Research on the Application of Big Data Technology and Collaborative Filtering Recommendation System in Accurately Guiding the Employment of College Students in the Context of COVID-19

Yini Liu[1,a*], Qirui Hao[1,b]

[1]*University of Electronic Science and Technology of China, Chengdu*
[*a]*liuyini@uestc.edu.cn*, [b]*haoqirui@std.uestc.edu.cn*
Corresponding Author: Yini Liu

ABSTRACT

Under the influence of COVID-19, the employment situation of college students is severe, and the employment guidance of college students is also facing new challenges and opportunities. Based on the analysis of the difficulties faced by college students' employment guidance in the context of the epidemic, this paper explores the application of big data technology and the role of collaborative filtering recommendation algorithm in student employment, to accurately guide college students' employment and improve the effect of employment guidance, alleviate the impact of the epidemic, and effectively improve the employment situation of college students.

Keywords: *Big data, collaborative filtering recommendation system, Guiding the employment*

1. INTRODUCTION

In 2020, a public health emergency that has occurred in China with the fastest spread, the widest range of infections, and the most difficult prevention and control since 1949. So far, the impact of the COVID-19 epidemic on China, including the employment of college students, has been far-reaching. Under the background of the epidemic, the employment situation of college students continues to be severe.

2. THE EMPLOYMENT SITUATION OF COLLEGE STUDENTS UNDER THE BACKGROUND OF THE EPIDEMIC

The first is the dilemma caused by the sharp reduction of job opportunities and the continuous increase in the number of graduates. Due to the negative impact of the COVID-19 pneumonia epidemic, China's macro economy has been seriously affected on both sides of supply and demand, economic growth has slowed down, coupled with the intensifying trade disputes between China and the United States, etc.,the demand for labor continued to decrease. The macroeconomic situation had a negative impact on "stabilizing employment". Xi Jinping pointed out that "the impact of the epidemic on economic operation will continue, mainly in terms of aggravating the production and operation difficulties of industrial enterprises, especially small and medium-sized enterprises...increasing the pressure to stabilize employment"[1]. And with the continuous development and progress of the economy and science and technology, as well as the influence of the education system, the national education level has been continuously improved, and the number of graduates has been increasing. The decrease in employment opportunities and the increase in the number of graduates have caused graduates' competitive pressure for job hunting has further expanded, and the degree of difficulty in finding a job has continued to increase. According to statistics, since 2015, the growth of the number of graduates from ordinary colleges and universities in China is as follows:

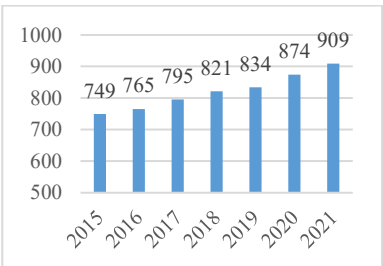

Figure 1: Number of undergraduate and postgraduate graduates from 2015 to 2021 (unit: 10,000 people).

A. El-Hashash et al. (Eds.): IEIT 2022, ACSR 100, pp. 754–760, 2023.
https://doi.org/10.2991/978-94-6463-058-9_118

In addition, due to the downward trend of the overall economic environment, many companies have experienced layoffs and salary cuts. In order to reduce labor costs such as training, employers have also raised overall quality requirements for graduates' academic qualifications, additional skills, and related vocational certificates. The requirements of all aspects of students are bound to be more stringent, making some graduates with weak comprehensive ability more passive in the job market .

The second is the limitation brought by the change of recruitment methods. Affected by the epidemic, many opportunities for offline recruitment in the past have been transferred to online, which has become the main way for college students to find employment. Although online recruitment has the characteristics of fast, non-contact and flexible compared with traditional offline recruitment, to a certain extent, there are shortcomings such as poor accuracy and cognitive bias between enterprises and applicants, which leads to missed job opportunities.

Thirdly, under the influence of the epidemic, the psychological pressure of college students' employment has increased. The employment opportunities for ordinary undergraduates are mainly concentrated in the junior and senior years. Affected by the epidemic, many graduates are worried about successful graduation, and the postponement of the resumption of colleges and universities has led to the slow progress,etc..Some graduates are under increasing psychological pressure. Graduates' job-seeking mentality is gradually changing, which will affect the initiative of graduates in employment to a certain extent [2].

