

Research on Smart City Traffic Flow Optimization Based on Hidden Markov Models

Mingjie Zhou

Universiti Sains Malaysia, Penang, 11800, Malaysia

ABSTRACT

With the continuous advancement of smart transportation development and the "Transportation Powerhouse" strategy, traffic flow forecasting plays a crucial role in alleviating traffic congestion and enhancing road resource utilization efficiency. Among existing traffic flow prediction methods, traditional statistical models struggle to capture nonlinear characteristics, while machine learning models face challenges such as high data requirements and parameter sensitivity. Furthermore, existing studies lack systematic model integration and comprehensive comparative analysis. This paper focuses on Hidden Markov Models (HMMs), employing literature review and comparative analysis to systematically examine the application progress and technical characteristics of traditional statistical models, machine learning models, and HMMs in traffic flow forecasting since 1990. Findings indicate that HMMs effectively capture both the randomness and temporal patterns of traffic flows, demonstrating superior stability and adaptability in short-term traffic flow prediction and congestion state identification. Improved HMMs effectively mitigate overfitting through state number optimization and algorithmic enhancements, further elevating prediction accuracy in complex traffic scenarios. The conclusions of this study provide guidance for the rational selection of traffic flow prediction models and lay a theoretical foundation for subsequent research on multi-factor integrated HMM optimization.

KEYWORDS

Traffic flow prediction; Hidden markov model; Model comparison analysis; Intelligent transportation

1 Introduction

1.1 Research Background

In recent years, China's accelerated urbanization has driven sustained growth in motor vehicle ownership. By 2025, the national motor vehicle fleet exceeded 460 million units, placing immense pressure on urban road networks. Traffic congestion and low travel efficiency have become core bottlenecks constraining urban development. Against this backdrop, intelligent transportation—a core component of the "Transportation Powerhouse" strategy—aims to achieve dynamic traffic signal control and intelligent route recommendations through precise traffic flow forecasting, shifting from "passive management" to "proactive optimization."

However, existing traffic flow forecasting methods face multiple challenges in complex urban road networks: First, traffic flows are influenced by time, including temporal granularity (e.g., minute, hour, day), variations between peak and off-peak periods, and differences between weekdays and holidays. These factors determine the cyclical and phased changes in traffic conditions. Second, spatial and road characteristics—such as road type, segment topology, and traffic conditions on adjacent segments—influence the transition patterns of hidden states. Third, traffic operational features—including historical flow, speed, density, and sudden incidents—exhibit significant nonlinearity and randomness. On the other hand, traditional prediction models either fail to adapt to complex scenarios due to linear assumptions or struggle to meet real-time demands because of excessive data dependency. Therefore, systematically reviewing the technical characteristics of existing prediction methods and identifying more adaptable models has become an urgent theoretical and practical issue in the field of intelligent transportation.

1.2 Research Significance

This study systematically reviews the theoretical development and research progress of Hidden Markov Models (HMMs) in traffic flow forecasting. It deepens understanding of HMM applicability and limitations in modeling time-series data. By comparing different methods, it summarizes their strengths and weaknesses in parameter selection, state space optimization, and integration with other forecasting models, providing theoretical references and methodological insights for future research. By contrasting and synthesizing traffic flow forecasting methods, this study assists transportation authorities and engineering practitioners in selecting appropriate forecasting techniques. This enhances the scientific rigor and efficiency of urban traffic management and decision-making, offering insights for alleviating traffic congestion and improving travel experiences.

1.3 Research Methodology

1.3.1 Literature Review Method

Using keywords such as "traffic flow prediction," "Hidden Markov Model," "intelligent transportation," "ARIMA," and "LSTM," this study collects and organizes relevant domestic and international research findings on traffic flow prediction. It summarizes the applications and limitations of traditional statistical models, machine learning methods, and HMMs, providing a theoretical basis for the model construction in this paper.

1.3.2 Machine Learning Approach

By comparing the performance and applicability of different prediction models—including linear models (regression, ARIMA) and machine learning models (SVM, neural networks, LSTM)—this section focuses on analyzing the advantages of HMM and its improved model (IHMM) in optimizing traffic flow for smart cities.

