



Research and Analysis of Student Behavior Portrait Based on Big Data

Jiameng Zhang (✉)

Shandong Institute of Commerce and Technology, Jinan, Shandong, China
jiamengzhang112233@163.com

Abstract. The application of big data in the Internet, cloud computing and other fields is becoming more and more extensive, with the continuous increase of the network environment and information resources of students' campuses, the network structure of university libraries has also undergone tremendous changes, and the amount of data storage has also become larger and larger. How to use massive data mining technology to effectively classify and analyze students' campus information has become an urgent problem to be solved. In this paper, we mainly study the modeling and profiling of college students' behavior characteristics based on big data analysis and predictive training set. Firstly, this paper studies the theoretical techniques related to big data, including data mining algorithms and behavioral profiling overviews. Secondly, the analysis of student behavior portraits was carried out; Finally, the fusion verification of the student behavior portrait model was carried out, and the experimental results were analyzed and summarized.

Keywords: Big Data · Portrait of student behavior · Research and analysis

1 Introduction

Big data is a new, massive, time-sensitive and valuable information resource, while it has a high degree of data relevance, spatial distribution and multi-source heterogeneity. With the continuous development of Internet technology and advanced information technology such as cloud computing, the large database contains a large number of rich and diverse types of data information, which are intertwined and related to each other, making both valuable information and non-valuable characteristics in the large database discovered and utilized [1]. Therefore, the use of existing large sample sets to obtain the information required for student behavior portraits has become one of the urgent problems in the current society: First, this paper analyzes the student behavior model based on big data technology, and then establishes the corresponding concept and proposes related algorithms according to the attribute relationship between data objects, and finally analyzes the characteristics of big data, and proposes the application strategy of student behavior portrait on the basis of the model [2].

2 Research on the Theory and Technology Related to Big Data

2.1 Data Mining Algorithms

The purpose of data mining is to describe, extract and interpret the huge redundancy and cannot be described, extracted and interpreted with existing information, so as to help us understand the essence of things at a deeper level, which can provide users with better and more efficient decisions and help users achieve more value. The key to data mining is how to find a tool that is efficient and practical, easy to use, and can effectively use existing information to provide a better basis for decision makers. In the context of the big data era, when analyzing and processing massive complex content, it is also necessary to apply to a large number of effective and highly accurate data, and big data analysis technology can provide effective information for decision makers, so that we can understand the essence of things at a deeper level, and on the basis of discovering the essence of things, analyze massive information through data mining technology to help us understand and master complex events at a deeper level [3].

2.2 Behavioral Profiles Overview

Behavioral profiling refers to the analysis of data to obtain representative feature points or preference information for a specific target to form a decision tree diagram. As an effective tool for constructing target users, correlating user behaviors and predicting behavior trajectories, behavior profiles are an effective tool based on data mining and analysis, which establish models and predict potential target groups by integrating user behavior characteristic information. The essence of behavioral profiling is to organize behavioral data into visual behavioral data features, realize the automatic transformation of user profiles, and predict the characteristics of potential target groups according to the data mining results, so as to formulate corresponding measures [4].

Behavior profile is mainly found through the mining and analysis of students' behavior track, various characteristics of students in the mining process can be described by one label. In order to better understand students' personality characteristics, interests, etc., this chapter will construct students' behavior portrait model from the perspective of students' behavior trajectory, and use big data technology to analyze students' behavior trajectory. The information collected and predicted results are used to further uncover valuable and potentially useful information. Firstly, the K-Means clustering algorithm is used to preprocess the original data. Then the association rules mining method is used to build the student behavior portrait model. Finally, according to the results of different dimensions, the comprehensive evaluation index of students' behavior portrait is calculated.

User behavior profile is a virtual real user behavior data, first of all, the user behavior profile divides the user's different behavior label categories for real and specific behavior data; Second, describe the behavior characteristics of users according to their different behavior tag categories. Therefore, based on big data analysis technology can effectively predict and reveal user behavior characteristics. Finally, a data-based user profile is outlined, and the analysis of this data-based user profile can predict the user's behavioral characteristics, thereby providing effective suggestions for schools, teachers and students.

