

Analysis of Student Health Status based on Campus All-in-one Card Consumption Data

Nan Wang, Yin Qing

College of information science and technology (College of network security, Oxford Brooks College), Chengdu University of Technology China Chengdu 610059, China.

Abstract

In recent years, the number of students in school has increased year by year, and the school has entered overload management, which makes it difficult for administrators to discover students' health problems in time. The consumption of students in canteens and supermarkets is based on the campus card, which has accumulated a large amount of student consumption data. Behind these data are most of the students' daily behavior information. Through mining and analysis of daily behavior activities, we can understand students' State of health. This article is based on the consumption data of the all-purpose card when students are in school, analyzes by cluster analysis method, and applies the results to the analysis of the health status of students in the school, helping school administrators to provide timely help and psychological counseling to students with health problems and promote school Intelligent management.

Keywords

Cluster Analysis; One-Card Consumption Data; K-Means Algorithm; Student Health and Safety.

1. Introduction

In recent years, as colleges and universities continue to expand enrollment, the number of students entering colleges and universities has increased dramatically. Due to the obvious difference between college life and high school life, students need to adjust in many ways to adapt to college life [1]. However, affected by factors such as subjective initiative, autonomy, and weak self-control ability, some students are in poor physical condition and fail to meet the physical test standards every year, which has a huge negative impact on the future development of students. School administrators have also entered an overloaded work state, and it is impossible to have a clear grasp of the students' situation. How to discover students' health problems in time and provide them with help is a problem that every school administrator needs to solve urgently. The physical condition of students is affected by many factors. Generally speaking, before students have physical problems, they have been reflected in their daily lives, and the swiping data of the campus card has a record of most of the students' consumption activities in the school. This article takes this as the starting point. According to the increasingly mature application of the campus card in schools, the storage of consumption data is also becoming more and more perfect. The use of K-means and statistics methods are used to study student consumption levels and breakfast consumption. According to the results, it analyzes the health status of students, helps school administrators understand students' consumption levels and school life, provides timely help and psychological counseling to students with physical problems, and provides a data basis for improving school management and decision-making capabilities.

2. Data and methods

2.1 Data source

The data selected in this article comes from a college card system, which records the card number, consumption amount, consumption time and consumption location of the entire school. In the daily life of school students, more than 300,000 transactions per month are consumed, and these data will be stored in the database. This paper selects 36183 consumption flow data generated by card-swiping behaviors such as dining in the canteen and consumption in supermarkets during the period of 2019.10.24-2019.11.24 among 275 students majoring in electronic information engineering in this university in 2019. Before data analysis, perform data preprocessing such as data audit, data cleaning, data integration, data transformation, and data reduction to ensure the integrity and validity of the data.

2.2 Analysis method

2.2.1 Concept of cluster analysis

Clustering is to divide data into several groups, so that data points in the same group are more similar than data points in other groups [2]. In a large amount of data, we can treat the same type of data in the same cluster as a group, the distance between samples in the same cluster is reduced as much as possible, and the distance between samples in different clusters is increased as much as possible, so that the data simplification goal can be achieved.

2.2.2 Cluster analysis classification

Faced with different data distribution characteristics and different tasks, different clustering methods are used. At present, the common cluster analysis methods are grid-based, density-based, hierarchical, model-based, and partition-based [3]. This article will use the K-means partition algorithm to analyze student consumption data. The basic idea of the division algorithm is to first divide a data set containing n data objects into k groups, each group represents a class, and $k \leq n$, and then iteratively cluster the data continuously, and finally these data form k stable classes, that is, k clusters. This requires at least one data object in each class, and to satisfy that each data object can only be divided into only one class. At the same time, the similarity of the attributes of objects in the same class is relatively high, and the differences in object properties in different classes are more obvious.

2.2.3 K-means algorithm

K-means is one of the most widely used partition-based clustering methods, with the characteristics of fast data convergence [4]. It uses feature distance as the standard, that is, the smaller the distance between data objects, the higher their similarity, and the greater the possibility that they are classified in the same category. The basic idea is to randomly select the initial cluster center for a given cluster category number k , and then calculate the Euclidean distance between each sample and the cluster center, and classify each sample in the sample set to the nearest in the class represented by the class center, after all the samples are allocated, the mean value of each class sample attribute is calculated, the cluster center is rebuilt, and this process is repeated until the function convergence stops.

