A Trend Forecast of Import and Export Trade Total Volume based on LSTM

Qinglin Qu\textsuperscript{1}, Zhao Li\textsuperscript{2,3,4,*}, Juanjuan Tang\textsuperscript{4}, Shiwei Wu\textsuperscript{4} and Ruishuang Wang\textsuperscript{4}

\textsuperscript{1}Shandong Business Information Center
\textsuperscript{2}Qilu University of Technology (Shandong Academy of Sciences), Shandong Computer Science Center (National Supercomputer Center in Jinan)
\textsuperscript{3}Shandong Provincial Key Laboratory of Computer Networks
\textsuperscript{4}Big Data Engineering Technology Research Center of E-Government, Shandong

\textsuperscript{*}Corresponding author’s e-mail address: liz@sdas.org

Abstract. The monthly import and export data are usually with the challenging characteristics including large scale, nonlinear and hard to fit, leading to their development trends can be hardly predicted. To tackle this problem, we propose to leverage the LSTM-based recurrent neural network to forecast the development trend in this paper. We demonstrate the effectiveness of our approach based on the monthly import and export data of Shandong Province from January 2001 to June 2018. In particular, we achieve the MSE score of 124.39, which outperforms the traditional time series model less than 12.2%.

1. Introduction
Import and export trade reflects the opening up of a country (region) to a certain extent. At present, the development of foreign trade is in the key stage of accelerating the pace of structural adjustment and the transformation between the new and old energy in our province. Generally, the environment of development is beneficial to the import and export trade. However, the trade frictions have been obvious recently. Influenced by the thought of "anti-globalization" in western countries, the trend of protectionism is on the rise and still spreading. More and more countries abused the emergency measures of the trade and have launched an investigation into the anti-dumping and anti-subsidy of the imported products as well as imposed high import tariffs. Therefore, we need to pay attention to the development situation of import and export trade and adjust the variable parameters of import and export monitoring system which is related to the macro economy. It is necessary for us to make a right prediction for the trend of total import and export volume in a certain period of time in the future and to make dynamic information, which has important reference significance for management departments and enterprises to know about the dynamic information of import and export trade.

In recent years, some domestic scholars have also done a lot of research on the prediction of total volume of import and export. Luo Wenbin\textsuperscript{[1]} et al. predicted the monthly growth rate of China's total imports and exports from January 1990 to May 2006 by using the small data sets arithmetic, chaotic characteristic analysis of time series and prediction model based on radial basis function neural network. Li Jumeip\textsuperscript{[2]} et al. forecast the annual import and export data of China from 1984 to 2005 by using the time series analysis and established the corresponding model of time series ,then got the conclusion that the effect of short-term prediction of ARMA is better than the effect of long-term prediction.
Duan Peng [3] judged the single order and the integral order of monthly import and export data by using the method proposed by Beaulieu & Miron based on monthly data of import and export trade in 2009. He used the SARIMA model to forecast the data of import and export trade in the next two years. He concluded that the growth rate of export and trade surplus will slow down and the growth rate of import and the total volume of trade will accelerate in 2008. The growth rate of export and the trade surplus will increase; and the growth rate of import and the total volume of trade will increase in 2009. Bai Shuang [4] studied the related indicators of Shenzhen's import and export trade in 2013. Firstly, she extracted the four main factors with 99% resolution by adopting Factor Analysis. Then, ARMA, BP and SVM are used to predict according to the prediction results of sub-predictors, the weight of the three sub-predictors in the combined predictor is optimized by the combined prediction method, and the comprehensive prediction results with higher accuracy are obtained.

Wu Xin [5] et al. selected VAR model which is related to the import and export to analyze the structural characteristics of import and export trade and forecast the total volume of import and export according to the causal relationship and the lag relationship between the variables and the total volume of import and export. Ma Ming [6] used GM (1,1) model to make a gray forecast to the trade deficit of maize farming industry in China in 2016. Duan Lei [7] et al. used BP neural network and discrete GM combination forecasting model to forecast the annual data of Jiangsu's total volume of import and export from 2013 to 2015, and got some better results in 2017.

