Sales Demand Prediction Model of Gated Recurrent Unit Neural Network Based on Improved Shape Distance Loss Function

Haichuan Lou
Zhejiang SUPCON Technology Co., Ltd, Hangzhou, China
louhaichuan@supcon.com

Zhiwei Zhang*
Zhejiang SUPCON Technology Co., Ltd, Hangzhou, China
zhangzhiwei1@supcon.com

Baihui Zha
Zhejiang SUPCON Technology Co., Ltd, Hangzhou, China
zhabaihui@supcon.com

ABSTRACT

Under the background of diversification and refinement of chemical products, product demand prediction is playing a guiding role in production planning. In this paper, a new sales demand prediction model based on improved shape distance Loss function of Gated Recurrent Unit Neural Network (ISD_GRUNN) is proposed for the long-term prediction of the sales quantity of chemical products. The improved shape distance is determined by the change trend, amplitude and distance between the two points. Compared with MSE which only considers the difference between the corresponding time point sequence values as the loss function, the change trend and range of the time series will be taken into account in the improved shape distance as the loss function. The experimental results show that the improved shape distance as the loss function can be better used for sales long-term prediction.

CCS CONCEPTS

• Computing methodologies; **• Machine learning**; **• Machine learning approaches**; **• Neural networks**;

KEYWORDS

Improved shape distance, Loss function, Demand prediction, GRUNN

ACM Reference Format:

Haichuan Lou, Zhiwei Zhang, and Baihui Zha. 2021. Sales Demand Prediction Model of Gated Recurrent Unit Neural Network Based on Improved Shape Distance Loss Function. In *The 5th International Conference on Computer Science and Application Engineering (CSAE 2021), October 19–21, 2021, Sanya, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3487075.3487139

1 INTRODUCTION

The globalization of modern chemical business, demand differentiation and product diversification have greatly increased the difficulty of supply chain planning, which cannot rely solely on judgment. It is very important to predict the short-term or long-term future demand of enterprises through historical product data and other information, that can predict the market demand of products as accurately as possible, and provide better decision-making management for enterprises.

There are more than 200 methods proposed for demand prediction at home and abroad, of which more than 20 are widely used in theoretical research and practice [1]. Prediction methods can be mainly divided into qualitative forecasting [2], quantitative forecasting [3-6] and combined forecasting [7, 8]. Using neural network to realize demand forecasting is a common method in quantitative forecasting. Tao Yongcai et al. [4] Optimized BP neural network model through improved particle swarm optimization algorithm to predict the demand of spare parts in the spare parts supply model. Cao Dandan et al. [5] used long short term memory (LSTM) neural network model to predict the short-term demand of shared bicycles. Li Qiong et al. [6] predicted the demand of manufacturing products through GPU-BP combined neural network. Qiao Junfei et al. [9] proposed an improved cerebellar model neural network to solve the time series prediction problem of nonlinear dynamic systems. Li Minjie et al. [10] simulated and predicted the logistics demand of China's aquatic products cold chain from 2007 to 2016 by using RBF neural network model. Lin Youfang et al. [11] proposed a deep spatiotemporal convolution neural network model for civil aviation demand forecasting.

However, when using neural network for prediction, the above scholars use MSE (mean square error) or RMSE (root mean square error), in which the predicted time series value is close to the real time series value. In the product demand prediction, it is necessary to forecast the demand for a period of time in the future. Therefore, this paper introduces the concept of time series similarity into the loss function. The common algorithms for calculating curve similarity include DTW (dynamic time warping) [12] and SD (shape distance) [13]. DTW mainly uses the idea of dynamic programming to solve the similarity of unequal time series.SD mainly considers the difference of fluctuation trend and fluctuation degree before the two sequences, This paper hopes that the long-term prediction model can better predict the trend of future series, so the improved shape distance is used as the loss function. That is, in addition to the absolute distance between the forecast and the real series, it is also expected that the change trend between the forecast and the real series is more consistent. Therefore, in this paper the loss function is improved that not only consider the distance between sequences , but also the consistent of change trend and the change range of sequences. The improved loss function comprehensively considers these three and adds corresponding weights.

