



Data Mining of Chinese Language and Literature Education Based on Diversified Constraint Model

Xiaolan Wei^{1,2}(✉)

¹ Nanchang Institute of Science and Technology, Nanchang 330108, Jiangxi, China
565607996@qq.com

² School of Economics and Management, Guilin University of Electronic Technology,
Guilin 541000, China

Abstract. China culture. When teaching Chinese language and literature, teachers should vigorously and guide the teaching, so as to effectively enhance students' learning enthusiasm. All colleges and universities have established their own teaching management information systems, which have greatly improved the level of teaching and management, and accumulated a large amount of teaching and management data. But at present, most of these information systems are on-line transaction processing, lacking the ability of comprehensive analysis and decision-making. Data mining technology is produced to solve this problem. If the function of classifying and deep-level analysis of the collected data cannot be provided, the managers cannot understand the relevant information in time in the face of complex situations, and the leaders do not have practical data support in decision-making. As the core component in the field of digital education practice, diversity constraint model understands the structural form of Chinese language and literature education data mining in the field of learning practice with internal psychological mechanism and external behavior structure.

Keywords: Chinese language and literature education · Data mining · Multiple constraint model

1 Introduction

With the advent of the information age, data mining has become an important tool for research in various fields as an effective means. In the field of education, Chinese language education has always been a highly concerned field, and educators have been trying to explore the best educational methods through continuous practice to improve students' Chinese literacy.

However, the research on data mining in Chinese language education is relatively limited compared to other fields. This study will explore data mining in Chinese education based on a multi constraint model. The multivariate constraint model not only captures the complex relationships between data well, but also accurately controls the

correlation between data. Therefore, if successfully applied to data mining in Chinese language education, it will greatly improve the efficiency of education and teaching.

Therefore, data mining methods have become a hot topic in the field of information management and have attracted the attention of many scholars. Data mining originated in the mid-1990s. This is a young and active research field. It is a product of the combination of multiple disciplines and technologies. It has broad application value [1]. This also brings great convenience. The most commonly used is data mining based on multi constraint models. Establish a digital portrait of learners, explain the correlation logic between models, and provide intelligent teaching decision support for personalized learning and learners' comprehensive development.

At present, although it is necessary to have a full understanding of its target audience, namely students. At the same time, combining measures and strategies of characteristic education, form a management system with its own characteristics. The problem of course arrangement is not only about simply allocating courses within a reasonable time frame, but also considering the various needs of teachers and students in choosing courses [2]. In theory, it can enrich curriculum theory, deepen people's understanding of the relationship between curriculum reform, and clarify the goals and theoretical basis of curriculum reform. If we want to reflect the different requirements of teachers and students in the course arrangement results, then we need to set some constraints in the course arrangement process to make the course arrangement results meet our expectations.

This study aims to mine Chinese language education data through a multi constraint model, and adopt corresponding educational strategies based on the mining results to improve the quality of education and promote students' learning outcomes. This study will be based on a certain amount of Chinese language education data, including educational goals, learning materials, educational methods, etc., and use multiple constraint models to model and analyze these data.

Through research, we will discover which factors have the most significant impact on improving students' Chinese literacy, and propose corresponding educational strategies accordingly [3]. The research significance of this study is to provide a new approach for the field of Chinese language education, while also providing reference for other educational fields.

Through the experiments and explorations of this study, we are expected to achieve some remarkable results that may be applied in practical education. This will play an important role in the development of future Chinese education and the improvement of students' Chinese literacy. Therefore, this study has important theoretical and practical significance.

2 Data Mining Method Based on Diverse Constraint Model

2.1 Main Methods of Data Mining

Data mining is a process that utilizes computers to automatically analyze large-scale datasets, using various complex algorithms and techniques to extract valuable information from data. The choice of data mining methods depends on the type of data, the

nature of the problem, and the researcher's goals. Here are several common data mining methods:

1. Classification and regression analysis

Classification and regression analysis is a supervised learning method. They are based on labeled datasets and predict the labels or numerical results of new unlabeled data through classifiers or regression models [4].

2. Cluster analysis

Cluster analysis is an unsupervised learning method. It divides data into different sets or groups by searching for similarities.

3. Correlation analysis

Association analysis is a method of mining frequent patterns in data. It can reveal the connections between different elements in the data, such as shopping basket analysis, which can be used to find frequently purchased items together.

4. Abnormal detection

Anomaly detection is a method of determining whether data is abnormal or does not conform to expected patterns. It can detect data points that are different from other data and find data that is slightly different from the general public. Application scenarios such as fraud detection, determining whether equipment needs maintenance, etc.

