ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 7, Issue 12 , December 2020

The application of big data technology in the management of college students

LiufangYu , AijunZhou

Nanjing Normal University TaizhouCollege , Taizhou,225300,China

ABSTRACT: The rapid development of big data technology promotes the construction of intelligent campus, and provides a broad platform for the information, Scientific and intelligent management of college students. In this paper, the students of a university in Taizhou are taken as the research object, through collecting various kinds of application data in the intelligent campus information system, and using K-prototype algorithm of cluster analysis to classify the students' campus behavior characteristics, then the Apriori algorithm is used to analyze the correlation between students' behavior characteristics and their academic achievements. It is found that there is a close relationship between the characteristics of consumption behavior, work-rest behavior and study behavior of different groups of students and their academic achievements. Colleges and universities can take different management measures for different types of students, which will not only help to improve students' academic performance, but also help to further enhance the effectiveness of student management.

KEYWORDS: Big Data Technology; College Student Management; K-prototype Algorithm; Apriori Algorithm

I.INTRODUCTION

With the rapid development of science and technology, the era of knowledge economy comes quietly. With the rapid rise of Mobile Internet, cloud computing, artificial intelligence and big data technology, the development of all walks of life can not be separated from the support of science and technology, and the development of universities is no exception, the information construction of colleges and universities has been upgraded from the digital campus of educational information 1.0 to the intelligent campus of educational information 2.0. In recent years, the construction of intelligent campus in colleges and universities is advancing steadily and improving gradually. More and more application systems are used by students in colleges and universities. Various application systems, such as the one-card system, the wireless campus network, the Educational Information System and the student management system, produce a large amount of data every day. These data are like the "blood" of the intelligent campus, it provides a solid foundation for big data mining in education.

By introducing big data technology into student management, schools can make use of big data technology to analyze students' study, life, consumption and other habits in detail, and deduce students' preferences, ideological characteristics, etc. , it provides convenience for the university to carry out the student management work, promotes the student management information construction to be more scientific, enhances the information management the efficiency. At the same time, under the effect of big data technology, colleges and universities can also grasp more comprehensive information, match students' development with their own management mode, innovate working ideas and methods, and make clear the defects and deficiencies in management, reform and optimize the existing management model, improve the management of college students, precision, refinement, scientific level.

II. APPLICATION AND ANALYSIS OF BIG DATA TECHNOLOGY IN COLLEGE STUDENT MANAGEMENT

In the management of students in colleges and universities, we should scientifically and reasonably classify students' behaviors in schools, formulate corresponding management and service methods for different types of students, and provide individualized and accurate support measures, can Improve the student management and the teaching service work fine degree, promotes the talented person training the quality level. In the era of big data on the Internet, students' consumption data of one-card, attendance data in class, books lending data, Internet data, Pu data,

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 7, Issue 12 , December 2020

etc. , reflect the laws of their study and life in school, can dynamically and accurately map the behavior characteristics of students and behavior habits hidden in these data.

Based on the behavior data of four-year undergraduates in a university in Taizhou, this paper sets up the research variables of student behavior characteristics and student achievement, as shown in Table I. Among them, the consumption behavior data mainly comes from the card system, the rest behavior data mainly comes from the card system and the network authentication system; The data of learning behavior mainly come from the Educational Information System, the library lending system and the online authentication system. The raw data includes 7538456 cards and 895652 books. Because of redundancy and inconsistent structure in the data, the student samples with serious missing information and some outliers were eliminated through data processing, and finally 8672 student samples were obtained.

Table1.Research variables of student behavior characteristics and student achievement

Indicator name		Indicator code	Indicator definition
Consumer behavior	Average monthly consumption	X_1	Average monthly consumption of students in a school year.
	Average monthly consumption of students in a school year	X_2	Average monthly consumption of students in a school year.
	Monthly peak consumption	X_3	The highest percentage of the student's total monthly spending over the school year.
Routine behavior	Eating habits	X_4	Average number of days a student eats regularly each month during the school year.
	Early riser	X_5	Average number of days a student gets up early each month during the school year.
	Internet habits	X_6	Average amount of time a student spends online per month in a school year.
	Exercise physical habits	X_7	Average number of times a student exercises each month during the school year.
Learning behavior	Attendance rate	X_8	Students'attendance/attendance record during the school year.
	Book circulation	X_9	Number of books borrowed by students in a school year.
	Library visits	X_{10}	The average number of times a student goes to the library in a school year.
	Length of study	X_{11}	Average monthly study time of students in a school year.
Academic achievement	Students' academic performance	X_{12}	Students'scores for all courses in the academic year are weighted by credit weight, with less than 60 being poor, 60 to 85 being average, and more than 85 being good.

