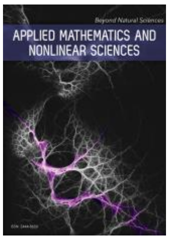
sciendo Applied Mathematics and Nonlinear Sciences, 9(1) (2024) 1-17



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

A Study of Informatization and Data-Driven Career Planning in Career Guidance in Colleges and Universities

Yuanyuan Yang[1], Zhanjun Wei[2,†]

1. Enrollment and Employment Office, Xinyang Normal University, Xinyang, Henan, 464000, China.
2. Business School, Xinyang Normal University, Xinyang, Henan, 464000, China.

Submission Info

Communicated by Z. Sabir
Received June 5, 2024
Accepted October 3, 2024
Available online November 11, 2024

Abstract

Traditional career planning education and employment guidance is based on lectures and text reading, with a large amount of subjective coloring, which can't achieve one person one strategy. This paper constructs a college career guidance information recommendation system based on collaborative filtering and K-means clustering to recommend suitable career information with a high potential success rate for students from both subjective and objective aspects. The weights of students' features are objectively calculated, and the K-means algorithm is improved with the minimum spanning tree method to spontaneously calculate the initial number of categories and the class center value from the data, avoiding the problem of random generation and enhancing the effect of clustering. The system is tested against other models in the dataset to examine its clustering performance. It is found that on the optimal K-mean-square error, the errors of the model in this paper are all less than 0.96, which is the lowest among the algorithms on all terms, and its clustering performance is again verified. The best results on the dataset are achieved by the proposed algorithm, with an HR@20 of 59.31%, which is almost twice the SVD recommendation method. The experiment confirmed the significance of each module in the recommender system. Compared with the recommendation of graduates' jobs, the system is easier to recommend industries for graduates, and the HR indicator is greater than 80% at the recommendation list lengths of 3 and 5, while the HR@5 of SVD, BPR is below 50%. This study provides useful insights into the integration of cutting-edge information technology career guidance and data mining to aid in career planning.

Keywords: Collaborative filtering; AK-means; Minimum spanning tree method; Career planning.
AMS 2010 codes: 94A16

†Corresponding author.
Email address: pdsxingzhe@126.com

ISSN 2444-8656
https://doi.org/10.2478/amns-2024-3155

1 Introduction

As an important part of public employment services, college career guidance is responsible for college students' career choice, employment and suitability. However, the employment guidance service in Chinese colleges and universities still exists problems such as insufficient service awareness, weak teacher strength, and poor service implementation [1-3], which is fundamentally due to the fact that colleges and universities, as a single supplying main body, with high inputs and low outputs, can no longer satisfy the needs of modernization of employment guidance as well as the growing personalized needs of students [4-6]. The new media technology breaks the limitation of time and space, which provides strong technical support for the informatization construction and practice of college students' employment guidance work and is conducive to providing more convenient employment guidance services for college students [7-9]. On the other hand, as the number of college graduates increases year by year, the employment situation of college students has become more severe, for which the employment guidance work of colleges and universities needs to be constantly innovated with the help of new media technology in order to better solve the problem of the difficulty of employment of college graduates, so as to help college students better adapt to the employment market and improve the quality of college students' employment [10-13].

For specific groups in colleges and universities, data-driven career planning needs to be combined with their characteristics and the work environment in which they are working to plan in all aspects, formulate career development goals that are in line with their development, and make constant adjustments and corrections according to the feedback results of the completion of the goals at different times and stages, and ultimately realize the process of achieving the goals [14-17].

Literature [18], on the basis of respecting the subjective position of students, combined with students and social relations to establish a "four-dimensional eight-framework" employment system based on the characteristics of the students, counseling strategy from two key perspectives and make an accurate working model. Literature [19] points out that China's economic development is contrary to the employment status of college graduates, and the difficulty of employment has become a serious problem in Chinese society. The related experience of innovation and entrepreneurship is borrowed, and the management mechanism of innovation and entrepreneurship education in colleges and universities is studied. Literature [20] shows that employment is an important way to realize the value of college students. Based on the background of "Internet +", innovative research on college students' employment guidance is carried out. It is emphasized that in the era of "Internet +", only by using the advantages of intelligent media to optimize the idea of employment guidance can we improve the employment situation of college students. Literature [21] studied the impact of informatization teaching on the development of college students' employment ability. The data of students from different academic degrees and majors were collected through a questionnaire survey, and after processing and analysis, it was concluded that informatization education has a great influence on the employment ability of college students. Literature [22] studied the career planning and employment guidance of Chinese college students, analyzed the problems existing in the current employment information management of college students in China, and put forward relevant suggestions with the help of the existing research results, which provided theoretical support for students' career planning and employment guidance. Literature [23] studied the integration path of college students' employment guidance based on Internet Civic Education. It also introduced the specific integration path factors, and college students' employment guidance factors. The results concluded that college students' employment guidance based on Internet Civic and Political Education has certain effects. Literature [24] discussed college students' career planning from the perspective of the Internet. It pointed out that in order to establish students' clear career planning, colleges and universities need to keep abreast of the times and appropriately adjust and optimize their education methods in order to correctly guide students' employment. Literature [25] emphasizes that the difficulty of college students' employment is largely related to the lack of career guidance education in colleges and universities, and the weak competitive ability of college students is also an important factor. The current situation of China's employment guidance education is analyzed, and it is pointed out that improving students' employment competitiveness in order to promote high-quality employment requires guidance and education in colleges and universities.

