Ai
CHINESE ROOTS

Original Paper

A BiGRU joint optimized attention network for recognition of drilling conditions

Ying Qiao [a, b], Hong-Min Xu [b], Wen-Jun Zhou [b], Bo Peng [b], Bin Hu [c], Xiao Guo [a]

[a] National Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu, 610500, Sichuan, China
[b] School of Computer Science, Southwest Petroleum University, Chengdu, 610500, Sichuan, China
[c] Information Management Center of Sinopec Southwest Oil and Gas Branch, Chengdu, 610500, Sichuan, China

ARTICLE INFO

Article history:
Received 9 July 2022
Received in revised form
20 February 2023
Accepted 31 May 2023
Available online 1 June 2023

Edited by Jia-Jia Fei

Keywords:
Drilling condition classification
BiGRU
Machine learning
Attention mechanism

ABSTRACT

The identification and recording of drilling conditions are crucial for ensuring drilling safety and efficiency. However, the traditional approach of relying on the subjective determination of drilling masters based on experience formulas is slow and not suitable for rapid drilling. In this paper, we propose a drilling condition classification method based on a neural network model. The model uses an improved Bidirectional Gated Recurrent Unit (BiGRU) combined with an attention mechanism to accurately classify seven common drilling conditions simultaneously, achieving an average accuracy of 91.63%. The model also demonstrates excellent generalization ability, real-time performance, and accuracy, making it suitable for actual production. Additionally, the model has excellent expandability, which enhances its potential for further application.

1. Introduction

Petroleum is a critical and strategic resource that is the most significant energy source for ensuring long-term global stability. Drilling is a vital link in oilfield development, and its efficiency and safety directly impact the economic benefits of oil firms. Over the years, oil companies have given high priority to ensuring safe and efficient drilling conditions. The drilling process involves human-computer interaction, where the driller provides instructions, and the required personnel operate the drill bit to carry out these instructions. The driller then examines the data collected by the logging equipment to understand the changing trends in downhole physical and chemical properties and records the actual execution outcomes. The increasing use of various types of sensors has led to a rise in the frequency of data collection by monitoring equipment, resulting in a large volume of data being created in a short period of time during drilling. However, the implicit relationship between diverse feature data is challenging to understand as a whole, and the intersection and overlap of working situations further complicate their identification. The drilling condition record is an essential tool to help the driller clarify the work plan and supervise the work progress. However, the traditional manual management method that is currently used to manage drilling condition records has limitations that include poor effectiveness, low efficiency, and strong subjectivity in classification. These limitations reduce management efficiency, decision-making precision, and responsiveness during oilfield development (Chen, 2021). It is crucial to track drilling operations to ensure safety, and therefore, drilling monitoring skills must keep up with drilling technology.

The development of drilling condition recognition has undergone a process from mathematical methods to machine learning methods. To categorize drilling conditions using the difference data, Arnaout et al. (2012) built a mathematical model based on polynomial approximation in 2012. In 2015, Caldwell and Hinton (2015) suggested that mining and using drilling data can help businesses better direct the execution of intervention plans and risk management capabilities. In the same year Khudiri et al. (2015) proposed a mathematical algorithm that uses real-time surface parameter data to determine the current operating status. By introducing surface data, they improved the algorithm's

Corresponding author.
Corresponding author. National Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu, 610500, Sichuan, China.
Corresponding author.
E-mail addresses: teachqiao@163.com (Y. Qiao), zhouwenjun@swpu.edu.cn (W.-J. Zhou), bopeng@swpu.edu.cn (B. Peng).

generalizability and integrated it into the drilling control software to realize the classification of drilling conditions for use in production. In an effort to separate the drilling conditions through pressure variations, Liu and Tao (2015) employed time series and gray system theory to anticipate the drilling pressure during drilling. The models obtained by traditional mathematical methods have limitations. When data characteristics change, mathematical formulas need to be reconstructed, which brings high development and maintenance costs.

Compared with mathematical methods, machine learning can effectively extract complex relationships in data and produce more accurate or better simulation combination results. Therefore, the topic of machine learning has attracted more scholars' interest. In 2014, Todorov and Thonhauser (2014) proposed using a simple, upgraded, fully connected network to anticipate drilling fluid pressure, which can assist drillers in effectively identifying pressure anomalies. In 2017, Zhao et al. (2017) came up with an improved SAX-based drilling data processing method. They used unsupervised machine learning methods to group time series of drilling data, which can help drillers improve how they do their work. In 2019, Ting et al. (2019) suggested a support vector machine working conditions classification approach. In 2022, Ge et al. (2022) introduced a model training approach that trained drilling site photographs utilizing U-Net, attention mechanism, and GAN networks. This algorithm partially resolved the issue of drilling condition detection.

To summarize the state of the art in drilling classification research, the majority of studies use machine learning methods relatively superficially, with small data volumes, simple network models, uncertain model generalization abilities, and no targeted design for data with time series characteristics, among other issues. However, some studies have shown promising results in utilizing advanced machine learning techniques to improve drilling condition recognition, such as (Ben et al., 2019) and (Liu and Zhang, 2021).

Considering the research scope of this paper, we propose a semi-empirical decision tree annotator based on actual data and design an improved two-way door control unit network combined with an attention mechanism. The attention mechanism was first introduced by Bengio's team when improving the seq2seq architecture (Chorowski et al., 2014). Based on this, the transformer framework developed by the Google team has received wide attention in various fields (Vaswani et al., 2017). Another key factor in the combined network is the GRU unit, also proposed by the Bengio team (Cho et al., 2014). Compared to other network units, GRU provides better computational efficiency without compromising the accuracy of the model.

2. Work overview

Fig. 1 illustrates the overall structure of the project. Initially, we studied existing papers and drilling manuals to summarize the basic discrimination formulas for seven drilling conditions: trip in, trip out, drilling, sliding drilling, circulate, stab pip, and ream. These formulas were first revised with the guidance of drilling professionals and used to construct a decision tree labeling model. To further expand the decision criteria of the discrimination formulas, we combined over 30 million actual drilling data and iteratively updated the decision tree model to ensure the accuracy of its labels. After obtaining the labeled data, we filtered and enhanced the data to construct the training set, validation set, and test set used in training. Subsequently, we built a neural network, utilizing a BiGRU network combined with an attention module to extract information from the data, and a fully connected neural network for classification. Finally, the resulting model will be tested for its generalization ability using the test set.

This work's contributions can be summarized as follows.

(1) To identify the key factors in drilling conditions, we used operational specifications and empirical formulas. We added to the empirical formulas based on real data characteristics to build an expert system that maps characterization parameters to drilling conditions, making the model more interpretable.
(2) For the seven drilling conditions in our study, we used BiGRU as the basic information extraction module, considering the characteristics of the time series data samples. We improved the traditional GRU unit to enhance the utilization of historical data and the expression abilities of the functions. After BiGRU, we added an improved attention mechanism network to improve the model's ability to build long-term dependencies, further ensuring the model's performance.
(3) We performed a sensitivity analysis on the combination of characterization parameters and hyperparameter settings to ensure the validity and desirability of the model.
(4) Using data that was not in the training set, we demonstrated that the BiGRU model with the attention mechanism is better at generalization and is more stable.

The rest of this article is structured as follows. Section 3 delves deeper into the seven drilling conditions to determine the critical characterization parameters that influence the drilling conditions. We clarify the engineering theory and essential data characteristics of the drilling conditions. In Section 4, we briefly describe BiGRU's architecture and how to incorporate attention mechanisms into machine learning models. We introduce the sample dataset expansion method and the model enhancement method in Section 5. Section 6 presents comparative experiments for various characterization parameters and hyperparameter settings. Finally, in Section 7, we discuss conclusions and provide recommendations for future research.

3. Data description and preprocessing

The dataset used in this paper was provided by the Information Management Center of the Southwest Oil and Gas Branch of Sinopec. It consists of feature data collected during drilling from six different wells, with each well containing 2 to 7 million data points, totaling approximately 30 million. Each entry contains 73 characterization parameters that describe physical, chemical, and location data recorded during the drilling process, with data typically collected every 5 s.

Based on field research at the drilling site, we determined that real-time capabilities were necessary for this study, with the model's frequency of identifying drilling conditions required to be less than 10 min per instance. Consequently, we utilized a sample of 60 time-series data points spanning approximately 6 min. This moderate sample size allows for a clear representation of the trend of each characterization parameter, making it easier for the neural network to extract features and establish long-term dependencies. Given that this project requires the use of supervised machine learning methods, the data needs to be labeled. However, manual labeling for several million data points is not feasible. Therefore, we established a decision tree model for automatic labeling.

