
Association Rule Mining for the Talents Introduction Strategy: A Case Study of Zhejiang University of Finance & Economics

Wang Qin, Zhang Kangkang, Chen Huiting

School of Economics, Zhejiang University of Finance & Economics, Hangzhou, China

Email address:

18010129619@163.com (Wang Qin), 914968197@qq.com (Zhang Kangkang), m15757905243@126.com (Chen Huiting)

To cite this article:

Wang Qin, Zhang Kangkang, Chen Huiting. Association Rule Mining for the Talents Introduction Strategy: A Case Study of Zhejiang University of Finance & Economics. *American Journal of Applied Mathematics*. Vol. 6, No. 2, 2018, pp. 55-61.
doi: 10.11648/j.ajam.20180602.15

Received: March 6, 2018; **Accepted:** April 17, 2018; **Published:** April 27, 2018

Abstract: In recent years, the issues of the talents introduction have attracted more and more researchers' and college administrators' attention. In the era of big data, data mining technology is widely used in various fields and has achieved remarkable results. The application of data mining technology in the introduction of university talents is in the ascendant. This paper uses the effective information of 245 teachers recruited by Zhejiang University of Finance & Economics since 2011 to explore and model the association rules. It preprocesses the raw information data by hierarchical clustering, and use Apriori algorithm to obtain a set of rules for the paper score and the situation of receiving the National Foundation of China (NFC) in 3 years. These rules will provide a constructive guiding significance for the introduction of talents in Zhejiang University of Finance & Economics.

Keywords: Association Rule Mining, Talents Introduction, Apriori Algorithm

1. Introduction

The basic national condition that China is the largest developing country in the world determines that China must take the strategy of "Reinvigorating China through Human Resource Development". Through implementing this strategy, it will strive to create more high-quality socialist builders and transform China from a large population resource country to a large human resource country. Only in this way, can China enhance its core competitiveness and comprehensive national strength, complete the historical task of building a moderately prosperous society in all respects and realize the great rejuvenation of the Chinese nation.

Higher education plays a very significant role in the development and progress of a country. Since the Reform and Opening-up, Chinese higher education has been developing rapidly. The vigorous development of education has made more universities recognize the key role of talents in the competition of universities, and the introduction of talents has become a strategic measure for the development of colleges and universities. However, the introduction of

college talents is a very complicated project. In this process, colleges and universities is often faced with all kinds of risks and various losses. This survey finds that the risks mainly focus on recruitment risks, matching risks, talent loss risks and so on. The first step to avoid these risks is to make more accurate selection in the process of recruiting talents. However, in view of a lot of resume information, it is extremely difficult to be able to select real talents with more precision. The reason is that the rapid development of information technology has brought tremendous conveniences to anthropic life, however, the development has also brought about explosive data growth in all fields. As the number of customer data collected and stored daily by the enterprise increases, LeeFlang et al. [1] pointed out that the so-called "data explosion" has been identified as one of the major challenges for marketers in both online and offline channels. Similarly, colleges and universities have established a perfect set of personnel information management system, which has accumulated a large amount of data. However, the system lacks trend analysis which is more important. What needs to be done in the big data era is

to dig up useful knowledge from huge and messy databases, and make it better for anthropic production and life. Based on the talent database about Zhejiang University of Finance & Economics, this article tries to find useful knowledge to explore the regularity of university human resources development and management using association rule analysis in R Language. It will not only provide more accurate basis for the subsequent talents introduction to find the real talents from a large number of candidates, but also provide support for the establishment of university human resources management policy.

The remainder of this article is organized as follows. In section 2, some related work is reviewed. What introduced in section 3 is the data preprocessing and association rule mining. The results and analysis of association rule mining are showed in section 4. And the conclusion and future works are showed in the last section.

