The Construction of AI-Enabled Mechanism for Preventing Medical Graduates' Employment Psychology Crisis Based on Association Rule Mining Algorithm

Lei Yue[1*] and Maoquan Su[2]

[1] School of Management, Weifang Medical University, Weifang 261053, China
18361263390@163.com
[2]Department of Dean's office, Weifang People's Hospital, Weifang 261044, China
wfyxysmq@163.com

Corresponding author: Lei Yue, 18361263390@163.com

Abstract. This paper proposes an association rule mining algorithm (MRFCM algorithm) for the Mechanism of medical graduates' employment psychological crisis prevention system. The algorithm determines the importance of items by setting different weights on items in the database and uses MapReduce data. The grouping idea divides the big data, uses the fuzzy C-means algorithm to cluster the quantified attributes, uses the compressed matrix model to dig out meaningful association rules, and demonstrates the whole process of the algorithm through examples. First, for the study of quantitative association rules, the concept and nature of quantitative association rules are introduced, as well as the k-means clustering algorithm. Aiming at the shortcomings of the Apdod algorithm, the algorithm is improved, and a method based on k-means' quantitative association rule algorithm introduces the algorithm's specific implementation process and analyzes the algorithm's effectiveness simultaneously. Afterward, given the shortcomings of the traditional Apriori algorithm, combined with the existing improvement strategies, this paper applies the new integrated algorithm MC-Apriori algorithm, which proposes corresponding improvements in reducing the number of database scans and reducing the number of candidate sets. On the one hand, the data records in the database are mapped and compressed, thereby reducing the number of scans of the database; on the other hand, in the process of generating k (k>2) item candidate item sets, the generated candidate item sets are performed multiple times Verification to reduce the number of invalid candidate item sets. The simulation experiment results show that the college employment management system requires analysis, overall design, module design, database design, and the MRFCM algorithm design and implementation of the college employment management system. The system analyzes employment data, unearths the correlation between student and employment information, and

provides two-way recommendations for students and employers. In addition, the system can effectively handle many complex tasks, provide accurate and real-time employment information, and ensure the quality of college employment management.

Keywords: Association rule mining; medical graduates; employment psychology; crisis prevention; AI-Enabled Mechanism.
DOI: https://doi.org/10.14733/cadaps.2024.S24.212-225

1 INTRODUCTION

With the continuous expansion of the enrollment scale of higher education institutions, the number of medical graduates is increasing daily, and the forms of employment are becoming increasingly severe. Facing the surge in employment data, some colleges and universities have established employment information management systems. Still, they can only perform simple query and modification operations and cannot effectively analyze employment data. It would be inappropriate to continue to use the past employment information management model [12]. Under the continuous development of computer technology, data mining technology has emerged to meet the needs of scientific management methods. It can effectively analyze employment data, mine valuable information, and provide data support for decision-makers [1],[13],[18]. At the same time, distributed technology has been rapidly developed and widely used. It can distribute employment data on a computer cluster for parallel processing and improve data processing efficiency. Therefore, it is of practical significance to combine data mining technology with distributed technology to discuss distributed association rule mining and its application in university employment data analysis [17].

With the rapid development of database technology, the employment data accumulated by colleges and universities has become larger and larger, which has resulted in a series of consequences, including increasing the workload and complexity of the management work of various colleges and universities [16]. More importantly, in the face of such a surge in data, people hope to conduct an in-depth analysis of it and unearth the hidden critical information. The traditional backward manual management model is increasingly complex in meeting the needs of modern data management [10],[3],[14]. Therefore, the college employment management system cannot be limited to simple query, modification, and deletion operations on massive data rows but should have specific auxiliary decision-making capabilities. With the rapid development of information network technology, database technology, and the wide application of database management systems, university employment information management has also become increasingly digital. How to use these information resources to improve university employment information management has become university employment for a research direction of data. At present, the employment information management system used by colleges and universities can carry out simple business operations. It cannot conduct multi-angle and multi-level analyses and mining employment data; it needs comprehensive analysis and decision-making support capabilities.

Given the large number of quantitative attributes and classification attributes in university employment information data, this paper proposes a k-means-based quantitative association rule mining algorithm KMSQAR, which first uses clustering algorithm k-means reasonably partitions the quantitative attributes, transforms the quantitative attributes into Boolean type, and then uses the improved Boolean association rule method to mine the association rules. The improved Boolean association rule mining method uses binary representation for the data items to be mined and realizes the functions of item set connection and pruning with bit operations. We select data items for data mining to form candidate transaction items. The Apriori algorithm generates frequent item sets according to the input minimum support and confidence. Then, we create association rules from

frequent item sets and analyze and organize the generated rules. It is described in natural language and used to guide youth entrepreneurship.

2 RELATED WORK

Given the above realities and problems, this article quantifies the twelve ability indicators of students. It no longer uses generally non-standard adjectives such as "excellent" and "good" to describe a student's ability. It can be more convenient and intuitive to understand a student's ability. The pros and cons of the abilities, combined with the relevant behavioral information of the students, use data mining technology to analyze and predict the students' graduation whereabouts so that students have a clearer understanding and mastery of their abilities and characteristics and have a choice of where to graduate [7].

