Paper Recommendation Using SPECTER with Low-Rank and Sparse Matrix Factorization

Panpan Guo[1], Gang Zhou[1*], Jicang Lu[1], Zhufeng Li[1], and Taojie Zhu[1]
[1] State Key Laboratory of Mathematical Engineering and Advanced Computing
Zhengzhou, Henan 450001 China
[e-mail: guopan14517wpw@163.com, gzhougzhou@126.com, lujicang@sina.com, 20086538@qq.com, zhutaojie158@163.com]
[*]Corresponding author: Gang Zhou

Abstract

With the sharp increase in the volume of literature data, researchers must spend considerable time and energy locating desired papers. A paper recommendation is the means necessary to solve this problem. Unfortunately, the large amount of data combined with sparsity makes personalizing papers challenging. Traditional matrix decomposition models have cold-start issues. Most overlook the importance of information and fail to consider the introduction of noise when using side information, resulting in unsatisfactory recommendations. This study proposes a paper recommendation method (PR-SLSMF) using document-level representation learning with citation-informed transformers (SPECTER) and low-rank and sparse matrix factorization; it uses SPECTER to learn paper content representation. The model calculates the similarity between papers and constructs a weighted heterogeneous information network (HIN), including citation and content similarity information. This method combines the LSMF method with HIN, effectively alleviating data sparsity and cold-start issues and avoiding topic drift. We validated the effectiveness of this method on two real datasets and the necessity of adding side information.

Keywords: paper recommendation, SPECTER, low-rank and sparse matrix factorization, heterogeneous information network.

1. Introduction

The surge in the number of papers has led to the problem of information overload, which has brought great inconvenience to scientific research. A paper recommendation method is one effective solution. Recommending articles to researchers is significant as it enables them to obtain relevant articles quickly and efficiently. These citation recommendation models are divided into content-based (CB) [1-3], collaborative filtering (CF) [4,5], and graph-based (GB) [6-8]. The CB model generates recommendations using descriptions and features from papers and user profiles. Its recommendation results tend to be similar to what users have liked. The CF-based model uses past user ratings and social networks. Due to limited user ratings, they often experience inaccurate predictions due to sparsity and cold start [9]. The GB model uses additional relationships between nodes in the network. However, traditional GB models view recommendation as a link prediction task. Therefore, these methods excessively weighted old and outdated nodes in the network [9,10]. In this study, we focus on CF models.

CF automatically predicts the interests of specific users based on the collective historical score records of similar users or items and has been extensively studied in the field of paper recommendation [11-13]. The most representative method of CF is matrix completion [12]. This method decomposes the original rating matrix into two low-rank matrices with joint latent factor spaces. One matrix represents the potential interests of users, while the other matrix represents the factors that contribute to the item's ownership. The inner product of the user vector and the item vector produces the recommendation result. This method typically encounters sparsity and cold-start issues, as the number of interactions between users and projects is usually limited. The appropriate solution to the above problem is to add more relevant information. In academic datasets, in addition to authors and papers, there are various types of nodes (such as publications, research fields, and keywords) and relationships (writing, publishing, co-authorship, and inclusion, for example). Therefore, datasets typically form heterogeneous networks. Other relationships have been jointly decomposed to use these types of relationships [4,14-17]. Some variants of CF have been proposed. However, the above methods still have shortcomings, that is, the introduction of noise when adding relevant information. Huang et al. [18] concluded that citation-based methods may introduce more noise into the reference graph and lead to topic drift [19]. Topic drift can be defined as deviating from the theme of a study. We believe that the research field is too broad, with a high overlap between keywords and publications, and cannot effectively represent the content of a paper. These methods may introduce more noise, leading to topic drift.

This study proposes a paper recommendation method using SPECTER [20] with low-rank and sparse matrix factorization (PR-SLSMF). We use SPECTER to learn the representation of paper content and calculate the similarity between papers. Using similarity as its weight, the proposed method adds links between similar papers and combines them with citation networks to form a weighted heterogeneous information network (HIN). We extract the composite relationship matrix from the weighted HIN and integrate these useful relationships into the learning process of Go Decomposition (GoDec) [21] in the form of deviation terms. This method can use the rich citation and paper content information in HIN. It also offsets the inherent sparsity and cold-start issues of traditional CF-based models. In addition, PR-SLSMF does not use information such as research keywords and publications, avoiding topic drift in recommendation results.

(1) We propose a new recommendation model called PR-SLSMF, which predicts scores by learning the sum of low-rank and sparse matrices. It improves the accuracy of

recommendations by reducing error propagation during the intermediate learning process.

(2) We use SPECTER to learn the representation of paper content. An abstract can effectively represent the content of a paper, avoid introducing noise, and effectively prevent topic drift.

(3) We explore a universal method combining CF with citation and text similarity information. This model effectively alleviates data sparsity and cold-start issues.

(4) We conduct experimental studies on two real datasets to verify the effectiveness of PR-SLSMF.

The remaining parts of this paper are organized as follows. Section 2 introduces the related work of paper recommendation and LSMF. Section 3 provides preliminaries and problem definitions. Section 4 introduces PR-SLSMF in detail. Section 5 introduces and analyzes the experimental research and summarizes our work. The last section discusses our research plans for future work.

2. Related Work

2.1 Paper recommendation

The CF-based model uses the opinions of user neighbors (explicit or implicit) for paper recommendation [4,17]. For this reason, McNee et al. [22] viewed the author's citation as a positive vote for the paper and applied CF to recommend scientific papers. Wang and Blei [17] proposed a collaborative topic regression (CTR) model, which combines the topic model with a collaborative filter to recommend articles. However, a CF-based model has issues with cold start and sparsity. Wang et al. [4] extended the CTR model by integrating network structure and user project feedback information using a principled hierarchical Bayesian model. Bansal et al. [23] proposed a CF-based paper prediction model. This model learns the article's content through a gated recurrent unit network. Both models use content and side information to alleviate the sparsity and cold-start issues that traditional CF-based models face.

Compared with CF, the CB model generates recommendations using corresponding papers and users' content, features, and descriptions [24]. For this reason, Bollacker et al. [25] proposed CiteSeer, the first CB-based academic paper recommendation system to use TFIDF vectors and citation relationships. Similarly, the CB model [26] uses latent Dirichlet assignment to generate latent representations of the textual content of research papers. Then, the model calculates the similarity between these representations to make the final recommendation. Sugiyama and Kan [27] established author profiles from a list of published papers and recommended academic papers by capturing authors' research preferences. The author's profile has been strengthened through publications that have cited the author's work. Khadka et al. [28] used their thematic information to generate a high-level representation of a paper to generate citation suggestions. However, the CB model can be plagued by cold start and over-specialization issues [29].

In the past decade, deep learning has been applied to paper recommendations [30], such as a multilayer perceptron [31], a convolutional neural network (CNN) [7,32], a recurrent neural network [23], and a confrontation generation network [8]. For example, Jeong et al. [33] used graph convolutional networks and bidirectional encoders from transformers to represent documents and contexts for context-aware paper citation recommendations. Dai et al. [34] proposed a gated relational stacked denoising autoencoder with localized author embedding that uses author information by introducing a novel author embedding method. Furthermore, this model uses three encoder–decoder neural network architectures to alleviate the scalability

problem of author embedding vectors faced by current global citation recommendation models [35]. Zafar et al. [8] proposed a network embedding model using generative adversarial networks (GCR-GAN) that uses SPECTER [20] and denoising autoencoder networks to capture the proximity of network structure and learn semantic preserving graph representations. Ali et al. [36] proposed a scientific paper recommendation model using SPECTER and storage networks. The model consists of three modules. The first is an embedding module that uses the SPECTER document embedding model to learn the representation of paper content that preserves context. The second is a personalized module that captures the preferences of researchers by studying research fields, author information, and citation information. The third uses long-term contextual information and significant factors to learn long-term dependent memory network modules.

Recently, complex graph and network representation techniques have used semantic relationships between graphs or network nodes to learn vector representations of corresponding nodes. Various paper recommendation models [37,38] use this embedding method to propose recommendations. For this reason, Gupta and Varma [39] used Doc2vec [40] and DeepWalk [41], respectively, to learn the embedding of paper content and network structure. Then, the model uses the similarity between learning representations to generate paper recommendations. Similarly, Kong et al. [42] used Paper2vec [43] and Struct2vec [44] embedding methods to generate recommendations by integrating text-based vector representation and structured embedding, respectively.

In contrast, Chen et al. [45] used Node2vec [46] to mine semantic relationships between heterogeneous bibliographic network objects and learn the embedding of participating nodes (i.e., papers, authors, content, and places). Last is using learned node embedding to make final recommendations for the queried paper. Current recommendation models based on network representation learning have achieved better results than traditional CB and CF models. However, these models may introduce noise when using semantic relationships to handle cold-start problems.

2.2 Low-rank and sparse matrix factorization

LSMF has attracted a growing amount of attention in many fields, such as movie and product recommendation [47,48], anomaly detection [49-51], image processing [52,53], video surveillance [54], low-rank textures [55], image processing and computer vision [56].

Research into the recommendation area also focuses on LSMF. Ning et al. [47] proposed a sparse linear method, which multiplies the sparse aggregation and scoring matrices to obtain the prediction matrix. Zhao et al. [57] proposed a low-rank and sparse matrix completion (LSMC) algorithm and verified that LSMC could be used to recommend products to users in food and movie datasets. LSMC relies only on the original interaction matrix to learn low-rank and sparse matrices. Thus, it can alleviate the error in an intermediate process.

