Department of Computing and Information Systems
The University of Melbourne

Masters Thesis

Large scale real-time traffic flow prediction using SCATS volume data

Author:
Rabindra Kumar Panda

Supervisor:
Dr. Vinh Nguyen

A thesis submitted in fulfilment of the requirements for the degree of Master of Science in Computer Science

May 2016

Declaration of Authorship

I , Rabindra Kumar PANDA, certify that

- this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person where due reference is not made in the text.
- the thesis is approximately 20000 words in length (excluding text in images, table, bibliographies and appendices).

Signed:

Date:

Abstract

In this thesis we apply extend an LSTM model for the application of short term traffic flow prediction. We analyse it using the SCATS volume data and suggest improvements.

Acknowledgements

I am deeply indebted to the following people for their invaluable support and comments without which I would not have been able to complete this work…

Contents

List of Figures

List of Tables

Abbreviations

ARIMA	**A**uto **R**egressive **I**ntegrated **M**oving **A**verage
LSTM	**L**ong **T**erm **S**hort **M**emory
RNN	**R**ecurrent **N**eural **N**etwork
SCATS	**S**ydney **C**oordinated **A**daptive **T**raffic **S**ystems

Dedicated to my parents

Chapter 1

Introduction

"As a reader I loathe introductions...Introductions inhibit pleasure, they kill the joy of anticipation, they frustrate curiosity."

Harper Lee, *To Kill a Mockingbird* (1960)

Road traffic congestion is a serious global issue, resulting in significant wastage of resources. While improving and extending the road infrastructure has reduced the issue to some extent, this is time consuming and limited in capacity. Thus in last few decades, for better planning and control of road traffic, adaptive traffic control systems have been deployed around the world. Still the role of adaptive control systems is not fully realised without predictive capabilities in the short term, without which the adapative traffic control systems only react to events at the real time. While this is the desired objective, the performance of these systems could be significantly improved by making them proactive (Smith and Demetsky [154]). Short term traffic flow prediction is not only helpful for these advanced traffic control systems, it is also useful for advanced traveller information systems.

1.1 Background

Predicting the future has always been a fascinating topic throughout the history of mankind. Instances of predicting the future through unconventinal means have been mentioned in various forms of literature such as mythologies, fantasy, science fiction etc. Even today, while the means of prediction have changed, we still try to predict almost everything in our day to day lives - from election polls to sports outcomes to financial results.

In the context of road traffic, prediction has always been a difficult task, merely due to the dynamic and complex nature of traffic. A lot of research has gone into short term traffic prediction. Various methods have been proposed in last few decades to provide more accurate predictions, yet none of those methods has been claimed to become the best.

1.2 Objectives and scope

Research objective is to propose a new model that can use the large amount of available traffic data to predict the traffic in the short term. More importantly this research tries to answer the following questions -

- How can we use the large amount of traffic data available for predicting short term traffic flow?
- Can the proposed deep learning models have better accuracy than exiting models?

Research scope - The scope of this research is to predict traffic volume in the short term. The use of the phrase 'short term' implies that we are only interested in the prediction within a very short horizon which typically ranges between few seconds to few hours in practice. The traffic parameters that are of usually of interest to be predicted are volume, time, speed and density. The scope of this research is limited to the prediction of only traffic volume. While doing so, we are only taking the past data into consideration and not taking the non-recurrent phenomena such as traffic accidents, weather or public events into consideration.

1.3 Thesis outline

Chapters – This thesis is divided into six chapters.

- Chapter 1: Introduction - In this chapter we present the background and research context, research objectives and scope.
- Chapter 2: Traffic Prediction: Literature Review - In this chapter we provide a reasonably thorough review of existing literature on short term traffic prediction.
- Chapter 3: SCATS Traffic Volume Data - In this chapter we provide the description of the traffic volume data collected by VicRoads using the SCATS systems.

Methods to deal with missing data are presented in this chapter. Finally we present some exploratory data analysis on the traffic data.

- Chapter 4: A Deep LSTM Network for Short Term Traffic Prediction - In this chapter we propose a long short term memory (LSTM) neural network model for short term traffic prediction.
- Chapter 5: Evaluation of the Model - In this chapter we evaluate the model.
- Chapter 6: Conclusions and Future Directions - In this chapter we conclude our thesis and provide inputs for future work.

Chapter 2

Traffic Prediction: Literature Review

"There is no way that we can predict the weather six months ahead beyond giving the seasonal average"

Stephen Hawking, *Black Holes and Baby Universes* (1993)

2.1 Introduction

In this chapter we provide a reasonably complete review of existing literature on short term traffic flow prediction. Research on short term traffic prediction has been going on since 1979, Ahmed and Cook [5]. After more than three decades of research, short term traffic flow prediction is still an interesting reserch subject for many professionals around the world. The simplest reseaon being the complex non-linear nature of traffic data and the effects of non-recurrent events(weather, public events, accidents etc.) on it. Critical reviews of existing literature on short term traffic flow have been presented in detail by Smith and Demetsky [154], Vlahogianni et al. [189], Van Lint and Van Hinsbergen [183] and Vlahogianni et al. [194]. The use of the phrase 'short term' limits the scope of traffic prediction in terms of the prediction horizon which usually varies between few seconds to few hours depending upon the approach and application.

The traffic parameter that are predicted can be - flow(number of vehicles per hour), time(minutes to travel between two points), speed(mean speed in km/hour) and density(number of vehicles per km).

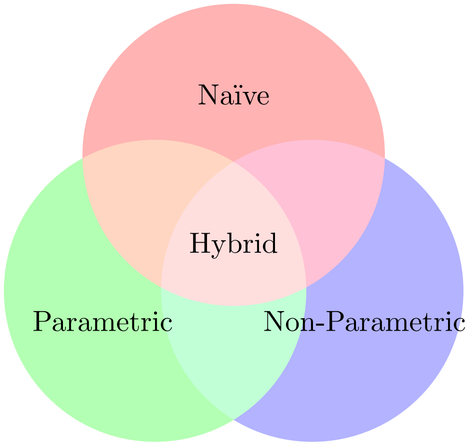


FIGURE 2.1: Methods used for short term traffic prediction

Selecting the right model for short term traffic prediction is a challenging task. A number of models have been suggested and yet there is no one-fits-all model. The various methods that have been suggested for short term traffic prediction can be categorised into four groups - naïve, parametric, non-parametric and hybrid as shown in figure 2.1.

2.2 Naïve methods

These are heuristics methods, and often used in practice because of their simplicity and the ease of implementations. In most cases these methods are used as baselines for comparison while creating more advances methods. The simplest naive approach in short term prediction would be to take the last observed value and this involves no computational effort. Another simple heuristic method known as the historical averages uses the average of past observed values.

2.3 Parametric models

In parametric models, we estimate the parameters from the training dataset to determine the function that classifies new unseen data. The number of parameters are fixed. The advantage of parametric models are that these perform quite well in situations where the large amount of data is not available. Some of the typical examples of parametric models include Linear and nonlinear regression, ARIMA models, Kalman filter, Linear SVM etc.

2.3.1 Classical regression

In machine learning and statistical applications, the use of linear models are predominant. These models are also important in time series domains such as traffic flow prediction. The primary idea behind the regression is to express the output variable as a linear combination of input vectors. We can express the linear regression in time series as an ouput influenced by a collection of inputs, where the inputs could possibly be an independent series

$$x_t = \beta_1 z_{t1} + \beta_2 z_{t2} + ... + \beta_q z_{tq} + w_t \tag{2.1}$$

where $\beta_1, \beta_2, ..., \beta_q$ are unknown regression coeffiecients and w_t is a random error.

Högberg [82] used non-liner regression for traffic prediction.

2.3.2 ARIMA

ARIMA(Auto Regressive Integrated Moving Average) is a class of parametric regression models. In this section we will introduce ARIMA and related methods such as exponential smoothing and moving averages. For an in depth understanding of these models the reader is encouraged to refer to to Tong [171], Brockwell and Davis [19] and Box et al. [17]. It is important to understand that ARIMA modelling works only with stationary time series data. A stationary time series is one whose properties do not depend on the time it is being observed. Trends and seasonality affect time series and hence make it non-stationary. Although this seems as a big restriction, in short term traffic prediction, ARIMA models have been very successful. Two basic models constituate ARIMA models - AR(autoregressive) and MA(moving average).

The main idea behind autoregressive models is that past values affect the present value, i.e. x_t can be expressed as a function of past p values $x_{t-1}, x_{t-2}, ..., x_{t-p}$, where p is the number of steps into the past. We can express an autoregressive model of order p as below

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + w_t \tag{2.2}$$

where x_t is stationary and $\phi_1, \phi_2, ..., \phi_p$ are constant parameters that are to be chosen. We have added the term w_t as a Guassian white noise with zero mean and variance σ_w^2.

In the MA model, the current value is dependent on the last q one-step forecast errors $e_{t-1}, e_{t-2}, ..., e_{t-q}$ and the white noise w_t. The expression for moving average is

$$x_t = -\theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q} + w_t \tag{2.3}$$

$\theta_1, \theta_2, ..., \theta_q$ are the parameters to be chosen.

Now proceeding to an ARMA(autoregressive moving average) model, we define an ARMA(p,q) model where the present value x_t is dependent on p past recent values and q past recent forecast errors and a white noise w_t.

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q} + w_t \tag{2.4}$$

When q is 0, the model becomes an autoregressive model of order p, AR(p) and when p is 0 the model is a moving average of order q, MA(q). We can rewrite 2.4 by using the backshift operator B^α, which is defined as $B^\alpha z_t = z_{t-\alpha}$,

$$\phi(B)x_t = \theta(B)e_t \tag{2.5}$$

where

$$\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p \tag{2.6}$$

$$\theta(z) = 1 - \theta_1 z - ... - \theta_q z^q \tag{2.7}$$

In practice, most time series data are non-stationary and so several approaches, for instance by differencing, are taken to make it stationary before applying the ARMA(p,q) model. By combining differencing with autoregressive and moving average we obtain the ARIMA model defined as below

$$x'_t = \phi_1 x'_{t-1} + \phi_2 x'_{t-2} + ... + \phi_p x'_{t-p} - \theta_1 e_{t-1} - \theta_2 e_{t-2} - ... - \theta_q e_{t-q} + w_t \tag{2.8}$$

where x'_t is the differenced series. Formally the model is denotes as ARIMA(p,d,q) where p is the order of autoregressive part, d is the degree of differencing and q is the order of moving average. This is also known as a non-seasonal ARIMA model.

The common method used to determine the parameters in an ARIMA(p,d,q) model is known as the Box-Jenkins approach (Box et al. [17]) which is three stage procedure. The three stages are identification, estimation and diagnostic checking. At the identification stage, the values p, d and q are determined by observing the autocorrelation and partial autocorrelation functions of the time series and its differences. At the estimation stage, the maximum liklihood estimates are determined for each model parameter. Finally in the dignostics stage, the residuals are analysed and model comparisions are done. If the model fits well then the standardised residuals behave as an i.i.d. with mean zero and variance one.

Ahmed and Cook [5] used Box-Jenkins method for short-term traffic forecast. The input data used was 166 sets of time series traffic data collected by freeway traffic surveillance systems in three locations - Los Angeles, Minneapolis and Detroit. The authors concluded an ARIMA(0,1,3) model, based on the autocorrealtion and partial autocorrelation functions, as a resonable fit for the short term prediction tasks for both traffic volume and occupancy. The model performance was evaluated against a moving average, a double smoothing average and a Trigg and Leach adaptive model. The comparisons suggest that the ARIMA model had better accuracy than the others. The authors used this model in detecting traffic incidents by comparing the real-time flow occupancy with the predicted value. Nihan and Holmesland [139] used the Box-Jenkins technique on monthly data collected at 15 minutes interval on a freeway segment from 1968 to 1976 to forecast for the year 1977. After examining several models they finally selecte an ARIMA(12,1,7) model. The forecast was done for average weekday volume with positive results.

Williams [207] used an ARIMAX model to use upstream traffic data along with the predicting location's traffic data while estimating the paramters of the ARIMA model. This is done using ARIMAX model which is an extension of the ARIMA model where an exogenous variable is used. The data was collected form four locations near Beaune, France. The data from three upstream locations were used for forecasting at the fourth location in Beaune. The same data were used in the proposed ATHENA and KARIMA models. The model was compared against the univariate ARIMA, ATHENADanech-Pajouh and Aron [42] and KARIMA(Van Der Voort et al. [177]) models. The results show tha the ARIMAX model consistently outperformed the ARIMA model. However the complexity of the ARIMAX model is more than the ARIMA model with as many as twice the parameters to estimate. Also in case of missing values the ARIMAX model performance degraded more than the ARIMA model.

Min et al. [130] proposed a dynamic Space Time ARIMA (STARIMA) model for short term traffic prediction. Their argument for the new proposed model was based on the factor that most of the existed model failed to take the spatial information of the transporatation system into account. The proposed dynamic STARIMA model combines STARIMA and Dynamic Turn Ratio Prediction (DTRP) model. Using DTRP they dynamically updated the static matric W_k in STARIMA model that contains the structural information of the transporatation network. The results of the study showed significant improvement in forecast accuracy. The authors later published another similar work (Min et al. [131]) that used the generalised STARIMA (GSTARIMA) model. The authors presented the results where this model has a small improvements over the STARIMA model. However the major drawbacks of the GSTARIMA model is the estimation of large number of parameters which significantly increases the computational time. It also suffers in performance if enough historical data is not available.

