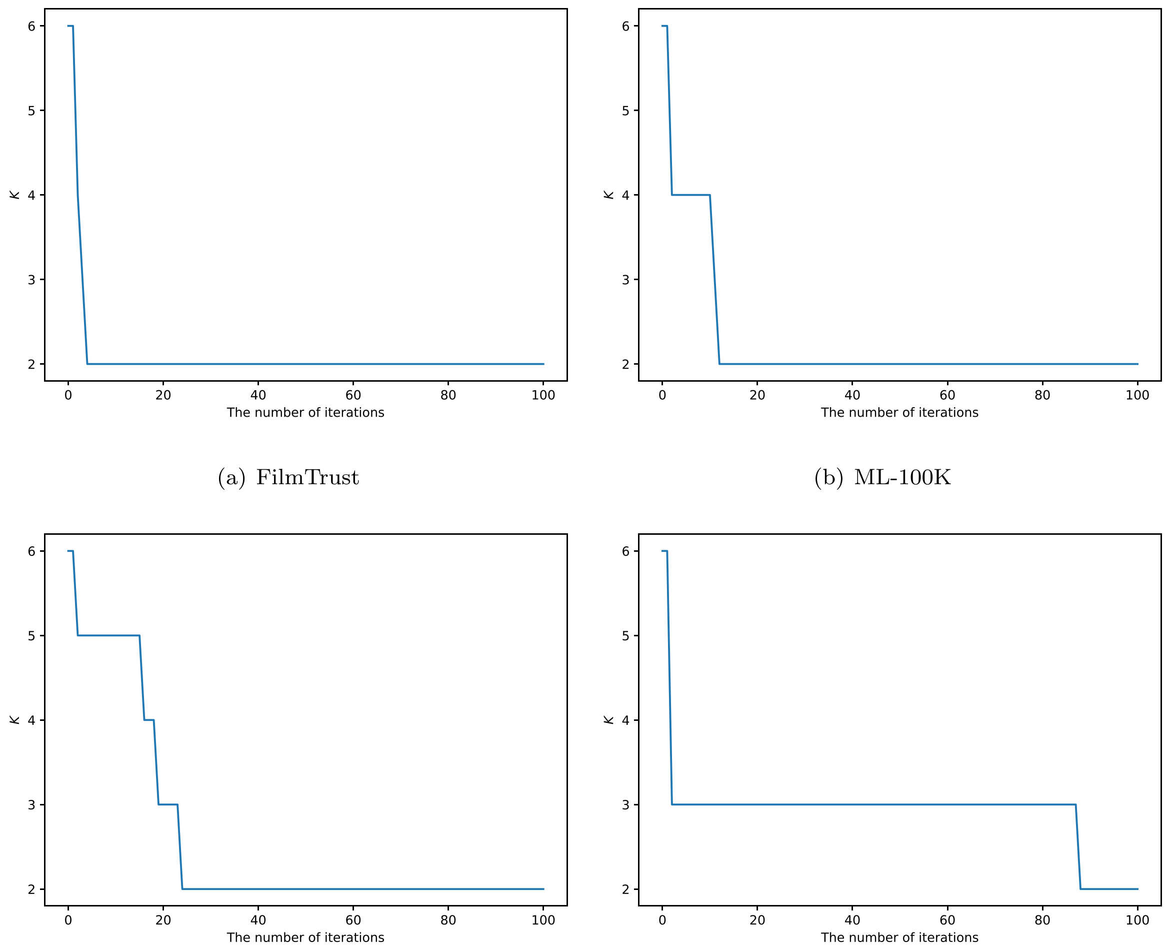
Table 6
Optimal MoEP parameters.

Dataset	K	$\mathbf{p}$	$\boldsymbol{\eta}$	$\boldsymbol{\pi}$
FilmTrust	2	[0.1, 0.5]	[13.11, 15.72]	[0.14, 0.86]
ML-100K	2	[0.1, 0.5]	[10.77, 14.20]	[0.12, 0.88]
ML-1M	2	[0.1, 0.5]	[10.86, 12.44]	[0.07, 0.93]
ML-10M	2	[0.1, 1]	[8.71, 24.28]	[0.22, 0.78]

latter two datasets were higher than those of existing SOTA algorithms, indicating that the model is unreliable.

