College Students' Learning Behavior from the Perspective of Deep Learning

Chen Jiao[a*] and Tianxingzi Gao[b]

School of Humanities and Education, Xi'an Eurasia University, Xi'an City, Shaanxi Province, China

[a]jiaochen@eurasia.edu, [b]gaotianxingzi@eurasia.edu

[*]corresponding author

Keywords: Deep Learning Horizon, Data Mining, K-Means Clustering Algorithm, Learning Behavior Analysis

Abstract: The purpose of deep learning is to cultivate high-level thinking ability and realize meaningful learning. Its core idea embodies the concepts in cognitive science such as criticism, understanding, integration, transfer, reflection and creation. Based on the perspective of deep learning, this paper analyzes and studies the learning behavior of college students. In the study, we selected 400 students of statistics major in 2018 and 2019 as the research object, and used K-means clustering algorithm in data mining to analyze the number of times students in and out of the library and the types of books borrowed, so as to analyze their learning behavior. This study found that 76.4% of the students in and out of the library were of medium or above in 2018, while 70% of the students in and out of the library in and out of the class of 2019 were medium or above, and the proportion of students borrowing professional books in 2018 was far greater than that in 2019. In addition, with the help of data mining and K-means clustering algorithm, this paper analyzes the academic performance of the two sessions of students from the perspective of deep students, from which we know that the more times we enter and exit the library, the higher the degree of learning effort, the better the academic performance.

1. Introduction

In recent years, deep learning has gradually become a research hotspot. Deep learning in computer science belongs to the category of machine learning, while the field of education science is deep learning related to shallow learning [1-2]. With the rapid development of information technology, more and more new technologies and emerging tools are used in research and practical teaching, and deep learning is one of the methods to make learning more targeted and effective [3-4]. Deep learning is a deep learning method based on reality and complex situations. It guides students to use knowledge (subject knowledge and interdisciplinary knowledge) to transfer to new scenes, use them on the basis of understanding, and use the learned knowledge and skills to solve practical problems, cultivate students' critical thinking, innovation ability, collaborative spirit and communication ability, and cognitive behavior in decision making [5-6].

Learning behavior analysis refers to the analysis of a large number of data generated by students in the learning process, through the discovery of students' behavior patterns, to master students' learning situation, and to guide students to learn better. At present, the main data sources of learning behavior analysis are: online training platform data, book borrowing data, consumption data, data degree and Internet data. [7-8]. In the context of deep learning, the analysis and research of students' learning behavior can timely understand the situation of students, predict the development of students' behavior, reasonably guide the development of teaching plan, teach according to students' ability, and guide students to better develop direction [9-10].

This paper analyzes and studies the learning behavior of college students from the perspective of deep learning. In the study, we first briefly described the data mining and K-means clustering algorithm, and used data mining and K-means clustering algorithm to select 400 students majoring in statistics as the research object for analysis. In the study, we analyzed the distribution of the

number of times of students entering and leaving the library and the types of books borrowed, analyzed the degree of students' learning efforts, and analyzed their academic performance through data mining and K-means clustering algorithm. This study found that the higher the degree of study effort, the better the results.

2. Data Mining and K-Means Clustering Algorithm

2.1 Data Mining

Data mining is to reveal hidden information from a large number of, incomplete, noisy, unclear, random, original and unknown data sets, to discover its potential value and finally to be understood. In data mining, we receive data from the data source according to the needs, carry out data integration, data cleaning, data conversion and data preprocessing protocols. We store the processed data in the data repository, and the data received by the data extraction engine is stored in the data repository. The processed data refers to the knowledge of knowledge-based regulatory standards and evaluates the knowledge. Finally, the results are displayed and presented in the desired form.

The main methods of data mining are as follows:

(1) Association analysis

Association analysis means that we should look for interesting indirect relationships between data, export data to data rules and obtain possible data to data relations.

(2) Forecast analysis

1) Regression analysis: analyzing the causal relationship between variables, identifying independent variables and dependent variables, and defining regression function according to the relationship between variables is essentially a functional evaluation problem. Replace the effective measurement data received in the function, and then solve the function to get the value of the dependent variable.

2) Classification: classification is an important data analysis model and a supervised learning process. Its purpose is to obtain a classification function (classifier) by learning the data of the training set, and map the data in the training set to the given category in.

(3) Cluster analysis

The unsupervised learning process divides the data into different categories. When the category of the data set is unknown, there must be some relationship between the data of the same category, and the data of different categories have great differences to classify the charges. Grouping is different from classification. When the category is unknown, the data is divided into categories.

