International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-17

https://doi.org/10.70917/ijcisim-2026-0013
Article

Design and Practice of Optimization Algorithm for College Career Planning Service Based on Cloud Computing Platform

Yadong Chai *

Student Affairs Department, Chengdu Sport University, Chengdu 610041, Sichuan, China; chaiyd@163.com

Abstract: Career planning is an important part of higher education, and effective career planning can not only help college students clarify their career goals, but also provide them with paths to realize their career goals. The article designs a cloud computing-based career planning recommendation system for college students, adopting the three-layer architecture of cloud computing, combining the student information model, the career information model and the collaborative filtering recommendation algorithm with improved CVAE and CPMF to realize the accurate matching of student and career information. In this paper, an empirical study is carried out with students from five colleges and universities. In this study, the gender difference of college students as a group has a certain impact on career, and generally speaking, the level of career planning of male students is higher than that of female students. The part-time experience is also reflected in the internship experience or not, and this aspect is the main influence on the level of career planning. The practical results show that it is feasible to carry out the career planning service for colleges and universities based on cloud computing platform.

Keywords: career planning; collaborative filtering; recommender system; cloud computing

1. Introduction

Career planning is the direction of development of a person engaged in a career, analyzing, planning, and adjusting their own careers, which can influence and determine the key factors of a person's lifestyle and ability to interact [1-2]. In career planning, it will go through the exploration period, development period and growth period. During the college years, students are in a critical stage of self-exploration, in which students discover their interests, hobbies, and work potentials and conduct a comprehensive assessment, and in conjunction with their personal career inclinations, students establish their goals and, as a result, develop an effective career plan to reach those goals [3-6].

Career planning plays a crucial role in the future employment development of college students. With the increase in the number of employed people and the changes in the job market every year, the employment situation has become more severe, which means that college students face more competitive pressure upon graduation, and the difficulty of finding an ideal job is also increasing [7-9]. In this employment environment, advance career planning can better help college students plan their future. Career planning can help college students establish correct career concepts and career ideals, help college students find career orientation to enhance employability, and help college students determine life goals to achieve life goals [10-12].

Currently, in career planning education, it does not follow the changes in the demand market to make adjustments, resulting in guidance failure. Moreover, most students carry out job search recommendations and career planning through the recommendation algorithm, and the underutilization of the employment service platform in colleges and universities is due to the fact that the school's service platform can not effectively capture the dynamic needs of students [13-14]. Cloud computing platform is a kind of Internet-based computing resource use and service management platform, which integrates computer, storage, network and other resources into a cloud computing resource pool through virtualization technology. And through systematic and automated management means, it realizes the efficient use of these resources, which can realize the fusion processing of different modal data and real-time computation, and provides assistance for students' dynamic career planning needs [15-16].

The article firstly designs a cloud computing-based career planning recommendation system for college students, and constructs the system model from the student information model, career information model and career recommendation algorithm. A recommendation algorithm based on improved CVAE and CPMF is proposed. An implicit layer is added on the basis of CVAE and random noise is introduced into the implicit layer, CVAE introduces the category of items as auxiliary information to supervise the encoding and decoding of item data, and the missing values of scoring data are reconstructed by learning the distributional features of the data. The reconstructed data is processed by CPMF algorithm. The performance of the algorithm is then tested through multiple rounds of comparison experiments. Finally, students from five colleges and universities were selected to conduct an empirical study, which introduced the practical process of college career planning based on cloud computing platform, and explored college students' career planning from the level of status quo assessment, the level of self-understanding, the level of career exploration, and the level of decision-making and action.

2. Recommendation Service System for Career Planning for Higher Education Students

2.1. Overall Architecture

Based on cloud computing to build a career planning recommendation system for college students, its architectural design strictly follows the three-layer architecture model of cloud computing, i.e., infrastructure layer, platform layer application layer. In the infrastructure layer, the AWS cloud computing platform is used to deploy the system using its elastic computing capability, and the AWS infrastructure supports availability and reliability, and can effectively utilize physical resources through virtualization technology. In the platform layer, AWS EC2 instances are used as the main computing resources, combined with RDS to manage and store structured data, such as student information, career databases, and system logs, etc., and AWS's AutoScaling and load balancing services are used to further enhance the performance and stability of the system in a highly concurrent environment. In the application layer, Web services are developed using the Django framework, which utilizes REST-fulAPI to process user requests, call the service interface provided by the platform layer, and return the results to the front-end for display [17]. Users can enter their personal information and career interests to view the recommendation results and perform related operations. The system also integrates a user feedback mechanism to continuously optimize the recommendation algorithm and service quality. Through the above three-layer architecture design, the career planning recommendation system for college students can make full use of the elastic scalability of cloud computing, big data processing capability and the ability to provide efficient services to provide accurate and efficient career planning recommendation services for college students.

2.2. System Modeling

The system model design is the core part of the career planning recommendation system for college students, which mainly includes three modules: student information model, career information model and career recommendation algorithm, and each module has different functions to realize the overall goal of the system through collaboration. The detailed design of each module and its role in the system is explained. The model design of the career planning recommendation system for college students is shown in Figure 1.

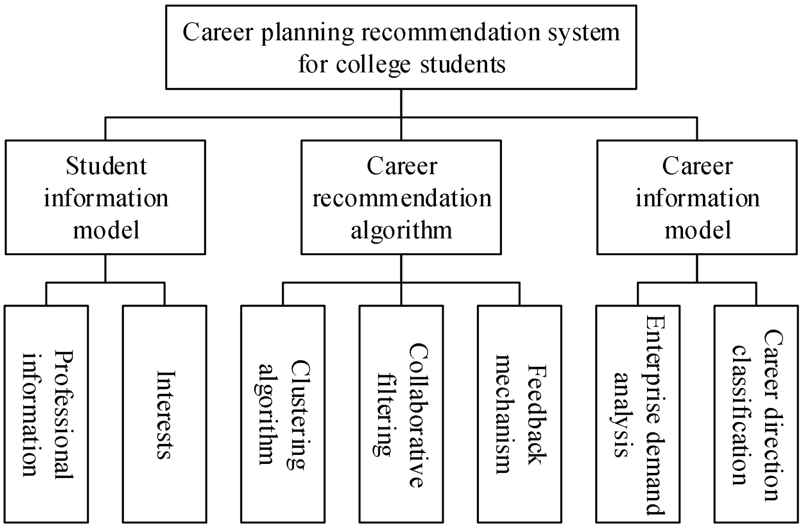

Figure 1. The design of the proposed system model of college students' career planning.

