Dean&Francis

ISSN 2959-6157

Short-term Road Traffic Flow Prediction based on LSTM and Transformer Models

Lubo Zhou *

School of Technology, Tianjin University of Finance and Economics, Tianjin,300000, China
Corresponding author: z15522352678@outlook.com

Abstract:

With the acceleration of urbanization, traffic congestion has become a prominent problem that hinders the operational efficiency of cities and affects the travel experience of residents. Accurate short-term traffic flow prediction is the support for intelligent transportation systems to achieve early prediction and congestion warning. Traditional prediction models have difficulty capturing the nonlinear and time-series dependent characteristics of traffic flow, resulting in limited prediction accuracy. This paper takes the 2-minute traffic flow data of the Creteil Ring Road in France from 7:00 to 9:00 as the research object. It adopts a Long Short-Term Memory (LSTM) and a Transformer model to preprocess traffic flow data, construct a standardized data set, and conduct comparative experiments on LSTM, Transformer models. The experimental results are presented intuitively through flow time series comparison diagrams, training loss curve diagrams, and model evaluation index comparison diagrams. The mean absolute error (MAE) of the LSTM model is 0.65, the root mean square error (RMSE) is 0.82, and the coefficient of determination(R^2) reaches 0.9061. The predicted values are highly consistent with the actual traffic flow fluctuations, and the error distribution is concentrated. Research shows that compared with the Transformer model, the LSTM model has more advantages in short-term traffic prediction tasks. These models can provide precise data support for practical applications such as traffic signal timing optimization and route planning, and provide technical references for the efficient operation of intelligent transportation systems.

Keywords: LSTM, Transformer, Short-term road flow.

1. Introduction

With urban traffic demand skyrocketing, transportation pressure surging, and urban populations expanding rapidly, city traffic networks have grown increasingly complex. Under the goal of sustainable urban development, urban traffic management faces unprecedented challenges. The surge in demand poses severe challenges to traffic system operations, making advanced passenger flow prediction crucial. Traffic systems continuously generate massive data streams around the clock [1]. Only by fully leveraging this data can people effectively optimize traffic light timing and provide congestion alerts. Intelligent transportation systems play a fundamental role in solving these challenges through accurate traffic prediction [2].

Traditional statistical models include historical average models, the least squares method, ARIMA, logistic regression, Kalman filtering and k-nearest neighbor models. Traditional prediction methods often fail to capture the complex, nonlinear and dynamic patterns in urban rail transit systems [3]. With continuous advancements in machine learning, deep learning—a key branch of this field—has achieved rapid development. Notable examples like recurrent neural networks (RNN), convolutional neural networks (CNN), and graph convolutional networks (GCN) have been widely adopted in natural language processing and pattern recognition [4]. These capabilities also unlock significant potential for deep learning models in transportation applications, enabling the development of more accurate predictive models.

LSTM and Transformer models are the most popular choices for traffic flow prediction. Sengupta et al. proposed a Markov-LSTM model capable of learning complementary features in traffic data for traffic flow prediction [5]. Zhang et al. built an LSTM model integrating transfer learning to effectively predict the traffic flow conditions of training samples with missing values [6]. Jang et al. explored more features and multi-layer models by using LSTM and GRU models to improve the accuracy of traffic flow prediction [7]. Shen et al. created TrafficPPT based on the Transformer model, aiming to model traffic flow as the distributed aggregation of trajectories [8]. Wang et al. proposed a spatiotemporal traffic prediction method based on the Transformer model, utilizing graph convolution layers to capture spatial correlations in traffic flows and identify short-term explosive correlations among users [9]. Liang et al. developed an interactive perception trajectory predictor based on Transformer models, which was further trained using limited interaction data through transfer learning. The predictor was then integrated into the path planning process [10].

Deep learning methods based on LSTM and Transformer models have been widely applied in urban traffic flow prediction. This study attempts to apply these two models to urban traffic flow forecasting research. Using vehicle travel data from the roundabout in Creteil, France, during the morning two-hour period (7:00-9:00), this paper compared two algorithmic prediction models. By analyzing the forecast results, this study provides a reference for travel planning. The research offers risk prediction for traffic management and travel directions, assisting drivers in planning routes in advance to avoid traffic congestion, thereby improving tourists‘ travel efficiency and convenience. It also helps authorities alleviate road congestion pressure, ensures normal road scheduling and operations, and enables the formulation and implementation of reasonable plans for high passenger flow. This holds significant meaning and value for building smart transportation systems, addressing urbanization challenges, and achieving efficient urban traffic operations.

