The Application of Big Data Analysis in the Employment of College Graduates

Wenjie Zhang and Jinfeng Li

Nanchang Institute of Science and Technology, China

email: 568740355@qq.com, email: 524892454@qq.com

Keywords: Big Data; Talent Training; Graduate Employment; Recommendation Model

Abstract: The era of big data on the Internet is impacting the field of employment in Colleges and universities, which makes the work of employment guidance in Colleges and universities face more severe challenges. It also puts forward higher requirements for the mode and standard of talent training in Colleges and universities. In order to improve the level of talent training in colleges and universities in China, and accurately position the direction of talent training, the method of big data analysis is now applied to the work of employment guidance for college graduates, and at the same time, it constructs An open employment information recommendation model is built, which combines universities, students and enterprises.

1. Principle of big data analysis

The core idea of big data is to use mathematical algorithm to calculate the massive data and predict the possibility of things. Big data can be used to predict the employment of college students. Big data can be used to analyze and predict which form college graduates will choose for employment, such as working in companies or starting their own businesses.

Data is the basis of big data analysis. If the data can become the basis of decision-making on the basis of prediction, then the data will become resources. Under the background of this era, the connotation of information will be various forms and types of information, no longer the traditional text data. For the employment staff in Colleges and universities, to adapt to the new way of information generation and dissemination, the employment work must change from simple data statistics to digital intelligence, undertake the task of information production, analysis and interpretation with diversified media, and make use of We should know the data, tap the potential of the data, and enhance the systematization, pertinence and timeliness of the employment work in Colleges and universities. Big data analysis is only to find the correlation between the two, but not to find out what causes the results. Based on cloud computing, it is the essence of big data analysis to store a large amount of data information and fully analyze and calculate. This paper analyzes the employment information of college students, obtains the employment data several years later, and uses the analysis results as the basis of policy-making, so as to apply it to the employment of college students.

2. Application principles of big data analysis

We should give full play to the role of big data analysis in the employment of college graduates. The following principles must be followed.

2.1 We must rely on a large number of data to analyze the results

If there are not enough data samples, it is difficult to analyze them, or it can only be applied to specific areas, not to the wider areas.

2.2 Accuracy of data is not mandatory

The employment situation of college students is very complex, which can not ensure that every detail is very accurate. When the total data is linked to a certain extent, the accuracy of the quasi determination will decline, at this time there will be a new bright spot, that is, the relevance of the

data is placed in the same important position as the accuracy. In the traditional mode, the amount of statistical employment data is relatively small. If a data is wrong, it will lead to obvious deviation of the analysis results. However, in the era of big data, the accuracy of data is not so important, because when the total amount of data exceeds a certain node, it is more important to quickly obtain the outline and development trend of the research object than to pursue the accuracy of data Of value.

2.3 All graduates are the subject of information data

In the field of employment, we usually follow the principle of "sample = total" by collecting all data instead of part data. The precondition of big data prediction is scale. Based on this transformation, in order to avoid random sampling errors, data mining needs to be more in-depth.

2.4 Information data does not care about causality, but more about correlation

In the era of big data, research on data relevance is gradually reflecting its value. In order to understand the data more easily, clearly and quickly, we can analyze the data and the information related to the data. In order to understand the data development trend, we can use the relevance of the data, rather than the connotation of the data itself. We emphasize the use of the correlation between the data to make reasonable prediction. The perspective of relevance research is more comprehensive and clear than that of causality research.

3. The construction and application of big data analysis method in the employment of college graduates

3.1 The construction of big data analysis method in the employment guidance of college graduates

At present, there are many kinds of construction methods, but there are only two most suitable ones, namely self construction and external cooperation. Self owned construction can be in a certain region, with administrative region as a point, there are many such points in the country, and then connect these points together to form an independent network. In this way, we can not only analyze the data of a certain node, but also the data of the whole country, so as to make the results more constructive. In addition, it can establish its own database for some nodes, which reflects the convenience of big data application.

