Research on Employment of College Students Based on Statistical Analysis of Data

Zhichao Lin

Beijing Jialian School, Beijing 102200, China

2813859367@qq.com

Keywords: Data statistics and analysis, Employment of college students, Employment issues, Data mining

Abstract: With the increasingly severe social employment situation in recent years, the employment difficulty of college students has become a hot topic. Faced with this employment situation, students' views on it, their employment situation, their employment confidence and how to improve their employment confidence are all issues worth studying. This study adopts the viewpoint that social capital is social network. At the individual level, it describes the social capital status of college graduates from four aspects: network scale, network differences, network top and network composition (ties with the leadership, managers and knowledge). It also tries to find out the influence of college graduates' personal characteristics and family background on their social capital and successful employment through quantitative analysis, and then analyzes the relationship between social capital and successful employment. The above findings indicate that, on the one hand, in China's current higher education labor market, the wage determining mechanism is increasingly returning to the neoclassical model; On the other hand, the unfair and unreasonable phenomena still exist in the field of college students' employment, and the reasonable allocation of labor force cannot be realized completely through the market mechanism.

1. Introduction

In recent years, the rapid development of China's higher education has attracted worldwide attention. With the expansion of college enrollment in 1999, the employment difficulty of college students has become an economic and social problem of great concern. According to the statistics of the Ministry of Education, on the one hand, the number of university graduates keeps increasing. The transition of college graduates' employment policy intensifies the appearance of graduates' “employment difficulties”, and various media reports about college students' “employment difficulties” frequently appear, which makes the employment situation of college graduates more and more severe[1]. In 2007, the number of college graduates reached 10,000, accounting for 55% of the newly increased 9 million jobs, exceeding half of the newly increased jobs in society for the first time. If we add about 30% of the graduates who failed to find employment in 2006, the graduates who need employment in 2007 will account for more than 65% of the newly increased 9 million jobs. Obviously, the employment of college graduates will become a very prominent contradiction in China's social employment[2]. By September 1st, 2008, the number of graduates from colleges and universities across the country had achieved employment was 3.51 million, although it was 540,000 more than the same period of last year, but there were still more than 1.4 million fresh college graduates who failed to find employment on schedule[3].

As the international financial crisis continues to spread and worsen, the overall demand for graduates in the job market will be significantly reduced. Enterprises, especially export-oriented enterprises, will certainly adjust their recruitment scale[4]. Facing the increasing number of graduates and the decreasing demand of the employment market, the graduates of 2009 will face more severe challenges than in previous years[5]. So does the decline in the enrollment growth rate mean that the expansion of China's universities has come to an end, and indicates that the severe employment situation of college graduates will be eased? In a statistical sense, the enrollment

growth rate is only a relative statistic to measure the growth rate of enrollment, but in terms of absolute statistics, China's higher education has formed a trend of scale development after several years of enrollment expansion. The slow growth of enrollment scale starting from a huge base is still rapidly promoting the expansion of China's colleges and universities, and the job seekers of college graduates entering the big social market will continue to expand[6]. The actual employment situation of college students has a lot to do with their employment expectations, and unreasonable employment expectations will lead to the low actual employment rate. Therefore, understanding the employment expectation and employment status of college students and grasping the gap between them are to help college students better understand the gap between employment expectation and employment status[7].

Due to historical accumulation, policy tendency and other reasons, the development level of higher education and employment choice of college students in different regions of China show different characteristics. In order to further understand the employment concept of contemporary college students, clearly reflect the employment requirements of contemporary college students, and coordinate the relationship among society, schools and students, it is feasible and necessary to study college students' employment satisfaction by quoting the theory and method of user satisfaction index.

In this paper, an association rule mining system based on students' employment information should be established. Through the mining of students' enrollment information, school performance and employment information, we can obtain useful information for the school to cultivate students and understand students, that is, we can apply data mining(DM), a new technology, to the management of students. By studying the practice and application of student information management system, this paper explores the technology of multidimensional association rule DM, improves algorithm, and realizes a practical and efficient method of association rule DM based on multidimensional frequent.

