Exploration and Practice of P2G Teaching Mode of Big Data Architecture and Mode Experiment Based on Clustering Algorithm

Qian Liu[1,a], Youfa Liu[1,b*], Jiawei Zhang[1], Fuhcuan Ni[1,c], Guoliang Li[1,d]

[1] *College of Informatics, Huazhong Agricultural University, Shizishan Street, Wuhan, China*
*[a]lq20200106@163.com, [*b]liuyoufa@mail.hzau.edu.cn, [c]fcni_cn@mail.hzau.edu.cn, [d]guoliang.li@mail.hzau.edu.cn*
Corresponding Author: Youfa Liu

ABSTRACT

According to the characteristics and teaching objectives of the Experimental Course of Big Data Architecture and Mode, a P2G mode teaching mode is proposed and practiced after analyzing and summarizing the problem in the teaching process. P2G teaching mode means point-to-group teaching mode, in which we can divide all students who need to learn the Experimental Course of Big Data Architecture and Mode into different groups.In this paper, we use clustering algorithm to complete the grouping of several forums. We try to select three kinds of clustering algorithms at different levels to complete this task, which are k-means algorithm based on partition, cure algorithm based on hierarchy and sting algorithm based on network.Then devise specific schemes for teaching, discussion and evaluation of the completion effect of the experiment based on different experimental projects. Finally, by analyzing and summarizing the teaching implementation and talent training effect, it can be concluded that the Experimental Course of Big Data Architecture and Mode based on P2G teaching mode can well combine the advantages of online and offline teaching models, which give full play to the autonomy and coherence of teachers' teaching and students' learning to the greatest extent. This provides a certain reference value for diversifying teaching mode and ensuring the practical ability of learning subjects and the evaluation of teaching quality.

Keywords: P2G mode; big data architecture and mode experiment; clustering algorithm

1. INTRODUCTION

In the Internet age, we are confronted with various kinds of information and data. Whether for individual or collective, the ability of processing information and data is particularly important. Big data architecture is an overall system for capturing and processing a large amount of data, so that it can be analyzed and processed for the purpose of various tasks. Based on the business needs of different organizations, the architecture can be regarded as the blueprint for big data problem solutions. Today, the amount of data available for analysis continues to increase every day, and there are more streaming media sources than ever before, including available data from traffic sensors, health sensors, transaction logs and activity logs. However, collecting data is only half the success. We also need to be able to understand various data in detail and use them efficiently to influence key decisions. Therefore, the big data architecture aims to handle the following types of work: (1) Batch processing of big data sources; (2) Real time processing of big data; (3) Predictive analysis and machine learning. The well-designed big data architecture can save money and reduce costs for different enterprises. Big data technologies such as Hadoop and cloud-based analysis can significantly reduce the cost of storing a large amount of data, and can also help enterprises predict future trends and needs and create new products [1]. For example, using big data can help enterprises evaluate customer needs and preferences. Of course, it can also facilitate the management of different departments to make good business decisions. For example, by using the flow component of big data architecture, enterprises can make decisions in time, which not only improves efficiency but also ensures quality. Finally, it can also help different groups design big data architectures that meet their needs and can solve practical problems. For example, various data in the scientific research field can

be more diversified and accurately utilized and managed if we design the architecture of the cloud platform of the scientific research management system [2].

The Experimental Course of Big Data Architecture and Mode is an important basis for cultivating hands-on project practice of undergraduate students majoring in computer, an effective supplement and appropriate expansion of the theoretical teaching process, and also the basis and way to solve practical problems better in the future professional work background. The common offline teaching method focuses on teaching theoretical knowledge and ignores the importance of practical ability. It is no longer suitable for the teaching of the Experimental Course of Big Data Architecture and Mode under the background of application-oriented specialty construction[3]. Computer science and technology and related majors should focus on the cultivation of students' ability to design and complete various related projects, and the setting and implementation of the Experimental Course of Big Data Architecture and Mode is a key step to ensure that students can complete various projects independently and improve their practical abilities. The content of the Experimental Course of Big Data Architecture and Mode is complex. It is difficult to ensure students' grasp of practical skills only by relying on the classroom experimental teaching in school. Moreover, due to the loose management of the experimental course compared with the usual offline course, some students can not complete the assigned experimental tasks alone [3].To make matters worse, due to the limited scale and teaching resources of most school laboratories, it is impossible to carry out large-scale comprehensive experiments and provide targeted guidance and explanation to each student.But the point-to-group teaching mode (P2G Teaching Mode) is a point-to-group teaching mode that expands the traditional offline point-to-point classroom learning into a combination of online learning and offline learning. This teaching model can be applied to the teaching process of the Experimental Course of Big Data Architecture and Mode.

