Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
 Volume 25, Issues 6, pp. 4913–4929, DOI 10.12694/scpe.v25i6.3404

A STUDENT EDUCATION DATA MINING METHOD BASED ON STUDENT SEQUENTIAL BEHAVIORS AND HYBRID RECURRENT NEURAL NETWORK

WEI LUO *AND QI WANG

Abstract. This research aims to propose a student education data mining method based on a hybrid recurrent neural network and improved support vector machine-decision tree algorithm through in-depth analysis of student behavior sequences. The method combines the feature extraction capability of a hybrid recurrent neural network and the nonlinear mining efficiency of support vector machine-decision tree algorithm to achieve efficient prediction of students' learning behavior and performance. Experimental results have shown that the designed method completed an APA of 91.3% and 89.2% for the HR-SDT model on the Student_1 and Student_2 datasets, respectively. The F1 score average values of the HR-SDT model reached 86.7% and 81.9%, respectively. The results indicate that the student behavior data mining method based on hybrid recurrent neural networks can accurately predict the learning behavior and performance of students, providing valuable insights and decision support for educators.

Key words: Sequence of student behavior; Recurrent neural network; Educational data mining; Learning behavior prediction; Decision tree

1. Introduction. In today's era, the rapid development of big data technology has brought unprecedented opportunities and challenges to the education sector. The advancement of information technology and continuous reforms in education have made the accumulation and storage of educational data easier and more convenient. An increasing number of educational institutions are employing student educational data for research and decision-making purposes [32]. However, traditional educational data mining (EDM) methods mainly focus on the static characteristics and academic performance of students and rarely consider the information of student behavior sequences (SBS) [16]. This information can reveal the learning status and characteristics of students from a more fine-grained perspective, understand the patterns and processes of student learning behavior sequences, and how to use these patterns for effective EDM. This is greatly important for improving the quality of education and optimizing the allocation of teaching resources [31]. Nowadays, the rapid development of deep learning technology provides powerful tools for EDM [24]. The hybrid recurrent neural network (HRNN) algorithm combines recurrent neural networks (RNNs) and other types of neural networks, which can simultaneously consider static features and sequence information, improving the expression ability and generalization performance of learning models [15]. Therefore, this study proposes a data mining model based on SBS and HRNN. This model predicts and explains student learning behavior and performance through in-depth analysis of SBS.

Compared with the research in the field of "EDM in the prediction and analysis of students' academic achievement", this study not only uses HRNN for feature extraction, but also combines support vector machine (SVM) and decision tree (DT) algorithms, i.e. SDT, to improve the efficiency of non-linear mining. Different from the former, which mainly focuses on the static characteristics and academic performance of EDM, this study pays more attention to the analysis of SBS information, which can reveal students' learning status and characteristics in more detail. By combining HRNN and various algorithms for in-depth analysis of SBS, it is found that the final EDM model can more accurately predict and explain students' learning behavior and performance. This model provides valuable insights and decision support for educators, optimizes teaching resource allocation, and improves education quality. The innovation of the research is mainly reflected in the following two aspects. Firstly, SVM is used to improve DT to shape the SDT algorithm, and the SDT data mining algorithm is constructed for SBS mining. Secondly, this study chooses HRNN instead of traditional

*Informatization Office, Henan Finance University, Zhengzhou 450046, China (Corresponding author, luowei1696@163.com)

RNN for SBS feature extraction, which can more accurately analyze and predict students' learning behavior and state. Through these two innovations, the HR-SDT model is designed in this study, which improves the ability to mine SBS and the accuracy and efficiency of prediction.

The main contributions of this study include the following three points. First, a new data mining model combining HRNN and an improved SDT algorithm is proposed, which realizes efficient prediction of students' learning behavior and performance through in-depth analysis of SBS. Second, based on the traditional RNN, HRNN is used for feature extraction, which can more accurately capture the time dependence and pattern of students' behavior. Third, the experimental results verify the effectiveness and high performance of this research method, which provides a new idea and tool for the field of EDM.

The structure of this article is divided into four parts. Part 1 introduces relevant research and summarizes the current development status of the technologies used in the study, providing theoretical preparation for this study. Part 2 is the research method, which constructs a mining model using current algorithms. Part 3 is the performance analysis, which involves conducting performance tests on the proposed model. Part 4 is the conclusion, summarizing the results and shortcomings of this study.

2. Related Works. In recent years, RNN has shown strong capabilities in processing sequence data, providing new ideas for EDM. Many researchers have begun to explore new applications of RNN and data mining. S. R. Sudharsan et al. [22] proposed a new framework for predicting customer churn through deep learning models, called Fast-RNN to address the issue of significant customer churn in the telecommunications industry. This framework predicted the possibility of Customer Churn (CC), which facilitates businesses to take various measures to retain customers who are about to be lost. This framework applied to the classification and prediction of CC and ordinary customers, and was superior to the current mainstream model in the prediction mechanism, so this model had certain progressiveness. M. Xia et al. [27] proposed an improved Stacked Gated Recursive Unit RNN (GRU-RNN) to predict renewable energy generation and load under univariate and multivariate scenarios. This method selected sensitive monitoring parameters based on correlation and forms input data, while utilizing improved training algorithms with AdaGrad and adjustable momentum to improve training efficiency and robustness. This method outperformed the advanced machine learning or deep learning methods in achieving accurate energy prediction for effective smart grid operation. S. P. Yadav et al. [29] researched the optimal architecture and algorithm usage of machine learning. After reviewing numerous literature, they summarized that the technologies discussed in machine learning are rapidly gaining ground, aiming to completely change the research and development field of speech and visual systems. In addition, the study also identified limitations and slow updates in the current field of machine learning integration, and proposed prospects for combining machine learning with mobile and embedded technologies. C. Liu et al. [12] established a data-driven model using stacked bidirectional long short-term memory RNNs to predict the remaining service life of super-capacitors. This model integrated time series processing algorithms based on traditional RNN, and the stacked network improved the data capacity and computational efficiency of the model to a certain extent. The proposed model had lower error values compared to ordinary RNNs, so the improvement of this model had practical significance. J. Zhu et al. [33] proposed a fault diagnosis model based on RNN. This model introduced a time series to handle equipment failures, thereby simulating the real-time status of equipment failures. After training, the model has shown good performance in fault diagnosis and was more advantageous compared to similar models. Therefore, this study had the potential to promote the development of intelligent detection of equipment faults. Y. Xu et al. [28] proposed a method for extracting and analyzing features of malicious domain names using deep neural networks (DNN). This method utilized the hierarchical structure of bidirectional RNN to extract effective semantic features while introducing a discriminator to detect malicious domain names, which was effective. This study also compared with other types of methods, proving that the method's detection performance was superior to most mainstream methods, so it was progressive.

Sequential data mining is also a commonly used research method in many fields today. Big data has gradually become a hot research field, and data mining has also been improved and optimized by many researchers. A. To address the application limitations of customized metal active sites and porous structures, Nandy et al. [14] proposed a Gaussian process and artificial neural network model using data mining methods. This model used natural language processing and image analysis to obtain over 2000 solvent removal stability indicators and 3000 thermal degradation temperatures. This model has enabled researchers to obtain important features of metal active sites and porous structures, making a certain contribution to the development of the industry. M. Zavarin et al. [30] proposed a comprehensive data modeling workflow, which includes search, access, interoperability, and database usage. The researchers in this process used a newly developed surface complexation/ion exchange (L-SCIE) database for centralized data processing. L-SCIE's data mining ability has been proven to be better than other data processing models, so this research is excellent. N. Aziz et al. [2] proposed a data mining method based on analysis bias. This method added a deviation correction module to the association rule mining method. This module constantly monitored whether the mining process generates outliers in the dataset processing, and obtained the same result of analysis bias by conducting numerous detection and recording operations before analyzing duplicates. After comparison, the proposed method had a smaller deviation between the results of data mining and the actual values, indicating its effectiveness. C. Sirichanya et al. [21] found that traditional data mining methods cannot interpret data at the semantic level, nor can they reveal the meaning of the data. Therefore, they proposed a framework that applies semantic data mining to data resource descriptions. This framework achieved and improved data mining performance by using domain ontology as background knowledge, breaking through the limitations of traditional data mining methods. This descriptive framework has made breakthroughs in understanding and analyzing natural language semantics and context, but there is still significant room for improvement in accuracy. P. Edastama et al. [7] conducted a study on data mining and processing using the activity and sales data of an eyewear store called Optik Nasional as an example. After using the Apriori algorithm to mine and analyze store sales data, the author found that many daily sales transactions, and even longer sales, of the store will increase due to activities. The results of this data mining would help improve marketing plans for sales and promotional products, and increase the growth of eyewear sales. H. K. Bhuyan et al. [4] attempted to use multi-objective models in data mining to address data privacy issues based on interactive computing, thus proposing an anonymity framework for data privacy. This framework achieved a balance between objects, reduced computational costs, and increased privacy. Based on the uniform distribution of noisy data and parameter α-cutoff values, the optimal values of framework parameters were obtained. The proposed model's solution controlled the amount of privacy that users can freely choose with maximum flexibility.

