Research on Voluntary Intelligent Reporting System of College Entrance Examination Based on Big Data Technology

Shu-xin Guo[1] and Li Lin[2(✉)]

[1] Jilin University of Finance and Economics, Changchun 130117, China
Cqjtzyxyl23@163.com
[2] School of Computer Engineering, Jimei University, Xiamen 361021, China
xd220210@163.com

Abstract. The college entrance examination application is a complex system project, which needs to collect many kinds of information. Aiming at the deficiency of the system research based on the analysis of the domestic mainstream platform, the reference system and Sina simulation system of college entrance examination application, the intelligent application system of college entrance examination application is designed based on big data technology. Considering the scores of examinees, the enrollment plan of colleges and universities, the enthusiasm of application, the prospect of professional development and other factors, the hardware structure of the intelligent filling system for college entrance examination is constructed. Through big data analysis and data mining, a large amount of real and valuable information for college entrance examination filling can be provided for the majority of examinees. It can be seen from the experimental verification results that the system fills in accurate results and has an ideal filling effect, which helps the candidates to apply for the ideal school and improve the admission rate.

Keywords: Big data technology · College entrance examination · Intelligent filling · Data mining

1 Introduction

College entrance examination application is a complex system engineering, which needs to collect many kinds of information, and comprehensively consider such factors as examinee score, college enrollment plan, enrollment enthusiasm, professional development prospect, examinee's personal interest and family situation [1]. Today, when the mobile Internet is highly developed, the Internet is full of various types of information about colleges, universities, majors, admission scores, etc. How to identify true and valuable information in the massive application information to suppress the troubles for the majority of candidates and parents [2]. Due to the lack of information and improper choice of information in college entrance examination, it is common for examinees to fail in high scores and get low marks. It is difficult for examinees and parents to evaluate accurately, and the phenomenon of high score falling out of the list

S. Liu and L. Xia (Eds.): ADHIP 2020, LNICST 347, pp. 98–111, 2021.
https://doi.org/10.1007/978-3-030-67871-5_10

and low score occurs every day [3]. Therefore, examinees and parents are in urgent need of comprehensive guidance on the application of college entrance examination.

The college entrance examination voluntary reference system based on college entrance examination scores or rankings is generally based on the candidates' own scores or rankings, combined with certain directly related information such as the admission scores of colleges and universities in a certain year or years. Candidates recommend some schools or majors [4]. At present, there are many such systems in China, such as the "sunshine college entrance examination" information platform of the Ministry of education, which is developed by the national college student information consultation and employment guidance center of the Ministry of education, the comprehensive reference system of college entrance examination filling and submitting, the Sina college entrance examination simulation filling and submitting system launched by sina.com, the reference system of college entrance examination filling and submitting launched by China Education online, etc. [5]. Some of these systems also provide historical information such as admission scores, enrollment numbers, and employment status in previous years, but they are usually limited to simple data query and statistics. Without in-depth data analysis function, it is difficult to find knowledge, and the laws hidden behind the data can not fundamentally solve the blindness of candidates when filling in the application form.

Based on the analysis of the deficiencies of the mainstream college entrance examination volunteer filling platform in China, this paper designs an intelligent college entrance examination volunteer filling system based on big data technology. The hardware structure of the system includes on-line analysis and processing server, data warehouse server and enrollment information data mining system. On the basis of hardware design, through functional module division, nearest neighbor search based on big data technology, database design Improve the data and recommendation results generation, design system software, so as to complete the system design, solve the problem of unclear candidates' voluntary filling, collect the data involved in the new college entrance examination, and calculate one or more volunteer filling candidate schemes recommended to examinees through big data combining the candidates' filling willingness and College entrance examination scores.

2 System Hardware Structure Design

The hardware structure of the system is as shown in Fig. 1. In terms of program platform, SSH framework of Java EE platform is adopted.

