Neurocomputing 329 (2019) 66–74

Contents lists available at ScienceDirect

Neurocomputing

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

Elastic-net regularized latent factor analysis-based models for recommender systems

Dexian Wang[a,1], Yanbin Chen[b,1], Junxiao Guo[b,1], Xiaoyu Shi[b], Chunlin He[c,1], Xin Luo[a,2,*], Huaqiang Yuan[a,*]

[a] School of Computer Science and Network Security, Dongguan University of Technology, Dongguan, Guangdong 523808, China
[b] Chongqing Engineering Research Center of Big Data Application for Smart Cities, and Chongqing Key Laboratory of Big Data and Intelligent Computing, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China
[c] Computer School of China West Normal University, Nanchong, Sichuan 637002, China

ARTICLE INFO

Article history:
Received 9 August 2018
Revised 9 October 2018
Accepted 22 October 2018
Available online 29 October 2018

Communicated by Dr. Nianyin Zeng

Keywords:
Big data
Recommender systems
Collaborative filtering
Latent factor analysis
Elastic-net
Regularization
Latent factor distribution

ABSTRACT

Latent factor analysis (LFA)-based models are highly efficient in recommender systems. The problem of LFA is defined on high-dimensional and sparse (HiDS) matrices corresponding to relationships among numerous entities in industrial applications. It is ill-posed without a unique and optimal solution, making regularization vital in improving the generality of an LFA-based model. Current models mostly adopt l_2-norm-based regularization, which cannot regularize the latent factor distributions. For addressing this issue, this work applies the elastic-net-based regularization to an LFA-based model, thereby achieving an elastic-net regularized latent factor analysis-based (ERLFA) model. We further adopt two efficient learning algorithms, i.e., forward-looking sub-gradients and forward-backward splitting and stochastic proximal gradient descent, to train desired latent factors in an ERLFA-based model, resulting in two novel ERLFA-based models relying on different learning schemes. Experimental results on four large industrial datasets show that by regularizing the latent factor distribution, the proposed ERLFA-based models are able to achieve high prediction accuracy for missing data of an HiDS matrix without additional computational burden.

1. Introduction

WITH THE RAPID DEVELOPMENT of the World-Wide-Web, people have entered the epoch of information overload. Emerged from 1990 s, recommender systems [1–4] have proven to possess the ability to connect people with their potential favorites based on their historical behaviors and become popular in solving the problem of information overload.

A collaborative filtering (CF)-based recommender is mostly seen in industrial applications suffering information overload owing to its high efficiency [4]. It takes a high-dimensional and sparse (HiDS) matrix as the data source, which quantizes certain relationships among a huge number of entities in the target application, like the user-item preferences among millions of users and commodities [5]. Due to the impossibility of observing full relationship mappings among numerous entities, such an HiDS matrix is filled with numerous missing data rather than zeroes in conventional sparse matrices [5–9]. For instance, on the biggest e-commerce platform TaoBao [10], a user can only access a very small subset of the entire item set and vice versa, making the resultant HiDS matrix describing their connections extremely sparse with most data unknown.

In general, a CF-based recommender can be implemented in two ways, i.e., nearest neighborhood analysis (NNA) [5–7] and latent factor analysis (LFA) [2–4, 8–11]. An NNA-based recommender computes the connection weights of users or items to select their nearest neighbors. Afterwards, it predicts a user's potential preference on an untouched item based on known data in the target HiDS matrix related to the highly relevant users or items. On the other hand, an LFA-based model originates from matrix factorization (MF) techniques [2–4,8–11]. It maps involved entities into a unique and low-dimensional latent factor (LF) space, constructs a series of loss functions based on known data of the target matrix retarding desired LFs, and then minimizes resultant loss functions with respect to desired LFs to obtain a low-rank model with fine representativeness to a target HiDS matrix [2–4,8–11]. Note that an HiDS matrix's known data are far less than its unknown ones.

* Corresponding authors.
E-mail addresses: sxy@cigit.ac.cn (X. Shi), luoxin21@cigit.ac.cn, luoxin21@dgut.edu.cn (X. Luo), yuanhq@dgut.edu.cn (H. Yuan).

[1] All the authors contributed equally to this work and should be considered co-first authors.

[2] Senior Member, IEEE.

https://doi.org/10.1016/j.neucom.2018.10.046
0925-2312/

Hence, an LFA-based model focusing its known data only is highly efficient in both computation and storage [2–4,8–11]. In addition, with appropriately designed objective functions and optimization algorithms, an LFA-based model can represent a target HiDS matrix precisely. Consequently, LFA-based models are widely adopted and studied by pioneer researchers [2–4,8–11].

To perform LFA on an HiDS matrix is ill-posed and mostly bilinear, where a global optimum cannot be achieved. Moreover, Due to the imbalanced distribution of known data of an HiDS matrix, the objective function of an LFA-based model is highly imbalanced, i.e., it depends heavily on LFs related to many instances and vice versa. When addressing such a problem, the model regularization is vital in guaranteeing its generality [12]. However, existing LFA-based models [1–4,8–11,14,24,29–34], in spite of their efficiency, mostly adopt a Tikhonov regularization method, following the experience by pioneer researchers [1–4,8–11,13].

In traditional linear/bilinear problems [15–20], it proves efficient to regularize a target systems' parameter distribution through l_1 norm or elastic-net-based regularizations. But prior work mainly focuses on general optimization problems, where we cannot find the specific characteristics of an LFA-based model with regularized LF-distributions. This work aims at investigating the characteristics of LFA-based models with regularized LF-distributions. To do so, we integrate an elastic-net-based method into an LFA-based model, thereby achieving an elastic-net-regularized latent factor analysis (ERLFA)-based model. For training its LFs, we further adopt two different learning schemes, i.e., forward-looking sub-gradients and forward-backward splitting (FOBOS) [18–20] and stochastic proximal gradient descent (SPGD) [21–23], to implement its LF training. Note that we choose these two different schemes owing to (a) their efficiency in solving a elastic-net-regularized problem, and (b) the desire to validate the effects of different but efficient learning schemes in an ERLFA-based model. The main contributions of work include:

1) Two ERLFA-based models for recommender systems. After building an elastic-net-regularized objective function for LFA on an HiDS matrix, we have adopted two different optimizers, i.e., FOBOS and SPGD, to solve it for achieving LFs which incorporate the regularization effects on both of their l_1 and l_2 norms. Detailed mathematical inferences are presented during such a process. Compared with existing works [1–4,8–11,14,24,29–34], it is for the first time to see ERLFA-based models based on such these two optimizers.
2) Empirical studies on four large, real datasets collected by industrial applications. Compared with an LFA-model with l_2-norm-based regularization only [1], the proposed ERLFA models are able to achieve higher prediction accuracy for missing data and highly close computational efficiency. The resultant LFs possess the characteristics of sparsity, which indicate the feasibility of improving the performance of an LF model through further regularizing its LF distribution.

To the authors' best knowledge, such efforts have been never seen in any previous work.

The rest of this paper is organized as follows. Section II gives the preliminaries. Section III presents the ERLFA-based models. Section IV reports the experimental results. Section V reviews related works. Finally, section V concludes this paper.

2. Preliminaries

Let $U=\{u_1, u_2 \ldots u_m\}$ and $I=\{i_1, i_2 \ldots i_n\}$ denote the sets of users and items in a recommender system, $R^{|U|\times|I|}$ denote the user-item rating matrix where each element r_{ui} represents user u's preference on item i, respectively. As mentioned before, R is an HiDS matrix with most data unknown. Let Λ and Ω denote the known and unknown entry sets of R, then we naturally have $|\Lambda|\ll|\Omega|$.

An LFA-based model seeks a rank-f approximation to R based on Λ with $f\ll\min\{|U|,|I|\}$. It maps each entity in U or I into a unique f-dimensional LF space, representing each entity with an f-dimensional LF vector. Thus, U and I are represented by LF matrices $P^{|U|\times f}$ and $Q^{|I|\times f}$, respectively. The desired rank-f approximation to R is therefore denoted by $R\hat{}=PQ^{\mathrm{T}}$ whose each element is formulated by:

$$\hat{r}_{u,i} = \sum_{k=1}^{f} p_{u,k}q_{i,k} \tag{1}$$

where $p_{u,k}$ and $q_{i,k}$ denote the kth element in the uth row of P and ith row of Q, respectively.

