Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 18 No. 9 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijim.v18i09.49291

PAPER

# Improvement of Student Interaction Analysis in Online Education Platforms through Interactive Mobile Technology and Machine Learning Integration

**Jinjin Wang**(✉)

School of Preschool Education, Shaanxi Vocational & Technical College, Xi'an, China

jinjinking2013@163.com

**ABSTRACT**

The emergence of online education platforms, driven by interactive mobile technology, has significantly reshaped traditional educational paradigms and underscored the critical need for advanced analysis and improvement of student interactions. Effective analysis of student interaction is crucial for enhancing teaching quality and optimizing the learning experience in these digitally enriched environments. Traditional analysis frameworks often face challenges such as inaccuracies in anomaly detection and inefficiencies in data handling, particularly when handling extensive datasets typical of online platforms. This study introduces a novel approach to enhancing student interaction analysis systems by leveraging the synergy between machine learning and advanced interactive mobile technologies. Initially, the study proposes an advanced anomaly detection method tailored for identifying irregular student interactions. This method utilizes a blend of machine learning algorithms and the real-time data processing capabilities of mobile technology. Furthermore, to address the complexities of data transmission in mobile-based online education ecosystems, a state-of-the-art congestion control algorithm has been developed. This algorithm optimizes data flow, significantly enhancing transmission stability and efficiency. The integration of interactive mobile technology with machine learning offers a robust and dynamic framework for analyzing student interactions, thereby facilitating a more engaging and effective online educational experience. This research contributes to the advancement of online education quality and efficiency by emphasizing the role of interactive mobile technology in shaping future learning environments.

**KEYWORDS**

interactive mobile technology, online education platforms, student interaction analysis, machine learning integration, anomaly detection, congestion control, mobile data transmission optimization

Wang, J. (2024). Improvement of Student Interaction Analysis in Online Education Platforms through Interactive Mobile Technology and Machine Learning Integration. *International Journal of Interactive Mobile Technologies (iJIM)*, 18(9), pp. 35–49. https://doi.org/10.3991/ijim.v18i09.49291

Article submitted 2024-02-13. Revision uploaded 2024-03-18. Final acceptance 2024-03-23.

## 1 INTRODUCTION

With the proliferation and development of online education, student interaction analysis systems on platforms have emerged as pivotal tools for enhancing teaching quality and learning efficiency. Particularly in the realms of massive open online courses (MOOCs) and remote instruction, the optimization of such systems plays a critical role in understanding student behaviors, enhancing engagement, and improving course retention rates [1–4]. Against this backdrop, the integration of machine learning and computer networking technologies offers new opportunities for optimizing student interaction analysis systems on online education platforms. The formidable capabilities of machine learning in processing large datasets and identifying patterns [5, 6], combined with the efficiency of computer networking technologies in data transmission and network management [7, 8], jointly facilitate the enhancement of student interaction analysis systems in online education platforms.

Currently, the focus of related research is on enhancing the system's responsiveness and precision to ensure effective analysis of student interaction data [9–11]. This is particularly important in vocational English education because it enables the timely identification of students' personalized needs in English learning, adjustment of teaching strategies, optimization of resource allocation, and an increase in student satisfaction with their English proficiency. On the other hand, with the increase in the number of users on online education platforms, particularly those for vocational English learning, there has been a significant rise in the volume of data generated. This not only demands higher data transmission speed and stability but also requires that the interactive analysis systems for vocational English education can process and transmit large volumes of data quickly and accurately. This is crucial for supporting real-time or near-real-time teaching feedback and data analysis, which can assist teachers in promptly understanding students' progress and challenges in various English skills [12].

However, existing research methods exhibit numerous flaws and limitations. For instance, in the detection of anomalous interactions, traditional approaches might not effectively differentiate between normal fluctuations and actual anomalies, leading to false positives or negatives [13–16]. Additionally, current data transmission optimization techniques often overlook the unique data characteristics of online education, such as the high temporal relevance of interaction data. This oversight makes it challenging to maintain efficient and stable transmission under high-load conditions. These issues constrain the effectiveness of student interaction analysis systems, which in turn adversely affects the quality of teaching and learning.

In response to these challenges, this study proposes a series of innovative optimization methods. Firstly, the method for detecting anomalous student interactions has been refined by integrating a quadratic detection approach for time series anomaly detection with a multifactorial time series forecasting algorithm, significantly enhancing the accuracy and efficiency of anomaly detection. Secondly, to address potential congestion issues during data transmission, a novel flow model for a congestion control algorithm based on a proportional load differential strategy has been introduced. This innovation effectively enhances the performance of data transmission within online education platforms. These optimization methods not only enhance the practicality and accuracy of student interaction analysis systems but also provide more stable and efficient data support for online education platforms, thereby improving the overall quality of educational instruction.

## 2 OPTIMIZATION OF ANOMALOUS STUDENT INTERACTION DETECTION METHODS IN ONLINE EDUCATION PLATFORMS

The purpose of detecting anomalous interactions among students on online education platforms primarily encompasses two aspects. First, identify abnormal patterns in student interaction data, such as extremely low engagement, unusual discussion frequency, or abnormal learning behaviors. These indicators suggest potential cognitive impairments, a lack of motivation, or other learning barriers. Timely identification of these conditions is crucial for guiding instruction and providing personalized education. Secondly, to ensure the quality and accuracy of data, it is important to address technical faults or operational errors that may result in incorrect data collection, such as duplicate posts or interaction records with inaccurate timestamps. In learning analytics, erroneous data can distort teachers' and researchers' true understanding of student interactions, thereby affecting the effectiveness of teaching decisions and intervention measures.