2.2.4 Features of K-means algorithm

- (1) The K-Means algorithm is usually only suitable for processing numerical data clustering, and not suitable for processing Boolean or mixed data clustering.
- (2) The logical structure is simple, the algorithm execution efficiency is high, there is no multiple iterations, and it has good scalability.
- (3) Sensitive to "noise" and easily affect the accuracy of clustering centers.
- (4) The number k of the divided categories must be given first. The number of categories is usually determined by the corresponding background knowledge of the field or combined with the actual situation. Usually $k=2,3,4,5,6,7$, etc., and then compare the distances between the categories of various situations.

(5) The calculation method of the average value of the class needs to be given first, and the Euclidean distance is commonly used as the measure of similarity.

(6) The results are interpretable and practical, which is convenient for users to understand and use [5].

2.2.5 Advantages and disadvantages of K-means algorithm

The K-means algorithm has the advantages of simple logic and high efficiency. Due to its iterative function, it can overcome the inaccuracy of a small number of sample clusters and reduce the time complexity. Secondly, the K-means algorithm is used for numerical data sets. It is easy to cluster, and the effect is better; finally, K-means also optimizes the unreasonable sample classification such as supervised learning.

The drawback of the K-means algorithm is that the selection of the k value is artificially determined in advance. Sometimes it is not known in advance how many categories a given data set should be divided into. The choice of the initial clustering center has a great influence on the result; Secondly, the K-means algorithm is sensitive to "noise" and is easily affected by abnormal data; thirdly, when the amount of data is very large, the time overhead of the algorithm is very large [6].

3. Result analysis

3.1 Analysis of student consumption level

In order to help poor college students successfully complete their studies, the state and schools will provide some assistance to poor students every year, such as providing scholarships, student loans, and part-time job-study positions[7]-[8]. However, a large number of students submit applications every year. Due to various factors, some students in need of financial assistance do not receive the help they deserve. This leads to their consumption in the cafeteria and can only consume the minimum amount of nutrition. In a sub-healthy state. For college students, with the exception of a few special circumstances, most students' daily basic consumption is on campus, that is, to a certain extent, the consumption level of students on campus all-in-one card can reflect their real consumption level and family status. Schools can use the student card consumption level, and under similar conditions, give priority to providing subsidies or part-time opportunities for students with low consumption levels to help students supplement nutrition, enhance their physical fitness, and keep their bodies in a healthy and active state.

Every time a student swipes a card, a record will be generated and saved in the campus all-in-one card database. Since the amount of each consumption will have a different record, to understand the student's consumption level for a month, you need to use the student's campus card number as a mark to consume Group sum of amounts. Table 1 shows part of the data of the total consumption of 275 students majoring in telecommunications in 2019.10.24-2019.11.24 in one month.

Table 1. Total consumption data of student card

Number	Card number	Total consumption (unit/yuan)
1	171915	321
2	171927	565.5
3	171403	651
4	171583	471.5
5	169937	501
6	170012	812
7	170034	637.5
8	171399	961
9	169994	732

In order to have a clearer understanding of the distribution of students' consumption amount, the total consumption amount is drawn into a distribution histogram as shown in Figure 1.