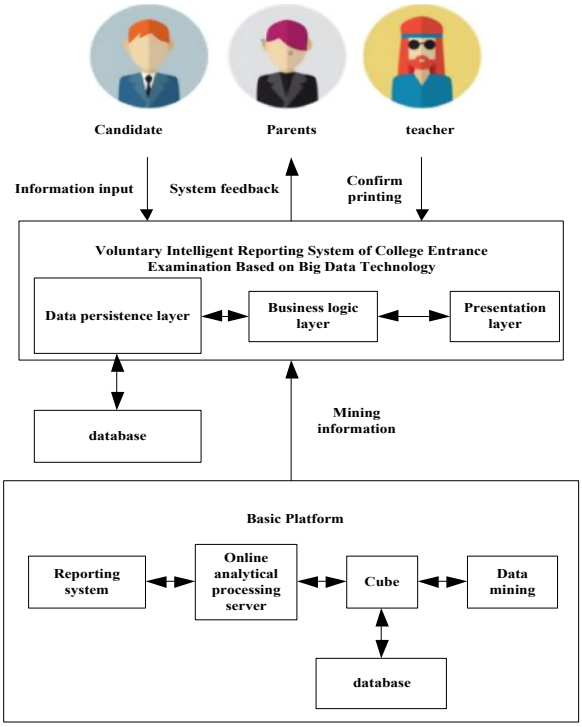

Fig. 1. System hardware structure

In terms of data storage, the SQL Server 2008 platform which has been tested by the enrollment data mining system is adopted. The system adopts a hierarchical structure, from the bottom to the top: basic data platform layer, SSH framework layer and function module layer [6].

2.1 Online Analytical Processing Server

Online analytical processing server is a fast software technology that shares multi-dimensional information, accesses and analyzes online data for specific problems. The online analytical processing server system is the most important application of the data warehouse system. It is specifically designed to support complex analytical operations, focusing on decision support for decision makers and senior management personnel. It can execute a large amount of data quickly and flexibly according to the requirements of analysts' complex query processing, and provide the query results to decision makers in an intuitive and understandable format [7–9]. The relationship between data warehouse and server is complementary. Modern server system is generally based on data warehouse, that is to extract a subset of detailed data from the data warehouse and store it in the server memory through necessary aggregation for front-end analysis tools to read [10, 11].

The server presents a multi-dimensional view to the user:

Dimension: it is a specific angle for people to observe data, and a kind of attribute when considering problems. Attribute sets form a dimension (time dimension, geographical dimension, etc.);

Dimension level: Observing a specific angle of big data (i.e. a dimension), there can also be various description aspects with different levels of detail (time dimension: date, month, quarter, year);

Member of dimension: a value of dimension, which is the description of the position of data item in a dimension. ("month day of a year" is a description of the position in the time dimension);

Metric: the value of multi-dimensional array;

The basic multidimensional analysis operations of the server include drilling, slicing and slicing, and rotation.

Drilling: It is to change the level of dimension and transform the granularity of analysis. It includes drill down and drill up/roll up. Drilling up is to summarize low-level detail data into high-level summary data on a certain dimension, or reduce the number of dimensions; while drilling up is the opposite, it drills down from the summary data to the detail data to observe or add new dimensions.

Slicing and slicing: it is concerned about the distribution of measurement data in the remaining dimension after selecting values in some dimensions. If there are only two remaining dimensions, they are slices; if there are three or more dimensions, they are slices.

Rotation: is to change the direction of the dimension, that is, rearrange the placement of the dimension in the table.

2.2 Data Warehouse Server

The warehouse control database contains the control tables necessary to store the metadata of the data warehouse center. In the updated version of the warehouse center, the warehouse control database must be a UTF-8 database. This requirement provides extended language support for the storage data warehouse center. If you try to log in to the storage data warehouse center using a database in a non-coding scheme format, the system will receive an error message that you cannot log in. The system can use the warehouse control database management tool to migrate metadata from the specified database to the new coding scheme database [12–14].

2.3 Design of Enrollment Information Data Mining System

The data mining system of enrollment information adopts a distributed system structure, which is based on the database of college entrance examination application, electronic data of general enrollment, application programming interface and large data files provided by some college network applications, using analysis, prediction, association rules, clustering and other mining methods, from a large number of incomplete and fuzzy practical application data, the relationship between colleges and majors hidden in it is found [15]. Analysis, statistics and reasoning of effective information such as school relationship, historical admission score line, enrollment plan, etc. can provide predictive suggestions for candidates to fill in the report. The architecture of enrollment information data mining system is shown in Fig. 2.