2 Literature Review and Technological Evolution of Traffic Flow Prediction Models

2.1 Traditional Statistical Models

Traditional statistical models, centered on the concept of "linear relationship modeling," became the early mainstream approach in traffic flow prediction due to their simplicity and strong interpretability. They primarily include two categories: regression models and ARIMA models.

2.1.1 Regression Models

Regression models achieve prediction by quantifying the linear relationship between traffic flow (dependent variable) and influencing factors (independent variables, such as time, weather, and number of lanes). Common types include linear regression, multiple linear regression, ridge regression, and lasso regression. Zhang Zhaokuan and Li Guohua (2023) proposed a regression-based method for repairing highway traffic flow data. Using linear regression, ridge regression, and lasso regression as base models, they selected the optimal submodel through training data screening ^[1]. Anderson M. D., Lee S., and Kim J. (2021) pioneered the integration of socioeconomic factors—including regional population density, commuting rates, and surrounding job opportunities—alongside temporal and meteorological variables into a multivariate linear regression model for suburban highways in Texas, USA. This approach quantified their linear relationships with traffic flow. Results indicate that "number of jobs (regression coefficient 0.62), morning peak period (0.81), and light rain (regression coefficient -0.35)" significantly influence traffic volume. The model achieves 82% prediction accuracy during off-peak hours ^[2], providing a linear modeling approach for forecasting long-term trends in suburban highway traffic flow.

2.1.2 ARIMA Model

Cui Jian et al. (2024) proposed an ARIMA-based method for short-term traffic flow forecasting on expressways. Results indicate that the ARIMA forecasting model demonstrates good adaptability and stability for expressway short-term traffic flow prediction, outperforming moving average and exponential smoothing models in forecasting accuracy ^[3]. Xu Cuicui (2023) proposed an ARIMA-based traffic flow forecasting model for urban intersections, accounting for temporal randomness and analyzing traffic flow forecasting principles ^[4]. However, traditional statistical models have limitations, as they can only capture linear relationships. Traffic is influenced by numerous nonlinear factors, making it difficult for these models to effectively capture the complex nonlinear characteristics of traffic flow.

In summary, traditional statistical models offer simplicity and strong interpretability but are limited to describing linear relationships, making it difficult to capture complex nonlinear characteristics.

2.2 Machine Learning Models

To overcome the linear limitations of traditional models, machine learning models have been widely adopted in traffic flow forecasting due to their "nonlinear feature extraction" capabilities. These primarily include four categories: Support Vector Machines (SVM), Neural Networks (NN), Long Short-Term Memory Networks (LSTM), and hybrid models. While achieving significant improvements in prediction accuracy, they also introduce issues of data dependency and insufficient interpretability.

2.2.1 Support Vector Machines (SVM)

SVMs map linearly inseparable traffic flow data to high-dimensional spaces using kernel functions (e.g., RBF kernel, polynomial kernel) to fit nonlinear relationships. They exhibit strong generalization capabilities in scenarios with small sample sizes. Jiang Xiaofeng et al. (2012) applied SVM to short-term traffic flow forecasting ^[5], demonstrating its suitability for such predictions by leveraging nonlinearity through kernel functions. Trinh Dinh Toan and Viet-Hung Truong (2020) proposed a short-term traffic flow forecasting method combining SVM with KNN-assisted training, significantly improving training efficiency while maintaining high prediction accuracy ^[6].

2.2.2 Neural Networks (NN) and LSTM

Traditional neural networks (e.g., BP neural networks) fit nonlinear relationships through multi-layer perceptrons but suffer from gradient vanishing issues, making them unsuitable for handling long-term traffic data. Long Short-Term Memory (LSTM) networks effectively resolve gradient vanishing through gating mechanisms (forgetting gate, input gate, output gate), establishing themselves as a mainstream approach for long-term traffic flow prediction. Zhang Yongyang (2021) addressed existing ITS prediction methods by focusing on vehicle flow forecasting—a key parameter in road traffic—using the TS-NN algorithm^[7], which combines time series analysis with neural networks. Hao Guifa, Ning (2023) overcame these challenges by integrating multiple simple recurrent LSTM neural networks with temporal features to predict traffic flow using deep gated stacked neural networks^[8].

