

RESEARCH ON EARLY WARNING SYSTEM OF COLLEGE STUDENTS' BEHAVIOR BASED ON BIG DATA ENVIRONMENT

Yang Chengyong¹, Liu Jiayi^{1, *}, Huang Sheng¹

1. Modern Educational Technology Center, Guilin University of Technology, No. 12 Jiangan Road, Guilin City, postcode: 541004,
ychy918@163.com

KEY WORDS: Internet of Things; Big data; Behavior early warning; Behavior analysis

ABSTRACT:

Because most schools have been using traditional methods to manage students, there is a lack of effective monitoring of students' behavioral problems. In order to solve this problem, this paper analyses the characteristics of big data in University campus, adopts K-Means algorithm, a traditional clustering analysis algorithm, and proposes an early warning system of College Students' behavior based on Internet of Things and big data environment under the mainstream Hadoop open source platform. The system excavates and analyses the potential connections in the massive data of these campuses, studies the characteristics of students' behavior, analyses the law of students' behavior, and clusters the categories of students' behavior. It can provide students, colleges, schools and logistics management departments with multi-dimensional behavior analysis and prediction, early warning and safety control of students' behavior, realize the informatization of students' management means, improve the scientific level of students' education management, and promote the construction of intelligent digital campus.

1 INTRODUCTION

In 2012, the United Nations issued the White Paper on Big Data, which elaborates on the coming of the big data era and its profound impact on economic and social development^[1]. In August 2015, the State Council promulgated the Notice of the Platform for Action to Promote the Development of Big Data, which has become the basic and forward-looking technology of our country and is the inherent need and inevitable choice for the country to implement the strategy of innovation-driven development^[2]. Relevant studies have found that about 20% of college students have psychological problems, of which 15% belong to general psychological problems and need psychological counselors and relatives and friends to provide psychological counseling. 3.5% of the students have mental disorders and often suffer from emaciation, insomnia and restlessness. 1.5% of the students are psychotic and lose self-control ability. They cannot distinguish reality from hallucination^[3]. At present, most of the behavior management methods of school students in our country are propaganda, sermon and regular investigation. Their management methods only focus on qualitative analysis afterwards. This kind of management means is very passive and backward. It is no longer suitable for the needs of modern school education management, nor for the direction of development in the era of big data.

Many colleges and universities use campus big data to carry out many applications for management and teacher-student service. Relying on big data mining methods to support campus management and decision-making in the field of education and the analysis of students' behavior rules has attracted wide attention. Many foreign online educational institutions, such as edX, Coursera and Udacity, aim at the high dropout rate of many students in the online education process. The data mining method is applied to the deep mining and analysis of students' online learning data, and the intrinsic causes of students dropping out are found. Based on the mining results, the strategies of supervision and guidance are formulated, which greatly improves the dropout problem of online education^[4]. Miller F G^[5] and others studied the

reliability and validity of the data obtained by three commonly used school-based behavioral assessment methods, and adaptively helped students establish a knowledge system for training. The data mining method is used to analyze the effective information under the students' learning data, and then to evaluate and feedback it. It can help students learn and also provide teaching strategies for teachers effectively^[6-8]. In view of the research on the mining method of a large number of consumption records data generated in the campus card application, document [9] takes HDFS as the data storage medium, combines Spark platform with data mining technology, and excavates data sources based on smart campus data. Literature [10] studied the rule extraction method based on the students' performance data, using decision tree mining method to predict students' curriculum performance, to help teachers understand students' performance, to provide appropriate learning suggestions, and to improve the quality of teaching. Document [11] Based on Hadoop platform, large-scale data in intelligent campus system are analysed and mined, and a collaborative filtering recommendation system based on student similarity is established to realize campus information sharing and recommendation.