In this paper, we constructed a college career guidance information recommendation system based on data mining to achieve more scientific and efficient career planning guidance. Firstly, the collaborative filtering algorithm is utilized to find the set of users most similar to the user by judging their similarity based on the user's historical data. Then, with reference to the kruskal minimum spanning tree method in data structure, the traditional K-Means algorithm is improved, and combined with the characteristics of students' employment recommendation, the student feature relevance is weighted to student similarity, and the improved collaborative filtering algorithm based on student feature attributes is designed. After completing the clustering analysis test on the dataset and verifying the improvement effect, the system is applied to the employment recognition recommendation.

2 Method

The original student data in the school is growing geometrically, and these data contain great value. In the face of the huge amount of data, if we can process it scientifically, "take the essence and remove the dross", and extract the information related to employment and career planning that is hidden in the data, then the data can become a resource for career guidance that can be used in colleges and universities. Colleges and universities can use it as a resource for career guidance. The collaborative filtering algorithm and the improved K-Means algorithm will be combined in this paper to create a employment guidance information recommendation system that is based on data mining for colleges and universities.

2.1 Collaborative Filtering Algorithm

2.1.1 User-based collaborative filtering algorithm

According to the principle of the collaborative filtering algorithm, the rating similarity is carried over to other items' evaluation, and the resulting ratings are more similar. [26]. The flow of the algorithm is expressed in the following 3 steps:

1) Data model description. Construct a matrix of order $a \times b$, a represents the number of users, b represents the number of items, and R represents the set of user ratings for the items, where $R = \left(r_{ij}\right)_{a \times b}$ and r_{ij} are the user i 's favorable ratings for item j, which is generally set to $r_{ij} \in \{0,1,2,3,4,5\}$.

2) Find the set of K nearest neighbor users. There are 3 commonly used similarity calculation methods such as pinch cosine, modified pinch cosine, Pearson's similarity, etc., and the equations are as follows:

$$sim(u,v) = \cos(\vec{u},\vec{v}) = \frac{\vec{u} \cdot \vec{v}}{|\vec{u}||\vec{v}|} = \sum_{i \in I_{uv}} r_{ui} r_{vi} \Bigg/ \sqrt{\sum_{i \in I_{uv}} r_{ui}^{2}} \sqrt{\sum_{i \in I_{uv}} r_{vi}^{2}} \tag{1}$$

$$sim(u,v)=\sum_{i\in I_{uv}}\left(r_{ui}-\overline{r_u}\right)\left(r_{vi}-\overline{r_v}\right)\Big/\sqrt{\sum_{i\in I_u}\left(r_{ui}-\overline{r_u}\right)^2}\sqrt{\sum_{i\in I_v}\left(r_{vi}-\overline{r_v}\right)^2} \tag{2}$$

$$sim(u,v)=\sum_{i\in J_{uv}}\left(r_{ui}-\overline{r_u}\right)\left(r_{wi}-\overline{r_v}\right)\Big/\sqrt{\sum_{i\in I_{uv}}\left(r_{ui}-\overline{r_u}\right)^2}\sqrt{\sum_{i\in I_{uv}}\left(r_{vi}-\overline{r_v}\right)^2} \tag{3}$$

3) Generate Top-N recommendation dataset. The expression to find the prediction score is as follows:

$$P(u,i)=\overline{r_u}+\sum_{v\in KNN(u)}sim(u,v)\left(r_{vi}-\overline{r_v}\right)\Big/\sum_{v\in KNN(u)}sim(u,v) \tag{4}$$

2.1.2 Project-based collaborative filtering algorithms

The algorithm for item-based collaborative filtering is to find the other items that are most similar to one's ratings of previous items and recommend the few with the highest similarity to the target user.

1) Data model description. Construct a matrix of order $a\times b$, a represents the number of users, b represents the number of items, and R represents the set of user ratings for the items, where $R=\left(r_{ij}\right)_{a\times b}$ and r_{ij} are the user i 's favorable ratings for item j, which is generally set to $r_{ij}\in\{0,1,2,3,4,5\}$.

2) Calculate the set of most similar items. Common methods for calculating the similarity of items i and j include angle cosine, modified angle cosine, Pearson's similarity, and so on.

3) Generate Top-N recommendation dataset. The user's predicted rating of the target has the following 3 functions:

$$P(u,i)=\sum_{j\in KNN(i)}r_{uj}\Big/\left|KNN(i)\right| \tag{5}$$

$$P(u,i)=\sum_{j\in KNN(i)}sim(i,j)r_{uj}\Big/\sum_{j\in KNN(i)}sim(i,j) \tag{6}$$

$$P(u,i)=\overline{r}_i+\sum_{j\in KNN(i)}sim(i,j)\left(r_{uj}-\overline{r}_j\right)\Big/\sum_{j\in KNN(i)}sim(i,j) \tag{7}$$

2.2 Improved AK-means feature clustering algorithm

2.2.1 Calculation of similarity of student employment information features

First, the similarity of students' employment information features is calculated, and each student's employment features are constructed into a vector table $X=\{X_1,X_2,\ldots,X_n\}$.