2.2.3 Ensemble Models

To address the limitations of single models, researchers proposed a "linear model + machine learning model" combination strategy, integrating the strengths of both types to enhance prediction accuracy. Li Jiansen, Shen Qi (2023) combined a backpropagation (BP) neural network model with an autoregressive integrated moving average (ARIMA) model to forecast short-interval traffic flow on urban roads. The BP neural network demonstrated strong nonlinear fitting capabilities, while the ARIMA model exhibited excellent linear fitting performance^[9].

2.3 Hidden Markov Model (HMM)

As a probabilistic statistical model, the Hidden Markov Model (HMM) captures both the randomness and temporal patterns of traffic flow by establishing probabilistic relationships between "hidden states" (e.g., free-flow, light congestion, heavy congestion) and "observed sequences" (e.g., traffic volume, vehicle speed, lane occupancy). This makes HMM an important supplement in traffic flow forecasting. Its development can be divided into two phases: traditional HMM and improved HMM (IHMM).

2.3.1 Traditional HMM

The core advantage of HMM lies in linking "unobservable states" (e.g., congestion levels) with "directly measurable observations" (e.g., traffic volume). Predictions are achieved through state transition matrices (describing temporal dependencies between states) and observation probability distributions (describing the relationship between states and observations).

Its applications primarily focus on three directions:

Congestion State Identification: Porikli and Li (2004) proposed a Gaussian Mixture Hidden Markov Model (GM-HMM) that extracts vehicle density features from MPEG video, achieving 89.3% congestion detection accuracy with a latency below 0.5 seconds, meeting real-time requirements^[10].

Short-term parameter prediction: Wang Man (2017) utilized California highway data to first determine traffic states via HMM, then combined this with random forest models for flow prediction. This approach reduced the Mean Absolute Error (MAE) by 23.6% compared to standalone random forests, demonstrating HMM's value in supplementing prior state information^[11].

Traffic propagation prediction: Priambodo et al. (2021) extended HMM's latent states to "section-state" combinations, accounting for adjacent section correlations. This reduced traffic propagation prediction error by 18.9% compared to traditional HMM, achieving a breakthrough from "single-point" to "network-wide coordination"^[12].

2.3.2 Improved HMM (IHMM)

Traditional HMMs exhibit two primary limitations: First, the number of states must be manually specified, and improper selection often leads to overfitting or underfitting. Second, observation probability distributions are typically assumed to follow a Gaussian distribution, which struggles to accommodate the non-normal characteristics of traffic flow observations. To address these shortcomings, researchers proposed the Improved HMM (IHMM). Zhao et al. (2019) automatically determined the optimal number of states using AIC/BIC criteria and improved parameter estimation through the Baum-Welch algorithm. This approach reduced the Mean Absolute Error (MAE) by 15.3% compared to traditional HMMs in urban intersection prediction, significantly lowering overfitting risks^[13]. She et al. (2025) integrated temporal features (peak hour indicators, holiday factors) into the state transition matrix of IHMM. In commuter vehicle trajectory prediction, the matching accuracy reached 87.5%, a 19.2% improvement over traditional HMM, with particularly outstanding performance during morning and evening rush hours, demonstrating IHMM's robustness to multi-factor interference. This research indicates that HMM-based improvement methods hold significant potential for optimizing smart city transportation^[14].

In summary, HMM and IHMM achieve dual characterization of traffic flow's "randomness" and "temporality" through probabilistic modeling. They demonstrate balanced performance in state identification and short-term forecasting while requiring less data than machine learning models, establishing themselves as crucial complementary methods for traffic flow prediction.

3 Research Findings

Existing research indicates that traffic flow prediction models have evolved from early traditional statistical methods—represented by linear regression and ARIMA—to a diversified technical framework integrating machine learning models and probabilistic statistical models. Different model types exhibit distinct characteristics in theoretical foundations, modeling approaches, and applicable scenarios, collectively forming the primary framework for current intelligent transportation forecasting research. Traditional statistical models retain value in scenarios with relatively stable traffic flow variations and fewer influencing factors due to their clear structure and strong parameter interpretability. However, their linear assumptions limit performance in complex urban traffic environments. Machine learning models significantly enhance prediction accuracy through nonlinear feature extraction and deep temporal modeling, proving particularly suitable for high-volatility, long-time-series traffic flow forecasting tasks. Nevertheless, their reliance on large-scale data and computational resources, coupled with limited interpretability, constrains their widespread adoption in certain practical traffic management scenarios. In contrast, Hidden Markov Models (HMMs) integrate the randomness and temporal nature of traffic flows into a unified probabilistic framework by incorporating unobservable traffic states. This approach demonstrates robust stability and practicality in short-term traffic flow forecasting and congestion state identification. Furthermore, improved HMMs mitigate overfitting issues through state number optimization and algorithmic enhancements, boosting prediction accuracy and robustness in complex scenarios like morning and evening rush hours. However, existing research remains limited in model system integration, unified evaluation framework development, and engineering adaptation—particularly in multi-factor fusion and large-scale road network applications—which warrant further exploration.

