Application Research of Big Data Technology in Intelligent Prediction of Urban Traffic Flow

Yihao Ning*

School of Converged Media Center, Hainan Vocational University of Science and Technology, Haikou, Hainan, 571126, China.
**Corresponding author: I24027287@student.newinti.edu.my*

Abstract: *With the acceleration of urbanization and the proliferation of information sensing technologies, the data generated by urban traffic systems exhibits typical big data characteristics, such as massive volume, high dimensionality, heterogeneity, and strong spatiotemporal dependencies. This presents both new opportunities and challenges for the field of traffic flow prediction. Traditional prediction models have inherent limitations in processing such data and capturing complex nonlinear spatiotemporal dynamics. This study aims to systematically investigate the application framework of big data technology in the intelligent prediction of urban traffic flow. The paper first analyzes the core characteristics of urban traffic big data and the evolutionary principles of intelligent prediction algorithms. Subsequently, it constructs a multi-source data fusion processing method tailored for prediction tasks, a spatiotemporal prediction model architecture based on deep learning (particularly graph neural networks and attention mechanisms), and a dynamic prediction integration system supporting real-time stream processing. Finally, it analyzes the application value of this technology in short-term high-precision prediction, simulation and deduction of congestion propagation, and system performance and uncertainty assessment. The research indicates that big-data-driven intelligent prediction methods can more effectively mine the underlying patterns of traffic systems, providing more accurate and robust technical approaches for dynamic traffic state perception and management decision support.*

Keywords: *Traffic flow prediction; Big data; Deep learning; Graph neural networks; Spatiotemporal data mining; Uncertainty analysis*

Introduction

The urban traffic system is a highly complex, dynamically evolving spatiotemporal network, whose operational efficiency directly impacts socio-economic activities and the quality of public life. Accurate traffic flow prediction constitutes the core prerequisite for achieving intelligent management and optimization of the traffic system. Traditional prediction methods are mostly based on mathematical statistics or simplified physical models. Their linear or quasi-linear core, limited data processing capabilities, and the difficulty in capturing long-range spatiotemporal dependencies encounter bottlenecks in prediction accuracy and generalization ability when confronted with modern traffic big data generated by multi-source sensors, mobile terminals, and the like. Therefore, researching how to apply big data technology, particularly cutting-edge artificial intelligence methods, to automatically learn from massive, heterogeneous real-time data and accurately predict the evolutionary patterns of traffic flow possesses significant theoretical necessity and application urgency. This study aims to systematically construct a complete technical framework, from data processing and model design to application analysis, and to explore the key methods and application potential of intelligent traffic flow prediction driven by big data technology, in order to provide a systematic reference for enhancing prediction accuracy, real-time performance, and reliability.

1. Theoretical Foundations of Big Data and Urban Traffic Flow Prediction

1.1 Analysis of Big Data Characteristics in Urban Traffic Flow Prediction

The urban traffic system is a complex spatiotemporal dynamic system, and the data generated during its operation inherently possesses the core attributes of big data, which constitutes the physical foundation for research on intelligent prediction methods. These data are not only massive in volume but are also characterized by their multi-dimensional, heterogeneous, and high-velocity generation.

Data from fixed sensors such as inductive loops, microwave radar, and video checkpoints provide highly accurate yet spatially discrete point observations, while floating car GPS trajectories and mobile signaling data offer continuous but potentially sparse spatial coverage. Furthermore, multimodal data, including social media events, weather conditions, and even special event information, collectively form complementary information sources for depicting the state of the traffic system.

The underlying characteristics of the aforementioned data impose specific requirements on prediction techniques. Its strong spatiotemporal dependence means that data values are not independent; they are influenced by cycles, trends, and unexpected events on the time axis, while on the spatial dimension, they are constrained by road network connectivity and the propagation patterns of upstream and downstream traffic waves. The data may have a low signal-to-noise ratio, along with issues of missing values, anomalies, and asynchronicity, which constitute a direct challenge for data preprocessing and quality enhancement methods. Understanding and formally defining these characteristics is a prerequisite for constructing prediction models capable of fully exploiting the potential value of the data, as it determines the formulation of subsequent data fusion strategies and the design direction of the model architecture.

1.2 Exploration of Traditional Models and Their Limitations in Traffic Flow Prediction

Before the widespread application of big data technology, the field of traffic flow prediction primarily relied on a model system based on classical theories and statistical methods. Statistical approaches, such as historical average and time series models, performed extrapolation based on the stationarity and linear relationships of the data; they were computationally simple but rested on strict assumptions. Macroscopic simulation models, grounded in the physical laws of traffic flow, attempted to describe the aggregate behavior of traffic streams through analogies with fluid dynamics. In contrast, microscopic simulation models simulated traffic flow evolution by defining behavioral rules for drivers, such as car-following and lane-changing[1].

