Journal of Network Intelligence
Taiwan Ubiquitous Information

Volume **10**, Number **2**, May **2025**

Employment Recommendation Method for College Students to Solve the Cold Start Problem: Affinity Propagation Clustering Algorithm

Ya Luo*

Chengdu Textile College, Chengdu 611731, P. R. China
34739142@qq.com

Di Luo

Kmart Australia Ltd., Mulgrave VIC 3170, Australia
luodi820@outlook.com

*Corresponding author: Ya Luo

Received May 8, 2024; revised September 25, 2024; accepted January 11, 2025.

Abstract. *Recently, the overall competition in the job-seeking market has shown a trend of strengthening, and university graduates are facing serious employment problems. Therefore, recommending the most suitable jobs for them is a significant current research direction. This article designs a college student employment recommendation method relied on the Affinity Propagation (AP) clustering algorithm for the problems of cold start and low accuracy of the existing research. Firstly, the AP algorithm is optimized based on the neighborhood density factor and the Mahalanobis distance to reduce the mutual interference between samples due to the influence of the number of samples. Secondly, the behavioral data of students are analyzed and counted to construct the scoring matrix of students and derive the interest preference, and then the background information is added to the basic information of students, and the optimized AP algorithm is adopted to divide the students into clusters with different backgrounds, and the job preference of each cluster is basically the same, which addresses the cold-start issue. Then the behavioral similarity between students and collaborative filtering algorithm are combined to populate the scoring data, and finally the scoring data are calculated and sorted according to the timeliness of the jobs, and TOP-N employment information is selected for group recommendation. The simulation outcome implies that compared with the existing methods, the suggested method not only improves the Hit Rate (HR) and the Mean Reciprocal Rank (MRR), but also enhances the accuracy of the jobs, which can offer a better recommendation experience for students.*
Keywords: Employment recommendation; Affinity propagation clustering; Neighborhood density factor; Marginal distance; Collaborative filtering

---

1. **Introduction.** As the Internet recruitment and the arrival of the big data era rapidly developing, the continuous growth of job information has brought the issue of information overload for college students' employment choice [1, 2]. College students are often tough to select in the face of the huge amount of job information, and they do not have a clear understanding of their own skills corresponding to which employment positions. In the face of this troubled employment phenomenon, there is a lack of a communication hub between college students and recruiting companies [3]. Recruiting enterprises use qualitative job requirements to recruit talents. While college students study according to

the courses carried out by colleges and universities, they lack knowledge of the positions of interest, and do not study deeply in a wide range of courses, resulting in the lack of more specialized knowledge and skill sets, which can not meet the requirements of recruiting enterprises, and the enterprises are unable to find college students who match their needs quickly [4, 5]. There is a serious information asymmetry problem between college students and recruiting enterprises, therefore, there is an urgent need to explore and study the characteristics of talents and positions to provide effective recommendations for college students' employment [6].

1.1. **Related work.** Hwang et al. [7] extracted features from behavioral sequence attributes and company description attributes of graduates, and made job recommendations through feature interaction and deep data fitting, but suffered from data sparsity. Liu et al. [8] recommended the most suitable jobs for college students relied on their job resumes and job postings. Hamzah et al. [9] established a fuzzy rule, which was based on the professional skill requirements of the jobs, and was adopted to measure the suitability of college students for the positions. Ochirbat et al. [10] found clusters of jobs that were of interest to students by collecting students' behaviors and formed a recommendation list. Mpia et al. [11] used a BERT model to capture semantic interactions between resumes and job descriptions, which work together to make job recommendations. Zhu et al. [12] offered a feature fusion-based representation learning model to predict the match scores between jobs and resumes, and utilized a bilinear module to obtain job and resume matches, but suffered from a cold-start issue.

Samin and Azim [13] used college students' skills as embedded features and fused machine learning and non-machine learning approaches to achieve better recommendation results. Daga et al. [14] used an artificial neural network based on the social networks of college students to determine whether a job opportunity should be recommended or not. Zhao et al. [15] recommended jobs for graduates through campus records, while introducing a Bayesian personalized ranking algorithm to calculate graduates' ratings of employment units, combining the ratings with personalized preferences for job recommendation. Mishra et al. [16] on the ground of decision tree employment portrait association rules, but the recommendation accuracy is not high.

To deal with the issues of sparsity of scoring matrix data and low recommendation accuracy. Huang and Liu [17] suggested a college student employment recommendation algorithm relied on a fuzzy clustering with singular value decomposition. Ouyang et al. [18] used the category information of items to design a recommendation algorithm relied on k-means clustering to mine the clustering information at the level of college students' interests and preferences. Wang et al. [19] proposed a collaborative filtering recommendation algorithm incorporating college students' interest preferences and clustering, by which the problem of data sparsity can be alleviated. However, for newly released positions or emerging industries, the recommendation system may not have enough historical data to evaluate the popularity of these positions or their matching with users, which makes it difficult to recommend these new positions to users who may be interested. Kumalasari and Susanto [20] clustered college students based on their preferences and decomposed the traditional rating data to obtain a brand-new rating matrix. In this way, the nearest neighbors of users are calculated to ensure the accuracy of recommendations. However, when new users (such as fresh graduates) join the recommendation system, the system lacks behavioral data about these new users, so it is difficult to accurately predict their interests and preferences, and thus it is difficult to provide personalized employment recommendations.

1.2. **Contribution.** It is clear from the analysis that current employment recommendation methods generally have the issues of cold start and unsatisfactory recommendation accuracy. This article suggests an employment recommendation method for college students based on Affinity Propagation (AP) clustering to address the above issues. Firstly, the AP clustering algorithm is optimized to replace the Euclidean distance with the Mahalanobis distance, and the neighborhood density factor is introduced into the nearest-neighbor propagation, so that the similarity matrix can more accurately reflect the relationship between the data. Then, statistically analyzing the students' behavioral data in Internet recruitment, construct the students' scoring matrix and derive the students' preferences, then add the background information to the basic information of the job-seeking students, cluster the background information of the job-seeking students by using the optimized AP clustering algorithm, divide them into different student clusters, and then fill in the scoring data by combining the behavioral similarity among the students and the collaborative filtering algorithm, finally, the scoring data are filled in according to the timeliness of the job requirements. Finally, the scoring data is calculated according to the timeliness of job requirements, and the results are used as the basis for sorting and selecting the TOP-N employment information for group recommendation.