5. Time series analysis

Time series analysis is a method of processing time series data. It can identify and predict trends and periodicity in time series, and is applied in fields such as stock market and weather forecasting.

The above are some common data mining methods, but there are actually many other methods, such as decision trees, support vector machines, etc. When selecting data mining methods [5], it is necessary to consider the specific scenario and problem nature, and comprehensively select the most suitable method.

Figure 1 shows the relationship between data mining and other fields.

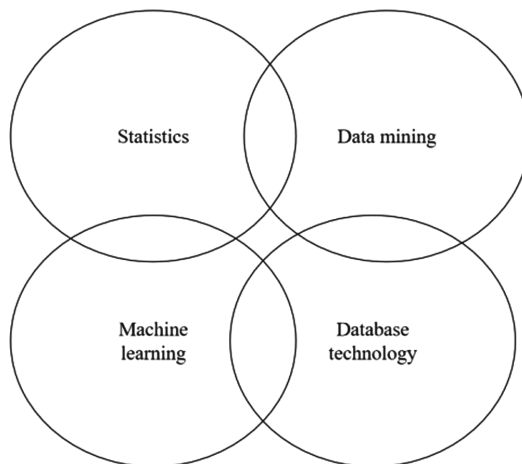


Fig. 1. Data mining is a collection of many disciplines

The methods and technologies of data mining mainly include statistical analysis method, association rule method, decision tree method and so on. To some extent, there is a problem of three-dimensional goal imbalance, but does not pay enough attention to the goal of ability, process and method, emotional attitude and values.

Statistical analysis method is a method to describe and infer the overall information and knowledge by analyzing the sample data in the population. These information and knowledge suggest the internal laws in the population. The statistical analysis method can be completely described as high-dimensional matching.

It is a simple but practical analysis rule. Assuming that the capacity lower bound constraint is not considered, we use Da algorithm to obtain a stable solution, but it is not satisfied with the capacity lower bound constraint [6]. Association rules are widely used in transaction data analysis to guide sales, catalog design and other market decision-making through the analysis results.

Decision tree method is derived from machine learning. It is an inductive learning method based on examples. Characterize the process of learners acquiring knowledge, standardize the data exchange of learners' information in different systems, standardize the learners' information with static features with semantics and syntax, and focus on the presentation of learners' internal relations.

2.2 Data Mining Process

Every semester, the school has to conduct classroom teaching evaluation survey, which has accumulated a lot of data. The knowledge level and mastery level of learners are described through the universal characteristics of knowledge extraction, but the solution of problems in the field of learning practice can not accurately describe the change of knowledge structure and specific cognitive state. We will introduce data mining based on diverse constraint models [7]. The basic model of association rule mining is shown in Fig. 2.

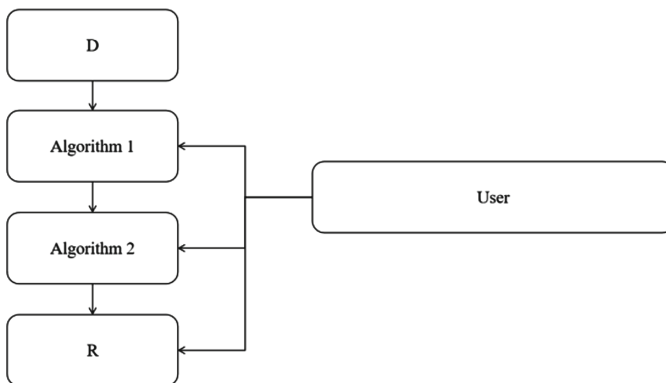


Fig. 2. Basic model of data mining

Preparation stage this stage includes problem definition, object understanding, data collection and other preparations, collects the whole space-time and multi-dimensional

data collection based on multimodal perception is realized, and the potential characteristics of learners such as “knowledge, cognition, emotion and interaction” are mined.

It is necessary to construct the initial feasible solution to meet all the hard constraints, which can usually be given by random sampling or experience.

Data mining includes algorithm design and analysis model establishment. Encode a specific problem into a string in binary form to generate the original feasible solution. Learner feature analysis model can be designed from six dimensions: basic information, learning style, emotional state, interest preference, social network and cognitive level. By integrating scene awareness modeling method and frequent sequence mining algorithm, learner feature values of each dimension in different learning scenarios can be calculated.

