Research and Application of Data Mining in College Students' Employment Based on Association Rules

Liu Lijuan[1] & Zhai Changliang[2]

[1] School of Information Engineering, Nanjing Normal University Taizhou College, Jiangsu, China

[2] Taizhou Public Security Bureau, Jiangsu, China

Correspondence: Liu Lijuan, School of Information Engineering, Nanjing Normal University Taizhou College, Jiangsu, China. Tel: 86-139-5261-9696. E-mail: 149110375@qq.com

Abstract

This paper takes the employment information data of university graduates as the research object, and takes Apriori algorithm as the main idea, develops a data mining system for early warning of student employment information and new functions of teacher guidance and employment, finds out some factors that affect the employment situation, and realizes the early warning to the students who may not be able to obtain employment smoothly and the teachers who may not have an ideal employment rate. So that managers can obtain more valuable knowledge and information, better management, improve employment rate, and enhance the competitiveness of institutions of higher learning.

Keywords: association rules, apriori algorithm, employment, data mining

1. Introduction

With the advent of the big data era, all colleges and universities have applied various advanced information systems to manage information, and employment is no exception. However, most of the employment management systems(Liu Yuhua, Chen Jianguo, & Zhang Chunyan, 2015) currently used, only provide some simple functions, such as information entry、modification andquery. How to dig out effective information from a large number of employment data, and further to guide the professional settings、 curriculum arrangements and employment management in Colleges and universities has become a problem worthy of study.

In this paper, the Apriori algorithm is used to study the employment information data of college graduates based on the theory of data mining, and the relevant data are analyzed. Finally, the meaningful rules are excavated from a large number of historical employment information to guide students in the university, pay attention to their comprehensive assessment of average results、participate in extracurricular social practice、 increase internship opportunities and participate in professional training outside the school etc.; teachers can carry out employment guidance in a targeted way according to the basic information of students. The management departments of colleges and universities can adjust their professional structure and training mode in a timely manner so as to promote the good development of higher education.

2 Data Mining Definition and Technical Analysis

2.1 Related Definitions

Data mining(ZhouXinshao,2014)is a high-level process of sorting out the implicit、 effective and credible information from massive amounts of data. It is a technology emerging in recent years with the development of large data and artificial intelligence.

There are many techniques of data mining, such as association rule analysis、data statistic analysis and decision tree etc., association rule analysis is the most frequently used and widely used one. Association rule mining (Ji Shanshan, Li Shufei, & Jiang Wuxue,2016) is based on a specific data set to mine the dependencies between data satisfying certain conditions. Assume that the data set is D, and each attribute of each record in the collection is recorded as a single item. With the association rule technique, at the same time to meet the set of items $X \Rightarrow Y$, the support degree is s%, and the confidence degree is c% , the association rules for the data set D are found.

Support refers to the strength or frequency of association rules, indicating the probability of recurring records in

different sets, expressed as support (). Confidence(Lu Xiaohua, & Liu Jing,2016) is a measure of the credibility of the association rules, it reveals the probability that another item will appear and appear when a project appears, which is expressed as confidence ().

2.2 Apriori Algorithm

Apriori algorithm (Zhang Liang,2016) is a classic data mining algorithm, by scanning the database several times and substituting it repeatedly, the latter set is generated on the basis of the previous set. The pseudo code for the Apriori algorithm is shown below:

Enter: transaction database D (Wang Xin,2006); minimum support threshold min_sup.

Output: FrequentitemsetsL in D.

method:

L_1 = find_frequent_1-itemsets (D);// first find all the frequent 1 items

For (k = 2; L_k-1! = Null; k + +)

{

   C_k = apriori_gen (L_k -1, min_sup);// loop generation candidate while pruning

For each transaction t in D// Count the D scans by looping

  {

    C_t = subset (C_k, t);// call to get a subset of t

    For each candidate c belongs to C_t

    c. count ++;

  }

  L_k = {c belongs to C_k | c.count> = min _ sup}

}

Return L = get all the frequent sets;

First, the algorithm generates a set of all frequent first entries, denoted as L1, and iteratively generates a set of second items on the basis of L1, denoted as L2, again, iterates over the L2 to generate L3, and continues to cycle in accordance with such rules until you can't find any frequent K item sets. When the candidate frequent itemsets are obtained, strong association rules can be generated for decision making.

3 System Structure and Design

3.1 System Structure Design

The system uses visual tools Visual Basic to develop user interfaces, and uses ADO to access the data of SQL Server database. The VB interactive interface is used for high-level interactive mining, which ensures the user's ease of operation.

Data acquisition is the first step of data mining, and this time mainly uses structured data of students' basic information.

Data preprocessing(Li Wei, Liu Guangming, Meng Xiangfei, & Zhang Zhenfa,2016) is divided into several steps, such as data cleaning、data integration、 data conversion and data reduction. Data cleaning can remove noise and reduce inconsistencies in data; data integration can combine data from multiple sources into unified data for storage; data transformation can standardize data; data reduction can compress data by means of aggregation、deletion redundancy and clustering. The design uses clustering to improve the Apriori algorithm, reduces the data repeatability, improves the time efficiency, and enhances the readability of the data. The structure of the whole system is shown in figure 1.