Let the n-dimensional eigenvectors of students u and v be $U=\{x_{u1},x_{u2},\ldots,x_{um}\}$, $V=\{x_{v1},x_{v2},\ldots,x_{vm}\}$, respectively, and measure the distance between students according to the Euclidean distance, then the eigendistances of students U and V are as follows, see Equation (8):

$$d_{uv}=\sqrt{\sum_{k}^{n}\left|x_{ut}-x_{vk}\right|^{2}} \tag{8}$$

The similarity of student characteristics is obtained by transforming the distance and similarity coefficients, see equation (9):

$$Sim_{uv}=1/\left(1+d_{uv}\right)=1\Bigg/\left(1+\sqrt{\sum_{k}^{n}\left|x_{\alpha}-x_{\alpha}\right|^{2}}\right) \tag{9}$$

Adding the weights of student features makes the cluster analysis more in line with the real situation. Let the weight of student features be $\theta=\{\theta_1,\theta_2,\ldots,\theta_K\}$, then the feature distance of students U and V after adding the weight is as follows, see equation (10):

$$d_{uv}=\sqrt{\sum_{k=1}^{n}\theta_k\left|x_{du}-x_{vk}\right|^{2}} \tag{10}$$

The similarity of student characteristics is obtained by transforming the distance and similarity coefficients, see equation (11):

$$Sim_{uv}=1/\left(1+d_{uv}\right)=1\Bigg/\left(1+\sqrt{\sum_{k=1}^{n}\theta_k\left|x_{d}-x_{vk}\right|^{2}}\right) \tag{11}$$

2.2.2 Improved AK-means clustering principle

Referring to the kruskal minimum spanning tree method in data structures, kruskal requires to let the spanning tree have no loops in the spanning tree, and the edges are minimized each time they are connected, and to ensure the similarity transfer, it is necessary to let the connected graph generate loops [27]. Therefore, this paper proposes the Adcanced-K-Means algorithm, or AK-means for short.

Let the set of students $S=\{s_1,s_2,\ldots,s_m\}$, the set of student features be $X=\{X_1,X_2,\ldots,X_n\}$, the categories of output clustering be $s_1,s_2\ldots,s_k$, the cluster center value of each class be $a_1,a_2\ldots,a_k$, and k be the number of clusters.The process of implementation of AK-means algorithm is:

1) Any two students u,v, $u,v\in S$, u,v are regarded as vertices, then the edges connected between students u and v are (u,v), and the student eigenvector is X, calculate the distance between students, assign the distance between students as a weight w to each edge, and express the student set and student feature distance as a weighted undirected connected figure $WN=(V,E)$, where V represents the vertice, E is the weighted edge, and the initial state time space is a subgraph $WN_0=(V,\{\})$ with m vertices and no edges, and each vertex is used as a connected component.

2) Select the edge with the smallest weight in turn, if the vertex of the edge falls in the WN and the vertex belongs to a different connected component, then put this edge into the WN and connect the two vertices, then select the edge with the smallest weight among the remaining edges until there is an edge with both vertices appearing in the connected component in the WN, then add this edge into the WN then make the edges in the WN form a loop, put these vertices into the same class S_i and delete these vertices in the WN, and then continue this selection process among the remaining vertices until all the vertices are assigned and the last vertex appears in the last connected component of the WN [28].

3) Repeat step 2) until all vertices have been assigned to a non-intersecting K set to form K classes S_i, $i \in [1,k]$. The center of the K class is then: $an = \sum_{s \in s_i} s / |s_i|$, where $|s_i|$ is the number of class S_i students. The resulting K categories are the initial number of clusters, and the resulting class centers are the initial centers of the clusters.

4) He knows the initial category center and the number of students, for each student $s \in S$, using the formula (4) to calculate the distance of students to the center of each class, and then with the shortest distance from the center of each category of students are divided into various categories until the end of the allocation of all students.

5) Calculate the mean value of each class again and assign that value to each class center.

6) Repeat steps 4) and 5) until the squared error criterion function $E = \sum_{i}^{k} \sum_{s \in s_i} |s - a_n|^2$ is 0 and the centers of the categories are not moving, ending the clustering.

2.3 Selection of Employment Information Characteristics

The data used in this study mainly includes the following information and records: records of students' personal information, records of students' academic performance, records of students' campus card purchases, data on demographics, regional economic and geographic distribution, and statistics on student employment.

1) Calculation of Regional Economic Index

Regional Economic Index (REI) indicates the degree of economic development of a region. The calculation formula is shown below:

$$REI = \sum_{i=1}^{n} a_i C_i \tag{12}$$

where all features as well as REI are normalized to interval $[0,1]$ using the normalization method Min-Max, which is shown in Equation (13):

$$X_{norm} = \frac{x - x_{\min}}{X_{\max} - X_{\min}} \tag{13}$$

2) Calculation of Regional Familiarity Index

The Regional Familiarity Index (RFI) reflects the degree of familiarity with a region.The formula for calculating the RFI is as follows:

$$RFI = \max\left(HFI, SFI\right) \tag{14}$$

Where HFI and SFI denote familiarity with the district relative to students in the birthplace and schoolplace, respectively, and are calculated as follows:

$$HFI = a^{-D(h,l)} \tag{15}$$

$$SFI = a^{-D(s,l)} \tag{16}$$

3) Calculation of Graduate Performance Index

The Graduate Academic Performance Index (GPI) is used to measure how well a student is performing academically, with larger values representing better academic performance. The value is estimated based on the grades of the courses taken by the students and is defined as follows:

$$GPI = \frac{\sum_{i=1}^{n} PI_i}{n} \tag{17}$$

4) Calculation of Graduate Family Economic Condition Index

The Graduate Family Economic Condition Index (GEI) is used to measure the economic condition of students' families, whose financial ability has a crucial impact on their employment choices. The calculation formula is shown in (18):

$$GEI = P\left\{Y = 0 \middle| X\right\} \tag{18}$$

The conditional probabilities in Equation (18) are calculated from the logistic regression function as shown in (19) below:

$$GEI = \frac{1}{1 + e^{-\left(w^{\mathrm{T}} X + b\right)}} \tag{19}$$

where the parameters to be solved w and b are modeled from the sample data.

