

# Application of Internet of Things Data Analysis in Innovation and Entrepreneurship

Feng wang

Hubei College of Chinese Medicine, 434020, China  
16121461@qq.com

## ABSTRACT

In order to deeply analyze the key factors and their potential interaction that affect the Job-hunting of college students in Yingjia, the author makes extensive statistics on the decision-making factors of the graduates of Nanjing Institute of information technology in recent years, such as recommending employment, choosing a job independently, starting a business independently or entering a higher school, waiting for a job and so on, and designs C4 with the support of data mining classification technology. 5 algorithm systematically analyzes various potential factors that affect the employment choice of graduates, so as to obtain the decision-making model that affects the employment rate of fresh graduates. The main significance of this study is that students can improve their knowledge structure and enhance their comprehensive quality and social competitiveness. In order to adapt to the increasingly severe employment situation and the needs of economic and social development, improve the success rate of employment contract. Fresh graduates can also refer to the algorithm model to compare their own quality, targeted selection of target units for career. Colleges and universities can make a vertical and horizontal comparative analysis of the employment index of graduates from all walks of life according to the algorithm model, and improve the teaching plan to make the graduates more in line with the needs of contemporary society and stimulate the rapid growth of employment rate

**Keywords:** Data mining; classification; decision tree; C4.5 algorithm; college students' job search

## 1 INTRODUCTION

In September 2014, Premier Li Keqiang first put forward the call of "mass entrepreneurship and innovation" at the summer Davos forum, and then set off a new wave of innovation and entrepreneurship across the country. The state and local governments at all levels have given strong support in terms of policies and funds. Numerous makers join in entrepreneurship, but only a few of them succeed in the end[1]. There are many reasons for the failure of entrepreneurship, such as creativity, decision-making, capital, technology, management ability and marketing strategy. Among many factors, correct decision-making is undoubtedly the key to the success of entrepreneurship, and the choice of decision-making technology has an important impact on making reasonable decisions[2]. According to the degree of certainty of decision conditions, decision-making can be divided into deterministic decision-making and non deterministic decision-making. The decision-making in the process of starting a business is faced with the changeable market environment, which belongs to the uncertain decision-making[3]. According to the risk preference of decision-makers, we can choose the method of selecting

the small from the large, the large from the large, the minimum of the maximum regret value and so on. If we can make an appropriate probability estimation of the future market environment, the decision tree rule is a better choice. The related concepts of personalized recommendation system are shown in Figure 1 below.

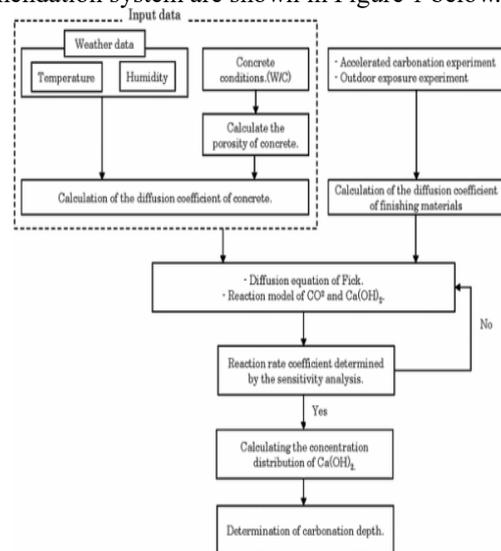


Fig. 1. personalized recommendation system

2 RELATED WORK

2.1 The concept of data mining

Data mining comes from knowledge discovery in database. Data mining is a hot research field at present. It is an information means to find valuable, unknown and potential data from massive real data, noisy data, variable data and mixed data. It involves machine learning, statistics, database, knowledge discovery and so on. It is considered to be one of the most effective means to solve the current situation of "data rich" and "knowledge poor" in enterprise information system. These important information can predict the development trend and provide important reference value in many fields such as information management, decision support, process control and so on[4].

