Research on Railway Passenger Volume Prediction Based on LSTM Neural Network

Yuping Xu¹,², Wei Xu¹,²* and Siwei Chen¹,²
¹East China Jiaotong University Transportation and Engineering Application Translation Research Center, Nanchang, Jiangxi, 330013, China
²East China Jiaotong University, Nanchang, Jiangxi, 330013, China
*Corresponding author’s e-mail: xwncl995@163.com

ABSTRACT. Accurate railway passenger volume prediction plays an important role in the development of railway passenger transport industry. Taking the national railway passenger volume as an example, based on the data of 2005-2017, LSTM neural network prediction model is established. The prediction model results are compared with the actual situation, and the relative error rate is analyzed. The results show that the prediction model of LSTM neural network has low error and good prediction effect.

1. Introduction

In recent years, railway plays a more and more important role in economic development and has a far-reaching impact on social development. Compared with other modes of transportation, railway transportation has many advantages, such as long transportation distance, strong transportation capacity, less limited by climatic conditions, high safety, low cost, etc. It is a convenient means of transportation. In July 2016, the National Development and Reform Commission, the Ministry of Transport and China Railway Corporation jointly issued the "Medium and Long Term Railway Network Planning", which outlines the grand blueprint of the "Eight Verticals, Eight Horizons" high-speed railway network in the new era. With the continuous increase of the total railway mileage and the gradual improvement of the railway network in our country, it is more convenient for people to choose railway travel than before. The forecast of railway passenger volume is helpful to make a general judgment on the development trend of railway passenger volume in the future, which plays a vital role in the development of railway passenger transport industry. Accurate forecast of railway passenger volume can not only discover problems in the development of railway passenger transport industry in China in time, but also provide constructive suggestions for relevant departments to improve railway passenger volume in China.

The commonly used methods for predicting railway passenger volume are mainly divided into two categories, one is parametric methods, such as ARIMA, multiple regression theory, exponential smoothing method, grey model, Markov model, and the other is non-parametric methods, such as k-nearest neighbor model, neural network, etc. [1]. Sun Li [2] used partial least squares regression model and grey GM(1,1) prediction model to construct IOWA combined prediction model based on the sum of squares of errors; Tang Yinying [3] constructed seasonal differential moving autoregressive model (SARIMA) to accurately predict railway monthly passenger volume in 2016; Ge Ling [4] used product seasonal model to build models, respectively. Eviews and R software are used to forecast the railway passenger volume; Wang Bin [5] establishes a combined forecasting model based on grey model and linear regression model to forecast the railway passenger volume; Li Li [6] used the improved four-stage
passenger flow forecasting method to predict the volume of Jiqing high-speed railway Wang Xiaofan[7] establishes a GRNN generalized regression neural network forecasting model to forecast the railway passenger volume of Qingdao City; non-parametric method relies on data to adjust the internal parameters of the model, and then obtains variables and variations. The relationship between quantities is less affected by users, so it has a broader scope of application. In non-parametric methods, neural networks are widely used because of their distributed storage, self-organization, self-regulation and non-linear fitting capabilities[8]. Although the neural network has the advantages mentioned above, its prediction ability for long time series is limited, so LSTM is proposed as an improved time recursive neural network (RNN), which can learn the long-term and short-term information of time series. Because it contains time memory units, it is suitable for processing and predicting interval and delay events in time series.

2. LSTM Neural Network
Long-Short Term Memory neural network, first proposed by foreign scholars Sepp Hochreiter and Jurgen Schmidhuber in 1997, is a specific form of RNN. The hidden unit of RNN in each step only performs a simple Sigmoid, Tanh or Relu operation. Its memory function needs complex training. However, due to the problem of gradient extinction, RNN can’t be trained in many cases. LSTM is a special form of RNN. Unlike RNN, LSTM neural network adds a "processor" (also known as cell), whose function is to judge the validity of information. Each cell contains an Input Gate, Forget Gate, and Output Gate, as shown in Figure 1.