4 Future Research Prospects

To address complex and dynamic real-world traffic environments, future research should deepen multi-factor fusion modeling within the improved HMM framework. This includes incorporating non-routine factors like sudden accidents, extreme weather, and holidays into state transition or observation probability modeling to enhance the model's ability to capture complex traffic state evolution. Concurrently, lightweight integration pathways between HMMs and machine learning methods should be explored. Introducing HMM-based state priors alongside lightweight learning models can reduce computational and deployment costs while maintaining prediction accuracy, thereby enhancing the model's practical implementability in real-world traffic systems. Furthermore, with the advancement of multi-source traffic sensing technologies, multimodal data from sources such as video surveillance, floating car GPS, and vehicle-to-everything (V2X) networks will provide richer information for traffic state recognition. How to effectively fuse multi-source heterogeneous data within the HMM framework to improve the accuracy of latent traffic state recognition and traffic flow prediction performance will become a key research focus in the future.

References

- [1] Zhang, Z.K., Li, G.H. (2024) Highway traffic flow data restoration method based on regression model. *Shanxi Architecture*, 50: 128–132.
- [2] Anderson, M.D., Lee, S., Kim, J. (2021) Multiple linear regression for traffic flow forecasting in suburban highways considering socioeconomic factors. *Journal of Transportation Engineering, Part A: Systems*, 147: 04021035.
- [3] Cui, J., Li, Z., Zhao, J. et al. (2024) Research on short-term traffic flow prediction method of expressways based on ARIMA. *Shandong Transportation Science and Technology*, 4: 89–93.
- [4] Xu, C.C. (2023) Modeling urban intersection traffic flow prediction based on ARIMA. *Electronic Design Engineering*, 30: 20–30.
- [5] Jiang, X.F., Xu, L., Zhu, Y. (2012) Short-term traffic flow prediction based on SVM. *Journal of Guangxi Normal University (Natural Science Edition)*, 4: 13–17.
- [6] Trinh, D.T., Truong, V.H. (2020) Support vector machine for short-term traffic flow prediction and improvement of its model training using nearest neighbor approach. *Transportation Research Record*, 2675: 1–12.
- [7] Zhang, Y.Y. (2021) Road traffic flow prediction based on TS-NN model. *Journal of Fujian University of Technology*, 19: 560–567.
- [8] Hao, G.F., Wang, N. (2023) Traffic flow prediction based on deep threshold stacked long short-term memory neural network. *Digital Technology and Application*, 41: 52–54.
- [9] Li, J.S., Shen, Q., Fan, X.Y. (2019) Short-term traffic flow prediction for urban roads. *Mathematics in Practice and Theory*, 49: 192–197.
- [10] Porikli, F., Li, X. (2004) Traffic congestion estimation using HMM models without vehicle tracking. In: *IEEE Intelligent Vehicles Symposium*. IEEE, New York. pp. 188–193.
- [11] Wang, M. (2017) Traffic state discrimination based on HMM and its application in short-term prediction of traffic flow parameters. *Southeast University, Nanjing*.
- [12] Priambodo, B., Ahmad, A., Kadir, R.A. (2021) Predicting traffic flow propagation based on congestion at neighboring roads using hidden Markov model. *IEEE Access*, 9: 85933–85946.
- [13] Zhao, S.X., Wu, H.W., Liu, C.R. (2019) Traffic flow prediction based on optimized hidden Markov model. *Journal of Physics: Conference Series*, 1168: 052001.
- [14] She, Y., Pei, Z., Liu, Q. et al. (2025) Improved hidden Markov prediction method for urban commuting vehicle trajectories. *Journal of Zhejiang University of Technology*, 4: 1006–4303.