In recent years, more and more people have been interested in the direction of educational data mining. There are many studies in this area internationally. Yan [20] uses decision trees to establish a predictive model to predict unknown academic performance based on the score data that students have achieved; Li [11] proposes the relationship between students' daily behaviors and learning outcomes according to students' performance before and during class. The changes in mental state after class and the evaluation of the course's feelings are used to predict students' scores in the subject using PLSA and LDA models. According to specific course data, Akinbade [2] uses an improved naive Bayes algorithm. The dropout rate of students in education schools is predicted; Ma [15] clustered the interaction behavior between teachers and students on an interactive learning forum using the k-means clustering algorithm, which can find the difficulties in learning. At the same time, it can also provide teachers with suggestions on managing student groups.

Aly [5] proposed a counter algorithm, which not only breaks the defect that the original Apdori algorithm is not suitable for finding multi-dimensional association rules but also requires only one database scan to find multi-dimensional association rules, which improves the processing of massive data. After using the Apfiofi algorithm, they set a weighted value between the classes to obtain a weighted confidence to filter out strong association rules. Regarding the problem of re-association after partitioning, the researcher proposed a method of partitioning quantitative attributes and finding out the best partition of the quantitative characteristics in an automated manner, and then using the ARCS algorithm to mine the quantitative association rules of the relational database [9],[8],[4]. Some scholars have proposed an improved method for measuring the distance between rules. First, we calculate the distance between items. The distance between item sets and the distance between rules, and finally, we use the DBSCAN algorithm to cluster association rules based on this distance. We analyzed the shortcomings of distance clustering and proposed an ISODATA algorithm suitable for association rule clustering based on the fuzzy ISODATA algorithm [19], [6].

3 THE CONSTRUCTION OF THE MECHANISM OF THE MEDICAL GRADUATES' EMPLOYMENT PSYCHOLOGICAL CRISIS PREVENTION SYSTEM BASED ON THE ASSOCIATION RULE MINING ALGORITHM

3.1 Association Rule Mining Algorithm Architecture

Association rule data mining can be applied to all transient data and information repositories such as relational databases, advanced database systems, general files, and the World Wide Web. The difficulty of mining varies from storage system to storage system. Currently, the data sources of data mining are mainly the Web, data warehouses, transaction databases, and relational databases. Figure 1 is the architecture of the association rule mining algorithm.

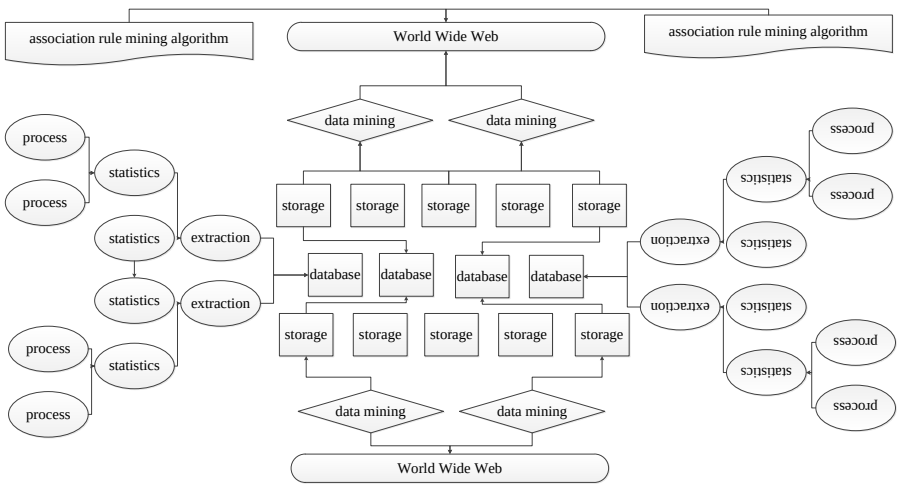

Figure 1: Association rule mining algorithm architecture.

Market forecasting is a classic case. Data mining uses historical market transaction records to unearth the hidden customers who have the greatest return on investment in the future. Data mining can also automatically predict the groups most likely to respond to specific events and predict bankruptcy problems.

$$w_i = w_i - \eta(t - y)x_i \tag{1}$$
$$b = b - \eta(t - y) \tag{2}$$

Data processing, including transformation and sorting, and storing the data information records in the data warehouse's data warehouse. The ability to access different types of data in terms of storage methods is a crucial process of data extraction tools, which should be able to generate other forms of scripts. Data conversion mainly includes calculating statistics and derived data, which assigns default values to missing data.

$$idf(w_i) = ln\frac{|D|}{|\{D_i : D_j \in w_j\} - 1|} \tag{3}$$

Metadata is a general term for the organization and establishment of data in a data warehouse. According to the standard of metadata usage, it can be divided into commercial and technical metadata. Business metadata is used to describe business business logic, including descriptions of business topics.

$$tf(w_i, D_i) = N_{w_1, D_j} \sum_{n=1}^{K} N_{w_n, D_j} N \tag{4}$$

Data marts are also called subject data or departmental data. They are a part of data independent of the data warehouse for specific application requirements. In the data warehouse implementation, the data mart can start from a particular table alone, associate with other tables to form several data table associations, and then use these data table associations to create a complete data warehouse.

$$\psi(x) = \frac{f(x,d) - f(x,m)}{f(x,d) + f(m,d)} \tag{5}$$

Data is usually regarded as people's source of knowledge. It can be structured data, such as data in relational databases, semi-structured data, such as text data, or even heterogeneous data on the Internet. There are many ways to discover knowledge: deductive or inductive, mathematical or non-

mathematical. The knowledge found is used in various fields, including process control, query optimization, information management, decision support, and data maintenance.