The student information model is the foundation of the entire system and is responsible for collecting and managing information about the student's various characteristics, including major information, interests and other characteristics. Major information records the student's major, course listings, and course credits, which help the system understand the student's academic background and match him/her with the right target direction in the career recommendation process. Direction of Interest records the student's interest in future careers, including inclination to specific industries or positions, which will be used as an important reference point in the recommendation process to generate more personalized career advice. Other features include students' academic performance, internship experience and other factors that can influence career choices. This information can further refine the accuracy of career recommendations and make them more in line with students' actual needs. The career information model extracts the corresponding professional skills by analyzing the demand of enterprises for talents, and divides the majors into different employment directions by combining with the professional courses set by the school to form career paths. The career information model first analyzes the needs of enterprises, extracts the key skills required in the recruitment information of enterprises, and maps these skills to the corresponding professional courses, and this process ensures that the professional knowledge learned by students matches the market demand. According to the market demand and the school's professional settings, the majors are divided into multiple employment directions, each of which corresponds to one or more career paths, in order to help students clarify their career development direction. On this basis, the requirements and development prospects of different careers are further refined to meet students' individual needs and provide them with clearer career planning suggestions. The career recommendation algorithm is the core of the system, which provides personalized career recommendations for students by matching student information with career information [18]. An adaptive K-means clustering algorithm is used to match students to the most suitable career clusters based on their feature information, and this algorithm ensures that the recommendation results are highly relevant to the students' personal background.

2.3. Feedback Mechanisms

The feedback mechanism is an important part of the career recommendation algorithm, which is able to continuously optimize the effectiveness of the algorithm through students' feedback. When a student accepts or rejects a recommended career path, the system records his/her feedback and takes it as a new data point to update the user-item matrix with the clustering data. Assuming that student u 's feedback on career cluster g is f_{ug}, the new rating matrix can be expressed as:

$$R'_{ug} = \left(1-\alpha\right)R_{ug} + \alpha f_{ug} \tag{1}$$

where α is the feedback weight, $0 \le \alpha \le 1$.

The system adjusts the clustering center and similarity calculation model according to the feedback information to further improve the accuracy of the recommendation, and the algorithm model will be re-trained after many feedbacks and data updates to ensure that the results of the career recommendation better meets the personalized needs of the students and helps them achieve the optimal career planning goals.

3. Recommendation Model for Career Planning Services in Higher Education Institutions

3.1. General Framework

The overall framework is shown in Figure 2:

(1) Improved CVAE: The improved CVAE encodes and decodes the input data under the supervision of the project auxiliary information, and can effectively alleviate the sparsity of the data by reconstructing the missing values of the scoring data.

(2) CPMF: The data processed by the improved CVAE is then submitted to the CPMF matrix decomposition process, which can optimize the feature extraction effect. Then the product of the feature matrices of the user and the item can get the user's prediction score of the item.

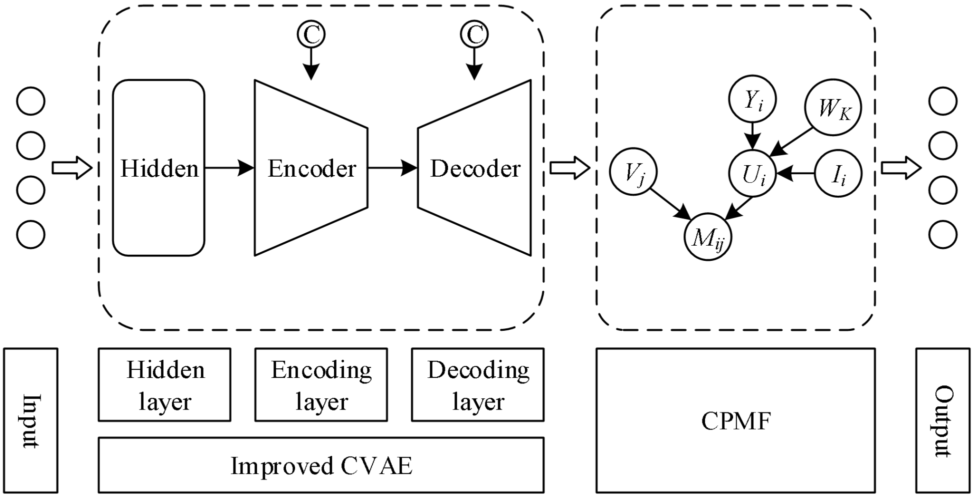

Figure 2. General framework diagram.

3.2. Improved Data Reconstruction for CVAE

The improved CVAE supervises the encoding and decoding of item data by introducing external auxiliary information in the encoding and decoding layers of the variational autocoder (VAE), which can effectively reduce the sparsity of the scoring matrix by reconstructing the missing values of the scoring data and can guide the output of the samples more efficiently through the processing of the implicit layer. The improved CVAE model is shown in Figure 3.

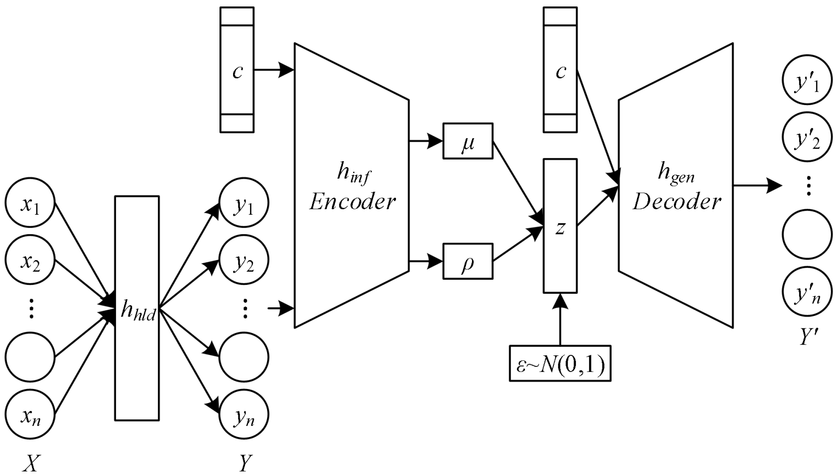

Figure 3. Improve CVAE model.

where X represents the original input data, h_{hid} represents the added implicit layer, Y represents the data processed by the implicit layer, h_{inf} represents the inference neural network of the latent spatial layer in the encoding stage, μ and ρ represent the mean and variance of the latent space distribution, z represents the implicit representation of the input data in the latent space layer, i.e., aligned to a normal distribution, h_{gen} represents the neural network that guides the encoding layer in the decoding stage to output new samples, and Y' represents the model processed the output data, and c represents the selection of item categories as auxiliary information to guide the encoding and decoding networks.

In this chapter, an implicit layer representation is added before the coding layer of CVAE, and random noise is added to the implicit layer to improve the model's anti-interference ability [19]. By processing the input data in the implicit layer, more accurate implicit features can be obtained, which can further improve the model generation capability. Assuming that the input variables are x and $x \in X$, x is processed through the implicit layer to obtain the transformed data y. Equation (2) represents the processing function of the implicit layer:

$$J_\theta(y \mid x) = f(y; x, \theta) \tag{2}$$

where f is the sigmoid nonlinear transformation function and θ is a parameter of the encoder representing random noise.