2. Research on Employment of College Students

2.1 A Summary of Relevant Research on Successful Employment of College Students

Talent cultivation in colleges and universities is a complex process, involving a series of factors and conditions such as teaching, curriculum and management[8]. In the Undergraduate Specialty Catalog of Ordinary Colleges and Universities issued by the Ministry of Education, the undergraduate specialty catalog includes 11 disciplines, including philosophy, economics, law, education, literature, history, science, engineering, agriculture, medicine, management, etc. There are 71 secondary disciplines and 249 specialty catalogs[9]. Because there is no special professional division standard for higher education levels other than undergraduate education, the professional division is conducted according to the training plan formed by the curriculum organized by each university according to the discipline classification[10].

In employment, subjectivity can be “himself”; Be able to actively and orderly manage your career and adapt to the development of society; Be able to have a comprehensive and sufficient understanding of their own personality characteristics, abilities and specialties, value orientation, as well as the nature of the recruiting unit, corporate culture and development opportunities they can provide, and integrate the above resources; Be able to choose your own occupation voluntarily and freely according to your own interests, hobbies and needs, and be willing to make efforts for it; Be able to work happily and actively, and give full play to one's specialty, especially professional specialty; Be able to give full play to your potential and creativity, and actively start a business by using all your learned strengths. The unfinished nature of human beings determines that there is great potential hidden in human beings, which makes their development have rich possibilities, and determines that human existence is an endless process of improvement. In addition to the stage and development of the career, it is impossible for a person to settle down in a specific career that will remain unchanged for life. This determines that people must learn all their lives. Only through constant self-education and learning can people master new knowledge and skills, improve

themselves, and seek new survival and development. The employment confidence can be quantified by using the index. This is because the index can not only be used to measure the general dynamics of a certain social phenomenon that cannot be directly added or compared, but also to measure the degree of influence of changes in various factors in the total changes of a certain social phenomenon, as well as the influence of various factors in the average comparison. With the help of the index, the evaluation of employment confidence can be more intuitive and scientific.

The use of social capital concept is at the general theoretical level, while the use of social resources is at the level of empirical research. The relationship between social resources and status acquisition is also studied, and the model shown in Figure 1 is put forward.

