Design and Implementation of English Teaching Analysis System Based on Data Mining

Bei Yu(✉)

Guangzhou College of Technology and Business, Guangzhou 510850, Guangdong, China
yongjiegui@sina.com.cn

Abstract. Driven by the current situation of teaching management, teaching managers at colleges and universities at all levels generally hope to build a more functional teaching analysis system by using advanced information technology, so as to provide more comprehensive historical data analysis functions for teaching managers. The design and implementation of English teaching analysis system based on data mining in this paper improves the management and teaching level of the whole school in the process of English teaching, and finally improves the ability of English subject management in teaching decision-making.

Keywords: Data mining · English teaching · Analysis system

1 Introduction

Since the new century, with the substantial adjustment of the national enrollment policy, China's higher education has been in the fast lane of adjustment and development. The development level of informatization and modernization in the field of education has also entered a new stage of development. Based on the basic theory of modern management information system, various colleges and universities have made great progress and achievements in management information. The most common is all kinds of information management systems for schools, students and teachers. Among the many information-based teaching systems used at present, there are many systems that can provide services for the daily teaching of the school. Therefore, how to realize the unified utilization of daily accumulated data, analyze and mine some laws in the data, and provide more real and specific basis and analysis results for decision-making and management in school teaching has become a problem and difficulty to be solved [1]. In this paper, the design and implementation of English teaching analysis system based on data mining is to improve the management and teaching level of the whole school in the process of English teaching, and finally improve the ability of English subject management in teaching decision-making.

With the development of globalization and the deepening of communication among countries, multilingual learning has been paid more and more attention, and it is very important to master a foreign language. Among them, English, as a widely used foreign language, has been included in the basic courses of education in Colleges and universities

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2022, LNICST 465, pp. 10–18, 2023.
https://doi.org/10.1007/978-3-031-23950-2_2

in China. In the process of English Teaching in Colleges and universities, due to the large number of students, weak interaction, single teaching mode and language environment, the imbalance between the written and oral education in the process of English education has been caused. Students often have high written test scores, but lack the oral ability to communicate in English, It is very important to strengthen the interactive process and oral learning in the learning process, and to explain words and sentences in an interesting teaching form. However, most of the current English teaching systems use simple audio-visual multimedia to assist teaching, which can not provide targeted learning programs for students' oral ability and classroom performance, and the interactivity is not strong.

To sum up, it is an urgent problem for technical personnel in the field to provide an English teaching system to assist teachers in multi-mode English teaching, which is highly interactive, can effectively improve teaching quality, accurately analyze students' oral ability, and assist teachers in targeted teaching.

2 Data Analysis System

2.1 Data Analysis System Concept

Nowadays, the widely used data analysis system, supported by a variety of modern disciplines and basic theories, provides intelligent support and assistance for various types of management processes by using comprehensive technologies such as computer, informatization and virtual simulation, aiming at solving semi-structured decision-making problems. A well-designed data analysis system can provide managers with more detailed supporting data, information and relevant background materials in use, and realize the identification of relevant problems through further clarifying the decision-making objectives; Then, the data analysis model can be modified and adjusted to provide alternative alternative schemes for decision-makers, evaluate and optimize different schemes, provide managers with all-round analysis, comparison and comprehensive judgment based on real-time human-computer interaction technology, and finally provide users with correct decision-making services and support.

Aiming at the problems and needs mentioned above, this scheme proposes an English teaching system for English learning, which can solve the above technical problems due to the following technical solutions. In order to achieve the above purpose, the invention provides the following technical scheme: an English teaching system for English learning, including: a signal acquisition and statistics module, an intelligent analysis module, a management module, a scenario simulation module, a power supply module and a wearable mobile terminal; The signal acquisition and statistics module includes a signal acquisition unit and a statistics unit. The signal acquisition unit is used to collect the start time and end time of the key signal and voice signal of the intelligent learning terminal. The signal acquisition unit sends the key signal and the collected time information to the statistics unit, The statistical unit is used to receive the key signal sent by the signal acquisition unit and the collected time information, and classify and count the signals and information; The intelligent analysis module is connected with the signal acquisition and statistics module. The intelligent analysis module is used to receive the data output by the statistics unit, perform intelligent analysis on the data, and generate the incentive learning scheme with the highest matching degree according to the incentive scheme in

the teaching database; The management module is used to receive intelligent analysis data and store it in the teaching database, evaluate teaching assessment scores, and manage intelligent learning terminals and wearable mobile terminal devices; The scenario simulation module is connected with the management module. The scenario simulation module is used to generate simulated dialogue roles and virtual scenes according to the input teaching content and the collected scenario simulation indoor object location information, indoor space information and student information.

