Research and Design of Information Teaching Management System Integrating Association Rule Mining Algorithm

Chengli Zhao(✉)

Ya'an Polytechnic College, Ya'an City 625000, Sichuan, China
zhaochengli1984@163.com

Abstract. The main purpose of this research is to design and implement an information teaching management system integrating association rule mining algorithm. The process is divided into two parts: the first part is about the design model, and the second part is about implementing it. Research and design of information teaching management system integrating association rule mining algorithm In this research, we are trying to develop an information teaching management system, which can be applied to college students studying in the field of computer science. The purpose is to provide teachers with a tool that can help them work. We also hope to use it as a learning tool for students so that they can acquire more knowledge about how computers work and what they need to learn when using computers. Combine the research and design of information teaching management system with association rule mining algorithm. This algorithm can be used to find the most important rules in a given dataset, which are related to each other. It can then be used to select the best rule from the rules found using this method. It will help teachers choose the best teaching methods or skills to improve their skills and knowledge.

Keywords: Teaching management · Association rules · Mining algorithm · Data mining

1 Introduction

The term "knowledge discovery from databases" (KDD) first appeared at the 11th International Joint Academic Conference on artificial intelligence held in 1989. After that, the research focus gradually shifted from discovery methods to system applications, focusing on the integration of multiple discovery strategies and technologies, and the mutual penetration of multiple disciplines. Among the database research projects of the National Science Foundation (NSF), KDD was listed as the most valuable research project in the 1990s. Scientists in the field of artificial intelligence research also generally believe that one of the important topics of the next artificial intelligence application will be large-scale database knowledge discovery with machine learning algorithm as the main tool [1]. Although data mining is still a very new research topic, its inherent

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2022, LNICST 465, pp. 543–553, 2023.
https://doi.org/10.1007/978-3-031-23950-2_59

potential to create huge economic benefits for enterprises has made it quickly have many successful applications. The representative application fields include market prediction, investment, manufacturing, banking, communications and so on [2].

American Iron and steel company and Kobe iron and steel company use ISPA system based on data mining technology to study and analyze product performance laws and carry out quality control, and have achieved remarkable results. General Electric Company (GE) and French aircraft engine manufacturing company (SNECMA) developed Cassiop by using data mining technology EE quality control system has been used by three European airlines to diagnose and predict the fault of crossover 737, which has brought considerable economic benefits. In 1996, the system won the European first-class creative application award [3].

Recently, a senior technology survey of Gartner Group listed data mining and artificial intelligence as the top of the "five key technologies that will have a profound impact on industry in the next three to five years", and also listed parallel processing system and data mining as the top two of the ten emerging technologies that will be the focus of investment in the next five years. According to the recent Gartner HPC research, "with the rapid development of data capture, transmission and storage technologies, users of large-scale systems will need to adopt new technologies to tap the value outside the market, and adopt broader parallel processing systems to create new business growth points." At present, several hotspots in the future include web site data mining, bioinformatics / genomics data mining and text mining.

At present, the key point in the management of colleges and universities is teaching management. Effective teaching management not only requires managers to set a good management system, but also should have an ideal information management system. Because the management methods and teaching modes of colleges and universities are different, the teaching management systems applied by colleges and universities are also different. Data information such as teacher information, student information, course information and achievement information is cumbersome for teaching managers, However, if the teaching management system combines such information data with different management directions according to their respective management characteristics, it can effectively reduce the pressure of managers and give full play to the effect of the teaching management system to the greatest extent [4]. As a new way of teaching management, information-based teaching management has been gradually introduced into the university management center, which can fundamentally simplify the work content of managers. Therefore, the design of information-based teaching management system has become the top priority of university management.

Data mining refers to extracting interested knowledge from the database. Association rule mining is the research focus in the field of data mining. Association rule refers to the rule of some relationship between a group of objects in the database. Its mining object is usually transaction database. Due to the continuous accumulation of a large number of historical data in the teaching management of colleges and universities, including a lot of valuable information, the application of association rule data mining in the information-based teaching management of colleges and universities can analyze and mine the useful information in the teaching management, and provide help for the teaching managers the foundation for the scientific and comprehensive management of college teaching.

