WHIOCE PUBLISHING PTE. LTD.
PROVIDING
FIRST-CLASS SCIENTIFIC INFORMATION
FOR TOP SCHOLARS

2025 Volume 3, Issue 11
Contemporary Education Frontiers
ISSN:3029-1879 (Online)
ISSN: 3029-1860 (Print)

Research on Methods for Detailed Analysis of Educational and Teaching Process Data in the Digital Age

Sijuan Xue*
Hainan Vocational University of Science and Technology, Haikou 571126, Hainan, China

Author to whom correspondence should be addressed.

Abstract: Digital technology facilitates the transition of educational data from extensive collection to refined analysis, providing data support for teaching optimization and personalized instruction. Based on the multi-source and dynamic nature of educational process data, this paper identifies the core requirements and application challenges of refined data analysis, explores a data analysis methodology tailored to teaching scenarios, and proposes actionable approaches across three dimensions: data preprocessing, analytical model development, and practical result application. The aim is to unlock the inherent value of teaching data through scientific refined analysis methods, enabling precise diagnosis and efficient optimization of teaching processes, thereby offering methodological insights for digital education reform.

Keywords: Educational and teaching process; Digitalization; Refined data analysis; Teaching data mining

1. Characteristics of educational and teaching process data and the foundation for detailed analysis

1.1. Multi-source classification and collection standards for educational teaching process data

Educational teaching process data originates from all stages of the instructional workflow, exhibiting distinct multi-source and heterogeneous characteristics. Based on data attributes, it can be categorized into three types: behavioral data, performance data, and feedback data. Behavioral data encompasses dynamic operational metrics such as classroom attendance, interaction frequency, and courseware viewing duration; performance data includes phased outcome metrics like unit tests and homework completion; while feedback data involves subjective evaluation sources, including teacher-student peer assessments and learning progress questionnaires. The collection channels and format standards vary significantly across different data types. Discrepancies in data generators, collection tools, and storage methods result in diverse characteristics in data dimensions, granularity, and update frequencies. For instance, classroom interaction data is collected in real-time with high frequency, whereas periodic test data is collected at lower intervals. These variations pose fundamental challenges for data integration required for subsequent refined analysis.

To ensure the availability of data throughout the educational process, unified collection standards must be established, specifying the collection dimensions, granularity criteria, and timeframes for different types of data. For behavioral data, core collection indicators should be defined to avoid excessive collection of invalid data, while data formats and storage paths should be standardized. For academic performance data, scoring criteria and data entry rules must be standardized to ensure data accuracy and comparability. For feedback data, standardized requirements for questionnaire design and evaluation dimensions should be established to minimize data bias caused by subjective factors.

1.2. The core meaning of refined data analysis in the digital era

The refined analysis of educational and teaching process data in the digital era differs from traditional extensive data statistics. Its core lies in being teaching demand-oriented, enabling deep data mining, precise interpretation, and personalized application. Essentially, through sophisticated analytical methods, it extracts valuable teaching patterns, student learning characteristics, and problem indicators from vast, multi-source teaching data. This approach breaks down data silos, transforms data into actionable value, and establishes data as a scientific foundation for instructional decision-making, learning diagnostics, and personalized guidance, rather than merely serving as a numerical record.

The core essence of detailed data analysis is also reflected in its hierarchical approach and targeted nature. It aims to achieve both a macro-level overview of the overall teaching situation and precise analysis at the micro-level of individual students and specific teaching components. At the macro level, data analysis enables the identification of the class's overall teaching progress, knowledge mastery status, and common challenges, providing guidance for optimizing collective instruction; at the micro level, it reveals students' individual learning habits, knowledge gaps, and differences in ability development, offering precise support for personalized teaching.

1.3. Technical support conditions for refined analysis of teaching process data

The refined analysis of teaching process data relies heavily on comprehensive digital technology support. Big data storage and management technologies serve as the foundational guarantee, enabling efficient storage, categorized management, and rapid retrieval of multi-source, massive, and heterogeneous teaching data. Technologies such as cloud storage and distributed databases address storage capacity challenges for teaching process data. Meanwhile, features like data indexing and classification tags facilitate precise management of data across different types and stages, providing efficient data retrieval channels for subsequent analysis and preventing analytical inefficiencies caused by disorganized data management [1].

