Analysis of Employment Pressure of Local University Graduates Based on Time Relaxation Take Computer Major as an Example

*Zhihong Zhong, Honglang Wei[a]

* Corresponding author: *zzhstudio@126.com, [a]2469994290@qq.com

Dept. of Computer and Information, Qiannan Normal University for Nationalities, Duyun, Guizhou, China

Abstract. Based on the employment datas of local college graduates in China, the time relaxation is defined by the initial employment time of graduates. With time relaxation as an indicator,the degree of employment pressure for graduates is divided into four levels: urgent, normal, loose, and comfortable,then,the employment pressure for graduates is classified. Based on various characteristics of graduates such as gender, ethnicity, poverty, major, political countenance, Normal student category, and nature of employment units, conduct a multidimensional analysis of employment pressure; Finally, the C4.5 decision tree data mining algorithm is used for principal component analysis to obtain the key features that affect the time relaxation, the results indicate that the key characteristics are political countenance, Normal student category, poverty category, and ethnicity. Through datas analysis, obtain numerous meaningful information that can guide graduates' employment decisions.

Keywords: Relaxation of employment time; Decision tree; Data mining; Chinese colleges and universities

1. INTRODUCTION

Since the implementation of the expansion policy in higher education in China in the late 1990s, the number of college graduates has repeatedly reached new highs, and the labor market has oversupplied. The employment situation for college graduates is very severe. From 2021 to 2023, the number of graduates was 9.09 million, 10.76 million, and 11.58 million, respectively; However, according to data from the Ministry of Human Resources and Social Security of China, the national urban unemployment rate from 2021 to 2023 was 5.1%, 5.5%, and 5.2%, respectively. In June 2023, the youth unemployment rate reached 21.3%. In recent years, Chinese universities have accumulated a large amount of management and teaching data in the process of informatization and digital development and upgrading; By applying data analysis and data mining techniques to relevant objective data, valuable information and knowledge can be obtained to guide the employment activities of college graduates. The application of data mining

ITEI 2023, November 24-26, Zhengzhou, People's Republic of China

DOI 10.4108/eai.24-11-2023.2343707

technology in the field of education and teaching is called educational data mining, which is widely used in the field of education as an intellectual support for policy decision-making and management methods. Literature [1-3] uses fuzzy genetic algorithm and decision tree to establish a model for analyzing and predicting the employment trend of college graduates; References [4-5] use the Apriori association rule algorithm to mine student performance and graduate employment data; Literature [6-7] designed a comprehensive analysis and mining system for comprehensive data of college graduates; Literature [8] aimed at the problem of low confidence in the effectiveness evaluation method of traditional college students' employment guidance strategy, an effectiveness evaluation method of college students' employment guidance strategy based on data mining is designed; Literature [9] used the PMC index model to screen the policy texts, obtains two perfect policy texts, and uses the Weibo comments to construct the evaluation model of policy measures support degree to analyze the social effects of employment promotion policies for college graduates; Literature [10] used the deep-seated neural network with strong learning ability and adaptability to predict college students' employment, so as to provide guidance for college students' employment. Literature [11] aimed to analyze the impact of Ideological and political education on the quality of employment by using association rules; [12]aimed to identify similar groups of students and patterns that might associate with these different cohorts of students by examining students' transcripts data.

This article takes local normal universities in China as an example, uses graduate employment data and student management tracking data, and uses the C4.5 decision tree algorithm based on information gain rate to mine the employment situation of computer major graduates. The relevant results can be applied to the employment guidance practice of graduates.

2. CONCEPT DEFINITION AND RESEARCH FRAMEWORK

2.1. Time Relaxation

The initial employment situation of college graduates is influenced by multiple factors, such as personal situation, school situation, and family situation. The time and quality of initial employment will affect the career development and life trajectory of graduates throughout their lives.

Define the initial employment value as (1).

$$V_{fb} = T_{fb} - T_{bs} \ (days) \tag{1}$$

- V_{fb} is value of initial employment, the unit is days.
- T_{fb} is the time of initial employmen.
- T_{bs} is the overall average initial signing time. Based on the overall contract signing time of Chinese university graduates, it can be determined as October of the year before graduation, and the job can be confirmed during the internship period, such as October 1st.

The relaxation of employment time is defined as the discrete value of V_{fb}, as (2).

$$D_{Relaxation} = \begin{cases} 1 & V_{fb} \leq 200 \\ 2 & 200 < V_{fb} \leq 290 \\ 3 & 290 < V_{fb} \leq 380 \\ 4 & 380 < V_{fb} \end{cases} \quad (2)$$

In formula (2), the discrete value range is based on the graduation time of college students as a reference point, and every three months is used as an evaluation value for employment relaxation. The values are: 1=urgent, 2=normal, 3=loose, and 4=comfortable.Then,the employment pressure for graduates is classified.