2. **Relevant theory and technology.**

2.1. **Affinity propagation clustering algorithm.** The AP algorithm is a clustering method of mutual information transfer between nearest neighbors [21], which has the advantages of dealing with missing data automatically and good clustering effect compared to algorithms such as k-means clustering and fuzzy clustering. The similarity matrix S of the whole sample set is constructed by calculating the similarity s between pairs of samples, and then the attraction g and attribution b are calculated in an iterative manner, and the termination condition of the algorithm is that the number of iterations reaches the preset upper limit, or the class representative points do not change after many iterations. The steps in detail of AP algorithm are implied as follows.

(1) Construct the similarity matrix S. Use Equation (1) to compute the similarity value $s(i,j)$ between x_i and x_j, and then complete the construction of sample set similarity matrix S.

$$s(i,j) = \begin{cases} -\|x_i - x_j\| & i \neq j \\ p(j) & i = j \end{cases} \tag{1}$$

(2) Transferring information. Update g and b. The information transfer continues through an iterative loop that alternates attraction and attribution values to produce high-quality class representatives. Where $g(i,j)$ denotes the attractiveness of x_j to x_i, and $b(i,j)$ denotes the degree of belonging of x_i to x_j. The updating methods for $g(i,j)$ and $b(i,j)$ are as bellow.

$$g(i,j) = s(i,j) - \max_{j' \neq j}\{b(i,j') + s(i,j')\} \tag{2}$$

$$b(i,j) = \begin{cases} \min\{0, g(j,j) + \sum_{i' \notin \{i,j\}} \max[0, g(i',j)]\} & i \neq j \\ \sum_{i' \neq j} \max[0, g(i',j)] & i = j \end{cases} \tag{3}$$

Since there is a certain amount of oscillation in the above information transfer process, it is necessary to introduce a damping factor μ, usually $\mu \in [0,1]$, which serves to correct the $g(i,j)$ and $b(i,j)$ of the 2 iterations before and after. Let the number of iterations be r. The corrected information transfer process is implied below.

$$g(i,j)^{r+1} = (1-\mu) \times g(i,j)^{r+1} + \mu \times g(i,j)^r \tag{4}$$

$$b(i,j)^{r+1} = (1-\mu) \times b(i,j)^{r+1} + \mu \times b(i,j)^r \tag{5}$$

where $g(i,j)^r$ and $b(i,j)^r$ denote the attraction and attribution of the r-th iteration, respectively.

(3) Determine the class representative point. Select the sample x_j that satisfies the maximum sum of $g(i,j)$ and $b(i,j)$ as the class representative point of x_i. The conditions satisfied by j are as follows.

$$j = \arg\max\{b(i,j) + g(i,j)\} \tag{6}$$

As indicated in Figure 1, the AP clustering algorithm optimizes the clustering results by propagating and maximizing the information relied on the factor graph.

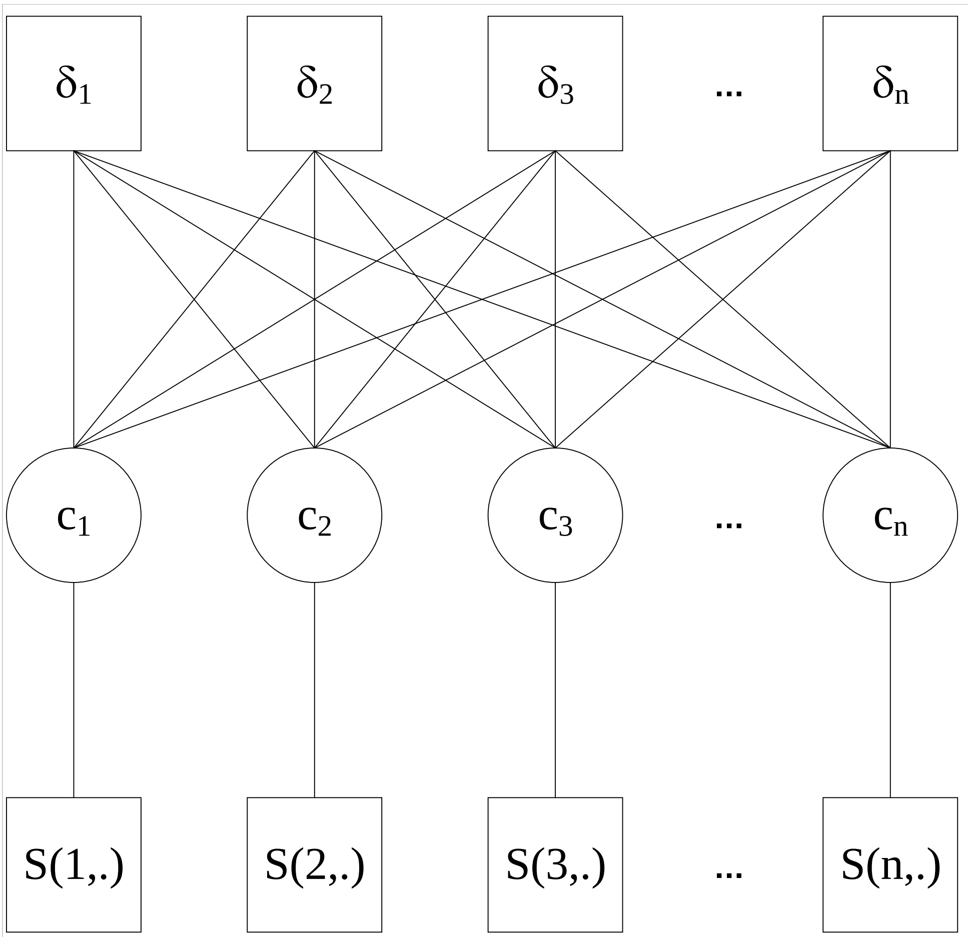

Figure 1. AP clustering algorithm

2.2. **Collaborative filtering techniques.** Collaborative filtering techniques are chiefly adopted to make recommendations relied on similar users or items. The similarity between different users and posts can be realized by various similarity formulas (e.g., cosine similarity) [22].