In the training stage of CVAE, $q_\phi(z \mid y, c)$ is used to denote the probability distribution of z in the hidden space obtained from the training, and $p_\theta(z \mid y, c)$ denotes the implicit probability distribution of z in the CVAE, i.e., "posterior probability", and the KL dispersion is used to denote the measure of similarity between the two, i.e.,:

$$\begin{aligned}
D_{KL}\left(q_\phi(z \mid y,c) \,\|\, p_\theta(z \mid y,c)\right) &= E_{q_\phi(z|y,c)}\left[\log \frac{q_\phi(z \mid y,c)}{p_\theta(z \mid y,c)}\right] \\
&= E_{q_\phi(z|y,c)}\left[\log \frac{q_\phi(z \mid y,c)}{p_\theta(z, y,c)}\right] + E_{q_\phi(z|y,c)}\left[\log p_\theta(y, c)\right] \\
&= E_{q_\phi(z|y,c)}\left(\log \frac{q_\phi(z \mid y,c)}{p_\theta(z, y,c)} + \log p_\theta(y, c)\right)
\end{aligned} \tag{3}$$

where $D_{KL}(q_\phi(z \mid y,c) \,\|\, p_{\boldsymbol{\theta}}(z \mid y,c))$ denotes the relative entropy between the approximate distribution and the posterior probability. A smaller relative entropy indicates a closer approximation between the approximate distribution $q_\phi(z \mid y,c)$ and the posterior probability $p_\theta(z \mid y,c)$. In order to generate true samples, it is necessary to ensure that the output sample probability is maximized as well as the relative entropy is minimized, then there are:

$$\begin{aligned}
L(\theta,\phi,y,c) &= -E_{q_\phi(z|y,c)}\left[\log \frac{q_\phi(z \mid y,c)}{p_\theta(z, y,c)}\right] \\
&= E_{q_\phi(z|y,c)}\left[\log p_\theta(z,y,c)\right] - E_{q_\phi(z|y,c)}\left[\log q_\phi(z \mid y,c)\right] \\
&= E_{q_\phi(z|y,c)}\left[\log p_\theta(y \mid z,c)\right] + E_{q_\phi(z|y,c)}\left[\log p_\theta(z,c)\right] \\
&\quad - E_{q_\phi(z|y,c)}\left[\log q_\phi(z \mid y,c)\right] \\
&= E_{q_\phi(z|y,c)}\left[\log p_\theta(y \mid z,c)\right] - D_{KL}(q_\phi(z \mid y,c) \,\|\, p_\theta(z,c))
\end{aligned} \tag{4}$$

where $L(\theta,\phi,y,c)$ represents the variational lower bound of the CVAE, and when the variational lower bound takes its maximum value, $D_{KL}(q_\phi(z \mid y,c) \,\|\, p_{\boldsymbol{\theta}}(z \mid y,c))$ is minimized. Then $L(\theta,\phi,y,c)$ is used as the optimization objective of CVAE to train CVAE.

3.3. CPMF

When dealing with a rating matrix with sparse data, the features of users with few ratings will approach the a priori mean, and the final predicted ratings will converge to the mean. Therefore, this chapter adopts CPMF to constrain users with fewer ratings when dealing with sparse matrices, which can effectively improve the accuracy of prediction and optimize the feature extraction effect. Where M represents the interaction matrix generated by users and items, U represents the user implicit spatial feature matrix, V represents the item implicit spatial feature matrix, W represents the user constraint matrix, I represents the user rating information, and Y represents the user compensation matrix.

It is known that there are N users and G items as well as the user-item interaction matrix M, the user and item implicit rating dimensions are D, and $M = U^T V$ denotes the inner product of the features of the user and item in the implicit space. The feature vector of user i in latent space can be expressed as:

$$U_i = Y_i + \frac{\sum_{k=1}^{G} I_{ik} W_k}{\sum_{k=1}^{G} I_{ik}} \tag{5}$$

where I_{ik} is the indicator function, 1 when user i has rated item k and 0 otherwise.

Assuming that the observation matrix M and the approximate rating matrix M' obey a Gaussian distribution with mean 0, the conditional distribution of the rating matrix is:

$$\begin{aligned} & p\left(M \mid Y, V, W, \sigma^2\right) \\ & = \prod_{i=1}^{N} \prod_{j=1}^{G} \left[N\left(M_{ij} \middle| g\left(\left[Y_i + \frac{\sum_{k=1}^{G} I_{ik} W_k}{\sum_{k=1}^{G} I_{ik}} \right]^T V_j, \sigma^2 \right) \right) \right]^{I_{ij}} \end{aligned} \tag{6}$$

Assuming that the potential eigenvectors of the items, the user compensation matrix, and the user constraint matrix also obey a Gaussian distribution with mean 0, we have:

$$p\left(V \mid \sigma_V^2\right) = \prod_{j=1}^{G} N\left(V_j \mid 0, \sigma_V^2 I\right) \tag{7}$$

$$p\left(Y \mid \sigma_Y^2\right) = \prod_{i=1}^{N} N\left(Y_i \mid 0, \sigma_Y^2 I\right) \tag{8}$$

$$p\left(W \mid \sigma_W^2\right) = \prod_{k=1}^{G} N\left(W_k \mid 0, \sigma_W^2 I\right) \tag{9}$$

According to the objective function formula of the probability matrix decomposition, the loss function of CPMF can be obtained as:

$$\begin{aligned} J = & \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{G} I_{ij} \left(M_{ij} - g\left(\left[Y_i + \frac{\sum_{k=1}^{G} I_{ik} W_k}{\sum_{k=1}^{G} I_{ik}} \right]^T V_j \right) \right)^2 \\ & + \frac{\lambda_Y}{2} \sum_{i=1}^{N} \left\| Y_i \right\|_{Fro}^2 + \frac{\lambda_V}{2} \sum_{j=1}^{G} \left\| V_j \right\|_{Fro}^2 + \frac{\lambda_W}{2} \sum_{k=1}^{G} \left\| W_k \right\|_{Fro}^2 \end{aligned} \tag{10}$$

where $g(x) = 1/(1+e^{-x})$, $\lambda_Y = \sigma^2 / \sigma_y^2$, $\lambda_V = \sigma^2 / \sigma_v^2$, $\lambda_W = \sigma^2 / \sigma_w^2$, and the Fro subscript represents $\|x\|_{Fro}^2 = X^T X$.

The training is performed by stochastic gradient descent optimization algorithm until convergence or maximum number of trainings. The process is as follows:

First solve for the negative gradient of the loss function:

$$\frac{\partial_J}{\partial_{Y_i}} = -\left(M_{ij} - g_{ij}\right) g'_{ij} V_j + \lambda_T Y_i = -e_{ij} g'_{ij} V_j + \lambda_T Y_i \tag{11}$$

$$\frac{\partial_J}{\partial_{V_j}} = -\left(M_{ij} - g_{ij}\right) g'_{ij} U_i + \lambda_Q V_j = -e_{ij} g'_{ij} U_i + \lambda_Q V_j \tag{12}$$

$$\frac{\partial_J}{\partial_{W_k}} = -\frac{\left(M_{ij} - g_{ij}\right) g'_{ij}}{\sum_{k=1}^{G} I_{ijk}} V_j + \lambda_W W_k = -\frac{e_{ij} g'_{ij}}{\sum_{i=1}^{G} I_{ik}} V_j + \lambda_W W_k \tag{13}$$

where $U_i = Y_i + \frac{\sum_{k=1}^{G} I_{ik} W_k}{\sum_{k=1}^{G} I_{ik}}$, $g_{ij} = g\left(U_i^T V_j\right)$, and then update the variables according to the negative gradient change:

$$Y_i = Y_i - \eta \frac{\partial_J}{\partial_{Y_i}} = Y_i + \eta\left(e_{ij} g'_{ij} V_j - \lambda_Y Y_i\right) \tag{14}$$

$$V_j = V_j - \eta \frac{\partial_J}{\partial_{V_j}} = V_j + \eta\left(e_{ij} g'_{ij} U_i - \lambda_Q V_j\right) \tag{15}$$

$$W_k = W_k - \eta \frac{\partial_J}{\partial_{W_k}} = W_k + \eta\left(-\frac{e_{ij} g'_{ij}}{\sum_{k=1}^{G} I_{ik}} Y_j - \lambda_W W_k\right) \tag{16}$$

where η is the learning rate. The final matrices U and V are obtained, and the ratings of the user item matrices are computed by means of $M = U^T V$ to calculate the predicted ratings.

