

Research Article

S²NMF: Information Self-Enhancement Self-Supervised Nonnegative Matrix Factorization for Recommendation

Ronghua Zhang,^{1,2} Yuanyuan Wang,² Fangyuan Liu ,³ Changzheng Liu ,² Yaping Song,² and Baohua Yu²

¹Intelligent and Distributed Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

²College of Information Science and Technology, Shihezi University, Shihezi 832000, China

³Graduate Business School, UCSI University, Kuala Lumpur 56000, Malaysia

Correspondence should be addressed to Fangyuan Liu; 1002164764@ucsiuniversity.edu.my and Changzheng Liu; liucz@shzu.edu.cn

Received 20 June 2022; Revised 27 July 2022; Accepted 6 August 2022; Published 30 August 2022

Academic Editor: Pengfei Wang

Copyright © 2022 Ronghua Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nonnegative matrix factorization (NMF), which is aimed at making all elements of the factorization nonnegative and achieving nonlinear dimensional reduction at the same time, is an effective method for solving recommendation system problems. However, in many real-world applications, most models learn recommendation models under the supervised learning paradigm. Since the recommendation performance of NMF models relies heavily on initialization, the user-item interaction information is often very sparse. In many cases, supervised information about the data is difficult to obtain, resulting in a large number of existing models for supervised learning being inapplicable. To address this problem, we propose an information self-supervised NMF model for recommendation. Specifically, this model is based on the matrix factorization idea and introduces a self-supervised learning mechanism based on the NMF model to enhance the sparse data information of sparse data, and an easily extensible self-supervised NMF model was proposed. Furthermore, a corresponding gradient descent optimization algorithm was proposed, and the complexity of the algorithm was analysed. A large number of experimental results show that the proposed S²NMF has better performance.

1. Introduction

In the age of information explosion, information overload has become a central issue faced by society. Recommender systems play a vital role in solving this problem, as they help determine what information to provide to individual consumers and allow online users to quickly find personalized information that suits their needs [1]. Recently, recommender systems have become ubiquitous on e-commerce platforms, such as Amazon for book recommendations, <http://Last.com/> for music recommendations, Netflix for movie recommendations, and CiteULike for references.

The main recommendation methods include collaborative filtering recommendation [2, 3], content-based recommendation [4], knowledge-based recommendation [5], and social network-based recommendation [6]. Collaborative filtering recommendation generally adopts the nearest neighbor technology, calculates the distance between users by using the historical preference information of users, and then predicts the preference degree of target users to specific products by using the weighted evaluation value of the nearest neighbor users, and the system makes recommendations to target users according to this preference degree. The maximum advantage of collaborative filtering is that it has no

special requirements on the recommended objects and can deal with unstructured complex objects, such as music and movies. Content-based recommendation is the continuation and development of information filtering technology. It makes recommendations based on the content information of the project, without relying on users' comments on the project. It is more necessary to use machine learning methods to get user interest information from the feature description of content. Knowledge-based recommendation can be regarded as a kind of inference technology, which is not based on the needs and preferences of users. Knowledge-based approaches differ markedly depending on the functional knowledge they use. Social network-based recommendation has previously been mostly domain-based approach. Firstly, the social network of raters was explored, and the scores of raters were aggregated to calculate the predicted scores. And then, find the raters' neighbors.

Learning high-quality user and item representations from interaction data is the core idea of collaborative recommendation. In early studies, such as matrix factorization (MF) [7, 8], a single ID of each user (or item) is projected into an embedding vector. Subsequent studies [9] enriched single IDs with interaction histories to learn better representations. Typically, nonnegative matrix factorization (NMF) [10, 11], a well-known dimensional reduction method in data representation, has also been successively applied to recommender system problems in recent years [12, 13]. Although NMF can be used for any nonnegative rating matrix (e.g., ratings from 1 to 5), its greatest interpretability advantage arises from the fact that users have the mechanism to specify the liking entry but not to specify a disliking entry. Such matrices include one-dimensional rating matrices or matrices in which nonnegative entries correspond to activity frequencies. These datasets are also referred to as implicit feedback datasets.

However, the NMF model is essentially a nonconvex optimization problem, and its sensitivity to initialization is unavoidable, i.e., the recommendation performance of the NMF model depends heavily on the initialization, and a poor initialization matrix can significantly degrade the recommendation performance. A general recommendation system uses only historical user-item interaction information (explicit or implicit feedback) as input, which poses two problems. First, in real-world scenarios, information about user-item interactions is often sparse. For example, a movie app may contain tens of thousands of movies, yet the average number of movies typed by a user may be only a few dozen. Using such a small amount of observed data to predict a large amount of unknown information can greatly increase the risk of overfitting the algorithm. For newly added users or items, the system does not have their historical interaction information, so it cannot accurately model and recommend them. This situation is called the cold start problem.

Moreover, most existing models learn recommendation models in a supervised learning paradigm [14–16], where the supervised signals are derived from observed user-item interactions. However, the observed interactions are

extremely sparse compared to the entire interaction space [8, 17], which makes it insufficient to learn quality representations. Moreover, in many cases, supervised information about the data is difficult to obtain, making a large number of existing models for supervised learning inapplicable.

Accordingly, this paper introduces a self-supervised learning mechanism based on the matrix factorization idea and NMF model. We propose an easily scalable self-supervised nonnegative matrix factorization recommendation model framework based on matrix decomposition.

Self-supervised nonnegative matrix factorization (S^2 NMF) was proposed, and a corresponding gradient descent optimization algorithm was further proposed. The complexity of the algorithm is analysed. Extensive experimental results show that the proposed S^2 NMF has better performance. The main contributions can be summarized as follows:

- (i) Based on the idea of matrix factorization, the self-supervised learning mechanism is introduced on the basis of the NMF model to realize the information enhancement of sparse data
- (ii) A self-supervised nonnegative matrix factorization recommendation model S^2 NMF is proposed, and a corresponding gradient descent optimization algorithm is further proposed, and the complexity of the algorithm is analysed
- (iii) Extensive experimental results demonstrate that the proposed S^2 NMF has superior performance on recommendation in comparison with comparison algorithm

The rest of this paper is organized as follows. Section 2 briefly reviews random-walk-based, factorization-based, and deep-learning-based approaches. Section 3 presents the proposed CAHE and the corresponding optimization algorithm. Section 4 analyses the performance of CAHE, including the experimental results of node classification and clustering. Finally, Section 5 concludes this work.