2. CHARACTERISTICS OF THE EXPERIMENTAL COURSE OF BIG DATA ARCHITECTURE AND MODE AND RELATED PROBLEMS IN TRADITIONAL TEACHING MODEL

The Experimental Course of Big Data Architecture and Mode has the following remarkable characteristics:

(1) The basic knowledge points are complex, experimental and expandable.

(2) The experimental items in all teaching schemes of this course are highly consistent, requiring students to operate each experiment, and have high requirements for personal hands-on practice ability.

(3) The database (HBase) used in this course is different from the traditional relational database. Students are required to focus on the differences between the two in analogy, so as to complete the experimental operation more accurately.

(4) It has strong connection with other professional courses of computer science and technology. With the further deepening of basic knowledge and the comprehensive application of front knowledge, students need to have a solid foundation of basic knowledge.

(5) This course requires students to master many frameworks and models with different degrees of difficulty. It is necessary to distinguish the ease of use, functions and other details of different frameworks.

Although the Experimental Course of Big Data Architecture and Mode plays a very important role in the computer teaching system and even the future practical work environment, due to the complexity of the basic knowledge points of the course and the uneven degree of experimental difficulty, there are many problems, such as the knowledge points are easy to understand, but the experiment is useless, not being able to use experimental tasks flexibly after completing, etc. Specifically, it can be attributed to the following aspects.

(1) Each experimental project and difficult points in it lack reasonable design planning and classification, which is divorced from the application in the practical background. The contents of the Experimental Course of Big Data Architecture and Mode are mostly common confirmatory experiments and design experiments. In most cases, a framework and model correspond to an experiment. The experimental contents lack some creativity and novelty. Some experimental codes can be found directly in textbooks or on the Internet, which is easy to make students lack space for thinking and imagination [4]. Of course, there is also a lack of experiments closely related to the practical application of the working environment, which leads to students' failure to solve complex practical problems after completing the experiment.

(2) The teaching methods of the whole experimental course are lack of diversification, and teachers' guidance to students is lack of comprehensiveness and pertinence. On the one hand, the traditional offline experimental teaching mode is not flexible enough, and students' actual mastery is not ideal. Students complete relevant experiments in class or after class, and wait until the final results are checked and accepted by teachers in the next class. When students complete an experiment, teachers cannot provide targeted guidance immediately, resulting in the reduction of teaching effect and practical ability. On the other hand, in general, the class hours of the experimental course are not enough, and some

students are not familiar with the basic knowledge of the previous course and their programming ability is weak, which leads to some difficulties in code writing, program implementation, optimization and program operation efficiency, thus enhancing the difficulty of completing the experimental task. And there is a lack of clear and appropriate guidance on the training program of students [4].

(3) Due to the limitation of experimental course hours and teaching resources, in fact, teachers can not provide a perfect assessment mechanism in the teaching cycle, and the evaluation standard of experimental results is not reasonable enough. In most cases, the evaluation of experimental course takes the advantages and disadvantages of experimental reports submitted by students as the main evaluation index, and there is a lack of assessment on the mastery of experiments, completion process and practical application scenarios. The evaluation criteria are mainly teachers' subjective evaluation, and the scientificity of evaluation indicators is inaccurate [5].

3. INTRODUCTION AND ADVANTAGES OF P2G MODE

In fact, at the beginning of the rise of the Internet industry, P2P mode represents Peer-to-Peer. With the advent of the era of big data and the continuous development of the Internet industry, this mode has gradually evolved into Point to Point mode, that is, Point-to-Point mode. Nowadays, P2P mode is more commonly regarded as Pointer to Pointer, PC to PC, etc., and of course, it is also endowed with richer practical significance [6]. In short, all data based on P2P mode no longer need to be transmitted through a specific server, but directly through different users in the same network. The data transmission parties only need to establish a Peer-to-Peer connection, which means that different hosts do not distinguish which is the service provider and which is the service requester in the communication process. As long as the host can run Peer-to-Peer connection software, peer-to-peer communication can be carried out. P2P technology belongs to the category of Overlay Network. It is a network information exchange mode relative to client / server (C/S) mode. In C/S mode, a specific server is used to receive and forward data, and clients in the whole network need to obtain the required data from this specific server [6]. The P2P model solves the problems such as the limitation of the number of servers, the limitation of network rate, bandwidth and delay, network utilization, service reliability and so on. For example, in general, the number of servers is only one. Even if multiple servers are allowed, considering the cost problem, the number is very limited, which leads to the situation that a single server peers many clients, thus limiting the synchronization of serviced clients.

