IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 16, 2019, accepted January 27, 2019, date of publication January 31, 2019, date of current version March 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2896621

An Enhanced LSTM for Trend Following of Time Series

YAO HU, XIAOYAN SUN, (Member, IEEE), XIN NIE, YUZHU LI, AND LIAN LIU
Information and Control Engineering College, China University of Mining and Technology, Xuzhou 221116, China

Corresponding author: Xiaoyan Sun (xysun78@ 126.com)

This work was supported by the National Nature Science Foundation of China, under Grant 61473298 and Grant 61876184.

ABSTRACT Mining and analysis of time series data (TSD) have drawn a great concern, especially in the TSD clustering, classification, and forecast. In the industrial field, e.g., the work condition monitoring and the environmental safety, it is crucial to follow the trend of the corresponding TSD for a safety forecast, and few studies have been devoted to such a trend following. Motivated by this, we propose a trend following the strategy of TSD by using a long short-term memory (LSTM) network for safety forecast, in which the training method aggregates the particle swarm optimization (PSO) algorithm with gradient descent (GD) to obtain more competitive model parameters. Three kinds of trend representations of TSD are first defined based on the corresponding research in stock option. Then, the LSTM optimized with the PSO-GD is developed to perform the trend following. From the viewpoint of safety forecast, the trends varied in different time length are further predicted and analyzed. The superiority of the proposed algorithm is experimentally demonstrated by applying it to the electromagnetic radiation intensity TSD sampled from an actual coal mine and PM2.5 in UCI repository.

INDEX TERMS LSTM, PSO, safety farecast, series data, trend following.

I. INTRODUCTION

Data mining have been widely used in our real life, and it refers to automatically extract intriguing, implicit, previous unknown, and potential helpful patterns or knowledge from large-scale mass data [1]. Time series data (TSD) are an important type of high dimensional data with high complexity, dynamics and noise. TSD are a series of data points indexed (or listed or graphed) in time order [2]. Examples of time series are stock market transactions, heights of ocean tides, counts of sunspots, and status collected from sensors [3]. Many fruitful data mining studies have been devoted to the classification [4], clustering [5], forecast [6] and visualization [7] of TSD. The existing research on TSD prediction in the industry field is further addressed here since it is closely related to our algorithm.

A large amount of diversified TSD is generated at high speeds from industrial equipment and formed as industrial big data. For the industry, environmental safety and equipment health are very important and can be forecasted based on their corresponding TSD, and machine learning methods have been well developed and widely applied. In literature [8], researchers proposed a Logistic regression based rock burst early-warning model, practice has proven the model achieved higher forecasting accuracy than conventional mathematics methods. The Markov Chain model was applied to forecast the lifecycle of expressway charging equipment based on the historical TSD in [9]. This method provides theoretical support for the life management of electromechanical equipment. [10] proposed a multi-objective genetic algorithm coupled artificial neural network to predict water quality. Compared with the SVM and artificial NNs trained with PSO, the proposed model exhibited a superior accuracy in processing water quality TSD. In the field of energy internet, literature [11] pointed out that the key to ensuring the safe and stable operation of the energy internet lies primarily in the accurate prediction of the load, which is a typical TSD. Deep learning based TSD prediction had been performed with great successes in the load TSD in guaranteeing the accuracy of load forecasting in [12].

Clearly, these works on TSD prediction for industrial big data have mostly emphasized on the accuracy. However, the variation trend within a certain period of time rather than a specific data point at a certain moment of the corresponding TSD is more significant for environmental safety or equipment health [13], [14]. For example, in water quality

The associate editor coordinating the review of this manuscript and approving it for publication was XXXXX.

monitoring, the conclusion that an artificial pollution source may exist can be obtained if a 13-month consecutive upward trend appears in the water dissolved oxygen TSD of an area. That is, even if we could get an accurate prediction of the water dissolved oxygen of each month, we cannot obtain valuable information for the pollution source unless we track the variation trends of the 13-month TSD.

The concept of trend following of financial TSD has been sufficiently studied in the market price or stock option, however, in the industry, especially for safety forecast, trend following of the industry TSD has a wide range of demands, but few studies have been reported to date. Therefore, it is necessary to develop powerful methods to solve the aforementioned problems. To this end, the trend of TSD should be first defined, which can be achieved by transferring the corresponding knowledge from the financial TSD. Once the trend of industrial TSD is defined, trend following becomes an essential new form of a TSD prediction.

LSTM provides a very elegant way to deal with TSD through time that embodies the correlations among the TSD in time order [15], [16], and lots of research have presented the superiority of LSTM on TSD. In literature [17], researchers embedded bottleneck vocal features in LSTM architecture to measure the pain level through voice characteristics, and the results showed that the classification accuracy are higher than other comparisons. Reference [18] utilized LSTM to predict the coding unit splitting, and the experimental results demonstrated its advantages in high efficiency. The trend following of a TSD is also a new form of a TSD prediction, therefore, LSTM is a natural choice here.

LSTM is outstanding in dealing with TSD only when its network parameters, e.g., weights and bias, are finely tuned or optimized. GD is the most popular training method for the LSTM due to its relatively easy implementation and has been met with great successes in practice [19]–[21]. GD training, however, may cause the 'vanishing/exploration gradient problem' which will greatly impede the network [22]. Designing more effective optimization method to train a LSTM is critical in applying it to TSD prediction. Evolutionary algorithms assisted optimization methods have proven to be feasible and powerful in traditional neural networks and thus such scheme is considered here.

Evolutionary algorithm (EA), which mainly contains PSO algorithm [23], Genetic algorithm [24], Ant colony [25], etc., is a generic population-based metaheuristic optimization algorithm. The algorithms mainly depend on stochastic optimization methods and do not need the process of gradient calculation. Compared with other evolutionary algorithms, the PSO algorithm is able to keep a memory of good solution to update the particles in the iteration process while the GA cannot [26]; meanwhile the algorithm is relatively easily implemented with no need of crossover or mutation operations. Because the algorithm intend to optimize the weights of LSTM, an algorithm that is efficient to solve continuous problem should be selected, thus the ant colony is not suitable for this problem [27]. Based on that, the PSO algorithm is chosen here.

Lots of research associating PSO with the NNs, e.g., in [28] and [29] the PSO-RNN, PSO-ESN are trained mainly by PSO instead of GD. Compared with the GD training methods, the PSO is less sensitive to weight initializations, less likely to be trapped in local optima, and independent of the activation function gradient [30]. Therefore, PSO is not susceptible to the gradient problem [31]. It seems that the PSO method is a potentially good deep learning approach, but research has indicated that the pure PSO training method struggles to train relatively large neural networks [32]. Reference [33] presented an integration framework of the PSO and GD for the LSTM when identifying handwriting. In his work, PSO was performed before the LSTM to arrive closer to the global minimum. The learning process of LSTM to fulfill the prediction of TSD is dynamic while pre-using the PSO is too complex for proper implementation. Therefore, new integration strategies for the PSO and GD must be developed.

Motivated by these factors, we present a novel PSO-GD-aggregated LSTM and show its effectiveness in performing trend following for industry TSD. Specifically, the trend definition of the TSD is first derived based on the volatility definition of the stock option [34]. Then, the trend following is converted into a trend TSD prediction and the LSTM is selected to accurately approximate the trend to provide a safety forecast. Different from traditional training methods, we propose a PSO-GD method by aggregating the GD with the PSO to train the LSTM. The trend of TSD in multiple time periods can be obtained using the proposed PSO-GD-LSTM. In the end, the proposed PSO-GD-LSTM model with different trend representations is first validated using the electromagnetic radiation intensity data sampled from a coal mine and the PM2.5 data from the US Embassy in Beijing, and is then applied for time various trend following of the two datasets for safety forecast.