The post-processing stage includes result interpretation, output, evaluation, analysis and use, and explains and evaluates the results. The constraints are easy to describe, and can usually approach or even reach the global optimal solution in specific calculation examples. Visual graphics and colors are used to represent learners’ cognitive process of knowledge, and adaptive knowledge navigation is used to allow learners to view and reflect on their mastery of knowledge, discover myth concepts and promote the improvement of metacognitive ability.

3 Establishment of Diversity Constraint Model and Analysis of Diversity as Upper and Lower Bound Constraints

3.1 Establishment and Analysis of Multiple Constraint Model

In the diversity of the same data attribute is fully considered. Therefore, unnecessary cumbersome data features can be reduced. The establishment of multiple constraints model collects the activity information of different learning service systems across platforms through unified description specifications and operation mechanism, so as to reflect learners’ social status and information sharing. The multi constraint model is shown in Fig. 3.

The following formula can be used to calculate the change parameters of different data attribute characteristics:

$$e(y, z) = \left[\sum_{j=1}^p |y_j - z_j|^s \right]^{\frac{1}{s}} \quad (1)$$

y --The amount of data in the database;

z --Number of attributes;

s --Difference of attribute characteristics.

The analytical mechanism of big data and learning analysis provides technical support for objectively understanding learners’ individual characteristics and establishing perceptual digital portraits of learners, and realizes the calculation and analysis of learners’ multi-dimensional characteristic values in the fields of cognitive level, knowledge structure, behavior and emotion integration.

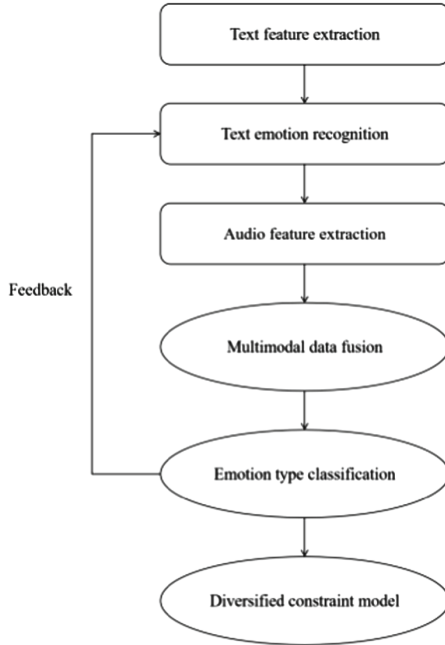


Fig. 3. Multiple constraint model

By processing the data in the database, the data variables can be obtained, and more than 90% of the database information characteristics can be obtained. The change of data is based on the collaborative change of sub model, showing the learning level of a certain stage and the overall degree of understanding things. According to the empirical data:

$$e(y, z) = \left[\sum_{j=1}^p b_{ij} |y_j - z_j|^s \right]^{\frac{1}{2}} \tag{2}$$

b_{ij} --The j and j data attributes in the database

Classify provide accurate data basis for data mining. If there is no way to collect the preference list of numerical values, we can artificially convert the order in the preference list into numerical values. Finding all frequent itemsets quickly and efficiently is the central problem of classification processing and the standard of data mining. Just define a descending function for sorting. The ranking result, as the initial value of the next level model, causes the internal structure of the next level model to adjust and update the collaborative relationship of sub-models. The dynamic structure and feedback mechanism of each level model constitute the dynamic system and iterative cycle of various constraint models, and the model transformation between levels is realized by the learning activity flow composed of a series of learning events.

3.2 Diversity as Upper and Lower Bound Constraint Analysis

As an optimization problem, we must first have quantitative optimization objectives. For each transaction, the subset function is used to find out all subsets of candidates in the transaction, and each such candidate is semantically calculated. Semantic computing combines semantic content and semantic links to establish attributes that can describe the function of knowledge elements, and produce knowledge elements with the smallest granularity. The knowledge element extraction is shown in Fig. 4.

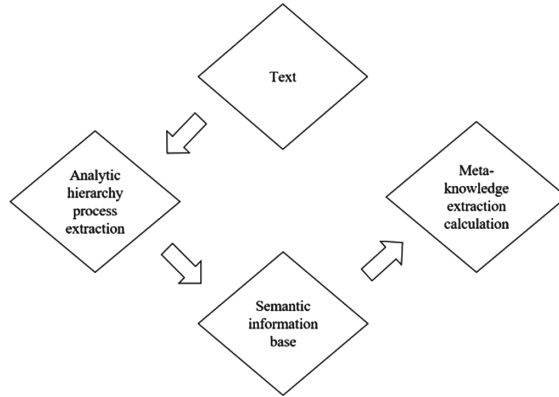


Fig. 4. Knowledge element extraction process

First, generate sets at the same time. The solvable set of linear gauge problems of unequal bundle is $R(A) = \{x: Ax \leq b, x \geq 0\}$. If A is a unitary matrix, then for any integer vector b , all vertices of $R(A)$ are integers. That is, in the constraints of different levels of particles, high-level coarse particles are transformed into low-level fine particles, and low-level fine particles form a richer understanding of the nature of things from different perspectives. The number of constraints is equal to the length of the maximum frequent itemset, so it is necessary to scan the database several times.

