Railway freight volume forecasting based on combined DWT-Bi-LSTM model

Yue-ying Qiu and Ming Lei*
Tianjin University of Technology, Tianjin 300384, China
834473682@qq.com

Abstract. Railway freight volume forecasting can provide information for National Railway Department to establish effective operation lines, and its research plays an important role in increasing the effective proportion of railway transport. In order to improve the prediction accuracy, this paper proposes a combined model based on wavelet transform (WT) and bidirectional long short-term memory (Bi-LSTM) for high-precision railway freight volume forecasting. In this method, experimental data are first denoised by wavelet transform to extract important signal features that can accurately express the data information, and then Bi-LSTM is utilized to learn from the historical denoised data and iteratively improve upon predictions. To verify the improved prediction performance of DWT-Bi-LSTM, results are compared with those of the single LSTM, GRU, Bi-LSTM model along with the LSTM, GRU model combined with wavelet transform. The experimental results show that the combined DWT-Bi-LSTM model proposed in this paper has higher accuracy in the prediction of railway freight volume than the other forecasting methods.

1. Introduction
Rail transport is an important part of the Chinese logistics chain as this method can transport large volumes of freight over long distances at high speed and low cost. With the continuous development and improvement of China’s economy, rail transport has been presented with more opportunities and more room for growth, as well as fiercer competition in the market from other methods of transportation. Accurately predicting the railway expected freight volume can yield improvements in the operational efficiency of the railway system, allow for reasonable planning of future railway construction and development, and play an important role in effectively increasing the proportion of rail transport in the market.

The methods commonly used for railway freight volume forecast are qualitative forecast and quantitative forecast. Qualitative forecast includes the expert meeting method and the Delphi method, among others, each of which requires substantial knowledge and experience to directly analyze the characteristics of the predicted objects while yielding a forecast result with poor reliability. Quantitative forecast studies the relationship between factor variables and predictive variables by collecting related historical data and establishing corresponding models, which can be roughly divided into the following three categories: 1) Time sequence model: In reference [1], ARIMA is used to analyze the railway freight volume and obtain higher accuracy. Reference [2] used the SARIMA model to make short-term extrapolation predictions for monthly railway freight volume and provides corresponding decision-making suggestions for railway enterprises based on the predicted results. However, the railway freight volume presents a complex nonlinear sequence; as the forecast period grows longer, it is hard to find a linear model with a height-fitting model to measure the relationship
among factors. In reference [3] and reference [4], the Holt-Winter model is used to predict the railway freight volume characterized by periodic fluctuations and linear trends. The model can appropriately filter out the effects of random fluctuations in the time series and obtain good results in short-term predictions. Due to the neglect of some socioeconomic factors, however, the accuracy is not high in long-term forecasts. 2) Causal model: The theory of grey system is introduced into the short- and medium-term forecast of railway freight volume in reference [5] and [6]. The factors affecting railway freight volume can be scientifically and objectively extracted and incorporated into the forecast. However, such models need a large amount of relevant historical data, and there is a large subjectivity in the establishment of the prediction system. At the same time, the long-term prediction stability of the grey model is poor. 3) Machine learning model: Reference [7] is based on the support vector machine model with three kinds of kernel function to conduct prediction experiments and obtain the ideal results respectively. However, the support vector machine has defects in multi-class prediction and is sensitive to the selection of kernel function and parameters, so it is difficult to obtain the global optimal solution with this model. Based on BP neural network, the AdaBoost algorithm is used to predict the effectiveness of China’s annual railway freight volume in reference [8]. In reference [9], the GA-BP prediction model was established by combining the grey model with BP neural network. However, BP neural network has the problems of slow convergence speed, ease of becoming “trapped” at a local minimum value, and sensitivity to initial weight and threshold training. Therefore, with the finite samples, it is necessary to be able to find an algorithm with high precision and ability to perform well over a long period to yield worthwhile forecasting results.

Deep learning is a process that simulates human cognition. It can automatically learn data and dig deeper features, which has great potential in prediction. As a deep learning method, Long Short-Term Memory Network (LSTM) [10] is good at dealing with multiple time-continuous variables. It is widely used in machine translation, speech and image recognition, text mining, and various predictions [11-13]. LSTM has a high learning ability for long span sequence processing and is more suitable for solving the nonlinear relationship between railway freight volume and influencing factors. Therefore, this paper proposes a DWT-Bi-LSTM hybrid model combining wavelet transform and bidirectional long short-term memory network to predict future monthly railway freight volume. Wavelet transform can extract important signal features that can accurately express data information from original data. The bidirectional long short-term memory network can make use of the contextual information in both directions of the time series to complete the cycle learning of the data, to obtain results closer to the real data, and provide an effective decision-making reference for the railway department.