In the novel method proposed in this manuscript, the DBSCAN clustering algorithm is initially utilized to classify student interaction data into distinct clusters by setting appropriate parameters. Interactions not fitting into any cluster are marked as potential anomalies. These represent isolated learning behaviors or irregular participation patterns. Subsequently, the extended forest-MIE model is employed to train and determine the optimal outlier ratio threshold, thereby identifying anomalous interactions within clusters, such as abnormal discussion rhythms or content.

When detecting anomalous interactions using the DBSCAN algorithm, the setting of the *Min Pts* parameter is crucial as it directly influences the density threshold of clustering, determining which interactions are considered anomalous. Given the complexity and multidimensionality of student interaction data, the selection of *Min Pts* should consider not only the dimensions of the data but also the diversity and density of interactions. Generally, *Min Pts* should be set to a value greater than the dimension of the data to ensure the effectiveness of clustering. In the analysis of student interactions, the "dimensions" encompass not only time points but also types of interactions, frequency, and the number of participants, among other factors. Therefore, the setting of *Min Pts* should be based on a comprehensive consideration of these dimensions. As interactions in an online learning environment are usually more complex than other time series data scenarios, such as electrical power time series data, parameter settings need to be more refined to reflect the diverse characteristics of student behaviors. Thus, the value of *Min Pts* is often higher than that in power time series data scenarios to capture subtle interaction patterns and potential anomalous behaviors. After determining the *Min Pts* parameter, a grid search combined with the Calinski-Harabasz (CH) score is used to find the optimal $\varepsilon$ value for the DBSCAN algorithm. Assuming the dataset is represented by $R$, the size of $R$ by $v_R$, the number of clusters obtained by grid search clustering by $j$, the set of points in cluster $w$ by $Z_w$, the cluster center of $w$ by $z_w$, the center of $R$ by $z_R$, and the number of points in $w$ by $v_w$. The covariance within a cluster is represented by $Q_j$, and the covariance between clusters by $Y_j$. The formulas are as follows:

$$Q_j = \sum_{w=1}^{j} \sum_{a \in Z_w} (a - z_w)(a - z_w)^S \tag{1}$$

$$Y_j = \sum_{w=1}^{j} v_w (z_w - z_R)(z_w - z_R)^S \tag{2}$$

In the student anomalous interaction analysis system for online education platforms, the training and evaluation process of the iForest-MIE model must specifically consider the characteristics of student interaction data. This data typically includes multidimensional interaction features such as frequency, duration, and content, rather than just one-dimensional time series data. The following describes the training and evaluation process of the iForest-MIE model:

Initially, the model's training process begins by randomly setting a proportion of outliers, which represents the expected proportion of anomalous interaction data. Subsequently, the tree is constructed by extending the multidimensional student interaction data to a dimensionality of *max_depth*. Here, *max_depth* is typically set to the logarithm of the number of sample points to account for the multidimensional nature of interaction data. This dimensionality includes not only quantitative aspects but also various dimensions, such as types of interaction and timing.

The dataset is divided into two subsets by randomly selecting a dimension from the dataset and then randomly choosing a split value within the range of that dimension. This process is recursively conducted until the tree reaches *max_depth* or the data can no longer be further divided. During this process, the length of the path traveled by the sample points is calculated, and the average path length is determined.

Subsequently, a threshold is determined based on the predetermined proportion of outliers. This threshold is determined by analyzing the distribution of average path lengths, usually selected from a high percentile of lengths, which enables the differentiation of anomalous interactions from normal ones. Additionally, the absolute difference between the critical normal values and critical outliers is calculated, serving as an important metric for subsequent evaluation. Figure 1 illustrates the separation between anomalous and normal data points.

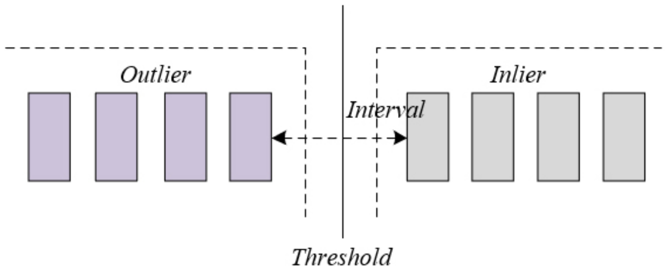



**Fig. 1.** Separation between anomalous and normal data points

Finally, by adjusting the preset proportion of outliers, the aforementioned training and calculation process is repeated. Each iteration aims to identify a new threshold and calculate the corresponding difference value. By comparing these difference values, an optimal proportion of outliers can be identified. This is the point at which the difference value reaches its maximum, signifying the model's enhanced ability to distinguish between anomalous interaction behaviors at this proportion.

