A novel Initialization method of Fixed Point Continuation for Recommendation Systems

Jianli Zhao[1*], Tianheng Zhang[1] , Qiuxia Sun[2], Huan Huo[3] ,Maoguo Gong[1,4]

[1] College of Computer Science & Engineering, Shandong University of Science and Technology
Qingdao, China

[2] College of Mathematics and Systems Science, Shandong University of Science and Technology
Qingdao, China

[3] Faculty of Engineering and Information Technology University of Technology Sydney, Australian

[4] Key Laboratory of Intelligent Perception and Image Understanding, Ministry of Education, Xidian University,
Xi'an 710071, China

*Corresponding author : Jianli Zhao

jlzhao@sdust.edu.cn, m17806237351@163.com, qiuxiasun@sdust.edu.cn,
Huan.Huo@uts.edu.au, mggong@mail.xidian.edu.cn

Abstract In recent years, the problem of matrix completion based on rank minimization has received widespread attention in machine learning. The tightest convex relaxation of this problem is the linearly constrained nuclear norm minimization. Fixed point continuation (FPC), as a representative nuclear norm relaxation matrix completion algorithm, has been proven to perform well in theories and experiments. However, the traditional FPC algorithm initializes the matrix to be completed by zero, which does not make full use of the shrinkage characteristics of the singular value shrinkage operator on the matrix elements and the known field data information, then will lead to slow convergence and poor accuracy. Aiming at this problem, this paper analyzes the shrinkage properties of matrix elements in the iterative process of the FPC algorithm. Combined with the known rating information in the recommendation systems, a new initialization method of overestimation based on FPC is proposed, and it is applied to the rating prediction in the recommendation systems. The experimental results show that the initialization method proposed in this paper greatly improves the algorithm efficiency and prediction accuracy.

Key words Matrix completion, Nuclear norm minimization, Fixed point continuation, Recommended systems, Initialization

1 Introduction

Recommendation systems (RSs) (Ko, Lee, Park, & Choi, 2022) are the most popular information filtering systems in the Internet world. They have been used in many areas such as e-commerce, music video sites, and online advertising. From the types of recommendation algorithms, it can be divided into collaborative filtering (Tang, Zhao, Bu, & Qian, 2021), content-based filtering (Pujahari, & Sisodia, 2022), and hybrid filtering (Biswas, & Liu, 2021). Collaborative filtering is the most successful and widely used recommendation algorithm, which analyzes the historical behavior data of users and explores their preferences in order to recommend items of interest to them. The main idea of the collaborative filtering method is to find N users whose ratings are very similar to U's when making recommendations for user U, and then estimate U's rating based on the ratings of these users. The advantage of this method is that it does not need to make the feature selection. The disadvantage is that it is difficult to find users who rated the same items due to the sparse rating matrix, and it can not effectively solve the cold-start problem. Content-based filtering is an algorithm for a recommendation based on item profiles. The main idea is to find some items that are similar to the items that user U ratings highly, and then recommend them to user U. The advantage of this method is that it does not need the data of other users, and makes recommendations based on the attribute information of items, so there is no cold-start problem and sparse problem. Its main disadvantage is that it is difficult to find appropriate features. In order to get the advantages of the above two methods, some researchers add content-based filtering to collaborative filtering and get a hybrid filtering method.

As one of the latent factor models in collaborative filtering, matrix completion (Ramlatchan, Yang, Liu, Li, Wang, & Li, 2018; Chen, & Wang, 2022) is widely used in personalized recommendation systems. It treats the users' ratings of items as a matrix, and completes the missing values according to the observed entries in the matrix. It is generally believed that the factors that affect users' ratings of items are limited, so the rank of the rating matrix is low. According to this characteristic, the matrix completion problem can be modelled as:

$$\begin{aligned} &\min rank(X) \\ &s.t. X_{ij} = M_{ij}, (i, j) \in \Omega \end{aligned} \tag{1}$$

Where X and M are the matrices of $m \times n$, and Ω is the index set of the observed entries in the matrix. The significance of this model is that after the missing entries in the matrix are completed, the matrix still maintains good properties, that is, the rank is as low as possible. Unfortunately, the time complexity of solving this problem is exponential, which is an NP-Hard problem. One method is to replace the rank of a matrix with the nuclear norm of the matrix, and obtain the nuclear norm relaxation model of matrix completion:

$$\begin{aligned} &\min \|X\|_* \\ &s.t. X_{ij} = M_{ij}, (i, j) \in \Omega \end{aligned} \tag{2}$$

Where $\|X\|_* = \sum_{k=1}^{n} \sigma_k$, and σ_k represent the kth singular value of the matrix X . Since the solution of the model involves complex singular value decomposition, it is still a challenging problem, and many scholars have conducted in-depth research on it.

Fixed point continuation (Ma, Goldfarb, & Chen, 2011) is an optimization algorithm for solving matrix completion problems. It transforms problem (2) into an unconstrained optimization problem for an iterative solution. It has been proved that, compared with the classical matrix completion algorithm such as the interior point method (Karmarkar, 1984), its efficiency is greatly improved. However, the FPC algorithm initializes the matrix to be completed by zero, which will affect the effect of the algorithm. Aiming at this problem, we improved the matrix initialization of the FPC algorithm from two aspects: The first is to study the change of matrix entries in the iterative process of the FPC algorithm, and apply this property to initialization to accelerate the convergence speed of the algorithm; The second is to improve the accuracy of the recommendation algorithm by integrating the rating prior information into initialization from the aspect of resource utilization.

The contributions of this paper can be summarized as follows:

• We analyze and obtain the contraction effect of the singular value shrinkage operator on the matrix entries. Using this property and combining the known rating information in the recommendation systems, an overestimation initialization method of the FPC algorithm is proposed. Compared with the original initialization, the prediction accuracy and algorithm efficiency are greatly improved.

• We take the upper limit of the rank as the parameter k and estimate it in advance, and only calculate the first k largest singular values during the iteration process, which greatly improves the efficiency of the algorithm.

The content structure of the paper is as follows: Section 2 introduces the research status and related work of matrix completion algorithms; Section 3 introduces the fixed point continuation algorithm of matrix completion and our initialization method; Section 4 introduces our experimental results and performs analysis; the last Section 5 is the summary and outlook of the work.

2 Related work

In the recommendation systems, the rating prediction model based on matrix completion has been widely used. The methods of matrix completion can be classified into rank minimization models, matrix decomposition models, and the models combined with the neural network.