The main contributions of our work are as follows: (1) The trend following for safety forecast is modeled as a TSD prediction based on three kinds of trend calculation methods with different time window lengths. (2) A PSO-GD-aggregated LSTM is presented by iteratively applying PSO and GD methods to optimize the structural parameters of the LSTM with lower calculation cost and higher accuracy. (3) The enhanced LSTM is applied to follow the trend of electromagnetic radiation intensity data sampled from a coal mine and the PM2.5 data from the US Embassy in Beijing for safety forecast.

The remaining of the paper is organized as follows. Three types of trend representations for TSD are presented in Section 2, while Section 3 illustrates the LSTM model for trend following. In Section 4, we present the detail of the training methods in aggregating the PSO and GD approaches. The application of the proposed algorithm in the electromagnetic radiation intensity data of a coal mine and PM2.5 data in Beijing together with the experimental analysis are demonstrated in Section 5. Finally, the conclusions are followed.

II. TREND REPRESENTATION OF TIME SERIES DATA

Times series data are a sequence of discrete-time instances taken at successive and equally-spaced time points. For simplicity, the industry TSD are denoted as $Y = \{p_t, t \in T\}$, $T = \{1, 2, \cdots M\}$, where T is the discrete-time index set with M time points. Considering a length l for the time window of Y, three trend representations of the k-th time span, i.e., the average $A_{k,l}$, the logarithm $r_{k,l}$ and the standard deviation $\sigma_{k,l}$, are presented here based on their similar definitions in the stock option.

For the T discrete-time index set, the number of points in the time span can be easily calculated as $N = \lfloor T/l \rfloor$, where $\lfloor . \rfloor$ represents the floor integer of an element. Then, the $A_{k,l}$ is defined in Eq. 1.

$$A_{k,l} = \frac{\sum_{t=(k-1)l+1}^{kl} p_t}{l}, k = 1, 2, \cdots, N \tag{1}$$

The logarithmic operation is universally applied in TSD processing, which simply uses the log differences between the starting and ending values of the considered length l of the time window to compresses the data scale. The logarithmic metric over the period l, is defined in Eq. 2.

$$r_{k,l} = \ln(\frac{p_{kl}}{p_{(k-1)l+1}}) = \ln(p_{kl}) - \ln(p_{(k-1)l+1}) \tag{2}$$

The logarithm of the trend definition is simple since only two points of a time period are used. However, the information of other data in this period are ignored, which cannot sufficiently reflect the overall trend of this time span. Therefore, the standard deviation of the logarithm based metric is further addressed according to the volatility definition of the market price in [34].

The logarithm of each pair of (p_t, p_{t+l-1}) is first calculated on Y as:(3)

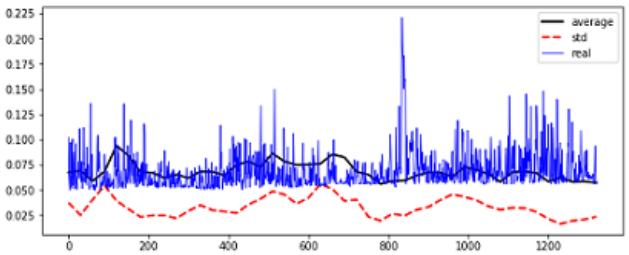
$$h_{k,l} = \ln(\frac{p_{t+l-1}}{p_t}) = \ln(p_{t+l-1}) - \ln(p_t) \tag{3}$$

Then, the standard deviation of the arithmetic or logarithmic mean over a k-th time window $\sigma_{k,l}$ can be calculated as follows.

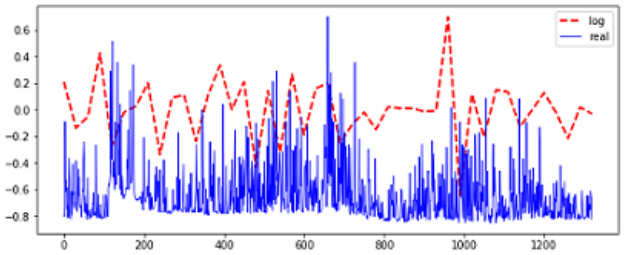
$$\sigma_{k,l} = \sqrt{\frac{\sum_{t=(k-1)l+1}^{kl} (h_{t,l} - \bar{h}_{k,l})^2}{l-1}}, k = 1, 2, \cdots, N \tag{4}$$

where $\bar{h}_{k,l}$ is the average of $h_{t,l}$ on the k-th time window.

The values of these three trend representations are illustrated in Figures 1 (a) and (b) by comparing with the TSD of the electromagnetic radiation intensity of a coal mine. A total of 1320 points are normalized or z-scored for comparability in the same coordinates, and the l is set to 30. The values of $A_{k,l}$ and $\sigma_{k,l}$ are plotted in Figure 1 (a) due to their similar variation scales. The same values for $r_{k,l}$ are given in Figure 1 (b). $\sigma_{k,l}$ and that of the average $A_{k,l}$ Several conclusions are observed from these two figures. (1) The values of the standard deviation (*std*) have similar



(a)



(b)

FIGURE 1. Visualization of the indicators. (a) Curves of $\sigma_{k,l}$,$A_{k,l}$ and the real data. (b) Curves of logarithmic $r_{k,l}$ and the real data.

variation trends, i.e., on the same time span points, the trends of the TSD on these two metrics are similar. (2) The curves of $\sigma_{k,l}$ and $A_{k,l}$ are relatively smooth, and the $\sigma_{k,l}$ is more stable. (3) The trend values calculated with the logarithmic $r_{k,l}$ vary greatly and may have different variations to the other metrics, i.e., between two adjacent time spans, the values of the logarithmic increase while that of the $\sigma_{k,l}$ and $A_{k,l}$ decrease.

Accordingly, the average method is simpler, and the logarithmic-based standard deviation is more stable, which should be more reliable than the logarithmic one in depicting the trend of the industry TSD. These defined trend representations will be further predicted by the LSTM strategy and further compared.

III. LSTM-BASED TREND FOLLOWING OF TSD

From the aforementioned trend representations, the trend of the TSD can be represented as $Tr = \{x_{k,l}, k = 1, 2, \cdots, N\}$ with l being the time window length. The values of $x_{k,l}$ are determined by the l, and the corresponding trend representation, i.e., the $x_{k,l}$ can be one of the form $A_{k,l}$,$r_{k,l}$ or $\sigma_{k,l}$. In this scenario, trend following of the TSD is to predict the values of $x_{N+1,l}, x_{N+2,l}, \cdots$, i.e., to predict the trend of the source TSD in the time span of $\{T+1, T+2, \cdots, T+l\}$,$\{T+l+1, T+l+2, \cdots, T+2l\} \cdots$. Clearly, trend following of the TSD is converted into a TSD prediction on $Tr = \{x_{k,l}, k = 1, 2, \cdots, N\}$. Accordingly, powerful machine learning methods and models in the TSD prediction can be applied to the trend following problem, besides, based on the analysis in the Introduction, LSTM is designed here to perform the TSD trend following.