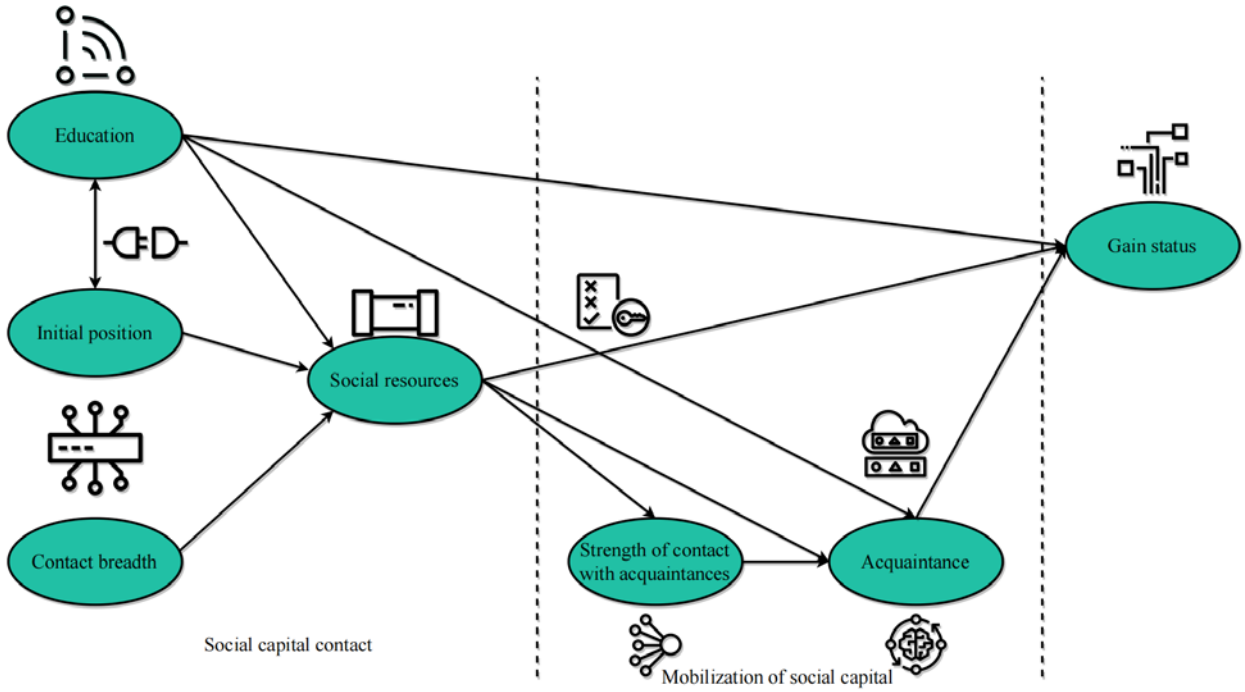

Fig.1 Social Capital and Status Acquisition Model

One process of this model focuses on the contact with social capital (the resources that an individual's self comes into contact with in a general social network). The assumption of this process is that status, together with education and initial status, will have a significant and important impact on the obtained professional status. On the contrary, the status of acquaintances is influenced by education, social resources and the strength of the connection between the individual self and acquaintances. The strength of the connection can be measured by perceived strength (such as the intimacy of the relationship) or role grouping (such as relatives, friends and acquaintances).

2.2 Systematic Analysis of Intelligent Analysis System of College Students' Employment Information

The system requirement analysis is the functional structure of the system, the basis of the process and the basic design of the successful system development. The requirement analysis is a very important prerequisite. When selecting the technical environment, the technical evaluation shall be carried out first. With the technical requirements of the two areas under consideration between business and data warehouse processing, the technical requirements of data in different environments and the management are also different. Therefore, under normal circumstances, the total analysis data is separated from the business data, and the analysis data is stored in a separate cluster, which is used to store the data warehouse. The aforementioned intelligent analysis system for college students' employment information sometimes does not need to use a data warehouse, but the data warehouse has been able to better prepare data for DM, and the knowledge mined better reflects the actual situation. Figure 2 shows the architecture of the data warehouse of the college students' employment information intelligent analysis system.

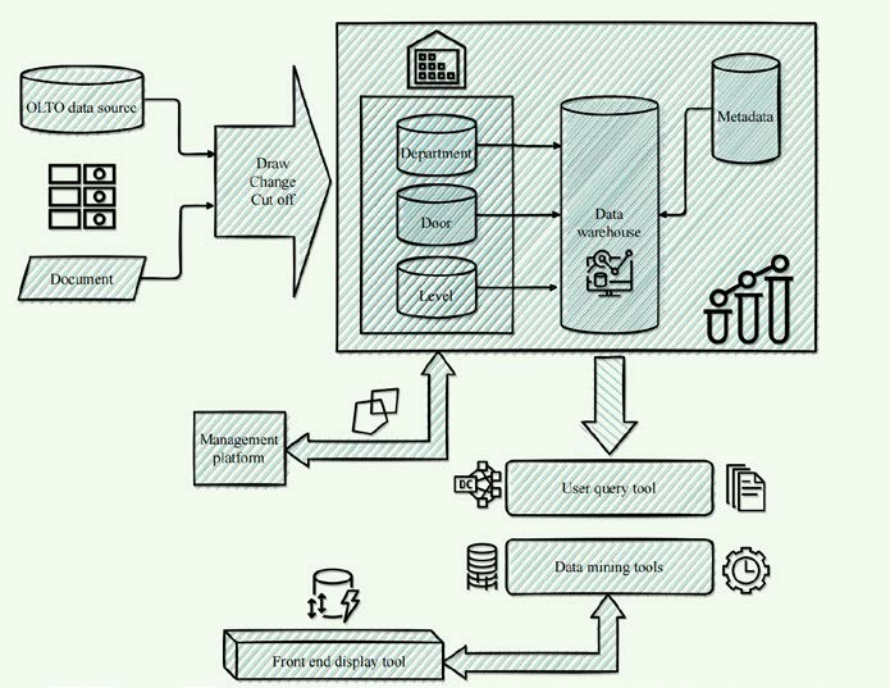

Fig.2 Data Warehouse Architecture of College Students' Employment Information Intelligent Analysis System

The source data of OLTP is extracted, converted, loaded, and transferred to the department level data mart. With the final formation of data warehouse metadata to organize ROLAP storage of relational data, users can access data directly from the data mart and the data warehouse of the management platform through query tools and DM tools.

