

A Summative Analysis of Deep Neural Networks in Traffic Flow Prediction

Xiao Hu *

School of Mathematics, University of Edinburgh, Edinburgh, UK

* Corresponding Author Email: S2161066@ed.ac.uk

Abstract. In the modern society, traffic congestion is a major problem that calls for the creation of intelligent transportation systems (ITS), of which precise traffic forecasting is an essential element. The use of deep learning architectures to the short-term traffic forecasting problem will be summarized in this study. Prior to thoroughly analyzing important deep learning models, such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNNs), and their hybrid variants, it first addresses the drawbacks of conventional statistical and machine-learning techniques. The analysis culminates in the discussion of Spatiotemporal Graph Neural Networks (STGNNs), arguing that deep learning, particularly through graph-based architectures capable of modeling complex non-Euclidean spatial relationships and temporal dynamics, has become the indispensable framework for accurate traffic forecasting. In its conclusion, the article discusses existing issues including computing demands and data quality and makes recommendations for future research, such as long-term prediction and dynamic graph learning.

Keywords: Traffic prediction, Deep learning, Spatiotemporal modeling, Graph Neural Networks, Intelligent Transportation Systems.

1. Introduction

In modern society, traffic has become one of the most important parts of a city, thus, traffic congestion has become a global challenge with profound economic, environmental, and social consequences. It leads to higher fuel use, more greenhouse gas emissions, billions of lost hours of productivity, and more stress for commuters. As urban populations continue to grow, the development of intelligent transportation systems (ITS) has become a critical priority for mitigating these issues. A cornerstone of effective ITS is accurate and timely traffic prediction, which enables proactive traffic management, dynamic route guidance, and improved urban mobility.

There are two types of traffic forecasts: long-term and short-term. Numerous outside variables, including weather and road construction, have an impact on long-term forecasting [1]. Therefore, short-term prediction is used more widely. Fig.1 illustrates the structure of current research on short-term traffic forecasting models.

This paper gives a summative analysis of the application of deep learning methods to the problem of traffic prediction. It will first delineate the limitations of traditional predictive methods before delving into a detailed examination of key deep learning models, including their operational principles, strengths, and weaknesses. The central thesis of this analysis is that deep learning, particularly through spatiotemporal graph architectures, has become the indispensable framework for accurate traffic forecasting, fundamentally transforming the capabilities of intelligent transportation systems.

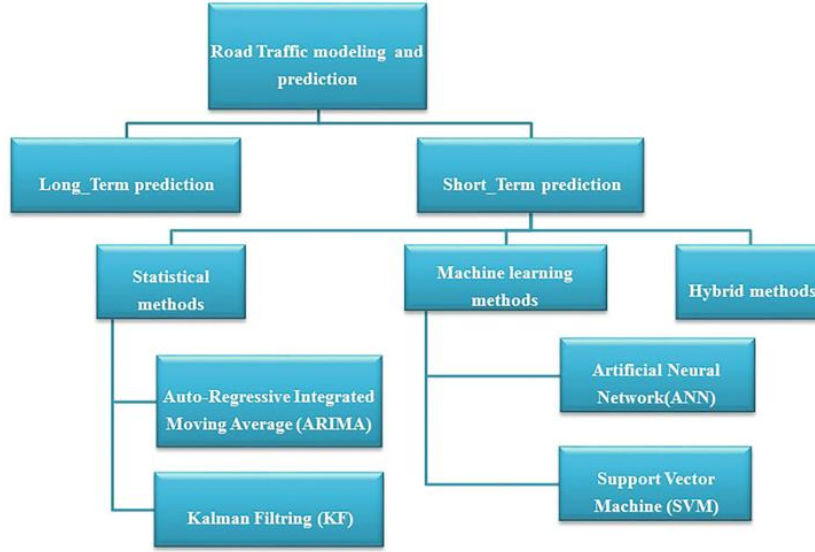


Fig 1. Structure of short-term traffic prediction

2. Limitations of Traditional Methods

The traditional methods of traffic prediction often use statistical models or machine-learning models. This section will give a brief introduction on both methods and discuss their limitations.