LSTM is composed of collections of cells and is one of the recurrent neural network (RNN) models developed to ease the dilemma of gradient vanishing or explosion in

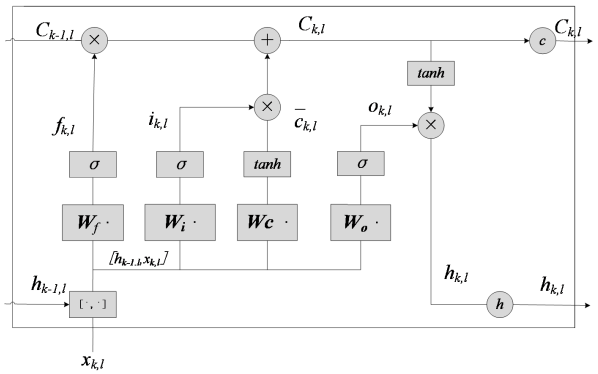


FIGURE 2. LSTM for trend following of a TSD.

the back-propagation algorithm when processing TSD with long-term dependency [35]. In our algorithm, a cell model of the LSTM is shown in Figure 2. For the presented trend following, the input of the LSTM is the trend values of $x_{k,l}$ and the training set $Ts = \{x_{k,l}, k = 1, 2, \cdots, L\}$ is obtained by selecting the first L elements from $Tr = \{x_{k,l}, k = 1, 2, \cdots, N\}$. The remaining $N - L$ elements of T_r are set as the testing samples. With the training set T_s, the trend following process can be implemented with Eq. 5 to Eq. 10, and the output $h_{k,l}$ in Eq. 11 is the forecasted trend of $\hat{x}_{k+1,l}$.

$$f_{k,l} = \sigma(W_f * [h_{k-1,l}, x_{k,l}] + b_f) \tag{5}$$
$$i_{k,l} = \sigma(W_i * [h_{k-1,l}, x_{k,l}] + b_i) \tag{6}$$
$$o_{k,l} = \sigma(W_o * [h_{k-1,l}, x_{k,l}] + b_o) \tag{7}$$
$$\bar{c}_{k,l} = \tanh(W_c * [h_{k-1,l}, x_{k,l}] + b_c) \tag{8}$$
$$C_{k,l} = f_{k,l} * C_{k-1,l} + i_{k,l} * \bar{c}_{k,l} \tag{9}$$
$$h_{k,l} = o_{k,l} * \tanh(C_{k,l}) \tag{10}$$
$$\hat{x}_{k+1,l} = W_{k,l}(o_{k,l} * \tanh(C_{k,l})) + b_{k,l} \tag{11}$$

The common training method in the LSTM is essentially the same as that used in traditional BP neural networks [35]. Compared with traditional shallow NNs, the LSTM is a deep learning model with significant growth in the number of parameters along with an increase of the cells. In our model, the input and output of the LSTM is 1-dimension. Supposing the LSTM has one hidden layer and H cells in each layer, the total number of model parameters can be $np = 4H^2 + 9H + 1$. For example, if $H = 5$, then $np = 146$, and if $H = 10$, then $np = 491$. Clearly, the optimization of the LSTM parameters is a critical but hard task. We present a PSO-GD method here to train the LSTM by integrating both the merits of the PSO in exploring and GD in exploitation and speed. The presented PSO-GD optimization will be further illustrated in the following section.

IV. IMPROVED LSTM WITH PSO-GD-AGGREGATED TRAINING

Particle swarm optimization (PSO) is a kind of population-based, meta-heuristic, optimization approach inspired by the social behavior of birds within a flock [23]. As previously addressed, PSOs have been successfully applied to optimize NNs by encoding the weights, bias or even structure parameters into particles [36]. For the LSTM, which is a kind of deep learning model, the PSO cannot be simply used as before due to the following two reasons. First, the number of optimized parameters is larger (e.g., 146, 491 or even bigger), which is a high dimensional optimization for the PSO and is difficult to obtain satisfactory solutions with an acceptable computational cost. Second, the LSTM is often designed to online fulfill the TSD, the computation should be much more rapid, which is quite difficult for the population based PSO. The GD is fast for training the LSTM, with the exception of local trapping. Therefore, developing an integration of the PSO and GD is a natural methodology.

Our purpose here is to optimize the weights and bias values for the trained LSTM, and the main idea of the PSO-GD method is as follows. We first train the LSTM network using the GD approach for a few iterations. Then, we send the GD obtained model parameters i.e., the weights and biases, of the LSTM network to the PSO procedure as a reference and informative solution to generate competitive initial particles for further exploring, which is expected to reduce the possibility of the ''vanishing/exploration gradient problem'' or the local optima problem. The model parameters of the LSTM optimized from the PSO are again fed into the GD. Such a loop can be repeated until the performance of the LSTM is satisfied. The detail aggregation of GD and PSO is developed as follows.

A. GD SESSION

(1) Set the hyper-parameters including the structure of the LSTM, the dimensions of input and the output of the LSTM, denoted as input_dim and output_dim, learning rate, time step and number of iterations.

(2) The gradient descent algorithm is selected and used to first train the LSTM for T iterations by inputting the calculated trend TSD, and the model parameters, i.e., weights and bias of the LSTM, are recorded as $W(t), t = 1, 2, \ldots T$.

B. PSO SESSION

(1) Calculate the total number of parameters to be optimized. For a standard LSTM, i.e., an input layer with input_dim inputs, one hidden layer with H cells and a full connect layer with output_dim outputs, the number of parameters can be calculated as $4 * ((H + input\_dim) * H + H) + (H + output\_dim)$ in our algorithm. Both the input_dim and output_dim are 1 and $H = 5$; therefore, a total number of 146 parameters are optimized. Real-number encoding is a good choice for such an optimization and can be expressed as $W = (W_f, W_i, W_c, W_o, b_f, b_i, b_c, b_o)$ according to the structure of LSTM.

(2) Initialize the PSO population based on $W(t)$ obtained from GD session. The PSO population is initialized here by considering the convergence information involved in $W(t)$ and the local escaping requirement. Several methods can be used to generate the initial population of the PSO. For

example, randomly selecting some optimized parameters as part of the initial particles or selecting parameters according to their prediction errors. In order to possibly cover more locals, we present an alternative selection method for generating competitive particles. T is divided equally into M batches, and the parameter sets $W(Ti/M), i = 1, 2, \ldots, M$ are selected as the bases of the initial particles, that's to say the weight parameters obtained from different iterations will be selected as the initial position of each particle. A 1/3 population size of the particles are directly copied from these parameter sets, and another 1/3 of the particles are generated by adding a Gaussian perturbation to the finally obtained parameter set $W(T)$. For the last 1/3 of the particles, we fully consider the process of weights convergence in GD thus randomly generate particles in the scope between the values of $W(T)$ and $W(T - T/M)$.

(3) Calculate the fitness of the particles. Each particle corresponds to a LSTM network, and we take the root mean square error (RMSE) between the LSTM output and the true value as the fitness of the particles.

(4) Assign the number of particles as I and iteration times in every particle as S, and then operate the PSO algorithm according to the following equations.

$$V_i(s+1) = \omega * V_i(s) + r_1 c_1 (pbest_i - W_i) + r_2 c_2 (gbest - W_i) \quad (12)$$

$$W_i(s+1) = W_i(s) + V_i(s+1) \quad (13)$$

where ω is the inertia weight, which controls the impact of the previous history on the new velocity, c_1 and c_2 are the acceleration terms that pull each particle X_i towards its own current best position and the population global best position, and r_1 and r_2 are two random numbers in the range [0,1], whose values can be found by trial and error.

(5) Output the best particle as the starting point of the next GD session.

V. APPLICATIONS

We apply our algorithm to two trend followings of TSDs for safety forecast, i.e., the electromagnetic radiation intensity data sampled in every minute from a Chinese coal mine and the PM2.5 data sampled in one hour from the US Embassy in Beijing.

