Research on Intelligent Management System for Personnel Resources of Medical School Teaching Platform Based on Cloud Computing

Libo Shi[1], Xiaoding Shang[1] and Dan Li[2,*]

[1] Mudanjiang Medical University, Mudanjiang, Heilongjiang, 157011, China
[2] Hongqi Hospital Affiliated to Mudanjiang Medical University, Mudanjiang, Heilongjiang, 157011, China

Corresponding authors: (e-mail: slbgyg6951998@163.com).

Abstract This paper proposes an intelligent management system for personnel resources of medical school teaching platform based on cloud computing, which integrates multi-dimensional personnel data with the architectural features of medical school teaching platform. With the help of Apriori algorithm and improved depth graph clustering model SDCN, personnel association rule mining and personnel resource clustering analysis are realized respectively. Experiments show that the system can effectively recognize four types of personnel groups, and the Apriori algorithm outperforms other algorithms in all performance indexes, with an accuracy rate of 96.28%, a time consuming of 37.65s, a rule coverage rate and a rule reliability of 90.27% and 93.66%, respectively. The results of cluster analysis show that the personnel of associate professors are missing in clusters 1 and 4, and the number of clusters 3 in the personnel of professors is significantly larger than the number of personnel in other positions. The application of the system in this paper helps to reveal the structural problems among positions and provides a technical framework for the intelligent management of medical education resources.

Index Terms cloud computing, medical school, personnel resource management, Apriori algorithm, SDCN model

I. Introduction

With the digital transformation of education, medical education is no exception. The background of digital medical education is the rapid accumulation and development of modern medical knowledge and technology, coupled with the promotion of national informatization construction, digital medical education has gradually become one of the priorities of medical education [1]-[3]. The modern medical field requires the mastery of a large number of basic science and clinical diagnostic skills, and students need to perform a large number of practices and experiments, so the demand for educational resources is huge. The emergence of digital medical education makes up for the shortcomings of the traditional teaching mode, provides richer, more intuitive and more targeted learning resources for medical students, promotes the improvement of students' skills, and makes medical education more efficient and scientific [4], [5]. In this context, medical education presents an imbalance in the teacher-student ratio, which leads to the phenomenon being more significant in the teaching of clinical and other practical operations [6], [7]. Moreover, the practical teaching of medical students is hindered by the underutilization of digital technology in teaching practice, such as the lack of experimental technology in simulated surgical environments [8]. Driven by personalized needs, the current management platform of medical schools, student data are separated from related teaching decisions, and resource management efficiency is low under platform access restrictions [9], [10].
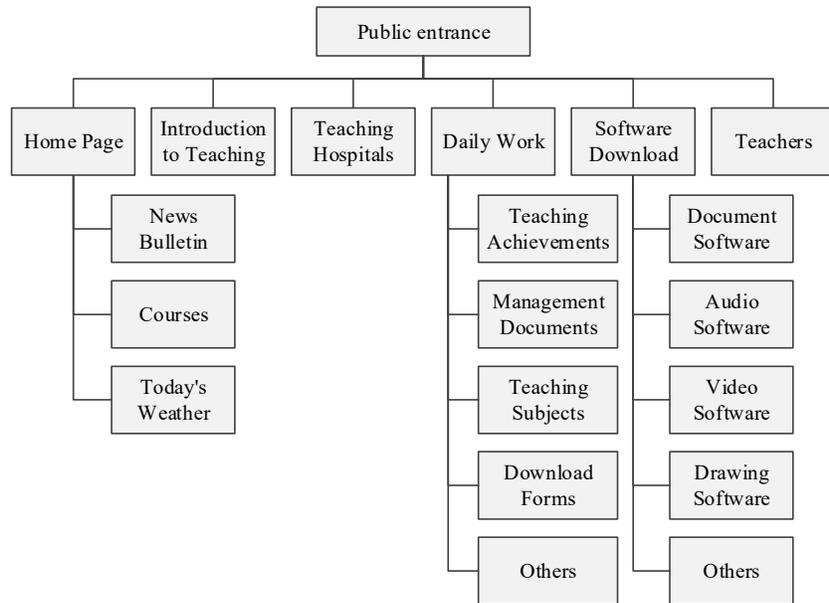
Cloud computing is an Internet-based information technology service model that provides on-demand, scalable, and flexible computing and storage resources through a shared pool of virtualized resources [11]. As an efficient information technology solution, the resource management of cloud computing plays a crucial role, which can realize the self-service of individual demand service of resources, dynamic resource allocation, convenient access, virtualized resources, measurable services, effective integration of resources distributed in different areas, and complete the shared use of terminal infrastructure, and optimization and automatic control of resources [12]-[14].

