LONG SHORT-TERM MEMORY RECURRENT NEURAL NETWORK FOR SHORT-TERM FREEWAY TRAFFIC SPEED PREDICTION

Junxiong Huang

A thesis submitted in partial fulfillment of the requirements for the degree of:

Master of Science

From the University of Wisconsin – Madison
Graduate School

Department of Civil and Environmental Engineering

April 2018

EXECUTIVE SUMMARY

Traffic congestions on Riverside Freeway, California, SR 91 usually happen on weekdays from 0:00 to 23:55 of January 3, 2011 to March 16, 2012. Traffic congestions can be detected from average mainline speed, the free flow speed fluctuates around 60 to 70 mph, the congestion speed fluctuate around 30 to 40 mph.

This research aims at predicting average mainline speed of the segment which is downstream to the merging road of a mainline and a ramp. The first objective is to determine the independent traffic features which can be used into the experiments, the second objective is to do data processing in order to form the data instances for the experiments, the third objective is to determine the predicted average speed with different time lags and to do the prediction, the fourth objective is to evaluate the performance of predicted average mainline speed.

This research data source is the Caltrans Performance Measurement System (PeMS), which is collected in real-time from over 39000 individual detectors. These sensors span the freeway system across all major metropolitan areas of the State of California. PeMS is also an Archived Data User Service (ADUS) that provides over ten years of data for historical analysis. It integrates a wide variety of information from Caltrans and other local agency systems includes traffic detectors, census traffic counts, incidents, vehicle classification, lane closures, weight-in-motion, toll tags and roadway inventory. For this research, the experimental data is 280 weekdays from January 3, 2011 to March 16, 2012, traffic speed and flow of each lane of mainline, traffic flow of each lane of ramp, the average mainline speed, the total mainline flow and the total ramp flow which was collected by one mainline detector and one on-ramp detector every 5 mins.

Previous studies proposed many different mathematical models, statistical models and machine learning models to predict traffic speed, flow and volume. The overall performance of the prediction is evaluated with overall accuracy values which measured with root-mean-square error (RMSE) and mean absolute percentage error (MAPE), besides, the predicted traffic speed is used to detect breakdown of congestions. However, the performance of existing traffic speed prediction models and methodologies has been criticized, many of the studies focus on simple data preparation and model application, traffic speed and flow are time series based data, the data set is very large, in order to improve the computation ability, feature selection, data processing and model selection are necessary for transportation research, in addition, the evaluation methods should also be considered to evaluate the performance of predictions.

Upon analyzing the data for this research, two external effects were immediately evident, the first of which was analysis of the traffic data set to identify the independent, the data processing to generate traffic data instances, cross validation to generate training and test sets for the prediction experiments. The second external effect was the structure of the architecture of the Long short-term Memory Recurrent Neural Network (LSTM) for the future 5 to 30 min average mainline speed prediction. The presence of traffic congestions caused speed to fluctuate, rise or fall in ways, these traffic congestions were not identified to distribute in the same time range of a day through the analysis.

It is difficult to capture the breakdown points of traffic congestions, especially the duration of traffic congestions is small, LSTM applications can capture most of the breakdown of traffic

congestions even minor rapid fluctuation of average mainline speed across the 280 days. According to the research, most of the breakdown starting points of predicted average mainline speed can match the real breakdown starting points of average mainline speed, besides, the overall RMSE and MAPE results achieved 7.5 and 4.56 for future 30 min average mainline speed prediction.

Acknowledgments

I would like to thank everyone who supported me throughout this research. Specifically, I would like to thank my advisor Dr. Bin Ran for his invaluable support and advise for my Master studies in University of Wisconsin-Madison. I would like to thank Dr. Sue Ahn for providing me her invaluable advice on Traffic Flow Theory and ideas of this research. I would like to thank Dr. Yang Cheng for being member of my thesis defense committee. A special thank you to my beloved family back home whose love and support are my motivation and the reason I stand here today.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Short-term Traffic Speed Prediction (STTSP) is an indispensable component of Intelligent Transportation System (ITS) which has received enormous attentions over the past two decades. Consequently, most effort has gone into developing methodologies that can be used to model traffic characteristics and produce anticipated traffic conditions. Existing literature is voluminous, and has largely used simple data collected from in-ground loop detectors, road sensors or GPS probes and has employed univariate or multivariate mathematical models to predict traffic speed or travelling times.

The purpose of highway traffic speed prediction is to use predicted traffic speed to identify the traffic congestions on the highway, so it is important to improve the efficiency of the existing highway network, with the accurate and real-time prediction of traffic speed for road segment can give travelers the ability to choose better routes and manage appropriate travelling time. Deep learning is part of a broader family of machine learning methods based on learning data representations, as opposed to forecasting algorithms, which can capture the sharp discontinuities in traffic speed that arise in large-scale networks.

Both the technological aspects of this analysis and the analytical have been the focus of countless research papers over the past few years [1,2,3], the two well-known modeling approaches for short-term prediction methods can broadly be classified into parametric and nonparametric techniques [4,5,6]. The goal of this paper is to model the temporal effects in recurrent and non-recurrent traffic congestion patterns and provide reliable traffic information in macroscopic way. These effects arise due to conditions at construction zones, weather, holiday, events, on road incident, off road incidents etc. Quantifying traveling time uncertainty requires real-time forecasts, travelers use the macroscopic information to adjust travel decisions on departure time and traveling route choice.

Deep learning of Long Short-term Memory Recurrent Neural Network (LSTM) forecasts congestion propagation given a bottleneck location with time lags, which can accurately capture the breakdown starting points of traffic congestion.

1.1 Statement of the problem

When people want to reach the destination quickly, they tend to choose the fastest route. However, because of the changing traffic speed, decisions cannot be made merely based on the spatial distances. Therefore, time-dependent traffic speed information between the origin and the destination associated with congestion situations are particularly important for travelers to plan flexible traveling route and traveling time.

Short-term traffic speed prediction can predict traffic speed of a certain segment in 5 to 30 min, for instance, a car will reach a highway road segment in 10 min based on the current traveling speed, but he has been informed that the highway average traffic speed of the segment will be decreased to 40 mph in 5 min, and the congestion will happen, the driver can adjust his traveling route ahead. Therefore, short-term traffic speed prediction is essential for travelers to plan appropriate traveling route and time.

1.2 Objectives

The primary objective of this research is to evaluate the relationship between highway mainline average traffic speed and backpropagation characteristics of highway mainline traffic congestions. The key objectives of this research are as follows:

- Determine the independent traffic features which can be used into the highway mainline traffic average speed prediction experiments.
- Propose data processing in order to form the data instances for the highway mainline traffic average speed prediction experiments.
- Determine the predicted highway mainline traffic average speed with different time lags for prediction.
- Evaluate the performance of predicted highway mainline average traffic speed and compare the results with different models.

1.3 Organization

This research is organized into five chapters, references and an appendix. The current Chapter (Chapter 1) comprises of an introduction to the research along with the hypotheses and objectives of the research. Chapter 2 represents a literature review relevant to this research. Chapter 3 describes the methodology used in the research, as well as the tests used to complete the analysis. Chapter 4 documents the results obtained and provides a detailed analysis of the acquired data. Chapter 4 describes the data and exploratory data analysis, Chapter 5 evaluates the results and measurements. Chapter 6 includes all the references used to complete this research.