2. Methods

2.1 Data Source

The experimental data for this study were sourced from the traffic operation system of the Creteil Ring Road in France. The data collection period was from 7:00 to 9:00 (a total of 2 hours). The original data consisted of traffic flow records at the second level, which were adapted to the time series modeling requirements of the model.

Firstly, the second-level original traffic trajectory data was resampled at a 2-minute interval to aggregate the cumulative number of passing vehicles within each time window. After selecting the target lane with the highest traffic volume, the traffic flow time series data set (unit: vehicles) was finally formed. Subsequently, data standardization processing was carried out. The flow sequence was scaled to the [0,1] interval using Min-Max normalization, and the sliding window method was used to construct the model input samples. The window length was set at 12, meaning that the traffic flow data of the previous 12 consecutive time steps were utilized to predict the traffic flow value at the 13th time step. Finally, the processed data set was divided into a training set, validation set and test set in a ratio of 7:1.5:1.5, which were used for model parameter training, training process optimization and final performance evaluation to ensure the reliability and comparability of the experimental results. Table 1 shows the information for the data.

ISSN 2959-6157

Table 1. Information for the data

Field name	Meaning	Data type
Time	2-minute timestamp	int
Vehicle ID	Lane marking	string
Total flow	Vehicles every 2 minutes	int
Average speed	Average speed of vehicles every two minutes(m/s)	float

2.2 Method Introduction

2.2.1 LSTM Model

Long Short-Term Memory (LSTM) has a keen ability to analyze temporal information and has become a core tool for analyzing traffic data (such as traffic volume, vehicle speed, congestion duration, etc.). The core of its method is to convert traffic temporal data into a form that the model can learn, analyze the gating mechanism, explore the temporal correlations in the data, and achieve traffic state prediction or anomaly identification.

The LSTM model adopts a stacked structure, suitable for the temporal prediction requirements of a single traffic flow feature. Input layer: receives temporal data, with dimensions corresponding to 12-time steps (2 minutes per step) and 1 traffic feature, matching the sample format generated by data processing. Two layers of long short-term memory units and two layers of dropout units are combined to enhance feature extraction and suppress overfitting. The first layer of long short-term memory units has 48 hidden units, outputs the complete temporal sequence, and configures a dropout ratio of 0.1. Then it is connected to a dropout layer with a dropout ratio of 0.25, randomly inhibiting redundant neuron connections; the second layer of long short-term memory units has 32 hidden units, outputs global temporal features, and also configures a dropout ratio of 0.1. The subsequent dropout layers (with a dropout ratio of 0.25) further optimize the regularization effect. The fully connected layer contains 16 neurons, uses the RELU activation function to enhance the nonlinear fitting ability, and introduces L2 regularization to suppress parameter redundancy. The output layer has 1 neuron, uses a linear activation function, and directly outputs continuous traffic flow prediction values. The model uses the Adam optimizer, monitors the loss of the validation set through early stopping to balance training efficiency and avoid overfitting. Figure 1 shows the LSTM model flowchart.

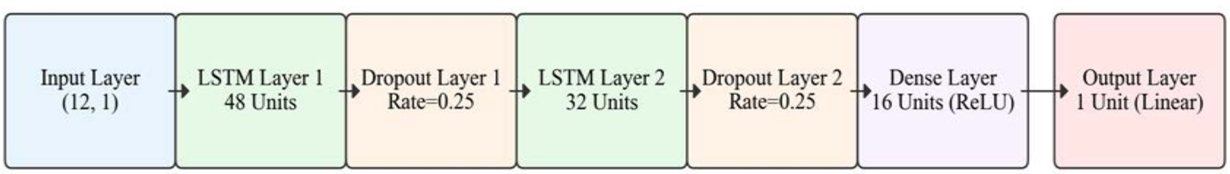

Fig. 1 LSTM Model Flowchart used in this article (Picture credit: Original)

2.2.2 Transformer Model

The Transformer model employs the self-attention mechanism, achieving global dependency modeling of temporal information through parallel computing, and adapting to the feature extraction of long sequence data. This study presents a lightweight improvement for short-term traffic flow prediction. The input layer is consistent with the LSTM model, receiving (12, 1) dimensional temporal input.