3.2 Employment Level Integration of Medical Graduates

Data mining for medical graduates at the employment level integrates database technology, artificial intelligence, and other fields from a new perspective. A deep mining model exists in the data and has potential commercial value. Data mining is divided into three parts: training data, test data, and application data. The key to data mining is to mine actual information from the original data, use test data as a theoretical basis for Verification and improvement, and apply knowledge to the data. Taxonomy is to classify data according to pre-defined criteria. It is divided into the rule induction method, neural network method, and decision tree induction method. Rule induction is to organize data through a series of if-then rules. The neural network method trains the neural network to identify different classes and then classify the data according to the neural network. The decision tree induction method is based on the data and organizes the data hierarchically into a tree structure. Figure 2 shows the degree of integration of graduates' employment levels.

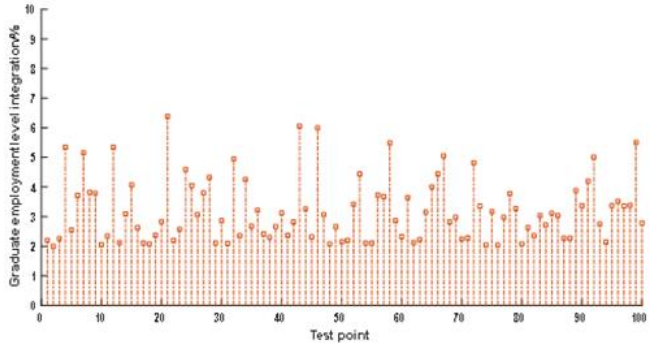

Figure 2: Graduate employment level integration.

The algorithm minimizes the sum of squares of the distances in the division interval. The degree of membership is used to determine the degree to which each data belongs to each cluster partition, and each group is obtained by dividing n data into k categories (1<k). The cluster center makes the value function of the dissimilarity index reach the minimum. The algorithm uses the sum of square errors (SSE) as the objective function to find the fuzzy partition matrix and the number of k cluster centers. We scan the transaction database D and use the binary representation for all transactions Ti to form a single item set binary table B1. The number of "1"s in the bit string of the statistical item (that is, the number of times the item appears in different transactions) is recorded as a count; if the count is less than the given minimum support count, delete the item; use sum to count the items contained in the transaction, if sum<2 (property 3.1), we delete the transaction, and get a frequent 1-item set L1. We perform the AND operation on the bit string of the item in L1 pairwise, calculate the count for the new binary table; if the count is less than the minimum support count, delete the item set; calculate the sum; if sum<3, delete the sum value. The mining of association rules is an essential issue in data mining research. It focuses on determining the relationship between different attributes in the data and which attributes meet the predetermined support and confidence threshold. The dependence relationship between the research is the relationship between Boolean attributes. It has attracted attention in many industries and has achieved good results in different fields. Mining association rules between quantitative and categorical attributes involves more than Boolean association rules.

3.3 Applicability Analysis of Employment Psychological Crisis

The execution process of the applicability analysis of employment psychology crisis is to divide the input big data into several data blocks, and each data block is assigned to a Map function for processing. The Map function takes a series of intermediate key/value pairs as the output result. The output result of the Mprint function is used as the function's input. The function merges the value parts of all intermediate key-value pairs with the same key value. The MapReduce model allows programmers to deal with the resources of extensive distributed systems without experience in distributed systems, so programmers only need to focus on how to write maps and reduce functions for distributed file systems, inter-machine communication, etc. Complicated issues in parallel computing are handed over to the background processing of the MapReduce operating system. Correlation refers to specific rules in the value of two or more variables. In general, there are three types of associations: simple associations, causal associations, and time series associations. Correlation analysis aims to find the correlation network hidden in the database. Usually, the correlation function of data is fuzzy and uncertain, so the rules generated by correlation analysis should be accompanied by confidence. Figure 3 is the applicability framework of the employment psychology crisis.