2. LITERATURE REVIEW

2.1 Applications of Intelligent Transportation System (ITS)

Intelligent Transportation System (ITS) is an advanced application which aims to provide innovative services relating to transport, traffic management and enable users to be better informed and make safer, sophisticated road and telecommunication infrastructure are used to communicate between vehicles and the highway in order to improve vehicle and road efficiency.

ITS is described as a generic concept that covers a wider range of technological systems such as Advanced Driver Assistance Systems (ADAS), In-Vehicle Information Systems (IVIS) and Roadside Telematics (RT) [7]. ITS is described as the use of information and communication technologies to facilitate the seamless transportation of people and goods on the European Commission's White paper about the European transport policy for 2010 [8].

The key applications of ITS are as follows:

- Adaptive Cruise Control (ACC) as an adaptation of the Advanced Driver Assistance System (ADAS) technology which exploits the radar technology to enhance the maintain longitudinal motion and speed among vehicles.
- Obstacle Warning technology utilizes the radar, ultrasound, infrared and laser knowhow to detect possible collision while the vehicle is in motion.
- Lane Detection estimates the direction of the road and the position of the moving vehicle within the lane.
- Collision notification and avoidance is an application of intelligent vehicle technology that is developed and design to detect and report the enormity as well as the exact location of incidents occurrence to agencies and services.

- ITS Applications on Infrastructure is a success on video based detection and mobile equipment using wireless communication techniques.

2.2 Overview on Short-term Traffic Prediction

Short-term traffic prediction has been integral of Intelligent Transportation System (ITS) and related research, which concerns the predictions made from few seconds to few hours into future based on current and past traffic information. Most of the research has been focused on developing models and methodologies to model traffic characteristics such as speed, volume, flow, density and travel time.

Most effort of short-term traffic prediction has gone as follows:

- Proposing data from motorways and freeways.
- Developing univariate or multivariate statistical models.
- Predicting traffic speed, volume, flow and travel time.
- Collecting data from single data source like loop detectors, GPS or probes.

2.3 Challenges of Short-term Traffic Prediction

2.3.1 Reliable algorithms and prediction schemes

Transportation research requires the predictions are robust to short and long term changes in traffic conditions, in cases of the unexpected conditions includes accidents, adverse weather conditions, work zones, constructions. Reliable algorithms and prediction schemes are very important, but difficult to structure, as the relationship between short-term traffic prediction and the non-recurrent events is complex and unclear, therefore, prediction algorithms that can incorporate the effect of non-recurrent conditions can help enhance the decision making ability.

Vlahogianni and Kamarianakis develop multi-regime models to account for shifts of traffic between congested and uncongested conditions [9,10], Van Lint, Castro-Neto, Fei and Min incorporate the effects of accidents and the adverse weather into predictions [11,12,13,14]. Results from reported, the more complex information of algorithms and prediction schemes anticipated into traffic conditions, the more robust structures of ITS applications can be realized.

2.3.2 Freeway, arterial and traffic network prediction

Most of short-term traffic predictions are to focus on a freeway, arterial or corridor, short-term prediction on urban arterial forms a more complex problem than freeway predictions due to the constraints such as signalization [15], data driven approaches can provide the structurally alternative to account for the signalization's unpredictability, but still a challenging task for traffic network prediction. The difficulty is the covering a sufficient part of the road network by sensors, as well as the complex interactions in densely populated urban road networks, which among the significant obstacles in short-term traffic prediction.

The ability of data driven approaches to develop spatio-temporal interrelations and predict traffic has been documented, Chen and Haworth develop multivariate kernel regression models are proposed to predict travel time in traffic network [16,17], Sun implements robust artificial intelligence algorithms for short-term traffic flow prediction for traffic network [18], however, traffic network prediction is still at early stage.

2.3.3 Traffic volume to travel time prediction

Travel time prediction has attracted many researchers and practitioners because it is the straightforward method to inform travelers of traffic conditions. Various univariate and multivariate methodologies have been proposed to model average travel time, however, the interactions between factors like heavy vehicles, speed, volumes, weather and travel time are hard to control, the existing literature incorporate data from different source to in travel time predictions, besides, artificial neural network as the robust technology has been proposed for travel time prediction [19,20,21].

2.3.4 Data resolution, aggregation and quality

Data collection technologies provide the opportunities for acquiring traffic data at a variety of resolutions to meet the requirement for traffic prediction, the higher resolution of the data (i.e. 30s), the larger portion of noise involved in the time series traffic data. Several approaches have been utilized for reducing noise from time series before processing traffic predictions, moving average, wavelets and fuzzy algorithms are proposed to aggregate traffic data [22,23].

2.3.5 Temporal and spatial dependence

Identifying temporal and spatial dependence of traffic variables has been an important consideration for short-term traffic prediction, which is usually introduced into the modeling phase through temporal and spatial correlations [24], the accurate temporal-spatial representation in the framework of prediction schemes is of importance in integrating ITS applications, attempting to integrate temporal-spatial information into the short-term prediction algorithms has its advantages and disadvantages, the advantages are to consider the data dependence, and the disadvantages are to make the traffic prediction be more complex and more difficult, the accuracy may not be improved as well.

2.3.6 Model selection

Short-term traffic prediction is an excellent field for developing and testing complex prediction algorithms because of the abundance of available traffic data and the resolutions of traffic data. Traffic prediction can be regarded as different aspects: time series problems [25], regression and approximation problems [26], clustering [27], pattern recognition [28] and combinations of different aspects [9]. To propose the experiments, statistical learning, mathematical methodologies, machine learning, deep learning etc., have been applied into. In the framework, there are two difficulties needed to be considered, first is the model selection, which follows the short-term traffic prediction characteristics to select the models that provides the most accurate predictions based on the collected traffic data and regardless of the traffic underlying's characteristics. The second difficulty is to select model's performance, most of the emphasis on discussion on findings and neglect the quality of models.

2.3.7 Realizing full potential of artificial intelligence

Artificial Intelligence (AI) is the key technology of many of today's transportation applications [29], the advantages of AI applications over the other alternatives lies in their interdisciplinary nature and ability to straightforwardly predictions. However, developing efficient AI transportation system is complex, and there has been increased interest among both researchers and practitioners for exploring the feasibility of applying AI in improving the efficiency, safety and environmental-protections of transportation system [30]. AI technology has been applied into all aspects of traffic predictions, which is a strong candidate that may provide novel and easily deployable data mining tools.

2.4 Connections with existing work

Short-term traffic speed prediction (STTSP) has a long history in transportation literature, deep learning as a form of machine learning is the application to learn tasks of artificial neural networks (ANNs) that contain more than one hidden layer. Adeli [1] and van Lint [31] reviewed neural network and AI applications to short-term traffic forecasting, collecting and analyzing the literature using such approaches. Vlahogianni provides an overview of realizing the full potential of artificial intelligence and shows AI is the key technology in many of today's transportation applications [32], the advantage of AI applications over alternatives lies in their interdisciplinary nature and ability to straightforwardly combine forecasts, ease of modeling and computing, and relative associated autonomy [33]. Therefore, there has been increased interest among both researchers and practitioners for exploring the feasibility of applying AI into improving efficiency, safety, reliability and environmental-protection of ITS.

