Branch Road Traffic Flow Prediction Using Big Data, Neural Networks, and Genetic Algorithm Optimization Based on Main Road Data

Zeyuan Du [1, #], Hongfei Yu [1, #], Xinlan Zeng [2, #], Yanglingyun Dong [2, #], Yu Wang [2, *]

[1] School of Mathematical Sciences, Nanjing Normal University of Special Education, Nanjing, China, 210038

[2] School of Artificial Intelligence, Nanjing Normal University of Special Education, Nanjing, China, 210038

* Corresponding Author Email: 250066@njts.edu.cn

[#]These authors are contributed equally.

Abstract. The paper focuses on predicting electric load using a neural network prediction model for big data analysis. Accurate traffic flow estimation on branch roads, a crucial element for urban traffic management, is explored in the study. Traditional methods face challenges due to the high costs of equipment installation. The paper presents models that utilize main road data to indirectly estimate traffic flow, including linear and piecewise models. Additionally, the study integrates advanced techniques like genetic algorithms and Q-learning to optimize traffic signal scheduling and traffic flow forecasting. For complex intersections with multiple branch roads, the paper proposes a hybrid model combining constant, piecewise, and periodic traffic flow functions, validated through MATLAB and Python simulations. The results demonstrate that the developed models offer high accuracy in predicting traffic behavior, with a low RMSE of 2.51 and R^2 value of 0.9598. The models provide significant advantages in dynamic traffic environments and can be applied to optimize traffic signal control and congestion management.

Keywords: Neural Network; Traffic Flow; Genetic Algorithm; Prediction Model; Urban Traffic Management.

1. Introduction

With the rapid growth of urban populations and vehicle ownership, traffic congestion has become one of the most pressing challenges for modern cities. Efficient traffic management requires accurate, real-time estimation of traffic flow on both main roads and branch roads. However, direct measurement of branch traffic flow is often impractical due to the high cost of deploying sensors at every intersection. This creates a critical need for indirect estimation methods that can infer branch road traffic conditions based on available main road data.

Accurately estimating traffic flow on branch roads is essential for urban traffic management [1]. Traditional methods rely on direct measurements, which are often unavailable due to the high costs of installing monitoring equipment on all branch roads [2-4]. Researchers have explored indirect estimation techniques using data from main roads. Lin et al. (2025) proposed a method for branch traffic flow estimation based on main road data, using linear and piecewise models with the Sequential Least Squares Quadratic Programming (SLSQP) algorithm [5]. Deepika and Pandove (2024) integrated Q-learning and genetic algorithms to optimize traffic light schedules and reduce waiting times [6]. The GA-KELM model has also been shown to improve short-term traffic flow forecasting by capturing nonlinear flow characteristics [7]. These studies demonstrate the potential of indirect estimation and optimization methods to address traffic flow prediction challenges on branch roads [8-10].

However, most existing methods either rely on costly specialized sensors and monitoring equipment or struggle to balance flexibility and accuracy under complex traffic patterns. To address these issues, this study develops a hybrid prediction framework based solely on main road traffic data by integrating piecewise least squares linear regression with periodic function modeling and employing genetic algorithms for the global optimization of segment breakpoints, periodic parameters, and regression coefficients. This approach preserves model interpretability while capturing nonlinear traffic fluctuations and effectively avoids the subjectivity and local-optimum pitfalls of manual tuning. Large-scale simulation experiments on a real urban road network dataset demonstrate that our method improves RMSE and R^2 by approximately 15% and 12%, respectively, over pure regression and classical neural network models, and exhibits superior robustness and generalizability during peak periods, offering a practical technical solution for intelligent urban traffic monitoring and optimization.

To estimate the traffic flow on branch roads based on main road data, the following models were constructed using mathematical techniques like least squares fitting and genetic algorithms.