3.2 The application of big data analysis method in the employment guidance of college graduates

3.2.1 Data collection

The basis of big data construction is to collect the standard data of employment and the comprehensive quality data of graduates. At present, colleges and universities mainly collect the standard data of employment through the official recruitment website and telephone inquiry of enterprises. In addition, they collect the comprehensive quality of graduates by means of student status files, curriculum achievements, social practice evaluation and teacher-student evaluation.

With the development of campus information, the daily behavior data of students in the information system is basically used to analyze and evaluate the characteristics of students, such as the behavior track data in the campus WiFi system, the consumption data of students' all-in-one card, the borrowing data in the book borrowing system, and so on. It can also be used to analyze and evaluate the characteristics of students on the official website and micro website of the school the data of blog, search software and wechat are automatically connected to the basic information system of graduates, and the system is used to capture the data.

3.2.2 Data cleaning

The purpose of data cleaning is to filter out and process the useful data of enterprises and graduates from the massive data, so that the two-way selection between enterprises and graduates

can be carried out smoothly. This process is also to make the employment analysis work of colleges and universities in China go on steadily. By deleting the noise data, improving the construction of the basic employment standards of enterprises, and improving the effective development of the quality survey of graduates.

This process is very important. In addition to cleaning out sensitive data involving privacy in big data, it is also necessary to evaluate relevant data quantitatively, such as WiFi for students After quantifying the positioning data in the system, we can get the proportion of students' staying time in the teaching area, dormitory, library, sports area, off campus and other areas, so as to analyze whether the students are sports type or learning type or sleeping type.

3.2.3 Data usage

Using the method of big data analysis to serve the employment of college graduates means using the algorithm to convert the obtained data into the employment force of graduates, and then using the employment force to match the employment positions of enterprises to help graduates and enterprises find suitable positions and employees. Based on the above content, we can summarize the current employment standards of social enterprises and the relevant situation of graduates. These data can be used to study and judge the employment situation, so as to define the work standards in the employment guidance of college students. The employment guidance teachers in Colleges and universities can also make the students' job-hunting behavior by analyzing the relevant data of enterprises, so as to improve the employment rate of colleges and universities in China. With the development of the times, the integrated use of big data has become very important. For the teaching of college teachers, only by fully mastering the indicators of enterprise employment standards can we promote the full employment of colleges and universities.

4. How to obtain employment big data

4.1 Crawler technology

Web crawler is a program that automatically grabs information on the network through a program. The specific process is to automatically obtain the information needed in the web page and save it in the local database by simulating the way of browsing the web page manually according to certain rules. Then, change the access conditions of the uniform resource locator (URL), and the above process is executed in a circular manner until the task is completed.

4.2 Data acquisition algorithm flow

The working data mainly comes from online recruitment websites on the Internet. Take 51job.com as an example to compile the web crawler. The data acquisition process is shown in Figure 1. Using the third-party library htmlagilitypack based on. Net, this paper crawls 4864 working data related to the employment of Geographic Information Science from 50 pages of the Internet. Htmlagilitypack is a third-party visual studio plug-in that integrates the functions of web page access to obtain return values, parsing HTML documents and so on. It is also an HTML parsing plug-in based on C ○ɔ mainstream development language at present. Read the working information of each page through the URL, store the working data of each page into the array, write them into the database in turn, change the page number value in the URL, and repeat the above steps to complete the data acquisition of the future worry free network.

4.3 Batch address decoding and visualization

The location data of universities and work units obtained from China Education online and employment information websites are stored in the form of text address. The addresses containing text geographic information are transformed into longitude and latitude coordinates. The batch address decoding function in Baidu map API is used for batch address decoding, and the crawled batch address encoded data is imported into ArcSDE spatial database. Because it is too messy to directly load a large number of point data on the map, in order to more succinctly display a large

number of point data on the page, we can choose kmeans clustering center point method, and use the open-source ESRI cluster feature layer to cluster and display all kinds of point set big data in the map.