At present, the research on the matrix completion model based on rank minimization has been very intensive. Fazel (2002) proved that the best convex approximation of matrix rank function is the nuclear norm of the matrix, which is similar to the skill of relaxing vector L_0 norm to vector L_1 norm in compressed sensing. In order to solve the problem easily, the rank function of the matrix is replaced by the nuclear norm. This kind of algorithm usually realizes the low-rank properties of the rating matrices through singular value shrinkage and completes the missing data through the closed-form optimal solution. Subsequently, a series of algorithms for solving the matrix completion model based on the relaxation of matrix nuclear norm has been proposed. Cai et al. (2010) proposed a simple first-order optimization method called the singular value threshold (SVT) algorithm; Tho et al. (2010) proposed an accelerated proximal gradient (APG) algorithm using the Nesterov technique. After using the continuation technique and line search technique, the convergence speed of the algorithm is much faster than that of SVT and other algorithms; Lin et al. (2010) regarded the matrix completion problem as a special case of matrix restoration, and proposed an inexact augmented Lagrangian multiplier (IALM) method by using imprecise strategy, which further improved the efficiency of matrix completion. Zhang & Yang et al. (2019) proposed a weighted nonconvex nonsmooth rank relaxation function to solve the problem of over shrinking the rank components in the matrix completion method based on nuclear norm minimization. Zhang & Wei et al. (2019) proposed a modified Schatten-p norm minimization algorithm that has a fast convergence speed.

Another recommendation algorithm based on low-rank matrix completion is to decompose the rating matrix into two factor-matrices, which are called user characteristic matrix and item characteristic matrix respectively. Koren et al. (2008) incorporated implicit feedback information into the matrix decomposition model to improve the accuracy of rating prediction. Xu et al. (2021) proposed a multi-armed bandit-based collaborative filtering recommender system on handling the dynamic changes in user preferences.

In recent years, the matrix completion method combined with neural network has been widely concerned in recommendation systems. Berg et al. (2017) employed graph autoencoders derived from graph convolutional network to retrieve the missing values in an incomplete matrix, where the matrix completion task was converted into the link prediction problem on graphs. Wang et al. (2019) presented a neural graph collaborative filtering (NGCF) framework which integrates user-item interactions into the GCN framework and explicitly leverages the collaborative signal. The advantages and disadvantages of the various related work approaches are compared in Table 1.

In the process of matrix completion, proper initialization can accelerate the convergence speed and improve the completion accuracy. Koren (2008) proposed a valuation method called "baseline estimation". The method considered that the factors that affect user i 's rating of item j are mainly divided into three points: the first one is the average value of all users' ratings on the item, the second one is the user's bias term p_i , and the third one is the bias term q_j of the item. Kannan et al. (2016) and Hsieh et al. (2017) used this method to initialize the matrix to be completed. They proposed the bounded low-rank matrix approximation (BMA) algorithm and the bounded matrix completion (BMC) algorithm respectively. These two algorithms were solved by block coordinate descent (BCD) and alternating direction method of multipliers (ADMM) respectively and achieved good results in the actual datasets.

Although the evaluation initialization method can improve the performance of the algorithm, the above initialization method does not combine the characteristics of the respective algorithm itself, which limits the effect of the initialization to a certain extent. Aiming at this problem, we study the iterative characteristics of FPC algorithm, which is widely used in the rating prediction tasks of recommendation systems. As a matrix completion algorithm based on nuclear norm minimization, FPC has an obvious characteristic in which matrix norms and entries changes regularly in the iterative process. Using this property and combining it with the known rating information of the recommendation systems, an overestimation initialization method of FPC algorithm is proposed. The experimental results on the MovieLens datasets (Harper, & Konstan, 2015) show that the initialization method given in this paper greatly improves the efficiency and prediction accuracy of the fixed point continuation algorithm, and is better than the above algorithms.

Table 1 Comparison of the various related work approaches with the advantages and disadvantages

Types of model	Algorithms	Advantages	Disadvantages
Rank minimization	SVT	Be efficient for large matrix completion problems.	Only works well for very low rank matrix completion problems.
	ISVTA	Has a faster convergence speed.	The time complexity of the algorithm is high.
Matrix factorization	SVD++	Considering implicit feedback, the information utilization rate is high.	There are certain requirements for data sets, and the complexity is high.
	BanditMF	Able to effectively handle the dynamic changes in user preferences.	Bandit algorithm does not consider contextual feature information.
Neural network	GCMC	Graph neural networks can make better use of structured external information.	The lack of efficient approximation schemes leads to poor scalability.
	NGCF	Able to effectively inject the collaborative signal into the embedding process in an explicit manner.	Compared with the baseline MF, the time consumption of the algorithm increases several times.

3 Overestimation initialization method of FPC algorithm

This section is mainly divided into two parts, preliminaries and methodologies. The first part mainly introduces the fixed point continuation algorithm used in this paper and the knowledge of matrix norms. The second part focuses on the initialization method of FPC algorithm given in this paper. The mathematical symbols used in this paper are summarized in Table 2.

Table 2 Summary of mathematical notations

Notations	Explanations
M	Observed rating matrix
X	Rating matrix to be completed
Ω	Index set of known entries in the rating matrix
$\mathcal{A}$	Linear transformation operator
b	A vector consisting of the entries observed in the rating matrix
$rank(\bullet)$	The rank of matrix $\bullet$
$\lVert\bullet\rVert_*$	The nuclear norm of matrix $\bullet$
$\lVert\bullet\rVert_F$	The Frobenius norm of matrix $\bullet$
$S_{\tau\mu}(\bullet)$	Singular value shrinkage operator
$trace(\bullet)$	The trace of matrix $\bullet$
$Diag(\bullet)$	Convert sequence $\bullet$ to a diagonal matrix
σ_i	The ith largest singular value

3.1 Preliminaries

3.1.1 FPC algorithm

The fixed point continuation algorithm for the matrix completion problem was proposed by Ma et al. (2011). He described the matrix completion problem as follows:

$$\begin{aligned} &\min \lVert X \rVert_* \\ &s.t. \mathcal{A}(X) = b \end{aligned} \tag{3}$$

Where $\mathcal{A}$ is a linear transformation and b is a vector composed of observed entries. The Lagrangian expression is as follows:

$$\min \mu \lVert X \rVert_* + \frac{1}{2} \lVert \mathcal{A}(X) - b \rVert^2 \tag{4}$$

Suppose X is the optimal solution of the problem, then:

$$0 \in \mu \partial \lVert X^* \rVert_* + g(X^*) \tag{5}$$

Among them:

$$\partial\|X\|_* = \left\{UV^T + W : U^T W = 0, WV = 0, \|W\| \le 1\right\}$$
$$g\left(X^*\right) = \mathcal{A}^*\left(\mathcal{A}(X^*) - b\right)$$

Eq. (5) is equivalent to:

$$0 \in \tau\mu\partial\left\|X^*\right\|_* + X^* - (X^* - \tau g(X^*)) \tag{6}$$

Let $Y^* = X^* - \tau g(X^*)$, then the equivalent form of Eq. (4) is obtained:

$$\min\ \tau\mu\|X\|_* + \frac{1}{2}\left\|X - Y^*\right\|_F^2 \tag{7}$$

The closed-form optimal solution of the problem is as follows:

$$X = S_{\tau\mu}\left(Y\right) = U_Y Diag\left(s_{\tau\mu}(\gamma)\right)V_Y^T \tag{8}$$

In summary, the FPC algorithm of the matrix completion problem can be described as:

$$\begin{cases} Y^k = X^k - \tau g(X^k) \\ X^{k+1} = S_{\tau\mu}(Y^k) \end{cases} \tag{9}$$

We can see that the FPC algorithm used in this paper is exactly the proximal gradient descent algorithm --- a classical first-order optimization algorithm for composite problems. The first line in Eq. (9) is the gradient step and the second line is the proximal operator for nuclear norm regularization.