If enrollment is the entrance of higher education, and factors such as enrollment scale and specialty setting affect the employment of college students in a potential way, then the management of college graduates' job hunting and employment is the export of higher education, which directly affects the employment of graduates. The employment guidance department of colleges and universities undertakes the main task of export management. The investigation and analysis of the work of the employment guidance department of colleges and universities is conducive to standardizing the export management of higher education, and is also conducive to the rational job search and effective employment of college graduates. The main contents of the employment guidance website generally include the information of graduates, employers, employment policies and regulations, employment guides, employment news, relevant notices, etc. The employment guidance departments of some colleges and universities regularly edit and publish their employment guidance manuals, interpret the latest employment policies, introduce job-seeking skills, matters needing attention in employment, etc. The Employment Guide for College Students regularly published by the Graduate Employment Guidance Office is the best publicity materials for college graduates in the region. The employment guidance departments of some colleges and universities regularly edit and publish their employment guidance manuals to interpret the latest employment policies, introduce job-seeking skills, matters needing attention in employment, etc. For example, the Employment Guide for College Students published regularly by the Graduate Employment Guidance Office of Guangxi University for Nationalities is the best publicity materials for college graduates in the region. At the same time, information correspondents can reflect the opinions, suggestions and demands of school employment on behalf of graduates.

3. Research and Method

3.1 Design and Implementation of Association Pattern Mining Model for Employment Information

With the expansion of the enrollment scale of colleges and universities in China, the problem of

students' employment has become increasingly prominent. Colleges and universities do everything possible to improve the employment rate. In addition to strengthening their own school conditions, university decision-makers want to know how to train students to improve the employment rate. Generally, colleges and universities have established student management systems, which hold a large number of historical data on student employment. How to find useful information from these data and provide it to decision-makers is our concern. Data mining technology is a good solution. It can transform some existing data into usable knowledge and mine valuable information about employment. Employment data has the characteristics of predictability and discreteness of classification. According to this feature, a decision tree algorithm is selected to establish a decision classification tree, and some rules of student employment are mined. Cluster analysis is conducive to dealing with random and subjective uncertainties, and achieves the classification effect. Factor analysis is a statistical method used to classify many influencing factors into several main indicators, so it is combined for the research and analysis of college students' employment satisfaction. The workflow of the model preprocesses employment data, selects decision attributes, implements mining algorithms and extracts rule knowledge, and points out which decision attributes determine the category of employment units by rule knowledge.

Decision tree method is one of the core algorithms of DM. It classifies a large amount of data purposefully to find out some potential and valuable information for decision-making. In which each node of the tree corresponds to a non-category attribute, each branch corresponds to each possible value of the attribute, and each leaf node of the tree represents a category. Decision trees can be easily converted into classification rules, and the path from the root to each leaf node corresponds to a classification rule. Starting from all the training samples at the root node of the tree, an attribute is selected to distinguish these samples. Each value of the attribute generates a branch, and the corresponding sample subset of the branch attribute value is moved to the newly generated child node. Let S be a set containing s data samples, and its category attributes take m different values, corresponding to m different categories $C_i(i=1,2,\cdots,m)$. Assuming that r_i is the number of samples in category C_i, the expected amount of information needed to classify a given data object.

$$I(S_{1j},S_{2j},\cdots,S_{mj}) = -\sum_{i=1}^{m} pijIb(pij) \quad (1)$$

Where: p_i is the probability that any data object belongs to the category C_i, and $p_i = \frac{r_i}{|s|}$.