![Figure 1. Internal structure of the LSTM memory block](image)

If the input sequence is \( (x_1, x_2, \cdots, x_t) \), then at \( t \):

\[
a'_t = \sum_{i=1}^{I} w_{i a} x'_i + \sum_{h=1}^{H} w_{h b} b'^{t-1}_h + \sum_{c=1}^{C} w_{c s} s'^{t-1}_c
\]  
(1)

\[
a'_c = \sum_{i=1}^{I} w_{i c} x'_i + \sum_{h=1}^{H} w_{h c} b'^{t-1}_h + \sum_{c=1}^{C} w_{c c} s'^{t-1}_c
\]  
(2)

\[
a'_c = \sum_{i=1}^{I} w_{i c} x'_i + \sum_{h=1}^{H} w_{h c} b'^{t-1}_h
\]  
(3)

\[
a'_c = \sum_{i=1}^{I} w_{i c} x'_i + \sum_{h=1}^{H} w_{h c} b'^{t-1}_h + \sum_{c=1}^{C} w_{c c} s'^{t-1}_c
\]  
(4)

\[
s'_c = b'_c s'^{t-1}_c + b'_t g(a'_c)
\]  
(5)
In the formula:

\[ w_{ij} \] - the weight value between neuron I and j; \\
\[ a_j^t \] - the input value of neuron J at t time; \\
\[ b_j^t \] - the output of neuron activation function at t time; \\
\[ l \] - the input gate; \\
\[ \varnothing \] - the forgetting gate; \\
\[ w \] - the output gate; \\
\[ c \] - the memory unit; \\
\[ w_{cl} \] - the connection weight value between memory unit and output gate; \\
\[ c^t \] - the state of memory unit at t time; \\
\[ f \] - the activation function of three gates; \\
\[ g \] - the activation function of memory unit; \\
\[ h \] - the output function of memory module; \\
\[ I \] - Number of input neurons; \\
\[ K \] - Number of output neurons; \\
\[ H \] - Number of hidden neurons. In the hidden layer, neurons are connected with other memory modules. The activation functions of memory modules, such as neuron state and gate activation functions, only act on the inner part of memory modules.

3. Data Processing

3.1. Experimental environment

In this paper, Keras is used as the experimental platform of LSTM neural network. The existence of Keras can greatly reduce the coding pressure of neural network algorithm. At the same time, the convenience of Keras makes it easy to be used in various fields. Keras's convenience is mainly embodied in its ability to generalize the complex Tensorflow grammar in a few simple lines of code.

| Experimental Platform | Design language | Operating System | System Hardware |
|-----------------------|-----------------|------------------|-----------------|
| Keras                 | Python          | Win7             | I7CPU 、 8GRAM  |

3.2. Loss function and optimizer

Keras constructs LSTM neural network with two important parameters, loss function and optimizer. Among them, the loss function is mainly used to measure the deviation between the output value of the model and the actual value, so as to measure the accuracy of the model. Since the railway passenger volume prediction belongs to the regression problem, this paper uses Mean_squared_error as the loss function to measure the difference between the forecast value and the real value of the model output. The definition of MSE is as follows:

\[
L(Y, f(X)) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(X_i))^2
\]  

In the formula, \( X_i \) is the input value of the model and \( Y_i \) is the output value of the model. Ideally, when \( X \) reaches the minimum value, the corresponding parameters are optimal, the model is optimal at this time.

The optimizer is the choice of the gradient descent method in the LSTM model. This article uses Adam as the optimizer. Adam can iteratively update the weight of the neural network based on the training data. The derivation formula of the Adam algorithm is as follows:

\[
m_t = \mu \times m_{t-1} + (1 - \mu) \times g_t
\]

\[
n_t = \nu \times n_{t-1} + (1 - \nu) \times g_t
\]

\[
\dot{m}_t = \frac{m_t}{1 - \mu}
\]

\[
\dot{n}_t = \frac{n_t}{1 - \nu}
\]

\[
\Delta \theta_t = -\frac{\dot{m}_t}{\sqrt{\dot{n}_t + \varepsilon}} \times \eta
\]

In the formula, \( m_t, n_t \) are the first moment estimation and second moment estimation of the gradient,
\( \hat{m}_t, \hat{n}_t \) are the corrections of \( m_t \) and \( n_t \), which can be approximated to the unbiased estimation of the desired, and \( \eta \) is the learning rate of the model.

### 3.3. Activation function

There are many forms of activation functions, and its main role is to make neural networks better solve more complex nonlinear problems. There are many forms of activation functions, the most commonly used are the Sigmoid function, the tanh function, and the Relu function. Among them, the Sigmoid function is calculated as follows:

\[
sigmoid(x) = \frac{1}{1 + e^{-x}}
\]

### 3.4. Establishment of LSMT neural network

The process of LSTM neural network training is actually a supervised learning process. It needs to design a training set data structure. According to the requirements of railway freight volume data for LSTM neural network.