2. Materials and Methods

This section provides a detailed description of the proposed model S^2 NMF and its model optimization algorithm and gives the pseudocode of the S^2 NMF model optimization algorithm and its time complexity analysis.

2.1. Problem Definition. Suppose there are M users $U = \{u_1, u_2, \dots, u_M\}$ and N items $V = \{v_1, v_2, \dots, v_N\}$. Let the scoring matrix array $R \in \mathbb{R}^{M \times N}$, where R_{ij} is the rating of user i for item j . If the rating is unknown, we set $R_{ij} = \text{unk}$. In detail, and the symbol descriptions of the work are shown.

There are usually two ways to construct the user-item interaction matrix $Y \in \mathbb{R}^{M \times N}$.

$$Y_{ij} = \begin{cases} 0, & \text{if } R_{ij} = \text{unk}, \\ 1, & \text{otherwise,} \end{cases} \quad (1)$$

$$Y_{ij} = \begin{cases} 0, & \text{if } R_{ij} = unk, \\ R_{ij}, & \text{otherwise.} \end{cases} \quad (2)$$

Based on literature [18], most researchers usually believe that all evaluations are equal to 1, and then, choose Formula (1) to construct the interaction matrix Y . In this paper, we choose Equation (2) to construct the interaction matrix Y , where the rating R_{ij} of user u_i for item v_j remains in the interaction matrix Y . Explicit recommendation rating is very complicated for recommendation. Here, we express the user's preference degree for a certain product by Formula (2) and mark the unknown rating as 0 by the method of implicit feedback without preference. Usually, recommendation systems are formulated as a problem of estimating the rating of each unobserved item in Y .

In order to better formalize the mathematical process of this work, the detailed notation is shown in Table 1. The tasks of a recommendation system can be divided into three types: scoring prediction, top-N prediction, and click prediction. Therefore, the proposed S²NMF model framework is summarized as follows.

- (i) *Input*. The observed user-item interaction matrix R
- (ii) *Output*. The predicted user-item interaction matrix \hat{R}

3. Model Framework

3.1. Classical NMF Model. Nonnegative matrix factorization (NMF) was proposed by Lee and Seung in 1999 in Nature [19] that can achieve nonlinear data dimensionality reduction and has strong interpretability. With the extensive attention of researchers, NMF has gradually become a mature and reliable multidimensional data processing model that is widely used in recommendation systems, pattern recognition, signal processing, computer vision, and network science. It is widely used in research fields such as recommendation systems, pattern recognition, signal processing, computer vision, and network science [20]. In addition, it can reveal the potential feature-to-feature relationship quite accurately and can also be used for other related relationships between features and for related tasks, such as node importance identification [21, 22], link prediction [23–25], and evolutionary analysis [26, 27].

In recent years, many researchers and scholars have applied NMF to recommender system discovery [14, 17, 28], which effectively improves the accuracy and efficiency of personalized recommendation results. Normally, the user product is represented as a data matrix $R \in \mathbb{R}^{N \times N}$. The matrix R can represent the interaction characteristics of users and products, such as the rating matrix and click-through rate matrix. NMF decomposes the matrix R into two nonnegative matrices and optimizes them iteratively such that $R \approx WH^T$, where $W, H \in \mathbb{R}_+^{N \times K}$, and K is the predetermined number of hidden features. In a normal situation, W denotes the basis matrix, while H denotes the data in the reduced feature space, also called the combined coef-

ficient matrix of the basis. In fact, W_{ik} can denote the probability that user i likes topic k , and H_{jk} can denote the propensity of topic k includes item j . So I have no reason to believe that $\hat{R}_{ij} = \sum_k W_{ik} H_{jk}$ can represent the probability that user i likes item j . Then, how can $\hat{R} = WH^T$ be made as close to R as possible? This involves the construction of the NMF model and the optimization process of solving it. The goal is to make WH^T as close to R as possible, and it may be assumed that $R_{ij} - \sum_k W_{ik} H_{jk}$ is a Gaussian distribution obeying a mean of 0 and a variance of σ .

$$R_{ij} - \sum_k W_{ik} H_{jk} \sim N(0, \sigma^2). \quad (3)$$

Assuming that $R_{ij} - \sum_k W_{ik} H_{jk}$ is independently and identically distributed, the likelihood function can be obtained from the Gaussian probability density function as

$$L = \prod_{i=1}^N \prod_{j=1}^N \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2} \left(\frac{R_{ij} - \sum_k W_{ik} H_{jk}}{\sigma}\right)^2\right). \quad (4)$$

The maximized likelihood can be transformed to maximize the log-likelihood, which is

$$\log L = \sum_{ij} \left(-\frac{1}{2} \left(R_{ij} - \sum_k W_{ik} H_{jk}\right)^2\right) + c = -\frac{1}{2} \left(\sum_{ij} \left(R_{ij} - \sum_k W_{ik} H_{jk}\right)^2\right) + c. \quad (5)$$

In Equation (5), c denotes a constant, since $\|R - WH^T\|_F^2 = \sum_{i=1}^N \sum_{j=1}^N (R_{ij} - \sum_k W_{ik} H_{jk})^2$ denotes the Frobenius norm and the Euclidean distance. Maximizing the log-likelihood translates into minimizing the Euclidean distance, which can then be expressed as

$$\min_{W, H \geq 0} O_1(W, H) = \min_{W, H \geq 0} \|R - WH^T\|_F^2. \quad (6)$$

