<u>Original Research Article</u>

Research on urban road traffic flow prediction based on wavelet denoising and multi-layer perceptron

.

ABSTRACT

Aims: Develop a novel traffic flow prediction model to improve the accuracy of traffic flow prediction, better assist intelligent traffic management, improve traffic efficiency, reduce traffic congestion, and thus better improve sanitation and quality of life.
Study design: Develop an urban road traffic flow prediction model with strong predictive power and excellent stability.
Place and Duration of Study: Southwest University of Science and Technology, between July 2021 and March 2022.
Methodology: Adopting wavelet threshold to denoise, first decompose the original data, then perform noise filtering on the subsequences obtained after decomposing, and finally reconstruct the denoised data. Use denoised data to train a multilayer perceptron and make predictions on future data. At the same time, several representative models are selected to compare with the proposed model to verify the competitiveness of the proposed model.
Results: The proposed model has the smallest prediction error in the two training sets with different temporal granularity. In addition, we are using the data after wavelet denoising for training and prediction results in a smaller prediction error than using the data without denoising.
Conclusion: The proposed prediction model has strong prediction ability and generalization performance in the field of traffic flow prediction. The wavelet denoising method can effectively improve the prediction accuracy of traffic flow prediction.

Keywords: {Traffic flow, wavelet threshold denoising, multi-layer perceptron, multi-step forecasting}

1. INTRODUCTION

With the development of the social economy, people's living standards are improving day by day, and the number of private cars is becoming more and more. The increase in the number of traveling vehicles has made the situation of road congestion more and more serious. On the one hand, traffic congestion will cause severe pollution to the environment, and on the other hand, it will greatly increase the incidence of traffic accidents and threaten people's lives and health. Traffic flow forecasting plays a fundamental role in the planning and developing traffic management and control systems. In recent years, traffic forecasting has become an essential task in the field of intelligent transportation[1]. Improving the accuracy of traffic flow prediction is conducive to intelligent traffic management to improve traffic efficiency and reduce congestion. However, traffic flow data is often highly volatile, and there is a lot of data noise. So far, traffic flow prediction is still a classic but challenging problem[2].

For a long time, a large number of researchers have been pouring into the field of traffic flow prediction, and achieved a large number of satisfactory results. Shengnan Guo et al. proposed a spatiotemporal-based convolutional network (ASTGCN) to solve the traffic flow prediction problem and tested coefficients in Caltrans performance. Finally, experiments on two real-world datasets from the Caltrans Performance Measurement System (PeMS)demonstrate that the proposed ASTGCN model outperforms the state-of-the-art baselines[3]. Jiawei Cao et al. applied Extreme Gradient Boosting (XGBoost) to short-term traffic flow prediction; the experiment results reveal the model's superiority by comparing it with the traditional prediction model[4]. Li Yuanyuan and Xu Weixiang used the improved KNN algorithm to reconstruct the phase space, wait until the input data, and then use the SVR model to predict the data. The best experimental results show that the model has better prediction performance[5]. JunShan Tian et al. proposed a random forest algorithm to train and learn data features to predict highway traffic flow. The results show that the proposed model has better prediction accuracy than traditional models like linear regression and GBDT[6]. Ximu Zeng et al. combined the prediction results of XGBoost, LightGBM, and CatBoost models through an ensemble strategy. The results show that the proposed model is more accurate and suitable for the short-term than single models such as ARIMA, LSTM, XGBoost, LightGBM, and CatBoost. Traffic state prediction[7]. Diogo David Oliveira et al. proposed an artificial neural network (ANN)-based vehicle traffic flow prediction model. They applied it to the traffic flow prediction of Interstate 87 in New York, USA. The results show that the model can be replicated in the same area. Or other roads in other states or countries[8]. Weihong Cai et al. proposed a PSO-ELM model based on particle swarm optimization for short-term traffic flow prediction, aiming at the nonlinear relationship affecting traffic flow prediction. The experimental results show that the performance of this model is significantly better than that of other comparative models[9].