3.1.2 The shrinkage property of matrix elements in the iterative process

The Frobenius norm of a matrix, F-norm for short, is a matrix norm, denoted as $\|\bullet\|_F$. The F-norm of matrix $A \in R^{m\times n}$ is defined as the square root of the sum of squares of the elements of matrix A, namely:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\left|a_{ij}\right|^2} \tag{10}$$

There is an identity relationship between the square root of the sum of squares of all singular values of matrix A and the F-norm of matrix A, namely:

$$
\begin{aligned}
\sqrt{\sum_{i=1}^{\min\{m,n\}} \sigma_i^2} &= \sqrt{trace(A*A^T)} \\
&= \sqrt{\sum_{j=1}^{n} a_{1j}^2 + ...+ \sum_{j=1}^{n} a_{mj}^2} \\
&= \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2} = \|A\|_F
\end{aligned} \tag{11}
$$

It can be seen from Eq. (3) that the nuclear norm of the rating matrix X is minimized by the FPC algorithm, and the core step of the algorithm is singular value shrinkage:

$$
s_v(\sigma) = \begin{cases} \sigma - v, & \sigma - v > 0 \\ 0, & o.w. \end{cases} \tag{12}
$$

Where v is the threshold of singular value shrinkage. Let $m<n$, and the matrix A has $k\,(k<m)$ singular values greater than v, then:

$$
\|A\|_F^2 = [(\sigma_1 - v) + v]^2 + ...+ \left[(\sigma_k - v) + v\right]^2 + \sigma_{k+1}^2 + ...+ \sigma_m^2 \tag{13}
$$

After completing the singular value shrinkage operation:

$$
\left\|\hat{A}\right\|_F^2 = (\sigma_1 - v)^2 + ...+ (\sigma_k - v)^2 \tag{14}
$$

Due to:

$$
\begin{aligned}
&\|A\|_F^2 - \left\|\hat{A}\right\|_F^2 \\
&= kv^2 + 2v\left[(\sigma_1 - v) + ...+ (\sigma_k - v)\right] + \sigma_{k+1}^2 + ...+ \sigma_m^2 > 0
\end{aligned} \tag{15}
$$

Therefore, the F-norm of the matrix is constantly decreasing with the shrinkage of singular values. In the recommendation systems, the users' ratings are generally specified as non-negative integers. For example, in the MovieLens datasets widely used in the recommendation systems, the users' ratings are integers between 1-5. In the iterative process, although some individual predicted ratings have negative values, most of the rating values are positive. Therefore, with the decrease of the F-norm of the rating matrix X, the elements in the matrix, that is, the users' ratings also have the tendency to become smaller.

To verify this rule, we construct two matrices $X_1 \in \mathbb{R}^{1500*1500}$ and $X_2 \in \mathbb{R}^{1500*1500}$ whose element values are random numbers between 0-1 and 1-5 respectively. Inspired by an effective singular value shrinkage strategy in reference (Cai, Candès, & Shen, 2010), we take the singular value shrinkage threshold as $0.8^k * 0.2\|X\|_2$ (where k represents the number of iterations). Figure 1 shows the changing trend of the nuclear norm, F-norm and the average value of all elements of matrices X_1 and X_2 with the shrinkage of singular values.

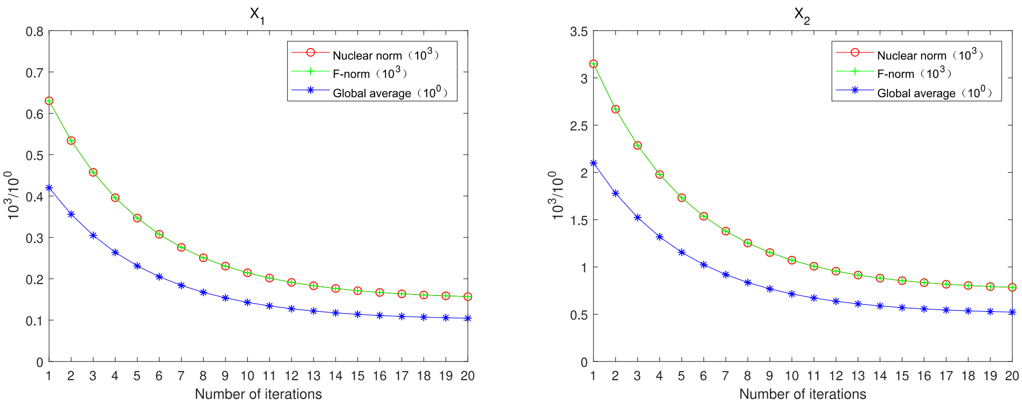

Fig. 1. The changing trend of the nuclear norm, F-norm, and the global average of matrices X_1 and X_2

It can be seen from Figure 1 that the nuclear norm, F-norm, and the global average of matrices X_1 and X_2 have the same changing trend, and they all decrease continuously with the shrinkage of singular values. Therefore, in the iteration process, the matrix elements have a shrinking property.

3.2 Methodologies

Baseline estimation (Koren, 2008) is proved to be effective when applied to matrix initialization. This paper uses a similar method to pre-estimate unknown ratings and construct an estimation matrix. The difference is that according to the shrinkage property of matrix elements in the iterative process, we initialize X as a matrix with larger elements values; that is to say, the values of all elements in the estimation matrix are expanded, which we call the overestimation initialization method.

When constructing the valuation matrix X_1, this paper fully captures the user characteristics and item features according to the known ratings. Assuming that the average of user i's existing ratings is p_i, then p_i is defined as the average of all rating records of user i in the training set, namely $p_i = \frac{\sum_{j \in N(i)} r_{ij}}{\sum_{j \in N(i)} 1}$; Suppose the average value of the existing ratings of item j is q_j, then q_j is defined as the average value of all the ratings of item j in the training set, namely $q_j = \frac{\sum_{i \in N(j)} r_{ij}}{\sum_{i \in N(j)} 1}$. We estimate user i's rating of item j as the average of p_i and q_j, that is, $X_1(i, j) = \frac{1}{2}(p_i + q_j)$. Since directly initializing the matrix X_1 will cause a great waste of storage resources, we initialize the matrix X_1 into two smaller matrices P and Q according to the idea of matrix decomposition (Srebro, 2004). After the evaluation matrix is constructed, the evaluation is expanded by ρ times, where $\rho \approx \frac{r_{\max}}{r_{mean}}$ ($r_{\max}$ represents the highest rating, r_{mean} represents the average of all known ratings). In the process of the experiment, we regard ρ as a variable parameter and adjust it around $\frac{r_{\max}}{r_{mean}}$ to get the optimal value.

The specific implementation steps are as follows:

(1) Set up the valuation matrix $X_1 = PQ^T$, where $P \in \mathbb{R}^{m \times 2}, Q^T \in \mathbb{R}^{2 \times n}$.