2.4 Personalized Employment Recommendation Algorithm

2.4.1 General framework

The framework of the university's employment guidance information recommendation system consists of two main parts: identification of graduate group clusters and recommendations for employment.

Graduate group identification involves inputting students' academic performance index and family economic condition index and using a clustering algorithm to classify graduate groups.

Graduate Employment Recommendation: This part first extracts the attributes of employment units from the student employment statistics and demographic, regional economic and geographic distribution data, as well as calculates the regional economic index and regional familiarity index of the regional characteristics of the employment units. Then, it calculates the student group employment choices with reference to the group division results; finally, it calculates the graduates' ratings of the employment units by combining the students' personalized preferences for the employment units, including the students' preferences for the attributes of the employment units, and their preferences for the locations of the employment units.

The improved AK-means clustering algorithm above is used in this study to divide the student groups.

2.4.2 Employment recommendation for graduates

The ratings of employers were based on three main components: group employer choice, students' preferences for employer attributes, and student preferences for employer location.

2.4.3 Choice of group employment units

In this paper, we use a Bayesian personalized ranking strategy to solve for group employment choices. Let $R \in \mathbb{R}^{M \times N}$ denote the graduate-employment unit interaction matrix, where M denotes the number of graduate groups and N denotes the number of employment units. R_{ui} denotes the number of times the graduate group u chooses the employment unit i. R_B is shown below:

$$\mathrm{R}_B = \left\{ (u,i,j) \middle| i \in \mathrm{R}(u) \text{ and } j \in \mathrm{R}(u)_{<i} \right\} \tag{20}$$

where $\mathrm{R}(u)$ represents the set of employment units for which group u has feedback, and $\mathrm{R}(u)_{<i}$ represents the set of all employment units for which group u has a rating less than that of employment unit i. Therefore, the ternary relationship (u,i,j) can be considered to represent the preference ranking of group u for employment unit i over employment unit j.

Let $U = [U_1, \cdots, U_D] \in \mathbb{R}^{M \times D}$ and $V = [V_1 \cdots, V_D] \in \mathbb{R}^{N \times D}$ denote the graduate cohort hidden factor matrix and the employment unit hidden factor matrix, respectively, so that a scoring matrix R that approximates R can be computed by the following equation:

$$\hat{R}_{ui} = U_u V_i^T \tag{21}$$

Where R_{ui} is the element of the u rd row and i th column of R, and U and V are the parameters to be solved. It can be solved by the objective function equation (22):

$$\arg\max_{U,V} \sum_{(u,i,j) \in R_B} \ln \sigma\left(\hat{R}_{uij}\right) - \lambda\left(\|U\|^2 + \|V\|^2\right) \tag{22}$$

Where σ is the sigmoid function, ie:

$$\sigma\left(\hat{R}_{uij}\right) = \frac{1}{1 + e^{-k_{uij}}} \tag{23}$$

$\hat{R}_{uij}$ is calculated as:

$$\hat{R}_{uij} = \hat{R}_{ui} - \hat{R}_{uj} \tag{24}$$

Where $\hat{R}_{ui}$ is calculated from equation (22).

2.4.4 Preferences for Attributes of Employment Units

Employment unit attributes should also be considered as an important factor in students' employment recommendations. This subsection introduces students' preferences for employment unit attributes based on the original group's employment unit selection method.

Let $A = [A_1, \cdots, A_L] \in \mathbb{R}^{M\times L}$ denote the matrix of group preferences for employment unit attributes, where M and L denote the number of groups and the number of employment unit attributes, respectively. $F = [F_1, \cdots, F_L] \in \mathbb{R}^{N\times L}$ denotes the matrix of employment unit attributes and N denotes the number of employment units. The attribute preference preference of group over employment unit scores is equal to $A_u F_i^T$, then equation (22) can be updated as:

$$\hat{R}_{ui} = U_u V_i^T + A_u F_i^T \tag{25}$$

Therefore, the optimization objective function is updated as:

$$\begin{aligned} \underset{U,V,A}{\arg\min} \sum_{(u,i,j)\in \mathbb{D}_B} & -\ln \sigma\left\{\left(U_u V_i^\tau + A_u F_i^\tau\right) - \left(U_u V_j^\tau + A_u F_j^\tau\right)\right\} \\ & + \lambda\left(\|U\|^2 + \|V\|^2 + \|A\|^2\right) \end{aligned} \tag{26}$$

Where U, V and A are the parameters to be solved.

2.4.5 Preferences for the location of the employing organization

Student's preference for place of employment should also be an important part of the employment recommendation algorithm. In this subsection, the method of solving for students' preference for place of employment will be given.