(1)

(2)

(3)

2.2 Data mining process

(1) Data selection and collection, the main task of this stage is to select the relevant data from the database or data warehouse, and establish a target data set. The storage format of the target data set needs to be processed, so that the data mining algorithm can directly access[5].

(2) The main reason of data preprocessing in this stage is the large amount of data, which makes the data incomplete, inconsistent and complex. By cleaning up the data, the data set can be integrated into a data set to store data. The transformation of data format or data planning can improve the effectiveness of the algorithm. Data filling, data redundancy, data clustering and other methods can talk about the analysis time. The purpose is to improve the efficiency and accuracy of data mining algorithm and reduce the time spent on data processing.

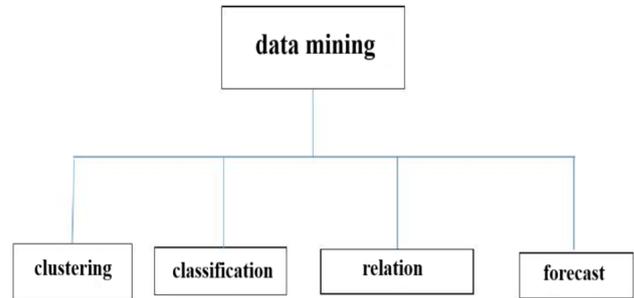


Fig. 2. Classification of data mining

(3) In this stage, we need to consider the type, method and efficiency of data mining. To decide which data mining algorithm to use, create a model describing a group of predetermined data classes or concepts, select the most appropriate data mining technology to operate through the analysis and classification of data sets, determine the technology to be used, and select the appropriate software to mine the data that has been processed. This process is at the core of data mining, It is also the key and difficult point in the process of knowledge method. The sample test data and sample training data in the construction model are used to evaluate the efficiency of the model.

(4) Results evaluation, this stage is mainly to explain and evaluate the classification results, this step is mainly used to evaluate the results of the established model and explain the value of the model. The result is to explain the patterns from the evaluation data mining and remove the inappropriate patterns and rules. In the application, due to the constant change of data and attributes, the final accuracy of the model will also change, and even show a large deviation, so the accuracy of the evaluation of the model results is not necessarily the best choice[6]. Clustering method, which is to analyze the follow-up data, using statistical classification method, makes the data in a class have high relevance, and the data between different classes have low relevance. There are five clustering methods: partition based, hierarchical, density based, grid based and model-based. Partition clustering method: there is a data set, which is composed of multiple objects. How many classes can these objects be divided into? Each class can contain one or more objects, but each object must belong to only one class. Hierarchical clustering method: for a given data set, the aggregation method is used to merge the objects of each data set one by one according to the relevance of attributes, or the splitting method is used to split the objects of the data set into a minimum class according to the minimum relevance. Density clustering method: if the density or correlation between various types exceeds the set value, the classes will be merged into one class,

and then continue to cluster. The network clustering method divides the data set into finite class units, and the clustering is carried out in the network structure formed by these small classes. Model clustering method: set a model for each class, and look at the data to see the correlation degree of the model.

### 3 DATA ANALYSIS

#### 3.1 The basic concept of association rule algorithm

In the field of data mining, association rules is an important research direction, which is proposed by Agrawal et al. The most important ability of association rules is to discover the potential association between attributes in the dataset, and then use these associations to create association rules[7]. The more famous case of association rules here is the supermarket's "drink beer and pee". These potential associations can't be found in the database, so association rules can help people deal with the market, make regulation, decision support and so on. The flow of association rule algorithm is shown in Figure 3 below.

The teaching content of innovation and entrepreneurship courses in some colleges and universities focuses on innovation, while the teaching content of innovation and entrepreneurship courses in some colleges and universities focuses on entrepreneurship. Those who focus on innovation may talk more about how to develop innovation potential; Those who focus on entrepreneurship may talk more about how to organize a team, how to carry out marketing and how to prepare a business plan. For students majoring in business administration, some colleges and universities make innovation management a compulsory course; For most students of other majors, innovation and entrepreneurship courses are generally set as elective courses, and many of these courses are linked to entrepreneurship competitions and other related competitions[8]. It seems that understanding and writing business plans have become the core content of elective courses.

