Adult and Higher Education (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/aduhe.2023.051005
ISSN 2523-5826 Vol. 5 Num. 10

Study on Student Dormitory Allocation

Rongjie Zhu*, Xinyue Zhang

Institute of Problem Solving, North China Electric Power University, Baoding, Hebei, China
**Corresponding author*

Keywords: Questionnaire survey, dormitory allocation for freshmen, K-means clustering algorithm, SPSS analysis

Abstract: Student dormitory is the main place for college students' activities, and the interaction between students in the dormitory plays an important role in the growth and development of college students. At present, most colleges and universities use the traditional dormitory allocation system, and some colleges and universities allocate dormitories according to the order of reports. These allocation methods are unscientific. Reasonable dormitory allocation for freshmen is of great significance to students' life and study. In this paper, the K-means clustering algorithm is used as the basic algorithm. During the experiment, the freshmen's living habits data provided by the school questionnaire are used for clustering, and the maximum impact factor determined by the SPSS online system analysis is used as the primary evaluation index. Combined with the obtained experimental data, it is shown that the application of this method has a better division effect, which can improve the rationality and scientificity of the dormitory allocation of freshmen.

1. Introduction

It shows the importance of rationality and scientificity, which is of great significance to freshmen who have just entered university.

In order to achieve a reasonable dormitory allocation model, it is first necessary to solve how to reasonably classify the massive freshman data. In fact, the largest impact factor determined by SPSS online system analysis is used as an evaluation index to measure the similarity of data of different freshmen's living habits and classify the data into different categories. This requires a scientific classification based on the characteristics of the relevant data of freshmen to ensure that freshmen in the same category have the same or similar living habits, personality traits, religious beliefs, etc. At the same time, the classification results should meet the requirements of high separation among different categories and high closeness within the same category [1].

At present, commonly used data analysis and classification techniques include heterogeneous analysis, association analysis, cluster analysis and evolution analysis, and there are some improved analysis methods based on these analysis methods. Among these methods, cluster analysis is highly efficient in clustering analysis of large-scale data sets, and the clustering results are independent of the input order, and it has been widely used in data analysis. A good visual representation of the classification effect [2].

2. Research Basis

Cluster analysis is an analysis process of grouping a collection of physical or abstract objects into multiple classes composed of similar objects, and is an important human behavior [3]. Cluster analysis aims to collect data on the basis of similarity for classification by classifying things with the same or similar nature into the same category and things with large differences in nature into different categories. Therefore, clustering-based methods have been widely used in many fields [4].

In the traditional K-means clustering algorithm, you first need to pay attention to the value of K, because the value of K directly determines the quality of the clustering effect. Generally speaking, the traditional K-means clustering algorithm determines the K value based on prior analysis and experience of the data set. When there is a lack of prior relevant knowledge, cross-validation can be used to determine the K value. After determining the number K of clusters, it is necessary to select K initial centroids. Since the position of the centroid will have a great influence on the clustering result and time complexity, it is necessary to select appropriate K centroids. Following are the steps of K-means clustering algorithm [5].

Suppose a set of sample data is given $D = \{X_i, X_2, \ , X_m\}$, the number of clusters is K , and the maximum number of iterations is N. That is, each sample data belongs to the same data set and the corresponding value is entered manually. The output is the corresponding cluster partition$C = \{C_1, C_2, \ , C_k\}$

(1)From the data set D, select K samples as the initial K centroid vectors: $Q = \{\mu_1, \mu_2, \ , \mu_k\}$.

(2)Use the enumeration method for each n=1,2,,,N ,

A. Divide the cluster into sets C, and initialize as: $C_t = \emptyset \ \ t = 1,2, \ , k$.

B. Use the formula: $d_{ij} = ||X_i - \mu_k||_2^2$, to calculate $d_{ij}$the distance between the sample and each centroid vector. $\mu_j(j = 1,2, \ , k)$The category corresponding to $\lambda_i$the smallest record d_{ij}will be marked X_i, and the value of the cluster under the corresponding $C_{\lambda i} = C_{\lambda i} \cup \{X_i\}$i and j will be updated , such that For $j = 1,2, \ m$, count $C_i$all sample points in the sample and recalculate the centroid $\mu_i = \frac{1}{|C_i|}\sum_{X \in C_i} X$.