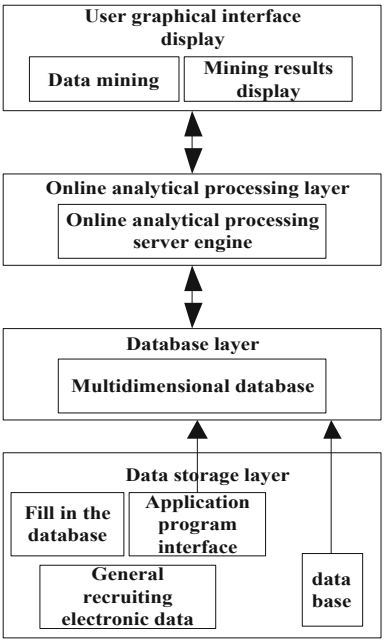

Fig. 2. Enrollment information data mining system

The data mining system of college enrollment information consists of front-end user interface, data preprocessing, data mining algorithm and other modules. The front-end user interface module mainly completes the interaction between the user and the system, and displays the mining results in easy-to-understand forms such as graphics and tables. The data preprocessing module mainly completes the extraction, cleaning, filtering and integration of a large number of complex redundant data in the data storage layer, and provides high-quality rule data for the data mining algorithm module.

3 System Development

3.1 Functional Module Division

The design goal of the college entrance examination voluntary intelligent reporting system is to provide candidates and parents with more accurate guidance for voluntary reporting. In addition, the system also needs to have security, stability and scalability. Therefore, the system is based on big data technology, and realizes six functional modules: human-computer interaction module, user management module, authority control module, voluntary management module, interest preference management module, voluntary recommendation module.

(1) Human-computer interaction module

It mainly provides convenient and beautiful user interface, and has friendly prompt for any operation of the user. At the front desk of the system, JavaScript dynamic scripting

technology is adopted to realize the page refresh operation, which can improve the user experience. The purpose of human-computer interaction module is to let users use the filling consultation system at the lowest cost, and obtain good experience effect.

(2) User management module

User management module is the main carrier of user registration and login system, which mainly includes user registration, adding and modifying basic personal information and details, password management, system login, etc. User management module is the main carrier of user registration and login system, which mainly includes user registration, adding and modifying basic personal information and details, password management, system login, etc.

(3) Authority control module:

The permission control module belongs to the background function, which complements the user management. This module is mainly used for the security control of the system. It can realize the system to judge the authority of the currently logged-in user, which can enable legitimate users to browse the appropriate content. The allocation of user rights can be managed in the background.

(4) Volunteer management module

Volunteer management belongs to the background function, mainly the addition, deletion and modification of colleges and majors. This module is used in the voluntary recommendation module, which belongs to the bottom data support. It caches the mining data at the database level and improves the operation efficiency of the system in the voluntary reporting guidance.

(5) Interest preference management module:

This module is a psychological test application module, which mainly includes candidates' professional preference, regional tendency, cost burden and other aspects. Among them, this module is also related to the user's "detailed information" in the user management module, including future, urban or rural, whether ethnic minorities, etc.

3.2 Nearest Neighbor Search Based on Big Data Technology

3.2.1 User Data Standardization

In order to facilitate data processing, according to the candidate's score information and intention information entered by the candidate, after the user attributes are determined, the data is standardized, and min-max standardization is used to standardize the data, as shown in Eq. 1:

$$s' = \frac{s - s_{\min}}{s_{\max} - s_{\min}} \tag{1}$$

In formula (1): s represents the original data of candidates; $s_{\max}$, $s_{\min}$ represents the maximum and minimum values of attributes respectively.

