ISSN 2663-8169 Vol. 3, Issue 6: 17-22, DOI: 10.25236/IJNDE.2021.030605

Big Data Processing and Intelligent Analysis System for Students' Learning State Based on PaddlePaddle

Jia Weifeng[a,*], Wang Defeng[b], Niu Honghui[c]

School of Software Engineering, Anyang Normal University, Anyang, China
[a] weifeng.jia@qq.com, [b]1225554566@qq.com, [c]738123804@qq.com
**Corresponding author*

Abstract: *This paper studies the construction idea, method and process of an intelligent learning state analysis system on the background of big data acquisition and processing. Based on the software and hardware of Baidu PaddlePaddle, the intelligent subsystem of feature extraction, face recognition and learning state recognition in the process of teaching process is built. Based on the idea of distributed scalability, big data processing architecture for large-scale data acquisition, storage and analysis is designed. The actual scene shows that this is an intelligent system used for students learning state analysis, which is supported by big data processing, can quickly carry out the teaching attendance, examination verification and class state detection, and can be widely used in learning state analysis of all kinds of schools at all levels.*

Keywords: *learning state analysis; artificial intelligence; big data; PaddlePaddle*

1. Introduction

As the higher education has entered a new stage of popularization, increasing attention has been paid to improving the quality of education and teaching. The analysis of undergraduate students' learning state is the key link [1]. The widely use of mobile Internet and smart equipment has greatly changed the classroom teaching mode. Under the traditional teaching scene, students' learning state data cannot be recorded and analyzed in real time, which only depends on Teachers' memory and attention. The whole teaching process cannot be effectively improved due to the loss of important information. In recent years, the development of Internet, big data and artificial intelligence technology makes it possible to collect and analyze classroom teaching data on a large scale. At present, the research on intelligent chemical situation analysis mainly involves computer vision, image processing and machine learning. Based on the existing cameras in the classroom, the real-time teaching images was sampled and used for students' learning status analyzing [2]. These studies did not solve the problem of large-scale data processing; the implementation of intelligent modules did not take into account the elastic cloud computing configured on demand. Generally speaking, there is a lack of scalability in terms of computing power and scale supporting. This paper focuses on solving these problems, and studies the design and implementation of a big data processing system, that can do intelligent students learning state analysis, based on *PaddlePaddle*. The overall design framework and implementation framework of the system was proposed, and the practice based on the real scene was conducted. Practice shows that this is an effective and scalable big data processing system, and the support scale of the system can be expanded according to the actual needs.

2. System Design

In the analysis of students' learning state in education, the recognition of students' learning emotion expression is an important way to judge students' learning behavior [3]. Paying attention to learners' learning emotion and then improving teaching mode is an important embodiment of student-centered thought [4]. Therefore, various expressions in the learning process can be identified in time, which is one of the key functions to be realized in the intelligent learning situation analysis system; On the other hand, the existing data processing systems mostly adopt the design idea based on relational database. The main problems are insufficient support for unstructured data, high scalability cost and limited system processing capacity [5]. An idea to improve the system processing capacity is to reduce the data scale through data sampling, but for some scenarios, this will affect the accuracy of data analysis [6].

ISSN 2663-8169 Vol. 3, Issue 6: 17-22, DOI: 10.25236/IJNDE.2021.030605

Literature [7] attempts to build a student expression database, but it has not been applied to practice for verification; In the real scene, literature [2] practiced intelligent learning state analysis, but did not use the design idea of big data system. The designed system did not introduce big data architecture, resulting in lack of scalability in the face of large-scale data processing problems. The traditional intelligent learning situation analysis system is based on a single architecture. Its application idea is to input the data obtained by other means into the system for off-line analysis, so as to obtain the learning situation indicators with teaching guiding significance. The algorithm adopted by the intelligent module is mainly self-developed and deployed locally, lacking computational flexibility and operation guarantee; The data processing ability is limited, and the common high concurrency requirements of teaching scenes cannot be well met; The real-time operation of the system is not its main concern, and involves a large amount of student information and a variety of analysis dimensions. The scale of data processing is infinitely expanded. It is urgent to apply big data system architecture in system design.

