BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 06/15 pp357–370
DOI: 10.26599/BDMA.2023.9020029
Volume 7, Number 2, June 2024

# A Novel Recommendation Algorithm Integrates Resource Allocation and Resource Transfer in Weighted Bipartite Network

Qiang Sun, Leilei Shi, Lu Liu, Zixuan Han, Liang Jiang, Yan Wu, and Yeling Zhao*

**Abstract:** Grid-based recommendation algorithms view users and items as abstract nodes, and the information utilised by the algorithm is hidden in the selection relationships between users and items. Although these relationships can be easily handled, much useful information is overlooked, resulting in a less accurate recommendation algorithm. The aim of this paper is to propose improvements on the standard substance diffusion algorithm, taking into account the influence of the user's rating on the recommended item, adding a moderating factor, and optimising the initial resource allocation vector and resource transfer matrix in the recommendation algorithm. An average ranking score evaluation index is introduced to quantify user satisfaction with the recommendation results. Experiments are conducted on the MovieLens training dataset, and the experimental results show that the proposed algorithm outperforms classical collaborative filtering systems and network structure based recommendation systems in terms of recommendation accuracy and hit rate.

**Key words:** cloud computing; bipartite graph network; recommendation algorithm; link prediction; cold start problem

## 1 Introduction

With the fast advancement of Internet technology, the total number of web servers throughout the world, as well as the quantity of online pages, continues to rise[1–3]. Meanwhile, the Internet is becoming increasingly cluttered with dark materials, and finding relevant information has become a major challenge for Internet technology[4–8]. That is why Google, Baidu, Microsoft, and Bing, among other specialized information search engines, have a vast user base. However, search engines need users to be able to identify their requirements and submit keywords, thus the feedback is confined to the user's known range of information, and cannot help users find the content that they are unfamiliar with but important or interesting. How to reduce the cost of information search for users and enable different users to obtain information resources that suit their needs quickly and accurately has become an inescapable reality. The recommendation system under cloud computing

• Qiang Sun, Leilei Shi, Zixuan Han, and Yan Wu are with the School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China. E-mail: 2222008053@stmail.ujs.edu.cn; 937109472@qq.com; 823460370@qq.com; 316119036@qq.com.
• Lu Liu is with the School of Computing and Mathematical Sciences, University of Leicester, Leicester, LE1 7RH, UK. E-mail: l.liu@leicester.ac.uk.
• Liang Jiang is with the School of Ocean College, Jiangsu University of Science and Technology, Zhenjiang 212003, China.
• Yeling Zhao is with the School of Finance and Economics, Jiangsu University, Zhenjiang 212013, China. E-mail: zhaoyeling2006@ujs.edu.cn.
∗ To whom correspondence should be addressed.
Manuscript received: 2023-02-15; revised: 2023-08-09; accepted: 2023-10-10

scheduling, as an effective information processing and personalised decision-making tool, has received widespread attention for filtering relevant information resources for users on the Internet platform.

To solve the problem of users getting the information they need quickly, the recommendation system is created, which can somewhat compensate for the deficiencies of the search engine[3]. Unlike traditional search engines, recommendation systems analyze users' historical behavior, create a model of user preferences, use algorithmic calculations to predict the weight of user preferences for unknown products, sort them by weight, and present users with a list of product recommendations that may be of interest to them[9]. This may seem to use the same recommendation algorithm as the bipartite "User-Commodity" e-commerce network. In fact, these are two related but not identical problems. Recommendation systems require that recommendations are given to every user, whereas link prediction in bipartite networks generally does not require this. Even if all the links predicted by the bipartite network are connected to one user, it does not matter. Because the bipartite network does not need to make recommendations for each user. In this sense, solving the algorithm design problem of the recommendation system does not solve the link prediction problem of the bipartite network. This is because there is no way to determine who is more likely to exist for the links recommended to different users. On the contrary, getting the estimated value of the probability of the existence of all connected edges solves, in principle, the problem of the recommendation algorithm, which can select a number of connections with the highest probability of existence with their neighbors for each user, and recommend them to each of these users. In theory, the estimated value of the likelihood of edge presence also solves the problem of recommendation algorithms, because multiple connections near each user with the highest likelihood of existence may be chosen and presented to each user individually[10, 11].

The recommendation algorithm is the heart and soul of the recommendation system, and a high-quality recommendation algorithm will have an impact on the overall performance and quality of the recommendations. Currently, the most common recommendation algorithms are content-based recommendation algorithms[12], collaborative filtering recommendation algorithms[13], and bipartite graph based recommendation algorithms[14], etc. A recommendation system provides a user with a product that the user could be interested in while also determining a user-commodity connection based on the user's historical behavior. This relationship can be expressed very clearly using a recommendation system based on a bipartite graph network topology.

Despite the fact that recommender systems have such powerful features that they have piqued the interest of many academics, there are still some issues in their development.

(1) **Issue of credibility:** In today's world of severe economic rivalry, e-commerce enterprises compete all the time. Recommendations from recommendation systems are based on the consumer's past consumption rating of the product. Fake score information is introduced to the scoring system by some online platforms to boost their interests, lowering the credibility of the recommendation system, lowering the accuracy and quality of the suggestion, and undermining the impartiality of the recommendation.

(2) **Rating weighting problem:** The edges linking users and goods are not enabled in the typical bipartite graph recommendation algorithm. Users' assessments of items, on the other hand, are discovered to represent users' subtle psychological reactions in the study, and these subtle manifestations will impact recommendations. However, the weighting of ratings in the present weighted bipartite graph algorithm[15] is not detailed enough to effectively assess user ratings, thus resulting in a decline in the accuracy of the recommendation system and affecting the suggestion quality.

(3) **Problem of sparseness:** The supporting data for the recommendation system come from user ratings, but nowadays there are more and more e-commerce online platforms, and more and more possibilities for users to choose different platforms for shopping. This has led to most users being less motivated to rate products, and the recommendation system is able to obtain very little valid rating information, and the similarity and predicted ratings calculated can be inaccurate, which has a significant impact on the accuracy of the overall recommendation system.

(4) **Complexity problem:** When recommending the target user in a traditional bipartite graph recommendation algorithm, all of the user information associated with it is used, and the trusted nearest neighbours are not screened, which not only increases

the algorithm's complexity but also reduces the accuracy of the recommendation.

Bipartite organisations based on mass diffusion and heat conduction[16], two common physical dynamics, have been presented to improve suggestion execution and have made tremendous progress. In these network-based recommendation algorithms, the initial resource configuration and resource allocation procedure between users and objects are two critical components.

The main contributions of this paper are as follows:

(1) The initial resource allocation vector and resource transfer matrix in the recommendation algorithm are improved based on the standard material diffusion algorithm, and the influence of user ratings on recommended products is taken into account, and an adjustment factor is added to solve the shortcomings of the standard material diffusion algorithm which treats all products equally.

(2) The average ranking score evaluation metric is introduced to quantify the user's satisfaction with the recommendation result and compare the accuracy of the recommendation algorithm more intuitively.

(3) The experimental results show that the proposed algorithm outperforms classical collaborative filtering systems and network structure based recommendation systems in terms of recommendation accuracy and hit rate.