From the literature review results, it can be seen that most scholars have devoted themselves to propose a model with solid predictive ability in recent years. Still, only a few scholars have improved the accuracy of prediction by correlating the data. Improving the model's prediction accuracy through data preprocessing is the core part of the data mining task, which can effectively increase the prediction results' accuracy. Therefore, we use wavelet threshold denoising in this study to preprocess the data and apply the denoised data to traffic flow prediction.

The rest of this article is as follows: Section.2 mainly introduces the basic principles of wavelet threshold denoising and multilayer perceptron and elaborates the technical means of this research. Section.3 presents and analyzes the experimental results of this paper and discusses the model and prediction method in detail. The main conclusions of this paper are given in Section.4. Section.5 presents some of the existing problems in this paper and proposes the next step of this research.

2. METHODOLOGY

2.1 Wavelet domain denoising(WD)

The wavelet domain denoising method was first proposed in 1992 by Professor Johnstone and Donoho[10]. It is a denoising method that can achieve approximately minimum mean square error and has the characteristics of the most straightforward implementation and the smallest amount of calculation.

The basic principle is: Orthogonal wavelet decomposition has the ability of time-frequency local deterioration. During signal processing, the amplitude of the useful wavelet component is large, which is in sharp contrast to the uniform realization of the noise in the high-frequency part. After wavelet decomposition, most of the coefficients with larger amplitudes are useful signals, while those with smaller amplitudes are generally noises. The wavelet transform coefficients of useful signals are larger than those of noise. The threshold denoising method is to find a suitable threshold, keep the wavelet coefficients larger than the threshold, process the wavelet coefficients smaller than the threshold accordingly, and restore the useful signal according to the processed wavelet coefficients.

Define:

$$F(t) = s(t) + e(t) \tag{1}$$

Among them, $s(t)$ represents the valuable signal, and $e(t)$ is the noise part. Do wavelet transform on both sides of Equation 1 at the same time:

$$WT_f(a,b) = WT_s(a,b) + WT_e(a,b) \tag{2}$$

Equation 2 shows that the wavelet transform of the actual measurement signal is equal to the sum of the wavelet transforms of multiple signals.

For Equation 1, after orthogonal wavelet transform, the correlation of the signal $F(t)$ can be removed to the greatest extent, and most of the energy is concentrated on a few wavelet coefficients with relatively large amplitudes. And the noise $e(t)$ will be distributed on all-time axes under each scale after wavelet transformation, and the amplitude is not very large. In each scale of the wavelet transform, the wavelet coefficients of the noise are reduced to the greatest extent, and then the processed wavelet coefficients are used to reconstruct the signal so that the noise can be suppressed.

2.2 Multi-Layer Perceptron

The multilayer perceptron (MLP) is the improvement and perfection of the feedforward neural network[11]. It mainly consists of an input layer, an output layer, and a hidden layer. Figure 1 shows a schematic diagram of a three-layer multilayer perceptron. Each layer in MLP has its fixed task, and the input layer is mainly responsible for receiving input samples. The hidden layer is the core part of the entire MLP, primarily processing and processing the input data. The output layer mainly performs activation processing and operations according to the required tasks such as prediction and classification.

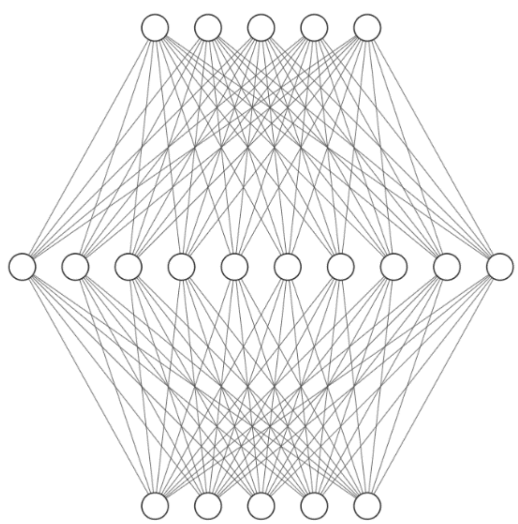


Figure 1 Schematic diagram of three-layer neural network

The following computations are performed on each neuron in the output and hidden layers:

$$O(x) = G(b(2) + W(x)h(x)) \tag{3}$$

$$h(x) = \Phi(x) = s(b(1) + W(1)x) \tag{4}$$

Where $b(1)$, $b(2)$ represent bias vectors, $W(1)$ and $W(2)$ represent weight matrices, and G and S are activation functions. It is worth mentioning that $W(1)$, $b(1)$, $W(2)$, $b(2)$ are parameters that need to be optimized. There are many choices of activation functions, as shown in Table 1:

Table 1 Several common activation functions

Function	Formula	Derivative
ReLU (Rectified linear unit)	$f(x) = \max(x, 0)$	$f(x)' = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$
sigmoide	$sigmoid(x) = \frac{1}{1 + e^{-x}}$	$sigmoid(x)' = \frac{e^{-x}}{(1 + e^{-x})^2}$
tanh	$tanh(x) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	$tanh'(x) = 1 - \left(\frac{e^z - e^{-z}}{e^z + e^{-z}}\right)^2$

2.3 Experiment

2.3.1 Data description

This study uses data derived from the OpenITS platform[12]. Whitemud Drive is an urban highway that runs through Edmonton, Alberta, Canada, with a total length of 28 kilometers. This study is based on the traffic flow of Whitemud Drive.

The row data is sorted into two granularities of 10 minutes and 30 minutes, and some statistical properties of the data set are analyzed. The results are shown in Table 2.

Table 2 Several features of the dataset

Time	Mean	Median	Std	Max	Min	Numsbers
10 minutes	835.86	924.00	552.89	1998.00	24.00	136
30 minutes	779.06	812.22	517.24	1924.00	42.00	141

What's more, to more intuitively understand the data characteristics and changing trends of the data set, the data trend graphs of the two data are displayed, as shown in Figure 2.

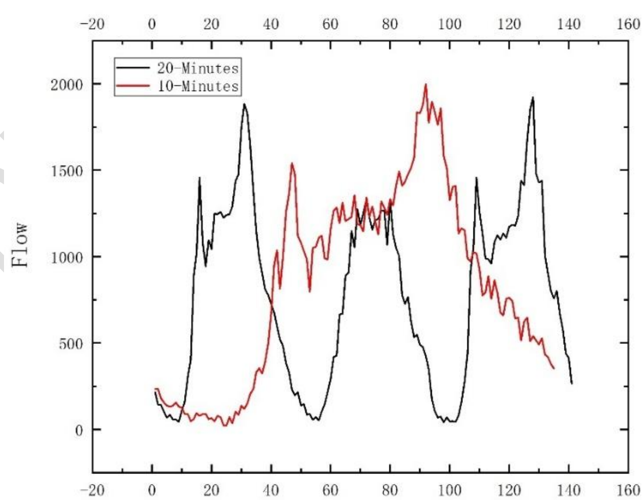


Figure 2 Data detailed trend graph

From the figure, we can observe that the traffic flow data of this expressway has several very prominent features: a) The data has substantial volatility. b) The data has a certain periodicity. c) There is a lot of data noise in the data.

2.3.2 Establishment of WD-MLP Hybrid Model

In the previous section, we learned that there is a lot of noise in the traffic flow data, and the data is very volatile. Therefore, we denoise the data through wavelet decomposition and reconstruction and input the denoised data into the MLP.

The main idea is divided into the following four steps:
Step 1: Initialize the wavelet parameters, and determine the type of wavelet to be used.
Step 2: Perform wavelet decomposition on the original sequence, perform noise filtering, and finally reconstruct the denoised data.
Step 3: Divide the denoising data and use the data to train MLP.
Step 4: Compare and analyze the results of WD-MLP and other models.

