Passenger Flow Prediction based on Recurrent Neural Networks and Wavelet Transform

Jing Huang¹, Fubo Shao¹² and Shuguo Yang¹*

¹School of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao, Shandong, 260061, China
²The Post-Doctor Station of the CRRC Corporation Limited, Beijing, 100067, China
*Corresponding author’s e-mail: ysg_2005@163.com

Abstract. It is a great significance for the accurate and real-time prediction of passenger flow in rail transit operation. In the process of passenger flow prediction, the new method of recurrent neural networks (RNNs) can well solve the problems of randomness and nonlinearity which cannot be solved by the existed linear models. In this paper, the long short-term memory (LSTM) and the gated recurrent unit (GRU) networks, which are methods of RNNs, are employed to predict the dayparting passenger flow and the raw passenger flow data is denoised by the wavelet transform. Experimental results show that LSTM and GRU networks can well predict the passenger flow. And compared to LSTM, GRU is better for passenger flow prediction.

1. Introduction

From the design to the operation, passenger flow prediction is always one of the most important problems of urban rail transit management, and the variation law of passenger flow is the premise of the safe operation of metro. However, there is a big gap between the actual passenger flow and the predicted passenger flow in the actual operation, the passenger flow of different routes is unbalanced. Therefore, how to use the existing data to accurately predict the passenger flow has become an important means to solve the problem of the overload operation and alleviate the pressure of the traffic operation department.

At present, there are many kinds of passenger flow forecasting models, which are generally divided into linear models [1-2] and non-linear models [3-5]. Linear prediction is represented by time series analysis, such as autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) [6]. They consider the continuity and randomness of events to predict the occurrence of future events. It can reflect the changing rules of periodic change and trend change of events. The methods to solve the problem of non-linear prediction include the method of system dynamics and the recurrent neural networks (RNNs) which have developed rapidly in recent years. RNNs can fully learn the features of passenger flow, especially long short-term memory (LSTM) and gated recurrent unit (GRU) networks. The forecast results of RNNs are better and more robust than the traditional time series models and econometric models [3,6].

The contributions of this paper mainly include the following two points. The first one is data denoising. Since the data used in this paper is the real passenger flow data, there is a big noise. Therefore, we use a wavelet filter to smoothen it before performing prediction [7]. Experiments show that the data denoising can improve the prediction accuracy of models. The second one is the use of LSTM and GRU to predict and analyze passenger flow [6,8]. We have also compared the effects of different wavelet
functions and activation functions in the experiments. Experiments show that GRU is better for passenger flow prediction.

The rest of this paper is organized as follows. Section 2 reviews the methods related to this paper. Section 3 introduces the used methods and model structures. Section 4 describes the experimental setup and discusses the experimental results of selected models. Experimental evaluation and conclusion are described in Section 5.

2. Introduction of related methods

2.1 Wavelet denoising

Traditional denoising methods, such as median filtering, cannot describe the signal with large fluctuation in detail, it may lose useful information. While wavelet denoising uses the stretching and translating of wavelet to describe signals, it adopts wavelet to decompose signals into different scale spaces, so that these signals can be observed more specifically, which is helpful to distinguish signal and noise [7,9].

Wavelet analysis uses a series of wavelets generated by the transformation of wavelet function to more accurately describe local features of signals (including non-stationary features such as spikes). The wavelet basis function $\phi(t)$ is transformed by scaling factor $a$ and shift factor $b$ to do scalar product with signal $x(t)$, as shown in equation (1).

$$f_a(b) = \int_{-\infty}^{\infty} x(t)\phi_a(t-b)/a \, dt, \quad a > 0.$$  \quad (1)

In this paper, the data of daily passenger flow can be regarded as a one-dimensional signal $S$ [9]. After one-level decomposition, the signal $S$ is divided into two parts. One is the approximate coefficient $cA_1$ obtained by low-pass filter, which describes the basic features of the signal, the other is the detail coefficient $cD_1$ obtained by high-pass filter, which describes the detailed features. Next, $cA_1$ is regarded as a new signal to be decomposed, the signal will be decomposed into a sequence $cA_1, cD_1, \ldots, cD_n$ with the decomposition of $n$ layers, which separates the basic feature from the detail part. Finally, the denoising data will be reconstructed by the low-frequency coefficient $cA_n$.

2.2 LSTM

Neural networks are suitable for non-linear prediction, but the traditional network models have no function for historical memory, while LSTM can learn the long-term dependency information and selectively forget invalid information and update effective memory information [3].

