

Study on Accurate Early Warning and Intervention Model for Online Learning based on Multi-source Heterogeneous Data Fusion

Qianqian Ge, Cuncun Wei, Dongyan Wu

School of Electronics and Information, Zhejiang Business Technology Institute, Ningbo 315012, China

Abstract

This paper proposes a precise early warning and intervention model for online learning based on multi-source data fusion. The model tracks and records students' multidisciplinary learning process data, analyzes learning characteristics using data mining methods, and makes quantifiable predictions about learning development, thus realizing personalized and accurate teaching services for the whole learning process data, facilitating analysis of the hidden learning trajectory behind the data, and effectively guiding personalized learning for personalized teaching and student management.

Keywords

Multi-source Heterogeneous; Data Fusion; Online Learning; Accurate Warning.

1. Background

In the post-epidemic era, education philosophy, teaching platform, teaching methods, teaching relationships, and teaching evaluation are facing comprehensive changes. Despite teachers' continuous optimization of teaching methods and teaching tools, students have a high dropout rate, low participation, difficulty in deep learning and other quality crises in the process of participating in online learning, which are mainly demonstrated in the following aspects.

(1) It is difficult to supervise students' learning status, and "pseudo-online-learning" is prevalent, which makes it more difficult to produce deeper learning. In online learning, teachers and students are in a semi-separate state, and learners set their own learning pace according to their own needs and actual situation, so the degree of freedom and autonomy is higher. However, in learning, students tend to feel lonely and are easily disturbed by life or affairs around them, so they often fail to focus on learning, think deeply, and systematically sort out the learning content, etc. It is difficult for them to spontaneously develop deep learning.

(2) Even though there is status data, there are many online course platforms, and students' learning data of the course is distributed on different teaching platforms, which is difficult to maintain and manage in a unified way. It is impossible to accurately understand the situation of learners, which makes it difficult to analyze the data of students' individual learning behaviors and guide personalized learning.

(3) The current online education learning platform cannot meet the requirements for multidisciplinary data statistics, information richness, timeliness, accuracy and reliability. Each platform has better single-subject data statistics and analysis. For example, it supports multiple forms of assessment such as in-class exercises, quizzes, homework and exams etc. Students may take multiple courses in a semester across multiple learning platforms, but it becomes difficult to aggregate a student's academic

data across multiple learning platforms. It is also lacking in accurate early warning and intervention with learner-centered data visualization from the perspective of multiple courses, classes, and schools.

(4) Multiple sources of heterogeneous data increase the complexity of learning analysis. Due to the different functions and architectures of each online teaching platform, a large amount of multi-source heterogeneous data is generated in the process of analysis and statistics of learning and teaching conditions. These multi-source heterogeneous data usually have the following characteristics.

Mixed data: including structured and unstructured data.

Discrete data: data distributed across different systems or platforms.

A large volume of data: the volume of data on every platform is essentially very large.

Uneven data quality: The quality of data is inconsistent across different platforms.

Multi-source heterogeneous data easily affects the accuracy of analysis, which is not conducive to accurate warning and personalized guidance. It is necessary to analyze students' online learning, and data fusion technology is needed to derive estimates and verdicts from multiple sources of data to increase the confidence and reliability of the data to provide discipline-wide learning analysis and guidance, as well as accurate data warning and intervention, so as to make personalized learning recommendations and guidance for online learning.

2. Domestic and International Research

In recent years, domestic and international researchers have collected behavioral data in teaching and learning processes and applied sophisticated analytical tools to explore and analyze them to improve learning and education, focusing on predicting behavior by building analytical models and intervening, optimizing, and warning the learning process. For example, Johnson, L. et al. argued that learning analysis can predict future learning performance and identify potential problems. Domestic researcher Li Yanyan et al. argued that learning analysis can be used to predict learner behavior, assess the learning status and effectiveness of learners, and then intervene in learning to improve the technique used to improve learners' learning performance. The collection of behavioral data on learners' learning processes and their individualized analysis, evaluation, prediction, and intervention are intrinsic to the interest of learning analysis research in promoting effective learning.