Moreover, a single fault point is easy to appear in the network path, which increases the difficulty of maintaining and optimizing the stability of the network and system. Therefore, in order to make full use of the huge terminal resources, the use of P2P network can make each node in the same network not only obtain the services and data provided by other nodes, but also provide services and data for other nodes. P2G mode is the expansion and extension of P2P mode. It further expands the advantages of the previous mode, that is, it expands the peer-to-peer network structure into a point-to-group multi-level network structure. In short, P2G mode comes from P2P mode and is obviously different from other general P2P modes and P2C modes. It is an important branch and component in the P2P field. A simple illustration of the above two modes is shown in the figure below.
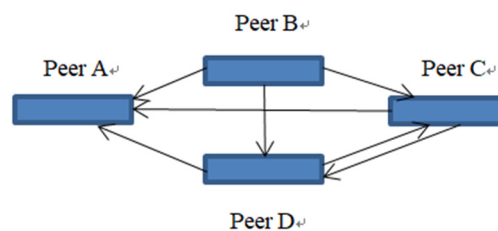

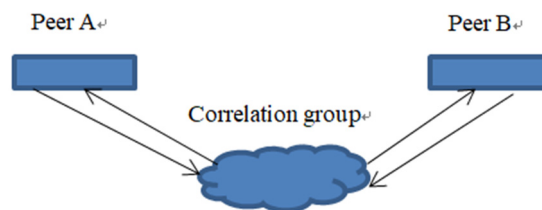**Figure 1**: Interaction between peers in P2P mode.

Figure 2: Interaction between peers in P2G mode.

Nowadays, P2G mode is mainly applied in the field of Internet finance. Its main idea is to graft the Internet financial service platform of private funds and government credit intervention projects, which aims to guide private funds to support the development of the real economy in a rational and scientific way. From the perspective of operation mode, P2G platform is no different from the common P2P online loan platform. It takes the platform as an intermediary to meet the needs of both investors and financiers. The reason why it is called P2G is that the platform mainly serves government direct investment projects, investment projects for which the government undertakes repurchase responsibility, factoring projects of state-owned enterprises, debt repurchase projects of state-owned financial institutions and other projects involving government credit. It is an internet financial investment model for government projects by private funds, supported by government credit [7]. Of course, it also highlights its significant advantages under the wide application of P2G mode. (1) The consistency of data is

easy to control, and the system is easier to manage; (2) The interaction of data, information and applications in the Internet is more flexible and efficient, which makes the database resources more abundant; (3) The scalability of the network is enhanced, it is no longer prone to a single point of failure, and the fault tolerance performance is improved.

Some of the advantages of P2G mode listed above are obvious. Can we apply this mode to the teaching process of big data architecture and mode experiment course? The answer is yes. Its core idea is to change the traditional offline single teaching mode and improve the "one-to-one" teaching mode into "one to many" teaching mode [8]. P2G mode in the teaching process, that is, teachers will carry out teaching work for their major and even all other student groups interested in the course, including but not limited to teaching and learning, experimental results and performance evaluation, Q&A, and no longer only for students in a class. Specifically, the school will design a special teaching system in which students and teachers can operate and communicate with relevant authorities. Teachers should set scientific and reasonable experimental contents and difficulty gradients for offline experimental courses, and students should complete corresponding experiments and obtain corresponding scores in each class. Most of the online teaching contents include the relevant theoretical knowledge of the big data architecture and mode experimental course, as well as the operation rules and precautions of the next offline experimental course. Students should complete the learning and exercises of the corresponding class hours within the specified time.

4. PRELIMINARY ATTEMPT OF P2G TEACHING MODE IN THE EXPERIMENTAL COURSE OF BIG DATA ARCHITECTURE AND MODE

Based on the corresponding problems raised in the first part of this paper, the following specific ways can be taken to improve the course teaching.

