Analysis of Contemporary College Students' Employment Model Based on Logit Model and Machine Learning

Huanhuan Li
1Xi’an Aeronautical Polytechnic Institute, Shanxi, Xian, 710089, China.

Abstract. With the emergence of Internet big data, cloud computing, artificial intelligence and other new technologies, employment positions have also undergone profound changes. Small, medium and micro enterprises, new business forms and new models put forward new requirements for employment services in Colleges and universities. Starting from the current situation of College Students' employment, this paper analyzes the reasons for the difficulty of College Students' employment and discusses how to solve the problem. In view of the group differences in the employment of college graduates, this paper puts forward some countermeasures and suggestions to speed up the development of service industry and promote employment, formulates the long-term strategy for the development of talent training institutions, and constructs an advanced account management system for college students. This paper analyzes the students' self-concept from the aspects of social needs, the change mode of school cultural concept and the lack of professional development. On the issue of College Students' employment, this paper analyzes the characteristics of College Students' employment and the needs of employers with the employment supply and demand as the main line. The results show that up to 26% of the graduates work in enterprises and at least 14% in administrative organs.

1 Introduction

With the acceleration of globalization and the rapid development of higher education, the employment situation of college students is very serious. There are some problems and defects in the employment of college students, which hinder the improvement of the quality of employment guidance.

With the continuous development of computer technology, many experts have studied the employment of college students. For example, some domestic teams have studied the prediction model of College Students' Employment Based on natural neighborhood classification algorithm, analyzed the formation of College Students' Employment Culture from four aspects of ideological culture, life culture, management culture and evaluation culture, and elaborated the necessity of College Students' employment culture construction. This paper puts forward the ways to construct the employment culture of college students, in order to play a guiding role in the employment of college students. This paper studies the concepts of employability and core employability, introduces the connotation and characteristics of College Students' core employability, and puts forward corresponding training measures. On the basis of investigation and literature search, this paper analyzes the current situation, characteristics and existing problems of the regional flow of College Students' employment, and puts forward some countermeasures to solve the problems. Through the questionnaire survey and analytic hierarchy process, this paper makes a quantitative analysis of the factors affecting the employment prospects of college graduates, which provides useful reference information for the employment direction of college students and the reform of school education. This paper analyzes the main reasons for the high employment expectation of college students, and expounds the research status of College Students' employment prediction at home and abroad. This paper introduces the definition and process of classification, several common classification algorithms and their advantages and disadvantages, and summarizes the mathematical model and calculation steps of factor analysis. Through factor analysis, this paper constructs a graduate description model and discusses the key factors affecting the quality of College Students' employment. The shortcomings of traditional KNN classification algorithm are pointed out, that is, the K value needs to be set manually according to historical experience. A classification algorithm based on natural neighborhood is proposed. The algorithm first uses the training sample weighting algorithm based on natural neighborhood to weight the training sample set, and then based on the idea of natural neighborhood, classifies and gives weight to the test samples according to the obtained training sample set [1]. Some experts have studied the construction and verification of the model of normal students' employment competency. Through job analysis, we can understand the structure of normal students' employability and the main tasks of College Students' employment. A semi open
questionnaire was used to investigate the normal students. Then, Bei interview method is used to interview normal university graduates, and the interview results are combined with the semi open questionnaire to construct the competency model of normal university graduates. The supervised learning algorithm (random forest) is used to construct a predictive classification model to classify the three states of students' employment (unemployed, employed and enrolled). With the help of slope and breakpoint, the linear change of students' behavior is measured; with the help of long-term memory network and short-term memory network, the time pattern of students' behavior is measured. After collecting big data such as campus card, preprocess the relevant data. Divide the data by week, and make a preliminary statistics of students’ weekly behavior frequency (such as the number of times to the library each week). Using the methods of statistical analysis and deep learning, this paper studies the law of students' behavior changing with time [2]. Some experts have studied the employment prediction model of Graduates Based on machine learning, elaborated the relevant theories of feature selection algorithm and prediction algorithm, deeply analyzed the Factors Affecting Graduates' employment, and summarized the characteristics of student data: large amount of information, high prediction accuracy, complex features, high dimension, and the attributes are often discrete, with discrete features and continuous features. And there are many redundant features. In view of these characteristics, this paper proposes the combination of xgboost feature selection algorithm, which provides a solution for the employment prediction of graduates. Compared with other common prediction algorithms, the algorithm is evaluated objectively [3]. Although there are many achievements in the research of College Students' employment, there are still some deficiencies in the analysis of contemporary college students' employment model based on logit model and machine learning.