2 Fusion Association Rule Mining Algorithm

2.1 Overview of Data Mining

Data mining, It was once called “fishing”, “snooping” and “fishing”. A considerable number of people regard data mining as another commonly used term, knowledge discovery in database or KDD (knowledge discovery in database). In short, data mining It is a process of extracting potentially useful information and knowledge hidden in a large number of incomplete, noisy, fuzzy and random practical application data that people do not know in advance. The object of data mining must be real, massive and noisy; What is discovered is the knowledge that the user is interested in; The discovered knowledge should be acceptable, understandable and applicable; Data mining technology involves database, artificial intelligence, neural network, prediction theory, machine learning and statistics [5].

With the maturity of database technology and the popularization of data application, the amount of data accumulated by human beings is growing rapidly at an exponential rate. In the 1990s, with the emergence and development of the Internet, the emergence and application of Intranet, extranet and VPN (virtual private network), the whole world is connected into a small global village. People can exchange data information and work together on the Internet across time and space. In this way, what is displayed in front of people is not limited to the huge database of the Department, the unit and the industry, but the vast ocean of information, and the data flood is coming to people [6]. When the amount of data grows extremely, if there is no effective method to extract useful information and knowledge with computer and information technology, people will feel helpless in the face of the information ocean like looking for a needle in a haystack. It is estimated that only 7% of the data in a large enterprise database is well used. In this way, compared with “data surplus” and “information explosion”, people feel “information poverty” and “data locked in a cage”, just as John naisbett exclaimed “well dropping in information, but standing for knowledge”.

Faced with the boundless data, people call for a technology and tool to remove the essence, eliminate the false and preserve the true from the vast ocean of data, that is, KDD. Knowledge discovery in databases. In this way, data mining (DM) is born. The general process of data mining technology is shown in Fig. 1.

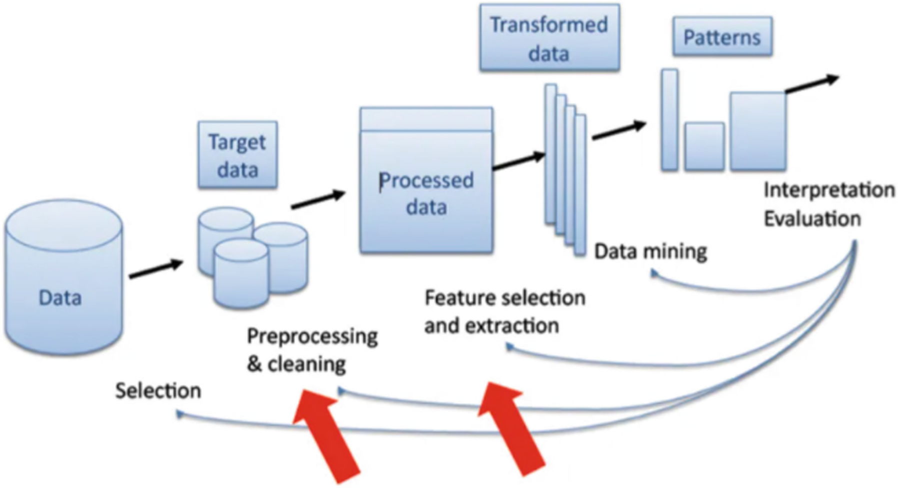

Fig. 1. General process of data mining technology

2.2 Association Rule Mining in Data Mining

Integrated association rules are rules involving multiple attributes or predicates. The mining of integrated association rules does not search for frequent item sets like the mining of single dimension association rules. In the mining of integrated association rules, we search for frequent predicate sets. The k-predicate set is a set containing K conjunctive predicates. Integrated association rule mining methods can be divided into the following three types according to their processing of quantitative attributes:

(1) Integrating association rules with static discretization mining based on quantitative attributes
i.e. discretizing the quantization attribute using a predefined concept hierarchy. This discretization is carried out before mining. The value of the numerical attribute is replaced by an interval, such as "O~20", "21~30", "31~40", etc., to replace the original value of the attribute. If the result data related to the task is stored in the relational table, the Apriori algorithm can find all the frequent predicates rather than the frequent itemsets (that is, by searching all the related attributes instead of only one attribute) with a slight modification [7]. Finding all frequent k-predicate sets will require K or K + 1 table scans. Other strategies, such as hashing, partitioning, and sampling, can be used to improve performance, as shown in Fig. 2.