Let a category attribute A take five different values of $\{a_1,a_2,\cdots,a_v\}$. The set S can be divided into V subsets $\{S_1,S_2,\cdots,S_v\}$ by using the category attribute A, where S_j contains the following samples; A data sample in which attribute A in the set takes the value of a_j. If attribute A is selected as the test attribute, that is, attribute is used to divide the current sample set, and set S_{ij} as the number of samples in subset S_j belonging to category C_i, the information needed to divide the current sample set by category attribute A can be calculated according to the following formula:

$$E(A) = \sum_{j=1}^{v} \frac{S_{1j}+S_{2j}+\cdots+S_{mj}}{|s|} I(S_{1j},S_{2j},\cdots,S_{mj}) \quad (2)$$

The information gain obtained by using the category attribute to partition the corresponding sample set of the current branch node is:

$$Gain(A) = I(r_1,r_2,r_m) - E(A) \quad (3)$$

Calculate the information gain rate of each attribute with formulas (1)~(3). Select the attribute with the highest information gain rate as the test attribute of a given set, create a node, mark it with this attribute, create a branch for each value of the attribute, and divide the sample.

The variance analysis of gender is shown in Table 1. The significant level of variance analysis in this case is P = 0.032c0.0, which indicates that it has reached an extremely significant level, indicating that there are significant differences in employment salary expectations among college students of different genders.

Table 1 Variance Analysis of Gender

	Sum of squares	Df	Mean square	F	Significance
Interblock	4.1652	1	3.223	5.258	0.032
Within the group	325.15	500	0.265		
Total	332.224	420			

Through correlation analysis, we can see that there is correlation between education background and five indicators of employment expectation. The correlation coefficient between education background and employment city type is 0.095, which is a weak correlation, indicating that education background has a certain impact on college students' choice of a city, but does not play a decisive role. The correlation coefficient between education background and the nature of the unit is 0.209, which is highly related. The higher the education background, the more likely they will choose state-owned enterprises or large enterprises, and the requirements for the nature of the unit are still high. The correlation between education background and employment salary is changed to 267, which is in line with the law of higher education investment income. The higher the education, the higher the cost of investment in human resources, and the graduates will hope to get higher remuneration. The awarding of college students is a very important factor, and this indicator is significantly linear with the other five indicators in employment expectation. The correlation coefficient r between awards and expected employment city type, nature of employment unit, salary expectation, development prospect of unit and individual, and individual potential exertion during school are 0.389, 0.133, 0.418, 0.170, 0.139 respectively, which are all positive correlation.

3.2 Data Statistics and Result Analysis

As for the measurement of graduates' individual human capital, respondents are required to make a pertinent evaluation of their own capital level in 11 core variables according to the actual situation. Through KMO test and spherical test, the KMO statistic is 0.904, the partial correlation among variables is weak, and Bertlett spherical test P=0.000, which indicates that the data is suitable for factor analysis. As the original data obtained by the investigation are a lot of chaotic ones, which can't be directly used for statistical analysis and inference, it is necessary to sort out the statistical data, that is, to review, group, summarize, describe and summarize the original data, so as to make them organized and convenient for statistical analysis and inference, so as to extract useful information from them and systematically and accurately reflect the essential characteristics and regularity. Multidimensional data sets are stored in cubes. Multidimensional data sets are divided into measurement values and latitudes. Measurement values are data used for statistics, and dimensions are the aspects from which statistics are made. It is the main object in online analytical processing, and it is a technology that can quickly access the data in the data warehouse. The core of multidimensional data analysis is to effectively calculate the aggregation on multiple dimension sets. In terms of terminology, these aggregations are called groupings.

Each packet can be represented by a cube, where the set of packets forms a lattice that defines the cube of the data cube. To realize the analysis of the number of employees, the first step is to establish a data cube based on the data warehouse. The number of employed people is analyzed from multiple dimensions, including gender dimension, class dimension, age dimension, school system dimension, student origin area, curriculum score dimension, skill certificate dimension, etc. Such multiple aspects of analysis can analyze the impact of different factors on employment in more detail, and what main impact aspects should be strengthened in management. With regard to the selection of rational employment units, it is impossible to see the characteristics of different groups of college students only from the overall sample. Therefore, more specific analysis is made below to clarify the correlation between them. The education background of the respondents will be introduced for discussion, and the results are shown in Figure 3 below.