Similarly, if $R_{ij} - \sum_k W_{ik} H_{jk}$ is assumed to obey Poisson distribution, its log-likelihood function is distance, which can then be expressed as

$$\log L = \sum_{i,j} \left(R_{ij} \log \frac{R_{ij}}{(WH)_{ij}} - R_{ij} + (WH)_{ij}\right) + c. \quad (7)$$

This corresponds to the K - L scatter, which also denotes the K - L distance. Then, maximizing the log-likelihood translates into minimizing the minimizing the K - L distance, which can be expressed as

$$\min_{W, H \geq 0} O_2(W, H) = \min_{W, H \geq 0} \sum_{i,j} \left(R_{ij} \log \frac{R_{ij}}{(WH)_{ij}} - R_{ij} + (WH)_{ij}\right). \quad (8)$$

In optimizing Equations (6) and (8), the objective functions O_1 and O_2 are nonconvex for both W and H , but they

TABLE 1: Symbol description.

| Symbol | Definition |
|-----------|---|
| M | Number of users |
| N | Number of projects |
| U | Set of users, $U = \{u_1, u_2, \dots, u_M\}$ |
| V | Set of projects, $V = \{v_1, v_2, \dots, v_N\}$ |
| R | User-item interaction matrix, the elements are generally nonnegative scores |
| \hat{R} | The predicted user-item interaction matrix |
| W | The basis matrix |
| H | The data in the reduced feature space, also called the combined coefficient matrix of the basis |
| C | The community indication matrix |
| B | The number of repetitions of matrix decomposition |

are convex if W or H are considered alone. Therefore, although it is not practical to find the global minimum of O_1 and O_2 , the local optimal solution can always be found by iterative optimization. Lee and Seung [15] proposed a corresponding update rule based on the gradient descent approach. For Equation (6), the update rule is

$$W_{ik} \leftarrow W_{ik} \odot \frac{(RH)_{ik}}{(WH^T H)_{ik}}, H_{jk} \leftarrow H_{jk} \odot \frac{(R^T W)_{jk}}{(H W^T W)_{jk}}. \quad (9)$$

For Equation (8), the update rule is

$$W_{ik} \leftarrow W_{ik} \odot \frac{\sum_j (R_{ij} H_{jk} / \sum_k W_{ik} H_{jk})}{\sum_j H_{jk}}, H_{jk} \leftarrow H_{jk} \odot \frac{\sum_i (R_{ij} W_{ik} / \sum_k W_{ik} H_{jk})}{\sum_i W_{ik}}. \quad (10)$$

By updating the rules, Equations (9) and (10) are iteratively updated to obtain locally optimal W and H . Then, usually by reconstructing R by W and H , we obtain the complementary matrix \hat{R} , which is

$$\hat{R} = WH^T. \quad (11)$$

There are many deformations of NMF methods [20], among which the more commonly used is symmetric NMF (SymNMF) [29, 30]. SymNMF decomposes the observation matrix A into two identical matrices, $A \approx HH^T$. SymNMF inherits the advantages of NMF, because the observed matrix A can fuse the similarity between data points and has fewer parameters. In addition, in 2013, Wang and Zhang [20] performed a systematic review of various expansion methods of NMF, and they classified NMF methods into the following: basic NMF, constrained NMF, structured NMF, and generalized NMF. In recent years, NMF-related models have been widely used by many researchers for graph image processing [31–33], complex network analysis [21, 34, 35], and recommendation systems.

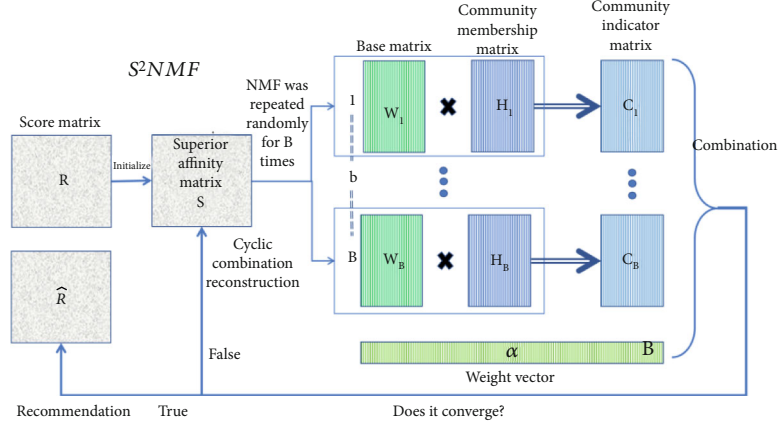
3.2. S^2 NMF Model. The proposed S^2 NMF model framework is shown in Figure 1. First, the super similarity matrix S is constructed by taking the score matrix R as input. Secondly, the NMF was repeated B times randomly, and B dimensionality reduction representations were analysed. Thirdly, B area indicator matrices are obtained by analytic strategy, and a new super similarity matrix S is obtained by combination and reconstruction. The above stochastic matrix factorization process is repeated to guide convergence to obtain the predicted scoring matrix. The S^2 NMF model framework proposed in this paper is an intelligent recommendation model with self-enhancement of information based on different types of NMF models and fusion of self-supervised information. Due to space limitations, this paper takes the classical NMF [19] as an example to introduce it in detail.

As mentioned earlier, NMF is required to solve a non-convex optimization problem that is sensitive to the initialization of variables. The details are shown in Figure 1. We propose self-supervised NMF (S^2 NMF). By exploiting the sensitivity of NMF, the model can gradually improve the recommendation performance without relying on any additional information. First, based on the classical NMF model, R is decomposed into two nonnegative matrices W^0 and H^0 . Based on the NMF basis introduced in the previous subsection, we assume that the $R_{ij} - (W^0 H^0)^T_{ij}$ errors obey a Gaussian distribution. Then, the model optimization problem at time t can be constructed as

$$\min_{W^0, H^0 \geq 0} \left\| R - W^0 H^0 \right\|_F^2, \quad (12)$$

where $R \in \{0, 1\}^{M \times N}$ is the scoring matrix, W^0 is the basis matrix of $M \times K$, and H^0 is the combined coefficient of $N \times K$ matrix.