Machine learning is evolved from the study of pattern recognition and computational learning theory in AI, which is a method of data analysis that automates analytical model building, using algorithms that iteratively learning from data and allows computers to find hidden insights without being explicitly programmed where to look. In the past years, K-nearest neighbors (KNN), decision tree, support vector machine (SVM), random forest, linear regression etc., which are very common machine learning algorithms and widely applied into traffic prediction. Zheng and Su develop an optimized KNN model for short-term traffic volume forecasting [34], which can control extreme values' undesirable impact. Crosby proposes decision tree for UK traffic network forecasting, cross validation is incorporated into training and the accuracy of the approach is 88.2% [35]. Random forest as an algorithm of splitting decision tree, which correct for decision tree's habit of overfitting to their training set, Hamner adopts context-dependent random forest for traffic flow forecasting [36].

ANN is computing system inspired by the biological neural networks that constitute animal brains, because of the powerful computing ability, ANN has been widely used into traffic prediction, Smith demonstrates the comparison of neural network and nonparametric regression approaches [37], Yin used a fuzzy-neural approach for urban traffic flow forecasting [38], Vlahogianni provides a strategy of optimization for traffic prediction [39], Zheng combines the forecasting from single neural network predictors according to an adaptive and heuristic credit assignment algorithm based on the theory of conditional probability and Bayes' rule for short-term freeway traffic prediction [40], Manoel proposes an online Support Vector Regression (SVR) approach for short-term traffic flow forecasting [41], and Tan aggregated moving average, exponential smoothing, autoregressive moving average and neural network models for short-term traffic flow forecasting [42].

Due to challenge of data resolution, aggregation and quality, the higher data resolution, the larger the portion of noise time series of traffic variables and consequently, several approaches have utilized for reducing noise from time series before proceeding with forecasting, these approaches are from simple smoothing to wavelets, fuzzy and deep learning of filtering algorithms, either approach performs well only in particular period, an improved approach is combined these single predictors together for forecasting in a span of periods. Sun proposes Bayesian network and Gaussian mixture model work well for urban traffic flow forecasting with complete and incomplete data [43], which have proved Bayesian network based model for traffic flow forecasting and traffic states estimation perform well. Polson reviews several analytical approaches to traffic flow modeling [44,45,46,47], and these approaches perform well on filtering and traffic states estimation. In addition, Polson proposes an architecture of deep learning algorithm of several filtering for traffic states transitions between free flow, breakdown, congestion and recovery, and shows deep learning theory architecture can capture the nonlinear spatio-temporal effects of transportation.

3. METHODOLOGY

3.1 Data Processing

3.1.1 Feature Selection

Machine learning aims to address larger, more complex tasks, the problems of focusing on the most relevant information in a potentially overwhelming quantity of data has become increasingly important, this becomes even more important when the number of features are very large, feature selection technique enables the machine learning algorithm to train faster, reducing time complexity of models and making it easier to interpret to improve the accuracy of models if the right subset is chosen and reduce overfitting.

The problem is to predict the average mainline speed L steps after the current time $v_m^A(t_c + L) := v_m^A(t_c + 5 : t_c + L)$. The research dataset is composed of 12 features, which are speed and flow of each lane of mainline, flow of each lane of ramp, the average mainline speed, the total mainline flow and the total ramp flow, the correlations of the 12 features are shown as follows:

$$v_m^A(t) = \frac{v_m^1(t)q_m^1(t) + v_m^2(t)q_m^2(t) + v_m^3(t)q_m^3(t)}{q_m^1(t) + q_m^2(t) + q_m^3(t)},$$

$$q_m^T(t) = q_m^1(t) + q_m^2(t) + q_m^3(t),$$

$$q_r^T(t) = q_r^1(t) + q_r^2(t) + q_r^3(t)$$

In order to propose independent features for research, the 4 cases of traffic features are selected shown as follows,

a. Average mainline speed

$\{v_m^A(t), t_0 \le t \le t_c\}$

b. Average mainline speed and total mainline flow

$\{v_m^A(t), q_m^T(t), t_0 \le t \le t_c\}$

c. Speed and flow of each lane of mainline

$\{v_m^1(t), v_m^2(t), v_m^3(t), q_m^1(t), q_m^2(t), q_m^3(t), q_r^1(t), t_0 \le t \le t_c\}$

d. Speed and flow of each lane of mainline and ramp

$\{v_m^1(t), v_m^2(t), v_m^3(t), q_m^1(t), q_m^2(t), q_m^3(t), q_r^1(t), q_r^2(t), q_r^3(t), t_0 \le t \le t_c\}$

3.1.2 Instances Generation

The goal of this research is to predict the average mainline speed L steps after the current time which is expressed as $v_m^A(t_c + L) := v_m^A(t_c + 5: t_c + L)$, in this case, the reliable data instances should be considered to train the models, therefore, traffic features with different time lags are considered to propose into research, stacking the traffic features of the past 90 min time lags are proposed into research, which is shown as follows,

a. Average mainline speed

$$X(t_c) = \begin{pmatrix} v_m^{(A)}(t_c - 90) \\ v_m^{(A)}(t_c - 85) \\ \vdots \\ v_m^{(A)}(t_c) \end{pmatrix}$$

b. Average mainline speed and total mainline flow

$$X(t_c) = \begin{pmatrix} v_m^{(A)}(t_c - 90) & q_m^{(T)}(t_c - 90) \\ v_m^{(A)}(t_c - 85) & q_m^{(T)}(t_c - 85) \\ \vdots & \vdots \\ v_m^{(A)}(t_c) & q_m^{(T)}(t_c) \end{pmatrix}$$

c. Speed and flow of each lane of mainline

$$X(t_c) = \begin{pmatrix} v_m^{(1)}(t_c - 90) & \dots & q_m^{(3)}(t_c - 90) \\ v_m^{(1)}(t_c - 85) & \cdots & q_m^{(3)}(t_c - 85) \\ \vdots & \vdots & \vdots \\ v_m^{(1)}(t_c) & \cdots & q_m^{(3)}(t_c) \end{pmatrix}$$

d. Speed and flow of each lane of mainline and ramp

$$X(t_c) = \begin{pmatrix} v_m^{(1)}(t_c - 90) & \dots & q_m^{(1)}(t_c - 90) & \dots & q_r^{(3)}(t_c - 90) \\ v_m^{(1)}(t_c - 85) & \cdots & q_m^{(1)}(t_c - 85) & \cdots & q_r^{(3)}(t_c - 85) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ v_m^{(1)}(t_c) & \cdots & q_m^{(1)}(t_c) & \cdots & q_r^{(3)}(t_c) \end{pmatrix}$$

Where $X(t_c)$ denotes the input traffic information at time of t_c.

Therefore, the data instance $X(t_c)$ is used to predict average mainline speed in 5 to 30 min as shown as follows,

$$X(t_c) \rightarrow v_m^A(t_c + 5: t_c + 30)$$

3.1.3 Training and Test Data Generation

5-fold Cross Validation [48] are proposed to generate training and test sets for the research, as mentioned above, there are 280 days in the dataset, we regard each day as the independent data instance, thus, there are 5 pairs of training and test datasets, each pair has 56 days as the test set, the rest 224 days as the training set, the training sets are used for modeling, the test sets are used for validation.