For constructing P and Q based on Λ, it is necessary to design an objective function describing the distance between Λ and corresponding entries in $R\hat{}$. An objective function with Euclidean distance is defined as [1–4,11,14,24,29–34]:

$$\varepsilon(P,Q) = \sum_{r_{u,i}\in\Lambda}\left(r_{u,i} - \sum_{k=1}^{f} p_{u,k}q_{i,k}\right)^2. \tag{2}$$

As indicated by prior research [1–4,11,14,24,29–34], SGD proves to be efficient in training an LFA-based model. By applying SGD to (2), we commonly consider the instant loss on each training instance, formulated by:

$$\varepsilon_{u,i} = \left(r_{u,i} - \sum_{k=1}^{f} p_{u,k}q_{i,k}\right)^2, \tag{3}$$

where we compute the stochastic gradient with respect to each decision parameter in $\{p_{u,k}, q_{i,k}|k=1\sim f\}$ to perform the SGD process for minimizing the instant loss $\varepsilon_{u,i}$. This process is taken on Λ iteratively for achieving a local optimum of $\varepsilon(P,Q)$.

Note that (2) is bi-linear and ill-posed. Consequently, it is vital to regularize (2) from overfitting. According to prior work [2, 10,14,24], an l_α norm-based regularization method can be helpful under such circumstances. With SGD as the training algorithm, such regularization can be applied on the instant loss (3) [11], resulting in the following regularized instant loss:

$$\rho_{u,i} = \varepsilon_{u,i} + \mu_{u,i}^{\alpha},\ \ \mu_{u,i}^{\alpha} = \lambda\left(\|p_{u,\cdot}\|_\alpha + \|q_{i,\cdot}\|_\alpha\right); \tag{4}$$

where $p_{u,\cdot}$ and $q_{i,\cdot}$ denote the uth row in P and ith row in Q, $\mu(P, Q)$ denotes the regularization function, λ is a positive constant controlling the regularization effects, and $||\cdot||_\alpha$ computes the l_α norm of the enclosed vector, respectively.

Existing LFA-based models commonly make $\alpha=2$ to achieve the following regularization function:

$$\mu_{u,i}^{2} = \lambda\left(\|p_{u,\cdot}\|_2^2 + \|q_{i,\cdot}\|_2^2\right), \tag{5}$$

which is actually a special case of (4) adopting a l_2-norm-based regularization method, a.k.a., a Tikhonov regularization method [1–4, 8–11, 14,24,29–33], to regularize the original loss function (2). Note that a solver like stochastic gradient descent (SGD) tends to minimize the sum squared error and l_2-norm of involved decision parameters simultaneously, thereby improving the generality of a resultant model. However, (5) cannot regularize the LF distribution of a resultant model.

On the other hand, prior research regarding sparsity constrained matrix factorization [15–20] indicates that by setting $\alpha=1$ in (4), i.e., applying an l_1-norm-basedregularization method to (4), is highly helpful in regularizing the LF distribution of a resultant LF model. With such principle, we achieve the following regularization function:

$$\mu_{u,i}^{1} = \lambda\left(\|p_{u,\cdot}\|_1 + \|q_{i,\cdot}\|_1\right). \tag{6}$$

However, compared with an l_2-norm-regularized objective function (5), an l_1-nrom-regularized objective function (6) cannot improve a resultant model's generality significantly, i.e., the resultant model can suffer accuracy loss when predicting missing data of an HiDS matrix with l_1-norm-based regularization only [1–4,8–11,14,24,29–33].

Next we present elastic-net-regularized LFA-based models, which regularizes the LF distribution and improve the model generality simultaneously.

3. Elastic-net regularized LFA-based models

3.1. Elastic-net regularized learning objective

An elastic-net-based regularization method integrates a convex combination of l_1 and l_2 norms of the decision parameters into an objective function [17–23,25–28], which is called the elastic-net penalty. In our context, it formulates the regularization terms as follows:

$$\mu_{u,i}^{e} = \lambda_1\left(\|p_{u,\cdot}\|_1 + \|q_{i,\cdot}\|_1\right) + \lambda_2\left(\|p_{u,\cdot}\|_2^2 + \|q_{i,\cdot}\|_2^2\right). \tag{7}$$

where the positive constants λ_1 and λ_2 balance the effects by l_1 and l_2 norms of decision parameters. With (7), we transform the generalized loss (2) into the following regularized objective:

$$\rho_{u,i} = \left(r_{u,i} - \sum_{k=1}^{f} p_{u,k}q_{i,k}\right)^2 + \lambda_1\left(\|p_{u,\cdot}\|_1 + \|q_{i,\cdot}\|_1\right) + \lambda_2\left(\|p_{u,\cdot}\|_2^2 + \|q_{i,\cdot}\|_2^2\right). \tag{8}$$

Note that (8) turns into an l_2-norm-regularized objective with $\lambda_1=0$, and an l_1-norm-regularized objective with $\lambda_2=0$. Hence, λ_1 and λ_2 both are in the scale of (0,1). With (8), we achieve the learning objective for an ERLFA-based model. Next we show how to acquire desired LFs from (8).

Note that a standard SGD algorithm cannot solve (8) correctly since the l_1-norm-based regularization terms are non-differentiable at the zero point. Next we present two ERLFA-based models via FOBOS and SPGD, respectively.

3.2. An ERLFA-based model relying on FOBOS

FOBOS is an effective optimization framework for empirical loss minimization with regularization [18–20,25–28]. Given a regularized objective function, it split the whole optimization task into three steps, i.e., (a) unconstrained stochastic sub gradient descent, (b) regularization incorporation, and c) forward sub gradient descent of regularization.

Following such principle, we firstly consider minimizing the original objective function (2) with respect to $p_{u,k}$ through stochastic gradient descent. Let $p_{u,k}^{(t)}$ denote the status of $p_{u,k}$ after the tth iteration. Thus, for the $(t+1)$th iteration, by applying SGD to (3) with respect to $p_{u,k}$ we achieve the following interim results:

$$\frac{\partial \varepsilon_{u,i}}{\partial p_{u,k}} = -q_{i,k}^{(t)}\left(r_{u,i} - \hat{r}_{u,i}^{(t)}\right) \Rightarrow p_{u,k}^{(t+1/2)} = p_{u,k}^{(t)} + \gamma\left(r_{u,i} - \hat{r}_{u,i}^{(t)}\right)q_{i,k}^{(t)}, \tag{9}$$

where $\hat{r}_{u,i}^{(t)} = \sum_{k=1}^{f} p_{u,k}^{(t)}q_{i,k}^{(t)}$ denotes the approximation to $r_{u,i}$ after the tth iteration, $p_{u,k}^{(t+1/2)}$ denote the interim results by the unconstrained stochastic sub gradient descent, γ denotes the learning rate, respectively. Let $e_{u,i}^{(t)} = r_{u,i} - \hat{r}_{u,i}^{(t)}$, we simplify (9) into the following form:

$$p_{u,k}^{(t+1/2)} = p_{u,k}^{(t)} + \gamma e_{u,i}^{(t)}q_{i,k}^{(t)}. \tag{10}$$

With (10), we rebuild the optimization objective by incorporating the regularization function into the learning process with the following guidance: (a) the optimization objective should stay close to the interim results by the unconstrained stochastic sub gradient descent; and (b) regularization effects should be controlled with care, avoiding keeping the resultant model from the original objective. Following such principle, we reformulate the objective function as follows:

$$p_{u,k}^{(t+1)} \leftarrow \arg\min_{p_{u,k}}\left(\left(p_{u,k} - p_{u,k}^{(t+1/2)}\right)^2 + \gamma\mu_{u,i}^{e}\right), \tag{11}$$

where the learning rate γ keeps the update magnitude by the regularization consistent with that by the unconstrained stochastic sub gradient descent. Naturally, only a small part of terms in $\mu_{u,i}^{e}$ are connected with $p_{u,k}$. Hence, we select them out of $\mu_{u,i}^{e}$, reformulating (11) as:

$$p_{u,k}^{(t+1)} \leftarrow \arg\min_{p_{u,k}}\left(\left(p_{u,k} - p_{u,k}^{(t+1/2)}\right)^2 + \gamma\mu_{u,i}^{e}(p_{u,k})\right),$$
$$\mu_{u,i}^{e}(p_{u,k}) = \left(\lambda_1|p_{u,k}|_{abs} + \lambda_2\left(p_{u,k}\right)^2\right); \tag{12}$$

where $|\cdot|_{\mathrm{abs}}$ computes the absolute value of an enclosed parameter. Note that in (12) we rewrite the symbols by applying the vector operators with respect to each single element, i.e., $|\cdot|_{\mathrm{abs}}$ for $\|\cdot\|_1$ and $(\cdot)^2$ for $\|\cdot\|_2$.