(4) Outlier detection

In a sense, it is closely related to classes (clusters). Firstly, the data sets are grouped, and the distance between data objects and classes (commonly used euclidean distance) is calculated, and the relationship between extreme value comparison and threshold value is determined. Outliers usually belong to classes other than the main class (cluster), or data objects that do not belong to any class (cluster).

2.2 K-Means Clustering Algorithm

K-means clustering algorithm is also called K-means or K-means clustering algorithm. This algorithm is widely used in text classification, ocean, image processing, astronomy and many other fields. It is a well-known, typical and commonly used partition method. It is an unsupervised learning clustering algorithm. For each cluster of data, a central point can be selected, so that the distance between all the points in the cluster and the center point of other clusters is less than that to other clusters.

The implementation steps of K-means clustering analysis are as follows:

(1) The data set that needs to be clustered is $S=(x_1,x_2,\ldots,x_n)$, and the k clustering centers are $(c_1,c_2,\ldots,c_k)$.

The (X_i,X_j) euclidean distance of two individuals in K-means algorithm is the square root of

the sum of squares of the difference between the values of K variables of two individuals

$$d(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \ldots + (x_{ip} - x_{jp})^2} \tag{1}$$

(2) The average distance of all sample points is calculated as follows:

$$Meandist(S) = \frac{2}{n(n-1)} \times \sum_{i \neq j, i, j=1}^{n} d(x_i, x_j) \tag{2}$$

Where n is the total number of sample objects in the dataset, and $d(x_i, x_j)$ is the euclidean distance of sample points xi and xj.

(3) The objective function is the square error criterion function, and the formula is as follows:

$$E = \sum_{i=1}^{k} \sum_{j \in N_i} \left\| x_j - c_i \right\|^2 \tag{3}$$

The steps of K-mean algorithm can be described as follows: the first step is to select sample K as the initial center point or divide all samples into initial center point class k, and then use these average K values as the starting point. The second step is to classify the samples outside the center point, and classify each sample in the order closest to the center point. The center point of the class is updated and becomes the current average until all samples are sorted. In the third step, repeat the second step until the termination of the treaty is satisfied. One class is based on the number of repetitions; the other is based on the degree of repetition. When two termination conditions are met, repetition can be terminated.

3. Research Design

(1) Research object

This paper takes 400 students majoring in statistics in our university in 2018 and 2019 as the research object, analyzes the book reading amount and the times of entering and leaving the library in one year, so as to analyze their learning behavior in school, that is, the degree of learning effort. Among them, there are 220 students in 2018 and 180 students in 2019. Through the cluster analysis of these data, we can know the learning behavior of students, analyze the current situation of college students to the library and borrow books, and provide valuable information for the library managers.

(2) Data sources

The original data of this paper is directly exported from the campus card database in batch, and the data is true and reliable.

(3) Data preprocessing

In order to ensure the validity of the data in this paper, we preprocess the original data, delete and clean the dirty data which is obviously invalid, including the processing of missing value and abnormal value.

(4) Research Indicators

In this paper, two years of students majoring in statistics were studied

4. Analysis and Research on College Students' Learning Behavior from the Perspective of Deep Learning

4.1 Students' Study Effort

In the research, we need to analyze the degree of students' learning effort. In Colleges and universities, the degree of learning effort is closely related to the number of Library in and out and the type of books borrowed. Therefore, in the research, we use data mining and K-means clustering algorithm to analyze the number of students in and out of the library and the types of books

borrowed from the perspective of deep learning.

(1) Analysis on the times of students entering and leaving the library

This paper analyzes the number of times that the students of statistics major enter and leave the library. First of all, according to the K-means clustering algorithm, we divide the number of students in and out of the library into five levels, which are high, high, medium, low and extra low, respectively. Then the analysis results of the situation of students in and out of the library in 2018 and 2019 in one year are shown in Table 1 and Figure 1.

Table 1. The number of students entering and leaving the library

Grade	2018 Session		2019 Session	
	Number of people	Proportion	Number of people	Proportion
High	20	9.1%	16	8.9%
Higher	62	28.2%	45	25%
Secondary	86	39.1%	65	36.1%
Low	34	15.5%	32	17.8%
Very low	18	8.1%	22	12.2%

