International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:12, No:9, 2018

Analysis of Cooperative Learning Behavior Based on the Data of Students' Movement

Wang Lin, Li Zhiqiang

***Abstract*—**The purpose of this paper is to analyze the cooperative learning behavior pattern based on the data of students' movement. The study firstly reviewed the cooperative learning theory and its research status, and briefly introduced the k-means clustering algorithm. Then, it used clustering algorithm and mathematical statistics theory to analyze the activity rhythm of individual student and groups in different functional areas, according to the movement data provided by 10 first-year graduate students. It also focused on the analysis of students' behavior in the learning area and explored the law of cooperative learning behavior. The research result showed that the cooperative learning behavior analysis method based on movement data proposed in this paper is feasible. From the results of data analysis, the characteristics of behavior of students and their cooperative learning behavior patterns could be found.

***Keywords*—**Behavior pattern, cooperative learning, data analyze, K-means clustering algorithm

I. Introduction

COOPERATIVE learning has risen in the United States since the 1970s. It has become a widely used classroom teaching model and teaching strategy. Cooperative learning refers to the grouping of students according to the principle of "heterogeneity within groups, homogeneity between groups" on the basis of group teaching and individual student learning. Through the activity of cooperative learning, students accomplish individual and group tasks, then achieve their academic goals and collaborative skills goals [9]. A large number of researches show that, compared with individual independent learning and competitive learning, cooperative learning includes not only teacher-student interaction, but also the interaction between teachers and the interaction between students. The creative use of interaction between students is one of the reasons for the success of cooperative learning [8]. Relevant literatures show that college students' active cooperative learning can break through the limitations of simple individual cognition and behavioral ability. The more active they are to participate in group discussion and cooperative inquiry, especially adopt the methods of questioning, argument and consultation, the better their learning effect will be and the more they will gain from their studies [4]. For students living in the context of globalization, they must have a sense of cooperation, learn cooperate skills, master communication skills, and establish team-work spirit [2].

Traditional cooperative learning is based on the disciplines of organizational behavior and psychology, and mostly uses qualitative and subjective quantitative analysis methods, such as interview, questionnaire, examination sheet, etc. Then, the behaviors, consciousness, psychological changes, learning effects and other aspects of individual and groups are studied by the method of combination of theoretical analysis and statistical analysis. These methods can reflect objective facts to a certain extent, but they are easy to be influenced by subjective factors. Therefore, the analysis of students' cooperative learning behavior is not detailed and objective [1], [7].

With the rapid development of information, science and technology, students' cooperative learning behavior can be explored from the perspective of data analysis. As a kind of behavior, cooperative learning can be studied from many aspects, such as participants' learning behaviors, communication behaviors, sports behaviors. Based on the students' actual movement data, this paper explores the cooperative learning behavior through the analysis of the data [5], [6].

II. Basic Definitions

In this paper, the K-means clustering algorithm is used to analyze the students' movement data. Through the analysis of the characteristics of student's activity, the behaviors with similar characteristics are classified into one category, so as to identify students' cooperative learning behavior. The K-means algorithm uses distance as a similarity evaluation index. It is believed that the smaller the distance between two objects, the greater the similarity [10]. Suppose that there is a data set of size N, the algorithm's framework is:

First, the number of clustering centers is given as k, and select k points randomly as the initial clustering center Zj(I), where I=1, j=1,2,3,…,k;

Then, calculate the distance D(xi, Zj(I)) between each object and the clustering center separately, where i=1,2,3,…,N, and j=1,2,3,…,k;

If D(xi, Zm(I))=min{D(xi, Zj(I)), j=1,2,3,…,k, m∈(1,k)}, then xi should be divided into the class whose clustering center is Zm(I). Place each object into the correct class according to the above rules.

After dividing all the objects, the sum of squared errors criterion for each class can be calculated: $J_c(I) = \sum_{j=1}^{k}\sum_{u=1}^{n_j}\left\|x_u^{(j)} - Z_j(I)\right\|^2$, where nj is the number of objects included by class j, and $\sum_{j=1}^{k} n_j = N$, $x_u^{(j)}$ is an object which belongs to class j;

F. Wang Lin is with the School of Reliability and Engineering, Beihang University, Beijing, China (phone: +86 13466697531; e-mail: nilgnaw@126.com).