The fully connected layer increases the input feature dimension from 1 to 16, providing sufficient feature expression space for the self-attention mechanism. Using the attention mechanism, the attention output is then connected to the Dropout layer (dropout rate = 0.2), and through a residual connection (Add) and layer normalization (to maintain gradient stability). The first fully connected layer contains 32 neurons, the second layer maps back to a 16-dimensional feature space, and after output, it is connected to the Dropout layer (dropout rate = 0.2), also through residual connection and layer normalization to optimize training stability. Global average pooling is used to aggregate temporal features, and the output prediction value is generated through a single neuron fully connected layer. The model optimizer, loss function, and training strategy are consistent with the LSTM model to ensure the fairness of the comparison experiment. Figure 2 shows the Transformer model flowchart.

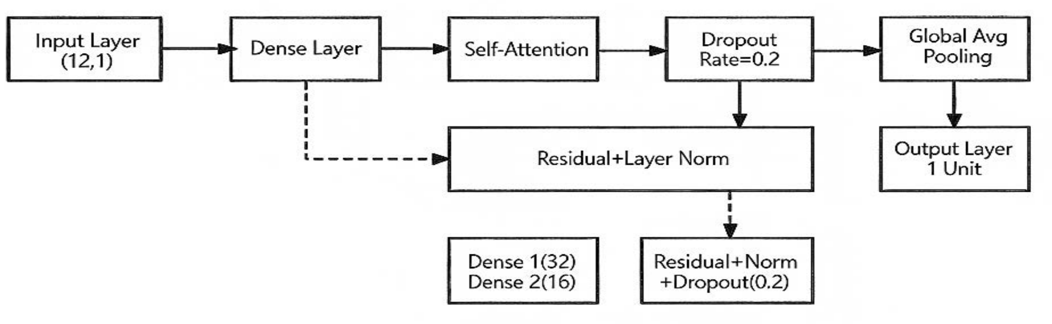

Fig. 2 Transformer Model Flowchart used in this article (Picture credit: Original)

3. Results and Discussion

3.1 Model training process

In the early stage of LSTM model training, both the training loss and the validation loss decreased rapidly. After training for 5 rounds, the loss curve gradually stabilized, and the difference between the training loss and the validation loss was small, indicating that the model did not fit and had good generalization ability. The Transformer model overfitted in the small sample training, resulting in a decline in performance. The LSTM model performed better in short-term traffic flow prediction. Figure 3 shows the comparison of losses for the two models.