In addition to RNN, other deep learning models are also widely used in different fields. N. Sharma et al. [19] proposed a Siamese convolutional neural network (CNN), aiming to complete the offline signature verification task by using this network model. The method processed similar and dissimilar images and calculated the Euclidean distance between them through a network of twin cell structures with shared weights and parameters. By reducing the distance of the same signature and increasing the distance of different signatures, the authenticity of the signature was verified. The results showed that the model performed well on various data sets. In addition, custom CNN also showed significant advantages in natural language recognition. Gurumukhi script was a complex writing system that could effectively deal with word segmentation through a holistic approach to offline handwritten word recognition. T. P. Singh et al. [20] employed a CNN architecture of five convolutional layers and three pooled layers to process a dataset of 24,000 images from 500 authors of different ages and occupations. The accuracy of training and verification on this dataset reached 97.03% and 99.50%, respectively. Online learning systems have expanded significantly in recent years, especially during the COVID-19 pandemic, which has seen a significant increase in enrollment of online learners. To improve the online education environment and reduce the dropout rate, R. Kaur et al. [9] proposed an intelligent portrait system based on the user interface. By collecting parameters such as personal information, educational background, and knowledge level of learners, the system created a portrait of learners and recommended appropriate courses based on user feedback. The results showed that this method had a good performance in improving user adaptability and personalized recommendations.

In summary, data mining algorithms and neural networks have wide applications in various fields, and different algorithms are suitable for different tasks and data types. The arrival of the big data era has brought new challenges and opportunities to data mining, such as processing high-dimensional data, incomplete, and inaccurate data, etc. Therefore, this study aims to promote the development of the smart education industry by constructing a more comprehensive data analysis model built on the optimized DT and HRNN to analyze students' daily behavior.

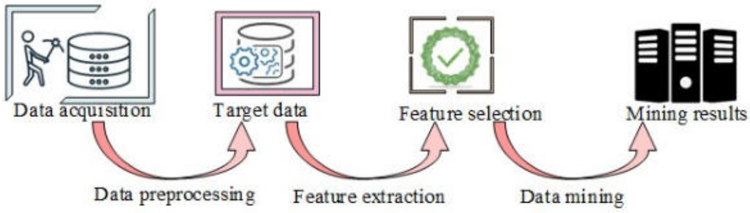

Fig. 3.1: Common Data Mining Flowchart

3. Construction of a SBD analysis model based on improved DT and HRNN. The advancement of educational technology has accumulated abundant student behavior data (SBD), but how to effectively utilize this data to provide targeted educational support remains a challenge. Therefore, this study proposes an HR-SDT model that combines DT and HRNN, aiming to explore potential patterns and patterns in SBD and provide personalized learning recommendations.

3.1. A data mining method based on DT improvement. Data mining is of great significance to the education industry, as it can help educational institutions better understand student needs, optimize curriculum design, improve teaching quality, and formulate more effective education policies. The current common data mining methods for data processing mainly include several steps: data collection, preprocessing, feature selection, data mining, and model evaluation. Figure 3.1 is the specific mining process.

The methods of data collection are relatively diverse, such as through sensors, questionnaire surveys, etc., and the methods of data collection vary in different application scenarios. Data preprocessing has a significant impact on subsequent data mining work. Especially in cases where the original data is missing or noisy, data preprocessing can greatly improve mining efficiency and quality by repairing and organizing the data. The preprocessing of data includes data cleaning, classification, integration, and data transformation, among which data transformation usually requires data normalization. The normalization calculation formula for this study is equation (3.1).

$$\mathrm{y} = \frac{\mathrm{x_i} - \mathrm{x_{min}}}{\mathrm{x_{max}} - \mathrm{x_{min}}} \tag{3.1}$$

In equation (3.1), y represents the normalized output value. $\mathrm{x_i}$ is the input raw data. i is the sequence number. $\mathrm{x_{min}}$ represents the min-value in the input raw data. $\mathrm{x_{max}}$ is the max-value in the input raw data. A mong various data mining algorithms, the finer the data preprocessing, the better the mining effect. The feature selection is the process of selecting a subset of features from the feature space, and the results of data mining vary depending on the selected features. The data mining process is the most crucial step, and different mining algorithms have different ways of processing data. This study uses DT to mine data. In the DT algorithm, each node represents a feature or attribute, each branch is a decision rule, and each leaf node means a classification or regression result. The structure diagram of the DT algorithm is shown in Figure 3.2.

The selection of partitioning attributes plays a crucial role in data mining and classification processes. This is because different attribute partitioning methods may have a significant impact on the representation of data and the performance of classifiers. In the DT algorithm, selecting the optimal partition attribute is closely related to the decrease in information entropy (IE). IE is a core concept in information theory, used to describe the uncertainty and randomness of information. In data mining and classification, IE is used to measure the degree of confusion among different categories in the dataset. If the division of an attribute can make the categories in the dataset clearer and more orderly, then this attribute has a higher IE. In the DT algorithm, selecting the optimal partitioning attribute is usually based on indicators such as information gain or Gini index. These indicators evaluate the quality of partitioning by calculating the change in IE after partitioning different attributes. The definition formula of IE is equation (3.2).

$$S(x) = -\sum P(x) \cdot \log^{P(x)} \tag{3.2}$$

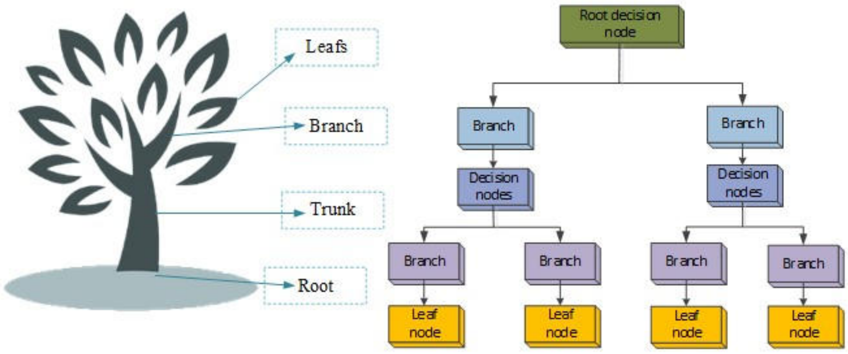

Fig. 3.2: Schematic diagram of DT algorithm structure

In equation (3.2), $S(x)$ represents the IE function. x represents the independent variable. P(x) represents the amount of information obtained for target classification. The IE from the root node to the leaf node is a decreasing process, and the IE of the leaf node is 0 . The selection of non-leaf nodes is achieved by comparing the increment of IE. The formula for IE gain is equation (3.3).