2.3 Data Mining Process Model

Process model is a mathematical method to describe the characteristics of things. It takes data as the input variable, calculates and analyzes its change rule with a specific algorithm. This paper establishes a student behavior portrait model based on big data in the practice of student behavior portrait. This modeling technology can obtain results by comprehensively processing the information related to students' behaviors collected. When encountering a large number of complex problems or failing to make correct judgments, it is necessary to analyze and solve the problems or provide solutions with the help of process models to achieve the expected purpose.

1) Correlation analysis studies

Correlation analysis research refers to the use of statistical analysis indicators to represent the correlation between random data objects, through the analysis of these data, so as to study the correlation between different categories of samples to reveal the inherent law between things [5]. Common correlation analysis methods include plotting scatter plots, scatter plot matrices, and calculating correlation coefficients. In order to more accurately determine and describe the degree of correlation between data objects, we analyze it by calculating the correlation coefficient between data objects. Commonly used methods in the correlation analysis of data objects are pearson correlation coefficient and Spyarman rank correlation coefficient.

Pearson correlation coefficient is generally used to analyze the correlation between the behavioral data of data objects, and some laws are obtained by analyzing the relationship between data, which provides reference for the classification and research of students' behavioral characteristics and psychological structure in the future [6]. The calculation of the Pearson correlation coefficient is shown in Eq. (1):

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \tag{1}$$

wherein, the range of correlation coefficient r is $-1 \leq r \leq 1$, and the different values of r indicate the degree of correlation between the behavioral data of the data object, and the result is to reveal the relationship between the data to a certain extent.

Spearman rank correlation coefficient method: Pearson linear correlation coefficients are generally used to process large sample data and are used to describe and analyze student behavior characteristics. Predict student behavior by statistically modeling and predictive modeling of the collected information. Pearson correlation analysis is a data mining method that statistically models the collected samples to obtain a feature that is highly relevant but highly controllable. The Spearman rank correlation coefficient is also known as the rank correlation coefficient and is calculated as shown in Eq. (2).

$$r_s = 1 - \frac{6 \sum_{i=1}^n (R_i - Q_i)^2}{n(n^2 - 1)} \tag{2}$$

2) CRISP-DM data mining process model

CRISP-DM Data mining process model is a student behavior analysis model based on big data. It is mainly used to process and mine information of potential users and convert it into effective customers to realize effective marketing. CRISP-DM analysis model is used to compare student behavior data with other users, so as to discover different demand characteristics. The process model for CRISP-DM data mining is shown in Fig. 1. The first is to build a set of information sources with multiple variable information and distributed structured data sources based on large sample features, and use this model to mine the hidden layer information contained in the data. Secondly, the method is used to classify the original valid samples, establish the corresponding correlation relationship according to different categories, and map the association rules to the high-dimensional space to analyze the connection between the properties of high-dimensional objects and the hidden things, and extract the implicit information according to the association rules, forming a student behavior portrait model based on big data analysis. Finally, by mining the potential value of the hidden data, and proposing corresponding solutions according to the model, the enterprise can provide a basis for decision-making [7].

The main feature of the CRISP-DM data mining process model is to visualize data by extracting, transforming, analyzing, and modeling a large amount of business data in a commercial database. This method collects, extracts and analyzes a large amount of student information, and uses this data to build a corresponding model. It uses data mining methods to model and analyze student behavior models to help understand their individual needs, which can provide targeted recommendations to the education sector.