*Corresponding author: zhangzhiwei1@supcon.com

CSAE 2021, October 19–21, 2021, Sanya, China

ACM ISBN 978-1-4503-8985-3/21/10…$15.00
https://doi.org/10.1145/3487075.3487139

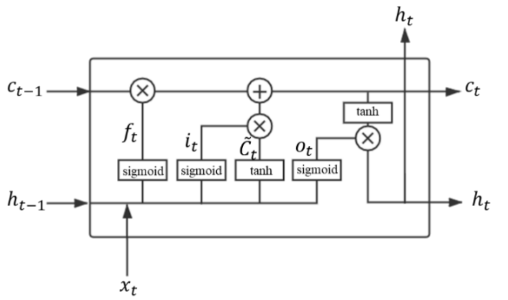

Figure 1: Unit Structure of Long Short Term Memory Neural Network.

2 CONVENTIONAL DEMAND PREDICTION MODEL

2.1 Autoregressive Moving Average Model

Autoregressive integrated moving average model (ARIMA) is a univariate time series prediction model. Its essence is to fit the stable series after difference with ARMA model. ARIMA (P, D, q) model structure is described as follows:

$$\begin{cases} \Phi(B)\nabla^{\mathrm{d}}x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s) = 0, s \neq t \\ E(x_s\varepsilon_t) = 0, \forall s < t \end{cases} \tag{1}$$

Where,$\nabla^d = (1-B)^d$,$B$is the delay operator, $\Phi(B) = 1 - \phi_1 B - \cdot s - \phi_{\mathrm{p}} B^p$is the autoregressive coefficient polynomial, and $\Theta(B) = 1 - \theta_1 B - \cdot s - \theta_{\mathrm{q}} B^q$is the moving smoothing coefficient polynomial.

2.2 Long Short Term Memory Neural Network

In order to solve the long-term dependence problem in the circular neural network, Hochreiter et al. [14] proposed a long short term memory (LSTM), which structure is shown in Figure 1

The forgetting gate mainly determines how much information to discard from storage unit C, and is realized by sigmoid function, the formula is:

$$f_t = \sigma(W_f x_t + U_f h_{t-1}) \tag{2}$$

The storage unit of the current time C_t is obtained by adding the result of multiplication of the storage unit and forgetting gate at the previous time C_{t-1} and the result of multiplication of hidden layer unit $\widetilde{C_t}$ and input gate. The formula is:

$$\begin{aligned} i_t &= \sigma(W_i x_t + U_i h_{t-1}) \\ \widetilde{C_t} &= \tanh(W_C x_t + U_C h_{t-1}) \\ C_t &= f_t * C_{t-1} + i_t * \widetilde{C_t} \end{aligned} \tag{3}$$

The storage unit C_t is multiplied by the output gate after the activation function tanh, and the implicit output is obtained. The calculation formula is as follows:

$$\begin{aligned} \mathrm{o}_t &= \sigma(W_o x_t + U_o h_{t-1}) \\ h_t &= o_t * \tanh(C_t) \end{aligned} \tag{4}$$

2.3 Gated Recurrent Unit Neural Network

Cho et al. [15] proposed the gated recurrent unit (GRU) neural network in 2014. Compared with LSTM, it has fewer parameters and faster calculation speed, but also has the function of LSTM to control the forgetting of old information and add new information. The hidden layer output at the current time in GRU $h_t$is the weighted sum of the hidden layer output at the previous time $h_{t-1}$and the newly generated hidden layer output $\widetilde{h_t}$ The formula is:

$$\begin{aligned} z_t &= \sigma(W_z x_t + U_z h_{t-1}) \\ r_t &= \sigma(W_r x_t + U_r h_{t-1}) \\ \widetilde{h_t} &= \tanh(W x_t + r_t * U h_{t-1}) \\ h_t &= (1 - z_t) * \widetilde{h_t} + z_t * h_{t-1} \end{aligned} \tag{5}$$

In addition, the improved models based on LSTM and GRU, such as Bi LSTM and Bi GRU, calculate the loss value and update the weight according to the loss function when using gradient descent to train the neural network. The improved shape distance loss function this paper proposed calculates the absolute distance, variation trend and variation range of the real sequence and the predicted sequence. Compared with the absolute distance, it can better fit the time series.

3 IMPROVED SHAPE DISTANCE DEMAND PREDICTION MODEL

In the past, MSE or RMSE was used as the loss function when using neural network to establish demand prediction model. In this paper, a improved shape distance between the predicted and the real time series value was used as the loss function.