(2) Initialize matrix $X_0 = \rho X_1$, where $\rho \approx \frac{r_{\max}}{r_{mean}}$.

$$X_1 = \overset{P}{\begin{bmatrix} p_1 & 1/2 \\ p_2 & 1/2 \\ \vdots & \vdots \\ p_{m-1} & 1/2 \\ p_m & 1/2 \end{bmatrix}} \times \overset{Q^T}{\begin{bmatrix} 1/2 & 1/2 & \cdots & 1/2 & 1/2 \\ q_1 & q_2 & \cdots & q_{n-1} & q_n \end{bmatrix}}$$

Fig. 2. Construction of estimation matrix

Convergence analysis

Regarding the convergence of FPC algorithm, the author has given a detailed proof in reference (Ma, Goldfarb, & Chen, 2011), which will not be repeated here. In this section, we mainly analyze and compare the convergence speed of the proposed method and the traditional initialization method.

Each iteration of the FPC algorithm can be divided into the following four steps:

• Update the elements in the known entries set Ω of the matrix;

• Singular value decomposition of the matrix;

• Shrinkage of singular values of the matrix;

• Multiply the left and right singular matrices and the singular value matrix back.

The detailed process is shown in Figure 3:

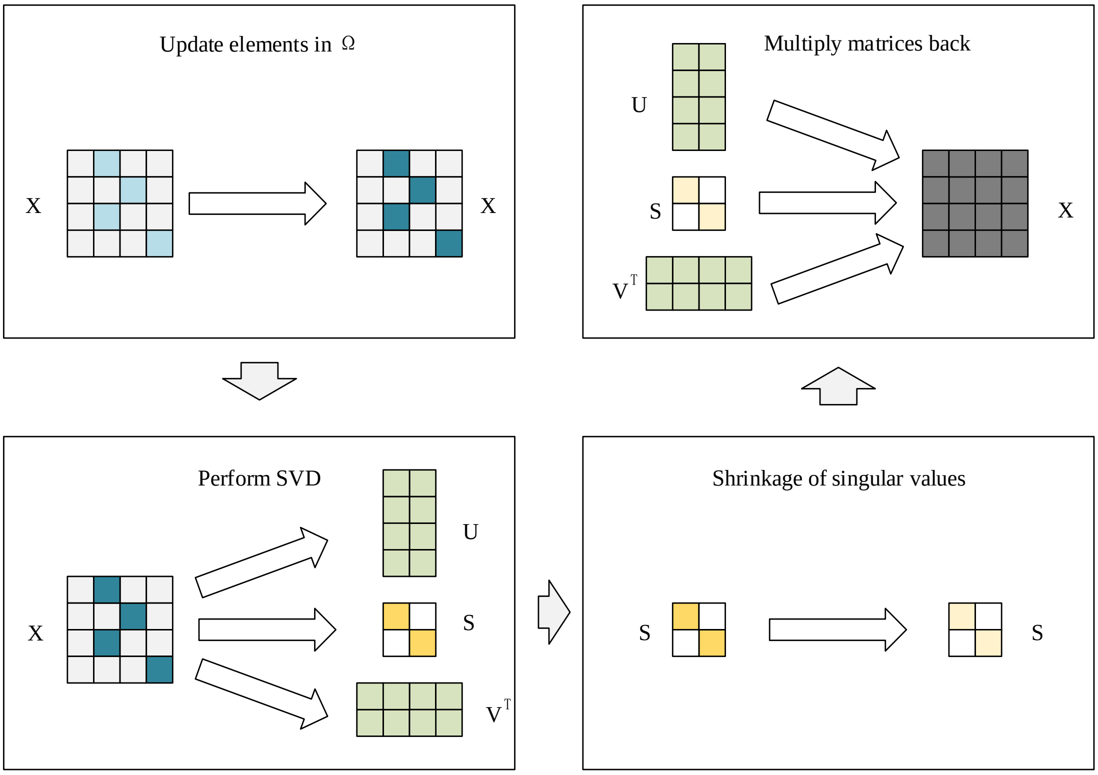

Fig. 3. The detailed process of each iteration of the FPC algorithm

In the first step, the update to the matrix X can be expressed as $X = X - \tau\mathcal{A}^*(\mathcal{A}(X) - b)$. Where $\mathcal{A}$ represents a projection transformation, and $\mathcal{A}(X)$ maps the elements of matrix X in Ω to a vector in a certain order; b is a vector composed of known elements in the matrix (ie the ratings in the training set); $\mathcal{A}^*$ is the inverse transformation of $\mathcal{A}$, and $\mathcal{A}^*\left(\mathcal{A}(X) - b\right)$ maps the vector $\mathcal{A}(X) - b$ back to the matrix. Because of the parameter $\tau > 0$, when the overestimation initialization method is used, taking into account $\mathcal{A}(X) > b$, after this step is executed, the elements values of matrix in Ω are reduced; When using the original initialization method (that is, initializing all the elements of the matrix to 0), considering $\mathcal{A}(X)<b$, after this step is executed, the elements values of the matrix in Ω are increased.

The second step is the singular value decomposition, namely $X=U\Sigma V^*$. This step does not make any changes to the values of the elements in matrix X, so there is no difference between the two initialization methods.

The third step is the singular value shrinkage of the matrix X, namely $S_{\tau\mu}(\Sigma)=diag\left(\{\sigma_i-\tau\mu\}_+\right)$. It can be seen from Section 3.1.2 that the shrinkage of singular value will lead to the decrease of matrix elements values. Therefore, after this step is performed, the values of the elements in the matrix will be reduced by using two initialization methods.

The fourth step is the operation of multiplying left and right singular matrices and the singular value matrix back to X after the singular value shrinkage in the third step, namely $X=US_{\tau\mu}(\Sigma)V^*$. Similar to the second step, this step does not change the elements of the matrix X, so there is no difference between the two initialization methods.

The process of approaching the fixed point in one iteration between the overestimation initialization method and the original initialization method is shown in Fig. 4:

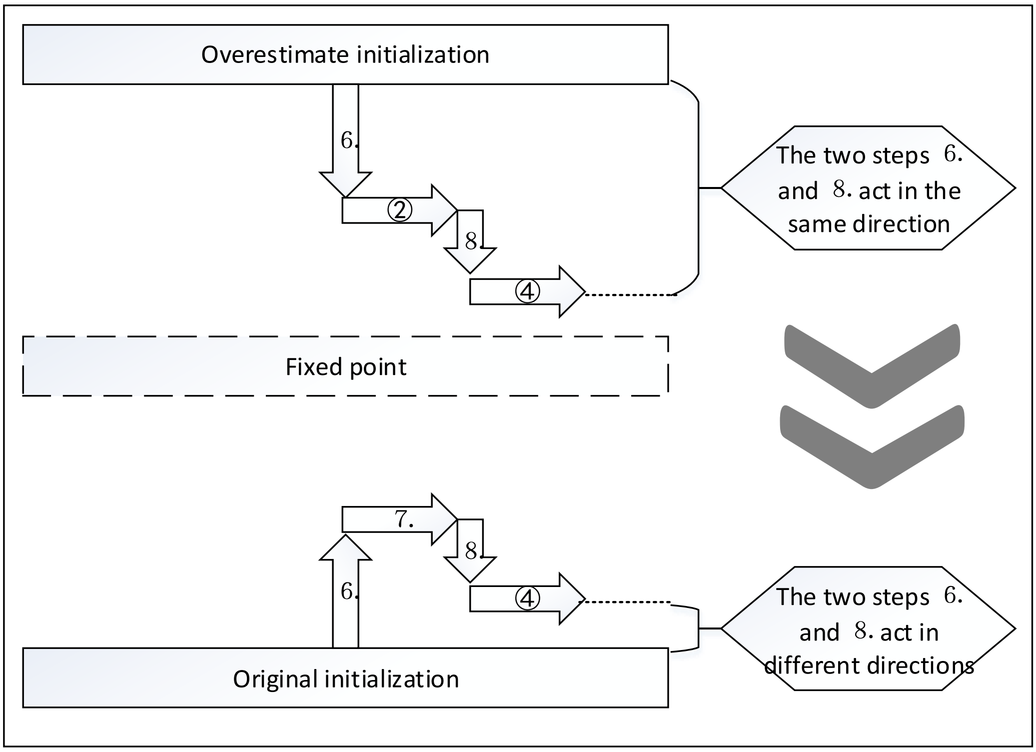

Fig. 4. The process of two initialization methods approaching the fixed point

It can be seen from Figure 4 that the overestimation initialization method will converge to the fixed point faster.

Pseudo code

Algorithm 1

Input: $X_0, M, b, \mu_0, \mu_{\min}, \tau, \eta, k$

Output: X

Initialize: $X = X_0, b = \mathrm{P}_{\Omega}(M), \mu = \mu_0, \tau = 1.99, \eta = 1/5$

While $\mu \geq \mu_{\min}$ **do**

While not converged, **do**

- **Compute** $Y = X - \tau \mathcal{A}^*\left(\mathcal{A}(X) - b\right)$
- **Compute** $\left(U, Diag(\sigma), V\right) = Partial\_svd(Y, k)$
- **Compute** $X = UDiag\left(s_{\tau\mu}(\sigma)\right)V^T$

End while

- $\mu = \eta\mu$

End while

Time complexity

The most time-consuming step of the algorithm is the singular value decomposition of matrix Y. For a matrix of size $m \times n$, the time complexity of singular value decomposition is $\mathcal{O}\left(\min\left\{mn^2, m^2n\right\}\right)$, which will seriously affect the efficiency of the algorithm. Inspired by the truncated nuclear norm (Hu, Zhang, Ye, Li, & He, 2012), in the solution process, we only calculate the singular values of the former k, among which $0 < k << \{m, n\}$. In the experiment, we will take k as the parameter for processing, and the time complexity is $\mathcal{O}\left(kmn\right)$. Compared with the previous, the algorithm has a higher execution efficiency.

Space complexity

The memory space consumed by the algorithm in the running process mainly lies in the rating matrix X. If the matrix X of size $m \times n$ is processed as a whole, the space complexity of the algorithm is $\mathcal{O}(mn)$, which will seriously limit the scalability of the algorithm. In this paper, the matrix X is initialized as a matrix P of size $m\times 2$ and a matrix Q of size $n\times 2$. In the iteration process, matrix X is represented as matrix U with size $m\times r$, matrix S with size $r\times r$ and matrix V with size $n\times r$, where $r\le k$. The space complexity of the algorithm is $\mathcal{O}(r(m+n+r))$, which greatly reduces the consumption of storage resources.

4 Experiment and analysis

In order to test the effectiveness and efficiency of the initialization method proposed in this article, three parts of experiments are conducted in this section. The first part is the comparison with the FPC algorithm using the original initialization method; The second part is the comparison with the FPC algorithm which uses the random initialization method commonly used in recommendation systems; The third part is the comparison with three classic matrix completion algorithms SVT, IALM, BPMF and two state-of-the-art matrix completion algorithms BMC and GCMC. All experiments in this article are performed on a PC with Intel Core i5-6200U and 4GB memory under the Matlab R2018b environment.

4.1 Experimental settings

4.1.1 Datasets

The datasets used in this paper are Movielens100K and Movielens1M, which are commonly used in the recommendation field. The specific information of the datasets is shown in Table 3.

Table 3 Statistics of the datasets

Dataset	Users	Items	Ratings	Scale	Density
MovieLens100K	943	1682	100,000	1-5	6.30%
MovieLens1M	6040	3952	1,000,209	1-5	4.19%

4.1.2 Data division and evaluation metrics

In this paper, the datasets are divided into training sets and test sets according to the ratio of 8:2. In order to reduce the chance of the experiment, we adopt a 5-fold cross-validation method, repeat each group of experiments five times, and take the average of the five experimental results as the final experimental result. This article uses the root mean square error (RMSE) and mean absolute error (MAE) commonly used in recommendation systems as the evaluation metrics of the algorithms' prediction accuracy and uses the time to measure the efficiency of the algorithms. The definition of RMSE and MAE are as follows:

$$RMSE = \sqrt{\frac{\sum_{r_{ui} \in T} \left(r_{ui} - \hat{r}_{ui} \right)^2}{|Test|}}$$

$$MAE = \frac{1}{|Test|} \sum_{r_{ui} \in T} |r_{ui} - \hat{r}_{ui}|$$

Among them, $|Test|$ represents the number of test ratings, r_{ui} represents the actual rating of user u on item i, and $\hat{r}_{ui}$ represents the predicted rating of user u on item i.

4.1.3 Comparison algorithm

In order to verify the effectiveness of our initialization method based on the FPC algorithm, this paper compares the following matrix completion algorithms:

FPC (Fixed Point Continuation): A fixed point iterative algorithm for solving the nuclear norm minimization problem (Ma et al., 2011).

SVT (Singular Value Threshold): A novel algorithm to approximate the matrix with minimum nuclear norm among all matrices obeying a set of convex constraints (Cai et al., 2010).

IALM (Inexact Augmented Lagrangian Multiplier): A scalable and fast algorithm for solving the robust principal component analysis problem (Lin et al., 2010).

BPMF (Bayesian Probabilistic Matrix Factorization): A fully Bayesian treatment of the probabilistic matrix factorization model in which model capacity is controlled automatically by integrating over all model parameters and hyperparameters (Salakhutdinov et al., 2008).

BMC (Bounded Matrix Completion): A bounded matrix completion algorithm which imposes bounded constraints into the standard matrix completion problem (Hsieh et al., 2017).

GCMC (Graph Convolutional Matrix Completion): A graph auto-encoder framework for the matrix completion task in recommender systems (Berg et al., 2017).

4.1.4 Parameter settings

In order to further improve the efficiency of the algorithm and reduce the time consumption of SVD in the iterative process, we use the propack package of Larsen et al. (Larsen, 2022) to obtain the first k large singular values. Specific parameter settings are shown in Table 4. All hyper parameters in the compared methods are based on the suggestions of their corresponding papers.