2. Related Work

As is known to all, data mining technology is a relatively advanced research direction in the world. It has a leading position in databases, data warehouses and information decision fields. Among the well-known IT technologies, data mining technology has a great impact on human life and its application in many fields has made remarkable achievements. The ultimate goal of data mining application is to find valuable information, which mainly includes association analysis, classification and prediction, clustering analysis and so on. Altaf et al. [2] argued that association rule mining is an effective data mining technique which has been used widely in health informatics research since it was introduced. The concept of association rule mining was first proposed by Agrawal et al. [3]. Its general form is $A \Rightarrow B$ (s, c), s and c are two important indicators of rule measurement, the former is called support and the latter is the confidence. The Apriori algorithm is one of the most popular algorithms for association rule mining, which has been applied in many areas. This study refers to relevant studies as researching this issue below and its review here is relatively brief. Through critical analysis of applications of association rule mining literature about health informatics, Altaf et al. [2] explored that the Apriori algorithm is still a widely used frequent itemset generation technique for application of association rule mining for health informatics. Chalmers et al. [4] used Apriori algorithm to find the association rules between preseason scores and injuries of football players. Tripathi et al. [5] introduced a new architecture for web fraud detection using Apriori algorithm for association rule mining and phishing database in web advertising network. To improve the service ability of the library, Zhang [6] used Apriori algorithm in library personalized service field. Lakshmi and Vadivu [7] suggested a novel approach to pick up association rules from medical records by choosing the best association rule mining algorithm using multiple-criteria decision analysis.

New technologies should serve for education optimization.

In recent years, data mining technology has been applied to university research, but the basic problems involved are varied. Zhang et al. [8] employed rule mining methods to excavate the relationship among all kinds of curriculums, and aimed to realize the goal of reasonable curriculum arrangement and improve the quality of education. Romero et al. [9] used a variety of data mining methods to improve the prediction of students' final performance. Asif et al. [10] used data mining technology to study the performance of undergraduate students in order to enhance the management level of the school and the quality of the educational processes. Zhu [11] used different data mining approaches to build models to predict the engineering students' performance. Zhang et al. [12] used various data mining methods to dig up the information of educational management system, in particular, the Gradient Boost Decision Tree model performs outstanding in predicting the grades of graduation thesis from university course grades. Sael et al. [13] proposed a new approach to help in learning evaluation by applying clustering and association rule mining techniques to analyze learners' behaviors.

Among them, data mining technology is also applied to human resources management in universities. Liu et al. [14] analyzed the data mining technology and processes applying on college human resources management, which is the closest to this research. Based on an actual dataset of college human resources, they used association rule analysis to show the relationship among teaching, research and social practices. Sael et al. [13] adopted the feature selection process and multilevel clustering to confirm the importance of these new data preprocessing methods. This literature you are reading benefited from the above approach since it attempt to find possible relationships between the different parts of the relevant content.

Have to admit it, however, this research cannot perfectly solve all problems, and the deficiencies will be showed in the section of conclusion. The justification for contributing to this literature is that it will provide a constructive guiding significance for the introduction of talents in Zhejiang University of Finance & Economics.

3. Data Preprocessing and Association Rule Mining

This section is aim to explore the attributes that universities should consider when introducing talents. The original data used in this case contains information of 245 teachers who were introduced by Zhejiang University of Finance & Economics from 2011 to 2017, including their numbers of papers published and their situation of getting the National Foundation of China (NFC) in 3 years.

3.1. Data Preprocessing and Discretization

In order to protect the personal information of teachers, it is essential to conduct a desensitization treatment. Then the schools were transformed into C9, 985 (excluding C9), 211

(excluding C9 and 985) and ordinary university to improve the accuracy of mining association rules. According to the index system of academic scholarship selection of Zhejiang University of Finance & Economics, it is suitable to make a score of 8: 2 for the first author or corresponding author (note the first author and correspondent are treated as such), and the second author, and the ratio of scientific research ability and comprehensive ability to national, provincial and university grade is 5: 3: 2 to calculate the score; and the total scores of the published papers of SCI, SSCI, EI, IC, 1A, 1B, 2A, 2B are used to measure the writing ability of the introduced talents. Then in order to make the data conform to the data type of association rule mining in R Language, the hierarchical clustering method was used to discretize the continuity values. The clustering results of some important variables are shown in Table 1.