3.2.2 Calculate User Proximity

Nearest neighbor search is to calculate the similarity between candidates and their preferences based on their attributes. On the basis of standardized candidate attribute data, the Pearson correlation coefficient is used to perform nearest neighbor search, and the distance between candidates is calculated to express the similar proximity between candidates, as shown in Eq. 2:

$$s = \frac{\sum_{iek} \left(t_{u,i} - \overline{t_u}\right) * \left(t_{v,i} - \overline{t_v}\right)}{\sqrt{\sum_{iek} \left(t_{u,i} - \overline{t_u}\right)^2 * \left(t_{v,i} - \overline{t_v}\right)^2}} \tag{2}$$

In formula (2): $t_{u,i}$ represents the i attribute value of user u; $t_{v,i}$ represents the i attribute value of user v; $\overline{t_u}$ represents the average value of all attributes of user u; $\overline{t_v}$ represents the average value of all attributes of user v.

3.3 Database Design

Database design plays an important role in website development and construction. A good website must be supported by a safe and high-performance database. In the process of candidates applying for a test, the safety, effectiveness, and real-time nature of the data are directly related to whether the candidates can be admitted to their favorite schools. Based on big data, the database of college entrance examination application is constructed by this system, which can obtain the latest enrollment information of colleges and majors through the data mining system of enrollment information. Table 1 and Table 2 are field definitions and descriptions of some data tables.

Table 1. College information table

Field name data type primary key remarks	Field name data type primary key remarks	Field name data type primary key remarks	Field name data type primary key remarks
COLLEGENAME	varchar (50)	Y	University name
PROVINCEID	varchar (50)	Y	Province
BATCHID	tinyint	N	Batch
COLLEGETYPEID	tinyint	N	Category
PLAYTYPEID	tinyint	N	Nature of plan
IS211	tinyint	N	Whether 211 institutions
IS985	tinyint	N	Whether 985 colleges
ISFOREIGN	tinyint	N	Chinese and foreign

Table 2. Batch control score

Field name data type whether primary key remarks	Field name data type whether primary key remarks	Field name data type whether primary key remarks	Field name data type whether primary key remarks
CLASS_ID	varchar (50)	Y	Family ID
BATCH_ID	varchar (50)	Y	Batch ID
YEAR	int	N	Year
CULTURE_SCORES	int	N	Cultural achievements
PROFESSION_SCORE	int	N	Professional performance
PROVINCE_ID	tinyint	N	Province ID
UP_NUM	int	N	Number of people on line

3.4 Complete Material

After entering the system, you need to first improve relevant information, such as personal details, interest preferences, etc. First click on "Personal Center" and enter the "Modify Details" page, you need to follow the prompts to improve all information, the most important of which is the candidate's urban and rural category (rural or urban), the previous category (fresh or past), whether it is a minority This information will be used as a reference when recommending volunteers. Then, click "Report Volunteer" to enter the preference setting mode.

The first step is to select the major categories of interest, which can be multiple choices;

The second step is to choose your own preferred area according to your geographical preference.

The third step is to choose the expense range that you can bear;

The fourth step is to choose your own reporting psychology, which is mainly divided into insurance or sprint types and strong emphasis on professional compliance. The default is to list all possibilities. After preference setting, enter the score setting and select science or liberal arts page. The relevant information set by the user is used as the basis for the voluntary recommendation of the system. If you modify the user profile, you can perform voluntary recommendation analysis in different states.

3.5 Recommendation Result Generation

The recommendation result is based on the admission of universities and professions of neighboring users. First of all, the neighboring users should be determined. Through the calculation and search of similar proximity, the user with the nearest neighbor of 0 is regarded as the nearest neighbor of the target user, and the neighbor is regarded as the neighbor. Corresponding institutions of successful admission are added to the

recommendation set. Due to the different difficulty of the test paper each year, the value of the score as the reference value is not stable, so it is more strict and accurate to take the ranking of users as the main reference.

According to the ranking segment of the target user's score, select the x users with the smallest distance as neighbor users, and arrange the universities enrolled by the x users in descending order according to the number of people. And then add the recommendation set in turn, and then screen one by one according to the preferences of candidates. After screening, the recommendation content is put into the new recommendation set until the number of the new recommendation set reaches the target of 50, forming the final recommendation set.