3.2 Model Selection

Model selection is the task of selecting a model from a set of candidate models with the given dataset, in most cases, model selection is the fundamental task of scientific inquiry, which determines the principle that explains a series of observations, especially the observations of dataset. This research dataset is time series based dataset, short-term traffic prediction concerns time continuities and discontinuities a lot, besides, there are many features we need to consider, thus, the computation ability of the models should be considered.

3.2.1 Compared Machine Learning Models

In this research, several popular machine learning models are proposed for experiments, each of them has been proved as robust model in research or practice.

Decision Tree [49], as a decision support tool that uses a tree-like graph of modeling the decisions to identify the possible consequences. Random Forest [50], which operates as multiple decision trees, and get the mean value of the outcomes. Linear Regression [51] is a linear approach for modeling the relationship between scalar dependent variables. Bagging [52] is an ensemble method to reduce variance and help to avoid overfitting, it can be used for decision tree and many other models. Boosting [53] is an ensemble method to reduce bias and variance to help a set of weak learners to create a single strong learner.

3.2.2 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is inspired by interconnected neurons in biological system which consists of simple processing units, each unit receives a number of real-value input, and each unit produces a single real-valued output.

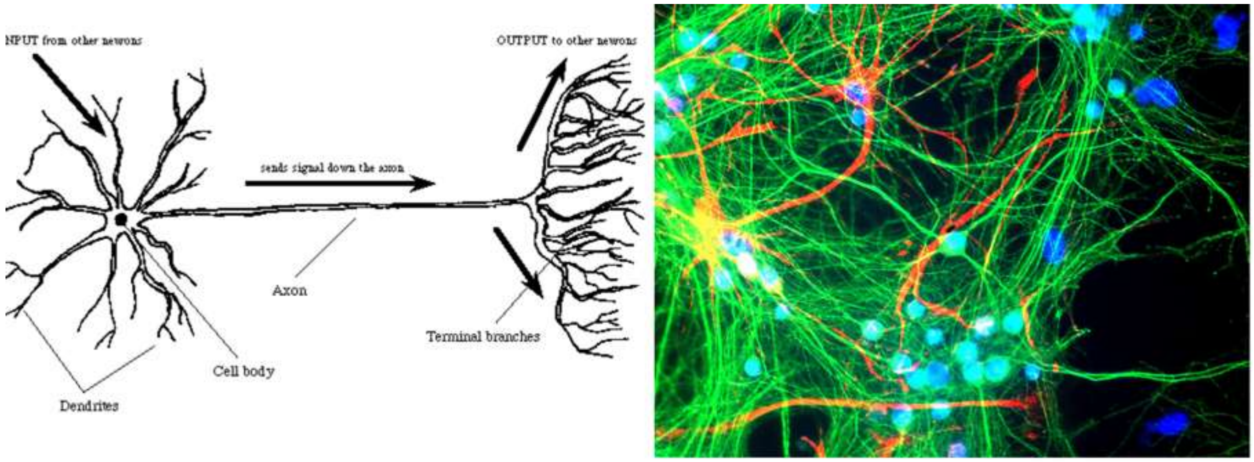

Figure 1: Artificial Neural Network (ANN) Inspired by Biological System

An ANN is a structure of interconnected units of large number of neurons, each neuron in the network is able to receive input signals to process them, and send an output signal out, it consists of a set of weighted synapses, an adder for summing the input data weighted by the respective synaptic strength, after that, an activation function for limiting the amplitude of the output of neurons is proposed.

An ANN consists of an input layer, one or more hidden layers and one output layer in its architecture, which is from input layer to hidden layer(s) and from hidden layer(s) to the output layer, each layer consists one or more neurons, every neuron in a layer is connected with all the neurons in the following layer, these connections are not equal, each connection may have a different weight. Figure 3 illustrates the architecture of an ANN.

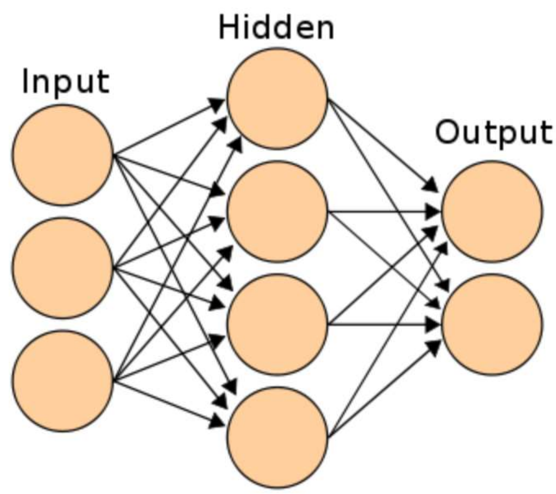

Figure 2: Architecture of an Artificial Neural Network (ANN)

From the architecture, we can see that the source nodes in the input layer supply respective elements of the activation pattern to constitute input signals applied to neurons in the hidden layer, the output signals of neurons in the hidden layer are used as an input to the output layer, there is no connection among neurons in the same layer, the set of output signals of the neurons in the output layer of the network constitute the overall response of the network to the activation pattern.

Based on the above architecture of an ANN, there are 3 input nodes in the input layer, 4 hidden nodes in the hidden layer and 2 output nodes in the output layer, assume input training features associated with training labels are paired as follows,

$D = \left\{\left(x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, y^{(1)}\right), \left(x_1^{(2)}, x_2^{(2)}, x_3^{(2)}, y^{(2)}\right), \ldots, \left(x_1^{(m)}, x_2^{(m)}, x_3^{(m)}, y^{(m)}\right)\right\} \in R^{m\times 3}$, where m is the number of instances of the training dataset, the subscript represents the index of features, the superscript represents the index of training dataset instances.

The weight matrix from input layer to hidden layer is $W_1 \in R^{3\times 4}$, the weight matrix from hidden layer to output layer is $W_2 \in R^{4\times 2}$.

The sigmoid function, as an activation function is used for limiting the amplitude of the output of neurons is proposed as follows,

$$\sigma \coloneqq f(x) = \frac{1}{1+e^{-x}}$$

Figure 4 illustrates the range of output values calculated by sigmoid function.

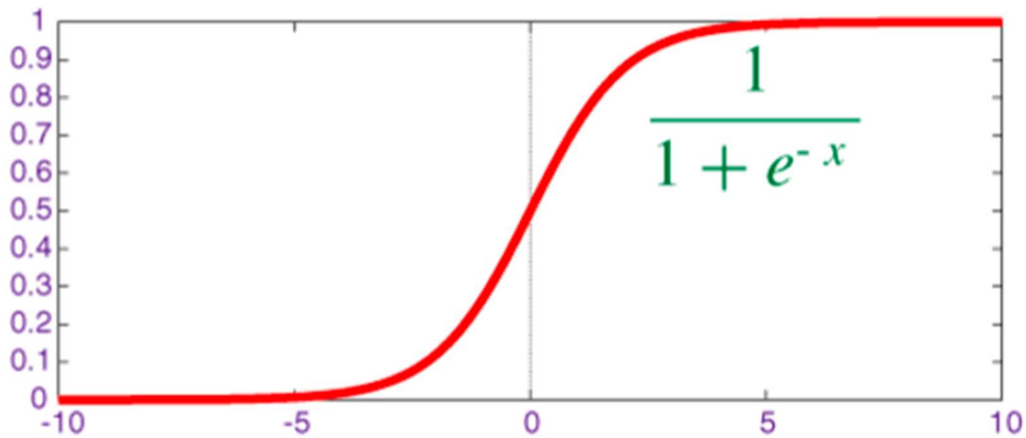

Figure 3: Sigmoid Function

Therefore, the forward computation of an ANN is shown as follows,

$$\mathrm{o} = \sigma(DW_1W_2) = \frac{1}{1+e^{-DW_1W_2}}\text{, where } \mathrm{o} \in R^{n\times 2}$$

3.2.3 Backpropagation (BP)

Backpropagation (BP) is a method used in ANN to calculate a gradient that is needed in the calculation of the weights to be used in the network [54], it is commonly used to train neural network, gradient descent optimization algorithm is used to adjust the weight of neurons by calculating the gradient of the loss function, and the error is calculated at the output and distributed back through the network layers.