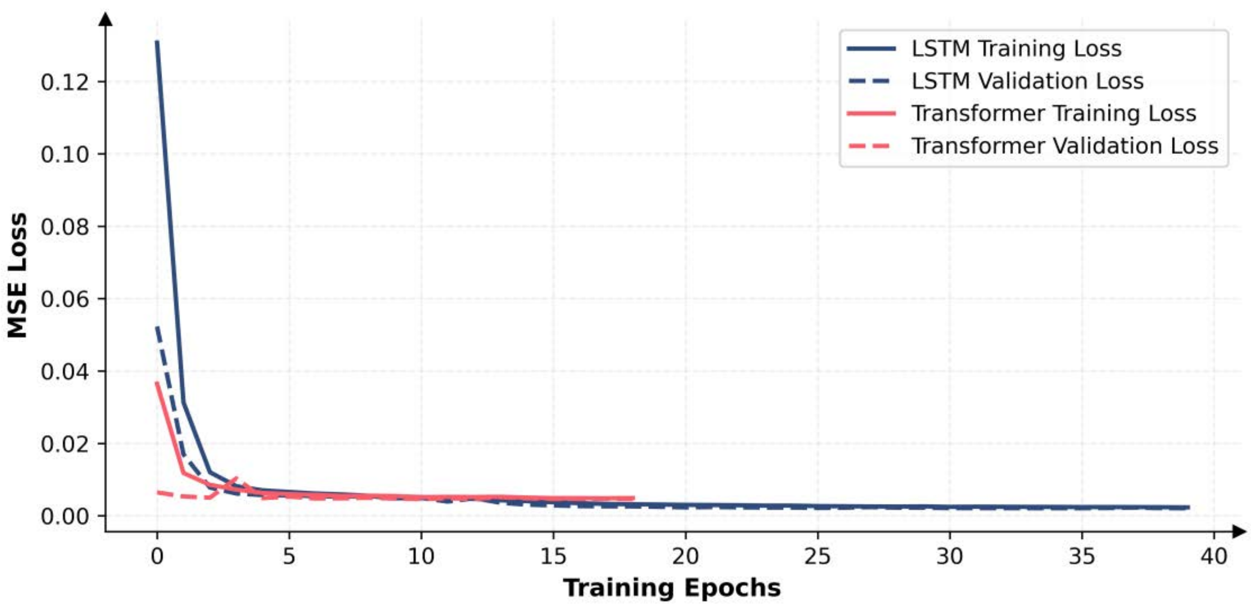

Fig. 3 Comparison of Model Loss (Picture credit: Original)

3.2 Traffic flow forecast results

The following two figures show the traffic flow prediction results of the test set for the LSTM model and the Transformer model. The red curve represents the actual flow, and the blue curve represents the predicted flow. It can be seen that the fluctuation trends of the prediction curves of the two models are highly consistent with the actual curves, and they can accurately capture the peak and trough changes of the flow. However, the fitting ability of the LSTM model is slightly better than that of the Transformer model. These two figures also verify the capturing ability of the two models for the temporal characteristics of traffic flow. Figures 4 and 5 show the traffic flow forecast results of two models.

ISSN 2959-6157

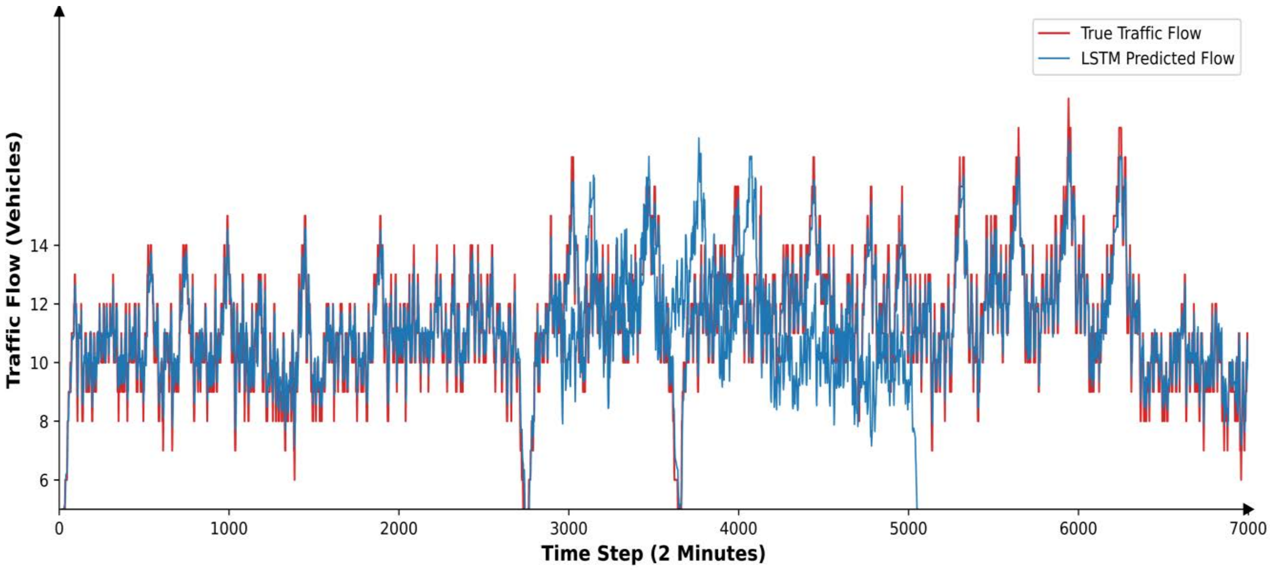

Fig. 4 Prediction of LSTM Traffic Model (Picture credit: Original)