3.2 Data Table Design

The database platform of this system uses SQL Server to establish the basic information database of employment students. The database stores several data tables based on employment information, which involves many years'historical data of college students' employment information. According to actual demand, these data can be excavated、 analyzed and processed. It contains 10 two-dimensional tables, such as student basic information、 student achievement、 student award information、 student employment information、 employment information of department, etc., stores about 60 million tuples of data, sets the support and confidence, creates the association

rule base(Judith,2004,pp.67-77), uses the Apriori algorithm to create the association rule two-dimensional table. Through the screening of the original data, the student employment comprehensive information summary table with the student number as the primary key is obtained, which prepares for the data mining.
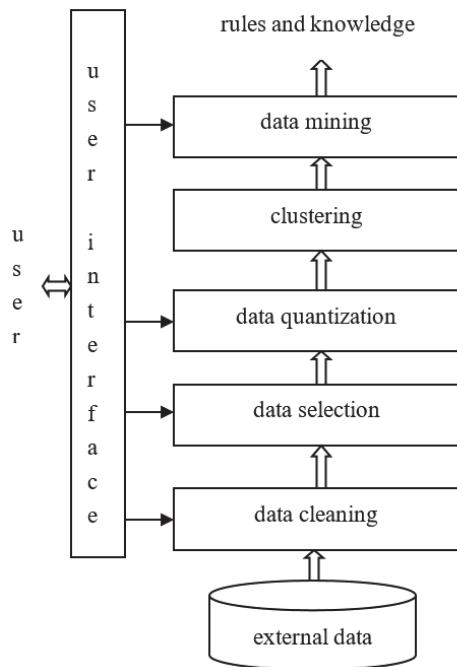

Figure 1. System structure diagram

4. Data Mining Application Analysis

4.1 Data Preparation

Taking the 1128 graduates of our school as the research object, through the analysis of the corresponding data mining algorithm, we find out the internal relations between the factors that affect the employment rate. It provides data decision support for the students' successful employment、 the school's employment rate and the quality of education. Table 1 gives the student employment comprehensive information summary table with the student number as the primary key.

Table 1. Summary of student employment information

No.	Gender	Major	Score	Win the prize	Family	Obtain employment	…
01080104	male	law	85	yes	rich	go abroad	…
05080305	female	primary-school education	91	no	poor	the entrance	…
02080212	female	dance	70	no	poor	teacher	…
07080422	female	foreign languages	95	no	poor	interpreter	…
09080134	male	physics	76	yes	rich	teacher	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
06080321	female	philosophy	72	no	rich	village official	…

4.2 Data Cleaning and Conversion

15 of the 1128 graduates in the school did not graduate successfully for various reasons, so the records are cleared from the database,the rest of the data is set without null values.

In order to facilitate data mining, it is necessary to convert data and complete quantitative work(Tao Ying, 2007) of data. Concretely as: gender quantification, 1 for girls, 0 for boys; professional quantitative, popular professional value of 1, ordinary professional value of 0; results are quantified by 1 above the average and 0 below the average; award information is based on participation in various professional competitions to receive

awards statistics, winning the value of 1, otherwise 0; family economic conditions mean that the parents have the conditions and agree to pay the student's high internship value of1, otherwise 0; whether employment is based on employment professional and professional counterparts, the corresponding value of 1, no employment or incorrect value of 0.The quantized data is shown in Table 2.

Table 2. Summary of comprehensive employment information of students after quantification

No.	Gender	Major	Score	Win the prize	Family	Obtain employment	…
01080104	0	0	1	1	1	1	…
05080305	1	1	1	1	0	1	…
02080212	1	0	0	0	0	1	…
07080422	1	1	1	0	0	1	…
09080134	0	1	0	1	1	0	…
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
06080321	1	0	0	0	1	0	…

4.3 Generation of Association Rule Tables

The process of computing candidate frequent itemsets is to operate the table 2 generated by the data transformation phase. After the same record is compressed, a new data processing table is generated, and a clustering operation is completed, then each record in the new table is computed for support and confidence in the next step. When user enters the minimum support, the program compares it to the support obtained from theresult, if the calculated value is greater than or equal to the minimum support specified by the user, the result is displayed, and the resulting set is the frequent itemsets. After multiple adjustments, with minimum support of 20% and minimum confidence of 75%, a final association rule table is formed and both mining results, as shown in table 3.