3. CHALLENGES FACING THE CURRENT EMPLOYMENT GUIDANCE WORK IN COLLEGES AND UNIVERSITIES

At present, the epidemic not only brings many obstacles and difficulties to the employment of college students, but also poses a huge challenge to the employment guidance of colleges and universities.

Firstly the function of the employment information system in colleges and universities is not perfect, and the preventive construction is insufficient. And not only the communication method with students has changed to online, but the connection with the job market has also become more online. It began to break, and gradually changed to a mode of "point-to-point" link between students and employers or a third-party employment platform to replace colleges and universities as a "transit point" link. How to use platform resources and policy resources to grasp market demand, broaden employment channels, or establish and improve precise push employment service mechanism, has become the key and difficult point of university employment guidance work under the background of epidemic.

4. APPLICATION OF BIG DATA TECHNOLOGY IN ACCURATELY GUIDING COLLEGE STUDENTS' EMPLOYMENT

Xi Jinping once pointed out: the use of big data to promote security and improve people's livelihood. Employment is one of the key issues of people's livelihood. Big data technology can break through the limitations of time and space in the past, so that the employment of college students can continue to develop in a positive direction.

4.1. Collaborative-filtering recommendation applied to student employment

In the era of big data, schools have all kinds of data about each student's study, life and activities during school. These data contain the main information of students. Big data mining and functional analysis can be used to generate a portrait of students' career aspirations, and accurate employment guidance can be provided based on the portrait.

The rich data resources in universities provide a perfect chance for the application of collaborative filtering system. The core idea of collaborative filtering is to make recommendations based on the similarity between behaviors, while the user-based collaborative filtering recommendation system seeks for the similarity between behaviors of users. After obtaining similar users, the products favored by one user will be recommended to other similar users, thus building a personalized recommendation system.

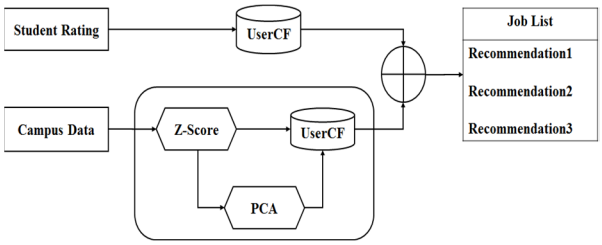

Figure 2: General algorithm design.

The algorithm is shown in Figure 2. Considering that students will have personalized job preferences independent of campus data, the algorithm uses campus data to calculate students' similarity in the cold start-up phase, and takes students rating on jobs into account after they evaluate them. The above two factors were filtered together, and then the results were combined to make recommendations.

In addition, due to the job competition between prospective graduates, the algorithm mainly makes recommendations based on the similarity between graduates and prospective graduates, so as to ensure that

it will help rather than hinder students' employment.

4.2. UserCF based on student ratings

Student job ratings are a good indicator of students' job preferences. User Collaborative Filtering (UserCF) can find graduates with similar preferences and recommend them to jobs they are interested in but not yet aware of.

Measuring the similarity of student u and v with Jaccard similarity coefficient[3]:

$$r_{uv}^{1} = \frac{S_u \times S_v}{\|S_u\| + \|S_u\| - S_u \times S_v} = \frac{\sum_{j=1}^{m}(s_{ju} \times s_{jv})}{\sqrt{\sum_{j=1}^{m} s_{ju}{}^2} + \sqrt{\sum_{j=1}^{m} s_{jv}{}^2} - \sum_{j=1}^{m}(s_{ju} \times s_{jv})} \quad (1)$$

S_i is the preference vector of students, and S_{ij} is the preference of students for a specific job.

Then, according to the similarity of students' preferences, the work recommendation vector ζ_i^1 (i =1,2... , n) could be established for each student:

$$\zeta_i^1 = \sum_{k=1}^{n}(r_{ik}^1 \times S_k) - S_i \quad (2)$$

Taking the five students shown in Figure 3(a) as an example, they evaluate some of the five jobs respectively. Calculate their similarity through formula 1 to get the similarity matrix shown in Figure 3(b). Among these students, s_1 and s_2 are quasi-graduates, and s_3, s_4 and s_5 are graduates. Therefore, recommendations are given to s_1 and s_2 according to the preference of s_3, s_4 and s_5, as Table 1. The recommendation degree of each work for s_1 can be visualized as Figure 3(c).