2 .DESIGN OF EARLY WARNING SYSTEM FOR COLLEGE STUDENTS' BEHAVIOR BASED ON BIG DATA ENVIRONMENT

2.1 Workflow of College Student Behavior Early Warning System Based on Big Data Environment

As shown in Figure 1, based on the data of student consumption, performance, attendance, access control, book borrowing and other sources, this paper analyses student behavior and predicts the law of student life. Firstly, data is preprocessed, multi-source data is fused, and data is stored in HDFS, which ensures data consistency with relational database, and makes it easy to convert data between distributed storage and relational database. At the same time, through the parallelization of machine learning algorithms such as clustering analysis, association rules mining, collaborative filtering on Hadoop, and the application of this technology to

*Corresponding author: Liu Jiayi; E-mail: 2840313589@qq.com;

Project fund: Promotion Project for Young and Middle-aged Teachers in Guangxi Universities (2018KY0252)

the distributed processing of historical data, we can complete the clustering and subdivision of students' behavior, the prediction and warning of students' learning behavior, and the

classification and prediction of students' activity paths. Finally, using Java Web technology, the analysis results are visualized and friendly user query interaction functions are provided.

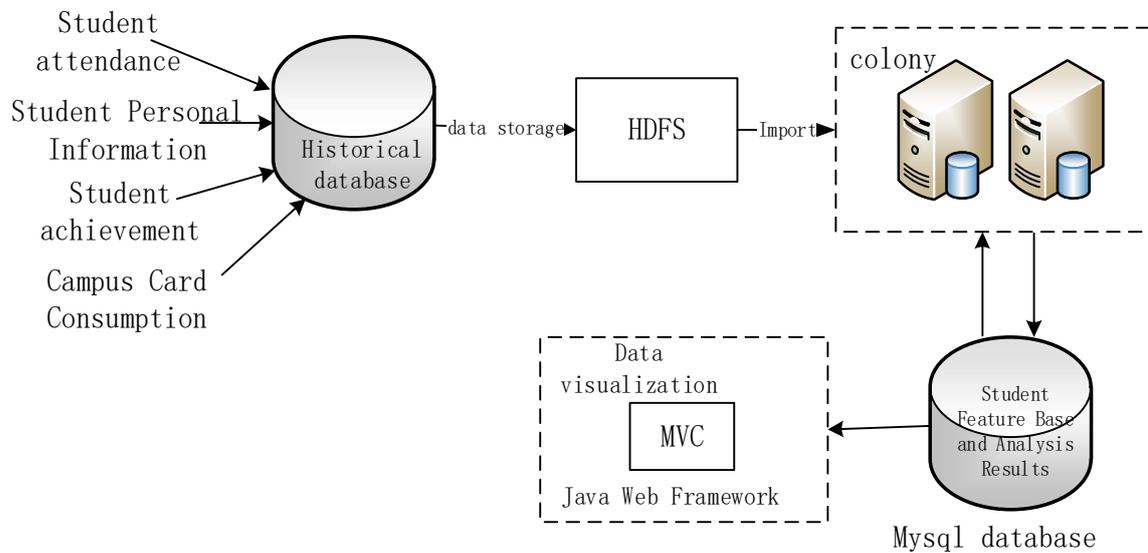


Figure 1. Systems Overall Business Process

2.2 Structural Design of College Students' Behavior Early Warning System Based on Big Data Environment

The early warning system of College Students' behavior in the Internet of Things and big data environment adopts a typical big data technology framework, which is divided into three levels: data source layer, data processing layer and application layer. Data source layer includes structured, unstructured and semi-structured data. Structured data includes data collected automatically by "one-card"

system, attendance system and financial system. Unstructured and semi-structured data mainly come from the Internet, documents and so on. Data processing layer includes data acquisition (stream data processing, network crawler, etc.), data storage (main data warehouse), data operation (distributed database, Hadoop platform, etc.). The application layer includes data analysis, intelligent early warning, real-time monitoring and so on, as shown in Figure 2.