LSMF cannot use valuable link information in HIN. Candès et al. [58] decomposed the original interaction matrix into a special matrix with low-rank characteristics and others with sparse characteristics. The rank sum of the low-rank matrix and the basis of a sparse matrix are automatically obtained using a convex optimization algorithm, which cannot control the complexity of the model. Dai et al. [59] extended the original GoDec with link information in heterogeneous scientific networks for a paper recommendation. Integrating the author affinity matrix and paper affinity matrix into GoDec's learning process effectively alleviates sparsity and cold-start issues to a certain extent.

3. problem definition

Definition (*Heterogeneous Information Network*): $G = (\mathcal{V}, \mathcal{E})$ is a special case of HIN [8] with two mapping functions, namely node type mapping function $\phi: \mathcal{V} \rightarrow \mathcal{A}$ and relation mapping function $\psi: \mathcal{E} \rightarrow \mathcal{R}$. Each node $v \in \mathcal{V}$ pertains to a specific node type, and each edge $e \in \mathcal{E}$ pertains to a specific relation type. In a HIN, we have $|\mathcal{A}| + |\mathcal{R}| > 2$.

Problem Statement: Given the seed researcher a and HIN $G = (\mathcal{V}, \mathcal{E})$, the proposed model aims to recommend the top-N relevant papers for a.

4. Methodology

Fig. 1 shows the architecture of the proposed model, which has a three-step process. First, it learns the embedding of the paper content using the pre-trained SPECTER model. Next, it uses the paper's citations and content similarity relation to construct HIN to extract the composite matrix. Finally, matrix decomposition is used to recommend the top-n relevant papers for the seed researcher. The components responsible for these steps are discussed in detail in the following sections.
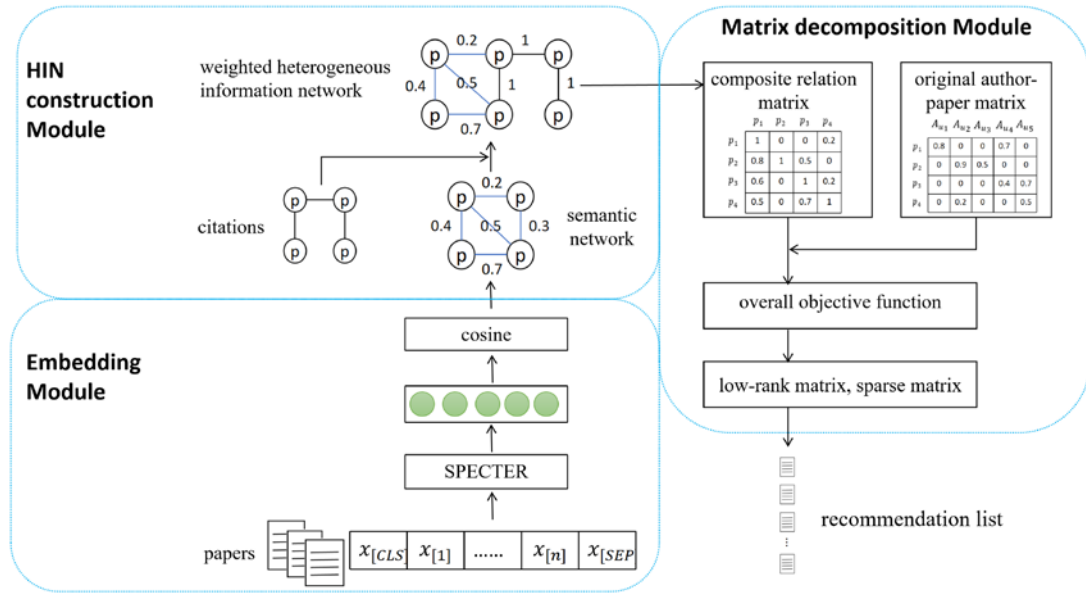

Fig. 1. PR-SLSMF framework.

4.1 Content-based paper embedding

This section explains how SPECTER learns the representation of scientific papers. To learn semantic perceptual embedding, SPECTER uses a reference perceptual converter to adjust the embedding learned through SciBERT [60]. Traditional language models such as Doc2vec and SBERT cannot capture more relevant contextual information corresponding to scientific literature and do not consider the correlation between literature based on citation relationships. SciBERT used a scientific literature library for training to address these issues. To learn the representation of c_{p_i} in the paper, SPECTER first uses Transformer LM to encode the connected text of the paper, which is defined as follows:

$$c_{p_i} = Transformer(input)_{[CLS]}, \tag{1}$$

where Transformer represents the forward function of the transformer. The model takes the concatenated WordPieces (of title and abstract) and the $[CLS]$ token as inputs, which are separated using the $[SEP]$ token. In addition, SPECTER uses citation relations as a relatedness signal to enrich the vector representations learned by SciBERT. In addition, the SPECTER model uses "hard negatives" and "simple negatives" to learn more optimized embeddings. The SPECTER model learns nodes' CB embeddings by optimizing the following margin loss objective:

$$\mathcal{L} = max\{(d(p^Q, p^+) - d(p^Q, p^-) + w, 0)\}, \tag{2}$$

where p^Q is used for the query paper, p^+ denotes the relevant paper, and p^- represents an irrelevant paper. Additionally, d denotes the Euclidean norm distance, and w represents the margin. The function of w is to ensure that the value of p^+ is at least w closer to p^Q than p^-. During the training process, the purpose of the model is to minimize the distance between p^Q and p^+ while maximizing the distance between p^Q and p^-. During inference, for an input paper p_i, the model learns a CB paper embedding c_{p_i} by activating the transformer pool output of the SPECTER model. In this way, the model successfully captures the contextual information of the paper.

4.2 HIN construction module

The composite relation matrix contains two types of relations: citation and semantic correlation. Therefore, we consider adding content-related links to the citation network. The specific steps are as follows.

First, for any two articles p_i and p_j, we use the cosine formula to calculate their semantic similarity $\mu(p_i, p_j)$. The semantic similarity (content–content similarity) between p_i and p_j is calculated using a cosine formula; that is:

$$\mu(p_i, p_j) = \frac{\overrightarrow{c_{p_i}} \cdot \overrightarrow{c_{p_j}}}{\left\| c_{p_i} \right\| \cdot \left\| c_{p_j} \right\|}, \tag{3}$$

where the most similar top q values are selected to join the function representing document relations.

Next, we construct a weighted HIN $G = (V, E, W, \psi, \theta)$ without direction. $\chi = P$ indicates that there is only one type of vertex: papers. $R = (p_{cit}p, p_{sim}p)$ contains two types of relations: $p_{cit}p$ represents citing relation, and $p_{sim}p$ represents semantic similarity relation. $\omega = \{\omega_{p_{cit}p}, \omega_{p_{sim}p}\}$ is the weight of R. $\psi: E \rightarrow R$ is a mapping function whose variables are vertex types, and $\theta: W \rightarrow \omega$ is a mapping function whose variables are attribute value types. We need to limit the number of links between two articles. There are three cases in the relation between any two articles: (1) citation relations, (2) semantic similarity relations, and (3) both citation and semantic similarity relations. The weight values that define the relation are as follows:

$$\omega = \begin{cases} \mu(p_i, p_j), \omega = p_{sim}p \\ 1, others \end{cases}. \tag{4}$$

When the relation type is a citation similarity relation or both citation and semantic similarity relations, the weight is assumed to be 1. When there are only semantic similarity relations between two articles, their semantic similarity $\mu(p_i, p_j)$ is taken as the final weight.

After constructing the network, we adjust the number of links to obtain different composite relation matrices.

4.3 Matrix decomposition module

4.3.1 Basic GoDec

We chose GoDec because it efficiently and robustly decomposes an interaction matrix $X \in \Re^{n \times r}$ into a low-rank matrix $L \in \Re^{n \times r}$ and a sparse matrix $S \in \Re^{n \times r}$ and it takes less time with bilateral random projection [61]. The objective function of GoDec given by Zhou et al. [21] can be expressed as follows:

$$arg\min_{L,S}\|X - L - S\|_F^2$$
$$s.t.\, rank(L) \leq r, card(S) \leq k, \qquad (5)$$

where $\|\cdot\|_F$ represents the Frobenius norm of a matrix, and $\|\cdot\|_F^2$ represents the square of the Frobenius norm of a matrix. r and k are the maximum values of the rank of L and the cardinality of S, respectively. The objective function (5) only contains original author–paper interaction information. In the next section, we extract the composite relation matrix from the constructed HIN and then integrate it into the original GoDec.