On online education platforms, the data on students' interactive behaviors constitutes a multidimensional, complex, and dynamically changing time series dataset. These data not only include the timestamps and durations of students' learning activities on the platform but also encompass various interaction features such as their activity levels, engagement, and homework submission status. Traditional singular models often struggle to capture all relevant features or to adequately adapt to the dynamic changes in complex time series data, which results in limited predictive accuracy. To improve the accuracy and robustness of detecting anomalous student interactions, this paper explores the network structure of the convolutional neural network-long short-term memory (CNN-LSTM) model with an attention mechanism. It then develops a combined predictive model that includes attention-TCN and XGBoost. The attention-TCN model, with its attention mechanism, can effectively identify and assign varying levels of importance to different time steps. This capability is particularly crucial for analyzing temporal sequence features in student interaction behaviors. In contrast, the XGBoost model excels in analyzing the significance of features across dimensions and their nonlinear relationships. Through the inverse error method, the predictive results of these two models are combined with weights. A greater weight is assigned to the model with a smaller prediction error, while a lesser weight is assigned to the model with a larger prediction error. This optimization enhances the predictive performance of the combined model. Specifically, suppose the real-time time series prediction value of a certain variable factor is represented by $d_{IT}(s = m, 2 \ldots v)$, with the observation time denoted by $s$. The weighting coefficients of the combination method are represented by $Q = [q_1, q_2]$, where the weighting coefficient in the $u$-the model is denoted by $q_u$, and the final weighted output of the combined model is denoted by $d_s$, then the calculation process is as follows:

$$d_s = q_1 d_{1s} + q_2 d_{2s}, s = 1, 2, \ldots, v \tag{3}$$

$$q_1 = \frac{\omega_2}{\omega_1 + \omega_2} \tag{4}$$

$$q_2 = \frac{\omega_1}{\omega_1 + \omega_2} \tag{5}$$

## 3 OPTIMIZATION OF STUDENT INTERACTION BEHAVIOR DATA TRANSMISSION ON ONLINE EDUCATION PLATFORMS

To optimize the transmission of student interaction behavior data on online education platforms, it is essential to thoroughly consider the distinctive characteristics of data transmission on such platforms. For instance, the immediacy requirements for real-time interactive data are higher than those for non-real-time data; therefore, a higher transmission priority can be assigned to the former. The adopted methods must scientifically address challenges of fairness and efficiency in data transmission while enhancing the stability and reliability of data transfer, thereby optimizing the user experience. From a technical implementation perspective, the selection of congestion control algorithms can enhance compatibility with existing network architectures and facilitate integration with educational platform systems. For this purpose, a congestion control algorithm based on a proportional load differential strategy is proposed. This strategy dynamically adjusts data flows based on the current level of network congestion and the priority of the data streams, allocating bandwidth resources accordingly.

The proposed flow model utilizes explicit congestion notification (ECN) markings as a timely feedback mechanism, dynamically adjusting transmission rates based on the real-time load of data streams in the network and preset priorities. Therefore, this study first introduces the application of a flow model based on ECN markings in this scenario. ECN is a network congestion avoidance mechanism that, by setting specific markings in the ECN field of the Internet protocol (IP) header, allows network devices to notify the sender and receiver of data before congestion occurs, rather than simply dropping packets. This is particularly important for delay-sensitive online education platforms. It can reduce packet loss and retransmissions caused by congestion, thereby improving the efficiency and stability of data transmission.

The primary distinction between online education platforms and other data transmission optimization scenarios lies in the real-time requirements of interactive data. During teaching processes, such as live video streaming and real-time discussions, data packets must arrive as quickly as possible to avoid impacting teaching quality and student experience. Therefore, congestion control algorithms must be able to respond promptly to changes in network status and prioritize data streams that are sensitive to delays. This necessitates that congestion control algorithms not only effectively alleviate congestion but also adjust without compromising the quality of real-time interactions.

The traditional transmission control protocol (TCP) manages congestion by adjusting the congestion window size using the additive increase multiplicative decrease (AIMD) algorithm to control the data transmission rate. However, in online education settings, this control mechanism can lead to delays in the transmission of real-time interactive data, which can impact the quality of teaching. A congestion control rate mechanism based on ECN markings provides an improved congestion handling solution for online education platforms, as detailed in Figure 2. The implementation steps of this mechanism are as follows: a) The sending end transmits data packets, such as video streams and real-time Q&A, into the network. b) Network devices set queue length thresholds and monitor the data flow passing through them. c) When the data flow causes the device queue length to exceed the threshold, switches mark the passing data packets as congestion experienced (CE) instead of discarding them. This is particularly important for online education platforms because packet loss directly impacts the continuity and quality of teaching activities. d) The receiving end, upon receiving data packets marked as CE, includes an ECN marking in the acknowledgment (ACK) and sends it back to the sending end. e) Upon receiving the ACK with the ECN marking, the sending end calculates the marking ratio to assess the degree of network congestion and adjusts the size of the congestion window. This delicate adjustment process is crucial for online education platforms as it supports the rapid and reliable transmission of real-time interactive data. f) Based on the level of congestion, the sending end adjusts the transmission rate to prioritize the smooth transmission of real-time interactive data while ensuring overall network stability and preventing service quality degradation caused by congestion.

Suppose $d_1(\beta)$ and $d_2(\beta)$ are functions of the congestion level $\beta$. Specifically, the sliding window increases by $(1 - d_1(\beta))$ per round-trip time ($RTT$) when there is no congestion in the network. In the presence of congestion, the sliding window decreases by a factor of $d_2(\beta)$. The following formula illustrates the window change in the ECN marking-based congestion avoidance phase:

$$q = \begin{cases} q + (1 - d_1(\beta)) \text{ without congestion} \\ q + (1 - d_1(\beta)) \text{ with congestion} \end{cases} \tag{6}$$

Assuming there are $V$ senders, such as teaching servers, and teacher and student terminals, connected to a receiving end through an ECN-marking-supportive switch, with the receiving end being a central server. Each sender calculates its congestion window based on the formula above.

