The Application of Data Mining Technology in Employment Analysis of University Graduates

Hui Yu
School Office of Northwestern
Polytechnical University
xi'an 710072，China
E-mail:yuhui@nwpu.edu.cn

Zhong-qiu Zhang
Ming de college of Northwestern
Polytechnical University
xi'an 710129，China
E-mail:281431225@qq.com

Abstract—**The employment of university graduates is the last link of university student education and a new starting point for every student's career. It is a very important thing for both the students themselves and the university. Therefore, it is an important task for colleges and universities to guide students to achieve high-quality employment, the whole school needs to be involved and complete. The core of data mining is to search useful data from data-intensive algorithms.In this paper, data mining technology is applied to analyze the employment data of university graduates to find out useful information and relationships, so as to promote student education and employment guidance.**

Keywords—Data mining;student education; Mining models; Association rules

I. INTRODUCTION

Student education is the fundamental task of a university. The quality of student education directly affects the reputation and social influence of a university. In fact, the entire cycle of talent cultivation includes three major aspects:enrollment, training and employment. Each university attaches great importance to the first two links. For example, each year, students and their parents will check the network through the internet to take scores in recent years and choose college entrance voluntary. This requires that colleges and universities must attach importance to enrollment, pay attention to college entrance examination scores analysis, so as to improve enrollment strategies, with a view to gather more outstanding students. For another example, colleges and universities will spend a lot of resources and energy on the revision of student education programs, the reform of education, the updating of teaching methods, and so on, putting forward practical standards for student education and promoting the quality improvement of student education. Although colleges and universities report the annual employment situation in accordance with the requirements of the Ministry of Education, the Employment data have not been well analyzed in depth. Therefore, strengthening the employment data mining [1] is an important measure of promoting student education and employment guidance. The traditional method of data analysis is difficult to accurately analyze the massive information. In the context of big data, making full use of data mining technology is an important technical means to do a good job in this area.

II. DATA MINING TECHNOLOGY INTRODUCTION

Data mining, also known as knowledge discovery in the database, is a process that uses various analytical tools to discover the relationships between models and data in massive data. These models and relationships can be used by companies to analyze risks and make predictions. In the field of enterprise applications, data mining technology can be used to support key decision-making of enterprises and the formulation of market strategies. For example, it has good applications in the fields of commerce, finance, and insurance. In business applications, data mining technology can enhance the company's competitive advantage, shorten the sales cycle, reduce production costs and help formulate strategies. In the application of scientific research, data mining technology can extract useful information for humans from a large number of irregular data.

A. The Features of data mining

Data mining is an interdisciplinary subject which involves many technical fields such as machine learning, neural network, pattern recognition and database. Knowledge discovered through data mining is usually expressed in the form of concepts, rules, patterns, constraints and visualizations. This knowledge can be provided directly to policy makers to assist in decision-making [2]. As a cross-disciplinary, data mining has its own unique features, which are also generated by cross-disciplinary.

- When data mining and machine learning are combined, the characteristics reflected are the emphasis on the characteristics and distribution of data. There are strict principles and methods, mainly reflected in the application of machine learning in large databases.
- When we are doing data mining, we need to get data from the data warehouse and then put it into a special database or data mart for mining. This feature can be summarized as indirect.
- Data mining and statistics are important methods for model discovery and prediction and have been widely used. In contrast, data mining is data-driven and more

IEEE computer society

suitable for information acquisition in the context of big data.

· Compared with intelligent decision-making, data mining is a problem-oriented comprehensive data utilization technology with predictability.

B. The process of data mining

Data mining process includes data preparation, data preprocessing, data mining and interpretation of results interpretation process. And these processes are constantly repeated, as shown in Figure 1.
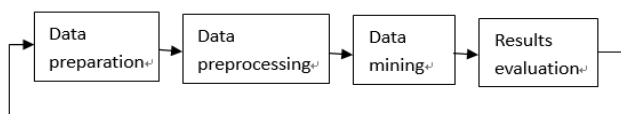

Fig 1.The Process of Data Mining

· Step 1: Data preprocessing. According to the analysis object, abstract the required feature information in the data analysis, and then select the appropriate information collection method and store the information in the database. Data from different sources, formats, and characteristics are organically integrated logically or physically to provide comprehensive data sharing for companies.

· Step 2: Data preprocessing. Data preparation usually includes data specification, data cleaning, and data transformation. The data reduction technique is used to get a representation of the data set, which is much smaller than the original data but maintains the integrity of the original data. The task of data cleaning is to process incomplete, noisy, inconsistent data, and store complete, correct, and consistent data information in the database. Data transformation is to convert data into a suitable form for data mining through smooth aggregation, data generalization, and normalization.

· Step 3: Data mining. According to the data information in the data warehouse, select appropriate analysis tools, and apply statistical methods, decision trees, fuzzy sets and other information processing methods to obtain useful information.

· Step 4: Results evaluation. On the one hand, the correctness of the mining information is verified, and on the other hand, the acquired information is presented to the user in a visual manner.

The data mining process is a repeated cycle process. If each step does not reach the desired goal, it needs to go back to the previous steps, and readjust and execute it.

