Traffic Prediction Methods

Ta Jiun (Jeff) Ting

August 4, 2020

1 Introduction

1.1 Motivation

Traffic congestion incurs a significant cost on the economic prosperity of a region. This cost not only materializes as opportunity cost to the travellers due to time delays but also as negative externality due to increased accident rate and vehicle emissions [1]. In the Greater Toronto and Hamilton Area, the total cost of traffic congestion is $6 billion in 2006, and this cost will be over $15 billion by 2031 [1]. Clearly, measures to reduce traffic congestion can result in significant economic benefits.

One of the main goals of intelligent transportation systems (ITS) is to alleviate road congestion through better utilization of the road network. ITS attains this goal through the use of real-time traffic management technology, such as variable speed limit and adaptive traffic signal control. Accurate traffic prediction is essential in achieving this goal as it allows these management technologies to be proactive rather than reactive. In adaptive traffic signal control, the adaptive system uses the predicted traffic information to optimize the phase timing, thus accurate predictions allow the adaptive system to respond optimally [2]. Similarly, in the case of variable speed limits, accurate prediction allows the system to better assess the impact of its decisions and set the optimal speed limit [3]. Accurate prediction also allows users to make more informed travel decisions and make timely adjustments in response to traffic incidents and congestion. Altogether, accurate prediction leads to reduced road congestion by better managing the demand for a transportation system.

1.2 The Traffic Prediction Problem

Urban traffic is a highly-dynamic process evolving over space and time. For a stretch of highway, traffic conditions at a given location influence its downstream sections as traffic moves forward. Meanwhile, traffic conditions at a given location also influence its upstream sections as traffic congestion propagates backwards via shock waves, a phenomenon known since the 1950s [4]. These two basic processes are constantly present on any road network, evolving continually in response to the changes in traffic conditions. In the urban setting, the presence of intersections and traffic signals further influences the dynamics of traffic patterns and complicates the prediction problem.

We can perform traffic prediction at the level of individual vehicles or the level of road links. The prediction methods at the two different levels are very different. At the individual vehicle level, the predictor models the behaviour to predict the future trajectories of each vehicle. Meanwhile, at the link level, the predictor only focuses on the macroscopic properties on each road and predicting their evolution over time. For this document, we define the prediction problem as link-level prediction and the task of predicting individual vehicle properties is considered out of scope.

We can represent the traffic properties on each link with a variety of variables, including:

- flow (number of vehicles per unit time)
- speed (distance covered per unit time)
- travel time (time to traverse a given link)
- occupancy (percentage of time the road is occupied by vehicles)
- density (number of vehicles per unit distance)
- headway (distance or time between vehicles)

However, typically only a small subset of these properties are available to the transportation agency depending on the sensors installed. Therefore, traffic prediction typically only predicts flow or speed depending on the source of data. Loop detectors installed in the road can produce accurate vehicle count data that allows us to calculate flow by counting the number of vehicles passing through in a unit time; however, it is impossible to obtain information such as speed or density with a single loop detector since they require observation over a distance rather than a single point. Meanwhile, GPS or Bluetooth data contains locations and timestamps that allows us to calculate average vehicle speed by dividing distance traversed over time elapsed; however, flow and density information from this data can be inaccurate since some vehicles may lack the technology and would be unaccounted for in the data.

The time horizons of the predicted values depends heavily on the time granularity of the available data (they are the same in the majority of cases). Even though traffic patterns evolve continuously on a road network, data are usually aggregated and discretized over time as they are collected. This is done to reduce the size of the dataset and to reduce variability in the data by increasing the sample size of each period. Consequently, the time horizons of the prediction model need to be a multiple of the time granularity of the dataset; otherwise, we cannot easily verify the accuracy of the prediction model using the available data.

1.3 Objectives and Evaluation Metrics

The end objective of a traffic prediction method is to generate accurate predictions of future traffic states by employing currently available information. This information may include past observations, road network characteristics, as well as any useful external information such as weather and the presence of traffic events. The output of the model is the predicted state values of each location in the road network over each time slice in the prediction horizon.

The goal of a predictive model is to minimize the difference between the predicted state values and the actual state values. To quantify this numerically, researchers typically use a combination of the following metrics to assess the performance of predictive models: mean absolute error (MAE), mean absolute percentage error (MAPE), mean-square error (MSE), and root-mean-square error (RMSE). Given a sequence of N predictions $\{\hat{X_1}, ..., \hat{X_N}\}$ and the actual observed value $\{X_1, ..., X_N\}$, MAE is the average of the absolute error while MAPE is the average of the absolute relative error. The calculation is shown below in equations (1) and (2):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|X_i - \hat{X}_i\right| \tag{1}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{X_i - \hat{X}_i}{X_i}\right| \tag{2}$$

Alternatively, MSE is the average of squared errors while RMSE is the square root of MSE. The calculation is shown in equations (3) and (4):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{X}_i)^2 \tag{3}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \hat{X}_i)^2} \tag{4}$$

1.4 Outline

This paper divides traffic prediction methods into three sections. Section 2 outlines the tradition methods that derive the model from observations. Afterwards, Section 3 examines the data-driven statistical models of traffic prediction. Lastly, Section 4 discusses the more elaborate deep learning methods in recent literature.