(1) All students of this major are randomly divided into several forum groups. Each forum has its own number and name, and can be connected and communicated in the school system.

As for how to complete the grouping of several forums in P2G teaching mode, we can choose the relevant clustering algorithm for grouping. We try to select three kinds of clustering algorithms to complete the grouping task, which are k-means algorithm based on partition, cure algorithm based on hierarchy and sting algorithm based on network.

1. Grouping method based on K-means algorithm:

a.K sample points are randomly selected from all student samples as the initial mean vector $\{\lambda_1,\lambda_2,\lambda_k\}$.

b.Cycle the following steps until the end.

c.Order $\beta_i=\phi(1\leq i\leq k)$.

d.Ask all students to calculate the distance between the sample points and the K mean vectors, select the shortest distance and the corresponding mean vector to mark the cluster mark of the sample point, and finally add the sample point to the corresponding cluster mark β_i.

e.Calculate their new mean vector for each cluster marker $\lambda_i =1 \mid \beta_i \mid \sum x \in \beta_i$ x,if the new mean vector changes greatly, it will be updated, and vice versa.

2.The hierarchical cure algorithm realizes the grouping method:

a.Take a random sample K from all the students.

b.The random sample K is divided into a group of partitions, and each partition is locally clustered.

c.Remove irrelevant feature points in random sample K, such as students who do not belong to this major or related majors.

d.The local classes are clustered once, and the central point in each newly formed class is based on a contraction factor defined by us α Shrink or move toward the center of the class.

e.Use class tags to mark and distinguish each different forum group.

3.Grouping method based on network sting algorithm:

a.All students are randomly divided into different grids.

b.All information in each grid cell is counted and expressed in a simplified method.

c.Select grid cells with high density based on the internal information of each grid.

d.All connected high-density grid cells are combined and recognized as cluster vectors to complete grouping.

Through the above three grouping methods, we can get the desired forum grouping, but these three methods have their own advantages and disadvantages.The following table compares their advantages and disadvantages.

Table 1 Comparison table of clustering algorithms selected for grouping

Algorithm	Grouping effect	Advantages	Disadvantages
K-means Algorithm	Good	Easy to implement and fast processing speed.	The computational overhead is large and the convergence is slow on large-scale data.
CURE Algorithm	Commonly	With less computational overhead.	Incorrect classification cannot be corrected.
STING Algorithm	Excellent	1. The processing speed is very fast, which is usually independent of the number of records in the target database, but only related to how many grid cells the data space is divided into. 2. The grid structure is conducive to parallel processing and incremental update.	The quality of clustering depends on the granularity of the lowest layer of the grid structure.

To sum up, we can choose the algorithm according to the sample size, demand and expected effect of grouping. Due to the moderate number of students majoring in relevant majors in our university, we chose the web-based sting algorithm to complete the task of dividing forum groups.

(2)The experimental tasks are divided into basic, advanced and challenge as difficulty gradients. The basic and advanced experiments are designed by the teachers of the teaching group, while the experiment of the challenge group is designed by the students of each forum group. The final winning forum group is determined in the form of "challenge arena competition" and is included in the teaching assessment. At the end of the semester, it is ranked according to the total points of each student.

(3)The teaching methods of experimental courses can be divided into three categories, and the specific contents are shown in the table below.

Table 2 Classification of experimental teaching methods

Type of experiment	Teachers are responsible for the content	Learning style	Teaching methods
Online experiment	Analyze the experimental principle,answer questions online and share divergent content	Autonomous	Personality guidance
Classroom experiment	Analyze the experimental scheme, reply in class and rate the experimental results	Discussion type	Project driven
Forum experiment	Investigate and guide the experimental design and later application, evaluate the experimental process, and summarize the experimental results and gains and losses	Mutual assistance	Project guidance

(4)When the experimental course content encounters problems, it can be published in the system forum for everyone's help. Students in different forum groups can freely join the discussion. If it still can't be solved, they can seek the help of teachers.