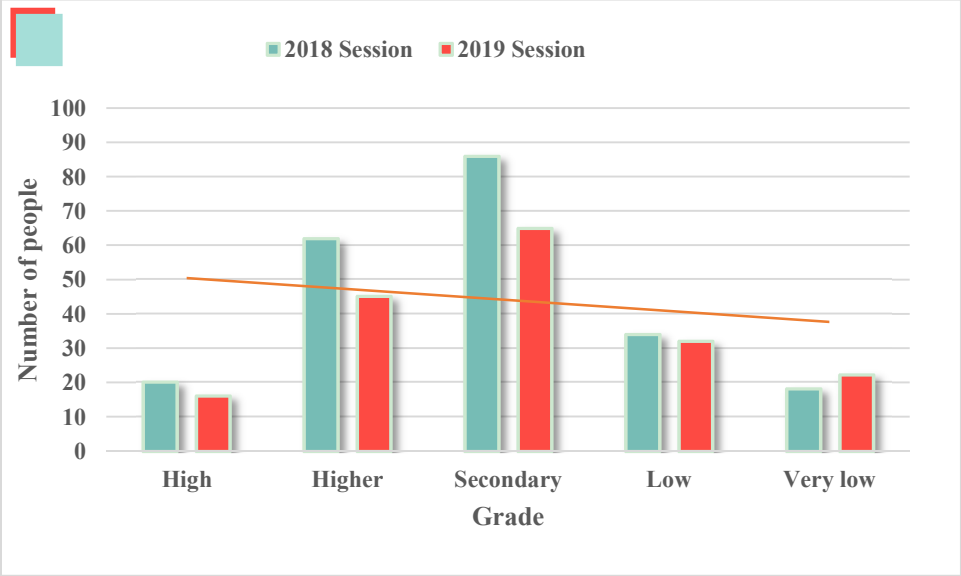

Figure 1. Analysis of the number of students entering and leaving the library

It can be seen from Table 1 and Figure 1 that in statistics major, there are 20 students in the class of 2018, accounting for 9.1%, and 16 students of class 2019, accounting for 8.9%. Within one year, there are 62 students in the class of 2018, accounting for 28.2%, and 45 students in the class of 2019, accounting for 25%. Among the students in and out of the library, there are 86 students in the class of 2018, accounting for 39.1%, and 65 students in the class of 2019, accounting for 36.1%. The proportion of students in the middle level is the highest among the five grades, and the same is true for the two classes. In addition, in the lower and lower grades, the number of students in class 2018 was 34 and 18, accounting for 15.5% and 8.1% respectively, and the number of students in class 2019 was 32 and 22, accounting for 17.8% and 12.2% respectively. To sum up, the number of students in and out of the library in the two sessions are concentrated in the medium level, but in general, the situation of students in and out of the library in 2018 is better than that in 2019.

(2) Analysis on the types of students' books borrowing

This paper analyzes the types of books that students of two years of statistics went to the library to borrow. The results are shown in Table 2.

Table 2. Analysis of the types of student books borrowing

Book type	2018 Session		2019 Session	
	Number of people	Proportion	Number of people	Proportion
Professional books	148	62.3%	98	54.4%
Other books	72	32.7%	82	45.6%

It can be seen from Table 2 that among the students of 2018, the types of books that students went to the library for borrowing were mainly professional books, and the number of people who mainly borrowed professional books was 148, accounting for 62.3%. In addition, among the students of 2019, 98 students mainly borrow professional books, accounting for 54.4%, while 82 students mainly borrow other books, accounting for 45.6%.

4.2 Student Achievement Analysis

We analyze the students' academic performance. In the study, we divided the students' academic performance into four grades, namely excellent, good, medium and poor. Through data mining and K-means clustering algorithm, we analyzed the results of the two sessions of students in the perspective of deep learning. The results are shown in Figure 2.