2.2. RESEARCH FRAMEWORK

This article mainly studies the employment pressure of local college graduates based on the relaxation of initial employment time, and conducts an overall frequency analysis from dimensions such as gender, nationality, poverty, major, political outlook, type of teacher training students, and nature of employment units to understand the impact of various characteristics on employment relaxation; Furthermore, based on the C4.5 decision tree algorithm with information gain rate, a classification model based on employment relaxation is constructed to analyze key features; And predict the possibility of their initial employment situation for a new sample of graduates, and provide employment guidance and learning suggestions. The basic research framework is shown in Figure 1.

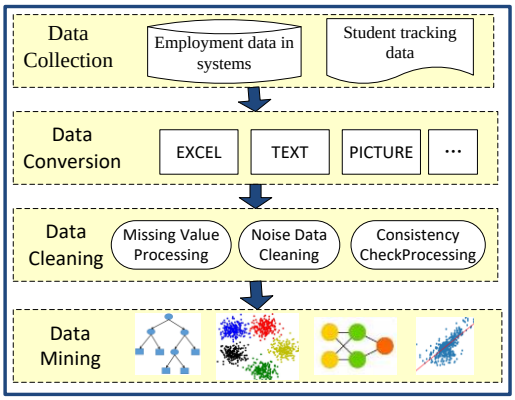

Figure 1 The basic research framework

3. DATA PREPARATION

Accurate, reasonable, and high-quality datas is a prerequisite for data analysis and data mining. The analysis datas comes from employment data of college graduates and various tracking data in student management.

3.1. Data Collection and Preprocessing

The analysis datas comes from different systems and includes database tables, Excel files, CSV files, text documents, paper documents, image images, and so on. In order to conduct comprehensive analysis and mining of data, it is necessary to use multiple methods and tools for datas cleaning of various structured and unstructured formats, such as scanners, ORC text recognition software, Python, Access, SPSS, Kettle, etc.

3.2. Basic information of analysis datas

In order to meet the needs of data analysis and data mining algorithms, it is necessary to format the data and ultimately obtain the employment table for graduates majoring in computer science. Important variable information is explained in Table 1.

Table 1 Explanation of important variables

Variable	***Data type***	***Value and meaning***
Sno	String	10 characters
Sname	String	20 characters
Gender	Integer	1- "male", 2- "female"
Major	Integer	1- "Computational Science", 2- "Information Management", 3- "Software Engineering", 4- "Digital Media", 5- "Network Applications", 6- "Intelligent Science and Technology"
Normal	Integer	1- "normal students", 2- "non normal students"
Political	Integer	1- "Party member", 2- "Non Party member"
Poverty	Integer	1- " Poverty ", 2- "not Poverty "
Nationality	Integer	1- "Han", 2- "Minority"
T_{fb}	Datetime	Registration time for graduates' first employment
V_{fb}	Integer	Unit (day)
$D_{Relaxation}$	Integer	1- "Urgent", 2- "Normal", 3- "Relaxed", 4- "Comfortable"
…	…	…

4. BASIC STATISTICAL ANALYSIS

The employment table for computer major graduates includes a total of 1164 employment data samples for computer major graduates in the past 5 years (2019-2023), for global frequency statistical analysis. Under multiple dimensions, the average initial employment relaxation value is averaged and evaluated corresponding to the corresponding employment relaxation degree.

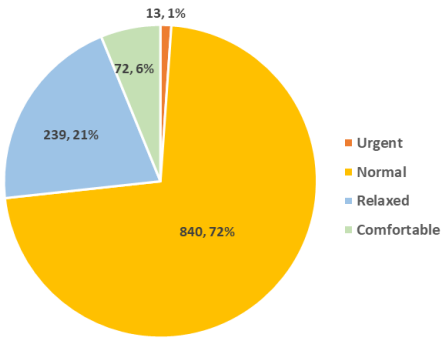

Figure 2 Distribution of students with different degrees of relaxation

In Figure 2, the majority of graduates have a normal degree of employment relaxation (840 people, accounting for 72%), indicating that they have had their first employment within 3 months before and after graduation; Only 1% of students choose to quickly employment as soon as possible, which is in Urgent; graduates with a comfortable relaxation(6%) mainly include those who have failed the postgraduate entrance examination and are preparing to retake the exam, or those who hope to return to their hometown for independent employment.