2 Traditional Methods

Traffic prediction and modelling has been studied since the 1950s. This section covers the traditional methods which model traffic dynamics explicitly using the observed behaviours of traffic. In contrast, the newer methods model traffic by employing statistical techniques such as regression analysis, time series analysis, or deep learning. The methods covered in this section can be roughly divided into two categories: macroscopic approach and microscopic approach.

2.1 Macroscopic Approach

The macroscopic approach to traffic modelling treats the stream of traffic on a road link as a collective, homogeneous entity evolving over time due to influences from upstream and downstream links. In 1956, Lighthill and Whitham [5] pioneered this approach by creating the kinematic wave model. This model observes the phenomenon of traffic waves propagating slowly along a road as similar to the kinematic waves in fluid dynamics. Consequently, this model applies the kinematic wave equations to describe the change in flow, density, and speed along a road. The notion of traffic shock waves augments this model by explaining the cause of congestion accumulation and dissipation on a road [4]. One drawback to the macroscopic approach is the view of a traffic stream as a homogeneous entity, which falls apart under light traffic conditions where vehicles can move at different speeds. Nevertheless, the simplicity of this model allows a numerical solution to be described by the cell transmission model [6], which is also applicable to the analysis of complex road networks [7].

In the 1970s, there were attempts to improve the first-order kinematic wave model by approximation with a second-order fluid dynamics model [8], [9]. The second-order model introduces a momentum equation capture the finer dynamics of the fluid-like behaviour of traffic. However, the fundamental differences between traffic and fluids cause second or higher-order fluid dynamics models to produce unrealistic results [10]. Nevertheless, there is ongoing research in improving second-order models by relaxing the assumptions and constraints applicable only to fluids [11].

2.2 Microscopic Approach

The microscopic approach to traffic modelling considers the collective behaviour of a traffic stream stems from the behaviour of individual vehicles within the stream. A car-following model achieves this by describing how vehicles act relative to its preceding vehicle. Mainly, each vehicle will maintain a gap with its preceding vehicle by changing its speed based on the behaviour of the leading car. [12]–[14] are examples of this model, and the modelling results are in agreement with the macroscopic models discussed in Section 2.1 [15]. This model is more descriptive than the kinematic wave model since it relates the macroscopic traffic properties to the behaviour of individual vehicles. However, this also increases model complexity significantly, and it is difficult to tune the model parameters to achieve a realistic driver behaviour.

3 Statistical Models

As data collection and processing technology improved over the years, a new class of traffic prediction models emerged when researchers started using statistical models to describe traffic in the 1970s. In contrast to the methods described in Section 2, these methods do not model traffic dynamics explicitly. Instead, these methods use the available data to estimate the parameters of the statistical model.

3.1 Time Series Analysis

We can view the macroscopic traffic properties on a road evolving over time as time series. Naturally, researchers have applied the techniques of time series analysis to traffic modelling and prediction.

3.1.1 Spectral Analysis

Traffic patterns exhibit a complex periodic behaviour. By assuming this periodic behaviour of traffic to be a combination of modal functions of different frequencies, we can discover the underlying modal functions using spectral analysis. We can then specify a traffic prediction model by estimating the covariance of the modal functions. [16], [17].

3.1.2 Autoregressive Models

In contrast to the frequency domain methods, autoregressive models for time series analysis considers each observation to be correlated with its previous observations. The autoregressive moving average (ARMA) model is a class of models that is extensively researched in this category.

The ARMA model combines the autoregressive component and the moving average component. The autoregressive component assumes the observation for location X at time t is a linear combination of past p observations at the same location, plus a constant and a white noise error term. This is described in Equation (5), where c is the constant, ϵ_t is the error term, and ψ_i are model parameters. Similarly, the moving average component assumes the observation for location X at time t is a linear combination of past q error terms, plus the current white noise error term. This is illustrated in Equation (6), where ϵ_t represents the error term and ω_i are model parameters. Finally, the full ARMA model combines the two component equations, shown in Equation (7). [18] describes the method to determine p and q, as well as calculating the model parameters.

$$X_t = c + \sum_{i=1}^{p} \psi_i X_{t-i} + \epsilon_t \tag{5}$$

$$X_t = \sum_{i=1}^{q} \omega_i \epsilon_{t-i} + \epsilon_t \tag{6}$$

$$X_t = c + \sum_{i=1}^{p} \psi_i X_{t-i} + \sum_{i=1}^{q} \omega_i \epsilon_{t-i} + \epsilon_t \tag{7}$$

The ARMA model describes a stationary stochastic process. However, traffic patterns often exhibit indications of non-stationarity. Therefore, researchers typically use the autoregressive integrated moving average (ARIMA) model, where an initial differencing step is applied to eliminate the non-stationarity [19]–[21]. Similarly, there is also extensive research on seasonal ARIMA (SARIMA), which combines the long term periodic trend of traffic with the ARIMA model [22], [23].