In this paper, we first start from the platform architecture design to analyze the supportive role of public function module and permission hierarchy mechanism for data collection. Aiming at the correlation and clustering needs in personnel management, Apriori algorithm and improved SDCN model are adopted respectively to realize group feature recognition through rule extraction and attribute graph embedding. Based on the 2022-2024 data of a medical university, the heterogeneous characteristics of different job groups are revealed. Preprocess the data to improve the data quality. Examine the advantages of Apriori algorithm performance performance through controlled
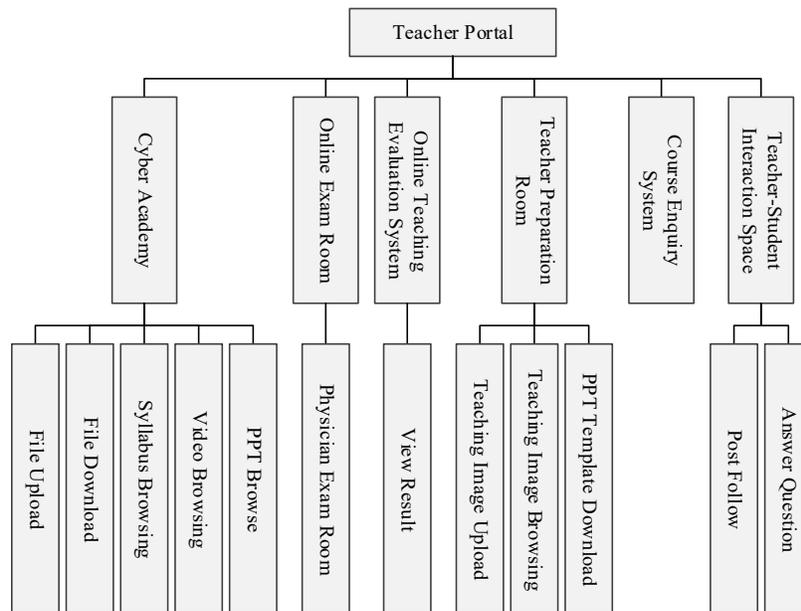
experiments. Utilizing SDCN model to achieve personnel clustering and reveal the structural imbalance of faculty positions in this university.

II. Intelligent management of staff resources for teaching platforms based on cloud computing
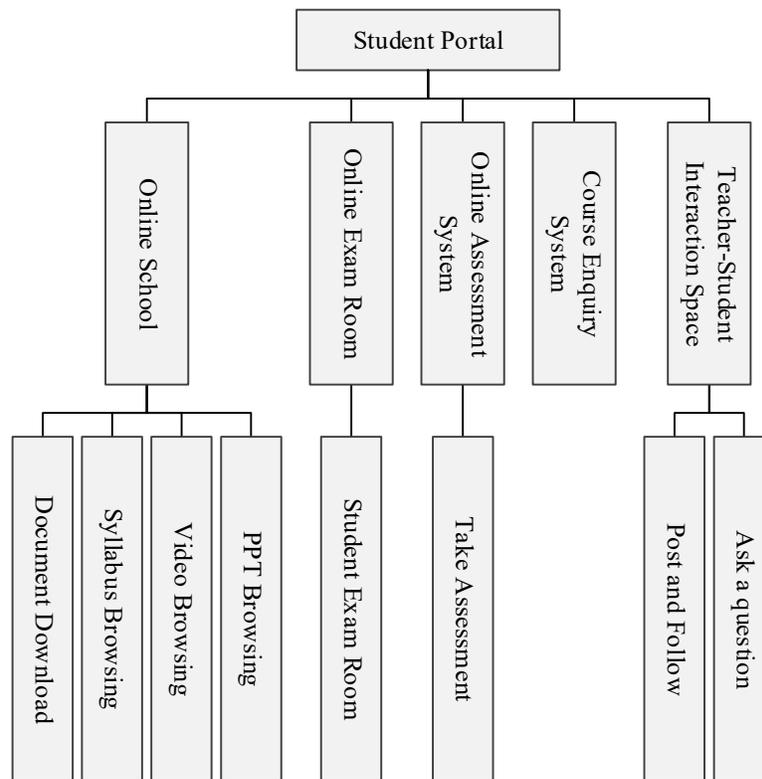
With the digital transformation of medical education, medical school teaching platforms are facing challenges such as low resource allocation efficiency and complexity of personnel roles. The traditional management mode relies on manual experience, which is difficult to dynamically adapt to multi-dimensional talent characteristics and teaching needs. This paper focuses on cloud computing technology to build an intelligent personnel resource management system, and proposes a solution that integrates association rule mining and depth graph clustering.

(a) Public entrance

(b) Teacher's Entrance

(c) Student Entrance

Figure 1: Structural framework

II. A. Architecture of the teaching platform in medical schools and colleges

II. A. 1) Structural framework

The web platform is divided into a public function section and a privileged use section, specifically divided into three modes: "Public Entrance", "Teacher's Entrance" and "Student's Entrance". If you are not logged in, you can view the entire contents of the site, but you are not authorized to use it. After logging in, there are two types of statuses: teachers and students, who can browse all the contents of the "Public Portal", but can also use different modules. For the above three cases, the three model frameworks are shown in Figure 1(a~c).