Fayyad data mining process model. The Fayyad data mining process model treats the pattern of knowledge in a database as a multi-step process. In this process, users can analyze and process the data, and they can obtain the corresponding information according to different characteristics. For example, large sample knowledge, user behavior preferences and personal privacy, etc., users can analyze this information to obtain user behavior preferences, so as to provide personalized services for schools, students and

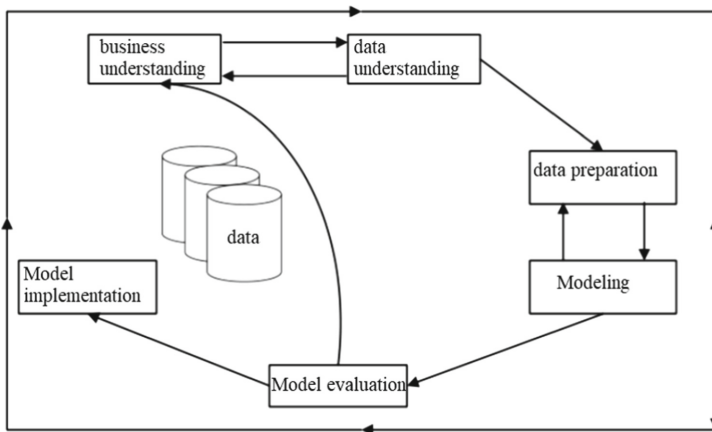


Fig. 1. CRISP-DM data mining process model

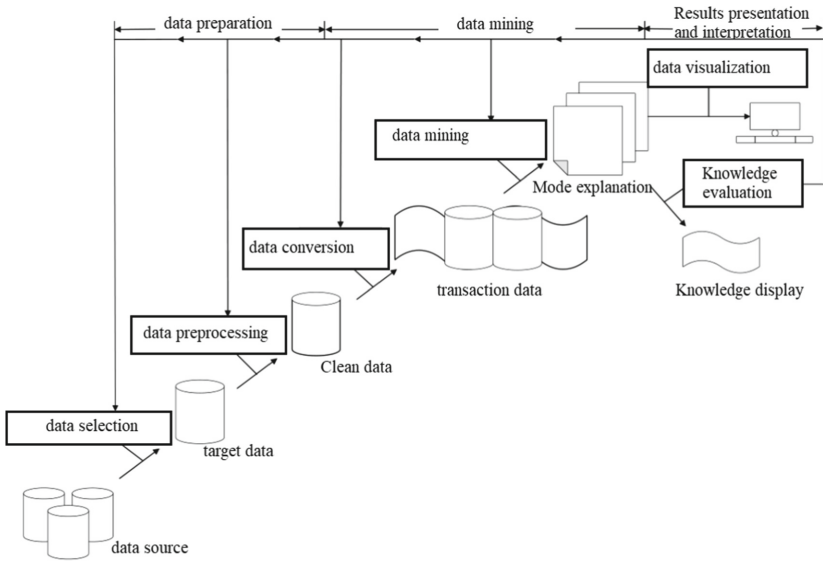


Fig. 2. Fayyad data mining process model

other groups, thereby improving students’ learning efficiency. The Fayyad data mining process model is defined as a model inclined to information processing technology, which predicts and analyzes user behavior based on the results of data mining, thereby realizing information extraction and combining it with the decision-making process [8]. The Fayyad data mining process model is shown in Fig. 2.

3 Student Behavior Profile Analysis

3.1 Student Behavior Profiling Process

Student behavior analysis refers to the process of collecting and analyzing the behavioral characteristics, motivation and psychological states of students in school, and analyzing and comparing them, which is the study of various activities and behavioral characteristics of college students in learning and life, as well as the relationship between students. Firstly, by analyzing the clustering categories of various types of behavioral data, a data-based set of behavioral labels is constructed, and the characteristics of student behavior are analyzed through data mining, and then the neural network training algorithm is used to predict them according to the differences between different categories. Secondly, the data-based label set analysis is used to obtain the scenario-based behavioral characteristics and verify and correct various behavioral labels, and the predictive model is used to analyze the students’ behavioral characteristics, obtain the factors affecting their psychological and emotional states, and make corresponding suggestions for these reasons [9]. The specific process of student behavior analysis is shown in Fig. 3.