4.3.2 Integrating composite relation matrix into GoDec

We integrate the extracted composite relation matrix ω_{PP}. For convenience, we use M to represent ω_{PP}. In the sparse matrix S in (5), each of its columns can be treated as an author's rating for papers. A new author–paper incidence matrix MS can be obtained by left-multiplying S by the composite relation matrix M. **Fig. 2** shows the process of adding side information to a sparse matrix; it also shows how MS is obtained. The element in MS is a linear combination of a row vector in M and a column vector in S. MS can be thought of as a completion of S. From the perspective of the graph, S can be thought of as a bipartite network with few links. The new non-zero elements in MS represent the weight of the new links. The results are interpretable. First, the original links of S (position of numbers in bold italics in matrix S and blue lines in the author–paper network) are preserved in MS. Second, the new author–paper incidence matrix MS (orange numbers in MS and orange lines in the author–paper network) is derived from the relevance of the papers. It can be seen that MS is a new interaction matrix completed by M and S. Therefore, it is reasonable for MS to have the same low-rank and sparse features as the original matrix X. Therefore, we can obtain the following objective function based on MS:

$$arg\min_{L,S,N}\|MS - L - N\|_F^2$$
$$s.t.\, rank(L) \leq r, card(S) \leq k_1, card(N) \leq k_2, \qquad (6)$$

where N is the loss error of MS and L, and its cardinality is controlled within the range of no more than k_2.
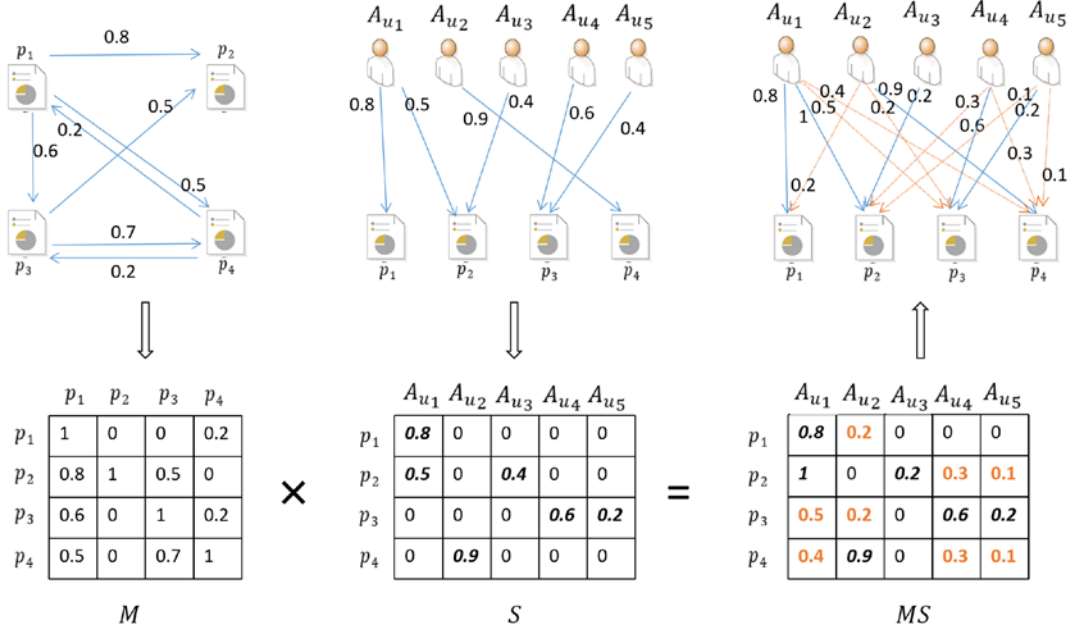
Fig. 2. The process of adding side information to the interaction matrix S.

5.3.3 Overall objective function

The proposed PR-SLSMF method consists of decomposing the original author–paper matrix and the new author–paper matrix. The decomposition of the original author–paper matrix uses the basic GoDec. Considering the new author–paper matrix, we can integrate it into GoDec as an offset term. We put these two parts together to obtain the following overall objective function of PR-SLSMF:

$$arg\min_{L,S,N}\|X - L - S\|_F^2 + \alpha\|MS - L - N\|_F^2$$
$$s.t. rank(L) \leq r, card(S) \leq k_1, card(N) \leq k_2, \tag{7}$$

where α is a regularization parameter to control the weight of the offset term.

5.3.4 Estimation process

The solution in (7) can be equivalent to that of the three sub-problems in (8). Each parameter is estimated alternately in each iteration until they converge.

$$\begin{cases} L_t = arg\min_{rank(L)\leq r}\|X - L - S_{t-1}\|_F^2 + \alpha\|MS_{t-1} - L - N_{t-1}\|_F^2 \\ S_t = arg\min_{card(S)\leq k_1}\|X - L_t - S\|_F^2 + \alpha\|MS - L_t - N_{t-1}\|_F^2 \\ N_t = arg\min_{card(N)\leq k_3}\|MS_t - L_t - N\|_F^2 \end{cases} \tag{8}$$

Next, we explain in detail the updating rules of L, S and N at t-th iteration respectively.
Updating rule for L

We first derive the objective function of L at t-th iteration. The function of L in (8) can be converted in the following ways:

$$
\begin{aligned}
J_L &= \|X - L - S_{t-1}\|_F^2 + \alpha\|MS_{t-1} - L - N_{t-1}\|_F^2 \\
&= tr[(X - L - S_{t-1})^T(X - L - S_{t-1})] + \alpha tr[(MS_{t-1} - L - N_{t-1})^T(MS_{t-1} - L - N_{t-1})] \\
&= tr[[(X - S_{t-1})^T(X - S_{t-1})] + L^T L - (X - S_{t-1})^T\mathrm{L} - L^T(X - S_{t-1})] \\
&\qquad + \alpha tr[(MS_{t-1} - N_{t-1})^T(MS_{t-1} - N_{t-1}) + L^T L - (MS_{t-1} - N_{t-1})^T\mathrm{L} \\
&\qquad - L^T(MS_{t-1} - N_{t-1})] \\
&= tr[[(X - S_{t-1})^T(X - S_{t-1})] + \alpha[(MS_{t-1} - N_{t-1})^T(MS_{t-1} - N_{t-1})] - [(X - S_{t-1})^T \\
&\qquad + \alpha(MS_{t-1} - N_{t-1})^{\mathrm{T}}]L - L^T[(X - S_{t-1}) + \alpha(MS_{t-1} - N_{t-1})] + (1 \\
&\qquad + \alpha)L^T L] + \mathrm{C} \\
&= tr[\left(\frac{X - S_{t-1} + \alpha MS_{t-1} - \alpha N_{t-1}}{1 + \alpha} - L\right)^T \times (\frac{X - S_{t-1} + \alpha MS_{t-1} - \alpha N_{t-1}}{1 + \alpha} - L)] + \mathrm{C} \\
&= \left\|\frac{X - S_{t-1} + \alpha MS_{t-1} - \alpha N_{t-1}}{1 + \alpha} - L\right\|_F^2, \qquad (9)
\end{aligned}
$$

where C is a constant generated in the intermediate process and ignored in the final result. For convenience, we use Z to represent $\frac{X - S_{t-1} + \alpha MS_{t-1} - \alpha N_{t-1}}{1 + \alpha}$, that is

$$
Z = \frac{X - S_{t-1} + \alpha MS_{t-1} - \alpha N_{t-1}}{1 + \alpha}. \qquad (10)
$$

The function in (10) can be settled by estimating L_t via singular value hard thresholding of Z:

$$
L_t = \sum_{i=1}^{r} \lambda_i U_i V_i^T, svd(Z) = U \Lambda V^T, \qquad (11)
$$

where $svd(\cdot)$ denotes singular value decomposition.

Updating rule for S

Next, we give the update rule of S at t-th iteration. The function of S in (8) can be converted in the following ways:

$$
\begin{aligned}
J_S &= \|X - L_t - S\|_F^2 + \alpha\|MS - L_t - N_{t-1}\|_F^2 \\
&= tr[(X - L_t - S)^T(X - L_t - S)] + \alpha tr[(MS - L_t - N_{t-1})^T(MS - L_t - N_{t-1})] \\
&= tr[[(X - L_t)^T(X - L_t)] + S^T S - (X - L_t)^T S - S^T(X - L_t)] + \alpha tr[(-L_t - N_{t-1})^T(-L_t \\
&\qquad - N_{t-1}) + (MS)^T(MS) - (-L_t - N_{t-1})^T MS - (MS)^T(-L_t - N_{t-1})] \\
&= tr[(X - L_t)^T(X - L_t) + \alpha(-L_t - N_{t-1})^T(-L_t - N_{t-1}) + (I + \alpha M^T M)S^T S \\
&\qquad - [-(X - L_t)^T + \alpha(-L_t - N_{t-1})M]S - S^T[-(X - L_t) - \alpha M^T(-L_t \\
&\qquad - N_{t-1})]] + \mathrm{C} \\
&= tr[\left(\frac{X - L_t + \alpha M^T L_t - \alpha M^T N_{t-1}}{I + \alpha M^T M} - S\right)^T \times (\frac{X - L_t + \alpha M^T L_t - \alpha M^T N_{t-1}}{I + \alpha M^T M} - S)] + \mathrm{C} \\
&= \left\|\frac{X - L_t + \alpha M^T L_t - \alpha M^T N_{t-1}}{I + \alpha M^T M} - S\right\|_F^2, \qquad (12)
\end{aligned}
$$

where C is a constant that deserves no attention. For convenience, we use D to represent $\frac{X - L_t + \alpha M^T L_t - \alpha M^T N_{t-1}}{I + \alpha M^T M}$, that is

$$
D = \frac{X - L_t + \alpha M^T L_t - \alpha M^T N_{t-1}}{I + \alpha M^T M}. \qquad (13)
$$

The function in (10) can be settled by estimating S_t via entry-wise hard thresholding of D:

$$S_t = P_\Omega(D), \Omega: |D_{i,j\in\Omega}| \neq 0 \; and \; \geq |D_{i,j\in\bar\Omega}|, |\Omega| \leq k_1, \quad (14)$$

where $|D|$ is the l_0 norm of D. $P_\Omega(D)$ indicates projection matrix to the entry set D. Ω is the nonzero entry set with the top k_1 values in D.

Updating rule for N

After removing the irrelevant constant, we get the objective function of N:

$$N_t = arg \min_{card(N)\leq k_3} \|MS_t - L_t - N\|_F^2. \quad (15)$$

According to the function in (15), we can estimate N as follows.

$$N_t = P_\Psi(MS_t - L_t), \Psi: |(MS_t - L_t)_{i,j\in\Psi}| \neq 0 \; and \; \geq |(MS_t - L_t)_{i,j\in\bar\Psi}|, |\Psi| \leq k_2, \quad (16)$$

where Ψ is the nonzero entry set with the top k_2 values in $MS_t - L_t$.