Williams and Hoel [208] poposed for the acceptance of seasonal ARIMA models for short term traffic prediction. A seasonal ARIMA $(p, d, q)(P, D, Q)_s$ for a time series x_t is one where s is the period, d and D are nonnegative integers. The time series theorem known as the World decomposition is used as the theoritical justification of applying seasonal ARIMA model to univariate time series with stationarity. Data from two freeway locations, one each from the United States and the United Kingdom were used for evaluating the model. The performance of the models were compared against three heuristics approaches - historical averages, random walk and deviation from historical avarages. The results show that for both the locations the seasonal ARIMA has better performance than the three methods mentioned earlier. However the authors did not present whether a non-seasonal ARIMA model would have similar performance. The only other model that was considered for comparison was the KARIMA model, which did not perform as good as the seasonal ARIMA model. Kumar and Vanajakshi [113] also used a seasonal ARIMA in the context of limited data for short term traffic prediction. They used data collected over three days from an arterial road in Chennai, India for the study. The model was validated on 24 hours ahead forecast . Thier resluts were positive when compared with historical avereges and naïve methods. They argued when availability of large traffic dataset is a constraint seasonal ARIMA method is a better choice. Szeto et al. [164] used a hybrid SARIMA model with cell transmission model for multivariate traffic prediction. The authors reasoned the use of multivariate models captuered the spatial characteristics of the transporatation network and hence are the natural and better choice over an univariate model. The model was validated against data collected form the city center in Dublin, Ireland. The results at two junctions were compared against real observations and had MAPE of 4.45 and 10.6. The authors however did not provide comparison against other univariate models or multivariate models which could present the model's relative performance.

The major defficiency of tha ARIMA models is that they do not take the extremes into consideration and focus on the means. This is in contrast to the nature of the traffic data. ARIMA models are also have the inability to perform will with missing data as pointed out by Smith and Demetsky [154].

2.3.3 Kalman filter

Okutani and Stephanedes [141] used Kalman filtering in traffic prediction in an urabn network and extended it for freeways.

Wang and Papageorgiou [200]

Xie et al. [213]

Guo and Williams [70]

Guo et al. [68]

The main advantage of Kalman filtering is the state variable is updated continuously.

2.3.4 Exponential smoothing

In exponential smoothing method the forecast is the weighted average of past observations, while the weights decrease exponentially for older observations. For time series data with no trends and seasons single exponential smoothing is usually used.

2.4 Non-Parametric models

In nonparamtric models the parameters are not fixed, and vary with the amount of data available. Usually more data is required for this models than parametric models. The advantage of these models is that they can model the complex non-linear data better. Some of the widely used non-parametric models are - k-Nearest Neighbour, Non-parametric regrssion and Neural Networks

2.4.1 K-nearest neighbour

The basic process of the k-nearest neighbour algorithm is described in figure 2.2.

Lv et al. [121]

Myung et al. [133]

Zhang et al. [227]

Meng et al. [127]

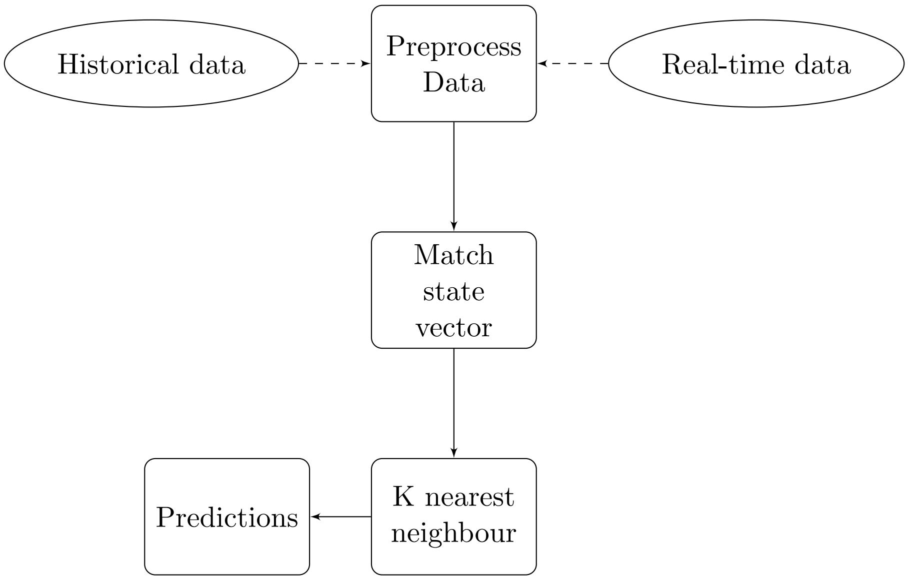


FIGURE 2.2: K nearest neighbour process flow

2.4.2 Neural networks

Artificial Neural Networks(ANN) were mathematical models (McCulloch and Pitts [126], Rosenblatt [149]) designed to provide a representation of how the human brain works. It is obvious now that these mathematical models bear little resemblance to the structure of brain, yet they have been hugely successful. Because they were initially inspired by the biological brain, the term neural is associated with such kind of mathematical models. A basic artificial neural network consists of a set of nodes connnected by edges with weights. We can say that the nodes represent the biological neurons and the edges represent the synapses. The conections among the nodes can be cyclic or acyclic. The former is known as a feedforward neural network and the later as a recurrent network. We describe about these neural networks in more details in chpater 4.

Several variations of artificial neural networks have been used in short term traffic prediction. Some well known examples include - *Multilayer perceptrons, Radial basis function networks, Kohnen maps* and *Hopfield networks.*

2.4.3 Fuzzy logic

[229]

2.4.4 Bayesian networks

[22]

2.5 Hybrid Methods

In recent years many hybrid methods have been tried in short term traffic prediction with mixed results.

A hybrid method by combining kohonen maps with ARIMA model was proposed by Van Der Voort et al. [177]. The model known as KARIMA, used the same data(collected near Beaune, France) that was used in the ATHENA model for an accurate comparison with the later.

Chen et al. [33] proposed an ARIMA-GARCH model for short term traffic prediction. The performance of this hybrid model when compared to the standard ARIMA model did not yield positive results.

2.6 Comparisons

Chapter 3

SCATS Volume Data

"There is no order in the world around us, we must adapt ourselves to the requirements of chaos instead."

Kurt Vonnegut, *Breakfast of Champions* (1973)

3.1 Introduction

SCATS(Sydney Coordinated Adaptive Traffic System) is an adaptive traffic control system. It was developed by the Department of Main Roads in the 1970's. SCATS operates in real-time by adjusting signal timings in response to changes in traffic demand and road capacity. All major and minor cities in Australia and New Zealand use SCATS. Few other cities around the world such as Hong Kong, Kuala Lumpur, Sanghai and Singapore also have adopted SCATS over other adaptive traffic control system. In Melbourne and surrounding cities, SCATS controls more than 3,900 sets of traffic signals

There are three main parameters that SCATS user to achieve traffic signal coordination:

- Cycle time: The total time of all signal sequences in a cycle
- Phase split: The proportion of the cycle time allocated to each phase
- Offset: The time relationship between the starting and finishing of the green phases of succesive sets of signals within a coordinated system

The desicion making of the SCATS system occurs at two levels - *strategic* and *tactical*.

3.2 Volume data

Traffic loop detectors are embedded in the raod pavement and located in each lane near the stop line at traffic intersections. These detectors collect traffic volume and the time it takes a vehicle to clear the loop. In this research we used the data collected from sensors at 1084 homogeneous links(a road segment where the volume for all traffic along that link is collected). Traffic volume was collected for every 15 minutes interval from 01/01/2008 to 25/07/2013, a total of 195168 observations.

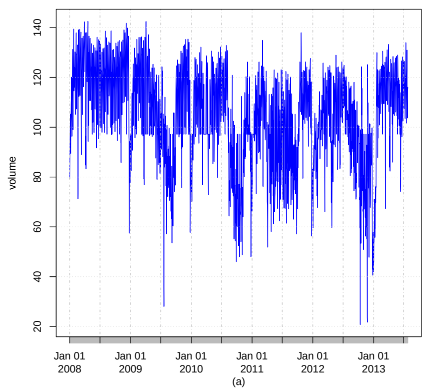
3.2.1 Handling missing data

One of the major difficulties with traffic sensor data is missing data, that can be caused by several factors.

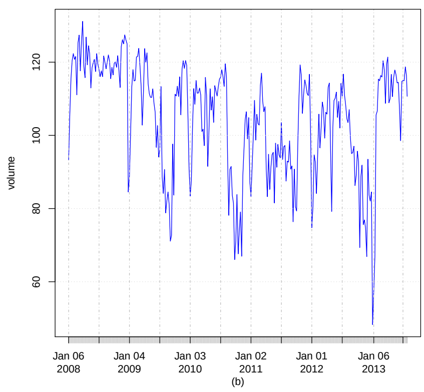
3.3 Exploratory analysis

In this section, we will perform exploratory analysis on the traffic volume data at a homogenous segment.For this purpose we will use both the south and north bound traffic data at Nicholson street (north of Melbourne's CBD)between Gertrude street and Victoria Parade during the period 01/01/2008 to 25/07/2013.

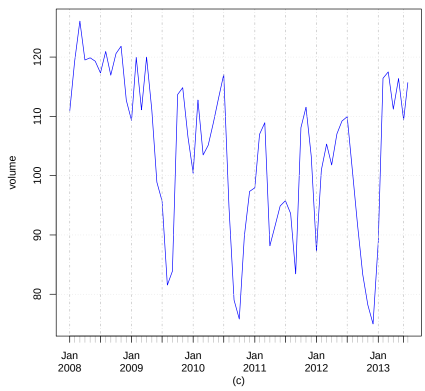
Figure 3.1 shows the daily, weekly, monthly and yearly average traffic volume at a site location.Figure 3.2 shows how a typical day of the week on average looks like at a homogeneous link.



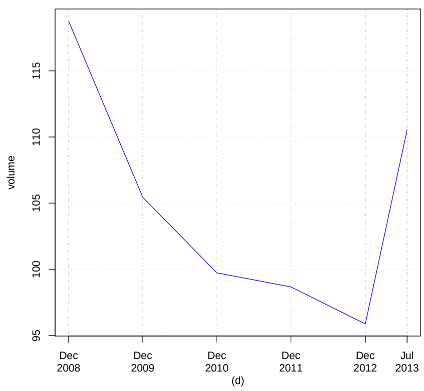
(A) Daily



(B) Weekly

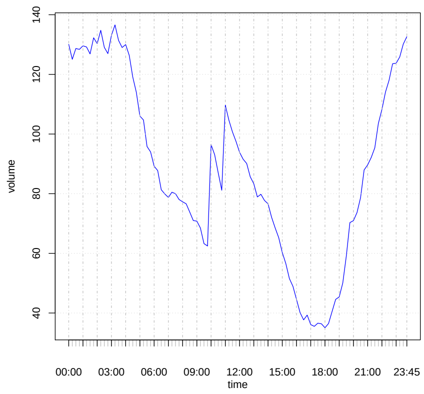


(C) Monthly

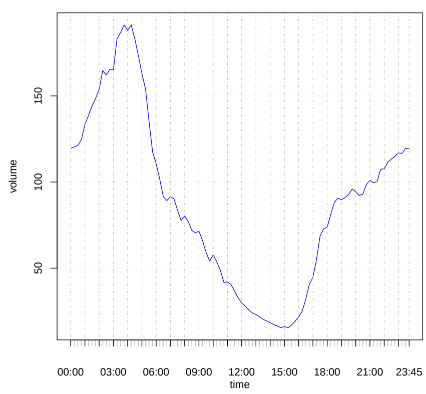


(D) Yearly

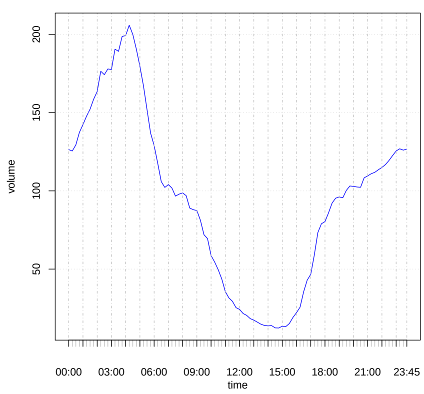
FIGURE 3.1: (a) daily, (b) weekly, (c) monthly and (d) yearly average of traffic volume (15 mins interval) at a site location from the period 01/01/2008 to 26/07/2013