(1) Content-based collaborative filtering technique [23]. The basic idea of this algorithm is to annotate the feature information of users and items, extract the keywords of users or items as annotations, and then calculate the similarity between the annotations or

personalized recommendation through feature matching, the process in detail is implied in Figure 2.

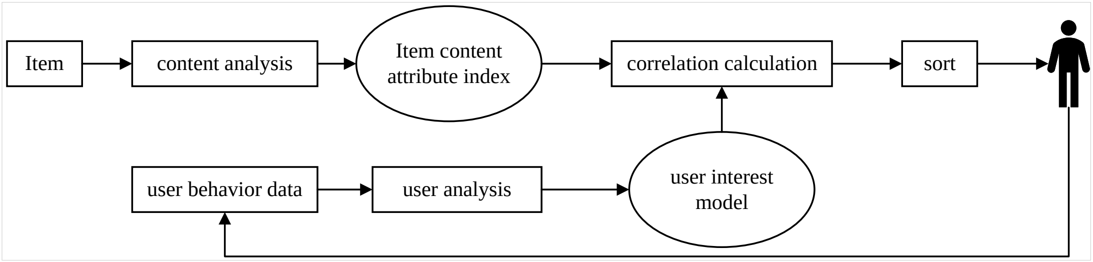

Figure 2. Content-based collaborative filtering

(2) User-based collaborative filtering technique [24]. Neighboring users are found based on their preferences for items and then the items preferred by the neighboring users are recommended to the current user. The neighboring users are decided based on the similarity between users' preferences for items, where users' preferences for items are historical records.

3. **Optimization of AP clustering algorithm based on neighborhood density factor and marginal distance.** Intending to the limitations of the distance metric of AP clustering algorithm and the low accuracy of the clustering algorithm for large-scale datasets, an AP clustering algorithm relied on the neighborhood density factor and the Mahalanobis distance is offered. In the similarity measure of the samples, Mahalanobis distance [25] replaces the Euclidean distance to reduce the mutual interference between samples due to the influence of the number of samples; the idea of neighborhood density factor [26] is introduced into the AP and combined with the pairwise constraints to improve the similarity of the data, making the similarity matrix can more accurately reflect the relationship between the data.

Assuming that the input dataset is $X = \{x^i\}(i = 1, 2, \dots, m)$, initialize the parameters in the computation process and compute the similarity matrix $[s(i, j)]_{(n^2-n)\times 3}$. The equation for the inter-sample similarity is as follows.

$$s(i, j) = (x_i - x_j) \sum^{-1} (x_i - x_j)^T \tag{7}$$

The similarity between the data is first adjusted using the known constraint information. If two samples come from known constraint pairs, and these two samples come from the same category, i.e., their similarity is 0, then these two samples are added to the Mustlink set. If these two samples come from different categories, i.e., their similarity is $-\infty$, then these two samples are added to the Cannot-link set.

$$\begin{cases} s(i, j) = 0 \quad \& \quad s(j, i) = 0, \quad \{(x_i, x_j)\} \in M \\ s(i, j) = -\infty \quad \& \quad s(j, i) = -\infty, \quad \{(x_i, x_j)\} \in C \end{cases} \tag{8}$$

Then the similarity between other samples is adjusted. Extend the known Must-link set to get a new Must-link set, and add the new Must-link set to the Must-link set. If there is a pair of samples that does not belong to the Must-link set, the pair of samples can be grouped into the Must-link set.

$$\begin{aligned} &(x_i, x_l) \notin M\& \ (x_i, x_j) \in M\& \ (x_i, x_l) \in M \Rightarrow \\ &s(i, l) = 0\& \ s(l, i) = 0\& \ M = (x_i, x_l) \cup M \end{aligned} \tag{9}$$

Therefore, the improved similarity calculation equation is adjusted relied on the above constraint information and Mahalanobis distance as follows.

$$B_{ij} = -exp\left[-\frac{d^2(x_i, x_j)}{\left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i\right)^2}\left\{1+\frac{|\delta_i-\delta_j|}{\delta_{max}}\right\}\right]+1 \tag{10}$$

where δ_i denotes the magnitude of the Mahalanobis distance between point x_i and its k-th nearest neighbor point, and δ_i, δ_{max} are defined as bellow.

$$\delta_i = d^2(x_i, x_k) = (x_i - x_k)\sum^{-1}(x_i - x_k)^T \tag{11}$$

$$\delta_{max} = \{\max|\delta_i - \delta_j|; i = 1, 2, \ldots, n; j = 1, 2, \ldots, n\} \tag{12}$$

Here the similarity calculation is positively correlated with the Marginal distance and density difference of the two points, which is a more realistic representation of the similarity between the multi-scale datasets.

Secondly, based on the calculated Mahalanobis distance, the neighborhood density factor of each sample point is calculated using Equation (13), and the sample dataset is divided into a core point set P_{core} and a non-core point set P_{marg} according to the definitions of core and non-core points [27].

$$F_{ND}(k, x) = \frac{|\{x_i \in X \mid x_j \in N_k(y)\}|}{|\{x_i \in X \mid d(x_i, x_j) \le r, x_i \ne x_j\}|} \tag{13}$$

The K-nearest neighbor method is used to construct the network graph for P_{core}, and the sample points in P_{marg} look for the nearest core points to establish a connection relationship, thus completing the construction of the entire undirected connectivity graph, and using Equation (11) to assign weights to the network graph constructed above to obtain the weights matrix V. The similarity matrix S is obtained by using Equation (10), and S is updated continuously and iteratively by using Equation (2) to Equation (5) before the iteration of the algorithm stops. Finally, the optimal class center point in the clustering is chosen adopting Equation (6), and the proximity is assigned according to the size of the distance between other data points and the optimal class center point.