S. Li Zhiqiang is with the School of Reliability and Engineering, Beihang University, Beijing, China (e-mail: 1428206712@qq.com).

Finally, one needs to judge whether the criterion Jc(I) is converged. That is, for a given value δ, if $|J_c(I) - J_c(I-1)| < \delta$, the algorithm comes to the end, otherwise I=I+1, $Z_j(I) = \frac{1}{n}\sum_{i=1}^{n_j} x_i^{(j)}$, j=1,2,3,…,k, the clustering centers and the distance D have to be recalculate [3].

The above algorithm can be implemented by the help of R software. It is a free software that integrates statistical analysis and graphic display. It can run on multiple operating systems and provides variety of data analysis tools. What's more, its syntax is easy to understand and easy to use.

Before performing cluster analysis, the original data need to be cleaned. Data cleaning includes the handling of quantity and quality. Its purpose is to find and correct the invalid and missing data, remove irrelevant data, and ensure the consistency of data. Therefore, the role of data cleaning is crucial in the entire data analysis process, and it is also a time-consuming step in the whole process. If data cleaning is appropriate and reasonable, the results of the analysis will be more realistic.

III. Analysis of Students' Behavior Patterns

The data used in this analysis is the trajectory of daily activities provided by 10 first-year graduate students (from November to December, 2016). It was recorded by GPS software. The sampling interval was set as 1 s. The collected data included basic information such as latitude, longitude, altitude, and time.

After extracting and merging all the data from files of different formats, certain rules needed to be made to clean the data. The main problems of raw data were: discontinuity of recording time, unstable sampling intervals, and coordinate error caused by signal drift. Aiming at these problems, the data were firstly segmented according to the duration of the record. When the time interval between two recording points was considered to be too long, they were thought to belong to different record segments. The segments with short durations could be deleted because they could not reflect the purpose of students. Finally, the data with obvious changes on the coordinate within a short time could be eliminated.

A. Initial Cluster Analysis

The cleaned data mainly included information on longitude, latitude, date, time, and students' number. A cluster analysis was done to the data of all people at all times according to the geographical location, and it pointed out the main activity areas of the students. Suppose that there were 10 clustering centers, the k-means algorithm was used. With the help of R software, the clustering result was plotted, as shown in Fig. 1. The abscissa and ordinate of the graph are longitude and latitude. The number 0-9 are located at the centers of each cluster, points have different colors because they belong to different clustering centers.

The position coordinates of the clustering centers were shown in Table I. The location names were given by Google Earth software according to latitude and longitude. The quantity of members of each class and the quadratic sum of distance were also included in this table.