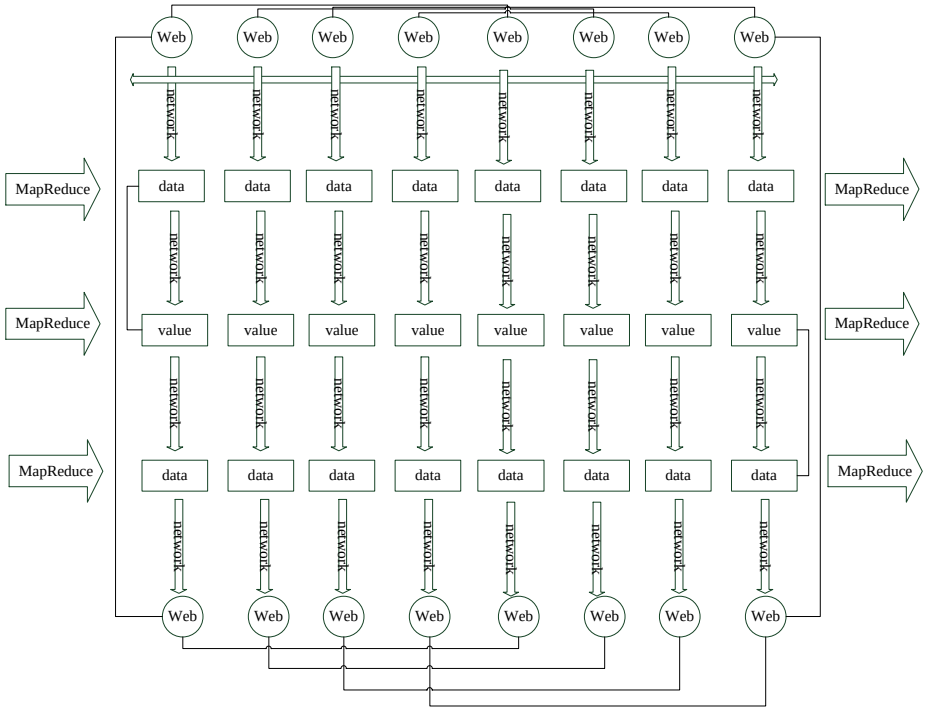

Figure 3: Applicability framework of employment psychology crisis.

According to the information storage format, the objects used for mining include transaction databases, relational databases, data warehouses, advanced database systems, data streams, general files, and the World Wide Web. Digging problems and techniques may vary from storage system to storage system. The clustering algorithm classifies data with similar characteristics into one category by comparing variables. Therefore, after clustering, the data set is transformed into a cluster. In the cluster, the data in the same category have similar variable values, and the variable

values in different categories are not identical. The main task of logistic regression is to use the inherent logistic function to estimate the possibility of things, that is, to predict and evaluate the relationship between several independent and dependent variables. Logistic regression models are mainly used in medicine, for example, in predicting and discriminating the incidence of patients, searching for potential risk factors of a disease, and so on. The basic steps of the logistic regression model are mainly divided into three: one is first to obtain the H function based on sigmoid, that is, the hypothesis function; the other is to construct the J function, that is, the loss function, by the least square method; finally, we use the gradient descent method to find the minimum of the cost function and find the relevant parameters. The process of data mining includes distinguishing different types. These classes are not defined in advance but are obtained fully automatedly through a clustering algorithm.

3.4 Iteration of the Weight of the Prevention System Mechanism

Analyzing the basic business process of the employment management system in colleges and universities, the system involves three types of users: administrators, medical graduates, and employers. First, the administrator sets up user permissions, adds users, and imports the database after logging in. The central databases include the employment information of medical graduates and basic information of medical graduates; users of medical graduates can browse personal details and modify personal contact information after logging in. We add professional skills, browse recruitment positions, obtain employment consultation, and perform employment recommendation operations; after logging in, employer users can browse unit information, modify unit contact information, post job positions, invite interviews, and screen and recommend medical graduates. At the same time, the administrator can review the login logs and the recruitment positions issued by the employer to ensure the system's security. Frequent item sets are generated using a pruning strategy to compare the minimum support threshold for candidate item sets. The itemsets are traversed by the iterative method of layer-by-layer search, and new candidates are generated according to the frequent itemsets of the traversal results. Figure 4 shows the fitted distribution of the weights of the prevention system mechanism.

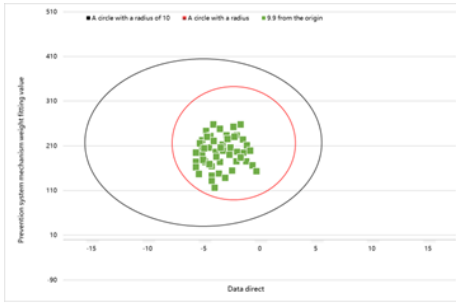

Figure 4: Fitting distribution of the weight of the prevention system mechanism.

Aiming at the shortcomings of the Apfiori algorithm, this paper will adopt an Apriori algorithm based on the two-system system. The algorithm uses binary representations of the data items to be mined and bit operations (AND, XOR, etc.) to implement itemset connection, pruning, and other steps. In this way, the database only needs to be scanned once, and no candidate item sets need to be generated. First, we select k initial centers from the sample data set N, define the similarity according to the distance calculation formula of the space or attributes, and classify the remaining part of the data sample into one of the initial centers through the similarity comparison, and the one with the highest similarity. The data samples are grouped into the same cluster class, and the above steps

are repeated until the standard function converges. At the same time, the binary system is used to store tuples and item sets, which can solve the problem of execution time caused by the partitioning of quantified attributes and categorical attributes. Personal quality evaluation belongs to the categories summarized by YBC experts and has nearly ten years of practical experience. It has a one-vote veto effect on whether to obtain YBC funding.