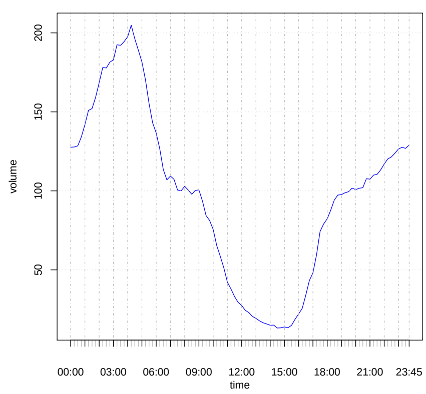
(A) Monday



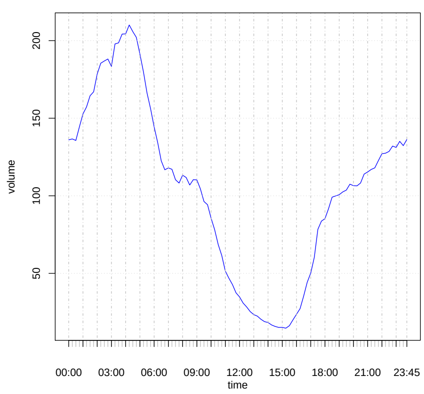
(B) Tuesday



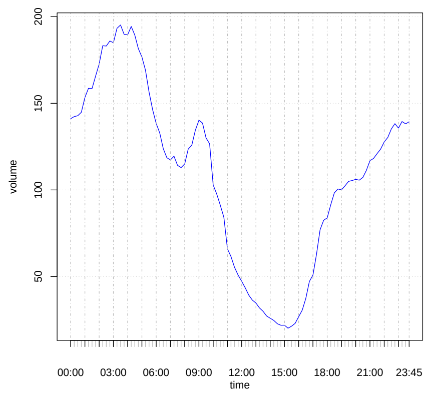
(C) Wednesday



(D) Thursday

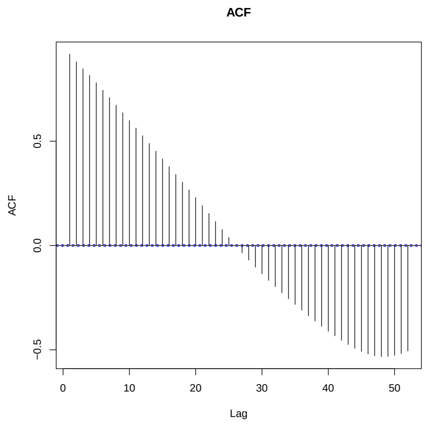


(E) Friday

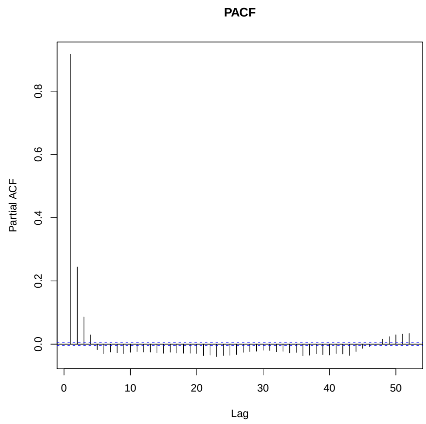


(F) Saturday

FIGURE 3.2: Average traffic grouped by every day of the week.



(A) ACF



(B) PACF

FIGURE 3.3: Plots of the autocorrelation and partial autocorrelation functions

Chapter 4

A Deep LSTM Network for Short Term Traffic Prediction

"I am a brain, Watson. The rest of me is a mere appendix."

Arthur Conan Doyle, *The Adventure of the Mazarin Stone* (1921)

In section 2.4.2, we presented a brief introduction to artificial neural networks and reviewed existing literature in short term traffic prediction that used various types of neural networks. In the following sections we present a bried overview of deep learning. We then describe deep feedforward networks, deep recurrent networks with emphasis on the Long Short Term Memory(LSTM) networks which are a redesigned version of recurrent networks. Later we present how we can we can use these kind of networks for short term traffic prediction.

4.1 Introduction

Today we live in a world where almost every interaction of ours with the external world uses some form of computing. Computers have become an inseparable part of human lives. In the earlier days when computers were built, people began to ponder whether they could achieve human level of intelligence. Even though at that point the answers seemed optimistic, it has taken quite some time and understanding on our part to make significant achievements in the field of artificial intelligence. One of the approaches was to use knowldge base systems, where computers reason about real world concepts, that were defined in hard-coded formal langauges, using logical inference rules. These systems led to little success. The difficulties faced in the knwoledge based appproach made us built computers to learn automatically from data, an approach we know as machine learning.

A large number of real world problems could eaily be tackled using machine leraning. However for the machine learning algorithms to perorm well they need to be provided with proper representaion of data. For example, in a problem where we would like to detect humans in images, it is difficult to represent various shapes of human body in terms of raw pixels. Finding a proper representation from data is a challenge and sometimes become very difficult. A class of machine learning algorithms called representation learning, tackles this problem by learning the representaions as well. Autoencodes are such types of algorithms. Again the problem with representation learing is that it is not easy to find the representations due to the presence of various factors of influence (Bengio et al. [10]). Deep learning solves this problem in representation learning by taking a layered approach by expressing representations in terms of simpler representations. The mapping from the input to output is done through a series of hidden layers, where each layer is an abstraction on the previous layer. The depth of the model can be viewed as the depth of the computational graph, i.e. the number of sequential instructions that need to be executed to map an input to output.

4.2 Feedforward neural networks

Deep feedforward networks are the most important deep learning models. The main goal of a deep feedforward network is to approximate a function f^* that maps an input $\mathbf{x}$ to an output y. As the name implies, the information in these models flow in the forward direction. These are the basis of several models used in commercial applications such as the convolutional networks, which are extensions of the feedforward networks, have been very successful in image recognition. With the addition of feedback connections to feedforward networks, recurrent networks are created. Feedforward networks consist of a chain of layers, which is simply done by composing functions for instance we can compose three functions as to map an input $\mathbf{x}$ to an output y, $y = f(\mathbf{x}) = f^3(f^2(f^1))$. Function f^2 acts as the hidden layer that maps the output from the input layer f^1 to the input of the output layer f^3.

The diagram 4.1 illustrates a simple feedforward neural network with 3 nodes in the input layer, 3 and 4 nodes in two hidden layers and a single node output layer. Information propagates from the input layer through the hidden layer to the output layer, known as the forward pass of the network. This kind of feedforward network is called a multilayer perceptron. Multilayer perceptrons are good at classification.

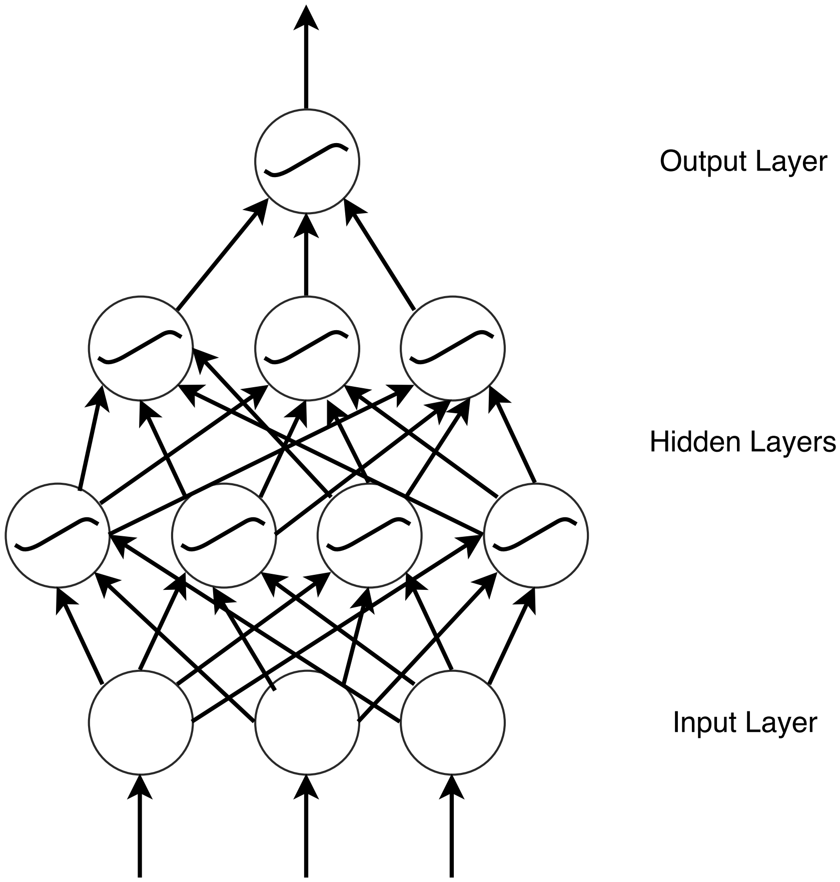


FIGURE 4.1: A feedforward neural netwrok with two hidden layers, this network is also known as a multilayer perceptron. The S-shaped curves denote the sigmoidal function.

Let's consider a simple multilayer perceptron with I nodes in the input layer. For an input vector $\mathbf{x}$, where length of $\mathbf{x}$ is I. Each node in the hidden layer gets a weighted sum of the units in the input layer. The output of each hidden unit a_h is then applied to an activation function θ_h to produce the activation b_h

$$a_h = \sum_{i=1}^{I} w_{ih} x_i \tag{4.1}$$

$$b_h = \theta_h(a_h) \tag{4.2}$$

There are several choices for the activation functions with sigmoidal and hyperbolic tan fucntions are the most common choices. The reason of these choices is nonlinearity of these functions. Recently the recommended activation function for feedforward neural networks is the *rectified linear unit* or ReLU (Nair and Hinton [134]), defined as $f(x) = max 0, x$, as they allow faster and efficient training of deep neural network architectures.

The activations flow through the rest of the hidden layers in similar fashion. For instance the l^{th} hidden unit in layer H_l

$$a_h = \sum_{h' \in H_{l-1}} w_{h'h} b_{h'} \tag{4.3}$$

$$b_h = \theta_h(a_h) \tag{4.4}$$

In the output layer, the activation function is applied on the output from the hidden layer to produce the output y. The input a_k to the output unit is given by

$$a_k = \sum_{h \in H_L} w_{hk} b_h \tag{4.5}$$

where L is the number of hidden layers in the network. The number of units in the output layer and the type of activation function are chosen based on the problem task at hand. For binary classification a single unit with logistic sigmoid activation function is primarily used. For classification with k ¿ 2 classes, k output units are used and the outputs are normalised using the *softmax* function. A very common example of this is the hand-written digits classification, where the output layer consists of 10 units.

4.3 Recurrent neural networks

As mentioned earlier, we can create a recurrent neural network by adding feedback connections to a feedforward network. Several types of recurrent neural networks have been proposed over the years, some of which are - *echo state networks, time delay networks, jordan networks.* At first the difference between a feedforward and a recurrent network may not be obvious and seem trvial but recurrent networks are very powerful in the sense that they can retain the history and thus forming a memory in their feedback connections.

4.4 Network training using gradient descent

Training neural networks is no different than any other machine learning models with a loss function and gradient descent algorithm. However the difficulty is that the non-linear characteristics of neural networks causes to the loss functions to become non-convex. So the training procedure usually involves small iterative gradient descent algorithm to get

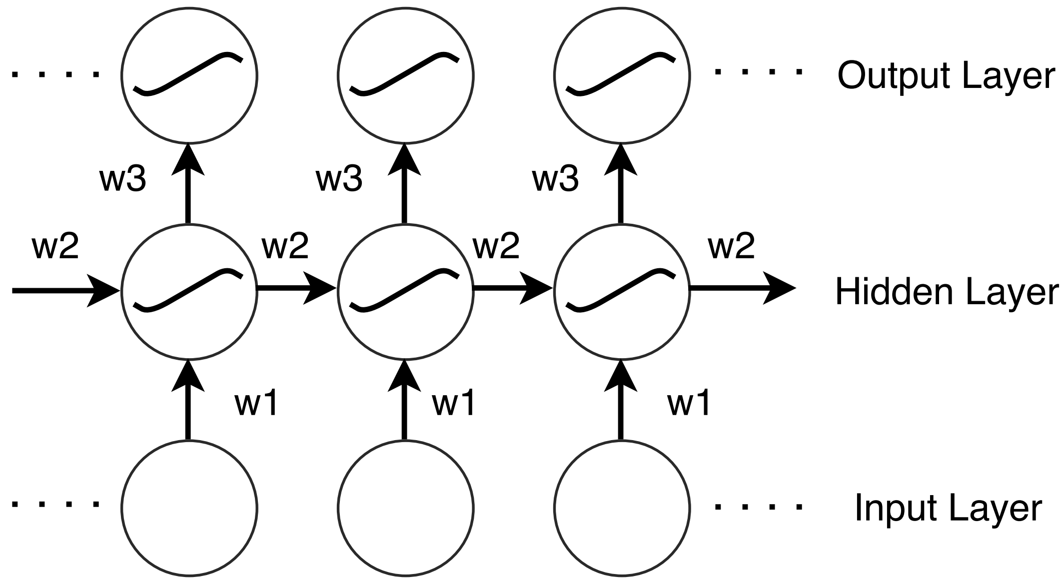


FIGURE 4.2: An unfolded recurrent neural network. w1, w2 and w3 are weighted connections.

a very low value of the cost function. For feedforward networks the weights are initialised with very small random numbers and the biased may be initilised to zero or very small values as well.