A. BACKGROUND

The dynamic coal or rock disasters, which mean large amounts of coal and rock burst out from the mass in a short period accompanying with a strong dynamic phenomenon, can cause significant personal casualties and property losses [37]. Electromagnetic radiation intensity (ERI) can reflect the stress state and deformation level of coal and its amplitude shocks with the dynamic energy changes [38]. The variation trend of the ERI-TSD is an important indicator for judging the occurrence of dynamic coal or rock disasters for underground coal mines [39]. Three cases are usually considered for safety forecast of the dynamic damage according to the trends of the ERI-TSD. Mines are considered safe if the variation trend of the ERI-TSD remains relatively stability, i.e., no large or sudden changes in the trend values. A high probability for the occurrence of dynamic disasters will appear if a continuous rising tendency of the ERI-TSD occurs, and the safety forecast should be activated with a high level of concern. The most dangerous scenario is that a sudden high change of the ERI-TSD happens and holds for a period. In such case, it is possible that the coal or rock may have exploded, and preventive measures should be taken immediately for life safety [37], [40]. Literature [41] defined a variable coefficient as the division between the standard deviation and mean value of ERI-TSD over the same time period, and took the coefficient as the trend variation over the time period. Once the coefficient value reached the threshold given by practice, the dynamic disaster was thought broke out. Similarly, if the predicted trend value is greater than a threshold, a pre-warning will be activated.

Atmospheric aerosol particles, also known as particulate matter (PM), are microscopic solid or liquid matter suspended in Earth's atmosphere. Sources of PM can be natural or anthropogenic and impact the climate and human health. PM2.5 refers to particles with aerodynamic equivalent diameters less than or equal to 2.5 microns. A study [42] revealed that there was no safe level of particulates, and the smallest PM2.5 were particularly deadly, with a 36% increase in lung cancer per $10\mu g/m^3$ since they can penetrate deeper into the lungs. Thus, if we can obtain the trend of PM2.5 over a certain period ahead of time, more targeted environmental protection measures and more accurate safety recommendations can be implemented.

It is clear that the single value of the TSD on each sampled time point has no significance towards the judgement of disaster occurrences or performing measurements and cannot provide reliable information for safety forecast. Only the variation trend of these different types of TSD is extracted and forecasted, the safety status together with its changes can be well measured and predicted. Therefore, trend following for such problems are important and can be adopted to verify the performance of the proposed algorithm.

B. EXPERIMENTAL SETTINGS

The specific experimental environment is the Anaconda platform using Python, TensorFlow 1.4 and Keras 2.1.3. Two groups of experiments are conducted to validate the proposed algorithm and forecast the trend for the real ERI and PM2.5 TSDs. Total 41,696 data points of the these two TSDs are separately obtained and used for our algorithm.

The forecast horizon of the TSD trend is specified by the calculations on the trend metrics, i.e., the values of $A_{k,l}$, $r_{k,l}$ and $\sigma_{k,l}$. Taking the ERI for example, the variation trend of the ERI in 15, 30 or 60 minutes are often concerned for the short-, medium- or long-term safety forecast. Therefore, three different trend metrics on these three kinds of time window lengths are calculated for forecasting. As for the PM2.5-TSD,

trend in 6 hours and 24 hours are considered. Thus, the training data sets Tr can be obtained with equations 1, 2 and 4 by calculating $A_{k,l}$, $r_{k,l}$ and $\sigma_{k,l}$. The sizes and elements of Tr on the ERI with different time window lengths are listed in Table 1. The Tr of the PM2.5-TSD can be similarly specified.

TABLE 1. ***Tr*** **with different time window lengths.**

l	$A_{k,l}$	$r_{k,l}$	$\sigma_{k,l}$
15	$A_{1,15},\cdots,A_{2779,15}$	$r_{1,15},\cdots,r_{2779,15}$	$\sigma_{1,15},\cdots,\sigma_{2778,15}$
30	$A_{1,30},\cdots,A_{1389,30}$	$r_{1,30},\cdots,r_{1389,30}$	$\sigma_{1,30},\cdots,\sigma_{1388,30}$
60	$A_{1,60},\cdots,A_{649,60}$	$r_{1,60},\cdots,r_{649,60}$	$\sigma_{1,60},\cdots,\sigma_{648,60}$

Two groups of experiments are conducted here. First, we validate the feasibility of the aggregated PSO-GD method by comparing the prediction accuracy and time consumption between the PSO-GD algorithm and two state-of-the-art GD based methods, i.e., Adam and RMSprop, since these two metrics are mainly chased by the LSTM. Specifically, the LSTM network trained by the traditional PSO, Adam, RMSprop, PSO-Adam and PSO-RMSprop are used to perform the trend following TSD problems with $l = 30$ and compared. Second, we change the time window length l for trend following of the ERI-TSD and PM2.5-TSD to find a reliable safety forecast under different time spans.

Four indicators, i.e., Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percent Error (MAPE), and Max Error are shown in Eq. 14 to Eq. 17 are used to measure the accuracy of our algorithm in trend following.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(x_{k+t,l} - \hat{x}_{k+t,l})^2}{n}} \tag{14}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|x_{k+t,l} - \hat{x}_{k+t,l}\right| \tag{15}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x_{k+t,l} - \hat{x}_{k+t,l}}{x_{k+t,l}}\right| \tag{16}$$

$$Max\_Error = Max\left|x_{k+t,l} - \hat{x}_{k+t,l}\right| \tag{17}$$

where $x_{k+t,l}$ is the true trend following value, $\hat{x}_{k+t,l}$ is the predictive value, t is the sampling time, and n is the number of predicted values.

C. EXPERIMENT I. FEASIBILITY OF PSO-GD LSTM

We first conduct experiments to validate the feasibility of the proposed PSO-GD based LSTM for trend following. Specifically, $\sigma_{k,l}$ and $A_{k,l}$ are chosen, then the prediction accuracy and time consumption of the LSTMs trained with PSO-Adam, PSO-RMSprop, Adam, RMSprop and PSO are compared with our PSO-GD on the ERI-TSD and PM2.5 problems, respectively.

Take the trend following of ERI-TSD as example, the $Tr = \{x_{1,30}, x_{2,30}, \cdots, x_{1388,30}\}$ are used in the experiments, and the first 1,250 points are selected as Ts, i.e., $Ts = \{x_{1,30}, x_{2,30}, \cdots, x_{1250,30}\}$. The remaining points are used as the testing set. The first 20 points of Ts are selected to predict the 21st value [43]. Specifically, we take $X_1 = \{x_{1,30}, x_{2,30}, \cdots, x_{20,30}\}$, $X_2 = \{x_{2,30}, x_{3,30}, \cdots x_{21,30}\}, \cdots$, $X_{1368} = \{x_{1368,30}, x_{1369,30}, \cdots x_{1387,30}\}$ as the inputs to predict $Y_1^*, Y_2^*, \cdots Y_{1368}^*$, respectively. To obtain a higher accuracy, the true values of the predicted trend on $Y_1 = \{x_{21,30}\}$, $Y_2 = \{x_{22,30}\}, \cdots, Y_{1388} = \{x_{1388,30}\}$ are also taken as the inputs of LSTM. In summary, the inputs of the model are $f_{in} = \{(X_1, Y_1), (X_2, Y_2), \cdots, (X_{1368}, Y_{1368})\}$. And the rest experiments are carried out follow this example.

The optimization objective function of PSO is defined as Eq. 18:

$$f_{loss} = \sqrt{\frac{1}{t}\sum_{t=1}^{N}(Y_t - Y_t^*)^2}, N = \lfloor T/l \rfloor \tag{18}$$

The parameters of all the compared methods are given in Table 2. Among the methods, PSO-GD-LSTM is the abbreviation of LSTM trained with the proposed aggregated optimization method, while the GD-LSTM and PSO-LSTM are the LSTM respectively optimized by only GD and standard PSO.