4. Employment recommendation method for college students based on AP clustering algorithm.

4.1. Pre-processing and weighting of student behavioral information. Focusing on the issue that the recommendation accuracy of the existing methods is not ideal, this article suggests the employment recommendation method for college students relied on AP clustering algorithm, firstly, by analyzing and counting the behavioral data of students, constructing the scoring matrix and arriving at the interest preference of the students, secondly, adding the background information in the basic information of the students, using optimized AP clustering algorithm to cluster the students, dividing them into different clusters of students, and combining with the collaborative filtering algorithm to the scoring data is filled in, and finally the scoring data is calculated and sorted according to the timeliness of the jobs, and the TOP-N employment information is selected for group recommendation.

The behavioral information of students in the Internet recruitment website mainly includes clicking browsing, collecting, canceling collecting and delivering. These behaviors can reflect the degree of students' preference for enterprises, assuming that $C_s = \{c_{s_1}, c_{s_2}, c_{s_3}, c_{s_4}\}$ denotes the behavioral data of student s, where c_{s_1} denotes the number of clicking browsing behaviors, c_{s_2} denotes the number of collecting behaviors, c_{s_3} denotes the number of canceling collecting behaviors, and c_{s_4} denotes the number of delivering behaviors. The behavioral data is analyzed to extract the number of times that student s have clicked and browsed. By analyzing the behavioral data and extracting the student behavior information, the following spatial model X is constructed.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix}_{n\times m} \tag{14}$$

Considering that the cancellation of collection is considered to be a decrease in student preference, its weight is naturally set as small as possible to conform to the system settings, and is standardized using Equation (15); while all behaviors other than it are positive indicators of student preference, and it is chosen to be standardized using Equation (16), as implied below.

$$x'_{ij} = \frac{\max\{x_{ij},\dots,x_{nj}\} - x_{ij}}{\max\{x_{ij},\dots,x_{nj}\} - \min\{x_{ij},\dots,x_{nj}\}}, \quad (i = 1,2,\dots,n; \quad j = 1,2,\dots,m) \tag{15}$$

$$x'_{ij} = \frac{x_{ij} - \min\{x_{ij},\dots,x_{nj}\}}{\max\{x_{ij},\dots,x_{nj}\} - \min\{x_{ij},\dots,x_{nj}\}}, \quad (i = 1,2,\dots,n; \quad j = 1,2,\dots,m) \tag{16}$$

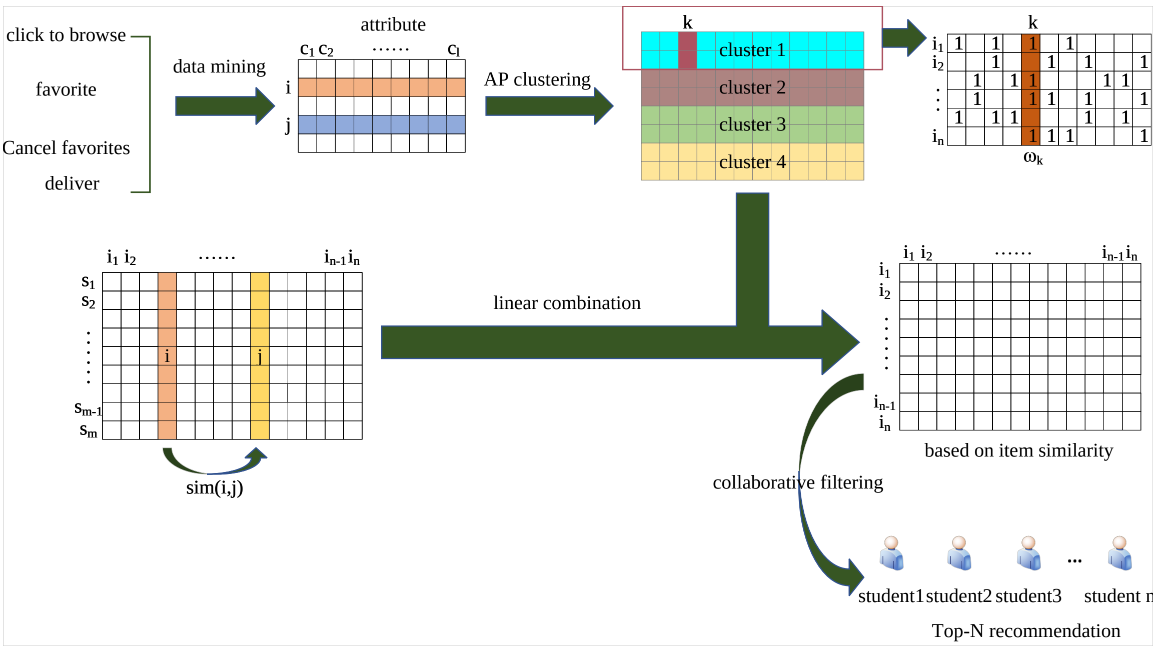

Figure 3. The model of the suggested employment referral methods

After the above normalization process, a new matrix X' is obtained, according to X' which the weight $q_{ij} = x'_{ij} \big/ \sum_{i=1}^{n} x'_{ij}$ of the j-th behavioral attribute under the i-th student is calculated, and the information entropy of the corresponding behavior is calculated according to q_{ij}, as implied below.

$$e_j = -\frac{1}{\ln(n)} \sum_{i=1}^{n} q_{ij} \ln(q_{ij}) \tag{17}$$

where e_j denotes the information entropy corresponding to the j-th behavior, and $e_j > 0$. Relied on the calculated information entropy, the corresponding weight of each behavior is calculated as bellow.

$$v_j = \frac{1 - e_j}{\sum_{j=1}^{m}(1 - e_j)} \tag{18}$$

where v_j is the weight corresponding to the j-th behavior.