$$G(D, A) = G'(D) - G'(D \mid A) \tag{3.3}$$

In equation (3.3), $G(D, A)$ represents the information gain of feature A on dataset $D.G'(D)$ represents the IE of $D \cdot G'(D \mid A)$ represents the conditional entropy of D under the given conditions of A. In the process of constructing DT, the optimal branch node is selected by calculating the IE increment of each branch node. When all the sample data of a branch node belong to the same class, the IE increment of the branch node is 0 , and classification can be terminated at this time. In addition, if there are no remaining features to further divide sub-DT, classification can also be terminated [5]. Although DT algorithm is widely used, its advantages in processing time series are not obvious due to its poor handling of continuous data and missing values, inability to handle nonlinear relationships, and other shortcomings. Given the aforementioned shortcomings of the DT algorithm, this study introduces the SVM algorithm for its optimization. SVM, as a common data classifier, uses a hinge loss function to calculate empirical risk and incorporates regularization terms in the solving system to optimize structural risk. It is a classifier with sparsity and robustness [1]. At the same time, this algorithm has a very sound mathematical theoretical foundation and can be used for mining nonlinear and continuous data, making up for the shortcomings of the DT algorithm. The SVM algorithm classifies positive and negative classes based on the optimal hyperplane, and the formula for calculating the optimal hyperplane is equation (3.4).

$$\varphi(\mathrm{x}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \alpha_{\mathrm{i}} \mathrm{y}_{\mathrm{i}} \mathrm{x}_{\mathrm{i}} + \mathrm{b} \tag{3.4}$$

In equation (3.4), α_{i} represents the support vector point. x, y represents the sample set, which is the data that needs to be classified. n is the gross data. b represents the bias vector. After the optimal hyperplane calculation is completed, the SVM algorithm can perform inner product operations in the designated feature space, thereby accurately classifying the data [25]. The SVM algorithm can handle both linear and nonlinear data, so the hyper-planes of the algorithm are shown in Figure 3.3.

By integrating the DT algorithm and SVM algorithm to construct SDT, the fused data mining algorithm not only retains the powerful classification ability of the DT algorithm, but also compensates for the shortcomings of the DT algorithm in handling nonlinear relationships. The proposal of the SDT algorithm undoubtedly brought revolutionary changes to the field of education. This algorithm provides a solid foundation for subsequent feature extraction by finely classifying student behavior.

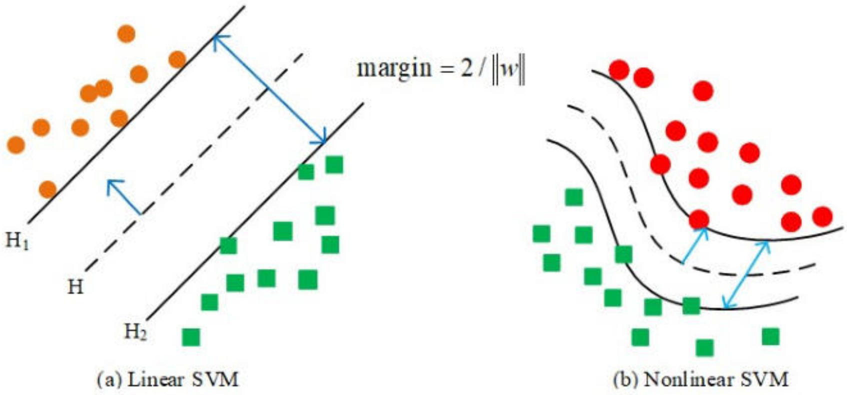

Fig. 3.3: Schematic diagram of SVM hyperplane

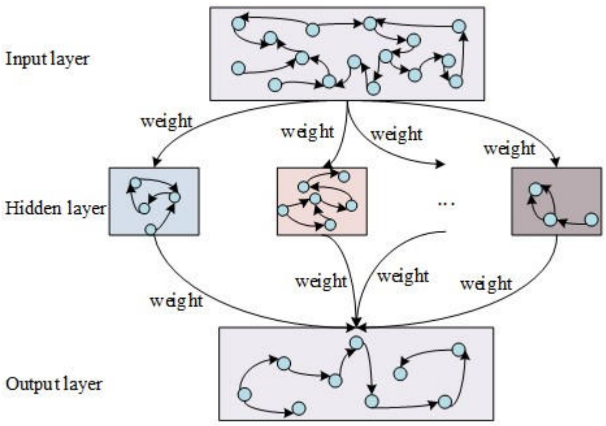

Fig. 3.4: Structure diagram of HRNN

3.2. Feature extraction algorithm based on HRNN. Due to the insufficient feature extraction ability of the above model for SBS, this study introduces a neural network algorithm to further optimize it, to better assist the model in analyzing student data and formulating more reasonable management policies. RNN has the shortcomings of being unable to capture long-term dependencies and having a large number of parameters. Therefore, this study adopts the HRNN algorithm as the core algorithm for feature extraction [17]. The HRNN algorithm is a hybrid algorithm that combines RNN and deep neural network. This algorithm is an important model for processing sequence data and can be used for tasks such as time series prediction, image recognition, speech recognition, and text generation [23]. HRNN also has strong modeling capabilities, capturing temporal dependencies, extracting features, flexibility, and efficiency, which can better handle long-term dependencies and time interval changes in SBS. HRNN consists of three common layers of neural networks, namely input, output, and hidden. The structural relationship diagram of each layer is shown in Figure 3.4.

The input layer of the HRNN algorithm contains a specific input sequence, which is a time series transformed from standardized input data. To improve the effect of the HRNN model in SBD mining, gated cycle unit (GRU) technology is used to solve the gradient disappearance problem. As a variant of RNN, GRUs can effectively capture long-time series dependencies and solve the problem of gradient disappearance of traditional RNN models when processing long-series data. Specifically, GRU not only simplifies the structure of LSTM through

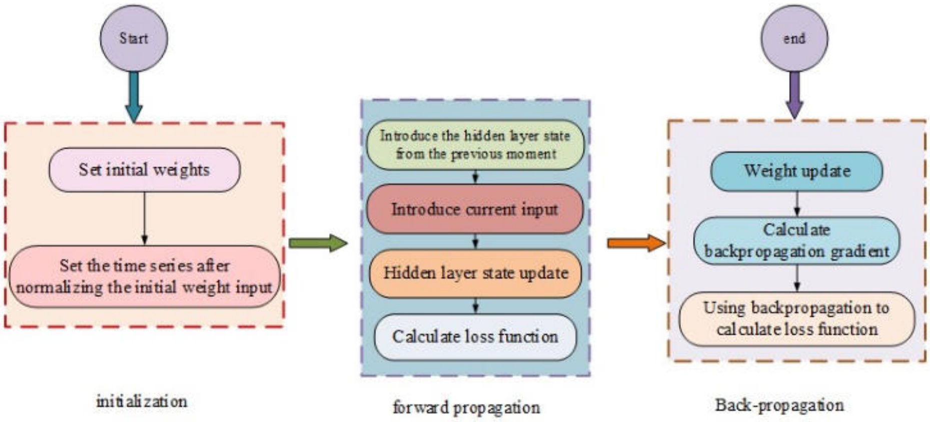

Fig. 3.5: Weight sharing flowchart

the mechanism of update gate and reset gate but also reduces the computation and has powerful time series modeling capability. In the process of HRNN model training, the GRU is introduced into the HRNN model, and then the training efficiency and model robustness are further improved by the AdaGrad optimization algorithm and mobilizable quantity mechanism. The data standardization calculation formula used is equation (3.5).

$$\mathrm{y}' = \frac{\mathrm{a_i} - \mathrm{a_{mean}}}{\mathrm{a_{stan}}} \tag{3.5}$$

In equation (3.5), y' is the standardized data output. $\mathrm{a_i}$ represents the input data of the HRNN algorithm. a_{mean} is the mean of the input dataset. a_{stan} means the standard deviation of the input dataset. The hidden layer is the core structure of HRNN, which includes specific activation functions, weight matrices, and output units for feature extraction. Due to HRNN processing time series, the hidden layer needs to be updated over time. The expression for updating the hidden layer is equation (3.6).