2. Algorithms introduction

2.1. Wavelet transform
Wavelet transform is a new signal-adaptive time-frequency analysis method [14] conducted after Fourier transform. This method overcomes the shortcomings of the Fourier transform that cannot describe the local characteristics of signals in the time domain, and it has a good processing effect on the transient signal. Wavelet transform is used to carry out multi-scale localization subdivision of data signal through translation and contraction operations, and decomposes a group of original signals \( f(x) \) containing comprehensive information layer by layer into an approximation information sequence \( A \) that reflects the real trend of signals and a sequence \( D \) that represents the small range of details:

\[
A_{j+1,k} = \sum_m H (m - 2k)A_{j,m}
\]

\[
D_{j+1,k} = \sum_m G (m - 2k)A_{j,m}
\]

\[
f(x) = A_j + D_1 + D_2 + \cdots + D_j
\]

where \( H \) and \( G \) represent the low-pass filter and high-pass filter respectively, \( j \) represents the decomposition scale, and \( k \) and \( m \) represent the translation variables.
The noise component of the decomposed signal exists in the high-frequency coefficient part. Each time the low-frequency coefficient part is decomposed again, the noise signal is partly stripped away and the peak curve of the signal gradually becomes smooth, thereby achieving the effect of signal denoising.

2.2. Long short-term memory network

Long short-term memory (LSTM) network is an improved recurrent neural network with a strong ability to learn long-term dependencies within data and demonstrates superior accuracy in the learning modeling of time series data. LSTM introduces a complex gate structure and storage unit on the basis of recurrent neural network (RNN), which effectively solves the gradient vanishing problem in RNN [15-16]. Compared with other models, LSTM can better save and learn past information through its special gate structure and realize the “forgetting” and “retention” of data dependencies. In recent years, LSTM has been widely used in activity identification and prediction.

LSTM consists of an input layer, an output layer, and a hidden layer. Different from the traditional network model, LSTM adds a gate structure in the hidden layer for data selection learning. LSTM neurons contain a storage unit, which is composed of three ”gate” structures [17-18], including a forget gate, an input gate, and an output gate along with a loop connection unit. Through the gate structure, the LSTM network can selectively learn and decide whether to learn long-term stored information [19]. Figure 1 shows the internal structure of LSTM, where $x_t$ is the input of the cell at time $t$ and $h_{t-1}, h_t$ is the output at time $t - 1, t$, respectively.

![Figure 1. The internal structure of LSTM.](image)

First, the forgetting gate selects the information that was discarded at the last moment:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$  \hspace{1cm} (4)

Then, the new information is determined through the input gate. At the same time, a new candidate vector $\tilde{C}_t$ is created through the $\tanh$ layer:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + bi)$$  \hspace{1cm} (5)

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$  \hspace{1cm} (6)

Next, the cell state is updated to discard the old state information and add the new state information:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$  \hspace{1cm} (7)

Finally, the output result $h_t$ is obtained, which is based on the interaction between the output gate and the current state of the cell:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$  \hspace{1cm} (8)

$$h_t = o_t \cdot \tanh(C_t)$$  \hspace{1cm} (9)

where $\sigma$ is the sigmoid function and $\tanh$ is the hyperbolic tangent function. The terms $W_f, W_i, W_c,$ and $W_o$ are the weight matrices for each forget gate, input gate, state transition, and output gate, respectively. The terms $b_f, b_i, b_c,$ and $b_o$ are the corresponding offset vectors for each forget gate, input gate, state transition, and output gate, respectively.
2.3. Bidirectional long short-term memory network

Bidirectional long short-term memory (Bi-LSTM) network consists of a pair of forward LSTM and reverse LSTM, and connects the same input and output, as shown in figure 2. For a set of time series $S: x_1, x_2, x_3, ..., x_n$, when the forward LSTM receives the sequence $S$, the information of an element $x_t$ in the sequence is based on the above information $x_1, x_2, ..., x_{t-1}$, and the following information $x_{t+1}, x_{t+2}, ..., x_n$ is ignored. Similarly, when the reverse LSTM receives the sequence $S$, the information of an element $x_t$ in the sequence is based on the following information $x_{t}, x_{t+1}, ..., x_n$, and the above information $x_1, x_2, ..., x_{t-1}$ is ignored. In summary, for a unidirectional LSTM, part of the information will be lost when it receives the time series for training, resulting in inaccurate final training results. Therefore, in order to solve this problem, the bidirectional long short-term memory network is proposed, which trains each time series according to the two-way training pattern that will be forward and backward, so that the output can fully capture every piece of information of the context and obtain the sufficient and stable training results.