The detailed establishment process of the hybrid model is shown in Figure 3

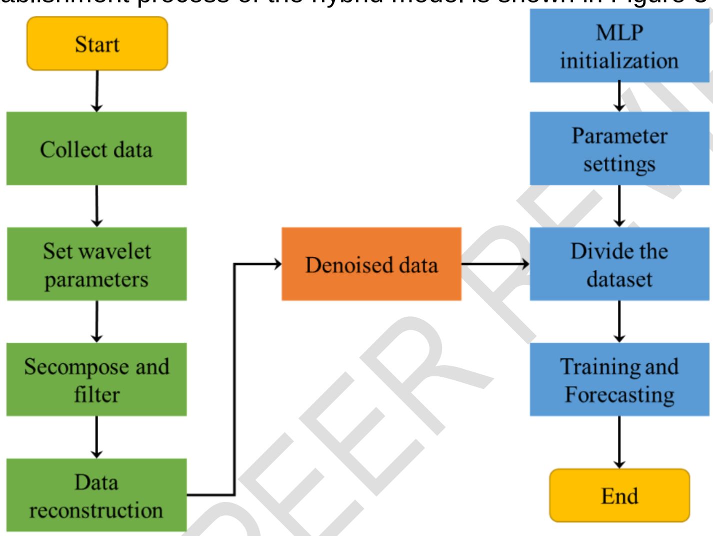


Figure 3 The construction process of the hybrid model

2.3.2 Evaluation metrics

To effectively test the predictive performance of the proposed model, in this study we introduce mean absolute percentage error (MAPE) to evaluate the prediction results of the model, and its calculation formula is shown in Equation 5. Where N is the number of samples, and $\hat{y}_i$ and y_i are the predicted and actual values, respectively.. A MAPE less than 10% indicates an excellent model, a MAPE greater than 20% indicates an inferior model, and the smaller the MAPE, the better the model's prediction performance.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \quad (5)$$

More importantly, to verify the model's competitiveness, we introduced five other machine learning models to compare with the proposed model. All comparative models are listed in Table 3.

Table 3 Models for comparison

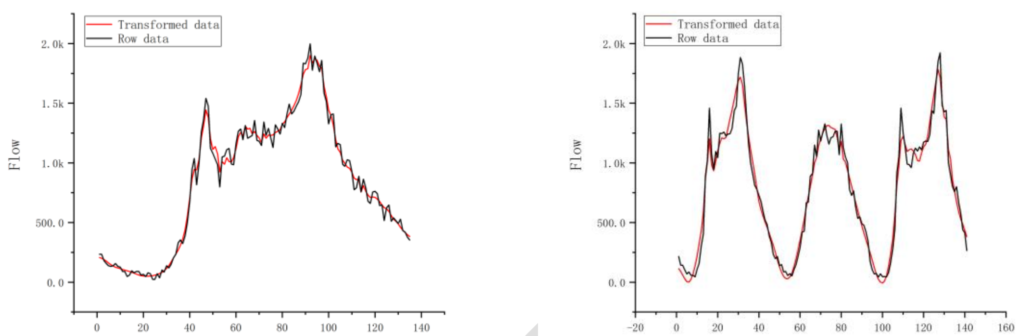
Model	Abbreviation	Literature	Invention time
Extreme gradient boosting	XGBoost	[13]	2016
Random forest	RF	[14]	2001
Light gradient oosting machine	LightGBM	[15]	2014
Least square support vector machine	LSSVM	[16]	1999
Linear regression	LR		1805

3. RESULTS AND DISCUSSION

3.1 Wavelet denoising result

In the above chapters, we have discussed the relevant principles of wavelet threshold denoising and the method if wavelet is used. Therefore, this section, analyzes and displays the results obtained by wavelet denoising.

On two datasets with different temporal granularity, we use the wavelet threshold denoising method to denoise them. The Daubechies8 wavelet is used in the denoising operation, and the threshold is set to 0.5, and the denoised result is shown in Figure 4.



(a) 10 minutes (b) 30 minutes

Figure 4 Wavelet denoising result graph

The figure can intuitively observe that after denoising, our dataset becomes significantly more stable and smooth.that after denoising, our dataset becomes significantly more stable and smooth. Compared with the original data, the periodicity of the data still exists, but there is no very drastic fluctuation.