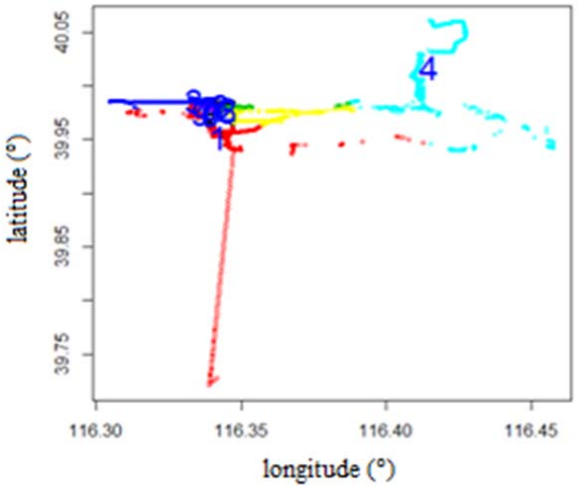

Fig. 1 Initial cluster analysis results

TABLE I
Clustering Result

No.	Position	Location name	Quantity of members	Quadratic sum of distance
0	39.97422,116.3396	Zhichun Road Hotel	19749	0.02690
1	39.95230,116.3422	Beijing Jiaotong University	18805	2.03222
2	39.97900,116.3450	Beihang New Main Building G	19627	0.05526
3	39.98687,116.3337	JiXin Building	15098	0.34864
4	40.01768,116.4149	Beiwei Homeland	19607	14.17503
5	39.97924,116.3402	Beihang Stadium	110850	0.40191
6	39.97668,116.3451	Beihang Weishiyuan	46063	0.79295
7	39.97693,116.3382	Dayun Village Apartment 3#	127760	0.10374
8	39.98170,116.3430	Beijing Qiushi Plaza	37001	0.09349
9	39.97524,116.3358	Zhichun Road Subway Station	236503	0.16133

From Table I, we can see that the quantity of members in class 5, 7, and 9 is significantly more than that of the other classes, indicating that the students often stayed in these areas. The quadratic sum of distance of class 1 and 4 is much higher than the other classes, which means that these classes have high dispersion. These cluster centers cannot represent these points well. Based on the above analysis, the scope of geographic location can be further reduced to the area near the school.

B. Analysis of Each Activity Area

Cluster analysis was carried out again to deal with the data of the area near the school. The results showed that narrowing the range to the vicinity of the school and considering the influence factors such as data intensity, major function and ease of division, it was reasonable to further divide this region into 13 areas. The relevant information of these areas, including the quantity of members, the number of participants, the number of participating days, was shown in Table II. The percentage of members is the ratio of the quantity of this region to all regions. The percentage of participating days is the ratio of the days to the maximum value of that column. At the same time, it was considered that the contributions of these three factors to the regional popularity were the same, that is, the weights were all 1, then the popularity index and the ranking of each region were obtained.

TABLE II
IMFORMATION OF REGIONAL DATA

No.	Location name	Quantity and percentage of members		Number of participants	Number and percentage of participating days		Popularity index	Rank
1	Academy Building	54193	8.88%	9	5	68.63%	1.68	5
2	Beihang New Main Building	32149	5.27%	10	2	92.16%	1.97	2
3	Beihang Main M Building	17446	2.86%	9	8	47.06%	1.40	8
4	Beihang Library	11172	1.83%	9	6	66.67%	1.58	6
5	Beihang Green Park	3107	0.51%	7	12	25.49%	0.96	12
6	Beihang Second Canteen	9037	1.48%	9	7	64.71%	1.56	7
7	Beihang First Canteen	26001	4.26%	10	4	68.63%	1.73	4
8	Beihang Stadium	33781	5.53%	10	3	82.35%	1.88	3
9	Dayun Village Apartment	271376	44.45%	10	1	100.00%	2.44	1
10	Zhichun Road Hotel	98709	16.17%	7	9	47.06%	1.33	9
11	Zhichun Road Subway Station	36560	5.99%	6	11	33.33%	0.99	11
12	Zhanchunqiao North	13311	2.18%	2	13	9.80%	0.32	13
13	Beihang Northwest Gate	3720	0.61%	7	10	52.94%	1.24	10

According to the results of the analysis above, the most frequently visited places for students were 9 Dayun Village Apartment, followed by 2 Beihang New Main Building, 8 Beihang Stadium, and 7 Beihang First Canteen. Fig. 2 showed the ratio of popularity of different regions. Based on the assumption that the importance of all factors was equal, some areas with small amounts of members were much more popular than the areas with larger amounts of members, such as areas 8 and 10. This was related to the size and nature of the region.