Using the above weights, it is possible to calculate the student rating data for the position based on the student's behavioral information, as indicated below.

$$o_{sq} = \sum_{j=1}^{m} v_j \times c_{sj} \tag{19}$$

where o_{sq} represents the rating information of student s for post q, and c_{sj} represents the number of times student s operates on the j-th behavior of post q.

4.2. **Classification of college students' information based on AP clustering algorithm.** After sorting the jobs, this paper needs to construct the background information vector of the job-seeking users first when clustering students based on the optimized AP algorithm in the previous section. The student information vector is constructed using student attributes such as gender, age, household nature, education level, marital status, and majors studied. Then, the similarity calculation is performed to find the most similar set of students and make job recommendations based on the employment status of this set of students, the steps in detail are as bellow.

(1) Let the student attribute dataset be $x = \{x_1, x_2, ..., x_n\}$. Initialize the attraction matrix $g(i, j)$ and support matrix $b(i, j)$ as zero matrices, set the current number of iterations already ($curN$), the maximum number of iterations ($maxN$), the damping factor (μ), and the number of stabilization ($staN$).

(2) The similarity matrix between any two students' characteristic attributes i, j is established using the Mahalanobis distance as bellow.

$$s(i, j) = 1 - \exp\left[-d^2(x_i, x_j)\frac{|\delta_i - \delta_j|}{\delta_{max}}\right] \tag{20}$$

where δ_i denotes the magnitude of the Mahalanobis distance between point x_i and its k-th nearest neighbor point.

(3) The diagonal value $s(i, j)$ in the similarity matrix is given a uniform assignment of P. And update the attractiveness matrix $g(i, j)$ and support $b(i, j)$ using Equation (2) and Equation (3) respectively.

$$P = \frac{\sum_{i,j=1,i\neq j}^{N} s(i, j)}{N \times (N - 1)} \tag{21}$$

(4) Judge whether the iteration is stable or not; take the $\max\{b(i, k) + g(i, k)\}$ vertex j with the largest sum of attraction $b(i, k)$ and support $g(i, k)$ of vertex i as its clustering representative point. After t iterations, we get the clustering center set Et, if $Et = Et-1$, then the number of stable times η is added 1. Otherwise, η goes to 0, if $\eta = staN$, then it is finished, otherwise, jumping to step (5).

(5) $curN = curN + 1$, if $curN = \max N$, the clustering ends and clusters of students with similar characteristics are output, otherwise skip to step (3).

After clustering the students as described above, all the job-seeking students are divided into K different clusters. Each cluster has similar attributes. Next, this paper populates the job scoring data in the same cluster by firstly counting all the jobs in the student similarity set for which the students have scoring data, then removing the jobs for which the students have scoring data, and constituting the remaining jobs into the job populated set Q', and finally calculating the populated scoring data in the job populated set in turn.

$$o'_{iq} = \sum_{j \in S'} \alpha_{ij} o_{jq}, \quad q \in Q' \tag{22}$$

where o'_{iq} denotes student i's data for position q populated ratings, o_{jq} denotes student j's data for position q ratings, S' denotes student clusters, and α_{ij} denotes the similarity between student i and student j.

4.3. **Predictive scoring and timeliness recommendations.** After the first filling of the ratings, the above partially filled ratings matrix is obtained, and this ratings matrix is used as input to the second filling of the ratings using the content-based collaborative filtering recommendation algorithm. First, the cosine similarity is adjusted to calculate the similarity between jobs and find out the similarity set of jobs.

$$sim(i,j) = \frac{\sum_{s=1}^{m} (o_{si} - \overline{o}_s)(o_{sj} - \overline{o}_s)}{\sqrt{\sum_{s=1}^{m} (o_{si} - \overline{o}_s)^2} \sqrt{\sum_{s=1}^{m} (o_{sj} - \overline{o}_s)^2}} \tag{23}$$

where $sim(i,j)$ represents the similarity between job i and job j, o_{si} represents the rating data of student s for job i, $\overline{o}_s$ represents the average of all rating data of student s, and m represents the number of student users.

Then, according to the student clusters clustered by the AP algorithm, find out the top n jobs that are most similar to the jobs in the cluster, eliminate the jobs that have been submitted, and merge the rest of the jobs into a job set Q^*, and then use the preference formula to predict the ratings of the student clusters to be recommended to the jobs in the job set, so as to complete the group recommendation.

$$o''_{si} = \overline{o}_i + \frac{\sum_{j \in Q^*} sim(i,j) \times (o_{sj} - \overline{o}_j)}{\sum_{j \in Q^*} |sim(i,j)|} \tag{24}$$

where o''_{si} denotes the predicted rating of student s for position i, $\overline{o}_i$ denotes the average rating for position i, $\overline{o}_j$ denotes the average rating for position j, and o_{sj} denotes the rating of student s for position j.

Finally, the scoring data within the same cluster were calculated and ranked according to the timeliness of the jobs. Since the chances of students being successfully hired by enterprises are negatively correlated with the demand for the number of jobs and time, the time factor is added, and the time factor of job recruitment is calculated according to Equation (25), T. Then the scoring matrix is calculated according to Equation (26), and the jobs are sorted in descending order according to the calculated values, and the TOP-N is finally recommended to students in the same cluster. students within the same cluster.

$$T_i = \frac{1}{1 + (t_n - t_{ib})} \tag{25}$$

$$O^*_{si} = O_{si} \times T_i \tag{26}$$

where O^*_{si} denotes the final timeliness score value of student s for position i, T_i denotes the time factor of position i, t_n denotes the current time, and t_{ib} denotes the time when recruitment for position i began.

5. Performance testing and analysis.

5.1. Comparative analysis of experimental results.
This article takes the data of 7,285 college students in the 2022 session of a university in China as the experimental object to evaluate the employment recommendation method suggested in this article, and the dataset contains the basic information, cadre information, and employment destinations of the students. And the crawler technology is used to collect and clean the behavioral data of the student samples browsing the online recruitment website, and the processed data are clustered into user backgrounds to obtain 267 user clusters with similar backgrounds. For the convenience of analysis, the method in the literature [17] is denoted as PMO_FM, the method in the literature [18] is denoted as CSJ_KM, the method in the literature [20] is denoted as RSI_CM, and the algorithm in this article is denoted as ERM-AP, and all the experiments are carried out under the Ubuntu 14.10 operating system and Python 3.10 programming environment.