The ARIMA model is a univariate model, thus road segments are decoupled from one another and analyzed individually through time in the ARIMA model. ARIMAX model ameliorates the ARIMA model by incorporating data from neighbouring links to generate predictions [24]. Moreover, we can also extend the ARMA model into vector ARMA (VARMA) and space-time ARIMA (STARIMA) to model multiple variables and capture the correlations among multiple roads [25], [26]. Meanwhile, the autoregressive conditional heteroskedasticity (ARCH) models the patterns of error term variation in a time series, [27] is an application of the ARCH model in traffic prediction.

3.2 Nonparametric Regression

Regression models predict future traffic properties by performing regression analysis on the past observation. In the simplest case of linear regression, we can compute future state values as a linear combination of multiple explanatory variables, such as past observations of neighbouring links. In contrast with the time series models, regression models assume no relationship between the predicted variables (dependent variables) and the past observations (independent variables).

Nonparametric regression refers to the type of regression analysis where the model makes no assumptions about the nature of the relationship between the independent variables and the dependent variables. Although we can certainly apply parametric regression models such as linear regression to traffic prediction, we encountered no work that proposes parametric regression models as a novel model. However, they have been listed as baseline models for performance comparison with other models [28]. Nonparametric regression models with application in traffic prediction include regression trees, k-nearest neighbour regression, and support vector regression.

3.2.1 Regression Trees

The regression tree model partitions the available data recursively based on the values of its independent variables, constructing a tree-like decision diagram. In its simplest form, it divides data into two partitions based on the value of a single independent variable. This is repeated until the desired number of leaf nodes is reached. At each leaf node of the tree, a simple model describes the data only within its partition. To generate predictions, the model traverses the regression tree based on the values of the independent variables. There is a test about the independent variables at each internal node of the tree, and the answer to the test determines which child branch should be visited next. Once the model reaches a leaf node, the model in the leaf node generates the dependent variables (prediction) based on the independent variables. Figure 1 is an example of a regression tree.

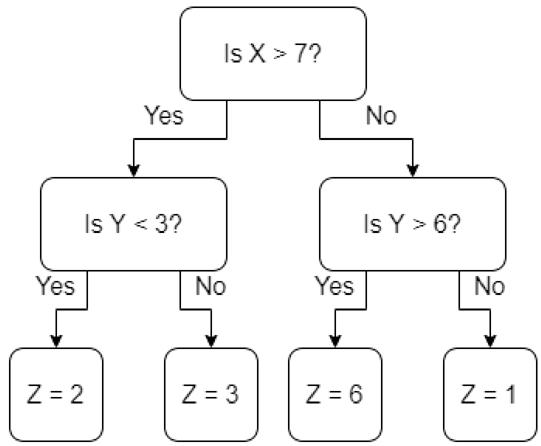

Figure 1: A simple regression tree. X and Y are the independent variables while Z is the dependent variable of this model.

Regression tree model can approximate complex functions using relatively simple functions with enough leaf nodes. Typically, traffic prediction models of this category use the ensemble method that contains more than one decision tree. An example of the ensemble method is the gradient boosting method developed by Friedman [29], which constructs each subsequent tree with an emphasis on instances where the previously constructed trees performed poorly. Similar to the parametric regression models, we encountered no work that proposes regression trees as a novel model for traffic prediction, only citations of them as powerful methods for model comparison purposes [28], [30].

3.2.2 k-Nearest Neighbour

The k-nearest neighbour method maintains the independent variables of all past observations in a feature space. During prediction, the model examines the independent variables of the new input and finds the k closest past observations in the feature space. The model then generates predictions by combining the dependent variables of the k observations.

The first evidence of applying the k-nearest neighbour method in traffic prediction is the work of Davis and Nihan [31], which uses the arithmetic mean of the k dependent variables to generate the prediction. Additionally, [32] shows that the model outperforms a 2-layer fully-connected neural network. Since then, this model continues to be improved [33], [34] and compared with other prediction models [35], [36]. To this day, the k-nearest neighbour model remains a powerful tool for the traffic prediction problem.

3.2.3 Support Vector Regression

Support vector regression (SVR) is a machine learning technique for regression based on the concept of support vectors, originally developed by [37]. The goal for SVR is to find a linear hyperplane that fits all data within an error threshold. In its simplest form, it can be described mathematically with following equation:

$$\begin{aligned} \min_{w} \quad & \frac{1}{2}||w||^2 \\ \text{subject to} \quad & |y_i - \langle w, x_i \rangle - b| \leq \epsilon \end{aligned} \tag{8}$$

Here w is the model parameters, b is a constant, ϵ is the threshold, while y_i are the dependent variable and x_i are the independent variables of each sample. The inner product $\langle w, x_i \rangle + b$ represents the model prediction for sample i, and the constraint is satisfied when the prediction is within the ϵ threshold of dependent variable y_i.

This form of SVR can only model linear dependence between the independent and dependent variables, and is in fact a parametric regression model. However, SVR can be extended to capture non-linear dependence in the input dimensions by mapping the input data to a high-dimensional feature space through a kernel function. A popular kernel function is the radial basis function (RBF) kernel that transforms the SVR to a nonparametric regression model.