The choice of a cost function is somehow important and usually these are same as the linear models.

4.5 LSTM networks

In previous section we learn that using a recurrent neural networks we can store information in form of activations in the feedback connnections. The major disadvantage with recurrent neural networks is their inability to ratain information for a long period of time. This is caused by an effect known as *vanishing gradient problem*(Bengio et al. [11], Hochreiter et al. [79]). The vanishing gradient problem is depicted in the figure 4.3. Number of attempts were made in the 1990's to resolve this issue. Hochreiter and Schmidhuber [80] proposed a redesigned network called Long Short Term Momory(LSTM) to address this problem.

4.5.1 Architecture

An LSTM network is a set of recurrently connected LSTM blocks, also known as memory blocks, where each memory block has one or more memory cells and three units (input,

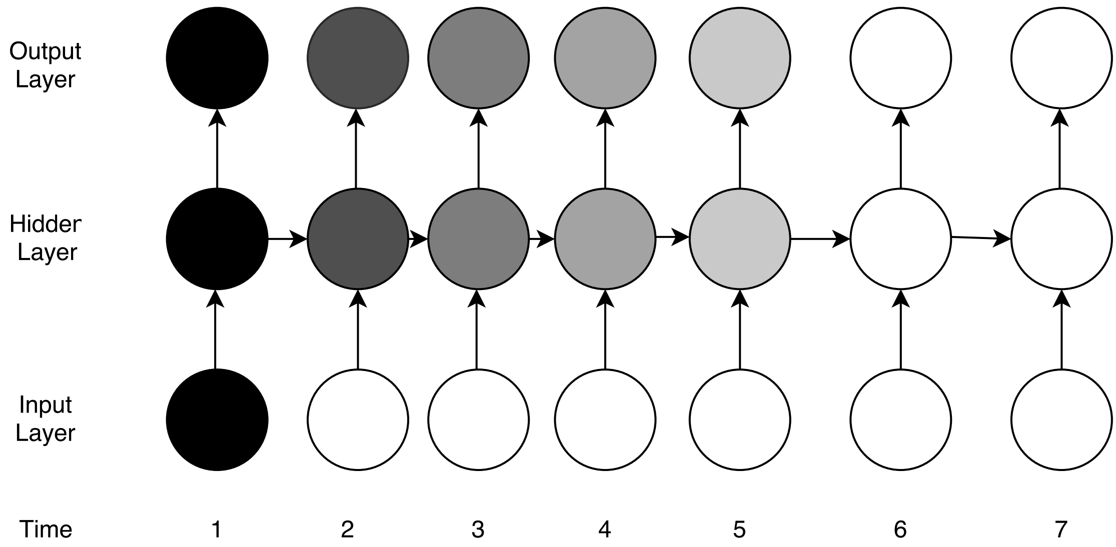


FIGURE 4.3: The problem of vanishing gradient in recurrent neural networks. The sensitivity, as indicated by the shading, gradually diminishes with time

output and forget gates) that perform the read, write and reset operations. A basic LSTM block with one memory cell is depicted in the figure 4.4. The multiplicative units allow the LSTM to store information for a long time and thus addresses the problem of vanishing gradient. An LSTM network is shown in figure 4.5, the hidden layers contains the LSTM blocks.

4.5.2 Training

4.6 A Stacked LSTM network for short term traffic prediction

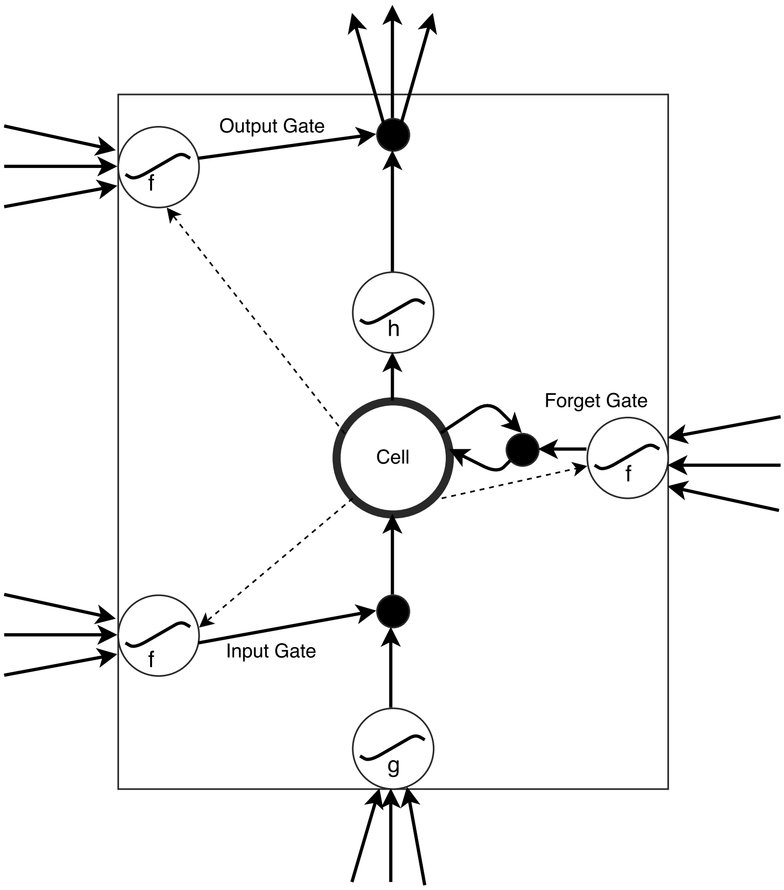


Figure 4.4: An LSTM block with one cell. The three units collect activations from both inside and outside of the block. The small black circles represents mulitipications by which the gates control the memory cell. The gate activation function is f, usually a logistic sigmoid. The cell input and output functions are g and h, usually tanh or logistic sigmoid. The dashed lines represent the weighted peephole connections from the cell to the gates. All other connections are not weighted. The only outputs from the block to the rest of the network is from the output gate multiplication.

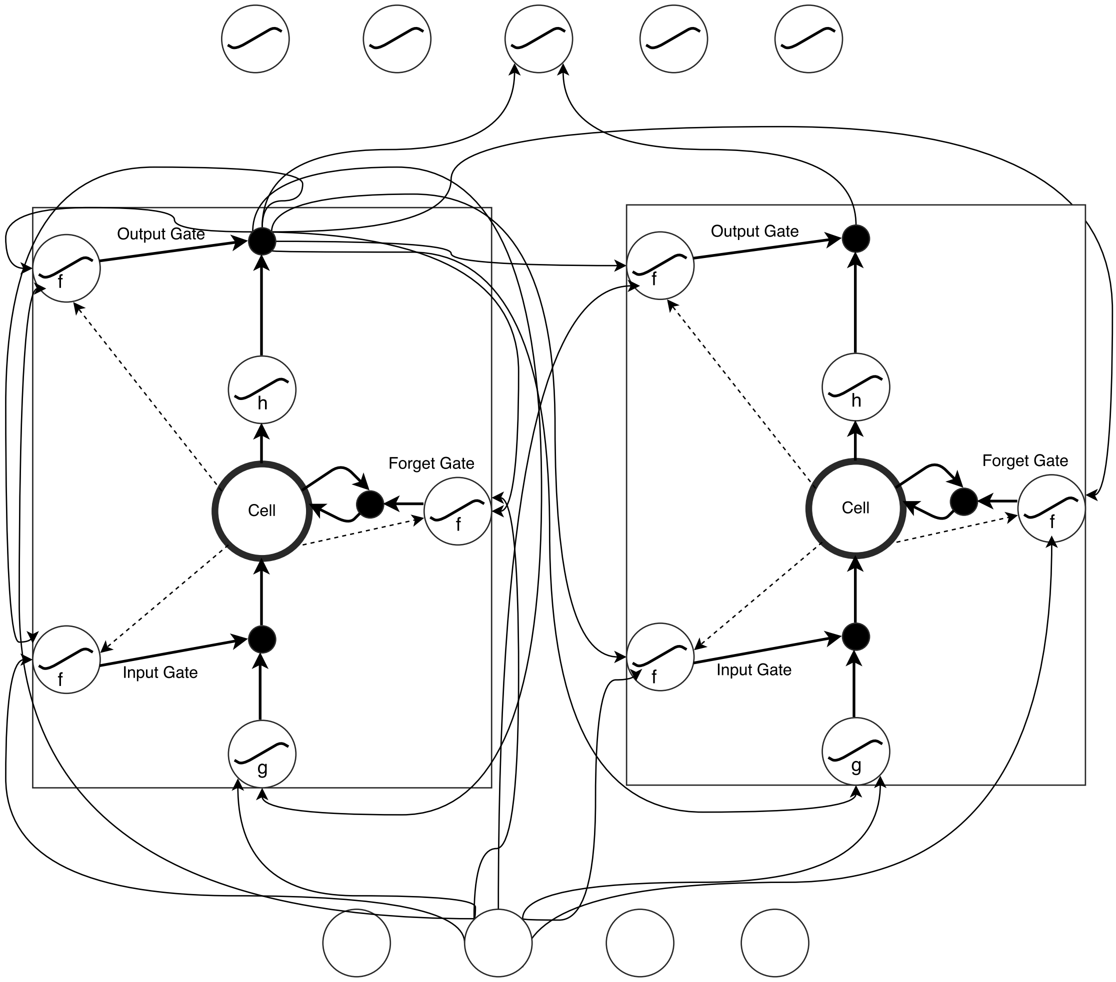


Figure 4.5: An LSTM network with one hidden layer with two memory blocks. The input layer consists of four input units and the output layer consists of five output units. Not all connections are shown in the figure. There is only one output from the block.

Chapter 5

Evaluation of the Model

"Science, my boy, is made up of mistakes, but they are mistakes which it is useful to make,because they lead little by little to the truth"

Jules Verne, *Journey to the Centre of the Earth* (1864)

5.1 Experimental setup

We chose a subset of the obtained traffic data for experimentation as shown in the figure 5.1. The region boundary of the subnetwork is denoted by the red line.

5.1.1 Training details

We trained our stacked LSTM network(4.6)

5.2 Results

For comparison purose we used the following methods that have been in used predominantly in short term traffic prediction - Naïve, Linear regression, ARIMA, Exponential smoothing and Feedforward neural network. In figure 5.2, we present the predictions of these models on test data. The input sequence was set to 96 observations(1 day) and the prediction was done for next 15 minutes(1 step ahead).

The results of the LSTM network are shown in the figure 5.3

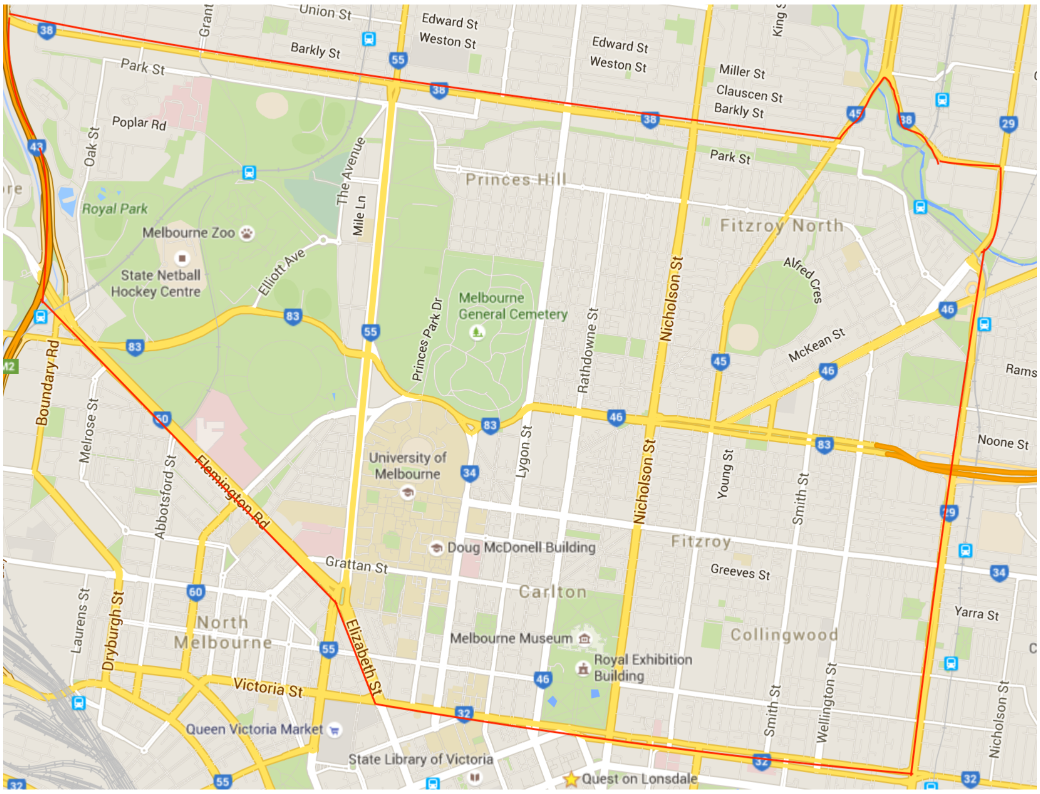


FIGURE 5.1: The traffic region used in this experiment. The boundary is dentoed by the red line.

