ATLANTIS PRESS
Advances in Intelligent Systems Research, volume 168
International Conference on Modeling, Analysis, Simulation Technologies and Applications (MASTA 2019)

Short-term Urban Rail Transit Passenger Flow Forecasting Based on Empirical Mode Decomposition and LSTM

Zi-ji-an WANG[1], Chao CHEN[1,*], Xiao-le LI[2] and Jing LI[1]

[1]School of Traffic and Transportation, Beijing Jiaotong University, Haidian District, Beijing, People's Republic of China

[2]School of Electronic and Information Engineering, Beijing Jiaotong University, Haidian District, Beijing, People's Republic of China

*Corresponding author

Keywords: EMD, LSTM, Urban rail transit, Regression-forecast extension.

Abstract. This paper proposed a method to forecast the short-term passenger flow, which is a vital component of urban rail transit system. We used a hybrid EMD-LSTM prediction model which combines empirical mode decomposition (EMD) and long short-term memory (LSTM) to forecast the short-term passenger flow in urban rail transit system. EMD can extract the variation trend of passenger flow, then LSTM can make the prediction to prove the accuracy. The experimental results indicate that the EMD-LSTM model used in this paper has better prediction accuracy than the LSTM model alone. Besides, the amount of data used in this experiment is small, and there is no need to consider additional features except temporal factor. According to what we have learned, this is the first time to combine EMD and LSTM to make short-term prediction in the urban rail transit system.

Introduction

Short-term passenger flow forecasting is a vital component of urban rail transit system. The forecasting results is an important basis for urban rail transit feasibility study and design, and also the main basis of project construction. In the recent studies, linear forecasting method and non-linear forecasting method are used. Grey System Theory and ARIMA are the represent of linear forecasting methods. LSTM [1], deep learning [2] and spatio-temporal deep learning [3] are the represent of non-linear forecasting methods.

Urban rail transit passenger flow has the characteristics of non-linear, periodicity and random, and it is inapplicability for short-term passenger flow forecasting. Moreover, some factors, like emergency, which affect passenger flow, are hard to acquire or forecast. So as to solve this problem, hybrid EMD-LSTM prediction model is used. Firstly, the passenger flow data of Beijing subway Line 10 is used, considering only the time characteristics of the data, then the hybrid EMD-LSTM prediction model is used. The EMD is used to decompose the original passenger flow data, and statistical method is used to select each component, then LSTM is used to predict each component separately. Finally, the prediction results of each component are added to the final result.

Methodology

Empirical Mode Decomposition

Empirical mode decomposition (EMD) [4] is a signal decomposition algorithm, which is suitable for non-liner and non-stationary signal. The original time series signal can be decomposed into a small number of oscillatory modes which can be expressed as some intrinsic modals functions (IMF) and a residue. The residue retains a non-periodic trend of the original signal, and any periodic fluctuation in original signal will be decomposed into IMFs. IMFs must satisfy the following two conditions [4]:

1. In the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one.
2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The essence of EMD algorithm is shifting IMFs from original signal. The EMD algorithm is described as follows [4,5]:

Step1: Find all the local extrema in signal $x(t)$.

Step2: Use cubic spline line to get the upper and lower envelopes, $e_{max}(t)$ and $e_{min}(t)$.

Step3: Generate the mean envelopes as below.

$$m_1(t) = e_{max}(t) + e_{min}(t))/2. \tag{1}$$

Step4: Calculate the difference between $x(t)$ and $m_1(t)$ as a proto-IMF,

$$h_1(t) = x(t) - m_1(t). \tag{2}$$

Step5: check whether $h_1(t)$ is an IMF. If $h_1(t)$ satisfy all the requirements of an IMF, $h_1(t)$ is denoted as the ith IMF$c_i(t)$, and replace $x(t)$ with residue $r_1(t)$.

$$r_1(t) = x(t) - h_1(t). \tag{3}$$

Otherwise, $h_1(t)$ is not an IMF, replace $x(t)$ with $h_1(t)$.

Step6: Repeat Steps 1–5. The sifting process stops when the residue meet one of the termination conditions.

By using the algorithm above, the original signal $x(t)$ can be decomposed into n modes and a residue as follows:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t). \tag{4}$$

In the formula, n is the number of IMFs, $c_i(t)$ represents IMFs, and $r_n(t)$ is the residue.