In traffic prediction, SVR allows the complex spatial dependencies among road segments to be captured. [38], [39] are early examples of applying the SVR model to traffic prediction. Additionally, [23] and [40] provide evidence that SVR with the RBF kernel is competitive with the ARIMA model in traffic prediction.

3.3 Bayesian Network

A Bayesian network encodes the probability distribution over random variables in the form of a graph. A Bayesian network expresses random variables as nodes in the graph, and dependencies between two random variables as directed edges. Bayesian networks express complex probability distributions in a compact format by specifying conditional independence with edges. This reduces the number of parameters in the model when compared with a joint probabilistic distribution over all random variables and provides computational time savings during inference. In contrast with time series analysis and regression methods, Bayesian networks specify the dependencies among variables as probability distributions.

Section 2 establishes that the traffic properties at a given location is influenced by its upstream and downstream links. Therefore, a Bayesian network can model this by first representing the traffic states at each location as a random variable, then specifying dependencies between every location and its neighbours. Additionally, by assuming the traffic properties at a given time is dependent on the traffic properties at the previous time step, additional dependencies can be added in the Bayesian network to model the transition of the random variables over time to create a dynamic Bayesian network.

The first application of a Bayesian network in traffic prediction is [41], which is later expanded upon by [42]. This work models the conditional probability distributions of the dependencies using a Gaussian mixture model. This model assumes the underlying probability distributions to be stationary, which may break down for the process of evolving traffic patterns under some conditions. [43] attempts to solve this by specifying a different model for each characteristic traffic condition.

4 Deep Learning Methods

Since the turn of the century, the rise of artificial neural networks and deep learning gave researchers a new tool for traffic prediction. This section discusses the various deep learning methods and their application in traffic prediction. Similar to the methods discussed in Section 3, these methods represent traffic dynamics implicitly. Additionally, while statistical models subsume deep learning models, its focuses on feature learning from raw data and explosion in recent literature warrant a separate section for discussion. Deep learning methods for traffic prediction can be broadly divided into four categories: feedforward neural networks, recurrent neural networks, convolutional neural networks, and graph convolution networks.

4.1 Feedforward Neural Networks

Feedforward neural networks are the first and simplest type of artificial neural network. The connections among the nodes do not form a cycle and information in this network travel only in the forward direction, from the input nodes to the output nodes. A feedforward neural networks with a single hidden layer can approximate any continuous function in Euclidean space, making them universal approximators [44]. The notion of this network has been around for decades, but they require significant computational power and data to train. Therefore, its rise to prominence in research as well as application to traffic prediction is fairly recent.

In traffic prediction, traffic properties on a link are influenced by the recent states of nearby links, but the exact relationship is complex and unknown. Since a feedforward neural network can approximate complex functions, they are suitable for this task. The network can generate predictions for a link by using a combination of recent traffic states of neighbouring links and the target link itself as the input. Early adopters of this model in traffic prediction include [45]–[48].

Later, efficient training technique proposed by [49], [50] allows increasingly deeper neural networks to be built and trained. [51] is an example of a deep feedforward neural network in traffic prediction. The potential of deep learning also prompted the development of more complex neural network architectures, their details and applications to traffic prediction are discussed in the later sections.

4.2 Recurrent Neural Networks

In contrast to feedforward neural networks, a recurrent neural network (RNN) is designed to process sequential data. In an RNN, the connections of successive layers of nodes form a temporal sequence. Additionally, the parameters in the model are shared among time steps and the output of each layer is used as the input to the next, repeated until the end of the sequence. As such, this architecture can process sequential data of different lengths provided that the feature dimensions are constant for every entry in the sequence. Furthermore, an RNN can generate a sequence of outputs by continuously appending the latest output to the input sequence.

Basic RNNs are essentially deep feedforward neural networks with shared weights and the depth increases with sequence length of the data. This depth causes the vanishing gradient problem to arise during training [52]. The long short-term memory (LSTM) architecture is proposed to address this issue by incorporating gates to the network to limit information propagation [53], which is later simplified by the gated recurrent unit (GRU) architecture [54].

Recurrent neural networks has been actively used in traffic prediction research to capture the temporal dynamics of evolving traffic since the work of [55] in 2015. This framework represents the recent history of traffic states as a sequence and uses an RNN to generate traffic state predictions. Later improvements to this model include stacking multiple RNNs [56] and incorporating spatial structure of the road network in the RNN [57].

While RNNs can be used as standalone traffic prediction models, they are ultimately designed to capture the dynamics within a sequence. However, we cannot easily represent the complex spatial influences within a traffic network as a sequence. Therefore, some newer prediction models use RNNs to capture the temporal patterns of traffic, but also include a separate module to capture the spatial dependencies within a traffic network. These types of models are discussed in the later sections and revisits the RNN idea.

4.3 Convolutional Neural Network

A Convolutional neural network (CNN) is a deep learning method where the convolution operations is the primary means of propagating information from one layer to the next. In addition, the parameters of the convolution operations within a layer are shared, as opposed to the the parameter sharing between different layers in an RNN. Each convolution operation is restricted to a small region in the feature space known as the receptive field. The receptive fields from multiple convolution operations partially overlap to cover the entire feature space. As a result, CNN excels at extracting simple patterns and assembling them into increasingly complex features that are informative to the task.

