Vol. 14 (2026): Proceedings of the 2nd International Conference on Artificial Intelligence, Modern Engineering and Environmental Sustainability (IC-AIMEES 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/ ic-aimees.20260447453

Short-term Traffic Flow Prediction at Urban Intersections Based on Quantile Regression and Bayesian Networks

Yuanzhuo Xie*

School of Transportation Engineering, Chang'an University, Shanxi, China

**Corresponding author: Yuanzhuo Xie.*

Abstract

With the advancement of urbanization, intersections—key nodes in road networks—frequently become hotspots for traffic congestion and accidents. Their multi-phase signal control, lane turning diversion, and upstream-downstream interaction lead to traffic flow with complex spatiotemporal correlations and nonlinear fluctuations. This study proposes a hybrid model integrating quantile regression and Bayesian networks to enhance prediction accuracy and robustness under complex conditions. Raw data from Wuchang District's traffic monitoring system undergo cleaning, missing-value imputation, and normalization. Quantile regression models fit 0.1, 0.5, and 0.9 quantiles to capture low, medium, and high traffic states, reflecting predictive distributions. Simultaneously, a Bayesian network analyzes causal relationships between external factors (weather, holidays, accidents) and traffic flow, quantifying influences via conditional probabilities. The two components are integrated into a framework accounting for both flow distribution and external disturbances. Methodologically, quantile regression provides preliminary estimates, which are dynamically adjusted by the Bayesian network's conditional probabilities to address anomalous fluctuations. Experimental results show the hybrid model outperforms traditional linear regression and time-series models in handling fluctuations, especially during peak periods, improving forecasting precision and robustness while supporting traffic management decisions in complex scenarios.

Keywords

traffic flow forecasting, quantile regression, Bayesian networks, abnormal fluctuations, peak traffic

1. Introduction

With the accelerated urbanization process, the number of motor vehicles has grown rapidly, making urban traffic congestion an increasingly prominent problem [1]. Intersections, as critical nodes connecting different road segments, bear the pressure of traffic flow convergence and diversion [2]. Their multi-phase signal control, lane function division, and the mutual influence of upstream and downstream traffic flow lead to complex and volatile traffic characteristics [3-5]. Short-term (usually within 10 minutes) accurate prediction of intersection traffic flow is crucial for optimizing traffic signal timing, guiding travel routes, and reducing congestion and accident risks [6-8]. However, the nonlinearity, spatiotemporal correlation, and susceptibility to external factors (e.g., weather, holidays, traffic accidents) of traffic flow pose significant challenges to prediction accuracy [9,10].

Existing short-term traffic flow prediction methods based on time series and machine learning have achieved results in road segment traffic flow prediction [11-13], but the spatial bifurcation and phase control characteristics in intersection scenarios have not been fully considered [14]. The prediction accuracy and robustness of models in peak and abnormal fluctuation situations still have room for improvement [15]. Addressing this challenge, this study proposes a hybrid prediction framework combining quantile regression and Bayesian networks based on real-time monitoring data of typical intersections in Wuchang District, Wuhan [16-18].

The purpose of this study is to address the limitations of existing traffic flow prediction models in handling complex and volatile traffic conditions and propose a more accurate and robust hybrid prediction model [19,20]. Short-term traffic flow prediction provides real-time data support for traffic management departments, enabling proactive signal control and traffic guidance, which improves road traffic efficiency, reduces residents' travel time costs, and alleviates urban traffic congestion [21,22].

Traffic flow short-term prediction is a core element of optimal traffic signal timing, which can make traffic control more intelligent and proactive [23,24]. According to relevant research, short-term traffic flow prediction can achieve high prediction accuracy when the time interval is within 10 minutes [25]. The urban traffic network is a highly complex and dynamically changing system—traffic flow changes not only have periodic laws but also show strong randomness due to factors such as weather changes and traffic accidents [26,27]. This characteristic is consistent with the advantages of Bayesian networks in describing complex dependencies and handling uncertainty [28,29].

Traffic flow prediction is divided into long-term and short-term prediction based on time scales [30]. Short-term traffic flow prediction, focusing on predicting traffic conditions within 15 minutes, is the core of real-time traffic management and has become a research focus [31]. Current prediction methods mainly include parametric, nonparametric, and hybrid prediction methods [32]. Parametric methods such as ARIMAX and vector autoregressive (VAR) models rely on mathematical model fitting, with low prediction errors but poor adaptability to abnormal data [33]. Nonparametric methods represented by k-nearest neighbor (KNN) and support vector machines (SVM) are data-driven, with good robustness but high complexity and strong data dependence [34]. Hybrid methods such as the combination of ARIMA and artificial neural networks integrate the advantages of multiple single models, achieving higher prediction accuracy [35].