Since the NMF model factorization has some randomness, the factorization operation is repeated randomly B times in this paper. $\{W_b^0 \in \mathbb{R}^{M \times K}\}_{b=1}^B, \{H_b^0 \in \mathbb{R}^{N \times K}\}_{b=1}^B$, where K represents the number of hidden features and the number of associations. In terms of physical meaning, it represents that users can cluster into K groups of similar

FIGURE 1: Schematic of S²NMF model.

hobbies. Therefore, the community delineation can be divided by resolving the community affiliation matrix H_b^0 to obtain the community indication matrix for each user.

$$C_{b,mk} = \begin{cases} 1, & \text{if } k = \arg \max_{k \in \{1,2,\dots,K\}} H_{b,mk}^0, \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

Considering that this community indication matrix C_b is more discriminative than the scoring matrix R , this paper constructs a super similarity matrix S as

$$S = \sum_{b=1}^B \alpha_b C_b C_b^T, \quad (14)$$

where α_m is the m th element of the vector $\alpha \in \mathbb{R}^{B \times 1}$. This weight matrix is mainly used to balance the contribution of each association degree of each association. The obtained supersimilarity matrix S can be resolved as a recommendation indicator matrix. Again, using the super similarity matrix S as input, a new community affiliation matrix can be obtained by the NMF model, and the experiment is repeated several times to obtain a better recommendation. The experiment is repeated several times to obtain better recommendation results. This process is repeated until the stopping criterion is reached or the maximum number of iterations. We represent the above process as a constrained optimization model.

$$\min_{W_b, H_b, S, \alpha} \sum_{b=1}^B \alpha_b \|S - W_b H_b^T\|_F^2 \quad \text{s.t. } W_b, H_b \geq 0, \forall b, \alpha 1 = 1, \alpha \geq 0, \quad (15)$$

where $1 \in \mathbb{R}^{B \times 1}$ denotes the full 1 vector. Clearly, by minimizing Equation (15), a better set of W_b and H_b will result in a smaller $\|S - W_b H_b^T\|_F^2$, and accordingly, a larger weighting factor α_b will be assigned. Thus, the value of α_b can measure the quality of W_b and H_b , and by resolving H_b , S can be constructed. However, there is a nonnegative constraint on α .

Equation (15) imposes an implicit weighted L_2 parametrization on α . This may lead to a rather sparse solution in the optimization of Equation (15); most elements of α are equal to or close to zero. Since our goal is to combine the contributions of multiple clusters, the extreme sparsity α is not a perfect choice. For this reason, a hyperparameter τ is introduced to control the distribution of α terms, and the final model is rewritten as

$$\min_{W_b, H_b, S, \alpha} \sum_{b=1}^B (\alpha_b)^\tau \|S - W_b H_b^T\|_F^2 \quad \text{s.t. } W_b, H_b \geq 0, \forall b, \alpha 1 = 1, \alpha \geq 0, \quad (16)$$

where τ belongs to $(1, +\infty)$. When τ is close to 1, only a few elements of α are valid. When τ tends to $+\infty$, the process of minimizing the equation causes equal weights to be assigned to α .

Therefore, τ should not be too large or too small, and the size needs to be appropriate. In this paper, we empirically set τ to 2.

By solving the equation, the final community indication matrix $C_b (b \in [1, B])$ can be obtained. Meanwhile, a better super similarity matrix S and contribution vector α are determined.

3.3. Model Optimization Algorithm. To solve the objective Equation (16), an alternating iteration strategy is proposed in this paper. First, using the fixed of S and multiple random nonnegative initialization matrices W^0 and H^0 , solving W and H and α for the objective equation, we obtain the following:

$$\min_{W_b, H_b, \alpha} \sum_{b=1}^B (\alpha_b)^\tau \|S - W_b H_b^T\|_F^2 \quad \text{s.t. } W_b, H_b \geq 0, \forall b, \alpha 1 = 1, \alpha \geq 0. \quad (17)$$

In this paper, a simple and effective criterion for adaptive termination Algorithm 1 is proposed.

In this paper, a simple and effective criterion for adaptive termination Algorithm 1 is proposed, and the pseudocode is


```

Input: Observed user-item interaction matrix  $R$   $\tau = 2B$ ;
Output: Observed user-item interaction matrix  $\hat{R}$ ;
Initialize:  $N_{iter} = 1$ ,  $maxIter = 10$ ,  $S = R$ ;
while  $N_{iter} < maxIter$  do
    Update according to algorithm 2:  $W_b, H_b, \alpha_b, b \in \{1, \dots, B\}$ ;
    Update according to equation (3):  $S$ ;
    if satisfy stop condition then
        Break;
     $N_{iter} = N_{iter} + 1$ ;
 $\hat{R} = \sum_{b=1}^B (\alpha_b)^\tau W_b H_b^T$ ;
Return  $\hat{R}$ ;

```

ALGORITHM 1: Optimization algorithm of S^2 NMF objective Equation (16).

given in Algorithm 1. It is reasonable to assume that in the first few iterations, the association detection of all partitions can be gradually improved and the consensus between them can also be increased. When the maximum consensus is reached, the consensus among them will remain at such a high level that it may even decrease and fluctuate due to the randomness of variable initialization in the iterations. Therefore, we use a different partition for the degree of agreement between them to construct the stopping criterion.

For the objective equation of the S^2 NMF model (see Equation (17)), the derivation of the update rule is similar to that of NMF and can be found in [15]. Since this objective equation has W , H , and α multiple parameters to be optimized, it belongs to a nonconvex optimization problem.