Urban rail passenger flow data can be equated to a time series signal, which can be deposed by EMD into a series of IMFs and a residue.

The number of IMFs decomposed from the original passenger flow signal n is determined by its own local characteristic time scale, and these IMFs components are independent of each other. More importantly, the trend information in the original passenger flow signal can be obtained in the residue.

EMD End Effect

The end effect, produced in EMD sifting process, can distort the decomposed sub-series and affect the following sub-series modeling process [6]. The edge effect is shown in the Fig. 1.

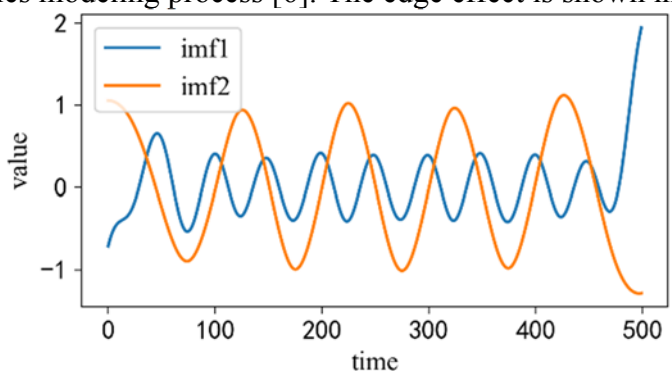

Figure 1. The edge effect

Fig. 1 shows two IMFs which have the obvious edge effect. Through observation, it can be found that the IMFs component does not perform well in the two edges, and has a drift.

In order to solve end effects, some scholars put forward Mirror method [7], Coughlin′s method [8], Slope-based method [9], and Rato′s method [10]. These methods extended the original signal, and then decomposed the original passenger flow signal by EMD, finally delete the extended part of the decomposed signal to remove the end effects. This paper proposes a regression-forecast extension method for the edge effect in EMD of urban rail transit passenger flow.

The historical data of the urban rail transit passenger flow can be obtained, but the future data cannot. Therefore, for the left edge of the original passenger flow signal, it can be extended by historical data, while for the right edge of the signal, a method is needed to predict the passenger flow

data of one cycle. This paper uses the data of the same period (for example, the data of all Mondays in the original passenger flow data) to predict the right edge of the signal by linear regression roughly. The advantage of using regression prediction is that the original general trend can be retained as much as possible and the operation is easy to deal with.

LSTM

Long short-term memory (LSTM) [11] is a special recurrent neural network (RNN) suitable for predicting continuous time series. Compared with traditional RNN, LSTM overcomes its long-term dependencies. For many tasks, the RNN with LSTM structure (shown in Fig. 2) performs better than the traditional RNN.

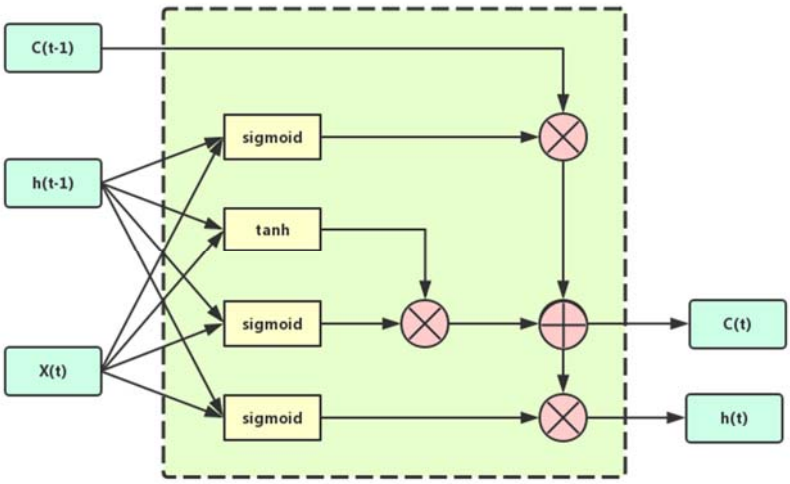

Figure 2. LSTM neuron structure

Fig. 2 shows the structure of LSTM. Compared with RNN, LSTM adds a structure called gate, including input gate, forgetting gate and output gate.

A gate is an operation using a sigmoid neural network and a bitwise multiplication, which together form a gate structure.