3.2 Forecasting result

This section applies the proposed and comparative models to both datasets separately. On a dataset with a time granularity of 10 minutes, the prediction step represents the prediction of road traffic flow 10 minutes into the future. On a dataset with a time granularity of 30 minutes, the prediction step represents the prediction of road traffic flow 30 minutes into the future. This study predicted road traffic flow 50 and 150 minutes into the future, respectively. The detailed prediction results are shown in Table 4.

Table 4 **MAPE(%) of prediction results for both datasets**

Dataset	Model / Step	WD-MLP	MLP	XGBoost	RF	LightGBM	LSSVR	LR
10 minutes	1-step	5.015	8.647	48.418	43.941	36.956	9.554	9.403
	2-step	5.487	8.977	49.899	46.434	44.248	8.393	12.575
	3-step	5.923	8.771	50.932	47.874	46.971	9.406	14.939
	4-step	6.504	9.139	52.098	49.144	49.094	13.679	17.866
	5-step	7.291	9.255	53.595	50.575	51.1381	15.926	20.658
30 minutes	1-step	9.166	9.442	15.893	13.438	14.814	9.873	17.108
	2-step	9.606	10.684	17.531	16.161	18.111	13.058	22.007

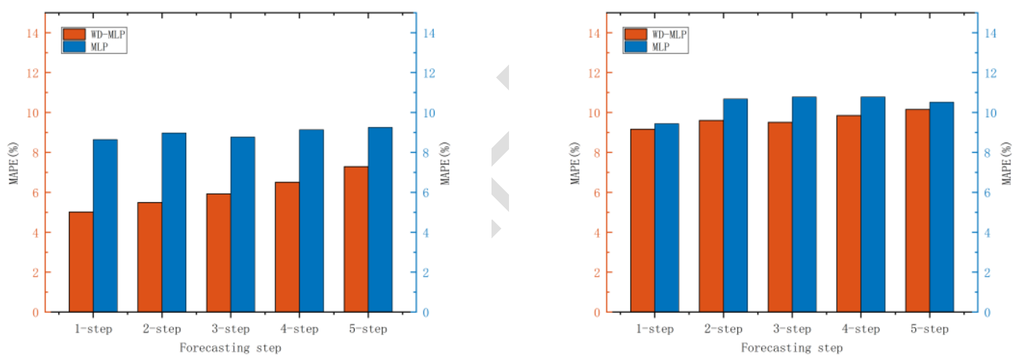
	3-step	9.508	10.784	20.832	19.166	23.894	17.142	27.327
	4-step	9.848	10.783	25.621	23.239	30.921	23.776	33.755
	5-step	10.156	10.522	33.626	29.071	40.594	30.155	39.338

From the table, we can see that in the dataset with the time granularity of 10 minutes, WD-MLP has an absolute advantage. The MAPE of WD-MLP is the smallest among all comparison models in one to five-step prediction. These conclusions also hold for the dataset with a time granularity of 30 minutes.

It is worth mentioning that from Figure 2 and Table 2, we can see that datasets with different time granularities have different statistical characteristics and different data trends. But the proposed model has the best performance in both datasets, which is enough to show that the proposed model has excellent stability and generalization in addition to good prediction performance.

3.3 Discussion

This section mainly discusses the improvement effect of the wavelet threshold denoising method on the model. In all comparative experiments, we discuss the prediction results of both WD-MLP and MLP models separately. Figure 5 shows the prediction results of the WD-MLP and MLP models on the two datasets, respectively.



(b) 10 minutes (b) 30 minutes

Figure 5 Comparison of prediction results of WD-MLP and MLP

From the figure, we can see that the wavelet threshold denoising has vastly improved the model. In the data set with a time granularity of 10 minutes, WD-MLP is 2.29% smaller than the total average MAPE of MLP's five-step prediction. In the dataset with a time granularity of 30 minutes, WD-MLP is 0.79% smaller than MLP. According to the summary of the results on the two datasets, we can conclude that the wavelet threshold denoising is very effective in improving the model's prediction performance.