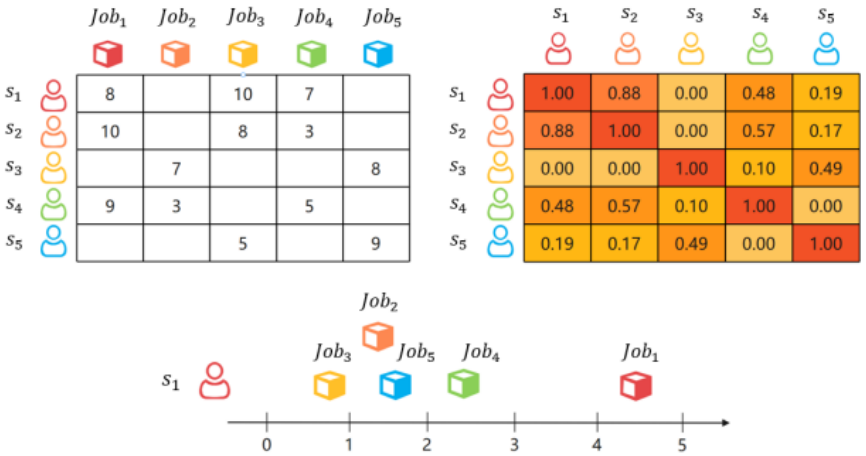

Figure 3(a): A grading matrix of five students on five jobs (out of 10). Figure 3(b): Similarity matrix among five students. Figure 3(c): The recommendation degree of the first two students based on the five jobs obtained by the last three students.

Table 1: Recommendation levels for each job

	Job 1	Job 2	Job 3	Job 4	Job 5
Student 1	4.357	1.452	0.929	2.421	1.673
Student 2	5.164	1.721	0.837	2.869	1.506

4.3. UserCF based on school data

4.3.1. Z-score standardization of input data

Due to the huge difference in the range of each student's features, the eigenvalues are firstly standardized and mapped to the interval [0,1].

Considering that there may be a few outliers among students, for example, some students who love reading may borrow books more often than other students, discrete standardization is not applicable to the data processing of this project. However, most behaviors of students as a group conform to normal distribution, so it is quite appropriate to use Z-Score regularization.

Therefore, z-Score transformation as shown in Formula () is performed for each feature.

$$x_{ij} = \frac{x_{ij}^0 - \mu_j}{\sigma_j} \quad (i = 1,2,\dots,n;\ j = 1,2,\dots,m) \quad (3)$$

Here n is the total number of students and j is the characteristic dimension. x_{ij}^0 is the original eigenvalue, x_{ij} is the mapped eigenvalue, μ_j, σ_j are the mean value and variance of each feature.

4.3.2. PCA for course selection data

As for the data of students' course selection and corresponding performance, students can choose a lot of courses, so if the course selection of each course is taken as a separate feature, the feature dimension is too high, which greatly affects the performance of the algorithm. Therefore, Principal Component Analysis (PCA) is firstly conducted on students' course selection, and then dimension reduction is performed before input into the model.

The basic idea of principal component analysis is to transform the original variable linearly and project it into a new coordinate system. The first major variance after projection is on the first coordinate (called the first principal component), the second major variance is on the second coordinate (the second principal component), and so on[4]. Thus, several principal components can be used to represent the information in the original high-dimensional data.

First, the scores of each course were regarded as a dimension, and the data matrix X of course selection with the dimension of n × q was constructed to calculate its correlation coefficient matrix R:

$$\gamma_{ij} = \sum_{k=1}^{n} \left(\frac{X_{ki} - X_{kj}}{n} \right) \quad (j = 1,2, \dots, q) \tag{4}$$

Therefore, the Jacobian method is used to find Q non-negative eigenvalues of matrix R: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q \geq 0$, and the eigenvector corresponding to the eigenvalue λ_{i}: $C_{ik} = (C_{i1}, C_{i2}, \dots, C_{iq}) \quad (i = 1,2, \dots, q)$

Then q new indexes can be obtained by linear combination of each index X_{k}:
The magnitude of the action of principal component is: $Y_1 > Y_2 > \cdots > Y_q$, the contribution rate of the ith principal component is $(\lambda_i/q) \times 100\%$.