Euclidean distance is used to analyze the difference between two time series. This method has the following disadvantages [13]: (1) the shape difference cannot be verified. (2) Unable to effectively identify the similarity of trend dynamic changes between two time series. (3) It can not reflect the difference between different analysis frequencies, so Dong et al. [13] proposed shape distance to measure the similarity of time series. When calculating the shape distance, segment the two time series at the same segmentation point firstly. According to the positive and negative conditions of the slope of each segment after segmentation, it is marked as 1, - 1 or 0, that is, the slope is positive, set as 1, negative as - 1 and unchanged as 0. After segmentation, the two time seriesS_1,S_2 can be expressed as,:

$$S_1 = \{(\mathrm{m}_{11},\mathrm{t}_1), (\mathrm{m}_{12},\mathrm{t}_2), \cdot s, (\mathrm{m}_{1n},\mathrm{t}_n)\} \tag{6}$$

$$S_2 = \{(\mathrm{m}_{21},\mathrm{t}_1), (\mathrm{m}_{22},\mathrm{t}_2), \cdot s, (\mathrm{m}_{2n},\mathrm{t}_n)\} \tag{7}$$

Where $t_j$represents the $\mathrm{j}th$ time interval of sequence $S_1$and $S_2$and m_{ij}represents the change trend of the jth time period of the sequence$i$. The division example of sequence $S_1$and $S_2$is shown in Figure 2, which can be expressed as:

$$S_1 = \{(1,\mathrm{t}_1), (-1,\mathrm{t}_2), (1,\mathrm{t}_3)\} \tag{8}$$

$$S_2 = \{(1,\mathrm{t}_1), (1,\mathrm{t}_2), (-1,\mathrm{t}_3)\} \tag{9}$$

LetA_1,A_2be the corresponding amplitude change sequenceS_1,S_2, respectively, which can be expressed as:

$$A_1 = \{(\Delta\mathrm{y}_{11},\mathrm{t}_1), (\Delta\mathrm{y}_{12},\mathrm{t}_2), \cdot s, (\Delta\mathrm{y}_{1n},\mathrm{t}_n)\} \tag{10}$$

$$A_2 = \{(\Delta\mathrm{y}_{21},\mathrm{t}_1), (\Delta\mathrm{y}_{22},\mathrm{t}_2), \cdot s, (\Delta\mathrm{y}_{2n},\mathrm{t}_n)\} \tag{11}$$

Where$\Delta\mathrm{y}_{ij}$is the difference between the sequence values $S_i$corresponding to the two ends of the segment$j$, then the shape

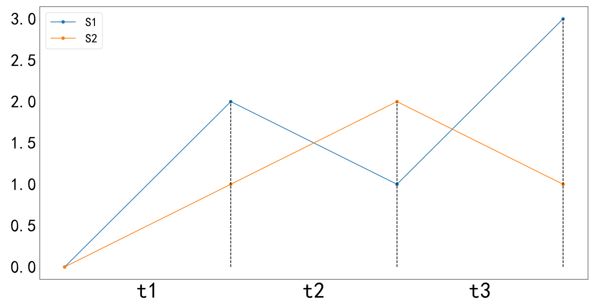

Figure 2: The Example of Sequence.

distance between the two time series is defined as:

$$D(S_1, S_2) = \sum_{j=1}^{n} t_{wj} \bullet \left|m_{1j} - m_{2j}\right| \bullet \left|A_{1j} - A_{2j}\right| \tag{12}$$

Where, $t_{w_j} = t_j/t_n$, t_{wj} represents the time length of the $j$mode and $t_n$represents the total time length, A_{ij}=Δy_{ij}.

The original shape distance is mainly determined by the product sum of the variation trend difference $(m_{1j} - m_{2j})$ and variation range difference $(A_{1j} - A_{2j})$ of each period of the two time series. If the variation trend difference of each period of the two time series is 0, that is, the variation trend of each period is the same, which can not well measure the distance between the two series. Therefore, when the original shape distance is used as the loss function, the model weight can not be updated well. The improved shape distance is mainly determined by the sum of variation trend difference $(m_{1j} - m_{2j})$ and variation range difference $(A_{1j} - A_{2j})$ and the value difference between the two sequences. The absolute distance, variation trend and variation range between the two sequences are considered, and the weight $\alpha$is added to the variation trend and variation range to control its impact on the distance between the two sequences, that taking the improved shape distance as the loss function. It can better measure the distance between the predicted sequence and the real sequence, so as to update the model weight better. When calculating the loss function, $S_1$and S_2represent the predicted sequence $\widehat{Y}$ and the real sequence Y respectively. Therefore, the common segment point of the two sequences is the time point j. The improved shape distance loss function is described as follows:

$$\begin{aligned} D\left(Y, \widehat{Y}\right) = \alpha \cdot \sum_{j=2}^{n} \left[\left(m_j - \widehat{m}_j\right)^2 + \left(A_j - \widehat{A}_j\right)^2\right] \\ + \sum_{j=1}^{n} \left(y_j - \widehat{y}_j\right)^2 \end{aligned} \tag{13}$$

Where, y_j represents the value of the real sequence at the time point j, $\widehat{y}_j$ represents the value of the prediction sequence at the time point j, $A_j$represents the difference of $y_j$and y_{j-1}, and $\widehat{A}_j$ represents the difference of $\widehat{y}_j$ and $\widehat{y}_{j-1}$,m_j is described in the following formula, $\widehat{m}_j$ is the same.

$$\mathrm{m}_j = \begin{cases} -1 y_j - y_{j-1} < 0 \\ 0 y_j - y_{j-1} = 0 \\ 1 y_j - y_{j-1} > 0 \end{cases} \tag{14}$$

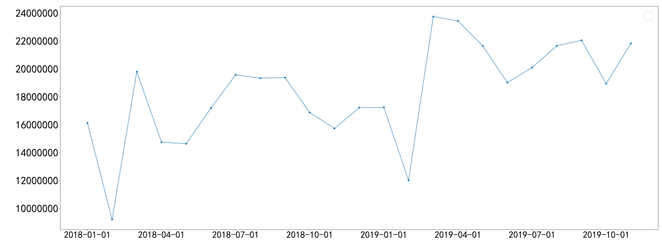

Figure 3: Total Sales Trend.

4 CASE STUDY

4.1 Dataset and Data Processing

The data used in this paper are the sales quantity of 23 products of a chemical plant from January 2018 to November 2019, including nutrition, Raw materials, spices and new materials. Sum these sales quantity data to calculate the total quantity from January 2018 to November 2019. The trend chart is shown in Figure 3. There uses the total quantity to build the demand prediction model.

The characteristics of product, market and time are selected. In the product category, the proportion of nutrition, raw material, spice, new material, total logarithm of nutrition price, total logarithm of API price, total logarithm of spice price and total logarithm of new material price are considered. In the market category, the closing price of New York oil and the rise and fall of the closing price of New York oil are also considered. The characteristics of the month and quarter are considered in the time category. Feature codes and meanings are shown in Table 1. And continuous features are Z-score standardized.

4.2 Data Processing

Because the feature dimension is larger than the number of data, this paper constructs cart regression tree to filter features. The tree node division method is MSE, the maximum depth of the tree is 5, and the minimum number of samples of leaf nodes is 2. Other parameters are default. The division characteristics and their scores are shown in Table 2

ARIMA only needs to train the total quantity from January 2018 to August 2019, and use the trained model to predict the total quantity from September to November 2019. When establishing the long-term prediction model using LSTM and GRUNN, the data should be reconstructed. Here, the window size is considered to 3, that is, the prediction of the next three days is predicted using the data of the first three days, and the sliding step is set to 1. Because the value of the dependent variable is too large, the model should not converge during training, so the logarithm of the value of the next three days is taken minus 16.

4.3 Experimental Results and Analysis

Take the data from January 2018 to August 2019 as the training set and the data from September 2019 to November 2019 as the test set. ARIMA, LSTM, GRUNN and GRUNN (ISD_GRUNN) based on the improved shape distance are selected in the prediction model. The selection of super parameters is determined by grid search and manual parameter adjustment.