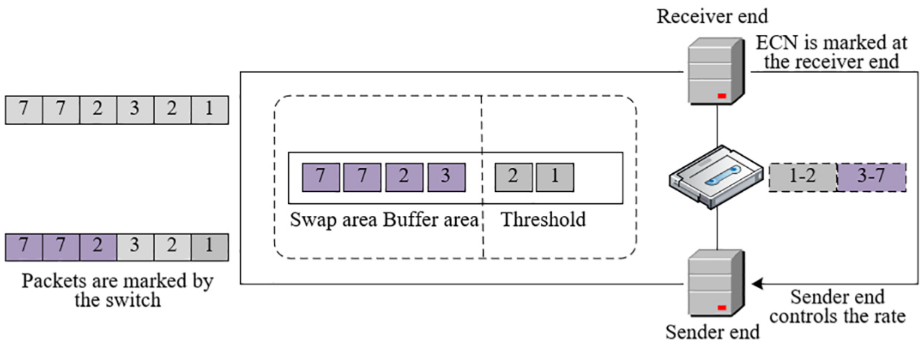



**Fig. 2.** Schematic of the congestion control rate mechanism based on ECN markings

In this model, the adjustment of the congestion window follows Eq. (7), where the increase of the congestion window is a function related to the degree of network congestion $\beta$, influenced by each *RTT*, represented as $(1 - d_1(\beta))$. The reduction of the congestion window is another function related to the congestion level $\beta$, represented as $q_u(s)d_2(\beta)$, multiplied by the probability function of congestion occurrence $o(s - E^*)$. Here, if $o(s - E^*) = 1$, it signifies that an ACK with an ECN marking was received within one RTT prior, indicating congestion; if $o(s - E^*) = 0$, it implies that no ACK with ECN marking was received, indicating no network congestion. In the context of online education platforms, where the smooth flow of real-time interactive data is prioritized, the functions $d_1(\beta)$ and $d_2(\beta)$ can be specifically designed to be more sensitive in detecting signs of congestion, thereby reducing the delay in real-time interactive data.

$$\frac{dq_u}{ds} = \frac{1 - d_1(\beta)}{E(s)} - \frac{q_u(s)d_2(\beta)}{E(s)} o(s - E^*) \tag{7}$$

Equation (8) describes the change in the switch queue length, which is the sum of the window sizes of all sending ends divided by their respective RTT minus the switch's processing rate $Z$. In online education platforms, the length of the switch queue should not exceed a certain threshold to prevent packet loss and delay, which is essential for ensuring the seamless flow of real-time interactions. Assuming the packets sent to the switch are represented by $\sum_{u=1}^{V} q_u(s)/E(s)$, and the switch queue length is represented by $\sum_{u=1}^{V} q_u(s)/E(s) - Z$, the calculation formula is as follows:

$$\frac{dw}{ds} = \sum_{u=1}^{V} \frac{q_u(s)}{E(s)} - Z \tag{8}$$

Lastly, Equation (9) indicates that when the switch queue length meets a certain marking condition $D(w(s))$, the switch will mark the packets with ECN $(D(w(s)) = 1)$; otherwise, it will not $(\hat{o}(s) = 0)$. Switches on online education platforms can set more sensitive queue length thresholds to be triggered before the queue length reaches a critical point that could impact the quality of data transmission.

$$\hat{o}(s) = 1_{D(o(s))==1} \tag{9}$$

Based on the aforementioned process, a more detailed description is presented for the flow model of the congestion control algorithm utilizing a proportional load differential strategy. Assuming there are $V$ concurrent data streams transmitting data to the same receiving node through a bottleneck link with a capacity of *Z pkts/sec*. $d_1(\beta)$ and $d_2(\beta)$ are both set to $\beta^*s/s_{MAX}$, indicating that the increase and decrease of the congestion window are proportional to the degree of congestion. Substituting $d_1(\beta)$ and $d_2(\beta)$ into Equation (7) yields, a new model for congestion window adjustment. When there is no congestion, i.e., no ACK with ECN marking, the window increases by $(1 - \beta^*s/s_{MAX})/E(s)$ per RTT. This means that the increase in window size is inversely proportional to the level of congestion. When the network is not congested, $\beta$ is small, and the window increases faster; conversely, it increases slower. When congestion is detected, i.e., when there is an ECN marking in the ACK, the window decreases by an amount calculated as $q_u(s)^*\beta^*s/s_{MAX}/E(s)$. The decrease in this context is directly proportional to the level of congestion; the more severe the congestion, the larger the reduction in the window size. Assuming the congestion window size of $FL_u$ is represented by $q_u$. The deadline of $FL_u$ is represented by $s_u$, and the congestion level of $FL_u$ is represented by $\beta_u$. The switch queue length is represented by $w$. RTT is represented by $E$, then the calculation formulas are as follows:

$$\frac{dq_u}{ds} = \frac{1 - \beta_u(s)\frac{s_u}{s_{MAX}}}{E(s)} - \frac{q_u(s)\beta_u(s)\frac{s_u}{s_{MAX}}}{E(s)} o(s - E^*) \tag{10}$$

$$\frac{d\beta_u}{ds} = \frac{h}{E(s)}(o(s - E^*) - \beta_u(s)) \tag{11}$$

For online education platforms, managing the switch queue length is particularly important because excessive queue lengths can cause delays, impacting the experience of real-time interactions. Therefore, it is necessary to maintain the queue at a reasonable length, which can be achieved by adjusting the size of the sending window.

$$\frac{dw}{ds} = \sum_{u=1}^{V} \frac{q_u(s)}{E(s)} - Z \tag{12}$$

In the congestion control algorithm based on a proportional load differential strategy, the switch determines whether to mark passing packets with ECN based on the disparity between the current queue length and the target queue length. When the current queue length exceeds the target, the switch starts marking the passing packets, signaling the sending end to decrease its window size, thereby easing congestion.

$$\hat{o}(s) = 1_{\hat{w}(s)>1} \quad (13$$

In the congestion control algorithm based on a proportional load differential strategy, $s_{MAX}$ is a critical parameter that represents the maximum transmission time the network system can tolerate. The optimization of student interactive behavior data transmission on online education platforms needs to consider real-time and interactive aspects. Therefore, when selecting the value of $s_{MAX}$, the specificity of classroom interactions must be considered comprehensively. Since online education emphasizes real-time interactions, such as live video streaming or real-time online testing, the value of $s_{MAX}$ should be set lower to ensure that data transmission delays can still be controlled at a relatively low level, even in the event of network congestion. This will ensure the continuity and interactivity of educational activities.