We present the experimental results for the SG model [8]. The SG model computes the MGBR by assuming that uncertainty follows a single normal distribution with a given standard deviation σ. The model parameters were selected from those listed in Table 7. Given $\sigma = 0.8$, the MGBRs of MAE and RMSE for the FilmTrust dataset were 0.5261 and 0.7320, respectively. Given $\sigma = 0.9$, the MGBRs of the MAE and RMSE for the ML-100K dataset were 0.5957 and 0.8557, respectively. Given $\sigma = 0.9$, the MGBRs of the MAE and RMSE for the ML-1M dataset were 0.5945 and 0.8541, respectively. Given $\sigma = 0.8$, the MGBRs of the MAE and RMSE for the ML-10M dataset were 0.5286 and 0.7484, respectively.

We present the experimental results for the MoG model [9]. The MoG model calculates the MGBR by fitting the uncertainty to a Gaussian mixture distribution. Given $K = 2$, the MGBRs of MAE and RMSE for the FilmTrust dataset were 0.5295 and 0.6848, respectively. Given $K = 3$, the MGBRs of the MAE and RMSE for the ML-100K dataset were 0.5684 and 0.8278, respectively. Given $K = 3$, the MGBRs of the MAE and RMSE for the ML-1M dataset were 0.5649 and 0.8188, respectively. Given $K = 4$, the MGBRs of the



(a) FilmTrust (b) ML-100K

(c) ML-1M (d) ML-10M

Fig. 4. Variation of K with respect to the number of iterations.

Table 7
Comparison of MoEP with other MGBR models and SOTA algorithms in terms of MAE and RMSE.

Dataset	Method	MAE	RMSE
FilmTrust	SOTA	0.6300 [38]	0.7371 [39]
	SG [8]	0.5261 (σ = 0.8)	0.7320 (σ = 0.8)
	MoG (K = 2)	0.5295	0.6848
	MoEP ($K = 2$)	**0.5154**	**0.6687**
ML-100K	SOTA	0.7254 [17]	0.9234 [17]
	SG	0.5957 (σ = 0.9)	0.8557 (σ = 0.9)
	Said [7]	N/A	0.8544
	MoG (K = 3)	0.5684	0.8554
	MoEP ($K = 2$)	**0.5224**	**0.5545**
ML-1M	SOTA	0.6724 [17]	0.8554 [39]
	SG	0.5945 (σ = 0.9)	0.8541 (σ = 0.9)
	Said	N/A	0.8786
	MoG (K = 3)	0.5649	0.8188
	MoEP ($K = 2$)	**0.5136**	**0.5335**
ML-10M	SOTA	0.6121 [40]	0.8088 [40]
	SG	0.5286 (σ = 0.8)	0.7484 (σ = 0.8)
	Said	N/A	0.8547
	MoG (K = 4)	0.5006	0.7206
	MoEP ($K = 2$)	**0.5113**	**0.7032**

*Note that N/A indicates that the model cannot calculate the corresponding metric.

MAE and RMSE for the ML-10M dataset were 0.5006 and 0.7206, respectively.

4.5. Comparison with SOTA algorithms

The MGBRs and SOTA algorithms are compared in terms of MAE and RMSE in Table 7.

According to Xu et al. [38] and Bithika et al. [39], the lowest MAE and RMSE for FilmTrust are 0.6300 and 0.7371, respectively. Tang et al. [17] found that the lowest MAE and RMSE values for ML-100K are 0.7254 and 0.9234, respectively. According to Tang et al. [17] and Bithika et al. [39], the lowest MAE and RMSE on ML-1M are 0.6724 and 0.8554, respectively. According to Jiang et al. [40], the lowest MAE and RMSE on ML-10M are 0.6121 and 0.8088, respectively.

Here, we present the experimental results of the MoEP model. The MoEP model calculates the MGBR by automatically determining the number of components and key parameters. The MGBRs of the MAE and RMSE for the FilmTrust dataset were 0.5154 and 0.6687, respectively, with $K = 2$. The MGBRs of the MAE and RMSE for the ML-100K dataset were 0.5224 and 0.5545, respectively, with $K = 2$. The MGBRs of the MAE and RMSE for the ML-1M dataset were 0.5136 and 0.5335 with $K = 2$, respectively. The MGBRs of the MAE and RMSE for the ML-10M dataset were 0.5113 and 0.7032, respectively, with $K = 2$.

The MoEP coincided with other models. The MGBRs estimated using SG, MoG, and MoEP models were less than or close to the dataset's minimum MAE and RMSE. Consequently, the MoEP estimation model is considered trustworthy.