C. This process judges whether the K centroids have changed. If all K centroid vectors have not changed, go to step (3).

(3)The output is the corresponding cluster partition $C = \{C_1, C_2, \ , C_k\}$.

3. Research Methods

In this article, in order to find out the center of each cluster more easily, that is what we think of as the dormitory category. We define the number of students as several elements. Of course, the number of students will not be a huge value, because factors such as gender, major and even class are considered. By repeating the K-means clustering algorithm until the K centroids do not change, we can get the final bedroom data we want [6].

3.1. Data Source

The data studied in this paper come from the number of people who participated in the questionnaire survey designed in this experiment. Combined with the actual situation, the contents of the questionnaire are all closely related to students' life and study, including: student number, gender, place of origin, wake-up time and other related information, and some contents have different options [7].

3.2. Data Cleaning

Because some of the users who participated in filling out the questionnaire did not fill in all of them, the data obtained in the experiment was incomplete, and of course it was easy to lead to the appearance of some irrelevant data. Therefore, based on the characteristics and actual requirements of the data, the obtained data is preprocessed in this paper, and this part of incomplete data is eliminated, and only the data filled in by the user is selected to ensure the authenticity, scientificity, and integrity of the data source, reliability. Indirectly ensure that the experimental results are convincing [8].

3.3. Data Analysis

Before designing the questionnaire, we have fully understood what conditions students are most concerned about, and set the weight of this item artificially to reflect its importance. For example, most students care about the time of getting up, so we can divide the time of getting up into five stages at intervals of one hour, and assign different weights to each of the five stages. The purpose of allocating dormitories is to take diversity into account while ensuring identity. We know that SPSS is a commonly used statistical data analysis software, widely used in data analysis, decision analysis, precision marketing, etc. Since we used weights to replace options in the process of designing the questionnaire, we used the SPSS online system to analyze the degree of commonality of the data, so that the factors with greater influence were considered as the primary evaluation indicators [9].

4. Research Results and Discussion

Table 1: SPSS analysis results.

project	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	common degree
Question 2: What is your gender?	-0.01	0.76	-0.25	0.13	0.01	0.657
Question 3: Your Place of Origin	0.30	-0.11	-0.17	-0.10	0.59	0.495
Question 4: When do you usually get up in your study life?	0.11	0.17	-0.49	0.14	-0.55	0.606
Question 5: Do you usually have the habit of taking a nap?	0.14	-0.03	0.03	0.57	0.04	0.348
Question 6: When do you usually go to bed at night?	0.17	0.10	-0.11	0.22	0.52	0.375
Question 7: What is your personality traits?	-0.10	0.02	0.07	0.71	-0.01	0.513
Question 9: What are your hobbies?	0.03	-0.03	0.75	0.18	-0.17	0.625
Question 10: Do you like to turn on the air conditioner?	0.16	0.60	0.34	-0.26	0.13	0.589
Question 11: After entering university, what are your expectations for your university studiesyes?	0.65	-0.28	-0.11	-0.03	0.30	0.603
Question 12: To enter university, what are your requirements for your academic performance?	0.58	0.11	-0.07	0.26	0.13	0.440
Question 13: Do you intend to study further when entering university?	0.65	0.26	0.26	-0.15	0.04	0.578

At this time, from the preliminary data analysis, it can be found that among the items that students participated in filling out, the commonness of wake-up time is relatively high, and the commonness of sleep time is relatively low, indicating that most students have relatively the same

wake-up time and different sleep time. According to the principle of dormitory allocation, diversity is considered on the basis of identity, so these two items are used as the main evaluation indicators in the experiment as the basis for classification. As shown in Table 1.

The core idea of the algorithm is to find K cluster centers, so that the sum of the squares of each sample point and its nearest cluster center is minimized. In the case of meeting the research needs of this paper, we import the experimental data into the pre-programmed program, and constantly change the number of clusters K and the maximum number of iterations N to find the maximum clustering score. We found that when the number of initial clusters is 5 and the number of iterations is 100, the score of clustering is found to be the largest. It shows that when K=5, the effect of clustering should be the most appropriate, and the classification is also the most independent, and the best clustering effect can be obtained. Therefore, the number of clusters is selected as 5, that is, the dormitories that need to be allocated are 5 categories. As shown in Figure 1.