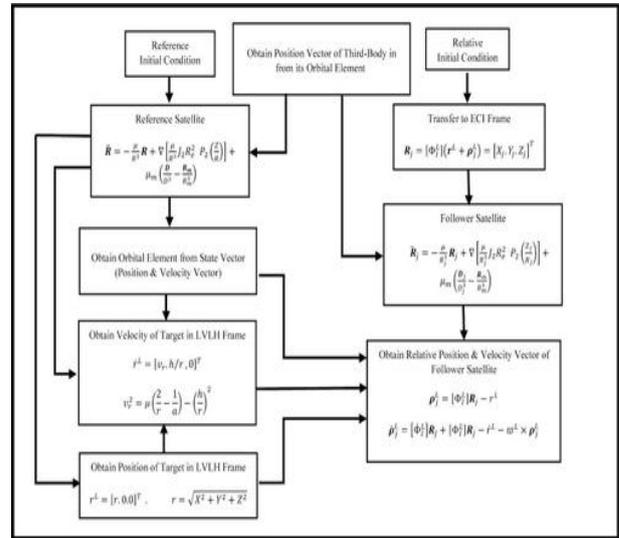


Fig. 3. Association rule algorithm flow

Association rule mining is mainly divided into two problems: a find out all itemsets in transaction database d that are greater than or equal to the minimum support specified by the user[9]. The itemset with the minimum support is called the maximum itemset, or large itemset for short. The support of an itemset is the number of items that contain the itemset. b. Use the largest project set to generate the required association rules. Find all non empty subsets a of a for each maximum itemset. If the ratio sup port (a) / support (a) ≥ minconfidence, generate the association rule A-A). Support () / support (a), that is, the confidence of rule a (A-A).

$$(4)$$

$$(5)$$

In fact, because step 2 is relatively simple, the key to mining association rules is to find frequent sets. The central idea of Apriori algorithm is to scan the transaction database to find all items with support not less than the minimum support, that is, frequent i-item set. The next work is a cycle. Each cycle is divided into three steps: a connection, which connects the items in the frequent k-item set. The first condition is that the first k-1 items must be the same. b. Pruning, in this step of pruning, we mainly filter the connected itemsets according to the idea that any subset of a frequent itemset should be complex, delete those itemsets whose subset is not a frequent set, and obtain the candidate (K + 1) itemset. c. Scan the database, calculate the support of

candidates, delete candidates with support less than the minimum support from the candidate set, and then obtain the frequent (K + 1) item set. The termination condition of the loop is that the frequent K itemset is empty, that is, the associated item can no longer be found.

### 3.2 Evaluation of Apriori algorithm

APIO  $\pi$  I algorithm is based on frequent itemsets and adopts the iterative method of hierarchical search. The algorithm is simple, clear, easy to implement, without complex theoretical induction. However, there are some insurmountable shortcomings. Too many scans of the database[10]. In the description of apron algorithm, it is known that whenever a candidate set is created, a comprehensive search of the database will be performed. To create an empty set of entries with a maximum length of N, the database must be scanned n times. If a large amount of transaction data is stored in the database, the system memory load will increase significantly due to the limited memory capacity, and the efficiency of retrieving the database each time is very low. The evaluation of Apriori algorithm is shown in Figure below.4

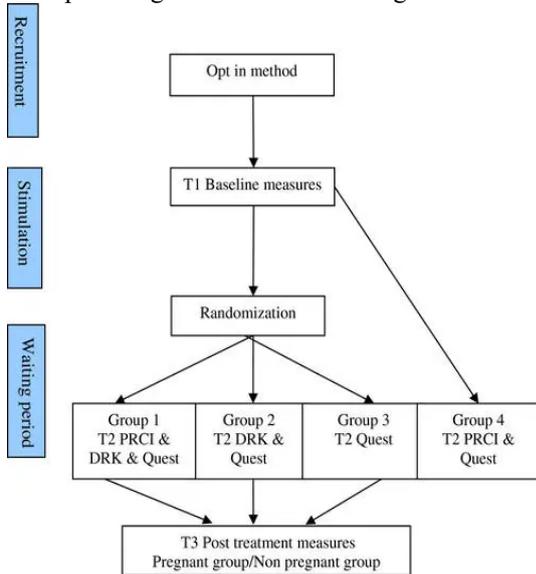


Fig. 4. Evaluation of Apriori algorithm.