First, we summarize the typical relationship between drilling conditions and representative parameters under macro conditions from relevant literature (Wei, 2014) and operational manuals. Then, these judgment conditions are organized into empirical formulas to guide the construction of decision tree models, as shown in Table 1. This version of the decision tree criteria is mainly based on the

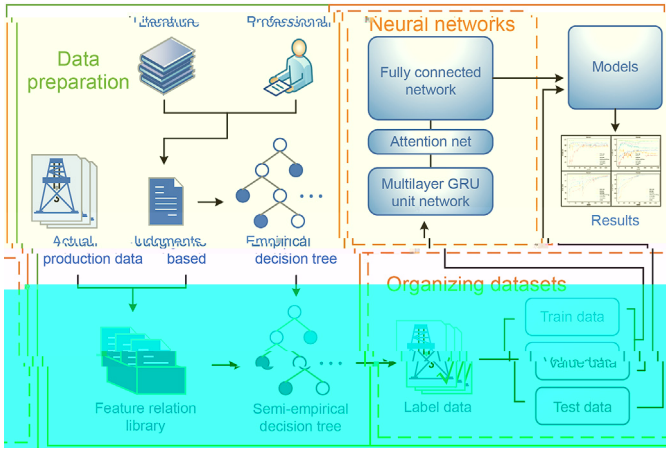

Fig. 1. Workflow diagram.

seven characteristic parameters in the table. Among these parameters, the pump parameters contain changes from three pumps, resulting in a total of nine parameters being involved in the work condition judgment.

The second step involves adapting the decision tree model to accommodate variations in the data features of each well. Since the same condition can be represented differently in different wells, the judgment conditions need to be modified accordingly to better fit the reality. We worked closely with drilling experts and conducted multiple rounds of modifications to the decision tree model. The result is an expert system that links drilling conditions to representative parameters, as shown in Table 2. The final version of the decision tree model includes an additional parameter that indicates the location of the working condition occurrence. In total, 10 parameters are involved in the decision tree judgment.

The final version of the decision tree was compared with the initial version, and both were used to annotate each drilling data record. The annotation results were manually reviewed, and the comparison results are shown in Table 3. Based on the comparison results, the average accuracy of the empirical decision tree for each drilling data table was 82.50%. The average accuracy of the semi-empirical decision tree was 97.09%, which represents a 14.59% improvement in accuracy, demonstrating the effectiveness of this approach.

The aforementioned work has demonstrated the decision tree labeling model's ability to identify and label seven drilling conditions quickly and accurately, with a comprehensive recognition accuracy rate exceeding 95%. However, during the sorting of drilling data, it was observed that many of the unutilized characterization parameters potentially contained valuable information, and summarizing the rules for these characteristics through visual observation was difficult. Additionally, considering the complex and constantly changing drilling environment, the expansion of new working conditions, and the model's applicability in different regions, relying solely on the fixed criteria matched by the current ten key parameters to identify working conditions would require rebuilding the criteria set when test data characteristics do not match existing criteria. This approach would incur high development and maintenance costs, limited flexibility and expandability, and wasted data resources. To make better use of more feature parameters, mine more hidden information in the data, and improve the model's generalization ability, this paper selects a neural network as the model responsible for data abstraction, feature extraction, and classification.

4. Drilling multi-conditions identification model

In this section, we describe the construction of a two-way gate control unit network with an attention mechanism. Specifically, we explain the structure of the GRU unit and the implementation principle of the attention mechanism.

4.1. BiGRU neural network

Previous studies on drilling condition classification mainly relied on traditional machine learning methods such as SVM and

Table 1
The relationship between drilling conditions and feature changes: traditional criteria.

Priority order	Working condition	Analyzing conditions						
		Standard well depth	Drill bit position	Turntable speed	Top drive rpm	Drilling pressure	Standpipe pressure	Pump
1	Stab pipe	→	↘	–	–	–	–	–
2	Ream	→	↘↗	≠ 0	≠ 0	–	–	–
3	Drilling	↗	↗	≠ 0	≠ 0	≠ 0	–	–
4	Sliding drilling	↗	↗	≈0	≠ 0	≠ 0	–	–
5	Trip out	→	↘	–	–	–	–	–
6	Trip in	→	↗	–	–	–	–	–
7	Circulate	–	–	–	–	–	≠ 0	≠ 0

Table 2
The relationship between drilling conditions and feature changes: expert system.

Priority order	Working condition	Analyzing conditions							
		Standard well depth	Drill bit position	Occurrence location	Turntable speed	Top drive rpm	Drilling pressure	Standpipe pressure	Pump
1	Stab pipe	→↗	↘↗	5–40 m	≈0	≈0	≈0	≈0	≥2
2	Ream (case1)	→	↘↗	≤10 m	» 0	» 0	≈0	≈0	≥2
3	Ream (case2)	→	↘↗	≥120 m	» 0	» 0	≈0	≈0	≥0
4	Drilling	↗	↗	–	≥10	≥10	» 0	≥15	≥2
5	Sliding drilling	↗	↗	–	≈0	–	» 0	≥15	≥2
6	Trip out	→	↘↘	–	≈0	≈0	≈0	≈0	= 0
7	Trip in	→	↗↗	–	≈0	≈0	≈0	≈0	= 0
8	Circulate	–	–	–	–	–	≠ 0	≠ 0	≥2

Table 3
Comparison of traditional and improved annotators for accuracy in labeling.

Well number	Empirical formula decision tree	Semi-empirical formula decision tree
A1	77.99%	98.23%
A2	85.34%	95.18%
A3	81.52%	96.65%
A4	86.74%	95.92%
A5	79.68%	97.87%
A6	83.74%	98.70%

MLP. However, these methods required converting time series data into the input structure required by the model, leading to difficulties in recognizing the relationships between parameters and capturing time series features. Recursive neural networks, on the other hand, are well-suited for processing time series data, as their specified data input format corresponds to time series data, enabling them to establish long-term dependencies. Among the most widely used models for processing time series data are the improved RNN, LSTM, GRU and their derivatives. GRU and LSTM can record the correlation of long-sequence data, effectively suppressing gradient disappearance and better extracting long-time data features. Unlike LSTM, the GRU unit uses an update gate instead of input and forget gates, as shown in Fig. 2. This simplifies the calculation of the hidden state in the network and saves more time when the training data is large.

Before training the GRU network, a two-dimensional matrix is used to construct samples as input data, containing important features that determine the working condition at a given time step, as well as hidden features that may play a role in determining the condition. At each time step t, the individual GRU cell in the network takes the previous hidden state h_{t-1} and the current working condition characteristic data x_t as inputs. The GRU cell calculates the output y_t at the current time node, and the hidden state h_t is passed to the next step. Fig. 3 illustrates the data flow and changes between cells from time step $t-1$ to $t+1$, using a multi-layer GRU cell network as an example.

Inside the GRU cell, the reset gate and update gate gating states are calculated first. Both gates take the working condition characteristics x_t and the hidden state h_{t-1} from the preceding time step as input. The Sigmoid function transforms the gating data into a value between 0 and 1. The equations for calculating the reset gate r_t and update gate z_t are given as follows, where x_t represents the work characteristic input at time t and h_{t-1} represents the hidden state at the previous time step:

$$r_t = \sigma(x_t W_r + h_{t-1} U_r + b_r) \tag{1}$$

$$z_t = \sigma(x_t W_z + h_{t-1} U_z + b_z) \tag{2}$$

Here, r_t represents the gating of the reset gate at time step t, while z_t represents the gating of the update gate at time step t. W_r, W_z, U_r, and U_z are weight parameters, while b_r and b_z are bias parameters. The hidden state at time $t-1$ is represented by h_{t-1}, and σ denotes the Sigmoid function.

The GRU unit then proceeds to compute the reset gate's information. $\tilde{h}$ represents a potential hidden state that will influence the future hidden state. When the gating information r_t approaches 0, the hidden state element is reset, indicating that the previously stored information is discarded. Conversely, if the gating information r_t approaches 1, it means that the previous information is still relevant, and the hidden information from the previous time step is retained. After resetting the gate, the input data from the working condition characteristics is used to determine whether the hidden state should be updated or not. The tanh function is used to

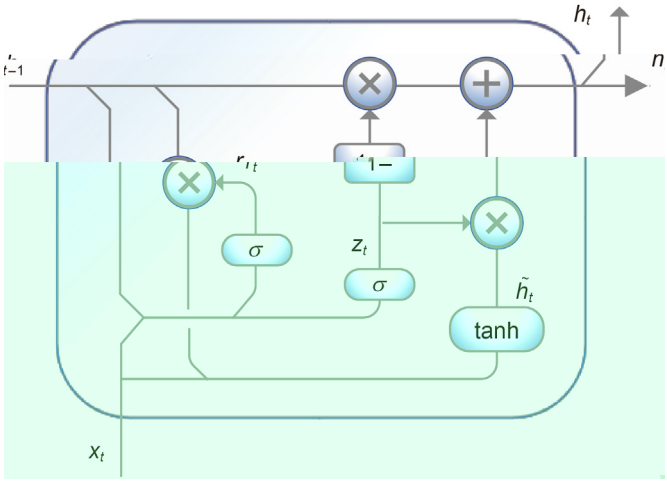

Fig. 2. GRU cell structure.

compute the hidden candidate states. For time step t, the candidate hidden states are calculated as follows:

$$\tilde{h} = \tanh(x_t W_h + r_t \odot h_{t-1} U_h + b_h) \tag{3}$$

The weight parameters are represented by W_h and U_h. b_h represents the bias parameter. Multiply by element is abbreviated as $\odot$. The information about the working condition attributes x_t input at the tth moment and the concealed state information at the $t-1$ moment after processing is contained in the $\tilde{h}$ in the formula.