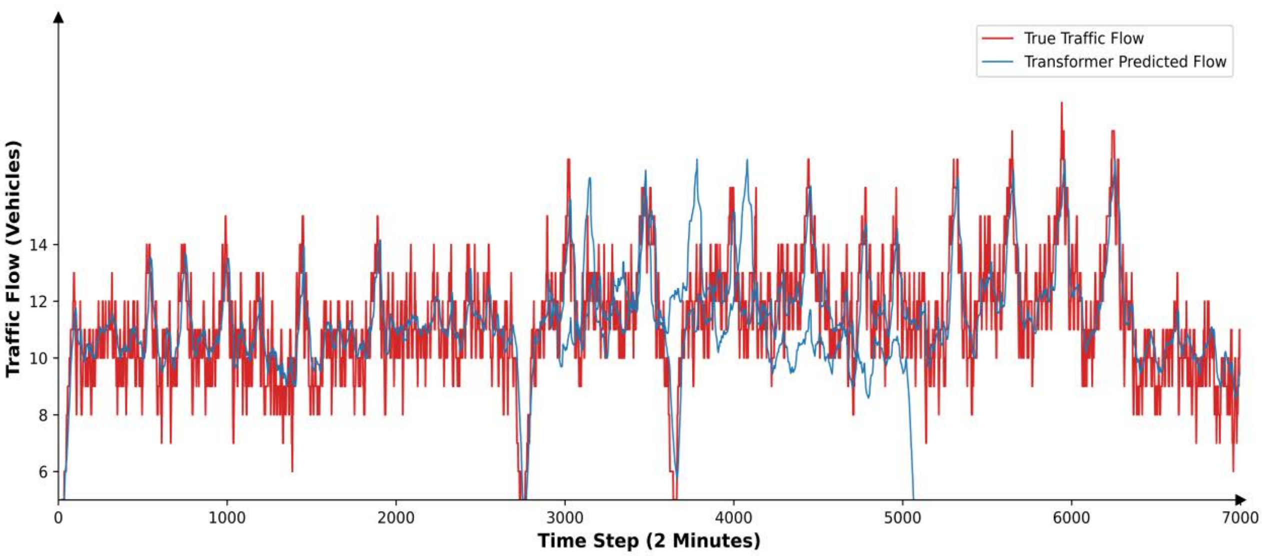

Fig. 5 Prediction of Transformer Traffic Model (Picture credit: Original)

3.3 Prediction Error Distribution

The figure below shows the comparison of the error distribution of the test sets for the two models. The errors of the LSTM model are mainly concentrated within the range of [-1, 1], and the number of samples near 0 is the largest, indicating that the prediction results of the model are overall close to the true values, with a relatively concentrated error distribution and no obvious deviation. The error distribution of the Transformer model is more scattered, and the probability of large errors is higher. Overall, the prediction accuracy and stability of the LSTM model are better. Figure 6 shows the prediction error distribution of two models.

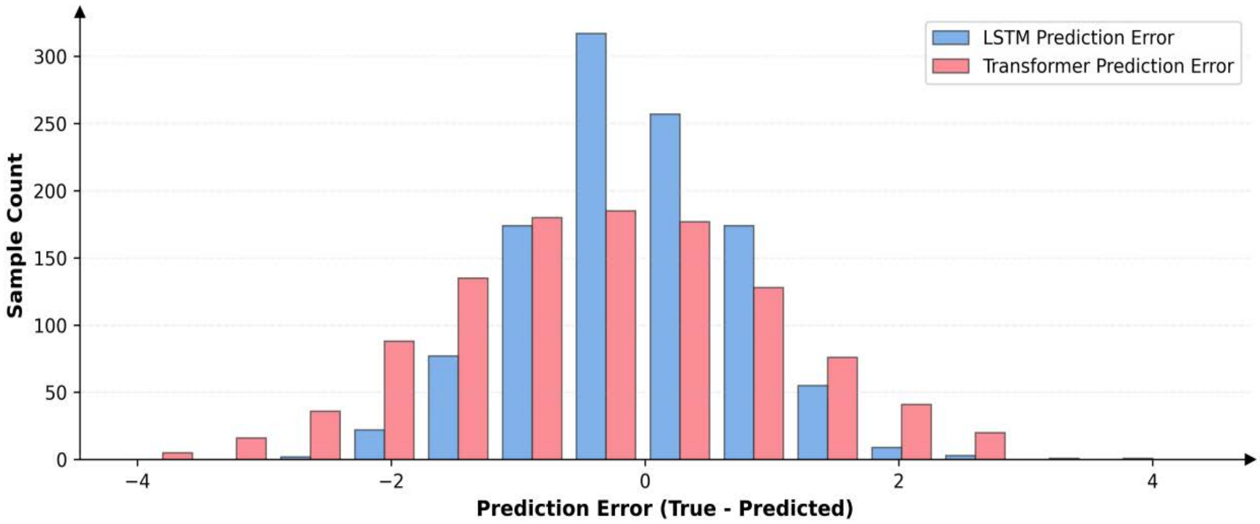

Fig. 6 Comparison of Model Error (Picture credit: Original)