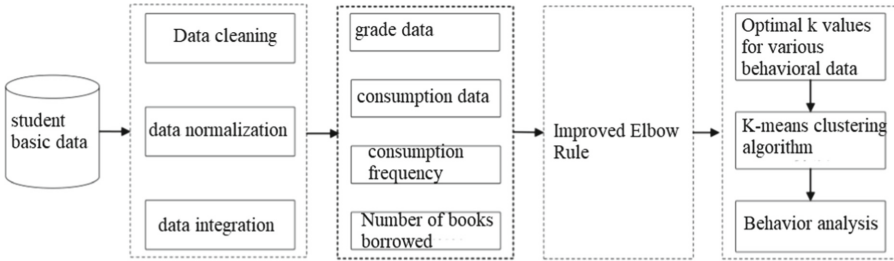


Fig. 3. The specific process of student behavior analysis

3.2 Student Behavior Label Collection

A collection of student behavior labels refers to classifying and tagging them, dividing the data into different types according to certain rules, thus forming a repository of information with specific attributes. Behavioral labels are classified according to different industry application scenarios and business backgrounds, but their data volume is limited, so when applying, they must select corresponding labels according to the behavioral characteristics of different types of students. Commonly used behavioral label classification methods include fact labels, attribute labels and customized labels; Fact-based labels refer to a type of label that uses numerical statistics to mine the actual behavior data of users, and attribute-type labels refer to some similarity between students' behavioral characteristics and data; Custom class labels use some specific demand points encountered or possible in the process of use as classification objects [10].

Student behavior portrait labels mainly use attribute labels to discretize and classify various types of behavioral data, and combine them according to the characteristics of the data to form a student behavior portrait with specific characteristics, so as to realize the adjustment of the relationship between students and others in the social living environment. In this paper, the K-means clustering algorithm is used to cluster the behavior data of students, and the behavior of students is classified according to different characteristics. The analysis results show that the label set with great difference can be generated when learning activities are carried out in the big data environment, and can provide reference for other students. Survey activities in a big data environment can produce a largely different set of labels, and it can make predictions about the adjustment of relationships between students in different behavioral categories.

4 Student Behavior Portrait Model Fusion Verification

4.1 Experimental Program

The integration of student behavior portrait model is mainly based on data mining, and comprehensive analysis is carried out in combination with the behavior characteristics of students, so as to realize the integration of college students' campus information resources. The student behavior portrait model used in this experiment is a fusion model based on the classification algorithm of multi-frequency pattern tree, which is divided into three steps, the first step is to mine the discrete label collection of student data, and

extract features through the data; The second step is to apply the student's behavior profile to the objects with large amount of information, high structure and low similarity, and then analyze the multi-frequency pattern tree of student data, and apply it to the objects with large amount of information, low structure and high similarity by analyzing the student's behavior model. The third step is to organize the multi-frequency patterns that exist in the student data, form a frequent pattern tree based on the multi-frequent patterns of student data, and analyze its characteristic attributes according to the student's behavior pattern tree. In addition, the model mainly uses the deformation structure of the FP-growth algorithm to diverge the structure of the student behavior pattern tree, and by clustering the data, so as to achieve the difference in student characteristic attributes between different categories for the same type of object, and through the clustering of data, the relationship between different types of students is predicted.

4.2 Experimental Results and Analysis

In this paper, the application of student behavior portrait based on the classification algorithm model of multi-frequency mode is mainly used, and the behavior profile of students in big data environment is analyzed by the deformation structure of the FP-growth algorithm, and the relationship pattern between feature structure features and features is determined through observation and experimental results. According to the model, different types of students are processed using the same set of decision software, so as to derive what connections exist between students in different decision-making behavior characteristics, and analyze the data model. Finally, by analyzing the results and comparing them with the differences and correlation factors in the behavioral profiles of others, the tree structure of the student data is obtained, as shown in Fig. 4.