C. Algorithms involved in data mining

Data mining is not only a user-centered, human-computer interaction exploration process, but also the comprehensive use of different algorithms and the repetition of a variety of professional knowledge. With the continuous development of data mining technology, data mining methods are also more and more, there is no way to deal with all users of data mining to go, it is because no algorithm can accomplish all kinds of data mining tasks at the same time [3]. Several methods commonly used in data mining include Decision tree algorithm, Clustering Algorithm and Association Rules Algorithm.

1) Decision tree algorithm [4]*:* It is the classification and regression algorithms provided by SSAS, mainly used for predictive modeling of discrete and continuous attributes. For discrete attributes, the algorithm predicts with the relationship between input columns in a data set. For continuous attributes, the algorithm uses linear regression to determine the split position of the decision tree. Decision tree model must contain a key column, a number of input columns and a predictable column.

2) Clustering Algorithm: It is the segmentation algorithm provided by SSAS, which use iterative techniques to group similar feature data instances. The clustering algorithm provides two methods to calculate the fitness of the points in the classification, including Expectation Maximization and K-Means. Clustering analysis model must contain a key column and a number of input columns, while also support specific input column content types and predictable column content types.

3) Association Rules Algorithm: It uses two parameters to illustrate the itemsets and the rules generated by the algorithm. The parameters are support and probability. The relevance algorithm uses the importance parameter to describe the importance of the rule. The association model must contain a key column, multiple input columns, and one predictable column, and the input column must be a discrete column.

D. The structure of data mining system

The structure of data mining system generally includes five parts, as follow.

1) Database and data warehouse: It reflects that data mining objects consist of information such as databases, data warehouses [5], and data forms. Data cleaning and data integration operations are often required to perform preliminary processing of these data objects. The database or data warehouse server is responsible for reading the relevant data according to the user's data mining request.

2) Knowledge base: The knowledge base is used to store the relevant knowledge needed for data mining. This knowledge will be used to guide the data mining search process or to help evaluate the mining results.

3) Data Mining Engine: It is the most basic part of the data mining system and usually includes a set of mining function modules to complete the mining functions such as qualitative induction, correlation analysis, classification induction, evolution calculation, and deviation analysis.

4) Pattern knowledge assessment module: The module can help the data mining module focus on mining more meaningful model knowledge based on interesting criteria. At

the same time, whether the module can be integrated with the data mining module is related to the mining algorithm used.

5) Visual user interface: It can help users communicate with data mining systems. On the one hand, users can submit their own mining requirements or tasks to the mining system through this module, and provide relevant knowledge needed for mining and searching. On the other hand, the system can show or explain the results of data mining or intermediate results to the user through this module.

III. APPLICATION OF DATA MINING TECHNOLOGY IN EMPLOYMENT ANALYSIS OF GRADUATES IN UNIVERSITY -A

The article applies data mining technology to the employment data of the 2017 graduates in University-A, focusing on the analysis of the employment destination of graduates. Meanwhile, make them evaluate the employment of the graduates through the questionnaire survey of employing units.

A. Data mining model

In order to more clearly reflect the whereabouts of college students, according to the standard set by the Ministry of Education on the statistics of employment information, the contracting units were classified according to the type of industry, the type of area and the type of city in this article.

1) Data preparation: The employment data for this paper come from University-A's annual report on the 2017 graduates' employment released on the Internet. According to the data mining step, the required data are first extracted, cleaned and verified. At this stage, it needs to establish a basic information table and employment information form and build employment view structure table. And the Employment view is shown in Table 1.

TABLE.1 The Employment View

Field name	The table	Field description
Xh	xjjbxx	Student ID
Xm	xjjbxx	Name
Xb	xjjbxx	Gender
Yxdm	xjjbxx	Department code
Xsh	xjjbxx	Department name
Zydm	xjjbxx	Professional code
Zyh	xjjbxx	Professional title
Dqszj	xjjbxx	Grade
Syszd	xjjbxx	Health source location
Dwmc	Jiuye	Company name
Dwdz	jiuye	Unit address

2) Design of Data Mining Model: Data mining model design includes multidimensional data set design, decision tree mining model design and association rule mining model design.

a) Multidimensional data set design. Star mode is one of the ways in which multidimensional data can be rendered. It represents the data as a collection of two types of data: facts and dimensions. The center of the star is the fact table which stores the data to be examined, and the dimension table

outside the fact table. This article needs to analyze the students' basic information, employment type and employment destination. Therefore, the student employment information is used as a fact table, faculty name, professional name, grade, employment categories and other industries is used as a dimension table.

b) Mining model design. Here mainly used decision tree mining model design and association rules mining model design. According to the algorithm requirements and data mining model analysis, the Employment view is selected as the key table and nested table of association rules mining, and the key table is used as the mining of decision tree. At the same time, the usage of the columns is also separately stipulated shown in Table 2 and Table 3.