Denote by $P = [P_1, \cdots, P_N] \in \mathbb{R}^{G\times N}$ the matrix of students' regional preferences for employment units, where G and N denote the number of students and the number of employment units, respectively. P_{gi} denotes the regional preference of graduates g for the employment unit i, which is calculated as follows:

$$P_{gi} = f\left(RFI_{gi}, REI_i; \theta\right) \tag{27}$$

Eq. f is a preference calculation function for the location of the employment unit with inputs RFI_{gi} and REI_i, and REI_i denotes the regional economic index of the region where the employment unit i is located. RFI_{gi} is the index of familiarity of graduates g with the region where the employment unit i is located.

f is a function on two variables RFI and REI. In this paper, we use a binary normal distribution function to fit the function f. The binary normal distribution is a special form of the multivariate normal distribution, which is a generalization of the one-dimensional normal distribution to the multidimensional one. denotes the probability density of two variables RFI and REI. The formula is as follows:

$$f(x;\theta)=\frac{1}{\sqrt{(2\pi)^2|\Sigma|}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \tag{28}$$

Where $x=(RFI,REI)^T$, $\theta=(\mu,\Sigma)$. μ and Σ denote the mean and covariance matrices of x, respectively. There are $\mu=\begin{pmatrix}\mu_{RFI}\\ \mu_{REI}\end{pmatrix}$, $\Sigma=\begin{pmatrix}\sigma_{RFI}^2 & \rho\sigma_{RFI} & \sigma_{REI}\\ \rho\sigma_{RFI} & \sigma_{REI} & \sigma_{REI}^2\end{pmatrix}$, where ρ is the correlation coefficient between RFI and REI, and the variance $\sigma_{RFI}>0$, $\sigma_{REI}>0$. then, equation (28) can be written as:

$$\begin{aligned} f(RFI,REI;\theta)=&\frac{1}{2\pi\sigma_{RII}\sigma_{REI}\sqrt{1-\rho^2}}\exp\\ &\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(RFI-\mu_{RFI})^2}{\sigma_{RFI}^2}+\frac{(REI-\mu_{REI})^2}{\sigma_{REI}^2}\right.\right.\\ &\left.\left.-\frac{2\rho(RFI-\mu_{RFI}(REI-\mu_{REI}))}{2\sigma_{RFI}\sigma_{REI}}\right)\right) \end{aligned} \tag{29}$$

2.4.6 Choice of Employment Units for Graduates

Combine $\hat{R}_{ui}$ and P_{gi} to calculate the graduate g 's rating $\hat{R}_{gi}$ of the employing organization i with the following formula:

$$\hat{R}_{gi}=\hat{R}_{ui}+\eta\cdot P_{gi} \tag{30}$$

Where η is the regulating parameter. Depending on the magnitude of the rating $\hat{R}_{gi}$, units of employment are deposited for students g in descending order.

2.4.7 Employment Recommendation Algorithm Solution

The detailed process of graduate employment recommendation is given above, but in solving the rating $\hat{R}_{gi}$, it is necessary to specify the value of $\hat{R}_{ui}$ and the value of $\hat{R}_{ui}$ depends on the variables U, V and A. In this section, the process of solving the rating $\hat{R}_{gi}$ of graduates g on the employment organization i and the method of solving the objective function are discussed.

Stochastic Gradient Descent (SGD) is used to solve the objective function.

The objective function is denoted by J and all parameters to be solved U, V, A are denoted by Θ. The gradient update formula is shown in (31):

$$\Theta = \Theta - \alpha \cdot \Delta_{\Theta} \tag{31}$$

Solving for the gradient of parameters U, V, and A, respectively, has:

$$\Delta_{U_u} = -\frac{e^{-\hat{R}_{uij}}}{1+e^{-\hat{R}_{uij}}} \cdot \left(V_i^T - V_j^T\right) + \lambda U_u \tag{32}$$

$$\Delta_{V_i} = -\frac{e^{-R_{uij}}}{1+e^{-k_{uij}}} \cdot U_u + \lambda V_i \tag{33}$$

$$\Delta_{V_j} = -\frac{e^{-\hat{R}_{uij}}}{1+e^{-\hat{R}_{uij}}} \cdot \left(-U_u\right) + \lambda V_j \tag{34}$$

$$\Delta_{A_u} = -\frac{e^{-R_{uij}}}{1+e^{-R_{uij}}} \cdot \left(F_i^T - F_j^T\right) + \lambda A_u \tag{35}$$

3 Results and Discussion

3.1 Real dataset clustering test

In order to compare the advantages and disadvantages of this paper's model and other clustering evaluation metrics, this paper compares different clustering evaluation metrics through the ability to determine the optimal K of the clustering evaluation metrics. By determining the optimal K, we mean that we use different Ks as parameters in clustering and then determine which K is the number of clusters of the true category labels by the clustering evaluation metrics of different clustering results.

In this paper, we use 19 datasets in UCI, which have sample numbers from 200 to 1800, cluster numbers from 3 to 26, and attribute dimensions from 4 to 90.

The ability of different clustering evaluation metrics to find the optimal K is compared by comparing the optimal K hit rate and the optimal K mean square error. The K hit rate is determined by the number of clusters in the run. The optimal K hit rate is the percentage of cases where the clustering algorithm is run a number of times, and the optimal K determined by the computational clustering evaluation metric is equal to the actual number of clusters. The optimal K mean square error is the error between the optimal K determined by running the clustering algorithm thousands of times and calculating the number of clusters determined by the clustering evaluation metrics and the actual number of clusters each time.

The optimal K hit rate of this paper's algorithm and other clustering evaluation indexes is shown in Table 1, from which it can be seen that this paper's algorithm outperforms other clustering evaluation indexes on all items, and the hit rate values are higher than 0.55, which shows the superiority of this paper's algorithm relative to other clustering evaluation indexes.