Moreover, according to the requirements of the YBC Code, the personal quality evaluation should reach 80 points or more to be qualified, so the guiding significance of this item for entrepreneurial success is still apparent. The following will integrate the personal quality items to delineate their essential order. A new cluster center is generated by calculating the arithmetic average of all samples in this cluster. At the same time, the convergence function is calculated according to the above formula. The binary representation of the data item is as follows: first, we determine the length of the binary bit string according to the number of transactions in D; then, we sort the transactions; each transaction only corresponds to one bit in the binary, for the data item in D, if it appears in transaction Ti. Then the corresponding bit of Ti in the binary bit string takes "1"; on the contrary, it takes "0".

4 APPLICATION AND ANALYSIS OF THE MECHANISM OF MEDICAL GRADUATES' EMPLOYMENT PSYCHOLOGICAL CRISIS PREVENTION SYSTEM BASED ON ASSOCIATION RULE MINING ALGORITHM

4.1 Association Rule Mining Data Preprocessing

The research purpose of this experiment is to calculate and generate twelve ability indicators based on the general standards of national engineering education certification and use the k-means clustering algorithm to divide more than 400 computer science and technology students into reasonable clusters and analyze the students. The overall graduation trend of students in each cluster is compared, the difference in the ability indicators of each cluster of students is reached, and the relationship between students in different clusters and their graduation destinations is studied. The data items A1 and A2 are the results of the classification attributes of the academic qualifications in the employment database according to the actual situation; that is, the classification attributes of the educational qualifications are divided into undergraduates and graduates. Ordinary and excellent students result from FCM clustering on English scores, professional scores, and quantitative characteristics of comprehensive assessment. First, it uses the FCM algorithm to cluster the data, set the fuzzification factor to 2, and the classification number to 2. The iteration termination condition is the minimum change in membership degree of 0.001, and the membership degree obtained is as follows. Figure 5 shows the association rule mining data clustering.

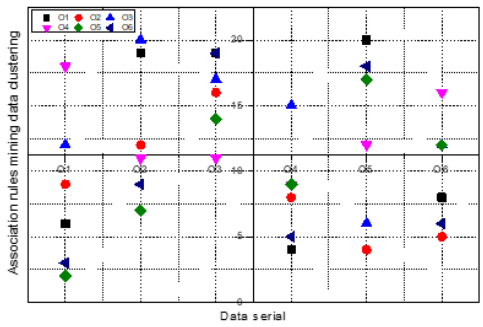

Figure 5: Association rule mining data clustering.

In this experiment, we use the Silhouette Coefficient method to select the optimal number of clusters for the clustering algorithm. The k value that can make the contour coefficient larger is the optimal number of clusters (k value) to be found. The coefficient method combines the concepts of cohesion and separation. The degree of aggregation and the degree of separation are not independent. The sum of the two is a constant. For the parameters of the k-means clustering algorithm, selecting the appropriate number of clusters (k value) is the most crucial step in the experiment. The groups can be reasonably divided only by determining the optimal number of clusters. The similarity within the same category of clustering results is high to the limit, the difference between different categories is noticeable, and the group division is apparent to analyze the objects of other clusters better. This rule states that 30% of graduate students have computer skills and are excellent students in English, professional, and comprehensive assessments. This shows employers are more interested in graduate students who are outstanding in English performance, professional performance, and thorough evaluation. Therefore, schools should strengthen not only the training of English and professional skills but also the comprehensive training of students' abilities.

4.2 Mechanism Simulation of Employment Psychological Crisis Prevention System

In this experiment, each student has a unique student number, and each subject has a unique course number. The student's basic information, four-year undergraduate academic performance, award experience, and other attributes can be integrated according to the student number. The academic performance and ability items are matched. In the integration process, redundant information can be deleted, or incomplete data may be encountered, such as lack of student source information, missing academic performance, etc., to ensure the quality of the data and the accuracy of the results of data analysis and mining, it needs to be eliminated. In addition, removing duplicate or redundant data based on the primary key is necessary. The redundant and duplicate data will affect the results of table association and data mining. Finally, we must consider the problem of data conflicts. Data conflicts are essential to data mining results, and conflicting data needs to be detected and processed. There is yet to be a unified standard for evaluating a student. You can only judge whether a student meets the graduation requirements from the student's academic performance, foreign language level, science and technology awards, scholarships, etc., so these inconsistent data need to be converted into corresponding indicators through the evaluation of the various abilities of students, find out the solid and weak skills, and make more reasonable choices for graduation. Figure 6 is the distribution of employment psychological crisis prevention evaluation.

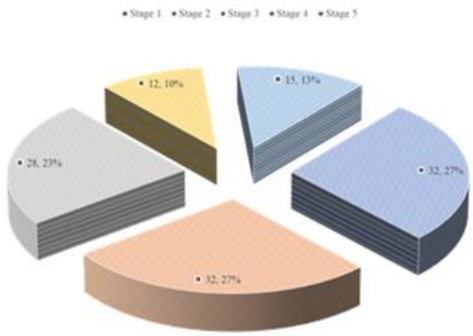

Figure 6: Distribution of employment psychological crisis prevention evaluation.