5.3.5 Acceleration procedure of PR-SLSMF With BRP

Low-rank approximation with BRP is optimal and an efficient operation. To speed up the iteration of L, we replace the SVD decomposition method used in (11) with BRP. We obtain a new matrix $\tilde{Z}$ as follows.

$$\tilde{Z} = (ZZ^T)^q Z, \quad (17)$$

where q represents a power exponent. Z and $\tilde{Z}$ have the same singular vector and the singular value of $\tilde{Z}$ decays faster than Z. In view of these two points, we calculate the BRP of $\tilde{Z}$ instead of Z.

Therefore, the optimized low-rank matrix can be denoted by:

$$\tilde{L} = Y_1(A_2^T Y_1)^{-1} Y_2^T, \quad (18)$$

where $Y_1 = \tilde{Z}A_1$ and $Y_2 = \tilde{Z}^T A_2$ are the BRP of $\tilde{Z}$. $A_1 \in \Re^{n\times r}$ and $A_2 \in \Re^{m\times r}$ are random matrices. The error of (18) can be reduced by increasing the value of q. We compute the QR decomposition of Y_1 and Y_2:

$$Q_1 R_1 = Y_1, Q_2 R_2 = Y_2, \quad (19)$$

We acquire the low-rank approximation of $\tilde{Z}$ with (17), (18) and (19):

$$L = (\tilde{L})^{\frac{1}{2q+1}} = Q_1[R_1(A_2^T Y_1)^{-1} R_2^T]^{\frac{1}{2q+1}} Q_2^T. \quad (20)$$

After changing the decomposition method, it is reduced from $O(m^2 n)$ to $O(r^2 m)$ that the time complexity required for each iteration. The estimation process of PR-SLSMF is described in Algorithm 1. L, S and N are updated from (14) to (16) iteratively until they converge. At the same time, after the iteration process is completed, we can obtain the optimal solutions L, S and N of the overall objective function in (10). By adding the obtained matrix L and S, the predicted author–paper incidence matrix $\tilde{X}$ obtained. Finally, the papers with top-n most relevant with the author in $\tilde{X}$, which are zeros in the X, will be recommended to corresponding

authors.

Algorithm 1 can make the total objective function converge to a local minimum value.

Algorithm 1 Estimation process of PR-SLSMF.
Input: author–paper incidence matrix X, composite relation matrix M , rank r, cardinality k_1 and k_2, bias term α.
Output: low-rank matrix L, sparse matrix S
Initialize: $L0 := X, S0 := 0, N0 := 0, t0 := 0$
1. for $\alpha = 0$ to 1 do
2.   while $t <= max\_iter$ do
3. $t := t + 1;$
4. $Z = \frac{X - S_{t-1} + \alpha M S_{t-1} - \alpha N_{t-1}}{1 + \alpha}$
5. $\tilde{L} = (ZZ^T)^q Z$
6. $Y_1 = \tilde{L} A_1, A_2 = Y_1$
7. $Y_2 = \tilde{L}^T A_2 = Q_2 R_2, Y_1 = \tilde{L} Y_2 = Q_1 R_1$
8.   if rank $(A_2^T Y_1) < r$
9. $r := $ rank $(A_2^T Y_1)$, go to the first step;
10. end if
11. $L = Q_1 [R_1 (A_2^T Y_1)^{-1} R_2^T]^{\frac{1}{2q+1}} Q_2^T$
12. $S_t = P_\Omega \left(\frac{X - L_t + \alpha M^T L_t - \alpha M^T N_{t-1}}{1 + \alpha M^T M} \right)$, Ω is the nonzero entry set with the top k_1 values in $\left| \frac{X - L_t + \alpha M^T L_t - \alpha M^T N_{t-1}}{1 + \alpha M^T M} \right|$.
13. update L_t according to (20);
14. update S_t according to (14);
15. update N_t according to (15);
16. end while
17. end for

6. Experiments and results

6.1 Dataset

We used the ACL anthology network (AAN) and DBLP[1] datasets for experiments. The data[2] we used is the data in Chen's paper. Some statistics of these datasets are given in **Table 1**. We selected papers published between 1965 and 2012 in the AAN dataset and between 2014 and 2019 in the DBLP dataset. Because citations, titles, and abstracts are used to obtain the composite relation matrix, we need to preprocess the data. We kept those with all the needed information, such as citations, titles, and abstracts. If a paper missed any information, we discarded it. We combine the title and abstract of each paper into a text and perform tokenization, lemmatization, and removal of the stop words. The AAN dataset contains 3936 authors and 13,375 papers; the DBLP dataset contains 3768 authors and 14,545 papers. According to the statistics, there are 56,002 and 10,921 pairs of writing relations. We regard the writing relation as an interactive relation. The sparsity of the interaction matrix was 99.89% and 99.98%. We adopt the method of stratified random sampling to determine the test sets.

---

[1] https://www.aminer.cn/billboard/citation.
[2] https://github.com/AHULiuYang/WHIN-CSL.

Table 1. Statistics of the datasets

Datasets	Authors	Papers	Citation relations	sparsity	Proportion of cold-start data in the test set	Release time
AAN	3936	13275	56002	99.89%	6.25%	2013-02
DBLP	3768	14545	10921	99.98%	20%	2020-04

6.2 Evaluation indicator

We divided the interactive data into 10 groups, randomly taking one group as the test set and the rest as the training set. To verify that the model can improve the cold-start recommendation effect, we took a certain proportion of cold-start researchers as test samples. The ratios of warm-start researchers to cold-start researchers were 15:1 and 4:1 for AAN and DBLP, respectively. We first learned two satisfactory matrices from the training set and then submitted the top N papers in the approximate interaction matrix to the authors in the test set. To reduce error, we changed the test set to do cross-validation experiments.

(1) Recall. Calculate the ratio of the number of references in the recommended list to the total number of references actually cited. Recall uses this percentage to check model recommendations. The higher the recall value, the higher the coverage of the model. We get the Recall by

$$Recall = \frac{1}{Q}\sum_{j=1}^{Q}\left(\frac{R_p \cap T_p}{T_p}\right) \tag{21}$$

where Q is the quantity of queries. Given a query in the test set, R_p is the list of Top-N documents obtained by querying the target article p for recommendation, and T_p is the group of documents that cite p.

(2) NDCG. NDCG evaluate the ranking of the truly relevant papers recommended by the model [30] in the first k papers. It is obtained by dividing the cumulative gain of loss by the maximum cumulative gain. The formula for calculating NDCG is

$$NDCG@N = \frac{1}{Q}\sum_{j=1}^{Q}\left(\sum_{i=1}^{k}\frac{2^{r_i}-1}{\log_2(i+1)}\Big/IDCG@N\right) \tag{22}$$

where Q is the quantity of queries and N is the quantity of the recommended documents. r_i is the score of the i-th document in the recommendation list. $r_i = 1$ if the document is relevant, and $r_i = 0$ if it is not. NDCG@N is the excellent ranking of the results such that NDCG@N = 1 if an ideal ranking is returned.

6.3 Baselines

We selected the following model as the baseline:

(1) NMF [62]: For any interaction matrix A, NMF finds two non-negative matrices U and V to satisfy $A = UV$.

(2) ConvCN [7]: ConvCN uses CNN-based knowledge graph embedding to encode citation relationships between papers, improving the performance of GB citation recommendation methods by capturing citation relationships. We have used the default parameter settings for all parameters.

(3) LSMFPRec [59]: LSMFPRec combines the low-rank and sparse matrix factorization method with the paper–author affinity matrix and paper–paper affinity matrix to improve the recommendation performance of the model. The paper affinity matrix includes citation relationships, indirect relationships (keywords, publications, and authors), and semantic correlation relationships. Author affinity matrices are extracted from heterogeneous scientific

networks. All parameters are the same as the experimental settings of the LSMFPRec model.

6.4 Model performance analysis

The performance of these methods on the AAN and DBLP datasets is shown in **Table 2** and **3**. PR-SLSMF is much better than others in all cases. We provide more detailed analyses in the following text.

As can be seen from **Table 2** and **3**, PR-SLSMF is much better than NMF, ConvCN, and LSMFPRec. The performance of NMF and ConvCN is poor, which may be caused by excessively sparse data. LSMFPRec outperforms not well, which may be due to introduced noise.

Taking Recall@300 in the AAN dataset as an example, Recall@300 of PR-SLSMF is 0.0341 larger than that of PR-DLSMF and 0.0718 larger than that of LSMFPRec. Regarding NDCG@300, NDCG @300 of PR-SLSMF is 0.0272 larger than that of PR-DLSMF and 0.0618 larger than that of LSMFPRec. Therefore, we can conclude: (1) PR-SLSMF alleviates the data sparsity problem and improves a model's performance. (2) Appropriate side information effectively improves model performance, while inappropriate side information may introduce noise and reduce model performance.

This section discusses the different parameter settings used in PR-SLSMF and their effect on the final recommendation.

Table 2. Performance evaluation between different models regarding Recall and NDCG on AAN.

Method	Recall @10	Recall @20	Recall @30	Recall @40	Recall @50	Recall @100	Recall @150	Recall @200	Recall @250	Recall @300
NMF	0.0051	0.0051	0.0051	0.0103	0.0128	0.0231	0.0282	0.0282	0.0308	0.0333
ConvCN	0.1333	0.2133	0.2533	0.2567	0.2767	0.3300	0.3467	0.3467	0.3533	0.3633
LSMFPRec	0.1846	0.2821	0.3359	0.3667	0.3974	0.4436	0.4667	0.4846	0.4897	0.4923
PR-SLSMF	0.2103	0.3359	0.4103	0.4436	0.4692	0.5256	0.5436	0.5538	0.5615	0.5641
Method	**NDCG @10**	**NDCG @20**	**NDCG @30**	**NDCG @40**	**NDCG @50**	**NDCG @100**	**NDCG @150**	**NDCG @200**	**NDCG @250**	**NDCG @300**
NMF	0.0170	0.0170	0.0170	0.0362	0.0435	0.0671	0.0771	0.0771	0.0818	0.0864
ConvCN	0.2260	0.3132	0.3538	0.3571	0.3768	0.4277	0.4432	0.4432	0.4493	0.4586
LSMFPRec	0.2788	0.3780	0.4295	0.4582	0.4863	0.5278	0.5483	0.5640	0.5685	0.5707
PR-SLSMF	0.3058	0.4295	0.4980	0.5278	0.5505	0.5996	0.6150	0.6238	0.6303	0.6325

Table 3. Performance evaluation between different models regarding Recall and NDCG on DBLP.