TABLE 2. **Parameters of compared methods.**

Parameter \ Category	PSO-GD-LSTM	GD-LSTM	PSO-LSTM
Optimization methods	Adam/RMSprop	Adam/RMSprop	–*
Number of cell_units	5	5	–*
Learning rate	0.0001	0.0001	–*
Initial training times	400	750	–*
Retraining times	200	–*	–*
Dimension of the input data	1	1	–*
Dimension of the output data	1	1	–*
Population size (pN)	150	–*	150
Population dimension(dim)	146	–*	146
Learning factor and learning(c_1, c_2, w)	c_1 = c_2 =2,w =0.8	–*	c_1 = c_2 =2,w =0.8
PSO iteration times(N)	50	–*	500

–* Indicating corresponding method does not involve that parameter.

TABLE 3. **Experimental results of trend following on ERI-TSD with different methods.**

Methods \ Error indicators	Train Process		Test Process			
	RMSE	Time Loss(s)	RMSE	MAE	MAPE	Max_error
PSO	0.00114	759.3	0.0191	0.0356	0.2501	0.0063
Adam	0.00186	**339**	0.0086	0.0055	0.101	0.0039
RMSprop	0.00193	375.9	0.0087	0.0056	0.0993	**0.0033**
PSO-Adam	0.00082	461.9	**0.0071***	**0.0048**	**0.089***	0.0034
PSO-RMSprop	**0.00079**	457.3	0.0076	0.0052	0.093	0.0037
PSO	0.00345	744.5	0.0245	0.0143	0.1511	0.0079
Adam	0.00092	**278.7**	0.0097	0.0058	0.059	0.0023
RMSprop	0.00126	333	0.0099	0.0058	0.0663	0.0038
PSO-Adam	**0.00021**	514.7	**0.0094**	**0.0057**	**0.0571**	**0.0014***
PSO-RMSprop	**0.00021**	530.6	0.0095	**0.0057**	0.0583	0.0035

The experimental results of trend following are shown in Tables 3 and 4. All the experiments are executed for twelve times, and their average values are calculated as experimental results. The best values of the results are bolded. If the value is significantly better than the compared ones by T-testing, it will be marked with a star.

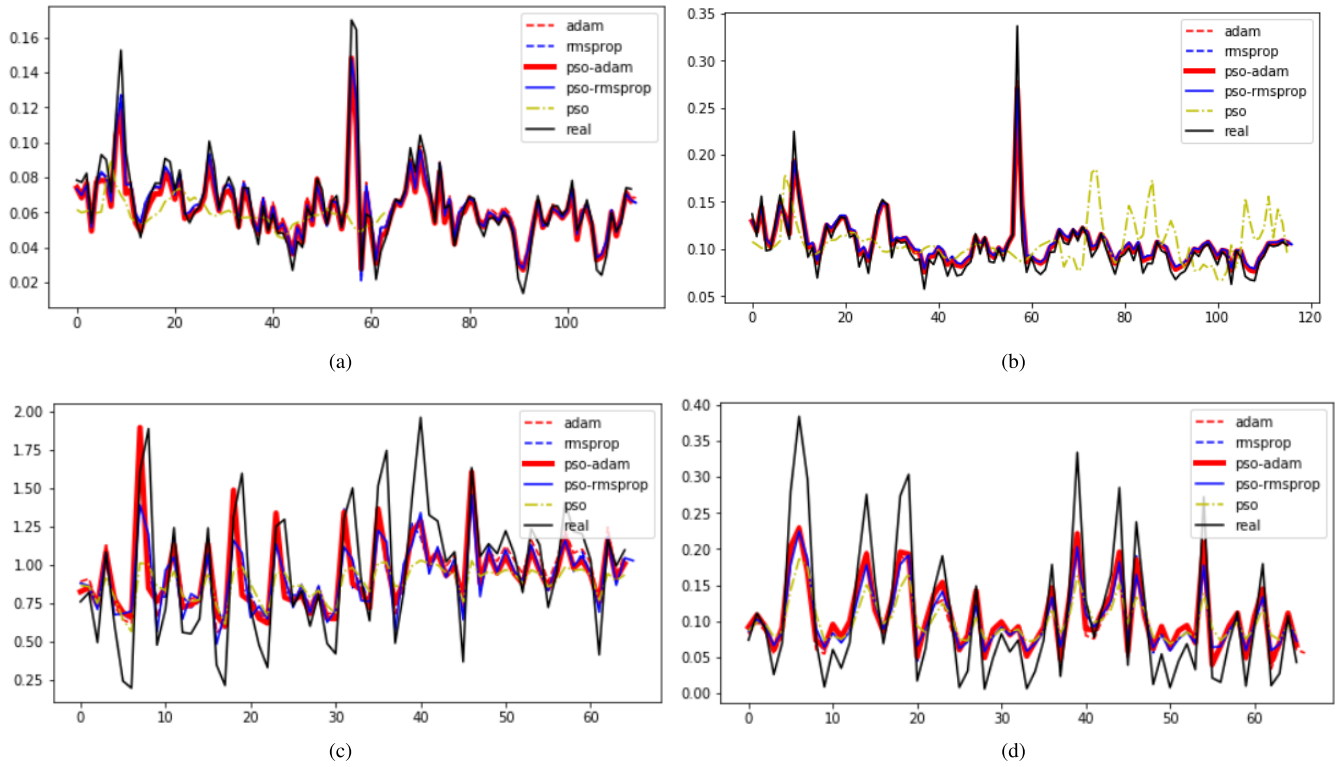


FIGURE 3. Trend-following results of the ERI-TSD (Row 1) and PM2.5-TSD (Row 2) data sets. (a) $A_{k,l=30}$. (b) $\sigma_{k,l=30}$. (c) $A_{k,l=30}$. (d) $\sigma_{k,l=30}$.

TABLE 4. Experimental results of trend following for the PM2.5-TSD with different methods.

Methods \ Error indicators	Train Process		Test Process			
	RMSE	Time Loss(s)	RMSE	MAE	MAPE	Max_error
PSO	0.0048	719.5	0.3754	0.3104	0.4272	0.2033
Adam	0.0043	**321**	0.2819	0.2059	0.3117	0.1207
RMSprop	0.0044	335.5	0.2795	0.1978	0.299	0.1279
PSO-Adam	**0.0042**	466.2	**0.245**	**0.1952**	0.3011	**0.0767**
PSO-RMSprop	**0.0042**	471.5	0.2541	0.1996	**0.2984**	0.1073
PSO	0.1390	700.9	0.0641	0.0742	1.433	0.0439
Adam	0.1237	**322**	0.049	0.0583	1.046	0.0511
RMSprop	0.1192	329	0.0493	0.0375	1.039	0.0309
PSO-Adam	**0.1169**	468.9	**0.0482**	**0.0371**	1.042	0.0549
PSO-RMSprop	0.1176	453.2	0.0486	0.0374	**1.038**	**0.0307**

The trend following process of the two datasets with different methods are given in forms of graph as shown in Figure 3.

From the above experimental results, the following conclusions can be made: (1) Significant differences in the MSE indicator exist between the PSO-Adam and the other methods, and the LSTM network trained by the PSO-Adam reaches the highest accuracy in the process of trend following for both ERI-TSD and PM2.5-TSD. (2) Compared with the GD-LSTM or PSO optimized LSTM, the LSTM trained by the PSO-GD (PSO-Adam, PSO-RMSprop) algorithm has a better performance in the training process and achieves smaller training errors, meanwhile generally performs better on accuracy in the test process. Although the PSO-GD is somewhat time consuming, the PSO has the highest time cost with the lowest accuracy. (3) The PSO-GD-LSTMs can accurately track the trends calculated with the average and standard deviation metrics. The indicators for the standard deviation trend following are the better between the two trend calculations. These results demonstrate that the PSO-GD-aggregated training method for LSTM is superior and feasible.