The administrator user is mainly responsible for the maintenance and safe operation of the system. Its main functions include user rights management, database management, medical graduate information management, employer information management, recruitment information

management, employment guidance and consultation, and review of user login logs. This system mainly has seven tables: user table, log-in log table, basic information table of medical graduates, employment information table of medical graduates, basic information table of employers, recruitment information table, and score information table of medical graduates. The main code of the basic information table of medical graduates is StudentID: the outer code of StudentID in the primary employment information table of medical graduates is the StudentID in the basic information table of medical graduates. The outer code of StudentlD in the table is the StudentID in the basic information table of medical graduates; the main code of the employer basic information table is CompanyID; the outer code of CompanyID in the recruitment information table is the basic information table of the employer CompanyID. Figure 7 shows the distribution of employment psychological crisis prevention profile coefficients.

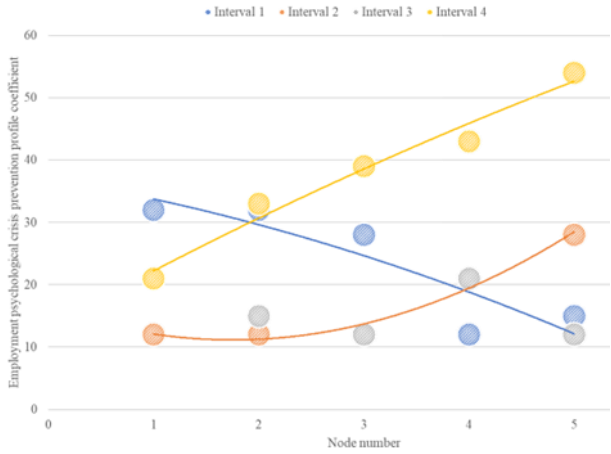

Figure 7: The distribution of the employment psychological crisis prevention profile coefficient.

The twelve ability indicators of more than 400 students majoring in computer science and technology are divided into clusters. The contour coefficient method has determined the optimal number of clusters as 5, and the k-means algorithm separates the students into five categories. It can be seen that the contour coefficient achieves the maximum value of 0.4 when the k value is 7, but in actual experiments, dividing the number of clusters into seven categories will cause the number of specific categories to be too small, so the number of clusters is selected to maximize the contour coefficient 7 is not a reasonable number of clusters. Therefore, the k value 5, which has the second largest contour coefficient, is selected in the experiment. In the experiment, the results obtained by dividing the five groups of groups are relatively large in the amount of data in each category. Hence, the value of k is five, which is more reasonable. Therefore, the optimal number of clusters obtained by the contour coefficient method should be five instead of 7.

4.3 Example Application and Analysis

Converting the four-year university academic performance into ability values is necessary to assess student abilities more conveniently and intuitively. Since one course corresponds to multiple ability items, numerous courses support each ability item, and the number differs. To make the ability value easy to read and intuitively compare and to prevent the occurrence of singular values, it is necessary to return the ability value. The first part is the student's basic information, including name, student ID, gender, primary data of the student's place of origin, science and technology awards, scholarships, and outstanding student titles obtained during the university. The second part is the

student's performance data, mainly the four-year undergraduate results of computer science and technology students (including all the results of regular exams, make-up exams, and retakes). There are about 400 students and 75 courses. The third part is the student's graduation information, which mainly includes graduation destination (including going abroad, entering a higher education, employment, waiting for employment), graduation destination, name of the unit, and nature of the unit (including state-owned enterprises, institutions, party and government agencies, etc.). The fourth part is students' graduation requirements, including the correlation matrix between the college's computer school curriculum system and graduation requirements; that is, each course corresponds to the student's graduation ability support level (divided into high H, medium M, and low L). Figure 8 is the evaluation of the employment level of medical graduates

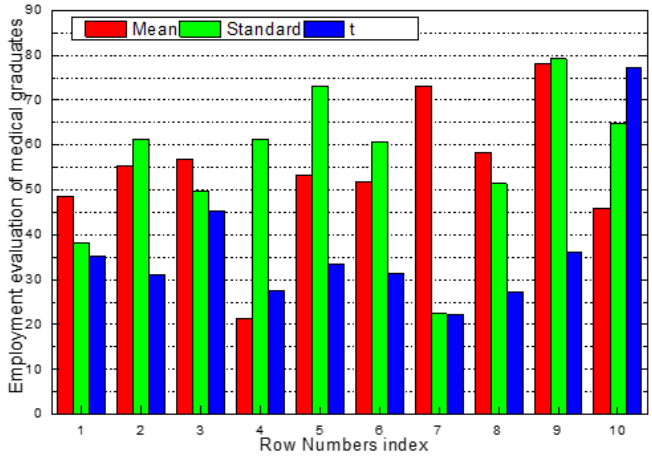

Figure 8: Evaluation of the employment level of medical graduates.