4.6. Discussions

The MoEP estimation model addresses the same problem as the SG [8] and MoG [9] models. Two new features are presented:

1. The new assumption is simpler to comprehend and more intuitive. The SG model [8] presupposes that user uncertainty has a normal distribution. Similar to the MoG [9] model, the MoEP model presupposes that user uncertainty follows an arbitrary distribution.
2. As the new model is more sophisticated, it can deal with the real data better. Owing to rating uncertainty, the SG model [8] requires user-specified variance, which is challenging, if not impossible, to obtain. In contrast, the MoEP model, like the MoG model [9], determines this parameter from the data itself. The difference is that the MoEP model requires fewer components to fit the arbitrary distribution than the MOG model [9].

Based on the aforementioned analysis, we can answer the questions proposed at the beginning of this section.

1. The MoEP model requires fewer components than the MoG model [9] to approximate any distribution of the rating datasets. The MoEP model's components are set to $1 \leq K \leq 2$ when the rating level is 5, while the MoG [9] requires a setting of $2 \leq K \leq 4$.
2. In comparison with the SG [8] and MoG [9] models, the MoEP model's estimated MGBRs are more consistent with those of SOTA algorithms; they are marginally less than the SOTA algorithms' MAE and RMSE. From another perspective, SOTA recommendation algorithms achieved satisfactory MAE or RMSE with little room for further optimization on the two metrics.

5. Conclusion and further work

The proposed MoEP model was used to estimate the MGBR in terms of MAE and RMSE and achieves excellent performance in recommender systems. Compared with other theoretical methods, the proposed model can automatically determine the number and size of key parameters and fit various uncertainties well. The estimated MGBR can be used to determine whether the current recommendation algorithm is worth improving. Furthermore, it can be used to evaluate the quality of data collected by recommender systems.

In the future, we will design novel algorithms that consider auxiliary information (such as content). We will also estimate MGBRs over more metrics (such as recall, accuracy, and F1-measure). In addition, we will consider the processed data as input (such as embedding) and attempt to find solutions with less complexity. We hope that this research will contribute to the theoretical issues in recommender systems.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Heng-Ru Zhang: Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing. **Ying Qiu:** Software, Writing – original draft, Writing – review & editing. **Ke-Lin Zhu:** Software, Writing – review & editing. **Fan Min:** Supervision, Writing – review & editing.

Data Availability

The authors do not have permission to share data.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61902328), the Applied Basic Research Project of Science and Technology Bureau of Nanchong City (SXHZ040), and Central Government Funds of Guiding Local Scientific and Technological Development (2021ZYD0003).