D. EXPERIMENT II. TREND FOLLOWING WITH DIFFERENT TIME WINDOW LENGTHS

In practical applications, we usually need to observe the trend of a TSD in different time periods, i.e., long-term or short-term trends. To this end, we can adjust the length of time window l to perform trend following in different time periods.

We carry out trend following of the ERI-TSD data sets with $l = 15$ (short-term) and 60 (long-term) minutes, and PM2.5-TSD with $l = 6$ (short-term) and 24 (long-term) hours. All of these are completed by the PSO-Adam-LSTM. The *Tr* set and corresponding parameters of the PSO-Adam-LSTM for $A_{k,l}$, $r_{k,l}$ or $\sigma_{k,l}$ under different time window lengths are given in Table 5. The results of the trend following are given in the form of graphs, and the error indicators are given in the tables.

The trend following process of the ERI-TSD are presented in Figure 4 and Table 6, and that of PM2.5-TSD are shown in Figure 5 and Table 7. From Figure 4, Figure 5, Table 6 and Table 7, the following observations can be reached: (1) The

TABLE 5. Parameters of PSO-Adam-LSTM for different trend following methods with different time window lengths.

Parameter \ Trend indicators	$A_{k,l}$				$r_{k,l}$				$\sigma_{k,l}$			
	ERI-TSD		PM2.5		ERI-TSD		PM2.5		ERI-TSD		PM2.5	
	l=15	l=60	l=12	l=24	l=15	l=60	l=12	l=24	l=15	l=60	l=12	l=24
Training_nums(Ts)	2715	630	3300	1650	2715	630	3300	1650	2715	630	3300	1650
Number of cell_units	5				5				5			
Learning rate	0.0001		0.0005		0.0001				0.0001		0.0005	
Initial training times	200		200		400		200		200		200	
Retraining times	100	150	200		200				150		200	
Population size(pN)	150				150				150			
c_1,c_2,w	c_1=c_2=2,w=0.8				c_1=c_2=2,w=0.8				c_1=c_2=2,w=0.8			
Iteration(max_iter)	50				50				50			

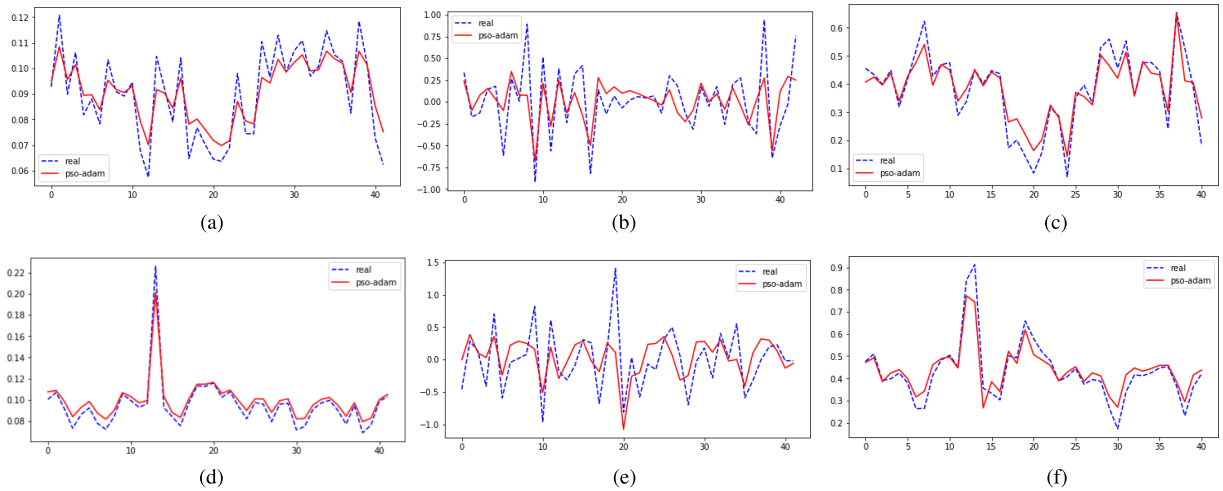


FIGURE 4. Trend following of ERI-TSD with different time window lengths. (a) $A_{k,l=15}$. (b) $r_{k,l=15}$. (c) $\sigma_{k,l=15}$. (d) $A_{k,l=60}$. (e) $r_{k,l=60}$. (f) $\sigma_{k,l=60}$.

TABLE 6. Average indicators of ERI-TSD trend following.

Trend type \ Error indicators	Train Process		Test Process			
	RMSE	Time Loss(s)	RMSE	MAE	MAPE	Max_error
$A_{k,l=15}$	**0.00079**	715	**0.0074**	**0.0061**	**0.0726**	**0.0128**
$r_{k,l=15}$	0.1227	**689**	0.2786	0.2098	1.4031	0.5106
$\sigma_{k,l=15}$	0.0144	702	0.0482	0.0362	0.1572	0.0981
$A_{k,l=60}$	0.00085	**271**	0.0486	0.0360	0.0971	0.0164
$r_{k,l=60}$	0.1371	275	0.3736	0.2988	3.2193	0.0399
$\sigma_{k,l=60}$	**0.00073**	277	**0.007**	**0.0056**	**0.0627**	**0.0019**

TABLE 7. Average indicators of PM2.5-TSD trend following.

Trend type \ Error indicators	Train Process		Test Process			
	RMSE	Time Loss(s)	RMSE	MAE	MAPE	Max_error
$A_{k,l=12}$	0.3719	817	0.3676	0.2698	0.7743	1.2281
$r_{k,l=12}$	0.7192	**782**	1.0656	0.8941	//	1.2539
$\sigma_{k,l=12}$	**0.0820**	817	**0.2581**	**0.1666**	**0.3105**	**0.2187**
$A_{k,l=24}$	0.1204	586	0.4186	0.3422	0.7307	0.2205
$r_{k,l=24}$	0.9383	**550**	1.365	1.2079	//	0.8983
$\sigma_{k,l=24}$	**0.1130**	588	**0.304**	**0.2182**	**0.3689**	**0.0504**

// Meaning indicators can't be calculated due to the existance of 0.

average indicator $A_{k,l=15}$ achieve lowest training error and performs the best for the ERI-TSD when the time window length is 15 minutes. The standard deviation method derived from the logarithmic, i.e., $\sigma_{k,l}$, however, is superior to all other indicators in the other cases with different time window length for both ERI-TSD and PM2.5-TSD. Accordingly, the standard deviation $\sigma_{k,l}$ is more reliable for trend following and safety forecast with the proposed PSO-Adam-LSTM. (2) For the trend of ERI-TSD, high fluctuations in testing samples can be found in the short-term trend following results, indicating some sudden changes occur; and from the long-term trends we can get these changes occur only on some points instead of keeping for a long period, there is no continuous rising tendency. All analysis implicates the status of the current coal mine may occur small rock mass disaster but has no obvious danger. (3)And the trend of PM2.5-TSD shows a relative high variation of concentration, this may be related to the temporal weather factor like the wind scale, wind direction, etc. But we should recognize sometimes the needed adjustment beyond the nature power, thus some efficient actions and timely life tips should be taken by the government when the PM2.5 concentration variation is relative high.