3.4. Algorithmic flow

The input of the algorithm in this chapter is the user's original rating and the output is the user's predicted rating, the specific steps of the algorithm are described as follows:

Input: target user i , user rating matrix M .

Output: predicted score of target user i .

Step1: Add the rating variable x in the matrix M to the implicit layer of the formula for processing, and introduce random noise in the implicit layer for constraints to get the transformed data y .

Step2: The category of the item c is introduced into the CVAE as auxiliary information to supervise the encoding and decoding of the item data, and the transformed data y is input into the coding part of the improved CVAE for training, to obtain the probability distribution of z in the implicit space and to express the KL dispersion with respect to the posterior probability with the formula. In order to ensure that the output sample probability is maximized as well as the KL scatter is minimized, the loss function for the improved CVAE training is obtained [20].

Step3: The missing values of the scoring data are reconstructed by decoding the decoding part of the CVAE, and finally the reconstructed scoring matrix M_2 is obtained from the CVAE.

Step4: For the reconstructed matrix M_2 , initialize the user compensation matrix Y , the constraint matrix W , and the implicit spatial feature matrix of items V .

Step5: Assume that the observation matrix M_2 and the approximate scoring matrix $M2'$ obey a Gaussian distribution with mean 0 to obtain the conditional distribution of the scoring matrix.

Step6: According to the objective function of probability matrix decomposition, the loss function of CPMF can be obtained.

Step7: Train by stochastic gradient optimization algorithm, first solve the negative gradient of the loss function according to the formula, and then update the variables according to the negative gradient change.

Step8: When the model is trained to the maximum number of times or the value of the loss function no longer changes, stop training. Finally, we get the user potential spatial feature matrix U and the project potential spatial feature matrix V. The final user-item rating matrix $M3$ is output by $M = U^T V$.

4. Practice and Analysis of Career Planning Services for University Students

4.1. Algorithm Performance Test Experiment

The proportion of non-zero elements in the training set is shown in Figure 4. From the figure, it can be seen that based on this paper's algorithm than the traditional User-CF recommendation algorithm has a higher proportion of non-zero recommendation results, which to a certain extent indicates that based on this paper's algorithm to make more recommendation results.

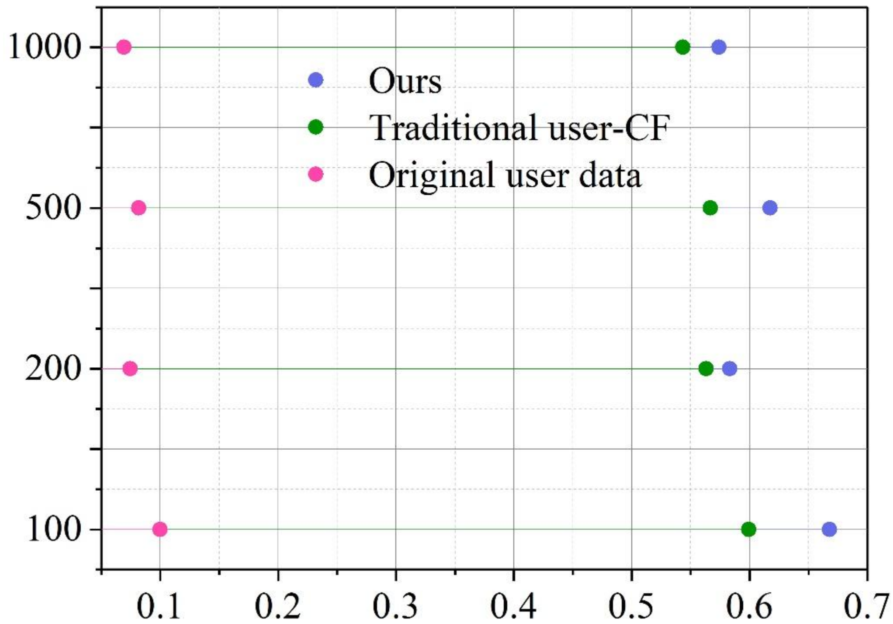

Figure 4. Training set non-zero element ratio.

The changes of mode length before and after 100 user vectors are shown in Fig. 5, and the changes of mode length before and after 200 user vectors are shown in Fig. 6. Among them, the larger the value of user vector mode length represents to a certain extent the greater the number of products evaluated by the user, indicating that the user is more active. From the curve of these two graphs can be seen from, based on the recommendation algorithm in this paper can be made for active users to make effective recommendations, the recommendation results produced by most of the users of the vector mode length has increased by a certain percentage, so as to give the active user to make more recommendations.

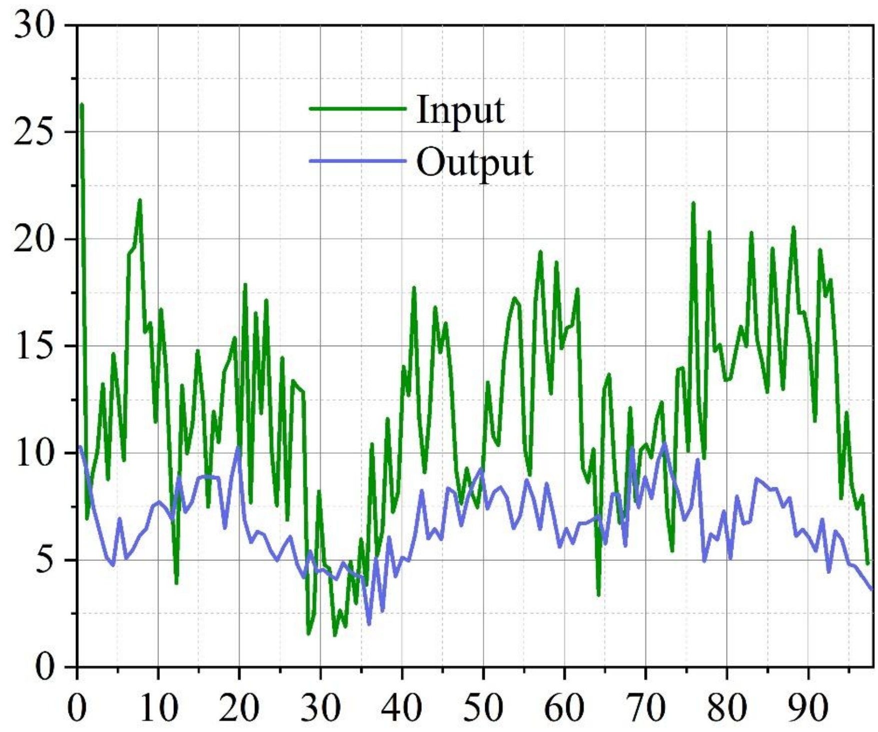

Figure 5. The 100 user vectors are recommended to change the pattern.