Considering the characteristics of online education, the configuration of $s_{MAX}$ should also consider the priority of various types of interactive data. For example, real-time audio and video data should have a higher priority and lower $s_{MAX}$ value compared to post-lesson homework submissions or forum discussions, as the former are more sensitive to delays. Additionally, the setting of $s_{MAX}$ should flexibly consider students' geographical locations, network conditions, and network congestion at different times to ensure all students receive a relatively fair and high-quality learning experience.

In the algorithm, each data stream's congestion control window $q_u$ and congestion indication variable $\beta_u$ can reach a fixed point, where all data streams are in a state of congestion, i.e., $\beta_u = 1$. In this state, all packets will be marked to indicate congestion, and each data stream will maintain a fixed congestion window size. In the flow model of the congestion control algorithm based on a proportional load differential strategy, the fixed point possessed by a flow is represented by $(q_u, \beta_u) = (s_{MAX} - s_u, 1)$. Assuming $V$ streams start simultaneously ($s_{MAX} > s_1 > s_2 > \ldots > s_v$), where the deadlines of streams $FL_1$~$FL_v$ are represented by $s_1$~$s_v$. The queue length at moment $s$ is represented by $w(s)$, then:

$$\frac{s_{MAX} - s_1}{s_1} + \frac{s_{MAX} - s_2}{s_2} + \ldots + \frac{s_{MAX} - s_v}{s_v} < w(s) + Z \times RTT \tag{14}$$

For $V$ data streams, if the size of the congestion window per RTT does not exceed 1, then it is satisfied that $w(s) < J + V$. Assuming $s_{MAX} > s_1 > s_2 > > s_v$, then $s_{MAX}$ can be calculated using the following formula when implementing the proportional load differential strategy:

$$s_{MAX} < \frac{s_1 \times (J + Z \times RTT + 2V)}{V} \tag{15}$$

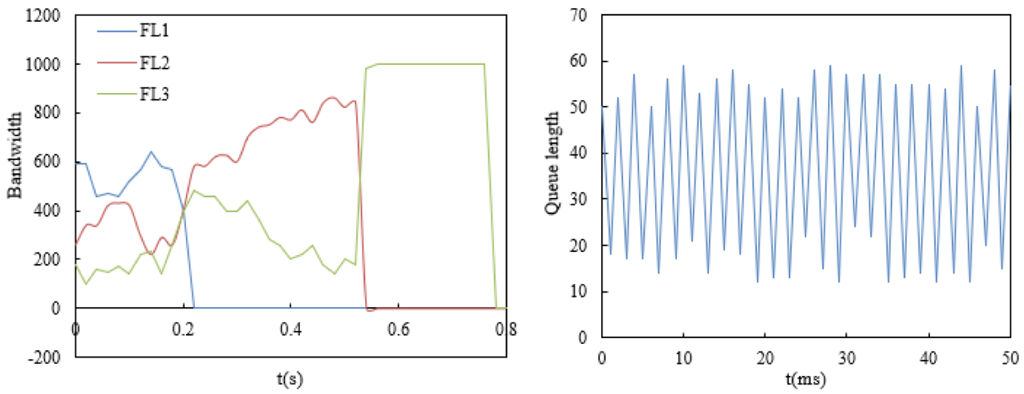
## 4 EXPERIMENTAL RESULTS AND ANALYSIS

Table 1 clearly demonstrates the performance comparison of different methods for detecting anomalous student interactions on online education platforms. With measurement metrics including mean absolute error (MAE), mean squared error (MSE), root mean absolute error (RMAE), and mean absolute percentage error (MAPE), the model proposed in this study exhibits significant superiority compared to other methods. The error metrics for LSTM, RNN, and Attention-RNN models are higher, while the TCN model incorporating the attention mechanism and the model combining XGBoost with Attention-TCN show improvements across all four metrics. Particularly, the XGBoost+Attention-TCN+XGBoost+Attention-TCN model outperforms the standalone Attention-TCN model in all metrics. However, the model proposed in this study further reduces errors across all metrics. For instance, it lowers MAE to 22.87, MSE to 1124.89, RMAE, and MAPE to 32.57 and 0.37% respectively. This indicates the higher accuracy and reliability of this model in detecting anomalous student interactions. Upon analyzing the data from Table 1, it can be concluded that the model proposed in this paper surpasses existing methods in both accuracy and efficiency in detecting anomalous student interactions. Lower MAE and MSE indicate smaller prediction errors for this model, while lower RMAE and MAPE reflect the better robustness of the model against data fluctuations and outliers.

These enhancements are credited to the model's combination of the robust learning capability of XGBoost with the temporal processing advantage of Attention-TCN, providing a more sophisticated approach for anomaly detection. Especially in the context of online education, this model is capable of accurately and promptly identifying anomalous student interactions. It provides the opportunity for timely intervention measures, which can significantly enhance the quality of interaction and teaching effectiveness in online education.