Table 1: Feature Name and Meaning

Feature category	Feature name	Characteristic meaning
product	Proportion of nutrition sales	Proportion of nutrition sales in total sales
	Proportion of raw material sales	Proportion of raw material sales in total sales
	Proportion of spice sales	Proportion of spice sales in total sales
	Sales proportion of new materials	Proportion of new material sales in total sales
	Logarithm of nutrition price	Logarithm of total price of nutrition
	Logarithm of raw material price	Logarithm of total price of raw material price
	Logarithm of spice price	Logarithm of the total price of spices
	Logarithm of new material price	Logarithm of total price of new materials
market	New York oil closing price	New York oil closing price
	Oil prices rose and fell in New York	Oil prices rose and fell in New York
time	month	Month period
	quarter	Quarter period

Table 2: Selected Characteristics and Corresponding Scores

Feature name	fraction
January	0.65236
Logarithm of spice price	0.21839
Proportion of raw material sales	0.06093
Proportion of nutrition sales	0.02464
Second quarter	0.02411
Logarithm of nutrition price	0.01958

Table 3: Prediction Model Results (Prediction Period for Three Months)

Model	RMSE
ARIMA	1693143.22
LSTM	1488810.96
GRU	1388734.57
ISD_GRU	1103027.35

It can be seen from table 3 that the RMSE of multivariable prediction model (LSTM and GRUNN) is better than that of univariate prediction model (ARIMA), the result of GRUNN is better than LSTM, and the GRUNN model with improved loss function is better than that based on MSE.

The trend chart of sequence value and real value fitted by each prediction model is shown in Figure 4. From January 2018 to August 2019 is the fitted value of the training set of the prediction model, and from September to November 2019 is the predicted value of the test set, so as to build a model for predicting the next three months. It can be seen from the figure that the change trend of the predicted value of ARIMA, LSTM, GRUNN and ISD_GRUNN is consistent with that of the real value, which decreases first and then increases, but the predicted value of ISD_GRUNN is closer to the real value than other models.

In order to study the prediction of each model in a longer prediction period, the data from January 2018 to July 2019 are used as the training set and the data from August to November 2019 are used as the test set to build a model for predicting the next four months.

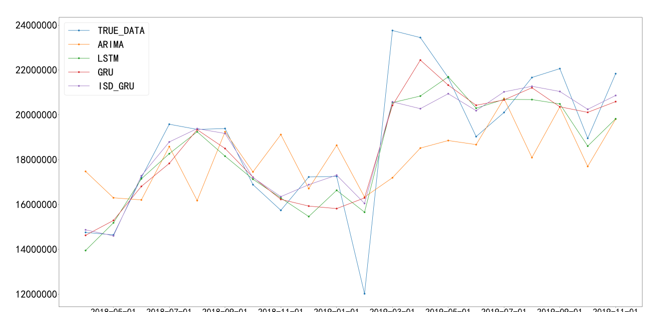

Figure 4: Predicted Value of Each Prediction Model (Predicted for Three Months).

The results of each model are shown in Table 4, which is consistent with the results of each model at 3 months. The multivariable prediction model is better than the univariate prediction model, the result of GRUNN is better than LSTM, and the GRUNN model

Table 4: Prediction Model Results (Prediction Period for Four Months)

Model	RMSE
ARIMA	3234557.99
LSTM	2016872.43
GRU	1613077.01
SD_GRU	409533.91

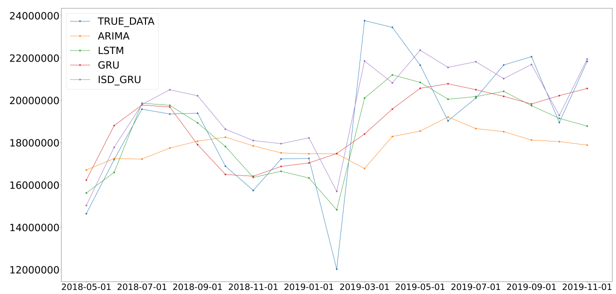

Figure 5: Predicted Value of Prediction Models (Prediction Period for Four Months).

that improving the loss function (ISD_GRUNN) is better than the GRUNN based on MSE.

The trend chart of sequence value and real value fitted by each prediction model at 4 months is shown in Figure 5. The change trend of real sequence is first rising, then falling and then rising. However, the change trend of predicted value of other prediction models is inconsistent with the change trend of real value except ISD_GRUNN. Besides, the bias between the predicted value and the real value of ISD_GRUNN is closer than that of other models.