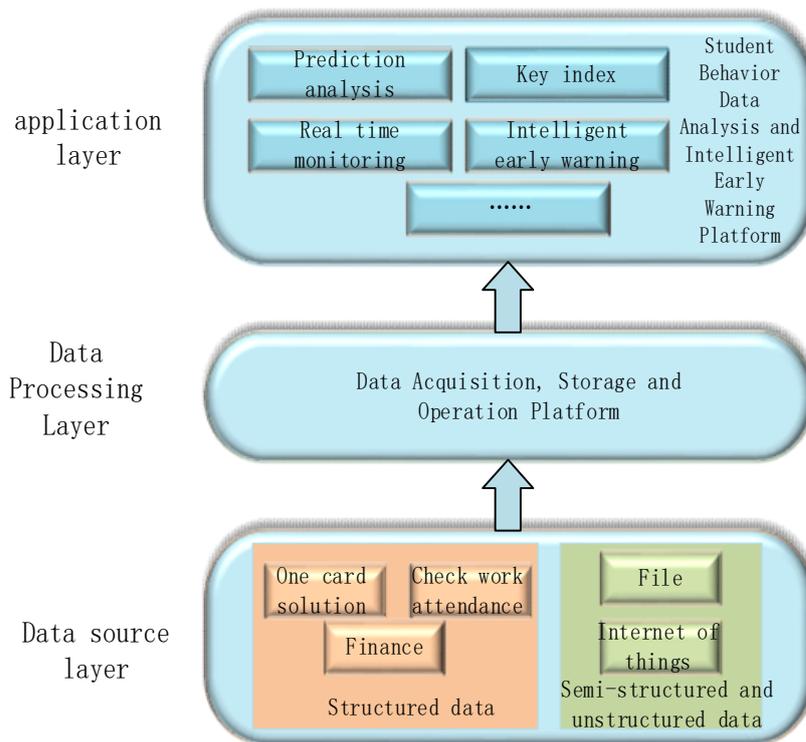


Figure 2. Technical Architecture of Platform

3 APPLICATION OF K-MEANS CLUSTERING ALGORITHMS IN COLLEGE STUDENTS' BEHAVIOR EARLY WARNING SYSTEM BASED ON BIG DATA ENVIRONMENT

This paper applies traditional K-Means clustering algorithm to analyze students' behavior in the system. This algorithm is an indirect clustering method. Its main idea is to divide data sets into different classes through iteration process, and to analyze, cluster and predict students' behavior trails, attendance, consumption hotspots and interpersonal relationships.

Means algorithm^[12] firstly chooses K clustering targets from the sample set, then calculates the distance between the sample data objects according to the requirements of the rule algorithm, and then analyses the grouping situation of the sample data objects according to the calculation results, and obtains K clustering results through iterative calculation until the cluster center remains unchanged. The implementation process of K-Means clustering algorithm is as follows:

Input: Set the number of data objects m and clustering K in sample set N.

$$N = \{s_1, s_2, \dots, s_m\}, K = \{a_1, a_2, \dots, a_k\}$$

Output: K clusters satisfying minimum criteria of criterion function. The processing flow is as follows:

(1) K samples are selected arbitrarily from sample set N as initial clustering centers.

(2) According to the mean value of each clustering sample, the distance d between each sample and the central object is calculated (represented by Euclidean distance), and then the corresponding objects are re-divided according to the calculated minimum distance. Let the distance between two q-dimensional sample data points $x_i = (x_{i1}, x_{i2}, \dots, x_{iq})$ and

$x_j = (x_{j1}, x_{j2}, \dots, x_{jq})$ be defined as a formula (1):

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{iq} - x_{jq})^2} \quad (1)$$

The average distance of all sample data points is (2):

$$\bar{d}(N) = \frac{2}{m(m-1)} \times \sum_{i \neq j, i=1, j=1}^m d(x_i, x_j) \quad (2)$$

(3) Recalculate the mean of each sample (central

object).

(4) Cycle the above steps (2) and (3) until the value of the objective function remains unchanged or less than the specified threshold. The objective function is the square error criterion function, which is defined as formula (3):

$$\text{average}_i = \sqrt{\frac{\sum_{i=1}^{m_i} (x_i - c_i)^2}{|C_i| - 1}} \quad (3)$$

Whereas c_i is the centroid of the same class of data objects, the centroid is defined as formula (4):

$$c_i = \frac{1}{|C_i|} \sum_{x_j \in T_i} x_j \quad (4)$$

Where c_i denotes the center of the first cluster

and $|C_i|$ denotes the number of data objects in C_i .