Based on the gradient descent approach, only the other parameters can be fixed separately to optimize the current parameters. Similar to Equation (9), it is easy to obtain W_{ik} and H_{jk} to update the equation as

$$H_{b,ik} \leftarrow H_{b,ik} \odot \frac{(SH_b)_{ik}}{(W_b H_b^T H_b)_{ik}}, \quad (18)$$

$$H_{b,jk} \leftarrow H_{b,jk} \odot \frac{(S^T W_b)_{jk}}{(H_b W_b^T W_b)_{jk}}. \quad (19)$$

Unlike the classical NMF, which fixes the parameters W and H and optimizes α , the objective equation can be rewritten as

$$\min_{W_b, H_b, \alpha} \sum_{b=1}^B (\alpha_b)^\tau \chi_b \quad \text{s.t. } \forall b, \alpha_b = 1, \alpha_b \geq 0, \quad (20)$$

where $\chi_b = \|S - W_b H_b^T\|_F^2$. Then, the Lagrangian function of Equation (20) can be expressed as

$$\mathcal{L} = \sum_b (\alpha_b)^\tau \chi_b - \lambda \left(\sum_b \alpha_b - 1 \right) \quad \text{s.t. } \alpha_b \geq 0. \quad (21)$$

Taking the first-order partial derivative of α_b and setting

it to 0 yields the following:

$$\alpha_b = \left(\frac{\lambda}{\tau \chi_b} \right)^{1/\tau-1}, \forall b. \quad (22)$$

Since $\sum_b \alpha_b = 1$, λ can be expressed as

$$\lambda = \left(\frac{1}{\sum_b (\tau \chi_b)^{1/1-\tau}} \right)^{\tau-1}. \quad (23)$$

Then, bringing λ into Equation (22), we obtain α as

$$\alpha_b = \frac{(\tau \chi_b)^{1/1-\tau}}{\sum_b (\tau \chi_b)^{1/1-\tau}}. \quad (24)$$

Clearly, the numerator and denominator of Equation (24) are greater than 0. Then, we have that $\alpha_b > 0$ is always greater than 0, which satisfies both the nonnegative constraint of α_b . The solution of Equation (24) satisfies the KKT (Karush-Kuhn-Tucker) condition of Equation (20), and it is a locally optimal solution. However, since the solution of Equation (20) is a convex problem, Equation (24) is a globally optimal solution of Equation (20).

In summary, the detailed optimization process of model S^2 NMF objective Equation (16) is summarized in Algorithm 2, and the pseudocode is given in Algorithm 2. The algorithm stops iterating if the difference in the maximum change of variables during two adjacent iterations is less than 0.001, e.g., $\max(\|W_b^t - W_b^{t-1}\|_\infty, \|H_b^t - H_b^{t-1}\|_\infty, \|\alpha^t - \alpha^{t-1}\|_\infty) < 10^{-3}$. W , H , and α are updated iteratively by updating the rules until the objective function converges.

3.4. Model Complexity Analysis. For the computational complexity of Algorithm 2, one of the two alternating iterative solutions W_b , H_b , and α_b are fixed first. For the solution of W_b and H_b , the computational complexity is $O(MNKB)$, and the complexity of the solution of α_b is $O(B)$. Therefore, the complexity of each iteration of Algorithm 2 is $O(MNKB)$.

Algorithm 1 solves Algorithm 2 repeatedly with the computational complexity $O(MNKB N_{iter})$, where N_{iter} is the maximum number of iterations of Algorithm 2. The

```

Input:  $S \tau = 2 B$ ;
Output:  $W, H, \alpha$ ;
Initialize:  $N_{iter} = 1$ , Randomly initialize a nonnegative matrix  $W$  and  $H$ 
while  $N_{iter} < 500$  do
  For  $b \in \{1, \dots, B\}$  do
    Update according to equation (18):  $W_b$ ;
    Update according to equation (19):  $H_b$ ;
    Update according to equation (24):  $\alpha_b$ ;
    if the convergence of objective function then
      Break;
   $N_{iter} = N_{iter} + 1$ ;
Return  $W, H, \alpha$ ;

```

ALGORITHM 2: Optimization algorithm of objective Equation (17).

computational complexity of constructing S is $O(MNK)$ of S . Therefore, the complexity of each iteration of Algorithm 1 is $O(MNKBN_{iter})$.

4. Results and Discussion

To verify the effectiveness of the S^2 NMF recommendation model proposed in this paper, this section designs comparison experiments on several standard data sets, focusing on the experimental setup, experimental results, and discussion of parameter sensitivity. The computer configuration used in our experiment is as follows: CPU: I5 6500, graphics card: Sotai 1600, and memory: 16 G.

4.1. Experimental Settings. To test the performance of the S^2 NMF recommendation model, four widely used datasets in recommender systems are selected to evaluate the model in this section: MovieLens 100 K (ML-100k), MovieLens10M (ML-1 M) (<https://grouplens.org/datasets/movielens/>), Amazon music (Amusic), and Amazon movies (Amovie) (<http://jmcauley.ucsd.edu/data/amazon/>). The statistics of the four datasets are listed in Table 2.

4.2. Comparison Algorithm

- (i) *ItemPop*. This is a ranking of items based on their popularity and the number of interactions they have. It is a nonpersonalized method and usually uses performance as a benchmark for personalization methods
- (ii) *ItemKNN* [36]. This is a standard item-based collaborative filtering method used commercially by the Amazon method
- (iii) *BPR* [37]. It is a generalized personalized ranking recommendation algorithm derived from the Bayesian analysis of the problem of the maximum a posteriori estimate

4.3. Evaluation Indicators. To comprehensively evaluate the effectiveness of the model proposed in this chapter, the experiments in this chapter use five evaluation metrics to evaluate the algorithm: recall, mrr, ndcg, hit, and precision.

These metrics examine the recommendation accuracy of the algorithm.