In order to study the analysis of contemporary college students’ employment model based on logit model and machine learning, this paper studies the college students' employment model and finds the feature selection algorithm. The results show that logit model and machine learning are helpful to analyze the employment mode of contemporary college students.

2 Method

2.1 Employment Model of College Students

2.1.1 Machine learning

With the rapid development of big data technology, new requirements and conditions are put forward in the field of machine learning. Supervised learning is a kind of machine learning[4]. There are also unsupervised learning and reinforcement learning, which is beyond the scope of discussion. When we know the type of data set to be trained, we input some labels manually, train the data set to meet our expectations, and predict the data set with unknown labels[5]. Generally speaking, supervised learning algorithms can be divided into two categories: regression and classification[6]. The purpose of machine learning is to acquire the knowledge and rules behind the data, so as to peel off the data itself[7]. People use machine learning to extract models from a large number of data and solve practical problems based on models[8]. Generally speaking, machine learning models can be divided into two categories[9]. One is predictive, which is mainly used to predict future trends; the other is descriptive, which is mainly used to extract data hiding rules, which can help people acquire knowledge through description[10].

2.1.2 On the comprehensive ability of College Students

Having a good career development plan can make you pay more attention to the study of relevant knowledge. The evaluation of College Students’ employability is a comprehensive evaluation, including many aspects of College Students' learning ability and working ability. It is often fuzzy, so it is more appropriate to use the fuzzy comprehensive evaluation method. However, the weight of each index is often subjective, so AHP can be used to aggregate and quantify the weight of each index. To make a targeted quantitative evaluation of the individual factors that affect the quality of College Students' employment, quantitative assessment efforts. Using the results of comprehensive evaluation, we can sort the comprehensive performance of graduates quantitatively, and prepare the data for their directional recommendation of employment. Students who have served as student leaders are excellent students with good ideological and moral quality, willing to contribute and all-round development. Many employers take the employment situation as an important indicator to measure the comprehensive quality of graduates. Many students have no accurate orientation of their interests and work goals, blindly obtain various certificates, and ignore the improvement of professional knowledge, humanistic quality and other comprehensive ability, which is not conducive to career development to a certain extent. Many vocational colleges lack professional career guidance teachers, which makes career planning education in vain. Graduates with strong comprehensive quality are more favored by employers and easier to find jobs. There is no good career planning ability, establish professionalism. Only by mastering the core skills can we effectively deal with various problems related to career development. Colleges and universities should provide more internship and practice opportunities for students majoring in economic management.

2.2 Feature Selection Algorithm

The process of feature selection is to extract all the features from the data set, record the total number of features as N, select the most effective feature set from n features according to certain rules, and record the size of the optimal feature subset as n (n < n). Among the total number of features n, select an optimal feature subset with
the size of N, the number of possible choices is shown in equation (1):

$$Q = C_n^d = D![[D - d]!]^{n}(d = n, D = N)$$  \tag{1}$$

If we divide the standard deviation of \(X\) and \(Y\) by the covariance, we can get the coefficient and write \(\rho\). Pearson correlation coefficient can be used to evaluate the degree of linear correlation. The calculation formula is shown in equation (2-3):

$$\text{cov}(X,Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$  \tag{2}$$

$$\rho_{x,y} = \text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x \sigma_y}$$  \tag{3}$$