In table 5.1, the performance of the LSTM network along with the compared models are given.

Several accuracy measures exist to evaluate a model. In below secions we describe the accuracy mesures and use those to evalute our proposed model against the benchmark models. For defining the accuracy measures let us denote x_i be the i^{th} observation and $\hat{x}_i$ be the prediction of x_i.

Scale-dependent errors The prediction error is simply given by $e_i = x_i - \hat{x}_i$, which is in the same scale as of the original data. So accuracy measures that depend on e_i are scale dependent and can not be used across multiple series on different scales. The two most used scale-dependent accuracy measures are mean absolute error and root mean squared error defined as below

$$MAE = mean(|e_i|) \tag{5.1}$$

$$RMSE = \sqrt{mean(e_i^2)} \tag{5.2}$$

MAE is easy to understand and popular in usage when using a single dataset.

Percentage errors Percentage errors are scale-independent and thus used across multiple datasets on different scales. The percentage error is given by $p_i = 100 * e_i/x_i$. The most commonly used percentage measure is Mean Absolute Percentage Error(MAPE) which is given by the below formula

$$MAPE = mean(|p_i|) \tag{5.3}$$

There are however few shortccomings of the MAPE, for instance when x_i is 0 or very large. Another shortcoming is that they put heavier penalty on negative error values than positve error values.

Scaled errors Hyndman and Koehler [94] proposed scaled errors to be used as an alternative in place of percentage errors. The proposed Mean Absolute Scaled Error(MASE) is defined as

$$MASE = mean(|q_i|) \tag{5.4}$$

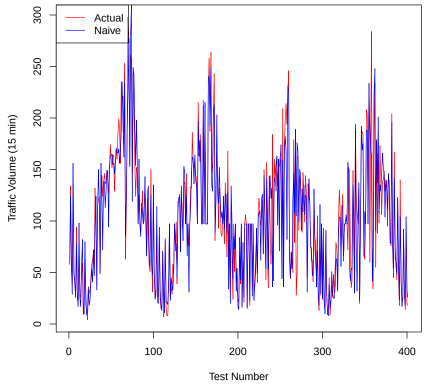
where

$$q_i = \frac{e_i}{\frac{1}{T-1}\sum_{t=2}^{T}|x_t - x_{t-1}|} \tag{5.5}$$

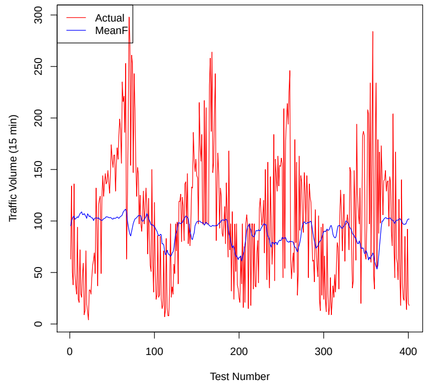
A scaled error is less than one if it is better than the average naïve forecast computed on the training data and vice versa.

Model	**ME**	**MAE**	**RMSE**	**MPE(%)**	**MAPE(%)**	**MASE**
15 minutes						
Mean Forecast	10.73	48.13	48.13	-63.61	98.34	1.57
Naive	-0.18	15.25	15.25	-5.65	21.24	0.49
Linear Regression	9.75	50.30	50.30	-62.20	97.84	1.65
ARIMA	1.87	15.10	15.10	-6.05	21.30	1.02
Eponential Smoothing	0.04	15.33	15.33	-5.48	20.68	1.04
Neural Network AutoRegression	5.61	19.40	19.40	-5.51	24.29	1.30
Deep LSTM(single location)						
Deep LSTM(multiple locations)						
30 minutes						
Mean Forecast	10.29	47.13	48.31	-63.63	98.27	1.54
Naive	-0.62	16.65	18.16	-6.35	23.02	0.54
Linear Regression	9.30	49.30	50.38	-62.67	98.30	1.62
ARIMA	2.42	16.56	18.02	-7.18	23.67	1.12
Eponential Smoothing	-0.61	16.66	18.08	-5.74	22.20	1.13
Neural Network AutoRegression	6.61	19.54	20.98	-4.51	25.06	1.33
Deep LSTM(single location)						
Deep LSTM(multiple locations)						

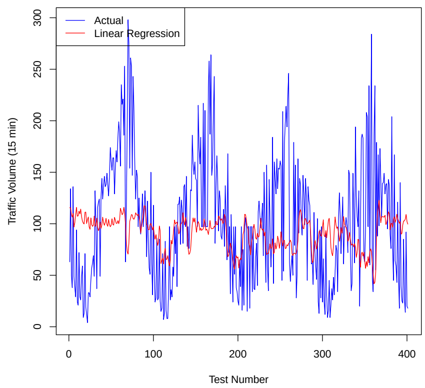
Table 5.1: Accuracy measures for the evaluted models. The scores are calculated for prediction horizon of 15-minutes(top half) and 30-minutes(bottom half). Mean 15-minutes traffic volume is 104.4



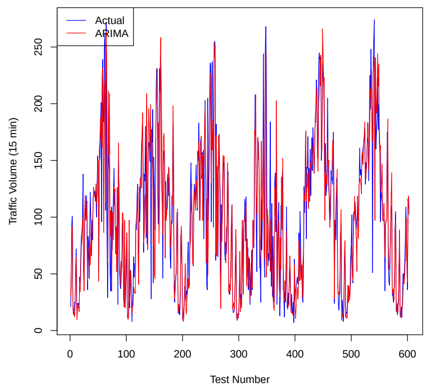
(A) Naïve



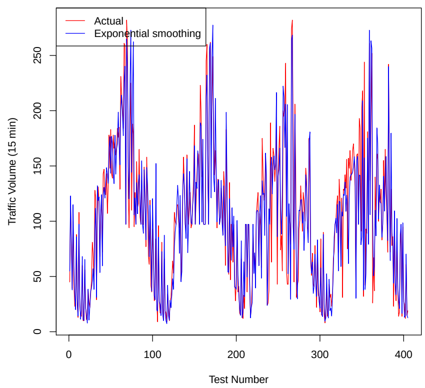
(B) Mean Forecast



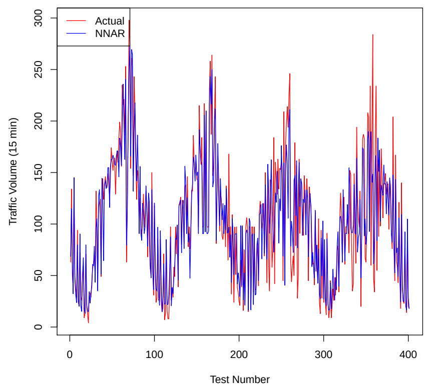
(C) Linear Regression



(D) ARIMA

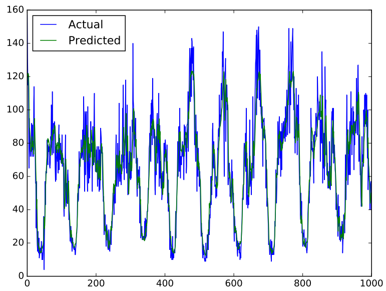


(E) Exponential smoothing state space model

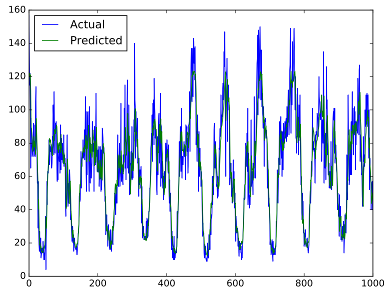


(F) Neural Network AutoRegression

FIGURE 5.2: Acutual vs Predictions - linear regression, ARIMA, feed forward neural network one hidden layer and exponential smoothing using state space model. The models were trained on trafiic data from one homogeneous road segment. The plots show the actual vs predictions(15 mins) on 400 test examples, for one of those road segment (Nicholson street between Gertrude street and Victoria Parade).



(A) Single location



(B) Multiple locations

FIGURE 5.3: Acutual vs Predictions - The left figure is from the model trained using data from a single homogeneous road segment only. The figure in right is from the model trained on 67 homogeneous road segements in the chosen subnetwork(5.1). Both the plots show the actual vs predictions(15 mins) on 400 test examples, for one of those road segment(Nicholson street between Gertrude street and Victoria Parade).

Chapter 6

Conclusions and Future Directions

"Everything should be made as simple as possible but not simpler."

Albert Einstein

6.1 Conclusions

In this work, we reviewed the existing literature on short term traffic prediction and proposed how a long short term momory recurrent neural network can be used for this task.

6.2 Future works

Appendix A

Appendix Title Here

Write your Appendix content here.

Bibliography

[1] Abdi, J., Moshiri, B., Abdulhai, B., and Sedigh, A. K. (2012). Forecasting of short-term traffic-flow based on improved neurofuzzy models via emotional temporal difference learning algorithm. *Engineering Applications of Artificial Intelligence*, 25(5):1022–1042.

[2] Abdulhai, B., Porwal, H., and Recker, W. (2002). Short-term traffic flow prediction using neuro-genetic algorithms. *ITS Journal-Intelligent Transportation Systems Journal*, 7(1):3–41.

[3] Abu-Lebdeh, G. and Singh, A. K. (2011). Modeling arterial travel time with limited traffic variables using conditional independence graphs & state-space neural networks. *Procedia-Social and Behavioral Sciences*, 16:207–217.

[4] Adeli, H. (2001). Neural networks in civil engineering: 1989–2000. *Computer-Aided Civil and Infrastructure Engineering*, 16(2):126–142.

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

[6] Alecsandru, C. and Ishak, S. (2004). Hybrid model-based and memory-based traffic prediction system. *Transportation Research Record: Journal of the Transportation Research Board*, (1879):59–70.

[7] Alecsandru, C.-D. (2003). *A hybrid model-based and memory-based short-term traffic prediction system*. PhD thesis, Citeseer.

[8] Antoniou, C., Koutsopoulos, H. N., and Yannis, G. (2013). Dynamic data-driven local traffic state estimation and prediction. *Transportation Research Part C: Emerging Technologies*, 34:89–107.

[9] Anwar, T., Liu, C., Vu, H. L., and Leckie, C. (2014). Spatial partitioning of large urban road networks. *EDBT*.

[10] Bengio, Y., Goodfellow, I. J., and Courville, A. (2016). *Deep learning*. MIT Press.

[11] Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *Neural Networks, IEEE Transactions on*, 5(2):157–166.

[12] Beskos, D., Michalopoulos, P., and Lin, J. (1985). Analysis of traffic flow by the finite element method. *Applied mathematical modelling*, 9(5):358–364.

[13] Bhaskar, A., Chung, E., and Dumont, A.-G. (2011). Fusing loop detector and probe vehicle data to estimate travel time statistics on signalized urban networks. *Computer-Aided Civil and Infrastructure Engineering*, 26(6):433–450.

[14] Bing, Q., Gong, B., Yang, Z., Shang, Q., and Zhou, X. (2015). Short-term traffic flow local prediction based on combined kernel function relevance vector machine model. *Mathematical Problems in Engineering*, 2015.

[15] Bishop, C. (2007). Pattern recognition and machine learning (information science and statistics), 1st edn. 2006. corr. 2nd printing edn.

[16] Boto-Giralda, D., Díaz-Pernas, F. J., González-Ortega, D., Díez-Higuera, J. F., Antón-Rodríguez, M., Martínez-Zarzuela, M., and Torre-Díez, I. (2010). Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks. *Computer-Aided Civil and Infrastructure Engineering*, 25(7):530–545.

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

[18] Boyce, D. (2012). Predicting road traffic route flows uniquely for urban transportation planning. *JSRSAI*, 42(1):77–91.

[19] Brockwell, P. J. and Davis, R. A. (2006). *Introduction to time series and forecasting*. Springer Science & Business Media.

[20] Buch, N., Velastin, S., Orwell, J., et al. (2011). A review of computer vision techniques for the analysis of urban traffic. *Intelligent Transportation Systems, IEEE Transactions on*, 12(3):920–939.

[21] Busseti, E., Osband, I., and Wong, S. (2012). Deep learning for time series modeling. Technical report, Technical report, Stanford University.

[22] Castillo, E., Menéndez, J. M., and Sánchez-Cambronero, S. (2008). Predicting traffic flow using bayesian networks. *Transportation Research Part B: Methodological*, 42(5):482–509.

[23] Castro-Neto, M., Jeong, Y.-S., Jeong, M.-K., and Han, L. D. (2009). Online-svr for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert systems with applications*, 36(3):6164–6173.

[24] Celikoglu, H. B. (2013). An approach to dynamic classification of traffic flow patterns. *Computer-Aided Civil and Infrastructure Engineering*, 28(4):273–288.

[25] Cetin, M. and Comert, G. (2006). Short-term traffic flow prediction with regime switching models. *Transportation Research Record: Journal of the Transportation Research Board*, (1965):23–31.