The following is the next arrangement of the paper: Section 2 covers and assesses the state of research on the bipartite graph recommendation algorithm in the United States and internationally. The bipartite network's essential principles and structural aspects are introduced in Section 3. Section 4 delves into the fundamentals of several widely used bipartite graph recommendation algorithms. Section 5 presents ideas for improving the conventional drug diffusion algorithm, as well as a special enhanced method that takes into account the user's assessment of the product. Section 6 evaluates the experimental results. Section 7 summarizes the full article.

## 2 Related Work

With the development of computer network technology and social network, recommendation system has attracted more and more attention and praise[17–19]. At the ACM Special Interest Group on Knowledge Discovery and Data Mining (ACM SIGKDD) International Conference, a graph-based collaborative filtering recommendation algorithm was first suggested, which is more optimized in terms of suggestion accuracy than typical collaborative filtering recommendation algorithms[20].

The bipartite network structure is used in Ref. [21] to define the user-commodity interaction and to investigate the impact of the bipartite graph's integrated nature on the recommendation algorithm. The research of weighting the projection of the bipartite graph was proposed by Zhou et al.[22], and the bipartite graph was turned into a one-part graph for research, reducing the complexity of the method. Later, Zhou et al.[23] introduced a recommendation method based on material diffusion and heat conduction knowledge and a bipartite graph network topology. Zhou et al.[24] discovered through the study on e-commerce platforms that customers frequently purchase hot-selling items, resulting in the "long-tail" problem of recommendations, in which the diversity of suggestions reduces dramatically as popularity rises. As a result of the bipartite graph recommendation algorithm, they can control the initial resources of popular products while leaving more resources for less popular products, allowing them to delve deeper into popular products while reducing the likelihood of popular products being widely selected.

Wang and Jian[25] broke the original bipartite graph network into sub-networks, promoted people in the sub-network, established a threshold of user similarity, and determined the similarity between two users in the system. Users will be eliminated from the sub-network if the threshold is not met, and these two users will not be included in the final suggestion. The testing findings suggest that the method decreases the complexity of the algorithm while saving calculation time. Liu et al.[26] included the user-commodity correlation into the Collaborative Filtering (CF) algorithm and utilized the bipartite network's features to adaptively adjust the parameters to increase the algorithm's accuracy and minimize its complexity.

He et al.[27] advocated adding feedback modification to the recommendation system in light of the project's appealing effect on users, and they were able to boost suggestion novelty in numerous datasets. Zhou et al.[28] used the bipartite network to execute a one-dimensional projection to produce the user-user connection graph, and then used the Jaccard similarity function[29] to generate a recommendation list based on the resources that users transferred to each other.

Rao et al.[30] proposed the Weighted Network-Based

Inference (WNBI) method, which is based on the Network-Based Inference (NBI) algorithm proposed by Zhou et al.[22] The weight is determined by the user's rating of the item, and the resource allocation matrix is calculated based on the weight. This improves recommendation accuracy without adding time or space overhead. According to the Ref. [31], independently related information similarity replaces the prior Pearson similarity[32], which addresses data sparsity and enhances recommendation accuracy. In order to increase accuracy and variety, Beckett[33] proposed a better recommendation algorithm based on a weighted network structure under the requirement of differentiating high and low scores and incorporated the ratio of item degree to the total of item weights.

Li and Shi[34] included the bipartite graph network structure into the collaborative filtering algorithm and used the grey correlation weighting approach when computing the user similarity to anticipate the items that the target user could be interested in Ref. [35]. Experiments contain the findings suggesting that the algorithm enhances recommendation accuracy and dependability. When resources are randomly distributed in a bipartite graph network topology, Codling et al.[36] considered the angular distribution and movement speed and developed the probability density function of the spatial distribution to create efficient suggestions and increase recommendation accuracy.

To increase the recommendation accuracy, the link prediction method is added to the bipartite graph network structure in Ref. [37], and the maximum likelihood is applied while computing the node similarity to choose the nearest neighbour set for the target user. According to Ref. [38], Ratcliffe and Arandjelović found that the user's interests are changing, therefore they suggested an interest drift detection technique and employed bipartite graph projection and random walk methods in the recommendation link in response to this change. The method increases accuracy while lowering complexity.

## 3 Basic Concept and Structural Feature of the Bipartite Network

### 3.1 Basic concept

The bipartite network, also called the bipartite graph, is a network with special composition characteristics[37]. Speaking of an undirected simple network $G(V, E)$ as a bipartite network, at least a pair of node sets $X$ and $Y$ should exist, satisfying

(1) $X \cap Y = \varnothing$;

(2) $X \cup Y = V$;

(3) Any edge in $E$ must have exactly one vertex in the set $X$ and the other vertex in $Y$.

Many common networks are bipartite networks. For example, all trees are bipartite networks, and tetragonal lattices are also bipartite networks. In fact, it can be further proved that for all planar graphs, if every face is an even-sided shape, it is a bipartite network.

Many natural bipartite graphs are also actual networks. For example, the opposite sex sexual relationship network is a bipartite network with two distinct sets of men and women[39], the metabolic network is a bipartite network with two distinct sets of chemical substances and chemical reactions[24], and the cooperative network is a bipartite network with two distinct sets of participants and events[40]. The Internet telephone network is split into two halves, each with its own set of computers and phone numbers[41]. The foundation of an e-commerce network is two distinct groups of consumers and items. The human illness network is a bipartite network of different sets[42], with two independent sets of physical and mental dysfunction symptoms and disease-causing genes[43].

### 3.2 Structural characteristics of bipartite network

Figure 1 is a schematic diagram of a bipartite network. This is an e-commerce network, the set on the top is the user, the set on the bottom is the product, and the link represents the purchase relationship. For example, the user $i$ purchases commodities $\alpha$, $\beta$, and $\gamma$. Like general undirected simple graphs, the degree of a node is the number of its associated links. If this bipartite network is not authorized, then for a user, the degree is the type and number of goods user purchased, and for a product, the degree is how many different users it sells.

When discussing the two part of the network

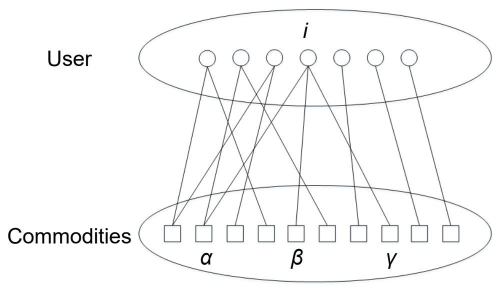



**Fig. 1 A schematic diagram of a bipartite network.**

distribution, the degree distribution of nodes in the set $X$ and the degree distribution of nodes in the set $Y$ are often analyzed separately. In general, these two-degree distributions are not the same. Take the user-commodity second part of the network as an example, Lambiotte and Ausloos[44] analyzed the music library data. It is considered that the distribution of the product is a power law, and the user's degree distribution is index. Shang et al.[42] were more careful and defined a new metric called collaborative similarity. However, the user-degree distribution is more suitable for portraying the latency index[45], and their dataset is a typical user-commodity dichotomous network. Recent empirical studies display that Wikilens[46] users distribution and product distribution are more suitable for portraying the delay index distribution, and the MovieLens user distribution is close to the index, the product distribution is a typical delay index distribution. Not only the form of distribution is not the same, but also the degree distribution itself is not necessarily stable[47].

