

Research and Practice of Online Learning Behavior Analysis Model from the Perspective of Big Data

Yougui Zhao and Dongmei Xia
Chongqing institute of engineering, Chongqing, China

Abstract—By using the analysis results of online learning behavior, teachers can understand and identify students' learning dynamics, analyze the difficulties and blind spots of the course, reasonably adjust the course structure and video content. Meanwhile, teachers can get inspiration from students' feedback and expand their teaching ideas. For platform managers, learning data analysis can indirectly show the application effect of platform functions and provide guidance for managers to improve and optimize platform functions. From the perspective of teaching model reform and innovation, the study of learners' learning behavior brings new ideas for educational researchers to reform teaching model.

Keywords—big data; online learning platform; online learning behavior

I. DEFINITION OF CORE CONCEPTS

Big data: international data center IDC defines big data as a new era technology and system, which can obtain valuable information through data collection, processing and analysis.

Big data is a data set with extremely large volume and data category, and such a data set cannot be captured, managed and processed by traditional database tools. "Big data" first refers to large data volume and refers to large data sets, generally about 10TB in size. However, in practical application, many enterprise users put multiple data sets together, which has formed a data volume of PB level. Secondly, there are large categories of data, and the data comes from various data sources. Data types and formats are becoming increasingly rich, which has broken through the previously restricted category of structured data, including semi-structured and unstructured data. Then, the data processing speed is fast. In the case of very large data volume, the real-time data processing can also be achieved. The last characteristic is the high authenticity of data. With the interest of new data sources such as social data, enterprise content, transaction and application data, the limitations of traditional data sources are broken, and enterprises increasingly need effective information power to ensure its authenticity and security.

Online Learning platform: online Learning platform, also known as online Learning platform, supports students' online course Learning by providing an open and Shared teaching environment. It is called e-learning platform in foreign countries. Li jiahou (2007) defined it from the perspective of the service function provided by the online learning platform, namely, the online learning system that can organize, send and present resources, track and manage the learning of the

platform users, and promote human interaction and human-computer interaction.

Online learning behavior: the general term for learners' autonomous learning and interactive activities in a technical environment created by the information network, utilizing the learning tools and support services provided by the platform. It can be defined in a broad sense and a narrow sense. In a broad sense, it refers to the sum of all information acquired by learners from the platform. In a narrow sense, learners choose courses in the online learning platform according to their own learning motivation and needs, and use the existing course resources and auxiliary tools in the platform to achieve learning effect. This study is mainly based on a narrow perspective.

II. RESEARCH STATUS

A. Research Status of Big Data Abroad

Big data was first proposed by TOFFLER(2006) in his classic book the third wave. Nature launched a special issue of Big Data in 2008. McKinsey has released a scientific report on big data, pointing out that all sectors of society are inseparable from big data. IDC has found that the amount of data generated in recent years has far exceeded previous data and is doubling every two years. At the same time, governments represented by the United States are also vigorously carrying out big data strategic plans. In the field of education, purdue university adopts big data technology to build a learning early warning mechanism by collecting students' data in the course. At present, the core focus of foreign scholars' big data research is to extract value from the data set.

B. Research Status of Big Data in China

With "big data" and "education" as key words, the retrieval results in the Chinese academic journal network were 659, indicating that China has been gradually on the right track in the field of education big data research. However, the retrieval results with "big data" and "learning behavior" as keywords were only 42, indicating that the study on the analysis of learning behavior in the context of big data in China is still in the initial stage and there is still a lot of room for development. Zhu zhiting (2013) discussed the application of big data in the field of education technology and different educational technology paradigms, transformed the main body of teaching into students, and proposed the construction idea of personalized adaptive learning system. Hu shaoxing (2013) built a learning analysis model to mine online education data and find out the key factors that affect the quality of education,

realizing the application of big data technology in the field of education.

C. *Research Status of Online Learning Behavior Abroad*

There are abundant research results in this field abroad. Kenneth(2016) et al studied the influence level of learning behavior and reflective learning in online business courses. Prior(2016) et al analyzed the influence of learning attitude, information literacy and self-efficacy on online learning behavior. Chyan(2013) et al. constructed a multi-level latent class model to explore the behavior patterns and regional differences of platform users. Shimada et al. (2015) used machine learning and classification method to analyze the relationship between learning behavior and validity in class test. Fakir(2014) et al analyzed the data stored in Moodle with the help of the research data algorithm to understand the pattern of learners' online learning behavior. Durksen(2016) established a probability analysis model of psychological needs for learners in online learning platform by using bayesian network.