We can specify an error measure that is a function of our weight vector *w* as $E(w) = \frac{1}{2}\sum_{d \in D}\left(y^{(d)} - o^{(d)}\right)^2$, this error measure defines a surface over the hypothesis (weight) space, gradient descent is an iterative process aimed to find the minimum in the error surface, on each iteration, current weights are defined as a point in the space, finding direction in which error surface descends most steeply and taking a step. Figure 5 illustrates the iteration steps of gradient descent.

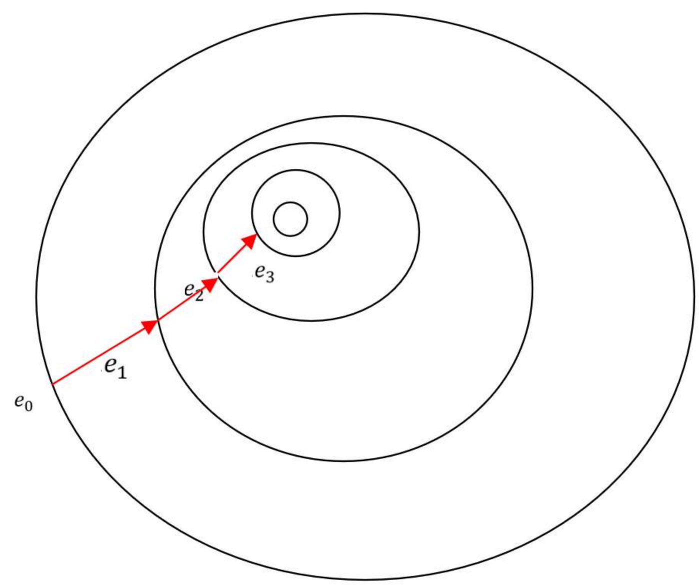

Figure 4: Iterative Steps of Gradient Descent

3.2.4 Recurrent Neural Network (RNN)

The Recurrent Neural Network (RNN) is a class of neural network whose connections of units form a directed cycle, the ability of this nature is to work with temporal data, the recurrent is defined as there exists a path of one or more cycles from a unit back to itself [55], Elman introduces multiple hidden layers, there are connections between hidden layers to hidden layers [56], which is the most common architecture of RNN, the expression is denoted as follows,

$$h^t = \sigma(W_h X_t + W_r h^{t-1})$$

$$y_t = \sigma(W_y h^t)$$

Where X_t denotes input at time of *t,* W_h denotes weight matrix from input layer to the hidden layer and W_r denotes the weight of recurrent computation, W_y denotes weight from hidden layer to output layer, h^t denotes values of hidden nodes at time of *t, y* denotes the value of output node at time of *t.* Figure 6 illustrates the unfold architecture of RNN.

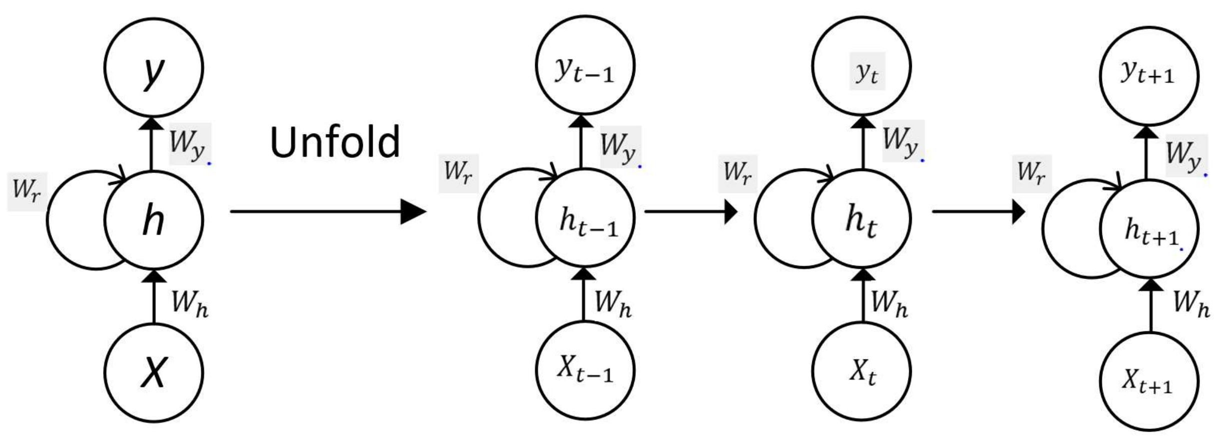

Figure 5: Unfold Architecture of RNN

3.2.5 Backpropagation Through Time (BPTT)

The recurrent architecture makes traditional backpropagation infeasible and backpropagation cannot stop, Backpropagation Through Time (BPTT) is introduced to unfold the recurrent architecture and expanded as Feed Forward Neural Network (FFNN) with certain time steps, which applies traditional Backpropagation (BP) onto unfolded neural network. BPTT will optimize hidden to hidden weight at the same time from the current time point to the previous time point [57,58,59]. In addition, BPTT is also trained online, or in a batch or mini-batch like BP. The loss function is implemented as the squared error, which is given by the following equation,

$$E_t = \sum (y_t - y)^2$$

Where E_t is the recursive error, y_t represents the predicted value at time of *t, y* denotes the real value.

3.2.6 Long Short-term Memory Recurrent Neural Network (LSTM)

The architecture of Long short-term Memory Recurrent Neural Network (LSTM) is like Artificial Neural Network (ANN), but hidden units of LSTM contains connections, which is the general architecture of Recurrent Neural Network (RNN), besides, each hidden unit is used for memorizing and forgetting goal, this research propose the architecture of two hidden layers, each layer consists 4 hidden units, which can capture temporal features of traffic speed accurately. Figure 7 illustrates the architecture of LSTM in this research.