## 4 Common Recommendation Algorithm

### 4.1 Content-based recommendation algorithm

The collaborative filtering technique is the continuation and development of the original Content-based Recommendation (CR) algorithm. It does not have to be based on the user's appraisal of the item, but rather on the product content information that the user has chosen to determine user similarity and then provide appropriate suggestions.

With advances in machine learning and other technologies, the current content-based recommendation system may create distinct configuration files for users and items[12], as well as create or update user configuration files by evaluating previously bought material (or browsed). The system may compare the user's configuration file to that of the product and propose the product that is most comparable to the user's configuration file. For example, in movie recommendation, the content-based system examines the common qualities of movies that users have viewed (actors, directors, styles, and so on)[48], and then suggests movies that are substantially comparable to the movies that users are interested in. The gathering and filtering of information are at the heart of a content-based recommendation system. Due to the advanced research in the collection and filtering of relevant textual information about films, many existing content-based film recommendation systems offer recommendations based on film titles, categories, etc.

### 4.2 Collaborative filtering algorithm

In collaborative filtering technology, first of all, it is necessary to calculate the similarity between the user and the target item. Secondly, according to the similarity between the user and the target user, the evaluation of the item, and the score of these evaluations, the recommendation score for the user is predicted. Finally, select $n$ items with the largest predicted recommendation scores as the recommendation result, that is, recommend product items based on the similarity between users[49], and make corresponding comparisons. The specific algorithm idea is as follows:

$$S_{ij} = \frac{\sum_{l=1}^{n} a_{li} a_{lj}}{\min\left\{d(u_i), d\left(u_j\right)\right\}} \tag{1}$$

Among them, $a_{li}$ is user $u_i$'s preference for item $l$, $a_{lj}$ is user $u_j$'s preference for item $l$, $d(u_i) = \sum_{l=1}^{n} a_{li}$ is the degree of user $u_i$, and $d\left(u_j\right) = \sum_{l=1}^{n} a_{lj}$ is the degree of user $u_j$. $S_{ij}$ represents the similarity between target user $i$ and user $j$. If you want to estimate the user $u_i$'s score on the item $o_j$, the predicted score is given as

$$v_{ij} = \frac{\sum_{l=1, l\neq i}^{m} s_{li} a_{li}}{\sum_{l=1, l\neq i}^{m} s_{li}} \tag{2}$$

for any user $u_i$, all non-zero prediction scores $v_{ij}$ with $a_{ij} = 0$ are sorted in descending order, and the items ranked first will be recommended to the user.

### 4.3 Mass diffusion recommendation algorithm

The material diffusion method behaves similarly to the random walk process' resource allocation mechanism[14]. Based on the user-commodity bipartite graph, it is assumed that each good has a certain amount of some recommending power, i.e., the goods selected by a user have some abstract ability to recommend other goods to that user, and the goods take resources and pass more resources to their more preferred goods.

Since the bipartite graph itself is an unweighted network, the resources in the $X$ node will be equally distributed to the neighbors in each $Y$ node. Similarly, the resources in the $Y$ node will also be equally distributed to the neighbors in each $X$ node.

Taking Fig. 2 as an example, the initially recommended resources for the four $X$ nodes are $a$, $b$, $c$, and $d$. The resource allocation process is: First, the initial resources in the $X$ node are transferred to the $Y$ node according to the connection relationship of the bipartite graph. The transfer process is similar to the process of material diffusion. Resource dilution, for example, the degree of $a$ is 2, then the end point of each edge will get $a/2$; then the resources accumulated in the $Y$ node are transferred from the $Y$ node back to the $X$ node, and the allocation principle is the same as the previous step. The two sets realize the redistribution of resources through the common connection relationship between them. The greater the degree of the node, the smaller the energy transmitted by the connected edges related to it. The resource transfer of the two processes is shown in Fig. 2.

Finally, the resources allocated to the three $X$ nodes are $w$, $x$, $y$, and $z$, and Eq. (3) represents the matrix mapping relationship between the final resource and the original resource.

$$\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{6} & \frac{2}{9} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} & \frac{2}{9} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{6} & \frac{2}{9} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \tag{3}$$

where the matrix is a column-normalized weight

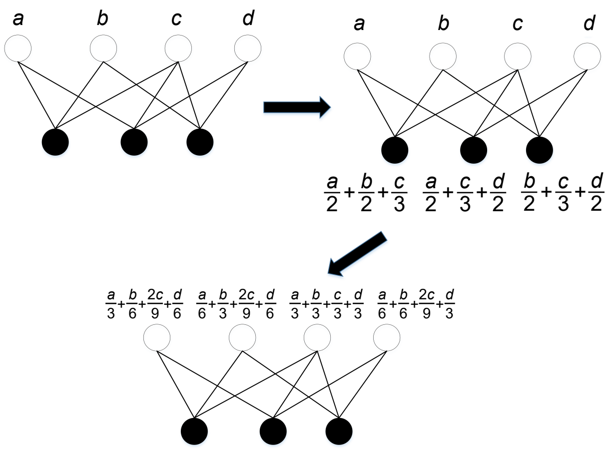



**Fig. 2 Schematic diagram of material diffusion.**

adjacency matrix. Assuming that there are $M$ users and $N$ kinds of commodities in the recommendation system, the bipartite graph has $M + N$ nodes, the user set $X = \{x_1, x_2, \ldots, x_m\}$, the product set $Y = \{y_1, y_2, \ldots, y_n\}$, and $E$ is the user's historical consumption record collection. It is assumed that all products selected by user $x_i$ have the ability to recommend other products to $x_i$. This abstract ability can be regarded as a kind of divisible resource on related products-products with resources that will give more resources to products that they prefer. If the user $x_i$ purchases, selects, or evaluates the product $y_k$, then $x_i, y_k \in E$. The initial resource of $x_i$ is $f(x_i) \geqslant 0$. The first step is to transfer the recommendation force resource from the $X$ set to the $Y$ set, then the resource of the commodity $y_k$ is Eq. (4).

$$f(y_k) = \sum_{i=1}^{m} \frac{a_{ik} f(x_i)}{k(x_i)} \tag{4}$$

where $k(x_i)$ is the user degree of user $x_i$, and $a_{ik}$ is a matrix of $M \times N$. Equation (5) can describe whether user $x_i$ and product $y_k$ have a choice relationship.

$$a_{ik} = \begin{cases} 1, & x_i, y_k \in E; \\ 0, & x_i, y_k \notin E \end{cases} \tag{5}$$

To put it another way, if the user $x_i$ selects the product $y_k$, then $a_{ik} = 1$, otherwise, $a_{ik} = 0$.