Note that (12) is convex in $p_{u,k}$. Hence, we analytically solve it with respect to $p_{u,k}$ as follows:

$$\frac{\partial\left(\left(p_{u,k} - p_{u,k}^{(t+1/2)}\right)^2 + \gamma\mu_{u,i}^{e}(p_{u,k})\right)}{\partial p_{u,k}} = 0,$$
$$\Rightarrow p_{u,k} - p_{u,k}^{(t+1/2)} + \gamma\frac{\partial\mu_{u,i}^{e}(p_{u,k})}{\partial p_{u,k}} = 0. \tag{13}$$

Considering the forward sub gradient descent of regularization in (13), from (12) we see that it depends on the derivative of an absolute value function. Hence, we need to adopt the soft-thresholding strategy to analyze it in detail. Firstly, we separate the differentiable part of $\mu_{u,i}^{e}(p_{u,k})$ from the whole regularization function, resulting in:

$$p_{u,k} - p_{u,k}^{(t+1/2)} + \gamma\frac{\partial\left(\lambda_1|p_{u,k}|_{abs} + \lambda_2\left(p_{u,k}\right)^2\right)}{\partial p_{u,k}} = 0,$$
$$\Rightarrow p_{u,k} - p_{u,k}^{(t+1/2)} + \gamma\lambda_2 p_{u,k} + \gamma\lambda_1\frac{\partial|p_{u,k}|_{abs}}{\partial p_{u,k}} = 0. \tag{14}$$

Note that in (14) we slightly abuse the symbols by folding the constant two into λ_2. With (14), we achieve the following rule:

$$p_{u,k} = \frac{1}{1+\gamma\lambda_2}\left(p_{u,k}^{(t+1/2)} - \gamma\lambda_1\frac{\partial|p_{u,k}|_{abs}}{\partial p_{u,k}}\right). \tag{15}$$

Note that in (15), the sign of $p_{u,k}$ depends on the interim results $p_{u,k}^{(t+1/2)}$. Following the principle of soft-thresholding, we discuss the update rule of $p_{u,k}$ as follows:

$$p_{u,k} > 0 \Rightarrow p_{u,k} = \frac{1}{1+\gamma\lambda_2}\left(p_{u,k}^{(t+1/2)} - \gamma\lambda_1\right) > 0 \Rightarrow p_{u,k}^{(t+1/2)} > \gamma\lambda_1,$$
$$p_{u,k} < 0 \Rightarrow p_{u,k} = \frac{1}{1+\gamma\lambda_2}\left(p_{u,k}^{(t+1/2)} + \gamma\lambda_1\right) < 0 \Rightarrow p_{u,k}^{(t+1/2)} < -\gamma\lambda_1. \tag{16}$$

From (16), we see that when $p_{u,k}^{(t+1/2)}$ falls in the interval of [-$\gamma\lambda_1$,$\gamma\lambda_1$], $p_{u,k}$ tends to be zero. Thus, for an ERLFA-based model with (8) as the objective function, we achieve its update rule for $p_{u,k}$ relying on FOBOS as follows:

$$p_{u,k}^{(t+1/2)} = p_{u,k}^{(t)} + \gamma e_{u,i}^{(t)} q_{i,k}^{(t)},$$
$$p_{u,k}^{(t+1)} \leftarrow \begin{cases} \frac{1}{1+\gamma\lambda_2}\left(p_{u,k}^{(t+1/2)} - \gamma\lambda_1\right), & p_{u,k}^{(t+1/2)} > \gamma\lambda_1, \\ \frac{1}{1+\gamma\lambda_2}\left(p_{u,k}^{(t+1/2)} + \gamma\lambda_1\right), & p_{u,k}^{(t+1/2)} < -\gamma\lambda_1, \\ 0, & -\gamma\lambda_1 \le p_{u,k}^{(t+1/2)} \le \gamma\lambda_1. \end{cases} \tag{17}$$

Analogously, we derive its update rule for $q_{i,k}$ as follows:

$$q_{i,k}^{(t+1/2)} = q_{i,k}^{(t)} + \gamma e_{u,i}^{(t)} p_{u,k}^{(t)},$$
$$q_{i,k}^{(t+1)} \leftarrow \begin{cases} \frac{1}{1+\gamma\lambda_2}\left(q_{i,k}^{(t+1/2)} - \gamma\lambda_1\right), & q_{i,k}^{(t+1/2)} > \gamma\lambda_1, \\ \frac{1}{1+\gamma\lambda_2}\left(q_{i,k}^{(t+1/2)} + \gamma\lambda_1\right), & q_{i,k}^{(t+1/2)} < -\gamma\lambda_1, \\ 0, & -\gamma\lambda_1 \le q_{i,k}^{(t+1/2)} \le \gamma\lambda_1. \end{cases} \tag{18}$$

With (17) and (18), we achieve an ERLFA-based model relying on FOBOS, named ERLFA-F hereafter.

3.3. An ERLFA-based model relying on SPGD

SPGD is an extension of SGD. It generally works similar to SGD, i.e., it also trains the decision parameters on each instant loss, i.e., (4) in our context. However, it is able to address an optimization problem with the l_1-norm-based regularization through analyzing its Lipschitz condition and second-order Taylor expansion efficiently.

Following the principle of SPGD, we firstly reformulate the regularized instant loss (8) as follows:

$$\kappa_{u,i} = \left(r_{u,i} - \sum_{k=1}^{f} p_{u,k} q_{i,k}\right)^2 + \lambda_2\left(\|p_{u,\cdot}\|_2^2 + \|q_{i,\cdot}\|_2^2\right),$$
$$\rho_{u,i} = \kappa_{u,i} + \lambda_1\left(\|p_{u,\cdot}\|_1 + \|q_{i,\cdot}\|_1\right). \tag{19}$$

where $\kappa_{u,i}$ denotes the differentiable part of $\rho_{u,i}$. Moreover, $\rho_{u,i}$ is bi-convex with p_u and q_i, but convex with each single LF in the set of $\{p_{u,k}, q_{i,k} | k=1\sim f\}$ if we alternatively fix the other LFs in this set. As indicated in prior work [1–4,8–11,14,24,29–33], we can fix them at the initial state of each iteration. For instance, for the $(t+1)$th iteration, let $p_{u,k}$ be the active LF, then we fix the remaining LFs in p_u and q_i as follows

$$\forall j \in \{\{1, 2, ..., f\} - \{k\}\} : p_{u,j} = p_{u,j}^{(t)},$$
$$\forall j \in \{1, 2, ..., f\} : q_{i,j} = q_{i,j}^{(t)}; \tag{20}$$

where $p_{u,j}^{(t)}$ and $q_{i,j}^{(t)}$ denote the status of $p_{u,j}$ and $q_{i,j}$ after the tth iteration, respectively. Naturally, only part of the regularization terms are connected with the active LF under such circumstances. Thus, we can reformulate (19) as follows:

$$\kappa_{u,i}\left(p_{u,k}\right) = \left(r_{u,i} - p_{u,k} q_{i,k}^{(t)} - \sum_{j\neq k}^{f} p_{u,j}^{(t)} q_{i,j}^{(t)}\right)^2 + \lambda_2 p_{u,k}^2,$$
$$\rho_{u,i}\left(p_{u,k}\right) = \kappa_{u,i}\left(p_{u,k}\right) + \lambda_1 |p_{u,k}|_{\text{abs}}. \tag{21}$$