3. Accurate Early Warning and Intervention for Online Learning based on Multi-source Data Fusion

This paper explores the accurate early warning and intervention model of online learning based on multi-source data fusion, tracking and recording students' learning process data in multiple disciplines, analyzing students' learning characteristics using data mining methods, and making quantifiable predictions of learning development. On the one hand, the study was instrumental in identifying different categories of learning groups in the post-epidemic era for online learning, providing more scientific and rational, differentiated instructional decisions, adapting to changing learning needs, and improving learning outcomes. On the other hand, it can promote the organic combination of teaching top-level design and practical exploration, form personalized and accurate teaching services based on the whole learning process data, help analyze the hidden learning trajectory behind the data, effectively guide students' learning, resolve learning crisis, and promote personalized teaching and student management.

3.1 Build a Multi-source Data Convergence and Sharing Platform

Starting from the exchange, storage, sharing, service and security of multi-platform data, researches are conducted on the multi-mode exchange method of multi-source heterogeneous academic data to realize the support of "data integration" at the logical layer. Researches are conducted on the efficient storage and indexing technology of data to realize the support of "datafication" and "efficient query". Researches are conducted on unified resource authority control technology to realize support for "multi-user" access. Researches are conducted on user-oriented dynamic data services to realize

support for "data development" at the service level. By establishing a standard specification system and a data security assurance system, the data aggregation and sharing operation process is standardized, business continuity and data security are guaranteed, and a multi-source heterogeneous data aggregation and sharing platform system is formed.

3.2 Conduct Accurate Early Warning of Online Learning

According to the data convergence sharing platform, the students are provided with a variety of learning traces left in the course learning process, such as the task submission and implementation of the interactive evaluation. By collecting these log data and filtering them, three dimensions of structured data were selected, including students' learning status (including 5 data items such as time to complete homework, time to complete evaluation, and total login time etc), learning interaction (including 6 data items such as the total number of posts, the total length of posts, and the total number of replies to others), and academic level (including 3 data items such as homework scores, quiz scores, and test scores), and a Naive Bayes classifier was selected as an early warning model for the study to identify students at learning risk.