In this method, the independent variable dependent variable is not a continuous variable, and the number of independent variable dependent variable values is determined. Suppose that the number of expected values of independent variable and dependent variable are \(n\) and \(m\) respectively. When the value of independent variable is 1, the difference between the number of samples with the value of dependent variable \(J\) and the expected value is shown in equation (4):

$$x^2 = \frac{\sum (A - E)^2}{E} = \frac{1}{E} \sum (A_i - E_i)^2 = \frac{1}{n} \sum (A_i - np_i)^2$$  \tag{4}$$

Mutual information is the correlation value between two random variables, which shows that the known value of one random variable reduces the value range of the other random variable, that is, the value certainty increases. Mutual information can be used to evaluate the correlation between random variables, as shown in equation (5):

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$  \tag{5}$$

3 Experience

3.1 Experimental Object Extraction

In the original data set, the performance of the prediction model constructed by xgboost algorithm is evaluated through the test set. This paper makes a detailed investigation on the Factors Affecting Graduates' employment. The employment of college graduates is related to many attributes. The attributes involved in different levels will affect graduates' employment results according to their correlation. In order to make the comparison more intuitive, the random forest algorithm is added as a comparison. Random forest algorithm is to establish an independent multi decision tree, classify and predict the data set of each tree. In the career evaluation module, the model is used to define the requirements and implementation of each sub module. It can perform predefined operations independently according to the information it carries. In the career evaluation module of each model, we need to know which factors have an impact on students' career evaluation, so as to formulate corresponding evaluation problems according to these factors. The final classification prediction result is determined by the voting result of each tree. By comparing the experimental results of the two algorithms, the superiority of xgboost algorithm is proved.

3.2 Experimental Analysis

Whether the algorithm process is terminated or not depends on the comparison result between the evaluation function h of the selected feature set and the termination condition of the set. If the set conditions are met, the feature selection process ends. Otherwise, the corresponding subset search strategy will continue to repeat until the conditions are met. In the process of feature selection algorithm, the termination condition is usually set as follows: if the size of the selected feature subset is greater than the threshold set at the beginning, the algorithm process will be terminated; if the set conditions are not met, and the number of search cycles of the subset search strategy is greater than the threshold set at the beginning, the algorithm process will be terminated; if the value obtained by the evaluation function is greater than the threshold set at the beginning, the algorithm process will be terminated. The selection of search strategy and the setting of evaluation function are very important to the whole search process. Appropriate search strategy can get the best search results in the shortest time and eliminate redundant information to the maximum extent. The quality of feature subset selection plays a key role in the accuracy and performance of the algorithm.

4 Discussion

4.1 Data Collection

Using the method of random sampling, this paper tests the public management students who graduated from work and the public management students in school. The questionnaire was filled in by mobile phone and distributed on site. A total of 150 questionnaires were collected. Among them, 5 questionnaires with incomplete answers and obviously unqualified answers were excluded. 145 valid questionnaires were obtained with a recovery rate of 97%. The detailed composition of the subjects is shown in Table 1.

| Graduates | attribute | category | number | Proportion |
|-----------|-----------|----------|--------|------------|
| Gender    | male      | 26       | 45%    |            |
|           | female    | 32       | 55%    |            |
| Nature of work | government sector | 15 | 17% |        |
|           | administrative organ | 11 | 14% |        |
|           | government-affiliated institutions | 16 | 18% |        |

Table 1. Basic information of graduates
It can be seen from the above that 17% of the graduates work in government departments, 14% in administrative organs, 18% in institutions, 26% in enterprises and 25% in other jobs. The results are shown in Figure 1.

| enterprise | 23 | 26% |
|------------|----|-----|
| other      | 22 | 25% |

Figure 1. Basic information of graduates

It can be seen from the above that 26% of graduates work in enterprises, 14% in administrative organs and 1% more in public institutions than in the government.

4.2 Influencing Factors of College Students’ Employ-ability

Through literature analysis, it is concluded that the factors affecting the employability of college students are students’ own factors, school factors, social factors and family factors. From the analysis of social employment situation and university development in recent years, the local talent demand and employment policy are relatively stable, and the employment rate of university graduates in Chongqing has been at a stable level in recent five years, as shown in Table 2.