The wearable mobile terminal is used to collect students' identity information and perform login verification. The wearable mobile terminal is wirelessly connected with the scenario simulation module to assist students in oral English scenario simulation learning. Further, the signal acquisition unit generates a signal object and a voice acquisition object according to the current teaching mode selected by the user, the signal object is used to express the key attribute and the key signal source number, the voice acquisition object is used to express the attribute and the voice signal source number of the voice signal, and sorts the key signals corresponding to the signal object according to the timestamp order, And recording the start time and end time of the voice signal corresponding to the voice acquisition object, and sending the sorting data and the recorded time data to the statistical unit for classification statistics. The specific steps of the classification statistics are as follows:

Step S1: monitor the statistical signal. When the statistical signal is monitored, judge whether the current statistical signal is used for key signal statistics. When it is used for key signal statistics, the key signals with different attributes and different key signal source numbers are counted according to the order of key signals to form the first attribute sorting list;

Step S2: otherwise, the start time and end time of voice signals with different attributes are statistically classified according to the voice signal source number, and the voice signal source number with voice overlapping time period greater than 30% is marked to form a second attribute sorting list;

Step S3: send the acquired first attribute sorting list and the second attribute sorting list to the intelligent analysis module for learning interest analysis, and display the classification results on the display in a visual form to assist teachers in teaching.

2.2 Development Analysis of Data Analysis System

The modern data analysis system, which has been widely used in various fields, originates from the emergence and development of management information system. Here, based on the database platform, the management information system can provide services for all kinds of managers, and deal with all kinds of events in the real environment with the support of modern information and network technology; Accordingly, the data analysis system can provide more efficient decision-making assistance for managers and users at all levels according to various data stored in the system [2]. Since the 1970s, artificial intelligence and computer technology have made great progress, making the auxiliary management of various strategies and development plans based on such technology widely used. In the process of auxiliary decision-making, there are many problems to be considered and the factors involved are more complex, so it is necessary to consider

the comprehensive results of various schemes, so as to It also promotes the emergence and rapid development of data analysis system.

In fact, the concept of data analysis system originated in the 1980s and has been developed for more than 30 years. As early as the 1980s, some scholars proposed a data analysis system based on three-tier structure, and the three-tier structure of this system corresponds to the dialogue components, data components and model components of the whole system respectively. Among them, the data component is the main component The data analysis system based on three-tier structure further defines the basic composition of the data analysis system, and gives the related technologies that can play an important role in the data analysis system, such as common model base management technology and complex components It is precisely because of the emergence of the three component structure that further promotes the considerable development and progress of the data analysis system supporting decision-making. After that, some scholars put forward the data analysis system based on the three system structure, namely language system, problem processing system and knowledge system. The data analysis system with this structure is more widely used It is a problem processing system, or knowledge system, which is different from the expert system based on artificial intelligence.

Driven by the increasing complexity of the problems that the data analysis system needs to deal with, the number of various models used to provide services is also increasing, which makes it more difficult to realize the whole between different service models; at the same time, due to the continuous improvement of researchers' understanding of information processing concepts and development laws in the field, in the context of more diverse changes in reality Driven by the environment, data analysis systems with higher levels and functions are needed to provide services, which further stimulates people's enthusiasm for the research of data analysis systems. Especially since this century, the scale of databases and the amount of data stored in various information systems have also shown an increasing trend; based on these huge amounts of data, all kinds of personnel responsible for management We hope to find and find deeper information and knowledge, and then serve the decision-making process.

3 Data Mining Technology

3.1 Data Mining Concept

Data mining refers to the process of extracting hidden, unknown but potentially useful information and knowledge from a large number of incomplete, noisy, fuzzy and random actual data. Data mining is also called knowledge discovery.

Data mining has become a useful and practical tool to discover the potential relationship between patterns and the amount of existing data. In addition, the combination of machine learning, statistical analysis, modeling technology and database technology, data pattern discovery, data relations and derivation rules used in data mining can help predict future trends, conditions and behaviors, and enable enterprises to make aggressive and knowledge driven decisions. Typical enterprises apply data mining technology, including market analysis, customer analysis, fraud detection, evaluation of retail promotion and credit risk analysis. A complete data mining tool set not only supports multiple

methods of pattern discovery and relationship, but also provides tools to evaluate the statistical quality and confidence [3].