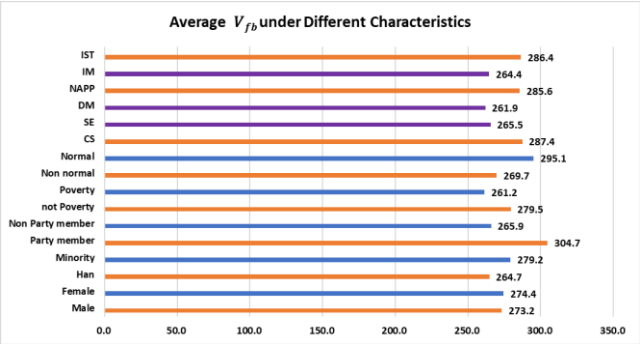

Figure 3 Average V_{fb} under Different Characteristics

Figure 3 compares the average relaxation value (V_{fb}) from multiple characteristics of graduates, and the following conclusions can be drawn.

1) From the gender characteristics of students. Gender differences do not show significant differences in employment relaxation (273.2 for men and 274.4 for women), which indicates that employment of male and female graduates is fair under factors such as employment choice, pressure and opportunities.

2) From the national characteristics of students. From the national characteristics of students. Han students are more willing to work (Han: 264.7, minority: 279.2).

3) From whether the students are Chinese Party members. Employment of Party members and students V_ Fb was significantly higher than that of non party students (party members: 304.7, non party members: 265.9), with a gap of 40 days; Through the examination of the nature

of employment units, it is found that party members and students mostly choose state-owned enterprises, institutions and government offices, and these positions have a relatively long recruitment time.

4) From the characteristics of student poverty. Poor students showed a stronger willingness to work, and the overall average employment relaxation value was 14 days higher than that of non poor students.

5) From whether it is the characteristics of normal students. The employment of teachers' students lags behind that of non teachers' students (teachers': 295.1, non teachers': 269.7), which indicates that the competition for teachers' posts provided by society is fierce, and the employment pressure of teachers' graduates is high.

6) From the professional characteristics of students. Information Management (IM), Digital Media (DM) and Software Engineering (SE) graduates are more competitive in the job market.

5. C4.5 DECISION TREE DATA MINING

The degree of relaxation in initial employment is influenced by factors such as gender students' majors, types of normal school students, politic countenance, poverty categories, ethnic groups, and provinces of origin. In order to further identify important relational factors, this article uses the C4.5 decision tree data mining algorithm based on information gain rate for principal component analysis.

The C4.5 algorithm is one of the most important algorithms in data mining decision tree algorithms. It is an improvement of the ID3 decision tree algorithm, introducing the concept of information gain rate, adopting pruning techniques, and K-th iteration cross validation. The selection of decision tree node attributes in C4.5 algorithm is based on information gain rate, which to some extent avoids errors caused by too scattered feature values.

Firstly, calculate the information entropy of category attributes, then calculate the expected information entropy of non category attributes, and obtain the information gain rate through information gain and segmentation information. The attribute with the maximum information gain rate is used as a node in the decision tree to construct the decision tree. If there is an attribute D, T is its set of values, then Gain (D, T) represents the information gain, and SplitInfor (D, T) is used to measure the breadth and uniformity of the attribute split data. The information gain rate GainRatio (D, T) is defined as (3).

$$GainRatio(D,T) = \frac{Gain(D,T)}{SplitInfor(D,T)} \tag{3}$$

The time relaxation classification data table for analysis is shown in the table 2.

Table 2 Decision Tree Analysis Data Example

ID	*Gender*	*Major*	*Normal*	*Political*	*Poverty*	*Nationality*	$D_{Relaxation}$
1	1	1	0	5	0	0	2
2	1	0	0	3	0	1	2
3	1	0	0	5	0	0	2

ID	*Gender*	*Major*	*Normal*	*Political*	*Poverty*	*Nationality*	$D_{Relaxation}$
4r	1	1	0	5	0	0	3
5	2	1	0	2	1	0	2
6	1	1	1	3	0	0	3
7	1	1	1	5	0	1	3
8	1	1	0	3	0	0	2
9	1	1	0	3	0	0	2
10	1	1	0	2	0	1	3
11	1	1	0	3	0	0	3
12	2	1	0	3	0	0	3
…	…	…	…	…	…	…	…

Using the Weka data mining analysis tool, execute the J48 (C4.5) decision tree classification algorithm on Table II to obtain the decision tree shown in the figure 4.