In the second step, the recommended force resources are passed from the $Y$ set back to the $X$ set, then the resources of the user $x_i$ can be expressed as Eq. (6).

$$f'(x_i) = \sum_{k=1}^{n} \frac{a_{ik} f(y_k)}{k(y_k)} = \sum_{k=1}^{n} \frac{a_{ik}}{k(y_i)} \sum_{j=1}^{m} \frac{a_{jk} f(x_i)}{k(x_j)} \tag{6}$$

Through the two-step material diffusion, the resource contribution from $x_j$ to $x_i$ can be expressed as

$$f'(x_i) = \sum_{j=1}^{m} w_{ij} f(x_j) \tag{7}$$

where $w_{ij}$ can be described as follows:

$$w_{ij} = \frac{1}{k(y_i)} \sum_{l=1}^{m} \frac{a_{il} a_{jl}}{k(x_l)} \tag{8}$$

The resource allocation matrix is $W=\{w_{ij}\}_{m \times n}$, $w_{ij}$ is the ratio of the initial resource of $j$ to $i$, and the final form of resource allocation of the original collection. It can be considered as the importance of node $j$ to node $i$. $W$ is an asymmetric matrix with diagonal elements greater than 0. Suppose that for a specific user $x_i$, an $n$-dimensional 0 and 1 personalized vector $f$ of the user's historical consumption record is defined. The commodity resource allocation selected by $x_i$ is initialized to unit quantity 1, and the rest are 0, that is $f\left(y_j\right)=a_{ji}\mathrm{e}^{a_{ik}}$, where it represents that the initial resource allocation structure for this user is personalized. For different users, the initialization vector $f$ is different. The final resource allocation vector is $f'$, $f' = wf$, and Eq. (9) is the component $f'\left(y_j\right)$ of $f'$:

$$f'\left(y_j\right)=\sum_{i=1}^{m} w_{jl} f\left(y_l\right)=\sum_{i=1}^{m} w_{jl} a_{li} \mathrm{e}^{a_{ik}} \tag{9}$$

Finally, all products $y_j$ ($1 \leqslant j \leqslant n$, $a_{ji} = 0$) that user $x_i$ did not select are sorted in descending order, and the product with the most resources is recommended.

## 5 Improvement of the Algorithm for Scoring of Substance Diffusion Products

Every product rated by the user is given the same weight in the conventional substance diffusion method. Even though the algorithm has a greater accuracy rate than the collaborative filtering algorithm, the practice of considering all goods identically ignores the impact of the user's rating on the recommendation effect. Many recommendation systems enable users to score products purchased or seen, making the system a scoring system[41], such as Douban's movie recommendation system, which allows users to rank movies from 1 to 10 points. The higher the score, the more people enjoy the film. The recommendation algorithm based on the network structure only judges whether the user has selected an item, and does not distinguish the impact of high and low scores on the recommendation results.

However, the user's score does, to some extent, reflect the user's liking for the item, and the lack of differentiation between high and low ratings may result in information loss. Another example is several e-commerce platforms, which allow customers to rate items depending on their level of pleasure after using them, which are separated into five categories. The higher the star rating, the more satisfied the customer is with the goods. When the average score is calculated using the old technique, the majority of the low ratings are ignored, and people with low scores are frequently exposed to the product's genuine quality. Users may also set certain high ratings depending on their preferences or traits. Ignoring the high score may also miss the same user's pursuit of the product.

As a result, the impact of product ratings on the recommendation effect is improved, and the high and low scores of goods are discriminated, indicating the user's preference for the product, using the substance diffusion recommendation algorithm. The first point to examine is that the user's average score is high, indicating that the user's desire for it is normally high. It should not be given the same resources as other items with low average ratings during the initial resource allocation. As a result, in the typical substance diffusion method, the initial resource allocation vector may be enhanced using Eq. (10):

$$f=a_{ji}\varpi=a_{ji}\mathrm{e}^{a_{ik}\left(r_j-\bar{r}\right)} \tag{10}$$

where $r_j$ represents the user's rating of product $j$, and $r$ represents the average rating of all products. When $r_j=\bar{r}$, the score of commodity $r_i$ is equal to the average score of all commodities, then commodity $r_i$ is a very common commodity, and the initial resource allocation vector $f$ degenerates to the initial resource allocation vector in the standard material diffusion algorithm. When $r_j-\bar{r}<0$, the adjustment factor has a weakening effect on the initial resource allocation, which means that the user's rating of the product is less than the score of all products, that is, the product is not very popular with users. It is a relatively unpopular product. Therefore, it should obtain fewer initial resources. On the contrary, popular products get more initial resources, and the adjustment factor can enhance the algorithm.

Another improvement point in the improved algorithm is the improvement of the resource transfer matrix $W_{ij}$, which also considers the impact of the user's rating of the product on the recommendation algorithm. The improved resource transfer matrix is

$$w'_{ij}=\frac{1}{k_j}\sum_{l=1}^{m}\frac{a_{il}a_{jl}}{k_l}\cdot\psi=\frac{1}{k_j}\sum_{l=1}^{m}\frac{a_{il}a_{jl}}{k_l}\cdot\mathrm{e}^{\frac{\bar{r}+\bar{r}_{li}-\bar{r}_{lj}}{\bar{r}}-1} \tag{11}$$

$$\psi=\mathrm{e}^{\frac{\bar{r}+\bar{r}_{li}-\bar{r}_{lj}}{\bar{r}}-1} \tag{12}$$

where $k_j$ is the degree of commodity $j$, that is, how many users choose commodity $j$, and $a_{il}a_{jl}$ indicates that user $l$ selects commodity $i$ and $j$ at the same time. $\psi$ is the adjustment factor. $\bar{r}_{li}$ and $\bar{r}_{lj}$ represent the average ratings of users for commodity $i$ and $j$, respectively. The same user may choose the same product more than once, so the average score of the score should be taken. When the average scores of the two commodities are equal, the algorithm degenerates to the standard substance diffusion algorithm, which means that the recommendation ability of commodity $j$ to commodity $i$ meets the average resource allocation. When $\bar{r}_{li} < \bar{r}_{lj}$, the adjustment factor is less than 1, which plays a weakening role, that is, the recommendation strength of commodity $j$ to commodity $i$ is reduced. On the contrary, the recommendation becomes stronger, and the adjustment factor is greater than 1, so that $i$ can obtain greater recommendation resources from $j$.

Through the resource transfer process of the material diffusion algorithm, the recommended resource vector obtained by the final product can be obtained as

$$f'_j = w'_{ij} f_j \tag{13}$$

where $f'_j$ is the final vector of referral resources obtained for commodity $j$. $f_j$ is the initial resource allocation vector for commodity $j$.

## 6 Experiment

### 6.1 Dataset

To evaluate the performance of the algorithm, benchmark datasets MovieLens 100K and MovieLens 1M are used for testing. The datasets used are widely used to test the efficiency of recommendation algorithms, and Table 1 summarizes the detailed data and sparsity of the two datasets. Each dataset is randomly divided into two parts: 80% of the data are used as the training set, and 20% of the user data are used as the test set. The reference link to the dataset is as follows: https://files.grouplens.org/datasets/movielens/.