4. CONCLUSION

Improving the accuracy of traffic flow prediction is the most critical part of an intelligent transportation system. Accurate traffic flow prediction can effectively prevent traffic jams, reduce traffic accidents, and improve sanitation. In this study, we combine wavelet threshold denoising and MLP to propose a WD-MLP hybrid model and apply it to AC traffic flow prediction and achieve very satisfactory results.

Through numerical simulation experiments[17], we can draw the following conclusions: a) The prediction performance of WD-MLP is better than other competitors on all prediction steps, which shows that the hybrid model has strong prediction ability. b) On two datasets with different time granularities, the WD-MLP model has absolute advantages, which indicates that the hybrid model has better stability and generalization performance. c) Comparing the two models of WD-MLP and MLP, WD-MLP always has higher prediction accuracy, which shows that wavelet threshold denoising can effectively increase the accuracy of prediction results in traffic flow prediction.

5. OUTLOOK

IThis study combines wavelet theory with a multilayer perceptron model to propose a novel prediction model and apply it to traffic flow prediction. However, it is worth noting that although the multilayer perceptron model has excellent prediction performance, its stability is poor, and the parameters have a significant influence on the model. Therefore, in the follow-up work, we will introduce a new parameter tuning method to optimize the hyperparameters of the model. In addition, in the research, we also found that the generalization ability of a single model has limitations, so in the follow-up work, we can consider using a combined prediction method to improve further the prediction accuracy and generalization performance through the method of model fusion.

REFERENCES

1. Lippi, M., Bertini, M., & Frasconi, P. (2013). Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *IEEE Transactions on Intelligent Transportation Systems*, *14*(2), 871-882.

2. Yang, H. F., Dillon, T. S., & Chen, Y. P. P. (2016). Optimized structure of the traffic flow forecasting model with a deep learning approach. *IEEE transactions on neural networks and learning systems*, *28*(10), 2371-2381.

3. Guo, Shengnan, et al. "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 33. No. 01. 2019.

4. Cao, Jiawei, et al. "Short-Term Highway Traffic Flow Forecasting Based on XGBoost." *2020 15th International Conference on Computer Science & Education (ICCSE)*. IEEE, 2020.

5. Yuanyuan, Li, and Xu Weixiang. "Short-term traffic flow forecasting based on SVR." *Advances in Engineering Research* 166 (2018): 57-61.

6. Tian, JunShan, et al. "Expressway Traffic Flow Forecasting based on SF-RF Model via ETC Data." International Conference on Frontiers of Electronics, Information and Computation Technologies. 2021.

7. Zeng, Ximu, et al. "Short-Term Traffic Flow Prediction Based on Ensemble Machine Learning Strategies." *2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS)*. IEEE, 2021.

8. Oliveira, Diogo David, et al. "Forecasting vehicular traffic flow using MLP and LSTM." *Neural Computing and Applications* 33.24 (2021): 17245-17256.

9. Cai, Weihong, et al. "PSO-ELM: A hybrid learning model for short-term traffic flow forecasting." IEEE access 8 (2020): 6505-6514.

10. Donoho, David L., and Jain M. Johnstone. "Ideal spatial adaptation by wavelet shrinkage." biometrika 81.3 (1994): 425-455.

11. Raj, Pethuru, and Preetha Evangeline. *The Digital Twin Paradigm for Smarter Systems and Environments: The Industry Use Cases*. Academic Press, 2020.

12. Peng,L. &OpenITS Org. OpenData V7.0-Canada Whitemud Drive highway data https://www.openits.cn/openData1/700.jhtml (2021). Accessed: 2021-12-12

13. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
14. Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.
15. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, *30*.
16. Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, *9*(3), 293-300.
17. Mahdy, A. M. S., et al. "Numerical solution and dynamical behaviors for solving fractional nonlinear rubella ailment disease model." *Results in Physics* 24 (2021): 104091.