4 Simulation System

4.1 System Login

First, enter the URL to enter the first page of the system, click "register new user" to register, after entering the relevant information, the system will automatically send a verification email to verify activation. Then use the approved account and password to log in, and enter the correct dynamic verification code, you can enter the guidance system.

4.2 System Voluntary Recommendation

At present, the system realizes three batches of voluntary recommendation, one batch of undergraduate, two batches of undergraduate and three batches of undergraduate. Each batch is divided into six volunteer. The function of voluntary recommendation includes shortcut mode and advanced mode. The difference is that the advanced mode can be modified for voluntary selection.

In any batch and any voluntary item, click "Add College Major". In the pop-up page box, you can choose "School Priority" and "Professional Priority". Major priority refers to the arrangement of the most likely major to be admitted, and then to the recommendation of colleges and universities with relevant majors, from high to low; college priority refers to the recommendation of the most likely college, and then to the selection of major. Regardless of whether it is an institution or a major, the interface of selecting an institution will prompt the institution's admission to similar conditions in the past three years.

4.3 System Simulation

Personal details: rural areas are selected for urban and rural areas.

Interest tendentiousness: professional tendentiousness chooses computer, network and technology, regional tendentiousness chooses southwest region, expense interval chooses 4001–10000, mental state system chooses sprint.

Relevant application requirements: The application category is science and the original score is 580 points. After inputting the data, the first volunteer is recommended

with the priority of College recommendation. According to the input conditions, the recommended institutions are shown in the Table 3 below.

Table 3. Simulated voluntary recommendation table

School name	2019	2018	2017	Select
List of eligible schools for three consecutive years				
XX University	562.0	544.0	505.0	√
XX University	556.0	533.0	510.0	√
XX University	547.0	525.0	506.5	√
XX University	564.0	520.0	565.0	√
List of eligible schools for two consecutive years				
XX University	565.2	540.0		√
List of schools that meet the standard for one year in a row				
XX University	565.0	598.0	554.0	√

Then, you can choose a college as your first volunteer, and so on, you can choose a second volunteer or other batch of volunteers. Finally, it can print the guidance form for filling in the application form for the convenience of reference.

Design test cases and conduct system tests based on the test cases. First fill in the candidate's ranking, batches, grades of three public courses in Chinese, mathematics and foreign language, and 7 of 3 subjects and corresponding scores, and automatically generate total scores, as shown in Fig. 3.

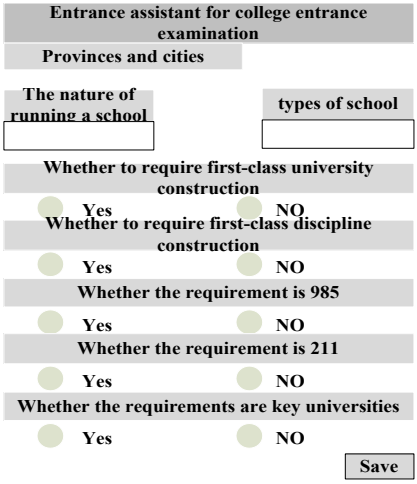

Fig. 3. Candidates' scores

Then, candidates fill in their willingness to report their intentions, including the excluded provinces, municipalities and autonomous regions, the nature and type of school running, whether they require first-class university construction and whether they require first-class discipline construction, and whether they require 985, 211 and key universities.

According to the data in Fig. 3, the intelligent recommendation platform for college entrance examination volunteer filling in is calculated by the server, and finally the recommendation result list is pushed to the Android client, as shown in Fig. 4.