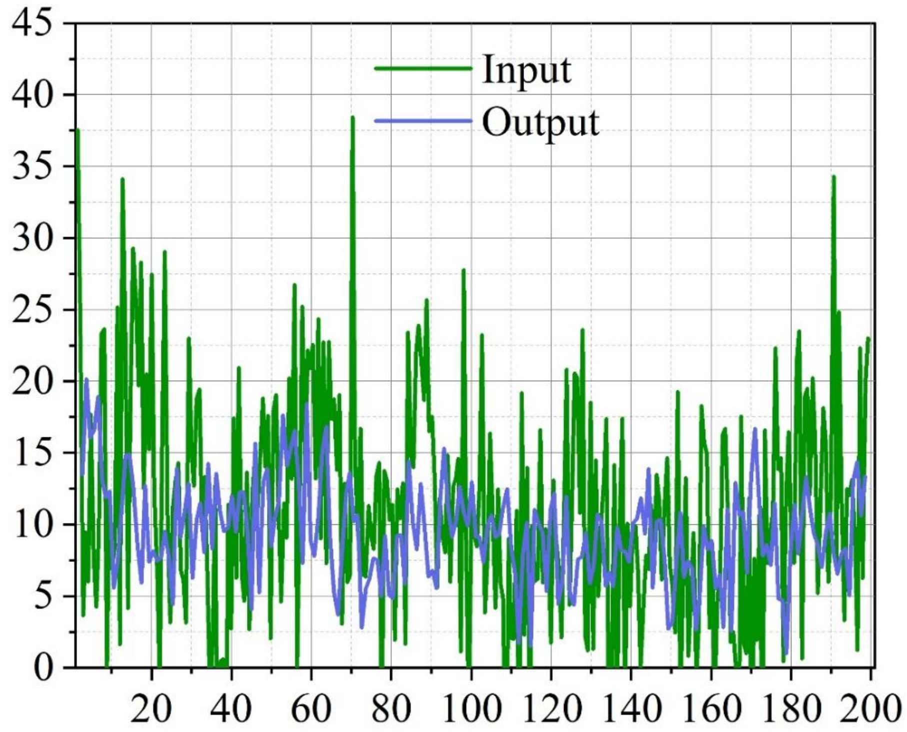

Figure 6. The 200 user vectors are recommended to change the pattern.

The comparison curves of accuracy, recall and coverage of the two algorithms are shown in Fig. 7 to Fig. 9. From the data in the figure, it can be seen that when the K value is selected as 80, the accuracy and recall rate reach the highest, which are 25.22% and 12.13% respectively. And the coverage rate shows a decreasing trend with the increasing K value. It can be seen that the recommendation accuracy and recall of the algorithm in this paper are not linearly related to the K value. In the dataset, selecting around K=80 will get a relatively high accuracy and recall rate. And due to the increase in the number of users with similar interests in reference, the coverage rate of the recommended list will be reduced instead, so the coverage rate and the value of K have an inverse correlation. From the figure, it can be seen that the algorithm based on this paper has a certain improvement in accuracy and recall compared to the traditional User-CF recommender system. Due to the balanced K value selection factor, the coverage rate of this method reaches the average level of traditional User-CF, i.e., around 25%.

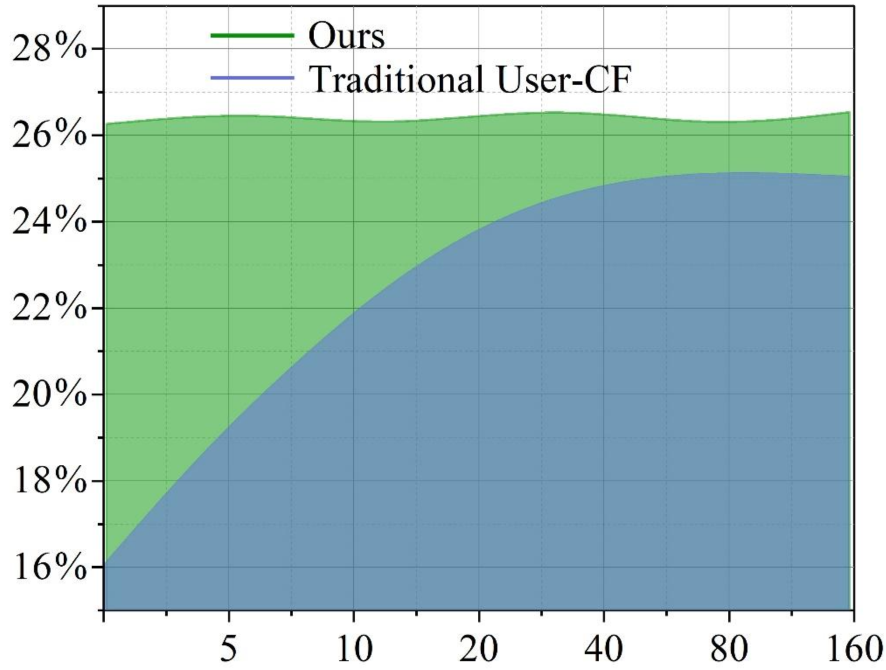

Figure 7. Two algorithms accuracy comparison curves.

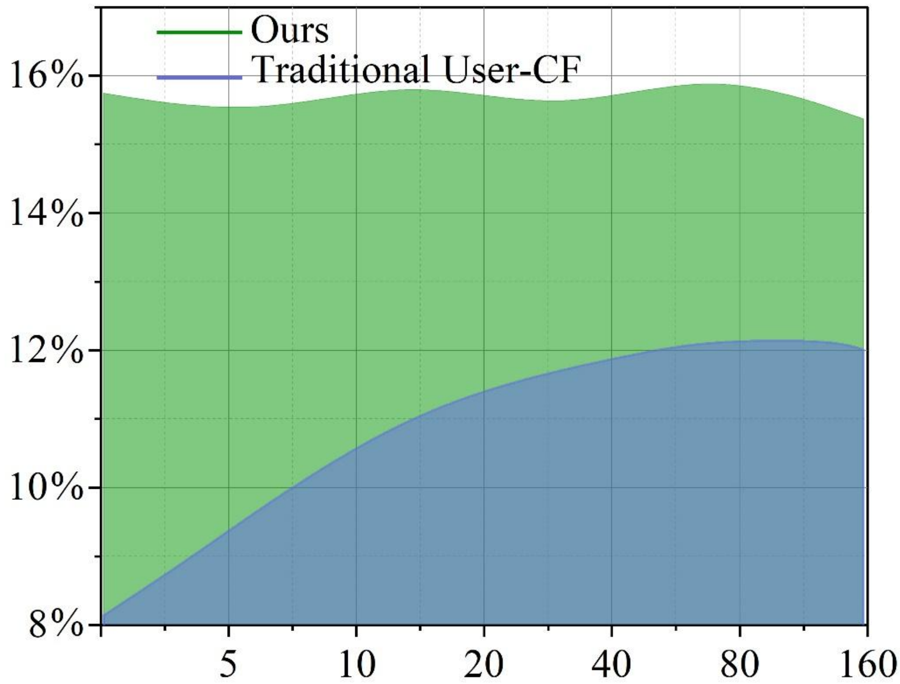

Figure 8. Two algorithms recall ratio curves.