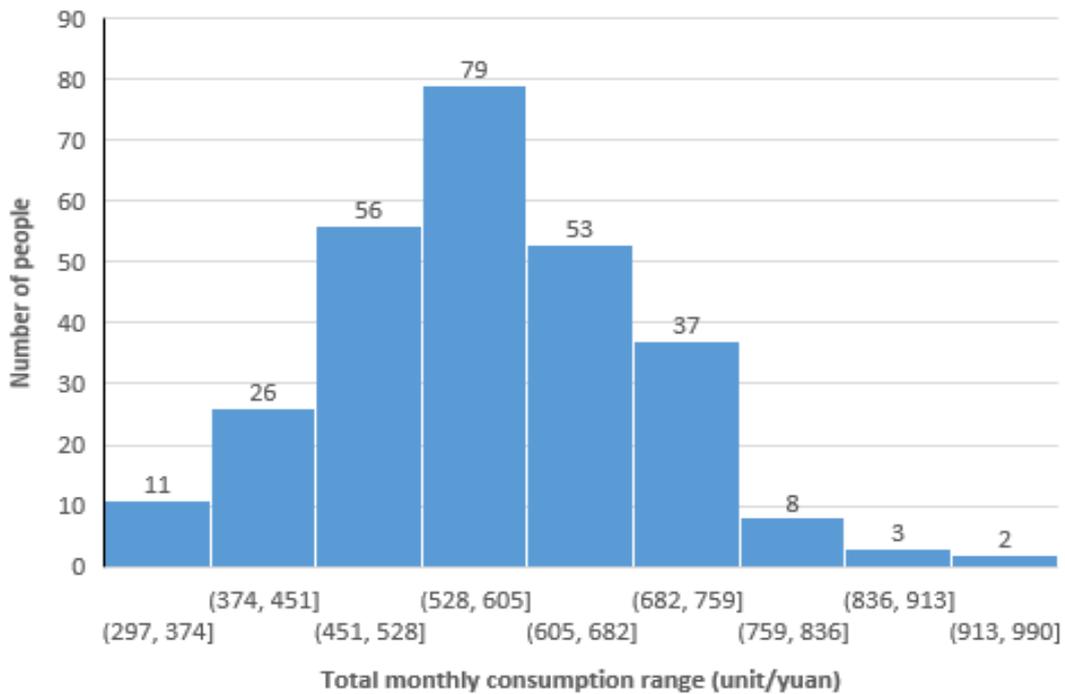


Figure 1. Distribution of total monthly consumption of students

It can be seen from Figure 2 that the total monthly consumption of students is mainly distributed in the interval [451,682]. This paper uses the K-means clustering method to divide the consumption of students into three levels: high-level consumption, medium-level consumption, and low-level consumption. Therefore, we need to set the value of the number of cluster centers k to 3. After experimental research, it is found that when the random seed is set to 4, the error square and SSE in the cluster are the smallest, that is, the effect is the best. At this time, the obtained cluster center is 460.5, 571.9, 736.7, the clustering effect is shown in Table 2.

Table 2. Cluster centers and number of cases

Amount of consumption	Cluster		
	1	2	3
Cluster center	460.5	571.9	736.7
Number of cases in cluster	37	188	50

According to the cluster analysis, we can see that all samples are clustered into three clusters, one of which contains 13.45% of the samples, and the cluster center is 460.5, which means that this kind of students averaged during the month from October 24 to November 24 Consumption is around RMB 460.5, which is a low-level consumption; the other cluster contains 68.36% of samples, and the cluster center is 571.9, indicating that these students spent an average of 571.9 in the month from October 24 to November 24. The last cluster contains 18.19% of the samples, and the corresponding cluster center is 736.7, indicating that the average consumption of such students in the month from October 24 to November 24 is about 736.7 yuan, which is a high level of consumption.

According to the result of clustering, the student consumption amount is converted into the corresponding consumption level. Part of the data is shown in Table 3.

Table 3. Consumption level of students

Card number	Total consumption (unit/yuan)
171915	Low
171927	Medium
171223	High
171583	Medium
170224	Medium
170012	High
170034	Medium
171399	High
171339	Low

Through the cluster analysis of the consumption of the 2019-level telecommunications students on campus all-in-one card, it can be found that 13.45% of the students have a low level of consumption. In the face of this situation, school administrators or relevant counselors can take the initiative to learn about some of the students with low consumption levels. For students who really need help, they will give priority to the allocation of student grants and work-study job opportunities. With the support of real data, the school can make relevant decisions efficiently and accurately, so that students can get the help they deserve truthfully and effectively. This not only materially meets the nutritional needs of students with low consumption levels, but also allows students to get mental and physical exercise, so that students have a healthy body and good mental state.

3.2 Analysis of students' breakfast consumption behavior

As the first meal of three meals a day, breakfast can not only protect your health, but also enhance your physical fitness. Generally speaking, there is more than ten hours between breakfast and the dinner of the previous day. During this period, the food in the stomach gradually decreases, and it is finally in an empty stomach state. If there is no breakfast to supplement food, the blood sugar in the body will continue to drop, which will lead to a decline in physical fitness and lethargy. A large number of studies at home and abroad have found that whether or not to eat breakfast has a little effect on the health of students. Skipping breakfast will increase the probability of getting sick. On the contrary, a healthy breakfast can help students improve their physical fitness [9].