Finally, the GRU unit computes the current hidden node output y_t and the hidden state h_t at time t. This operation aims to discard some of the content in the hidden state information passed from the previous time step and add the hidden state information from the current node's output as a supplement. The hidden content is compiled and transmitted to the next step. The update gate z_t is used to determine the weights of discarded and additional data. The range of update gate z_t is 0–1. The closer z_t gets to 0, the more irrelevant information in the hidden message is forgotten. More memories are passed on when z_t approaches 1. The following formula is used to compute the hidden state at time step t:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h} \tag{4}$$

In this study, we have made targeted improvements to the traditional calculation method of the GRU unit in two aspects. First, since the sample data used in this study contains 60 pieces, we aim to ensure that the GRU unit makes the most of all the available data. However, due to the characteristic of the sigmoid function, the parameters of the update gate and the reset gate may fall into a range very close to 0, resulting in the loss of historical data. To address this issue, we restrict the range of the sigmoid function and add a minimum threshold to ensure the preservation of historical information. Second, in the update gate, we use the softsign

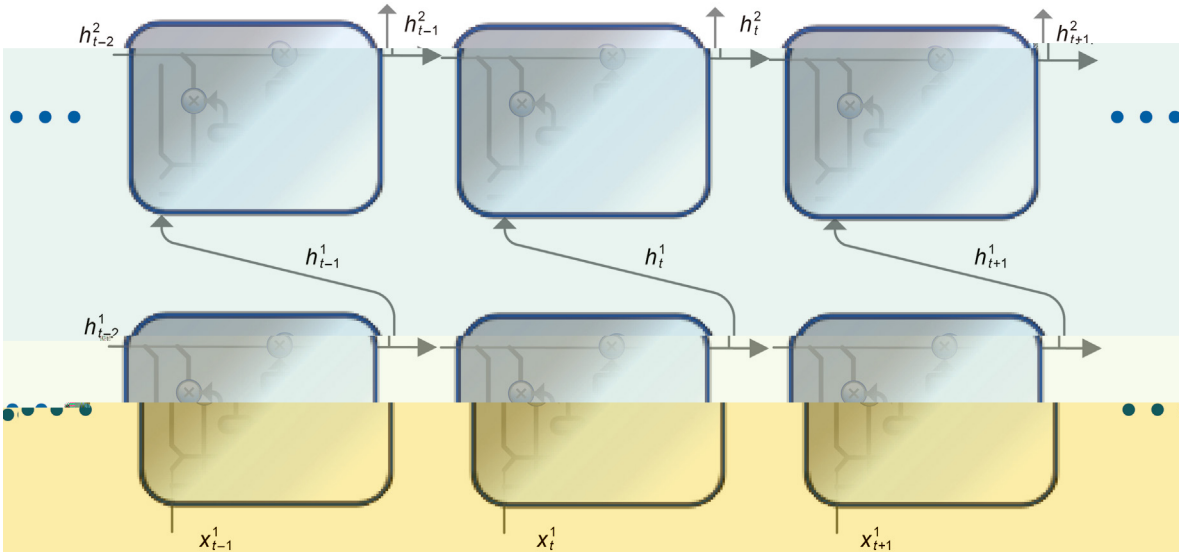

Fig. 3. Multi-layer GRU data flow.

function instead of the traditional tanh function in the GRU unit. The softsign function has a wider non-saturation interval and a smoother gradient change trend compared to the tanh function. Although it will require more computational power, the effect will be better. Fig. 4 shows the improved GRU unit, and the updated calculation formulas for the update and reset gates are presented below.

$$r_t = 0.9\sigma(x_t W_r + h_{t-1} U_r + b_r) + 0.1 \tag{5}$$

$$z_t = 0.9\sigma(x_t W_z + h_{t-1} U_z + b_z) + 0.1 \tag{6}$$

$$\tilde{h} = \text{softsign}(x_t W_h + r_t \odot h_{t-1} U_h + b_h) \tag{7}$$

with the same parameter settings, this paper constructed both the traditional GRU + MLP classification network and the improved GRU + MLP classification network, used the same training data for five repetitions, and selected the best-performing model for comparison. Fig. 5 shows the curve of the average recall rate measured by the two models with the number of training iterations for the seven drilling conditions studied in this paper. The comprehensive accuracy of the optimal model in the test set was 69.78% for the traditional GRU and 73.16% for the improved GRU. This study demonstrated that the improved GRU achieved a 3.38% performance improvement compared to the traditional GRU.

The BiGRU is composed of two unidirectional GRUs that operate in opposite directions, forming an additional hidden layer (Hu and Xue, 2019). The main difference between BiGRU and GRU is the extra layer of hidden states. The neural network architecture consists of an input layer, a forward hidden layer, a backward hidden layer, and an output layer, as shown in Fig. 6. The input data x_t is fed simultaneously to the forward and backward hidden layers. The forward and backward cells receive the input and the previous time step's forward hidden state $\overrightarrow{h}_{t-1}$ and backward hidden state $\overleftarrow{h}_{t-1}$, respectively. The current forward and backward hidden states $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ are calculated and then combined to produce the present hidden state h_t. The entire process can be represented as follows:

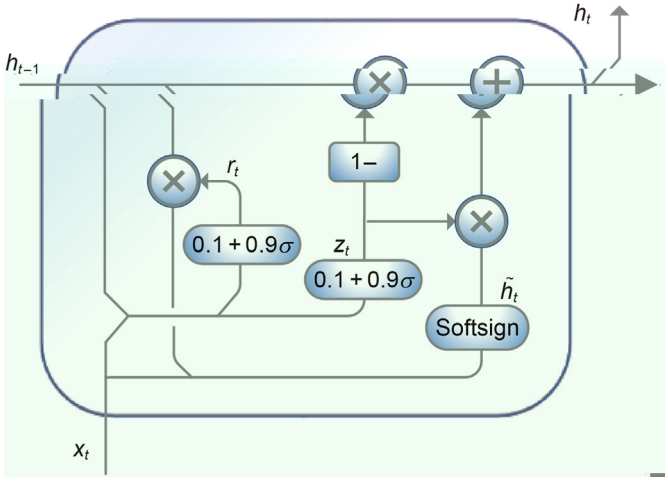

Fig. 4. Improved GRU cell structure.

$$\overrightarrow{h}_t = GRU\left(x_t, \overrightarrow{h}_{t-1}\right) \tag{8}$$

$$\overleftarrow{h}_t = GRU\left(x_t, \overleftarrow{h}_{t-1}\right) \tag{9}$$

$$h_t = W_{\overrightarrow{h}_t} \overrightarrow{h}_t + W_{\overleftarrow{h}_t} \overleftarrow{h}_t + b_t \tag{10}$$

After passing through BiGRU, the timing features of each working condition will be discarded based on the importance of the information, and the features will be extracted. This time-series characteristic is influenced by both historical and future data. The extracted features of the working conditions are then passed through an attention mechanism.

4.2. Attention mechanism

In time series learning tasks, attention mechanisms have been shown to significantly enhance performance, as demonstrated by

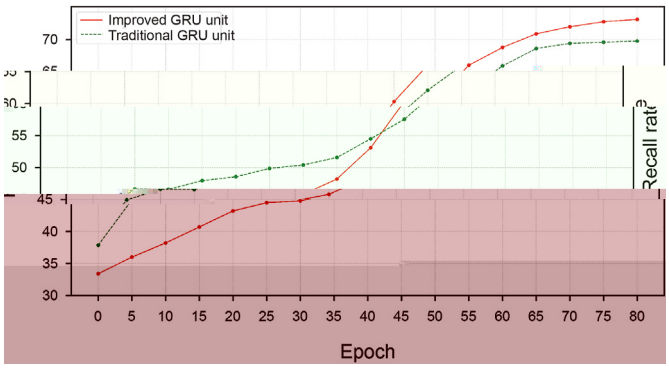

Fig. 5. Comparison of recall rate between traditional GRU unit and improved GRU unit.

Cinar et al. (2017). The basic idea is similar to the principle of human attention: each element in a sequence carries different amounts of information and requires different levels of attention. In working condition recognition, numerous physical features describe specific scenarios, and extracting key parameters is critical for model generalization and establishing long-term dependencies. The attention mechanism addresses this by allowing the model to focus more on specific feature sections that require more attention.