### 6.2 Algorithm evaluation criteria

(1) Average ranking score is abbreviated as $\langle r \rangle$[50]. The recommendation algorithm will give a user's recommendation list in descending order according to the calculated user's preference for items. A good recommendation algorithm should try to rank the items that users like in the first place, and the item ranking results given by the recommendation algorithm should best match the user's preference for items. $\langle r \rangle$ represents the average ranking position of the edges belonging to the test set $E^P$ in the recommended sequence. For any connection $l_{ij}$ in the test set $E^P$, if the ranking position of the item $o_i$ in the recommendation sequence of user $u_j$ is $\text{rank}_{ij}$, and the number of items that have not been selected by user $u_j$ is $N_j$, then the average ranking score $\langle r \rangle$ is defined as

$$\langle r \rangle = \frac{\sum_{l_{ij} \in E^P} \frac{\text{rank}_{ij}}{N_j}}{\left| E^P \right|} \tag{14}$$

where $E^P$ represents the total number of records in the test set. Obviously, the smaller the value of $\langle r \rangle$, the better the accuracy of the recommendation algorithm.

(2) Precision[51] is abbreviated as $P$. The accuracy rate is used to measure the precision rate of the recommendation system, and is an important index to evaluate the accuracy of the recommendation algorithm. The ratio of the number of items that users are interested in to the total number of recommended items in the recommendation sequence is known as recommendation accuracy. For any user $u_i$ (denote user $u_i$ by $T_i(L)$), when the number of edges of test set $E^P$ in the recommendation sequence is $L$ items, the recommendation accuracy of user $u_i$ is $\frac{T_i(L)}{L}$. By averaging the accuracy rates of all users, the accuracy rate $P$ of the entire system can be obtained as

$$P = \frac{\sum_{i \in U} \frac{T_i(L)}{L}}{|U|} \tag{15}$$

where $U$ is the user set, and $|U|$ represents the total number of users.

(3) The recall rate[51] represents the recall rate of the recommendation system, and is another important metric for evaluating the accuracy of the recommendation algorithm. The recommendation recall rate indicates the probability that the item of interest to the user appears in the recommendation sequence. The recommendation recall rate of user $u_i$, is measured by the ratio of the number of accurately recommended items $T_i(L)$ in the recommendation sequence to the total number of candidate items in the test set, that is $\frac{T_i(L)}{E^P}$. Then when $L$ items are recommended, the recall rate of the entire system is

**Table 1 Statistical properties of the two datasets.**

| Dataset | Number of users | Number of items | Rating | Sparsity (%) |
|---|---|---|---|---|
| MovieLens 100K | 943 | 1682 | 100 000 | 93.7 |
| MovieLens 1M | 6040 | 3952 | 1 000 209 | 95.8 |

$$\text{Recall}(L) = \frac{\sum_{i=1}^{m} T_i(L)}{\left|E^P\right|} \tag{16}$$

The accuracy of the recall rate is not independent of each other, but there is a relationship that affects each other. The larger the value, the better the recommendation effect.

### 6.3 Algorithm comparison

In order to verify the effectiveness of the proposed algorithm in this paper, the below algorithms are used as benchmarks.

(1) **CR:** Content-based recommendation algorithm.

(2) **CF:** Collaborative filtering algorithm. CF is a classic recommendation algorithm and it recommends items based on the similarity between users.

(3) **MD:** Mass diffusion recommendation algorithm.

(4) **IMD:** Improvement of the algorithm for scoring of substance diffusion products.

### 6.4 Experimental result and analysis

In the movie rating table (Table 2), the first column indicates the user's ID, and the second column indicates the user's rating of the movie. These scoring data are all obtained from data preprocessing in the original dataset. The training set and test set of the data set have been cross-validated 5 times. Calculation of user and film weights using the IMD algorithm. Finally, the recommendation list of User 1 is shown in Table 3.

In the experiment, the precision, recall, and $\langle r\rangle$ of the four comparison algorithms were compared and analyzed on the two datasets.

The experiment sets different lengths of recommendation lists, observes the impact of various recommendation list lengths on the algorithm evaluation indicators, and determines the appropriate length of recommendation lists to achieve optimal personalized recommendations.

**Table 2 Part of the score sheet for Movie 1.**

| User ID | Rate |
|---|---|
| 2 | 4 |
| 4 | 4 |
| 10 | 4 |
| 14 | 5 |
| 21 | 3 |
| 22 | 4 |
| 24 | 3 |
| 25 | 4 |
| 26 | 4 |
| 28 | 5 |

**Table 3 Partial recommendation list of User 1.**

| Movie ID | Rate |
|---|---|
| 1 | 5 |
| 75 | 4 |
| 242 | 3 |
| 333 | 4 |
| 523 | 4 |
| 536 | 3 |
| 676 | 3 |
| 856 | 4 |
| 885 | 4 |
| 921 | 3 |
| 996 | 5 |

From the comparison of the precision of each algorithm on the two datasets in Fig. 3, the CR algorithm has the lowest precision, the traditional collaborative filtering algorithm is slightly lower than the standard material diffusion recommendation algorithm, and the improved material diffusion recommendation algorithm proposed in this paper has the highest precision. It can be seen that when the length of the recommendation list is 10, the accuracy of each recommendation algorithm reaches its peak, and decreases as the length of the recommendation list increases. Comparing the dataset of MovieLens 1M, we can also see that as the amount of data becomes larger and larger, the accuracy of the recommendation continues to decrease. When the length of the data recommendation list is small, the accuracy of the recommendation algorithm is higher.

Therefore, the experimental results mentioned above prove that our proposed algorithm can improve the prediction accuracy of the algorithm. Regarding the average prediction accuracy, this algorithm has increased by 1.8%, 1.6%, and 1.2%, respectively over the collaborative filtering algorithm, content-based recommendation algorithm, and standard substance diffusion algorithm.

From the comparison of the recall rate of each algorithm on the two datasets in Fig. 4, it shows the opposite result to the precision. Generally speaking, the smaller the precision, the greater the recall rate. Although the recall rate of the algorithm is the highest when the length of the recommendation list is 20, the