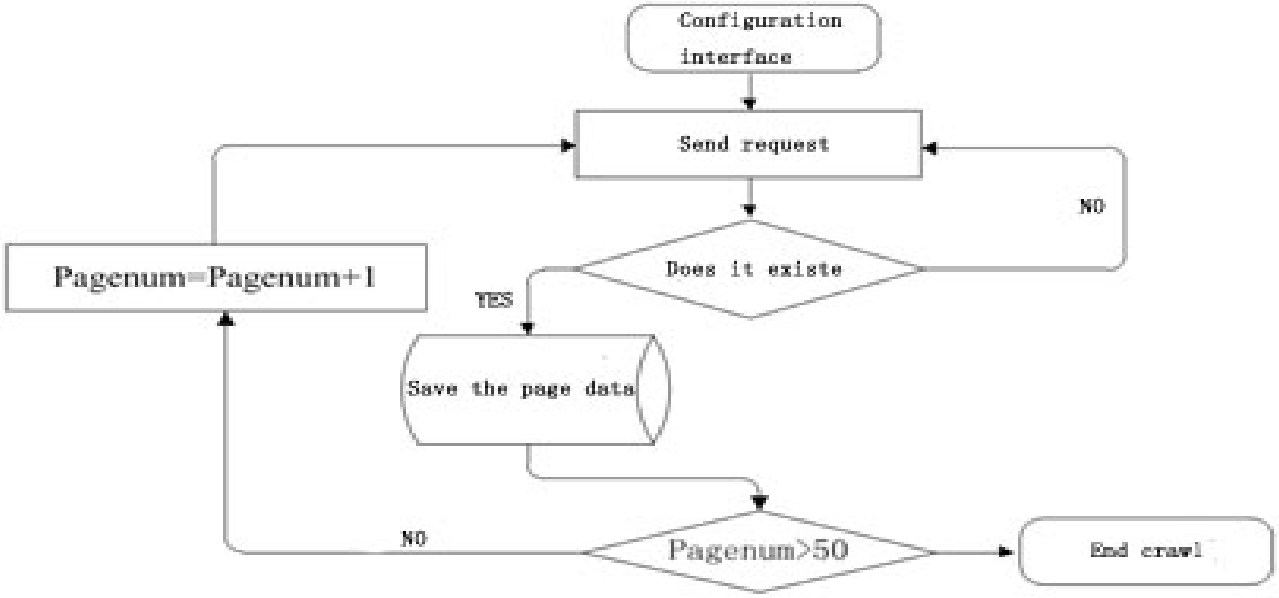

Figure 1. Algorithm flow of obtaining employment data

5. Analyze big data and build model

5.1 Work recommendation model based on multiple linear regression

In order to understand the characteristics of graduates who adapt to various jobs in society, a matching model between working conditions and individual conditions of employees is constructed, as shown in Figure 2.

In Figure 2, work is a dependent variable and personal condition is an independent variable. Take various conditions of work as a whole, including salary, welfare, bonus, address, etc., and quantify them according to certain rules to get a quantitative value, i.e. salary reference value. The independent variable of user's personal conditions is affected by many factors, some of which have little change, some of which have great change, and the degree of change is different. For example, the school ranking is 1-200. If the initial value is used directly, the proportion of each factor will seriously deviate from the actual situation and affect the accuracy of the results. Therefore, all factors are normalized and standardized to make all data between [0,1], and the formula is as follows:

$$X=（x\text{-}Min）/（Max\text{-}Min）$$

In the formula, Max is the maximum value of each group of sample data, and min is the minimum value of each group of sample data. Through the calculation of a large number of real data, the proportion value of personal conditions to work reference value is obtained, and then the model is constructed to recommend the corresponding work to users. The specific construction steps are as follows.

5.1.1 Construct salary reference value

Reference value = sample monthly salary - local five insurances and one fund + supplement