The relevant transaction database must first be transformed into a Boolean matrix to use the matrix-based Apriori algorithm for data mining. After data preprocessing, each student's mental health problem analysis table is a transaction Ti (TID), the transaction set T={T1, T2, T3,..., Ti}. Each psychological dimension factor is an item set. If the nine psychological factors corresponding to each transaction are symptomatic, it corresponds to 1; if there is no symptom, it corresponds to 0. Assuming the minimum support minsup=20%, the minimum support count is sup_count=minsup|T|. Specifically, comparing medical graduates whose birthplace and graduation destination are the same or not, the study found that male students who choose to return to their birthplace are more likely to be students in state-owned enterprises, military units, and party and government agencies than male students who choose to return to their birthplace. Because of the characteristics of computers, many male students who choose non-local will choose Internet companies; for girls, it is found that whether it is girls who decide to return to their birthplace, they still select other places. The proportion of choosing Internet companies is higher than that of boys, and the proportion of choices in education and public institutions is much higher than that of boys. In comparison, the proportion of boys who choose state-owned enterprises and party and government agencies is higher. Figure 9 shows the level distribution of association rule mining capabilities.

Calculating ability value requires students' four-year undergraduate academic performance and grade for each subject's corresponding ability item. According to the correlation matrix between the curriculum system and the university's computer school's graduation requirements, the ability item grades are divided into high H, medium M, and low L.

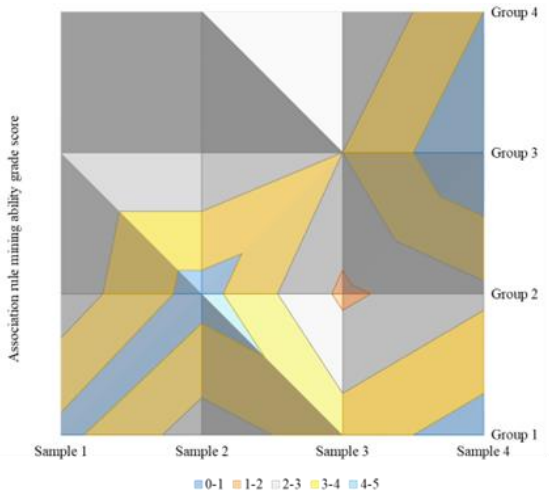

Figure 9: Association rule mining ability level distribution.

The level must be converted into a numerical value and academic performance for calculation. To make the calculation result of the ability value meet the normalization process, first, we divide the ability item level proportionally because the ability item and communication ability item using modern tools is only medium, and the proportion division is set to 0.55, 0.30, and 0.15. Then, the student's academic performance should be scaled and reduced to [0, 1.0], where 10 points or less (not including 10 points) are defined as 0, 10-19 is 0.1, 20-29 is 0.2, 30- 39 is 0.3, 40-49 is 0.4, 50-59 is 0.5, 60-65 is 0.6, 66-75 is 0.7, 76-85 is 0.8, 86-95 is 0.9, 96-100 is mapped to 1.0. The employment recommendation function is also one of the core functions of this system. The recruitment information management module of this system involves this function. During the system's operation, students can input their English proficiency, course scores, performance evaluation, a combination of the two, etc. The system will match the analysis results stored in the database, that is, the information in the rule table with the grades entered by the students, and give employment recommendations with higher matching degrees.

5 CONCLUSIONS

This paper uses the k-means clustering algorithm to divide the students of computer science and technology into reasonable clusters, analyze the overall graduation trend of the students, compare the differences in the ability indicators of each cluster of students, and study the students of different clusters. Secondly, using the Apriori association rule mining algorithm, we can dig out the relationship rules between student ability indicators and graduation destination and the relationship rules between student behavior information and ability indicators and graduation destination and analyze the most vital influencing factors of students' graduation destination. The results are compared and evaluated through model evaluation indicators, and the most suitable predictive algorithm is selected to predict the student's graduation fate. The importance of various characteristic factors is compared and analyzed. The demand analysis, function design, and database design of the college employment information management system are carried out, combined with the KMSQAR algorithm, and the college employment information management system is designed and implemented. The system can mine existing data to find out the relationship between the student's education attributes and employment attributes, and at the same time, match the students' scores with the results of the rules and provide medical graduates with higher matching

employment recommendations. In addition, the system can automatically process and manage a large amount of complex and repetitive data in an orderly manner, provide dynamic, accurate, and complete employment information in real time, and provide decision-making support for relevant departments of colleges and universities. Data mining technology can perform a higher-level analysis of current entrepreneurial data, find potential connections between various entrepreneurial factors and entrepreneurial success, determine the decisiveness of each factor, and, in turn, guide the future development trend of youth entrepreneurship. Implementing this AI-enabled system ethically and effectively would require a delicate balance between leveraging technology for personalized assistance and maintaining ethical standards in handling graduates' data and mental health considerations.