(5) At the end, K clusters were obtained. The implementation process of K-Means algorithm is shown in Figure 3.

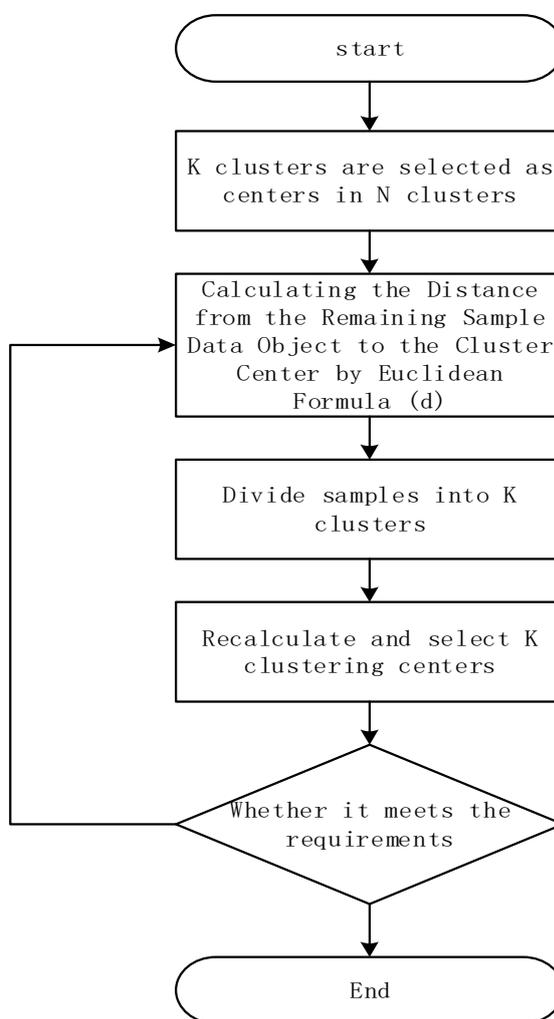


Figure 3. K-Means algorithm implementation process

4 REALIZATION OF COLLEGE STUDENTS' BEHAVIOR EARLY WARNING SYSTEM BASED ON BIG DATA ENVIRONMENT

4.1 System Development Environment and Configuration

Based on the Internet of Things and big data environment, the platform of the early warning system for college students' behavior is built as follows: Firstly, this paper installs and configures VMWare virtual software platform on the physical machine, and then builds Hadoop distributed cluster on the platform (this paper builds three virtual nodes). By using NAT network mode to communicate among nodes, each virtual node is configured with RAM of 4G, external memory of 1T and computing processor of 2.56Hz. The IP address of NameNode node in the Hadoop cluster is 192.168.80.122; the other two are DataNode nodes. The corresponding IP is 192.168.80.123 and 192.168.80.124 respectively.

4.2 The Implementation of Student Early Warning Analysis Algorithms in College Student Behavior Early Warning System Based on Big Data Environment

This paper takes the systematic student early warning analysis algorithm as an example, through the compilation of MapReduce function to analyze student behavior, get the early warning situation of students. The specific algorithm analysis steps are as follows:

1. Referring to the package needed for Hadoop project development, MapReduce functions perform the required classes Job, Mapper and Reducer; And the basic class GenericOptions Parser, which parses command-line parameters in Hadoop framework, can identify some standard command-line parameters and make it easy for applications to specify namenode, job tracker, and other additional configuration resources.

2. Implementing Map static class, which inherits Mapper class, needs to implement Map function, which splits file data in HDFS file system, reads data information and expresses it in the form of key/value, and generates many pairs of key-value pairs. Through the context write function, these key values are output to get the student's student number and the student's early warning record. The student number is the key and the early warning record is the value.

3. Implementing Reduce static class, which inherits Reducer class, needs to implement reduce function. The Iterator < IntWritable > values parameter of the function parameter receives the result of key pair of the school number and early warning situation obtained by map function, and reduces the function to process the key pair to get the early warning information of each student.