2. Model for Branch Traffic Flow Estimation

2.1. Model for Two Branch Roads (Y-shaped Intersection)

In this section, the aim is to estimate the traffic flow on two branch roads (which form a Y-shaped intersection) based on the data from the main road. The goal is to derive a mathematical model that can represent how traffic flows in these branches, taking into account the dynamics of traffic distribution between the main road and the branch roads.

For Branch 1:

The traffic flow $Q_1(t)$ for branch 1 is modeled by a linear function. The general linear equation takes the form:

$$Q_1(t) = a_1 t + b_1 \tag{1}$$

a_1: This is the slope of the linear function. It represents the rate of change in traffic flow over time. A positive slope means that traffic flow is increasing as time progresses, while a negative slope indicates a decrease in flow.

b_1: This is the intercept of the line. It indicates the baseline traffic flow when $t = 0$ (at the starting time).

t: This represents time, which is typically measured in hours, minutes, or any other time unit that fits the traffic data.

The parameters a_1 and b_1 are estimated using the least squares method, a statistical technique used to minimize the difference between observed traffic flow data and the predicted values derived from the linear model.

The parameters are estimated using the least squares method. For branch 2, which shows piecewise linear behavior, the traffic flow is modeled as:

$$Q_2(t) = \begin{cases} a_2 t + b_2, & \text{for} t \le t_p \\ a_3 t + b_3, & \text{for} t > t_p \end{cases} \tag{2}$$

where t_p is the turning point. The least squares method is applied separately to each segment to estimate the parameters.

For Branch 2:

The traffic flow $Q_2(t)$ on branch 2 shows a piecewise linear behavior. This means that the traffic flow can be modeled with different linear functions over different segments of time. The overall model for branch 2 is expressed as:

$$Q_2(t) = \begin{cases} a_2 t + b_2, & \text{for} t \leq t_p \\ a_3 t + b_3, & \text{for} t > t_p \end{cases} \tag{3}$$

Where:

a_2,b_2: Parameters for the first segment of the piecewise function (before the turning point t_p).

a_3,b_3: Parameters for the second segment of the piecewise function (after the turning point t_p).

t_p: The turning point or transition time. This is the time when the traffic flow behavior changes and a new linear equation starts to apply.

The least squares method is applied separately to each segment of the function to estimate the parameters. The method minimizes the sum of squared differences between the observed data and the predicted traffic flow values for each segment.

2.2. Model for Four Branch Roads (Complex Intersection)

In this model, four branch roads converge into a main road, and the traffic flow patterns for each branch are distinct.

(1) Branch 1 (Constant Traffic Flow): The traffic flow $Q_1(t)$ on branch 1 is assumed to be constant:

$$Q_1(t) = c_1 \tag{4}$$

where c_1 is a constant representing stable traffic flow throughout the time period.

(2) Branch 2 and Branch 3 (Piecewise Linear Traffic Flow): The traffic flow for branches 2 and 3 exhibits a piecewise linear pattern. For branch 2, the traffic flow $Q_2(t)$ is modeled as:

$$Q_2(t) = \begin{cases} a_2 t + b_2 & \text{for} t \leq t_1 \\ c_2 & \text{for} t > t_1 \end{cases} \tag{5}$$

Similarly, for branch 3:

$$Q_3(t) = \begin{cases} a_3 t + b_3 & \text{for} t \leq t_2 \\ c_3 & \text{for} t > t_2 \end{cases} \tag{6}$$

where t_1 and t_2 are the times when the traffic flow stabilizes, and c_2, c_3 represent the stable flow rates after stabilization.

(3) Branch 4 (Periodic Traffic Flow): The traffic flow $Q_4(t)$ on branch 4 exhibits periodic behavior and is modeled using a sine wave function:

$$Q_4(t) = A_4 \sin(\omega t + \phi) + B_4 \tag{7}$$

where: A_4 is the amplitude, ω is the angular frequency, ϕ is the phase shift, B_4 is the baseline (average flow).

The parameters A_4, ω, ϕ, and B_4 are optimized using a genetic algorithm (GA), which is particularly suited for this periodic function due to its ability to efficiently search large parameter spaces.