Table 4 Parameters in the Algorithm.

Dataset	k	$\mu_{\min}$	η	μ_0	τ
MovieLens100K	7	6	1/5	$\eta\left\Vert\mathcal{A}^*(b)\right\Vert_2$	1.99
MovieLens1M	14	12	1/4	$\eta\left\Vert\mathcal{A}^*(b)\right\Vert_2$	1.99

4.2 Experimental results and analysis

First of all, through the estimation method of parameter ρ in Section 3.2, after calculation, the values of ρ on MovieLens100K and MovieLens1M are 1.42 and 1.40 respectively. Then, we adjusted the parameters in the vicinity of these two values. The experimental results are shown in Figures 5.

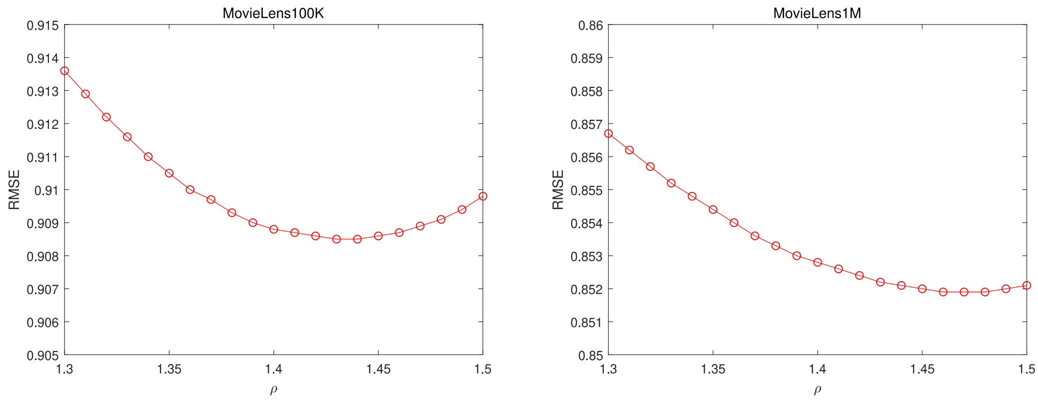

Fig. 5. Experimental results of parameter adjustment on two datasets

It can be seen from the results of the tuning experiment that in the MovieLens100K and MovieLens1M datasets, when ρ is 1.43 and 1.46 respectively, the best experimental results are obtained. Similarly, when MAE is used as the evaluation metric, the best experimental results are obtained when ρ on two datasets is 1.45 and 1.48 respectively. After determining the final initialization method, we compared it with the traditional FPC algorithm that initializes the matrix X by zero. The experimental results are shown in Figure 6, Figure 7 and Table 5.

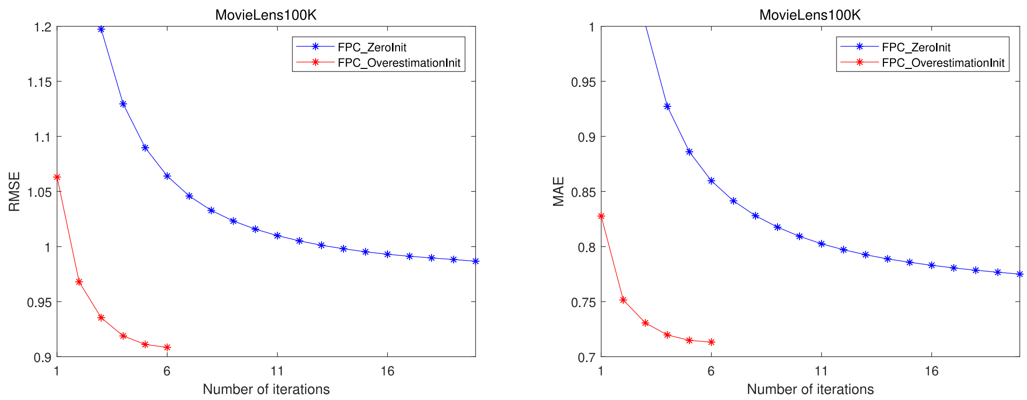

Fig. 6. experimental results on Movielens 100K

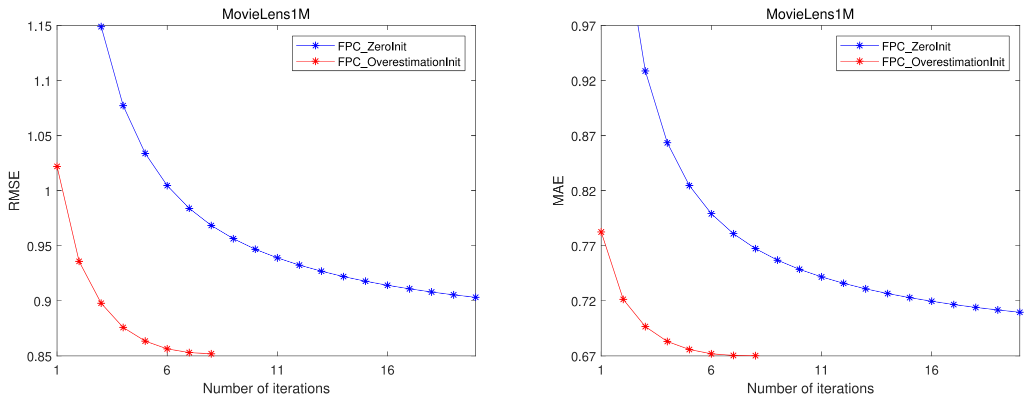

Fig. 7. experimental results on Movielens 1M

Table 5 Comparison of specific experimental results of FPC using original initialization method and proposed initialization method.

		FPC_ZeroInit	FPC_OverestimationInit
	RMSE	0.9574	0.9085
MovieLens100K	MAE	0.7483	0.7133
	Time(s)	19	2
	RMSE	0.8681	0.8519
MovieLens1M	MAE	0.6856	0.6702
	Time(s)	211	15

The experimental results of Fig. 6 and Fig. 7 show that the fixed point continuation algorithm after using the initialization given in this paper has a faster convergence rate than before, and it only takes a few iterations to achieve higher prediction accuracy. From the detailed experimental results in Table 5, we can see that on the MovieLens100K, RMSE and MAE are reduced by 5.11% and 4.68% respectively; on the MovieLens1M, RMSE and MAE are reduced by 1.87% and 2.25% respectively. Therefore, compared with the original initialization method, the prediction accuracy of the proposed initialization method is improved. And the lapsed time of the algorithms shows that the FPC using the proposed initialization method has a higher filling efficiency.

Random initialization (Rendle, & Schmidt-Thieme, 2008) is a commonly used initialization method in recommendation systems. It initializes the user feature matrix and item feature matrix to small random numbers. We compared this initialization method with the proposed initialization method. Since random initialization is affected by the dimension d of the feature matrices, we compared random initialization methods of multiple dimensions. The experimental results are shown in Table 6.

Table 6 Comparison of experimental results of FPC using random initialization method and proposed initialization method.