Further, the specific steps of intelligent analysis include:

Step S1: record the collected key signal of the same student as R, the start time of the voice signal as T1, and the completion time of the voice signal as T2, then the duration of the voice signal is TC = T1–T2;

Step S2: record the intelligent learning terminal number of the input voice signal of the same student, the number of times N1 of the voice module used in the intelligent learning terminal and the number of hours T1 of the voice module used, and the interval of the use hours C1, and record the number of the intelligent learning terminal of the input key signal R of the same student, the number of times N2 of the key module used in the intelligent learning terminal and the number of hours T2 of the key module used, The interval of using class hours is recorded as C2. The interval of class hours C1 and C2 respectively represent the interval of using key module and voice module by the same student;

Step S3: calculate the duration TB of the historical average speech signal of the same student, the average values CB1 and CB2 of the class interval, and assign the preset weight values W1, W2, W3, W4 and W5 for the number of class hours T1, the number of class hours T2, the duration TB of the historical average speech signal, the average value CB1 of the class interval and the average value CB2 of the class interval, W1 = W2 "W4 = W5" W3;

Step S4: according to the formula P = (t1*w1 + t2*w2 + tb*w3 + cb1*w4 + cb2*w5) /5, a matching value p is generated, and the corresponding learning interest level is obtained according to the generated matching value P. from the learning interest level, the highest matching encouraged learning scheme is obtained from the encouraged learning scheme in the teaching database.

Further, the management module comprises a teaching management unit for managing teaching information, student information and scoring classroom assessment, and a device management unit for managing intelligent learning terminals and wearable mobile terminals;

The teaching management unit includes a mode selection module, a teaching resource management module, a teacher information management module and an assessment scoring module. The mode selection module is used for teachers to select ordinary teaching modes and scenario simulation teaching modes. The teaching resource management module is used for teachers to upload independently edited English teaching resources, English teaching content, Download ar data templates and the students' classroom learning data, The teacher information management module is used to verify the identity of the teaching teacher and record the teaching content introduction, teacher name and professional title. Mode selection and equipment control can be carried out only after the authentication is successful. The assessment scoring module is used to obtain the matching value p of each student according to the preset scoring division range, and give the assessment score according to the scoring division range of the matching value p;

3.2 Data Mining Classification

The knowledge that can be found by data mining is as follows:

(1) Generalized knowledge refers to the general description knowledge of category characteristics. Finding universal, meso and macro knowledge according to the micro nature of data, reflecting the common nature of similar things, is the generalization, refinement and abstraction of data.
(2) Classified knowledge is the characteristic knowledge that reflects the common nature of similar things and the knowledge of attribute differences between different things. For example, decision tree algorithm is a commonly used classification method.
(3) Clustering knowledge, clustering is to achieve "birds of a feather flock together", not the same kind. The whole database is divided into different classes. The data between the same class are as similar as possible, and the differences between classes should be obvious.
(4) Correlation knowledge reflects the dependence or correlation between one event and other events. The information of another data object can be inferred from the information of one data object.
(5) Predictive knowledge, which infers future data based on historical and current data, can be considered as time series data.
(6) Deviant knowledge, the description of differences and extreme special cases, reveals the abnormal phenomenon of things deviating from the Convention.

3.3 Data Mining Steps

The whole data mining process is composed of several steps. The main steps are:

(1) Data cleaning: its function is to remove data noise and eliminate incomplete and inconsistent data.
(2) Data integration: its function is to combine data from multiple data sources and store them in a consistent data store.
(3) Data conversion: its function is to convert data into a data storage form suitable for data mining.
(4) Data mining: select the appropriate mining algorithm to obtain the corresponding rules. Its function is to use mining algorithm technology to mine relevant knowledge laws.
(5) Result analysis: the function of interpreting and evaluating the results is to screen out effective potential knowledge from the mining results according to certain evaluation criteria.
(6) Knowledge representation: its function is to integrate the knowledge obtained from analysis into the organizational structure of information system by using visualization and knowledge expression.

The specific process of data mining is shown in Fig. 1.