Lei Yue, https://orcid.org/0009-0002-0662-5434
Maoquan Su , https://orcid.org/0009-0002-3669-8968

REFERENCES

[1] Adnan, K.; Akbar, R.: Limitations of Information Extraction Methods and Techniques for Heterogeneous Unstructured Big Data, International Journal of Engineering Business Management, 11, 2019, 1871. https://doi.org/10.1177/1847979019890771
[2] Akinbade, D.; Ogunde, A. O.; Odim, M. O.: et al. An Adaptive Thresholding Algorithm-Based Optical Character Recognition System for Information Extraction in Complex Images, J. Comput. Sci., 16(6), 2020, 784-801. https://doi.org/10.3844/jcssp.2020.784.801
[3] Al-Temeemy, A. A.: Human Region Segmentation and Description Methods for Domiciliary Healthcare Monitoring Using Chromatic Methodology, Journal of Electronic Imaging, 27(2), 2018, 023005. https://doi.org/10.1117/1.JEI.27.2.023005
[4] Aly, S.: An Effective Human Action Recognition System Based on Zernike Moment Features, Innovative Trends in Computer Engineering, 2019, 52-57. https://doi.org/10.1109/ITCE.2019.8646504
[5] Aly, S.; Sayed, A.: Human Action Recognition Using Bag of Global and Local Zernike Moment Features, Multimedia Tools and Applications, 78(17), 2019, 24923-24953. https://doi.org/10.1007/s11042-019-7674-5
[6] Bian, Q.; Zhang, G.: Fast Moving Human Detection using Depth Information, Intelligent Control and Automation, 2018, 1227-1232. https://doi.org/10.1109/WCICA.2018.8630371
[7] Bilal, M.; Maqsood, M.; Yasmin, S. et al. A Transfer Learning-Based Efficient Spatiotemporal Human Action Recognition Framework for Long and Overlapping Action Classes, The Journal of Supercomputing, 2021, 31-36. https://doi.org/10.1007/s11227-021-03957-4
[8] Ding, S.; Qu, S.; Xi, Y.: et al. A Long Video Caption Generation Algorithm for Big Video Data Retrieval, Future Generation Computer Systems, 93, 2019, 583-595. https://doi.org/10.1016/j.future.2018.10.054
[9] Du, S.; Liu, Y.; Ye, M.: et al. Single Image Deraining Via Decorrelating the Rain Streaks and Background Scene in Gradient Domain, Pattern Recognition, 79, 2018, 303-317. https://doi.org/10.1016/j.patcog.2018.02.016
[10] Li, G.; Li, C.: Learning Skeleton Information for Human Action Analysis Using Kinect, Signal Processing: Image Communication, 84, 2020, 115814. https://doi.org/10.1016/j.image.2020.115814
[11] Li, J.; Wang, Q.; Li, S.: et al. Face Replacement and Image Animation System in Cultural Experience, Journal of Physics: Conference Series, 2010(1), 2021, 012144. https://doi.org/10.1088/1742-6596/2010/1/012144

[12] Liu, L.; Jiao, Y.; Meng, F.: Key Algorithm for Human Motion Recognition in Virtual Reality Video Sequences Based on Hidden Markov Model, IEEE Access, 8, 2020, 159705-159717. https://doi.org/10.1109/ACCESS.2020.3020591
[13] Liu, Y.; Lin, Z.; Yang, Z.: et al. Review on Biologic Information Extraction Based on Computer Technology, Big Data Analytics for Cyber-Physical-Systems, 2020, 1235-1241. https://doi.org/10.1007/978-981-33-4572-0_177
[14] Liu, Z.; Xie, Q.; Lu, Y.: et al. Skeleton-based Action Recognition with Two-Branch Graph Convolutional Networks, Journal of Physics: Conference Series, 2030(1), 2021, 012091. https://doi.org/10.1088/1742-6596/2030/1/012091
[15] Ma, Q.; Wang, Y.: Application of Embedded System and Artificial Intelligence Platform in Taekwondo Image Feature Recognition, Journal of Ambient Intelligence and Humanized Computing, 2021, 10-12. https://doi.org/10.1007/s12652-021-03222-9
[16] Qian, H.; Zhou, J.; Lu, X.: et al. Recognizing Human Activities Using Appearance Metric Feature and Kinematics Feature, Journal of Electronic Imaging, 26(3), 2017, 033015. https://doi.org/10.1117/1.JEI.26.3.033015
[17] Qian, H.; Zhou, J.; Mao, Y.: et al. Recognizing Human Actions from Silhouettes Described with Weighted Distance Metric and Kinematics, Multimedia Tools and Applications, 76(21), 2017, 21889-21910. https://doi.org/10.1007/s11042-017-4610-4
[18] Shukur, W. A.; Abdullah, W. N.; Qurban, L. K.: Information Hiding in Digital Video Using DCT, DWT and CvT, Journal of Physics: Conference Series, 1003(1), 2018, 012035. https://doi.org/10.1088/1742-6596/1003/1/012035
[19] Wang, Y.; Yang, Y.; Li, Y.: Recognition and Difference Analysis of Human Walking Gaits Based on Intelligent Processing of Video Images, Traitement du Signal, 37(6), 2020, 32. https://doi.org/10.18280/ts.370621
[20] Yan, X.; Fan, Y.: Foreground Extraction and Motion Recognition Technology for Intelligent Video Surveillance, International Journal of Pattern Recognition and Artificial Intelligence, 34(10), 2020, 2055021. https://doi.org/10.1142/S0218001420550216