$$H_t = F\left(H_{t-1}, X_t\right) \tag{3.6}$$

In equation (3.6), F(.) is the activation function, which is related to the current state of the hidden layer. $\mathrm{X_t}$ is the input at time t. $\mathrm{H_{t-1}}$ means the hidden layer state (HLS) at time $\mathrm{t}-1$.
The HLS is usually composed of one or more neurons, which are responsible for transforming input data into meaningful feature representations. The expression for calculating the HLS in HRNN is equation (3.7).

$$\mathrm{H_t} = \mathrm{f_{tanh}}\left(\mathrm{w_h H_{t-1}} + \mathrm{w_{h,O} X_t} + \mathrm{B_h}\right) \tag{3.7}$$

In equation (3.7), $\mathrm{w_h}$ represents the weight matrix of the HLS. $\mathrm{w_{h,0}}$ is the weight matrix from the hidden layer to the output layer. B_h is the bias vector of the hidden layer. f_{tanh} (. represents the Tanh(.) function, which is a nonlinear function. The HLS at a certain moment is influenced by the output at that moment and the HLS from the previous moment [10]. According to this time series relationship, there is a certain connection between the output of the algorithm at a certain moment and all the inputs before that moment, but the degree to which this connection affects the results will be affected by the weight [26, 6]. The HRNN algorithm has the feature of weight sharing in the time dimension, so it requires a small number of parameters and does not require high computing power and caching of devices. The weight sharing process of this algorithm is shown in Figure 3.5.

From Figure 3.5, weight sharing is implemented in the HRNN algorithm through backpropagation. In this process, the error of each neuron is calculated by the cross-entropy function and conveyed to the pertinent upper-layer neuron in succession, after which it is transmitted back to the initial time point. Finally, the weights are updated by gradient descent. After the weight update is completed, the data features of the next time step are extracted and the features of the previous time step are output. The cross-entropy of error calculation is similar to that of IE calculation, and the expression of this function is equation (3.8).

$$\Delta \mathrm{H} = -\sum \left(\mathrm{P_i} \cdot \log^{\mathrm{q_i}}\right) \tag{3.8}$$

In equation (3.8), P is the true probability distribution, and q represents the predicted probability distribution. log represents the natural logarithm. P_i and q_i represent the probability of the true class i and the probability of the predicted class i . The error value calculated at this moment is first transmitted to the upper-level neurons that are only related to the weight value, and then propagated back to the initial time. Finally, the weight value is updated using gradient descent. After the weight is updated, it enters the feature extraction of the data at the next time step and outputs the features from the previous time step. At this time, the HRNN hidden layer output expression is equation (3.9).

$$\mathrm{y_h} = \mathrm{F_h}\left(\mathrm{w_h} + \mathrm{w_{h,0}} \cdot \mathrm{X_t}\right) \tag{3.9}$$

In equation (3.9), $\mathrm{y_h}$ represents the output of the hidden layer. $\mathrm{F_h}$ represents the hidden layer activation function. The output of the hidden layer participates in the calculation of the final output result of the algorithm. After each weight update, the expression of the final output result of the algorithm is equation (3.10).

$$\mathrm{y_0} = \mathrm{f_o}\left(\mathrm{w_0} \cdot \mathrm{y_h}\right) \tag{3.10}$$

In equation (3.10), y_0 is the output of the output layer. f_0 represents the output layer activation function, which is usually a linear function.

3.3. Construction of a student behavior analysis model integrating DT and HRNN. When analyzing student behavior, real-time and convenient data acquisition is required. Therefore, this study selects student performance ranking and card swiping behavior for analysis. As an indispensable part of students' campus life, the card swiping data can to some extent reflect the behavioral characteristics of students in various aspects such as learning, life, and consumption. By obtaining student card swiping data and using corresponding data processing techniques, the daily behavior habits of each student can be analyzed, providing valuable reference information for schools, students themselves, and relevant management departments. The data types and formats exported from the One Card Management System are shown in Figure 3.6.

After exporting data, it is necessary to preprocess and clean the data. The preprocessing, cleaning, and other processes of data can eliminate duplicate and noisy data, which is of great significance for subsequent feature extraction and mining work [8]. To facilitate data preprocessing, this study redefines various sequences. Assuming sequence $\alpha = [\mathrm{x_1}, \mathrm{x_2}, \cdots, \mathrm{x_i}, \cdots \mathrm{x_p}]$ is an SBS and $\mathrm{x_i}$ represents the i -th behavior of q devices, the expression of the relationship term between student behavior and campus devices is equation (3.11).

$$R_{p,q} = \left(\gamma_{i,j}\right)_{p,q} \tag{3.11}$$

In equation (3.11), p is the total quantity of students. q is the total amount of campus devices. $\gamma_{\mathrm{i,j}}$ indicates that student i has performed a card swiping operation on the j device.

In the above assumption, the input data is divided into student set P, campus device set Q, and SBS α. The output data is defined as the mapping relationship between the input sequence and student grades, as shown in equation (3.12).

$$\xi = \mathrm{R_{p,q}} \rightarrow \mathrm{G} \tag{3.12}$$

In equation (3.12), $R_{p,q}$ represents the relationship between student behavior and campus equipment. G represents the student's grade level. In addition to data preprocessing, data dimensions are also an important factor affecting the operation of HRNN models, so further processing of data dimensions is needed. This study utilizes the pooling method of CNN for dimensionality reduction. In CNN, pooling operations are divided into

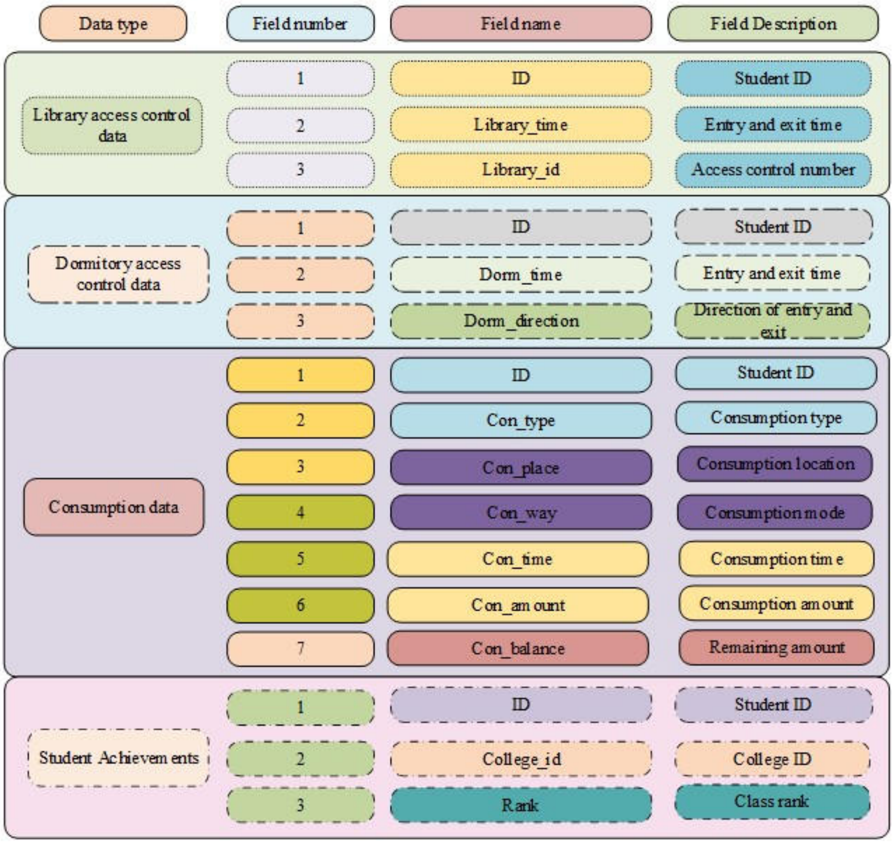

Fig. 3.6: Research data types and formats

maximum pooling and average pooling. The former preserves the most prominent features of the data, while the latter preserves the overall features of the data [3]. The calculation of maximum pooling is equation (3.13).