Table 1. The clustering results of some important variables.

Attributes	Discretization results
SCI score	0; (0, 2.4]; (2.4, +∞)
SSCI score	0; (0, 1]; (1, +∞)
EI score	0; (0, 2.4]; (2.4, +∞)
IC score	0; (0, 0.8]; (0.8, 2.4]; (2.4, +∞)
1A score	0; 0.2; [0.8, +∞)
1B score	0; (0, 2.8]; (2.8, +∞)
2A score	0; (0, 1.4]; (1.4, 2.6]; (2.6, +∞)
2B score	0; (0, 1]; (1, 4]; (4, +∞)
Score	0; (0, 52.8); (52.8, +∞)
Scientific Research Award Score	0; (0, 0.6]; (0.6, +∞)
Comprehensive Award Score	0; (0, 0.7]; (0.7, 1.6]; (1.6, +∞)
Age	[28, 32]; [33, 40]; [41, 44]; [49, 53]

After data preprocessing and discretization, it is easy to obtain the data containing 245 teachers' information and 29 attributes, including the situation of getting the NFC in 3 years and their paper scores.

3.2. Association Rule Mining

In this part, this paper uses the association rule analysis to further excavate the clustering results. By using the Apriori breadth-first algorithm based on the two-phase frequency-set theory, this part uses the frequent itemsets' nonmonotonicity to search the entire data source from the bottom to the top and iteratively retrieves all frequent itemsets of the transaction database, and then uses the frequent itemsets to construct the association rules that meet the preset conditions. The association rule has the form of $A \Rightarrow B$, where $A \in I$, $B \in I$, $A \cap B = \emptyset$, and I is the set of items. The support and confidence are two key parameters for association rule mining. The support of the association rule $A \Rightarrow B$ is the probability that both A and B occur in all observations. Thus, the support of the association rule $A \Rightarrow B$ is equal to the support of the association rule $B \Rightarrow A$. The confidence of association rule $A \Rightarrow B$ is the probability that both A and B occur in the observations that contain the item A . The mathematic equations of support and confidence in association rule mining are showed in equation (1) and equation (2), respectively [15]. In this article, item A is a collection of various teachers' attributes, in particular the

scores published for various types of papers, item B for the NFC of 3 years, and the ability to publish papers.

$$\text{Support}(A \Rightarrow B) = P(A \cap B) \tag{1}$$

$$\text{Confidence}(A \Rightarrow B) = P(B | A) \tag{2}$$

Based on the introduction of support and confidence above, it is noticeable that there is no point in finding rules of support or confidence that are too small because of the high probability of misjudgment. After a series of tests, this paper sets the support and confidence to 0.04 and 0.5 respectively when the consequence is Receive the NFC in 3 years, and to 0.02 and 0.3 when the consequence is Score. In addition, lift is another indicator of association rules that is calculated in Equation (3) [15]. According to equation (3), when the lift of association rule $A \Rightarrow B$ is equal to 1, there is no correlation between A and B . Therefore, the subset that the degree of lift is greater than one from the rule is suitable. Finally, after considering these three indicators, it is preciseness to select the appropriate rules.

$$\text{Lift}(A \Rightarrow B) = P(B | A) / P(B). \tag{3}$$

Then, there is a map of item frequency (relative) to understand the basic situation of teachers of Zhejiang University of Finance & Economics. From Figure 1 which shows the top 10 frequent items, you can see that there are nearly half of female teachers, and over 70 percent of teachers don't have the experience of overseas study. These phenomena have so strong similarities with other universities in China that the database used in this article is universal.