Since the cumulative contribution rate of the first P components is $\sum_{i=1}^{p}(\lambda_i/q) \times 100\%$, so when observing the results, if the cumulative contribution rate of the first several principal components has reached more than 70%, then they are taken as the main component.

PCA is applied to the standardized data of student course selection, and four principal component variables are obtained, whose standard deviation, variance contribution rate and cumulative contribution rate are as follows:

Table 2: Principal components

Index	Comp.1	Comp.2	Comp.3
Standard deviation	2.56836	1.43662	1.22947
Variance Proportion	0.36398	0.15759	0.10077
Cumulative Proportion	0.36398	0.52157	0.62234
Standard deviation	2.56836	1.43662	1.22947

It is not difficult to see that the cumulative contribution rate of the four principal components has reached 70%, which can well represent the actual factors. Therefore, they are used as independent variables to adjust the model, while the original independent variables can be represented by the four principal components through addition and subtraction transformation.

4.3.3. Recommendations Making Based on Collaborative Filtering

Calculate the similarity of student feature representation X_i according to the Jaccard similarity coefficient constant:

$$r_{uv}^2 = \frac{X_u \times X_v}{\|X_u\| + \|X_u\| - X_u \times X_v} \tag{5}$$

Then, according to the preference matrix S of graduates and its similarity coefficient matrix R with prospective graduates, the job recommendation vector ζ of prospective graduates is calculated:

$$\zeta_i^2 = \sum_{k=1}^{n} (r_{ik}^2 \times S_k) - S_i \tag{6}$$

4.4. Recommendations that combine student ratings and campus data

The recommendation for each student is the combination of recommendation based on student ratings and campus data, as shown in formula 7.

$$\zeta_i = w_1 \zeta_i^1 + w_2 \zeta_i^2 \tag{7}$$

According to above, employment intention can be divided into study abroad, employment in first-tier cities, service projects in the western China, etc.; Making full use of big data technology can make employment guidance work more effective and discover many relationships that hardly found before. Furthermore, it can give the really personalized recommendations[5].

5. EXPERIMENT

5.1. Dataset

Apply experiment based on the postgraduate employment of 2016, 2017 and 2018 in the School of Computer Science, University of Electronic Science and Technology of China.

In terms of student rating, the preference data of 2016 graduate students were obtained by means of return interview, while 2017 and 2018 students graded their jobs independently in the employment process.

In terms of school data, students' course selection and scores, ranking, library usage time, the number of books borrowed from the library, and consumption amount of one-card are selected as the characteristic factors of students.

5.2. Evaluation Indicators

Precision rate, recall rate, F index, coverage and prevalence are used to evaluate the recommendation results.

5.2.1. Precision rate and recall rate

Precision rate is the percentage of highly rated jobs in the system's list of recommended jobs that are recommended. And recall rate is defined as the ratio of highly rated jobs in the recommended list to all

Jobs highly rated by the system. They are two measurements widely used in information retrieval and statistical classification to evaluate the quality of results.

5.2.2. F-index

The biggest problem of using precision rate and recall rate to evaluate recommendation system is that they must be used together to fully evaluate the algorithm. Therefore, by integrating the two, F index is put forward.

$$\mathrm{F} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{8}$$

Because F-index unify the accuracy rate and recall rate into one index, it has been widely used. If F-index is high, the system is efficient.

5.2.3. Coverage rate

Coverage is defined as the percentage of all jobs that are recommended. A system with low coverage can only evaluate a limited number of jobs, so it is of little significance to students. Only high coverage can find as many jobs as students are interested in. Coverage must also be used in conjunction with accuracy, because recommendation systems should not give poor accuracy just to improve coverage[6].

$$\mathrm{Coverage} = \frac{\sum \mathrm{Job}|_{\beth>0}}{\sum \mathrm{Job}} \tag{9}$$

5.3. Parameter Settings and Results

Based on the graduate employment data of school of Computer Science of UESTC in recent three years, the number of relevant students is set as 20, and the number of recommended jobs is set as 10 each time. Based on the data of graduates, the calculation accuracy, recall rate and F indicators are shown in table 3.