Table 1. Optimal K percentage

Dataset	ARI	NMI	NVD	Purity	CSI	PSI	CHI	This model
Balance	0.19	0.35	0.23	0.39	0.19	0.27	0.17	0.75
Car	0.27	0.2	0.31	0.28	0.16	0.3	0.16	0.73
Chart-input	0.21	0.36	0.39	0.16	0.29	0.14	0.14	0.57
Conn	0.19	0.31	0.22	0.19	0.16	0.28	0.26	0.57
Dermatology	0.31	0.12	0.21	0.35	0.25	0.11	0.3	0.68
Ecoli	0.2	0.2	0.22	0.35	0.16	0.13	0.34	0.79
Glass	0.24	0.25	0.26	0.26	0.14	0.25	0.19	0.56
Image	0.26	0.19	0.37	0.18	0.39	0.17	0.12	0.6
Letter	0.12	0.27	0.23	0.28	0.3	0.21	0.27	0.76
Libras	0.32	0.33	0.18	0.27	0.36	0.36	0.28	0.72
Optdigits	0.29	0.24	0.33	0.32	0.27	0.12	0.4	0.59
Pen	0.38	0.22	0.26	0.38	0.23	0.26	0.25	0.56
Pendigits	0.28	0.31	0.11	0.22	0.18	0.37	0.31	0.74
Segment	0.15	0.35	0.15	0.16	0.39	0.12	0.38	0.7
Statlog-image	0.24	0.35	0.33	0.29	0.25	0.37	0.12	0.58
Statlog-vehicle	0.15	0.28	0.28	0.12	0.26	0.33	0.39	0.8
Steel-plates	0.21	0.18	0.15	0.32	0.26	0.34	0.24	0.72
Thyroid	0.32	0.17	0.38	0.26	0.25	0.13	0.36	0.59
Yeast	0.1	0.33	0.27	0.13	0.34	0.38	0.27	0.73

A comparison of the optimal K-mean-square errors is shown in Table 2. It can be seen that the model errors in this paper are all less than 0.96, which is the lowest among the algorithms in all terms, and its clustering performance is again verified.

Table 2. Optimal K error

Dataset	ARI	NMI	NVD	Purity	CSI	PSI	CHI	This model
Balance	11.1	9.96	11.93	8.54	1.7	11.9	14.58	0.48
Car	9.12	13.47	7.62	6.44	1.8	15.69	7.77	0.95
Chart-input	14.29	4.32	8.2	6.28	5.84	4.43	5.07	0.67
Conn	14.13	10.18	7.74	4.46	8.22	4.39	7.87	0.72
Dermatology	8.03	6.41	4.99	7.88	12.99	12.15	2.76	0.88
Ecoli	8.3	4.75	4.15	12.37	12.71	2.2	14.36	0.54
Glass	8.4	6.62	10.81	6.19	15.86	7.32	2.81	0.62
Image	11.68	11.47	13.33	14.61	7.8	5.55	1	0.45
Letter	10.53	1.88	5.94	3.24	12.81	5.99	9.73	0.64
Libras	8.43	4.21	12.86	5.56	7.14	6.4	3.72	0.56
Optdigits	9.3	4.26	13.94	15.4	14.44	8.99	11.74	0.85
Pen	8.62	12.1	13.36	11.07	12.97	15.21	15.32	0.88
Pendigits	8.52	14.07	1.57	4.15	4.53	2.76	12.4	0.79
Segment	8.82	4.55	8.28	1.42	4.45	7.47	13.45	0.51
Statlog-image	5.04	9.91	9.97	8.23	13.83	10.88	13.9	0.82
Statlog-vehicle	16.06	6.87	4.91	6.43	8.51	15.34	15.16	0.75
Steel-plates	7.82	11.75	11.98	2.56	15.65	9.39	6.54	0.9
Thyroid	1.57	11.88	13.18	2.11	8.69	14.88	2.21	0.82
Yeast	15.18	8.45	13.98	12.94	10.47	2.29	4.07	0.47

3.2 Employment Identification Recommendation Application

In order to evaluate the performance of this paper's method in real-world applications, this study compares it with the following representative baseline models: singular value decomposition matrix decomposition (SVD), Bayesian personalized ranking (BPR), and item-based collaborative filtering (Item-CF). Collaborative filtering-based baseline methods are applied to the population obtained from the clustering algorithm computed in the previous section since graduates do not have historical employment records.

Two widely used ranking-based metrics are used to evaluate all the models, namely HR and MRR.MRR average inverse rank evaluates the performance of the recommender system by considering the rank of correctly recommended items in the recommendation results.HR hit rate measures whether the actual job is included in the recommendation list. The model is mainly used to perform three recommendation tasks, namely: recommendation for graduate jobs recommendation for graduate The model is used to accomplish three recommendation tasks, namely: recommendation of graduate jobs, recommendation of graduate industries, and recommendation of graduate job types. Multiple baseline models are used to compare the performance on each of the three tasks to validate the overall performance of this paper's model on graduate job recommendations.

1) Graduate job recommendation

   In terms of recommendations for graduate jobs, this study uses four indicators, HR@20, MRR@20, HR@50 and MRR@50, and the overall performance of each model is shown in Table 3. It can be observed from the table:

   (1) The item-CF method, which relies on item similarity, performs poorly compared to most methods, with HR more than 40% lower and MRR more than 10% lower than the model in this paper.

   (2) P2CF outperforms item-CF, BPR and SVD methods. This illustrates that methods that consider users' preferences for job locations can effectively improve HR and MRR in job recommendations.

   (3) Two models, SR1-Rec and SR2-Rec, based on a single social relationship network have lower performance than this paper's model, and this comparison shows that the information obtained from social networks is not as valuable as the valuable information obtained from integrating two social networks. Therefore, this paper's model integrates more than one social network that a user uses in order to obtain a better recommendation.