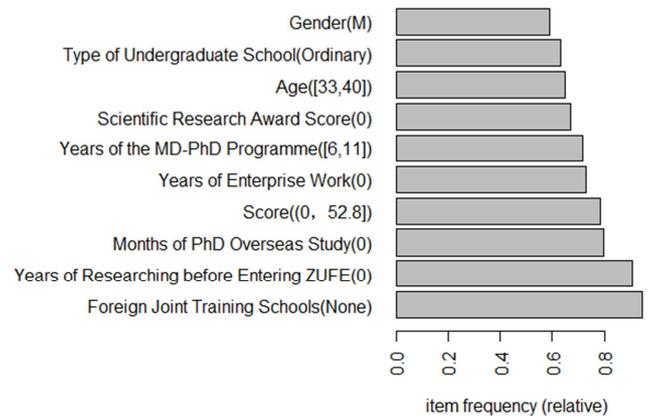


Figure 1. The top 10 most frequent 1-itemsets.

4. Result and Analysis

In the process of considering the introduction of talents, it is emphatically considered what qualities can prompt them to get the NFC more likely in 3 years, and what qualities can make the paper score higher. This study respectively, therefore, uses Receive the NFC in 3 years and Score as consequents in R Language, draw two scatterplots in Figure 2 and Figure 3. The X-axis shows the support and the Y-axis shows the confidence. Each point represents an association rule and different colors represent different degrees of lift. Generally speaking, when the

lift is high, the support is low. Because unique qualities often influence things to a greater extent, while ordinary qualities do not have such charm. Specifically, as shown in Figure 2, when lifts of the association rules are bigger than 1.6, their supports are around 0.1, probably because only some talents have distinctive qualities that are vital for access to the NFC in 3 years. There are also some points with high degree of lift between 1 and 1.5 have high support, probably because the qualities that most talents have cannot largely determine whether these talents will be able

to obtain the NFC in 3 years. According to the distribution of the rules in Figure 2, for Receive the NFC in 3 years, this paper sets the following support as 0.04 and the confidence as 0.5. As shown in Figure 3, when lifts of the association rules are bigger than 4.5, their supports are around 0.05. The obvious difference indicates that only rare qualities can determine whether more or less outstanding papers can be published or not. According to the distribution of the rules in Figure 3, for Score, the following support will be set as 0.02 and the confidence will be set as 0.3.

Scatter plot for rules

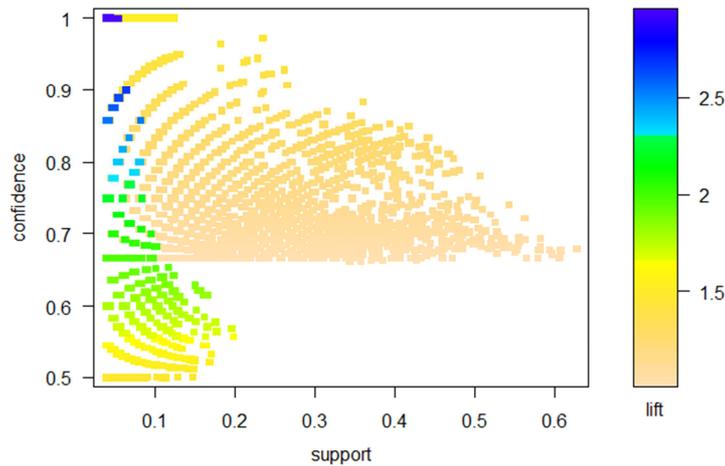


Figure 2. Rules of talents being awarded the NFC in 3 years.