A) K-prototype Algorithm for clustering analysis of student behavior

The K-prototype algorithm is a typical algorithm for handling mixed attribute clustering. The idea of Kmean Algorithm and Kmode algorithm is inherited, and the dissimilarity formula between prototype and mixed attribute data is added. The general definition: X = { X_1, X_2, X_3... ... ... Xn } represents the data set (containing n data) , where the data has m attributes. Data Xi = { X_{11}, X_{12}, X_{13}... ... . X_{1m} } , where A_j stands for attribute j, and where Dom (A_j) stands for domain j: for numeric attributes, domain dom (A_j) stands for range; for categorical attributes, domain dom (A_j) stands for set. X_{ij} represents the j attribute of Data I. Similarly, Data Xi can also be expressed as

$$x_i = (A_1=X_{11}) \Lambda (A_2=X_{12}) \Lambda (A_3=X_{13}) \ldots\ldots \Lambda (A_m=X_{1m}) \quad (1)$$

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 12 , December 2020

Data a total of M attributes, may wish to set the first P attributes for numerical attributes (r) , after m-r attributes for classification attributes (c)

$$\left[x_{i,1}^{r}, x_{i,2}^{r} \ldots, x_{i,\ p}^{r}, x_{i,\ p+1}^{c}, x_{i,\ p+2}^{c} \ldots, x_{i,\ m}^{c} \right] \quad (2)$$

The K-prototype algorithm sets an objective function, similar to the SSE (sum of squares of errors) of the KMEAN, and iterates until the objective function is constant. At the same time, K-prototype algorithm proposed a prototype of mixed attribute cluster, we can understand that the prototype is the centroid of numerical attribute cluster. There are numerical attribute and classification attribute in the mixed attribute, whose prototype is defined as the mean value of all attribute values in the attribute for the numerical attribute prototype, and the classification attribute prototype is the attribute with the highest frequency of selecting attribute values. Together, it's a prototype.

Before clustering the data, the data is standardized and dimensionless. Then cluster analysis is carried out by using the software of SPSS Modeler, and a new data stream is established.

A.1. CLUSTER ANALYSIS OF STUDENTS CONSUMPTION BEHAVIOR

The K-prototype algorithm is used to cluster the students' consumption data. According to the evaluation criterion of clustering Algorithm, the best clustering effect is obtained when the number of clusters is set to 5. According to the actual situation of students' consumption, the average value of each cluster and the average value of students' overall index are compared, and h is higher than the average value of students' overall index, l is below the average for the student population. The results of cluster analysis on consumption behavior of students in different clusters are shown in table2.

Table2.Results of cluster analysis on consumption behavior of students

Consumption type	Student share （%）	Average monthly consumption (yuan)	Monthly average frequency of consumption (frequency)	Monthly peak consumption (yuan)	Comparison results
1	10.36	472.36	108.5	598.64	LHL
2	15.88	680.24	73.5	1058.31	LLH
3	35.62	715.63	85.2	869.25	HHL
4	22.31	876.28	93.7	1178.64	HHH
5	15.83	505.3	42.6	645.87	LLL
Mean value		687.35	80.91	904.90	

As can be seen from table2, the characteristics of students' consumption behavior can be divided into five categories. The first group has the lowest monthly consumption level and the lowest monthly consumption peak, but the consumption frequency is frequent and belongs to the group of low consumption level;

A.2. CLUSTER ANALYSIS OF STUDENTS 'WORK-REST BEHAVIOR

K-prototype algorithm was used to analyze the habit of diet, Internet, getting up early and exercising in the index of Students' daily behavior. According to the cluster average criterion, when the cluster number is 3, the students' proportion and the average index of each cluster are as shown in Table3

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 7, Issue 12 , December 2020

Tabe3.Results of cluster analysis on students' daily behavior

Sleep pattern	Student share （%）	Frequency of regular Diet (frequency)	Number of early risers (frequency)	Length of Internet access (hours)	Number of physical exercises (frequency)	Compariso n results
1	35.43	22.5	20.1	142.32	52.1	HHLH
2	11.29	9.6	7.2	210.97	7.8	LLHL
3	53.28	14.2	18.3	198.65	19.4	LHHL
Mean value		16.21	17.68	180.08	29.68	

As can be seen from table Ⅲ, the first group of students often get up early every month, eat regularly in the school canteen, surf the Internet for a long time, often participate in physical exercise; The second group of students stayed in bed more every month, had irregular meals in the school canteen, spent the longest time on the Internet and had very little physical exercise, but go to school canteen meal is not regular, Internet time is longer, take part in the number of physical exercise is not much.