II. A. 2) Public Functional Use section

The module can be browsed without logging in. Among them, the home page has the most basic "News Bulletin" and "Fine Courses" columns, through the "News Bulletin" timely release of relevant news and notices, etc., which is convenient for users to understand the latest news at the first time, while the "Fine Courses" column is highly focused on displaying teaching achievements. The "News Bulletin" releases relevant news and notices in a timely manner, which is convenient for users to understand the latest news, while the "Fine Courses" section highly concentrates on displaying the teaching achievements. The "Teaching Introduction" section mainly introduces the existing teaching hardware and equipment in the medical school, which is convenient for users to make self-service inquiries on teaching resources. The understanding of teaching hospitals also helps to allocate and adjust teaching tasks, and helps teachers and students to further grasp teaching resources, and this part of the introduction and related links are summarized in detail in the "Teaching Hospitals" section. The "Daily Work" section includes information on teaching achievements and topics, downloadable forms and related management documents. In order to allow users to download and use the latest software tools more conveniently, a special "Software Download" area has been opened, which can be uploaded freely and released to the platform for download after strict examination and classification. Students and outsiders can learn more about the composition of the faculty of the School of Medicine through the "Famous Teachers" section.

II. B. Association rule-based data mining

Association rules refer to the discovery of relationships between items in a data set. For example, during shopping in a mall, if a customer buys shoes, then he is likely to buy socks, and if he buys bread, then he is likely to buy milk.

In commercial sales, merchants can apply similar management rules as above to bundle goods in order to obtain higher revenues; in the insurance business, if a combination of claims from users is very rare, then an in-depth investigation is needed to prevent fraudulent insurance behavior; in the medical field, the existence of potential treatment combinations can enable patients to speed up the recovery process; in the banking industry, customers are analyzed in terms of their income and consumption and then analyze the customer's income and consumption, and then offer him or her the right relevant business.

A rule is something like "If...Then...", where the former is the condition and the latter is the result. For example, a customer, if he buys bread, then he may also buy milk, this is an association rule, indicating that the goods "bread" and "milk" between the existence of a combination of purchases.

Association rule is a potential relationship between different data items, this relationship is reflected in the information value of the data item, so you can use the data value of one object to estimate the data value of another object through the association relationship. For example, in shopping basket analysis, Milk $\Rightarrow$ Bread [Support: 3%, Confidence: 40%]. A support level of 3% means that out of every one hundred shoppers, three shoppers will buy both milk and bread. A confidence level of 40% means that of the customers who buy milk, 40% are likely to buy bread again. Association rule mining is the use of some algorithm to compute relationships between certain individual items in a given set of data so that the subsequent work can be served by these relationships.

Definition 1 (Support): support s denotes the ratio of transactions containing $A \cup B$ in the transaction database D, i.e., the probability $P(A \cup B)$, denoted as $\sup port(AB) = P(A \cup B)$, which denotes the probability that the concatenated set of A and B occurs in all the transactions.

Definition 2 (Confidence): confidence denotes the rate, i.e., the probability $P(B \mid A)$, that the transaction database D contains both A and B transactions, and is denoted as $confidence(AB) = P(B \mid A)$.

Definition 3 (Frequent itemsets): the set of items whose support is not less than the minimum support threshold given by the user. All frequent 1-itemsets are denoted as L1.

As an example, for a purchase record, the number of transactions in which a customer buys an Apple is 4, while the number of transactions in which he buys both an Apple and a Banana is 2.

Confidence indicates the degree of trustworthiness of a rule. For example, the confidence level of "If Apple then Banana" is calculated. Since 2 out of 4 customers who purchased Apple also purchased Banana, the confidence level is 0.5.

Support indicates the probability of having both A and B. Among the above customer purchase records, there are 2 records that have both Apple and Banana. The support of this rule is $2/5 = 0.4$.

For the rule "If Band Cthen A", 33.33% of the people who bought both B and C would buy A. And the support level of A is 0.45, which means that 45% of customers will buy A. This rule is not meaningful enough to be used by the mall to guide customers in their shopping.

In order to measure whether a rule has practical guidance or not, the concept of enhancement is introduced. It indicates how much the rule can be improved if it is used in response to the fact that the rule does not apply:

$$Lift(A \Rightarrow B) = \frac{Confidence(A \Rightarrow B)}{Support(B)} = \frac{Support(A \Rightarrow B)}{Support(A) \times Support(B)} \tag{1}$$

In the above example, $Lift(If\ Band\ C\ then\ A) = 0.05/(0.15*0.45) = 0.74$, while $Lift(If\ A\ then\ B) = 0.25/(0.45*0.42) = 1.32$. It means that if the merchant recommends product B to the person who has purchased product A, the probability of the customer purchasing product B is 1.32 times higher than that of the random recommendation.