Note that $\rho_{u,i}(p_{u,k})$ is not differentiable at the zero point, i.e., $p_{u,k} = 0$. However, $\kappa_{u,i}(p_{u,k})$ is convex in $p_{u,k}$ and thus satisfies the following Lipschitz condition:

$$\forall\left(p'_{u,k}, p_{u,k}\right) : \left|\kappa'_{u,i}\left(p'_{u,k}\right) - \kappa'_{u,i}\left(p_{u,k}\right)\right|_{\text{abs}} \le L\left|p'_{u,k} - p_{u,k}\right|_{\text{abs}}, \tag{22}$$

where $p'_{u,k}$ denotes a stationary point of $p_{u,k}$, $\kappa'_{u,i}(p'_{u,k})$ denotes the first-order derivative of $\kappa_{u,i}(p_{u,k})$ at the stationary point of $p'_{u,k}$, and L denotes the Lipschitz constant, respectively. Naturally, we can make $p'_{u,k} = p_{u,k}^{(t)}$. Thus, the quadratic approximation to $\kappa_{u,i}(p_{u,k})$ at the point of $p_{u,k}^{(t)}$ is given by:

$$\kappa_{u,i}\left(p_{u,k}\right) \cong \kappa_{u,i}\left(p_{u,k}^{(t)}\right) + \kappa'_{u,i}\left(p_{u,k}^{(t)}\right)\left(p_{u,k} - p_{u,k}^{(t)}\right) + \frac{L}{2}\left(p_{u,k} - p_{u,k}^{(t)}\right)^2$$
$$= \frac{L}{2}\left(p_{u,k} - z_{u,k}\right)^2 + \varphi\left(p_{u,k}^{(t)}\right),$$
$$z_{u,k} = p_{u,k}^{(t)} + \frac{1}{L} e_{u,i}^{(t)} q_{i,k}^{(t)} - \frac{\lambda_2}{L} p_{u,k}^{(t)},$$
$$\varphi\left(p_{u,k}^{(t)}\right) = \kappa_{u,i}\left(p_{u,k}^{(t)}\right) - \frac{1}{2L}\left(\kappa'_{u,i}\left(p_{u,k}^{(t)}\right)\right)^2; \tag{23}$$

where we introduce an auxiliary matrix $Z^{|M|\times f}$ corresponding to P whose element $z_{u,k} = p_{u,k}^{(t)} + \frac{1}{L} e_{u,i}^{(t)} q_{i,k}^{(t)} - \frac{\lambda_2}{L} p_{u,k}^{(t)}$. Note that in (23) φ is not connected with the decision parameter $p_{u,k}$. So it can be treated as a constant during the minimization process of $\kappa_{u,i}(p_{u,k})$. Thus, with (23) we have the following inferences:

$$p_{u,k}^{(t+1)} \leftarrow \arg\min_{p_{u,k}}\left(\kappa_{u,i}\left(p_{u,k}\right) + \lambda_1 |p_{u,k}|_{\text{abs}}\right)$$
$$= \arg\min_{p_{u,k}}\left(\frac{L}{2}\left(p_{u,k} - z_{u,k}\right)^2 + \varphi\left(p_{u,k}^{(t)}\right) + \lambda_1 |p_{u,k}|_{\text{abs}}\right), \tag{24}$$

whose analytic solution is given by:

$$\frac{\partial\left(\frac{L}{2}\left(p_{u,k} - z_{u,k}\right)^2 + \varphi\left(p_{u,k}^{(t)}\right) + \lambda_1 |p_{u,k}|_{\text{abs}}\right)}{\partial p_{u,k}} = 0$$
$$\Rightarrow L\left(p_{u,k} - z_{u,k}\right) + \lambda_1 \frac{\partial |p_{u,k}|_{\text{abs}}}{\partial p_{u,k}} = 0$$
$$\Rightarrow p_{u,k}^{(t+1)} \leftarrow z_{u,k} - \frac{\lambda_1}{L} \cdot \frac{\partial |p_{u,k}|_{\text{abs}}}{\partial p_{u,k}}. \tag{25}$$

Following the principle of soft-thresholding, we discuss the update rule of $p_{u,k}$ in (25) as follows:

$$p_{u,k} > 0 \Rightarrow p_{u,k} = z_{u,k} - \frac{\lambda_1}{L} > 0 \Rightarrow z_{u,k} > \frac{\lambda_1}{L},$$
$$p_{u,k} < 0 \Rightarrow p_{u,k} = z_{u,k} + \frac{\lambda_1}{L} < 0 \Rightarrow z_{u,k} < -\frac{\lambda_1}{L}. \tag{26}$$

From (16), we see that when $z_{u,k}$ falls in the interval of $[-\lambda_1/L, \gamma\lambda_1/L]$, $p_{u,k}$ tends to be zero. Thus, we achieve the update rule for $p_{u,k}$ relying on SPGD as follows:

$$z_{u,k} = p_{u,k}^{(t)} + \frac{1}{L} e_{u,i}^{(t)} \cdot q_{i,k}^{(t)} - \frac{\lambda_2}{L} p_{u,k}^{(t)},$$
$$p_{u,k}^{(t+1)} \leftarrow \begin{cases} z_{u,k} - \frac{\lambda_1}{L}, & z_{u,k} > \frac{\lambda_1}{L}, \\ z_{u,k} + \frac{\lambda_1}{L}, & z_{u,k} < -\frac{\lambda_1}{L_1}, \\ 0, & -\frac{\lambda_1}{L_1} \le z_{u,k} \le \frac{\lambda_1}{L}. \end{cases} \tag{27}$$

Similarly, let $Y^{|N|\times f}$ be an auxiliary matrix corresponding to Q, we derive the update rule for $q_{i,k}$ as follow:

$$y_{i,k} = q_{i,k}^{(t)} + \frac{1}{L} e_{u,i}^{(t)} \cdot p_{u,k}^{(t)} - \frac{\lambda_2}{L} q_{i,k}^{(t)},$$
$$q_{i,k}^{(t+1)} \leftarrow \begin{cases} y_{i,k} - \frac{\lambda_1}{L}, & y_{i,k} > \frac{\lambda_1}{L}, \\ y_{i,k} + \frac{\lambda_1}{L}, & y_{i,k} < -\frac{\lambda_1}{L_1}, \\ 0, & -\frac{\lambda_1}{L_1} \le y_{i,k} \le \frac{\lambda_1}{L}. \end{cases} \tag{28}$$

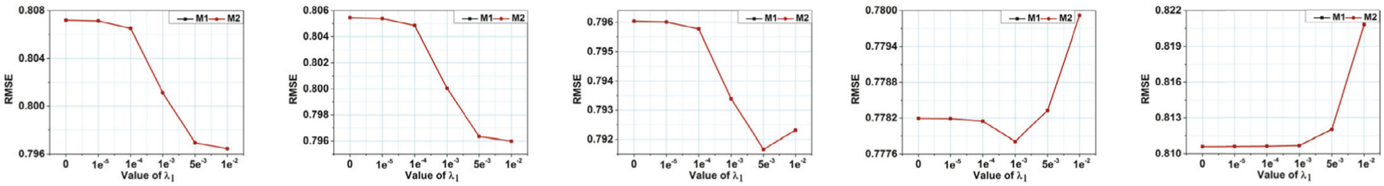
With (27) and (28), we achieve an ERLFA-based model relying on SPGD, named ERLFA-S hereafter.

4. Experimental results and analysis

4.1. General settings

4.1.1. Evaluation metrics

We mainly concern the accuracy and efficiency of tested models. For accuracy, we adopt root mean squared error (RMSE)



(a) λ_2=0 (b) λ_2=1e^{-4} (c) λ_2=1e^{-3} (b) λ_2=1e^{-2} (e) λ_2=1e^{-1}

Fig. 1. RMSE of M1 and M2 as λ_1 and λ_2 change on D1.

Table 1
Details of experimental datasets.

No.	Description	|U|	|I|	$\|\Lambda\|+\|\Gamma\|$	Density
D1	ML20M [35]	138,493	26,744	20,000,263	0.54%
D2	Flixter [36]	147,612	48,794	8,196,077	0.11%
D3	NetFlix [1]	480,189	17,770	100,480,507	1.18%
D4	Douban [12]	129,489	58,540	16,830,839	0.21%

[35,36] as the metric:

$$RMSE = \sqrt{\left(\sum_{r_{u,i}\in\Gamma} \left(r_{u,i} - \hat{r}_{u,i}\right)^2\right)/|\Gamma|}, \tag{29}$$

where $|\Gamma|$ denotes the size of a validation dataset Γ. Naturally, we have $\Gamma \cap \Lambda = \phi$.