### 4. Empirical analysis

Before the LSTM neural network trains the railway passenger traffic data, it is necessary to preprocess the original data. Normally, the data of the LSTM neural network is normalized data, and its function is to prevent the input data from being too large or too small, resulting in poor training effect of the model. The normalized data range is \([-1, 1]\). Import the MinMaxScaler module in the sklearn toolbox and call the fit_transform() function to normalize the training data.

When training in machine learning, we must turn the problem into a supervised learning problem. The data processed by the general neural network can be divided into two parts: input (X) and output (Y). The feature that the neural network needs to find is between the observation of the previous time step (t-1) and the current time step (t). relationship. We can use the shift function of the Pandas library in Python to achieve the purpose of the conversion.

According to the analysis, the results of the railway freight volume data prediction using the LSTM neural network with the above parameters are shown in Figure 2 and Table 2.

![Figure 2. Prediction results of LSTM neural network](image)
Table 2. Comparison table between prediction results and actual values

| Month    | 2018.04   | 2018.05   | 2018.06   | 2018.07   | 2018.08   | 2018.09   |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Actual   | 28900     | 26827     | 27834     | 32276     | 34340     | 28253     |
| Predictive | 28816     | 26889     | 24778     | 33993     | 30590     | 29652     |
| Relative | 0.29%     | 0.23%     | 10.98%    | 5.32%     | 10.92%    | 4.95%     |

| Month    | 2018.10   | 2018.11   | 2018.12   | 2019.01   | 2019.02   | 2019.03   |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Actual   | 30467     | 25177     | 25164     | 28342     | 29112     | 27860     |
| Predictive | 27268     | 24464     | 23584     | 26854     | 27188     | 25266     |
| Relative | 10.50%    | 2.83%     | 6.28%     | 5.25%     | 6.61%     | 9.31%     |

Combined with Table 3-1 and Figure 3-1, the LSTM neural network can roughly describe the trend of railway passenger traffic. The average error is 5.94%, the maximum relative error is 10.98% in June 2018, and the minimum relative error is May 2018. The 0.23% indicates that the simulation model has good learning ability, which can make the prediction result more accurate and is suitable for the prediction of railway passenger traffic.

**Foundation**

Jiangxi Social Science Planning Fund Program (18YJ16); Jiangxi Provincial Graduate Innovation Special Fund Project (YC2018-S252); 2018 Annual Research Base of Humanities and Social Sciences of Colleges and Universities in Jiangxi Province - Tendering Project of Communication and Engineering Application Translation Research Center of East China Jiaotong University (JD18104); 2015 Humanities and Social Sciences Planning Fund of the Ministry of Education (15YJAZH091).

**References**

[1] Li, W., Feng, F.L., Jiang, Q.W. (2018) Prediction for railway passenger volume based on modified PSO optimized LSTM neural network. Journal of Railway Science and Engineering, 15(12):260-266.

[2] Sun, L., Mou, H.B. (2018) A High-Speed Railway Traffic Forecast based on IOWA Combined Model, Railway Transport And Economy, 40(09):74-79.

[3] Tang, Y.Y., Zhu, X.L., Li, L. (2019) Monthly Railway Passenger Traffic Volume Forecasting Based on SARIMA Model, Journal of Transportation Engineering and Information, 17(01):25-32.

[4] Ge, L., Zhang, J. (2018) The Forecast of Railway Passenger Traffic Volume Based on Multiple Seasonal ARIMA, Journal of Chongqing Technol Business Univ, 35(03):18-25.

[5] Wang, B. (2017) Forecast of Railway Passenger Traffic Based on Grey Linear Regression Combined Model, Journal of Chongqing University of Technology, 31(11):230-234+240.

[6] Li, L. (2016) Traffic Volume Forecast for Ji’nan-Qingdao High-Speed Railway, Railway Transport And Economy, 38(9):45-49.

[7] Wang, X.F., Zhu, Y.Q., Pan, F.Q. (2019) Prediction of Railway Passenger Volume Based on GRNN Neural Network, Journal of Luoyang Institute of Science and Technology, 29(01):40-43+50.