The most representative algorithm for association rule mining is Apriori, which employs an iterative layer-by-layer search where k-itemsets are used to search for $(k+1)-$itemsets. First, find the set of frequent 1-term sets. This set is denoted as L1. Use L1 to search to the set L2 of frequent 2-item sets, use L2 to search to L3, and so on until no frequent k-item sets can be searched.

II. C.Personnel management based on attribute graph clustering algorithm

In the era of big data, the performance of traditional graph clustering algorithms is limited in the face of large-scale, high-dimensional and structurally complex data. Deep learning makes it possible to deal with large-scale high-dimensional complex graph data by virtue of its powerful representation capability. Therefore, deep clustering algorithms have also become a hot research direction in the field of unsupervised learning. Researchers have started to combine deep learning and traditional clustering algorithms to implement clustering tasks to improve clustering performance. However, some of these methods implement node embedding learning and clustering tasks step by step, which makes the node embedding not necessarily adapted to the subsequent clustering task and limits the performance of the algorithms. In order to leverage the powerful representation capabilities of neural

networks to improve clustering performance, researchers have proposed a number of end-to-end neural network clustering algorithms that jointly optimize the clustering process and node embedding learning through gradient descent algorithms in a unified framework. Such algorithms introduce a clustering loss function to pre-train node embeddings $Z \in \square^{n \times d}$ in a self-supervised manner, and perform a one-time clustering of node embeddings using traditional clustering algorithms to obtain the initial clustering centers $C \in \square^{k \times d}$, where n, k, and d respectively denote the number of nodes, the number of clusters, and the potential feature dimensions in the graph. Then the soft distribution distribution between the node embedding and the initial clustering center is calculated, followed by generating the target distribution through the soft distribution distribution, and finally using the KL dispersion to constrain the two distributions as the clustering loss of the algorithm, and the clustering loss of the whole model can be expressed as shown in Eq. (2):

$$L = \alpha L_c + \beta L_n \tag{2}$$

where L_c and L_n denote the clustering loss and the neural network loss, respectively, L_n is used to learn the node features, and L jointly optimizes the node embedding and clustering.

Deep graph clustering utilizes techniques such as graph neural networks to obtain the embeddings of nodes in the attribute graph and then classify them into different class clusters, the process is shown in Equation (3):

$$Z = F (G_A) = F (X, A) \tag{3}$$

where $X \in \square^{n \times d}$ denotes the attribute matrix of the node, $A \in \square^{n \times n}$ denotes the adjacency matrix, $Z \in \square^{n \times d}$ denotes the graph neural network learning the node embeddings, and F denotes the graph neural network. After obtaining the node embeddings, the node embeddings are classified using the clustering algorithm C as shown in Equation (4):

$$\hat{y} = C(Z, k) \tag{4}$$

where k denotes the number of class clusters and $\hat{y} \in \square^n$ denotes the predicted cluster ID vector. After the clustering process, the performance of the algorithm is evaluated using the evaluation method M as shown in Equation (5):

$$s = M (y, \hat{y}) \tag{5}$$

where $s \in \square$ denotes the clustering score and $y \in \square^n$ denotes the groundtruth label vector.

The node attribute information and graph structure information in the attribute graph are conducive to mining graph information and can improve the accuracy of clustering algorithms, and thus are more popular in deep graph clustering. Attribute graph clustering usually uses node attributes and graph topology to obtain graph embeddings, and optimizes the clustering results through self-supervised clustering iterations. Several typical algorithms for attribute graph clustering are described below.

The DAEGC algorithm is an early implementation of an end-to-end trained attribute graph clustering framework, which consists of a graph attention self-encoder module and a self-supervised clustering module. In DAEGC algorithm a modified graph attention network is used to encode the graph considering the higher order neighbor information of the nodes to get the node embeddings, and then reconstruct the topological information by dot product decoder, the loss function of the graph attention self-encoder module is shown in Eq. (6):

$$L_{re} = \sum_{i=1}^{N} loss(A_{i,j}, \hat{A}_{i,j}) \tag{6}$$

The node embedding is clustered by applying the traditional clustering algorithm k-means, obtaining the initial clustering center μ, calculating the soft assignment distribution Q and the target distribution P, and minimizing the KL dispersion between the two, and the final clustering loss is shown in Eq. (7):

$$L_c = KL(P \square Q) = \sum_{i} \sum_{u} p_{iu} \log \frac{p_{iu}}{q_{iu}} \tag{7}$$

where p_{iu} and q_{iu} denote the target and soft assignment distributions, respectively, and are computed as shown in Eq. (8) and Eq. (9):

$$q_{iu} = \frac{\left(1 + \|z_i - \mu_u\|^2\right)^{-1}}{\sum_k \left(1 + \|z_i - \mu_k\|^2\right)^{-1}} \tag{8}$$

$$p_{iu} = \frac{q_{iu}^2 / \sum_i q_{iu}}{\sum_k \left(q_{ik}^2 / \sum_i q_{ik}\right)} \tag{9}$$

$$L = L_{re} + \gamma L_c \tag{10}$$

Finally, the whole DAEGC model is jointly optimized by Eq. (10), where $\gamma \geq 0$ is used to balance the reconstruction loss and clustering loss of the graph selfencoder.