2.1. Statistical methods

The statistical methods build statistical time-series models using existing dataset or predicted data [2]. One of the most significant statistical models for traffic prediction is the autoregressive integrated moving average (ARIMA) model. Researchers use this model not only to predict but also as a measure model [3]. There are also other variants of the ARIMA model such as the seasonal ARIMA (SARIMA) introduced by Williams and Hoel in 2003. Their research proved that SARIMA gives a better performance and accuracy than ARIMA [3]. However, the statistical methods can only perform on small linear datasets and only use for single-location traffic predictions, while the reality is more complicated and non-linear. This leads to a less reliable forecast using only the statistical methods.

2.2. Machine-learning methods

Conventional machine-learning methods could handle more complex datasets and perform better than statistical methods. It includes broader categories such as Neural Network, Support Vector Regression (SVR) and Spinning network etc. [3]. The main topic of this paper, deep Learning, is also a branch of ML-learning methods, but it is not traditional methods and thus will not be included in this part.

3. Deep Learning Models for Traffic Predictions

3.1. Recurrent Neural Networks (RNNs)

The recurrent neural networks (RNNs) is an iterative learning machine that uses the same weights recurrently when handling sequential data [4]. Thus, the sequential nature of traffic data makes RNNs a natural starting point. It effectively maintains the prior input and permits information to persist from one time step to the next. Particularly, the RNNs model uses a transfer function to iterate. The general formula is showed below:

$$h_t = f(W_x(x_t + b_x) + W_h(h_{t-1} + b_h)), \quad (1)$$

where x_t is the input vector at time step t and h_{t-1} is the previous hidden state. W_x is the weight matrix for connecting the current input and the current hidden state. W_h is the weight matrix for connecting the previous hidden state and the current hidden state. b_x and b_h are bias vectors. Since the activation function f is typically non-linear, it ensures that the model can handle more complicated sequential input [4]. Fig. 2 shows an example of an RNNs:

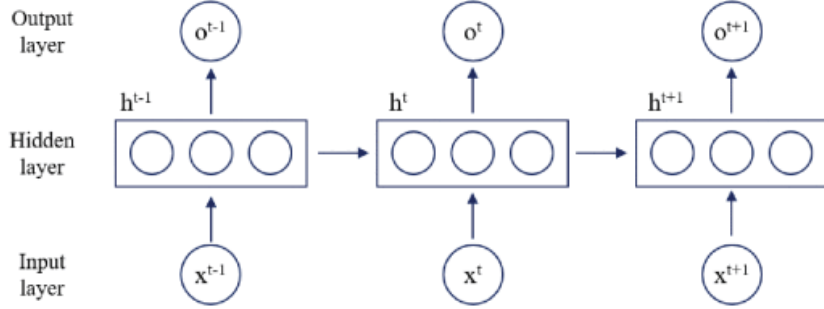


Fig 2. An example of RNNs

However, the vanishing gradient problem makes it difficult for RNNs to learn long-term dependencies. Hochreiter and Schmidhuber introduce Long Short-term Memory (LSTM) as a solution to this issue. This method uses memory blocks instead of hidden units and control the overall behavior within them [5]. Fig. 3 presents what a memory block is like.

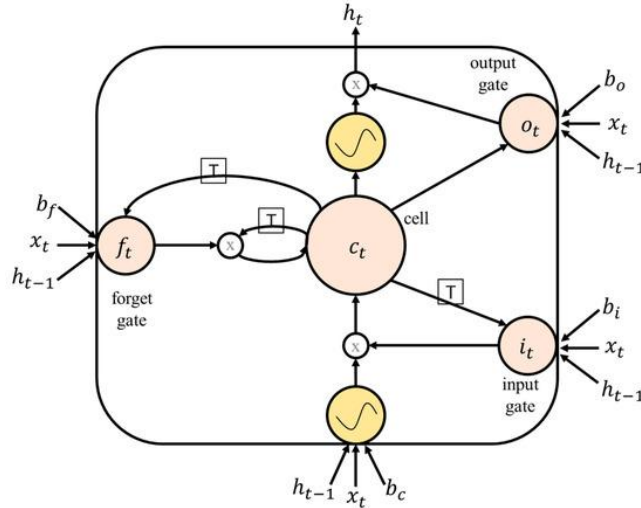


Fig 3. An illustration of a LSTM unit

The formulas for LSTM are:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \quad (2)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \quad (3)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad (4)$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C), \quad (5)$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t, \quad (6)$$

$$h_t = o_t \tanh(C_t), \quad (7)$$

where x_t is the input vector at time step t and h_{t-1} is the hidden state at time step $t - 1$. f_t , i_t , and o_t are the outputs of the forget gate, input gate and output gate [6].