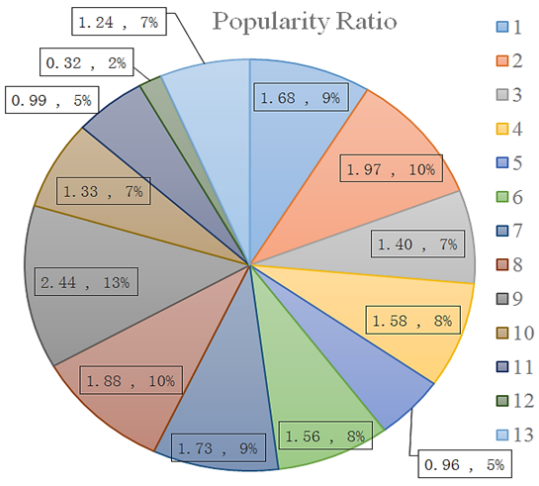

Fig. 2 Regional popularity

C. Analysis of Students' Individual Behavior Patterns

Among the 13 areas divided in the previous section, 1-4 belonged to the teaching area, where area 1 is the seat of the academy, areas 2 and 3 are teaching buildings, and area 4 is a library. Detailed analysis was performed for these areas to understand students' individual learning behavior patterns. The data in these areas were statistically analyzed in time dimension, and the scatter plot was plotted as shown in Fig. 3. The abscissa is time, which ranges from 0:00 to 24:00, and the ordinate is the frequency of the data record at that moment. The distribution of movement data in each region can be clearly seen from the figure.

The academy building is the location of the students' labs and can be accessed within 24 hours. In Fig. 3 (a), the data frequency started to increase greatly around 9:00, and the data volume continued to be large between 11:00 and 23:00. The points were especially dense in the evening, and records occasionally appeared in the early morning. The Beihang new main building is composed of eight teaching buildings. There are many classrooms, new facilities, good environment and geographical location. Many courses are located here. It can be seen from Fig. 3 (b) that the frequency suddenly increased significantly at the time points of 10:00, 12:00, 14:00 and 23:00, and the amount of points during the day was more than that in the evening. Fig. 3 (c) shows that the points in this area were not so much, and only at 10:00, 12:00, and 18:00，slightly more records appeared.

From the analysis above, we can speculate on the pattern of students' learning behaviors. In the absence of mandatory requirements, students are more inclined to start studying after 9:00. The points in Fig. 3 (b) where the frequency had increased sharply, were just the time between classes and at the end of self-study. The decrease of the amount of data between these time points, maybe that related to GPS dormancy or poor GPS

signal caused by no displacement for a long time. The amount of data was large in the area 1 at night, and in the area 2 during the day. It means that after the end of the daytime class, students would like to go to the lab to learn professional knowledge or self-study in the evening, and some students even study until midnight. The data in the library area appeared before and after the class. It might be left by the way to the classroom, and it did not reflect the real learning behavior.