This paper mainly studies and solves two kinds of problems: intelligent learning situation analysis and big data processing. The former is the core function of the system, which mainly serves the learning state discrimination such as face capture, face segmentation, face recognition and expression recognition based on video stream in practical teaching activities. The results will be used for learning state analysis. The latter is the supporting foundation of the system, which mainly solves the application problems of the above functional modules in large-scale data acquisition, storage, analysis, mining and visualization.

2.1. Design of the Architecture

The system design architecture is shown in Figure 1. The whole system is shown to users using the intelligent emotion analysis subsystem. From the perspective of functional module design, it includes face image segmentation, face search and recognition, expression analysis and recognition, et al. The operation of these modules needs to face the problems of throughput improvement and user experience caused by a large amount of data.

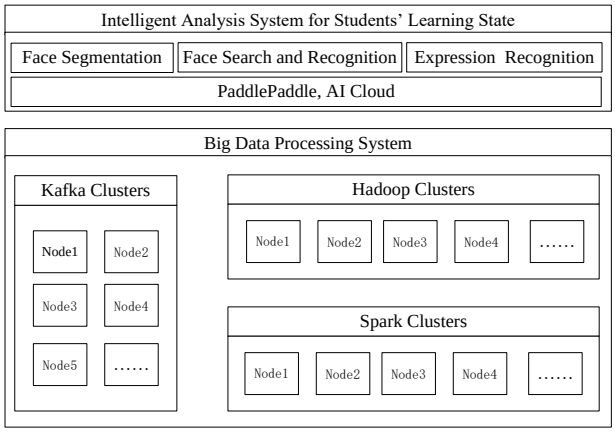

Figure 1: System Design Architecture

Therefore, the author designs a supporting subsystem for big data processing. The system has the ability to process big data scenes from data collection at the source to data storage and calculation. In the big data processing subsystem, Kafka cluster is used to connect with the face video acquisition module, accept the input of a large amount of video data, and do a good job of corresponding management according to the Kafka topics; Hadoop cluster is adopted as the storage platform for massive data; Spark cluster is used for real-time analysis and calculation of large-scale data.

2.2. Subsystem for Data Analysis

The input data of the intelligent situation data analysis subsystem comes from the video stream of the camera. It is obtained in real time by the cameras deployed in the classroom or examination room, so the data is obviously distributed. According to the actual situation of classroom implementation, this kind of data has obvious large flow characteristics. On the other hand, although not all of the analysis of learning big data needs real-time, it involves special application scenarios that need to achieve real-time results (such as examination room verification), which requires that there must be a module in the system that can realize real-time calculation.

ISSN 2663-8169 Vol. 3, Issue 6: 17-22, DOI: 10.25236/IJNDE.2021.030605

In the intelligent learning state analysis system, personalized learning reports are made for learning individuals for parents and students to refer to and improve their learning. This is the main requirement that the system needs to meet. The real-time acquisition of face information based on video stream is the key link in the system. In order to quickly detect the face in video stream and make the system have the characteristics of elastic computing, the system uses *PaddlePaddle* platform Intelligent Cloud face acquisition interface to achieve relevant functions. The purpose of face recognition is to obtain the tag number of the learning individual. On the basis of the tag number, the learning situation analysis also needs to identify the learning state in the learning process through other information, such as listening attentively, dozing off or doing something unrelated to learning. This is essentially a classification problem in pattern recognition. The recognition idea of this link is to intercept n consecutive pictures, recognize the learning state through posture, and record, count and analyze it.

2.3. Subsystem for Data Analysis

Compared with traditional system data processing, the architecture design of big data processing system is based on the idea of distributed scalability. In fact, it is an unlimited architecture expansion to deal with the unlimited surge of data, so as to ensure the acquisition and processing of large-scale data from the mechanism. This idea originally came from the distributed application deployment of search engine system storage server and computing server [8]. In engineering practice, it has low expansion cost and good fault tolerance, and shows obvious advantages in the calculation of super large-scale data. In this system, the big data processing subsystem is faced with the video acquisition equipment distributed in every classroom in a specific time, which is uniformly sent to the data processing center for intelligent analysis and identification. Big data processing module is divided into big data acquisition module, big data storage module, and big data analysis and mining module. The big data acquisition module uses Kafka to connect with the video input terminal (producer), which can improve the ability of the system to collect video information by expanding the number of Kafka nodes; The information collected by the system is stored in the distributed file system HDFS, which has the characteristics of unlimited expansion and distributed storage; For the face recognition module of learning situation analysis, the system interfaces with the propeller and uses API call to obtain elastic computing resources and quickly complete the recognition of intelligent information on the same scale; The system uses spark memory computing to analyze and mine big data based on the results of intelligent recognition, and gives learning data, including personalized learning report and learning guidance.