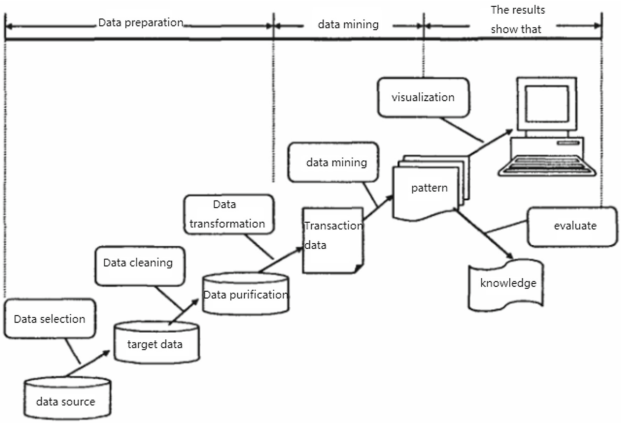

Fig. 1. Data mining process

4 Design and Implementation of English Teaching Analysis System Based on Data Mining

The existing college English teaching data analysis system is based on the database accumulated in the existing various information management systems of the school, and makes an in-depth analysis of the contents of the school in each link of the English teaching process with the help of data warehouse, online analysis, data mining and other technologies widely used in the field of information processing. For various processing processes, the one that focuses on the operation process can be called transaction processing process, which is also a common operation in the online operation and use of the database. It mainly queries and modifies the records in the database, and more serves the specific application of the school in the implementation process of English teaching. The focus is mainly on the response time of the operation Data security and integrity of processed data. It is because there are different operation modes in form and content that the demand for data in the system is also different. Specifically, the data analysis system used to assist decision-making realizes the analysis and query process based on data [4]. The system stores a large amount of massive data in different periods. Due to the lack of analysis process supported by historical data, it can not get accurate results. At the same time, considering that the composition of various historical data in the system is the integration of information data in various business systems before the school, it is difficult to accurately estimate the time consumed by the data analysis system in query and analysis, which usually requires continuous operation for multiple hours and consumes more system resources. Different from the data analysis system, in the operation system, the data is mainly the process data used in the system operation. Each operation in the business operation process is recorded. The operated data is also mainly the dynamic update data, and the relevant data needs to be processed in real time. Therefore, it is precisely because of this fundamental difference in data requirements that there are fundamental differences in the process of data management. Therefore, in order to effectively solve this difficult contradiction, it is necessary to divide the two data

management systems with different properties, and construct their corresponding data application environment on the basis of the division, and then provide support for different levels of applications. As shown in Fig. 2, it is the specific application architecture of such solutions in the field of education.

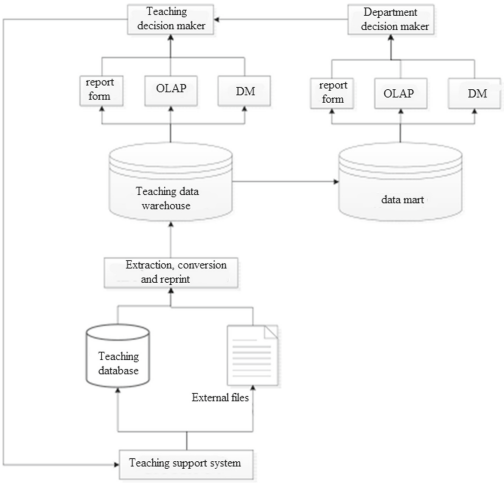

Fig. 2. Structure of English teaching analysis system

Using the system architecture shown in Fig. 2, external data can be introduced into the analysis environment, and through the extraction and conversion of data in the database, the errors and formats in the data can be corrected and unified to ensure that the data can meet the data format and quality standards required by the data warehouse. Then, the data can be stored in the data warehouse. In use, for different functional institutions and departments of the school, we can make full use of the data mart to realize daily management and decision-making.

5 Conclusion

Based on the comprehensive summary and analysis of the research background and current situation of relevant topics, this paper makes a targeted research on the data analysis system, and constructs a teaching data operation center and data analysis system with relatively perfect functions, so as to provide a more comprehensive historical data analysis function for English teaching managers. In this way, the school's English teaching managers can timely understand the information at all levels, so that the historical data stored in the database can play a full role; Through the multi-dimensional analysis and mining of multi-source data, some potential conclusions and laws are obtained, which can assist the implementation process of English teaching, improve the English teaching level of the whole school, and finally improve the ability of discipline management in teaching management and decision-making.

Acknowledgement. Guangxi Higher Education Undergraduate Teaching Reform Project in 2020, Research and Practice on flipped Classroom teaching of Business English Listening and Speaking course based on mobile platform,serial number 2020JGB456.

References

1. Maier, C., Laumer, S., Wirth, J., Weitzel, T.: Technostress and the hierarchical levels of personality: a two-wave study with multiple data samples. Eur. J. Inf. Syst. **28**(5), 496–522 (2019)
2. Zhang, H., Liang, Y., Su, H., Liu, C.: Event-driven guaranteed cost control design for nonlinear systems with actuator faults via reinforcement learning algorithm. IEEE Trans. Syst. Man Cybern. Syst. **50**(11), 4135–4150 (2020)
3. Shen, B., Wang, Z., Wang, D., Li, Q.: State-saturated recursive filter design for stochastic time-varying nonlinear complex networks under deception attacks. IEEE Trans. Neural Netw. Learn. Syst. **31**(10), 3788–3800 (2020)
4. Yao, D., Li, H., Lu, R., Shi, Y.: Distributed sliding-mode tracking control of second-order nonlinear m systems: an event-triggered approach. IEEE Trans. Cybern. **50**(9), 3892–3902 (2020)