Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/75/20240544

Traffic flow prediction based on nonlinear weight decreasing PSO-SVR univariate time series prediction algorithm

Hongyi Li

College of traffic & transportation, Chongqing Jiaotong University, Chongqing, 91803, 400000

Corresponding author's e-mail: lihongyi9999@gmail.com

Abstract. The aim of this paper is to investigate the application of car flow prediction in the field of transport in order to solve the problem of urban traffic congestion. For this purpose, we adopt the nonlinear weight decreasing PSO-SVR univariate time series prediction algorithm to predict the car flow, and divide the data set into training set and test set according to the ratio of 7:3. By analysing the scatterplot and line graph between the predicted and actual values of the training and test sets, we find that the prediction effect is better, but there is a certain deviation. Specifically, the scatter plot shows a Y=X distribution, but the actual values have a wide range of variation relative to the predicted values. Meanwhile, the line graph shows that the actual traffic flow and the predicted traffic flow have the same trend, but the actual traffic flow value changes more, while the predicted traffic flow value changes less. This may be due to the long time span resulting in too many cycles, and shortening the time span can reduce the number of cycles and thus improve the prediction accuracy. Further analysis of the MAE values for the training and test sets shows that the MAE values for both the training and test sets are relatively small, 1.8437 and 2.6408, respectively, where the MAE for the test set is on the large side, but the overall prediction results of the model are better. Therefore, the nonlinear weight decreasing PSO-SVR univariate time-series prediction algorithm used in this paper can provide powerful decision support for traffic management departments and help them to better formulate traffic planning and management strategies.

Keywords: PSO-SVR, Time series, Traffic flow prediction.

1. Introduction

Traffic flow prediction is an important research direction in the field of transport, and its background mainly stems from the problem of urban traffic congestion. With the continuous acceleration of urbanisation, the expansion of urban road networks and the increase in the number of private cars, traffic congestion has become an inevitable problem in urban development [1]. Therefore, in order to better solve the traffic congestion problem, accurate prediction of traffic flow is needed. Traffic flow prediction can help traffic management departments to rationally plan road construction, optimise signal light control, adjust public transport routes, etc., so as to alleviate traffic congestion.

Time series algorithms play an important role in traffic flow prediction. Time series algorithm is a method of forecasting based on historical data, which predicts the future trend by analysing the laws of trend, periodicity and seasonality in historical data [2,3]. In traffic flow forecasting, time series algorithms can use the traffic flow data in the past period to infer the traffic flow situation in the future period. Specifically, time series algorithms can make predictions by building ARIMA models (Autoregressive Moving Average Models) or using neural network models [4].

The ARIMA model is a commonly used method for time series analysis and is also widely used in traffic flow forecasting.The ARIMA model can be used to build a forecasting model by analysing historical data and determining its autoregressive, differential and moving average orders [5,6]. The model can predict the traffic flow in the future period and give the corresponding confidence intervals. Neural network model is another commonly used time series forecasting method, which is also widely used in traffic flow prediction [7]. The neural network model can be used to build a prediction model by learning from historical data, and then the model is used to predict the traffic flow in the future period. This method has a strong nonlinear fitting ability, and can better adapt to the changing law of traffic flow in the complex traffic environment.

Traffic flow prediction is one of the important means to solve the problem of urban traffic congestion. And time series algorithm, as a commonly used prediction method, plays an important role in traffic flow prediction, this paper is based on nonlinear weight decreasing PSO-SVR univariate time series prediction algorithm for car flow prediction, which can help the traffic management department to better formulate traffic planning and management strategies.

2. Introduction to the dataset

The traffic flow datasets used in this paper record traffic flow data over a continuous period of time, including information on the number of vehicles, pedestrians, or other modes of transport passing through a particular location or roadway, as well as their speed. Covering information from different cities, roads or transport hubs, this dataset can be used for in-depth study of the changing patterns of traffic flow, peak hours and congestion, and can provide an important reference for urban planning, traffic management policy making and optimisation of intelligent transport systems, so as to improve the efficiency of urban traffic operation and enhance the travelling experience.

The dataset used in this paper is from the publicly released dataset of Kaggle, the dataset contains two variables, time and traffic flow, and records the traffic flow data from 2017/5/25 0:00:00 to 2017/6/30 23:00:00, the data interval is 1 hour, and the change of the traffic flow is shown in Figure 1.