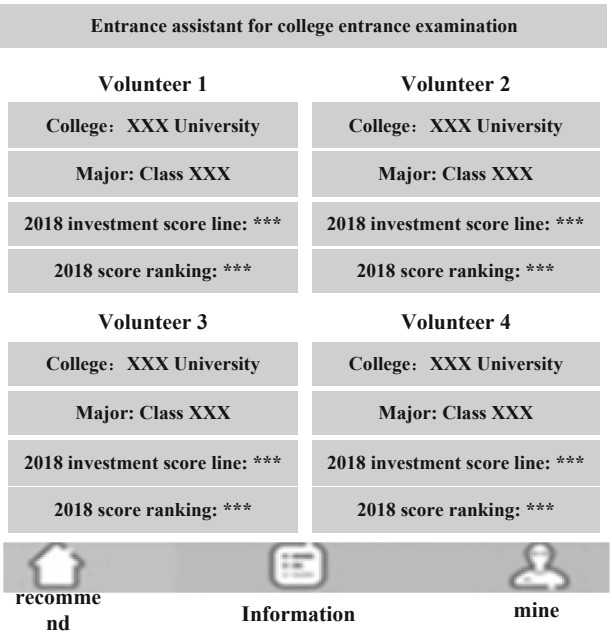

Fig. 4. List of recommended results

5 Analysis of Experimental Results

Based on the big data technology, this paper studies the intelligent filling system of college entrance examination, and takes the practical application of the display system as an example to carry out experimental verification analysis. The specific process of the example is to simulate the examinee or parents to use the system, and input relevant information according to the actual operation process, including account registration, login, personal information improvement, performance and other necessary information input, and display the relevant human-computer interaction screenshots.

Based on the analysis of the mainstream domestic college entrance examination voluntary reporting platform W1, the college entrance examination voluntary reference system W2, the Sina college entrance examination simulated voluntary reporting system W3 and the big data technology-based college entrance examination voluntary intelligent reporting system W4, the results of the comparison and analysis are shown in Fig. 5 As shown.

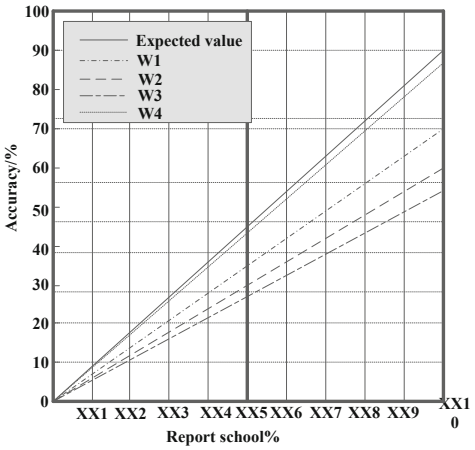

Fig. 5. Comparison and analysis of filling accuracy between the two systems

It can be seen from Fig. 5: The use of the mainstream domestic college entrance examination voluntary reporting platform W1, the college entrance examination reporting voluntary reference system W2, and the Sina college entrance examination simulation voluntary reporting system W3 lack in-depth data analysis functions, unable to find dynamic data, resulting in the system not reaching the actual predicted value Fundamentally solve the blindness of candidates when they fill in the report. Although the intelligent filling system of college entrance examination based on big data technology can't exactly match the actual predicted value straight line, it is also very close. Therefore, the intelligent filling system of college entrance examination based on big data technology has accurate and intelligent filling effect.

In order to further verify the effectiveness of the system, this paper compares and analyzes the student satisfaction of the system and the traditional system, and the results are shown in Fig. 6.

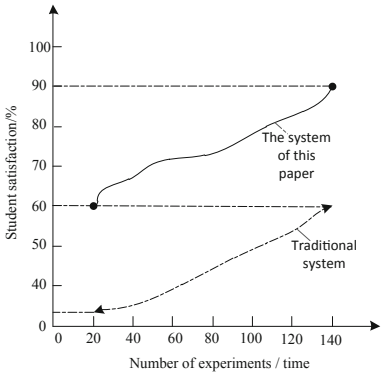

Fig. 6. Comparison of student satisfaction results

According to Fig. 6, students' satisfaction can reach up to 90% with the application of this system, while only 60% with the traditional system, indicating that the application of this system can make students more satisfied in the intelligent filling of college entrance examination.

6 Conclusion

There are unstable factors and links in voluntary reporting. Most examinees and parents rely on subjective analysis and teachers' relevant historical experience to make voluntary reporting. The whole process does not comprehensively combine the objective factors such as scores, policies, school conditions, personal and family conditions to make comprehensive analysis. Such reporting is often one-sided and subjective, which makes examinees' voluntary reporting Newspaper is full of gambling and blindness. Therefore, the research on the intelligent reporting system for college entrance examination based on big data technology is designed.