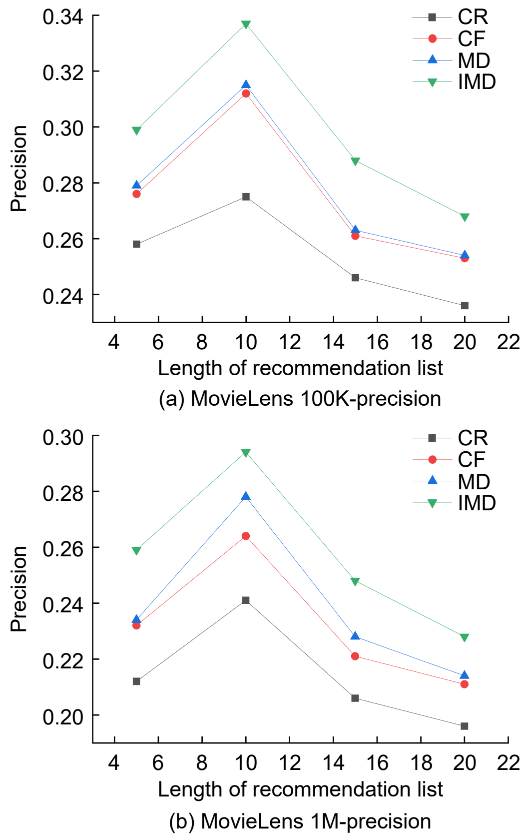



**Fig. 3 Comparison of the precision of CR, CF, MD, and IMD algorithms in two datasets.**

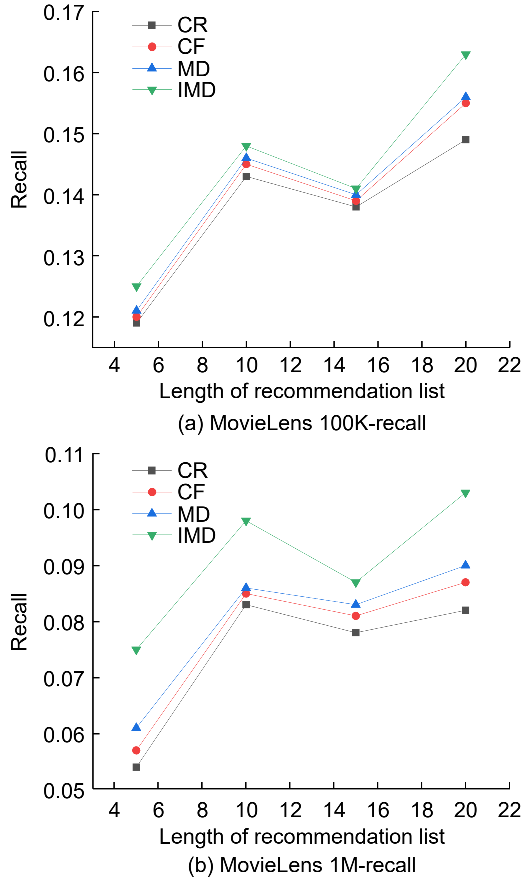



**Fig. 4 Comparison of the recall of CR, CF, MD, and IMD algorithms in two datasets.**

precision in Fig. 3 is the lowest. When the length of the recommendation list is 10, the combination of precision in Fig. 3 and recall rate of the algorithm is the highest. Therefore, from the overall point of view of the algorithm, the performance of the algorithm is the best when the length of the recommendation list is 10, and the improved substance diffusion recommendation algorithm proposed in this paper is better than the other three comparison algorithms.

In Fig. 5, the average ranking score index in the comparison algorithm shows a trend that is proportional to the length of the recommendation list in both datasets. This is in line with a general problem with our current recommendation algorithms. The algorithm proposed in this paper still has a good advantage in the index of the average ranking score. This is because after considering the user's score, it distinguishes between high scores and low scores, and increases the allocation of initial resources for projects with high user scores. For items with low user ratings, the allocation of initial resources is reduced, which plays a good role in the personalized recommendation.

In addition, this paper also considers the time cost of each comparison algorithm. CR needs to compute the weight vector of each item with time complexity $O(md)$, where $m$ is the number of items and $d$ is the number of features. CF needs to compute the similarity of each user or item with time complexity $O(n^2)$, where $n$ is the number of users or items. MD and IMD facilitate all lattice points one by one and compute the diffusion for each lattice point with time complexity O($N$), where $N$ is the diffusion region within is the number of lattice points in the diffusion region. Overall, MD and IMD have the lowest time complexity, and CR has a lower time complexity compared to CF.

## 7 Conclusion and Future work

This paper mainly studies the recommendation algorithm based on the bipartite graph network structure, and proposes an improved recommendation algorithm based on the weighted bipartite network structure graph. Firstly, on the basis of the standard

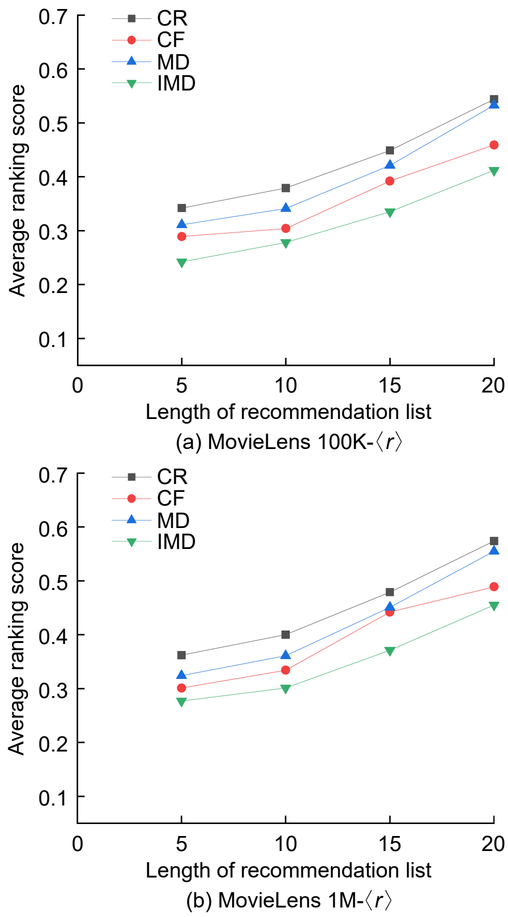



**Fig. 5 Comparison of the $\langle r \rangle$ of CR, CF, MD, and IMD algorithms in two datasets.**

substance diffusion algorithm, taking into account the impact of user ratings on recommended products, the initial resource allocation vector in the recommendation algorithm is improved, and the adjustment factor $\varpi$. For commodities whose scores are greater than the average score of all commodities, they are defined as popular commodities and should be recommended more strongly. On the contrary, it is an unpopular product, and the recommended resource should be smaller. In addition, when new products are added, the corresponding initial recommendation resources can also be obtained, thus solving the cold start problem, scalability problem, and diversification of product recommendation in the standard substance diffusion algorithm.

Secondly, on the basis of the standard substance diffusion algorithm, taking into account the impact of user ratings on recommended products, the resource transfer matrix in the recommendation algorithm is improved, and the adjustment factor $\psi$ is added. For products with greater similarity, that is, products with similar user ratings, more recommended resources will be obtained from the other party. Conversely, products with far different similarities, that is, when the ratings differ greatly, will receive more recommended resources. The similarity is the greatest when the two scores are the same, thus degenerating to the standard substance diffusion algorithm. In addition, when a new user joins, the value of the adjustment factor is 0, because there is no record of the selection of the product, and the algorithm degenerates to the standard substance diffusion algorithm, i.e., it treats all users who want to join equally, thus solving the cold start problem and the scalability problem of the standard substance diffusion algorithm.

The research methodology in this paper still needs to be improved. Although improvements in the similarity algorithm have improved the accuracy of the algorithm in a data sparse environment, the problem of data sparsity has not been solved, and future work is to use dense vectors to solve the sparsity problem. Dense vectors are able to extract important features by adaptively selecting the weight of each element through a learning process. Also regarding the cold start problem for users and projects, future work is to consider using migration learning to train this model by using the optimised recommendation model parameters as initial values.