A.3 CLUSTER ANALYSIS OF STUDENTS' LEARNING BEHAVIOR

The K-prototype algorithm is used to analyze the four indexes of the students' learning behavior, which are the attendance rate of class, the amount of books borrowed, the number of library visits and the length of study. The students' proportion and average value of each cluster are shown in table 4

Table 4.Results of cluster analysis on learning behavior of students

Type of learning	Student share （%）	Attendance rate（%）	Book Circulation (copies)	Visits to the library (frequency)	Length of study (hours)	Comparison results
1	31.47	94.87	7.8	31.5	165.7	HLHH
2	11.63	85.32	3.5	12.3	89.4	LLLL
3	13.02	92.03	15.7	25.9	135.8	HHHH
4	43.88	90.14	8.2	17.2	117.2	LLLL
Mean value		91.31	8.5	22.3	131.7	

Table4 shows that the first group of students has the highest class attendance rate, the less borrowing books, the most frequent visits to the library and the longest study time; the second group of students has the lowest class attendance, the least borrowing books, the least visits to the library and the shortest study time.

B) USING APRIORIALGORITHM TO ANALYZE THE RELATIONSHIP BETWEEN STUDENTS' BEHAVIOR IN SCHOOL AND THEIR ACADEMIC ACHIEVEMENT

The above cluster analysis has divided the students' behavior into three categories: consumption behavior, work-rest behavior and study behavior. In order to further study the relationship between students' behavior characteristics and their academic achievements, and to find out whether there is a certain relationship between students' behavior characteristics in school and their academic achievements, Apriori Algorithm is selected to carry out correlation analysis, mining Hidden Correlations and laws from big data.

Apriori algorithm is a kind of frequent itemsets algorithm for mining association rules, one of the most influential algorithms for mining frequent itemsets of association rules in Bourg. The core idea is to mine frequent itemsets through two stages: candidate set generation and downward closed detection of plots. Its core is a recursive algorithm

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 7, Issue 12 , December 2020

based on the idea of two-stage frequency set. The association rules belong to single dimension, single layer and Bourg association rules in classification. Here, all the support is greater than the minimum support of the item set known as the frequent item set, referred to as the frequency set. Apriori algorithm has been widely used in business, network security and other fields. The Apriori algorithm adopts the iterative method of searching layer by layer. The Algorithm is simple and clear, there is no complicated theoretical derivation, and it is easy to realize. Algorithm idea: The basic idea of the Algorithm is: First find all the frequency sets, these itemsets appear at least as frequently as the predefined minimum support. Then strong association rules are generated from the frequency set, which must satisfy the minimum support and the minimum confidence. Then use the frequency set found in step 1 to generate the desired rule, producing all rules that contain only the items of the set, with only one item on the right of each rule. Once these rules are generated, only those rules that are greater than the minimum confidence level given by the user are left. In order to generate all frequency sets, a recursive method is used.

（1） L1 = find_frequent_1-itemsets(D);
（2） for (k=2;Lk-1 ≠Φ ;k++) {
(3）Ck = apriori_gen(Lk-1 ,min_sup);
（4） for each transaction t ∈ D {//scan D for counts
（5） Ct = subset(Ck,t);//get the subsets of t that are candidates
（6） for each candidate c ∈ Ct
（7）c.count++;
（8） }
（9）Lk ={c ∈Ck|c.count≥min_sup}
（10） }
（11） return L= ∪ k Lk;

A large number of candidate sets may be generated, and may need to repeat scanning database, Apriori Algorithm are two major shortcomings.

In the parameter setting of Apriori Algorithm Model, five types of students' consumption behavior, three types of students' work and rest behavior, four types of students' study behavior and three types of students' academic achievement are set, is set to the preceding and following variables of the association rule. Set the support degree to 10% , the confidence degree to 80% , carries on the association rule analysis, altogether obtains 24 association rules. According to the goal of this study, the association rules with the degree of improvement greater than 1 are selected, in which the latter is the association rules of students' academic achievement, as shown in table 5.