To validate the performance of the recommendation method ERM_AP proposed in this paper, the experiments set the threshold value to 5, the weighting factor to 0.4, and N stands for the length of the recommendation list, which is set to an initial value of 1, a step size of 1, and a maximum value of 100. The Hit Rate (HR) and the Mean Reversed Rank (MRR) [28] are adopted as the evaluation metrics.

Table 1. Recommended Performance Comparison for Different Metrics

Method	**TOP5**		**TOP10**		**TOP20**	
	HR	MRR	HR	MRR	HR	MRR
ERM_AP	43.15%	37.12%	74.91%	45.83%	90.26%	50.16%
PMO_FM	19.28%	8.36%	36.58%	17.12%	55.79%	3.73%
CSJ_KM	28.33%	17.15%	49.72%	23.84%	72.81%	19.57%
RSI_CM	35.24%	22.39%	65.44%	37.46%	85.66%	42.48%

As can be seen from Table 1, ERM_AP achieves the best performance in both HR and MRR metrics when TOP5, TOP10, and TOP20 are taken, and in particular, when TOP10 is taken, the HR and MRR of the ERM_AP method are 74.91% and 45.83%, which are 38.33% and 28.71% better than PMO_FM, respectively, and 9.47% and 8.37% better than CSJ_KM 25.19% and 21.99% respectively, and 9.47% and 8.37% respectively over RSI_CM. PMO_FM performs the worst because it analyzes students' intentions only through historical search behavior and does not fully consider students' preferences. CSJ_KM performs the second worst because it clusters jobs only using fuzzy clustering and does not cluster students with similar preferences, which results in inefficiency. RSI_CM performs the next best, which clusters nearest-neighboring students based on their preferences, but lower than ERM_AP because it does not take into account the students' own background information. The suggested method ERM_AP not only utilizes the improved AP algorithm to cluster users with similar background information, but also considers the timeliness, which ensures the improvement of recommendation performance.

In addition, the RMSE value is also an indicator to verify the performance of the recommendation model, the MAE value is a measure of the root mean square difference between the learner's predicted score and the actual score, if the RMSE value is smaller, it indicates that the recommendation model score situation is closer to the actual score. A comparison of the RMSE values for each employment recommendation method is implied in Figure 4 below. As can be seen from Figure 4, the four employment recommendation methods have similar patterns of change with the number of clusters, and the overall trend is that the RMSE gradually decreases with the increase of the number of clusters.

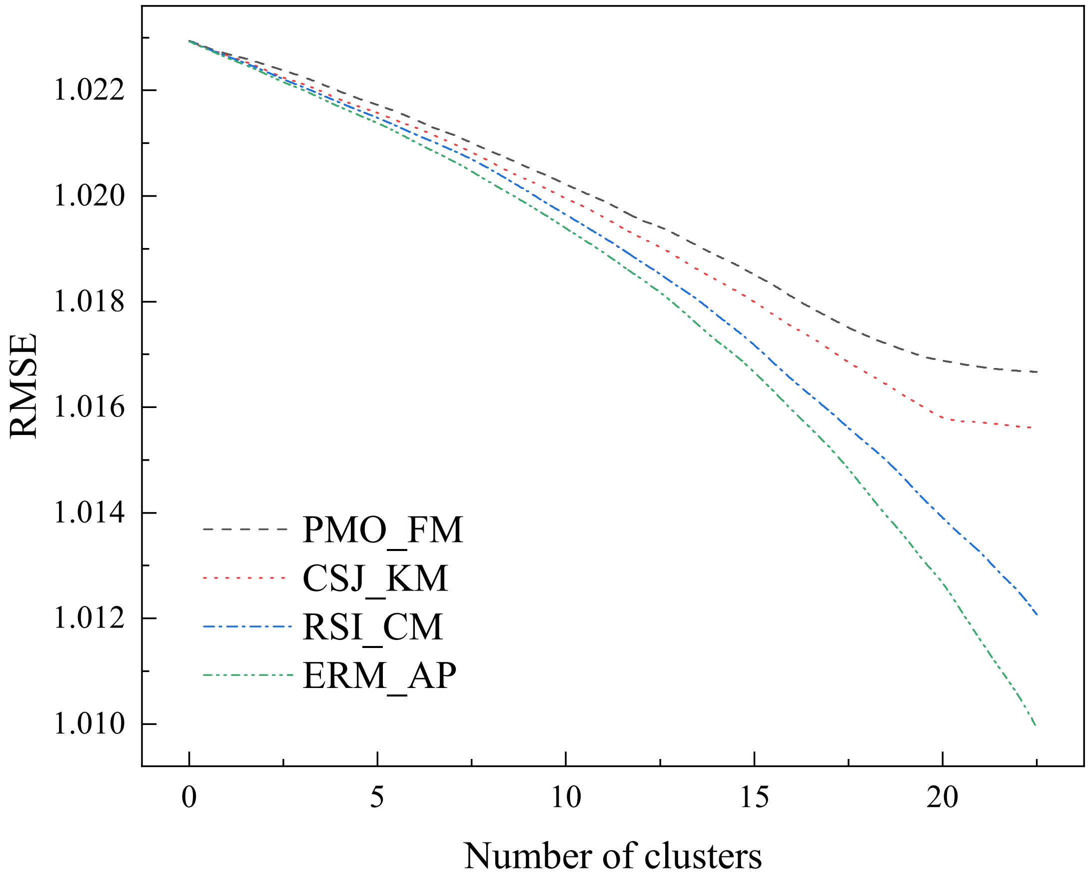

Figure 4. Comparison of RMSE for different recommendation methods

The RMSE value of the ERM_AP method is smaller than that of the other three methods because ERM_AP not only considers students' preferences, but also utilizes the optimized AP algorithm to cluster students with similar backgrounds, so the proposed method has good prediction accuracy.