Scatter plot for rules

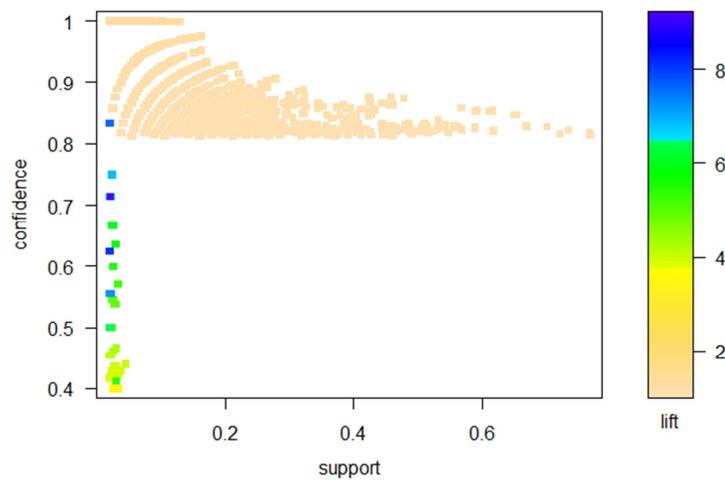


Figure 3. Rules of Score.

Figure 4 and Figure 5 are graphs of association rules about Receive the NFC in 3 years and Score. Each of these circles represents a rule. The color of the circle represents the lift of the rule, and the deeper it is, the higher the lift. The size of the circle indicates the support, and the larger the circle, the higher the support. Because there are tens of thousands of rules for each consequence, it is inevitable to choose the first few rules that with high support. As shown in Figure 4, the rules of receiving the NFC in 3 years have less support and higher lift than not receiving the NFC in 3 years. This phenomenon may

be due to the difficulty of receiving the NFC, which only a few people can achieve. In terms of access to the NFC, those with comprehensive award score below 0.7 and a faculty title of associate professor have more opportunities than those with comprehensive award score above 0.7, possibly because associate professors have more pressure on their subjects, and they are more motivated to study the subjects. As shown in Figure 5, the majority of candidates applying for jobs scored more than 0, less than 52.8. There are few people whose Score is 0 or more than 52.8. This indicates that only a few have

excellent writing skills or weak writing skills. Paper score of talents whose educational background is engineering is higher than those whose educational background is economics and management, this may be because talents who study

engineering have higher modeling skills or talents who study economics or management have more difficulties in publishing papers in direction of economics and management in recent years.

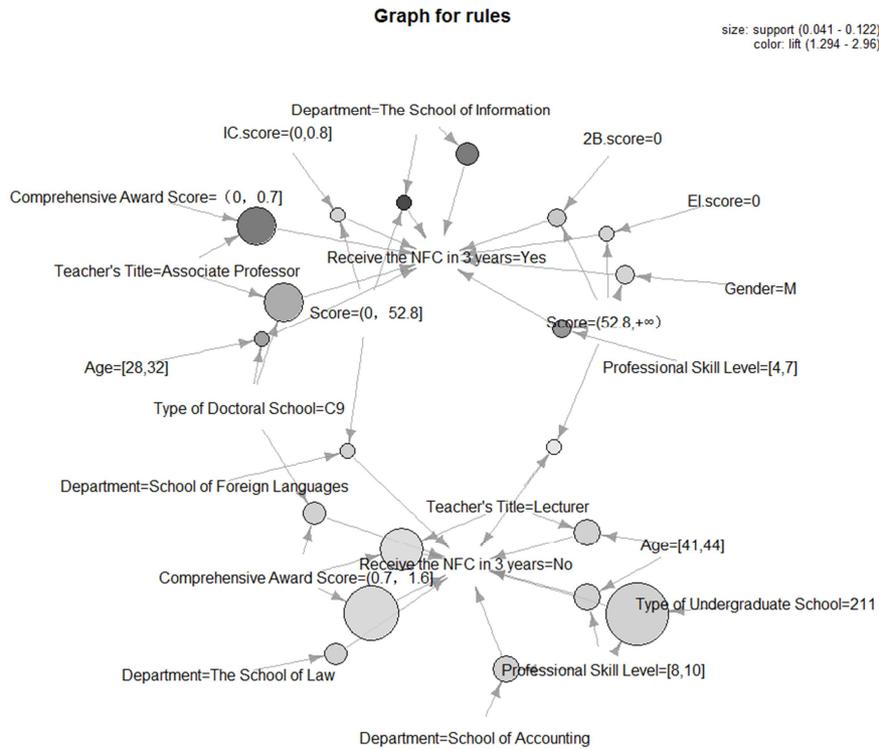


Figure 4. Graph for rules of talents being awarded the NFC in 3 years.