Table 3. Association rules table

Serial number	Association rules	Sup	Conf
1	good family =>obtain employment	35	86
2	obtain employment =>good grades	42	82
3	winning ^ good family => obtain employment	21	87
4	obtain employment ^ winning => good family	22	80
5	good major^ good family => obtain employment	24	83
6	good major ^ obtain employment => good family	27	78
7	good grades ^ good family => obtain employment	28	84
8	good grades ^ obtain employment => good family	23	82
9	good grades => good major	43	81
10	good grades ^ good family => good major	21	77
11	good grades ^ winning => good major	24	82
12	good grades ^ obtain employment => good major	22	78

4.4 Impact of Association Results

4.4.1 The Role of Early Warning for Students' Employment

A conclusion can be drawn from the above rules, students want to successfully employment, achievements、 family、 major and award-winning experience are related factors ,not entirely in accordance with the results to determine employment, but combine the various factors that are closely related to students to students' employment choice factors. For the more prominent students, you can consider continuing to study for further study; family conditions on the impact of student employment plays a leading role, if students before graduation, can be funded by qualified families to participate in the professional training, the employment will be both early and early to determine; professionals are more popular, job opportunities will be more; during the school award, indicating that students play a more professional foundation, for some units to apply for additional points.

Students during the school shouldstrive to learn professional, and participate in social practice、competition to improve hands-on ability, conditional permission to participate in the relevant professional training in advance, each term concern about personal achievement in a timely manner, from many aspects to find suitable for their own employment advantage, to ensure that on time、high quality and smooth employment.

4.4.2 Tips for Educators

With the reform of the educational system, the initiative of employment comes from the larger part of the students, while the leading role of the school is second. In the recruitment of students propaganda, we must remind the parents, according to the actual situation, choose their own children's school, professional, and finally achieve the goal of employment after graduation.

For the professionalism of the school, the popular professional compared to the general professional, there is not much advantage, which requires the school's education decision-makers to re-examine the professional settings of the school, set the professional needs of the community, adjust the professional structure, reasonable arrange the course to determine the advantages of the school professional, not stick to the rules, nor blind with the new.

Students' employment can't be separated from the guidance of faculty when they exert themselves. Therefore, from the students enrollment, teachers should take employment as the ultimate goal, according to the relevant factors of the association rules, rational organization, give students the correct guidance, avoid the combination of adverse factors, and in the work of the link is fully aware of the process, each factor may become factors restricting the employment problem, not from single hand to look at the problem, should consider the problem, continue to explore the impact of employment rules, and effectively improve the employment rate of college students.

5. Discussion

This paper is based on the employment information of college students, uses Apriori algorithm to transform、mining (Zhang Guohua,2016), draws relevant rules, makes guidance for the employment of students and tips for the educators to set up majors、courses, so as to improve the students' employment rate , the education reform is more scientific、more in-depth、more in line with the current needs of the community、more targeted to solve the scientific and rational use of talent.

Acknowledgments

References

Ji, S. S., Li, S. F., & Jiang, W. X. (2016). Association analysis of hot topics in mobile classroom WeChat APP based on Apriori algorithm. *Wireless Internet Technology, 11*, 67-69. https://doi.org/10.3969/j.issn.1672-6944.2016.11.029

Judith (2004). Practical data mining. Beijing: Electronic Industry Press.

Li, W., Liu, G. M., Meng, X. F., & Zhang, Z. F. (2016). Application and optimization of parallel Apriori algorithm in massive medical document data mining. *Journal of Beijing Normal University (Natural Science Edition), 52*(4), 420 -424.https://doi.org/10.16360/j.cnki.jbnuns.2016.04.003

Liu, Y. H., Chen, J. G., & Zhang, C. Y. (2015). Two-way recommendation system of domestic college students' employment information based on data mining. *Journal of Shenyang University (Natural Science Edition), 27*(3), 226-232. https://doi.org/10.3969/j.issn.2095-5456.2015.03.011

Lu, X. H., & Liu, J. (2016). Application of Apriori algorithm in automatic recommendation of learning resources on network teaching platform. *Modern Industrial Economy and Information Technology, 123*, 101-104. https://doi.org/10.16525/j.cnki.14-1362/n.2016.15.43

Tao, Y. (2007). Research on the application of improved association rules algorithm to the employment of students. (Master's thesis, Liaoning University of Engineering and Technology, Liaoning,China).

Wang, X. (2006). Research on the training and employment guidance of college students based on data mining. (Master's thesis, Southwest Jiaotong University, Sichuan, China). https://doi.org/10.7666/d.y883730

Zhang, G. H. (2016). Application and practice of data mining in early warning ofindependent college curriculum. *Modern Electronic Technology, 39*(17), 136-139.https://doi.org/10.16652/j.issn.1004-373x.2016.17.034

Zhang, L. (2016). Research on university employment early warning system based on Apriori algorithm. *Microcomputer Applications, 32*(7), 44-46. https://doi.org/10.3969/j.issn.1007-757X.2016.07.013

Zhou, X. S. (2014). Application of data mining based on Apriori algorithm in employment guidance of college graduates. *Journal of Hunan City University (Natural Science Edition), 23*(3), 76-78. https://doi.org/10.3969/j.issn.1672-7304.2014.03.019

Copyrights