References

[1] F. Ricci, L. Rokach, B. Shapira, Recommender systems: introduction and challenges, Signal Process. 172 (2015) 1–34, doi:10.1007/978-1-4899-7637-6_1.
[2] X. Amatriain, J.M. Pujol, N. Oliver, I like it... i like it not: evaluating user ratings noise in recommender systems, in: International Conference on User Modeling, Adaptation, and Personalization, vol. 5535, 2009, pp. 247–258, doi:10.1007/978-3-642-02247-0_24.
[3] A. Said, B.J. Jain, S. Narr, T. Plumbaum, Users and noise: the magic barrier of recommender systems, in: International Conference on User Modeling, Adaptation, and Personalization, 2012, pp. 237–248, doi:10.1007/978-3-642-31454-4_20.
[4] M. Hsu, M. Bhatt, R. Adolphs, D. Tranel, C.F. Camerer, Neural systems responding to degrees of uncertainty in human decision-making, Science 310 (5754) (2005) 1680–1683, doi:10.1126/science.1115327.
[5] W.T. Koh, Making decisions in committees a human fallibility approach, J. Econ. Behav. Organ. 23 (2) (1994) 195–214, doi:10.1016/0167-2681(94)90067-1.
[6] Chernilo, Daniel, Reflexivity: margaret archer, 2017, 181–205, 10.1017/9781316416303.008.
[7] A. Said, A. Bellogín, Coherence and inconsistencies in rating behavior: estimating the magic barrier of recommender systems, User Model. User-Adapted Interact. 28 (2) (2018) 97–125, doi:10.1007/s11257-018-9202-0.
[8] H.-R. Zhang, F. Min, Y.-X. Wu, Z.-L. Fu, L. Gao, Magic barrier estimation models for recommended systems under normal distribution, Appl. Intell. 48 (12) (2018) 4678–4693, doi:10.1007/s10489-018-1237-8.
[9] H.-R. Zhang, J. Qian, H.-L. Qu, F. Min, A mixture-of-Gaussians model for estimating the magic barrier of the recommender system, Appl. Soft Comput. 114 (2022) 108162, doi:10.1016/j.asoc.2021.108162.
[10] M.G. Rahman, M.Z. Islam, Adaptive decision forest: an incremental machine learning framework, Pattern Recognit. 122 (2022) 108345, doi:10.1016/j.patcog.2021.108345.
[11] D.C. Mocanu, H.B. Ammar, L. Puig, E. Eaton, A. Liotta, Estimating 3D trajectories from 2D projections via disjunctive factored four-way conditional restricted Boltzmann machines, Pattern Recognit. 69 (2017) 325–335, doi:10.1016/j.patcog.2017.04.017.
[12] Z.-H. Zhuang, H.-F. Tao, Y.-Y. Chen, V. Stojanovic, W. Paszke, Iterative learning control for repetitive tasks with randomly varying trial lengths using successive projection, Int. J. Adapt. Control Signal Process. 36 (5) (2022) 1196–1215, doi:10.1002/acs.3396.
[13] L.A. Dalton, E.R. Dougherty, Optimal classifiers with minimum expected error within a Bayesian framework—Part I: discrete and Gaussian models, Pattern Recognit. 46 (5) (2013) 1301–1314, doi:10.1016/j.patcog.2012.10.018.
[14] M. Teye, H. Azizpour, K. Smith, Bayesian uncertainty estimation for batch normalized deep networks, Statistics 80 (2018) 1467–5463. https://proceedings.mlr.press/v80/teye18a.html
[15] R.-P. Shen, H.-R. Zhang, H. Yu, F. Min, Sentiment based matrix factorization with reliability for recommendation, Expert Syst. Appl. 135 (2019) 249–258, doi:10.1016/j.eswa.2019.06.001.
[16] X.-L. Xin, Y.-D. Tu, V. Stojanovic, H. Wang, K.-B. Shi, S.-P. He, T.H. Pan, Online reinforcement learning multiplayer non-zero sum games of continuous-time Markov jump linear systems, Appl. Math. Comput. 412 (2022) 126537, doi:10.1016/j.amc.2021.126537.
[17] H. Tang, G.-S. Zhao, X.-X. Bu, X.M. Qian, Dynamic evolution of multi-graph based collaborative filtering for recommendation systems, Knowledge-Based Syst. 228 (2021) 107251, doi:10.1016/j.knosys.2021.107251.
[18] T.T. Nguyen, D. Kluver, T.-Y. Wang, P.-M. Hui, M.D. Ekstrand, M.C. Willemsen, John.Riedl, Rating support interfaces to improve user experience and recommender accuracy, in: Recsys, 2013, pp. 149–156, doi:10.1145/2507157.2507188.
[19] X.-Y. Cao, Y. Chen, Q. Zhao, D.-Y. Meng, Y. Wang, D. Wang, Z.B. Xu, Robust low-rank matrix factorization under general mixture noise distributions, TIP 25 (10) (2016) 4677–4690, doi:10.1109/TIP.2016.2593343.
[20] C. Scricciolo, Posterior rates of convergence for Dirichlet mixtures of exponential power densities, Electron. J. Stat. 5 (2011) 270–308, doi:10.1214/11-EJS604.
[21] H.-R. Zhang, F. Min, Z.-H. Zhang, S. Wang, Efficient collaborative filtering recommendations with multi-channel feature vectors, Int. J. Mach. Learn. Cybern. 10 (5) (2019) 1165–1172, doi:10.1007/s13042-018-0795-8.
[22] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Item-based collaborative filtering recommendation algorithms, in: WWW, 2001, pp. 285–295, doi:10.1145/371920.372071.
[23] Y. Koren, S. Rendle, R. Bell, Advances in collaborative filtering, 2022, 91–142, 10.1007/978-1-0716-2197-4_3.
[24] Z. Huang, D. Zeng, H. Chen, A comparison of collaborative-filtering recommendation algorithms for e-commerce, IEEE Intell. Syst. 22 (5) (2007) 68–78, doi:10.1109/MIS.2007.4338497.
[25] Z.-D. Zhao, M.S. Shang, User-based collaborative-filtering recommendation algorithms on Hadoop, in: KDD, 2010, pp. 478–481, doi:10.1109/WKDD.2010.54.