In summary, the models used by scholars for forecasting the date of import and export can be divided into three types: 1. Traditional time series models include auto-regressive (AR), moving average (MA), auto-regressive moving average (ARMA), auto-regressive integrated moving average (ARIMA) or SARIMA with seasonal characteristics. This kind of model mainly considers the fluctuation and trend of quantity of import and export. 2. Regression analysis: Screening other economic variables which are related to total volume of import and export in order to analyze the most suitable influence factors. Using multiple regression, stepwise regression or VAR models or simultaneous equation models. This kind of model analyzed the influence of other factors on import and export in order to deduce the change direction of import and export. But the influence factors of imports and exports are not well determined. 3. Nonlinear prediction: This kind of method mainly includes the gray prediction, SVM model, BP neural net, wavelet transform and so on. The strategies used in forecasting can be divided into two types: 1. single prediction; 2. combination prediction of two or more methods.

But most scholars use the annual data of import and export in the statistical yearbook. According to the actual situation of data, the annual data is much smoother than the monthly data, and the fluctuation is small. It is easy to get a better trend of fitting. But it is difficult to fit if the fluctuation of monthly data is great and quick. However, for the monitoring and early warning of import and export, the use of forecasting cycle of the annual data is longer. In recent years, in-depth learning has developed rapidly, and significant progress has been made in the accuracy of time series prediction. We are very good at dealing with nonlinear data with relatively long intervals and delays. Therefore, this paper forecast the trend of monthly data of import and export over the years by using Long Short-Term Memory (LSTM) network and compares the prediction results with those traditional analysis.

2. The prediction of trend based on LSTM

2.1. The introduction of LSTM

Before introducing LSTM, I will introduce recurrent neural network (RNN) first. The results are from the input layer to the hidden layer and then to the output layer in the process of being fully connected with neural network. Layers and layers are fully or partially connected, and the nodes between each layer are not connected. But the nodes between hidden layers of RNN are connected. The input of the hidden layer includes not only the input of the input layer, but also the output of the hidden layer of the previous moment. It not only learns the information of the current moment, but also depends on
the previous sequence information. Therefore, it can describe the relationship between the current output and the previous information, and it has unique advantages in processing the time series data.

The typical structure of RNN is shown on the left of the Figure 1. From the figure, we can see that the main structure, S of RNN not only comes from the input layer X, but also has a circular edge to provide the state at the current moment. At the same time, the state of S will be transferred from the current step to the next step. On the right of the Figure 1, the loop is expanded so that we can see the transformation of the information between the hidden layers clearly.

But RNN can only memorize several parts of the sequence, so its performance in long sequence is much worse than that in short sequence. The effect of RNN becomes worse when facing the phenomenon of "long-distance dependence". Because in the process of gradient return, every return must be multiplied by a rectangular with W matrix and calculated repeatedly, which will bring some problems in training, such as When the weightiness W is less than 1, the error of back propagation will be smaller and smaller, which may cause gradient dispersion. When the weightiness W is larger than 1slightly, the error of back propagation will become larger and larger, which will cause gradient explosion. In this case, no matter what numerical operation is performed on the gradient, the parameters cannot be updated by using the gradient and then LSTM is born at that right moment.

LSTM [8-10] is a special net structure with three “gates”: “forgetting gate”, “input gate” and “output gate”. The forgetting gate is mainly used to remember the state of memory cells and forget the information in memory cells selectively. Input Gates also act on memory cell states, selectively recording new information into new cell states. The output gate acts on the input and the output of the hidden layer, so that the final output includes both the state of cell and the input, and the result is updated to the next hidden layer. Through these three gates, LSTM can determine which information should be forgotten and which information should be retained more effectively. The cell structure of LSTM is shown in Figure 2, in which C is the cell state, X is the input and “a” is the output of each layer.
According to the structure of LSTM, the calculating formula of each LSTM cell is as follows: 
\[
\Gamma_f^{(t)} = \sigma(W_f[a^{(t-1)}, x^{(t)}] + b_f) 
\]
\[
\Gamma_i^{(t)} = \sigma(W_i[a^{(t-1)}, x^{(t)}] + b_i) 
\]
\[
\tilde{c}^{(t)} = tanh(W_c[a^{(t-1)}, x^{(t)}] + b_c) 
\]
\[
c^{(t)} = \Gamma_f^{(t)} \odot c^{(t-1)} + \Gamma_i^{(t)} \odot \tilde{c}^{(t)} 
\]
\[
\Gamma_o^{(t)} = \sigma(W_o[a^{(t-1)}, x^{(t)}] + b_o) 
\]
\[
a^{(t)} = \Gamma_o^{(t)} \odot tanh(c^{(t)}) 
\]

LSTM trained with Back Propagation (BPTT). To calculate the errors between the output value and the real value of each LSTM neuron. The gradient of each weight is calculated according to the corresponding errors, and the gradient descent algorithm is used to update the weight.