4. Complete the map/reduce job configuration, define the output data type, divide the output data into small data blocks splits, provide an implementation of Recordreder, and then provide an implementation of RecordWriter, responsible for data output. Finally, the input and output file directories are set. When the system calls the map and reduce functions, the files are read from the input directory for processing, and the results are stored in the output file directory.

5. The above Java program is executed in Hadoop cluster environment, so as to realize the analysis of massive student data and obtain student number and early warning record information.

In this way, other Mapreduce analysis functions can be written to realize the analysis of students' behaviors in different situations, and then the Java data structure is compiled to store the analysis results in the Mysql database, so as to facilitate subsequent interface display.

4.3 realization of some functions of early warning system of College Students' behavior based on big data environment

Student early warning

user	gender	student_no	major	classify	details	results	date
jacky	male	3130757129	Philosophy	3	breakfast in the morning	confirm	2018-10-13 08:35:51
aron	male	3130757130	Economics	3	breakfast in the morning	confirm	2018-10-13 08:21:10
ceres	female	3130757131	Law	3	breakfast in the morning	confirm	2018-10-13 08:20:12
eve	male	3130757132	Political Science	3	breakfast in the morning	confirm	2018-10-13 08:18:15
brayton	female	3130757133	International Relations	3	breakfast in the morning	confirm	2018-10-13 08:17:33
dorris	male	3130757134	Psychology	3	breakfast in the morning	confirm	2018-10-13 08:15:20

Fig

Figure 4. Student Early Warning Analysis

Student attendance

user	gender	student_no	major	classify	attendance_status	results	date
antonio	female	20180358110	Science of Physical Culture and Sports	3	normal	confirm	2018-10-13 11:20:54
primrose	male	20180358111	Literature	3	normal	confirm	2018-10-13 11:19:30
carver	female	20180358112	Journalism and Communication	3	normal	confirm	2018-10-13 11:19:28
edsel	female	20180358113	History	3	normal	confirm	2018-10-13 11:18:25
raina	female	20180358114	Natural Science	3	normal	confirm	2018-10-13 11:18:23
federico	male	20180358115	Biology	3	normal	confirm	2018-10-13 11:18:22