Through the preliminary trial operation of the project, it shows that the establishment of filling and consulting system based on enrollment data mining is of high feasibility and application value. Summary of the current work, although achieved some results, but there is still a lot of room for improvement, improve or create mining algorithm, in order to get better results of enrollment data mining, it may be necessary to do further research on the algorithm and propose a suitable algorithm.

Acknowledgements. 1. The "Thirteenth Five-Year Plan" Science and Technology Project of Jilin Province Education Department "Research on the Service System of Volunteer Filling for College Entrance Examination Based on Data Mining" (Project Contract Number: JJKH20180466KJ).

2. The 2017 General Planning Project of the "13th Five-Year Plan" of Educational Science in Jilin Province "Research and Practice of Senior High School Students' Career Planning from the Perspective of the New College Entrance Examination Reform" (Project approval number: GH170348).

3. The 2017 Higher Education Scientific Research Key (Self-funded) Project of Jilin Province Higher Education Society "Research on the Decision Model of College Students' Career Planning" (Project Number: JGJX2017C43).

References

1. Ying, Y.: Research on college students' information literacy based on big data. Cluster Comput. **22**(2), 3463–3470 (2018). https://doi.org/10.1007/s10586-018-2193-0
2. Zhang, J., Wang, L., Xing, L.: Large-scale medical examination scheduling technology based on intelligent optimization. J. Comb. Optim. **37**(1), 385–404 (2018). https://doi.org/10.1007/s10878-017-0246-6
3. Bu, F.: An efficient fuzzy c-means approach based on canonical polyadic decomposition for clustering big data in IoT. Future Gener. Comput. Syst. **88**, 675–682 (2018)
4. Jin, S., Peng, J., Xie, D.: A new MapReduce approach with dynamic fuzzy inference for big data classification problems. Int. J. Cogn. Inform. Nat. Intell. **12**(3), 40–54 (2018)

5. Nie, M., et al.: Advanced forecasting of career choices for college students based on campus big data. Front. Comput. Sci. **12**(3), 494–503 (2018). https://doi.org/10.1007/s11704-017-6498-6
6. Chu, S.C., Dao, T.K., Pan, J.S., et al.: Identifying correctness data scheme for aggregating data in cluster heads of wireless sensor network based on naive Bayes classification. EURASIP J. Wirel. Commun. Netw. **2020**(1), 1–15 (2020)
7. Liu, S., Bai, W., Liu, G., et al.: Parallel fractal compression method for big video data. Complexity **2018**, 2016976 (2018). https://doi.org/10.1155/2018/2016976
8. Kumar, P.M., Lokesh, S., Varatharajan, R., et al.: Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier. Future Gener. Comput. Syst. **86**, 527–534 (2018)
9. Shuai, L., Weiling, B., Nianyin, Z., et al.: A fast fractal based compression for MRI images. IEEE Access **7**, 62412–62420 (2019)
10. Liu, Z., Hu, L., Wu, C., et al.: A novel process-based association rule approach through maximal frequent itemsets for big data processing. Future Gener. Comput. Syst. **81**, 414–424 (2018)
11. Hu, Z.W., Niu, X.B., Liu, B., et al.: The design and implementation of voluntary filling assistant system under the new college entrance examination reform. Intell. Comput. Appl. **009**(002), 175–179 (2019)
12. Hua, T.: Model for evaluating the classification modes of the China's college entrance examination with hesitant fuzzy information. Int. J. Knowl. Based Intell. Eng. Syst. **21**(4), 265–272 (2017)
13. Lu, M., Liu, S.: Nucleosome positioning based on generalized relative entropy. Soft. Comput. **23**(19), 9175–9188 (2018). https://doi.org/10.1007/s00500-018-3602-2
14. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. Entropy **21**(9), 902 (2019)
15. Yao, Y., Zhang, Z., Cui, H., et al.: The influence of student abilities and high school on student growth: a case study of Chinese National College Entrance Exam. IEEE Access **7**, 148254–148264 (2019)