When the hidden layer state h_t is obtained at a given time from the output of BiGRU, it is fed into a single layer perceptron to produce u_t. The dimensionality of the input is not changed by this fully connected network; it simply represents the h_t hidden layer as u_t. In this paper, we improved upon the multilayer attention approach proposed by Pappas and Popescu-Belis (2017) by using a softsign activation function instead of a tanh activation function. This change makes it more difficult for data to fall into the saturated domain of the function, and while it may result in increased computational difficulty, it enables more efficient learning. The updated formula is as follows.

$$u_t = \text{softsign}(W_w h_t + b_w) \tag{11}$$

In the following formulas, W_w represents the weight parameter, and b_w represents the bias parameter. To assess the significance of each element, a matrix is randomly initialized to represent the information meaning of the data segment, and it is multiplied with each feature in the data segment to calculate the similarity. The resulting matrix is then normalized using the softmax operation to obtain the attention weight matrix α_t.

$$\alpha_t = \frac{\exp(u_t^T u_w)}{\sum_t \exp(u_t^T u_w)} \tag{12}$$

After obtaining the attention weight matrix, multiplying h_t with α_t and summing them yields the weighted vector s. The formulas are as follows.

$$s = \sum_t \alpha_t h_t \tag{13}$$

After the attention mechanism is applied, the resulting vector s represents the overall information contained in the data segment. This vector s is then passed to a two-layer fully connected neural network, which further extracts and classifies the information present in the data.

5. Model improvement and data organization

5.1. Organization of data sets

The input samples of the neural network were annotated by the decision tree annotator, and 60 data points were kept unchanged during input. In this study, drilling data from 6 wells were used, and the training set and validation set consisted of data from 5 wells, which were split and decomposed in a 4:1 ratio. The remaining well data was used as the test set to evaluate the model's generalization ability. Table 4 displays the sample sizes of the training and test sets after decision tree labeling.

Table 4 reveals that some operating conditions occur less frequently than others and there may be gaps between the same operating conditions that are too large to be expanded using a sliding window approach. Therefore, we need to use a reasonable method to solve the problem of sample imbalance while organizing

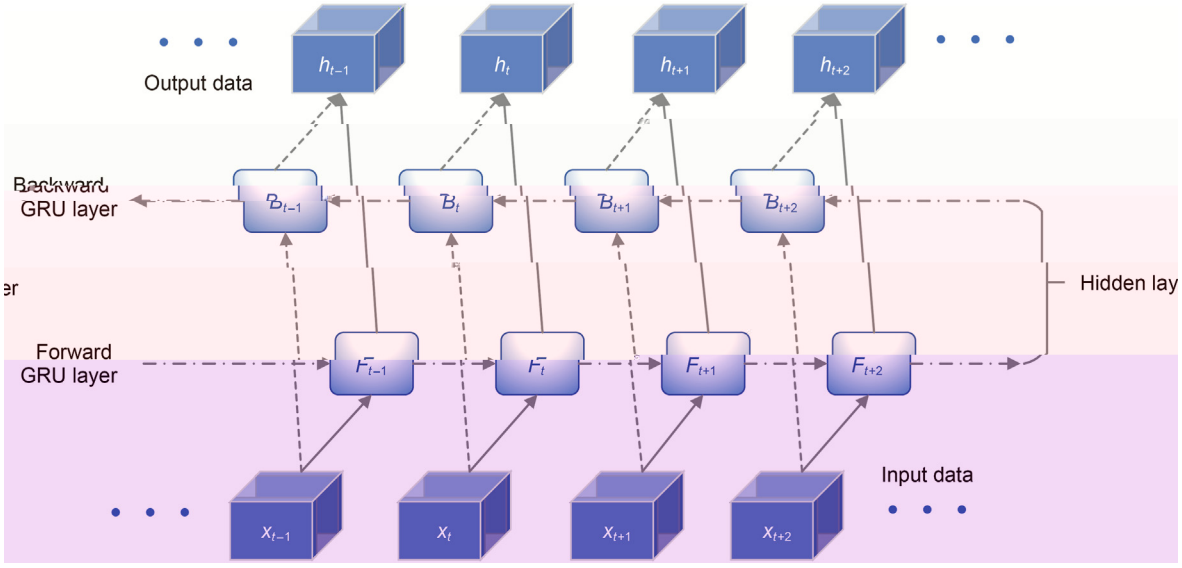

Fig. 6. BiGRU network structure.

the dataset. To address these issues, we use two methods for data expansion. Firstly, we use the sliding window approach to slice and sample continuous operating conditions, with a step size of 3–5 data, resulting in a 20% increase in sample size compared to using the decision tree samples directly. Secondly, for those operating conditions that do not occur continuously, we use adjacent operating conditions for interval sampling and synthesis while ensuring that the interval sampling follows time and parameter change threshold constraints. The parameter and time constraints are listed in Table 5. If the sample time interval is too long, it may have a negative impact on the model training. Fig. 7 depicts the sliding window and interval sampling methods. After data enhancement and sample size balance, Table 6 displays the total number of samples and the number of samples involved in training and testing. It is worth noting that both sliding window and interval sampling methods are employed to enhance the number of samples under the same working condition.

Table 5
Range of parameter variation constraints used in sample augmentation.

Descriptive parameters	Variation limit
Standard well depth	±1
Dill bit position	±1
Turntable speed	±5
Top drive rpm	±5
Drilling pressure	±3
Standpipe pressure	±10
Pump	±5
Time	≤3 min

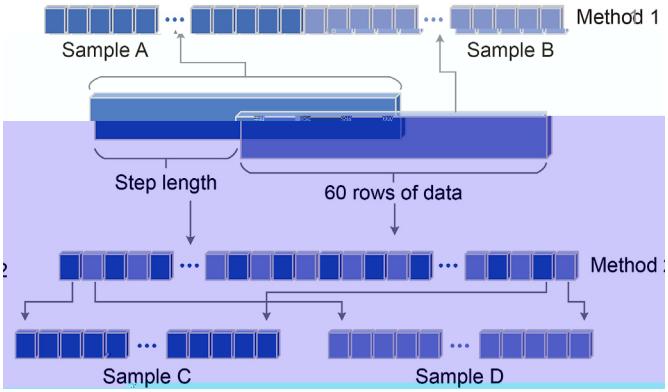

Fig. 7. Sliding window and interval sampling data augmentation.

5.2. Model improvement

The training process includes various methods for improving the network's ability to extract information, which can enhance model performance. Weight initialization (Wang, 2019), normalization (Mittal et al., 2021), batch normalization (Keskomon et al., 2020), and adaptive learning rate adjustment (Li et al., 2010) are all part of this process.

Since each feature involved in the training had different magnitudes and units, the data was normalized to remove the magnitude effect, as sending unprocessed data to the network for training would affect the model's effectiveness. In this experiment, min-max normalization (Tang, 2017) was used, which maps the data of each feature column to the range [0, 1] using a linear transformation method. The normalization formula for a particular characteristic column is shown as follows:

$$x_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{14}$$

here, $x_{\min}$ is the minimum value of the current feature column, and $x_{\max}$ is the maximum value of the current feature column.

To speed up model convergence, the model weights should be initialized. Different weight initialization methods such as Xavier, He, and orthogonal initialization are commonly used, and each method targets a different audience. We used He initialization on the fully connected layer, and orthogonal initialization is used to initialize the matrix in the BiGRU section. Xavier initialization is primarily used for fully connected networks and is used to initialize the single perceptron layer in the attention network module.

To solve the problem of gradient disappearance and explosion in the multilayer fully connected structure at the end of the network structure, a batch normalization layer is introduced. It normalizes the n inputs first, then scales the translation. It improves the network's flow gradient, increases training speed (Keskomon et al., 2020), and improves the network's generalization ability.

Additionally, an adaptive learning rate adjustment method is used to automatically adjust the learning rate based on the training effect. This method adjusts the learning rate more effectively than fixed-value learning rates, step-based learning rate adjustment methods, or cosine annealing methods.

After the above preparatory work, the preliminary network model structure is constructed as shown in Fig. 8. The model is trained after the parameters have been set, and the network parameters are fine-tuned based on the training results.

6. Experiments and analysis

6.1. Training evaluation metrics

Our experiments will use two evaluation metrics, namely cross-entropy loss and recall, in order to effectively evaluate the methods used during the training period. The cross-entropy loss is calculated in two steps during the training process. First, the sigmoid function is used to scale the output result to be between 0 and 1, as shown in Eq. (15). Then, the negative log-likelihood loss function (NLLLoss) is used to obtain the results, as shown in Eq. (16). After each epoch is completed, the recall is calculated using Eq. (17). The recall rate is an essential indicator for judging the effectiveness of the model.

$$\mathrm{sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{15}$$

Table 4
Distribution of labeled samples for different drilling conditions in the training and testing datasets.

Well number	Drilling conditions						
	Trip out	Trip in	Drilling	Sliding drilling	Stab pipe	Circulate	Ream
A1~A5 (Training)	14887	21743	62877	10366	18431	6394	35331
A6 (Testing)	1830	3425	15231	1328	2027	653	7496

Table 6
Sample distribution across drilling conditions in the training and testing datasets after sample augmentation.