$$\mathrm{y_{i,j}} = \max_{\mathrm{n}=0}^{\mathrm{k}-1} \max_{\mathrm{m}=0}^{\mathrm{k}-1} \mathrm{x}_{\mathrm{i+m},j+\mathrm{n}} \tag{3.13}$$

In equation (3.13), k represents the size of the pooling kemel. x is the input matrix. $\mathrm{y_{i,j}}$ is the output matrix. Maximum pooling cannot preserve all the features of the data, so this method is suitable for feature extraction models with low accuracy requirements [13]. However, in the case of learning unified management, there is relatively little difference in behavior among students, so the maximum pooling method is not suitable. Therefore, this study adds an average pooling module to the input layer of the HRNN model to achieve the effect of preserving detailed features and reducing dimensionality. The mathematical expression for the average pooling method is equation (3.14).

$$y_{i,j} = \frac{1}{k_h \cdot k_w} \sum_{u=0}^{k_\mathrm{h}-1} \sum_{v=0}^{k_{w-1}} x_{i+u,j+v} \tag{3.14}$$

In equation (3.14), x is the input matrix. $\mathrm{y_{i,j}}$ is the output matrix. $\mathrm{k_h}$ and $\mathrm{k_w}$ represent the step size of the pooling window in the vertical and horizontal directions, respectively. u and v are intermediate variables in the calculation process. After adding the average pooling module, the structure diagram of the SBD analysis model (HR-SDT) based on improved DT and HRNN is shown in Figure 3.7.

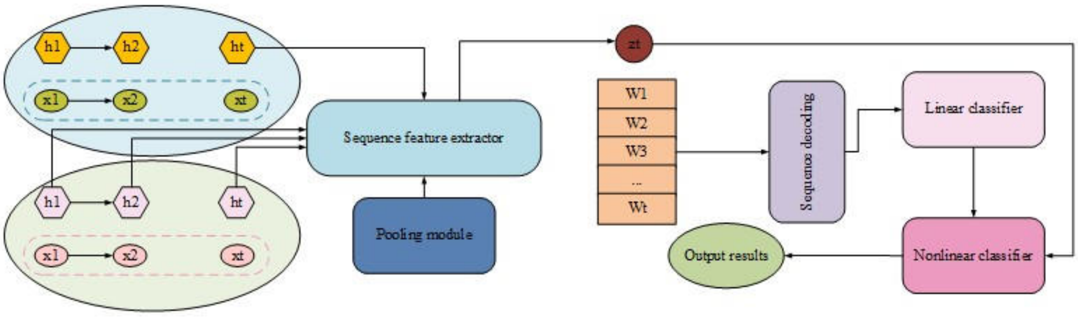

Fig. 3.7: HR-SDT Structural Diagram

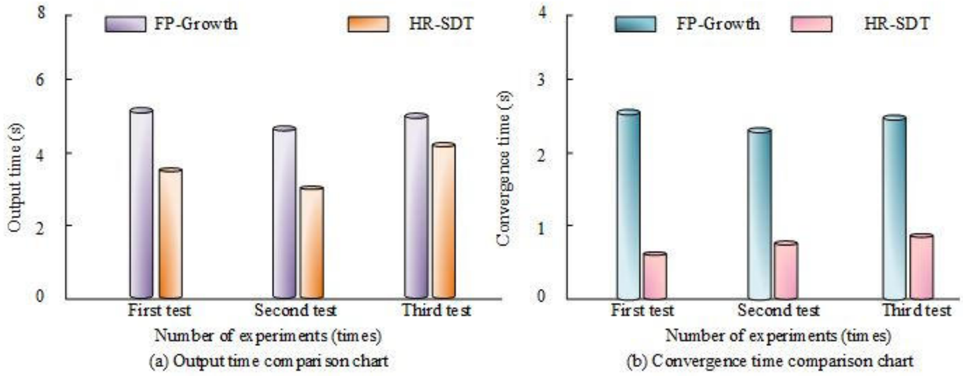

Fig. 4.1: Comparison of model output time and convergence time

The feature selection module in Figure 3.7 uses the improved HRNN method for feature processing, while the mining module uses the SDT model for data mining. This model combines the advantages of different algorithms and can quickly and accurately analyze the behavior sequence of students. The method proposed in this study first preprocesses SBS, including data cleaning, feature extraction, and sequence annotation. Then, the HR-SDT model is used to model the processed data to capture temporal dependencies and patterns in SBS. Finally, the learning behavior and performance are predicted and explained through training the model. The proposed data mining method based on SBS and HRNN provides new ideas for data mining in the field of education.

4. Performance testing and analysis of a student behavior analysis model integrating DT and HRNN. The performance verification experiment equipment environment is a desktop computer with Windows 1164 bit operating system, DDR4 16GB of memory, and NVIDIA GTX 1660 graphics card installed. The development environment is Python 1.5. The datasets used in the experiment are Student_1 and Student_2 datasets exported from a third-party database, with FP-Growth and DBSCAN models selected as control models. The convergence time and output time can indirectly reflect the data processing efficiency of the model. This study compares the convergence time and output time of the HR-SDT model with the FP-Growth model on the Student_1 and Student_2 datasets, as displayed in Figure 4.1 (a) shows the comparison of output time between two models on the Student_1 dataset. The average output time of the HR-SDT model is 3.92 seconds, which is 2.26 seconds less than the FP-Growth model. 4.1 (b) shows the comparison of convergence time between two models on the Student_2 dataset. The average convergence time of the HR-SDT model is 0.51 seconds, which is 1.63 seconds less than the FP-Growth model.

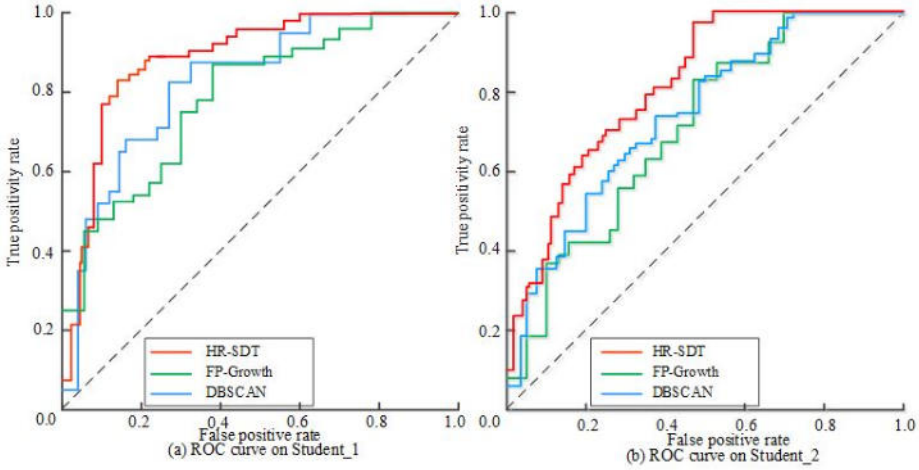

Fig. 4.2: Schematic diagram of model ROC curve

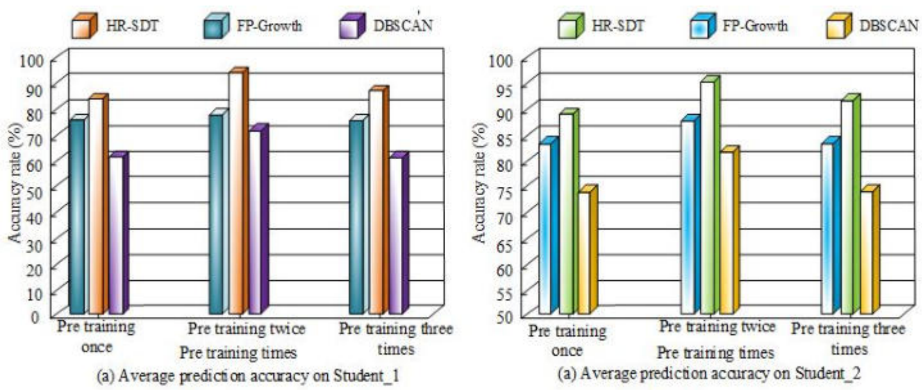

Fig. 4.3: Average prediction accuracy on different datasets

The ROC curve is a comprehensive indicator of sensitivity and specificity as continuous variables. It calculates a series of sensitivity and specificity by setting multiple different critical values for continuous variables, and usually the larger the area under the curve, the higher the accuracy. To analyze the ROC curves of the proposed model, this study compares the ROC curves of HR-SDT, FP-Growth, and DBSCAN models on Student_1 and Student_2, as shown in Figures 4.2 (a) and (b). The area enclosed by the ROC curve and reference line of the HR-SDT model on both datasets is larger than that of the other two models, therefore the HR-SDT model has higher performance.