First of all, compared with lunch and dinner, the general breakfast time is between 6:20 and 8:00. Students who can insist on eating breakfast often have the habit of getting up early, their self-discipline is generally stronger, and their lives are more regular. Secondly, there are basically no takeaway merchants in the morning, and time is relatively tight. The possibility of students ordering takeout is very small. Students can only eat breakfast in the cafeteria or supermarket. Therefore, the dining data stored by the campus card is more comprehensive, which can be analyzed by analyzing the student's breakfast situation. Study students' eating patterns.

The school's morning class time is generally at 8:10 in the morning. Since the canteen and supermarket are nearly 15 minutes away from the classroom, students eat breakfast normally, and generally need to swipe their card to eat in the canteen and supermarket before 7:45. Therefore, before 7:45 in the morning It is normal for students who come to eat in the cafeteria or supermarket. Taking the card-swiping records of the students in the campus card data between 6:20 and 7:45 in the morning

as the basis for the normal meals of the students, the statistics of the students' breakfast conditions are shown in Figure 2.

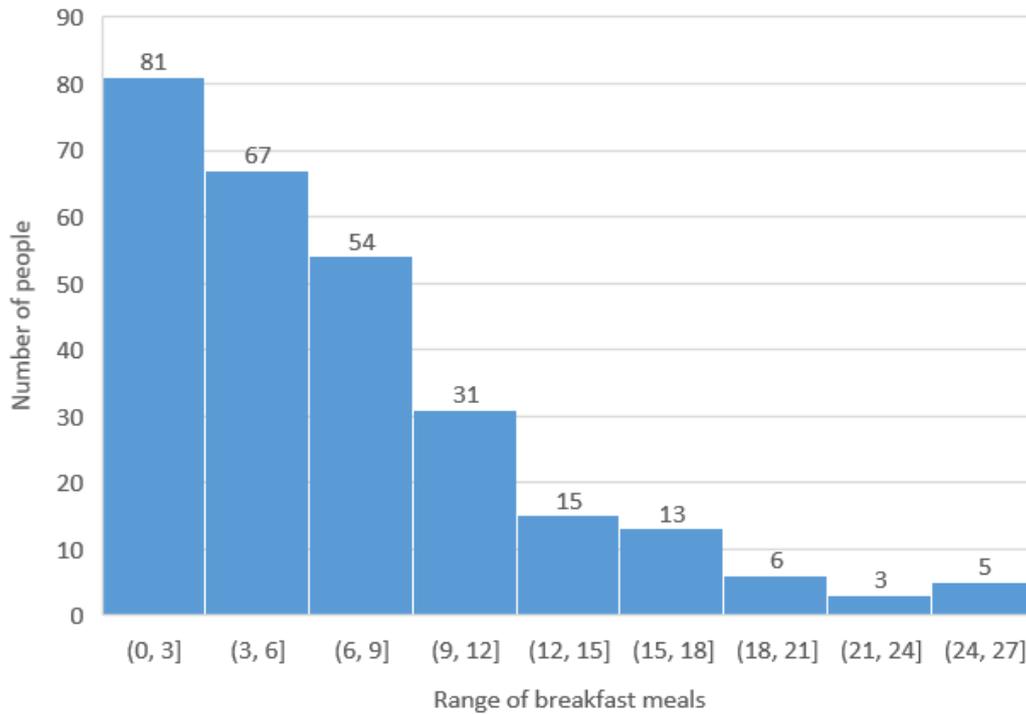


Figure 2. Distribution of average monthly breakfast times for students

Since many students rarely eat breakfast on weekends, except for the weekend, there are 22 days left. Considering that students may have no classes in the morning from one to two days a week, this paper uses the K-means algorithm to divide the number of breakfasts for students into two. The grade is the breakfast regularity and the breakfast irregularity. Therefore, we need to set the value of the number of cluster centers k to 2. After experimental research, it is found that when the random seed is set to 5, the error square and SSE in the cluster are the smallest, that is, the effect is the best. At this time, the obtained cluster center is 8, 17, the clustering effect is shown in Table 4.