D. *Research Status of Online Learning Behavior in China*

In recent years, domestic research in this field has also made some achievements. From the perspective of literature, most domestic researchers focus on the theory and framework research of online learning behavior analysis model, and there is still a lot of room for the relevant research on specific technologies and methods such as collection, prediction and analysis of online learning behavior. For example, li shuang et al. (2016) constructed an analysis framework of online learning behavior investment from the perspective of behavioral investment and selected measurement indicators of online learning behavior investment. Jiang zhuoxuan (2015) et al. classified online learning characteristics of platform users and discussed the relationship between learning behavior and learning effect. In terms of analysis methods and modeling of learners' learning behavior data, some scholars have optimized and innovated them from different perspectives. Fan chao (2016) et al studied the characteristics of learners' online learning behaviors based on human dynamics, revealing that learners' online learning behaviors showed significant heterogeneity. Zong Yang et al. (2016) used logistic regression method to reveal the correlation between online learning behavior and effect, and found out the key indicators that affect the effect of online learning. Hu yiling et al. (2014) discussed online learning behaviors of learners from the perspectives of data mining, learning analysis technology and analysis modeling. Huang zhicheng (2014) introduced markov model to evaluate and diagnose online learning behaviors, which can effectively identify users' learning behaviors.

In summary, it can be seen from the literature review that foreign research and development should precede domestic research and the development and application of online learning platforms are relatively complete, but the research on online learning behavior analysis in the context of big data is still in the initial stage. In theory and practice research in recent years, domestic also have certain achievements, but there are still some shortcomings, such as data acquisition for online learning behavior lack of pertinence, at the same time for online learning behavior analysis model analysis methods have

limitations, not according to the actual situation in big data situation choose the right analysis comes in. Therefore, this project is dedicated to the establishment of an online learning behavior analysis model in the context of big data. Based on the interdisciplinary integration of multiple theoretical disciplines, it builds a multidimensional and universal online learning behavior analysis model with the help of big data technology, gradually making up for the lack of research in the field of education informatization in China.

III. AGENT-BASED NETWORK TEACHING MODEL

Real world problems is extremely complex, and the function of a single Agent is extremely limited, a single Agent generally difficult to complete a given task, at this moment need through appropriate architecture to organize multiple Agent to form multiple Agent system, to share a task, to make up for the shortage of single Agent, makes the ability of the whole system more than a single Agent. System based on multi Agent technology refers to multiple Agent communication and coordination with each other, each other mutual finishing homework tasks in the system, it not only has the general distributed system of resource sharing, easy expansion, strong reliability, flexibility, good real-time characteristic, and each Agent can solve the problem of large-scale complex through mutual coordination, the system has strong robustness, reliability and self-organizing ability. In the multi-agent system, a single Agent is a physical or abstract entity that can act on itself and the environment, manipulate part of the representation of the environment, and communicate with other agents, with the basic functions of perception, communication, action and control and reasoning ability. These characteristics of multi-agent technology make it have a broad application prospect in dealing with knowledge problems based on the Internet.

The outstanding characteristic of network teaching is that students are the main body of learning. Teachers should, through curriculum design, create problem scenarios, online discussion, summary, evaluation and motivation, stimulate students' interest in learning and learning motivation, encourage them to understand and master the knowledge system, cultivate the spirit of innovation, and conduct extensive and in-depth learning.