**Table 1.** Comparison of detection errors among different anomalous student interaction detection methods on online education platforms

| Model | MAE | MSE | RMAE | MAPE |
|---|---|---|---|---|
| LSTM | 74.56 | 10214.26 | 101.23 | 1.14% |
| RNN | 62.31 | 6985.23 | 81.24 | 0.98% |
| Attention-RNN | 51.24 | 5874.23 | 75.36 | 0.78% |
| TCN | 72.58 | 8526.31 | 91.45 | 1.24% |
| Attention-TCN | 46.59 | 3895.36 | 62.58 | 0.77% |
| XGBoost+Attention-TCN | 25.68 | 1547.23 | 41.23 | 0.42% |
| XGBoost+Attention-TCN+ XGBoost+Attention-TCN | 24.56 | 1325.25 | 36.87 | 0.41% |
| The model proposed in this study | 22.87 | 1124.89 | 32.57 | 0.37% |




**a) Bandwidth** **b) Queue length**

**Fig. 3.** Flow differentiation test in a real testing environment

In tests conducted in a real environment, the student interaction behavior data transmission within online education platforms was optimized through the implementation of the in-cast congestion control for TCP (ICTCP) protocol, which is based on data center TCP (DCTCP), and a congestion control algorithm that relies on a proportional load differential strategy in the Linux kernel version 3.2.61. Within a small data center network environment consisting of two switches and six servers, the experiment simulated data transmission scenarios with different traffic deadlines. Flow sizes were set to 10MB, 30MB, and 50MB, corresponding to deadlines of 300ms, 600ms, and 900ms, respectively. Figure 3 displays the bandwidth variation over time for each flow and the queue length on the switches during the experiment. The results indicate a clear distinction in bandwidth usage among flows with different deadlines, and the switch queue length was effectively maintained below the

38KB threshold. This demonstrates the effectiveness of the optimization strategy in controlling congestion and ensuring data transmission performance. Based on the experimental data, it can be concluded that the congestion control algorithm utilizing a proportional load differential strategy can effectively distinguish between traffic with varying deadlines in an actual network environment. This ensures the efficient transmission of student interaction data on online education platforms. This differentiation ensures that critical interaction data can be delivered on time, meeting the deadline, which is crucial for enhancing students' real-time interaction experience. Additionally, the algorithm can limit the length of the switch queue, reducing packet delays and losses caused by congestion, thereby enhancing the reliability of data transmission. These results confirm that the proposed optimization strategy not only meets the real-time requirements of student interaction data on online education platforms but also maintains network stability and efficiency under high load conditions, thereby supporting the operation of an effective student interaction analysis system.

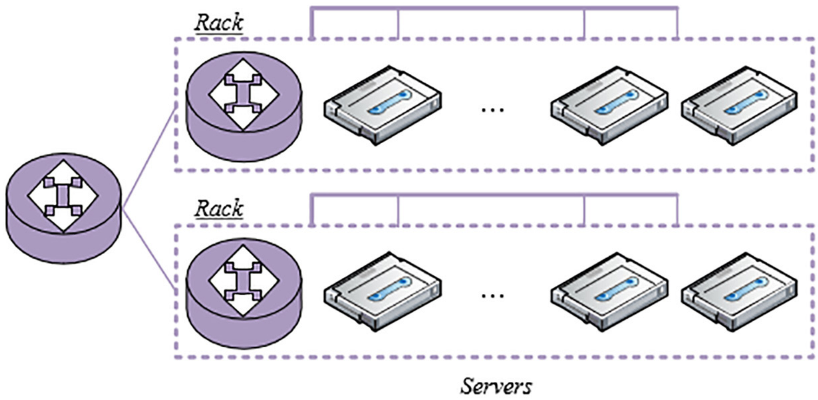



**Fig. 4.** Topology of the online education platform testbed used in the experiments

In the small online education platform testbed topology shown in Figure 4, performance tests were conducted on the congestion control algorithm based on a proportional load differential strategy. During the experiments, connections were randomly established between servers to send data streams. The occurrence of streams missing their deadlines, including average, maximum, and minimum values, was recorded to evaluate the effectiveness of the congestion control algorithm. The experimental results (see Figure 5a) indicate that, when implementing this algorithm, the proportion of streams missing their deadlines significantly decreased compared to traditional ICTCP and DCTCP. There was a reduction of about 50% compared to DCTCP and 30% compared to ICTCP. Moreover, in simulated online learning interaction data application scenarios, the performance of the congestion control algorithm based on a proportional load differential strategy (see Figure 5b) also surpassed traditional algorithms. It exhibited about a 30% performance improvement over DCTCP and approximately a 20% improvement over ICTCP. It can be concluded that the congestion control algorithm based on a proportional load differential strategy exhibits significant advantages in handling student interaction behavior data on online education platforms, especially in scenarios with high loads and streams of varying deadlines. The algorithm's penalty mechanism effectively distinguishes between streams with varying deadlines, guaranteeing the successful transmission of more data streams before their deadlines, even under highly congested network conditions. This is crucial for providing real-time interactions and high-quality user experiences on online education platforms. It ensures that critical interaction data can arrive within the required timeframe, thereby enhancing the accuracy and timeliness of data transmission. In summary, this congestion control algorithm not only enhances the data transmission performance of online education platforms but also strengthens the system's stability and reliability under various network conditions.