Artificial intelligence and data analysis algorithms serve as the core technological foundation for refined data analysis in the teaching process. Algorithms such as descriptive statistics, association rule mining, and machine learning prediction provide technical solutions tailored to various analytical objectives. For instance, machine learning algorithms can be employed to develop student performance prediction models, enabling precise forecasting of learning outcomes; association rule mining facilitates the identification of intrinsic relationships between teaching components and learning effectiveness.

2. Advanced preprocessing method for educational and teaching process data

2.1. Cleaning and duplicating removal optimization strategies for multi-source teaching data

During the collection of multi-source teaching process data, contaminated data can easily arise due to issues such as tool malfunctions, manual entry errors, or system synchronization discrepancies. Data cleaning thus becomes the primary step in preprocessing, with the core objective being to eliminate invalid data and correct erroneous entries to ensure accuracy. For numerical anomalies and formatting inconsistencies caused by input errors, data validation rules should be established, combining manual review with automated system checks to rectify errors. Fields missing or containing corrupted data resulting from collection tool failures must be directly removed to prevent invalid data from compromising analytical outcomes.

Data deduplication is a critical step in multi-source data integration. Since the same teaching activity may be recorded by different collection tools, duplicate data can easily occur. For example, student attendance data may be collected simultaneously by both the classroom management system and smart classroom devices, leading to data redundancy. To perform data deduplication, it is first necessary to define core deduplication criteria, such as student ID, activity time, and data type, and then identify and remove duplicate records through a combination of system-based automatic matching and manual review. Additionally, deduplication rules for array merging must be established to ensure the integrity and consistency of merged data, preventing the loss of valuable information due to simplistic deduplication methods and providing a clean, unique dataset for subsequent analysis.

2.2. Standardization and normalization methods for isomorphic teaching data

Heterogeneity is a defining characteristic of teaching process data, with significant variations in the dimensions, units, and value ranges of different data types. For example, grade data are expressed as percentage values, interaction frequency is counted as integers, and feedback data are provided as rating evaluations. This heterogeneity can lead to biases in data analysis results, making standardization an essential step. The core of data standardization lies in converting heterogeneous data with different dimensions and formats into standardized data with unified formats and evaluation criteria. For qualitative data such as rating evaluations, they can be transformed into quantitative data through value assignment methods; for quantitative data with varying value ranges, format standardization can be achieved by aligning metric definitions, thereby enabling comparability among heterogeneous data.

Data normalization is a processing technique that maps data to a fixed value range based on standardization. Its core objective is to eliminate analytical bias caused by differences in data dimensions, thereby enhancing the accuracy and stability of data analysis models. Common normalization methods include minimum-maximum normalization and Z-score standardization, which can be selected according to the type and distribution characteristics of the teaching data [2]. For example, minimum-maximum normalization can be applied to score data with fixed ranges to map them to the 0–1 interval; for interaction data following a normal distribution, Z-score standardization is used for normalization.

2.3. Missing value imputation and outlier identification in teaching process data

During the collection of teaching process data, missing values can easily occur due to factors such as students not participating in operations, delayed equipment recording, or omissions in manual entry. Failure to promptly fill these missing values reduces the sample size for data analysis and leads to biased results; therefore, a scientific supplementation method should be selected based on the type and distribution characteristics of the missing values. For numerical data with random missing values and a low missing rate, the mean or median supplementation method can be employed; for sequential data with inherent correlations, interpolation is recommended; for qualitative data or data with a high missing rate, manual verification combined with reasonable inference should be used to ensure the rationality and authenticity of the supplemented data.

An outliers refer to extreme data points that deviate from the overall distribution characteristics of the data. Their origins include collection errors or exceptional teaching behaviors, and failure to identify and address them can significantly compromise the accuracy of data analysis results. Outlier identification should be conducted within the context of actual teaching scenarios, employing a combination of statistical analysis and business logic judgment. Common statistical methods include the 3σ principle and the box plot method; these methods identify extreme values in the data, which are then verified against teaching business logic to determine whether they represent genuine exceptional teaching behaviors.

3. Construction of a detailed analysis model for educational teaching process data

3.1. Descriptive statistical analysis method based on student learning diagnosis

Descriptive statistical analysis based on learning situation diagnosis serves as the foundational method for refined analysis of teaching process data. Its core lies in using statistical descriptions of teaching process data to characterize current teaching practices and learning characteristics, enabling precise diagnosis of overall learning conditions. This approach focuses on key indicators during instruction, such as knowledge module mastery rates, homework completion rates, and classroom interaction participation levels, by calculating statistics like means, medians, standard deviations, and frequency distributions. These metrics allow quantitative analysis of class-wide and student-group learning characteristics, clarifying the overall distribution of knowledge mastery, learning engagement, and learning outcomes, while identifying common challenges and strengths in the teaching process.