   (4) As can be seen from the table, the proposed algorithm achieves the best results on the dataset, with an HR@20 of 59.31%, which is about twice that of the SVD recommendation method. Experiments verify the importance of each module in the recommender system.

Table 3. Compare with the employment recommendations of his benchmark approach

Model	HR@20 (%)	MRR@20 (%)	HR@50 (%)	MRR@50 (%)
Item-CF	12.67	36.1	8.48	31.7
SVD	30.24	10.5	19.72	8.82
BPR	6.47	6.87	21.04	7.64
P2CF	22.18	28.46	26.6	13.14
SR1-Rec	35.76	5.4	26.35	38.23
SR2-Rec	39.77	29.16	26.57	31.01
MSR-Rec	30.31	7.57	20.22	6.93
MP-Rec	19.66	27.23	35.95	25.3
This Model	59.31	48.65	61.22	58.78

2) Graduate job industry recommendation

The performance comparison for graduate job industry recommendation is shown in Table 4. Since there are 19 types of unit industries, in order to provide industry recommendations for graduates more accurately, six types of indexes, HR@1, MRR@1, HR@3, MRR@3, HR@5, MRR@5, are selected for the analysis and comparison of several models. Table 4 shows that compared to the recommendations of graduates’ jobs, the system in this paper is easier for graduates to recommend their industry, and its HR@3 and HR@5 both reach more than 91%. It’s possible that the industry chosen by graduates is influenced by their classmates

Table 4. Industry recommendation performance comparison

Model	HR@1 (%)	MRR@1 (%)	HR@3 (%)	MRR@3 (%)	HR@5 (%)	MRR@5 (%)
Item-CF	26.81	55.59	66.15	56.18	44.79	57.94
SVD	69.79	24.61	44.47	49.03	23.14	67.74
BPR	26.73	24.85	43.43	54.75	28.33	55.75
P2CF	15.01	20.08	36.15	18.39	69.6	21.58
SR1-Rec	52.39	19.68	39.41	34.52	53.18	21.3
SR2-Rec	44.01	58.47	54.66	26.07	67.86	62.97
MSR-Rec	60.1	22.25	54.75	66.67	53.14	57.36
MP-Rec	68.24	35.07	36.19	58.06	31.33	66.81
This Model	92.46	91.37	92.29	93.06	91.87	91.14

3) Graduate job type recommendation

Table 5 displays the performance comparison results for graduate job type recommendations, which include 17 types of optional job types. To provide more accurate job type recommendations for graduates, HR@1, MRR@1, HR@3, MRR@3, HR@5, and MRR@5, which are six types of indicators, are selected for the analysis and comparison of several models. In this paper, the HR index is greater than 80% when the length of the recommended list is 3 and 5, and the performance of recommending job types for graduates is better, while the HR@5 of several traditional recommendation algorithms, SVD and BPR, is below 50%, which indicates that the algorithm in this paper has obvious improvement effect compared with the traditional algorithms.

Table 5. Recommendation of type of post' performance comparison

Model	HR@1 (%)	MRR@1 (%)	HR@3 (%)	MRR@3 (%)	HR@5 (%)	MRR@5 (%)
Item-CF	20.12	52.27	27.88	24.54	12.31	33.84
SVD	27.75	26.47	40.41	49.43	26.33	45.04
BPR	45.6	43.62	46.13	19.32	4.51	12.1
P2CF	55.96	45.45	28.89	34.39	56.7	31.39
SR1-Rec	27.83	33.22	38.27	16.5	16.13	33.61
SR2-Rec	42.94	41.99	14.94	33.57	38.62	52.62
MSR-Rec	30.64	61.44	57.74	42.36	17.14	52.78
MP-Rec	47.59	40.08	13.17	23.43	23.99	34.17
This Model	83.16	87.49	81.91	86.24	86.78	82.45

4 Conclusion

In this paper, we constructed a recommendation system based on data mining to help informatize employment guidance information in colleges and universities, thereby achieving more scientific and efficient career planning guidance. This paper's algorithm demonstrates superiority over other clustering evaluation indexes, outperforming them on all terms and achieving hit rate values higher than 0.55. We again confirmed the clustering performance of this paper's model by verifying that it achieves a clustering performance of less than 0.96 on the optimal K-mean-square error, the lowest among all algorithms. The proposed algorithm is the best on the dataset, with an HR@20 of 59.31%, which is almost twice the SVD recommendation method. The experiment validated the importance of each module in the recommendation system. The system's recommendation of industries for graduates is more straightforward than that of graduate jobs, with HR@3 and HR@5 both exceeding 91%. When the HR indicator exceeds 80% in the recommended list lengths of 3 and 5, the system recommends a better job type for graduates. However, when compared to several traditional recommendation algorithms, such as SVD and BPR's HR @ 5, the difference is less than 50%, indicating that this paper's algorithm significantly outperforms these traditional algorithms.

References

[1] Yang, W. (2023). The intervention of Chaihu Hehuan Jieyu Tang combined with career planning guidance on employment anxiety patients of college students. CNS Spectrums, 28(S2), S49-S49.

[2] Udagama, P., Wijayanama, C., & Vithanapathirana, M. (2019, November). An innovation in career guidance in higher education: effectiveness and sustainability of institutionalization of service learning in the university of colombo. In 2019 From Innovation to Impact (FITI) (pp. 1-5). IEEE.