Based on DAEGC, this paper proposes the SDCN algorithm, whose model framework is shown in Fig. 2. The SDCN algorithm consists of a DNN module, a GCN module, and a dual self-supervision module. The SDCN combines attribute information and structural information to realize deep graph clustering. The DNN module learns the attribute information through the selfencoder to alleviate the oversmoothing problem of the GCN. The GCN module learns the structural information through the constructed KNN graph to learn structural information and passes the representation learned by the DNN module to the corresponding GCN layer through the pass operator. The dual self-supervised module connects the GCN module and the DNN module together to guide the update of the whole model. The DNN module learns the node features through attribute reconstruction with a loss function as shown in Eq. (11):

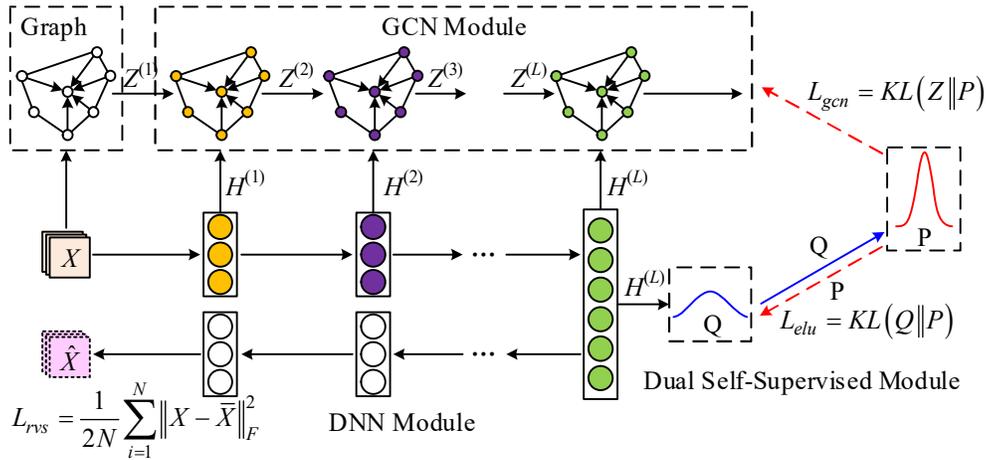$$L_r = \frac{1}{2n} \sum_{i=1}^{n} \|x_i - \hat{x}_i\|_2^2 = \frac{1}{2n} \|X - \hat{X}\|_F^2 \tag{11}$$

Figure 2: SDCN Framework

The transfer operator in the SDCN model refers to the splicing of the representation $H^{(l)}$ obtained by the DNN module and the representation $Z^{(l)}$ obtained by the GCN module as inputs to the next layer of the GCN, and the splicing is shown in Eq. (12):

$$\tilde{Z}^{(l-1)} = (1 - \eth)Z^{(l-1)} + \eth H^{(l-1)} \tag{12}$$

The probability of each node belonging to each node is obtained using the softmax function in the last layer of the GCN module as shown in Equation (13):

$$Z = soft\max\left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} Z^{(L)} W^{(L)}\right) \tag{13}$$

In the dual self-supervision module, the representations obtained from DNN are first clustered to obtain the initial clustering centers, and then the soft distribution Q and the target distribution P are computed as shown in Eqs.

(8) and Eq. (9), and the target distribution P supervises the probability distributions Q and the probability distributions Z of the GCN module as shown in Eqs. (7) and Eq. (14):

$$\mathrm{L}_{gcn} = KL(P \Vert Z) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{z_{ij}} \tag{14}$$

The final objective function of the SDCN algorithm is shown in Equation (15):

$$\mathrm{L} = \mathrm{L}_r + \alpha \mathrm{L}_c + \beta \mathrm{L}_{gcn} \tag{15}$$

III. Empirical Analysis of Intelligent Management of Personnel Resources on Teaching Platforms in Medical Colleges and Universities

This study is based on the operational data of the teaching platform of a medical university for the years 2022-2024, covering the records of four categories of personnel: lecturers, associate professors, professors, and external experts. The data were obtained from log files, performance appraisal forms and questionnaires from the platform database, which were anonymized and used for analysis. The research objective is to construct dynamic and precise personnel resource management through association rule mining and attribute graph clustering algorithms to optimize the efficiency of human resource management on the teaching platform.

III. A. Association rule-based personnel data mining
III. A. 1) Analysis of association rules
Apriori algorithm was used to set the minimum support level of 30% and confidence level of 70%, and the variables were selected: position, education, gender, age and research output index. Positions were categorized into lecturers, associate professors, professors, and external experts, focusing on the relationship between positions and other variables. Meaningful rules were picked from the results run out by Clementine software and the results of the correlation analysis are shown in Table 1.