When these traditional models confront the complexity of modern urban traffic, their theoretical limitations have gradually become apparent. They typically rely on highly simplified abstractions of the system, making it difficult to accurately characterize its true nonlinear, stochastic dynamics. Model parameters often require complex calibration based on specific scenarios, resulting in limited generalization capability. More importantly, their model architectures struggle to effectively incorporate and process multi-source, high-dimensional, and massive real-time data inputs, leading to low information utilization efficiency. In terms of expressing the complex spatiotemporal correlations within large-scale road networks, traditional methods are mostly confined to local neighborhoods or predefined simple rules. They are unable to automatically learn deep, long-range spatiotemporal dependency patterns from the data, which has become the primary bottleneck hindering the improvement of prediction accuracy.

1.3 Evolutionary Trajectory and Technical Principles of Intelligent Prediction Algorithms

To address the bottlenecks of traditional models, the evolutionary trajectory of prediction algorithms has shown a development trend from relying on manual feature engineering to end-to-end automatic representation learning. Early machine learning methods, such as support vector machines and gradient boosting trees, enhanced predictive capabilities by introducing nonlinear kernels and ensemble strategies; however, their performance ceiling was bounded by the quality of feature construction. The rise of deep learning marked a new stage in intelligent prediction, with its core lying in the automatic extraction of hierarchical feature representations from raw data through multi-layer neural network structures.

Convolutional neural networks, through their mechanisms of local perception and weight sharing, map the road network topology onto grid data, efficiently capturing local spatial patterns and translation-invariant features. Recurrent neural networks and their variants, such as gated recurrent units, are specifically designed to process sequential data, capturing long-term dependencies in temporal dynamics through internal state memory. In recent years, graph neural networks have emerged as a cutting-edge direction in this field. They directly model the traffic network as a graph structure, with nodes representing road segments or regions and edges representing connections. Through message-passing mechanisms, they explicitly model spatial dependencies between any nodes, thereby depicting the fundamental nature of spatial propagation in traffic flow more naturally and accurately. The essence of this evolutionary trajectory lies in the continuously enhanced representational capacity of prediction models, enabling them to directly learn from big data and simulate the intrinsic complex dynamics of urban traffic systems.

2. Construction of Big Data Technology Methods for Intelligent Traffic Flow Prediction

2.1 Collection and Fusion Processing of Multi-Source Heterogeneous Traffic Data

The construction of high-quality prediction models depends on the systematic preprocessing and information fusion of raw data. The data collection phase requires the integration of diverse data streams from physical sensors, mobile terminals, and the cyberspace information domain. This includes structured data, such as frequency data from inductive loops and spatiotemporal trajectory coordinates of floating cars, as well as unstructured data, such as video streams from traffic cameras and text-based event reports. Raw data commonly suffer from issues like mismatched spatiotemporal scales, inconsistent sampling rates, heterogeneous formats, and uneven quality; direct utilization would introduce significant noise. Therefore, constructing a standardized data cleaning and augmentation pipeline is crucial. The process encompasses outlier detection and correction, missing value imputation, as well as data resampling and alignment for different temporal granularities.

The core objective of data fusion is to generate a consistent, complete, and information-rich unified spatiotemporal feature representation. This typically involves multi-level fusion strategies: at the data level, observations from different sources are mapped onto a unified road network partition and timestamp through spatiotemporal matching and coordinate transformation; at the feature level, embedding techniques in deep learning are utilized to transform categorical information, spatial coordinates, and temporal context into dense numerical vectors, which are then concatenated with continuous sensor readings; at the decision level, certain architectures design specific fusion modules to dynamically weigh the reliability contributions of different data sources under particular spatiotemporal contexts. Effective fusion processing not only improves the signal-to-noise ratio and dimensional consistency of the input data but also reveals, at a deeper level, the complex correlations in traffic states that are difficult to capture from a single data source, thereby laying a high-quality data foundation for subsequent modeling[2].

2.2 Deep Learning-Based Traffic Flow Prediction Model Architecture

The prediction model architecture, centered on deep learning, aims to capture the complex spatiotemporal dependencies governing traffic flow evolution directly from fused multi-source data through end-to-end learning. Current mainstream architectural paradigms primarily evolve around how to more effectively model spatiotemporal graph-structured data jointly. One widely adopted architecture is the spatiotemporal convolutional network, which typically couples graph convolutional networks for extracting spatial features with one-dimensional temporal convolutions or gated recurrent units for capturing temporal dynamics. Graph convolutional networks explicitly learn spatial dependencies between nodes by defining convolutional operations in the spectral or spatial domain on the topological structure of the traffic network, thus overcoming the limitations of traditional convolutional neural networks that rely on Euclidean grid data.