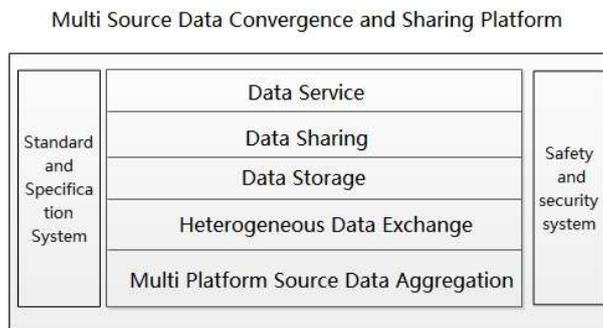


Figure 1. Multi-source Heterogeneous Data Aggregation and Sharing platform

3.3 Realize Effective Intervention of Online Learning Difficulties

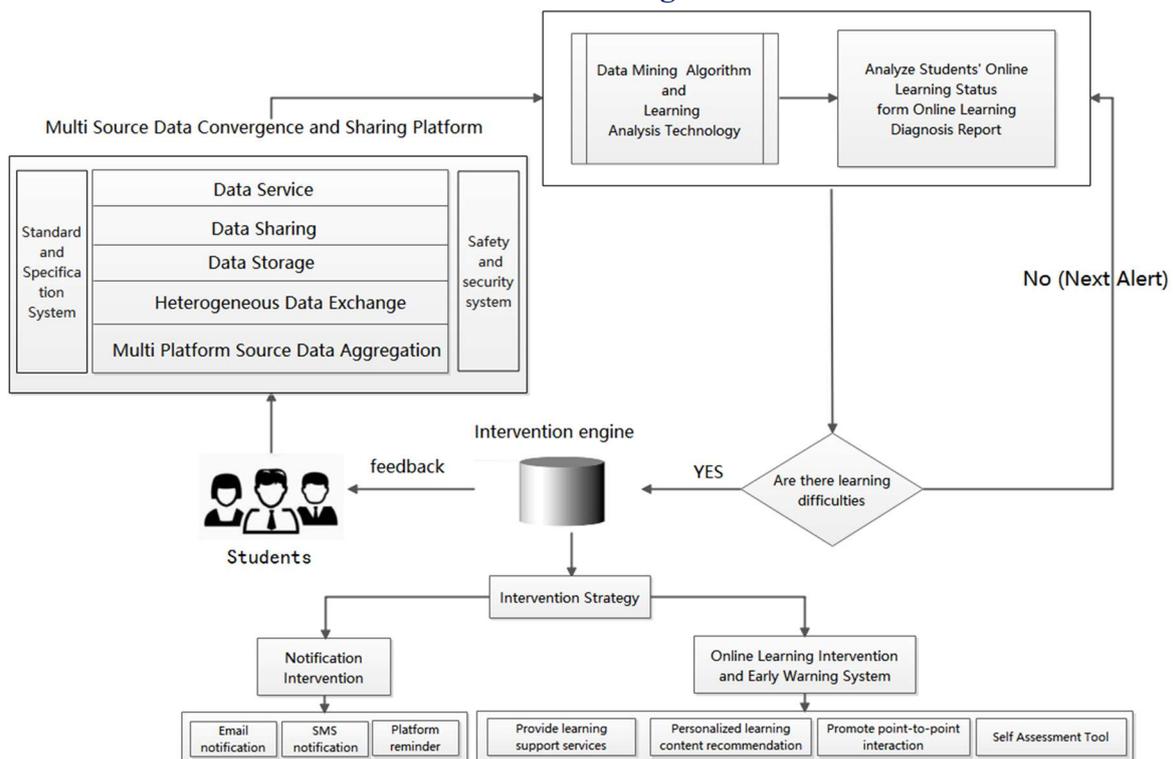


Figure 2. Online learning difficulty intervention model

Based on the early warning of learning situation, the online learning crisis intervention model is designed by combining the features of the online learning environment. The early warning model identifies whether a student has a learning crisis. If there is no diagnosis, the next round of diagnosis is continued and the learner's personal online learning diagnosis information is updated in real time. If a learning crisis is diagnosed, an intervention notification and an online learning support environment intervention will be sent to the student.

3.4 Visualization of Learning Data

The data related to the learner's learning situation is visually presented. Based on the data aggregation and sharing platform, the learning status recorded by these data is understood and interpreted through analytical models to assess the future performance of students and intervene to provide effective motivation and guidance for learner learning, promote self-monitoring and improvement of learner behavior, and achieve personalized self-adaptation of learning guidance. The data to be presented in this process are diverse, including learning behavior, time, interactions, emotions, interests, and habits etc. Learning analysis visualization presents results that can directly affect the object of use, such as the progress of the learner, their learning effect, and their learning obstacles.

4. Summary

The paper integrates and unifies the learning data of multi-course platforms by aggregating, extracting, cleaning, converting and merging multi-source heterogeneous learning data, and establishes standardized data aggregation and sharing. The analysis of multidisciplinary data provides a comprehensive quantitative description of students' learning behaviors, academic diagnosis, accurate early warning, and personalized intervention, which helps to accurately identify students with learning difficulties and provide appropriate teaching services.

Acknowledgments

This work was supported by the National vocational education teachers teaching innovation team project of China (Grant: YB2020020103). "Spatial Effect and Mechanism of the Impact of High-skilled Talents Gathering on Strategic and Emerging Industries: A Case of Dawan District of Zhejiang Province" Philosophy and Social Science Planning Project of Zhejiang Province, 2020. (Subject approval No.:21NDJC192YB).

References

- [1] Data fusion in cyber-physical-social systems: State-of-the-art and perspectives[J]. Puming Wang, Laurence T. Yang, Jintao Li, Jinjun Chen, Shangqing Hu. Information Fusion .2019.
- [2] libPLS: An integrated library for partial least squares regression and linear discriminant analysis[J]. Hong-Dong Li, Qing-Song Xu, Yi-Zeng Liang. Chemometrics and Intelligent Laboratory Systems. 2018.
- [3] The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability[J]. Simon Elias Bibri. Sustainable Cities and Society . 2018.
- [4] A disease diagnosis and treatment recommendation system based on big data mining and cloud computing[J]. Jianguo Chen, Kenli Li, Huigui Rong, Kashif Bilal, Nan Yang, Keqin Li. Information Sciences. 2018.
- [5] Expanding LogP: Present possibilities[J]. Chrysoula Vraka, Sanja Mijailovic, Vanessa Frhlich, Markus Zeilinger, Eva-Maria Klebermass, Wolfgang Wadsak, Karl-Heinz Wagner, Marcus Hacker, Markus Mitterhauser. Nuclear Medicine and Biology. 2018.
- [6] Estimating Individual Treatment Effect in Observational Data Using Random Forest Methods[J]. Min Lu, Saad Sadiq, Daniel J. Feaster, Hemant Ishwaran. Journal of Computational and Graphical Statistics . 2018 (1).