		MovieLens100K		MovieLens1M	
		RMSE	MAE	RMSE	MAE
FPC_RandomInit	d = 10	0.9551	0.7456	0.8675	0.6843
	d = 20	0.9214	0.7284	0.8626	0.6802
	d = 30	0.9183	0.7235	0.8575	0.6778
	d = 40	0.9207	0.7263	0.8598	0.6791
FPC_OverestimationInit		0.9085	0.7133	0.8519	0.6702

It can be seen from the experimental results in Table 6 that when the feature dimension $d = 30$, the FPC using the random initialization method achieves the best prediction accuracy. The overestimation initialization method given in this article is always better than random initialization, which shows that it is necessary to initialize the matrix with the known information of the data.

Subsequently, we compared the prediction accuracy and efficiency of the FPC algorithm using the proposed initialization method with three classic matrix completion algorithms SVT, IALM and BPMF, and the currently more advanced bounded matrix completion algorithms BMC and GCMC. The experimental results are shown in Figure 8, Figure 9, Figure 10 and Table 7 (the FPC algorithm using the proposed initialization method is recorded as FPC in the Figures and Table):

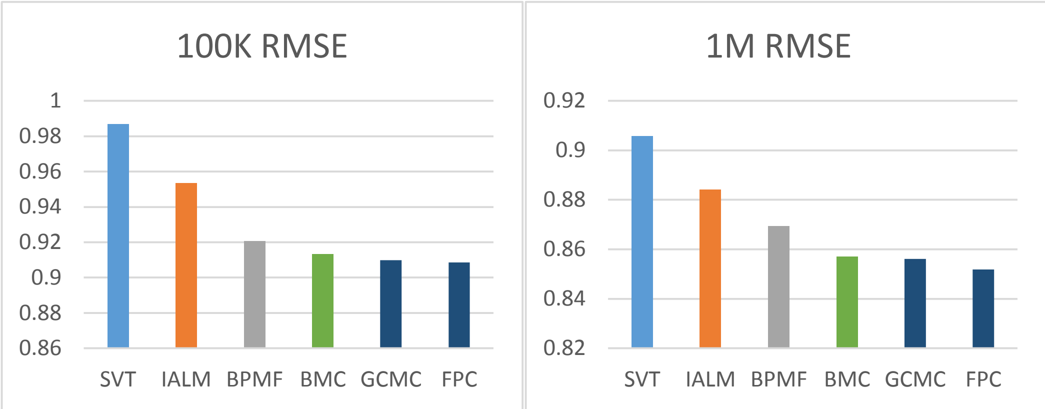

Fig. 8. Comparison of RMSE on two datasets

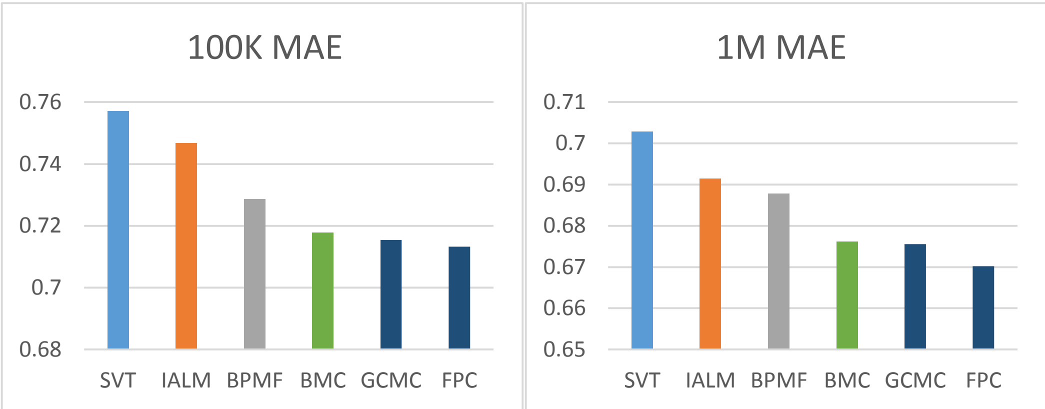

Fig. 9. Comparison of MAE on two datasets

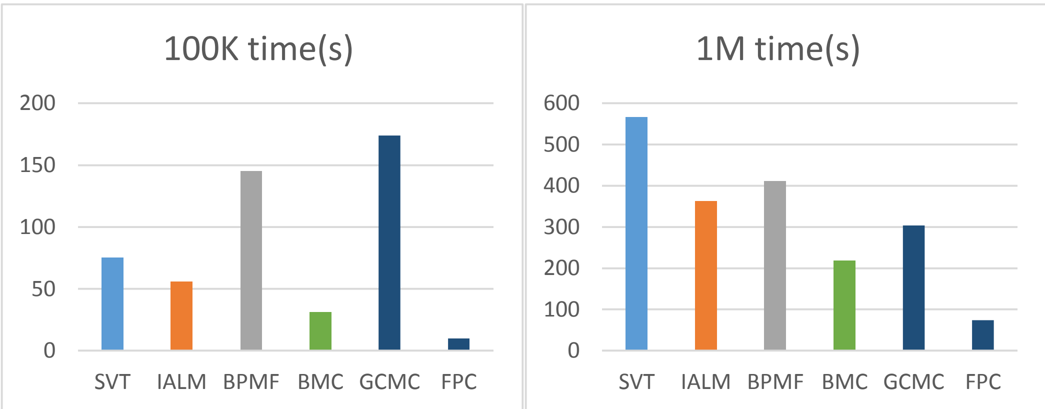

Fig. 10. total time of five-fold cross-validation on two datasets

Table 7 Specific experimental data on two datasets

	MovieLens100K			MovieLens1M		
	RMSE	MAE	Time(s)	RMSE	MAE	Time(s)
SVT	0.9868	0.7570	75	0.9058	0.7029	567
IALM	0.9535	0.7467	56	0.8841	0.6914	363
BPMF	0.9208	0.7287	145	0.8693	0.6878	412
BMC	0.9133	0.7178	31	0.8572	0.6762	218
GCMC	0.9098	0.7154	174	0.8561	0.6756	304
FPC	0.9085	0.7133	10	0.8519	0.6702	74

It can be seen intuitively from the experimental results that the improved initialization of the fixed point continuation algorithm in this paper has greatly improved the prediction accuracy and algorithm efficiency compared with the two classic matrix completion algorithms SVT and IALM. Compared with bounded matrix completion BMC and neural network-based matrix completion GCMC, the prediction accuracy is less improved, but the lapsed time of the algorithm is reduced a lot.

From all the above experimental results, whether compared with the other initialization methods of the FPC algorithm, or compared with other recommendation algorithms based on matrix completion, the FPC algorithm using the proposed initialization method has shown an outstanding effect. This shows that initialization has a huge impact on the FPC algorithm. Therefore, it is necessary to fully mine the known data information and combine the characteristics of the algorithm during the initialization process.