The analysis shows that the distribution of the characteristics of the different position groups shows significant heterogeneity. In the group of professors, the combination of doctoral degree and high research output (>0.8) accounted for 70.38% of the total, with a high confidence level of 94.27% (Rule 1), while associate professors were more concentrated in the group of 30-40 years old and with doctoral degree (Rules 2 and 8), and external experts with doctoral degree of 26-30 years old accounted for 40.85% of the total (Rule 9). 52.57% of the lecturers were in the age of 25-30 years old and with insufficient scientific output with a confidence level of 82.55% (Rule 5).

Table 1: Results of Correlation Analysis

Serial number	Antecedent	Consequent	Support/%	Confidence/%
1	Educational background=Doctoral degree, Research output index>0.8	Professor	70.38	94.27
2	Educational background=Doctoral degree, Age=30-40 years old	Associate professor	65.28	92.56
3	Age=30-40 years old, Research output index>0.8	Professor	50.78	87.58
4	Educational background=Doctoral degree, Research output index>0.8	External expert	32.85	85.26
5	Age=25-30 years old, Research output index<0.8	Lecturer	52.57	82.55
6	Age=26-30 years old, Research output index>0.8	Professor	57.58	80.17
7	Educational background=Doctoral degree, Gender=Female	External expert	52.66	77.36
8	Educational background=Doctoral degree, Research output index>0.8	Associate professor	54.29	75.17
9	Educational background=Doctoral degree, Age=26-30 years old	External expert	40.85	73.22
10	Education level=Doctoral degree, Gender=Female	Lecturer	40.39	70.68

III. A. 2) Data pre-processing

The personnel resources data preprocessing solution includes data cleansing, data transformation and data standardization. Data cleansing is to use mean value filling or deletion processing for missing values, and 3σ guidelines are used to identify and process outliers. The data conversion stage focuses on solving the data format unification problem. For category-based data, unique thermal coding is used for conversion, and for text data, feature vectors are extracted using the TF-IDF method. The standardization of the converted data is done using the Min-Max method, which maps the data to the [0,1] interval. The data preprocessing effect was evaluated using the data quality index system, including four dimensions of completeness, accuracy, consistency and timeliness, and the evaluation results are shown in Table 2. Through data cleansing, data completeness increased from 79.66% to 97.39%, and accuracy, consistency and timeliness reached 98.66%, 97.28% and 99.65%, respectively. The standardized coverage of the processed data reached 99.8%, setting the foundation for subsequent classification analysis. The results of the evaluation showed that the preprocessing programme had significantly improved data quality and that all indicators had met the expected goals.

Table 2: Evaluation of Data Preprocessing Effect (%)

Evaluation dimension	Before processing	After processing	Increase amplitude
Integrity	79.66	97.39	17.73
Accuracy	83.24	98.66	15.42
Consistency	75.17	97.28	22.11
Timeliness	88.64	99.65	11.01

III. A. 3) Modeling of association rules

The rules were evaluated using three indicators of support, confidence and enhancement, and the results of association rule analysis are shown in Figure 3. The results show that after data preprocessing, the association rules generated by mining the four types of job rules for lecturers, associate professors, professors, and external experts have an accuracy rate of 90.56%, a recall rate of 89.78%, and an F1 value of 90.17%, which verifies the effectiveness of data preprocessing.
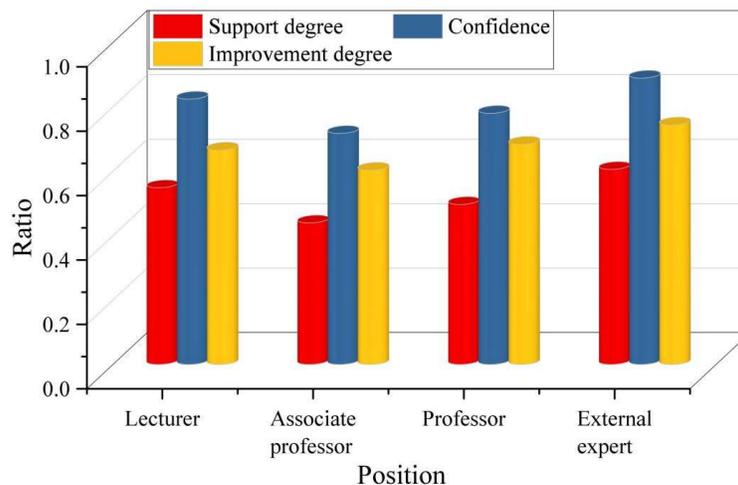

Figure 3: Analysis results of association rules

III. A. 4) Comprehensive comparison of performance levels

In order to verify the effectiveness of Apriori algorithm, LCM and FP-Growth are selected as control algorithms, and the results of the comparison of the performance of different algorithms are shown in Table 3. Apriori algorithm is better than the other algorithms in all performance indexes, and in the aspect of the accuracy, Apriori algorithm reaches 96.28%, which improves by 11.11% compared to LCM algorithm, and improves by 6.06% compared to FP-Growth algorithm. Growth algorithm by 6.06%. In terms of time efficiency, when processing 250,000 pieces of data, Apriori algorithm takes 37.65s, which is 80% higher than LCM algorithm and 60.3% higher than FP-Growth algorithm. Rule quality analysis shows that the rule set generated by Apriori algorithm has a significant advantage in reliability

and coverage, with 90.27% rule coverage and 93.66% rule reliability. Apriori algorithm has only 506MB memory occupation, which is much lower than LCM and FP-Growth algorithms, while the performance of Apriori algorithm is optimal. Algorithm stability test results show that the Apriori algorithm maintains stable performance under different data sizes and parameter settings, and the standard deviation is controlled within an acceptable range.