Method	Recall @10	Recall @20	Recall @30	Recall @40	Recall @50	Recall @100	Recall @150	Recall @200	Recall @250	Recall @300
NMF	0.0026	0.0053	0.0079	0.0079	0.0106	0.0185	0.0237	0.0343	0.0369	0.0396
ConvCN	0.1768	0.2216	0.2401	0.2454	0.2480	0.2480	0.2507	0.2586	0.2612	0.2639
LSMFPRec	0.1821	0.2111	0.2348	0.2427	0.2480	0.2559	0.2639	0.2665	0.2691	0.2718
PR-SLSMF	0.2216	0.285	0.3193	0.3404	0.3456	0.3826	0.4037	0.4248	0.438	0.4406
Method	**NDCG @10**	**NDCG @20**	**NDCG @30**	**NDCG @40**	**NDCG @50**	**NDCG @100**	**NDCG @150**	**NDCG @200**	**NDCG @250**	**NDCG @300**
NMF	0	0.0173	0.0283	0.0283	0.0370	0.0573	0.0686	0.0883	0.0929	0.0973
ConvCN	0.2708	0.3181	0.3368	0.3421	0.3448	0.3448	0.3474	0.3553	0.3579	0.3606
LSMFPRec	0.2765	0.3072	0.3315	0.3395	0.3448	0.3527	0.3606	0.3632	0.3658	0.3684
PR-SLSMF	0.3181	0.3812	0.4142	0.4341	0.4390	0.4732	0.4924	0.5114	0.5232	0.5255

Although using the SPECTER model to learn the content-based representation of the paper, we do not discuss the impact of SPECTER parameters on the model. SPECTER uses 768 dimensional node embedding. The loss amplitude parameter is w=1. For training, SPECTER uses negative samples containing two hard negative and three easily negative samples. The batch size is 16, and the learning rate is 0.001. In order to embed the content of the study paper, we used a research paper abstract with a size of 517.

To determine the optimal number of similarity links between papers, for AAN, we uniformly set the other parameters to $\alpha = 0.3$, $r = 0.3 \cdot rank(X)$, and $k_1 = k_2 = 0.3 \cdot size(X)$. It can be seen from **Fig. 3** that the blue line performs best. We believe that a link count of 10 is optimal. The fuchsia and orange lines perform poorly due to insufficient side information. In contrast, the purple and lime lines do not perform well, possibly due to the introduction of noise when adding information. For DBLP, we uniformly set the other parameters to $\alpha = 0.5$, $r = 0.5 \cdot rank(X)$, and $k_1 = k_2 = 0.5 \cdot size(X)$. It can be seen from **Fig. 3** that the blue and purple lines perform better than the other lines. As k increases, the blue line stabilizes while the purple line still shows an upward trend. Therefore, we believe that a link count of 30 is optimal. The fuchsia and orange lines have poor performance due to insufficient side information. The lime line performs poorly because the side information contains noise.

To determine the approximate range of the parameters α, r, k_1 and k_2, for AAN, we uniformly set these four parameters to be equal, and the value range is from 0.1 to 0.5. It can be seen from **Fig. 4** when k<100, the recall of the lime, orange, and purple lines is greater, and when k>100, the recall of the lime, blue, and purple lines is greater. In summary, the parameters represented by the purple and lime lines make the model perform better than the other lines. Therefore, we believe that the parameters can be 0.3 or 0.4. For DBLP, we uniformly set these four parameters to be equal, and the value range is from 0.4 to 0.8. It can be seen from **Fig. 4** that the lime line performs best. We believe that the parameters can be 0.6.
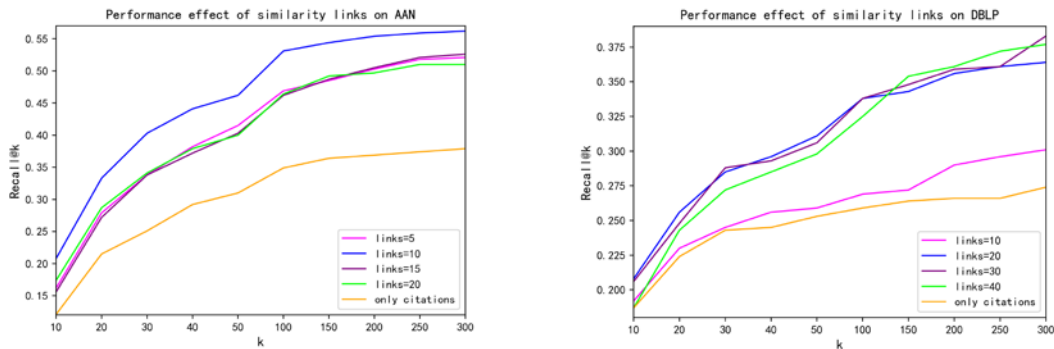

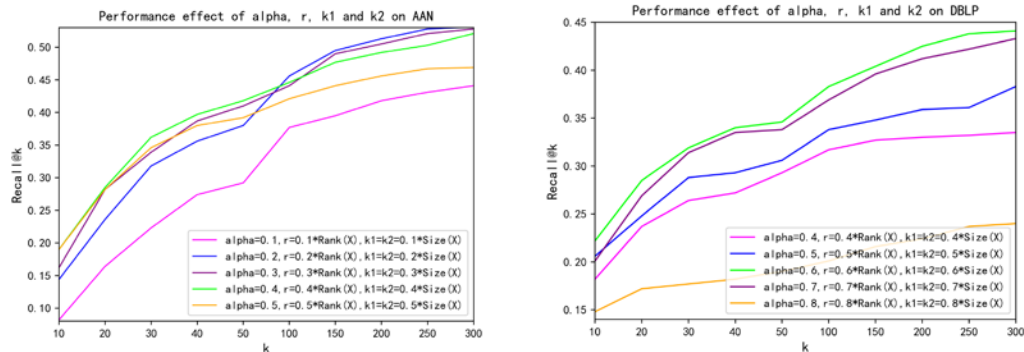**Fig. 3.** Performance effect of similarity link on datasets.

Fig. 4. Performance effect of α, r, k_1 and k_2 on datasets.

To determine the optimal value of α of the model, for AAN, we uniformly set the other parameters to $links = 10$, $r = 0.3 \cdot rank(X)$ and $k_1 = k_2 = 0.3 \cdot size(X)$. It can be seen from **Fig. 5** that the lime and blue lines almost overlap, and the lime line slightly higher than the blue line. Therefore, we believe that $\alpha = 0.4$ is appropriate. For DBLP, we uniformly set the other parameters to $links = 30$, $r = 0.6 \cdot rank(X)$, $k_1 = k_2 = 0.6 \cdot size(X)$. It can be seen from **Fig. 5** that the blue line performs best and is significantly better than the others, we believe that $\alpha = 0.6$ is appropriate.

To determine the optimal value of r of the model, for AAN, we uniformly set the other parameters to $links = 10$, $\alpha = 0.4$, and $k_1 = k_2 = 0.3 \cdot size(X)$. It can be seen from **Fig. 6** that the purple line is much higher than the other lines. We believe that $r = 0.3 \cdot rank(X)$ is appropriate. For DBLP, we uniformly set the other parameters to $links = 30$, $\alpha = 0.6$, and $k_1 = k_2 = 0.6 \cdot size(X)$. It can be seen from **Fig. 6** that the blue line performs best. We believe that $r = 0.6 \cdot rank(X)$ is appropriate.

To determine the optimal value of k_1 and k_2 of the model, for AAN, we uniformly set the other parameters to $links = 10$, $\alpha = 0.4$, and $k_1 = k_2 = 0.3 \cdot size(X)$. It can be seen from **Fig. 7** that the purple line is much higher than the other lines. We believe that $k_1 = k_2 = 0.3 \cdot size(X)$ is appropriate.
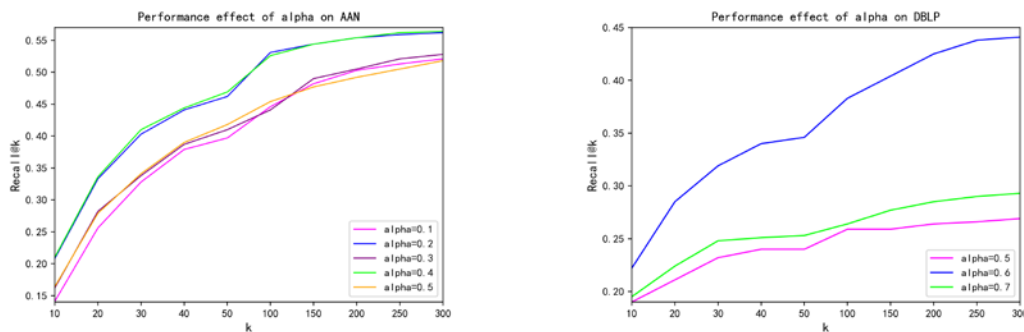

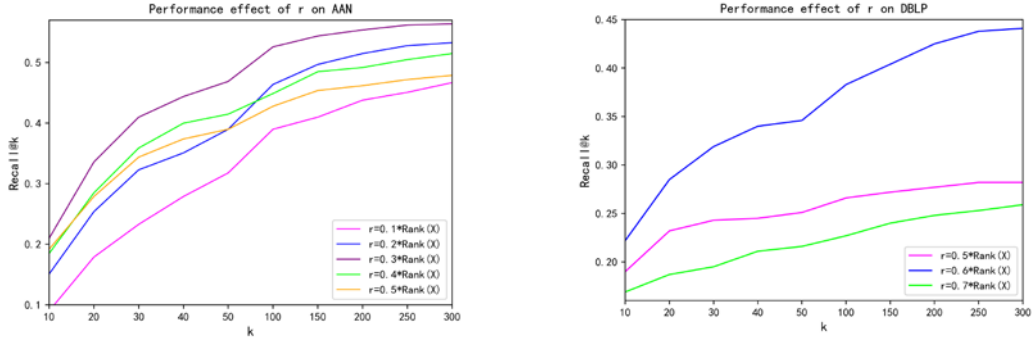**Fig. 5.** Performance effect of different values of α on datasets.