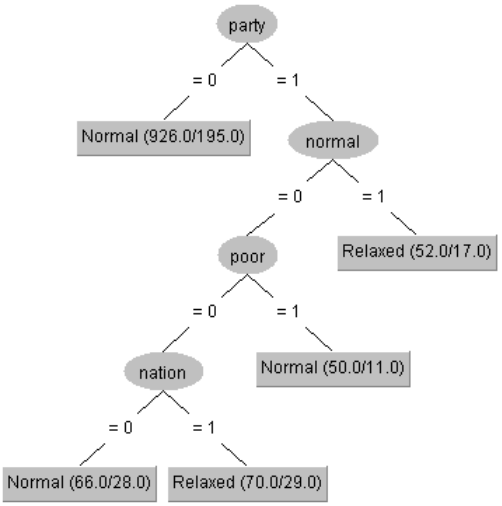

Figure 4 C4.5 Decision Tree

The classification rules are as follows:

1) Non Party member→ Normal

2) Party member→ normal students→ Relaxed

3) Party member→ non normal students→Poverty → Normal

4) Party member → non normal students → not population → Minority → Retired

5) Party member → non normal students → not Poverty → Han → Normal

From the above rules, we can be seen that the key characteristics of graduates include political outlook, the category of normal school students, poverty category, and nationality, which determine the type of employment slack, while gender and professional characteristics have a weaker impact.

6. CONCLUSION

This article integrates the employment data of graduates with various tracking data of student management, and analyzes the current employment situation of computer major graduates in normal universities in ethnic minority areas of China. The relaxation degree of employment time is defined by the initial employment time of graduates, based on various characteristics of graduates, multi-dimensional employment pressure analysis is carried out. Finally, the C4.5 decision tree data mining algorithm is used for principal component analysis, Thus, key characteristics that affect the relaxation of employment time can be obtained, and numerous meaningful information can guide graduates' employment decisions can be obtained.

Analyzing the employment data samples of 1164 graduates majoring in computer science from 2019 to 2023, the conclusions are as follows:

1) The vast majority of graduates (72%) have a normal degree of employment slack, indicating that they have started their first employment within 3 months before and after graduation; Only 1% of students choose to find employment as soon as possible, which is in an urgent area.

2) Gender differences have no significant impact on employment pressure and choices; The students of Han have a more positive desire for employment; Party members and students are more focused on employment positions in state-owned enterprises, public institutions, and government agencies; Poor students and normal school students face significant employment pressure; In terms of computer science, graduates from information management, digital media, and software engineering majors are more competitive in the job market.

3) The C4.5 decision tree data mining results indicate that among the classification attributes of employment slack, the key attributes are {political appearance, normal students, poverty category, nationality}.

Funds: The 2022 Guizhou Provincial "Golden Course" educational reform project: Introduction to Database Systems, and the 2021 Guizhou Provincial Teaching Content and Curriculum System Reform Project (project number: 2021339).

REFERENCES

[1] HASHMIA H, SIMI I, JUBILANT J.Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algo-rithm [J]. Procedia Technology, 2016, 25 : 326 - 332.

[2] WANG C,WANG J Z.Cloud-service decision tree classification for education platform[J].Cognitive Systems Research,2018,52:234-239.

[3] LI Z F,XIONG X Z,TANG Y,et al.Research on employment trend analysis method of college graduates based on neighbor joining phylogenetic trees[J].Journal of Sichuan University of Science & Engineering(Natural Science Edition),2021,34(5):86-93.

[4] YIN F D, YAO H Z, HE G.Studying and Ap-plication of Apriori Algorithm in Analysis of Students'Achievement [J]. Applied Me-chanics and Materi-als, 2014, 3172(529) : 759 - 762.

[5] LIU Li na.An Intelligent R ecommendation System for Graduate Employment Based on Improved Apriori Algorithm[J].Journal of Eastern Liaoning University(Natural Science Edition),2021,28(2):130-135.

[6] ZHOU Li min,ZHANG Li jing. Research on Employment Quality Evaluation System of College Graduates Based on Big Data Mining[J].Journal of Lanzhou Petrochemical Polytechnic,2021,21(2):40-44.
[7] LI Zuojun.The Application of Interestingness Mining of Association Rules in Employment Analysis in the Era of Big Data[J].SOFTWARE ENGINEERING,2018,21(11):25-27.
[8] Dong Fei, Jinghan Song. Evaluation Method of Effectiveness of College Students' Employment Guidance Strategy Based on Data Mining[J]. Journal of Education and Culture Studies, 2021, 5, 4.
[9] Chen Tinggui, Rong Jingtao. Analysis of Social Effects on Employment Promotion Policies for College Graduates Based on Data Mining for Online Use Review in China during the COVID-19 Pandemic[J]. Healthcare, 2021,9, 7, 846.
[10] Cheng Ping. Employment Data Screening and Destination Prediction of College Students Based on Deep Learning[J]. Wireless Communications and Mobile Computing, 2022 .
[11] Zhang Ying. The Influence of Ideological and Political Education on Employment Quality of College Students based on Association Rule Analysis[J]. Journal of Physics: Conference Series, 2021,1744, 4, 042169.
[12] Maqsood Rabia;Ceravolo Paolo;Ahmad Muhammad;Sarfraz Muhammad Shahzad.Examining students' course trajectories using data mining and visualization approaches[J]. International Journal of Educational Technology in Higher Education. 2023,20 ,1.