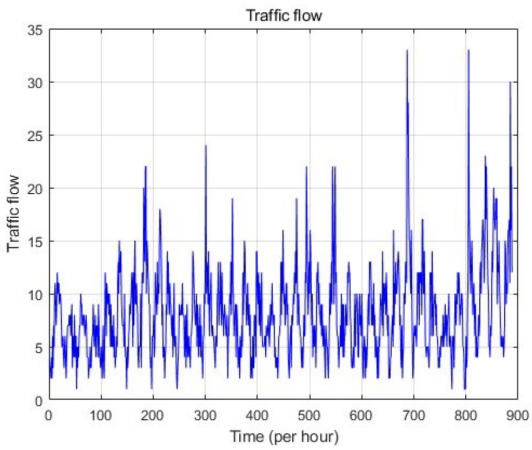

Figure 1. The change of the traffic flow.
(Photo credit : Original)

3. Related work

Time series analysis methods are widely used for traffic flow forecasting, ARIMA model, neural network model and deep learning model are the methods used for traffic flow forecasting.

DOI: 10.54254/2755-2721/75/20240544

3.1. ARIMA
ARIMA model is a commonly used time series analysis method, which is also widely used in traffic flow forecasting. For example, researchers from the University of Chinese Academy of Sciences used the ARIMA model to predict the traffic flow of a main road in Beijing for the coming week and got better results.

3.2. Neural Network
Neural network model is another commonly used time series analysis method, which is also widely used in traffic flow prediction. For example, researchers at the University of Waterloo in Canada used a neural network model to predict traffic flow in the city of Toronto in the next 15 minutes and compared it with other methods.

3.3. Deep Learning
Deep learning models have also been widely used in traffic flow prediction in recent years. For example, researchers at Nanjing University in China used a Long Short-Term Memory (LSTM) network to predict traffic flow in the city of Nanjing within the next hour, and achieved better results.

4. Method
PSO Optimisation SVM for Time Series Forecasting is a forecasting model based on Particle Swarm Optimisation algorithm and Support Vector Machines. The model is learnt and trained on historical data to obtain a forecasting model which is used to forecast future time series data.

4.1. SVM Fundamentals
SVM is an optimisation algorithm for binary classification problems, the basic idea is to map the samples into a high-dimensional space, where an optimal hyperplane is found to partition the samples of different categories [8]. Among them, the hyperplane is selected by maximising the interval between two different categories. In practice, SVMs can handle nonlinear problems with kernel functions.

4.2. PSO optimised SVM time series forecasting model
In the PSO optimised SVM time series prediction model, the historical time series data are first pre-processed, including denoising, smoothing and normalisation. Then the preprocessed data is divided into training set and test set [9]. In the training stage, this paper uses the PSO algorithm to optimise the parameters of the SVM model, including the kernel function type, kernel function parameters and penalty factor. Through continuous iterative optimisation, the optimal SVM model is finally obtained [10].

In the prediction stage, the trained SVM model is used to predict future time series data. Specifically, the time series data to be predicted are input into the SVM model, and the corresponding prediction results are obtained through calculation. Finally, the prediction results are compared with the actual values to evaluate the prediction accuracy.

5. Result
Firstly the dataset is read and divided in the ratio of 7:3, 70% of the data is used for training and 30% of the data is used for testing. This experiment was run on matlab R2022a with 32G of memory. after inputting the data it was normalised and transposed for ease of processing.

For the experimental setup, the pso parameter local search capacity was set to 1.5, pso parameter global search capacity was set to 1.7, maximum number of evolutions was set to 50, maximum number of populations was set to 5. The model was evaluated using R2, MAE, MBE and MAPE to output the predicted value-actual value line graph.

In this paper, we use a standalone experimental equipment to conduct experiments, the CPU is Intel Core i7, the memory is 32GB RAM, for large-scale datasets, 32GB or more memory will be more favourable, and the graphics card is NVIDIA GeForce GTX series.

Firstly output the predicted-actual value scatterplot of the training set as shown in Fig. 2. Secondly output the predicted-actual value scatterplot for the test set as shown in Fig. 3.