Table 3: Comprehensive Comparison of Performance Levels

Performance index	Apriori	LCM	FP-Growth
Accuracy rate/%	96.28	85.17	90.22
Time consumption/s	37.65	188.32	94.83
Memory usage/MB	506	983	691
Rule coverage rate/%	90.27	79.38	84.22
Rule reliability/%	93.66	84.52	89.38

III. B. Personnel resource management based on the SDCN model

Four variables, namely education, gender, age and research output index, were selected for cluster analysis using SDCN algorithm, and four typical personnel were selected to serve as the initial center of mass. The number of clustering classes K was set to 4, the number of iterations was set to 20, and then the test was conducted, and the final clustering center and some of the clustering members were shown in Tables 4 and 5, respectively. The SDCN algorithm successfully clustered the teachers of this medical university into 4 categories, and the distance between each clustered personnel and the clustering center was calculated.

Table 4: Final Clustering Centers

	Clustering			
	1	2	3	4
Educational background	2	4	6	5
Gender	2	0	1	0
Age	3	3	4	4
Research output index	1	7	8	5

Table 5: Clustering Member Situation (Part)

Case Number	Clustering	Distance	Case Number	Clustering	Distance
1	1	1.673	11	4	0.934
2	1	1.246	12	4	1.208
3	1	0.937	13	4	0.836
4	2	1.265	14	2	0.901
5	2	0.736	15	2	1.037
6	2	0.883	16	3	1.184
7	2	1.108	17	3	1.209
8	3	1.119	18	3	0.937
9	3	1.023	19	1	0.922
10	2	0.937	20	4	0.835

For the second experiment, this paper attempts to verify whether there is a large effect of certain variables on the results of clustering. After comprehensive consideration, it was decided not to use gender as a condition for verification. Through arithmetic, the clustering results were obtained as shown in Table 6. Comparing with the results in Table 4, it can be found that the final center of mass of the second clustering is basically the same as that of the most second clustering, indicating that gender does not have a great influence on the results of the final clustering.

Table 6: Results of the second clustering

	Clustering			
	1	2	3	4
Educational background	2	4	5	5
Age	3	3	4	4
Research output index	1	6	8	5

Through the implementation of the verification SDCN algorithm can complete the analysis of human resource management research, in order to facilitate the staff to more intuitive observation of the distribution of personnel, this paper will be the position as a horizontal coordinate, the cluster number as a vertical coordinate, the number of personnel in each cluster results as a value, the clustering visualization results as shown in Figure 4. It can be very intuitively seen that the personnel of associate professors have only cluster numbers 2 and 3, and cluster numbers 1 and 4 are missing. The number of cluster number 3 in the personnel of professors is significantly larger than the number of personnel in other positions. From this, it can be judged that the existing personnel distribution has both the problem of structural imbalance between positions and the imbalance of personnel structure within positions, and it is urgent for the personnel resource management department to put forward a reasonable proposal for personnel adjustment.
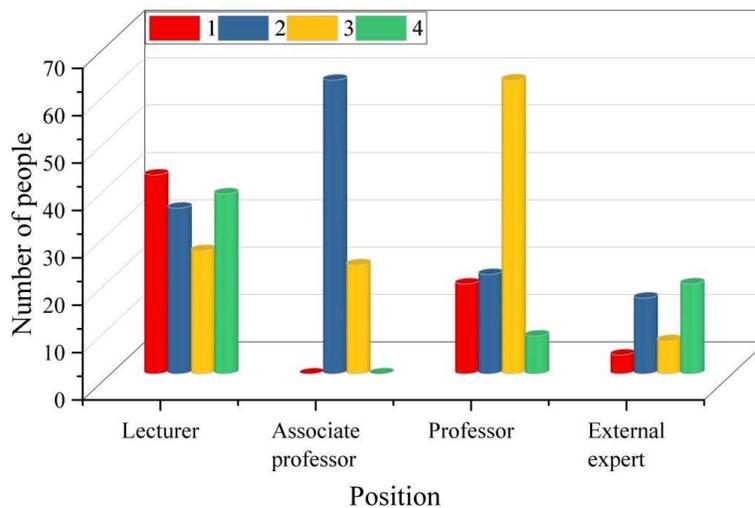

Figure 4: Visualization results of clustering

IV. Conclusion

In this study, a personnel resource management system adapted to the needs of medical schools was constructed through the integration of cloud computing architecture and intelligent algorithms.