The weight sharing of the convolution operation requires the receptive fields to be congruent with one another. This is typically a grid structure where the receptive field can be translated freely in the feature space. Therefore, CNNs are commonly used for image processing since images possess a convenient grid structure. Meanwhile, CNNs cannot be readily applied to a road network to capture its spatial dependencies since road networks generally do not have a perfect grid structure.

There are multiple attempts at converting the road network into a two-dimensional lattice graph, creating an image-like matrix that allows a CNN to be used in the analysis. [28], [58], [59] are examples of adapting CNN to capture the spatial dependencies of the road network. In particular, [58] combines a CNN with an RNN to capture the transitions in the temporal dimension, and the results show that this type of model can outperform ARIMA by a significant margin. However, road networks are rarely perfect lattice graphs. The required modifications usually involve dividing a road network into grid-like zones that dismantles the topological structure of the road network [59]. As a result, this framework cannot generate predictions on a single road segment, which is needed for tasks such as traffic signal control.

We can also use CNNs to analyze a stream of data, since we can translate the receptive fields on the time axis. There were numerous attempts of applying this in traffic prediction to capture the temporal dependencies of traffic patterns [60], [61]. Similar to recurrent neural networks, this type of convolutional framework are usually integrated with separate

Approach	Category	Setting	State	Prediction Horizon	Evaluation Metric
STGCN [60]	Spectral-based graph convolution	Highway/ Urban	Speed	15/30/45 minutes	MAE/ MAPE/ RMSE of speed
DGCNN [64]	Spectral-based graph convolution	Highway/ Urban	Speed	15/30/45 minutes	MAE/ RMSE of speed
ASTGCN [61]	Spectral-based graph convolution	Highway	Flow	60 minutes	MAE/ RMSE of Flow
DCRNN [63]	Spatial-based graph convolution	Highway	Speed	15/30/60 minutes	MAE/ MAPE/ RMSE of speed
DeepTransport [65]	Spatial-based graph convolution	Urban	Traffic rating (1-4)	15/30/45/60 minutes	Cohen's kappa of traffic rating
MRes-RGNN [66]	Spatial-based graph convolution	Highway	Speed	Unspecified	MAE/ MAPE/ RMSE of speed
ST-MetaNet [67]	Spatial-based graph convolution	Highway	Speed	15/30/60 minutes	MAE/ RMSE of speed
GRNN [30]	Spatial-based graph convolution	Urban	Speed	10 minutes	MSE of speed

Table 1: Main characteristics of selected graph convolutional network traffic prediction methods

module to capture the spatial dependencies within a traffic network.

4.4 Graph Convolutional Networks

Unlike images, graphs have no fixed shape; thus, we cannot use convolutional kernels with fixed shapes to analyze them. However, graph convolutional networks [62] generalized the convolution operation to the graph domain recently. Graph convolutional networks can be separated into two main categories: spectral-based and spatial-based graph convolution.

Since 2017, graph convolutional networks have been increasingly applied to traffic prediction with improved performance over recurrent neural networks and feedforward neural networks [60], [63]. However, there has been no direct comparison between the two categories of graph convolutional networks in this setting. This section outlines the differences between the three categories and discusses their applications to traffic prediction. Table 1 provides an overview of some traffic prediction methods that use graph convolutional networks.

4.4.1 The Graph perspective of Road Networks

We can absrtact a road network into a graph with intersection and road segments using Definition 1. However, the relevant traffic properties consist in road segments rather than intersections. Therefore, [30] proposed the linkage network that converts the road network into a graph with road segment as nodes and road connections through intersections as edges. Definition 2 and Figure 2 illustrate this process.

Definition 1 ***Road Network.*** *A road network $G(\mathcal{V}, \mathcal{E})$ is defined as a directed graph where $\mathcal{V}$ is a set of intersections and $\mathcal{E}$ is a set of links.*

Definition 2 ***Linkage Network.*** *A linkage network $G'(\mathcal{V}', \mathcal{E}')$ is a directed graph where vertices $\mathcal{V}'$ represent the road segments while directional edges $\mathcal{E}'$ denote the linkages between contiguous segments. A directional edge from segment v_i to v_j exist if and only if the traffic rules allow vehicle transfers from v_i to v_j directly.*

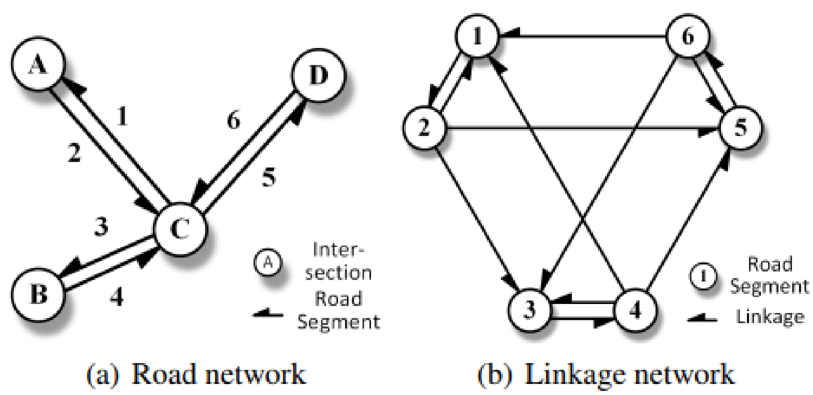

Figure 2: The relationship between the road network and the linkage network, adapted from [30]. Each road segment numbered from 1 to 6 is represented as a node in the linkage network. A directed edge in the linkage network signifies that traffic can move from the source road segment to the destination road segment directly through an intersection. Note that U-turn movements are permitted in this example.