[26] Chan, K. Y., Dillon, T., Chang, E., and Singh, J. (2013a). Prediction of short-term traffic variables using intelligent swarm-based neural networks. *Control Systems Technology, IEEE Transactions on*, 21(1):263–274.

[27] Chan, K. Y., Dillon, T. S., and Chang, E.-J. (2013b). An intelligent particle swarm optimization for short-term traffic flow forecasting using on-road sensor systems. *Industrial Electronics, IEEE Transactions on*, 60(10):4714–4725.

[28] Chan, K. Y., Dillon, T. S., Singh, J., and Chang, E. (2012a). Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and levenberg–marquardt algorithm. *Intelligent Transportation Systems, IEEE Transactions on*, 13(2):644–654.

[29] Chan, K. Y., Khadem, S., and Dillon, T. S. (2012b). Optimization of neural network configurations for short-term traffic flow forecasting using orthogonal design. In *Evolutionary Computation (CEC), 2012 IEEE Congress on*, pages 1–7. IEEE.

[30] Chan, K. Y., Khadem, S., Dillon, T. S., Palade, V., Singh, J., and Chang, E. (2012c). Selection of significant on-road sensor data for short-term traffic flow forecasting using the taguchi method. *Industrial Informatics, IEEE Transactions on*, 8(2):255–266.

[31] Chang, H., Lee, Y., Yoon, B., and Baek, S. (2012). Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences. *IET Intelligent Transport Systems*, 6(3):292–305.

[32] Chang, T.-H., Chueh, C.-H., and Yang, L.-K. (2011). Dynamic traffic prediction for insufficient data roadways via automatic control theories. *Control Engineering Practice*, 19(12):1479–1489.

[33] Chen, C., Hu, J., Meng, Q., and Zhang, Y. (2011a). Short-time traffic flow prediction with arima-garch model. In *Intelligent Vehicles Symposium (IV), 2011 IEEE*, pages 607–612. IEEE.

[34] Chen, C., Wang, Y., Li, L., Hu, J., and Zhang, Z. (2012). The retrieval of intra-day trend and its influence on traffic prediction. *Transportation research part C: emerging technologies*, 22:103–118.

[35] Chen, H. and Grant-Muller, S. (2001). Use of sequential learning for short-term traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, 9(5):319–336.

[36] Chen, H., Grant-Muller, S., Mussone, L., and Montgomery, F. (2001). A study of hybrid neural network approaches and the effects of missing data on traffic forecasting. *Neural Computing & Applications*, 10(3):277–286.

[37] Chen, L. and Chen, C. P. (2007). Ensemble learning approach for freeway short-term traffic flow prediction. In *System of Systems Engineering, 2007. SoSE'07. IEEE International Conference on*, pages 1–6. IEEE.

[38] Chen, Y., Yang, B., Meng, Q., Zhao, Y., and Abraham, A. (2011b). Time-series forecasting using a system of ordinary differential equations. *Information Sciences*, 181(1):106–114.

[39] Cheng, T., Haworth, J., and Wang, J. (2012). Spatio-temporal autocorrelation of road network data. *Journal of Geographical Systems*, 14(4):389–413.

[40] Chrobok, R., Kaumann, O., Wahle, J., and Schreckenberg, M. (2004). Different methods of traffic forecast based on real data. *European Journal of Operational Research*, 155(3):558–568.

[41] Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting*, 5(4):559–583.

[42] Danech-Pajouh, M. and Aron, M. (1991). Athena: a method for short-term interurban motorway traffic forecasting. *Recherche Transports Sécurité*, (6).

[43] Davarynejad, M., Wang, Y., Vrancken, J., and Van den Berg, J. (2011). Multi-phase time series models for motorway flow forecasting. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, pages 2033–2038. IEEE.

[44] Dendrinos, D. S. (1994). Traffic-flow dynamics: a search for chaos. *Chaos, Solitons & Fractals*, 4(4):605–617.

[45] Dia, H. (2001). An object-oriented neural network approach to short-term traffic forecasting. *European Journal of Operational Research*, 131(2):253–261.

[46] Dimitriou, L., Tsekeris, T., and Stathopoulos, A. (2008). Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow. *Transportation Research Part C: Emerging Technologies*, 16(5):554–573.

[47] Dion, F. and Rakha, H. (2006). Estimating dynamic roadway travel times using automatic vehicle identification data for low sampling rates. *Transportation Research Part B: Methodological*, 40(9):745–766.

[48] Djuric, N., Radosavljevic, V., Coric, V., and Vucetic, S. (2011). Travel speed forecasting by means of continuous conditional random fields. *Transportation Research Record: Journal of the Transportation Research Board*, (2263):131–139.

[49] Dougherty, M. S. and Cobbett, M. R. (1997). Short-term inter-urban traffic forecasts using neural networks. *International journal of forecasting*, 13(1):21–31.

[50] Du, L., Peeta, S., and Kim, Y. H. (2012). An adaptive information fusion model to predict the short-term link travel time distribution in dynamic traffic networks. *Transportation Research Part B: Methodological*, 46(1):235–252.

[51] Dunne, S. and Ghosh, B. (2011). Regime-based short-term multivariate traffic condition forecasting algorithm. *Journal of Transportation Engineering*, 138(4):455–466.

[52] El Faouzi, N.-E., Leung, H., and Kurian, A. (2011). Data fusion in intelligent transportation systems: Progress and challenges–a survey. *Information Fusion*, 12(1):4–10.

[53] Fei, X., Lu, C.-C., and Liu, K. (2011). A bayesian dynamic linear model approach for real-time short-term freeway travel time prediction. *Transportation Research Part C: Emerging Technologies*, 19(6):1306–1318.

[54] Fowe, A. J. and Chan, Y. (2013). A microstate spatial-inference model for network-traffic estimation. *Transportation Research Part C: Emerging Technologies*, 36:245–260.

[55] Fries, R. N., Gahrooei, M. R., Chowdhury, M., and Conway, A. J. (2012). Meeting privacy challenges while advancing intelligent transportation systems. *Transportation Research Part C: Emerging Technologies*, 25:34–45.

[56] Fusco, G. and Colombaroni, C. (2013). An integrated method for short-term prediction of road traffic conditions for intelligent transportation systems applications. *7th WSEAS European Computing Conference, Dubrovnik*, pages 25–27.

[57] Gan, C. and Canghui, Z. (2009). Modeling and simulation of freeway short-term traffic flow prediction. In *Advanced Computer Control, 2009. ICACC'09. International Conference on*, pages 46–50. IEEE.

[58] Gentili, M. and Mirchandani, P. (2012). Locating sensors on traffic networks: Models, challenges and research opportunities. *Transportation research part C: emerging technologies*, 24:227–255.

[59] Gers, F. (2001). *Long short-term memory in recurrent neural networks*. PhD thesis, Universität Hannover.

[60] Gers, F. A., Schmidhuber, J., and Cummins, F. (2000). Learning to forget: Continual prediction with lstm. *Neural computation*, 12(10):2451–2471.

[61] Ghosh, B., Basu, B., and O'Mahony, M. (2007). Bayesian time-series model for short-term traffic flow forecasting. *Journal of transportation engineering*, 133(3):180–189.

[62] Ghosh, B., Basu, B., and O'Mahony, M. (2009). Multivariate short-term traffic flow forecasting using time-series analysis. *Intelligent Transportation Systems, IEEE Transactions on*, 10(2):246–254.

[63] Ghosh, B., Basu, B., and O'Mahony, M. (2010). Random process model for urban traffic flow using a wavelet-bayesian hierarchical technique. *Computer-Aided Civil and Infrastructure Engineering*, 25(8):613–624.

[64] Graves, A. (2012). *Supervised sequence labelling*. Springer.

[65] Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

[66] Guo, F., Polak, J. W., and Krishnan, R. (2010). Comparison of modelling approaches for short term traffic prediction under normal and abnormal conditions. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pages 1209–1214. IEEE.

[67] Guo, J., Huang, W., and Williams, B. M. (2012). Integrated heteroscedasticity test for vehicular traffic condition series. *Journal of Transportation Engineering*, 138(9):1161–1170.

[68] Guo, J., Huang, W., and Williams, B. M. (2014). Adaptive kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification. *Transportation Research Part C: Emerging Technologies*, 43:50–64.

[69] Guo, J., Huang, W., and Williams, B. M. (2015). Real time traffic flow outlier detection using short-term traffic conditional variance prediction. *Transportation Research Part C: Emerging Technologies*, 50:160–172.

[70] Guo, J. and Williams, B. (2010). Real-time short-term traffic speed level forecasting and uncertainty quantification using layered kalman filters. *Transportation Research Record: Journal of the Transportation Research Board*, (2175):28–37.

[71] Guo, J., Williams, B., and Smith, B. (2008). Data collection time intervals for stochastic short-term traffic flow forecasting. *Transportation Research Record: Journal of the Transportation Research Board*, (2024):18–26.

[72] Hamad, K., Shourijeh, M. T., Lee, E., and Faghri, A. (2009). Near-term travel speed prediction utilizing hilbert–huang transform. *Computer-Aided Civil and Infrastructure Engineering*, 24(8):551–576.

[73] Hamed, M. M., Al-Masaeid, H. R., and Said, Z. M. B. (1995). Short-term prediction of traffic volume in urban arterials. *Journal of Transportation Engineering*, 121(3):249–254.

[74] Haworth, J. and Cheng, T. (2012). Non-parametric regression for space–time forecasting under missing data. *Computers, Environment and Urban Systems*, 36(6):538–550.

[75] Haykin, S. S., Haykin, S. S., Haykin, S. S., and Haykin, S. S. (2009). *Neural networks and learning machines*, volume 3. Pearson Education Upper Saddle River.

[76] Heilmann, B., El Faouzi, N.-E., de Mouzon, O., Hainitz, N., Koller, H., Bauer, D., and Antoniou, C. (2011). Predicting motorway traffic performance by data fusion of local sensor data and electronic toll collection data. *Computer-Aided Civil and Infrastructure Engineering*, 26(6):451–463.

[77] Herrera, J. C., Work, D. B., Herring, R., Ban, X. J., Jacobson, Q., and Bayen, A. M. (2010). Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4):568–583.

[78] Hobeika, A. G. and Kim, C. K. (1994). Traffic-flow-prediction systems based on upstream traffic. *Vehicle Navigation and Information Systems Conference, 1994. Proceedings., 1994*.

[79] Hochreiter, S., Bengio, Y., Frasconi, P., and Schmidhuber, J. (2001). Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.

[80] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

[81] Hodge, V. J., Krishnan, R., Austin, J., Polak, J., and Jackson, T. (2014). Short-term prediction of traffic flow using a binary neural network. *Neural Computing and Applications*, 25(7-8):1639–1655.

[82] Högberg, P. (1976). Estimation of parameters in models for traffic prediction: a non-linear regression approach. *Transportation Research*, 10(4):263–265.

[83] Hong, W.-C. (2011). Traffic flow forecasting by seasonal svr with chaotic simulated annealing algorithm. *Neurocomputing*, 74(12):2096–2107.

[84] Hong, W.-C. (2012). Application of seasonal svr with chaotic immune algorithm in traffic flow forecasting. *Neural Computing and Applications*, 21(3):583–593.

[85] Hong, W.-C., Dong, Y., Zheng, F., and Lai, C.-Y. (2011a). Forecasting urban traffic flow by svr with continuous aco. *Applied Mathematical Modelling*, 35(3):1282–1291.

[86] Hong, W.-C., Dong, Y., Zheng, F., and Wei, S. Y. (2011b). Hybrid evolutionary algorithms in a svr traffic flow forecasting model. *Applied Mathematics and Computation*, 217(15):6733–6747.

[87] Horvitz, E. and Mitchell, T. (2010). From data to knowledge to action: A global enabler for the 21 st century. *Computing*.

[88] Horvitz, E. J., Apacible, J., Sarin, R., and Liao, L. (2012). Prediction, expectation, and surprise: Methods, designs, and study of a deployed traffic forecasting service. *arXiv preprint arXiv:1207.1352*.

[89] Hosseini, S. H., Moshiri, B., Rahimi-Kian, A., and Araabi, B. N. (2012). Short-term traffic flow forecasting by mutual information and artificial neural networks. In *Industrial Technology (ICIT), 2012 IEEE International Conference on*, pages 1136–1141. IEEE.

[90] Hu, J., Zong, C., Song, J., Zhang, Z., and Ren, J. (2003). An applicable short-term traffic flow forecasting method based on chaotic theory. In *Intelligent Transportation Systems, 2003. Proceedings. 2003 IEEE*, volume 1, pages 608–613. IEEE.

[91] Hu, S.-R., Peeta, S., and Chu, C.-H. (2009). Identification of vehicle sensor locations for link-based network traffic applications. *Transportation Research Part B: Methodological*, 43(8):873–894.

[92] Huang, S. and Sadek, A. W. (2009). A novel forecasting approach inspired by human memory: the example of short-term traffic volume forecasting. *Transportation Research Part C: Emerging Technologies*, 17(5):510–525.

[93] Huisken, G. (2006). *Inter-urban short-term traffic congestion prediction*. University of Twente.

[94] Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688.