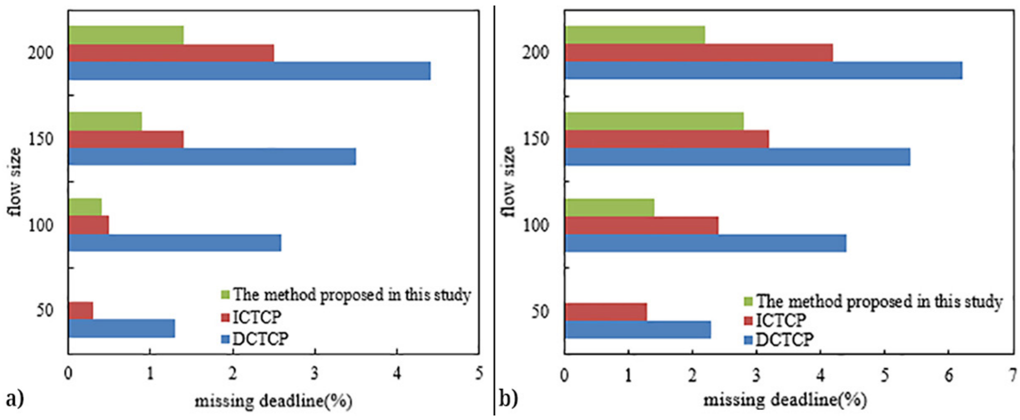



**Fig. 5.** Comparison of missing deadlines between different data transmission optimization methods for behavioral data

In the process of conducting in-depth testing of the congestion control algorithm based on a proportional load differential strategy, experimental data were obtained by configuring each server to transmit data streams to random destination nodes for a duration of two minutes, repeated 20 times. The results, presented in Figure 6a, show that compared to traditional DCTCP and ICTCP algorithms, this strategy could reduce the average completion time of streams by approximately 20% and 10%, respectively. Moreover, when testing the performance of this strategy in simulated online learning interaction data application scenarios, query requests were sent from the root node server to endpoint nodes. The endpoint nodes sent a response data stream of size *K* upon receiving the requests, where *K*'s size increased from 50KB to 200KB. Observations from the results in Figure 6b indicate that the congestion control algorithm could further reduce the average completion time of streams, achieving reductions of 20% and 10% compared to DCTCP and ICTCP, respectively. This performance improvement is attributed to the penalty function adopted by the algorithm. It can more precisely differentiate between traffic of various sizes and priorities under network congestion, effectively controlling congestion even under high-load conditions. It can be concluded that for optimizing student interaction behavior data transmission on online education platforms, the congestion control algorithm based on a proportional load differential strategy has a significant optimization effect. This strategy, through its unique penalty function, offers more detailed traffic differentiation when network congestion occurs. This ensures that critical interaction data can be transmitted quickly and accurately, even in high-load environments. The experimental results demonstrate that this optimization strategy not only significantly reduces the average completion time of data streams but also maintains stable performance improvements under various network load conditions. For online education platforms, this entails offering a smoother and more efficient student interaction experience. It ensures that data requiring robust real-time interaction, such as video, audio, and real-time feedback, is processed promptly. This significantly enhances the quality and effectiveness of online education.

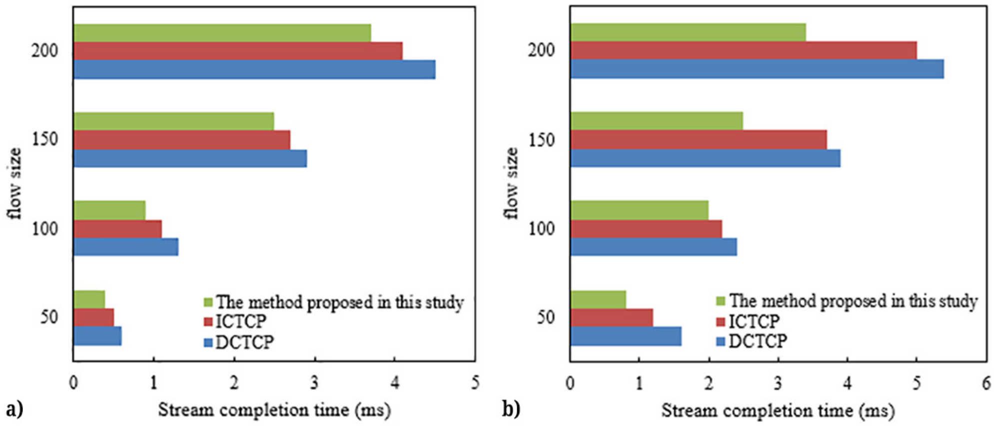



**Fig. 6.** Comparison of stream completion time between different data transmission optimization methods

## 5 CONCLUSION

This study has addressed two critical technical issues in online education platforms: the detection of anomalous student interactions and congestion control during data transmission. By integrating quadratic detection for time series anomaly detection with multifactorial time series forecasting algorithms, a significant enhancement in the accuracy and efficiency of detecting anomalous student interaction behaviors was achieved. This enhancement is crucial for ensuring the quality of online educational interactions and enabling timely intervention. Furthermore, the congestion control algorithm based on a proportional load differential strategy proposed in this study effectively reduced the average completion time of data streams and maintained good performance under high load conditions. This enhancement further improves the stability and user experience of online education platforms.

The methods proposed in this study not only optimized existing technical challenges but also provided systemic improvement solutions for real-time interactions and data transmission on online education platforms. The enhanced accuracy in detecting anomalous interactions aids in precisely identifying student needs, enabling timely adjustments to teaching strategies. Simultaneously, enhancements in the congestion control algorithm ensure efficient data communication, delivering higher-quality services to online education. However, despite the exemplary performance in experiments, actual deployment might be affected by the complexity of real-world environments, such as network instability and diverse student behavior patterns. These factors could lead to discrepancies between actual effects and experimental results. Moreover, the computational complexity and resource consumption of the model also need further consideration in practical applications.