## Acknowledgment

The work was supported by the National Natural Science Foundation of China (No. 62302199), China Postdoctoral Science Foundation (No. 2023M731368), Natural Science Foundation of the Jiangsu Higher Education Institutions (No. 22KJB520016), Ministry of Education in China (MOE) Youth Foundation Project of Humanities and Social Sciences (No. 22YJC870007), 2022 Jiangsu University Undergraduate Student English Teaching Excellence Program, and Ministry of Education's Industry-Education Cooperation Collaborative Education Project (No. 202102306005).

## References

[1] J. Mi, Research on the coupling and coordination of China's tourism economy and Internet development, in *Proc. 6th Int. Conf. Economics, Management, Law and Education* (*EMLE 2020*), Krasnodar, Russia, 2021, pp. 61–70.

[2] T. P. Liang, Y. F. Yang, D. N. Chen, and Y. C. Ku, A

semantic-expansion approach to personalized knowledge recommendation, *Decis. Support. Syst.*, vol. 45, no. 3, pp. 401–412, 2008.

[3] T. F. Tavares and L. Collares, Ethnic music exploration guided by personalized recommendations: System design and evaluation, *SN Appl. Sci.*, vol. 2, no. 4, pp. 1–9, 2020.

[4] D. Miao, L. Liu, R. Xu, J. Panneerselvam, Y. Wu, and W. Xu, An efficient indexing model for the fog layer of industrial Internet of Things, *IEEE Trans. Ind. Inform.*, vol. 14, no. 10, pp. 4487–4496, 2018.

[5] X. Bai, L. Liu, M. Cao, J. Panneerselvam, Q. Sun, and H. Wang, Collaborative actuation of wireless sensor and actuator networks for the agriculture industry, *IEEE Access*, vol. 5, pp. 13286–13296, 2017.

[6] L. Liu, N. Antonopoulos, M. Zheng, Y. Zhan, and Z. Ding, A socioecological model for advanced service discovery in machine-to-machine communication networks, *ACM Trans. Embed. Comput. Syst.*, vol. 15, no. 2, pp. 1–26, 2016.

[7] L. Shi, Y. Wu, L. Liu, X. Sun, and L. Jiang, Event detection and identification of influential spreaders in social media data streams, *Big Data Mining and Analytics*, vol. 1, no. 1, pp. 34–46, 2018.

[8] D. Mumin, L. L. Shi, L. Liu, and J. Panneerselvam, Data-driven diffusion recommendation in online social networks for the Internet of people, *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 52, no. 1, pp. 166–178, 2022.

[9] M. H. Yang and Z. M. Gu, Personalized recommendation based on partial similarity of interests, in *Advanced Data Mining and Applications*, X. Li, O. R. Zaïane, and Z. Li, eds. Berlin, Germany: Springer, 2006, pp. 509–516.

[10] L. L. Shi, L. Liu, Y. Wu, L. Jiang, M. Kazim, H. Ali, and J. Panneerselvam, Human-centric cyber social computing model for hot-event detection and propagation, *IEEE Trans. Comput. Soc. Syst.*, vol. 6, no. 5, pp. 1042–1050, 2019.

[11] L. L. Shi, L. Liu, Y. Wu, L. Jiang, J. Panneerselvam, and R. Crole, A social sensing model for event detection and user influence discovering in social media data streams, *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 1, pp. 141–150, 2020.

[12] H. S. Park, J. O. Yoo, and S. B. Cho, A context-aware music recommendation system using fuzzy Bayesian networks with utility theory, in *Proc. 3rd Int. Conf. Fuzzy Systems and Knowledge Discovery*, Xi'an, China, 2006, pp. 970–979.

[13] Y. L. Chen and L. C. Cheng, A novel collaborative filtering approach for recommending ranked items, *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2396–2405, 2008.

[14] D. E. Chen and Y. L. Ying, A collaborative filtering recommendation algorithm based on bipartite graph, *Adv. Mater. Res.*, vols. 756–759, pp. 3865–3868, 2013.

[15] Y. Zhang, X. Duan, X. Yue, and Z. Chen, A new efficient algorithm for weighted vertex cover in bipartite graphs based on a dual problem, in *Proc. 2018 9th Int. Conf. Information Technology in Medicine and Education* (*ITME*), Hangzhou, China, 2018, pp. 20–23.

[16] J. V. Sengers, Mass diffusion and thermodiffusion in multicomponent fluid mixtures, *Int. J. Thermophys.*, vol. 43, no. 4, pp. 1–10, 2022.

[17] L. L. Shi, L. Liu, Y. Wu, L. Jiang, and A. Ayorinde, Event detection and multi-source propagation for online social network management, *J. Netw. Syst. Manag.*, vol. 28, no. 1, pp. 1–20, 2020.

[18] L. Jiang, L. Shi, L. Liu, J. Yao, B. Yuan, and Y. Zheng, An efficient evolutionary user interest community discovery model in dynamic social networks for Internet of people, *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9226–9236, 2019.

[19] L. L. Shi, L. Liu, L. Jiang, R. Zhu, and J. Panneerselvam, QoS prediction for smart service management and recommendation based on the location of mobile users, *Neurocomputing*, vol. 471, pp. 12–20, 2022.

[20] S. H. Liao and C. A. Yang, Big data analytics of social network marketing and personalized recommendations, *Soc. Netw. Anal. Min.*, vol. 11, no. 1, pp. 1–19, 2021.

[21] Z. Huang, D. D. Zeng, and H. Chen, Analyzing consumer-product graphs: Empirical findings and applications in recommender systems, *Manag. Sci.*, vol. 53, no. 7, pp. 1146–1164, 2007.

[22] T. Zhou, J. Ren, M. Medo, and Y. C. Zhang, Bipartite network projection and personal recommendation, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, vol. 76, p. 046115, 2007.

[23] T. Zhou, R. Q. Su, R. R. Liu, L. L. Jiang, B. H. Wang, and Y. C. Zhang, Accurate and diverse recommendations via eliminating redundant correlations, *New J. Phys.*, vol. 11, no. 12, p. 123008, 2009.

[24] T. Zhou, M. Medo, G. Cimini, Z. K. Zhang, and Y. C. Zhang, Emergence of scale-free leadership structure in social recommender systems, *PLoS One*, vol. 6, no. 7, p. e20648, 2011.

[25] F. H. Wang and S. Y. Jian, An effective content-based recommendation method for web browsing based on keyword context matching, *Journal of Informatics and Electronics*, vol. 1, pp. 49–59, 2006.

[26] J. G. Liu, T. Zhou, H. A. Che, B. H. Wang, and Y. C. Zhang, Effects of high-order correlations on personalized recommendations for bipartite networks, *Phys. A Stat. Mech. Appl.*, vol. 389, no. 4, pp. 881–886, 2010.

[27] X. He, M. Gao, M. Y. Kan, and D. Wang, BiRank: Towards ranking on bipartite graphs, *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 1, pp. 57–71, 2017.

[28] Y. Zhou, D. Wen, F. Yuan, J. Li, and M. Li, Research of online water quality monitoring system based on zigbee network, *Advances in Information Sciences and Service Sciences*, vol. 4, no. 5, pp. 255–261, 2012.

[29] P. Jaccard, Comparative study on the distribution of flowers in some areas of the Alps and Jura, (in French), *Bull. Soc. Vaudoise Sci. Nat.*, vol. 37, pp. 547–579, 1901.