Fig. 6. Performance effect of different values of r on datasets.

To determine the optimal value of k_1 and k_2 of the model, for AAN, we uniformly set the other parameters to $links = 10$, $\alpha = 0.4$, and $k_1 = k_2 = 0.3 \cdot size(X)$. It can be seen from **Fig. 7** that the purple line is much higher than the other lines. We believe that $k_1 = k_2 = 0.3 \cdot size(X)$ is appropriate. We uniformly set the other parameters to $links = 10$, $\alpha = 0.4$, $r = 0.3 \cdot rank(X)$, and $k_2 = 0.3 \cdot size(X)$ to obtain the optimal value of k_1. It can be seen from **Fig. 8** that the purple lin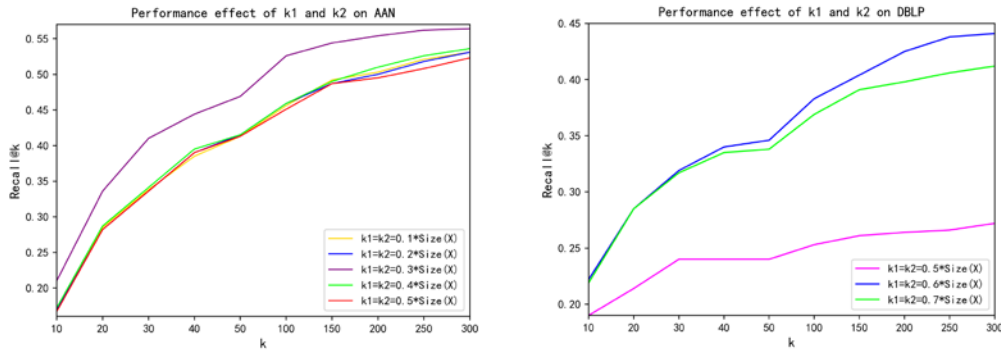e is clearly much higher than the other lines. let $\alpha = 0.4$, $r = 0.3 \cdot rank(X)$, $k_1 = 0.3 \cdot size(X)$ to obtain the optimal value of k_2. It can be seen from **Fig. 9** that the purple line is clearly much higher than the other lines.

Fig. 7. Performance effect of different values of k_1 and k_2 on datasets.
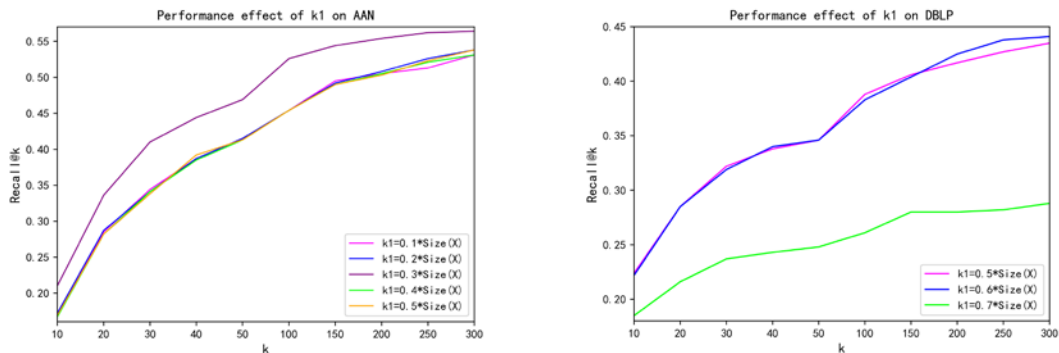

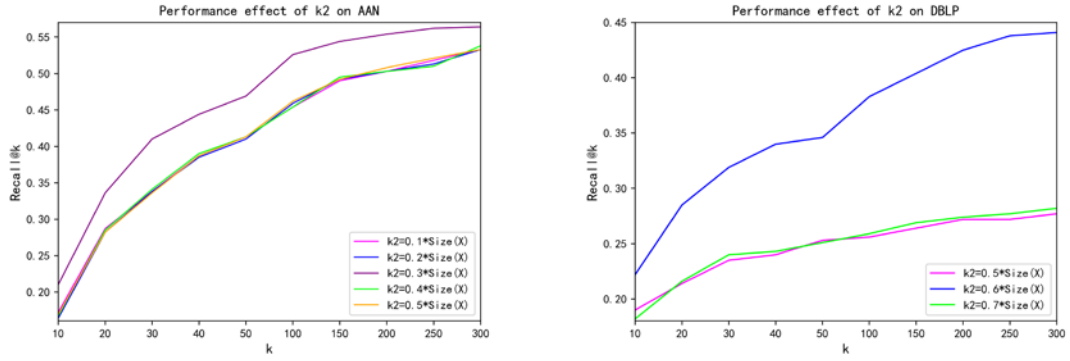**Fig. 8.** Performance effect of different values of k_1 on datasets.

Fig. 9. Performance effect of different values of k_2 on datasets.

For DBLP, we uniformly set the other parameters to $links = 30$, $\alpha = 0.6$, and $k_1 = k_2 = 0.6 \cdot size(X)$. It can be seen from **Fig. 7** that the blue line performs best. We believe that $k_1 = k_2 = 0.6 \cdot size(X)$ is appropriate. We uniformly set the other parameters to $links = 30$, $\alpha = 0.6$, $r = 0.6 \cdot rank(X)$, and $k_2 = 0.6 \cdot size(X)$ to obtain the optimal value of k_1.

All in all, we have obtained the optimal parameters: $links = 10$, $\alpha = 0.4$, $r = 0.3 \cdot rank(X)$, and $k_1 = k_2 = 0.3 \cdot size(X)$ are appropriate for AAN, $links = 30$, $\alpha = 0.6$, $r = 0.6 \cdot rank(X)$, and $k_1 = k_2 = 0.3 \cdot size(X)$ are appropriate for DBLP.
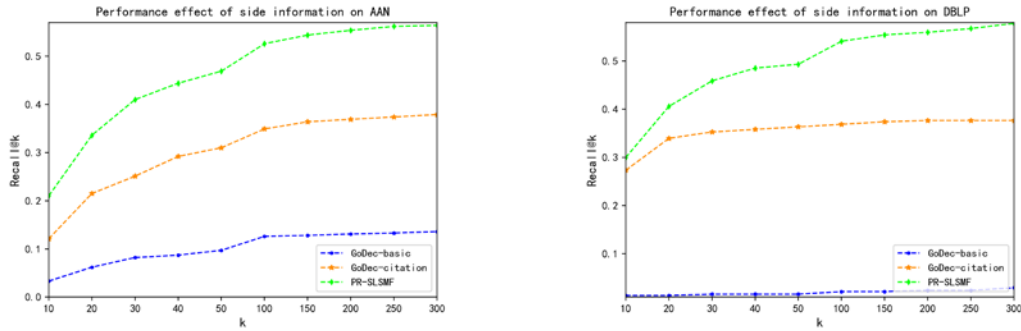

Fig. 10. Performance effect of side information on datasets.

6.5 Analysis of recommendation results

First, we designed experiments to verify that PR-SLSMF alleviates data sparsity and prevents topic drift. We compared PR-SLSMF with basic GoDec and citation information extended GoDec. In **Fig. 10**, GoDec-basic uses only the original interaction matrix. At the same time, GoDec citation represents the extension of GoDec through reference relationships in the offset term, with an objective function equal to (7). However, in the GoDec citation, the composite relationship matrix in (7) must be replaced by a reference relationship matrix. PR-SLSMF performed the best on both datasets, followed by GoDec citation, and GoDec-basic performed the worst. We conclude that side information can improve the model's performance on both datasets. More specifically, the performance of PR-SLSMF and GoDec citation was significantly better than that of GoDec-basic, indicating that adding side information effectively alleviated data sparsity problems. The performance of PR-SLSMF was significantly better than that of GoDec citation, indicating that adding only citation

information was insufficient. Adding appropriate content information limited the topic to a certain range to avoid topic drift and improved the model's performance.

Second, we verified that PR-SLSMF alleviated the cold-start problem. We compared the top 10 recommended results between PR-SLSMF and GoDec citations. Here we give only the documents recommended for one author, and there is more evidence to support this conclusion. There is no author named "Esther Duflo" in the training set, so it is a cold-start problem to recommend papers for him. The corresponding "Esther Duflo" paper in the test set is "Gossip: Identifying Central Individuals in a Social Network." In Table 4, PR-SLSMF provides the correct paper in the results, while the GoDec citation does not. At this point, we have verified that PR-SLSMF alleviated the cold-start problem.