4.4.2 Spectral-based Graph Convolution Networks

Spectral graph convolution uses the mathematical foundations of graph signal processing and spectral graph theory to define the convolution operation for graphs. The normalized graph Laplacian matrix of a graph is defined as shown in (9), where L, A, and D are the normalized Laplacian matrix, adjacent matrix, and degree matrix of a graph G respectively. I_N is an identity matrix whose dimension N equals to the number of vertices in G. The normalized graph Laplacian matrix is a real positive semi-definite matrix. Therefore, it can be factored using (10), where Λ is a diagonal matrix whose diagonal elements are eigenvalues of L and U is a unitary matrix consists of eigenvectors of L.

$$L = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \tag{9}$$

$$L = U \Lambda U^T \tag{10}$$

Additionally, in graph signal processing, the Fourier transform for graph signals is defined in (11) and the inverse Fourier transform is defined in (12).

$$\mathscr{F}(x) = U^T \cdot x \tag{11}$$

$$\mathscr{F}^{-1}(\hat{x}) = U \cdot \hat{x} \tag{12}$$

Finally, the convolution theorem states that the convolution between two signals is the point-wise product of their Fourier transforms. We can view the graph input as a signal and the convolutional filter as another signal. Ultimately, we can write the graph convolution of a graph input x with a filter g as

$$x * g = \mathscr{F}^{-1}(\mathscr{F}(x) \odot \mathscr{F}(g)) = U(U^T x \odot U^T g) \tag{13}$$

In traffic prediction, this framework can be used to extract the spatial dependencies of traffic, and it is first applied in [60]. This work uses an approximated version of spectral graph convolution [68] to speed up the eigendecomposition operation in (10). The overall traffic prediction model combines the spectral graph convolution module with a separate convolution module to analyze the temporal dimension, and testing results show a modest improvement on highway data over existing methods. Later attempts to improve this model include using a dynamic graph Laplacian matrix [64] and adding attention layers [61].

4.4.3 Spatial-based Graph Convolutional Networks

In contrast with spectral graph convolution, a spatial-based graph convolution uses the graph structure directly with localized receptive fields, similar to a CNN. This framework propagates node information along edges to nearby nodes instead of converting the graph to frequency domain. The convolution filters aggregate the propagated information to produce a hidden representation for every node, which can then be used in subsequent convolutional layers or to produce outputs. Spatial-based graph convolution networks have less robust mathematical foundation, but are more computationally efficient and generalizeable when compared with their spectral-based counterparts [69].

There are many forms of spatial-based graph convolution, differing with one another regarding the propagation process mentioned earlier. Since each node in the graph may have an arbitrary number of neighbours, we need to specify a propagation procedure for each neighbour of the target node. The convolution operation then aggregates the propagated information for every node to determine its hidden representation. For example, diffusion-convolutional neural networks [70] model the propagation as a diffusion process and the diffusion probability determines how much each node propagates to nearby nodes. Meanwhile, PATCHY-SAN [71] uses a fixed receptive field similar to a CNN and a node selection process determines which neighbouring nodes belong in the receptive field. Other definitions of graph convolution also exist in literature with different methods of information propagation.

Similar to spectral graph convolution, we can also use spatial-based graph convolution to capture the spatial dependencies of traffic, which is first proposed by [63]. This work integrates the diffusion convolution operation into an encoder-decoder recurrent neural network to capture the temporal dependencies of traffic, and the results shows a modest improvement on highway data over existing methods. Later, [66] improves this model by adding residual connections in the recurrent neural network. We can also cite [30], [65], [67] for using other forms of spatial-based graph convolution in traffic prediction.

References

[1] HDR Corporation, "Costs of road congestion in the greater toronto and hamilton area: Impact and cost benefit analysis of the metrolinx draft regional transportation plan", Greater Toronto Transportation Authority, Tech. Rep., 2008.

[2] P. Mirchandani and L. Head, "A real-time traffic signal control system: Architecture, algorithms, and analysis", *Transportation Research Part C: Emerging Technologies*, vol. 9, no. 6, pp. 415–432, 2001.

[3] P.-W. Lin, K.-P. Kang, and G.-L. Chang, "Exploring the effectiveness of variable speed limit controls on highway work-zone operations", in *Intelligent transportation systems*, Taylor & Francis, vol. 8, 2004, pp. 155–168.

[4] P. I. Richards, "Shock waves on the highway", *Operations research*, vol. 4, no. 1, pp. 42–51, 1956.