Table 3: Experimental results based on graduates of UESTC, and two comparative experiments

	Pre	Rec	F-ind	Cov
E	0.3451	0.1677	0.2257	0.9542
C 1^6	0.3013	0.0606	0.1010	0.5582
C 2^7	0.3174	0.1083	0.1615	-

The First comparison experiment comes from movie recommendations based on the MovieLens dataset[6], and comparison experiment 2 comes from book recommendation based on BookCrossing dataset[7].

It is not difficult to see that the recommendation accuracy and F value of the system are high, and the job recommendation for graduate students is well realized. In addition, the system has a good coverage rate and can basically recommend all jobs to students, which plays a role of employment guidance.

6. COUNTERMEASURES FOR USING BIG DATA TECHNOLOGY TO IMPROVE THE PRECISE GUIDANCE OF COLLEGE STUDENTS' EMPLOYMENT UNDER THE BACKGROUND OF THE EPIDEMIC

6.1. Improve the infrastructure of the employment information system

The situation of "remote, online and decentralized" employment of college students caused by the epidemic has put forward higher requirements for the construction of employment informatization systems. Based on the analysis of the actual employment needs during the epidemic prevention and control period, the employment guidance should fully consider the students' employment intentions and then make corresponding suggestions. Using the above methods, it is convenient for employment teachers to grasp the job search dynamics in a timely manner, so as to follow up and provide services.

6.2. Deepen the use of data mining and analysis techniques

In order to mine information related to students' future graduation under the background of big data and provide important support for the actual employment guidance work, colleges and universities should first conduct a comprehensive record of student information, and should dynamically grasp the talent market information. The content of the employment guidance service is integrated into the management of the students' mobile platform, and the online evaluation and comprehensive analysis of the learning tendency of the college students are carried out, and then the corresponding normative guidance is carried out, so as to ensure the integrity and scientific nature of the employment guidance work.

7. CONCLUSION

In the current environment, the COVID-19 epidemic and the wide application of big data technology have had a profound impact on the employment of college students and the employment guidance behavior of colleges and universities. The platform data model can improve the construction level of the employment information system as soon as possible, and on the basis of keeping up with the employment market, conduct objective, real and detailed targeted analysis of the employment situation of college students, personalized employment guidance and full-process tracking, which can help graduates " Apply what you have learned", find your ideal job, and give full play to your talents.

REFERENCES

[1] Xi Jinping.Speech at the meeting of the Standing Committee of the Political Bureau of the Central Committee on the study of the response to the new coronavirus pneumonia epidemic [J]. China Labour Movement, 2020(05):4-7.

[2] Lv Zhijuan, Chen Xudong. Exploration on the career guidance work path for college counselors in the post-epidemic era [J].Modern Commerce and Industry, 2022, 43(03): 72-74. DOI: 10.19311/j.cnki.1672-3198.2022.03.028 .

[3] Bhagirathi Nayak,Rajesh K. Ojha,P. S. Subbarao, VijayaBatth. Machine Learning Finance: Application of Machine Learning in Collaborative Filtering Recommendation System for Financial Recommendations[J]. International Journal of Recent Technology and Engineering (IJRTE),2019,8(1).

[4] J. Sangeetha,V. Sinthu Janita Prakash. Opinion based Memory Access Algorithms using Collaborative Filtering in Recommender Systems[J]. International Journal of Innovative Technology and Exploring Engineering (IJITEE),2019,8(12).

[5] Liu Wei. Precise analysis of college students' employment guidance under the background of big data [J]. Theoretical Research and Practice of Innovation and Entrepreneurship, 2021,4(18):145-147.

[6] Meng-Hui Chen,Chin-Hung Teng,Pei-Chann Chang. Applying artificial immune systems to collaborative filtering for movie recommendation[J]. Advanced Engineering Informatics,2015,29(4).

[7] Hou Ting. Collaborative filtering recommendation algorithm in the application of the intelligence library [J]. Journal of information science and technology, 2021 (33) : 149-151. The DOI: 10.16661/j.carol carroll nki. 1672-3791.2111-5042-0189.