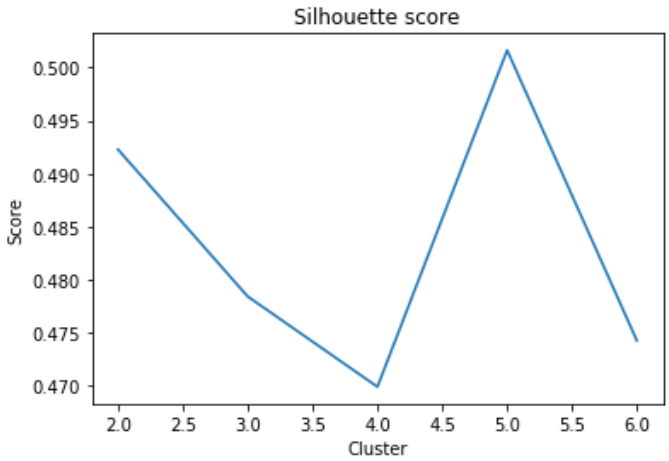

Figure 1: Cluster score map.

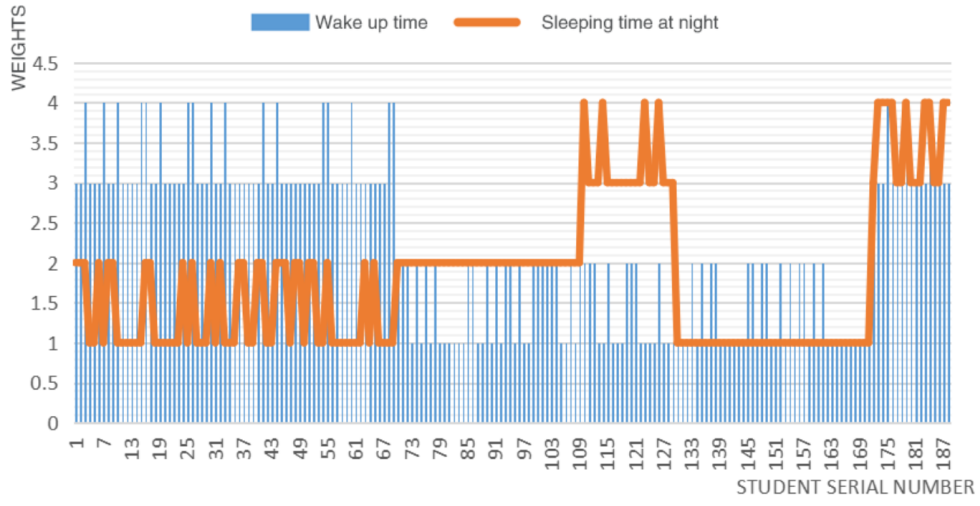

Figure 2: Experimental result table.

According to this method of division, we found that students who get up early and go to bed late at night should be assigned to the same dormitory as much as possible, students who get up early and go to bed early at night should be assigned to the same dormitory as much as possible, and students who get up early and go to bed at night Students with earlier time should be assigned to the same dormitory as much as possible, students with late wake-up time and early bedtime should be

assigned to the same dormitory as much as possible, students with late wake-up time and late bedtime should be assigned to the same dormitory as much as possible [10]. While considering the two main influencing factors, other factors such as personality characteristics, hobbies, etc. are used as secondary influencing factors to allocate dormitories. Such allocation will be more conducive to the harmony of the dormitory atmosphere and the growth of students. As shown in Figure 2.

5. K-means and Other Clustering Algorithms

At present, the commonly used clustering algorithms include: partition method, hierarchical method, density algorithm, graph theory clustering method, etc., and there are some improved algorithms based on these algorithms. Among them, when clustering a large amount of data, the partition method is widely used in the field of data analysis because of its good clustering effect and efficient data processing. And thus derived many efficient algorithms, such as K-means clustering algorithm, K- medoids algorithm, Clarans algorithm, etc. Below is a comparison of some other clustering algorithms with the K-means clustering algorithm.