3. System Implementation

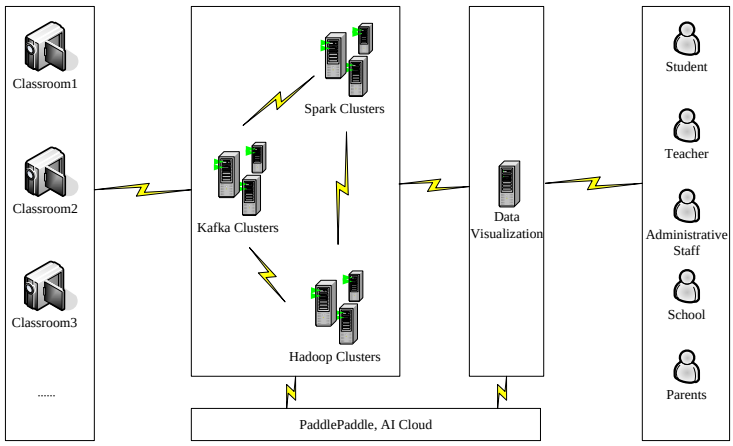

Figure 2: System Implementation Architecture

3.1. Implementation Architecture

The implementation architecture of the system is shown in Figure 2, which is composed of the following parts: (1) Teaching information collection cluster: flume is used for teaching information collection to automatically collect information to Kafka cluster for teaching videos transmitted in the classroom; (2)Teaching information storage cluster: for the storage of teaching information, flume automatically collects it to Hadoop cluster through another pipeline; (3)Intelligent information

ISSN 2663-8169 Vol. 3, Issue 6: 17-22, DOI: 10.25236/IJNDE.2021.030605

recognition cluster (Baidu Intelligent Cloud): for intelligent teaching information recognition, such as face recognition, through the API interface, call the Baidu intelligent cloud to obtain AI resources that can support the operation of the system. (4)Learning situation information analysis cluster: Based on the data collected and stored in the above links, combined with spark computing engine, through big data analysis technology, obtain the learning situation data for schools, majors, classes, classrooms, teachers, students and even parents, and provide improvement strategies. (5)Data visualization system: for all kinds of information collected and analyzed by the system, it is displayed to users in corresponding roles in friendly report style or graphic image.

3.2. Implementation Technologies

The system can deal with the collection and analysis of super large-scale data. The key technologies include the following categories, which are briefly introduced:

Flume is developed based on Java and provides powerful big data collection function. In terms of architecture, flume uses agent to deploy collection points. Each agent includes source, channel, sink and other components, supports the expansion of distributed architecture, and theoretically provides unlimited expansion ability to meet the needs of massive data collection in the era of big data. The essence of flume and other big data collection tools is to provide a "buffer" between data source and data storage.

Kafka is a popular big data tool, which can be used for big data collection or as a data pipeline. It has a wide range of application scenarios, supports the expansion of distributed architecture, and theoretically provides unlimited expansion capacity. Some concepts in Kafka, such as producer, Kafka cluster (broker), consumer, topic, partition, publish and subscribe mechanism, and the coordination role of zookeeper, are helpful to do a good job in the distribution center of teaching information.

Baidu Intelligent Cloud provides a variety of intelligent identification schemes, which can be identified online or off-line. There are software identification based on interface calls, and hardware identification based on degree products. According to the system requirements, video analysis box is selected to connect the video information stream collected in the classroom, and the facial expression image recognition information is obtained based on the video information stream.

Spark computing engine is a memory based fast computing engine, which can connect with Kafka cluster topics, analyze students' learning status in real time, and get the corresponding analysis results.

4. System Operation

Using the above technology, this project implements a prototype system and applies it to the fields of teaching attendance and learning state identification to analyze the learning and teaching in school. This section briefly introduces the operation effect of the prototype system and the solutions to key problems.

4.1. Teaching Attendance

Figure 3: Teaching Attendance Recording

Figure 3 shows the operation effect of the system in the teaching scene: in the captured classroom class image, the key segmented information such as face is extracted for intelligent recognition, which can accurately and quickly determine the classroom attendance results, and play a good role in urging students to actively participate in basic classroom teaching activities.