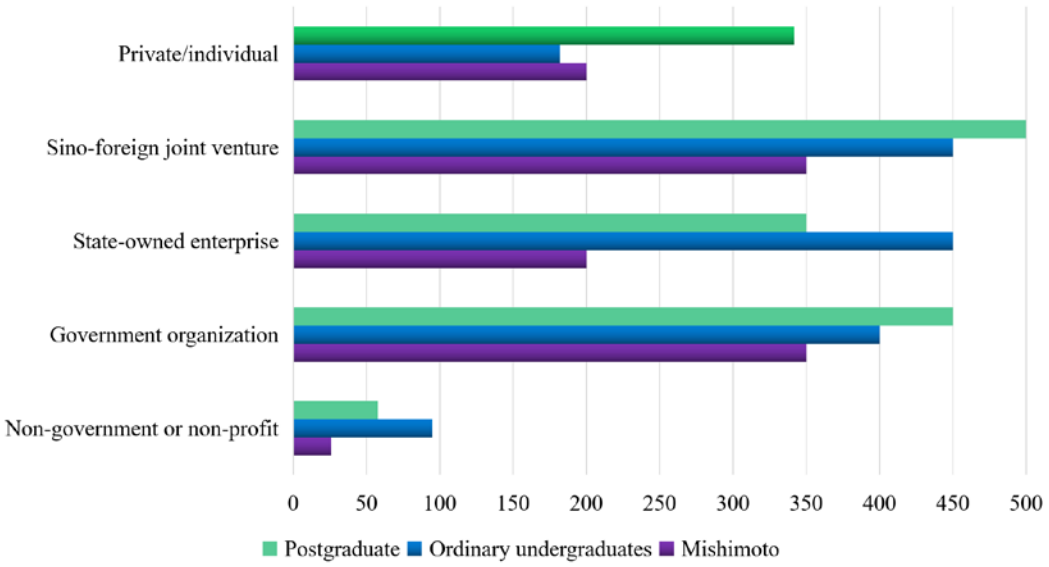

Fig.3 Proportion of Employment Units Selected by College Students with Different Degrees

As shown in Figure 3, the order in which the three students choose units is: Sino-foreign joint venture, private/individual, government agency, state-owned enterprise, non-government or non-profit. The choice of ordinary undergraduates is: Sino-foreign joint ventures, state-owned enterprises, private/individual, government agencies, non-profit or non-government. The expected employment units of graduate students are: state-owned enterprises, Sino-foreign joint ventures, private/individual, government agencies, non-government or non-profit.

According to the results of the correlation analysis, the six indicators of social capital are used as independent variables for regression analysis. The results show that the network size, network differences and ties with the knowledge layer all enter the regression equation, and the predictive regression coefficients of these three variables have passed the significance test at 0.05 level, indicating that the network size, network differences and ties with the knowledge layer have a significant regression effect on the perceived competitiveness of college students outside their primary employment units. As shown in Figure 4.

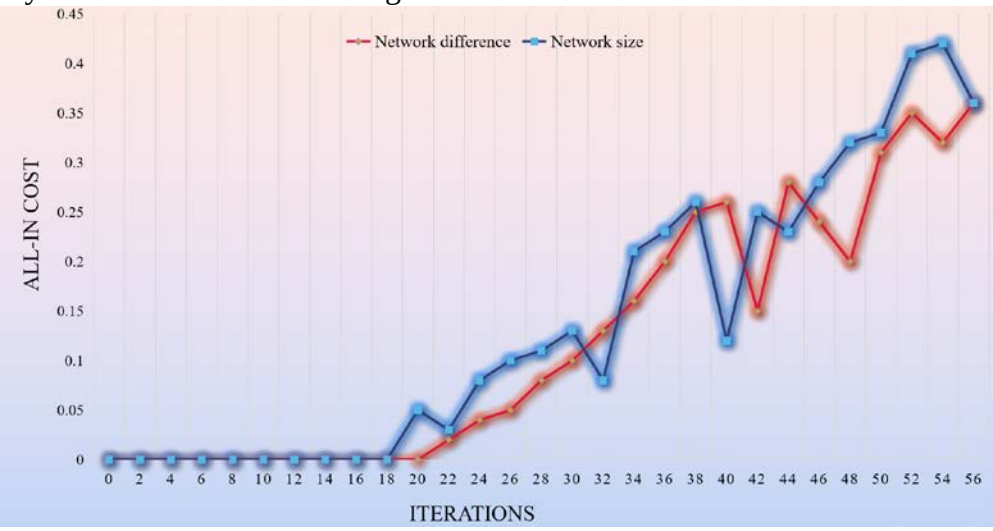

Fig.4 Regression Analysis Results of College Students' Social Capital on Their Perceived External Competitiveness of First-Time Employment Units