To address the inherent challenges of long-term dependencies and dynamic spatiotemporal correlations in traffic prediction, more advanced model architectures have introduced attention mechanisms and adaptive graph learning capabilities. The spatiotemporal attention mechanism allows the model to dynamically focus on relevant moments in the historical time steps and significantly influential spatial nodes within the road network, rather than treating all spatiotemporal inputs equally. This enhances the modeling capability for non-stationary patterns such as traffic peaks and unexpected events. The adaptive graph learning module further discards the fixed graph structure that relies on prior road network topology, enabling the model to infer potential, dynamic correlation strengths between nodes from the data itself. Consequently, it can capture implicit spatial relationships revealed by the traffic flow patterns themselves. Through stacking and combination, these architectural components form powerful nonlinear mapping functions, capable of distilling robust spatiotemporal representations from massive historical data.

2.3 Integration of Real-Time Data Stream Processing and Dynamic Prediction Algorithms

The ultimate application value of an intelligent prediction system is reflected in its capability to process real-time data streams and perform online dynamic prediction. This necessitates the deep integration of trained prediction models with stream computing frameworks to construct a low-latency, high-throughput online service pipeline. The data stream processing layer is typically implemented based on a distributed stream processing engine. It is responsible for continuously ingesting raw data streams from various sensors, executing parallel data parsing, filtering, cleaning, and preliminary window aggregation calculations, and then feeding the processed structured streams to the prediction model with extremely low latency.

The core of dynamic prediction algorithms lies in achieving online updates of model parameters and real-time execution of prediction inference. As traffic patterns may drift slowly over time, the performance of statically trained models will gradually degrade. Therefore, it is necessary to design online learning or periodic incremental learning mechanisms, enabling the model to fine-tune its parameters using the latest data streams to maintain prediction accuracy. At the inference level, the model needs to be highly optimized to achieve millisecond-level response times, which involves engineering techniques such as model pruning, quantization, and dedicated hardware acceleration. The entire integrated system requires the design of efficient resource scheduling and fault-tolerance mechanisms to ensure the stability and reliability of the prediction service under high-concurrency data input, thereby forming a closed-loop intelligent decision support capability, from real-time data inflow to prediction result output[3].

3. Analysis of Intelligent Traffic Flow Prediction Applications Driven by Big Data

3.1 High-Precision Short-Term Traffic Flow Prediction and Spatiotemporal Pattern Mining

The core application value of intelligent prediction models based on big data technology is primarily reflected in their capability for high-precision estimation of future short-term traffic flow. These models can perform rolling predictions of traffic parameters for the next five minutes to one hour with unprecedented spatiotemporal resolution. Their output is not merely a point estimate but typically also includes interval predictions or probability distributions for key indicators such as traffic volume, speed, and occupancy rate. The direct function of this high-precision prediction is to provide foresight for dynamic traffic state perception, enabling the system to identify trends of speed decline or flow saturation on critical road sections in advance. The improvement in prediction accuracy stems from the model's more accurate characterization of the nonlinear dynamics of microscopic traffic flow, for example, its ability to effectively capture complex fluctuations caused by signal phase changes, temporary bottlenecks, or the platooning and dispersion of vehicle convoys.

More profoundly, high-quality prediction results themselves constitute the data foundation for mining the intrinsic spatiotemporal patterns of the traffic system. Through comparative analysis of a large number of historical prediction sequences and actual data, combined with the representations learned internally by the model, one can inversely analyze the spatiotemporal evolution patterns of traffic flow. This includes identifying the recurring sources and propagation pathways of congestion within the road network, quantifying the statistical regularities of traffic states under different time periods and weather conditions, and discovering latent traffic correlations between specific regions. These patterns transcend traditional empirical knowledge, providing data-driven insights into the relationship between the macroscopic behavior and microscopic mechanisms of urban traffic systems. Consequently, prediction is transformed from a mere tool for 'foretelling' into an analytical means for 'understanding' the complexity of the system.

3.2 Simulation and Deduction of Traffic Flow Propagation in Complex Road Network Environments

Based on the capability for high-precision short-term prediction, a further functional extension of intelligent prediction models lies in the simulation and deduction of traffic flow propagation processes within complex road network environments. This is not traditional micro-simulation but rather utilizes the trained deep learning model as a simplified 'digital twin' kernel. Given the initial road network state and boundary conditions, it performs multi-step forward iterative prediction, thereby simulating the spatiotemporal evolution process of traffic congestion or unexpected events. This deductive capability enables analysts to assess system-level impacts under specific disturbances, for example, simulating how congestion forms, spreads, and ultimately reaches a new equilibrium on surrounding alternative routes after the closure of a major intersection[4].

The key to this process lies in the spatiotemporal dependency modeling capability of the model.