As network teaching is a complex, huge and unpredictable information system, which generally covers all aspects of teaching, it should be divided into sub-problems and constructed multiple agents with certain functions, which can deal with sub-problems respectively. When the subproblems are interdependent, the agents in the system must be able to control the dependency through cooperation. Based on the above analysis, the multi-agent-based network teaching model we constructed is shown in figure 1. Figure 1 multi-agent-based network teaching model the whole model is divided into three parts: interface Agent, analysis Agent and suggestion Agent. Interface agent: responsible for direct communication with the user. The function of inquiry is to record the user's basic information. Presentation is the delivery of textbook content to the user through an output port. Diagnosis is to judge the user's learning ability, knowledge structure. The function of recording

is to record all the interactions between learners and the system. Interface agent from the interface database to obtain a user dialogue mode and screen. Analysis agent: analyze the user information obtained from the three sub-agents including diagnosis, record and inquiry in the interface agent, and then store the results into the basic database of students, learning achievement database and learning database. Suggestion agent: the suggestion agent obtains the contents of the analysis agent database, and then evaluates the agent to analyze the differences among learners. The suggestion agent provides different teaching methods for different learners according to these differences. In this multi-agent system, in order to reduce the direct communication between agents, the mutual communication information is placed in the global database of interface agents, and the main controller is responsible for notifying relevant modules. Of course, the agent of the main controller can also form a separate module. There are five basic databases involved in the model, namely learning resource database, system global knowledge base, teaching strategy database, interface database and student model database. 3 basic users, namely teachers, students and managers; There are 6 types of agents, namely, teacher Agent, student Agent, manager Agent, human-computer interaction interface Agent, teaching Agent and management Agent. The following part of the specific introduction of Agent functions.

A. *The Student Agent*

After students log in the network teaching system, the system will automatically generate a student Agent. On the one hand, it should provide interactive interface for the corresponding students to guide their learning. In the process of learning, according to the actual situation of students, the teaching Agent should select appropriate strategies from the teaching strategy library to guide students and return the learning results to the student model library after the end of learning. On the other hand, the student Agent should also call the current login student's basic information and learning records to check the students' previous learning conditions, and present the initial learning materials for the students according to these records. During the whole learning period of the user, the student Agent should continuously analyze the learning state of the student through the human-computer interaction interface Agent, so as to make corresponding preparations for the user's next learning. At the same time, I am also responsible for returning the final analysis results of this study to the student information base so as to provide materials for the next study

B. *The Teacher Agent*

After the teacher logs on the network teaching system, the system will automatically generate a teacher Agent. On the one hand, the teacher Agent is responsible for the interaction between the teacher and the network teaching system. Through the teaching Agent, the teacher guides and monitors the teaching process, understands the learning process and reaction of students, points out which students' understanding or answer is the best and makes explanation. If there is any deviation in learning direction, correct it in time. In addition, relevant questions can be extracted from the question bank for

discussion to enhance understanding. On the other hand, the teacher Agent is also the database and active collector of professional knowledge, which can provide the largest professional resource data for each learner and establish relevant courses or curriculum framework for the teaching Agent to choose. In addition, the knowledge base can be adjusted and expanded actively according to the intervention of teachers and the response of students, and relevant information can be obtained from the network actively to become more meaningful knowledge.

C. *The Manager Agent*

Generally speaking, education and teaching activities include a series of management, such as curriculum management, student registration management, performance management and so on, so it is necessary to establish a management Agent in the network teaching system to be responsible for the intelligent coordination of the whole system. After the administrator logs into the network teaching system, the system will automatically generate an administrator Agent. The management Agent is mainly responsible for macro-control of the whole teaching situation through the management Agent. The management Agent actively obtains the data and materials of other agents, and automatically generates relevant management data, such as learners' learning time, regional distribution, learners' level statistics, teachers' work statistics, etc., to assist the manager in effective and rapid response. At the same time, the management Agent should also take on the responsibility of adding and deleting management, directory and address management and communication chain management of other agents.

IV. REALIZE THE SYSTEM KEY TECHNOLOGY

The network teaching system based on multi-agent technology is a complex, huge and unpredictable information system, and its design and implementation will involve many key technologies. Among them, the most important is the organization of data model, the construction of Agent and the selection of development technology.

A. *Organization of Data Models*

In the teaching system, mainly involves the following several objects, considering the extensibility and portability of the system, in the object modeling, we refer to the "distance education specification" and the relevant requirements of the new curriculum reform, and each object has the corresponding extension.

B. *Knowledge Point Model*

Knowledge point is a complete teaching unit that describes the knowledge in the field of teaching. If knowledge point A must be mastered before knowledge point B, then B is called the precursor knowledge point of A, while A is called the successor knowledge point of B. The collection of knowledge points and their relationships is called a knowledge tree.