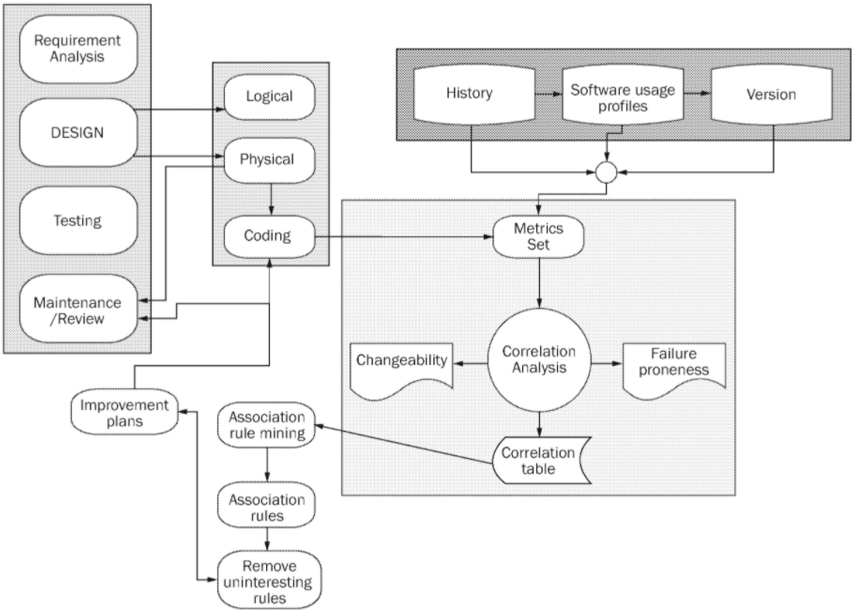

Fig. 2. Integrated association rule mining algorithm flow

(2) Mining quantitative association rules
Quantitative association rules are integrated association rules, in which numerical attributes are dynamically discretized to meet certain mining standards. This method discretizes the quantitative attributes into "boxes" according to the distribution of data. These boxes may be further combined during excavation. For example, the method used by the system arcs (association rule clustering system) is to map the quantitative attribute pairs to the 2-D grid that meets the given classification attribute conditions, and then search for the clustering of grid points to generate association rules.

(3) Mining distance based association rules
This method is the discretization of quantitative attributes to closely follow the semantics of interval data, and does not allow the approximation of data values. This process takes into account the distance between data points, so it is called distance based association rules. The two pass algorithm can be used to mine such association rules. The first time use clustering to find intervals or clusters. The second pass searches for frequently occurring cluster groups to obtain distance based association rules.

3 Information Teaching Management System Integrating Association Rule Mining Algorithm

3.1 Overall System Structure Design

The information-based teaching management system is composed of three key parts: client, server and database. Its structure is shown in Fig. 2. Among them, providing users with presentation and interaction functions is the core function of the client, and completing the logical operation of the application is realized through the core function of the server. The client applies the C / S structure mode in this system, focusing on the smart phone and mobile terminal commonly used at present; The server can apply B / s and C / S structure modes in the system, and the communication protocol can be compatible with HTTP and socket. According to the analysis of Fig. 3, the information-based teaching management system is developed on the basis of the layering principle. The system database uses the relational online processing and analysis (ROLAP) method to store all data [8]. After effectively mining the data by using the fusion association rule mining algorithm, various functions of the system information management are presented through the main interface of the client, such as score management, course management, course selection management, etc., The security of the system and data information is improved by means of firewall in the system network.