The key to using hierarchical clustering algorithm is that a large number of objects and clusters need to be inspected and estimated during the clustering process, and this clustering method does not have good scalability. The complexity of the algorithm is$O(n^2)$, it is obviously not suitable for the case where n is very large. In contrast: The complexity $O(n)$of K-means clustering calculation is linear. Although the experimental data in this experiment is not huge, it is undoubtedly better to choose the K-means clustering algorithm in terms of computer performance.

Compared with the SOM clustering algorithm, the K-means clustering algorithm needs to determine the number of classes in advance, that is, the K value. In this experiment, since the number of students participating in filling out the questionnaire is not huge, the problem of finding the K value can be solved by using loop statements in programming. Secondly, after K-means finds the most similar class for each input data, only the parameters of this class are updated, so the accuracy of the results obtained by the K-means clustering algorithm is higher than that of the SOM clustering algorithm. On this basis, we believe that it is a better choice to choose the K-means clustering algorithm for dormitory allocation.

Compared with the K-means clustering algorithm, the biggest disadvantage of the FCM clustering algorithm is that when processing each text to be classified, it must calculate its distance to all known samples, so as to obtain its K nearest neighbors. This increases the computational load of the program, and the computer takes too long to run. Another disadvantage is that the algorithm calculates "imbalance". That is, when the sample size of one class is large, while the sample size of other classes is small, it may cause that when a new sample is input, the samples of the large-capacity class among the K neighbors of the sample account for the majority. When the K-means clustering algorithm processes student data, the running time of the program will be shorter due to the fewer iterative updates. In addition, the K-means clustering algorithm counts all sample points and recalculates the centroid every cycle, which well avoids the shortcomings of the FCM clustering algorithm. Therefore, we choose the K-means clustering algorithm.

6. Conclusion

At present, most colleges and universities use the traditional dormitory allocation system, and some colleges and universities allocate dormitories according to the order of reports. These allocation methods are somewhat unscientific. Reports about incidents in college dormitories also confirm the irrationality of traditional dormitory allocation, such as the Fudan University poisoning case. The traditional dormitory allocation is done mechanically according to class division, and cannot be screened by region. Moreover, the dormitory allocation based on initial letters does not

take into account the characteristics, living habits, religious beliefs and other factors of different students, which is not scientific enough, resulting in The event that conflicts with the problem occurs. Therefore, this paper uses the combination of K-means clustering algorithm and questionnaire survey to allocate dormitories, which improves the rationality and scientificity of dormitory allocation for freshmen.

References

[1] Sun Huiting, Ma Jian. Research on Intelligent Dormitory Allocation in Universities Based on K-means Algorithm. Computer and Telecommunications, 2019 (5): 4.

[2] Zhao Ying, Zhai Yuanwei, Chen Junjun, et al. Research on Student Dormitory Allocation Based on Constrained K-means Clustering Algorithm. The 23rd Network New Technology and Application Year of China Computer Users Association Network Application Branch in 2019 (6).

[3] Gu Zhengqing. Application of K-means algorithm in intelligent dormitory allocation of college students. Digital User, 2019.

[4] Xie Jingxuan, Xu Hongda, Qin Li, et al. The problem of dormitory allocation based on K-means algorithm. Modern Computer, 2021 (21): 4.

[5] Liu Guocheng, Zhang Yang, Huang Jianhua, et al. Image segmentation and recognition of spider mites based on K-means clustering algorithm. Acta Entomology, 2015 (12): 1338-1343.

[6] Lu Minglei, Liu Dongmei, Zeng Zhiyong, et al. Image Retrieval Method Based on Improved K-means Algorithm. Computer Applications, 2013, 33 (A01): 4.

[7] Zhang Shiyong. Research on K-means Algorithm and Parallelization Based on Hybrid PSO. Chongqing University. 2019 (21): 4.

[8] Zhu Zhengguo. Application of K-means Algorithm to Analyze Campus Network User Behavior. Journal of Anqing Normal University: Natural Science Edition, 2017, 23 (1): 4.

[9] Zhao Ying, Zhai Yuanwei, Chen Junjun, et al. Research on Student Dormitory Allocation Based on Constrained K-means Clustering Algorithm. The 23rd Network New Technology and Application Year of the Network Application Branch of China Computer Users Association in 2019 (10): 4.

[10] Zhang Qin. Research on clustering problems based on k-means algorithm. Digital Design. CG WORLD, 2020, 009 (008): P. 182-182.