From the experimental results, it can be seen that, taking the improved shape distance as the loss function, the absolute distance between the real sequence and the test sequence, whether the change trend is consistent and the change range are considered. Compared with the mean square error, only the absolute distance is considered, the improved shape distance can learn the change trend and change range of the real sequence, It makes the prediction sequence more close to the real sequence. In addition, because the fluctuation of the series in period for three months is small, taking the improved shape distance and mean square error as the loss function, the prediction series can learn the fluctuation trend of the series, while the trend fluctuation increases in period for four months. The improved shape distance can learn the fluctuation trend of the real series better than the mean square error. Moreover, this experiment does not use the improved loss function to model the improved LSTM and GRU, the proposed improved loss function not only can improve these LSTM or GRU, but also makes the gradient descent algorithm better train the model weight.

5 CONCLUSION

Aiming at the long-term prediction of chemical product sales quantity, this paper proposes a sales demand prediction model of GRUNN based on improved shape distance loss function. The improved shape distance is determined by the change trend, change amplitude and distance between points of the two time series. Compared with using MSE as the loss function that only considering the bias between the series values of the corresponding time points, the improved shape distance loss function will take into account the change trend and change range of the time series. The experimental results show that the model based on the improved shape distance (ISD_GRUNN) as the loss function is better than that based on MSE, but the change trend of the two results is the same as the real results. When predicting the series value in the next four months, the ISD_GRUNN model based on the improved shape distance is not only better than that based on MSE as the loss function , but also the change trend of the predicted value is consistent with the real value. It indicates that the ISD_GRUNN model with improved shape distance as the loss function can better predict the long-term.At present, the demand prediction model is established for the total sales quantity of 23 products, and the prediction granularity is relatively large, but it proves the effectiveness of the improved loss function. In the future, the improved loss function will be used to build the demand prediction model for each product, and the prediction effect will be observed for a long time.

ACKNOWLEDGMENTS

This work is supported by National Key R&D Program of China (Integrated demonstration of intelligent manufacturing plant operating system in process industry, 2020YFB1711205).

REFERENCES

[1] Bin Lei, Rui Zhao, Weihong Wang, etc (2018). Review of inventory demand forecasting methods. Statistics and decision making, 37(3), 58-62.
[2] Mingtao Wang (1997). Accuracy and effectiveness analysis of nal of qualitative prediction methods. Prediction, 2, 45-47.
[3] Xiaomeng Luo, Jianbin Li, Peng Hu (2014). E-commerce inventory optimization strategy based on time series prediction. Systems engineering, 32(6), 91-98.
[4] Yongcai Tao, Chen Yang, Jianhong Ma, etc (2021). Research on a spare parts supply model combined with IPSO-BP neural network. Small microcomputer system, 1-9.
[5] Dandan Cao, Shurui fan, Yan Zhang, etc (2020). Short term demand forecasting of shared vehicle based on long short-term memory neural network model. Science, technology and engineering, 20(20), 8344-8349.
[6] Qiong Li, Jia Chen, Bin Xu (2021). Research on manufacturing product demand forecasting under the condition of multi product mutual restriction. Computer application and software, 38(3), 59-69.
[7] Fengjiao Cheng, Ruoying Sun (2018). Combined inventory demand forecasting method based on demand characteristics. Journal of Beijing University of information technology (NATURAL SCIENCE EDITION), 33(3), 61-65.
[8] Xiaoping Wang, Fei Yan (2018). Study on influencing factors and prediction model of cold chain logistics demand for agricultural products in Beijing Tianjin and Hebei. Fujian Jouragriculture, 33(8), 870-878.
[9] Junfei Qiao, Jingjiao Dong, Wenjing Li (2021). Improved cerebellar model neural network and its application in time series prediction. Journal of Beijing University of technology, 47(6), 598-606.
[10] Minjie Li, Jian Wang (2020). Research on logistics demand prediction of aquatic products cold chain based on RBF neural network. China's agricultural resources

and zoning, 41(6), 100-109.
[11] Youfang Lin, Youyin Kang, Huaiyu Wan, etc (2018). Civil aviation demand forecasting based on deep spatiotemporal convolution network. Journal of Beijing Jiaotong University, 42(2), 1-8.
[12] Sakoe H (1990). Two-level dp-matching–a dynamic programming-based pattern matching algorithm for connected word recognition. Readings in Speech Recognition, 27(6), 180-187.
[13] Xiaoli Dong, Chengkui Gu, Zhengou Wang (2006). Research on Shape-Based Time Series Similarity Measure. Machine Learning and Cybernetics, 2006 International Conference on.
[14] Hochreiter S, Schmidhuber J (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780.
[15] Cho K, Merrienboer B V, Gulcehre C, etc (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. Computer Science.