The refinement of descriptive statistical analysis is manifested in the diversity of analytical dimensions and the precision of analytical granularity. It requires moving beyond the traditional holistic statistical approach and conducting stratified and categorized statistical analyses across multiple dimensions. Analyses can be conducted by dimensions such as knowledge modules, teaching stages, and student groups. For example, identifying students' weak points by knowledge module, evaluating the effectiveness of classroom interactions and homework exercises by teaching stage, and analyzing learning differences among groups based on students' learning abilities.

3.2. Association rule analysis method for mining teaching patterns

Association rule analysis focused on uncovering teaching patterns fundamentally involves identifying inherent correlations among various indicators in teaching process data to reveal hidden instructional and learning patterns, thereby providing a scientific basis for teaching optimization. Built upon extensive teaching process data, this method employs association rule mining algorithms such as Apriori to analyze relationships between different teaching behaviors, learning behaviors, and learning outcomes. For example, it examines the correlation between classroom interaction frequency, homework completion quality, and knowledge mastery rates, identifying a strong association rule: "high interaction frequency + high-quality homework" correlates with high knowledge mastery rates, thereby revealing effective teaching and learning methodologies [3].

Refining association rule analysis requires emphasizing the effectiveness and practicality of the rules, avoiding the identification of ineffective associations that are detached from actual teaching practices, while enhancing the granularity and specificity of rule mining. During the mining process, appropriate support and confidence thresholds should be established based on teaching business logic to select association rules with genuine instructional significance. Additionally, hierarchical association rule mining can be conducted across different knowledge modules, student groups, and teaching stages to uncover personalized teaching patterns applicable to various scenarios. For example, association rules linking teaching behaviors with learning outcomes for lower-grade students, or those connecting teaching methods with mastery rates for challenging knowledge modules.

3.3. Construction of a predictive analysis model for personalized teaching

The predictive analysis model focused on personalized teaching fundamentally utilizes machine learning algorithms to construct predictive models based on historical teaching process data, enabling precise forecasting of individual students' learning outcomes and knowledge acquisition trends, thereby providing forward-looking guidance for personalized instruction. This model employs students' historical learning behavior data, academic performance data, and feedback data as feature variables, with subsequent learning outcomes as the target variable. By selecting appropriate machine learning algorithms such as linear regression, decision trees, and random forests, and through model training and optimization, it develops a highly accurate learning prognosis model capable of accurately identifying students' knowledge gaps and learning effectiveness.

The refinement of predictive analysis models lies in their personalization and dynamic optimization. This requires moving away from traditional unified model frameworks to develop tailored predictive models for individual students and specific knowledge modules, while enabling dynamic updates and optimizations. During model construction, personalized predictive models should be built for different student groups based on their learning abilities and habits to enhance prediction accuracy. Since teaching process data is continuously updated, a dynamic optimization mechanism must be established to incorporate new data into model training promptly, thereby continuously refining model parameters and ensuring its timeliness and accuracy.

4. Practical application of detailed analysis results from educational teaching process data

4.1. The transformation path from analytical results to precise diagnosis of teaching issues

Transforming refined analytical results into precise diagnoses of teaching issues constitutes the primary step in applying these findings practically. The core lies in establishing a clear correspondence between analytical outcomes and specific teaching challenges, achieving an accurate mapping from data metrics to actionable instructional questions. First, it is essential to clarify the educational implications of each analytical metric. For instance, a low homework completion rate may indicate insufficient student initiative or poorly designed assignments, while low mastery rates in a particular knowledge module could reflect inappropriate teaching methods or inaccurate identification of key instructional priorities. By interpreting these metric implications, quantified analytical indicators can be translated into qualitative descriptions of teaching issues.

To effectively translate analytical findings into precise diagnoses of teaching issues, it is essential to establish a multi-dimensional, multi-level diagnostic framework that integrates both macro and micro-level analysis results for comprehensive and accurate identification of instructional challenges. At the macro level, this involves analyzing overall student performance to identify common issues in the teaching process, such as inappropriate pacing or unsuitable instructional methods. At the micro level, it entails examining individual student characteristics and specific teaching segments to pinpoint personalized and detailed problems, such as particular knowledge gaps in individual learners or inadequate interactive design in certain classroom activities [4].