2.2. The empirical analysis of LSTM

2.2.1. Description and Pre-process of Data. The data which is used in this paper comes from the monthly import and export data of China Customs of Provincial Department of Commerce in Shandong Province. The data spans from January 2001 to June 2018, there are 210 samples with a unit of 100 million yuan in total.

We can see that the fluctuation of monthly data of import and export is great in the Figure 3, and it shows an overall growth trend, which is not a stable but volatile time sequence. Before 2008, imports and exports grew rapidly, showing an exponential growth trend. The financial crisis in 2008 had a great impact on imports and exports, the total volume is decreased by 50%. The total imports and exports in January 2009 were only half of those in August 2008. It slowed down considerably after 2010.

The difference of raw data. The advantage of LSTM is to deal with time sequence with certain correlation. The auto-correlation of the original sequence is not strong. LSTM carries out differential
processing. It can be seen the sequence after difference has 12-order correlation from the autocorrelation and partial autocorrelation plots of the difference sequence, which means the current quantitative value is related to the value of previous 12-period.

Normalization of input data. Because LSTM is sensitive to the size of the input data, especially when using sigmoid (default) or tanh activation functions, the data need to be normalized first. In this paper, we will scale the data between the maximum and minimum by standardizing the maximum and minimum. To transform the sequence, \( \ldots \), it's necessary to normalize the data. The transformation formula is:

\[
y_t = \frac{x_t - \min_{1 \leq j \leq n}(x_j)}{\max_{1 \leq j \leq n}(x_j) - \min_{1 \leq j \leq n}(x_j)}
\]

After transformation, the original sequence is scaled between [0,1].

To transform the time sequence into the problem of supervising learning. According to the autocorrelation, we predict the data of the 13th month by using the data of every 12 months. There are 197 samples after processing the original data.

To divide training set and test set. The first 80% of the samples are taken as training set and the remaining 20% as test set without disturbing the original order of the data. Thus, there are 157 samples in the training set and 40 samples in the test set.

To build input data that can be put into the LSTM. The data format required by LSTM is tensor format. The predictive variable (X) must be a three-dimensional array, its dimensions are samples, timesteps, and features. The first dimension represents the number of samples; the second dimension is the time step (lag order); and the third dimension represents the number of predictor variable (1 is a single variable, \( n \) is multivariate). The input form of training set is \((157, 12, 1)\), and the input form of test set is \((40, 12, 1)\).

2.2.2. The construction and prediction of model. The LSTM model used in this paper includes an input layer, two LSTM layers which are as a hidden layer and an output layer, and each hidden layer contains a Dropout layer to prevent over-fitting. After several adjustments, the input dimension of the input layer is selected to be 1. There are 128 units contained in the first hidden layer, the output dimension of the second hidden layer is 256, and the output dimension of the output layer is 1. The loss function used in compiling is mean square error (MSE), the optimizer is Adam, and epochs is 200. Keras framework is used to train the model.

To train the model on the training set and to predict on the test set. The predicted values of the training set and the test set are output respectively. We obtained the mean square errors of the training set and the prediction set by processing the normalization and inverse difference. It is as shown in Figure 4. The mean square error of training set is 84.49 and test set is 124.39.

![Figure 4. The MSE of the training set and test set.](image)
To combine the prediction results of training set with that of test set and to observe the fitting effect of LSTM in the same graph with the original data set. The effect is shown in Figure 5. We can see that the model fits the trend change of import and export data very well.

![LSTM imitative effect graph](image)

**Figure 5.** The fitting effect of the import and export of LSTM.

2.2.3. *The cubic smoothing prediction of index.* Classical time sequence model [11-14] is used to forecast the data of import and export. We choose the cubic exponential smoothing method here. Although the method is simple, the effect of prediction is better in practice. The cubic exponential smoothing algorithm can predict time series with both trend and seasonality. It retains seasonal information on the basis of quadratic exponential smoothing, so that it can predict time sequence with seasonality. A new parameter $P$ is added in order to express the trend after smoothing. It has two methods, the accumulation and the multiplication.