![Figure 2. The structure of Bi-LSTM.](image)

In the Bi-LSTM network model, the state of the hidden layer of the forward LSTM can be represented as $h_t$, and the state of the hidden layer of the reverse LSTM is $\bar{h}_t$. In order to capture the context information at the same time, the model integrates these two states to obtain a new hidden layer state $\bar{h}_t$, and weights $h_t$ to get the prediction result $y_t$:

$$h_t = \sigma(W_h \cdot [h_{t-1}, x_t] + b_h)$$  \hspace{1cm} (10)
$$\bar{h}_t = \sigma(W_{\bar{h}} \cdot [h_{t+1}, x_t] + b_{\bar{h}})$$  \hspace{1cm} (11)
$$h_t = h_t \oplus \bar{h}_t$$  \hspace{1cm} (12)
$$y_t = W_y h_t + b_y$$  \hspace{1cm} (13)

where $W_h, W_{\bar{h}},$ and $W_y$ are the weight matrices of the current neuron, $b_h, b_{\bar{h}},$ and $b_y$ are offset vectors, and $\oplus$ is the integration operation.

3. Hybrid DWT-Bi-LSTM prediction model

This paper proposes a railway freight volume prediction model that combines wavelet transform and Bi-LSTM. The specific steps are as follows:

Step 1: In order to prevent the difference in input and output data levels from being too large to affect the training weight and lead to the deviations, min-max normalization is performed on the sample data to map the data within the range of (0,1) and easier convergence to the optimal solution is achieved.

Step 2: The input sequence $\{x_1, x_2, ..., x_n\}$ can be transformed by multiscale k-layer wavelet transform according to equations (1)-(3) to obtain $\{X'_1, X'_2, ..., X'_n\}$, in which $X'_i = (A_{k,i}, D_{1,i}, ..., D_{k,i})$, $i = 1,2, ..., n$, where $A_{k,i}$ is the approximation signal and $D_{1,i}, ..., D_{k,i}$ are the detailed signals from layer one to layer k.

Step 3: The data set $\{X'_1, X'_2, ..., X'_n\}$ is divided into a training set $\{X'_1, X'_2, ..., X'_n\}$ and a test set $\{X'_{m+1}, X'_{m+2}, ..., X'_n\}$, $m < n, m, n \in N$. 


Step 4: The training set is fed into the first layer of Bi-LSTM neurons. The states of the forget gate \( f_t \), the input gate \( i_t \), the candidate vector \( c_t \), and the output gate \( o_t \) in the forward and backward layers, respectively, are calculated according to equations (4)-(8).

Step 5: The forward output \( h_t \) and backward output \( \bar{h}_t \) of the current neuron are calculated according to equation (9) and the existing state.

Step 6: The output integration state \( h_t \) is obtained according to equation (12) and the next layer of Bi-LSTM neurons is entered.

Step 7: Steps 4-6 are repeated until the last layer of Bi-LSTM neurons to obtain the output result \( y_t \).

Step 8: The parameters and number of layers of the Bi-LSTM model are adjusted based on the analysis between the predicted results and the actual results until the predicted results are the best, and then the model is saved after iteration.

Step 9: The test set is entered into the saved model to predict results.

4. Experimental analysis

4.1. Data
For this experiment, railway freight volume prediction was conducted for the Shanghai area, with a data set comprised of eight terms that are each tabulated monthly by the Shanghai Statistics Bureau. The data set covers the 68 months from January 2014 to September 2019, and includes total retail sales of consumer goods \( X_1 \), total volume of imports and exports \( X_2 \), GDP \( X_3 \), secondary industry \( X_4 \), tertiary-industry \( X_5 \), output of crude steel \( X_6 \), waterway freight volume \( X_7 \), and highway freight volume \( X_8 \). In the experiment, the first 80% of the data set (52 months of data from January 2014 to April 2018) was used as a training sample for building and training of the model, and the remaining 20% (16 months of data from May 2018 to August 2019) was used as a test sample for verification and comparison of DWT-Bi-LSTM predicted results to real-world observed results.