LSTM has an internal mechanism called gate that can regulate the flow of information, which can memorize temporary information by adding a structure called memory unit. LSTM controls the use of historical information by using three kinds of gate structures, they are update gate, forget gate and output gate. The update gate is used to update the state of the units. Firstly, the hidden state $a^{t-1}_{t-1}$ of the previous layer and the current input $x^{t-o}$ are passed to the $\sigma$ function to decide what needs to be updated. At the same time, $a^{t-o}_{t-o}$ and $x^{t-o}$ are passed to tanh function to get the temporary memory unit $c^{t-o}$ of the current moment, so as to take the value between -1 and 1 to help adjust the network. The function of forget gate is to control the forgetting of the previous layer with a certain probability. The hidden state $a^{t}_{t-1}$ of the previous layer and the current input $x^{t-o}$ are passed to the $\sigma$ function, its value is between 0 and 1. When this value is close to 0, it means forgetting the state of the previous layer, and when the value is close to 1, it means keeping the state of the previous layer. Next, the update gate and the forget gate are combined by the temporary memory unit $c^{t-o}$ and the memory unit $c^{t-1}_{t-1}$ of the previous layer to obtain the current memory unit $c^{t-o}$. Finally, the current hidden state $a^{t-o}$ is determined by the output gate and the current memory unit $c^{t-o}$ [2,5,8]. The process can be expressed by equation (2).
\[
\tilde{c}^{t+1} = \tanh(W_c [a^{t-\delta} \times c^t] + b_c), \quad \Gamma_u = \sigma(W_u [a^{t-\delta} \times c^t] + b_u), \quad \Gamma_f = \sigma(W_f [a^{t-\delta} \times c^t] + b_f),
\]
\[
\Gamma_o = \sigma(W_o [a^{t-\delta} \times c^t] + b_o), \quad c^{t+1} = \Gamma_u \tilde{c}^{t+1} + \Gamma_f c^t, \quad a^{t+1} = \Gamma_o \sigma c^{t+1},
\]

where, \( \Gamma_u, \Gamma_f \) and \( \Gamma_o \) are update gate, forget gate and output gate, respectively. \( t \) represents the moment or position of the current data. \( l \) represents the time lag, which means that the state of the current moment depends on the state of being at a time in the past. \( W \) and \( b \) are parameter matrices, \( \sigma \) is a sigmoid function.

2.3 GRU

GRU is a variant of the LSTM, it is a simpler network than the LSTM [10]. There are only two kinds of gates in the GRU, namely the update gate and the relevance gate. The function of the update gate of the GRU is similar to the combination of the update gate and the forget gate in the LSTM, which determines how much information from the previous moment is carried into the current state. The larger value of the update gate, the more information of the previous moment is brought in. Relevance gate \( \Gamma_r \) are used to control the degree of ignoring the information of the previous moment. The smaller value of the relevant gate, the more information is ignored [8]. The process can be expressed by equation (3).

\[
\tilde{c}^{t+1} = \tanh(W_c [\Gamma_r \times a^{t-\delta} \times c^t] + b_c), \quad \Gamma_u = \sigma(W_u [a^{t-\delta} \times c^t] + b_u), \quad \Gamma_o = \sigma(W_o [a^{t-\delta} \times c^t] + b_o).
\]

3. Recurrent neural networks and wavelet transform

In this paper, the combination of wavelet transform and RNNs is used to predict the daily passenger flow. Firstly, the passenger flow data is decomposed into low-frequency and high-frequency components by using wavelet transform. Then, we reconstruct the low-frequency coefficient to obtain new data which is smooth. Finally, LSTM and GRU are used to predict passenger flow. We have carried out experiments on them and compared the experimental results in Section 4.

Due to passenger flow data is one-dimensional, it is easy to produce over-fitting when using deep network structure to learn the features of data. Therefore, the LSTM and GRU used in this paper contain an input layer, a hidden layer and an output layer, the number of neurons in the hidden layer is 64. A random inactivation unit is added between the hidden layer and the output layer, its parameter is 0.2.

People's work and rest are carried out in a cycle of one week in daily life, the regularity of their activities must be reflected in the change of daily passenger flow in a week. Therefore, the time lag \( l \) in the model is set to 7, which preserves the dependence between data in the short term [3]. The network structure and parameters of the LSTM and GRU used in this paper are shown in figure 1 and figure 2.

4. Experiments

4.1 Data description & Evaluation indexes

In this paper, the daily passenger flow data of Nanjing metro will be analyzed. Considering that the opening of new metro lines will have an impact on the total passenger flow, the time period selected in
this paper is from May 26, 2018 to September 9, 2019. During this period, all the operating metro lines in Nanjing have been opened. In order to eliminate the time dependence of the data, the data from July 15, 2018 to September 9, 2019 is selected as the training set, and the rest is used as the test set.

In our experiment, the mean absolute percent error (MAPE) [3] and root mean squared error (RMSE) [5] are selected as the evaluation indexes, as shown in equation (4). MAPE represents the average of the percentage of the absolute value of the actual deviation to the observed value. RMSE represents the square root of the ratio of the square deviation to the number of observations.

\[ MAPE(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% , \quad RMSE(y, \hat{y}) = \left( \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \right)^{1/2}. \] (4)

### 4.2 Passenger flow forecast

In order to eliminate the influence of weather conditions and major events on the data, this paper uses wavelet transform to process data, as shown in figure 3. Figure 3 depicts a broken-line chart of passenger flow data. The blue line with a circle corresponds to the original data, and the yellow line with a five-pointed star corresponds to the data reconstructed by the low-frequency coefficient with db9 function. Figure 4 and figure 5 are the predictive curves of the test set. Comparing the two figures, it can be seen that the results of using processed data is better than that of using raw data. This means that the fluctuation of data has a great influence on the prediction results, and the wavelet transform is suitable to deal with this situation. Next, we will use the processed data to forecast the passenger flow.