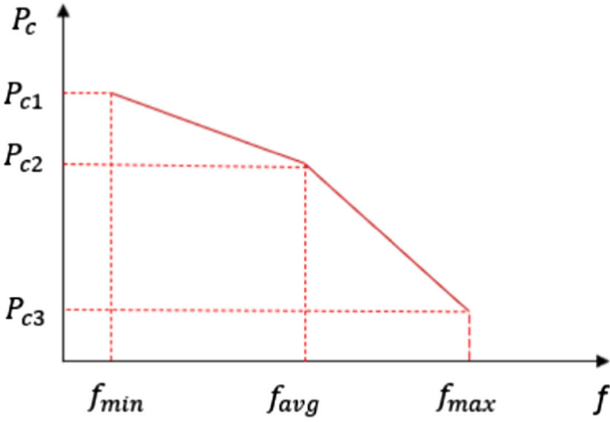
$$\max \sum_{i,j} w_{ij} x_{ij} \quad (3)$$

w_{ij} --The ij first preference scoring matrix;

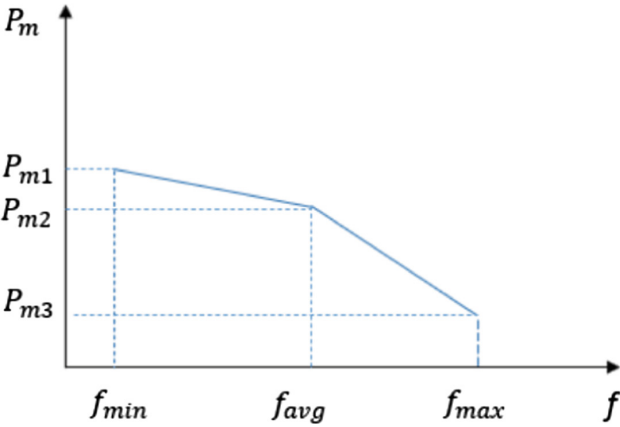
x_{ij} --Final match.

The types of knowledge points are presented by semantic annotation. The inheritance, generalization and combination of knowledge points are completed by semantic nodes. The adaptive curve is shown in Fig. 5.

For the case where both local and global constraints appear in the extracted sub matrix, we can multiply the row corresponding to the local constraint by—1 and add it to the global constraint (the elementary transformation does not change the determinant). In this way, we transform the sub matrix into a form with only local constraints without changing the determinant of the sub matrix. Local constraints need to generate candidate



(a) The trend of PC changes



(b) The changing trend of pm

Fig. 5. Adaptive curve

itemsets of frequent itemsets. When the support threshold is small, a large number of “long” association rules will appear, so the cost of the algorithm is very high, and it is easy to have the problem of combinatorial explosion.

4 Experimental Results and Analysis

The experiment of this study aims to explore the effectiveness of data mining in Chinese education based on multiple constraint models, and draw corresponding conclusions from the experimental results. The data set used in the experiment is a group of students' Chinese learning data, including students' gender, age, learning time, academic performance and other factors.

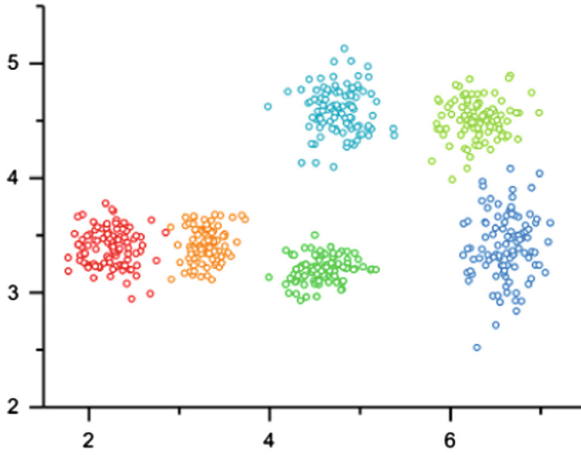
Before conducting the experiment, data preprocessing and feature extraction were first carried out. After cleaning, deduplicating, and filling in missing values on the data, multiple features such as student gender, age, study time, homework completion, and exam scores were selected as input to the model.