LSTMs are highly effective for predicting the traffic state like speed or flow at a single sensor location by learning the temporal patterns of that location's historical data. A significant drawback of standard LSTMs is that they process data from each sensor in isolation. They are not able to

incorporate spatial information from related nodes in the road network, treating the entire traffic system as a collection of independent time series.

3.2. Convolutional Neural Networks (CNNs)

The convolutional Neural Networks (CNNs) is often used in image processing and suitable for extracting spatial features by applying learnable filters across input data. Unlike RNNs or other algorithms that can only deal with numerical data, CNNs could process visual data from traffic camera or GPS etc. directly.

Convolution and pool sampling are the central modules of CNN, a three-layer learning neural network with fully linked layers [7]. Fig. 4 shows an overall structure of CNNs.

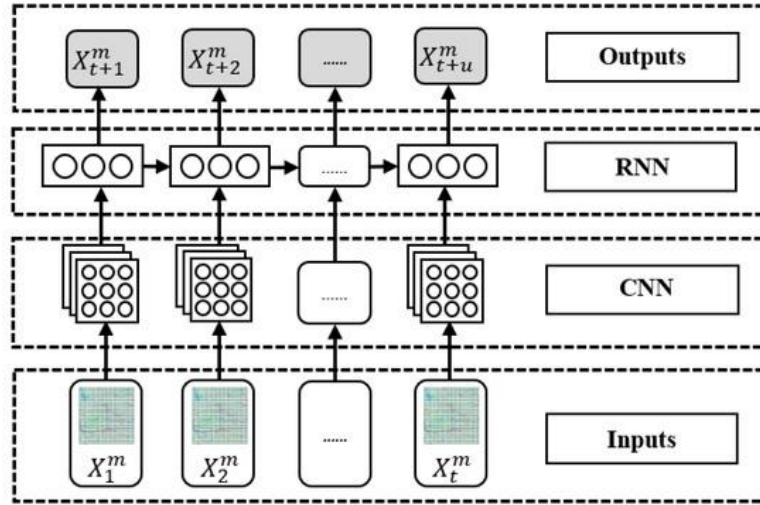


Fig 4. Structure of CNNs

The convolutional layer will supply the input layer's output neurons, which will include the pixel values of the visual data, by computing the scalar product between their weights and the region corresponding to the input volume. The input's spatial dimensionality and parameter count will thereafter be compressed by the pooling layer. Direct connections exist between neurons in the two adjacent layers and neurons in the fully-connected layers [7].

The primary limitation of this approach is its reliance on a Euclidean grid structure. Road networks are inherently non-Euclidean. They are graphs where the relationship between nodes (intersections) is not based on physical pixel distance but on connectivity and travel time. Standard CNNs struggle to model these irregular, graph-based relationships effectively.

3.3. Hybrid and Graph-based Architectures

When forecasting traffic, it is not enough to use just one single model to process the data, since the reality is much more complicate. Researchers began combining the models to increase forecast efficiency and dependability. For instance, by combining CNNs and RNNs, researchers might take use of CNNs' effectiveness in handling spatial information and RNNs' superiority in processing time series data [4]. Fig. 5 shows an example of the combined model.