The correlation rules showed that the combination of doctoral degree and high research output (>0.8) accounted for 70.38% of the professorial group with a high confidence level of 94.27% (Rule 1), while associate professors were more concentrated in the group of 30-40 years old and with doctoral degree (Rule 2 and 8), and external specialists with doctoral degree of 26-30 years old accounted for 40.85% of the total number of professors (Rule 9).52.57% of the lecturers were in the age of 25-30 years old and with insufficient research output with a confidence level of 82.55% (Rule 5).

Through data cleansing, the data integrity was improved from 79.66% to 97.39%, and the accuracy, consistency and timeliness reached 98.66%, 97.28% and 99.65%, respectively. The accuracy of the generated association rules reaches 90.56%, the recall rate is 89.78%, and the F1 value is 90.17%. The Apriori algorithm outperforms other algorithms in all performance indicators, with an accuracy of 96.28%, a time consuming of 37.65s, a rule coverage rate, and a rule reliability of 90.27%, and 93.66%, respectively. The Apriori algorithm has an optimal performance in the At the same time, the memory occupation is only 506MB, which is much lower than LCM and FP-Growth algorithms.

Using the SDCN algorithm for cluster analysis, there are only clusters 2 and 3 for associate professors, and clusters 1 and 4 are missing. The number of cluster number 3 in the personnel of professors is significantly larger

than the number of personnel in other positions. From this, it can be judged that the existing personnel distribution has both structural imbalance between positions and imbalance of personnel structure within positions.

References

[1] Kuhn, S., Frankenhauser, S., & Tolks, D. (2018). Digital learning and teaching in medical education: Already there or still at the beginning?. Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz, 61, 201-209.

[2] Speidel, R., Schneider, A., Körner, J., Grab-Kroll, C., & Öchsner, W. (2021). Did video kill the XR star? Digital trends in medical education before and after the COVID-19 outbreak from the perspective of students and lecturers from the faculty of medicine at the University of Ulm. GMS Journal for Medical Education, 38(6), Doc101.

[3] Han, E. R., Yeo, S., Kim, M. J., Lee, Y. H., Park, K. H., & Roh, H. (2019). Medical education trends for future physicians in the era of advanced technology and artificial intelligence: an integrative review. BMC medical education, 19, 1-15.

[4] Scott, K., Morris, A., & Marais, B. (2018). Medical student use of digital learning resources. The clinical teacher, 15(1), 29-33.

[5] Kyaw, B. M., Posadzki, P., Paddock, S., Car, J., Campbell, J., & Tudor Car, L. (2019). Effectiveness of digital education on communication skills among medical students: systematic review and meta-analysis by the digital health education collaboration. Journal of medical Internet research, 21(8), e12967.

[6] Schiekirka-Schwake, S., Anders, S., Von Steinbüchel, N., Becker, J. C., & Raupach, T. (2017). Facilitators of high-quality teaching in medical school: findings from a nation-wide survey among clinical teachers. BMC medical education, 17, 1-8.

[7] Chy, T. I., Das, A. K., Sarker, B. K., Akter, S., Azad, R. R., & Naznin, A. (2023). Teachers' and Students' Views Regarding Acute Shortage of Teachers of Basic Subjects in Medical Colleges of Bangladesh. The Planet, 7(01), 208-217.

[8] Zhu, H., Xu, J., Wang, P., Liu, H., Chen, T., Zhao, Z., & Ji, L. (2023). The status of virtual simulation experiments in medical education in China: based on the national virtual simulation experiment teaching Center (iLAB-X). Medical Education Online, 28(1), 2272387.

[9] Pappada, S., Owais, M. H., Aouthmany, S., Schneiderman, J., Toy, S., Schiavi, A., ... & Rega, P. (2022). Personalizing simulation-based medical education: the case for novel learning management systems. Int J Healthc Simul, 10.

[10] Dash, S. (2019). Google classroom as a learning management system to teach biochemistry in a medical school. Biochemistry and molecular biology education, 47(4), 404-407.

[11] He, S., Cheng, B., Wang, H., Huang, Y., & Chen, J. (2017). Proactive personalized services through fog-cloud computing in large-scale IoT-based healthcare application. China Communications, 14(11), 1-16.

[12] Mousavi, S., Mosavi, A., Varkonyi-Koczy, A. R., & Fazekas, G. (2017). Dynamic resource allocation in cloud computing. Acta Polytechnica Hungarica, 14(4), 83-104.

[13] Wen, Y., Li, Z., Jin, S., Lin, C., & Liu, Z. (2017). Energy-efficient virtual resource dynamic integration method in cloud computing. IEEE Access, 5, 12214-12223.

[14] Nawrocki, P., & Reszelewski, W. (2017). Resource usage optimization in mobile cloud computing. Computer Communications, 99, 1-12.