Table 4. Cluster centers and number of cases

Breakfast times	Cluster	
	1	2
Cluster center	8	17
Number of cases in cluster	248	27

Therefore, when a student eats breakfast no less than 17 times a month, it is considered that the student eats breakfast regularly. As shown in Figure 3, the average number of times of having breakfast in a month is 17 as the threshold, the students are divided into two categories, one is regular breakfast, and the other makes breakfast irregular.

It can be seen from Figure 4 that most of the students majoring in telecommunications at the 2019 level have irregular breakfasts. The relevant managers should focus on understanding the daily habits of the students and take measures to guide the students to develop healthy work and rest habits and enhance their physical fitness.

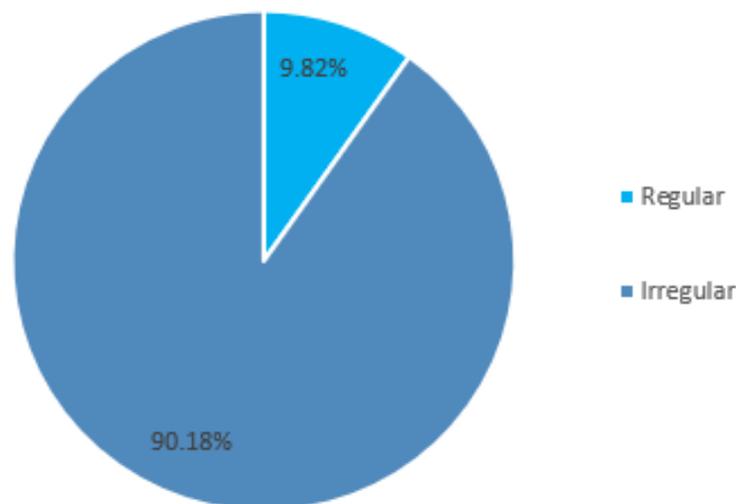


Figure 3. The proportions of the two types of students are shown in the figure

4. Conclusion

This article first introduced the concept and classification of cluster analysis, and then described the concept, characteristics, advantages and disadvantages of the K-means algorithm in detail. Taking student card consumption data as the research object, the consumption data is preprocessed through cleaning, integration, conversion, etc., and the original data is converted into statistical data containing student behavior characteristic fields. Combined with data mining technology, the K-means algorithm is used for students. The consumption amount and breakfast consumption are analyzed to obtain the behavior characteristics and consumption characteristics of different types of students. The analysis results scientifically and reasonably reflect students' consumption levels and breakfast conditions. School administrators can easily classify students through intuitive analysis graphs, effectively help and manage students, and improve campus intelligent management; students can also analyze. As a result, I understand my own rules of life and improve my daily routine, so that I have a healthy body to face the colorful life of the university and fulfill my studies.

References

- [1] Xiaodan Fei, Xinke Dong, Hui Zhang. Analysis of Poor College Students Based on Campus Card Consumption Data[J]. Computer Knowledge and Technology, 2014, 10(20): 4934-4936.
- [2] Guo Cheng. Analysis and application of campus card consumption data based on data mining [D]. Wuhan University of Light Industry, 2017.
- [3] Jing Guo, Su Hou. K-means algorithm optimal clustering number evaluation index research [J]. Software Guide, 2017, 16(11): 5-8.
- [4] Jianbing Chen. Using campus card data to optimize the identification system of poor students in universities [D]. University of Electronic Science and Technology of China, 2012.
- [5] Nan Jiang, Weisheng Xu. Analysis of student campus consumption behavior based on data mining technology[J]. Popular Science and Technology, 2015(1): 26~28.
- [6] Dan Zheng, Qianping Wang. K-means initial clustering center selection algorithm [J]. Computer Applications, 2012, 32(8): 2186~2188,2192.
- [7] Meiling Liu, Xi Li, Yongsheng Li. The application of data mining technology in teaching and management of colleges and universities [J]. Computer Engineering and Design, 2010, 31(05):1130~1133.
- [8] Decai Wang. Research and application of data mining in the analysis of campus card consumption behavior [D]. Harbin Engineering University, 2010.
- [9] Huaqun Luo. Mining and Application of Campus Card Data [D]. East China Normal University, 2009.