

Short-term Traffic Flow Prediction of Urban Roads Based on Random Forest

Luofeng Jiang

School of Software Engineering

Chengdu University of Information Technology

Chengdu,China

Abstract: In the intelligent traffic system, accurately grasp the intrinsic changes in the traffic flow law, and timely scientific prediction of the future traffic flow for a number of moments, will have a very important significance to traffic guidance work, traffic management work, traffic planning work, etc. In this paper, we propose a short-time traffic flow prediction model based on random forest regression(RFR) algorithm, and improve the performance of the model by adjusting the hyperparameters of the model. The performance of the random forest model was also compared with the support vector regression (SVR) model and the decision tree regression(DTR) model, and the RFR model ultimately yielded the best predictions.

Keywords: Prediction; Random forest; Regression; Traffic flows;Model

1 INTRODUCTION

With the development of society and economic progress, a series of traffic problems have become increasingly serious problems that many cities have to face, and the application of intelligent transportation systems to ease traffic congestion has become possible. Among them, short-term traffic flow prediction, which is one of the important foundations of intelligent transportation system, is the key to real-time, accurate and rapid traffic management, inducement and control [1]. Therefore, the study of short term traffic flow prediction on roads is of great significance to improve the level of urban traffic management. At present, there are many models and methods that can be used for traffic flow prediction, commonly used methods such as historical mean model [2], time series model [3], Kalman filter model [4] and other prediction methods and models based on traditional mathematics, as well as prediction methods based on neural networks [5], wavelet theory [6], non-parametric regression [7], support vector machine [8] and other non-linear theories.

This paper presents a model based on random forest regression to predict short term traffic flow on roads. Because random forests are characterized by nonlinear mapping ability, self-learning and adaptive ability,

generalization ability and fault tolerance, the predictive accuracy of the model is guaranteed while reducing model complexity, overfitting and computational volume.

2 RANDOM FOREST

Random forest has both regression and categorization, and when the study variable is a continuous variable, random forest regression is used for analysis, and when the study variable is a categorical variable, random forest classification is used for analysis. The traffic flow data samples in this paper are continuous variables and fall under the scope of regression methods.

Random Forest is an integrated learning method based on Bagging. The algorithm refers to the analytical prediction of a sample by constructing a combined model through multiple decision trees, each decision tree model will have a predictive value, and the predictive value of each tree will be aggregated, which will ultimately enhance the predictive effect of the model. The regression prediction will average the predicted value of each tree to get the final predicted value.

The random forest regression algorithm process is as follows.

(1) Random generation of sample subsets using

Bagging ideas.

- (2) K attributes are randomly selected from a large number of attributes, node splitting is performed, and a single regression decision sub-tree is constructed.
- (3) Repeat steps 1 and 2 to construct N regression decision sub-trees to form a forest.
- (4) The predicted values of the N decision sub-trees are averaged as the final predicted result.

In the above algorithmic process, no pruning is required during the splitting of each decision tree.

3 RESULTS AND ANALYSIS

This section begins with an introduction to data sources and data collection. The process of constructing the prediction model is then described, the parameter adjustments in the construction process are described in relation to it, and the performance of the prediction model is evaluated by the prediction results.

3.1 Experimental Data

In this paper, the subject of the study is an intersection in Mianyang City, the time period is November 4, 2019 to November 29, 2019, remove the weekend, a total of 20 days of traffic flow data, every 5 minutes is a time period, the whole data set is processed statistics to get 5760 sets of data, split it into training set and test set, training set 4608 sets, test set 1152 sets.

3.2 Model Performance Metrics

A common regression model accuracy is used as a metric for parameter tuning and performance analysis of random forest regression models. The meaning absolute error(MAE), R-squared(R^2) and root mean square error(RMSE) are included. the smaller the MAE and RMSE, the larger the R^2 , the higher the model accuracy.

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

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_c)^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (3)$$

In equation (1), equation (2) and equation (3), y_i is the true value; \hat{y}_i is the predicted value; \bar{y}_c is the mean of the true value; N is the sample size.

3.3 Model Parameters

After initializing the random forest regression model, a grid search was applied for parameter tuning and 10-fold cross-validation (Python based scikit-learn package) to get the parameters min_samples_split is 12, min_samples_leaf is 1, max_features is 29 and max_depth is None.