In terms of quantile regression applications, relevant studies have combined it with deep learning frameworks such as LSTM to handle passenger flow prediction and applied it to urban rail transit passenger flow prediction to address the impact of outliers [36]. However, research on applying quantile regression to intersection traffic flow prediction, especially combining it with other methods, is still insufficient [37]. For Bayesian networks, researchers have used Linear Conditional Gaussian Bayesian Networks (LCG-BN) to consider spatiotemporal correlations and traffic attribute associations, and combined Bayesian optimization with LSTM to improve prediction performance. Existing hybrid models for short-term traffic flow prediction include ICEEMDAN-KOSELM-ARIMA, VMD-LSTM, and SSA-LSTM-SVR, which have significantly improved prediction accuracy but lack integration of quantile regression and Bayesian networks.

Overall, existing studies have made progress in traffic flow prediction, but parametric models are poor at handling abnormal data, nonparametric models have high complexity, and single hybrid models often fail to simultaneously consider traffic flow distribution characteristics and external factor influences. The combination of quantile regression and Bayesian networks can complement each other's advantages, but relevant research on intersection traffic flow prediction is relatively scarce.

This study mainly includes four aspects: first, extracting raw data from the traffic flow monitoring system in Wuchang District, Wuhan, and completing data cleaning, missing value filling, and standardization; second, constructing a quantile regression model to capture traffic flow characteristics under low, medium, and high states through different quantiles; third, analyzing the causal relationships between external factors (weather, holidays, traffic accidents) and traffic flow based on Bayesian networks, and calculating corresponding conditional probabilities; fourth, organically combining quantile regression and Bayesian networks to form a hybrid prediction framework [38].

The core innovation of this study lies in proposing a hybrid prediction method combining quantile regression and Bayesian networks. Quantile regression can effectively handle abnormal fluctuations in traffic flow, while Bayesian networks enhance the model's adaptability to external factors through causal reasoning. Additionally, the study verifies the model using real traffic flow data from typical intersections in Wuchang District, Wuhan, ensuring the practicality of the model. The model can adapt to both normal and abnormal traffic flow conditions, maintaining high prediction accuracy even during peak periods or sudden events.

2. Model Principles

Quantile Regression (QR) aims to estimate different quantiles of the conditional distribution of the response variable, rather than just the mean in traditional linear regression. For observation data (xi,yi) (i=1,2,...,n), the core is to minimize the weighted absolute residual sum to estimate the conditional distribution of specific quantiles. The optimization objective is:

$$\mathrm{Min}(\beta)\Sigma(\rho\tau(yi-xiT\beta)) \tag{1}$$

where $\rho\tau(u)$ is the quantile loss function, defined as $\rho\tau(u)=\tau u$ if $u\geq 0$, and $\rho\tau(u)=(\tau-1)u$ if $u<0$. τ is the quantile parameter ($0 < \tau < 1$), with $\tau=0.5$ for median regression, $\tau=0.25$ for lower quartile regression, and $\tau=0.75$ for upper quartile regression.

The advantages of quantile regression include strong robustness to outliers, flexibility in predicting traffic flow at different quantiles, and adaptability to non-normal and heteroscedastic data. However, it also has limitations such as weak interpretability, sensitivity to multicollinearity, and high computational cost.

A Bayesian network is a probabilistic model based on a graphical model, consisting of nodes and directed edges. Nodes represent random variables, and edges represent conditional dependencies between variables, forming a Directed Acyclic Graph (DAG). Each node has a Conditional Probability Table (CPT) describing the probability distribution of the node given its parent nodes. Bayesian inference is based on Bayes' theorem:

$$\{\{P(A|B) = \{P(B|A)P(A)\}/\{P(B)\}\} \tag{2}$$

where P(A|B) is the posterior probability of A given B, P(B|A) is the likelihood function, P(A) is the prior probability of A, and P(B) is the marginal probability of B.

The advantages of Bayesian networks include intuitive expression of causal relationships, strong ability to handle uncertainty, and dynamic adaptability to new data. However, it also faces challenges such as complex structure construction, high computational complexity, and dependence on data quality.