[3] Alim, S. (2023). Social Media Use in Career Guidance Delivery in Higher Education in the United Arab Emirates: A Literature Review. Research Anthology on Applying Social Networking Strategies to Classrooms and Libraries, 1933-1955.

[4] Mporananayo, N., & Andala, H. O. (2018). Career Guidance and Students' Academic Performance in Higher Learning Institutions in Rwanda. A Case Study of Nyarugenge District. IOSR Journal of Humanities and Social Science, 23(12), 72-80.

[5] Hooley, T., & Rice, S. (2019). Ensuring quality in career guidance: A critical review. British Journal of Guidance & Counselling, 47(4), 472-486.

[6] Zahid, G., Hooley, T., & Neary, S. (2020). Careers work in higher education in Pakistan: Current practice and options for the future. British Journal of Guidance & Counselling, 48(4), 443-453.

[7] Westman, S., Kauttonen, J., Klemetti, A., Korhonen, N., Manninen, M., Mononen, A., ... & Paananen, H. (2021). Artificial Intelligence for Career Guidance--Current Requirements and Prospects for the Future. IAFOR Journal of Education, 9(4), 43-62.

[8] Gati, I., Levin, N., & Landman-Tal, S. (2019). Decision-making models and career guidance. International handbook of career guidance, 115-145.

[9] Ostroga, M. M. (2024). THE USE OF DIGITAL TECHNOLOGIES IN CAREER GUIDANCE ACTIVITIES. Publishing House "Baltija Publishing".

[10] Ge, J. (2023, March). Analysis of the Innovation Mechanism and Implementation Effect of College Students' Career Guidance Courses Based on Market Demand. In The International Conference on Artificial Intelligence and Logistics Engineering (pp. 822-832). Cham: Springer Nature Switzerland.

[11] Buchanan, R. (2017). Social media and social justice in the context of career guidance: is education enough?. In Career Guidance for Social Justice (pp. 109-124). Routledge.

[12] Hidayati, L., Rizal, F., & Syahril, S. (2021). Development of android-based career guidance media. Jurnal Teknologi Informasi Dan Pendidikan, 14(2), 99-105.

[13] Trach, O. (2022). Methods of Career Guidance Activity of the Department of Higher Education Institutions in Social Networks. In SCIA (pp. 40-51).

[14] Kraatz, S., Rübner, M., & Weber, P. (2021). Career guidance in the digital context: trends in Germany. The European Centre for the Development of Vocational Training (Cedefop), 201.

[15] Pascariati, P. S., & Ali, H. (2022). Literature Review Factors Affecting Decision Making and Career Planning: Environment, Experience and Skill. Dinasti International Journal of Digital Business Management, 3(2), 219-231.

[16] Fahmi, I., & Ali, H. (2022). Determination of Career Planning and Decision Making: Analysis of Communication Skills, Motivation and Experience (Literature Review Human Resource Management). Dinasti International Journal of Management Science, 3(5), 823-835.

[17] Chetana, N., & Mohapatra, A. D. (2017). Career planning and career management as antecedents of career development: A study. Asian Journal of Management, 8(3), 614-618.

[18] Tong, D., Ma, X., Guan, P., & Kang, Z. (2020). Precision and Systematization: On the Employment Guidance Service of Graduates from the Perspective of Connotation Construction in Colleges and Universities. Adult and Higher Education, 2(1), 3-11.

[19] Luo, J. (2021). Information-based education in college students' guidance of innovation and entrepreneurship. In Frontier Computing: Proceedings of FC 2020 (pp. 1621-1627). Springer Singapore.

[20] Hu, H. (2022). Research on the innovation of college students' employment guidance work from the perspective of "Internet+". In MATEC Web of Conferences (Vol. 365, p. 01052). EDP Sciences.

[21] Yi, C. (2021, April). Under the Background of Big Data, Information-Based Teaching has Great Influence on the Development of College Students' Employability. In Journal of Physics Conference Series (Vol. 1852, No. 4, p. 042096).

[22] Li, Y., Zhong, Z., Li, S., Wu, G., & He, D. (2023). Research on the Innovative Path of Informatization Management of College Students' Employment in China. Journal of Computing and Electronic Information Management, 11(3), 80-86.

[23] Li, P., & Deng, K. (2024). Integration Path of Employment Guidance for College Students Based on Internet of Things Education. Journal of Human Resource Development, 6(2), 155-165.

[24] Mingze, T. (2021). Analysis of college students' career planning education from the perspective of Internet plus. Advances in Vocational and Technical Education, 3(2), 72-76.

[25] Liu, D., & Gao, P. (2019, April). Research on the Path of Employment Guidance Education for College Students in the New Period. In 3rd International Conference on Culture, Education and Economic Development of Modern Society (ICCESE 2019) (pp. 1259-1262). Atlantis Press.

[26] Luong Vuong Nguyen. (2024). Advanced Clustering Techniques with Bio-Inspired for Collaborative Filtering Recommendation Systems.Vietnam Journal of Computer Science(prepublish).

[27] Ramdas Vankdothu,Mohd Abdul Hameed,Raju Bhukya & Gaurav Garg. (2024). Retraction Note: Entropy and sigmoid based K-means clustering and AGWO for effective big data handling. Multimedia Tools and Applications(prepublish),1-1.

[28] Aditya Kumar & Jainath Yadav. (2024). Minimum spanning tree clustering approach for effective feature partitioning in multi-view ensemble learning. Knowledge and Information Systems(prepublish),1-29.