Data set	Drilling conditions						
	Trip out	Trip in	Drilling	Sliding drilling	Stab pipe	Circulate	Ream
Total	17149	24037	77890	12716	21633	6632	43495
Train	14922	14922	9948	9948	9948	6632	14922
Test	2156	4205	17973	1674	2408	785	9295

$$H(p,q) = -\sum_{i=1}^{n} p(x_i)\log(q(x_i)) \tag{16}$$

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

TP and FN, respectively, represent true positives and false negatives. A higher recall and a lower cross-entropy loss value indicate a better model effect. Overall, these evaluation metrics are crucial in assessing the performance of the model. The cross-entropy loss measures the difference between predicted and actual values, while recall measures the ability of the model to correctly identify positive instances. By utilizing both metrics, we can obtain a comprehensive understanding of the model's performance and make informed decisions regarding improvements or adjustments to the model.

6.2. Feature expansion experiment

This experiment collected data on 73 different drilling characterization parameters during the drilling of six wells. This paper calculates the Pearson correlation coefficient, Spearman correlation coefficient, and Kendall correlation coefficient among features to reflect the positive and negative correlation between feature relationships. After obtaining the three coefficient matrices, the average value of the correlation coefficient of the characteristic parameters in the three matrices is calculated. If the absolute value of the average value is greater than 0.8, only one of the two features will be retained. Additionally, features with a correlation coefficient lower than 0.2 with other features are removed as they may interfere with training and affect the training effect. Finally, the remaining 24 characterization parameters are used to construct samples and handed over to the neural network for training. Fig. 9 shows the correlation coefficient matrix thermodynamic diagram.

However, these 24 characterization parameters do not guarantee that each parameter positively affects the neural network, thus requiring feature expansion experiments to confirm the parameters' validity. The experiments start with ten features used in the decision tree and gradually increase the number of characterization parameters used in training. The best combination of parameters is determined by comparison.

In summary, by using correlation coefficients to select and remove features, the dimensionality of the input data is reduced, which simplifies the training process and improves training efficiency. Additionally, by conducting feature expansion experiments, the optimal combination of parameters for the neural network can be determined, which will improve the accuracy of the model's predictions.

In the decision tree construction phase, we provided ten basic features, including standard well depth, drill bit position, turntable speed, top drive speed, drilling pressure, standpipe pressure, pump punch #1, pump punch #2, pump punch #3, and large hook load. To make subsequent feature screening easier, the 24 features under consideration are classified according to their physical descriptions, as shown in Table 7. As the experiment progresses, the model records more implicit connections of the characterization parameters through learning with the neural network.

During training, the models are saved once per epoch, and after training, each model is tested for generalization ability sequentially using drill data not used in training. Within the same graph, the recall of each model for each of the seven operating conditions is

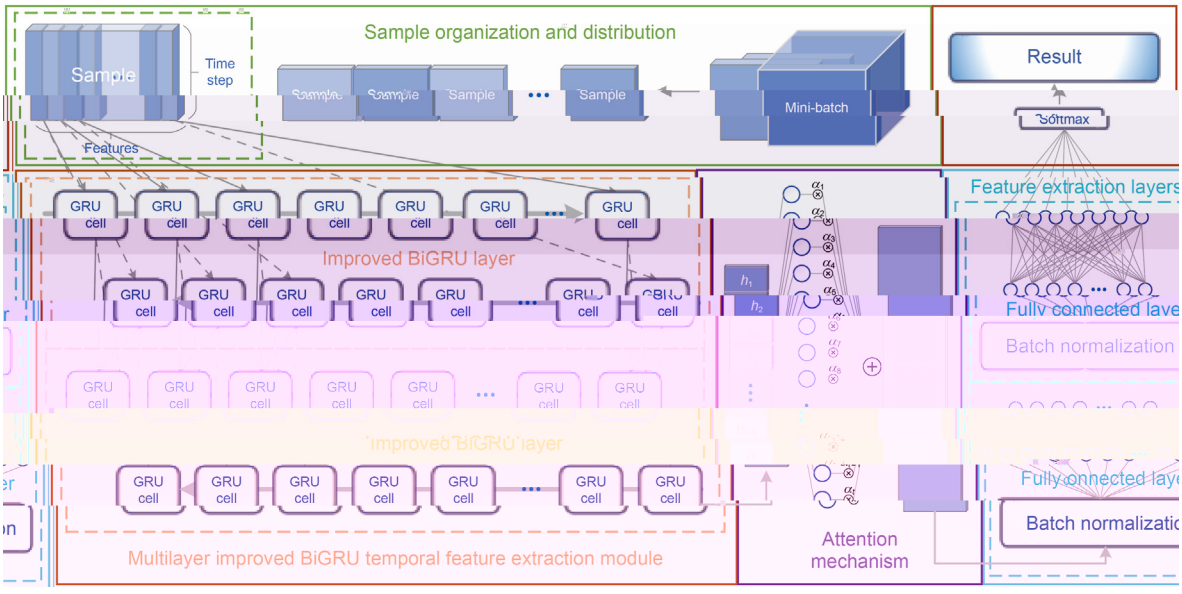

Fig. 8. Flowchart of drill condition recognition using enhanced BiGRU network with attention mechanism and fully connected layers.

plotted. In the following sections, the recall curves are plotted similarly.

Ten basic features are used as input for training the network. The models obtained after several rounds of training are evaluated for recall on the test set, and the effectiveness of each model is shown in Fig. 10(a). The recall performance of the best model on the test set, after its initial training, is shown in Table 8.

The initial experiment trained the network using ten fundamental features, which achieved a recall rate of 90% for drilling, sliding drilling, and scribing eye, but poor recall rates for other working conditions. To improve the recall rate for these conditions, additional characterization parameters were added to the training set. For example, outlet flow and inlet flow were added to distinguish start-up down-drilling and circulation from each other, resulting in improved recall rates as shown in Fig. 10(b). It was observed that starting and down drilling had interaction, requiring features that behave differently in these two conditions to be added. Adding the total pool volume and increment decrement descriptions to the training set further improved recall rates as depicted in Fig. 10(c). The effect of the drilling fluid's carbon dioxide, methane, and gas content on the recall rate was investigated by adding these characterization parameters to the training set and retraining the model, as shown in Fig. 10(d). The addition of four features related to the drilling fluid's physical properties did not positively influence the judgment of drilling conditions and were not added to the training set. Drill position-related features such as top drive torque and drilling time were added and re-trained, resulting in improved recall rates as shown in Fig. 10(f). The best model obtained from the feature expansion experiment was the optimal model, which achieved the recall rates listed in Table 9. Finally, the large hook position, which may influence the judgment of trip in, trip out, and drilling, was added, resulting in further improvements as depicted in Fig. 10(g). The characterization parameters used in each experimental training are listed in Table 10.

After comparing the performance of the best models for each training, it was found that the use of 19 characterization parameters produced the best results. These 19 parameters are: standard well depth, drill bit position, turntable speed, top drive speed, large hook load, drilling pressure, standpipe pressure, pump punch #1, pump punch #2, pump punch #3, outlet flow, inlet flow, total pool volume, increment and decrement of pool water volume, CO_2 content, methane content, gas content, top-drive torque, and drill time.

6.3. Parameter

After determining the best combination of characterization parameters, the focus shifts to optimizing the model's parameters for improved performance. The objective is to identify the parameters that can be adjusted to achieve the highest possible recognition rate for the model. The main parameters that are modified include the number of BiGRU layers, the number of BiGRU hidden layer units, the starting learning rate value, and the activation functions. In this experiment, traditional GRU units are used instead of improved GRU units to identify the parameters and provide evidence for the research's innovative points.

The control variable method is used to experiment with the number of BiGRU hidden layers, and it is found that the best results are obtained when the number is set to six. The number of hidden layer cells is related to the dimensionality of the input and output, and a one-to-one or one-to-many quantitative relationship is maintained (Zhang, 2017). The number of BiGRU hidden layer units is set to 152, and the number of fully connected layer neuron units is set to 57, based on the experimental results.

The learning rate is another important parameter that affects the training effect. After several attempts, it is found that the best model effect is achieved when the initial value of the learning rate is set to 0.003. Various activation functions, such as the sigmoid, tanh, relu (Liu and Liang, 2021), and elu (Bai and Pei, 2018), are also considered. Among them, the elu function yields the best performing model. The Adamax optimization function (Kingma and Ba, 2014) is chosen because it has a simpler bound range for the learning rate.

Fig. 11(a)–(c) depict the variations in the traditional BiGRU unit with the attention mechanism network, utilizing the optimal parameter settings on the training, validation, and test sets, respectively. On the other hand, Fig. 11(d)–(f) illustrate the recall rate variations in the training, validation, and testing using the improved GRU unit proposed in this paper. The recall rate of the optimal model of the neural network composed of the traditional BiGRU unit and the improved BiGRU unit in the test set under each operating condition is presented in Table 11. The results indicate that the improved BiGRU unit leads to an average recall rate improvement of 3.2% in the model, reaching 91.63%. The recall rate improvement of the drilling operation is particularly evident, reaching 14.46%, while other operating conditions also have a slight improvement in recall rate. Table 12 displays the data flow dimension alterations through the neural network's various layers, and Table 13 presents the optimal parameter settings.