Table 5.Results of association rule analysis

Posterior term	Antecedent	Percentage rule support（%）	Percentage of support（%）	Confidence percentage（%）
Academic achievement = Worse	Types of routine behavior = 2	8.387	10.332	82.341
Academic achievement = Normal	Types of consumer behavior = 5，Types of routine behavior = 3	9.652	11.852	83.458
Academic achievement = Normal	Types of routine behavior = 3	44.693	53.215	84.781
Academic achievement = Normal	Type of learning behavior = 4，Types of routine behavior = 1	18.994	23.569	80.258
Academic achievement =Good	Types of routine behavior =1，Type of learning behavior =3	8.369	10.327	81.356

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 7, Issue 12 , December 2020

As can be seen from table Ⅵ, the first rule reflects that 10.332% of the students are characterized by not getting up early, eating irregularly, surfing the Internet for a long time, exercising less, and 82.341% of these students are likely to have poor academic performance. The rule support rate was 8.387% which indicated that 8.387% of the students with 2 type of work-rest behavior and poor academic performance. The second rule reflects that 11.852% of the students are characterized by low monthly consumption, low peak consumption, low consumption times, frequent early rising, irregular diet, long Internet time, and less exercise times, there was an 83.458% chance of average academic performance. The rule support rate was 9.652% , which indicated that the proportion of students with 5 type of consumption behavior, 3 type of work-rest behavior and average academic achievement was 9.652% . The third rule reflects that 53.215% of the students are characterized by early rising, irregular diet, long time on the Internet and less exercise, and 84.781% of these students are likely to have average academic performance. The rule support rate was 44.693% , which indicated that 44.693% of the students with 3 type of work-rest behavior and average academic achievement. The fourth rule reflects that 23.569% of the students are characterized by average class attendance, less frequent visits to the library, less books borrowed, less study time, and often getting up early, eating regularly, surfing the Internet for a long time and exercising regularly, there was an 80.258% chance of average academic performance. The rule support rate was 18.994% , which indicated that 18.994% of the students had 4 type of study behavior, 1 type of work-rest behavior and average academic achievement.

III.THE WAY TO OPTIMIZE THE MANAGEMENT OF COLLEGE STUDENTS UNDER THE BIG DATA TECHNOLOGY

Based on the behavior data of undergraduates in a university in Taizhou, the K-prototype algorithm clustering analysis and Apriori Algorithm Association analysis were carried out.

A) USING BIG DATA TECHNOLOGY CAN CHANGE THE STUDENT MANAGEMENT MODEL FROM UNIFIED MANAGEMENT TO PERSONALIZED MANAGEMENT

In the current high-quality development of higher education background, the traditional management model of college students can not meet the needs of the development of the Times. "If we continue to teach today's children the way we taught yesterday, we are killing their future, " said Dewey, an American educator, the ways of obtaining knowledge and information of college students have become diversified, the external information they receive every day has become more and more, the mode of thinking and behavior of college students are changing unprecedentedly, and the need of satisfying their individuality has become more and more intense. Through big data technology, colleges and universities can master students' behavior habits in daily consumption, work and rest, study and so on, timely understand the individualized needs of different groups of students and formulate corresponding management measures, thus causes the student management pattern to change from the unified management to the personalized management.

B) USING BIG DATA TECHNOLOGY CAN CHANGE THE MANAGEMENT OF STUDENTS FROM PASSIVE MANAGEMENT TO ACTIVE MANAGEMENT

Big Data Technology brings about the change of management concept. As the cradle of talent cultivation, universities need to keep up with the pace of social development. All kinds of application systems, such as one-card system, digital campus system, student teaching information system and book lending system, provide abundant data resources for school management. By mining these precipitated big data, we can get the information of the whole school students' study, work and rest, so that we can change the way of students' management work from solving problems after they have appeared before to finding problems on our own initiative, potential hazards are then addressed or pre-empted

IV. CONCLUSION

In this paper, we use cluster analysis method to classify the behavior characteristics of four-year undergraduates in a university in Taizhou, at the same time, through the correlation analysis between the characteristics of students' behavior in school and their academic achievements, we find that there is a close relationship between the characteristics of consumption behavior, work-rest behavior and study behavior of different groups of students and their academic achievements, this provides a basis for the school to take different management measures for different

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 7, Issue 12 , December 2020

types of students, and on this basis, it gives some suggestions on how to use big data technology to improve student management under the background of smart campus construction. Colleges and universities should make full use of the achievements of educational information construction and process a great deal of acquired data information by means of big data technology and information technology, so as to provide rich data support for school decision-making and development.

REFERENCES

[1] Jian Huang. Challenges and countermeasures of College Student Management in Big Data Era [J] . Wit, 2019(14) : 115.
[2] BaiHu.A study on management strategies of college students 'problem behaviors. Chinese Journal of Education, 2015(S1) : 236-237.
[3] Can Wang,WenlianLi,BoSun.Role Shift, big data and innovation of Student Education Management Model [J] . Adult education, 2017(13) : 43-46.
[4]PeijiWang,YulinZhao,JianfengLv.Research on association rule data mining based on Apriori Algorithm [J] . Statistics and decision-making, 2011(23) : 19-21.