## 4 EXAMPLE ANALYSIS

### 4.1 Software implementation and testing

Software definition: Nowadays, there are many employment data in Colleges and universities, the utilization rate is low, there is no value of deep mining data, and the potential rules of data can not be found[11]. The

decision tree algorithm combined with data mining technology is used to realize the rule discovery of data, which is used to guide the employment of college students. The decision tree algorithm can quickly find the rules between the data. Requirement analysis: the software needs to sort out the test data and transform it into test data mode. ID3 based decision tree generation algorithm initializes the root node of the decision tree, presses into the stack while (the stack is not empty) to take out the top element of the stack, and pops up the entropy of each attribute corresponding to the element, and takes the attribute with the minimum entropy value for expansion[12]. If the attribute is the last attribute, extend all its attribute values to leaf nodes, otherwise traverse all possible values of the attribute; if the entropy corresponding to this value is 0, extend it to leaf nodes and store it in the decision tree; otherwise, extend it to non leaf nodes and store it in the decision tree and press it into the stack. The software implementation and testing are shown in Figure 5 below.

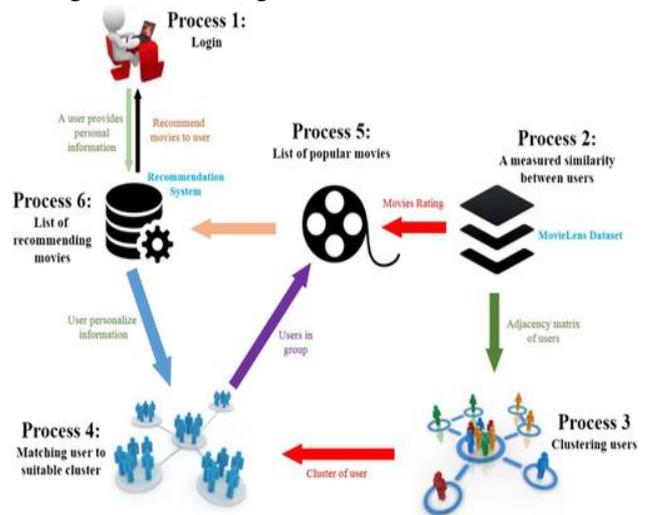


Fig. 5. Software implementation and testing

### 4.2 Complex multistage decision tree

In order to reduce the influence of future uncertainty and improve the accuracy of decision-making, more information about decision-making can be obtained through certain ways, such as market investigation and information consultation, which can effectively improve the accuracy of future prediction. Based on the previous example, in order to improve the accuracy of decision-making, Xiao Li considered purchasing relevant market forecast information from an information consulting company Y, with a consulting fee of 3000 yuan. Company Y reminds Xiao Li that it has 60% possibility to

predict the future market as "good" and 40% possibility to predict the future market as "bad"[13].

For the scheme branch of "don't buy market forecast information", the drawing of this part is correct, and then analyze the scheme branch of "buy market forecast information". Finally, if we choose to buy market information, then the subsequent decision-making must be based on the result of market forecast, that is, if the market forecast is "good", which store form to choose; if the market forecast is "bad", which store form to choose[14].

Innovation and entrepreneurship education should first cultivate people's awareness, thinking and spirit of innovation and entrepreneurship, then cultivate people's skills in innovation and entrepreneurship, and finally make the educated have a certain ability of innovation and entrepreneurship. Through innovation and entrepreneurship education, students' employment concept can be changed, so that students can take entrepreneurship as a choice of future career, so as to transport a large number of excellent innovation and entrepreneurship builders for China's economic development. Of course, qualified students can start a business immediately after graduation, but more students may start a business after working for several years and having certain social experience and industry knowledge accumulation, which should be the norm.