Considering the computational efficiency, we simply record the time cost of each tested model to measure it. Note that for achieving objective results, all experiments are conducted on a bare machine with a 2.4 GHz i5 CPU and 16 GB RAM. All models are implemented in JAVA SE 7U60 to check their suitability for industrial usage.

4.1.2. Compared models

Both of the proposed models, i.e., ERLFA-F and ERLFA-S, are compared in our experiment. In the experiments, we call ERLFA-F as M1, and ERLFA-S as M2, respectively. Note that when $\lambda_1 = 0$, both M1 and M2 rely on the l_2-norm-based regularization only and lose the regularization on LF distributions. Note that in all experiments we set $f = 20$ for all testing cases.

4.1.3. Datasets

The experiments are conducted on four datasets, whose details are shown in Table 2. As shown in Table 1, all datasets are (a) high-dimensional, (b) extremely sparse, and (c) collected by industrial applications currently in use. Hence, results on them are representative.

The known entry set of each HiDS matrix is randomly split into five equally-sized, disjoint subsets. In all experiments, we adopt the 80–20% train-test settings and five-fold cross-validations, i.e., each time we select four subsets as the training set Λ to train a model predicting the remaining one subset as the testing set Γ. This process is sequentially repeated for five times to obtain the final results. The training process of a tested model terminates if (a) the number of consumed iterations reaches a preset threshold, i.e., 1000, and (b) the model converges, i.e., the error difference between two consecutive iterations is smaller than 10^{-7}.

4.2. Results

Figs. 1–4 depict the RMSE of M1 and M2 as λ_1 and λ_2 change on D1-4. Tables 2 and 4 records the RMSE and time cost of M1 and M2 with optimal λ_1 and λ_2. Tables 3 and 5 record the he RMSE and time cost of M1 and M2 with $\lambda_1 = 0$ and optimal λ_2. Fig. 5 shows the LF distribution of M1 and M2 with optimal λ_2 as λ_1 changes. From them, we have the following findings:

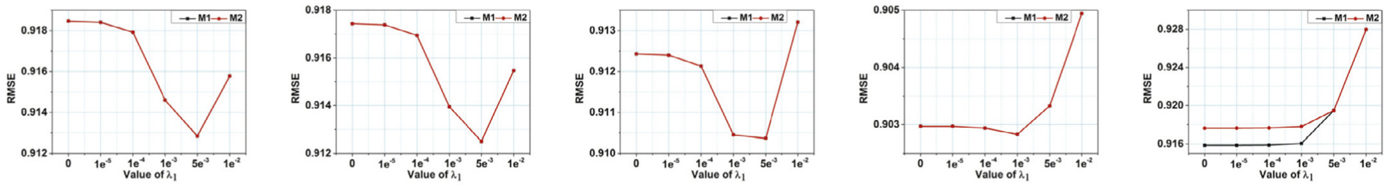
Table 2
RMSE of M1 and M2 with optimal λ_1 and λ_2.

Model	RMSE	λ_1	λ_2	Dataset
M1	0.7778	1e^{-3}	1e^{-2}	D1
M2	0.7778	1e^{-3}	1e^{-2}	
M1	0.9028	1e^{-3}	1e^{-2}	D2
M2	0.9028	1e^{-3}	1e^{-2}	
M1	0.8243	1e^{-4}	1e^{-2}	D3
M2	0.8243	1e^{-4}	1e^{-2}	
M1	0.7173	1e^{-2}	1e^{-2}	D4
M2	0.7173	1e^{-2}	1e^{-2}	

Table 3
RMSE of M1 and M2 with $\lambda_1 = 0$ and optimal λ_2.

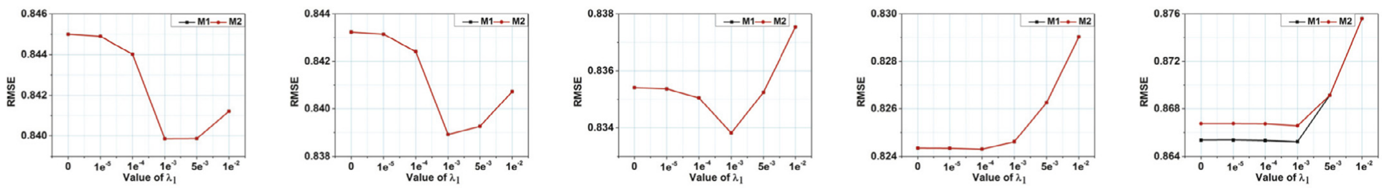
Model	RMSE	λ_1	λ_2	Dataset
M1	0.7782	0	1e^{-2}	D1
M2	0.7782	0	1e^{-2}	
M1	0.9030	0	1e^{-2}	D2
M2	0.9030	0	1e^{-2}	
M1	0.8244	0	1e^{-2}	D3
M2	0.8244	0	1e^{-2}	
M1	0.7190	0	1e^{-2}	D4
M2	0.7190	0	1e^{-2}	

a) Elastic-net-based regularization is effective in regularizing a resultant model's LF distribution. As recorded in Figs. 5–8, on all datasets, the LF distribution of M1 and M2 tends to be sparse (i.e., concentrate on zero). As mentioned before, sparse LF distribution makes the resultant model better describe the primary patterns while ignoring the less important ones. Results shown in Figs. 5–8 indicate that we can do so with the proposed ERLFA-based models.
b) To regularize the LF distribution through the elastic-net-based regularization is helpful in improving the prediction accuracy of an LFA-based model. As shown in Figs. 1–4, and Tables 2 and 3, M1 and M2's prediction accuracy for missing data of an HiDS matrix is higher than that with the pure l_2-norm-based regularization, i.e., cases with $\lambda_1 = 0$ recorded in Table 3. For instance, on D4, the lowest RMSE of M1 and M2 is 0.7173 with $\lambda_1 = 0.01$ and $\lambda_2 = 0.01$, about 0.24% lower than the RMSE at 0.7190 achieved with $\lambda_1 = 0$ and $\lambda_2 = 0.01$. Similar situations are also encountered on D1-D3. This phenomenon indicates that it is necessary to regularize the LF distribution of an LFA-based model for better representing known data from an HiDS matrix.
c) Both FOBOS and SPGD are efficient in implementing an ERLFA-based model. As recorded in Figs. 1–4 and Tables 2–3, both M1 and M2 achieve very close performance on D1-4. These results indicate that they enable very similar ERLFA-based models through different algorithms, i.e., FOBOS and SPGD.
d) Elastic-net-based regularization will not increase the computational burden. As recorded in Tables 4 and 5, M1 and M2's time costs are very similar for the following two cases: with



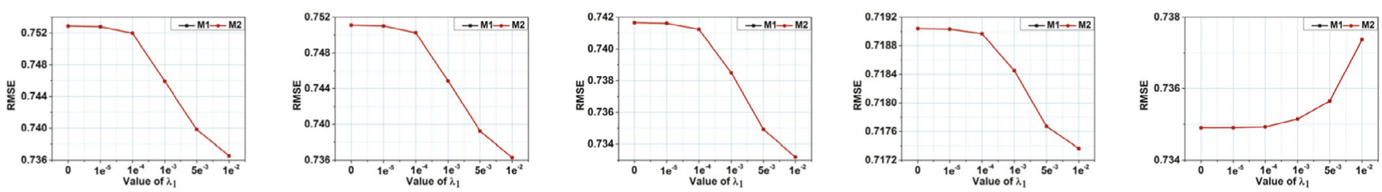
(a) λ_2=0 (b) λ_2=1e^{-4} (c) λ_2=1e^{-3} (b) λ_2=1e^{-2} (e) λ_2=1e^{-1}

Fig. 2. RMSE of M1 and M2 as λ_1 and λ_2 change on D2.