In summary, the PSO-Adam-LSTM model has an excellent performance for trend following with higher accuracy and can be used for trend following of a TSD with different

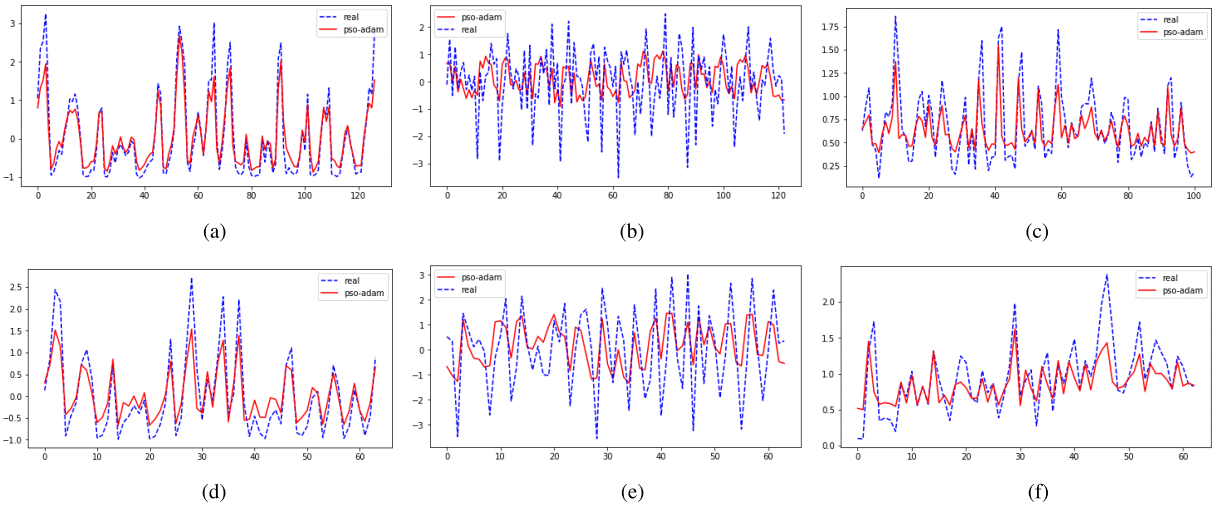


FIGURE 5. Trend following of PM2.5-TSD with different time window lengths. (a) $A_{k,l=12}$. (b) $r_{k,l=12}$. (c) $\sigma_{k,l=12}$. (d) $A_{k,l=24}$. (e) $r_{k,l=24}$. (f) $\sigma_{k,l=24}$.

time window lengths for creditable safety forecast. From the viewpoint of reliable and stable trend following, the standard deviation method derived from the logarithmic, i.e., $\sigma_{k,l}$, is recommended in practical applications.

VI. CONCLUSIONS

In industrial fields with time series data, trend extraction and forecasting are much more significant for work condition monitoring and environmental safety forecast, but few researches have been concerned about these problems. Consequently, we present our work on the trend following of TSD for safety forecast. The extraction of TSD is first addressed using three methods, i.e., the average values of a time span, logarithmic differences between the starting and ending points of a time span, and the standard deviation of the logarithmic values of the time span. With trend extraction, the trend following is converted into a novel TSD prediction. A PSO and GD aggregated optimization is developed to train a LSTM to perform the trend following of a TSD. We validate the feasibility of our algorithm by applying several PSO-GD based LSTMs to real problems, i.e., the trend following in the analysis of the electromagnetic radiation intensity TSD for the coal or rock dynamic damage safety forecast and the PM2.5-TSD for the environmental safety forecast. The experimental results demonstrate that PSO-GD methods can improve both the training effects and testing accuracy, the PSO-Adam-LSTM outperforms all the other compared methods and can get a most accurate prediction on the trend following of a TSD. Trend extraction with some latent learning methods and trend following with fewer parameters for stronger optimization-based models will be further studied in the future.

REFERENCES

[1] H. Xu, *Data Mining Methods and Applications*. Beijing, China: Tsinghua Univ. Press, 2014, pp. 12–13.
[2] E. Keogh and S. Kasetty, "On the need for time series data mining benchmarks: A survey and empirical demonstration," *Data Mining Knowl. Discovery*, vol. 7, no. 4, pp. 349–371, Oct. 2003. doi: 10.1023/A:1024988512476.
[3] H.-Y. Zeng and M. Yao, "Research on methodology of time serial data mining," *Comput. Eng. Des.*, vol. 25, no. 11, pp. 1999–2001, 2004.
[4] X. Zhu, S. Zhao, Y. Yang, H. Tang, Z. Wang, and J. Hao, "A real-time ensemble classification algorithm for time series data," in *Proc. IEEE Int. Conf. Agents*, Jul. 2017, pp. 145–150. doi: 10.1109/AGENTS.2017.8015322.
[5] A. Jha, S. Ray, B. Seaman, and I. S. Dhillon, "Clustering to forecast sparse time-series data," in *Proc. IEEE 31st Int. Conf. Data Eng.*, Apr. 2015, pp. 1388–1399. doi: 10.1109/ICDE.2015.7113385.
[6] S. Zhang, Y. Wang, M. Liu, and Z. Bao, "Data-based line trip fault prediction in power systems using LSTM networks and SVM," *IEEE Access*, vol. 6, pp. 7675–7686, 2018. doi: 10.1109/ACCESS.2017.2785763.
[7] J. Walker, R. Borgo, and M. W. Jones, "Timenotes: A study on effective chart visualization and interaction techniques for time-series data," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 1, pp. 549–558, Jan. 2016. doi: 10.1109/TVCG.2015.2467751.
[8] J.-B. Wu, E.-Y. Wang, X.-K. Ren and X.-R. Wang, "Rock burst early-warning for thick coal seam in deep mining based on Logistic regression," *Ind. Mining Autom.*, vol. 43, no. 6, pp. 42–46, Jun. 2017. doi: 10.13272/j.issn.1671-251x.2017.06.010.
[9] Q. Yu, Z.-Z. Guo, and R. Liang, "A prediction model of expressway electromechanical equipment life based on Markov chain," *J. Highway Transp. Res. Develop.*, vol. 35, no. 1, pp. 28–35, 2018. doi: 10.3969/j.issn.1002–0268.2018.01.005.
[10] S. Chatterjee, S. Sarkar, N. Dey, S. Sen, T. Goto, and N. C. Debnath, "Water quality prediction: Multi objective genetic algorithm coupled articial neural network based approach," in *Proc. IEEE 15th Int. Conf. Ind. Inform.*, Jul. 2017, pp. 963–968. doi: 10.1109/INDIN.2017.8104902.
[11] X. Liu, M. Ding, Y. Zhang, and N. Xu, "Dynamic economic dispatch for microgrids," *Proc. Chin. Soc. Elect. Eng.*, vol. 31, no. 31, pp. 77–84, 2011.
[12] T. Jarábek, P. Laurinec, and M. Lucká, "Energy load forecast using S2S deep neural networks with k-Shape clustering," in *Proc. IEEE 14th Int. Sci. Conf. Inform.*, Nov. 2017, pp. 140–145. doi: 10.1109/INFORMATICS.2017.8327236.
[13] Y. Wang, J. Zhou, K. Chen, Y. Wang, and L. Liu, "Water quality prediction method based on LSTM neural network," in *Proc. 12th Int. Conf. Intell. Syst. Knowl. Eng.*, Nov. 2017, pp. 1–5. doi: 10.1109/ISKE.2017.8258814.
[14] Y. Pan, Z. Mao, Q. Xiao, X. He, and Y. Zhang, "Discrete wavelet transform based data trend prediction for marine diesel engine," in *Proc. 6th Data Driven Control Learn. Syst.*, May 2017, pp. 782–787. doi: 10.1109/DDCLS.2017.8068173.
[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

[16] F. A. Gers and J. Schmidhuber, "Recurrent nets that time and count," in *Proc. IEEE-INNS-ENNS Int. Joint Conf. Neural Netw. (IJCNN), Neural Comput., New Challenges Perspect. New Millennium*, vol. 3, Jul. 2000, pp. 189–194. doi: 10.1109/IJCNN.2000.861302.