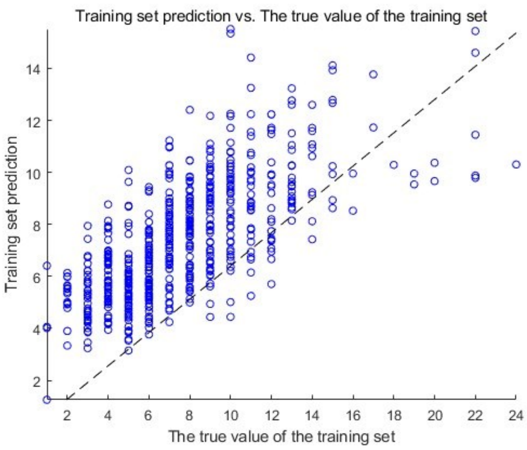

Figure 2. Y-X scatter plot.
(Photo credit : Original)

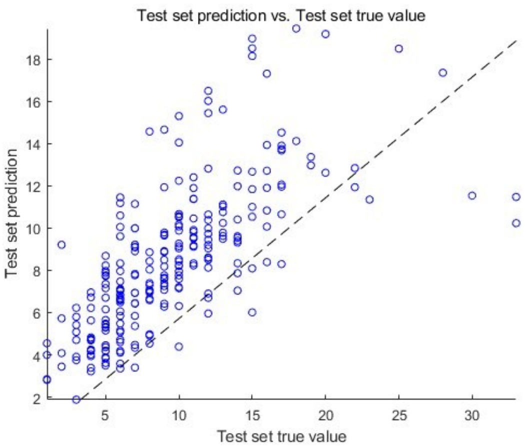

Figure 3. Y-X scatter plot.
(Photo credit: Original)

As shown by the scatterplot distribution of predicted-actual values in the training set and test set, the scatterplot is roughly distributed as Y=X, but slightly shifted, especially the range of change of the actual values relative to the predicted values is relatively large, and the prediction effect is better in general.

The line graphs of predicted-actual values in the training set and test set are output to analyse whether the change trends of predicted and actual data are the same, and the results are shown in Figures 4 and 5.

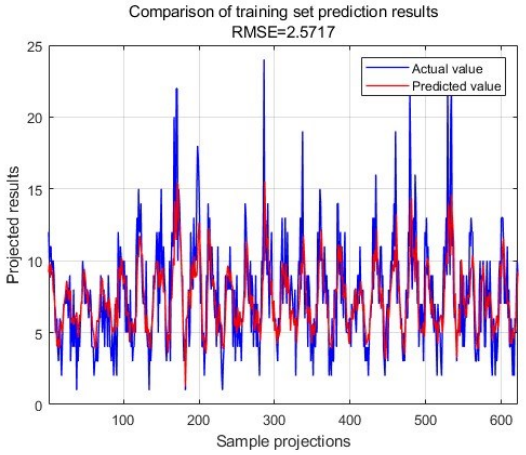

Figure 4. Training set predicted-actual value line plot.
(Photo credit : Original)

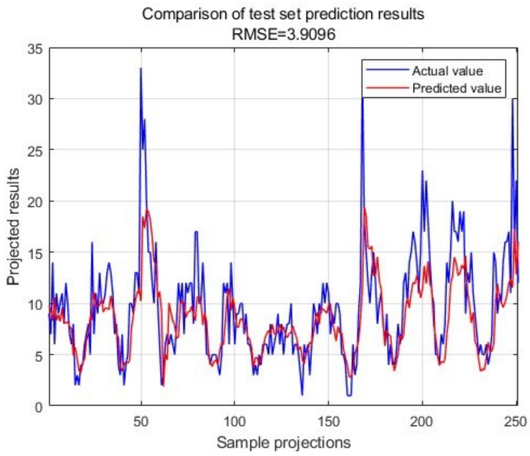

Figure 5. Test Set Predicted-Actual Line Chart.
(Photo credit : Original)

From the trend of the line graph of predicted-actual values in the training and test sets, it can be seen that the trend of actual traffic flow and predicted traffic flow is consistent and relatively overlapping, and the actual traffic flow value changes more. While the predicted traffic flow values change less, the analysis of the reason may be due to the time span is relatively large resulting in the number of cycles is too much, if the time span is shortened to reduce the number of cycles, the predicted traffic flow values will be closer to the actual traffic flow values.