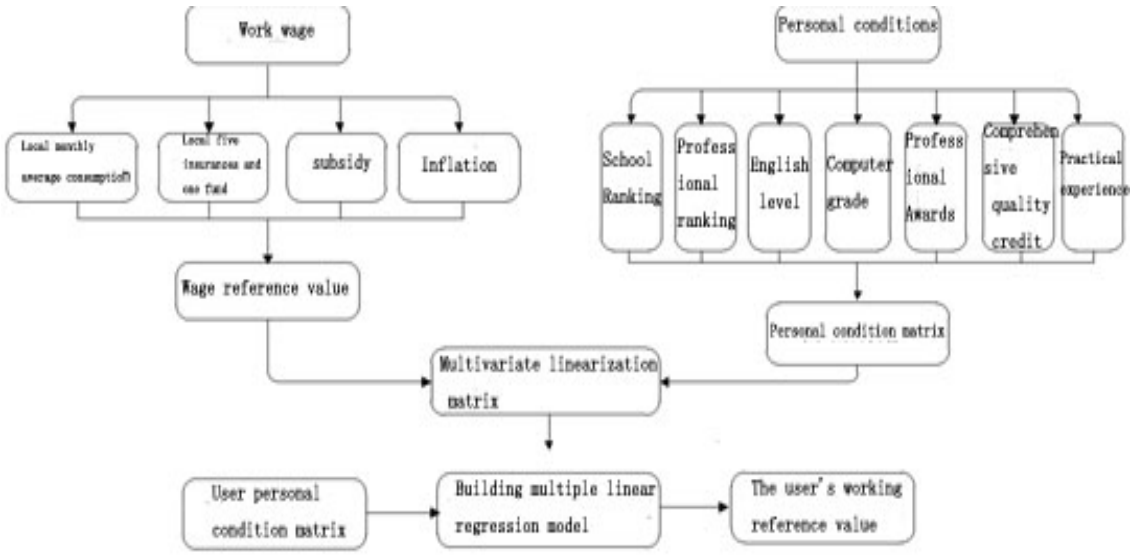

Figure 2. Work recommendation model

2. Building personal condition matrix

5.1.2 Building personal condition matrix

By collecting the employment situation of college graduates for many years, and then constructing the personal condition matrix according to the conditions of graduates. In order to quantify the personal conditions more reasonably, the condition matrix of users is constructed according to the entry time of users while the user's practical experience is added. The user's graduation College ranking, major ranking, English level, computer level, scholarship, professional award experience and other conditions are added into the condition matrix. The normalization equation is used to normalize each group of data in the sample. The processed value is the user's personal condition matrix.

5.1.3 Building multiple linear regression model

The multiple linear regression model is constructed by the work reference value and personal condition matrix successfully found by the user, as shown in formula (2). The input user condition matrix is used to predict the work within plus or minus 100 of the output reference value range. Finally, the appropriate work is recommended to the user according to the work direction of the user intention work.

$$Y=aX_1+b\ X_2+cX_3+\ldots\ldots$$

Where y is the reference wage value of the sample, x1, X2, X3 Is the personal condition matrix vector of the sample, a, B, C Is partial regression coefficient.

5.2 Key words extraction method of word segmentation

In order to let users understand the basic situation of each industry and show the hot requirements of related work in each industry, a large number of enterprise introduction and work requirements data need to be obtained through in-depth analysis. Because these data are all text types, we can't get rid of useless information directly, and get the development direction of each industry and the general requirements of work. Chinese word segmentation is a basic work of Chinese natural language processing, which plays an important role in information retrieval, text automatic classification and data mining. At present, the mainstream Chinese word segmentation technologies mainly include Jieba, thulac, snownlp and so on. Among them, Jieba word segmentation technology is widely used because of its powerful word segmentation ability. This model uses the Jieba word segmentation module to cut the complete sentences into a phrase set, screen and count the concentrated phrases, and display the high-frequency phrases in the company introduction and work requirements. The algorithm flow is shown in Figure 3.