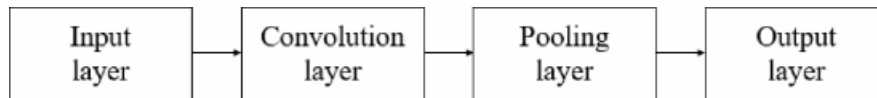


Fig 5. An example of RNNs + CNNs

The most significant advances have come from models designed to integrate spatial and temporal modeling from the outset, moving beyond treating them as separate problems. A straightforward yet powerful approach combines CNNs and LSTMs. The typical architecture uses a CNN layer to extract spatial features from a grid representation of the network at each time step. These spatial features are

then fed sequentially into an LSTM layer, which learns the temporal dynamics. This allows the model to learn how spatial features evolve over time.

The Graph Neural Networks (GNNs) is another popular model on dealing with graph-structured data. GNNs can be divided into Recurrent Graph Neural Networks (RecGNNs), Convolutional Graph Neural Networks (ConvGNNs), Graph Autoencoders (GAEs), and Spatial-Temporal Graph Neural Networks (STGNNs) [8]. GNNs is working by passing messages between nodes. After taking the message from adjacent nodes and computing a feature vector for each node, a readout function will give a feature vector of all nodes [9]. Spatio-Temporal Graph Neural networks (ST-GNNs) extends GNNs into traffic prediction [10].

4. Challenge and Future Directions

Despite the remarkable success of deep learning in advancing traffic prediction, it also has its limitations. The implementation of these models in real-world systems faces significant practical and theoretical challenges. First, the amount and caliber of data used to train deep learning models are related to their performance. However, the availability or quality of data is not always guaranteed, since sensor networks often have coverage gaps and the use of GPS or local data from mobile devices raises privacy issues. Also, large models such as GNNs require substantial computational resources for training. This can be barrier for resource-constrained transportation agencies and can limit the feasibility of real-time model updating on edge devices.

As mentioned at the beginning, current researches mostly focus on short-term predictions. This is a field that we still need to discover in the future. In addition, most current GNNs assume a static, pre-defined graph structure. However, traffic correlations are dynamic. Future models could learn the graph structure automatically from data adapting the spatial dependencies over time to reflect real-time conditions.

5. Conclusion

The transformative impact of deep learning on the subject of traffic prediction has been thoroughly summarized in this work. The core of this analysis detailed the deep learning revolution. We explored how Recurrent Neural Networks (RNNs), particularly LSTMs, mastered temporal dynamics, and how Convolutional Neural Networks (CNNs) began to address spatial feature extraction. The discussion culminated with the state-of-the-art: hybrid CNN-LSTM models and, more powerfully, Spatiotemporal Graph Neural Networks (STGNNs).

References

- [1] Alsolami B., Mehmood R., Albeshri A. Hybrid statistical and machine learning methods for road traffic prediction: A review and tutorial. *EAI/Springer Innovations in Communication and Computing*, 2019: 115–133.
- [2] Lohrasbinasab I., et al. From statistical- to machine learning-based network traffic prediction. *Transactions on Emerging Telecommunications Technologies*, 2021, 33(4).
- [3] Ma T., Antoniou C., Toledo T. Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transportation Research Part C: Emerging Technologies*, 2020, 111: 352–372.
- [4] He Y., et al. In-depth insights into the application of recurrent neural networks (RNNs) in traffic prediction: A comprehensive review. *Algorithms*, 2024, 17(9): 398.
- [5] Sameen M., Pradhan B. Severity prediction of traffic accidents with recurrent neural networks. *Applied Sciences*, 2017, 7(6): 476.
- [6] Li Wenqi, Liu Dongyu, Yang Menghua. A model of traffic accident prediction based on convolutional neural network. *Proceedings of the 2nd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, Singapore, 2017: 198–202.

- [7] O'Shea K., Nash R. An introduction to convolutional neural networks. 2015.
- [8] Piao H., et al. Complex relationship graph abstraction for autonomous air combat collaboration: A learning and expert knowledge hybrid approach. *Expert Systems with Applications*, 2023, 215: 119285.
- [9] Wu Y., et al. A hybrid deep learning based traffic flow prediction method and its understanding. *Transportation Research Part C: Emerging Technologies*, 2018, 90: 166–180.
- [10] Ta X., et al. Adaptive spatio-temporal graph neural network for traffic forecasting. *Knowledge-Based Systems*, 2022, 242: 108199.