The evaluation metrics R2, MAE, MBE and MAPE for the prediction accuracy of the output training set and test set are shown in Table 1.

Table 1. Indicators for model evaluation.

	R2	MAE	MBE	MAPE
Training set	0.48617	1.8437	-0.2399	0.30673
Test set	0.45085	2.6408	-1.0228	0.32466

The MAE values for both the training and test sets are small, 1.8437 and 2.6408, respectively, and the MAE for the test set is on the large side, with better predictions from the model.

DOI: 10.54254/2755-2721/75/20240544

6. Conclusion

Vehicle flow prediction is an important research direction in the field of transport, and its background mainly stems from the urban traffic congestion problem. In order to better formulate traffic planning and management strategies, this paper adopts the nonlinear weight decreasing PSO-SVR univariate time series prediction algorithm to predict the car flow, and analyses and compares the training set and test set, and draws some conclusions.

Firstly, we divide the dataset into training set and test set according to the ratio of 7:3. 70% of the data is used for training and 30% of the data is used for testing. By observing the scatterplot distribution of predicted-actual values in the training and test sets, it can be found that the scatterplot is roughly distributed as Y=X, but slightly shifted. In particular, the range of variation of the actual values relative to the predicted values is relatively large, which indicates that there is still some room for improvement in the prediction effect of the model.

Next, we analysed the trend of the line graph of predicted-actual values in the training and test sets. The results show that the trends of actual traffic flow and predicted traffic flow are consistent and relatively overlapping. However, the actual traffic flow values change more, while the predicted traffic flow values change less. We analyse that this may be due to the large time span resulting in too many cycles, if the time span is shortened and the number of cycles is reduced, the predicted traffic flow values will be closer to the actual traffic flow values.

Finally, we compared the MAE values of the training and test sets. The results show that the MAE values of both the training set and the test set are relatively small, respectively 1.8437 and 2.6408. where the MAE of the test set is on the large side, but overall the model has better prediction results.

Combining the above analyses, we can conclude that the nonlinear weight decreasing PSO-SVR univariate time series prediction algorithm used in this paper to predict the car traffic has achieved better results to a certain extent.

References

[1] Yinpu W ,Siping K ,Chengchuan A , et al.A Hybrid Framework Combining LSTM NN and BNN for Short-term Traffic Flow Prediction and Uncertainty Quantification[J].KSCE Journal of Civil Engineering,2023,28(1):363-374.

[2] Zhao Z ,Xiaohong J .A spatio-temporal grammar graph attention network with adaptive edge information for traffic flow prediction[J].Applied Intelligence,2023,53(23):28787-28803.

[3] Sun L ,Liu M ,Liu G , et al.FD-TGCN: Fast and dynamic temporal graph convolution network for traffic flow prediction[J].Information Fusion,2024,106102291-.

[4] Martínez I ,Heras G J ,Jardines A , et al.Predicting Air Traffic Flow Management hotspots due to weather using Convolutional Neural Networks[J].Engineering Applications of Artificial Intelligence,2024,133(PA):108014-.

[5] Zhao Z ,Yuan J ,Chen L .Air Traffic Flow Management Delay Prediction Based on Feature Extraction and an Optimization Algorithm[J].Aerospace,2024,11(2):

[6] Gao C ,Wang Z ,Wang S , et al.Mitigating oscillations of mixed traffic flows at a signalized intersection: A multiagent trajectory optimization approach based on oscillation prediction[J].Physica A: Statistical Mechanics and its Applications,2024,635129538-.

[7] Yong L ,Jianying Z ,Xiang W , et al.GT-LSTM: A spatio-temporal ensemble network for traffic flow prediction[J].Neural Networks,2024,171251-262.

[8] Zhang X ,Ren Q ,Zhang Y , et al.Expressway traffic flow prediction based on MF-TAN and STSA.[J].PloS one,2024,19(2):e0297296-e0297296.

[9] Saad H ,Ashadul I ,Kashif A , et al.Deep learning based multimodal urban air quality prediction and traffic analytics[J].Scientific Reports,2023,13(1):22181-22181.

[10] Du ,Kun,Liu , et al.Investigating the Slurry Fluidity and Strength Characteristics of Cemented Backfill and Strength Prediction Models by Developing Hybrid GA-SVR and PSO-SVR[J].Mining, Metallurgy Exploration,2022,39(2):1-20.