[30] T. R. Rao, S. K. Ghosh, and A. Goswami, Mining user-user communities for a weighted bipartite network using spark GraphFrames and Flink Gelly, *J. Supercomput.*, vol. 77, no. 6, pp. 5984–6035, 2021.

[31] M. S. Shang, L. Lü, W. Zeng, Y. C. Zhang, and T. Zhou, Relevance is more significant than correlation: Information filtering on sparse data, *Europhys. Lett.*, vol. 88, no. 6, p. 68008, 2009.

[32] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W. Y. Ma, Mining user similarity based on location history, in *Proc. 16th ACM SIGSPATIAL Int. Conf. Advances in geographic information systems*, Irvine, CA, USA, 2008, pp. 1–10.

[33] S. J. Beckett, Improved community detection in weighted bipartite networks, *R. Soc. Open Sci.*, vol. 3, no. 1, p. 140536, 2016.

[34] S. G. Li and L. Shi, The recommender system for virtual items in MMORPGs based on a novel collaborative filtering approach, *Int. J. Syst. Sci.*, vol. 45, no. 10, pp. 2100–2115, 2014.

[35] D. Zhang, F. Xie, D. Wang, Y. Zhang, and Y. Sun, Cluster analysis based on bipartite network, *Math. Probl. Eng.*, vol. 2014, pp. 1–9, 2014.

[36] E. A. Codling, R. N. Bearon, and G. J. Thorn, Diffusion about the mean drift location in a biased random walk, *Ecology*, vol. 91, no. 10, pp. 3106–3113, 2010.

[37] L. Lü and T. Zhou, Link prediction in complex networks: A survey, *Phys. A Stat. Mech. Appl.*, vol. 390, no. 6, pp. 1150–1170, 2011.

[38] C. C. Ratcliffe and O. Arandjelović, Tracking of deformable objects using dynamically and robustly updating pictorial structures, *J. Imaging*, vol. 6, no. 7, p. 61, 2020.

[39] J. P. Eckmann, E. Moses, and D. Sergi, Entropy of dialogues creates coherent structures in e-mail traffic, *Appl. Math.*, vol. 101, no. 40, pp. 14333–14337, 2004.

[40] P. P. Zhang, K. Chen, Y. He, T. Zhou, B. B. Su, Y. Jin, H. Chang, Y. P. Zhou, L. C. Sun, B. H. Wang, et al., Model and empirical study on some collaboration networks, *Phys. A Stat. Mech. Appl.*, vol. 360, no. 2, pp. 599–616, 2006.

[41] Q. Xuan, F. Du, and T. J. Wu, Empirical analysis of Internet telephone network: From user ID to phone, *Chaos*, vol. 19, no. 2, p. 023101, 2009.

[42] M. S. Shang, L. Lü, Y. C. Zhang, and T. Zhou, Empirical analysis of web-based user-object bipartite networks, *EPL Europhys. Lett.*, vol. 90, no. 4, p. 48006, 2010.

[43] K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A. L. Barabási, The human disease network, *Proc. Natl. Acad. Sci. USA*, vol. 104, no. 21, pp. 8685–8690, 2007.

[44] R. Lambiotte and M. Ausloos, Uncovering collective listening habits and music genres in bipartite networks, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, vol. 72, p. 066107, 2005.

[45] J. Laherrère and D. Sornette, Stretched exponential distributions in nature and economy: "Fat tails" with characteristic scales, *Eur. Phys. J. B Condens. Matter Complex Syst.*, vol. 2, no. 4, pp. 525–539, 1998.

[46] Z. Yang, Z. K. Zhang, and T. Zhou, Anchoring bias in online voting, *EPL Europhys. Lett.*, vol. 100, no. 6, p. 68002, 2012.

[47] B. Gonçalves and J. J. Ramasco, Human dynamics revealed through Web analytics, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, vol. 78, p. 026123, 2008.

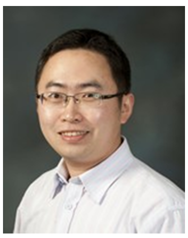
[48] Y. I. Chang, J. H. Shen, and T. I. Chen, A data mining-based method for the incremental update of supporting personalized information filtering, *J. Inf. Sci. Eng.*, vol. 24, no. 1, pp. 129–142, 2008.

[49] M. F. Aljunid and M. D. Huchaiah, Multi-model deep learning approach for collaborative filtering recommendation system, *CAAI Trans. Intell. Technol.*, vol. 5, no. 4, pp. 268–275, 2020.

[50] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender, Learning to rank using gradient descent, in *Proc. 22nd Int. Conf. Machine learning*, Bonn, Germany, 2005, pp. 89–96.

[51] D. M. W. Powers, Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation, arXiv preprint arXiv: 2010.16061, 2020.

**Qiang Sun** received the bachelor degree from Nanjing Sanjiang College, China in 2020. He currently is pursuing the PhD degree at the School of Computer Science and Communication Engineering, Jiangsu University, China. His research interests include data mining, social networks, link prediction, and recommender systems.



**Lu Liu** received the PhD degree from University of Surrey, UK in 2007. He currently is the head of the School of Computing and Mathematical Sciences University of Leicester, UK, and the adjunct professor at the School of Computer Science and Communication Engineering, Jiangsu University, China. He is the fellow of British Computer Society and member of IEEE. His research interests are in areas of cloud computing, social computing, data analytics, service-oriented computing, and peer-to-peer computing.

**Leilei Shi** received the BS degree from Nantong University, Nantong, China in 2012, the MS degree from Jiangsu University, Zhenjiang, China in 2015, and the PhD degree from Jiangsu University, Zhenjiang, China in 2020. He is a lecturer at the School of Computer Science and Communication Engineering, Jiangsu University, China. His research interests include event detection, data mining, social computing, and cloud computing.

**Zixuan Han** received the BS and MS degrees from Jilin Normal University, Siping, China in 2018 and 2021, respectively. He currently is pursuing the PhD degree at the School of Computer Science and Communication Engineering, Jiangsu University, China. His research interests include data mining, social networks, and cloud computing.

*Big Data Mining and Analytics, June* 2024, 7(2): 357−370

**Liang Jiang** received the MS and PhD degrees from Jiangsu University, Zhenjiang, China in 2011 and 2021, respectively. He currently is a lecturer at the School of Ocean College, Jiangsu University of Science and Technology, China. His research interests include Online Social Networks (OSNs), data mining, social computing, and network security.

**Yan Wu** received the MS degree from Shandong University of Science and Technology, Qingdao, China in 2009, and the PhD degree from Tongji University, Shanghai, China in 2014. He currently is a lecturer at the School of Computer Science and Communication Engineering, Jiangsu University, China. His research interests include formal methods, service-oriented computing, and cloud computing.

**Yeling Zhao** received the bachelor degree in finance management and the master degree in accounting from Jiangsu University, China in 2012 and 2015, respectively. She now serves as the lecturer at the School of Finance and Economics, Jiangsu University, China. Her research interests mainly cover big data mining in finance, information system for internal control, and e-commerce accounting.