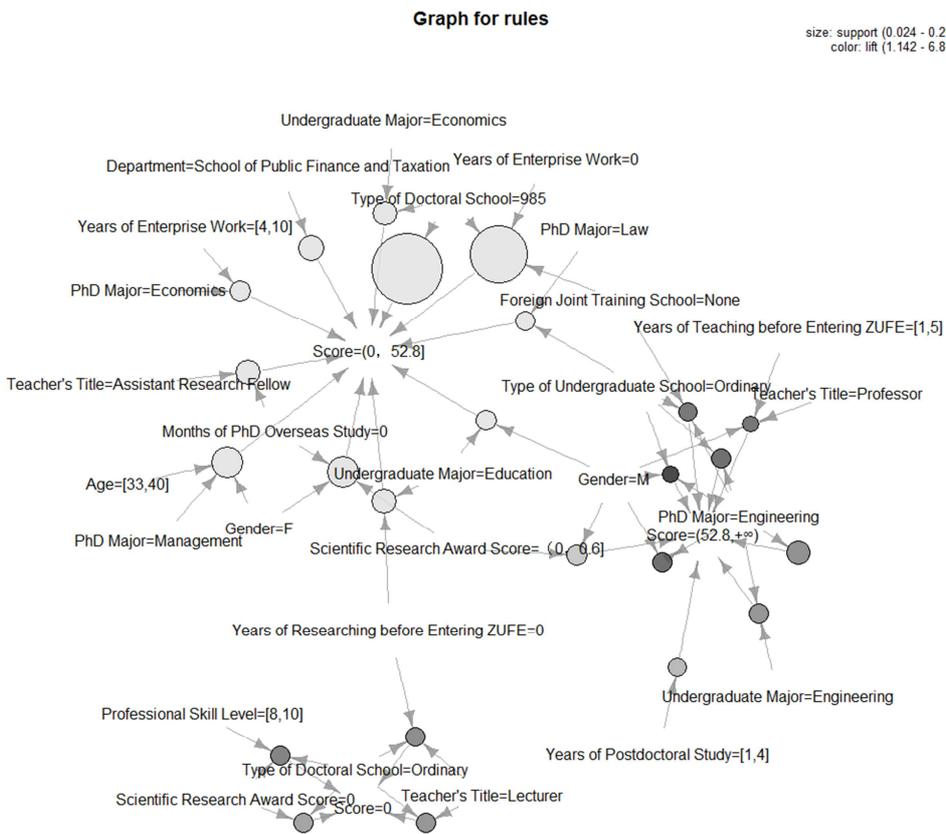


Figure 5. Graph for rules of Score.

The rules of the characteristics of the talents quality and the probability of obtaining the NFC or paper score are shown in Table 2 and 3. For example, two rules in Table 2 point out that among those whose paper score are 0-52.8, those who work in the school of information are more likely to obtain the NFC in 3 years than those who work in the school of foreign languages. This rule suggested that research on information technology is more likely to receive the NFC than those on language. Another rule states that people with professional

skills level below 8 but who have published a large number of papers are more likely to receive the NFC. It can be explained by the fact that for the research, high paper writing ability is more important than professional skills level. The few rules in Table 3 show that for high writing ability, as long as a talented person has a Ph.D. education background in engineering, the type of undergraduate school and the experience of foreign joint training are not important.

Table 2. The partial results of the association rules of talents being awarded the NFC in 3 years.