ISSN 2663-8169 Vol. 3, Issue 6: 17-22, DOI: 10.25236/IJNDE.2021.030605

4.2. Learning State Detection

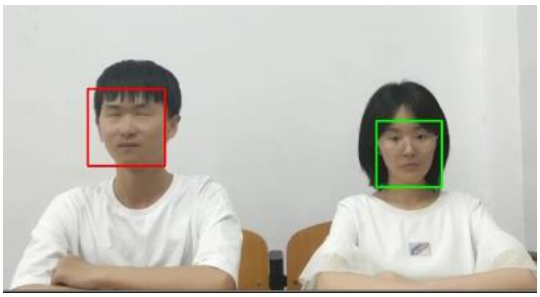

Figure 4: Student Learning State Detection

Figure 4 shows the operation effect of the system in class state detection, which can accurately predict the phenomenon of dozing, which is reflected in the later learning situation analysis report.

4.3. Key Problems and Solutions

In the teaching scene shown in Figure 5 and similar examination room verification and other environments, there is a problem that the resolution of the collected image is not high, resulting in the decline of the recognition rate. The idea to solve this problem is to apply the resolution enhancement technology to the system to improve the recognition rate. For example, as shown in Fig. 5, the student information that was originally in the back row of Fig. 3 and could not be marked can be recognized.

Figure 5: Teaching Attendance Record Improvement

5. Conclusions

According to the needs of academic situation analysis, this paper designs an extensible intelligent big data processing system for students learning situation analysis. The deployment and practice of the prototype system show that its architecture can recognize the learning state in the big data scenario, and provide learning analysis reports for different roles for decision-making. However, due to the lack of more extensive testing, although the system has unlimited expansion ability in theory, it does not run in a scenario of sufficient scale, which is an important aspect of system practice and evolution in the future. In addition, the privacy protection of data collection involved in the learning situation analysis data can also be considered in the subsequent system upgrade.

Acknowledgements

We thank Wang Wenlong, Zhang Nan and students of class 2 majored in *Data Science and Big Data Technology 2019, School of Software Engineering, Anyang Normal University*, for their work in system testing. This work was supported by Students' Creative Foundation of Anyang Normal University.

References

[1] Wang Yihan. Research on the Analysis of Undergraduates ' Learning Situation and the Reform of

ISSN 2663-8169 Vol. 3, Issue 6: 17-22, DOI: 10.25236/IJNDE.2021.030605

Teaching Mode Under the Background of Universal Higher Education [J]. Journal of Tianjin University (SOCIAL SCIENCE EDITION),2019,21(03):268-275.
[2] Xu Hongji,Zhao Wenjie, Li Menghe,Sun Yuxin. Intelligent Analysis System of Learning Situation in Class Based on Behavior Recognition [A]. Science and Engineering Research Center.Proceedings of 2019 3rd International Conference on Advanced Education and Management Science(AEMS 2019) [C]. Science and Engineering Research Center:Science and Engineering Research Center, 2019:5.
[3] Xu Xiaoqing, Zhao Wei, Liu Hongxia. Research on Application and Model of Emotional Analysis in Blended Learning Environment: From Perspective of Meta-analysis [J]. E-education Research, 2018,39(08):70-77.
[4] Zhang Dan, Wang Qing, Yuan Jinping, Wang Kun. Reform and Practice of Technology Enabled Teaching Model [J],2021(04):125-138.
[5] Wang W F,Yang L. Design and implementation of postal delivery of big data analytic system based on Hadoop[J].Journal of University of Chinese Academy of Sciences,2017,34(3) : 395-400.
[6] Qin Xiongpai, Wang Huiju, Du Xiaoyong, Wang Shan. Big Data Analysis—Competition and Symbiosis of RDBMS and MapReduce [J]. Journal of Software,2012,23(01):32-45.
[7] Xu Zhenguo, Liu Zhi, Kong xi, Dang Tongtong. The Design and Implementation of Learning Expression Database Oriented for Deep Learning [J]. Modern Educational Technology, 2021, 31(08):112-118.
[8] S. Ghemawat, H. Gobioff, S. Leung. "The Google file system," In Proc.of ACM Symposium on Operating Systems Principles, Lake George,NY, Oct 2003, pp 29–43.