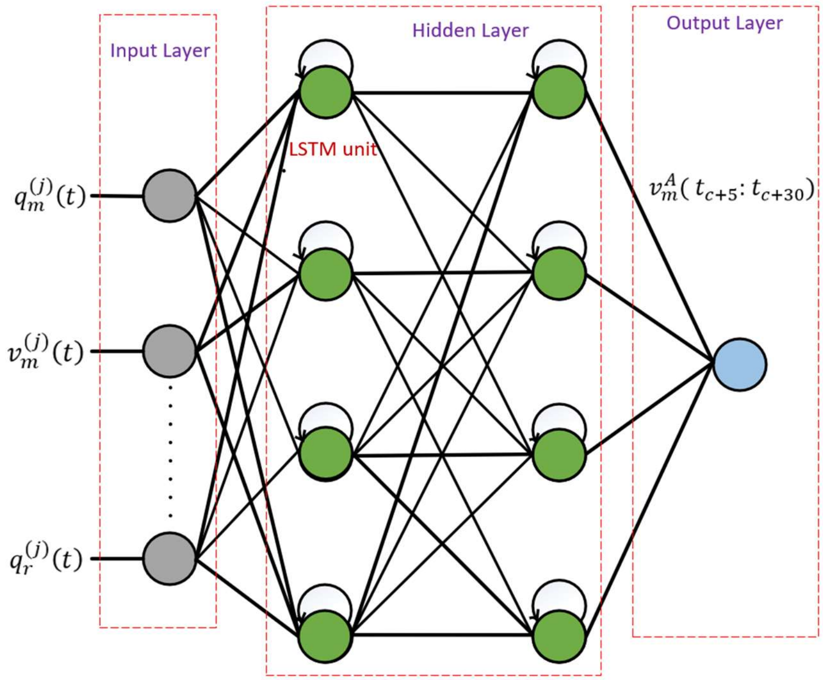

Figure 6: Developed Architecture of LSTM

For general-purpose of sequence modeling, LSTM as a special RNN structure has proven stable and powerful modeling long-range dependencies [60,61], the major innovation is its special unit called Memory Block or LSTM unit, which has self-connections storing the temporal state of the network, a LSTM unit is composed of an input gate, an output gate, a forget gate and memory cell which are used to control the flow of information, memory cell essentially acts as an accumulator of the state information, the cell is accessed, written and cleared by several self-parameterized controlling gates. Every time a new input comes, its information will be accumulated to the cell if the input gate is activated, and the past cell status will be forgotten in the process if the forget gate is on and activated. Whether the final state will be propagated to the final state controlled by output gate. Overcoming the vanishing too quickly from RNN model is a critical improvement for LSTM, the multivariate version of LSTM is where the input, cell output and states are all vectors, the key equations of LSTM unit are shown as follows, where $' \cdot '$ represents Hadamard product,

$$i_t = \sigma(W_{xi}x_t + U_{hi}h_{t-1} + V_{ci} \cdot c_{t-1} + b_i),$$

$$f_t = \sigma\left(W_{xf}x_t + U_{hf}h_{t-1} + V_{cf} \cdot c_{t-1} + b_f\right),$$

$$o_t = \sigma(W_{xo}x_t + U_{ho}h_{t-1} + V_{co} \cdot c_{t-1} + b_o)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot tanh(W_{xc}x_t + U_{hc}h_{t-1} + b_c)$$

$$h_t = o_t \cdot \tanh(c_t)$$

Figure 8 illustrates the architecture of LSTM unit is shown as follows,

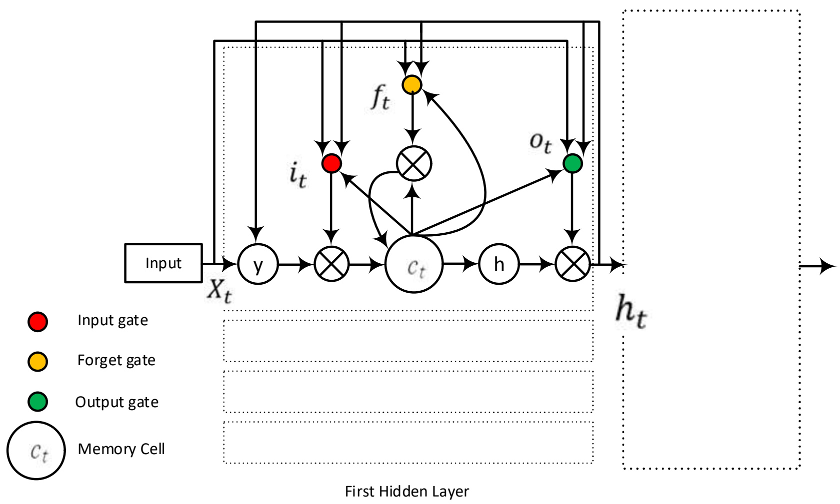

Figure 7: Architecture of LSTM unit

The three exit arrows form the memory cell to the 3 gates directly denotes the peephole connections, which denote the contributions of the activations of the memory cell at time of *t-1*, the single left-to-right arrow exiting the memory cell denotes the contribution of memory cell at time of *t*. Where $x_t \in R^d$ is the input vector to the LSTM unit at time of *t*, $i_t, f_t, o_t, c_t, h_t \in R^h$ denotes the input gate's, forget gate's, output gate's activation vector, memory cell status vector and output vector of LSTM unit at time of *t*, $W_{xi}, W_{xf}, W_{xo} \in R^{h \times d}$ denote weight matrices between input vector and input gate, forget gate and output gate, $U_{hi}, U_{hf}, U_{ho} \in R^{h \times h}$ denote weight matrices between output vector and input gate, forget gate and output gate, $V_{ci}, V_{cf}, V_{co} \in R^{h \times d}$ denote the weight matrices of peephole connection. σ is the sigmoid function.

4. DATA DESCRIPTION

Highway speed and flow data were extracted from one merging road segment of major east to west Riverside Freeway, State Route 91, both mainline and ramp have 3 lanes, the mainline detector and on-ramp detector collected the traffic data every 5 min on 280 weekdays from January 3, 2011 to March 16, 2012.

错误!未找到引用源。 shows the traffic data where the mainline detector and on-ramp detector collected, and Figure 1 shows the freeway which this research will focus on, we focus on the road segment which involves frequent traffic congestions (Bottleneck).

Table 1: Traffic Data Collected on one Merging Road Segment

Riveside Freeway Traffic Data	
Mainline	**Ramp**
Speed of Lane 1	Flow of Lane 1
Speed of Lane 2	Flow of Lane 2
Speed of Lane 3	Flow of Lane 3
Flow of Lane 1	Total Ramp Flow
Flow of Lane 2	
Flow of Lane 3	
Average mainline Speed	
Total Mainline Flow	