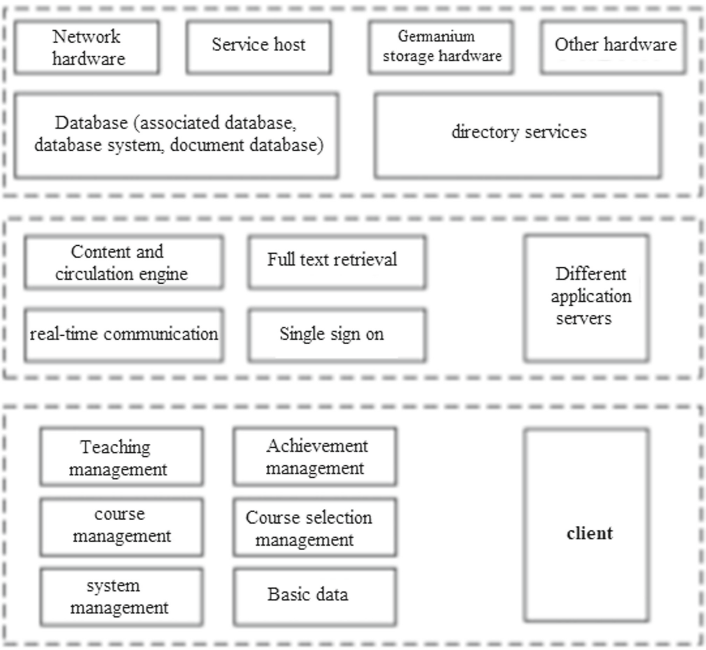

Fig. 3. Overall system structure diagram

3.2 Application of Association Rule Mining Algorithm

The data mining process determines the data type to be mined. The data in the database is transformed and cleaned up. The mining algorithm of association rules is used to destroy the data according to the correct data type. After mining, the mining results are analyzed and evaluated and displayed to users through customers. The detailed data mining process of the system is shown in Fig. 4.

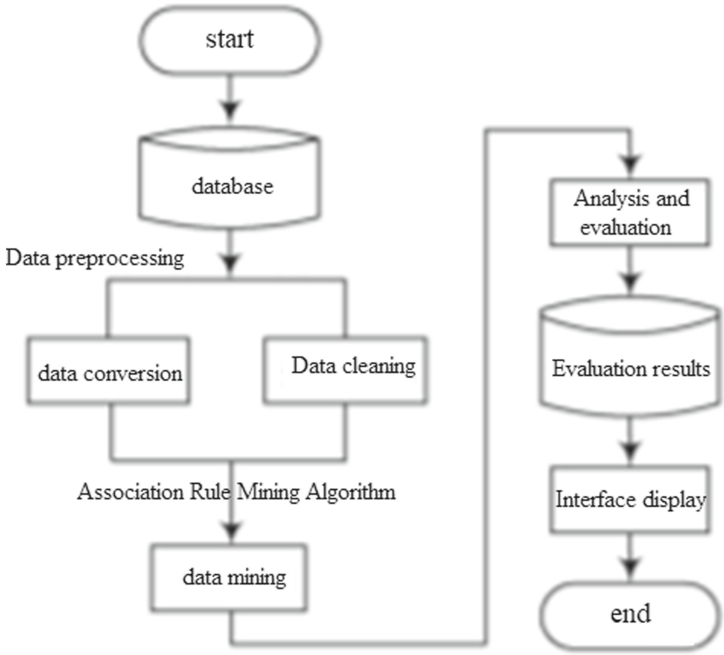

Fig. 4. System data mining process

(2) Apriori algorithm core.

In order to generate all frequent sets, Apriori algorithm uses recursive method, as follows:

Ki = (large I-itemsets}
For(1 = 2;K1-1, μ;1++) do begin
Ar = apriori-gen(K.-)
For all transactions ijB do begin
A, = subset(A,t)
For all candidates cj A, do
c.count++
End
K={cjA,| c.count minsup)
End
Reply= I ∪ Kr

The first frequent itemset K1 and the second frequent itemset K2 are generated successively until a U value makes K null, and the algorithm terminates. In the 1st cycle, the set al of candidate 1 item set is generated. The function of each set in Al is to generate the candidate set of frequency set. As shown in Fig. 5 below.