5 Conclusion and future work

In this paper, we analyzed the changing law of the matrix norm during the iterative process of the FPC algorithm, and then obtained the shrinkage characteristics of the matrix elements. Using this feature of the FPC algorithm and combining the known data information, an overestimation initialization method for the FPC algorithm is given, and the algorithm is analyzed in terms of convergence and space-time complexity. Experimental results on real datasets show that the proposed algorithm is better than the classical matrix completion algorithm and the current advanced algorithm.

The main theoretical contribution of this paper is to give an idea of adapting matrix initialization to algorithm iteration in matrix completion. Aiming at rating prediction in the recommendation field, the given matrix initialization method integrates the prior information of users' ratings. The work of this paper is helpful to stimulate researchers to study further the initialization of matrix completion method based on nuclear norm minimization, and mine more prior information to integrate into the initialization of matrix.

Although the initialization method proposed in this paper makes the matrix completion algorithm FPC has higher recommendation accuracy and faster speed. However, this method still has some limitations, mainly manifested as the following two points: First, the method fails to solve the problem of cold-start. Second, although it takes only a few iterations to achieve high accuracy, with the increase in the number of iterations, the accuracy will decline.

As a generalization of low-rank matrix completion from two-dimensional space to multidimensional space, low-rank tensor completion (Zhou, Lu, Lin, & Zhang, 2017) is an effective tool for high-order data analysis and has been widely concerned in the fields of context recommendation systems (Frolov, & Oseledets, 2017) and computer vision (Yang, Zhao, Ma, Ding, & Huang, 2020). Compared with matrix completion, tensor completion started late, and the completion technology needs to be explored. Therefore, in future work, we plan to study the recommendation algorithm based on low-rank tensor completion and extend the initialization method in this paper to the tensor completion algorithm.

Acknowledgements

This study was supported by Natural Science Foundation of Shandong Province (No. ZR2021MF104, No. ZR2021MF113), National Natural Science Foundation (No. 62072288), Key R&D Projects of Qingdao Science and Technology Plan (No.21-1-2-19-xx), and Qingdao West Coast New District Science and Technology Plan (No.2020-1-6).

Reference

Ko, H., Lee, S., Park, Y., & Choi, A. (2022). A survey of recommendation systems: recommendation models, techniques, and application fields. Electronics, 11(1), 141.

Tang, H., Zhao, G., Bu, X., & Qian, X. (2021). Dynamic evolution of multi-graph based collaborative filtering for recommendation systems. Knowledge-Based Systems, 228, 107251.

Pujahari, A., & Sisodia, D. S. (2022). Item feature refinement using matrix factorization and boosted learning based user profile generation for content-based recommender systems. Expert Systems with Applications, 117849.

Biswas, P. K., & Liu, S. (2021). A Hybrid Recommender System for Recommending Smartphones to Prospective Customers. arXiv preprint arXiv:2105.12876.

Ramlatchan, A., Yang, M., Liu, Q., Li, M., Wang, J., & Li, Y. (2018). A survey of matrix completion methods for recommendation systems. Big Data Mining and Analytics, 1(4), 308-323.

Chen, Z., & Wang, S. (2022). A review on matrix completion for recommender systems. Knowledge and Information Systems, 1-34.

Ma, S., Goldfarb, D., & Chen, L. (2011). Fixed point and Bregman iterative methods for matrix rank minimization. Mathematical Programming, 128(1), 321-353.

Karmarkar, N. (1984, December). A new polynomial-time algorithm for linear programming. In Proceedings of the sixteenth annual ACM symposium on Theory of computing (pp. 302-311).

Fazel, M. (2002). Matrix rank minimization with applications (Doctoral dissertation, PhD thesis, Stanford University).

Cai, J. F., Candès, E. J., & Shen, Z. (2010). A singular value thresholding algorithm for matrix completion. SIAM Journal on optimization, 20(4), 1956-1982.

Toh, K. C., & Yun, S. (2010). An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. Pacific Journal of optimization, 6(615-640), 15.

Lin, Z., Chen, M., & Ma, Y. (2010). The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint arXiv:1009.5055.

Zhang, H., Gong, C., Qian, J., Zhang, B., Xu, C., & Yang, J. (2019). Efficient recovery of low-rank matrix via double nonconvex nonsmooth rank minimization. IEEE Transactions on Neural Networks and Learning Systems, 30(10), 2916-2925.

Zhang, H., Qian, J., Zhang, B., Yang, J., Gong, C., & Wei, Y. (2019). Low-Rank Matrix Recovery via Modified Schatten-p Norm Minimization With Convergence Guarantees. IEEE Transactions on Image Processing, 29, 3132-3142.

Koren, Y. (2008, August). Factorization meets the neighborhood: a multifaceted collaborative filtering model. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, Las Vegas, Nevada, USA (pp. 426-434).

Xu, S. (2021). BanditMF: Multi-Armed Bandit Based Matrix Factorization Recommender System. arXiv preprint arXiv:2106.10898.

Berg, R. V. D., Kipf, T. N., & Welling, M. (2017). Graph convolutional matrix completion. arXiv preprint arXiv:1706.02263.

Wang X, He X, Wang M, Feng F, Chua T-S (2019) Neural graph collaborative filtering. In: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval, pp 165–174.

Kannan, R., Ishteva, M., Drake, B., & Park, H. (2016). Bounded matrix low rank approximation. In Non-negative Matrix Factorization Techniques (pp. 89-118). Springer, Berlin, Heidelberg.

Fang, H., Zhang, Z., Shao, Y., & Hsieh, C. J. (2017, August). Improved Bounded Matrix Completion for Large-Scale Recommender Systems. In IJCAI, Melbourne, Australia (pp. 1654-1660).

Harper, F. M., & Konstan, J. A. (2015). The movielens datasets: History and context. Acm transactions on interactive intelligent systems (tiis), 5(4), 1-19.

Srebro, N. (2004). Learning with matrix factorizations (PhD thesis, Massachusetts Institute of Technology).

Hu, Y., Zhang, D., Ye, J., Li, X., & He, X. (2012). Fast and accurate matrix completion via truncated nuclear norm regularization. IEEE transactions on pattern analysis and machine intelligence, 35(9), 2117-2130.

Salakhutdinov, R., & Mnih, A. (2008, July). Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In Proceedings of the 25th international conference on Machine learning (pp. 880-887).

Larsen, R. M., (2022, Mar). PROPACK – software for large and sparse SVD calculations. Available from http://sun.stanford.edu/~rmunk/PROPACK/.

Rendle, S., & Schmidt-Thieme, L. (2008, October). Online-updating regularized kernel matrix factorization models for large-scale recommender systems. In Proceedings of the 2008 ACM conference on Recommender systems, Lausanne, Switzerland (pp. 251-258).

Zhou, P., Lu, C., Lin, Z., & Zhang, C. (2017). Tensor factorization for low-rank tensor completion. IEEE Transactions on Image Processing, 27(3), 1152-1163.

Frolov, E., & Oseledets, I. (2017). Tensor methods and recommender systems. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7(3), e1201.

Yang, J. H., Zhao, X. L., Ma, T. H., Ding, M., & Huang, T. Z. (2020). Tensor train rank minimization with hybrid smoothness regularization for visual data recovery. Applied Mathematical Modelling, 81, 711-726.