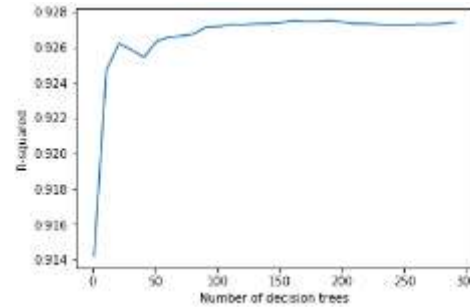


Figure 1. Learning curve a

Figure 1 shows the change in performance when the number of decision trees in the random forest regression model is increased by 10 decision trees at a time from 1-300, and it can be seen that the accuracy is highest when the number of decision trees is around 190. Figure 2 represents the change in performance when the number of decision trees in the random forest regression model is increased by one decision tree at a time over the range 182-201, it can be seen that the model has the highest accuracy when the number of decision trees is set at 186, the parameter n_estimators is 186.

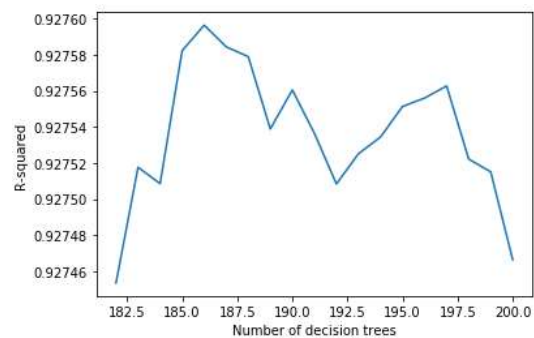


Figure 2. Learning curve b

3.4 Comparative Model Analysis

Calculate the MAE, RMSE and R^2 of the model according to the training results of the model on the experiment, and compare the experimental model with the support vector regression prediction model and the decision tree prediction model. The results are shown

in Table 1.

Table 1. Model comparison

	R^2	MAE	RMSE
RFR	0.927	17.213	26.436
SVR	0.926	18.437	26.669
DTR	0.888	20.528	32.798

By comparing the R^2 , MAE and RMSE of three models of decision tree regression, support vector regression and random forest, the prediction effect of random forest regression is finally shown to be better.

4 CONCLUSION

Random forest, as an efficient machine learning algorithm, has been widely used in many fields. However, there are few applied researches in traffic flow prediction. In this paper, through the research and implementation of the road short-term traffic flow prediction model based on random forest regression algorithm, it is found that the model has high prediction accuracy.

5 REFERENCES

- [1] VLAHOIANNI E I, GOLIAS J C, KARLAFTIS M G. Short-term Traffic Forecasting: Overview of Objectives and Methods [J]. Transport Reviews, 2004, 24 (5):533 – 557
- [2] Okutani I, Stephanedes Y J. Dynamic prediction of traffic volume through Kalman filtering theory [J]. Transportation Research Part B: Methodological,1984,18(1): 1-11.
- [3] KARLAFTIS M G , VLAHOIANNI E I. Memory Properties and Fractional Integration in Transportation Time Series [J] . Transportation Research Part C:Emerging Technologies, 2009, 17 (4) : 444 – 453.
- [4] GUO JH, WILLIAMS B M. Real Time Short Term Traffic Speed Level Forecasting and Uncertainty Quantification Using Layered Kalman Filters [J]. Transportation Research

Record, 2010, 2175: 28 – 37

- [5] Ledoux C. An urban traffic flow model integrating neural networks [J]. Transportation Research Part C: Emerging Technologies, 1997, 5(5): 287-300.
- [6] YANG Chun-xia , FU Yi-qin , BAO Tie-nan. Short-term Traffic Flow Prediction Based on Similarity [J]. Journal of Highway and Transportation Research and Development , 20015, 32 (10) : 124 – 128.
- [7] ZHANG Tao, CHEN Xian, XIE Mei-ping, et al. K-NN Based on Nonparametric Regression Method for Short Term Traffic Flow Forecasting [J]. System Engineering-Theory & Practice, 2010, 30 (2) : 376 – 384.
- [8] WU Q. A Hybrid-forecasting Model Based on Gaussian Support Vector Machine And Chaotic Particle Swarm Optimization [J]. Expert Systems with Applications, 2010, 37 (3): 2388 – 2394.