As can be seen from the figure, the network scale, network differences and ties with the knowledge layer have passed the significance test, which shows that these three indicators can predict the competitiveness of college students outside their first employment units. Among them, network difference has the greatest influence on external competitiveness, and it is the most important among these three forecasting variables.

4. Conclusion

The above statistical analysis shows that human capital and social capital factors play an important role in the current employment of college students in China. Specifically, for higher wages, college students' human capital plays a major role; However, both human capital and social capital have an important impact on obtaining employment opportunities, and the role of strong relationship in social capital is more obvious. In order to establish the employment data warehouse, this system has constructed a multidimensional data model of the data warehouse according to the characteristics of the employment data, completed the extraction, conversion and loading of the data. Based on the principle of simplicity and efficiency, we use the method of integrating and querying the data in the multidimensional data set of the data warehouse to display the results of multidimensional data analysis on the client. Through the practical application of data warehouse, this paper puts forward some suggestions on the employment of Beijing Polytechnic Institute. According to the results of cluster analysis, at present, college students' satisfaction with employment is generally not high, and 42/100 students are not satisfied with the above indicators. This requires all employment departments and universities to put graduates' employment on an important agenda, continue to open up new employment channels, further improve the policies to promote graduates' employment, and strive to form a long-term mechanism.

References

[1] Luo L, Guo M, Huang J, et al. Research on the Effect of an Entrepreneurial Environment on College Students' Entrepreneurial Self-Efficacy: The Mediating Effect of Entrepreneurial Competence and Moderating Effect of Entrepreneurial Education. Sustainability, vol.22,no.3,pp.15,2022.

[2] Liu X, Lin C, Zhao G, et al. Research on the Effects of Entrepreneurial Education and Entrepreneurial Self-Efficacy on College Students' Entrepreneurial Intention. Frontiers in Psychology, vol.10,no.5,pp.35,2019.

[3] Harrer MathiasAdam Sophia H. Baumeister HaraldCuijpers PimKaryotaki EiriniAuerbach Randy P. Kessler Ronald C.Bruffaerts RonnyBerking MatthiasEbert David D. Internet interventions for mental health in university students: A systematic review and meta‐analysis. International journal of methods in psychiatric research, vol.28,no.2,pp.33,2019.

[4] Yang J, Yifang H U. Research on the Influence Factors of College Students' "Internet+" Business Intention. Journal of Higher Education, vol.66,no.10,pp.20,2018.

[5] Xiaocui L I, Xing J. Research on the Vocational Interests of Tibetan College Students--Taking Tibet University as an Example. Journal of Higher Education, vol.446,no.11,pp.33,2019.

[6] Jie H U, Wei J. An Empirical Research on College Students' Needs and Expectations of Entrepreneurship Education Under Industrial Restructuring--Based on the Survey of College Students in Chongqing and Chengdu. Journal of Teacher Education, vol.47,no.3,pp.16,2017.

[7] Guo L, Ying S U, Feng J, et al. The Study on Influencing Factors of Major Satisfaction of College Students Based on Logistic Regression Model. Journal of Higher Education, vol.22,no.5,pp.70,2019.

[8] Eilers L, Paloyo A R, Bechara P. The effect of peer employment and neighborhood characteristics on individual employment. Empirical Economics, vol.63,no.10,pp.27,2022.

[9] Mao R, Xu J, Zou J. The labor force age structure and employment structure of the modern sector. China Economic Review, vol.52,no.1,pp.47,2018.

[10] Salami T, Lawson E, Metzger I W. The impact of microaggressions on Black college students' worry about their future employment: The moderating role of social support and academic achievement. Cultural Diversity and Ethnic Minority Psychology, vol.27,no.2,pp.19,2021.