[95] Innamaa, S. (2005). Short-term prediction of travel time using neural networks on an interurban highway. *Transportation*, 32(6):649–669.

[96] Innamaa, S. (2009). Self-adapting traffic flow status forecasts using clustering. *IET Intelligent Transport Systems*, 3(1):67–76.

[97] Ishak, S. and Al-Deek, H. (2002). Performance evaluation of short-term time-series traffic prediction model. *Journal of Transportation Engineering*, 128(6):490–498.

[98] Ishak, S. and Alecsandru, C. (2004). Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings. *Journal of Transportation Engineering*, 130(4):452–465.

[99] Jiang, X. and Adeli, H. (2004). Wavelet packet-autocorrelation function method for traffic flow pattern analysis. *Computer-Aided Civil and Infrastructure Engineering*, 19(5):324–337.

[100] Jiang, X. and Adeli, H. (2005). Dynamic wavelet neural network model for traffic flow forecasting. *Journal of transportation engineering*, 131(10):771–779.

[101] Jieni, X. and Zhongke, S. (2008). Short-time traffic flow prediction based on chaos time series theory. *Journal of Transportation Systems Engineering and Information Technology*, 8(5):68–72.

[102] Kamarianakis, Y., Gao, H. O., and Prastacos, P. (2010). Characterizing regimes in daily cycles of urban traffic using smooth-transition regressions. *Transportation Research Part C: Emerging Technologies*, 18(5):821–840.

[103] Kamarianakis, Y., Shen, W., and Wynter, L. (2012). Real-time road traffic forecasting using regime-switching space-time models and adaptive lasso. *Applied Stochastic Models in Business and Industry*, 28(4):297–315.

[104] Karlaftis, M. and Vlahogianni, E. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, 19(3):387–399.

[105] Karlaftis, M. G. and Vlahogianni, E. I. (2009). Memory properties and fractional integration in transportation time-series. *Transportation Research Part C: Emerging Technologies*, 17(4):444–453.

[106] Karpathy, A., Johnson, J., and Li, F.-F. (2015). Visualizing and understanding recurrent networks. *arXiv preprint arXiv:1506.02078*.

[107] Kaya, S., Kilic, N., Kocak, T., and Gungor, C. (2014). From asia to europe: Short-term traffic flow prediction between continents. In *Telecommunications (ICT), 2014 21st International Conference on*, pages 277–282. IEEE.

[108] Kerner, B. S., Klenov, S. L., Hermanns, G., and Schreckenberg, M. (2013). Effect of driver over-acceleration on traffic breakdown in three-phase cellular automaton traffic flow models. *Physica A: Statistical Mechanics and its Applications*, 392(18):4083–4105.

[109] Khan, A. M. (2012). Bayesian predictive travel time methodology for advanced traveller information system. *Journal of Advanced Transportation*, 46(1):67–79.

[110] Khosravi, A., Mazloumi, E., Nahavandi, S., Creighton, D., and Van Lint, J. (2011). Prediction intervals to account for uncertainties in travel time prediction. *Intelligent Transportation Systems, IEEE Transactions on*, 12(2):537–547.

[111] Kirby, H. R., Watson, S. M., and Dougherty, M. S. (1997). Should we use neural networks or statistical models for short-term motorway traffic forecasting? *International Journal of Forecasting*, 13(1):43–50.

[112] Krause, A., Horvitz, E., Kansal, A., and Zhao, F. (2008). Toward community sensing. *Proceedings of the 7th international conference on Information processing in sensor networks*.

[113] Kumar, S. V. and Vanajakshi, L. (2015). Short-term traffic flow prediction using seasonal arima model with limited input data. *European Transport Research Review*, 7(3):1–9.

[114] Lan, C.-J. and Miaou, S.-P. (1999). Real-time prediction of traffic flows using dynamic generalized linear models. *Transportation Research Record: Journal of the Transportation Research Board*, (1678):168–178.

[115] Li, C., Anavatti, S. G., and Ray, T. (2011). Short-term traffic flow prediction using different techniques. In *IECON 2011-37th Annual Conference on IEEE Industrial Electronics Society*, pages 2423–2428. IEEE.

[116] Li, Z., Li, Y., and Li, L. (2014). A comparison of detrending models and multi-regime models for traffic flow prediction. *Intelligent Transportation Systems Magazine, IEEE*, 6(4):34–44.

[117] Lin, S. (2011). *Efficient model predictive control for large-scale urban traffic networks*. PhD thesis, Delft University of Technology.

[118] Lippi, M., Bertini, M., and Frasconi, P. (2013). Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *Intelligent Transportation Systems, IEEE Transactions on*, 14(2):871–882.

[119] Liu, X., Fang, X., Qin, Z., Ye, C., and Xie, M. (2011). A short-term forecasting algorithm for network traffic based on chaos theory and svm. *Journal of network and systems management*, 19(4):427–447.

[120] Lu, C.-C. J. (2012). An adaptive system for predicting freeway travel times. *International Journal of Information Technology & Decision Making*, 11(04):727–747.

[121] Lv, Y., Tang, S., and Zhao, H. (2009). Real-time highway traffic accident prediction based on the k-nearest neighbor method. In *Measuring Technology and Mechatronics Automation, 2009. ICMTMA'09. International Conference on*, volume 3, pages 547–550. IEEE.

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

[123] Ma, Y., Chowdhury, M., Sadek, A., and Jeihani, M. (2012). Integrated traffic and communication performance evaluation of an intelligent vehicle infrastructure integration (vii) system for online travel-time prediction. *Intelligent Transportation Systems, IEEE Transactions on*, 13(3):1369–1382.

[124] Marfia, G., Roccetti, M., and Amoroso, A. (2013). A new traffic congestion prediction model for advanced traveler information and management systems. *Wireless Communications and Mobile Computing*, 13(3):266–276.

[125] McCrea, J. and Moutari, S. (2010). A hybrid macroscopic-based model for traffic flow in road networks. *European Journal of Operational Research*, 207(2):676–684.

[126] McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

[127] Meng, M., Shao, C.-f., Wong, Y.-d., Wang, B.-b., and Li, H.-x. (2015). A two-stage short-term traffic flow prediction method based on avl and aknn techniques. *Journal of Central South University*, 22:779–786.

[128] Miles, J. and Walker, A. J. (2006). The potential application of artificial intelligence in transport. In *IEE Proceedings-Intelligent Transport Systems*, volume 153, pages 183–198. IET.

[129] Min, W. and Wynter, L. (2011). Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, 19(4):606–616.

[130] Min, X., Hu, J., Chen, Q., Zhang, T., and Zhang, Y. (2009). Short-term traffic flow forecasting of urban network based on dynamic starima model. In *Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on*, pages 1–6. IEEE.

[131] Min, X., Hu, J., and Zhang, Z. (2010). Urban traffic network modeling and short-term traffic flow forecasting based on gstarima model. *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pages 1535–1540.

[132] Mu, T., Jiang, J., and Wang, Y. (2013). Heterogeneous delay embedding for travel time and energy cost prediction via regression analysis. *Intelligent Transportation Systems, IEEE Transactions on*, 14(1):214–224.

[133] Myung, J., Kim, D.-K., Kho, S.-Y., and Park, C.-H. (2011). Travel time prediction using k nearest neighbor method with combined data from vehicle detector system and automatic toll collection system. *Transportation Research Record: Journal of the Transportation Research Board*, (2256):51–59.

[134] Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814.

[135] Nanthawichit, C., Nakatsuji, T., and Suzuki, H. (2003). Application of probe-vehicle data for real-time traffic-state estimation and short-term travel-time prediction on a freeway. *Transportation Research Record: Journal of the Transportation Research Board*, (1855):49–59.

[136] Ng, M. (2012). Synergistic sensor location for link flow inference without path enumeration: A node-based approach. *Transportation Research Part B: Methodological*, 46(6):781–788.

[137] Nguyen, N. and Gaffney, J. (2006). Real-time traffic performance measures of adaptive traffic signal control systems. *Research into Practice: 22nd ARRB Conference*.

[138] Nicholson, H. and Swann, C. (1974). The prediction of traffic flow volumes based on spectral analysis. *Transportation Research*, 8(6):533–538.

[139] Nihan, N. L. and Holmesland, K. O. (1980). Use of the box and jenkins time series technique in traffic forecasting. *Transportation*, 9(2):125–143.

[140] Oh, C. and Park, S. (2011). Investigating the effects of daily travel time patterns on short-term prediction. *KSCE Journal of Civil Engineering*, 15(7):1263–1272.

[141] Okutani, I. and Stephanedes, Y. J. (1984). Dynamic prediction of traffic volume through kalman filtering theory. *Transportation Research Part B: Methodological*, 18(1):1–11.

[142] Ou, Q., Bertini, R. L., Van Lint, J., and Hoogendoorn, S. P. (2011). A theoretical framework for traffic speed estimation by fusing low-resolution probe vehicle data. *Intelligent Transportation Systems, IEEE Transactions on*, 12(3):747–756.

[143] Pan, T., Sumalee, A., Zhong, R.-X., and Indra-Payoong, N. (2013). Short-term traffic state prediction based on temporal-spatial correlation. *Intelligent Transportation Systems, IEEE Transactions on*, 14(3):1242–1254.

[144] Qiao, F., Yang, H., and Lam, W. H. (2001). Intelligent simulation and prediction of traffic flow dispersion. *Transportation Research Part B: Methodological*, 35(9):843–863.

[145] Qiu, C., Wang, C., Zuo, X., and Fang, B. (2011). A bayesian regularized neural network approach to short-term traffic speed prediction. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, pages 2215–2220. IEEE.

[146] Qu, L., Li, L., Zhang, Y., and Hu, J. (2009). Ppca-based missing data imputation for traffic flow volume: a systematical approach. *Intelligent Transportation Systems, IEEE Transactions on*, 10(3):512–522.

[147] Quek, C., Pasquier, M., and Lim, B. B. S. (2006). Pop-traffic: a novel fuzzy neural approach to road traffic analysis and prediction. *Intelligent Transportation Systems, IEEE Transactions on*, 7(2):133–146.

[148] Rice, J. and Van Zwet, E. (2004). A simple and effective method for predicting travel times on freeways. *Intelligent Transportation Systems, IEEE Transactions on*, 5(3):200–207.

[149] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.

[150] Shang, P., Li, X., and Kamae, S. (2005). Chaotic analysis of traffic time series. *Chaos, Solitons & Fractals*, 25(1):121–128.

[151] Sheu, J.-B., Lan, L. W., and Huang, Y.-S. (2009). Short-term prediction of traffic dynamics with real-time recurrent learning algorithms. *Transportmetrica*, 5(1):59–83.

[152] Simroth, A. and Zähle, H. (2011). Travel time prediction using floating car data applied to logistics planning. *Intelligent Transportation Systems, IEEE Transactions on*, 12(1):243–253.

[153] Smith, B. L. and Demetsky, M. J. (1994). Short-term traffic flow prediction models - a comparison of neural network and nonparametric regression approaches. *IEEE INTERNATIONAL CONFERENCE ON SYSTEMS MAN AND CYBERNETICS*, 2:1706.

[154] Smith, B. L. and Demetsky, M. J. (1997). Traffic flow forecasting: comparison of modeling approaches. *Journal of transportation engineering*.

[155] Smith, B. L., Williams, B. M., and Oswald, R. K. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, 10(4):303–321.

[156] Sohr, A. and Wagner, P. (2008). Short term traffic prediction using cluster analysis based on floating car data. In *15th World Congress on ITS*.

[157] Srinivasan, D., Chan, C. W., and Balaji, P. (2009). Computational intelligence-based congestion prediction for a dynamic urban street network. *Neurocomputing*, 72(10):2710–2716.

[158] Stathopoulos, A., Dimitriou, L., and Tsekeris, T. (2008). Fuzzy modeling approach for combined forecasting of urban traffic flow. *Computer-Aided Civil and Infrastructure Engineering*, 23(7):521–535.

[159] Stathopoulos, A., Karlaftis, M., and Dimitriou, L. (2010). Fuzzy rule-based system approach to combining traffic count forecasts. *Transportation Research Record: Journal of the Transportation Research Board*, (2183):120–128.

[160] Stathopoulos, A. and Karlaftis, M. G. (2003). A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies*, 11(2):121–135.

[161] Sun, S., Huang, R., and Gao, Y. (2012). Network-scale traffic modeling and forecasting with graphical lasso and neural networks. *Journal of Transportation Engineering*.

[162] Sun, S. and Xu, X. (2011). Variational inference for infinite mixtures of gaussian processes with applications to traffic flow prediction. *Intelligent Transportation Systems, IEEE Transactions on*, 12(2):466–475.

[163] Sun, S. and Zhang, C. (2007). The selective random subspace predictor for traffic flow forecasting. *Intelligent Transportation Systems, IEEE Transactions on*, 8(2):367–373.

[164] Szeto, W., Ghosh, B., Basu, B., and O'Mahony, M. (2009). Multivariate traffic forecasting technique using cell transmission model and sarima model. *Journal of Transportation Engineering*, 135(9):658–667.