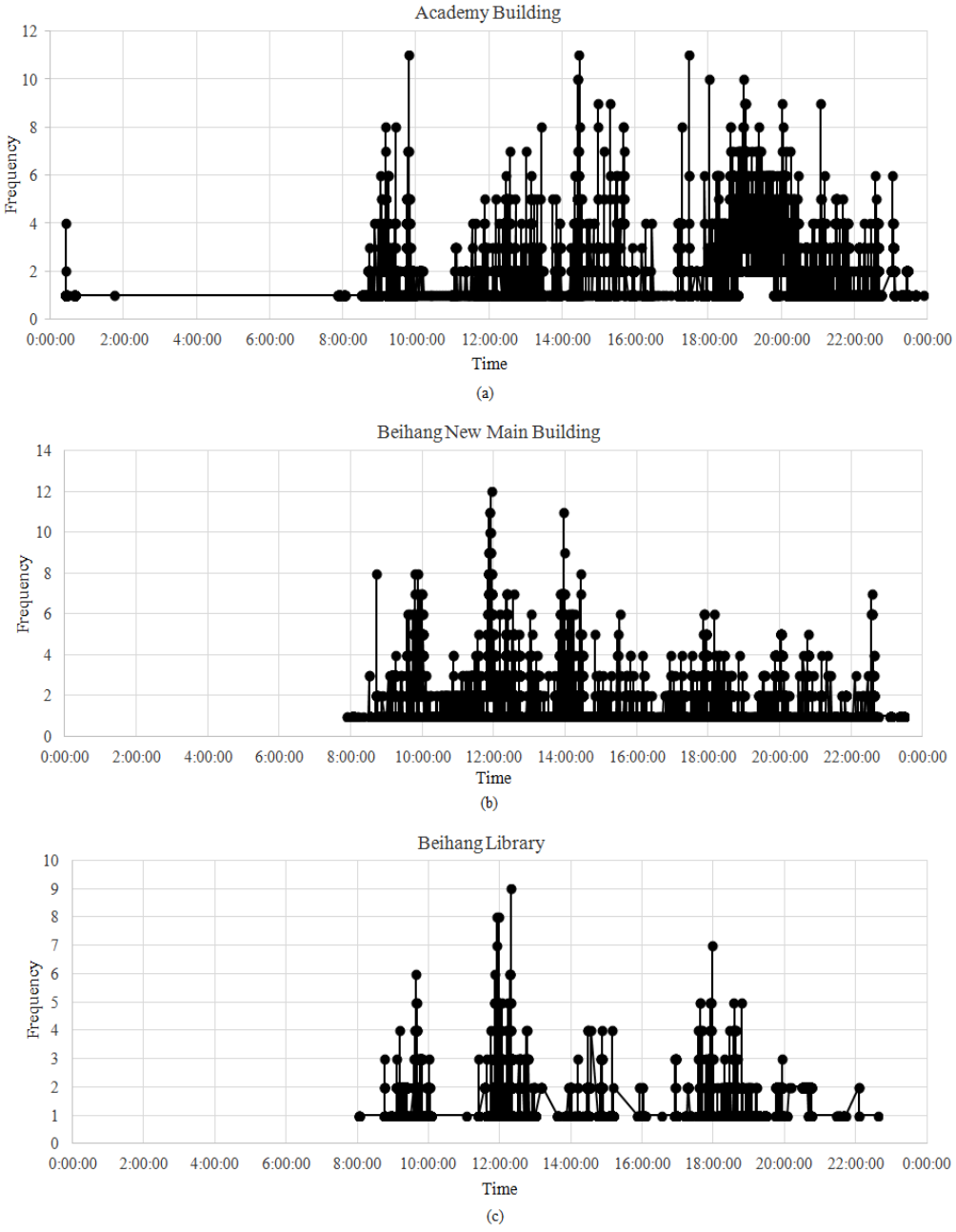

Fig. 3 Visiting time of teaching area

D.Analysis of Cooperative Learning Behavior Patterns

Through the data analysis, it was found that some classmates often appeared in the same geographical location at the same time. Taking one group of students as an example, they had joint activity records in 19 days, their footprints covered all of the 13 areas, and 49% of the total duration of joint activities was in the teaching area. Apart from the teaching area, the two students also often appeared in the area 9 and 7 which are living areas. The time distribution of their coming in and out of the teaching and living areas on different dates was shown in Fig. 4. The abscissa represents time, and the ordinate represents date. The teaching areas include area 1-4, and the living areas include area 7-9.