[5] M. J. Lighthill and G. B. Whitham, "On kinematic waves ii. a theory of traffic flow on long crowded roads", *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 229, no. 1178, pp. 317–345, 1955.

[6] C. Daganzo, "The cell transmission model. part i: A simple dynamic representation of highway traffic", 1992.

[7] C. F. Daganzo, "The cell transmission model, part ii: Network traffic", *Transportation Research Part B: Methodological*, vol. 29, no. 2, pp. 79–93, 1995.

[8] H. J. Payne, "Model of freeway traffic and control", *Mathematical Model of Public System*, pp. 51–61, 1971.

[9] G. B. Whitham, *Linear and nonlinear waves*. John Wiley & Sons, 2011, vol. 42.

[10] C. F. Daganzo, "Requiem for second-order fluid approximations of traffic flow", *Transportation Research Part B: Methodological*, vol. 29, no. 4, pp. 277–286, 1995.

[11] A. Aw and M. Rascle, "Resurrection of" second order" models of traffic flow", *SIAM journal on applied mathematics*, vol. 60, no. 3, pp. 916–938, 2000.

[12] E. Kometani and T. Sasaki, "On the stability of traffic flow (report-i)", *J. Oper. Res. Soc. Japan*, vol. 2, no. 1, pp. 11–26, 1958.

[13] L. A. Pipes, "Car following models and the fundamental diagram of road traffic", *Transportation Research/UK/*, 1966.

[14] G. F. Newell, "A simplified car-following theory: A lower order model", *Transportation Research Part B: Methodological*, vol. 36, no. 3, pp. 195–205, 2002.

[15] ——, "Nonlinear effects in the dynamics of car following", *Operations research*, vol. 9, no. 2, pp. 209–229, 1961.

[16] H. Nicholson and C. Swann, "The prediction of traffic flow volumes based on spectral analysis", *Transportation Research*, vol. 8, no. 6, pp. 533–538, 1974.

[17] T. T. Tchrakian, B. Basu, and M. O'Mahony, "Real-time traffic flow forecasting using spectral analysis", *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 519–526, 2011.

[18] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[19] M. S. Ahmed and A. R. Cook, *Analysis of freeway traffic time-series data by using Box-Jenkins techniques*, 722. 1979.

[20] C. Moorthy and B. Ratcliffe, "Short term traffic forecasting using time series methods", *Transportation planning and technology*, vol. 12, no. 1, pp. 45–56, 1988.

[21] S. Lee and D. B. Fambro, "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting", *Transportation Research Record*, vol. 1678, no. 1, pp. 179–188, 1999.

[22] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results", *Journal of transportation engineering*, vol. 129, no. 6, pp. 664–672, 2003.

[23] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning", *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 871–882, 2013.

[24] B. M. Williams, "Multivariate vehicular traffic flow prediction: Evaluation of arimax modeling", *Transportation Research Record*, vol. 1776, no. 1, pp. 194–200, 2001.

[25] Y. Kamarianakis and P. Prastacos, "Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches", *Transportation Research Record*, vol. 1857, no. 1, pp. 74–84, 2003.

[26] T. Ma, Z. Zhou, and B. Abdulhai, "Nonlinear multivariate time–space threshold vector error correction model for short term traffic state prediction", *Transportation Research Part B: Methodological*, vol. 76, pp. 27–47, 2015.

[27] B. Krithikaivasan, Y. Zeng, K. Deka, and D. Medhi, "Arch-based traffic forecasting and dynamic bandwidth provisioning for periodically measured nonstationary traffic", *IEEE/ACM Transactions on networking*, vol. 15, no. 3, pp. 683–696, 2007.

[28] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, and Z. Li, "Deep multi-view spatial-temporal network for taxi demand prediction", in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[29] J. H. Friedman, "Greedy function approximation: A gradient boosting machine", *Annals of statistics*, pp. 1189–1232, 2001.

[30] X. Wang, C. Chen, Y. Min, J. He, B. Yang, and Y. Zhang, "Efficient metropolitan traffic prediction based on graph recurrent neural network", *arXiv preprint arXiv:1811.00740*, 2018.

[31] G. A. Davis and N. L. Nihan, "Nonparametric regression and short-term freeway traffic forecasting", *Journal of Transportation Engineering*, vol. 117, no. 2, pp. 178–188, 1991.

[32] B. L. Smith and M. J. Demetsky, "Short-term traffic flow prediction models-a comparison of neural network and nonparametric regression approaches", in *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, IEEE, vol. 2, 1994, pp. 1706–1709.

[33] L. Zhang, Q. Liu, W. Yang, N. Wei, and D. Dong, "An improved k-nearest neighbor model for short-term traffic flow prediction", *Procedia-Social and Behavioral Sciences*, vol. 96, pp. 653–662, 2013.

[34] P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun, "A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting", *Transportation Research Part C: Emerging Technologies*, vol. 62, pp. 21–34, 2016.

[35] S. Clark, "Traffic prediction using multivariate nonparametric regression", *Journal of transportation engineering*, vol. 129, no. 2, pp. 161–168, 2003.

[36] B. Yu, X. Song, F. Guan, Z. Yang, and B. Yao, "K-nearest neighbor model for multiple-time-step prediction of short-term traffic condition", *Journal of Transportation Engineering*, vol. 142, no. 6, p. 04 016 018, 2016.