[17] F.-S. Tsai, Y.-M. Weng, C.-J. Ng, and C.-C. Lee, "Embedding stacked bottleneck vocal features in a LSTM architecture for automatic pain level classification during emergency triage," in *Proc. 7th Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Oct. 2017, pp. 313–318. doi: 10.1109/ACII.2017.8273618.

[18] Y. Wei, Z. Wang, M. Xu, and S. Qiao, "An LSTM method for predicting CU splitting in H. 264 to HEVC transcoding," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4. doi: 10.1109/VCIP.2017.8305079.

[19] D. S. Reddy and P. R. C. Prasad, "Prediction of vegetation dynamics using NDVI time series data and LSTM," *Model. Earth Syst. Environ.*, vol. 4, no. 1, pp. 409–419, 2018. doi: 10.1007/s40808-018-0431-3.

[20] W. Wu, K. Chen, Y. Qiao, and Z. Lu, "Probabilistic short-term wind power forecasting based on deep neural networks," in *Proc. Int. Conf. Probabilistic Methods Appl. Power Syst.*, Oct. 2016, pp. 1–8. doi: 10.1109/PMAPS.2016.7764155.

[21] P. Wang, Q. Song, H. Han, and J. Cheng, "Sequentially supervised long short-term memory for gesture recognition," *Cogn. Comput.*, vol. 8, no. 5, pp. 982–991, Oct. 2016. doi: 10.1007/s12559-016-9388-6.

[22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: Massachusetts Institute of Technology, 2016, pp. 230–289.

[23] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw. (ICNN)*, vol. 4, 1995, pp. 1942–1948. doi: 10.1109/ICNN.1995.488968.

[24] D. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston, MA, USA: Addison-Wesley, 1989, pp. 1–25. doi: 10.1111/j.1365-2486.2009.02080.x.

[25] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: Optimization by a colony of cooperating agents," *IEEE Trans. Syst., Man, Cybern., B (Cybern.)*, vol. 26, no. 1, pp. 29–41, Feb. 1996. doi: 10.1109/3477.484436.

[26] R. Hassan, B. Cohanim, and O. de Weck, "A comparison of particle swarm optimization and the genetic algorithm," in *Proc. 46th AIAA/ASME/ASCE/AHS/ASC Struct., Structural Dyn. Mater. Conf.*, 2004, pp. 1–13. doi: 10.2514/6.2005-1897.

[27] A. M. Adrian, A. Utamima, and K.-J. Wang, "A comparative study of GA, PSO and ACO for solving construction site layout optimization," *KSCE J. Civil Eng.*, vol. 19, no. 3, pp. 520–527, 2015. doi: 10.1007/s12205-013-1467-6.

[28] X. Cai, N. Zhang, G. Venayagamoorthy, and D. C. Wunsch, "Time series prediction with recurrent neural networks using a hybrid pso-ea algorithm," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, vol. 2, Jul. 2004, pp. 1647–1652. doi: 10.1109/IJCNN.2004.1380208.

[29] N. Chouikhi, B. Ammar, N. Rokbani, and A. M. Alimi, "PSO-based analysis of echo state network parameters for time series forecasting," *Appl. Soft Comput.*, vol. 55, pp. 211–225, Jun. 2017. doi: 10.1016/j.asoc.2017.01.049.

[30] N. Zeng, Z. Wang, H. Zhang, and F. E. Alsaadi, "A novel switching delayed PSO algorithm for estimating unknown parameters of lateral flow immunoassay," *Cogn. Comput.*, vol. 8, no. 2, pp. 143–152, 2016. doi: 10.1007/s12559-016-9396-6.

[31] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, Jan. 2010.

[32] A. Rakitianskaia and A. Engelbrecht, "Training high-dimensional neural networks with cooperative particle swarm optimiser," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2014, pp. 4011–4018. doi: 10.1109/IJCNN.2014.6889933.

[33] H. Moalla, W. Elloumi, and A. M. Alimi, "H-PSO-LSTM: Hybrid LSTM trained by PSO for online handwriter identification," in *Proc. Int. Conf. Neural Inf. Process.*, 2017, pp. 41–50.

[34] P. Christoffersen, *Element of Financial Risk Management*. New York, NY, USA: Academic, 2012.

[35] K. Gref, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2222–2232, Oct. 2017. doi: 10.1109/TNNLS.2016.2582924.

[36] B. Nakisa, M. Rastgoo, A. Rakotonirainy, F. Maire, and V. Chandran, "Long short term memory hyperparameter optimization for a neural network based emotion recognition framework," *IEEE Access*, vol. 6, pp. 49325–49338, 2018. doi: 10.1109/ACCESS.2018.2868361.

[37] E.-Y. Wang, X.-F. Liu, Z.-H. Li, and X.-Q. He, "Application of electromagnetic radiation technology in monitoring and warning on coal and rock dynamic disasters," *J. Liaoning Tech. Univ.*, vol. 31, no. 5, pp. 642–645, 2012.

[38] G.-J. Liu, C.-P. Lu, H.-Y. Wang, P.-F. Liu, and Y. Liu, "Warning method of coal bursting failure danger by electromagnetic radiation," *Shock Vib.*, vol. 2015, no. 6, pp. 1–9, 2015. doi: 10.1155/2015/583862.

[39] E. Wang, X. He, X. Liu, and W. Xu, "Comprehensive monitoring technique based on electromagnetic radiation and its applications to mine pressure," *Saf. Sci.*, vol. 50, no. 4, pp. 885–893, 2012. doi: 10.1016/j.ssci.2011.08.013.

[40] E.-Y. Wang, Z.-H. Li, X.-Q. He, and L. Chen, "Application and prewarning technology of coal and gas outburst by electromagnetic radiation," *Coal Sci. Technol.*, vol. 42, no. 6, pp. 53–57, 2014.

[41] L. Qiu *et al.*, "Characteristics and precursor information of electromagnetic signals of mining-induced coal and gas outburst," *J. Loss Prevention Process Industries*, vol. 54, pp. 206–215, Jul. 2018. doi: 10.1016/j.jlp.2018.04.004.

[42] O. Raaschou-Nielsen *et al.*, "Air pollution and lung cancer incidence in 17 European cohorts: Prospective analyses from the European study of cohorts for air pollution effects (ESCAPE)," *Lancet Oncol.*, vol. 14, no. 9, pp. 813–822, 2013.

[43] T. Zhou, S. Gao, J. Wang, C. Chu, Y. Todo, and Z. Tang, "Financial time series prediction using a dendritic neuron model," *Knowl.-Based Syst.*, vol. 105, pp. 214–224, Aug. 2016. doi: 10.1016/j.knosys.2016.05.031.

YAO HU is currently pursuing the master's degree in control science and engineering with the School of Information and Control Engineering, China University of Mining and Technology, China.

His current research interests include machine learning and multi-objective optimization.

XIAOYAN SUN received the Ph.D. degree in control theory and control engineering from the China University of Mining and Technology, in 2009.

She is currently a Professor with the School of Information and Control Engineering, China University of Mining and Technology. Her research interests include interactive evolutionary computation, big data, and intelligence optimization. She is an IEEE member.

XIN NIE is currently pursuing the master's degree in control engineering with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China.

His research interests include data mining, machine learning, and personalized recommendation.

YUZHU LI is currently pursuing the master's degree in control engineering with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China.

His research interests include machine learning, data mining, and fault diagnosis.

LIAN LIU is currently pursuing the master's degree in control engineering with the School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, China.

Her research interests include data reduction and machine learning.