The forgetting gate is to allow the RNN to "forget" previously information which is useless. As shown in Eq. 5, the forgetting gate will determine which part of the memory needs to be forgotten based on the current input x_t and the output h_{t-1}at the last moment.

$$f = sigmoid(W_f[h_{t-1}, x_t]). \tag{5}$$

The input gate is used to replenish the current memory of the circulatory neural network after it has "forgotten" the previous state of the part. As shown in Eq. 6, the input gate will determine which information is added to the state c_{t-1} according to $x_t$and h_{t-1} to generate a new statec_t.

$$i = sigmoid(W_i[h_{t-1}, x_t]). \tag{6}$$

The output gate generates the output of the current moment after calculating the new state c_t. As shown in the Eq. 7, the output gate will determine the output h_t at this moment based on the latest statec_t, the output h_{t-1} at the previous moment and the current inputx_t.

$$o = sigmoid(W_o[h_{t-1}, x_t]). \tag{7}$$

A Hybrid EMD-LSTM Prediction Model

The model applied in this paper is hybrid EMD-LSTM prediction model, which is composed of EMD and LSTM. The EMD-LSTM model consists of four stages:

Stage1: Extension Stage

In this stage, the original passenger flow signal is extended in order to deal with the EMD end effect. The left edge was extended by historical data of one cycle, and the right edge was extended by regression prediction of one cycle too.

Stage2: EMD Stage

In this stage, the extended signal is decomposed into IMF components and residual components by EMD. Based on the difference of volatility and time span, the number of IMF components decomposed are different. Generally, the more violent the passenger flow signal is and the longer the time span is, the more the IMF components will be.

Stage3: Component identification Stage

In this stage, we use statistic method to identify the significant component. The IMF components can be judged by calculating the correlation with the original signal and determining whether the components are meaningful. Strong correlation signal could be taken as input individually, other signal aggregated as a part of input. The statistical methods, such as Pearson product moment correlation coefficient, and Kendall rank correlation coefficient are used to test the correlation.

Stage4: LSTM Stage

In this stage, LSTM method is applied to predict each component after screening. In this paper, the prediction only relies on the component itself as the input of LSTM for prediction, without the need for other features in the prediction process.

LSTM neural network needs to determine the three main parameters of the in-put layer, the hidden layer and the output layer. The hidden neurons is deter-mined by experiments. The neurons in the input layer and the output layer are based on the rolling step and the forecasting step. Since we use historical data to forecast next day's data, the input neurons can choose a certain rolling step according to the data set, such as three or five days. The output neurons is selected as one day according to the prediction needs.

It should be noted that the number of neurons in the input layer and the hidden layer is not proportional to the prediction accuracy. Meanwhile, as the number of neurons increases, the computational efficiency decreases. Therefore, it is necessary to determine the input neurons and hidden neurons through repeated experiments by using RMSE as the selected indicator.

Experiment

The Data Set

The data set was collected from the daily passenger flow of Beijing subway line 10 between February 18, 2019 and March 31, 2019, a total of 42 days, excluding holidays.

The daily passenger flow of urban rail transit shows a cyclical change with a period of 7 days, which is generally manifested as a large passenger flow in week-days and a small passenger flow in weekends. According to the cyclical of urban rail transit passenger flow, the data are evenly divided into groups by week.

Group 1 is the extension group to the left side of the original passenger flow signal, groups 2 to 4 serves as training groups for LSTM neural networks, and groups 5 and 6 are used as the contrast groups.

Modeling

First, according to the EMD-LSTM model, extend the original passenger flow signal. The extension results are shown in Fig. 3:

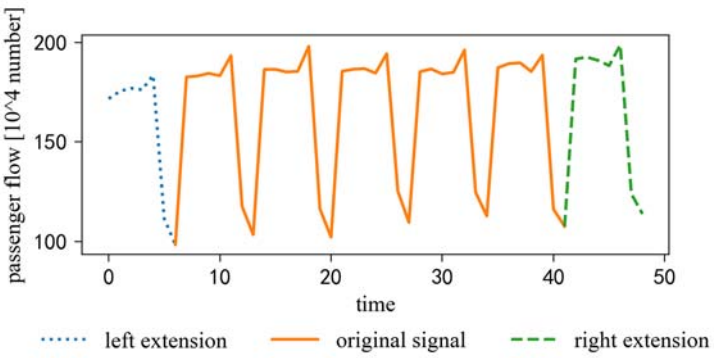

Figure 3. The extension results of original passenger flow signal

Secondly, decompose the extended signals by EMD, and the decomposition results are as follows:

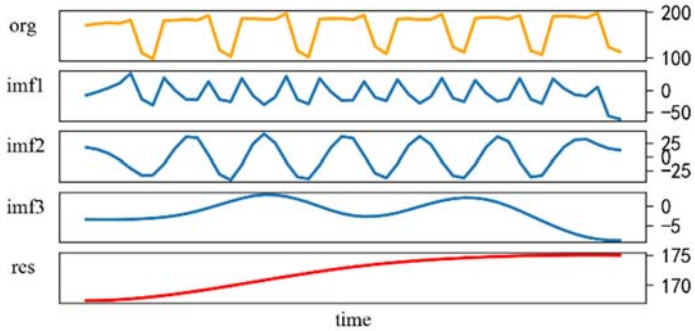

Figure 4. The extracted IMF components

EMD decomposed the original passenger flow signal into three IMFs components and one residue component. The three IMFs components are clearly cyclical and the residue component retains the trend of the original sequence.

Then, perform the correlation test using the Pearson product moment correlation coefficient, and the Kendall rank correlation coefficient. And the results are shown in the following tables.

Table 1. Pearson product moment correlation coefficient

Component	imf1	imf2	imf3	res	original signal
imf1	1.00	-0.12	0.00	-0.09	0.54
imf2		1.00	0.02	-0.12	0.77
imf3			1.00	-0.01	0.07
res				1.00	-0.10
original signal					1.00

Table 2. Kendall rank correlation coefficient

Component	imf1	imf2	imf3	res	original signal
imf1	1.00	-0.11	-0.02	-0.09	0.49
imf2		1.00	0.04	-0.07	0.35
imf3			1.00	0.00	0.15
res				1.00	0.09
original signal					1.00

According to the results of correlation test, the correlation between IMF components is very low. Pearson product moment correlation of imf1 and imf2 are 0.54 and 0.77, respectively, which indicates a strong positive correlation, and Kendall rank correlation coefficients of mf1 and imf2 are 0.49 and 0.35, respectively, which indicate imf1 and imf2 have higher ranks. The result of Pearson and Kendall correlation coefficients are consistent. Therefore, imf1 and imf2 are as separate inputs calling c_1 and c_2, imf3 and residue are aggregated as an input c_3.

Finally, LSTM is applied for prediction. The rolling step of each component and the number of hidden neurons were determined through experiments.

After repeated experiments, the optimal application scheme was determined by comparing RMSE of each input component on the training set. The best solution is shown in Table 3.

Table 3. Values of rolling step and hidden neurons

Component	Rolling horizon[number]	Hidden layer neurons[number]
c1(imf1)	7	5
c2(imf2)	4	6
c3(imf3+res)	7	7

Then we use LSTM to predict each component, and the final prediction is obtained by adding up the predicted results of each component. The final prediction effect is shown in the Fig. 5:

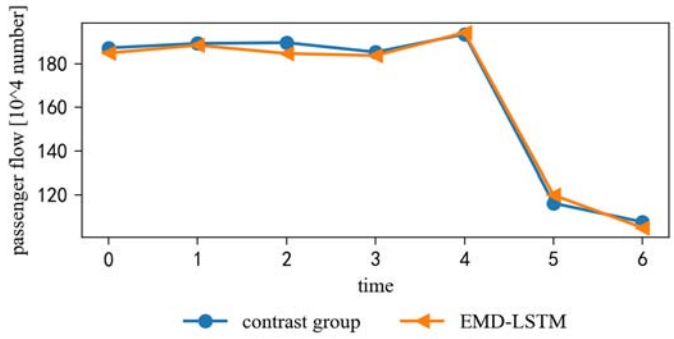

Figure 5. Predicted results

Results and Analysis

The first group data in contrast group is used to satisfy the rolling step, and the second group is used to contrast with the prediction value. Table 4 shows the true value contrast with prediction value of EMD-LSTM and prediction results of LSTM.