Rules	Support	Confidence	Lift
{Score=(0, 52.8], Department=The School of Information} => {Receive the NFC in 3 years=Yes}	0.0405	1	2.96
{Score=(52.8, +∞), Professional Skill Level=[4, 7]} => {Receive the NFC in 3 years=Yes}	0.0473	0.7	2.072
{Type of Doctoral School=C9, Age=[28, 32]} => {Receive the NFC in 3 years=Yes}	0.0405	0.6667	1.9733
{Age=[33, 40], Professional Skill Level=[4, 7]} => {Receive the NFC in 3 years=Yes}	0.1149	0.6296	1.8637
{Type of Doctoral School=C9, Professional Skill Level=[4, 7]} => {Receive the NFC in 3 years=Yes}	0.0946	0.6087	1.8017
{Score=(0, 52.8], Department=School of Foreign Languages} => {Receive the NFC in 3 years=No}	0.0405	1	1.5102
{Age=[41, 44], Teacher's Title=Lecturer} => {Receive the NFC in 3 years=No}	0.0608	1	1.5102
{Type of Undergraduate School=211, Professional Skill Level=[8, 10]} => {Receive the NFC in 3 years=No}	0.1216	1	1.5102
{IC.score=(0, 0.8], Score=(0, 52.8]} => {Receive the NFC in 3 years=Yes}	0.0405	0.5	1.48

Table 3. The partial results of the association rules of Score.

Rules	Support	Confidence	Lift
{Type of Undergraduate School=Ordinary, Ph.D. Major=Engineering, Gender=M} => {Score=(52.8, +∞)}	0.0245	0.75	6.8056
{Years of Teaching before Entering ZUFE=[1, 5], Gender=M, Teacher's Title=Professor} => {Score=(52.8, +∞)}	0.0245	0.5455	4.9495
{Type of Undergraduate School=Ordinary, Ph.D. Major=Engineering, Foreign Joint Training School=None} => {Score=(52.8, +∞)}	0.0286	0.5385	4.886
{Type of Doctoral School=Ordinary, Scientific Research Award Score=0, Professional Skill Level=[8, 10]} => {Score=0}	0.0327	0.3478	4.4851
{Type of Doctoral School=Ordinary, Years of Researching before Entering ZUFE=0, Teacher's Title=Lecturer} => {Score=0}	0.0327	0.32	4.1263
{Undergraduate Major=Engineering, PhD Major=Engineering} => {Score=(52.8, +∞)}	0.0327	0.4211	3.8207
{Ph.D. Major=Engineering} => {Score=(52.8, +∞)}	0.0449	0.44	3.9926
{Months of Ph.D. Overseas Study=0, Scientific Research Award Score=(0, 0.6], Gender=F} => {Score=(0, 52.8]}	0.0735	1	1.2312
{Type of Doctoral School=985, Foreign Joint Training School=None, Years of Enterprise Work=0} => {Score=(0, 52.8]}	0.1633	0.976	1.2011

5. Conclusion

The aims of this article of association rule analysis are to study what kind of quality helps talents obtain the NFC in 3 years, and what kind of characteristics will affect their level of published papers, so as to provide suggestions for future talents introduction and recruit more excellent talents. It uses the effective information of 245 teachers recruited from Zhejiang University of Finance & Economics since 2011 to explore and model the association rules. After preprocessing the raw information data by clustering and using Apriori algorithm, what comes out is a set of rules for the paper score and the situation of receiving the NFC in 3 years. Finally, according to Table 2 and Table 3, in part, talents who work in the school of information, obtain high teacher's title, have published some paper in the journal are more likely to receive the NFC in 3 years. In addition, a graduate has a Ph.D. education background is more capable of writing papers.

Based on association rules, this paper analyzes the talents introduction of the Zhejiang University of Finance & Economics, and gets some rules that have a great reference to the introduction of talents. There is no doubt that these rules

will play an important role in the future introduction of talents. Moreover, in this article, what is recognized is that data mining technology plays an extremely important role in the introduction of talents. If the data mining technology was applied to the introduction of talents and even more areas, it will greatly promote anthropic production and life.