In traffic flow prediction, Bayesian networks can effectively model the causal relationships between traffic flow and external factors (weather, holidays, traffic accidents) through directed acyclic graphs, and handle missing values and noise data. However, they lack accurate prediction capabilities for different traffic flow states. Quantile regression can avoid sensitivity to outliers by selecting specific quantiles (e.g., median, 90% quantile) to fit data, predict traffic flow at different quantiles, and adapt to skewed traffic flow data. However, it does not consider the causal relationships between data.

The combination of the two models can complement each other's shortcomings: quantile regression provides prediction results under different traffic flow states, and Bayesian networks adjust these results based on causal relationships between traffic flow and other variables, improving the comprehensiveness and accuracy of prediction.

3. Hybrid Model Design

3.1 Data Preprocessing

Traffic flow data usually come from multiple sensors. In the data preprocessing stage, missing value imputation (using mean imputation, linear interpolation, etc.) is first performed to ensure data integrity; then normalization is carried out to convert data from different sensors to the same scale, avoiding excessive influence of certain features on the model; finally, statistical methods are used to detect and handle outliers (such as abnormal traffic flow caused by traffic accidents) to avoid interference with the model.

3.2 Model Construction

The study selects 0.1, 0.5, and 0.9 quantiles for regression. Median regression (0.5 quantile) estimates the median value of traffic flow, suitable for regular traffic flow prediction; 90% quantile regression is used to capture peak traffic flow. By minimizing the weighted absolute error, the model is fitted to obtain regression coefficients, and the traffic flow values under different quantiles are predicted.

The Bayesian network nodes include traffic flow, weather, time, holidays, and other variables. Edges represent dependencies between variables. Conditional Probability Tables (CPT) are defined for each node, and the network topology and parameters are determined through structural learning algorithms and maximum likelihood estimation. For example, the conditional probability of traffic flow affected by weather, time, and holidays is calculated as:

$$P(F|W,T,H)=P(F|W)P(F|T)P(F|H) \tag{3}$$

The hybrid model integrates the two models through weighted summation. The preliminary prediction results of quantile regression are dynamically adjusted by the conditional probabilities of the Bayesian network to realize the dynamic correction of abnormal fluctuations. The mathematical logic is:

$$F\tau=\sum jP(F|Xj)\cdot F\hat{}\tau \text{where } P(F|Xj) \tag{4}$$

This is the conditional probability calculated by the Bayesian network, and F^τ is the prediction result of quantile regression.

3.3 Selection of Comparative Models

To highlight the superiority of the hybrid model, three models are selected for comparison: linear regression, decision tree regression, and random forest regression. Linear regression assumes a linear relationship between variables and fits data by minimizing squared errors; decision tree regression recursively divides data through splitting criteria (such as mean square error) to capture complex nonlinear relationships; random forest regression integrates multiple decision trees to average predictions, reducing overfitting.

4. Experiment and Results Analysis

4.1 Dataset Introduction

The data used in this study come from the traffic flow monitoring system of typical intersections in Wuchang District, Wuhan, including more than 60,000 groups of data such as time, weather, number of lanes, and traffic flow. The data are collected by multiple sensors (e.g., 313344, 313349, 313438) covering different time periods (peak hours, holidays) and traffic states. The target variable is the traffic flow of sensor 313349, and the feature variables include data from other sensors and external factors. The dataset is split into a training set (80%) and a test set (20%) using the train_test_split method.

4.2 Model Implementation

The quantile regression model uses 50% quantile regression to estimate the median of traffic flow and 90% quantile regression to capture peak traffic flow. The model is trained by minimizing the weighted absolute error, and the prediction accuracy is evaluated by Mean Absolute Error (MAE). The Bayesian network constructs nodes including sensor data and external factors, defines CPT, and uses Variable Elimination for inference. The final prediction value of the hybrid model is obtained by weighted summation of the quantile regression results and the Bayesian network adjustment part.

4.3 Results Analysis

The MAE values of each model are shown in the table below. The hybrid model has the smallest MAE (3.73), followed by linear regression (3.75), random forest regression (4.02), and decision tree regression (5.43).

Table 1: MAE in different model

Model	MAE
Decision Tree Regression	5.43
Random Forest Regression	4.02
Linear Regression	3.75
Quantile Regression + Bayesian Networks	3.73

The visualization results show that the prediction trend of the hybrid model is consistent with the actual traffic flow, and it can better capture peak traffic flow fluctuations. The prediction results of comparative models such as decision trees have larger deviations in peak periods. The hybrid model's superior performance is due to the complementarity of the two components: quantile regression handles abnormal fluctuations, and Bayesian networks integrate external factor influences, improving the model's robustness and accuracy.