[37] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, "Support vector regression machines", in *Advances in neural information processing systems*, 1997, pp. 155–161.

[38] A. Ding, X. Zhao, and L. Jiao, "Traffic flow time series prediction based on statistics learning theory", in *Proceedings. The IEEE 5th International Conference on Intelligent Transportation Systems*, IEEE, 2002, pp. 727–730.

[39] C.-H. Wu, J.-M. Ho, and D.-T. Lee, "Travel-time prediction with support vector regression", *IEEE transactions on intelligent transportation systems*, vol. 5, no. 4, pp. 276–281, 2004.

[40] B. Yao, C. Chen, Q. Cao, L. Jin, M. Zhang, H. Zhu, and B. Yu, "Short-term traffic speed prediction for an urban corridor", *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 2, pp. 154–169, 2017.

[41] G. Yu, J. Hu, C. Zhang, L. Zhuang, and J. Song, "Short-term traffic flow forecasting based on markov chain model", in *IEEE IV2003 Intelligent Vehicles Symposium. Proceedings (Cat. No. 03TH8683)*, IEEE, 2003, pp. 208–212.

[42] S. Sun, C. Zhang, and G. Yu, "A bayesian network approach to traffic flow forecasting", *IEEE Transactions on intelligent transportation systems*, vol. 7, no. 1, pp. 124–132, 2006.

[43] Y. Qi and S. Ishak, "A hidden markov model for short term prediction of traffic conditions on freeways", *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 95–111, 2014.

[44] G. Cybenko, "Approximation by superpositions of a sigmoidal function", *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303–314, 1989.

[45] M. S. Dougherty and M. R. Cobbett, "Short-term inter-urban traffic forecasts using neural networks", *International journal of forecasting*, vol. 13, no. 1, pp. 21–31, 1997.

[46] B. Park, C. J. Messer, and T. Urbanik, "Short-term freeway traffic volume forecasting using radial basis function neural network", *Transportation Research Record*, vol. 1651, no. 1, pp. 39–47, 1998.

[47] H. Dia, "An object-oriented neural network approach to short-term traffic forecasting", *European Journal of Operational Research*, vol. 131, no. 2, pp. 253–261, 2001.

[48] B. Abdulhai, H. Porwal, and W. Recker, "Short-term traffic flow prediction using neuro-genetic algorithms", *ITS Journal-Intelligent Transportation Systems Journal*, vol. 7, no. 1, pp. 3–41, 2002.

[49] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets", *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[50] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks", in *Advances in neural information processing systems*, 2007, pp. 153–160.

[51] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach", *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 865–873, 2014.

[52] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult", *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.

[53] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[54] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation", *arXiv preprint arXiv:1406.1078*, 2014.

[55] Y. Tian and L. Pan, "Predicting short-term traffic flow by long short-term memory recurrent neural network", in *2015 IEEE international conference on smart city/SocialCom/SustainCom (SmartCity)*, IEEE, 2015, pp. 153–158.

[56] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Deep bidirectional and unidirectional lstm recurrent neural network for network-wide traffic speed prediction", *arXiv preprint arXiv:1801.02143*, 2018.

[57] Z. Zhao, W. Chen, X. Wu, P. C. Chen, and J. Liu, "Lstm network: A deep learning approach for short-term traffic forecast", *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[58] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li, "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction", in *AAAI Conference on Artificial Intelligence*, 2019.

[59] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, "Dnn-based prediction model for spatio-temporal data", in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, 2016, p. 92.

[60] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting", *arXiv preprint arXiv:1709.04875*, 2017.

[61] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting", in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 922–929.

[62] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs", *arXiv preprint arXiv:1312.6203*, 2013.

[63] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting", *arXiv preprint arXiv:1707.01926*, 2017.

[64] Z. Diao, X. Wang, D. Zhang, Y. Liu, K. Xie, and S. He, "Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting", in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 890–897.

[65] X. Cheng, R. Zhang, J. Zhou, and W. Xu, "Deeptransport: Learning spatial-temporal dependency for traffic condition forecasting", in *2018 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2018, pp. 1–8.

[66] C. Chen, K. Li, S. G. Teo, X. Zou, K. Wang, J. Wang, and Z. Zeng, "Gated residual recurrent graph neural networks for traffic prediction", in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 485–492.

[67] Z. Pan, Y. Liang, W. Wang, Y. Yu, Y. Zheng, and J. Zhang, "Urban traffic prediction from spatio-temporal data using deep meta learning", 2019.

[68] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering", in *Advances in neural information processing systems*, 2016, pp. 3844–3852.

[69] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S. Y. Philip, "A comprehensive survey on graph neural networks", *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[70] J. Atwood and D. Towsley, "Diffusion-convolutional neural networks", in *Advances in Neural Information Processing Systems*, 2016, pp. 1993–2001.

[71] M. Niepert, M. Ahmed, and K. Kutzkov, "Learning convolutional neural networks for graphs", in *International conference on machine learning*, 2016, pp. 2014–2023.