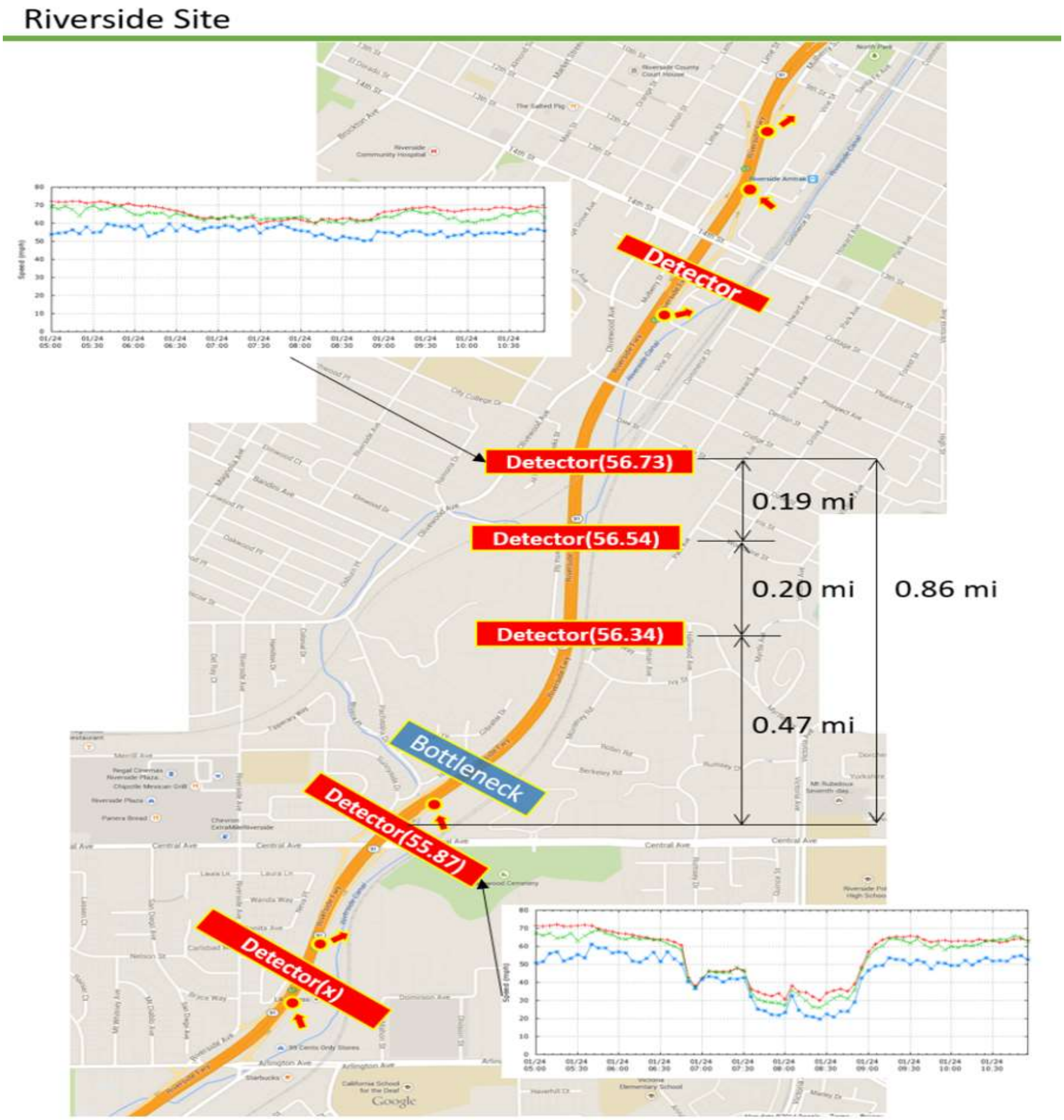

Figure 8: Marked Congestion Bottleneck at Riverside Freeway, SR 91, California

4.1 Caltrans Performance Measurement System (PeMS)

Caltrans Performance Measurement System (PeMS) traffic data is collected in real-time from over 39000 individual detectors. These sensors span the freeway system across all major metropolitan areas of the State of California.

PeMS is also an Archived Data User Service (ADUS) that provides over 10 years of data for historical analysis. It integrates a wide variety of information from Caltrans and other local agency systems includes:

- Traffic Detectors
- Census Traffic Counts
- Incidents
- Vehicle Classification
- Lane Closures
- Weight-In-Motion
- Toll Tags
- Roadway Inventory

4.2 Short-term Traffic Speed and Flow Data Applications

Without loss of generality, we define short-term traffic speed and flow data as follows:

For a given traffic speed series on mainline $\left\{v_m^{(j)}(t), t_0 \leq t \leq t_c, j = 1,2,3, A\right\}$, for a given traffic flow series on mainline $\left\{q_m^{(j)}(t), t_0 \leq t \leq t_c, j = 1,2,3, T\right\}$, for a given traffic flow series on ramp $\left\{q_r^{(j)}(t), t_0 \leq t \leq t_c, j = 1,2,3, T\right\}$, where *v* and *q* denote speed and flow, t_0, t_c are the indices of the beginning and the current time, the subscript *m, r* are the mainline and ramp variables, $j = 1, 2, 3$ represent Lane 1, Lane 2, Lane 3, $j = A$ represents average value of speed, $j = T$ represents total value of flow.

The problem is to predict the average mainline speed L steps after the current time. More specifically, our task is to predict the $v_m^A(t_c + L) := v_m^A(t_c + 5 : t_c + L)$.

4.3 Exploratory Data Analysis

The experiments of this research includes Exploratory Data Analysis (EDA) to explore the characteristics of freeway traffic data, researchers can be aware of the distribution, variance, value range of traffic data. Figure 1 illustrates the research process with EDA as follows,

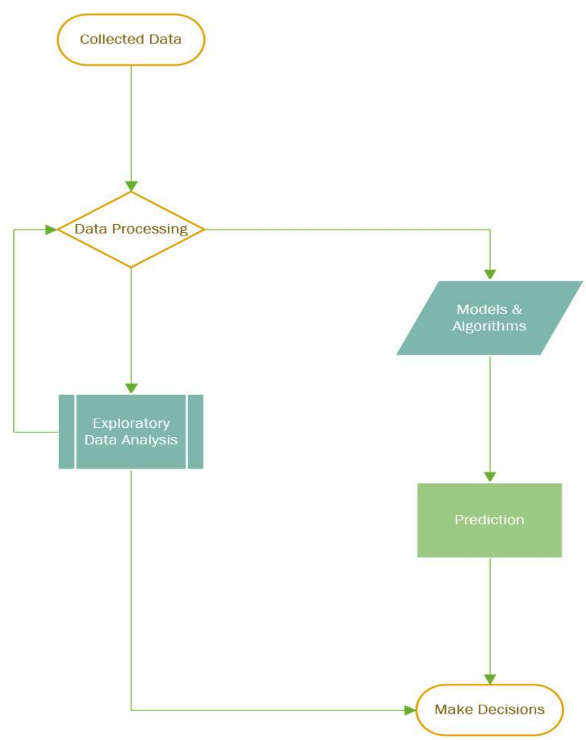

Figure 9: Short-term Traffic Speed Prediction (STTSP) Process

EDA is an approach to analyzing data sets to summarize their main characteristics, often with visual methods, which can tell the information beyond the formal modeling or hypothesis testing tasks. Based on the average value, variance and the range of the traffic data, we can get the intuitive understanding of the traffic data set. The EDA analysis is as follows:

- The range and variance of traffic data

  The range and variance of the traffic data set can help people to know the role of the importance and distribution of the traffic data set.