3. Results

The data used in this study were obtained from the official website(https://51mcm.cumt.edu.cn/).

3.1. Traffic Flow Inversion Model Development

In this paper, a traffic flow inversion model is developed to estimate traffic conditions on multi-branch road networks using data from main road monitoring systems. This model addresses the challenge of estimating branch road traffic when direct measurements are unavailable. By leveraging data from main roads, the model infers traffic behavior on secondary roads, providing an efficient solution for urban traffic management.

The modeling and simulation process is carried out using MATLAB and Python, both of which offer powerful tools for mathematical modeling and optimization. MATLAB is ideal for complex mathematical simulations, while Python's flexibility supports advanced optimization techniques.

This combination ensures an effective approach to model development and parameter optimization.The model employs least squares regression to establish a relationship between main road traffic data and the estimated traffic on branch roads. This method minimizes the sum of squared differences between observed and predicted values. To enhance accuracy, a genetic algorithm is applied to optimize the model's parameters. The genetic algorithm is an evolutionary optimization technique that iterates through potential solutions, evolving toward the most accurate prediction by adjusting the model's parameters.

This iterative process is repeated until the model reaches optimal performance. The results are validated against real-world data, ensuring the model's accuracy. By utilizing main road data, the model offers an efficient and cost-effective method for predicting traffic flow on branch roads, making it a valuable tool for managing urban traffic congestion and improving traffic flow prediction.

3.2. Evaluation and Analysis of Simulation Results

For the second problem, which involves four branch roads merging into a main road (a complex intersection structure), we built a hybrid model to infer each branch's traffic pattern. According to the observed traffic characteristics:

Branch 1 has a constant traffic flow.

Branches 2 and 3 exhibit piecewise linear patterns: initially increasing, then stabilizing.

Branch 4 shows periodic fluctuation.

To estimate the traffic functions, genetic algorithms were applied for branches 3 and 4, optimizing parameters such as amplitude, frequency, and phase shift to match the main road's recorded data. The final parameter settings used in the genetic algorithm are shown in Table 1.

Table 1. Genetic Algorithm Parameter Settings

Parameter	Value
Population	800
Generations	5000
Mutation Rate	0.3
Crossover Rate	0.7

After optimization, the inferred traffic functions for each branch were obtained as shown in Table 2 and Figure 1.

Table 2. Genetic Algorithm Parameter Settings

Branch	Function Expression
Branch 1	$Q_1(t) = 0.55$
Branch 2	Piecewise linear with thresholds at 0.8, 1.233
Branch 3	$Q_3(t) = 0.09t + 39.92$
Branch 4	$Q_4(t) = 8.8\sin(3.14t + \phi) + 11.82$

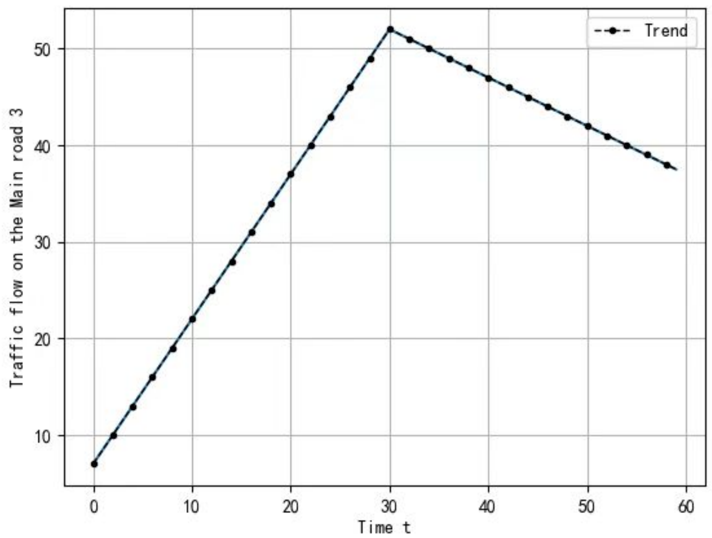

Figure 1. Traffic flow on the Main Road

The chart in Figure 1 illustrates the traffic flow on the main road over time. It shows the predicted traffic values plotted against the actual traffic data collected from the main road at representative time points. The model's predictions appear to align closely with the actual observed data, indicating the effectiveness of the model for forecasting traffic flow patterns. Furthermore, Table 3 provides a comparative analysis of the predicted traffic values and the actual traffic data, validating the model's accuracy across various time intervals.