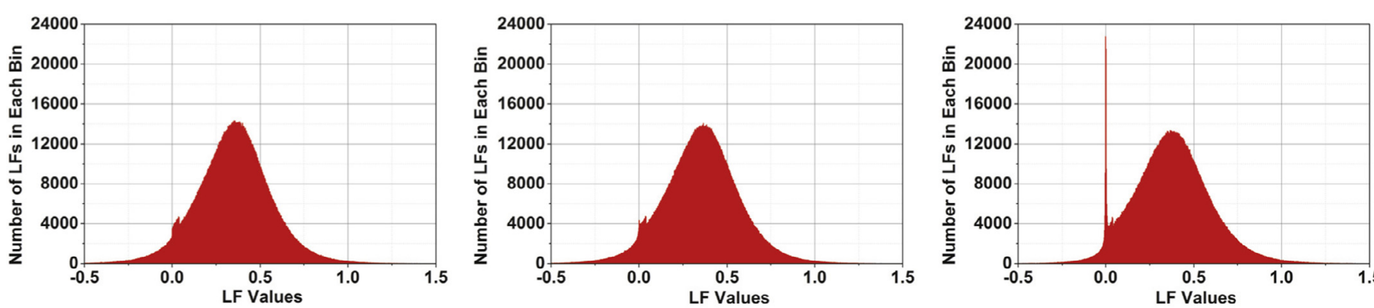
(a) λ_2=0 (b) λ_2=1e^{-4} (c) λ_2=1e^{-3} (b) λ_2=1e^{-2} (e) λ_2=1e^{-1}

Fig. 3. RMSE of M1 and M2 as λ_1 and λ_2 change on D3.



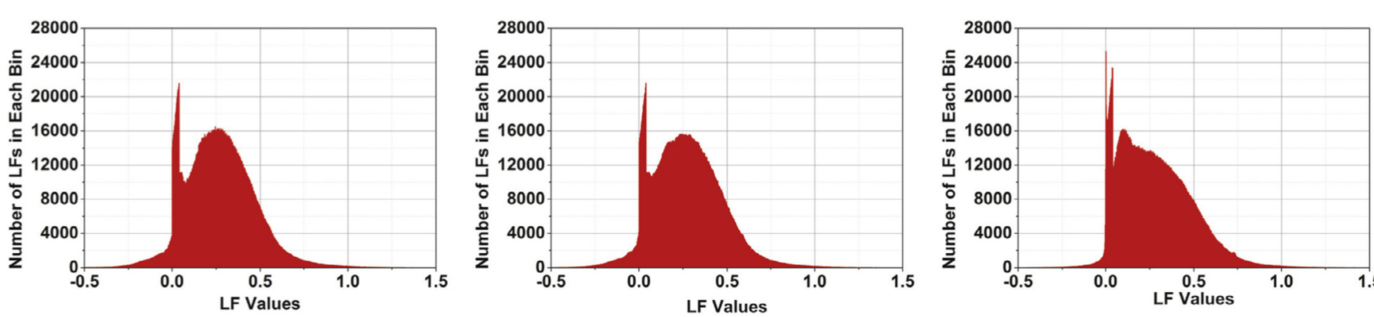
(a) λ_2=0 (b) λ_2=1e^{-4} (c) λ_2=1e^{-3} (b) λ_2=1e^{-2} (e) λ_2=1e^{-1}

Fig. 4. RMSE of M1 and M2 as λ_1 and λ_2 change on D4.



(a) λ_1=0 (b) λ_1=1e^{-3} (c) λ_1=1e^{-2}

Fig. 5. LF distribution of M1 and M2 as λ_1 changes on D1.



(a) λ_1=0 (b) λ_1=1e^{-3} (c) λ_1=1e^{-2}

Fig. 6. LF distribution of M1 and M2 as λ_1 changes on D2.

optimal λ_1 and λ_2, and with $\lambda_1 = 0$ and optimal λ_2. Note that when $\lambda_1 = 0$, an ERLFA-based model is decreased to an LFA-based model relying on pure l_2-norm-based regularization. Hence, records in Tables 4 and 5 indicate that the computational burden of an ERLFA-based model is very close to that of an LFA-based model.

e) To summarize, owing to the positive effects brought by the elastic-net-regularization and efficient training algorithms FOBOS and SPGD, the proposed ERLFA-F and ERLFA-S models are

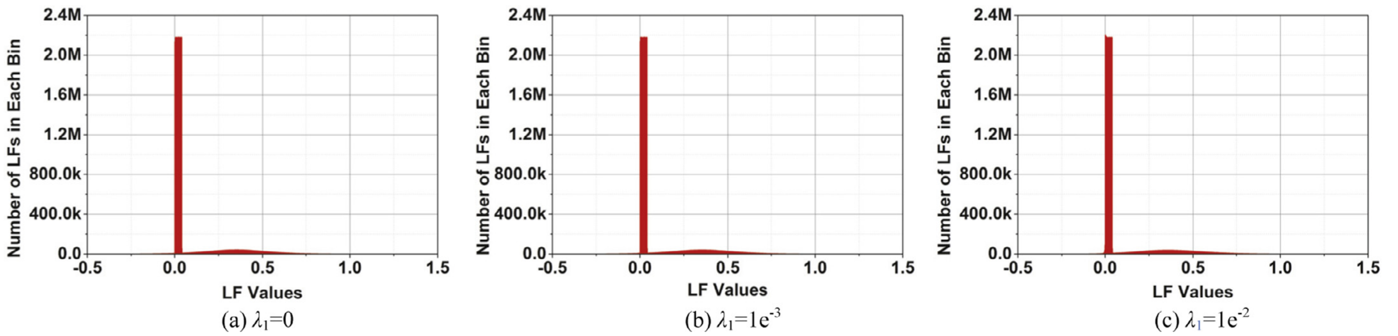


Fig. 7. LF distribution of M1 and M2 as λ_1 changes on D3.

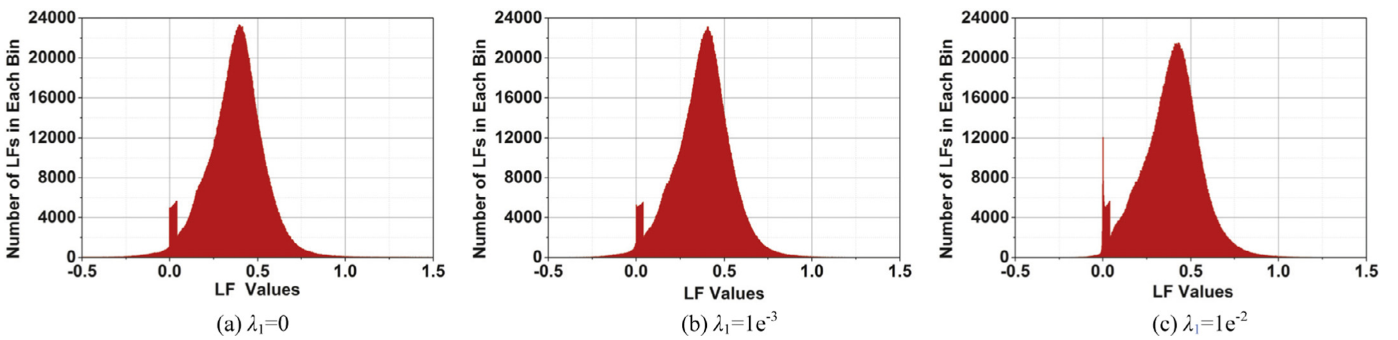


Fig. 8. LF distribution of M1 and M2 as λ_1 changes on D4.

Table 4
Time cost of M1 and M2 with optimal λ1 and λ2.

Model	Time Cost (ms)	λ_1	λ_2	Dataset
M1	402,292	1e^{-3}	1e^{-2}	D1
M2	400,936	1e^{-3}	1e^{-2}	
M1	210,501	1e^{-3}	1e^{-2}	D2
M2	214,936	1e^{-3}	1e^{-2}	
M1	5,047,575	1e^{-4}	1e^{-2}	D3
M2	4,990,826	1e^{-4}	1e^{-2}	
M1	381,324	1e^{-2}	1e^{-2}	D4
M2	381,324	1e^{-2}	1e^{-2}	

Table 5
Time cost of M1 and M2 with $\lambda_1 = 0$ and optimal λ_2.