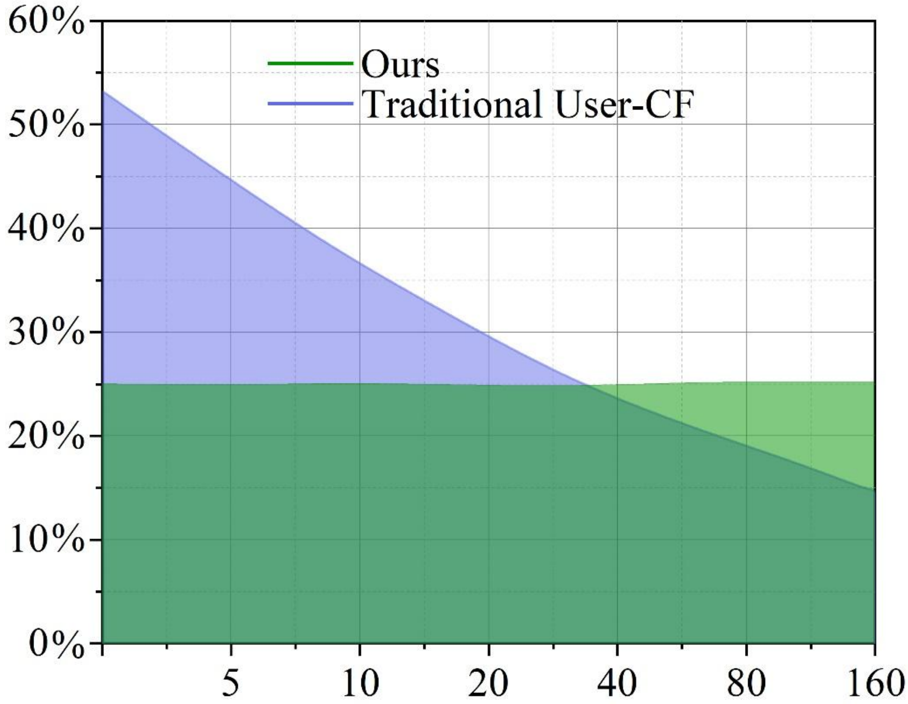

Figure 9. Two algorithm coverage curves.

4.2. Investigation of the Use of Cloud Computing Platforms in the Careers of College and University Students

4.2.1. Background to the Investigation

By investigating the issues related to career planning of college students, this paper aims to understand the application of the cloud computing platform based on this paper in the career planning of college students, to explore the differences in the career planning of college students from different backgrounds, and to discover the problems and reasons for the career planning of college students. To ensure the representativeness of the survey results, college students from five different types of colleges and universities, namely, engineering, liberal arts, teacher training, agriculture, and language, were selected as subjects for the questionnaire survey. In the selection of majors, the advantageous and characteristic disciplines of each university were considered and representative majors were selected as much as possible. The sampling method is stratified based on schools, grades, majors and so on. The questionnaire survey is mainly distributed and collected by mail, and the author contacted the teachers and students of the participating universities and asked them to help distribute and collect the questionnaires in their respective universities under the premise of explaining the precautions of the questionnaire survey to them. The time period of the questionnaire survey is from September to December 2024.

4.2.2. Pre-Test Results and Analysis

(1) Subjects

A total of 3205 questionnaires were distributed in the pre-test and 3000 were returned with a recovery rate of 93.6%. The pre-test sample is shown in Table 1.

Table 1. Presample Case.

Attribute	Specific Situation	Number	Percentage
School	School A	762	23.78
	School B	729	22.75
	School C	862	26.90
	School D	695	21.68
	School E	157	4.90
Gender	Man	1625	50.70

	Female	1580	49.30
Grade	Freshman year	806	25.15
	sophomore	911	28.42
	junior	639	19.94
	Senior year	849	26.49

(2) Exploratory factor analysis

The KMO test and Bartlett's spherical test are used to determine whether the sample data are suitable for exploratory factor analysis. When the value of KMO is larger, it indicates that there are more common factors among the variables, and it is more suitable for factor analysis.Generally speaking, if the value of KMO is less than 0.5, it is less suitable for factor analysis, and the common guideline for factor analysis is at least 0.6.The results of the KMO and Bartlett's test are shown in Table 2. As shown in the above table, the value of this KMO test is 0.855, which indicates that there are common factors between the topics and it is suitable for exploratory factor analysis. In addition, the chi-square value ($\chi 2$) of Bartlett's test of sphericity is 3505.422 with a degree of freedom of 300, and its probability of companionship is 0.000, which is less than the significance level of 0.05, indicating that the data are statistically significant. The results of the above two tests indicate that the data are suitable for exploratory factor analysis.

Table 2. The results of KMO and Bartlett.

Inspection index		Test result
KMO		0.855
Bartlett ball test	Approx.Chi-Squae	3505.422
	df	300
	P.	0.000

4.3. Analysis of the Current Situation of College Students' Career Planning

4.3.1. Analysis of the Overall Status of Career Planning for University Students

The overall status is shown in Table 3. As can be seen from the table, the calculated self-assessment level of the students' current situation scored an average of 31.46 points, which is higher than the average of 30 points corresponding to the “conformity”, indicating that the students have a relatively good understanding of career planning. Students' self-understanding level (including: interests, personality, values, skills) averaged 29.38 points, although already in the second place, but much lower than the average score of 33 points to achieve the medium level of “conformity”, indicating that although the students have a better awareness of career planning, but there is some confusion in the planning behaviors, which is derived from their own understanding of what they really want and for what. Finally, we see that the average score of career exploration and decision-making in the table is not more than 30, which is lower than the average score, indicating that the students are “all talk and no action”, lack of action, only aware of its importance, but lack of a decisive career line, and lack of a clear career goal, so no concrete action. In addition, the skewness and kurtosis indicators given in the table show that the distribution at all levels is skewed to the right.

Table 3. Overall condition.

	Sample		Mean Value	Standard Deviation	Degree Of Bias		Kurtosis	
	In Effect	Ineffectiveness	Numerical Value	Numerical Value	Numerical Value	Standard Error	Numerical Value	Standard Error
Status Quo	3205	0	31.46	5.546	0.656	0.105	0.368	0.205
Understand	3205	0	14.6	3.202	0.522	0.102	1.156	0.205
Explore	3202	3	29.38	5.608	0.417	0.105	1.028	0.208
Decision Making	3202	3	14.79	3.566	0.22	0.105	0.209	0.205

4.3.2. Differential Analysis of College Students' Career Planning

(1) Gender

As also mentioned above, gender is usually used as a salient variable when conducting many studies on human nature, as men and women differ very significantly in many aspects of their performance. The gender variance analysis is shown in Table 4. In this study, from the table, it can be found that the p-value of the t-test of the total table is 0.007, which is a significant difference. Also from the numerical scores, the average score of male students' career planning level is higher than the average score of female students. That is, there is a significant difference between male and female university students in all levels of career planning. In these four levels, numerically only the self-knowledge level is not a significant difference between male and female scores, in the rest of the levels except it: assessment of the current situation, career exploration, and decision-making actions, there is a significant difference between male and female students, and this difference for male students is generally higher than that of female students.