Table 4. Recommendation results for "Esther Duflo" whose author id is '30'

method	Paper id	Paper title
PR-SLSMF	14225	Friendships in Online Peer-to-Peer Lending: Pipes, Prisms, and Relational Herding
	3423	Containment control of networked autonomous underwater vehicles with model uncertainty and ocean disturbances guided by multiple leaders
	14431	Convergence of univariate non-stationary subdivision schemes via asymptotic similarity
	14063	ISEE Smart Home (ISH): Smart video analysis for home security
	14012	Detection of review spam: A survey
	13934	Satisfiability Modulo Transcendental Functions via Incremental Linearization
	3015	Cooperative output feedback adaptive control of uncertain nonlinear multi-agent systems with a dynamic leader
	14104	Improved Results on Passivity Analysis of Uncertain Neural Networks with Time-Varying Discrete and Distributed Delays
	1429	Long-Term Correlation Tracking
	14318	Gossip : Identifying Central Individuals in a Social Network
GoDec-citation	14098	DeepOrgan: Multi-level Deep Convolutional Networks for Automated Pancreas Segmentation
	14076	Nested Weighted Automata
	14361	Localizability and distributed protocols for bearing-based network localization in arbitrary dimensions
	10432	DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA: mRNA interactions
	10540	Spectral caustic rendering of a homogeneous caustic object based on wavelength clustering and eye sensitivity
	1911	Incorporating adaptability-related knowledge into support vector machine for case-based design adaptation
	14110	Oblivious Network RAM
	14026	Push Attack: Binding Virtual and Real Identities Using Mobile Push Notifications
	14440	Topology Identification of Directed Dynamical Networks via Power Spectral Analysis
	14053	Matched Gk-constructions always yield Ck-continuous isogeometric elements

7. Conclusion

In order to find the relevant research papers that researchers need, the paper recommendation model will introduce various auxiliary information, especially in cold technology papers and sparse scenarios. Most existing models overlook the negative impact of noise contained in auxiliary information on recommendations. Therefore, the recommendations provided by these models cannot be guaranteed to be what researchers truly need. To address these issues, we propose a matrix decomposition model called PR-SLSMF. PR-SLSMF uses the title and abstract that best represent the content information of a paper as auxiliary information. Then it uses the SPECTER to learn the content representation of the paper. Finally, the content information and citation information are integrated into Godec decomposition to generate recommendations. PR-SLSMF improves recommendation performance through three aspects: effectively alleviating data sparsity and cold start problems, effectively preventing topic drift, and reducing error propagation during matrix decomposition. The experimental results indicate that PR-SLSMF is superior to the baselines.

In the future, we will consider using some techniques or methods to remove or reduce noise contained in auxiliary information to improve the performance of recommendation models. At the same time, we will evaluate our model on other datasets to demonstrate its universal applicability and analyze its significance.

Acknowledgement

References

[1]   C. Bhagavatula, S. Feldman, R. Power, and W. Ammar, "Content-Based Citation Recommendation," in *Proc. of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol.1, pp.238-251, 2018. Article (CrossRef Link)

[2]   W. Zhao, Z. Yu, and R. Wu, "A citation recommendation method based on context correlation, " *Intelligent data analysis*, vol. 25, no. 1, pp. 225-243, 2021. Article (CrossRef Link)

[3]   S. Ma, C. Zhang, and X. Liu, "A review of citation recommendation: from textual content to enriched context," *Scientometrics*, vol. 122, pp. 1445-1472, 2020. Article (CrossRef Link)

[4]   H. Wang, and W. Li, "Relational collaborative topic regression for recommender systems," *IEEE transactions on knowledge and data engineering*, vol. 27, no. 5, pp. 1343-1355, 2015. Article (CrossRef Link)

[5]   K. Sugiyama, and M. Y. Kan, "Exploiting potential citation papers in scholarly paper recommendation," in *Proc. of the 13th ACM/IEEE-CS joint conference on Digital libraries*, pp. 153-162, Jul. 2013. Article (CrossRef Link)

[6]   X. Cai, J. Han, and L. Yang, "Generative adversarial network based heterogeneous bibliographic network representation for personalized citation recommendation," in *Proc. of the AAAI conference on artificial intelligence*, Vol. 32, No. 1, pp. 5747-5754, Apr. 2018. Article (CrossRef Link)

[7]   C. Pornprasit, X. Liu, P. Kiattipadungkul, N. Kertkeidkachorn, K. Kim et al., "Enhancing citation recommendation using citation network embedding," *Scientometrics*, vol. 127, pp. 233-264, 2022. Article (CrossRef Link)

[8]   Z. Ali, G. Qi, K. Muhammad, P. Kefalas, and S. Khusro, "Global citation recommendation employing generative adversarial network," *Expert Systems with Applications*, vol.180, 2021. Article (CrossRef Link)

[9]   J. Son, and S. B. Kim, "Academic paper recommender system using multilevel simultaneous citation networks," *Decision Support Systems*, vol.105, pp.24-33, 2018. Article (CrossRef Link)

[10] P. Kefalas, P. Symeonidis, and Y. Manolopoulos, "Recommendations based on a heterogeneous spatio-temporal social network," *World Wide Web*, Vol. 21, no. 2, pp.345-371, 2018. Article (CrossRef Link)

[11] R. Torres, S. M. McNee, M. Abel, J. A. Konstan, and J. Riedl, "Enhancing digital libraries with Techlens+," in *Proc. of the 4th ACM/IEEE-CS joint conference on digital libraries*, pp.228-236, Jun. 2004. Article (CrossRef Link)

[12] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," *Computer*, vol.42, no.8, pp.30-37, 2009. Article (CrossRef Link)

[13] C. Yang, B. Wei, J. Wu, Y. Zhang, and L. Zhang, "CARES: a ranking-oriented CADAL recommender system," in *Proc. of the 9th ACM/IEEE-CS joint conference on digital libraries*, pp.203-212, Jun. 2009. Article (CrossRef Link)

[14] S. Purushotham, Y. Liu, and C. J. Kuo, "Collaborative topic regression with social matrix factorization for recommendation systems," in *Proc. of the 29th international conference on international conference on machine learning*, pp. 691-698, Jun. 2012. Article (CrossRef Link)

[15] T. Dai, L. Zhu, X. Cai, S. Pan, and S. Yuan, "Explore semantic topics and author communities for citation recommendation in bipartite bibliographic network," *Journal of ambient intelligence and humanized computing*, vol. 9, no. 4, pp. 957-975, 2018. Article (CrossRef Link)

[16] Y. Lee, J. Yeom, K. Song, J. Ha, K. Lee et al., "Recommendation of research papers in DBpia: A Hybrid approach exploiting content and collaborative data," in *Proc. of the 2016 IEEE international conference on systems, man, and cybernetics*, pp. 002966-002971, Oct. 2016. Article (CrossRef Link)

[17] C. Wang, and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in *Proc. of the 17th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 448-456, Aug. 2011. Article (CrossRef Link)

[18] S. Huang, G. Xue, B. Zhang, Z. Chen, Y. Yu et al., "Tssp: A reinforcement algorithm to find related papers," in *Proc. of the IEEE/WIC/ACM international conference on web intelligence*, pp. 117-123, Sep. 2004. Article (CrossRef Link)

[19] A. Khadka, and P. Knoth, "Using citation-context to reduce topic drifting on pure citation-based recommendation," in *Proc. of the 12th ACM conference on recommender systems*, pp. 362-366, Sep. 2018. Article (CrossRef Link)

[20] A. Cohan, S. Feldman, I. Beltagy, D. Downey, and D. S. Weld, "Specter: Document-level representation learning using citation-informed transformers," in *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 2270-2282, Jul. 2020. Article (CrossRef Link)

[21] T. Zhou, and D. Tao, "GoDec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. of the 28th international conference on machine learning*, pp. 33-40, Oct. 2011. Article (CrossRef Link)

[22] S. M. McNee, I. Albert, D. Cosley, P. Gopalkrishnan, S. K. Lam et al., "On the recommending of citations for research papers," in *Proc. of the 2002 ACM conference on computer supported cooperative work*, pp. 116-125, Nov. 2002. Article (CrossRef Link)

[23] T. Bansal, D. Belanger, and A. McCallum, "Ask the GRU: Multi-task learning for deep text recommendations," in *Proc. of the 10th ACM conference on recommender systems*, pp. 107-114, Sep. 2016. Article (CrossRef Link)

[24] Z. Ali, I. Ullah, A. Khan, J. A. Ullah, and K. Muhammad, "An overview and evaluation of citation recommendation models," *Scientometrics*, vol. 126, pp. 4083-4119, 2021. Article (CrossRef Link)

[25] C. L. Giles, K. D. Bollacker, and S. Lawrence, "CiteSeer: An automatic citation indexing system," in *Proc. of the third ACM conference on digital libraries*, pp. 89-98, May. 1998. Article (CrossRef Link)

[26] M. Amami, G. Pasi, F. Stella, and R. Faiz, "An LDA-based approach to scientific paper recommendation," in *Proc. of the natural language processing and information systems*, pp. 200-210, Jun. 2016. Article (CrossRef Link)

[27] K. Sugiyama, and M. Y. Kan, "Scholarly paper recommendation via user's recent research interests," in *Proc. of the 10th annual joint conference on digital libraries*, pp. 29-38, Jun. 2010. Article (CrossRef Link)

[28] A. Khadka, G. I. Cantador, and M. Fernandez, "Exploiting citation knowledge in personalised recommendation of recent scientific publications," in *Proc. of the conference on language resources and evaluation*, pp. 2231-2240, May. 2020. Article (CrossRef Link)