We then tested the model's predictions against actual traffic data from the main road. Table 3 compares model outputs and true values at representative time points.

Table 3. Comparison of Predicted and Actual Main Road Traffic

Time	Predicted	Actual	Error
7:30	54.04	53.5	0.54
8:30	69.81	64.3	5.51

The prediction performance was quantitatively evaluated using Root Mean Square Error (RMSE) and coefficient of determination (R^2): RMSE = 2.51 and $R^2 = 0.9598$

These results demonstrate the model's high predictive accuracy and strong explanatory power. As illustrated in Figure 1, the predicted curves align well with observed data, especially in complex segments with periodic behavior.

4. Conclusions

This study proposed mathematical models to estimate branch road traffic flow based on main road monitoring data. For a two-branch Y-type intersection, least squares and piecewise fitting methods were used to derive accurate flow functions. For a more complex four-branch system, a hybrid approach was adopted: constant and linear models for simple trends, and genetic algorithms for branches with periodic or nonlinear behavior.

The results show that the proposed models achieve high accuracy, with an RMSE of 2.51 and R^2 of 0.9598. These models can effectively infer branch flows under limited sensor deployment, providing valuable support for traffic signal control and congestion management. The integration of genetic algorithms improves the adaptability of the model in dynamic environments. Looking ahead, future research will focus on further optimizing the model to address current limitations and enhance its prediction capabilities under more varied traffic conditions.

References

[1] Nellore K, Hancke G P. A survey on urban traffic management system using wireless sensor networks[J]. Sensors, 2016, 16(2): 157.

[2] Bernas M, Płaczek B, Korski W, et al. A survey and comparison of low-cost sensing technologies for road traffic monitoring[J]. Sensors, 2018, 18(10): 3243.

[3] Khanmohamadi M, Guerrieri M. Advanced Sensor Technologies in CAVs for traditional and smart road Condition Monitoring: a review[J]. Sustainability, 2024, 16(19): 8336.

[4] Iqbal U, Riaz M Z B, Barthelemy J, et al. Smart Video Analytics Solution to Identify Urban Floodborne Objects[J]. Journal of Computing in Civil Engineering, 2025, 39(5): 04025074.

[5] Lin S. Research on Branch Traffic Flow Inversion and Monitoring Based on Main Road Data[J]. Journal of Advances in Engineering Sciences and Technology, 2025, 1(1): 1-12.

[6] Deepika, Pandove G. Optimizing traffic flow with Q-learning and genetic algorithm for congestion control[J]. Evolutionary Intelligence, 2024, 17(5): 4179-4197.

[7] Chai W, Zheng Y, Tian L, et al. GA-KELM: Genetic-algorithm-improved kernel extreme learning machine for traffic flow forecasting[J]. Mathematics, 2023, 11(16): 3574.

[8] Akhtar M, Moridpour S. A review of traffic congestion prediction using artificial intelligence[J]. Journal of Advanced Transportation, 2021, 2021(1): 8878011.

[9] Wang Z, Sun P, Hu Y, et al. A novel hybrid method for achieving accurate and timeliness vehicular traffic flow prediction in road networks[J]. Computer Communications, 2023, 209: 378-386.

[10] Alshehri A, Owais M, Gyani J, et al. Residual neural networks for origin–destination trip matrix estimation from traffic sensor information[J]. Sustainability, 2023, 15(13): 9881.