6.4. Comparative experiment

Table 14 is utilized in this paper to compare different network models and further demonstrate the advantages of the proposed method for multiclassification identification of drilling data. The test set used for the experiments is the same as the one used in the previous section. The models selected for comparison include RNN, LSTM, GRU, BiLSTM, BiGRU, BiLSTM + Attention, and BiGRU + Attention. RNN was the first network model used for classifying and predicting temporal data segments. One of its key features is the ability to apply previously extracted information to the current task. However, establishing long-term dependence is difficult when using RNN to extract information from long texts. LSTM adds input gates, forgetting gates, and control gates to

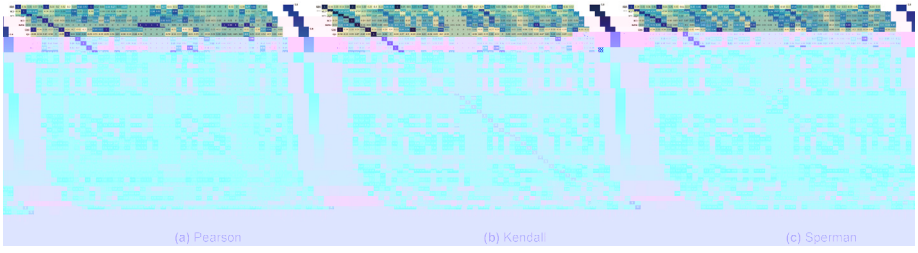

Fig. 9. Correlation heatmap matrix for dataset using Pearson, Kendall and Spearman coefficients.

Table 7
Categorization of drilling parameter features based on their physicochemical properties.

Type	Characterization parameter
Location related	Standard well depth, Drill bit position, Large hook load, Large hook position
Pressure related	Turntable speed, Top drive speed, Top-drive torque, Drill time
Pressure related	Drilling pressure, Standpipe pressure
Physical properties of drilling fluids	Export temperature, Entrance temperature, Export conductivity, Entrance conductivity
Drilling fluid volume related	Pump punch, Total pool volume, Increase or decrease in pool water volume, Outlet flow, Inlet flow
Contents of drilling fluid gas	CO_2 content, Methane content, Gas content

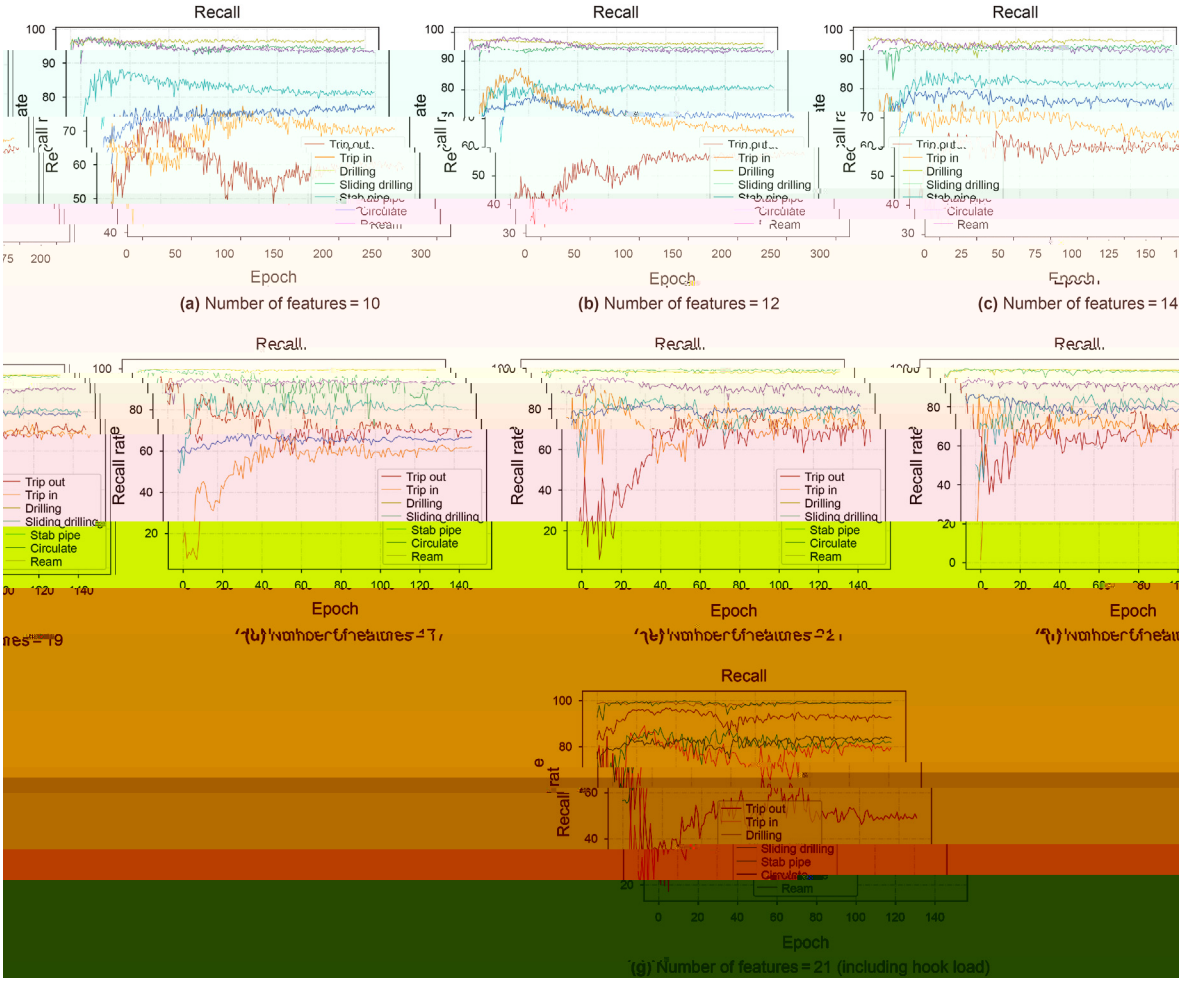

Fig. 10. Recall variation with different feature quantities for model training.

Table 8
Regression performance of initial optimal model using expanding drilling parameter features.

Evaluation metrics	Numerical value
Cross-entropy loss	0.4284
Trip out recall rate	65.29%
Trip in recall rate	65.36%
Drilling recall rate	95.81%
Sliding drilling recall rate	93.24%
Stab pipe recall rate	83.66%
Circulate recall rate	74.98%
Ream recall rate	93.16%
Total recall rate	81.64%

Table 9
Best performing model for drilling parameter feature expansion experiment.

Evaluation metrics	Numerical value
Cross-entropy loss	0.3941
Trip out recall rate	74.21%
Trip in recall rate	72.70%
Drilling recall rate	99.68%
Sliding drilling recall rate	98.72%
Stab pipe recall rate	84.28%
Circulate recall rate	80.36%
Ream recall rate	87.46%
Total recall rate	85.34%

Table 10
Features included in each round of drilling parameter feature expansion experiment.

Training number	Characterization parameters
1	10 Essential Features
2	All features in experiment 1, Outlet flow, Inlet flow
3	All features in experiment 2, Total pool volume, Increase or decrease in pool water volume
4	All features in experiment 3, CO_2 content, Methane content, Gas content
5	All features in experiment 4, Export temperature, Entrance temperature, Export conductivity, Entrance conductivity
6	All features in experiment 4, Top-drive torque, Drill time
7	All features in experiment 6, Large hook position

traditional RNN units to extract information and establish long-term dependencies more efficiently. The GRU unit curtails a control gate compared to the LSTM unit to improve computational efficiency and converge to the local optimum based on guaranteed performance. Traditional LSTM and GRU can only process the current task based on previous data. Bidirectional LSTM and GRU are similar to unidirectional ones in theory. They create a hidden layer by combining a forward network unit group and an inverse network unit group. The forward network unit can use historical data, while the reverse network unit can use data from the future. They collaborate to obtain current information, and this method

Table 11
Performance of the traditional and improved BiGRU network.