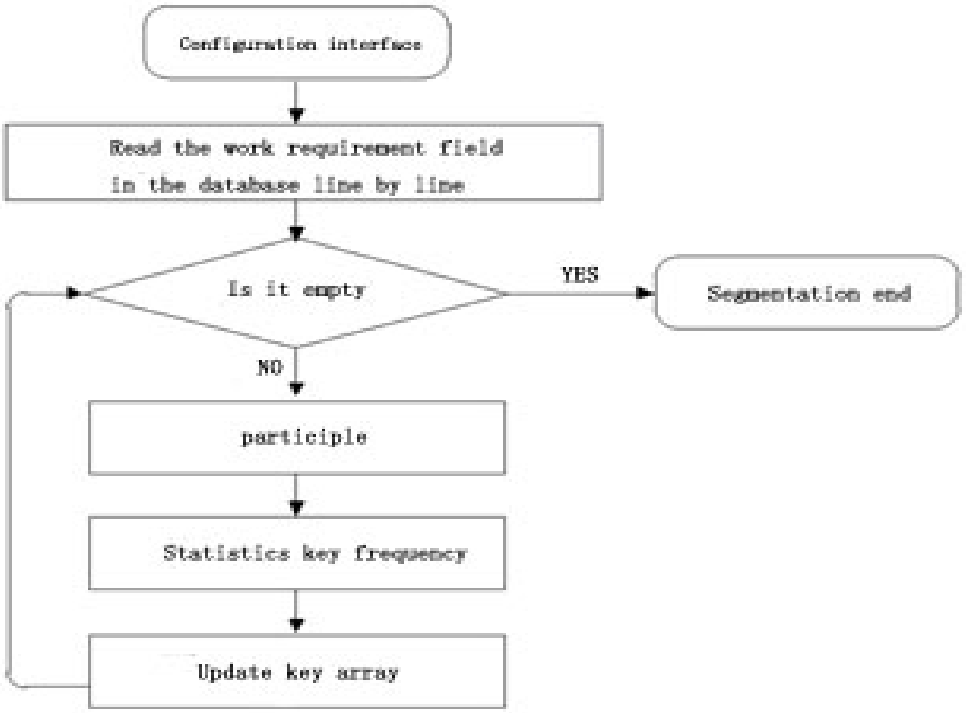

Figure 3. Algorithm flow of high frequency keyword extraction from segmentation

Summary

It is an inevitable trend of the times to apply the concept and method of big data in the field of college employment. We need to further explore the key technologies of data collection, storage, mining, analysis and decision-making. At the same time, we should make it clear that the practical significance of big data analysis is to provide reference and services for students' employment work, and its purpose is to provide empirical support for our decision-making, which can not replace artificial decision-making. Only by strengthening the study and research of big data theory, can we better integrate big data analysis and employment work, and promote the professional, personalized and systematic employment work in Colleges and universities. Colleges and universities should actively implement the requirements of national education opening to the outside world, give full play to the teaching characteristics and advantages of Chinese colleges and universities, take undergraduate education, graduate education, vocational education, school enterprise talent docking as the main content, cooperate with local colleges and universities to cultivate local talents and further expand education cooperation with foreign countries, while training professional talents in the industry, implement precise development We will build a "community of common destiny" for higher education, promote exchanges and cooperation in science, technology, culture and education between China and overseas industries, and serve the overall situation of China's diplomatic and economic development.

Reference

[1] Victor Meier scheenberg. Big data age [M]. Translated by Sheng Yangyan and Zhou Tao. Zhejiang People's publishing house, 2012.

[2] Wang Lei, Chen Min. Carrying out graduate employment guidance in different stages under complex employment environment [J]. brand (theoretical monthly), 2011 (5): 73.

[3] Lu Kui. The transformation of education brought by the era of big data and the enlightenment to the ideological and political education of college students [J]. China Science and education innovation guide, 2014 (4): 237-238.

[4] Chen Shiqi. Analysis of the employment challenges and coping strategies of college students in the era of big data [J]. China business, 2015 (5): 193-195.

[5] Victor Meier scheenberg. Big data age [M]. Translated by Sheng Yangyan and Zhou Tao. Zhejiang People's publishing house, 2012

[6] Xue Zheng. Some thoughts on strengthening the individualized guidance of College Students' employment [J]. Heihe journal, 2013 (196): 128-130.

[7] Chen Hongmin. Application of big data to promote full employment [J]. Modern management. 2014 (6): 76-77.

[8] Li Fengchun, Yang Bo, Zhang Lin. research on the application of information network in the employment and entrepreneurship guidance of graduates under the big data environment [J]. Information science, 2015, 33 (12): 111-115.

[9] Hou Jun, Yao Sihan, Shen Yanbing. Exploration and practice of graduate employment precision service system in the context of big data [J]. School party building and ideological education, 2018 (5): 69-71.

[10] Chen Zhixu. Analysis on the use of big data to achieve accurate employment of college students [J]. School party building and ideological education, 2017 (8): 81-83.