[26] A. Mnih, R.R. Salakhutdinov, Probabilistic matrix factorization, NIPS 20 (2007) 1257–1264, doi:10.5555/2981562.2981720.
[27] T. Silveira, M. Zhang, X. Lin, Y.-Q. Liu, S.P. Ma, How good your recommender system is? A survey on evaluations in recommendation, Int. J. Mach. Learn. Cybern. 10 (5) (2019) 813–831, doi:10.1007/s13042-017-0762-9.
[28] A. Bellogín, A. Said, A.P. de Vries, The magic barrier of recommender systems – no magic, just ratings, in: International Conference on User Modeling, Adaptation, and Personalization, 2014, pp. 25–36, doi:10.1007/978-3-319-08786-3_3.
[29] M.P. O'Mahony, N.J. Hurley, G.C. Silvestre, Detecting noise in recommender system databases, in: International Conference on Intelligent User Interfaces, 2006, pp. 109–115, doi:10.1145/1111449.1111477.
[30] G. Jawaheer, P. Weller, P. Kostkova, Modeling user preferences in recommender systems: a classification framework for explicit and implicit user feedback, ACM Trans. Interact. Intell. Syst. 4 (2) (2014) 1–26, doi:10.1145/2512208.
[31] K. Jasberg, S. Sizov, The magic barrier revisited: accessing natural limitations of recommender assessment, in: Recsys, 2017, pp. 56–64, doi:10.1145/3109859.3109898.
[32] T. Moon, The expectation-maximization algorithm, IEEE Signal Process. Mag. 13 (6) (1996) 47–60, doi:10.1109/79.543975.
[33] T. Huang, H. Peng, K. Zhang, Model selection for Gaussian mixture models, Stat. Sin. 1 (2017) 147–169, doi:10.5705/ss.2014.105.
[34] Y. Takane, F.W. Young, J.D. Leeuw, Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features, Psychometrika 42 (1) (1977) 7–67, doi:10.1007/BF02293745.
[35] G.-B. Guo, J. Zhang, N. Yorke-Smith, A novel Bayesian similarity measure for recommender systems, in: IJCAI, 2013, pp. 2619–2625, doi:10.5555/2540128.2540506.
[36] J.L. Herlocker, J.A. Konstan, L.G. Terveen, J.T. Riedl, Evaluating collaborative filtering recommender systems, TOIS 22 (1) (2004) 5–53, doi:10.1145/963770.963772.
[37] F.M. Harper, J.A. Konstan, The movielens datasets: history and context, ACM Trans. Interact. Intell. Syst. 5 (4) (2016) 1–19, doi:10.1145/2827872.
[38] S.-S. Xu, H.-Y. Zhuang, F.-Z. Sun, S.-Q. Wang, T.-H. Wu, J.W. Dong, Recommendation algorithm of probabilistic matrix factorization based on directed trust, Comput. Electr. Eng. 93 (2021) 107206, doi:10.1016/j.compeleceng.2021.107206.
[39] B. Pal, M. Jenamani, Kernelized probabilistic matrix factorization for collaborative filtering: exploiting projected user and item graph, Recsys (2018) 437–440, doi:10.1145/3240323.3240402.
[40] S.-H. Jiang, Z.-M. Ding, Y. Fu, Heterogeneous recommendation via deep low-rank sparse collective factorization, TPAMI 42 (5) (2019) 1097–1111, doi:10.1109/TPAMI.2019.2894137.

Heng-Ru Zhang received the M.S. degree from the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China, in 2002, and the Ph.D. degree from the School of Sciences, Southwest Petroleum University, Chengdu, in 2019. He is currently a professor with Southwest Petroleum University, Chengdu. He has published more than 30 refereed papers in various journals and conferences, including Applied Soft Computing, Information Sciences, and Knowledge-Based Systems. His current research interests include recommender systems, label distribution learning and granular computing.

Ying Qiu received the B.S. degree from the School of Computer Science, Southwest Petroleum University, Chengdu, China, in 2017. She is currently pursuing the postgraduate degree with the School of Computer Science, Southwest Petroleum University, Chengdu, China. Her current research interests include federated learning and recommender systems.

Ke-Lin Zhu received the B.S. degree from the School of Computer Science, Southwest Petroleum University, Chengdu, China, in 2019. He is currently pursuing the postgraduate degree with the School of Computer Science, Southwest Petroleum University, Chengdu, China. His current research interests include recommender systems and computer vision.

Fan Min received the M.S. and Ph.D. degrees from the School of Computer Science and Engineering, University of Electronics Science and Technology of China, Chengdu, China, in 2000 and 2003, respectively. He visited the University of Vermont, Burlington, Vermont, from 2008 to 2009. He is currently a professor with Southwest Petroleum University, Chengdu. He has published more than 100 refereed papers in various journals and conferences, including the IEEE Transactions on Systems, Information Sciences, International Journal of Approximate Reasoning, and Knowledge-Based Systems. His current research interests include data mining, recommender systems, active learning and granular computing.