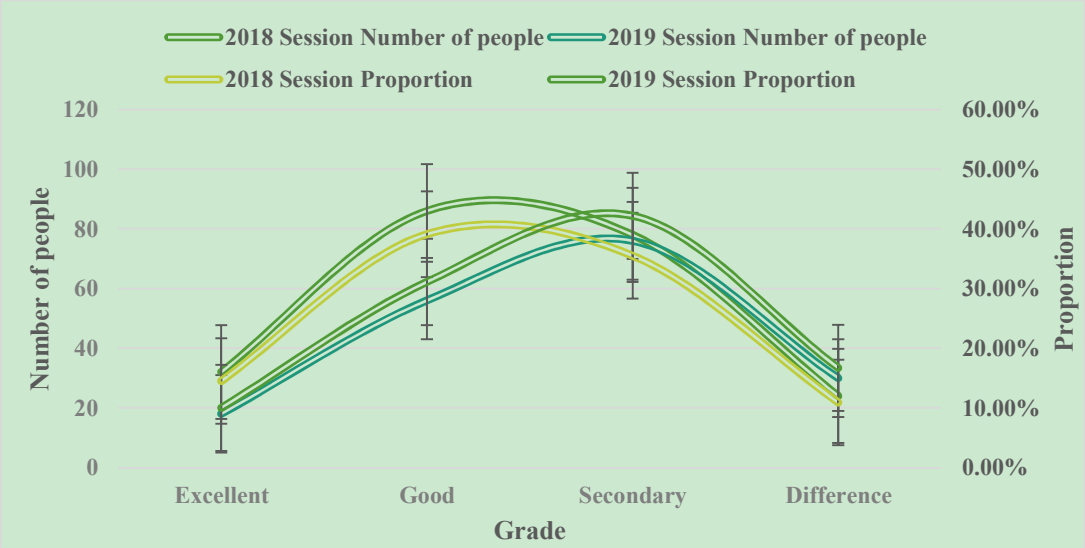

Figure 2. Analysis of students' academic performance

As can be seen from Figure 2, there are 32 students with excellent grades, accounting for 14.5%, 86 students with good grades, accounting for 39.1%, and 78 students with medium grades, accounting for 35.5%. There are 24 students with poor grades, accounting for 10.9%. It can be seen that the grade of students in 2018 is mainly concentrated in the two grades of good and medium, among which the good grade is the most, and the good rate of students in 2018 reaches 53.6%. In addition, there are 18 students with excellent grades, accounting for 10%, 56 students with good grades, accounting for 31.1%, and 76 students with medium grades, accounting for 42.2%. There are 30 students with poor grades, accounting for 16.7%. Although the grades of students in the class of 2019 are also concentrated in the good and medium grades, more students in the class of 2019 are in the medium level, with a good rate of only 41.1%. It can be seen that the performance of students in 2018 is better than that of students in 2019.

In general, from the perspective of deep learning, through the analysis of the learning behavior of the two sessions of students, we believe that the students of 2018 have higher learning efforts and better performance.

5. Conclusions

The students' learning behavior is closely related to their learning performance. How to effectively use the data of students' behavior, analyze their behavior characteristics through data,

realize the teaching according to their aptitude, intelligent management, improve the quality of education and teaching and work efficiency, which has certain practical significance. This paper analyzes and studies the students' learning behavior in the perspective of deep learning. We know that the number of students entering and leaving the library is closely related to the degree of students' study effort. Therefore, in the study, we use data mining and K-means clustering algorithm to analyze the number of students entering and leaving the library and the type of borrowing, and have done a research on the students' learning achievements. The research finds that the higher the students' study effort, the better the students' academic achievement.

Acknowledgements

Xi 'an Eurasia University University-level Key Curriculum Project in Education No.2019KC010;
Xi 'an Eurasia University University-level Key Curriculum Project in Pedagogy of pre-school curriculum design No. 2018KC011

References

[1] Kudoh S N. From the Perspective of Brain Function Current Situation and Expectation of Deep Learning and IoT. Journal of the Institute of Electrical Engineers of Japan, 2018, 138(5):272-275.

[2] Wang D, Zhang M, Li J, et al. Intelligent constellation diagram analyzer using convolutional neural network-based deep learning. Optics Express, 2017, 25(15):17150.

[3] Luo Q, Wang C, Zhao Y. Cultivation Strategy of College Students' Craftsman Spirit from the Perspective of Artificial Intelligence. Journal of Physics: Conference Series, 2020, 1575(1):012173 (7pp).

[4] Yin X, Chen Y, Bouferguene A, et al. A deep learning-based framework for an automated defect detection system for sewer pipes. Automation in construction, 2020, 109(Jan.):102967.1-102967.17.

[5] Naderi M , Salahinejad A , Ferrari M C O , et al. Dopaminergic dysregulation and impaired associative learning behavior in zebrafish during chronic dietary exposure to selenium. Environmental Pollution, 2018, 237(JUN.):174.

[6] Sang, Rak, Kim, et al. Analysis of Multi-group Structural Relationship among Transformational Leadership, Personal & Team Learning Behavior and Team Effectiveness. Journal of Agricultural Education and Human Resource Development, 2017, 49(1):85-111.

[7] Huang, Zhen, Zhang, et al. Vehicle Behavior Learning via Sparse Reconstruction with l (2) - l(p) Minimization and Trajectory Similarity. IEEE transactions on intelligent transportation systems, 2017, 18(2):236-247.

[8] Sugimoto M, Inada S, Matsufuji H, et al. An Experimental Study for Exploration-oriented Behavior in Maze-solving using Reinforcement Learning based on Communication Protocol. International Journal of New Computer Architectures and thr Applications, 2019, 9(2):31-37.

[9] Xiao F, Lin Z, Sun Y, et al. Malware Detection Based on Deep Learning of Behavior Graphs. Mathematical Problems in Engineering, 2019, 2019(PT.4):1-10.

[10]Zeng Y, Chen X, Ong Y S, et al. Structured Memetic Automation for Online Human-Like Social Behavior Learning. IEEE Transactions on Evolutionary Computation, 2017, 21(1):102-115.