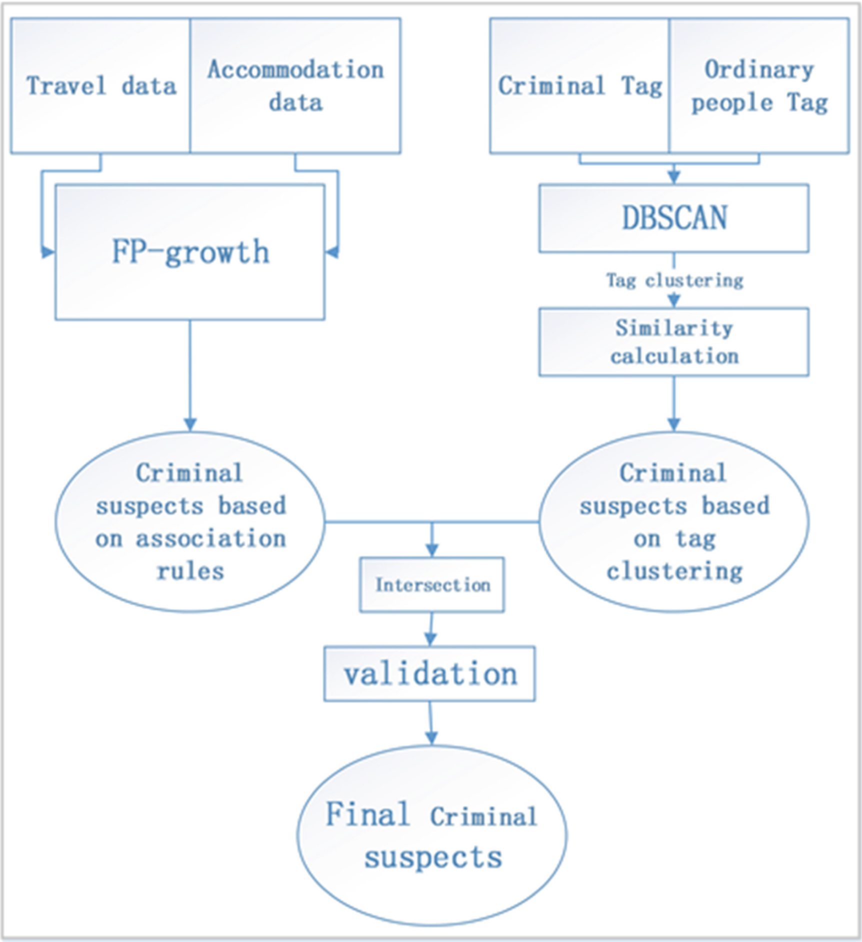

Fig. 5. Generate candidate sets of frequency sets

The generation of each set is completed by (I-2) connecting the KR-1 frequency set with only one different item, and the finally generated frequency set K must be a subset of A. Each element in a needs to be verified in the database to determine whether each element can add people to K. during verification, the database needs to be scanned several times. If there are at most 12 items in the frequency set, the database should be scanned 12 times, which requires a large I/0 load, and a large number of candidate sets will be produced several times. Therefore, the database still needs to be scanned many times during verification [9]. Therefore, the disadvantage of Apriori algorithm is that the amount of candidate set generation is too large and the scanning time is too long due to repeated scanning of the database, which is the bottleneck of Apriori algorithm.

4 Information Teaching Management System Based on Integrated Association Rule Mining Algorithm

4.1 Overall Objectives and Project Requirements of the System

The overall goal of the teaching management system based on DM is to establish a unified teaching management decision support system based on data warehouse, with the database of the teaching management system and other business systems as the data source, supplemented by the external data related to the business such as the student source database and the examination database, so as to realize the comprehensive and centralized management of the business data; Take the teaching management system as the platform, provide management reports, information query and data analysis services for personnel at all levels, and adopt or develop specific tools to provide information decision support for leaders of various departments and school administrators [10]. The teaching management decision support system based on data warehouse shall meet the following requirements:

(1) It can realize automatic data collection and centralized management
It can automatically collect, store, apply and manage raw data (data source) according to conditions. The design of the system is based on the B / S structure of the web. Users should be able to operate according to their different permissions to ensure the security of the whole system.
(2) Friendly human-machine interface for data input
The system can provide a flexible and convenient data entry method, and the new data can be input into the management analysis database through a friendly human-machine interface. Some data are entered manually.
(3) Flexible dynamic report function
Reports are indispensable auxiliary tools in financial business activities. The system should be able to generate reports flexibly and quickly according to user requirements and report requirements, so as to improve the work efficiency of business personnel.
(4) Decision analysis function
The decision analysis function is divided into three types of methods. In addition to data mining analysis and sensitivity analysis, it also includes graphical comparative analysis. Through the processing and integration of the original business report data, the deep and detailed business analysis is carried out on student information, teacher information, examination results, course selection information, etc.
(5) Scalability

The amount of data in teaching management is very large. In the process of construction, business expansion and analysis elements should be considered. Therefore, maintaining the relative independence and redundancy of the database is very beneficial to the expansion of the data warehouse in the future.

4.2 System Functional Architecture

The data mining system based on teaching management system is composed of five parts: basic system network, data source, warehouse management, data warehouse and data presentation. They interact with each other to form a teaching management decision support system, as shown in Fig. 6.