Table 4. The true value and prediction value of EMD-LSTM and LSTM

Date	Contrast group [10^4number]	EMD+LSTM [10^4number]	Relative error	LSTM [10^4number]	Relative error
3.25	187.32	184.97	-1.26%	186.98	-0.18%
3.26	189.37	188.53	-0.44%	189.31	-0.03%
3.27	189.74	184.75	-2.63%	186.01	-1.97%
3.28	185.40	183.80	-0.86%	186.61	0.65%
3.29	193.61	194.36	0.39%	195.97	1.22%
3.30	116.19	119.94	3.22%	128.99	11.02%
3.31	107.59	104.93	-2.47%	114.08	6.03%

The RMSE value of EMD-LSTM and LSTM are 1.06 and 2.15 respectively. According to Table 1.4 and RMSE, we can find that the prediction results of hybrid EMD-LSTM prediction model is better than that of LSTM alone. At the same time, the EMD-LSTM model can achieve higher prediction accuracy with less historical data, and only requires the passenger data sequence itself, without other features (such as weather, climate, etc.)

At present, the model is only suitable for using the previous data to predict next day's daily passenger flow in the future. If there is a need to predict the daily passenger flow in the next few days, there are two ways to do this: First, increase the number of LSTM output neurons to increases the predictable duration.

Secondly, use iteration prediction. The predicted passenger flow data can be used as historical data and repeat the iteration. According to the previous experience, the error of these methods will increase with the increase of the prediction period. This article does not carry on the thorough research to this kind of situation.

At the same time, if there are holidays' data in the historical data applied, they should be revised before making the prediction. If there are holidays during the predicted period, this method will no longer be applicable.

Summary

In this paper, EMD method is applied to decompose the original passenger flow signal of urban rail transit into IMFs components, then LSTM is applied to predict each component separately. The experimental results show that the short-term passenger flow predicted by hybrid EMD-LSTM prediction model is more accurate than that predicted by LSTM neural network directly. Meanwhile, the regression-forecast extension method proposed in this paper can effectively rectify the edge effect of EMD and improve the prediction accuracy. Another advantage of the EMD-LSTM model is that it can use a small amount of historical passenger flow data to achieve a high prediction accuracy without other features.

Acknowledgement

This research was financially supported by the National Key Research and Development Program of China (Project No. 2018YFB1201402).

References

[1] Zhang, Yong, J. Zhu, and J. Zhang. "Short-term Passenger Flow Forecasting Based on Phase Space Reconstruction and LSTM." International Conference on Electrical & Information Technologies for Rail Transportation Springer, Singapore, 2017.

[2] Polson, Nicholas G, and V. O. Sokolov. "Deep learning for short-term traffic flow prediction." Transportation Research Part C: Emerging Technologies 79(2017):1-17.

[3] Ke, Jintao, et al. "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach". Transportation Research Part C: Emerging Technologies 85(2017):591-608.

[4] Huang, N. E., et al. "The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis." Proceedings A 454.1971(1998):903-995.

[5] Wei, Yu, and M. C. Chen. "Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks." Transportation Research Part C: Emerging Technologies 21.1(2012):148-162.

[6] Xiong, Tao, Y. Bao, and Z. Hu. "Does Restraining End Effect Matter in EMD-Based Modeling Framework for Time Series Prediction? Some Experimental Evidences." Neuro-computing 123(2014):174-184.

[7] G. Rilling, P. Flandrin, and P. Gonçalvés. On empirical mode decomposition and its algorithms, IEEE-EURASIP Workshop on Nonlinear signal and Image Processing, 2003, pp. 1–5.

[8] Coughlin, K. T, and K. K. Tung. "11-Year solar cycle in the stratosphere extracted by the empirical mode decomposition method." Advances in Space Research 34.2(2004):323-329.

[9] Datig, Marcus, and T. Schlurmann. "Performance and limitations of the Hilbert–Huang transformation (HHT) with an application to irregular water waves." Ocean Engineering 31.14-15(2004):1783-1834.

[10] Rato, R. T., M.D. Ortigueira, and A. G. Batista. "On the HHT, its problems, and some so-lutions." Mechanical Systems & Signal Processing 22.6(2008):1374-1394.

[11] Hochreiter, Sepp, and Schmidhuber, Jürgen. "Long Short-Term Memory." Neural Com-putation 9.8(1997):1735-1780.