6. Further Work

The previous section shows the results of data mining and you can see a lot of interesting conclusions. However, there are still some deficiencies in this paper. For example, this paper uses hierarchical clustering method to perform data preprocessing. The certain subjectivity may affect the clustering result, and the weight of different levels of different journals lacks a more unified standard, which may affect the final results. In addition, since there are tens of thousands of rules for the mining, only some typical rules were analyzed. It would be interesting, of course, to continue to study these issues. After mining the association rules, it is worthwhile to envisage the clustering analysis of these rules, and then obtain more scientific and efficient conclusions.

References

- [1] P. S. H. Leeftang, P. C. Verhoef, P. Dahlström, et al., "Challenges and solutions for marketing in a digital era," *Eur. Manag. J.*, 2014, vol. 32, pp. 1-12.
- [2] W. Altaf, M. Shahbaz and A. Guergachi, "Applications of association rule mining in health informatics: a survey," *Artif. Intell. Rev.*, 2017, vol. 47, pp. 313-340.
- [3] R. Agrawal, T. Imielński, and A. Swami, "Mining association rules between sets of items in large databases," In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, Washington, D. C., ACM SIGMOD Record, May, 1993, pp. 207-216.
- [4] S. Chalmers, J. T. Fuller, T. A. DeBenedictis, et al., "Asymmetry during preseason functional movement screen testing is associated with injury during a junior Australian football season," *J. Sci. Med. Sport.*, 2017, vol. 20, pp. 653-657.
- [5] D. Tripathi, B. Nigam, and D. R. Edla, "A novel web fraud detection technique using association rule mining," *Proc. Comput. Sci.*, 2017, vol. 115, pp. 274-281.
- [6] C. Y. Zhang, "Research on library personalized service based on apriori algorithm," *Agro. Food. Ind. Hi. Tec.*, 2017, vol. 28, pp. 2555-2559.
- [7] K. S. Lakshmi and G. Vadivu, "Extracting association rules from medical health records using multi-criteria decision analysis," *Procedia. Comput. Sci.*, 2017, vol. 115, pp. 290-295.
- [8] W. Zhang, S. Xu, and S. Zhang, "Association rule mining for reasonable curriculum arrangement: a case study of Zhejiang University of Finance and Economics," *Int. J. Inf. Proc. Manag., RoMEO*, 2015, vol. 6, pp. 42-47.
- [9] C. Romero, M. I. López, J. M. Luna, et al., "Predicting students' final performance from participation in on-line discussion forums," *Comput. Educ.*, 2013, vol. 68, pp. 458-472.
- [10] R. Asif, A. Merceron, S. A. Ali, et al., "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, 2017, vol. 113, pp. 177-194.
- [11] K. Zhu, "Research based on data mining of an early warning technology for predicting engineering students' performance," *World. Trans. Eng. Technol. Educ., (WIETE)*, 2014, vol. 12, pp. 572-575.
- [12] W. Y. Zhang, H. L. He and S. Zhang, "Predicting the grades of students' graduation thesis using optimized gradient boost decision tree," *Int. J. Inf. Proc. Manag.*, 2017, vol. 8, pp. 9-16.
- [13] N. Sael, A. Marzak and H. Behja, "Multilevel clustering and association rule mining for learners' profiles analysis," *Int. J. Comput. Sci. Issues. (IJCSI)*, 2013, vol. 10, pp. 188-194.
- [14] P. Liu, L. Sun, J. Zhao, et al., "Study and application of data mining technologies on human resources management in colleges," (in Chinese), *Comput. Eng. Appl.*, 2008, vol. 44, pp. 201-204.
- [15] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," In *Proceedings of the 20th International Conference on Very Large Data Bases*, Santiago, Chile, September, 1994, pp. 487-499.