5. Conclusions

This study constructs a hybrid model integrating quantile regression and Bayesian networks, using traffic flow data from intersections in Wuchang District, Wuhan for verification. The results show that the hybrid model outperforms traditional models such as linear regression, decision tree regression, and random forest regression in prediction accuracy and robustness. It can effectively handle abnormal traffic flow fluctuations, especially in peak period prediction, and provide reliable support for traffic management departments in complex scenarios such as adverse weather and holidays.

In terms of the limitations of the study, the construction and reasoning process of Bayesian networks are relatively complex, requiring long computing time when processing large-scale datasets. Second, the model has high requirements for data quality, and prediction performance may be affected by missing values and noise. Third, although external factors such as weather and holidays are considered, the adaptability to sudden events (such as traffic accidents) is still limited.

Future research can optimize the structure of Bayesian networks, adopt approximate reasoning methods to reduce computational overhead; combine deep learning methods (such as LSTM, GCN) with the hybrid model to further improve prediction accuracy and generalization ability; deploy the model in traffic monitoring systems to realize real-time traffic flow prediction and dynamic scheduling, providing more intelligent decision support for traffic management.

References

[1] National Bureau of Statistics. China Statistical Yearbook 2022 [M]. Beijing: China Statistics Press, 2022: 945.

[2] Central Committee of the Communist Party of China and the State Council. Outline for Building a Transport Power [J]. Journal of Waterway and Harbor, 2019, 40(06): 628.

[3] Wang X M. Dynamic Real-time Feedback Control of Traffic Signals in Urban Oversaturated Road Networks [D]. Dalian: Dalian Maritime University, 2018.

[4] Xiang J P. Research on Regional Balanced Control Methods and Applications of Urban Road Traffic Signals [D]. Hefei: University of Science and Technology of China, 2018.

[5] Wei F, Li X, Guo Y, et al. Flow Direction Level Traffic Flow Prediction Based on a GCN-LSTM Combined Model [J]. Intelligent Automation & Soft Computing, 2023, 37(2): 2001–2018.

[6] Zhang J F. Research on Short-term Traffic Flow Prediction Methods at Urban Intersections [D]. Kunming: Kunming University of Science and Technology, 2018.

[7] Fan Y Y. Review of the Development and Prospect of Intelligent Transportation Systems [J]. Xinjiang Transportation Science & Technology, 2014, (4): 90-96.

[8] Shi Y F. Research on Traffic Flow Optimization Control Strategies Under Vehicle-infrastructure Cooperation [D]. Jinan: Shandong University, 2017.

[9] Tian X J. Research on Traffic Signal Control Optimization Methods Under Unsaturated Conditions [D]. Changchun: Jilin University, 2018.

[10] Williams B M. Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modeling [J]. Transportation Research Record: Journal of the Transportation Research Board, 2001, 01-3488: 194–200.

[11] Sun H, Liu H X, Xiao H, et al. Short term traffic forecasting using the local linear regression model[C]. In 82nd Annual Meeting of the Transportation Research Board, Washington, DC, 2003.

[12] Wang Y, Papageorgiou M. Real-time freeway traffic state estimation based on extended Kalman filter: a general approach [J]. Transportation Research Part B: Methodological, 2005, 39(2): 141-167.

[13] Chandra S R, Al-Deek H. Predictions of freeway traffic speeds and volumes using vector autoregressive models [J]. Journal of Intelligent Transportation Systems, 2009, 13(2): 53–72.

[14] Selvaraj Vasantha Kumar. Traffic flow prediction using kalman filtering technique[C]. 10th International Scientific Conference Transbaltica 2017: Transportation Science and Technology. Procedia Engineering, 2017, 187: 582 – 587.

[15] Marković H, Dalbelo Bašić B, Gold H, et al. GPS data-based non-parametric regression for predicting travel times in urban traffic networks [J]. Promet–Traffic & Transportation, 2010, 22 (1): 1-13.

[16] Habtemichael F G, Cetin M. Short-term traffic flow rate forecasting based on identifying similar traffic patterns [J]. Transportation Research Part C: Emerging Technologies, 2016, 66: 61-78.

[17] Chan K Y, Dillon T, Chang E, et al. Prediction of short-term traffic variables using intelligent swarm-based neural networks [J]. IEEE Transactions on Control Systems Technology, 2013, 21(1): 263–274.

[18] Yu C Y, Lam K C. Applying multiple kernel learning and support vector machine for solving the multicriteria and nonlinearity problems of traffic flow prediction [J]. Journal of Advanced Transportation, 2014, 48(3): 250–271.