[29] S. Khusro, Z. Ali, and I. Ullah, "Recommender systems: issues, challenges, and research opportunities," in *Proc. of the information science and applications 2016*, pp. 1179-1189, Feb. 2016. Article (CrossRef Link)

[30] Z. Ali, P. Kefalas, K. Muhammad, B. Ali, and M. Imran, "Deep learning in citation recommendation models survey," *Expert systems with applications*, vol. 162, pp. 113790. 2020. Article (CrossRef Link)

[31] W. Huang, Z. Wu, C. Liang, P. Mitra, and C. Giles, "A neural probabilistic model for context-based citation recommendation," in *Proc. of the AAAI conference on artificial intelligence*, pp. 2404-2410, Feb. 2015. Article (CrossRef Link)

[32] Z. Luo, Q. Xie, and S. Ananiadou, "CitationSum: Citation-aware Graph Contrastive Learning for Scientific Paper Summarization," in *Proc. of the ACM web conference 2023*, pp. 1843-1852, Apr. 2023. Article (CrossRef Link)

[33] C. Jeong, S. Jang, E. Park, and S. Choi, "A context-aware citation recommendation model with BERT and graph convolutional networks," *Scientometrics*, vol. 124, no. 3, pp. 1907-1922, 2020. Article (CrossRef Link)

[34] T. Dai, W. Yan, K. Zhang, C. Qiu, X. Zhao et al., "Gated relational stacked denoising autoencoder with localized author embedding for global citation recommendation," *Expert Systems with applications*, vol. 184, no. C, pp. 115359, Dec. 2021. Article (CrossRef Link)

[35] T. Dai, L, Zhu, Y. Wang, and K. M. Carley, "Attentive stacked denoising autoencoder with bi-LSTM for personalized context-aware citation recommendation," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 28, pp. 553-568, 2019. Article (CrossRef Link)

[36] Z. Ali, G. Qi, P. Kefalas, S. K. I. Khusro, and K. Muhammad, "SPR-SMN: Scientific paper recommendation employing SPECTER with memory network," *Scientometrics*, vol. 127, no. 11, pp. 6763-6785, 2022. Article (CrossRef Link)

[37] J. Wang, J. Zhou, Z. Wu, and X. Sun, "Toward paper recommendation by jointly exploiting diversity and dynamics in heterogeneous information networks," in *Proc. of the database systems for advanced applications: 27th international conference*, pp. 272-280, Apr. 2022. Article (CrossRef Link)

[38] L. Pan, X. Dai, S. Huang, and J. Chen, "Academic Paper Recommendation Based on Heterogeneous Graph," in *Proc. of the Chinese computational linguistics and natural language processing based on naturally annotated big data*, pp. 381-392, Nov. 2015. Article (CrossRef Link)

[39] S. Gupta, and V. Varma, "Scientific article recommendation by using distributed representations of text and graph," in *Proc. of the 26th international conference on world wide web companion*, pp. 1267-1268, Apr. 2017. Article (CrossRef Link)

[40] Q. Le, and T. Mikolov, "Distributed representations of sentences and documents," in *Proc. of the 31st International conference on machine learning*, pp. 1188-1196, Jun. 2014. Article (CrossRef Link)

[41] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proc. of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 701-710, Aug. 2014. Article (CrossRef Link)

[42] X. Kong, M. Mao, W. Wang, J. Liu, and B. Xu, "VOPRec: Vector representation learning of papers with text information and structural identity for recommendation," *IEEE transactions on emerging topics in computing*, vol. 9, no. 9, pp. 226-237, 2021. Article (CrossRef Link)

[43] S. Ganguly, and V. Pudi, "Paper2vec: Combining graph and text information for scientific paper
       representation," in *Proc. of the advances in information retrieval*, pp. 383-395, Apr. 2017.
       Article (CrossRef Link)

[44] L. F. Ribeiro, P. H. Saverese, and D. R. Figueiredo, "Struc2vec: learning node representations
       from structural identity," in *Proc. of the 23rd ACM SIGKDD international conference on
       knowledge discovery and data mining*, pp. 385-394, Apr. 2017. Article (CrossRef Link)

[45] J. Chen, Y. Liu, S. Zhao, and Y. Zhang, "Citation recommendation based on weighted
       heterogeneous information network containing semantic linking," in *Proc. of the 2019 IEEE
       international conference on multimedia and expo*, pp. 31-36, Jul. 2019. Article (CrossRef Link)

[46] A. Grover, and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proc. of the
       22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 855-
       864, Aug. 2016. Article (CrossRef Link)

[47] X. Ning, and G. Karypis, "Slim: Sparse linear methods for top-n recommender systems," in *Proc.
       of the IEEE 11th international conference on data mining*, pp. 497-506, Dec. 2011.
       Article (CrossRef Link)

[48] J. Wang, L. Zhu, T. Dai, Q. Xu, and T. Gao, "Low-rank and sparse matrix factorization with prior
       relations for recommender systems," *Applied intelligence*, vol. 51, no. 6, pp. 3435-3449, 2021.
       Article (CrossRef Link)

[49] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based
       Mahalanobis distance method for hyperspectral anomaly detection," *IEEE transactions on
       geoscience and remote sensing*, vol. 54, no. 3, pp. 1376-1389, 2015. Article (CrossRef Link)

[50] Q. Wang, K. Paynabar, and M. Pacella, "Online automatic anomaly detection for photovoltaic
       systems using thermography imaging and low rank matrix decomposition," *Journal of quality
       technology*, vol. 54, no. 5, pp. 503-516, 2021. Article (CrossRef Link)

[51] S. Feng, S. Tang, C. Zhao, and Y. Cui, "A hyperspectral anomaly detection method based on low-
       rank and sparse decomposition with density peak guided collaborative representation," *IEEE
       transactions on geoscience and remote sensing*, vol. 60, pp. 1-13, 2021. Article (CrossRef Link)

[52] X. Li, G. Cui, and Y. Dong, "Graph regularized non-negative low-rank matrix factorization for
       image clustering," *IEEE transactions on cybernetics*, vol. 47, no. 11, pp. 3840-3853, 2017.
       Article (CrossRef Link)

[53] M. Song, T. Yang, H. Cao, F. Li, B. Xue et al., "Bi-Endmember Semi-NMF Based on Low-Rank
       and Sparse Matrix Decomposition," *IEEE transactions on geoscience and remote sensing*, vol. 60,
       pp. 1-16, 2022. Article (CrossRef Link)

[54] Y. Mu, J. Dong, X. Yuan, and S. Yan, "Accelerated low-rank visual recovery by random
       projection," in *Proc. of the IEEE conference on computer vision and pattern recognition*, pp. 2609-
       2616, Jun. 2011. Article (CrossRef Link)

[55] Z. Zhang, A. Ganesh, X. Liang, and Y. Ma, "TILT: Transform invariant low-rank textures,"
       International journal of computer vision, vol. 99, no. 1, pp. 1-24, 2012. Article (CrossRef Link)

[56] R. S. Rezende, J. Zepeda, J. Ponce, F. Bach, and P. Perez, "Kernel square-loss exemplar macHINes
       for image retrieval," in *Proc. of the IEEE conference on computer vision and pattern recognition*,
       pp. 2396-2404, Jul. 2017. Article (CrossRef Link)

[57] Z. Zhao, L. Huang, C. Wang, J. Lai, and P. S. Yu, "Low-rank and sparse matrix completion for
       recommendation," in *Proc. of the international conference on neural information processing*, pp.
       3-13, Oct. 2017. Article (CrossRef Link)

[58] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the
       ACM*, vol. 58, no. 3, pp. 1-37, 2011. Article (CrossRef Link)

[59] T. Dai, T. Gao, L. Zhu, X. Cai, and S. Pan, "Low-rank and sparse matrix factorization for scientific
       paper recommendation in heterogeneous network," *IEEE Access*, vol. 6, pp. 59015-59030, 2018.
       Article (CrossRef Link)

[60] I. Beltagy, K. Lo, and A. Cohan, "SciBERT: A pretrained language model for scientific text," in
       *Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th
       International Joint Conference on Natural Language Processing*, pp. 3615–3620, Sep. 2019.
       Article (CrossRef Link)

[61] T. Zhou, and D. Tao, "Bilateral random projections," in *Proc. of the 2012 IEEE international symposium on information theory proceedings*, pp. 1286-1290, Jul. 2012. Article (CrossRef Link)

[62] D. Lee, and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. of the 13th international conference on neural information processing systems*, pp. 556-562, Jan. 2000. Article (CrossRef Link)

Panpan Guo received the B.S. degree from Zhengzhou institute of light industry, in 2016, and the M.S. degree from Zhengzhou University, in 2019. She is currently pursuing the Ph.D. degree with the College of Data and Target Engineering, Information Engineering University. Her research interests focus on recommendation systems.

Gang Zhou received his B.E. and M.A. Eng degrees from Information Engineering University in 1996 and 1999, Ph.D. degree from Beihang University in 2007. He was with the State key of Laboratory of Mathematical Engineering And Advanced Computing, Information Engineering University, as a research fellow. His research interests are big data , knowledge graph, and data mining.

Jicang Lu received the Ph.D. degree in Information Engineering University, Zhengzhou, China, in 2014. He has been an Assistant Professor with Information Engineering University, Zhengzhou, China. His research interests include data intelligence, social media mining and related applications.

Zhufeng Li received the Ph.D. degree of literature from PLAUFL，in 2017. His research interests include data mining, big data analysing and natural language processing.

Taojie Zhu received the B.E. degree in engineering from Henan University of Engineering, Zhengzhou, China, in 2020. He received the M.S. degree in University of Information Engineering Zhengzhou, China, in 2023. He is currently an assistant teacher of the State Key Laboratory of Mathematical Engineering and Advanced Computing, His main research interests include knowledge graph and data mining.