TABLE.2 The Columns Usage Of Association Rules Mining Model

Field name	Column usage
Dwtype	Enter the column
JG	Enter the column
Province	Prediction column
XH	Key column
Zy	Enter the column
XB	Enter the column

TABLE.3 The Columns Usage Of Model

Field name	Column usage
City_type	Prediction column、Enter the column
Dwtype	Enter the column
JG	Enter the column
Position	Enter the column
Province	Enter the column
Xh	Key column
Zy	Enter the column

3) Data mining results:

a) The result of Multidimensional data set: The article respectively describes the employment data according to the dimension model of "department-grade", "department-employment unit industry category", "profession-employment unit industry category" and "province of employment unit-employment unit industry type" that reflect the relationship between each other.

b) The result of association rules mining model: Through the mining of employment data, we get the relationship between the attributes of decision-making attribute set of the industries such as sex, origin, profession and employment unit and the attribute of the category of province where the employer is located.

c) The result of decision tree mining model: According to the demand analysis, the "employer-city category" is selected as the category identifier attribute, and "unit category", "student source", "area type of employment unit", "province where employment unit is located" are selected as the decision attribute set. The result shows that the urban categories of employment units are respectively related to the

types of units and regions, the provinces where the units are located, the types of units, student majors and students' origin according to the order of dependency intensity.

B. Analysis of data mining results

Based on the above data mining pattern, the employment data of the graduates of 2017 University-A have been deeply tapped and analyzed. Related results are analyzed as follows.

1) Analysis of the type of contractor industry: The results of the analysis show that there are differences in employment areas at different academic levels, and the proportions of the same fields are different. Sorting by the high and low proportions, undergraduates mainly to private enterprises and state-owned enterprises, master's graduates mainly to private enterprises, research and design institutes and state-owned enterprises, doctoral graduates mainly in the nature of employment units in universities and research institutes. There are two reasons for this situation, one is based on the students' objective judgments on their own strength; the other is the requirements of the employer on the academic qualifications of graduates according to the nature of the work. Colleges and universities can help students master these situations and help them to establish their own development goals.

2) Geographic distribution of contracted units: Graduate employment units are all over China, but there are relatively concentrated areas, provinces and cities. In terms of regions, the top two regions are the northwest and east of China. In terms of provinces, the top four provinces are Shaanxi, Guangdong, Beijing and Shanghai. In terms of cities, Xi'an is the preferred city for graduate employment, while Beijing, Shenzhen and Shanghai are the cities where graduates choose relatively more. There are two reasons for this. One is that University-A is located in Xi'an, and students have fallen in love with the city after several years' study and learning. Meanwhile, Xi'an now enjoys a strong momentum of development and offers many opportunities for young people. Secondly, Beijing, Shanghai and Guangzhou are places where all university students want to fight. These cities are highly inclusive, open-minded and have broad prospects for development.

3) Relationship between native place and employment area: As University-A is for enrollment across the country, so students come from all 31 provinces and autonomous regions in China. Thus, the relationship between student's native place and employment area is also one aspect of employment data mining. The results showed that students from Jiangxi Province preferred units of Jiangxi and those from Chongqing students preferred units of Chongqing and Sichuan. This is a great inspiration for the school employment service. One efforts of the school's employment work is to match the students with the enterprises in their native places so as to improve the pertinence and effectiveness of the employment work.

4) Factors affecting employment units: Students' choice of employment units involves many factors and is a complicated matter. For example, there is a certain relationship with the type of employer, province, student major, academic level and native place. Despite many factors involved, the strength of each factor's dependence is different, some degree of academic impact larger, some professional influence larger, and so on. In short, the situation of each student is different.

5) Employers feedback: The questionnaire survey shows that employers are satisfied with the employment service of University-A and spoke highly of the overall quality of graduates. University-A students' study majors, academic performance, quality of consciousness and social practice is the emphasis of employer. At the same time, employers hope that students in University-A need to further strengthen their knowledge structure, research ability and innovation ability. Employer feedback focused on two important information. One is to summarize the experience and practice of the employment service in University-A and form the traditional characteristics to persevere. The other is to combine the concerns of employers and the shortcomings of University-A graduates to reflect on the improvement of student education, and strive to develop more comprehensive and excellent students.

IV. CONCLUSION

This paper uses data mining technology to analyze the employment data of the graduates of 2017 University-A. Some useful information is excavated from the student's employment data that will be helpful to further the employment of graduates and personnel training. This information includes the distribution of student employment fields, geographical distribution, and factors affecting employment selection. This study has a strong practical value and promotion value, as for other universities.

REFERENCES

[1] Jiawei Han, Micheline Kamber. Data Mining: Concepts and Techniques. Higher Education Press, 2002.

[2] Wang Lin. Research on College Students Training and Employment Guidance Based on Data Mining. Southwest Jiaotong University, 2006

[3] Wang Chenxi.The Design and Implementation of University Students' Comprehensive Quality Evaluation System Based on Data Mining.University of Electronic Science and Technology, 2011.

[4] Lei Songze, Hao Yan. Employment data mining based on decision tree. Journal of Xi'an Institute of Technology 2005(10): 429-432.

[5] Xu JiePan. Data warehousing and decision support systems. Science Press. 2005.