Recall rate	Traditional	Improved	D-value
Cross-entropy loss	0.2833	0.2035	−0.0798
Trip out	85.60%	87.15%	+1.55%
Trip in	71.74%	86.59%	+14.85%
Drilling	98.92%	99.31%	+0.39%
Sliding drilling	99.60%	98.33%	−1.27%
Stab pipe	86.75%	92.65%	+5.9%
Circulate	81.67%	85.53%	+3.86%
Ream	94.70%	91.86%	−2.84%
Total	88.43%	91.63%	+3.2%

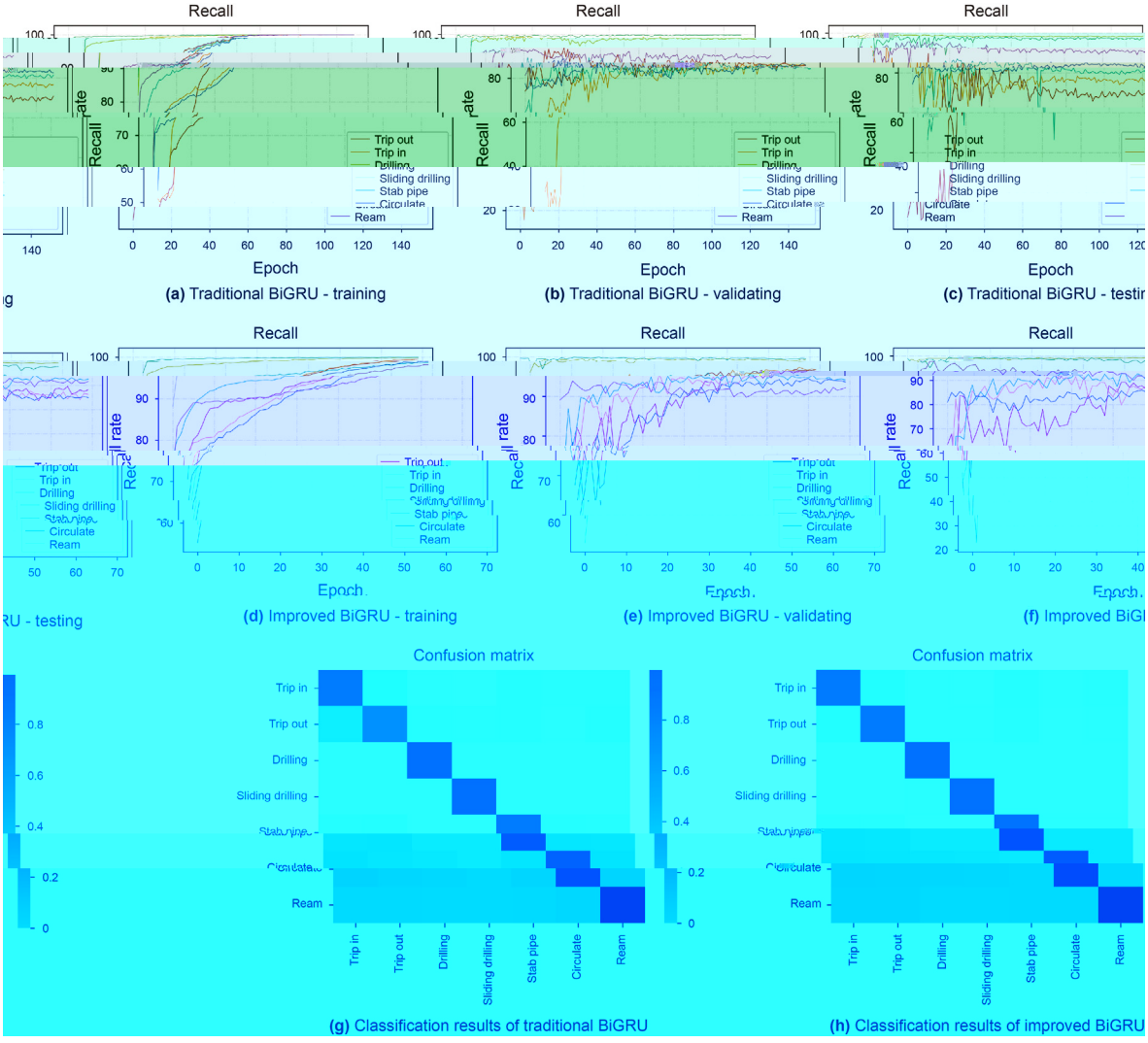

Fig. 11. Performance comparison between traditional and improved GRU models.

Table 12
Data dimensionality changes across neural network layers.

Dimensional change	
Mini-batch	128 × 60 × 19
BiGRU	128 × 60 × 152
Attention net	128 × 152
Batch normalization	128 × 152
Fully connected Network 1	128 × 57
Fully connected Network 2	128 × 7

produces a more accurate model than the one-way LSTM. The addition of the attention mechanism improves the model performance by assigning attention weights to the information extracted from the previous layer of units. In addition, we compared the performance of a traditional bidirectional GRU network with an attention mechanism and a neural network model composed of improved GRU units. These models used a consistent number of hidden units in the recurrent neural network, the same number of layers in the fully connected network, and slightly different settings for other hyperparameters to obtain better-performing models. The best performing models during training were selected for comparison in the experiments. Table 14 compares the performance of each network in the test set by comparing the above network structures. Fig. 12 shows the individual recall of each network's best model for each operating condition as a histogram.

The effective fusion network model demonstrates the superiority of the improved BiGRU + Attention model with the highest overall test recall rate of 91.63% and the lowest loss value of 0.2035. However, the use of the softsign function makes it slower than the traditional BiGRU-based network, placing it in second place. The recall rates of the improved BiGRU with attention mechanism model are over 85% for all seven working conditions, with a recall rate of over 90% for four of them, as shown in Table 11. These results highlight the feasibility and generalizability of the BiGRU with attention mechanism method based on semi-experience decision trees for drilling condition time series recognition. Moreover, the proposed method ensures a high level of accuracy and practical application value. By organizing the samples according to the proposed method, the data information of the samples is constrained within a reasonable range by the expert system, which ensures that the neural network learns within a controlled range during training.

Table 13
Neural network hyperparameter configuration.

Optimum model parameters	
BiGRU hidden layer	6
BiGRU hidden layer cell number	152
Fully connected layer	2
Fully connected units in the first layer	152
Fully connected units in the second layer	57
Learning rate initial value	0.003
Batch number	128
Drop out rate	0.3

7. Discussion

In many studies on work condition identification, decision trees based on empirical formulas are commonly used for data labeling. In this paper, we propose a data labeling method that combines empirical formulas with actual data to construct a library of relationships between drilling conditions and characterization parameters. Based on this library, a decision tree is prepared, which outperforms the traditional empirical formula in the data set of six wells, with an average accuracy improvement of 14.59%, ranging from 9.18% to 20.24%. By labeling data based on key characterization parameters, the interpretability of the final model is significantly improved.

This paper presents the first application of a neural network composed of BiGRU + Attention for drilling condition recognition, which is more sophisticated compared to previous drilling condition identification networks. While some previous networks required the weakening of the drilling data to meet the input requirements, the critical information about drilling conditions is actually in the temporal variation of the features. The GRU unit, which proposes the concepts of reset and update gates, does not use all of the historical information when processing the sequence task like the RNN unit does. Instead, it uses reset and update gates to control the processing of different data types. The reset gate regulates how much historical data is used and combines it with the input data to create a new hidden state, while the update gate orchestrates the fusion ratio of historical information with the new hidden state, resulting in the current time step's output. Such network units with the ability to filter information can effectively extract implicit information from time-series data, making long-term dependence easier to establish.

Moreover, this paper has made targeted improvements to the GRU unit based on the use of the BiGRU network, focusing on improving the ability to express and handle historical data. Experiments have demonstrated that the improved GRU unit can bring a 3.2% accuracy improvement to the model. By incorporating the attention mechanism, the proposed BiGRU + Attention model achieves a higher overall test recall rate of 91.63% and a lower loss value of 0.2035. Although the improved model is slower than the traditional BiGRU-based network due to the use of the softsign function, it still has high practical value. The experiments have proven that the proposed method based on semi-experience decision trees is feasible and generalizable for drilling condition time-series recognition, with a high level of accuracy and practical application value. The proposed data labeling method based on empirical formulas can also significantly improve the final model's interpretability by labeling data based on key characterization parameters.

The information extraction capability is further enhanced when

Table 14
Comparison of neural network models' performance on study dataset.

Model name	Total recall	Loss value	Best model appearance round
RNN (Elman, 1990)	58.09	0.9272	25 (unconverged)
LSTM (Hochreiter and Schmidhuber, 1997)	74.17	0.4967	52
GRU (Cho et al., 2014)	73.16	0.5131	49
BiLSTM (Schuster and Paliwal, 1997)	82.21	0.4084	73
BiGRU	83.02	0.3941	58
BiLSTM + Attention (Pappas and Popescu-Belis, 2017)	86.38	0.3371	63
Tradition BiGRU + Attention	88.43	0.2833	**37**
Proposed	**91.63**	**0.235**	47

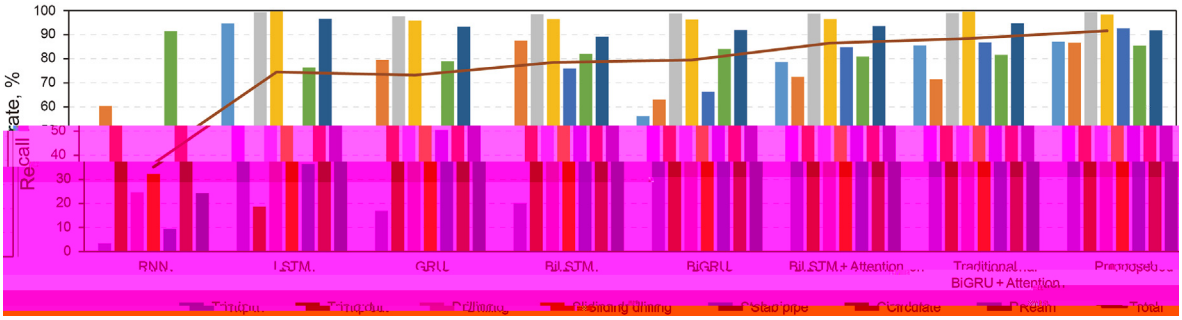

Fig. 12. Performance comparison of different network structures in identifying drilling conditions.

two GRU units with inverted directions are connected in parallel to form a BiGRU unit. This study's comparative experiments demonstrated the advantages of a bidirectional network, with LSTM and GRU networks using a bidirectional network improving average recall by 8.04% and 9.86%, respectively.