Figure 5. Analysis of student attendance

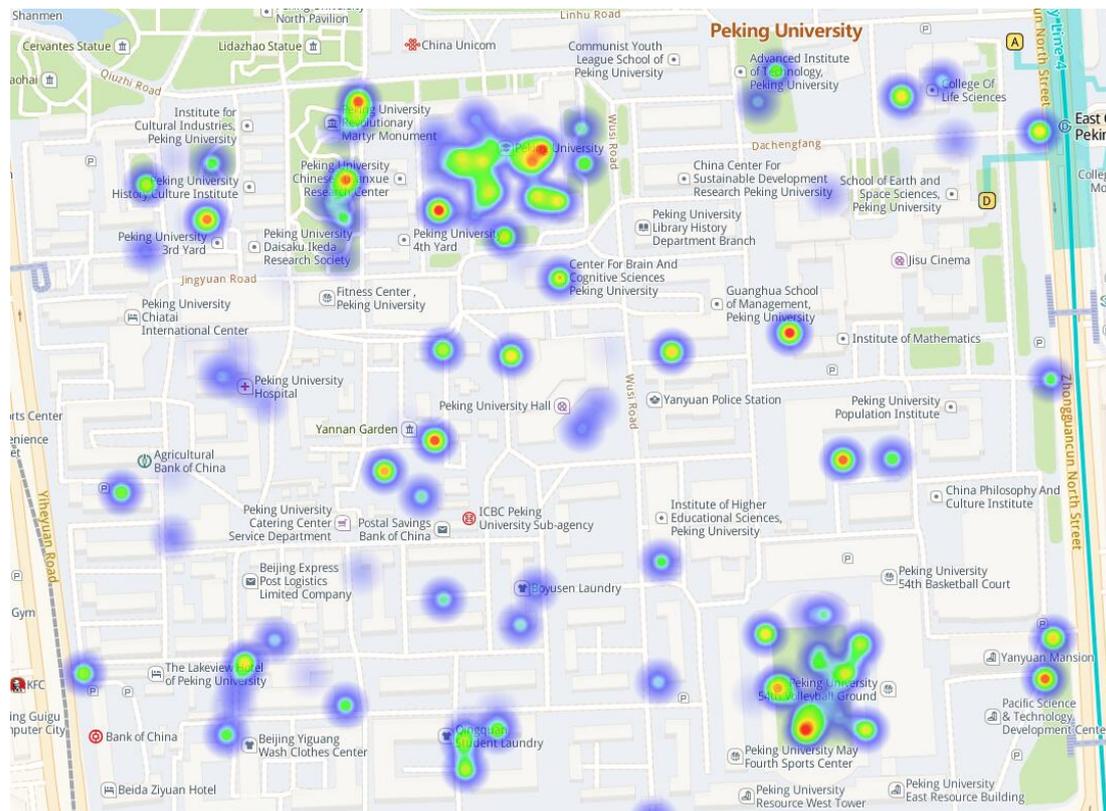


Figure 6. Hotspot Analysis of Student Activities

5 CONCLUDING REMARKS

The early warning system of College Students' behavior based on Internet of Things and big data environment adopts clustering analysis algorithm to analyze the hidden characteristics of data in the

system, and then designs and implements the early warning model of students' behavior prediction by establishing it. The system is redeveloped under the popular Hadoop open source platform. Firstly, the underlying data acquisition module is designed and developed, and the text storage module is used to

realize the collection and text culture of heterogeneous and non-heterogeneous data. Then, the analysis algorithm is used to realize each functional module, further mining and analyzing various data, and the mainstream architecture is used to complete the development of related modules, at the same time, the analysis results are displayed graphically and intuitively. The platform system can deeply excavate students' behavior and psychological problems, help school administrators make macro-decision of teaching management, assist teaching production safety management and control, realize a series of early warning functions such as early warning index model, early warning information generation, early warning information push, and truly realize intelligent digital campus.

REFERENCES

- [1] Zou T L, 2018. The transformation of school moral education and the improvement of teachers' Moral Education in the era of big data. *Chinese moral education*, 3, 30-33.
- [2] Ren F, Hao J Y, 2018. Explore the road of "future school" supported by big data: application and exploration of teaching diagnosis and quality monitoring platform. *information technology education in primary and secondary schools*, 6, 32-35.
- [3] Yu G L, Zhao F Q, Luo X L, 2017. Mental health education: cognition and evaluation of college students. *Heilongjiang higher education research*, 5, 109-112.
- [4] Rodrigues R L, Ramos J L C, Silva J C S, et al, 2016. Discovery engagement patterns MOOCs through cluster analysis. *IEEE Latin America Transactions*, 14(9), 4129-4135.
- [5] Miller F G, Johnson A H, Yu H, et al, 2018. Methods matter: A multi-trait multi-method analysis of student behavior. *Journal of School Psychology*, 68, 53-72.
- [6] Li H Y, 2017. Student learning behavior analysis based on learning platform log. *computer knowledge and technology*, 13 (6), 160-162.
- [7] Gowri G S, Thulasiram R, Baburao M A, 2017. Educational Data Mining Application for Estimating Students Performance in Weka Environment. *IOP Conference Series: Materials Science and Engineering*, 263, 032002.
- [8] Elbadrawy A, Polyzou A, Ren Z, et al, 2016. Predicting Student Performance Using Personalized Analytics. *Computer*, 49(4), 61-69.
- [9] Li H F, Bai S, Ma Q, et al, 2016. Data mining research on smart campus based on Spark. *intelligent computer and application*, 6 (6), 106-107.
- [10] Hu L L, 2017. Application and research of association rules and decision tree combination algorithm in student achievement analysis. *Qingdao Technological University*.
- [11] Hasbun T, Araya A, Villalon J, 2016. Extracurricular activities as dropout prediction factors in higher education using decision trees. *Advanced Learning Technologies (ICALT), 2016 IEEE 16th International Conference on*, 242-244.
- [12] Chen Y H, 2019. Data mining and early warning analysis of student behavior in university information system. *Hubei Nationalities University*.