Then, Chinese education data mining can be carried out based on a multi constraint model. The advantage of multivariate constraint models is that they can accurately control the correlation between data and avoid interference factors in data analysis. When implementing it, we used the Python programming language and the Scikit learn machine learning library.

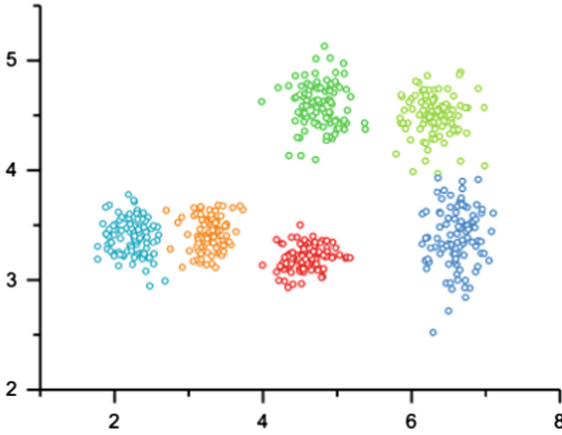
The experimental results indicate that the multi constraint model is very accurate in analyzing Chinese language education data. After using PCA algorithm for principal component analysis, we found that students' age, learning time, and homework completion were most significantly correlated with Chinese language grades, with learning time having the most significant impact on grades. By ranking the importance of different features, we can obtain an overall ranking of data feature importance, which can be used to propose corresponding educational strategies based on data mining results. The experimental scatter analysis results are shown in Fig. 6.

For example, according to the conclusion that students' learning time and homework completion have the greatest impact on their grades, strengthening homework guidance and organizing appropriate learning plans can be helpful in improving students' Chinese language grades. In addition, other features can also be sorted by importance through data mining, which can be used to guide the practical operation of Chinese language education.

In summary, the data mining of Chinese education based on multiple constraint models has achieved satisfactory experimental results. This model can accurately grasp the correlation between data, avoid interference factors in data analysis, and provide beneficial data analysis and decision support for the field of Chinese language education. In practical applications, it is necessary to further enrich the dataset and improve the algorithm model to improve the accuracy and value of data mining.



a. Set 1 original clustering partition



b. SGA clustering results

Fig. 6. Experimental scatter analysis results

5 Conclusions

That is, the “three-dimensional” curriculum goal must be qualitatively improved, the curriculum structure must match the professional quality of Chinese teachers, the curriculum content should be close to students’ life, the curriculum implementation method should highlight students’ subjectivity, and the curriculum evaluation should focus on students’ development. The universities contains a lot of useful information for students, teachers and school management departments. As explore the law of education from four aspects: student characteristic analysis, talent factor analysis, student employment prediction and student source quality analysis. The development of educational big data and

learning analysis technology promotes accurate and comprehensive portraits of learners in the digital age, and promotes the deepening of personalized support services. When dealing with diversity constraints, we first try to give a complete description of the problem in high-dimensional space. Then, the diversity is characterized by classifying the participants and controlling.

This paper proposes data mining education data mining based on multiple constraint models is the core of digital learning and educational big data analysis, and learning data elements provide strong support for the analysis of multiple constraint models. Data mine the teaching model from multiple levels and angles, to judge the data categories that meet the query conditions, and to query data according to different data categories, so as to realize the database optimization query.

References

1. Qing, C., Cheng, Z.: Data mining model of mobile terminal user preference under interval constraints. *Comput. Simul.* **37**(11), 4 (2020)
2. Tao, Z., Si, Z.: Learner model design and computational research in educational big data mining. *Electron. Educ. Res.* **41**(9), 7 (2020)
3. Peng, G.: Chinese language and literature education and humanistic quality education of college students. *New West: Late Ten Days Theory* (3), 2 (2016)
4. Qianqian, H.: Analysis of the current problems in Chinese language and literature education. *Commodities Q.* **000**(038), 280 (2017)
5. Chun'e, S.: Innovative research on chinese language and literature education. *Arts Lovers: Educ. Teach. Edn.* (2), 1 (2018)
6. Pengfei, X., Qinhuo, Z., Yaohua, C., et al.: Research on learner modeling in educational data mining. *China Distance Educ.* (6), 7 (2018)
7. Yumei, J., Hang, C.: The value positioning, realistic limits and effective strategies of educational big data mining. *Educ. Theory Practice* **41**(19), 6 (2021)