The attention mechanism is combined with BiGRU to further improve model performance. The softsign function is used instead of the tanh function in this paper to improve the perceptron activation function in the attention mechanism, which has a flatter curve and slower decreasing derivatives, making it more efficient for learning. The trends and numerical characteristics of each characterization parameter over the sample time are contained in the hidden states generated by BiGRU. When BiGRU's information from the original data block passes through the attention unit, the unit assigns attention weights to individual sequence elements based on their impact on working condition recognition. This weight will direct the model to pay more attention to the areas that need to be addressed, resulting in improved model performance. LSTM and GRU networks that used the attention mechanism improved average recall by 4.17% and 5.41%, respectively, demonstrating the effectiveness of this task.

During the training phase, various model enhancement methods are used, including weight initialization, data normalization, batch normalization, and adaptive learning rate. These methods improve the model's generalization ability and convergence speed by removing the effect of magnitude, disrupting the symmetry of the data, and improving the flow gradient through the network. The resulting model outperforms the other six networks in terms of average recall and convergence speed on the test set and the real-time and generalization capabilities required for practical use.

8. Conclusion

This paper presents a novel approach to classifying and identifying seven common drilling conditions simultaneously, using an improved BiGRU neural network algorithm with an attention mechanism. The proposed model outperforms existing networks in terms of recognition accuracy and generalization ability, achieving a model accuracy of 91.63% on the test set. The improved GRU units provide an average accuracy improvement of 3.1%, addressing the challenge of organizing high-dimensional data effectively. Compared to other drilling condition classification methods, the proposed technology is more mature, closely related to practical production, and has wide applicability and expandability. It is capable of meeting the real-time requirements of drilling condition recording and equipment monitoring during drilling, making it of significant practical value.

In the future, we plan to investigate other drilling conditions during drilling, optimize the proposed model, and explore its application across the entire drilling lifecycle.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was supported by open fund (PLN2021-23) of National Key Laboratory of Oil and Gas Reservoir Geology and Exploitation (Southwest Petroleum University).

References

Arnaout, A., Fruhwirth, R., Esmael, B., et al., 2012. Intelligent Real-Time Drilling Operations Classification Using Trend Analysis of Drilling Rig Sensors Data. SPE Kuwait International Petroleum Conference and Exhibition, Kuwait, USA. https://doi.org/10.2118/163302-MS.

Bai, Y., Pei, J., 2018. Application of an improved ELU convolutional neural network in SAR image ship detection. Surveying and Mapping Bulletin (1), 4. https://doi.org/10.13474/j.cnki.11-2246.2018.0024 (in Chinese).

Ben, Y., James, C., Cao, D., 2019. Development and Application of a Real Time Drilling State Classification Algorithm with Machine Learning. SPE/AAPG/SEG Unconventional Resources Technology Conference, Denver, Colorado, USA. https://doi.org/10.15530/urtec-2019-253.

Caldwell, B., Hinton, J., 2015. Data Drilling: Changing the Way the Oil and Gas Industry Manages Safety and Risk. SPE Health, Safety, Security, Environment, Denver, Colorado, USA. https://doi.org/10.2118/173497-MS.

Chen, Q., 2021. Improve drilling supervision and management to enhance wellbore quality control. China Petroleum and Chemical Standard and Quality 41 (17), 2. https://doi.org/10.3969/j.issn.1673-4076.2021.17.012 (in Chinese).

Cho, K., Van Merriënboer, B., Bahdanau, D., et al., 2014. On the properties of neural machine translation: encoder-decoder approaches. Computer Science. https://doi.org/10.3115/v1/W14-4012.

Chorowski, J., Bahdanau, D., Cho, K., et al., 2014. End-to-end continuous speech recognition using attention-based recurrent nn: first results. Eprint Arxiv. https://doi.org/10.48550/arXiv.1412.1602.

Cinar, Y.G., Mirisaee, H., Goswami, P., et al., 2017. Position-based content attention for time series forecasting with sequence-to-sequence RNNs. Neural Information Processing: 24th International Conference, ICONIP 2017. Guangzhou, China 24, 533–544. https://doi.org/10.1007/978-3-319-70139-4_54.

Elman, J.L., 1990. Finding structure in time. Cognit. Sci. 14 (2), 179–211. https://doi.org/10.1016/0364-0213(90)90002-E.

Ge, Y., Zhang, Y., Ge, S., et al., 2022. Development of Automatic Recognition and Recording System for Rig Jobs. In: SPE/IADC Drilling Conference and Exhibition. Galveston. Texas, USA. https://doi.org/10.2118/208726-MS.

Hochreiter, S., Schmidhuber, R.A., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Hu, C., Xue, W., 2019. Bi-GRU text emotion analysis model based on attention mechanism. World Scientific Research Journal 5 (9), 296–301. https://doi.org/10.6911/WSRJ.201909-5(9).0033.

Keskomon, N., Harnsomburana, J., 2020. Thai character-word long short-term memory network language models with dropout and batch normalization. et al. International Journal of Machine Learning and Computing 10 (6), 783–788. https://doi.org/10.18178/ijmlc.2020.10.6.1006.

Khudiri, M., Contreras, W., Sanie, F., et al., 2015. Saudi Aramco Real-Time Drilling Operation Activity Recognition. SPE Middle East Intelligent Oil and Gas Symposium. https://doi.org/10.2118/176799-MS.

Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. https://doi.org/10.48550/arXiv.1412.6980.
Li, Y.H., Tian, H.F., Zhang, Y., 2010. An Improved Gaussian Mixture Background Model with Real-Time Adjustment of Learning Rate. In: 2010 International Conference on Information, Networking and Automation (ICINA). Kunming, China. https://doi.org/10.1109/ICINA.2010.5636758.
Liu, B., Liang, Y., 2021. Optimal function approximation with ReLU neural networks. Neurocomputing 435, 216–227. https://doi.org/10.1016/j.neucom.2021.01.007.
Liu, G., Tao, Y., 2015. Research on prediction of drilling pressure based on time series. Petrochemical Industry Application 34 (4), 43–46+50. https://doi.org/10.3969/j.issn.1673-5285.2015.04.013 (in Chinese).
Liu, J., Zhang, T., 2021. Research on PCA-RF-based sticking prediction method. Journal of Beijing Information Science & Technology University 36 (1), 18–22. https://doi.org/10.16508/j.cnki.11-5866/n.2021.01.004 (in Chinese).
Mittal, A., Singh, A.P., Chandra, P., 2021. Weight and bias initialization routines for sigmoidal feedforward network. Appl. Intell. 51 (4), 2651–2671. https://doi.org/10.1007/s10489-020-01960-5.
Pappas, N., Popescu-Belis, A., 2017. Multilingual hierarchical attention networks for document classification. arXiv preprint. https://doi.org/10.48550/arXiv.1707.00896.
Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. IEEE Trans. Signal Process. 45 (11), 2673–2681. https://doi.org/10.1109/78.650093.
Tang, R., 2017. Research on the method of data normalization for improve SVM training efficiency. J. Shanxi Normal Univ. Nat. Sci. Ed. CNKI:SUN:SDZK.0.2016-04-012 (in Chinese).
Ting, S., Ying, Z., Jin, Y., et al., 2019. Real-time intelligent identification method under drilling conditions based on support vector machine. Petroleum Drilling Techniques 47 (5), 28–33. https://doi.org/10.11911/syztjs.2019033 (in Chinese).
Todorov, D., Thonhauser, G., 2014. Hydraulic monitoring and well-control event detection using model-based analysis. J. Petrol. Technol. 66 (9), 144–147. https://doi.org/10.2118/0914-0144-JPT.
Vaswani, A., Shazeer, N., Parmar, N., et al., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30. https://doi.org/10.48550/arXiv.1706.03762.
Wang, X., 2019. Prediction algorithm of higher education school scale based on weight initialization-multilayer convolutional neural network sliding window fusion. Information Technology and Informatization (10), 6. CNKI:SUN:SDDZ.0.2019-10-014 (in Chinese).
Wei, P., 2014. Research and Development of Data Processing and Display Module of Drilling Engineering Parameter Monitoring System. Chongqing University, MA thesis (in Chinese).
Zhang, S., 2017. Hidden layer node estimation algorithm of bp network based on simulated annealing. J. Hefei Univ. Technol. (Nat. Sci.): Natural Science Edition 40 (11), 4. CNKI:SUN:HEFE.0. 2017-11-010 (in Chinese).
Zhao, J., Shen, Y., Chen, W., et al., 2017. Machine learning-based trigger detection of drilling events based on drilling data. In: SPE Eastern Regional Meeting. Lexington, Kentucky, USA. https://doi.org/10.2118/187512-MS.