5.2. **Comparative analysis of recommendation accuracy.** In addition to visualizing the recommended performance of different methods through HR and MRR, further in-depth analysis of the classification effect evaluation indexes Accuracy, Precision, Recall, and F1 value [29] of each type of method is also needed to obtain more accurate information about the classification performance. The classification effect evaluation indexes of each model are summarized as implied in Figure 5.

As can be seen from Figure 5, the classification accuracies of both ERM_AP and RSI_CM are quite high, and their Accuracy values are all greater than 0.85, with the highest being ERM_AP with an Accuracy value of 0.908, followed by the RSI_CM method with an Accuracy value of 0.853, while the classification accuracies of PMO_FM and CSJ_KM are low, with their Accuracy values of 0.704 and 0.786. Looking at the accuracy alone can easily lead the model prediction into a misconception. F1 is the reconciled average of precision and recall, making the performance evaluation more comprehensive, neither falling into the trap of predicting too many negative samples as positive samples to boost the recall, nor falling into the misconception of predicting not enough positive samples. The strongest correct prediction is ERM_AP with an F1 value of 0.911, the RSI_CM method has an F1 value between 0.8 and 0.9, the CSJ_KM method has the next lowest F1 value of 0.773, and the PMO_FM method has the lowest F1 value of 0.712, which suggests that the PMO_FM method has the worst prediction performance among these four models.

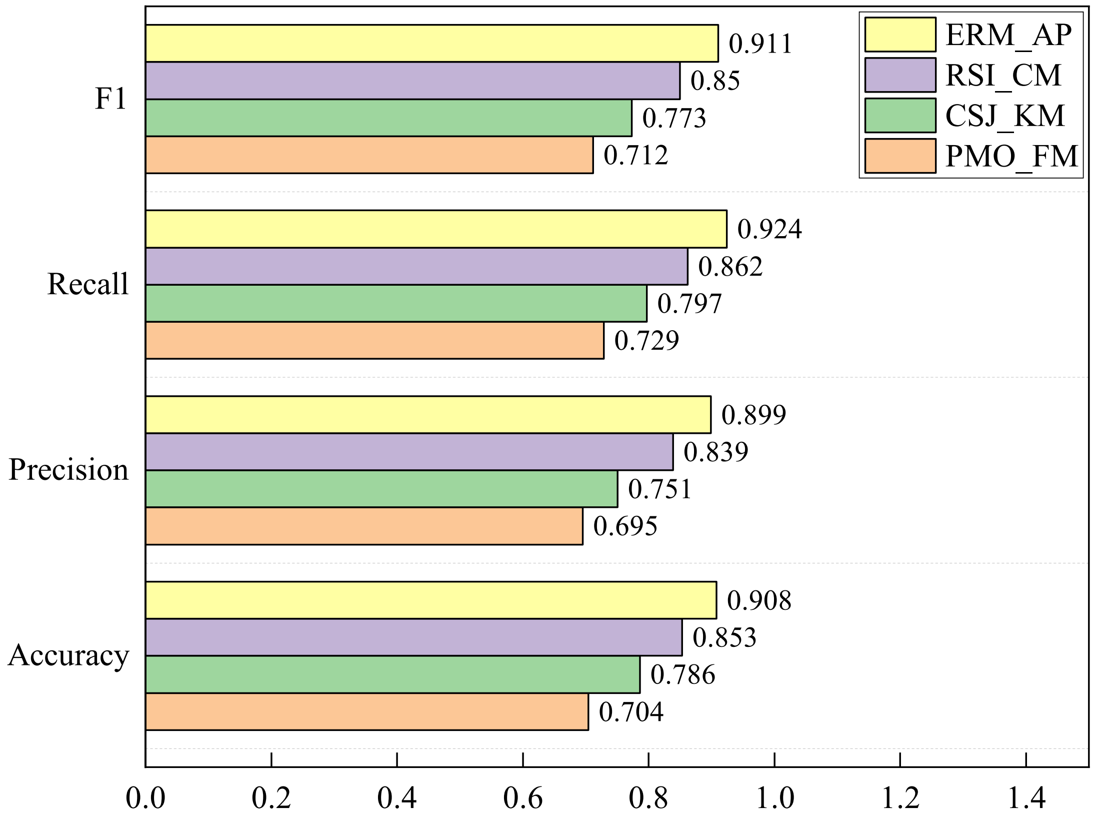

Figure 5. Comparison of the categorization performance of different methods

6. **Conclusion.** Aiming at the issues of hard employment and tough career choice of college graduates, this article suggests a college employment recommendation method based on AP clustering algorithm. Firstly, the Euclidean distance is replaced by the Mahalanobis distance, and the idea of neighborhood density factor is introduced into the nearest-neighbor propagation, so that the similarity matrix can more accurately reflect the relationship between the data, and achieve the goal of optimizing the AP algorithm. Secondly, the statistical analysis of students' behavioral data, the construction of students' scoring matrix and the derivation of students' preferences, and then added the background information in the basic information of job-seeking students, used the improved AP algorithm to cluster the students' background information and divide them into different student clusters, and then combined with the similarity of students' behaviors within the clusters and the collaborative filtering algorithm to fill in the scoring data, and finally, according to the timeliness of the job recruitment requirements. Finally, the scoring data is calculated according to the timeliness of the job requirements, and the results are used as the basis for sorting and selecting the TOP-N employment information for group recommendation. The experimental outcome implies that the suggested method has better performance in HR, MRR, RMSE and F1.2

REFERENCES

[1] Y.-M. Huang, C.-H. Liu, C.-Y. Lee, and Y.-M. Huang, "Designing a personalized guide recommendation system to mitigate information overload in museum learning," *Journal of Educational Technology & Society*, vol. 15, no. 4, pp. 150–166, 2012.

[2] G. K. S. Akorfu, "Measuring information overload within the private university system," *Journal of Information Technology Management*, vol. 24, no. 2, pp. 59–69, 2013.