[19] Kumar K, Parida M, Katiyar V K. Short term traffic flow prediction in heterogeneous condition using artificial neural network [J]. Transport, 2015, 30 (4): 397-405.

[20] Cai P L, Wang Y P, Lu G Q, et al. A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting [J]. Transportation Research Part C: Emerging Technologies, 2016, 62: 21-34.

[21] Zeng D, Xu J, Gu J, et al. Short-term traffic flow prediction using hybrid ARIMA and ANN models[C]. 2008 Workshop on Power Electronics and Intelligent Transportation System, Guangzhou, China, 2008: 621-625.

[22] Zhang Y, Liu Y C. Analysis of peak and non-peak traffic forecasts using combined models [J]. Journal of Advanced Transportation, 2011, 45(1): 21–37.

[23] Li L H, He S L, Zhang J, et al. Short-term highway traffic flow prediction based on a hybrid strategy considering temporal–spatial information [J]. Journal of Advanced Transportation, 2016, 50: 2029–2040.

[24] Liu Z, Guo J, Cao J, et al. A hybrid short-term traffic flow forecasting method based on neural networks combined with K-nearest neighbor [J]. Promet–Traffic & Transportation, 2018, 30(4): 445-456.

[25] Vlahogianni E I, Golias J C, et al. Short-term traffic forecasting: overview of objectives and methods [J]. Transport Reviews, 2004, 24(5): 533-557.

[26] HOSSEINI H. Traffic flow prediction using MI algorithm and considering noisy and data loss conditions: an application to Minnesota traffic flow prediction[J]. Promet–Traffic & Transportation, 2014, 26(5): 393-403.

[27] TIAN X, YU D, XING X, et al. Hybrid short-term traffic flow prediction model of intersections based on improved complete ensemble empirical mode decomposition with adaptive noise[J]. Advances in Mechanical Engineering, 2019, 11(4).

[28] QU W, LI J, YANG L, et al. Short-term intersection traffic flow forecasting[J]. Sustainability, 2020, 12(19): 8158.

[29] Qu W, Li J, Song W, et al. Entropy-Weight-Method-Based Integrated Models for Short-Term Intersection Traffic Flow Prediction [J]. Entropy, 2022, 24(7): 849.

[30] Ye B-L, Zhu S, Li L, et al. Short-term traffic flow prediction at isolated intersections based on parallel multi-task Learning [J]. Systems Science & Control Engineering, 2024, 12(1).

[31] Xiao Y, Yin Y. Hybrid LSTM Neural Network for Short-Term Traffic Flow Prediction [J]. Information, 2019, 10(3): 105.

[32] Liu S, Lin W T, Wang Y, et al. Convolutional Neural Network-Based Bidirectional Gated Recurrent Unit–Additive Attention Mechanism Hybrid Deep Neural Networks for Short-Term Traffic Flow Prediction [J]. Sustainability, 2024.

[33] Huang M-L. Intersection traffic flow forecasting based on ν-GSVR with a new hybrid evolutionary Algorithm [J]. Neurocomputing, 2015, 147: 343–349.

[34] Dai J, Sun M, Shi Q. Short-Term Traffic Volume Prediction Based on Gray Wolf Optimization Algorithm [M]//Smart Innovation, Systems and Technologies. Springer Singapore, 2021: 221–226.

[35] Wei F, Li X, Guo Y, et al. Flow Direction Level Traffic Flow Prediction Based on a GCN-LSTM Combined Model [J]. Intelligent Automation & Soft Computing, 2023, 37(2): 2001–2018.

[36] Chen K, Zhao S, Zhang D Y. Short-term Traffic Flow Prediction based on Data-Driven K-nearest neighbour Nonparametric Regression [C]. Journal of Physics: Conference Series, 2019, 1213.

[37] Qin X, Yin D, Dong X, et al. Passenger flow prediction of scenic spots in Jilin Province based on convolutional neural network and improved quantile regression long short-term memory network[J]. ISPRS International Journal of Geo-Information, 2022, 11(4): 509.

[38] Liu H, Ren J, Wang Z. Short-term passenger flow prediction strategy of urban rail transit based on ridge regression and LSTM[C]. 2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT), 2023: 530-534.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to thank the supervisor for their valuable guidance during the research process, the research team members for their cooperation in data collection and analysis, and the relevant staff of the Wuchang District Traffic Flow Monitoring System for providing data support. Thanks also to family and friends for their encouragement and support.

Open Access