After training, the $\alpha$ is 0.6023, it means that about 60% of the current level depends on historical observations; $\beta$ is 0; $\gamma$ is 0.6163, which means that the estimated 61.6% of the trend is dependent on historical data. The mean square error of training set is 83.19, and that of test set is 141.69.

To combine the prediction results of training set with that of test set and to observe the fitting effect in the same graph with the original data set. The effect is shown in Figure 6.

![Winter exponential smoothing imitative effect graph](image)

**Figure 6.** The fitting effect of import and export data of the cubic exponential smoothing method.

3. Conclusions
We adopt LSTM recurrent neural network and cubic exponential smoothing method of classical time sequence to forecast the trend of import and export data monthly in this paper. From the training error and test error of this two training models, we can see that the training error of the cubic exponential smoothing method is lower than that of LSTM, but the prediction error of the test set is higher than that of LSTM model. This phenomenon illustrate that the generalization ability of LSTM is stronger, we can also see it from the picture of fitting effect. Both of them can better fit the trend and fluctuation of data in training set, but from the date after January 2015, we can see that LSTM model is better than cubic exponential smoothing method to fit the trend and fluctuation of data. Therefore, LSTM provides a better method to fit monthly data of import and export.
Acknowledgments
This work has been supported by Key Research and Development Plan in Shandong Province, Grant/Award Number: 2018GGX101012.

References
[1] Luo Wenbin, Wang Haiyan, Zhao Lindu. Analysis and Prediction of Chaotic Characteristic of Time Sequence of Total Import and Export Volume [J]. Industrial Technology Economy, 2007, 26 (1): 115-119.
[2] Li Jumei, Fan Liang and Cui Junfeng. Forecasting the total volume of China's import and export by time sequence analysis method [J]. Chongqing Institute of Technology (Natural Science), 2008, 22 (12): 136-138.
[3] Duan Peng. Import and Export of Commodities in China: Prediction Based on SARIMA [J]. Journal of Central China Normal University (Humanities and Social Sciences Edition), 2009, 48 (1): 60-66.
[4] Bai Shuang. Research on Forecasting Model of Foreign Trade of Import and Export in Shenzhen Based on Combination Prediction [D]. Tianjin: Tianjin University, 2013.
[5] Wu Xin, Wang Wenjie. Prediction of Potential of China's Import and Export Trade [J]. Reference for Economic Research, 2016 (33): 34-39.
[6] Ma Ming. Analysis and Scale Forecast of Import and Export Trade of Maize Farming Industry in China [D]. Hangzhou: Zhejiang University, 2016.
[7] Duan Lei, Hu Mingli. The research on forecasting of total foreign trade imports and exports based on BP neural network and discrete GM combined model [J]. China Management Science, 2017, 25 (11): 650-655.
[8] Wang Xudong, Yan Ke, Lu Huijuan, Ye Minchao. Test of LSTM Univariate Demand of Short-term Household Electricity [J]. Journal of International Trade, 2018, 29 (2): 142-148.
[9] Lu Zenan, Shang Yulin. The prediction of Steel price based on LSTM neural network model [J]. Science and technology horizon, 2017:116-117.
[10] Sun Ruiqi. Research on the forecasting model of the trend of US stock index price based on LSTM neural network [D]. Beijing: Capital University of Economics and Trade, 2015.
[11] Luo Hongben. Research on Intelligent Hybrid Forecasting of Financial Time Series Based on Grey-ARIMA [J]. Financial Theory and Practice (Bimonthly), 2014, 35 (188): 27-34.
[12] Wang Shuhua. Study on Combination Forecasting Model Based on Time Sequence [D]. Qinhuangdao: Yanshan University, 2011.
[13] Li Su. Prediction of the development trend of total volume of China's import and export trade based on gray system [J]. Business Research, 2008 (371): 113-115.
[14] Shen Hanxi, Lin Jian. The prediction of import and export of China's foreign trade based on ARIMA l: 2006-2010 [J]. International trade issues, 2007 (6): 112-114.24-2.