Advanced graph neural network models, through learning, essentially encode the dynamic influence weights between nodes in the road network. When performing deduction, the model can automatically calculate, based on the current state of the entire network, the extent to which each road segment will be influenced by its upstream, downstream, and even more distant segments at the next moment, thereby achieving a realistic simulation of traffic wave propagation. This deduction, grounded in data-driven learning, circumvents the need for calibrating numerous detailed behavioral parameters required by traditional simulation models. Its deduction results more directly reflect the macroscopic propagation patterns inherent in historical observational data, providing a powerful computational experimentation platform for assessing road network resilience, analyzing the causes of bottlenecks, and predicting the potential effects of management measures.

3.3 Performance Evaluation and Uncertainty Analysis of Prediction Systems

The deployment and application of any intelligent prediction system are inseparable from systematic performance evaluation and rigorous analysis of the uncertainty in its outputs. Performance evaluation necessitates the construction of a multi-dimensional, scenario-specific indicator system that goes beyond a single mean absolute error or root mean square error. This includes examining the model's prediction stability across different temporal granularities and various traffic states, such as the differences in error distribution during off-peak versus peak periods, and under free-flow versus congested conditions. Evaluation in the spatial dimension is equally important; it requires analyzing the model's predictive performance on roads of different classes and nodes at different topological positions to identify potential weaknesses of the model. The evaluation process should also encompass consideration of the model's computational efficiency and resource consumption to ensure it meets the real-time requirements of online applications[5].

Uncertainty analysis of prediction results is the crucial bridge connecting model outputs with subsequent scientific decision-making. Given the inherent stochasticity of traffic systems, noise in the data collection process, and the simplifications of the model itself, any prediction is inevitably accompanied by a certain degree of uncertainty. Advanced intelligent prediction models should possess the capability to characterize this uncertainty quantitatively. This can be achieved, for example, by outputting prediction intervals at different confidence levels through quantile regression techniques, or by estimating the posterior probability distribution of predicted values using Bayesian deep learning frameworks. A thorough analysis of the primary sources of uncertainty is essential. These sources can typically be attributed to noise and missing data at the data quality level, approximation errors between the model structure and complex real-world relationships, and unforeseen external events within the traffic system that are difficult to anticipate or model. By systematically quantifying, attributing, and visualizing uncertainty, the credibility and practical value of prediction results can be significantly enhanced. This provides critical risk boundary information for downstream decisions based on predictions, such as route guidance, signal control, or demand management, thereby enhancing the overall reliability and scientific rigor of intelligent transportation system applications[6].

Conclusion

This study systematically elaborates on the application framework of big data technology in intelligent urban traffic flow prediction, covering the entire chain from data characteristic analysis and intelligent algorithm principles to the construction of specific technical methods and application analysis. The research indicates that the multi-dimensional heterogeneity and strong spatiotemporal correlation characteristics of traffic big data necessitate fundamental innovation in prediction methods. Model architectures centered on deep learning, particularly graph neural networks and spatiotemporal attention mechanisms, can effectively integrate multi-source data and automatically capture complex nonlinear spatiotemporal dependencies within road networks. This enables high-precision prediction of short-term traffic states and dynamic deduction of congestion propagation processes. However, the practical effectiveness of intelligent prediction systems depends on high-quality data fusion processing workflows, efficient real-time computational integration, and systematic performance evaluation and uncertainty quantification of prediction results. Future research directions may focus on developing more interpretable prediction models, exploring adaptive prediction capabilities in few-shot or zero-shot scenarios, further deepening the integrated application of prediction uncertainty in dynamic decision-making, and achieving the construction of efficient, scalable real-time prediction systems in larger-scale urban road networks. Through continuous technological integration and innovation, big-data-driven intelligent prediction is expected to become a key enabling technology for understanding and optimizing complex urban traffic systems.

References

[1] Tian, M. (2025). Research on the application of big data technology in smart city research and planning. Urban Construction Theory Research (Electronic Version), (26), 216-218.

[2] Wang, H. L. (2025). Research on the application of smart transportation in urban traffic management. Transportation Manager World, (13), 70-72.

[3] Liu, G. H., Duan, L. Z., & Hu, H. C. (2025). Design and research of an intelligent operation and maintenance system for rail transit signals based on big data technology. Inner Mongolia Science Technology & Economy, (07), 134-138.

[4] Li, W. Y. (2025). Design and application of an intelligent transportation system based on the Internet of Things and big data technology. China Telecommunications Trade, (02), 77-80.

[5] Peng, J. F. (2024). Research on urban traffic system congestion mitigation strategies based on big data technology. Transportation Manager World, (29), 58-60.

[6] Xiao, B. Y. (2020). Research on short-term traffic flow prediction for urban roads based on big data technology. China New Telecommunications, 22(15), 125-126.