4.2. Method for designing personalized teaching plans based on analytical results

Personalized teaching plans are designed based on refined analysis results, with the core focus on formulating targeted teaching strategies and tutoring programs around identified individual student differences and knowledge gaps to achieve differentiated instruction. For each student, the analysis identifies their weak knowledge areas, learning habit issues, and capacity development disparities, enabling the creation of tailored learning plans. These include specialized tutoring content addressing knowledge deficiencies, behavioral correction strategies for learning habits, and tiered learning tasks accommodating ability variations, ensuring that teaching plans precisely align with students' individual learning needs.

The design of personalized teaching plans should emphasize both practicality and systematic coherence, aligning with overall educational objectives and students' individual developmental needs to achieve an organic integration of differentiated instruction, personalized tutoring, and group teaching. During classroom instruction, students can be categorized into distinct levels based on learning abilities and knowledge mastery, enabling the formulation of tiered teaching objectives, content, and assignments that address the diverse learning requirements of each group. For after-class tutoring, tailored one-on-one guidance plans should be developed to address students' specific challenges, complemented by a tracking and feedback mechanism to facilitate timely adjustments to instructional strategies.

4.3. Dynamic optimization mechanism for teaching processes driven by analytical results

The core of establishing an analysis-driven dynamic optimization mechanism for the teaching process lies in integrating refined analytical insights throughout the entire instructional workflow, enabling real-time monitoring, dynamic adjustments, and continuous refinement of teaching practices. This creates a closed-loop system encompassing "data collection–analysis and diagnosis–teaching optimization–data feedback." It is essential to identify key optimization points and adjustment criteria for each phase of the teaching process, using data analysis results as the direct basis for adjusting teaching schedules, optimizing instructional methods, and designing teaching content. For example, teaching pacing should be promptly adjusted based on learning analysis, teaching methods should be refined to address students' knowledge gaps, and the difficulty and scope of teaching content should be modified in response to student learning feedback.

The dynamic optimization mechanism for the teaching process requires establishing efficient feedback channels and rapid adjustment mechanisms to ensure that analysis results are promptly communicated to teaching practitioners, enabling teaching optimization measures to be swiftly implemented and yield tangible outcomes. By leveraging digital teaching platforms for real-time dissemination of analysis results, teachers can promptly monitor changes in student learning conditions and identify teaching challenges. Concurrently, a tracking and evaluation mechanism for teaching optimization effectiveness should be established. This involves continuously collecting process data post-optimization, analyzing the implementation impact of these measures, and refining optimization strategies based on evaluation findings, thereby creating a closed-loop system for continuous improvement.

5. Conclusion

In the digital era, refined analysis of educational process data serves as a crucial tool for achieving precision teaching and enhancing educational quality, while also constituting a core component in driving educational digital transformation. This paper establishes a comprehensive analytical framework for educational process data tailored to practical teaching scenarios across four dimensions: analytical foundations, data preprocessing, model development, and result application. It specifies core operational methodologies and implementation pathways for each stage, enabling meticulous design throughout the entire workflow, from data collection to value realization. Grounded in the multi-source and heterogeneous nature of teaching process data, this framework is demand-driven, eschewing purely technical analyses detached from real-world teaching contexts, thereby ensuring both the scientific rigor of the analysis process and the practical applicability of its outcomes.

Disclosure statement

The author declares no conflict of interest.

References

[1] Liu K, 2025, Big Data Empowering Refined Management in Education and Teaching. Xinhua Daily, February 21, 2025(012).

[2] Song J, Zang J, Li Y, 2025, Digital Intelligence Empowering the New Process of Educational Digital Transformation and Development: Insights from the Beijing Regional Case Study on Educational Big Data. Information Technology Education in Primary and Secondary Schools, 2025(4): 29–31.

[3] Zhou D, Qiu D, Ren S, 2025, Leveraging the Potential of "Data Elements × Education" to Enhance Students' Core AI Competencies. Digital Teaching in Primary and Secondary Schools, 2025(3): 79–82.

[4] Chen X, 2025, Research on the Refined Development of Online Ideological and Political Education in Universities under the Big Data Context. Jilin Education, 2025(17): 71–73.