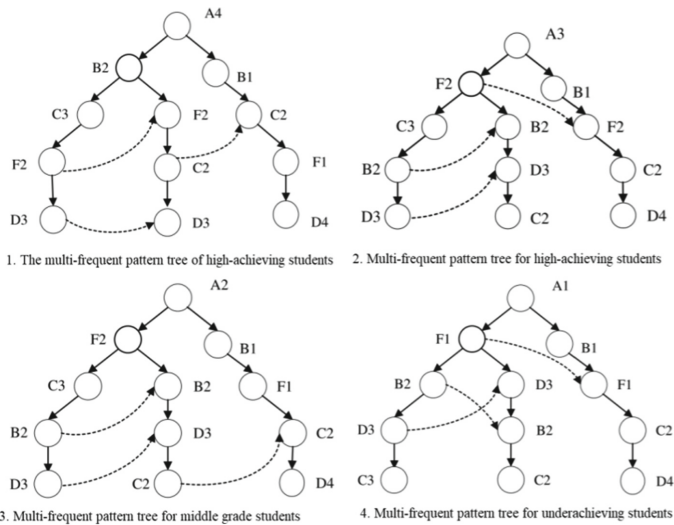


Fig. 4. Multi-frequent tree structure of student data

According to the above, it can be concluded that the behavioral characteristics of students affect their learning activities and learning methods, and are also directly related to the direction of their future career development. Therefore, it is very necessary to analyze and process students' information in the context of big data era.

5 Conclusion

Big data has gradually become the new darling of today's society, its application is becoming more and more extensive, it plays an important role in student learning, campus management and social activities, and the application of big data has also become an important topic in student learning research and campus culture construction. Through the collection and analysis of a large number of information on the behavior of college students, the statistical analysis of student learning behavior data, so as to understand and grasp the personality characteristics of contemporary college students, and in-depth analysis, which can help college administrators better understand and grasp the behavior characteristics of contemporary college students, so as to develop a suitable student development, it is conducive to the campus cultural environment managed by the school. Starting from big data, this paper collects and analyzes student behavior information, analyzes the characteristics of students' learning behavior, and constructs a campus cultural environment model based on big data to achieve college management and college students' mental health education through big data technology.

References

1. GuerreroPrado Jenniffer S.,AlfonsoMorales Wilfredo,CaicedoBravo Eduardo F. A Data Analytics/Big Data Framework for Advanced Metering Infrastructure Data[J]. *Sensors*,2021(16):21–23.
2. He Yi,Yang Shuo,Zhou Xiao,et al. An Individual Driving Behavior Portrait Approach for Professional Driver of HDVs with Naturalistic Driving Data[J]. *Computational Intelligence and Neuroscience*,2022:33–35.
3. Magdalena Cantabella,Raquel Martínez-España,Belén Ayuso,et al.Analysis of student behavior in learning management systems through a Big Data framework[J].*Future Generation Computer Systems*, 2019:262–263.
4. Christoph Klingler,Karsten Schulz,Mathew Herrnegger. LamaH | La rge-Sa m ple D a ta for H ydrology: Big data für die Hydrologie und Umweltwissenschaften[J]. *Österreichische Wasser- und Abfallwirtschaft*,2021:14–16.
5. Guan Weiling,Zhang Daolu,Yu Huang,et al. Customer Load Forecasting Method Based on the Industry Electricity Consumption Behavior Portrait[J]. *Frontiers in Energy Research*,2021:57–59.
6. Han Zhen-Hui,Chen Xing-Shu,Zeng Xue-Mei, et al. Detecting Proxy User Based on Communication Behavior Portrait[J]. *The Computer Journal*,2019(12):62.
7. Wang Aiqun,Yu Hongxiang. The Construction and Empirical Analysis of the Company's Financial Early Warning Model Based on Data Mining Algorithms[J]. *Journal of Mathematics*,2022:33–36.
8. Albrecht Anne,Ben-Yishay Elhanan,Richter-Levin Gal. Behavioral profiling reveals an enhancement of dentate gyrus paired pulse inhibition in a rat model of PTSD[J]. *Molecular and Cellular Neuroscience*,2021:111–113.

9. Ziv Ardi, Anne Albrecht, Alon Richter-Levin, et al. Behavioral profiling as a translational approach in an animal model of posttraumatic stress disorder[J]. *Neurobiology of Disease*, 2016:88–90.
10. Aleman-Meza Boanerges, Loeza-Cabrera Mario, Peña-Ramos Omar, et al. High-content behavioral profiling reveals neuronal genetic network modulating *Drosophila* larval locomotor program.[J]. *BMC genetics*, 2017(1):18.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