To study the prediction accuracy of the HR-SDT, this study compares the average prediction accuracy (APA) of the HR-SDT, FP-Growth, and DBSCAN models using the Student 1 and Student_2 datasets as inputs. The experimental results are exhibited in Figure 4.3 (a) and 4.3 (b) show the relationship between the APA and pre-training times of the model on Student 1 and Student 2. The HR-SDT model achieves the highest prediction accuracy of 91.3% on Student 1 and 89.2% on Student_2. In Figure 4.3, the best effect is achieved through two rounds of pre-training, and from the third round onwards, the accuracy begins to decrease after pre-training.

Mean absolute error (MAE) is an indicator utilized to evaluate the model's accuracy, which is the MAE between the predicted value and the true value. This study compares the MAE between the HR-SDT model

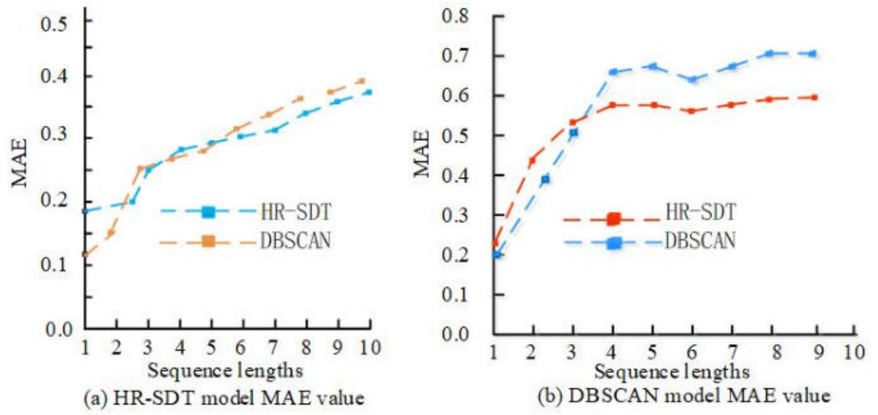

Fig. 4.4: Comparison of model MAE

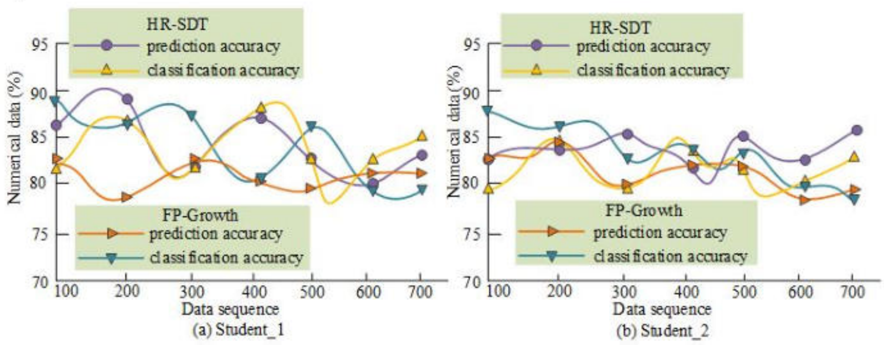

Fig. 4.5: Comparison of model prediction and classification accuracy

and the DBSCAN model on the Student_1 and Student_2 datasets, as shown in Figures 4.4 (a) and (b). The MAE of the HR-SDT model is lower than that of the DBSCAN model on both datasets.

To verify the accuracy of the proposed model in predicting and classifying student learning behavior, the Student_1 and Student_2 datasets are used as inputs. After repeated experiments, the A PA of the HR-SDT and the FP-Growth model is compared, as listed in Figure 4.5 (a) and 4.5 (b) show the comparison of prediction accuracy and classification accuracy between two models on Student_1 and Student_2. In Figure 4.5 (a), the APA of the HR-SDT model is significantly higher than that of the FP-Growth model, while the average classification accuracy is not significantly different. The trend of curve changes in Figure 4.5 (b) shows that the situation of the two models on Student_2 is similar to that on Student_1.

Loss rate is one of the commonly used indicators for model evaluation, which can reflect the error rate or proportion of loss that occurs when the model makes predictions. This study presents the relationship between the loss rate and training steps of HR-SDT, FP-Growth, and DBSCAN models on Student_1 and Student_2 in Figure 4.6 (a) and 4.6 (b) show the loss rates of the three models on the Student_1 and Student_2 datasets. The loss rates of the three models, in descending order, are DBSCAN model, FP-Growth model, and HR-SDT model.

During the operation of the model, noise signals are inevitably generated to interfere with the original data. Noise signals can cause certain interference in model feature extraction and recognition, and usually the less noise signals generated, the better the model. This study compares the noise signals generated by HR-SDT and DBSCAN models using Student_1 as input, as displayed in Figure 4.7 (a) and 4.7(b) represent the noise

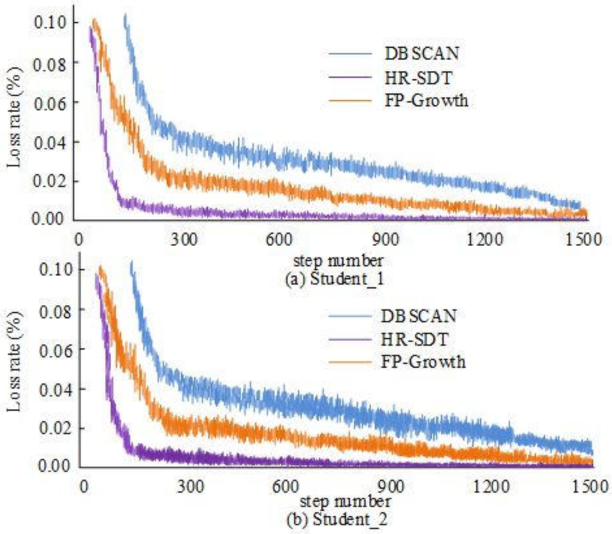

Fig. 4.6: Comparison of loss rates of various models

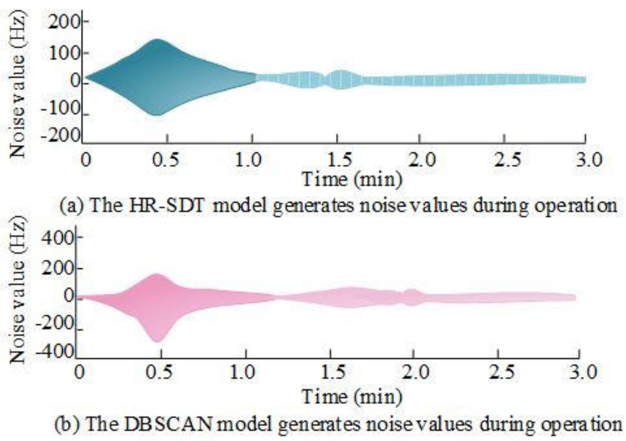

Fig. 4.7: Comparison of noise interference generated by different model operations

signals generated during the operation of the HR-SDT and DBSCAN models. The peak noise signal of the HR-SDT model is 167.45 Hz , while the DBSCAN model reaches 194.55 Hz . Figure 14 shows that the numerical values and noise fluctuations generated by the noise signal in the HR-SDT model are smaller than those in the DBSCAN model, therefore the HR-SDT model performs better.