Table 2. Initial employment rate of college graduates in recent five years

| particular year | 2016  | 2017  | 2018  | 2019  | 2020  |
|-----------------|-------|-------|-------|-------|-------|
| Number of graduates (ten thousand people) | 19.7  | 20.8  | 25.7  | 28.4  | 31.6  |
| Initial employment rate (%) | 15.6  | 16.4  | 20.5  | 22.5  | 25    |

It can be seen from the above that in 2016, there were 197000 university graduates, with the initial employment rate of 15.6%; in 2017, there were 208000 university graduates, with the initial employment rate of 16.4%; in 2018, there were 257000 university graduates, with the initial employment rate of 20.5%; in 2019, there were 284000 university graduates, with the initial employment rate of 22.5%; in 2020, there were 316000 university graduates, with the initial employment rate of 25%. The results are shown in Figure 2.

Figure 2. Initial employment rate of college graduates in recent five years

It can be seen from the above that the initial employment rate of college students increases with the year and the number of graduates. In 2020, the number of graduates will increase to 316000, and the initial employment rate will be 25%.

5 Conclusion

With the expansion of the scale of higher education, the number of college students entering the job market has increased greatly. At the same time, China has entered a new stage of accelerating the adjustment of industrial structure, and the contradiction between the social demand for talents and the supply of talents in Colleges and universities is prominent. The employment problem of college students has become increasingly prominent. Based on the survey of the supply and structure of college graduates, this paper analyzes the influence of college enrollment expansion, urbanization, industrialization and urban employment growth on the scale of college employment. Using the factor analysis method, this paper makes an empirical analysis on the relationship between urbanization and employment, and obtains the relationship between industrialization and the employment number of urban college students and the employment growth of urban college students. From the two aspects of specialty supply and industrial growth, it comes to the conclusion that the employment number of college students in the city is on the rise. This paper analyzes the current situation of College Students’ Employment Confidence, and reflects on the influence of College Students’ employment quality and employment expectation on their career. This paper is helpful to expand employment ideas, broaden employment vision, and do a good job in the employment of college graduates.

Reference

1. Gai Y , Qi Z . 76.Research on the Sports Humanistic Spirit of Contemporary Female College Students based on Big Data Method[J]. Boletin Tecnico/Technical Bulletin, 2017, 55(10):539-545.

2. Alzahrani A K A A K . Critical Theory Based on the communicative function: Analytical study of contemporary experience[J]. Journal of King Abdulaziz University-Arts and Humanities, 2018,
3. Davis J M, Topaloglu H, Williamson D P. Pricing Problems under the Nested Logit Model with a Quality Consistency Constraint [J]. Informs Journal on Computing, 2017, 29(1):54-76.

4. Byun H, Lee C Y. Analyzing Korean consumers' latent preferences for electricity generation sources with a hierarchical Bayesian logit model in a discrete choice experiment [J]. Energy Policy, 2017, 105(JUN.):294-302.

5. Nayeem, Ameen, Camille, et al. Forecast of Airport Ground Access Mode Choice with the Incremental Logit Model: Case Study of the AirTrain at John F. Kennedy International Airport [J]. Transportation Research Record, 2018, 2336(1):97-104.

6. Megaravalli A V. Management Science Letters Estimating growth of SMES using a logit model: Evidence from manufacturing companies in Italy [J]. Management Science Letters, 2017, 7(3):125-134.

7. Byun H, Lee C Y. Analyzing Korean consumers' latent preferences for electricity generation sources with a hierarchical Bayesian logit model in a discrete choice experiment [J]. Energy Policy, 2017, 105(JUN.):294-302.

8. Wang X, You S, Wang L. Classifying road network patterns using multinomial logit model [J]. Journal of Transport Geography, 2017, 58(JAN.):104-112.

9. Timothy, J, Gates, et al. Prediction of Driver Action at Signalized Intersections by Using a Nested Logit Model [J]. Transportation Research Record, 2018, 2463(1):10-15.

10. Davis J M, Topaloglu H, Williamson D P. Pricing Problems under the Nested Logit Model with a Quality Consistency Constraint [J]. Informs Journal on Computing, 2017, 29(1):54-76.