Model	Time Cost (ms)	λ_1	λ_2	Dataset
M1	400,460	0	1e^{-2}	D1
M2	403,178	0	1e^{-2}	
M1	209,857	0	1e^{-2}	D2
M2	211,698	0	1e^{-2}	
M1	5,460,568	0	1e^{-2}	D3
M2	5,520,017	0	1e^{-2}	
M1	372,259	0	1e^{-2}	D4
M2	380,121	0	1e^{-2}	

able to achieve higher prediction accuracy for missing data of HiDS matrices than an LFA-model with pure l_2-norm-based regularization. Such performance gain is achieved through regularizing the LF distribution of the resultant model, without additional burden in computation.

5. Related works

LFA-based models are becoming increasingly important for performing pattern analysis and knowledge discovery on HiDS data from recommender systems [1–4,8–11,14,24,29–33]. Takács et al. propose the biased, regularized, incremental and simultaneous matrix factorization model [3], which integrates linear biases into an LF model, regularizes the objective function with Tikhonov regularization, and adopts SGD as the learning scheme. Koren and Bell propose the singular value decomposition plus-plus model [1], which further integrates the global statistics into the model for refinement of the statistic bias hidden in the data. It also contains the neighborhood regularization part for reflecting the similarity among involved users and items. Salakhutdinov and Mnih propose the probabilistic matrix factorization model [30], which adopts the probabilistic graph to construct an objective function based on the known data of an HiDS matrix, and further adopts the momentum effect during the SGD-based training process for accelerating the resultant model's convergence. Luo et al. propose the incremental and regularized matrix factorization model [29], which carefully models the increment on involved LFs relying on newly-arrived ratings for achieving high computational efficiency and satisfactory prediction accuracy when addressing data flows in recommender systems. They further propose the non-negative latent factor model [2,10,14,24], which implements the single LF-dependent, non-negative and multiplicative update for training non-negative LFs on HiDS matrices with high efficiency in both computational and storage. For implementing second-order LFA on HiDS matrices, Luo et al. adapt the Hessian-free optimization technique to the problem of LFA, and propose the - Hessian-free optimization based second-order latent factor model [32,33], which can commonly achieve higher prediction accuracy for missing data than LF models relying on first-order solvers at the cost of additional computation burden.

To perform LFA on an HiDS matrix is ill-posed. Hence, regularization techniques are vital for an LFA-based model in its generality. Existing LFA-based models [1–4,8–11,14,24,29–33], mostly adopt a Tikhonov regularization, which cannot affect the distribution of resultant LFs. As indicated by prior work [15–20], in a linear (or bilinear) model each entity only correlates with a part but not all of the feature dimensions, implying that it belongs to a few but not all of the entity clusters. In our context, a user/item can only be closely connected with a few LF dimensions out of the LF space because it cannot belong to all of the potential user/item clusters. From this point of view, it appears necessary to further regularize the LF distributions for describing the cluster tendencies of involved entities more precisely, thereby achieving performance gain.

In traditional linear/bilinear problems [15–20], it has proven to be effective to regularize a target systems' parameter distribution through l_1 norm or elastic-net-based regularizations. For instance, Tibshirani [16] proposes least absolute shrinkage and selection operator where the objective function is regularized by l_1 norm, thereby making the decision parameters become sparse after training. Zou et al. propose the elastic net-based regularization for feature selection [17], where many resultant parameters become zeroes after training, which are designated as the trivial features. Mairal et al. [15] propose a sparse matrix factorization model under sparseness constraints, where the factorization process on full matrices like images is regularized with l_1 norm of the decision parameters, thereby making the resultant features sparse for better describing clusters of involved entities. However, prior work mainly focuses on general optimization problems, where we cannot find the specific characteristics of an LFA-based model with regularized LF-distributions. Note that the problem of LFA is defined on an HiDS matrix where most data are missing (but still meaningful) and only a tiny portion of data are given for defining the model and conducting the training. With l_1 norm or elastic-net-based regularizations, its model building and parameter training are greatly changed and should be carefully investigated.

The proposed model in this work, i.e., ERLFA-F and ERLFA-S, integrates the l_1 norm-based regularization effects into an LFA-based model for regularizing its LF distributions, thereby improving its performance. Although they rely on different solvers, i.e., FOBOS for the former and SPGD for the latter, they both work well in manipulating the LF distribution of the resultant model. Hence, they achieve higher prediction accuracy for missing data than an LFA-based model relying on only l_2 norm-based regularization. However, they further integrate a hyper parameter, i.e., λ_1, into the model for controlling the regularization effects on l_1 norm. This parameter should be tuned carefully, thus perplexing the parameter tuning process. How to make it self-adaptive remains an open issue. We plan to address it in the future.

6. Conclusions

This work focuses on implementing ERLFA-based models for recommender systems. To do so, an elastic-net-based regularization method is applied to an LFA model, thereby regularizing its LF distributions through the l_1-norm-based terms in the regularization scheme. We further adopt two different learning algorithms, i.e., FOBOS and SPGD to train desired LFs, thereby achieving ERLFA-F and ERLFA-S models. Detailed empirical studies on four large industrial datasets show that with the elastic-net regularization, the LF distribution of the resultant model is regularized, and the prediction accuracy for missing data of HiDS matrices is improved accordingly.

This work investigates the effects by elastic-net regularization in an LFA-based model relying on an Euclidean distance-based objective function. It is interesting to validate its performance in LFA-based models relying on other objective functions such as Kullback-Leibler divergence. Moreover, it is highly interesting to investigate the compatibility of the proposed models with the evolutionary computing methods [37–42]. We plan to address these issues in the future.

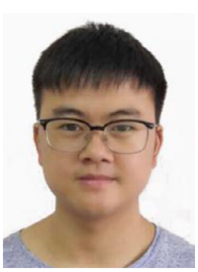
Acknowledgments

This research is supported in part by the National Natural Science Foundation of China under grants 61772493, 91646114 and 61602434, in part by the National Key Research and Development Program of China under grant 2017YFC0804002, in part by the Chongqing Cultivation Program of Innovation and Entrepreneurship Demonstration Group under grant cstc2017kjrc-cxcytd0149, in part by the Chongqing Research Program of Technology Innovation and Application under grant cstc2017rgzn-zdyfX0020, cstc2017zdcy-zdyf0554 and cstc2017rgzn-zdyf0118, in part by the Chongqing Overseas Scholars Innovation Program under Grant cx2017012 and cx2018011, and in part by the Pioneer Hundred Talents Program of Chinese Academy of Sciences.