[3] M. V. Vinichenko, S. A. Makushkin, A. V. Melnichuk, E. V. Frolova, and S. N. Kurbakova, "Student employment during college studies and after career start," *International Review of Management and Marketing*, vol. 6, no. 5, pp. 23–29, 2016.

[4] S. Park, "A Study on University Big Data-based Student Employment Roadmap Recommendation," *Journal of Korea Society of Digital Industry and Information Management*, vol. 17, no. 3, pp. 1–7, 2021.

[5] G. Farell, C. N. Z. Latt, N. Jalinus, A. Yulastri, and R. Wahyudi, "Analysis of Job Recommendations in Vocational Education Using the Intelligent Job Matching Model," *JOIV: International Journal on Informatics Visualization*, vol. 8, no. 1, pp. 361–367, 2024.

[6] W. Li, "Research on personalised recommendation algorithm for college students' employment," *Applied Mathematics and Nonlinear Sciences*, vol. 8, no. 2, pp. 1143–1154, 2022.

[7] A. Hwang, E. H. Kessler, and A. M. Francesco, "Student networking behavior, culture, and grade performance: An empirical study and pedagogical recommendations," *Academy of Management Learning & Education*, vol. 3, no. 2, pp. 139–150, 2004.

[8] R. Liu, W. Rong, Y. Ouyang, and Z. Xiong, "A hierarchical similarity based job recommendation service framework for university students," *Frontiers of Computer Science*, vol. 11, pp. 912–922, 2017.

[9] F. N. I. A. Hamzah, A. F. Saad, and I. Y. Panessai, "Career Finder System using Rule-Based Filtering for University Student Candidates," *International Journal of Artificial Intelligence*, vol. 10, no. 1, pp. 45–58, 2023.

[10] A. Ochirbat, T. K. Shih, C. Chootong, W. Sommool, W. Gunarathne, H.-H. Wang, and Z.-H. Ma, "Hybrid occupation recommendation for adolescents on interest, profile, and behavior," *Telematics and Informatics*, vol. 35, no. 3, pp. 534–550, 2018.

[11] H. Mpia, L. W. Mburu, and S. N. Mwendia, "CoBERT: A Contextual BERT model for recommending employability profiles of information technology students in unstable developing countries," *Engineering Applications of Artificial Intelligence*, vol. 125, p. 106728, 2023.

[12] G. Zhu, Y. Chen, and S. Wang, "Graph-community-enabled personalized course-job recommendations with cross-domain data integration," *Sustainability*, vol. 14, no. 12, p. 7439, 2022.

[13] H. Samin and T. Azim, "Knowledge based recommender system for academia using machine learning: a case study on higher education landscape of Pakistan," *IEEE Access*, vol. 7, pp. 67081–67093, 2019.

[14] B. Daga, J. Checker, A. Rajan, and S. Deo, "Computer science career recommendation system using artificial neural network," *International Journal of Computer Trends and Technology*, vol. 68, no. 3, pp. 84–88, 2020.

[15] F. Zhao, Y. Shen, X. Gui, and H. Jin, "SDBPR: Social distance-aware Bayesian personalized ranking for recommendation," *Future Generation Computer Systems*, vol. 95, pp. 372–381, 2019.

[16] T. Mishra, D. Kumar, and S. Gupta, "Students' employability prediction model through data mining," *International Journal of Applied Engineering Research*, vol. 11, no. 4, pp. 2275–2282, 2016.

[17] Z. Huang and G. Liu, "Prediction model of college students entrepreneurship ability based on artificial intelligence and fuzzy logic model," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 2541–2552, 2021.

[18] P. Ouyang, "The construction of college students' job recommendation model based on improved k-means-CF," *International Journal of Computational Systems Engineering*, vol. 7, no. 2–4, pp. 190–198, 2023.

[19] L. Wang, Y. Fu, and Y. Zhang, "A Career Recommendation Method for College Students Based on Occupational Values," *International Journal of Emerging Technologies in Learning (Online)*, vol. 18, no. 1, 201, 2023.

[20] L. D. Kumalasari and A. Susanto, "Recommendation system of information technology jobs using collaborative filtering method based on LinkedIn skills endorsement," *Sisforma*, vol. 6, no. 2, pp. 63–72, 2020.

[21] T.-Y. Wu, H. Li, S. Kumari, and C.-M. Chen, "A Spectral Convolutional Neural Network Model Based on Adaptive Fick's Law for Hyperspectral Image Classification," *Computers, Materials & Continua*, vol. 79, no. 1, pp. 19–46, 2024.

[22] T.-Y. Wu, A. Shao, and J.-S. Pan, "CTOA: Toward a Chaotic-Based Tumbleweed Optimization Algorithm," *Mathematics*, vol. 11, no. 10, p. 2339, 2023.

[23] T.-Y. Wu, H. Li, and S.-C. Chu, "CPPE: An Improved Phasmatodea Population Evolution Algorithm with Chaotic Maps," *Mathematics*, vol. 11, no. 9, p. 1977, 2023.

[24] J. Chen, X. Wang, S. Zhao, F. Qian, and Y. Zhang, "Deep attention user-based collaborative filtering for recommendation," *Neurocomputing*, vol. 383, pp. 57–68, 2020.

[25] K. Hwang and Y. S. Ko, "Improvement of the marginal reflex distance-1 in blepharoptosis surgeries," *Journal of Craniofacial Surgery*, vol. 27, no. 2, pp. 455–460, 2016.

[26] M. Baldassare, "The effects of neighborhood density and social control on resident satisfaction," *Sociological Quarterly*, vol. 23, no. 1, pp. 95–105, 1982.

[27] W. Vanhaverbeke, V. Gilsing, B. Beerkens, and G. Duysters, "The role of alliance network redundancy in the creation of core and non-core technologies," *Journal of Management Studies*, vol. 46, no. 2, pp. 215–244, 2009.

[28] C.-Y. Lin and H.-S. Chen, "Personalized channel recommendation on live streaming platforms," *Multimedia Tools and Applications*, vol. 78, pp. 1999–2015, 2019.

[29] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, pp. 1–13, 2020.