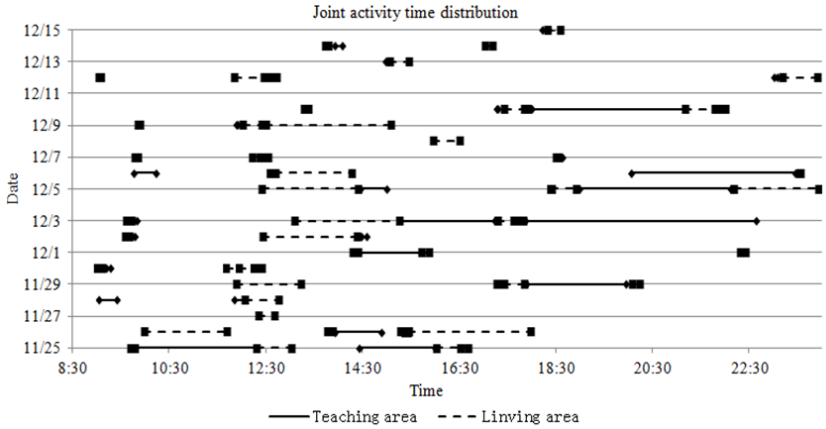

Fig. 4 Joint activity time distribution

When the students had joint activity, the time spent in the above two types of areas accounted for more than 90% of the total time. Nearly half of the time was spent in the teaching area, indicating that there was cooperative learning behavior among the students, and this was their main content of joint activity. It can be speculated that cooperative learning is a kind of behavior pattern which is conducive to improving the enthusiasm of the students on studying. As can be seen from Fig. 4, most of the beginning (termination) points of time period of the teaching area coincided with the termination (beginning) point of time period of living area. This suggests that, before and after the cooperative learning, students generally had other joint activities, such eating or returning to the dormitory together. This kind of activities promoted the communication between the students, and also reflected the basic idea of paying attention to the interaction between students in cooperative learning. Meanwhile, it also realized both academic and social goals.

IV. Conclusion

The analysis showed that it is indeed possible to study students' cooperative learning behavior from the perspective of data analysis. After the raw data were cleaned, it could provide a lot of effective information. It can be used not only to analyze the cooperative learning behavior of the students, but also to understand their behavior patterns; that points out the direction of providing a better learning and living environment for students.

Through the processing of the data in the above 13 regions, it was found that the regions 9 and 2 were frequent places for students, and they respectively corresponded to the living area and the teaching area. So, the fact that the facilities of these two areas are perfect and humanized will greatly influence students' study and life. At the same time, it was also found that although the area 8 contained stadiums, and it was very popular too; however, the track records were mostly left when passing through this area. Few physical exercise behaviors occurred. This is a noteworthy problem, and measures should be taken to guide students to strengthen exercise while studying.

The earliest record in the teaching area was at around 8:00, and the latest record was at around 23:00. This showed that the students were very diligent. After finishing the course in the daytime, they kept on learning in the evening. However, the first peak time in the teaching area was after 9:00, which means that for most of the students, it may be difficult to attend classes at 8:00 in the morning.

The data of the joint activities showed that there were indeed cooperative learning behaviors among the students. The students who worked together spent nearly half of the time in the teaching area, indicating that cooperative learning is conducive to improving the enthusiasm of the students, and there will be some other joint activities before and after the cooperative learning. This expanded the scope of cooperation and enhanced communication between students.

This paper was based on the data records of 10 students, and some records were not continuous. Therefore, it is necessary to obtain a larger amount of data to get more realistic and effective conclusions.

References

[1] Liu Yanfei. Cooperative Learning Community Research From the Perspective of Organizational Behavior (D). Shandong: Shandong Normal University, 2016.
[2] Chen Xiangming. Group Cooperative Learning Conditions (J). Tsinghua Journal of Education, 2003, 24(4):11-16.
[3] Zhu Ming. Data Mining-2nd Edition (M). Hefei: University of Science and Technology of China Press, 2008.
[4] Zhang Ping, Shen Yanhua, An Jing. The Mediating Role of Professional Interest and Deep Learning in the Relationship between College Students' Active Cooperative Learning and Academic Harvest (J). Heilongjiang Research on Higher Education, 2018(2):94-98.
[5] Jiang Yongchao. Research on Students Course Selection and Learning Behavior Analysis Algorithm Based on Data Mining (J). Modern Electronics Technique, 2016, 39(13):145-148.
[6] Liu Yan, Geng Qiang, Tian Li. Data Research on College Students' Classroom Learning Behavior Based on Data Mining(J). E-Business Journal, 2017(11):90-91.
[7] Chen Guoan. Research on College Students' Teamwork Learning Behavior in Experimental Teaching- Based on Psychological Perspective (J). New Education Era Magazine: Teacher Edition, 2017(33).
[8] Wang Tan. The Basic Concept of Cooperation Learning (J). Educational Research, 2002(2):68-72.
[9] Ling Jizheng. Study on Learning Behavior in Students' Cooperative Learning (J). Journal of Henan College of Finance & Taxation, 2015, 29(5):61-64.
[10] Sun Haiyan, Zhou Meng, Li Weiguo, Feng Wei. Mathematical Statistics

(M). Beijing: Beihang University Press, 2016.