## 5 CONCLUSION

In the era of data and information explosion, how to use big data, develop big data and explore big data has become the focus of today's research, and data mining technology is a hot research issue in this era of big data. Data mining technology has made great progress in supermarket shopping analysis, online shopping user analysis, sales forecast of fast-selling industry, market analysis, financial analysis, medicine and chemical industry. The purpose of data mining technology is to discover the hidden information and associated information in massive data, as well as the useful knowledge behind the data resources. The main research methods of data mining technology include genetic algorithm, artificial neural network, rough set, fuzzy set, decision tree, statistical analysis and so on. With the rapid development of data mining technology and the demand of massive data analysis, people's enthusiasm for its research will not decline. With the increasing enrollment of Chinese universities and colleges, the number of students going to school has increased sharply, and the employment problem of students has become increasingly prominent. In the era of big data, colleges and universities should make full use of this, use data min-

ing technology to analyze the employment data of students, find the relationship between the data and new knowledge between the data, which can not only help colleges and universities to improve the employment rate of graduates, alleviate the employment pressure of students, but also improve certain data support for colleges and universities in professional construction, It is very helpful for students' education and employment. For students, they can adjust their career planning and employment direction in time through the employment analysis of the school and the evaluation of the employment market, so as to speed up the connection with the employment market. Through the research on the application of data mining technology in students' employment, the direction of future efforts is roughly clear. There are many problems in the research process of this paper, which need to be further studied and improved in the future. In the analysis of the decision tree rules generated by the data test, there are some inconsistencies with the actual situation. The reason for the output error is that the data acquisition is not accurate enough. When the data is preprocessed, the important attributes are deleted or the attributes that have a greater impact on the rules are retained.

## REFERENCES

- [1] Jin Ying. Research on the application of decision tree algorithm in college students' employment [D]. Hefei University of technology, 2009
- [2] Li Xue. Performance analysis system based on decision tree algorithm [d]. Changchun University of technology, 2010
- [3] Zhang Guangrong. Analysis of student employment data based on decision tree algorithm and association rule analysis method [D]. Shaanxi Normal University, 2014
- [4] Exploration on the construction of smart campus application support platform layer under the background of high-level higher vocational school construction [J]. Yu Yongjia, Su Biao, Zhou Wei. Chinese Journal of multimedia and network teaching (zhongxunjian). 2020 (09)
- [5] On the application of data in the construction of smart campus in Colleges and universities [J]. Zhang Jie, Hu Bo. International public relations. 2020 (08)
- [6] Current situation of smart city construction in China and its impact on Colleges and universities [J]. Wang Xiaoshuang. Business story. 2016 (26)

- [7] On the application of Internet of things technology in smart campus [J]. Ran Deling. Network security technology and application. 2014 (08)
- [8] Research on personalized push system of smart campus [J]. Yan Zhenhao, Wang Hong. Computer products and circulation. 2020 (05)
- [9] Micro school application - Intelligent school management [J]. Hu Xiaomao. New curriculum (primary school). 2018 (07)
- [10] Design and practice of College Multimedia Classroom Based on smart campus multi-mode certification [J]. Dai Linxue. China high tech Zone. 2018 (09)
- [11] Design and implementation of smart campus education platform [J]. Ma Hongru. Electronic technology and software engineering. 2016 (01)
- [12] Design of smart campus system based on LabVIEW [J]. Tang Hui, Jiang Bin, Li Jian, Xu Weichuan. Science and technology vision. 2019 (28)
- [13] Technical architecture design and application of smart campus in Higher Vocational Colleges [J]. Jiang Zhiyong. Scientific and technological innovation and application. 2016 (35)
- [14] Xu Ling. Research on the application of classical decision tree algorithm in the association between College Students' learning and employment [D]. Inner Mongolia Agricultural University, 2014