- (i) *Recall*. It is an evaluation metric that is the same as accuracy and indicates the proportion of relevant content in the returned recommendation list to all relevant content, regardless of the order of the returned results. It is defined as the percentage of resources preferred by users in the test set that appears in the recommendation list. It reflects the proportion of correctly predicted content in the returned recommendation list as a percentage of all known content. $R=0$ when there are no accurately predicted resources in the user's test set at all, and $R=1$ when all resources in the user's test set are accurately predicted. Thus, a higher recall indicates a more comprehensive prediction of the user's preferred resources
- (ii) *mrr: Mean Reciprocal Rank (MRR)*. The indicator response is whether we find these items placed in a more obvious position of the user, emphasizing the location relationship in a sequential manner
- (iii) *ndcg*. The whole process is called normalized discounted cumulative gain (NDCG), which is used as an evaluation metric for ranking results and evaluates the accuracy of the ranking
- (iv) *Hit*. Hit rate (HR) reflects whether the recommended sequence contains the items that the user in the recommendation sequence, i.e., whether the item selected by the user is in the recommendation sequence. This value is 1 if it exists and 0 if not
- (v) *Precision*. Accuracy, which is a common unordered evaluation index, indicates the proportion of the returned results that indicate the proportion of relevant content in the returned results, without considering the order of the returned results. When the user's recommendation list $P=0$, there are no accurate resources in the user's recommendation list, and $P=1$ when all predictions are accurate, so the

TABLE 2: Basic statistical information about the dataset.

| Statistical dimension | ML-100 k | ML-1 M | Amusic | Amovie |
|------------------------|----------|-----------|---------|---------|
| Number of users | 944 | 6,041 | 339,232 | 426,923 |
| Number of items | 1,683 | 3,707 | 83,047 | 23,966 |
| Number of interactions | 100,000 | 1,000,209 | 500,176 | 583,933 |
| Sparsity | 93.70% | 95.53% | 99.99% | 99.99% |

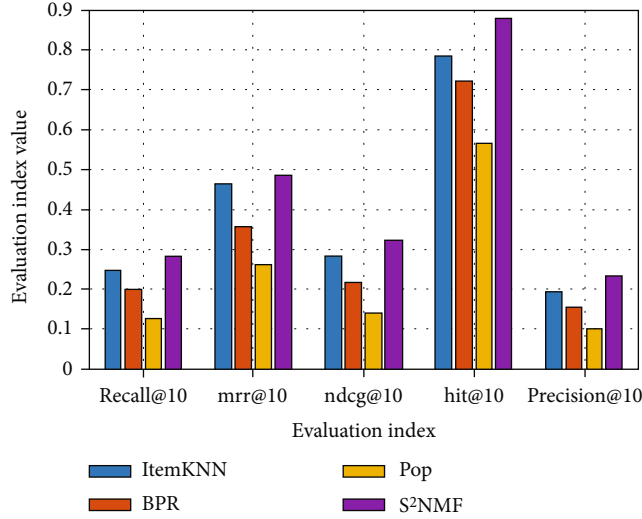


FIGURE 2: Comparison of experimental results for data set ML-100 K.

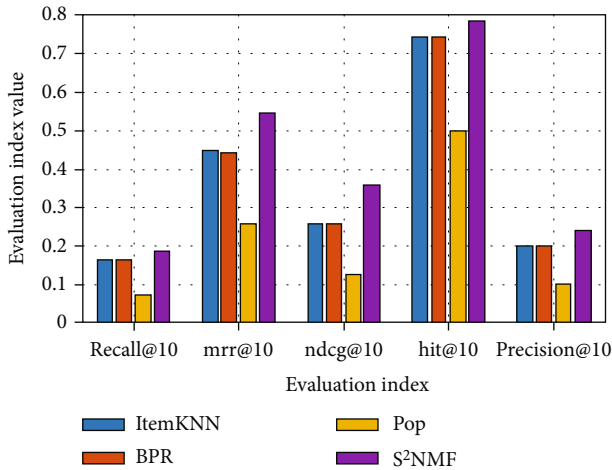


FIGURE 3: Comparison of experimental results for data set M1-1M.

higher the hit rate, the more the recommendation list matches the user's actual situation

4.4. Comparison Experiments. To fully validate the performance of the S²NMF model proposed in this chapter, the comparison results of the four data sets in Table 2 under three comparison methods and five evaluation metrics are given in this subsection, and the results are discussed and analysed in detail.

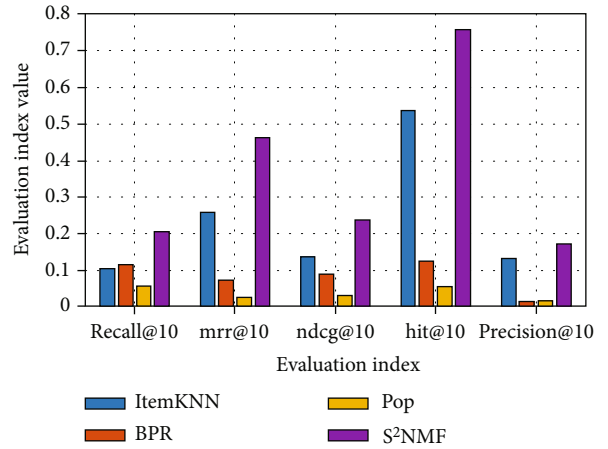


FIGURE 4: Comparison of experimental results for the Amusic dataset.

In Figure 2, the histogram of comparison results for the data set ML-100 K is given. Specifically, the horizontal coordinate represents the type of evaluation metrics, and the vertical coordinate represents the five evaluation metrics calculated by recommending items to users according to the top 10 items of the rating prediction value.

In addition, the four colours represent the four different model results. As seen overall from the figure, the results of S²NMF proposed in this chapter are all higher than the other three commonly used benchmark methods.

In Figure 3, a histogram of the comparison results for the data set M1-1 M is given. Again, the horizontal coordinates represent the type of evaluation metrics, and the vertical coordinate represents the top 10 items recommended to users according to the rating prediction value of the 5 calculated evaluation metrics. The overall figure shows that the proposed S²NMF results are higher than those of the other three commonly used benchmark algorithms.