Future research directions could include further optimizing algorithms to adapt to more dynamic and unstable network environments. Additionally, researchers could explore more complex student behavior patterns and educational scenarios. Additionally, research could be extended to other types of online platforms, such as virtual offices or remote medical services, which encounter similar technical challenges. Finally, the aspects of energy savings and sustainability of algorithms also warrant further exploration to reduce the operational costs of online platforms and achieve environmentally friendly online services.

## 6 REFERENCES


[1] A. K. Singh, S. Kumar, S. Bhushan, P. Kumar, and A. Vashishtha, "A proportional sentiment analysis of MOOCs course reviews using supervised learning algorithms," *Ingénierie des Systèmes d'Information*, vol. 26, no. 5, pp. 501–506, 2021. https://doi.org/10.18280/isi.260510

[2] W. Song, C. Zhang, and M. Gao, "Analysis method for teacher-student interaction in online English courses," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 9, pp. 170–183, 2022. https://doi.org/10.3991/ijet.v17i09.31371

[3] A. Almiman and M. T. B. Othman, "Predictive analysis of computer science student performance: An ACM2013 knowledge area approach," *Ingénierie des Systèmes d'Information*, vol. 29, no. 1, pp. 169–189, 2024. https://doi.org/10.18280/isi.290119

[4] Y. Q. Zhang and A. Mangmeechai, "Exploring the factors of undergraduate learners' engagement and knowledge sharing for sustainable HMOOC learning," *International Journal of Sustainable Development and Planning*, vol. 17, no. 3, pp. 1007–1015, 2022. https://doi.org/10.18280/ijsdp.170332

[5] A. B. Feroz Khan and S. R. A. Samad, "Evaluating online learning adaptability in students using machine learning-based techniques: A novel analytical approach," *Education Science and Management*, vol. 2, no. 1, pp. 25–34, 2024. https://doi.org/10.56578/esm020103

[6] W. Wang, "Student behavior simulation in English online education based on reinforcement learning," *International Journal of Interactive Mobile Technologies*, vol. 17, no. 22, pp. 136–151, 2023. https://doi.org/10.3991/ijim.v17i22.45303

[7] S. E. Ayman and A. Abo El Rejal, "Ontology and machine learning-based recommender system for teacher resource personalization," *Education Science and Management*, vol. 1, no. 3, pp. 145–157, 2023. https://doi.org/10.56578/esm010303

[8] M. Chen, X. Liang, and Y. Xu, "Construction and analysis of emotion recognition and psychotherapy system of college students under convolutional neural network and interactive technology," *Computational Intelligence and Neuroscience*, vol. 2022, 2022. https://doi.org/10.1155/2022/5993839

[9] Y. He and Y. Gong, "Improving the quality of online learning: A study on teacher-student interaction based on network multi-modal data analysis," in *Proceedings of the 5th International Conference on Big Data and Education*, Shanghai, China, 2022, pp. 311–318. https://doi.org/10.1145/3524383.3524388

[10] O. Oladipupo and S. Samuel, "A learning analytic approach to modelling student-staff interaction from students' perception of engagement practices," *IEEE Access*, vol. 12, pp. 10315–10333, 2024. https://doi.org/10.1109/ACCESS.2024.3352440

[11] Y. Liu and T. Ren, "Construction of an intelligent student management and evaluation information platform based on data evaluation algorithms," in *International Conference on Data Science and Network Security (ICDSNS)*, Tiptur, India, 2023, pp. 1–6. https://doi.org/10.1109/ICDSNS58469.2023.10245536

[12] V. S. Sheshadri, C. F. Lynch, and T. Barnes, "InVis: An EDM tool for graphical rendering and analysis of student interaction data," *Educational Data Mining (Workshops)*, vol. 1183, pp. 65–69, 2014.

[13] Z. Y. Zhang, Y. X. Li, Z. Y. Wang, and C. X. Li, "NIR model optimization study of larch wood density based on IFSR abnormal sample elimination," *Spectroscopy and Spectral Analysis*, vol. 42, no. 11, pp. 3395–3402, 2022.

[14] J. Wong, J. Nerbonne, and Q. Zhang, "Ultra-efficient edge cardiac disease detection towards real-time precision health," *IEEE Access*, vol. 12, pp. 9940–9951, 2024. https://doi.org/10.1109/ACCESS.2023.3346893

[15] J. Harischandra and U. Perera, "Virtual stomach visualization and a stomach tumor detection system," in *IEEE-EMBS Conference on Biomedical Engineering and Sciences*, Langkawi, Malaysia, 2012, pp. 377–382. https://doi.org/10.1109/IECBES.2012.6498053

[16] A. V. Savkin, S. C. Verma, and W. Ni, "Autonomous UAV 3D trajectory optimization and transmission scheduling for sensor data collection on uneven terrains," *Defence Technology*, vol. 30, pp. 154–160, 2023. https://doi.org/10.1016/j.dt.2023.03.020

## 7 AUTHOR

**Jinjin Wang** graduated from Northwest University of Political Science and Law with a Master's Degree in Translation, is working as the Vice Professor at Shaanxi Vocational and Technical College. Her main research focus is on English teaching in higher vocational college (E-mail: jinjinking2013@163.com; ORCID: https://orcid.org/0009-0007-0826-8430).