F1-score is a commonly used indicator for evaluating the performance of classification models, which comprehensively considers the accuracy and recall of the model. To analyze the F1-score of the proposed model, Student_1 and Student_2 are used as inputs in this experiment to record the relationship between the F1-score of HR-SDT, FP-Growth, and DBSCAN models and the sample sequence. Figure 4.8 shows the results. Figures Figure 4.8 (a) and (b) show the F1-score of HR-SDT, FP-Growth, and DBSCAN models on Student_1 and Student_2. Comparing Figure 4.8(a) and Figure 4.8 (b), it is found that the F1-score averages of the HR-SDT model on the Student_1 and Student_2 datasets reach 84.6% and 80.3%, respectively, which are higher than

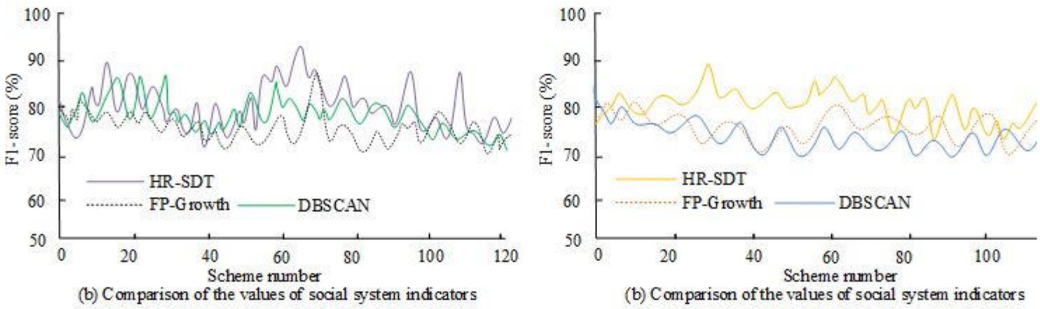

Fig. 4.8: Model F1-score comparison diagram

]

Table 4.1: Performance comparison results of different models

Dataset	Model	APA%	F1%	Convergence time /s	Output time /s
	HR-SDT	91.3	86.7	0.51	3.92
Student_1	FP-Growth	89.0	84.3	2.14	6.18
	DBSCAN	87.5	82.9	2.85	5.45
	HR-SDT	89.2	81.9	0.56	4.01
Student_2	FP-Growth	86.7	79.5	2.58	6.35
	DBSCAN	85.3	77.8	2.77	5.48
	HR-SDT	90.1	84.2	0.53	4.02
Student_3	FP-Growth	87.8	81.6	2.21	6.25
	DBSCAN	86.2	79.3	2.90	5.60
	HR-SDT	88.7	80.5	0.49	3.85
Student_4	FP-Growth	85.9	78.4	2.05	6.10
	DBSCAN	84.4	76.2	2.70	5.35

the other two models.

The above data validate that the SBD mining method based on HRNN can accurately predict students' learning behavior and performance, providing valuable insights and decision support for educators. In addition, error analysis and model inter-pretability studies are conducted to further verify the reliability and practicality of the method. To further prove that the proposed model has good generalization, more teaching data are collected on the basis of Student_1 and Student_2, and two new datasets Student_3 and Student_4 are created. Further, the comparison results of the APA, F1-score, convergence time, and output time of the three models on the four data sets are shown in Table 4.1.

From Table 4.1, the HR-SDT model outperforms the FP-Growth model and DBSCAN model on different data sets, showing higher APA and F1-score, and also has significant advantages in convergence time and output time. In the four data sets, the APA of the HR-SDT model is up to 91.3%, the F1-score is up to 86.7%, and the convergence time and output time are the shortest as low as 0.49 s and 3.85 s . The validity and robustness of the HR-SDT model in EDM are further proved by the experimental verification of various education data sets.

5. Discussion. The proposed SBD mining method based on HRNN and SDT has demonstrated excellent predictive performance and generalization ability on multiple data sets. First, compared with traditional data mining methods, the HR-SDT model performs well in processing long time series data and capturing complex behavioral patterns. By comparing the performance of different models, the study finds that the APA and F1-score of the HR-SDT model are superior to other methods on multiple datasets. Compared with the community resource method based on genome and metabolome data mining proposed by Schom et al. [18], the HR-SDT model shows higher flexibility and accuracy in data processing and application in the field of education. In addition, compared with the method from data mining to wisdom mining proposed by Khan and Shaheen [11], the HR-SDT model not only has advantages in data processing efficiency, but also shows stronger practicability and reliability in practical applications. Secondly, to provide deeper application insights and more accurately capture the phase relationship between signals, this paper further discusses how to improve the capture of spatial distribution features by fusing local information nodes. Although this method has shown a certain effect in capturing local information in SBD, it is still insufficient in practical application to effectively divide spatial distribution features. The study introduces GRU and AdaGrad to better capture and represent temporal and spatial features in SBD, thereby improving the prediction accuracy and robustness of the model. In terms of specific applications, such as in analyzing students' classroom engagement and extracurricular activities, the improved model can more accurately identify and predict student behavior patterns, providing more valuable decision support for educators. Finally, the function of task-related features and their application in constructing graph neural networks (GNN) and enhancing model generalization performance are not fully discussed in this paper. Future research could further explore how task-related features can be incorporated into HR-SDT models to improve their processing of complex educational data. For example, research can be conducted on how to combine GNN technology to better understand and utilize the graph structure features in SBD, thereby further improving the prediction accuracy and generalization performance of the model.

In summary, the results of this study have important implications for personalized learning. By accurately predicting student learning behaviors and performance, educators can personalize instruction by providing targeted guidance and support tailored to each student's specific circumstances. This method can not only help students better grasp the learning content, but also improve their learning interest and initiative, and ultimately improve the quality of education and learning effect.

6. Conclusion. In education today, student learning behavior is considered an important source of data that can reveal their learning status, habits, and potential. However, existing research mostly focuses on specific learning behaviors or adopts traditional data analysis methods, lacking in-depth understanding and mining of complex behavior sequences. Therefore, this study proposed a data mining method based on SBS and HRNN, which achieved an accurate prediction of student learning behavior and performance through in-depth analysis of SBD and model training. After verification, the proposed method achieved an APA of 91.3% and 89.2% for the HR-SDT model on the Student_1 and Student_2 datasets, respectively. The F1-score average values of the HR-SDT model reached 86.7% and 81.9%, respectively. The average output time of the HR-SDT model on Student_1 was 3.92 seconds, 2.26 seconds less than the FP-Growth model. The average convergence time of the HR-SDT model on Student_2 was 0.51 s, which is 1.63 s less than the FP-Growth model. Additionally, the peak noise signal of the HR-SDT was 167.45 Hz , which is much lower than the control model. The experiment proved that the designed method is progressiveness for SBS analysis. At the same time, this method provides valuable insights and decision support for educators, helping to improve the quality and efficiency of education.

This study proposes and validates an SBD mining method based on HRNN, which performs well on multiple educational datasets. However, there are still some aspects that need further study and improvement. Future research can be carried out from the following aspects. First, future research should focus on an in-depth analysis of the different components of the HR-SDT model, assessing their impact on overall performance, and applicability in different scenarios. Second, although the current HRNN model is excellent in many aspects, it still has some limitations, such as the processing of long-time series dependence and computational complexity. Future research could explore other advanced machine learning techniques to achieve better performance in different EDM tasks. Third, future work should also focus on the performance of the model in the long-term application, including its actual effect on the improvement of education quality, long-term impact on students' learning behavior, and feedback and improvement suggestions in actual teaching management.