4.2. Experimental results and analysis
The signal is decomposed into different frequency channels by wavelet transform, and the single signal decomposed is more stable than the original signal. Figure 3 shows the result of the input index \( X_1 \) after three-layer wavelet transform - the original information is decomposed into a low-frequency information group \( A = (a_3) \) and a high-frequency information group \( D = (d_1, d_2, d_3) \). As can be seen in figure 3, the low-frequency information group is similar to the original information, which represents the approximate information of the real trend of the signal. In contrast, the high-frequency information group presents more complex structural features, which represent the detailed information of the original signal in a small range, and has the characteristic of indefinite period. Therefore, wavelet transform can be used to separate the original information from the components that accurately represent the real data for the next step.

![Figure 3. The decomposition result of the total retail sales of consumer goods.](image-url)
Figure 4 shows the comparison between the prediction curve of the DWT-Bi-LSTM model and the actual curve. It can be seen that some points of the prediction curve will deviate from the actual results, but the overall prediction curve has a high degree of fitting. Figure 5 shows the absolute percentage error between true and predicted values. The error of each point is concentrated in a low value (less than 0.002), indicating that the model has a high prediction accuracy.

In order to further verify the prediction ability of the DWT-Bi-LSTM model, the following comparison experiments with the benchmark prediction model selected in this paper are needed: (1) Selection of three traditional machine learning prediction models: LSTM, GRU, and Bi-LSTM to predict railway freight volume and verify the prediction performance of the model algorithm. (2) Combining of wavelet transforms of three-layer, four-layer, and five-layer with these three single prediction models for comparison and prediction. MAPE, RMSE, and MAE are used as indicators to evaluate the prediction performance of the model. The calculation equation is as follows, where $y_i = \{y_1, y_2, ..., y_n\}$ is the target value and $y'_i = \{y'_1, y'_2, ..., y'_n\}$ is the predicted value:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - y'_i}{y_i} \right|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y'_i)^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - y'_i|$$

The experimental results are shown in figure 6 and table 1.

Figure 6 (a) and table 1 show that the three machine models each have a level of prediction ability in the estimation of railway freight volume, and that the Bi-LSTM model has the best fitting effect of the three. This shows the rationality of selecting Bi-LSTM as the prediction benchmark model for this experiment.
Figure 6 (b)(c)(d) shows the prediction results of wavelet transformation models of three-layer, four-layer, and five-layer, respectively. The comparison between figure 6 and table 1 shows that the signal fitting degree of the original data is improved by using the three-layer of Daubechies wavelet basis function, which indicates that the wavelet transform can significantly improve the prediction ability of the original model and has a strong ability to predict the extreme points. With the continuous decomposition of the signal, the finer the signal frequency, the better the stability and smoothness after the decomposition. However, compared with the three-layer wavelet transform, the errors after the four-layer and five-layer wavelet transform increase. This indicates that the excessive number of layers is likely to cause information loss during the transformation, which leads to larger errors in the prediction results. Therefore, an appropriate number of decomposition layers should be selected.

![Figure 6](image)

Figure 6. The prediction results of single models and wavelet transformation models of three-layer, four-layer, and five-layer.

Table 1. Prediction performance comparison of methods.

| model                      | MAPE(%) | RMSE  | MAE  |
|----------------------------|---------|-------|------|
| LSTM                       | 13.76   | 7.84  | 5.03 |
| GRU                        | 9.14    | 4.48  | 3.31 |
| Bi-LSTM                    | 7.95    | 3.71  | 2.81 |
| DWT-LSTM(three-layer)      | 9.04    | 5.16  | 3.33 |
| DWT-GRU(three-layer)       | 6.80    | 3.30  | 2.53 |
| DWT-Bi-LSTM(three-layer)   | **4.21**| **2.27**| **1.64** |
| DWT-LSTM(four-layer)       | 14.34   | 7.05  | 5.57 |
| DWT-GRU(four-layer)        | 17.28   | 9.28  | 6.67 |
| DWT-Bi-LSTM(four-layer)    | 9.38    | 4.44  | 3.56 |
| DWT-LSTM(five-layer)       | 11.98   | 5.58  | 4.65 |
| DWT-GRU(five-layer)        | 18.34   | 8.22  | 6.94 |
| DWT-Bi-LSTM(five-layer)    | 11.44   | 4.97  | 4.09 |
The results of the DWT-Bi-LSTM model depicted in table 1 are closest to the real value, which indicates that the prediction ability of this model is better than the ability of the other models tested. The MAPE value of the DWT-Bi-LSTM model is 4.21%, which represents reductions of 9.55% from the traditional LSTM prediction model of 3.71% from the original Bi-LSTM model. Moreover, the RMSE and MAE values are 2.27 and 1.64, which are both lower than those of the other prediction models, indicating that the model has higher prediction accuracy and can be used as an effective prediction model for railway freight volume.