(5)For different teaching methods, a variety of experimental assessment modes can be designed, and appropriate assessment modes can be adopted for the above three different experimental types. First, online or offline exploratory experiments can be automatically and intelligently evaluated through the official teaching platform. Secondly, the offline class experiment should be conducted in the form of preview and writing a short report, then completing the relevant experiment in class, and then submitting the experimental report and making the final defense. Finally, the forum experiment should adopt the investigation method of mutual rating between the Q&A parties and the summary of small papers that need to be submitted to the professional tutor and scored

by the tutor. All the above assessment results are finally included in the students' personal scores in the form of points, and the personal total points will be calculated and ranked during the final examination.

5. TEACHING IMPLEMENTATION AND TALENT TRAINING EFFECT

In recent years, undergraduate colleges and universities are constantly carrying out education and teaching reform based on practical application-oriented courses. Although the exploration and practice of teaching mode in the big data architecture and mode experimental course is only a preliminary attempt, students' understanding of big data architecture and mode design ability have indeed been effectively improved. Firstly, after several semesters of pilot teaching, the feedback effect of students on the course shows an upward trend. The average score of the big

data architecture and mode experimental course in recent year shows an upward trend as a whole, and the scores of some students in individual classes have reached a record high[9]. Secondly, from the questionnaire and graduation destination, we can see that the situation of students obtaining professional related certificates with high gold content is getting better and better, which leads to a good employment situation of college students in this major. Most students find their own satisfactory jobs after graduation and realize the counterpart of professional resources. Finally, after follow-up investigation and return visit, most personnel engaged in the field of big data architecture are basically able to apply the contents of the previous experimental courses to practical work to solve the relevant problems actually encountered in the professional field.

6. CONCLUSIONS

In general, the exploration and practice of P2G teaching mode in this experimental course is particularly important for the cultivation of project practice ability of students majoring in computer science and technology. The results show that the P2G teaching mode of big data architecture and mode experiment has obtained good feedback in practical teaching. This teaching model can not only give full play to the initiative and innovation of the experimental subject, but also enable teachers to obtain more immediate feedback and comprehensive supervision in the teaching process. Of course, the reform and optimization of teaching methods will not happen overnight, so colleges and universities should continue to develop their strengths and avoid their weaknesses in the future teaching process, so as to improve the professional ability and quality of their students [10].

ACKNOWLEDGMENTS

REFERENCES

[1] Ai Qiang.(2021).Discussion on computer network application architecture design in the era of big data.J. China new communications. 2,88-89.

[2] Cui Xiaolong.Zhang Min.Guo Xi.Wang Xiaomei.Zhang Dezheng.(2021).Multi level teaching design of big data engineering practice based on Hadoop.J. Laboratory research and exploration. 1,157-161.

[3] Cao Yi.(2017).Obstacles and Countermeasures in the development of p2g model in China.J.Journal of Harbin University. 11,43-47.

[4] Fu Ju.Sun Lianshan.Zhao Xiao.Liu Wei.Wang Xiaoxia.(2021).Exploration and practice of experimental teaching mode of data structure and algorithm based on ability training.J.Computer education. 3,99-103.

[5] KRISTINAP.SINAGA.MIIN-SHEN YANG.(2020). Unsupervised K-Means Clustering Algorithm.J.IEEE Access,8,80716-80727.

[6] Liu Fang.Dai Qin.Bi Jing.Zhao Liang.(2021).Construction and implementation of hybrid experimental teaching model of network interconnection technology.J. Computer education. 3,165-170.

[7] Lu Qingping.Chen Jiajun.Yang Liu.(2021).Exploration of online teaching scheme of public computer experiment course based on hybrid mode.J.Computer education. 3,51-54.

[8] Maurice Roux.(2018).A Comparative Study of Divisive and Agglomerative Hierarchical Clustering Algorithms.J.Journal of Classification.35,345-366.

[9] Wang Yaoyao.Zhan Hongfei.(2021).Analysis on the cloud platform architecture of scientific research management system in the era of big data.J.Science and technology entrepreneurship monthly.2,52-57.

[10] Wang Wei.(2021).P2P traffic identification method based on neural network.J.Microcomputer application.3,111-113.

[11] YangJianqiang.HangBo.ZhengYi.LiXuefeng.(2021).Practice of hybrid teaching mode based on OBE in the course of "Fundamentals of Information Security".J.Computer education.3,143-148.

[12] Zhou Zhonghe.Li Cheng.(2020).Design and application of intelligent education business architecture based on 5G.J.Journal of Hunan Post and Telecommunications vocational and technical college. 12,4-6.