3.4 Quantitative performance indicators

From the indicators, it can be seen that the MAE and RMSE of the LSTM model are both lower than those of the Transformer model, and the R^2 value is closer to 1, indicating that the prediction results of the LSTM model fit the true values more closely. This shows that its prediction accuracy is higher in the short-term traffic flow prediction task. Table 2 presents the quantitative evaluation indicators of the two models on the test set.

Table 2. Indicators for evaluation

	MAE	RMSE	R^2
LSTM	0.65	0.82	0.9061
Transformer	1.05	1.29	0.7688

4. Conclusion

This paper takes the 2-minute traffic flow data of the Creteil Ring Road in France from 7:00 to 9:00 as the research object, and builds a short-term traffic flow prediction model based on the LSTM model and the Transformer model. The experimental results show that the MAE of the LSTM model on the test set is 0.65, the RMSE is 0.82, and the R^2 is 0.9061. The predicted values are highly consistent with the time series fluctuation trend of the actual traffic flow, can effectively capture the changes in traffic peaks and troughs, and the error distribution is concentrated without systematic deviations. This fully validates the excellent ability of the LSTM model to capture the nonlinear time series characteristics of traffic flow in improving the generalization ability of the model. However, the self-attention mechanism of the Transformer model has not fully demonstrated its advantages in short sequence scenarios.

The principal limitations of this study include the difficulty in achieving real-time prediction and a susceptibility to overfitting. This study does not adequately account for the inter-lane traffic flow interactions, such as road segment congestion levels. The Transformer model exhibits delayed responsiveness to short-term traffic variations, while the generalization capability of the LSTM model requires further improvement.In the future, multiple features can be further integrated to improve the model architecture to enhance prediction accuracy and scene adaptability, providing more reliable technical support for the practical decision-making of intelligent transportation systems.

References

[1] Zhao S Z, Ni T H, Wang Y, et al. A new approach to the prediction of passenger flow in a transit system. Computers & Mathematics with Applications, 2011, 61(8),1968–1974.

[2] Chaoura C, Lazar H, Jarir Z. Traffic flow prediction at intersections: Enhancing with a hybrid LSTM-PSO approach. International Journal of Advanced Computer Science &

ISSN 2959-6157

Applications, 2024, 15(5).
[3] Kang Y. Short-term passenger flow prediction in urban rail transit based on hybrid deep learning models. Highlights in Science, Engineering and Technology, 2024.
[4] Xue Q, Zhang W, Ding M, Yang X, Wu J, Gao Z. Passenger flow forecasting approaches for urban rail transit: A survey. International Journal of General Systems, 2023, 52(8), 919–947.
[5] Sengupta A, Das A, Guler S I. Hybrid hidden Markov LSTM for short-term traffic flow prediction. arXiv preprint, 2023.
[6] Zhang Z, Yang H, Yang X. A transfer learning–based LSTM for traffic flow prediction with missing data. Journal of Transportation Engineering, Part A: Systems, 2023, 149(10),04023095.
[7] Jang H C, Chen C A. Urban traffic flow prediction using LSTM and GRU. Engineering Proceedings, 2024, 55(1),86.
[8] Shen S, Pan B, Xue G. A pretrained probabilistic transformer for city-scale traffic volume prediction. arXiv preprint, 2025.
[9] Wang F, Xin X, Lei Z, et al. Transformer-based Spatio-Temporal traffic prediction for access and metro networks. Journal of Lightwave Technology, 2024, 42(15),5204–5213.
[10] Liang J, Tan C, Yan L, et al. Interaction-aware trajectory prediction for safe motion planning in autonomous driving: A transformer-transfer learning approach. 2024.