Funding. The research is supported by In 2023, the special research project of wisdom teaching in undergraduate colleges in Henan Province: Application Research of Wisdom "Teaching" and "Learning" Based on 5G+ Holographic Technology (No.:68); Key scientific research project of colleges and universities in Henan Province: Research on Data-driven Teaching Quality Monitoring (Fund No.: 22A 520015).

REFERENCES

[1] K. Abrar, A. Mian, and S. Zaman, *How social gratification affects social network gaming habitual behavior. sequential mediation of flow experience and consumer satisfaction*, GMJACS, 12 (2022), pp. 112–128.

[2] N. Aziz and S. Aftab, *Data mining framework for nutrition ranking: Methodology: Spss modeller*, IJTIM, 1 (2021), pp. 85–95.

[3] K. Bhosle and V. Musande, *Evaluation of deep learning cnn model for recognition of devanagari digit*, Artif. Intell. Appl., 1 (2023), pp. 114–118.

[4] H. K. Bhuyan, V. Ravi, and M. S. Yadav, *Multi-objective optimization-based privacy in data mining*, CLUSTER COMPUT, 25 (2022), pp. 4275–4287.

[5] A. Bremer, M. Farag, W. Borcherds, I. Peran, E. Martin, R. Pappu, and T. Mittag, *Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains*, NAT CHEM, 14 (2022), pp. 196–207.

[6] ———, *Deciphering how naturally occurring sequence features impact the phase behaviours of disordered prion-like domains*, NAT CHEM, 14 (2022), pp. 196–207.

[7] P. Edastama, A. S. Bist, and A. Prambudi, *Implementation of data mining on glasses sales using the apriori algorithm*, INT J TECHNOL MANAGE, 1 (2021), pp. 159–172.

[8] M. Hasanvand, M. Nooshyar, E. Moharamkhani, and A. Selyari, *Machine learning methodology for identifying vehicles using image processing*, AIA, 1 (2023), pp. 170–178.

[9] R. Kaur, D. Gupta, M. Madhukar, A. Singh, M. Abdelhaq, R. Alsaqour, and N. Goyal, *E-learning environment based intelligent profiling system for enhancing user adaptation*, ELECTRONICS, 11 (2022), p. 3354.

[10] S. Khan and M. Shaheen, *From data mining to wisdom mining*, J INF SCI ENG, 49 (2023), pp. 952–975.

[11] ———, *From data mining to wisdom mining*, J INF SCI ENG, 49 (2023), pp. 952–975.

[12] C. Liu, Y. Zhang, J. Sun, Z. Liu, and K. Wang, *Stacked bidirectional lstm rnn to evaluate the remaining useful life of supercapacitor*, INT J ENERG RES, 46 (2022), pp. 3034–3043.

[13] N. Luo, H. Yu, Z. You, Y. Li, T. Zhou, Y. Jiao, and S. Qiao, *Fuzzy logic and neural network-based risk assessment model for import and export enterprises: A review*, JDSIS, 1 (2023), pp. 2–11.

[14] A. Nandy, C. Duan, and H. J. Kulik, *Using machine learning and data mining to leverage community knowledge for the engineering of stable metal–organic frameworks*, J. AM. CHEM. SOC, 143 (2021), pp. 17535–17547.

[15] K. Nazah, A. W. Ningsih, R. Irwansyah, D. R. Pakpahan, and S. D. Nabella, *The role of ukt scholarships in moderating student financial attitudes and financial literacy on finance management behavior*, JM, 6 (2022), pp. 2205–2212.

[16] L. Quinn-Evans, D. Keatley, M. Arntfield, and L. Sheridan, *A behavior sequence analysis of victims' accounts of stalking behaviors*, J INTERPERS VIOLENCE, 36 (2021), pp. 6979–6997.

[17] M. A. Schorn, S. Verhoeven, L. Ridder, F. Huber, D. D. Acharya, A. A. Aksenov, and J. J. van der Hooft, *A community resource for paired genomic and metabolomic data mining*, NAT CHEM BIOL, 17 (2021), pp. 363–368.

[18] ———, *A community resource for paired genomic and metabolomic data mining*, NAT CHEM BIOL, 17 (2021), pp. 363–368.

[19] N. Sharma, S. Gupta, H. G. Mohamed, D. Anand, J. L. V. Mazón, D. Gupta, and N. Goyal, *Siamese convolutional neural network-based twin structure model for independent offline signature verification*, SUSTAINABILITY, 14 (2022), p. 11484.

[20] T. P. Singh, S. Gupta, M. Garg, D. Gupta, A. Alharbi, H. Alyami, and N. Goyal, *Visualization of customized convolutional neural network for natural language recognition*, SENSORS, 22 (2022), p. 2881.

[21] C. Sirichanya and K. Kraisak, *Semantic data mining in the information age: A systematic review*, INT J INTELL SYST, 36 (2021), pp. 3880–3916.

[22] S. R. Sudharsan and E. N. Ganesh, *A swish rnn based customer churn prediction for the telecom industry with a novel feature selection strategy*, CONNECT SCI, 34 (2022), pp. 1855–1876.

[23] P. Sunhare, R. R. Chowdhary, and M. K. Chattopadhyay, *Internet of things and data mining: An application oriented survey*, J KING SAUD UNIV-COM, 34 (2022), pp. 3569–3590.

[24] R. Thornberg, C. Forsberg, E. H. Chiriac, and Y. Bjereld, *Teacher–student relationship quality and student engagement: A sequential explanatory mixed-methods study*, RES PAP EDUC, 37 (2022), pp. 840–859.

[25] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez, and A. Troncoso, *Deep learning for time series forecasting: a survey*, BIG DATA-US, 9 (2021), pp. 3–21.

[26] Y. Wang, Y. Wu, L. Long, L. Yang, D. Fu, C. Hu, and Y. Wang, *Inflammation-responsive drug-loaded hydrogels with sequential hemostasis, antibacterial, and anti-inflammatory behavior for chronically infected diabetic wound treatment*, ACS Appl. Mater. Interfaces, 13 (2021), pp. 33584–33599.

[27] M. Xia, H. Shao, X. Ma, and C. W. de Silva, *A stacked gru-rnn-based approach for predicting renewable energy and electricity load for smart grid operation*, IEEE T IND INFORM, 17 (2021), pp. 7050–7059.

[28] Y. Xu, X. Yan, Y. Wu, Y. Hu, W. Liang, and J. Zhang, *Hierarchical bidirectional rnn for safety-enhanced b5g heterogeneous networks*, IEEE T NETW SCI ENG and Engineering, 8 (2021), pp. 2946–2957.

[29] S. P. Yadav, S. Zaidi, A. Mishra, and V. Yadav, *Survey on machine learning in speech emotion recognition and vision systems using a recurrent neural network (rnn)*, ARCH COMPUT METHOD E, 29 (2022), pp. 1753–1770.

[30] M. Zavarin, E. Chang, H. Wainwright, N. Parham, R. Kaukuntla, J. Zouabe, and V. Brendler, *Community data mining approach for surface complexation database development*, Environ. Sci. Technol., 56 (2022), pp. 2827–2838.

[31] S. Zhang, Y. Wen, and Q. Liu, *Exploring student teachers' social knowledge construction behaviors and collective agency in an online collaborative learning environment*, INTERACT LEARN ENVIR, 30 (2022), pp. 539–551.

[32] X. Zhang, L. Liu, X. Chen, Y. Gao, S. Xie, and J. Mi, *Glc_fcs30: Global land-cover product with fine classification*

system at 30 m using time-series landsat imagery, EARTH SYST SCI DATA, 13 (2021), pp. 2753–2776.
[33] J. Zhu, Q. Jiang, Y. Shen, C. Qian, F. Xu, and Q. Zhu, *Application of recurrent neural network to mechanical fault diagnosis: A review*, J MECH SCI TECHNOL, 36 (2022), pp. 527–542.

Edited by: Zhengyi Chai
Special issue on: Data-Driven Optimization Algorithms for Sustainable and Smart City
Received: Feb 27, 2024
Accepted: Aug 6, 2024