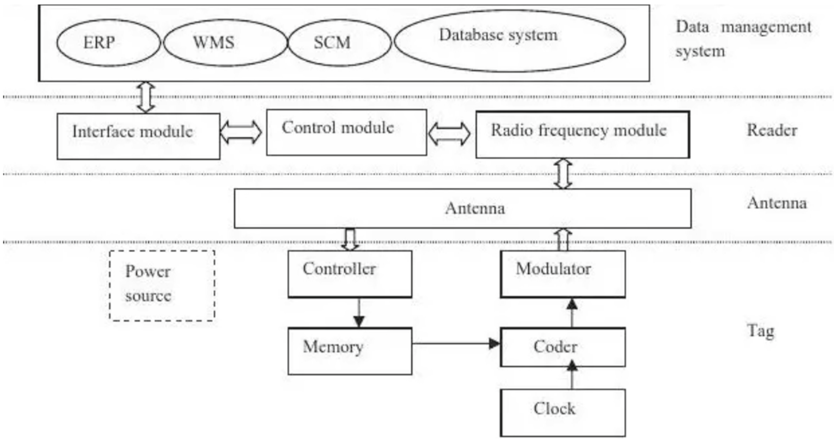

Fig. 6. System functional architecture

The lowest layer is the data acquisition layer, including a series of software and hardware data. Data sources include databases and non-traditional data. Due to the heterogeneity and autonomy of the underlying data, in order to maintain the data consistency and the data quality of data mining, it is necessary to clean up, extract and transform the data in the data source, generate a unified data and store it in the data warehouse.

The data storage layer is mainly used to store and manage the processed comprehensive data oriented to the decision theme, and reorganize it according to the requirements of the decision theme to provide a large amount of data basis for decision support. In order to mine the data of data warehouse, the data warehouse needs to provide complex data transformation functions.

The data processing layer includes data mining, OLAP and its corresponding management system.

The data access layer provides the decision maker with an access to interact with the system. After the decision-maker sends a decision request, the data mining tool mines the data related to the task from the data warehouse and stores it in the knowledge base. The online analysis tool further analyzes and processes these data, and generates data that is helpful to the decision-maker in combination with the knowledge base. The human-computer interaction system presents the results to the decision-maker, and the decision-maker solves the problem accordingly.

5 Conclusion

The research and design of integrated association rule mining algorithm in information teaching management system is a research problem, that is, in order to store information about teaching courses, students' learning ability and achievements, we need to use data mining technology. This is a new way for us to collect data from different sources, such as student records, teacher records and other related documents. It can also be used as a tool to analyze students' learning ability using statistical methods (such as association rules). In this paper, we will introduce a system design integration method based on association rule mining algorithm. The main purpose of this study is to find out how to use this algorithm in the process of information management, and what are its advantages and disadvantages. In order to answer these questions, we used a university database containing more than 20000 records to conduct experiments. For each record, we examined two types: people not affected by any rules and people affected by rules. We also compared them with other databases that do not use rules but only have data preprocessing methods (data cleaning).

References

1. Li, M., Li, J., Chong, R.: Research and application of flipped classroom in the information teaching module of interior design course. In: ICIMTECH 21: The Sixth International Conference on Information Management and Technology (2021)
2. Li, Y., Chen, L., Yu, D., et al.: Research and developing of evaluation information system using B/S structure and SQL server technology. J. Phys. Conf. Ser. **1952**(4), 042088 (2021). (6pp)
3. Zhang, X.: Research and Design of Intelligent Detection System in Teaching Manner Training of Normal School Students (2021)
4. Li, B.: Design and research of computer-aided english teaching methods. Int. J. Humanoid Robot. (2022)
5. Beidelschies, M., Cella, D., Katzan, I., D'Adamo, C.R.: Patient-reported outcomes and the patient-reported outcome measurement information system of functional medicine care and research. Phys. Med. Rehabil. Clin. North Am. **33**(3), 679–697 (2022)
6. Jing, R., Wang, H., Gao, G.: Research on integrated management of sales and inventory information in circulation enterprises based on case-based learning. Int. J. Inf. Technol. Manage. **21**, 153–168 (2022)
7. Park, H., Kang, S.H., Lee, Y.S., et al.: global trends of regional health information systems and suggested strategic utilization of their medical information. Healthcare Inform. Res. **27**(3), 175–181 (2021)
8. Lin, J.: Design and research of inverted classroom teaching of information technology integrating into the major of finance and economics. In: IPEC 2021: 2021 2nd Asia-Pacific Conference on Image Processing, Electronics and Computers (2021)
9. Jin, Z., Zou, W.: Research on the design of online teaching system of basketball basic technology. J. Phys. Conf. Ser. **1992**(3), 032080 (2021)
10. Yan, Y., Zhou, F., Zhang, X., et al.: Research and Design of ACM-oriented C Language Experimental Teaching (2021)