Knowledge point: knowledge point identification, knowledge point name, keywords, knowledge goal, emotional goal, ability goal, learning content, difficulty, teaching focus, importance, learning time, media resources, strong precursor

knowledge, weak precursor knowledge, weak successor knowledge, extension basis.

In network learning, the knowledge point is the basic learning unit, corresponding to the knowledge point coding and knowledge tree of a certain discipline. Such knowledge point coding and knowledge tree structure clearly reflect the relationship between knowledge points. Taking a certain knowledge point as an example, it is convenient to find the precursor and subsequent knowledge points of this knowledge point. Each knowledge point in the tree is given a unique code, which facilitates the query, addition and deletion of knowledge points. When learners learn a certain knowledge point, ITS can easily test ITS precursor knowledge and prompt ITS subsequent knowledge. According to the frequent access path mined from the Web log, the knowledge tree structure is adjusted to form a knowledge point relational network and navigation suitable for network learning.

C. Learning Resource Model

The media types of learning resources are divided into five categories: text, graphics (images), audio, video and animation.

Resource information: resource identification, knowledge point identification, keywords, description, effective learning time, size, use environment, feedback evaluation, extension basis.

The coding of resource identification is as follows: T begins with text, P begins with graphics (images), A begins with audio, V begins with video and F begins with animation. Definition of learner model: learner model refers to the detailed description of learner attributes, including six tuples of personal information, security information, academic information, preference information, relationship information and learning behavior. $Student_Model_Database = (Student_Model(I) \mid I = 1 \dots N) \ Student_Model = \{personal\ information, security\ information, academic\ information, preference\ information, relationship\ information, learning\ behavior\}$

a) *Personal information is the personal information that is not directly related to the measurement and record of learner performance and is mainly related to management. Generally speaking, such information is private and confidential.*

Personal information: : = learner id, name, gender, date of birth, phone, Email, extension base.

b) *Safety information is information about learner safety certificates.*

Security information: : = learner identity, login password, extension base.

c) *Academic information is some brief information related to learners' learning.*

Academic information: : = learner identification, course learning, completion, extension basis.

d) *Preference information description can promote the selection of parameters for human-computer interaction.*

Preference information: : = learner identity, emotion, interest, media preference, learning type, extension basis.

e) *Relational information is information that describes the relationship between learners and other system users (such as teachers and other learners).*

Relationship information: : = learner identity, relationship, degree, extension basis.

Learning behavior is to describe the relevant operational information of learners on the knowledge points (such as homework information, exercise information, test information, question information and media learning information).

Learning behavior information = {homework information, practice information, test information, question information, media learning information, extension base}.

V. CONCLUSION

Of course, the network teaching system based on multi-agent technology is still in the stage of research and experimental development, and the research in related fields needs to be further deepened. It is believed that in the near future, more humanized distance education systems that conform to China's national conditions and future learning models will be developed successfully.

VI. PROJECT FUNDS

This paper is funded by chongqing education science planning project (project no. 2017-gx-035).

Supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-M201901901)

REFERENCES

- [1] ALVIN TOFFLER. Third wave [M]. Translated by huang mingjian. Beijing: citic press, 2006:35-37.
- [2] MANYIKA J, CHUI M, BROWN B, et al. The Big data: The next frontier for innovation, competition, and The productivity [R]. Mckinsey Global Institute, 2011.
- [3] LUNCH c.b. data: How do your data grow? [J]. Nature, 2008, 455 (7209) : 28 and 29.
- [4] li shuang, wang zengxian, yu chenzhen. Online learning behavior input analysis framework and measurement index research -- learning analysis based on LMS data [J]. Open education research, 2016, 22 (2) : 77-88.
- [5] jiang zhuoxuan, zhang yan, li xiaoming. Analysis and prediction of learning behavior based on MOOC data [J]. Computer research and development, 2015, (3) : 614-628.
- [6] fan chao, zong liyong. Human dynamics analysis of MOOC online learning behavior [J]. Open education research, 2016, 22 (2) : 53-58.
- [7] zong Yang, sun hongtao, zhang xiangguo, zheng qinhua, Chen li. Logistic regression analysis of MOOCs learning behavior and learning effect [J]. China distance education, 2016, 36 (5) : 14-22.