[165] Tan, H., Feng, G., Feng, J., Wang, W., Zhang, Y.-J., and Li, F. (2013). A tensor-based method for missing traffic data completion. *Transportation Research Part C: Emerging Technologies*, 28:15–27.

[166] Tan, M.-C., Wong, S., Xu, J.-M., Guan, Z.-R., and Zhang, P. (2009). An aggregation approach to short-term traffic flow prediction. *Intelligent Transportation Systems, IEEE Transactions on*, 10(1):60–69.

[167] Tchrakian, T. T., Basu, B., and O'Mahony, M. (2012). Real-time traffic flow forecasting using spectral analysis. *Intelligent Transportation Systems, IEEE Transactions on*, 13(2):519–526.

[168] Teodorović, D. (2008). Swarm intelligence systems for transportation engineering: Principles and applications. *Transportation Research Part C: Emerging Technologies*, 16(6):651–667.

[169] Thomas, T., Weijermars, W., and Berkum, E. (2008). Variations in urban traffic volumes. *European Journal of Transport and Infrastructure Research*, 8(3):251–263.

[170] Tiriolo, M., Adacher, L., and Cipriani, E. (2014). An urban traffic flow model to capture complex flow interactions among lane groups for signalized intersections. *Procedia-Social and Behavioral Sciences*, 111:839–848.

[171] Tong, H. (2002). *Non-linear time series: a dynamical system approach*. Oxford University Press.

[172] Treiber, M. and Kesting, A. (2012). Validation of traffic flow models with respect to the spatiotemporal evolution of congested traffic patterns. *Transportation research part C: emerging technologies*, 21(1):31–41.

[173] Tsekeris, T. and Stathopoulos, A. (2009). Short-term prediction of urban traffic variability: Stochastic volatility modeling approach. *Journal of Transportation Engineering*, 136(7):606–613.

[174] Tsirigotis, L., Vlahogianni, E. I., and Karlaftis, M. G. (2012). Does information on weather affect the performance of short-term traffic forecasting models? *International Journal of Intelligent Transportation Systems Research*, 10(1):1–10.

[175] Turochy, R. E. (2006). Enhancing short-term traffic forecasting with traffic condition information. *Journal of Transportation Engineering*, 132(6):469–474.

[176] Van Arem, B., Kirby, H. R., Van Der Vlist, M. J., and Whittaker, J. C. (1997). Recent advances and applications in the field of short-term traffic forecasting. *International Journal of Forecasting*, 13(1):1–12.

[177] Van Der Voort, M., Dougherty, M., and Watson, S. (1996). Combining kohonen maps with arima time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies*, 4(5):307–318.

[178] van Hinsbergen, J. and Sanders, F. (2007). Short term traffic prediction models.

[179] Van Lint, J. (2006). Reliable real-time framework for short-term freeway travel time prediction. *Journal of transportation engineering*, 132(12):921–932.

[180] Van Lint, J. (2008). Online learning solutions for freeway travel time prediction. *Intelligent Transportation Systems, IEEE Transactions on*, 9(1):38–47.

[181] Van Lint, J., Hoogendoorn, S., and van Zuylen, H. J. (2005). Accurate freeway travel time prediction with state-space neural networks under missing data. *Transportation Research Part C: Emerging Technologies*, 13(5):347–369.

[182] Van Lint, J. and Hoogendoorn, S. P. (2010). A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways. *Computer-Aided Civil and Infrastructure Engineering*, 25(8):596–612.

[183] Van Lint, J. and Van Hinsbergen, C. (2012). Short term traffic and travel time prediction models, in artificial intelligence applications to critical transportation issues. *Transportation Research Circular, National Academies Press, Washington DC.*

[184] Van Lint, J. and Van Zuylen, H. (2005). Monitoring and predicting freeway travel time reliability: using width and skew of day-to-day travel time distribution. *Transportation Research Record: Journal of the Transportation Research Board*, (1917):54–62.

[185] Vlahogianni, E. and Karlaftis, M. (2013). Testing and comparing neural network and statistical approaches for predicting transportation time series. *Transportation Research Record: Journal of the Transportation Research Board*, (2399):9–22.

[186] Vlahogianni, E., Karlaftis, M. G., Golias, J. C., Kourbelis, N. D., et al. (2006a). Pattern-based short-term urban traffic predictor. In *Intelligent Transportation Systems Conference, 2006. ITSC'06. IEEE*, pages 389–393. IEEE.

[187] Vlahogianni, E. I. (2007). Prediction of non-recurrent short-term traffic patterns using genetically optimized probabilistic neural networks. *Operational Research*, 7(2):171–184.

[188] Vlahogianni, E. I. (2015). Optimization of traffic forecasting: Intelligent surrogate modeling. *Transportation Research Part C: Emerging Technologies*, 55:14–23.

[189] Vlahogianni, E. I., Golias, J. C., and Karlaftis, M. G. (2004). Short-term traffic forecasting: Overview of objectives and methods. *Transport reviews*, 24(5):533–557.

[190] Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. (2005). Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. *Transportation Research Part C: Emerging Technologies*, 13(3):211–234.

[191] Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. (2006b). Statistical methods for detecting nonlinearity and non-stationarity in univariate short-term time-series of traffic volume. *Transportation Research Part C: Emerging Technologies*, 14(5):351–367.

[192] Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. (2007a). Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks. *Computer-Aided Civil and Infrastructure Engineering*, 22(5):317–325.

[193] Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. (2008). Temporal evolution of short-term urban traffic flow: A nonlinear dynamics approach. *Computer-Aided Civil and Infrastructure Engineering*, 23(7):536–548.

[194] Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43:3–19.

[195] Vlahogianni, E. I., Webber, C. L., Geroliminis, N., and Skabardonis, A. (2007b). Statistical characteristics of transitional queue conditions in signalized arterials. *Transportation Research Part C: Emerging Technologies*, 15(6):392–403.

[196] Wang, J., Deng, W., and Guo, Y. (2014a). New bayesian combination method for short-term traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, 43:79–94.

[197] Wang, J., Shang, P., and Zhao, X. (2011). A new traffic speed forecasting method based on bi-pattern recognition. *Fluctuation and Noise Letters*, 10(01):59–75.

[198] Wang, J. and Shi, Q. (2013). Short-term traffic speed forecasting hybrid model based on chaos–wavelet analysis-support vector machine theory. *Transportation Research Part C: Emerging Technologies*, 27:219–232.

[199] Wang, Y. (2015). *On-line Distributed Prediction and Control for a Large-scale Traffic Network*. PhD thesis, Delft University of Technology.

[200] Wang, Y. and Papageorgiou, M. (2005). Real-time freeway traffic state estimation based on extended kalman filter: a general approach. *Transportation Research Part B: Methodological*, 39(2):141–167.

[201] Wang, Y., Papageorgiou, M., and Messmer, A. (2006). Renaissance–a unified macroscopic model-based approach to real-time freeway network traffic surveillance. *Transportation Research Part C: Emerging Technologies*, 14(3):190–212.

[202] Wang, Y., Van Schuppen, J. H., and Vrancken, J. (2014b). Prediction of traffic flow at the boundary of a motorway network. *Intelligent Transportation Systems, IEEE Transactions on*, 15(1):214–227.

[203] Washington, S. P., Karlaftis, M. G., and Mannering, F. L. (2010). *Statistical and econometric methods for transportation data analysis*. CRC press.

[204] Wei, D. and Liu, H. (2013). An adaptive-margin support vector regression for short-term traffic flow forecast. *Journal of Intelligent Transportation Systems*, 17(4):317–327.

[205] Weijermars, W. A. M. (2007). *Analysis of urban traffic patterns using clustering*. PhD thesis, University of Twente.

[206] Wild, D. (1997). Short-term forecasting based on a transformation and classification of traffic volume time series. *International Journal of Forecasting*, 13(1):63–72.

[207] Williams, B. (2001). Multivariate vehicular traffic flow prediction: Evaluation of arimax modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 1776(25):194–200.

[208] Williams, B. M. and Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal arima process: Theoretical basis and empirical results. *Journal of transportation engineering*, 129(6):664–672.

[209] Wu, C.-H., Ho, J.-M., and Lee, D.-T. (2004). Travel-time prediction with support vector regression. *Intelligent Transportation Systems, IEEE Transactions on*, 5(4):276–281.

[210] Xia, J., Huang, W., and Guo, J. (2012). A clustering approach to online freeway traffic state identification using its data. *KSCE Journal of Civil Engineering*, 16(3):426–432.

[211] Xiang, L., Ge, X.-H., Liu, C., Shu, L., and Wang, C.-X. (2010). A new hybrid network traffic prediction method. In *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, pages 1–5. IEEE.

[212] Xie, H. and Liu, Z. (2008). Short-term traffic flow prediction based on embedding phase-space and blind signal separation. In *Cybernetics and Intelligent Systems, 2008 IEEE Conference on*, pages 760–764. IEEE.

[213] Xie, Y., Zhang, Y., and Ye, Z. (2007). Short-term traffic volume forecasting using kalman filter with discrete wavelet decomposition. *Computer-Aided Civil and Infrastructure Engineering*, 22(5):326–334.

[214] Xie, Y., Zhao, K., Sun, Y., and Chen, D. (2010). Gaussian processes for short-term traffic volume forecasting. *Transportation Research Record: Journal of the Transportation Research Board*, (2165):69–78.

[215] Yang, M., Liu, Y., and You, Z. (2010). The reliability of travel time forecasting. *Intelligent Transportation Systems, IEEE Transactions on*, 11(1):162–171.

[216] Yang, Y.-n. and Lu, H.-p. (2010). Short-term traffic flow combined forecasting model based on svm. In *Computational and Information Sciences (ICCIS), 2010 International Conference on*, pages 262–265. IEEE.

[217] Yang, Z., Bing, Q., Lin, C., Yang, N., and Mei, D. (2014). Research on short-term traffic flow prediction method based on similarity search of time series. *Mathematical Problems in Engineering*, 2014.

[218] Yasdi, R. (1999). Prediction of road traffic using a neural network approach. *Neural computing & applications*, 8(2):135–142.

[219] Ye, Q., Szeto, W., and Wong, S. (2012). Short-term traffic speed forecasting based on data recorded at irregular intervals. *Intelligent Transportation Systems, IEEE Transactions on*, 13(4):1727–1737.

[220] Yin, H., Wong, S., Xu, J., and Wong, C. (2002). Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C: Emerging Technologies*, 10(2):85–98.

[221] Yuan, Y., Van Lint, J., Wilson, R. E., Wageningen-Kessels, V., Hoogendoorn, S. P., et al. (2012). Real-time lagrangian traffic state estimator for freeways. *Intelligent Transportation Systems, IEEE Transactions on*, 13(1):59–70.

[222] Yue, Y., Yeh, A. G., and Zhuang, Y. (2007). Prediction time horizon and effectiveness of real-time data on short-term traffic predictability. In *Intelligent Transportation Systems Conference, 2007. ITSC 2007. IEEE*, pages 962–967. IEEE.

[223] Zargari, S. A., Siabil, S. Z., Alavi, A. H., and Gandomi, A. H. (2012). A computational intelligence-based approach for short-term traffic flow prediction. *Expert Systems*, 29(2):124–142.

[224] Zeng, D., Xu, J., Gu, J., Liu, L., and Xu, G. (2008a). Short term traffic flow prediction based on online learning svr. In *Power Electronics and Intelligent Transportation System, 2008. PEITS'08. Workshop on*, pages 616–620. IEEE.

[225] Zeng, D., Xu, J., Gu, J., Liu, L., and Xu, G. (2008b). Short term traffic flow prediction using hybrid arima and ann models. In *Power Electronics and Intelligent Transportation System, 2008. PEITS'08. Workshop on*, pages 621–625. IEEE.

[226] Zhang, J., Wang, F.-Y., Wang, K., Lin, W.-H., Xu, X., and Chen, C. (2011). Data-driven intelligent transportation systems: A survey. *Intelligent Transportation Systems, IEEE Transactions on*, 12(4):1624–1639.

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

[228] Zhang, X., Onieva, E., Perallos, A., Osaba, E., and Lee, V. C. (2014a). Hierarchical fuzzy rule-based system optimized with genetic algorithms for short term traffic congestion prediction. *Transportation Research Part C: Emerging Technologies*, 43:127–142.

[229] Zhang, Y. and Ye, Z. (2008). Short-term traffic flow forecasting using fuzzy logic system methods. *Journal of Intelligent Transportation Systems*, 12(3):102–112.

[230] Zhang, Y., Zhang, Y., and Haghani, A. (2014b). A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model. *Transportation Research Part C: Emerging Technologies*, 43:65–78.

[231] Zheng, F. and Van Zuylen, H. (2013). Urban link travel time estimation based on sparse probe vehicle data. *Transportation Research Part C: Emerging Technologies*, 31:145–157.

[232] Zheng, W., Lee, D.-H., and Shi, Q. (2006). Short-term freeway traffic flow prediction: Bayesian combined neural network approach. *Journal of transportation engineering*, 132(2):114–121.

[233] Zhong, M., Sharma, S., and Lingras, P. (2005). Refining genetically designed models for improved traffic prediction on rural roads. *Transportation planning and technology*, 28(3):213–236.