References

[1] Y. Koren, R. Bell, Advances in collaborative filtering, in: F. Ricci, L. Rokach, B. Shapira, P.B. Kantor (Eds.), Recommender Systems Handbook, Springer US, New York, 2011, pp. 145–186.
[2] X. Luo, M.-C. Zhou, Y.-N. Xia, Q.-S. Zhu, A.C. Ammari, A. Alabdulwahab, Generating highly accurate predictions for missing QoS data via aggregating nonnegative latent factor models, IEEE Trans. Neural Netw. Learn. Syst. 27 (3) (2016) 524–537.
[3] G. Takács, I. Pilászy, Bottyán Németh, D. Tikky, Scalable collaborative filtering approaches for large recommender systems, J. Mach. Learn. Res. 10 (2009) 623–656 Mar.
[4] Y. Shi, M. Larson, A. Hanjalic, Collaborative filtering beyond the user-item matrix: a survey of the state of the art and future challenges, ACM Comput. Surv. 47 (1) (2014) 1–45.
[5] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, IEEE Trans. Knowl. Data Eng. 17 (6) (2005) 734–749.
[6] Y. Cai, H. Leung, Q. Li, H. Min, J. Tang, J. Li, Typicality-based collaborative filtering recommendation, IEEE Trans. Knowl Data Eng. 26 (2014) 766–779 Mar.
[7] L. Yao, Q.Z. Sheng, A.H.H. Ngu, J. Yu, A. Segev, Unified collaborative and content-based web service recommendation, IEEE Trans. Serv. Comput. 8 (3) (2015) 453–466.
[8] H. Thomas, Latent semantic models for collaborative filtering, ACM Trans. Inf. Syst. 22 (1) (2004) 89–115.
[9] K. Huang, N.D. Sidiropoulos, A. Swami, Non-negative matrix factorization revisited: uniqueness and algorithm for symmetric decomposition, IEEE Trans. Signal Process. 62 (1) (2013) 211–224.
[10] X. Luo, M. Shang, S. Li, Efficient extraction of non-negative latent factors from high-dimensional and sparse matrices in industrial applications, in: Proceedings of the 16th IEEE International Conference on Data Mining, 2016, pp. 311–319.
[11] D. Yu, Y. Liu, Y. Xu, and Y. Yin, "Personalized QoS prediction for web services using latent factor models." In Proceedings of the 11th IEEE International Conference Serv. Comput., pp. 107–114.
[12] H. Ma, D. Zhou, C. Liu, M.R. Lyu, I. King, Recommender systems with social regularization, in: Proceedings of the 4th ACM International Conference on Web Search and Data Mining, 2014, pp. 287–296.
[13] B. Jin, C. Che, K. Yu, Y. Qu, L. Guo, C. Yao, R. Yu, Q. Zhang, Minimizing legal exposure of high-tech companies through collaborative filtering methods, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 127–136.
[14] X. Luo, M. Zhou, Y. Xia, Q. Zhu, An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems, IEEE Trans. Ind. Inform. 10 (2) (2014) 1273–1284.
[15] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. 58 (1) (1996) 267–288.
[16] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc. 67 (2) (2005) 768-768.
[17] J. Mairal, F. Bach, J. Ponce, G. Sapiro, Online learning for matrix factorization and sparse coding, J. Mach. Learn. Res. 11 (1) (2009) 19–60.
[18] Y. Singer, J.C. Duchi, Efficient learning using forward-backward splitting, Adv. Neural Inf. Process. Syst. 22 (2009) 495–503.
[19] J. Duchi, S. ShalevShwartz, Y. Singer, Chandra, Tushar, Efficient projections onto the l_1-ball for learning in high dimensions, in: Proceedings of the 25th International Conference on Machine Learning, 2008, pp. 272–279.
[20] H. Li, N. Chen, L. Li, Error analysis for matrix elastic-net regularization algorithms, IEEE Trans. Neural Netw. Learn. Syst. 23 (5) (2012) 737–748.
[21] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (1) (2009) 183–202.
[22] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, 2009.
[23] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, Belmont, Massachusetts, 1999.
[24] X. Luo, M.-C. Zhou, S. Li, Z.-H. You, Y.-N. Xia, Q.-S. Zhu, A nonnegative latent factor model for large-scale sparse matrices in recommender systems via alternating direction method, IEEE Trans. Neural Netw. Learn. Syst. 27 (2016) 579–592.
[25] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, IEEE Comput. 42 (8) (2009) 30–37 Aug.

[26] Z. Lu, H. Bai, B. Sun, A gradient-based algorithm for optimizing sensing matrix with normalization constraint, in: Proceedings of the IEEE 11th International Conference on Industrial Electronics and Applications, 2016, pp. 2376–2380.
[27] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Stat. 32 (2) (2004) 407–451.
[28] J. Langford, L. Li, T. Zhang, Sparse online learning via truncated gradient, J. Mach. Learn. Res. 10 (2) (2008) 777–801.
[29] X. Luo, Y. Xia, Q. Zhu, Incremental collaborative filtering recommender based on regularized matrix factorization, Knowl. Based Syst. 27 (3) (2012) 271–280.
[30] R. Salakhutdinov, A. Mnih, Probabilistic matrix factorization, Adv. Neural Inf. Process. Syst. 20 (2008) 1257–1264.
[31] N. Ding, Y. Qi, R. Xiang, I. Molloy, N. Li, "Nonparametric bayesian matrix factorization by power-EP, in: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics, „ SardiniaItaly, 2010, pp. 169–176.
[32] X. Luo, M.-C. Zhou, S. Li, Z.-H. You, Y.-N. Xia, Q.-S. Zhu, H. Leung, An efficient second-order approach to factorizing sparse matrices in recommender systems, IEEE Trans. Ind. Inform. 11 (4) (2015) 946–956 Aug.
[33] X. Luo, M. Zhou, S. Li, Y. Xia, Z. You, Q. Zhu, H. Leung, Incorporation of efficient second-order solvers into latent factor models for accurate prediction of missing QoS data, IEEE Trans. Cybern. 48 (4) (2018) 1216–1228.
[34] X. Luo, M. Zhou, S. Li, Y. Xia, Z. You, Q. Zhu, H. Leung, Incorporation of efficient second-order solvers into latent factor models for accurate prediction of missing QoS data, IEEE Trans. Cybern. 48 (4) (2018) 1216–1228.
[35] J.A. Konstan, B.N. Miller, D. Maltz, J.L. Herlocker, L.R. Gordon, J. Riedl, Grouplens: applying collaborative filtering to usenet news, Commun. ACM 40 (1997) 77–87.
[36] J. Mohsen, E. Martin, A matrix factorization technique with trust propagation for recommendation in social networks, in: Proceedings of the Fourth ACM Conference on Recommender Systems, „ Barcelona Spain, 2010, pp. 135–142.
[37] W. Luan, G. Liu, C. Jiang, L. Qi, Partition-based collaborative tensor factorization for POI recommendation, IEEE/CAA J. Autom. Sin. 4 (3) (2017) 437–446.
[38] N. Zeng, Z. Wang, Y. Li, M. Du, X. Liu, A hybrid EKF and switching PSO algorithm for joint state and parameter estimation of lateral flow immunoassay models, IEEE/ACM Trans. Comput. Biol. Bioinform. 9 (2) (2012) 321–329.
[39] N. Zeng, Z. Wang, Y. Li, M. Du, X. Liu, Inference of nonlinear state-space models for sandwich-type lateral flow immunoassay using extended Kalman filtering, IEEE Trans. Bio-med. Eng. 58 (7) (2011) 1959–1966.
[40] N. Zeng, Z. Wang, H. Zhang, Inferring nonlinear lateral flow immunoassay state-space models via an unscented Kalman filter, Sci. China Inf. Sci. 59 (11) (2016) 112204.
[41] N. Zeng, Z. Wang, B. Zineddin, Y. Li, M. Du, L. Xiao, X. Liu, T. Young, Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach, IEEE Trans. Med. Imaging 33 (5) (2014) 1129–1136.
[42] N. Zeng, H. Zhang, Y. Li, J. Liang, A.M. Dobaie, Denoising and deblurring gold immunochromatographic strip images via gradient projection algorithms, Neurocomputing 247 (2017) 165–172.



Dexian Wang received the B.S. degree in computer science from China West Normal University, Nanchong, China, in 2016. He is currently pursuing a master's degree at China West Normal University and is a visiting student at the Chongqing Institute of Green and Intelligent Technology. His research interests include big data analysis, recommender system, and their applications.

Yanbin Chen is currently an undergraduate student at the School of Computer Science, Chongqing University. His research interests include big data analysis, recommender system, and their application.

Junxiao Guo received the B.S. degree in Physics from Drexel University, Philadelphia, the United States in 2017. He is currently visiting the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China, as a research assistant. His research interests include big data analysis, signal processing, machine learning, and neural networks.

Xiaoyu Shi received the B.S. degree in computer science from the PLA Information Engineering University, Zhengzhou, China, in 2007, and the Ph.D. degree in computer science from the University of Electronic Science and Technology of China, Chengdu, China, in 2015. He joined the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China in 2015, and is currently an associate Professor of computer and science and engineering. His research interests include recommender system, cloud computing, green computing and big data applications.

Chunlin He received the B.S. degree in computer science from Southwest University, Chongqing, China, in 1994, and the MS from China West Normal University, Nanchong, China, in 2006. He is currently a professor with School of computer science in China West Normal University. His research interests include big data analysis, image processing.

Xin Luo (M'14-SM'17) received the B.S. degree in computer science from the University of Electronic Science and Technology of China, Chengdu, China, in 2005, and the Ph.D. degree in computer science from Beihang University, Beijing, China, in 2011. He joined the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China in 2016, as a professor of computer science and engineering. In 2018, he joined the Dongguan University of Technology, Dongguan, Guangdong, China, as a distinguished professor of computer science. His research interests include big data analysis, artificial intelligence and intelligent control. He has published more than 90 papers (including 20 + IEEE Transactions papers) in his related areas.