5. Conclusion
In this paper, based on the non-linear characteristics of railway freight volume data, a hybrid DWT-Bi-LSTM forecasting model was constructed to predict the monthly railway freight volume in Shanghai. The prediction result of MAPE is 4.21%, indicating that the hybrid model has a high prediction accuracy and can effectively predict the future monthly railway freight volume. Compared with the LSTM and Bi-LSTM models, MAPE is reduced by 9.55% and 3.71%, respectively, indicating that the bidirectional long short-term memory network has the advantages of complete learning and iterative improvement from data. Furthermore, wavelet denoising can significantly improve the prediction accuracy of the original model, and the combined hybrid model can achieve better prediction results.

References
[1] Yin Z B, Zheng P H, Mei J N and Zhang E L 2008 Application of ARIMA model on forecast of railway freight volume Railway Freight Transport 6 8-11
[2] Liu M T and Yu J L 2015 Forecast analysis of China railway freight volume based on SARIMA model Journal of Guizhou Education University 31(12) 43-47
[3] Tong M R, Xue H X and Lin L 2007 Study on the forecast of railway freight traffic volume based on Holt-Winter model Railway Transport And Economy 29(1) 79-81, 86
[4] Tang Y L and Li L 2017 Prediction research about monthly traffic volume based on Holt-Winters model Journal of Transportation Engineering and Information 15(2) 1-5,13
[5] Zuo X Y and Huang X J 2016 Prediction of railway freight volume based on grey prediction model in China Logistics Sci-Tech 8 82-84.
[6] An Y, Bao X Y and Wang Q C 2016 Railway freight volume forecasting based on unbiased grey Verhulst model Journal of Railway Science and Engineering 13(1) 181-186
[7] Zhao C, Liu K and Li D S 2004 Research on application of the Support Vector Machine in freight volume forecast Journal of the China Railway Society 26(4) 10-14
[8] Li S, Xie Y L and Wang W X 2012 Application of AdaBoost_BP neural network in prediction of railway freight volumes Computer Engineering and Applications 48(6) 233-234
[9] Li P 2014 Railway freight volume forecasting method based on GA-BP model Lanzhou Jiaotong Univ 33 (3) 203–207
[10] Hochreiter S and Schmidhuber J 1997 Long short-term memory Neural Computation 9(8) 1735-1780
[11] Shahzad M and Afshin A 2019 Short-term load forecasts using LSTM network Energy Procedia 158 2922-2927
[12] Reddy B K and Delen D 2018 Predicting hospital readmission for lupus patients: an RNN-LSTM-based deep-learning methodology Computers in Biology and Medicine 101 199-209.
[13] Zhang X P, Ye M, Yang J Z, Zhang J F and Zhu Y 2018 Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas Journal of Hydrology 561 918-929
[14] Pang J F, Liu B, Zhang B, Zhang P P, Wang B and Zeng F J 2019 Study on the combined model of forecasting the days of sand-dust weather based on wavelet decomposition--taking the time series of dust weather in the transitional zone of qira desert-oasis during 2008-2016 as an example Meteorological Monthly 45(05) 651-658
[15] Wu Y X, Wu Q B and Zhu J Q 2019 Improved EEMD-based crude oil price forecasting using
LSTM networks *Physica A: Statistical Mechanics and its Applications* **516** 114-124

[16] Zheng Z, Chen H N and Luo X W 2019 Spatial granularity analysis on electricity consumption prediction using LSTM recurrent neural network *Energy Procedia* **158** 2713-2718

[17] Graves A and Schmidhuber J 2005 Framewise phoneme classification with bidirectional LSTM and other neural network architectures *Neural Networks* **18**(5) 602-610

[18] Hu R, Fang F and Pain C C 2019 Rapid spatio-temporal flood prediction and uncertainty quantification using a deep learning method *Journal of Hydrology* **575** 911-920

[19] Li Z B, Peng F, Niu B S and Li G Y 2018 Water quality prediction model combining sparse Auto-encoder and LSTM network *IFAC-PapersOnLine* **51**(17) 831-836