Table 4. Gender difference analysis.

	Gender	Sample		Mean value	Standard deviation	T test	Significance
		In effect	ineffectiveness	Numerical value	Numerical value	Numerical value	Numerical value
Status assessment	Man	1055	0	32.89	5.507	2.334	0.019*
	Female	2150	0	31.96	5.339		
self-understanding	Man	1054	1	17.88	3.163	2.352	0.019*
	Female	2150	0	16.64	2.992		
Career quest	Man	1055	0	27.19	5.845	1.887	0.053
	Female	2149	1	25.19	5.209		
Decision action	Man	1055	0	16.37	3.638	2.386	0.012*
	Female	2150	0	15.14	3.52		
Inventory of total	Man	1054	1	97.17	15.017	2.823	0.007**
	Female	2147	3	93.44	14.433		

(2) Specialization

Since the topic of this study is career planning for college students, it is natural that majors are worth discussing for college students, and the difference analysis of majors is shown in Table 5. The table yields a p-value of 0.691 for the total table t-test, which is not significant, indicating that the differences in majors (which are only divided into arts and sciences to be discussed here) do not lead to various differences in the level of career planning among college students.

Table 5. Analysis of the difference of specialty.

	Majors	Sample		Mean value	Standard deviation	T test	Significance
		In effect	ineffectiveness	Numerical value	Numerical value	Numerical value	Numerical value
Status assessment	Liberal arts	1936	0	31.8	5.628	-0.141	0.887
	Science	1269	0	31.64	5.526		
self-understandi	Liberal	1936	0	16.99	3.152	-0.127	0.898

ng	arts						
	Science	1269	0	16.75	3.376		
Career quest	Liberal arts	1936	0	27.22	5.985	-0.233	0.814
	Science	1268	1	27.2	5.182		
Decision action	Liberal arts	1935	1	16.12	3.505	-0.481	0.619
	Science	1269	0	16.32	3.637		
Inventory of total	Liberal arts	1935	1	94.59	14.979	-0.403	0.691
	Science	1267	2	95.08	14.233		

(3) Part-time and internship experiences

The difference analysis is shown in Table 6. Part-time jobs and internships are a way for school students to get in touch with the society and are necessary. It should be advocated in large quantities under the premise of ensuring safety. The t-test p-value for the total table is 0.008, a significant difference. In addition, it is also highlighted that in the three aspects of current situation assessment, self-understanding and career exploration, the P-value of t-test is 0.006<0.01, 0.005<0.01, 0.012<0.05 respectively, which indicates that students will participate in part-time jobs and internships more often is very helpful to students' personal employment and career planning, so that the students who have internships and part-time jobs experiences are bound to be stronger than the ones who do not have these experiences in many aspects. Students.

Table 6. Variance analysis.

	Part-time experience	Sample		Mean value	Standard deviation	T test	Significance
		In effect	ineffectiveness	Numerical value	Numerical value	Numerical value	Numerical value
Status assessment	Yes	2826	0	32.12	5.001	2.62	0.006**
	No	379	0	31.43	5.546		
self-understanding	Yes	2826	0	17.6	3.238	3.011	0.005**
	No	378	1	16.04	3.151		
Career quest	Yes	2826	0	27.41	5.197	2.564	0.012*
	No	377	2	16.79	5.332		
Decision action	Yes	2826	0	16.04	3.418	0.121	0.903
	No	379	0	17.25	3.499		
Inventory of total	Yes	2826	0	95.19	14.76	2.612	0.008**
	No	372	7	89.9	14.381		

4.3.3. Cluster Analysis

Each student has his own type and characteristics, in the career planning of college students to distinguish between different types of students, so that “the right medicine” is necessary. If at the beginning of the different types of students to adopt the training methods that are not suitable for him, but easily counterproductive, once wrong, will be in the future students in the graduation of normal employment will bring difficulties, especially some of the technical work, so in this study will still be more than 3,000 samples of students to carry out a categorization process. Before clustering and classification analysis, preliminary statistical analysis will be carried out, respectively, the students will be summarized in a simple list according to gender, major, part-time internship experience using descriptive statistical analysis, so that we can have a preliminary understanding of the situation of the sample personnel: next, we will use K-means and pam algorithms respectively for cluster analysis of the students, to dig out which are the key training groups, medium training groups, general training groups, and comparative analysis. K-means and pam algorithms will be used to analyze the students' clustering results to find out which are the key training group, medium training group, general training group, and to compare and analyze the clustering results of the two algorithms, and finally select one clustering algorithm result for subsequent analysis. According to the comparison of the two algorithms, the K-means algorithm produces a smaller difference in the size of the classes than the K-center point algorithm, but it is more sensitive to dirty data. pam algorithm is more robust here than the K-means algorithm, and it is able to better deal with the different types of data points, so it can be seen here that the pam algorithm is more optimized, which means that the students of this school need to be focused on the training of the career plan There are about 1,820 students, 1,362 students need medium training, and only 211 students need general training, that is to say, more than 93.78% of the students are in great need of quality career planning training, that is to say, the optimization mode of career planning service in colleges and universities based on cloud computing platform is feasible.

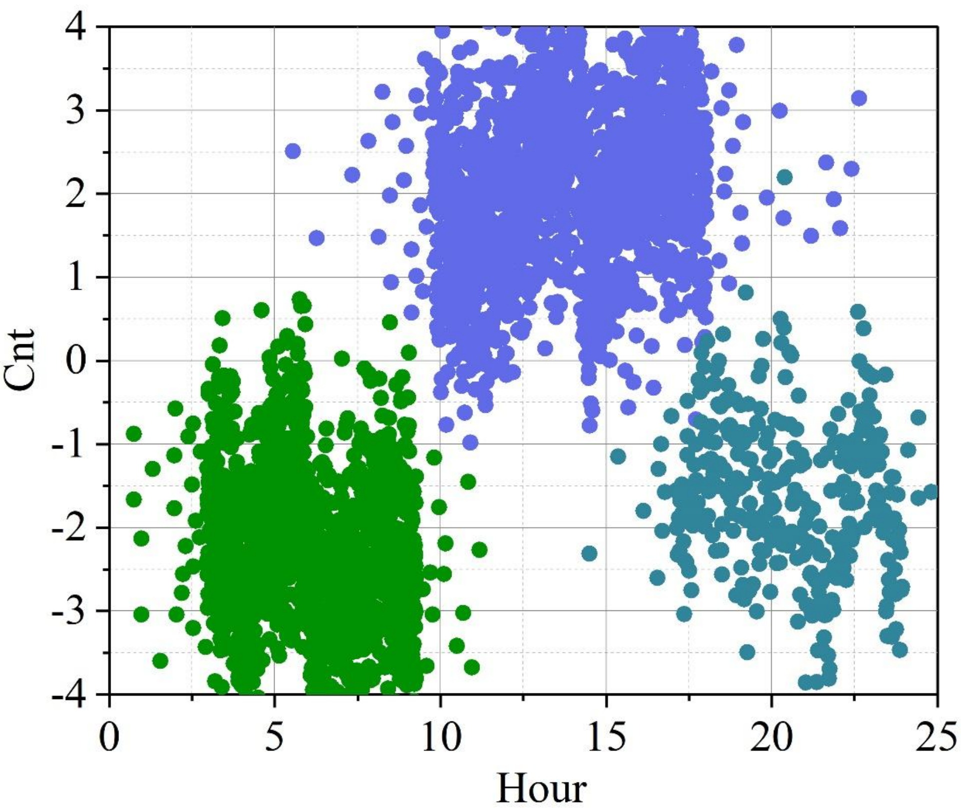

Figure 10. K-means clustering effect diagram.