In Figure 4, a histogram of the comparison results for the Amusic dataset is given. Again, the horizontal coordinate represents the type of evaluation metric, and the vertical coordinate represents the five evaluation metrics calculated by recommending items to users according to the top 10 items of the rating prediction value. It is generally seen from the figure that the S²NMF results proposed in this chapter are all higher than those of the other three commonly used benchmark algorithms.

In Figure 5, a histogram of the comparison results for the Amovie dataset is given. Again, the horizontal coordinate represents the type of evaluation metrics, and the vertical coordinate represents the five evaluation metrics calculated by recommending items to users according to the top 10 items of the rating prediction value. It is generally seen from the figure that the S²NMF results proposed in this chapter are all higher than the other three commonly used benchmark algorithms.

Figure 6 shows the sensitivity analysis of the parameters for the four data sets with respect to Algorithm 1. The horizontal coordinate represents the number of iterations, and

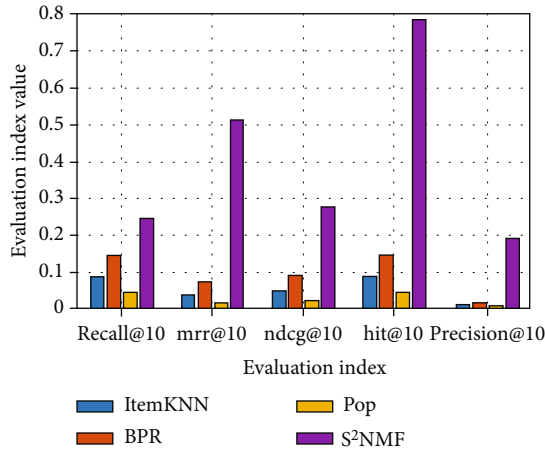


FIGURE 5: Comparison of experimental results for the Amovie dataset.

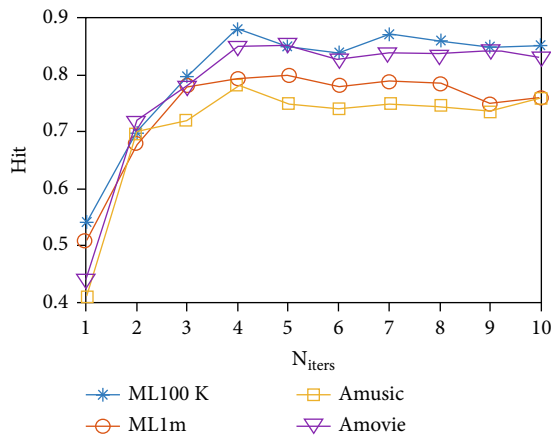


FIGURE 6: Experimental results for the number of iterations of Algorithm 1.

the vertical coordinate represents the number of iterations that hit the indicator value. From the figure, it can be generally seen that the S²NMF proposed in this chapter is relatively sensitive to the parameter (number of iterations) on the four data at less than 4, while it does not change much at greater than 4. And then, it becomes relatively stable as the number of iterations increases. Accordingly, the S²NMF model can choose a relatively small number of iterations 4 to effectively reduce the computational cost without affecting the model performance.

5. Conclusions

Based on the matrix factorization idea, this paper introduced a self-supervised learning mechanism based on the NMF model to achieve information enhancement of sparse data, proposed an easily scalable self-supervised nonnegative matrix factorization recommendation model framework S²NMF, further proposed a corresponding gradient descent optimization algorithm, and analysed the complexity of the

algorithm. Numerous experimental results showed that the S²NMF proposed in this paper has superior performance.

From the contributions of this paper, the sparse data problem of user-project interaction is solved, the interpretability of the recommendation model is enhanced based on the matrix factorization idea, and the self-supervised learning mechanism is introduced to realize the information enhancement of sparse data. However, determining the number of hidden features automatically is still an urgent problem. Likewise, exploring deep hidden features and expanding them to large-scale application scenarios is an urgent problem, which has important research significance in the field of recommendation systems.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 62141205 and the Fund Project of XJPCC (2022CB002-08, 2022CA007, 2019AB001, 2020DB005, and 2017CD010).

References

- [1] K. Krenc, V. Nimier, K. Rein, L. Snidaro, J. Biermann, and J. Garcia, "From situation awareness to intelligence: increasing complexity in fusion models for sense-making," in *In 2018 International Conference on Military Communications and Information Systems (ICMCIS)*, pp. 1–5, Warsaw, Poland, May 2018.
- [2] Z. Hu, G. Xu, X. Zheng et al., "SSL-SVD: semi-supervised learning-based sparse trust recommendation," *ACM Transactions on Internet Technology (TOIT)*, vol. 20, no. 1, pp. 1–20, 2020.
- [3] F. Wang, H. Zhu, G. Srivastava, S. Li, M. R. Khosravi, and L. Qi, "Robust collaborative filtering recommendation with user-item-trust records," *IEEE Transactions on Computational Social Systems*, vol. 9, no. 4, pp. 986–996, 2022.
- [4] L. Wu, C. Quan, C. Li, Q. Wang, B. Zheng, and X. Luo, "A context-aware user-item representation learning for item recommendation," *ACM Transactions on Information Systems (TOIS)*, vol. 37, no. 2, pp. 1–29, 2019.
- [5] E. Palumbo, G. Rizzo, and R. Troncy, "Entity 2rec: learning user item relatedness from knowledge graphs for top-n item recommendation," in *In proceedings of the eleventh ACM conference on recommender systems*, pp. 32–36, Como, Italy, 2017, August.
- [6] L. Berkani, S. Belkacem, M. Ouafi, and A. Guessoum, "Recommendation of users in social networks: a semantic and social based classification approach," *Expert Systems*, vol. 38, no. 2, p. e12634, 2021.