![Figure 3. Data curves.](image1)

![Figure 4. Predicted curve of raw data.](image2)

![Figure 5. Predicted curve of new data.](image3)

In order to compare the effects of different wavelet functions, we have carried out experiments on them respectively, wavelet functions used in this paper are db2, db4, db9, coif3 and sym7. Table 1 shows the prediction results of LSTM and GRU with different wavelet functions and activation functions, it shows that the effect of db9 is the best. The MAPE scores of LSTM and GRU are 1.81% and 1.60%, and the RMSE scores are 7.10 and 6.31, respectively. The effects of different activation functions on neural networks are also compared. It can be seen from table 1 that the choice of the activation function has a significant impact on the prediction results, the activation function used in this paper is relu.

| Data Denoising | Wavelet Function | Activation Function | MAPE (LSTM/GRU) | RMSE (LSTM/GRU) |
|----------------|------------------|---------------------|-----------------|-----------------|
| No             | -                | relu                | 5.31%, 4.44%    | 20.61, 19.40    |
| Yes            | db2              | relu                | 2.92%, 4.26%    | 11.71, 15.49    |
| Yes            | db4              | relu                | 2.75%, 2.23%    | 10.05, 8.27     |
| Yes            | db9              | relu                | 1.81%, 1.60%    | 7.10, 6.31      |
| Yes            | db9              | sigmoid             | 2.06%, 1.71%    | 7.49, 6.53      |
| Yes            | db9              | tanh                | 2.00%, 1.68%    | 7.90, 6.31      |
| Yes            | coif3            | relu                | 1.87%, 1.62%    | 7.37, 6.46      |
| Yes            | sym7             | relu                | 3.65%, 1.90%    | 12.74, 7.98     |

After comparison and analysis of the above experiments, GRU and LSTM are suitable for passenger flow prediction, but the error of the prediction results of GRU network is smaller, the MAPE and RMSE scores are 1.60% and 6.31, respectively. It shows that a simple gated loop neural network can meet the needs of passenger flow forecast.
5. Conclusion
In this paper, the combination of RNNs and wavelet transform is employed to predict the passenger flow and the results show that the method can effectively improve the prediction accuracy. In the process of data denoising, wavelet decomposition and single-reconstruction are employed to remove noise. And then, the LSTM and the GRU are used to predict the passenger flow. The experimental results show that wavelet denoising can well reduce the data fluctuation, and LSTM and GRU can learn and memorize the information of data. However, the data set adopted in the experiment is too small and the influence of holiday and other factors is not considered in this paper. In the future research, these factors will be taken into account on a larger and more standard passenger flow data set.

Acknowledgments
This work was supported by Shandong Province Key Research and Development Planning Project (2015GGX101020), in part by Shandong Province Graduate Education Innovation Planning Project (SDYY16010), in part by the Project of Shandong Province Education Science in 12th Five-Year Plan (YBS15014), in part by the Project of Shandong Province Instructing Ability Improvement for Graduate Supervisor (SDYY17042).

References
[1] Kumar, S.V., Vanajakshi, L. (2015). Short-term traffic flow prediction using seasonal ARIMA model with limited input data. European Transport Research Review, 7: 1-9.
[2] Li, G. S., Peng, L., Li, X., Wu, T. (2019). Short-Term Traffic Forecast of Urban Bus Stations Based on Long Short-Term Memory. Journal of Highway and Transportation Research and Development (English Edition), 13: 65-72.
[3] Liu, Y., Qin, Y., Guo, J.Y., Cai, C.J., Wang, Y.G., Jia, L.M. (2018). Short-term forecasting of rail transit passenger flow based on long short-term memory neural network. In: 2018 International Conference on Intelligent Rail Transportation (ICIRT). Singapore. 1-5.
[4] Wu, Y.K., Tan, H.C. (2016). Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework. arXiv preprint, arXiv:1612.01022.
[5] Hao, S.Y., Lee, D. H., Zhao, D. (2019). Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system. Transportation Research Part C, 107: 287-300.
[6] Xiong, Z., Zheng, J.C., Song, D.J., Zhong, S.B., Huang, Q.Y. (2019). Passenger Flow Prediction of Urban Rail Transit Based on Deep Learning Methods. Smart Cities, 2: 371-387.
[7] Zhao, S. Z., Ni, T. H., Wang, Y., Gao, X. T. (2011). A new approach to the prediction of passenger flow in a transit system. Computers & Mathematics with Applications, 61: 1968-1974.
[8] Fu, R., Zhang, Z., Li, L. (2016). Using LSTM and GRU neural network methods for traffic flow prediction. In: 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). Wuhan. 324-328.
[9] Ouyang, L.W., Zhu, F.H., Xiong, G., Zhao, H.X., Wang, F.Y., Liu, T.Z. (2017). Short-term traffic flow forecasting based on wavelet transform and neural network. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). Yokohama. 1-6.
[10] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint, arXiv:1406.1078.