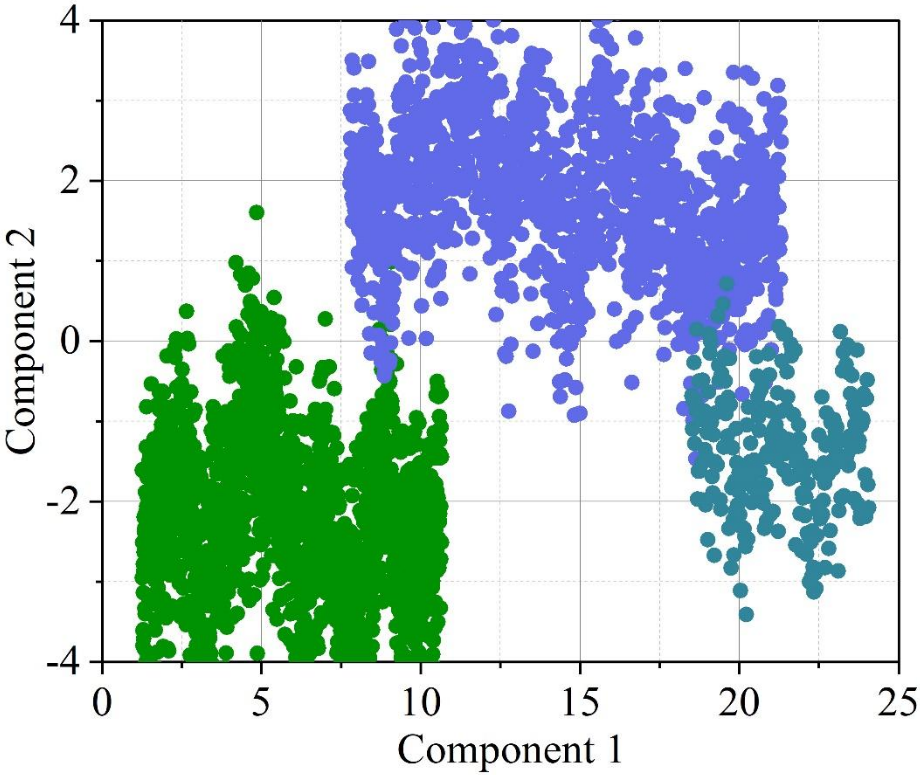

Figure 11. PAM algorithm clustering effect diagram.

5. Conclusion

The article designs a cloud computing-based career planning recommendation system for college students, accurately matches student and career information through improved collaborative filtering algorithms, and processes student data and industry trends in real time, in order to provide college students with career planning suggestions with high accuracy. The main conclusions of the article are as follows:

The algorithm in this paper has improved the accuracy and recall rate to a certain extent compared with the traditional User-CF recommender system, and the coverage rate reaches the average level of the traditional User-CF, i.e., it is around 25%.

By applying the method proposed in this paper to cluster analysis of more than 3,000 sample students, it is concluded that more than 93.78% of the students are in need of quality career planning training, thus indicating that the optimization mode of career planning service for colleges and universities based on cloud computing platform is feasible.

References

1. 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.
2. Sinsky, C. A., Dyrbye, L. N., West, C. P., Satele, D., Tutty, M., & Shanafelt, T. D. (2017, November). Professional satisfaction and the career plans of US physicians. In Mayo Clinic Proceedings (Vol. 92, No. 11, pp. 1625-1635). Elsevier.
3. 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.
4. Hirschi, A., & Koen, J. (2021). Contemporary career orientations and career self-management: A review and integration. Journal of Vocational Behavior, 126, 103505.
5. Quinlan, K. M., & Renninger, K. A. (2022). Rethinking employability: how students build on interest in a subject to plan a career. Higher Education, 84(4), 863-883.
6. Reardon, R. C. (2017). Enhancing self-help career planning using theory-based tools. Journal of Career Assessment, 25(4), 650-669.
7. Jackson, D., & Tomlinson, M. (2020). Investigating the relationship between career planning, proactivity and employability perceptions among higher education students in uncertain labour market conditions. Higher education, 80(3), 435-455.

8. He, H. (2023). The Practical Difficulties and Implementation Paths of Career Development Planning for College Students in the New Era. Advances in Vocational and Technical Education, 5(8), 59-64.
9. Hordósy, R., & Clark, T. (2018). 'It's scary and it's big, and there's no job security': Undergraduate experiences of career planning and stratification in an English red brick university. Social Sciences, 7(10), 173.
10. Serbes, M., & Albay, M. (2017). Importance of career planning and development in education. International Journal of Social Sciences & Educational Studies, 4(2), 149-154.
11. Teychenne, M., Parker, K., Teychenne, D., Sahlqvist, S., Macfarlane, S., & Costigan, S. (2019). A pre-post evaluation of an online career planning module on university students' career adaptability. Journal of Teaching and Learning for Graduate Employability, 10(1), 42-55.
12. Han, Y. O. (2024). Analysis of College Students' Life Planning and Career Class Effectiveness. The Journal of the Convergence on Culture Technology, 10(3), 37-45.
13. Gu, X. (2022). Effect of deep learning on college students' career planning. Mathematical Problems in Engineering, 2022(1), 1573635.
14. Gu, L., Zhou, Z., & Zhu, Y. (2018). Design and Implementation of Employment Recommendation Service Platform for College Graduates. Educational Sciences: Theory & Practice, 18(5).
15. Ozyurt, O., Gurcan, F., Dalveren, G. G. M., & Derawi, M. (2022). Career in cloud computing: exploratory analysis of in-demand competency areas and skill sets. Applied Sciences, 12(19), 9787.
16. Hu, L., & Yang, F. (2023). Design of a Cloud Computing Based Employment Platform for College Students. vol, 4, 131-140.
17. Yang Jiang. (2025). Design of dynamic career path recommendation system based on Markov decision process. Journal of Computational Methods in Sciences and Engineering,25(3),2427-2441.
18. Yizhou Zhou,Yong Zhang,SijiaYu& Naijie Liu. (2024). Research on the Design of an AI Career Path Recommendation System Based on MBTI from a Cross-Cultural Perspective. Artificial Intelligence Technology Research,2(1).
19. Tianyu Wang & Dong Ge. (2025). Research on Recommendation System of Online Chinese Learning Resources Based on Multiple Collaborative Filtering Algorithms (RSOCLR). International Journal of Human–Computer Interaction,41(3),1771-1781.
20. Nan Xiang & Kayo kanato Xiaolan. (2022). Design and Implementation of a Personalized Tourism Recommendation System Based on the Data Mining and Collaborative Filtering Algorithm. Computational Intelligence and Neuroscience,2022,1424097-1424097.