- [7] Z. Kang, C. Peng, and Q. Cheng, "Top-n recommender system via matrix completion," in *Proceedings in Thirtieth AAAI Conference on Artificial Intelligence*, vol. 30, Phoenix, Arizona, USA, 2016no. 1.
- [8] F. O. Isinkaye, "Matrix factorization in recommender systems: algorithms, applications, and peculiar challenges," *IETE Journal of Research*, pp. 1–14, 2021.
- [9] X. He, Z. He, J. Song, Z. Liu, Y. G. Jiang, and T. S. Chua, "Nais: neural attentive item similarity model for recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 12, pp. 2354–2366, 2018.
- [10] W. Wu, Y. Jia, S. Wang, R. Wang, H. Fan, and S. Kwong, "Positive and negative label-driven nonnegative matrix factorization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 7, pp. 2698–2710, 2020.
- [11] C. Peng, Z. Zhang, Z. Kang, C. Chen, and Q. Cheng, "Nonnegative matrix factorization with local similarity learning," *Information Sciences*, vol. 562, pp. 325–346, 2021.
- [12] O. Gouvert, T. Oberlin, and C. Fevotte, "Ordinal non-negative matrix factorization for recommendation," in *In Proceedings of the 37th International Conference on Machine Learning*, pp. 3680–3689, Vienna, Austria, 2020.
- [13] X. Zhang, L. Chen, Y. Wang, and G. Wang, "Improving incremental nonnegative matrix factorization method for recommendations based on three-way decision making," *Cognitive Computation*, pp. 1–19, 2021.
- [14] Y. Pei, N. Chakraborty, and K. Sycara, "Nonnegative matrix tri-factorization with graph regularization for community detection in social networks," in *In Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2083–2089, Buenos Aires, Argentina, 2015.
- [15] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *In Advances in Neural Information Processing Systems*, vol. 13, pp. 556–562, 2001.
- [16] A. Lancichinetti, S. Fortunato, and F. Radicchi, "Benchmark graphs for testing community detection algorithms," *Physical Review E*, vol. 78, no. 4, p. 046110, 2008.
- [17] J. Yang and J. Leskovec, "Overlapping community detection at scale: a nonnegative matrix factorization approach," in *In Proceedings of the sixth ACM International Conference on Web Search and Data Mining*, pp. 587–596, New York, United States, February 2013.
- [18] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. S. Chua, "Neural collaborative filtering," in *In Proceedings of the 26th international conference on world wide web*, pp. 173–182, Perth, Australia, April 2017.
- [19] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [20] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: a comprehensive review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 6, pp. 1336–1353, 2012.
- [21] P. Jiao, W. Yu, W. Wang, X. Li, and Y. Sun, "Exploring temporal community structure and constant evolutionary pattern hiding in dynamic networks," *Neurocomputing*, vol. 314, pp. 224–233, 2018.
- [22] S. Yang, Q. Ning, G. Zhang, H. Sun, Z. Wang, and Y. Li, "Construction of differential mRNA-lncRNA crosstalk networks based on ceRNA hypothesis uncover key roles of lncRNAs implicated in esophageal squamous cell carcinoma," *Oncotarget*, vol. 7, no. 52, pp. 85728–85740, 2016.
- [23] L. Lü, L. Pan, T. Zhou, Y. C. Zhang, and H. E. Stanley, "Toward link predictability of complex networks," *Proceedings of the National Academy of Sciences*, vol. 112, no. 8, pp. 2325–2330, 2015.
- [24] W. Yu, W. Cheng, C. C. Aggarwal, H. Chen, and W. Wang, "Link prediction with spatial and temporal consistency in dynamic networks," in *In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 3343–3349, Melbourne, Australia, 2017.
- [25] X. Xu, B. Liu, J. Wu, and L. Jiao, "Link prediction in complex networks via matrix perturbation and decomposition," *Scientific Reports*, vol. 7, no. 1, p. 14724, 2017.
- [26] W. Yu, C. C. Aggarwal, and W. Wang, "Temporally factorized network modeling for evolutionary network analysis," in *In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 455–464, New York, United States, 2017.
- [27] E. G. Tajeuna, M. Bouguessa, and S. Wang, "Modeling and predicting community structure changes in time-evolving social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 6, pp. 1166–1180, 2018.
- [28] X. Wang, P. Cui, J. Wang, J. Pei, W. Zhu, and S. Yang, "Community preserving network embedding," in *In Thirty-First AAAI Conference on Artificial Intelligence*, pp. 203–209, 2017.
- [29] D. Kuang, C. Ding, and H. Park, "Symmetric nonnegative matrix factorization for graph clustering," in *In Proceedings of the 2012 SIAM International Conference on Data Mining*, pp. 106–117, California, United States, 2012.
- [30] P. Jiao, H. Lyu, X. Li, W. Yu, and W. Wang, "Temporal community detection based on symmetric nonnegative matrix factorization," *International Journal of Modern Physics B*, vol. 31, no. 13, p. 1750102, 2017.
- [31] R. Sandler and M. Lindenbaum, "Nonnegative matrix factorization with earthmover's distance metric for image analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1590–1602, 2011.
- [32] Q. Liao and Q. Zhang, "Local coordinate based graph-regularized NMF for image representation," *Signal Processing*, vol. 124, pp. 103–114, 2016.
- [33] S. Wang, C. Deng, W. Lin, G. B. Huang, and B. Zhao, "NMF-based image quality assessment using extreme learning machine," *IEEE Transactions on Cybernetics*, vol. 47, no. 1, pp. 232–243, 2017.
- [34] X. Ma and D. Dong, "Evolutionary nonnegative matrix factorization algorithms for community detection in dynamic networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 5, pp. 1045–1058, 2017.
- [35] X. Ma, D. Dong, and Q. Wang, "Community detection in multi-layer networks using joint nonnegative matrix factorization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 2, pp. 273–286, 2019.
- [36] M. Deshpande and G. Karypis, "Item-based top-N recommendation algorithms," *ACM Transactions on Information Systems (TOIS)*, vol. 22, no. 1, pp. 143–177, 2004.
- [37] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," 2012, <http://arxiv.org/abs/1205.2618>.