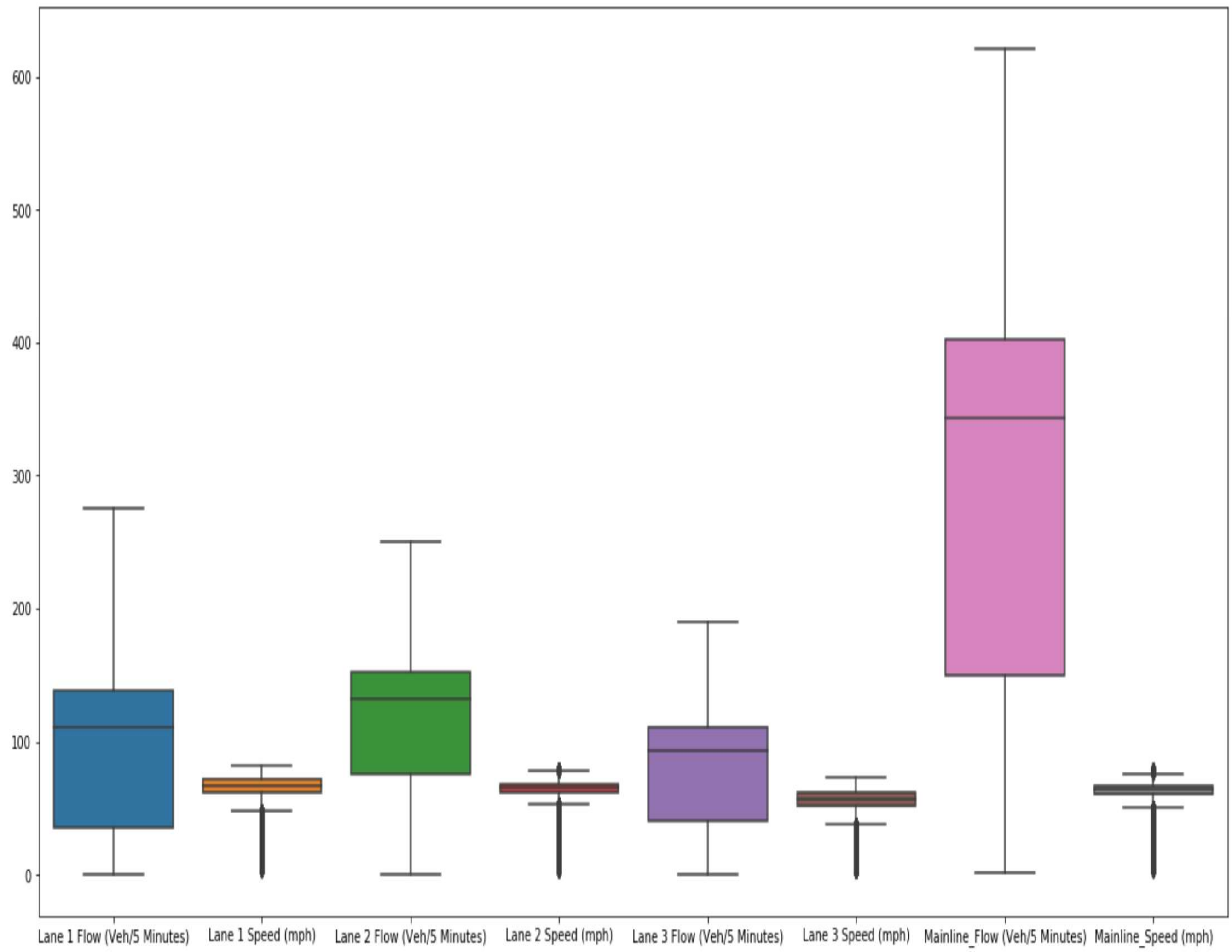

Figure 10: The Range and Variance of Mainline Data

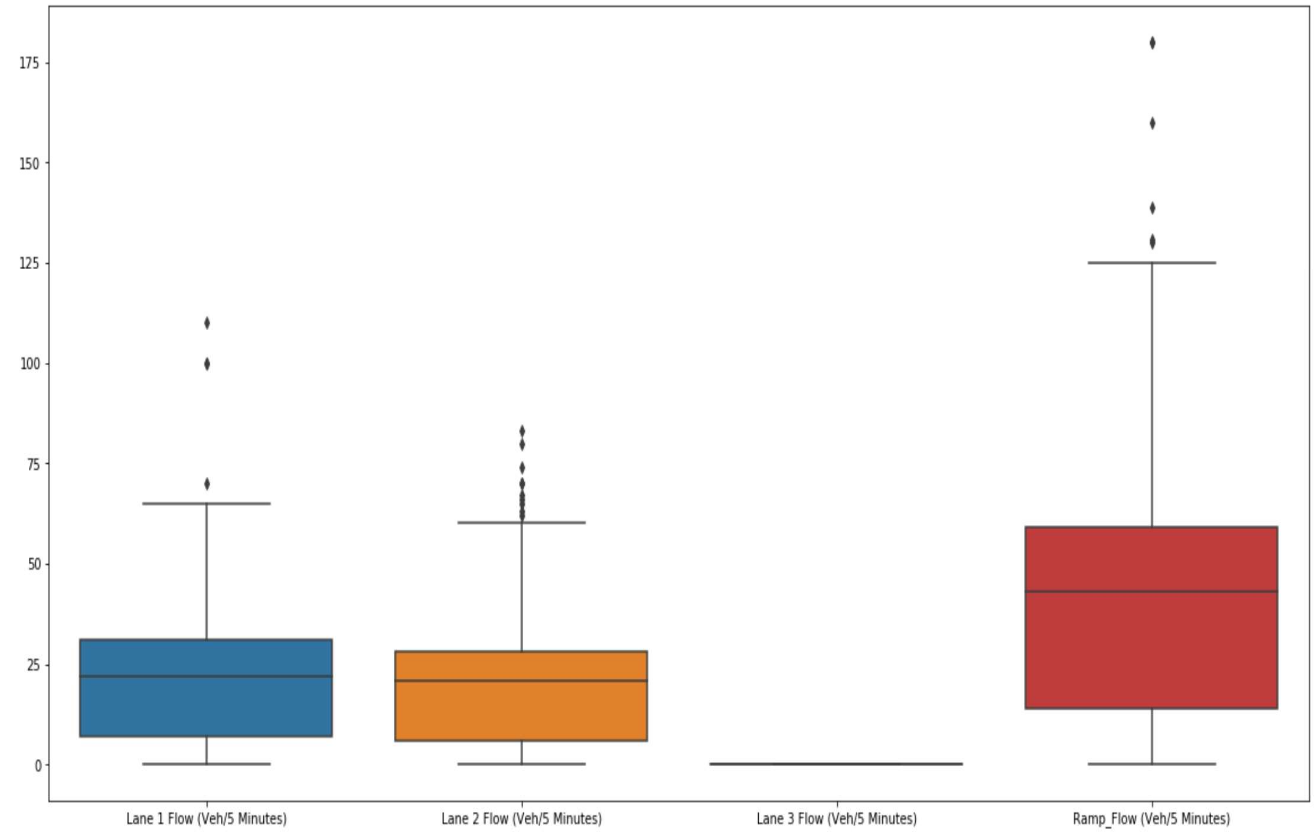

Figure 11: The Range and Variance of Ramp Data

From the Figure 2 and Figure 3, we can get to know the range and variance of flow is much larger than that of speed, the flow of 3rd lane of ramp is always 0, the characteristics of speed of each lane of mainline is similar, the flow of each lane of mainline is similar as well.

- Scatterplot Analysis

A scatter plot is type of mathematical diagram using Cartesian coordinates to display the values for typically each two traffic variables for the data set, the points are color-coded, and the data set is displayed as a collection of points, the average mainline speed is determined by the other traffic variable on the horizontal axis. Assume each of the variables in the data set is independent for average mainline speed, Figure 4 illustrates the relationship between speed of each lane of mainline and average mainline speed, Figure 5 illustrates the relationship between flow of each lane of mainline, total flow of mainline and average mainline speed, Figure 6 illustrates the relationship between flow of each lane of ramp, total flow of ramp and average mainline speed.

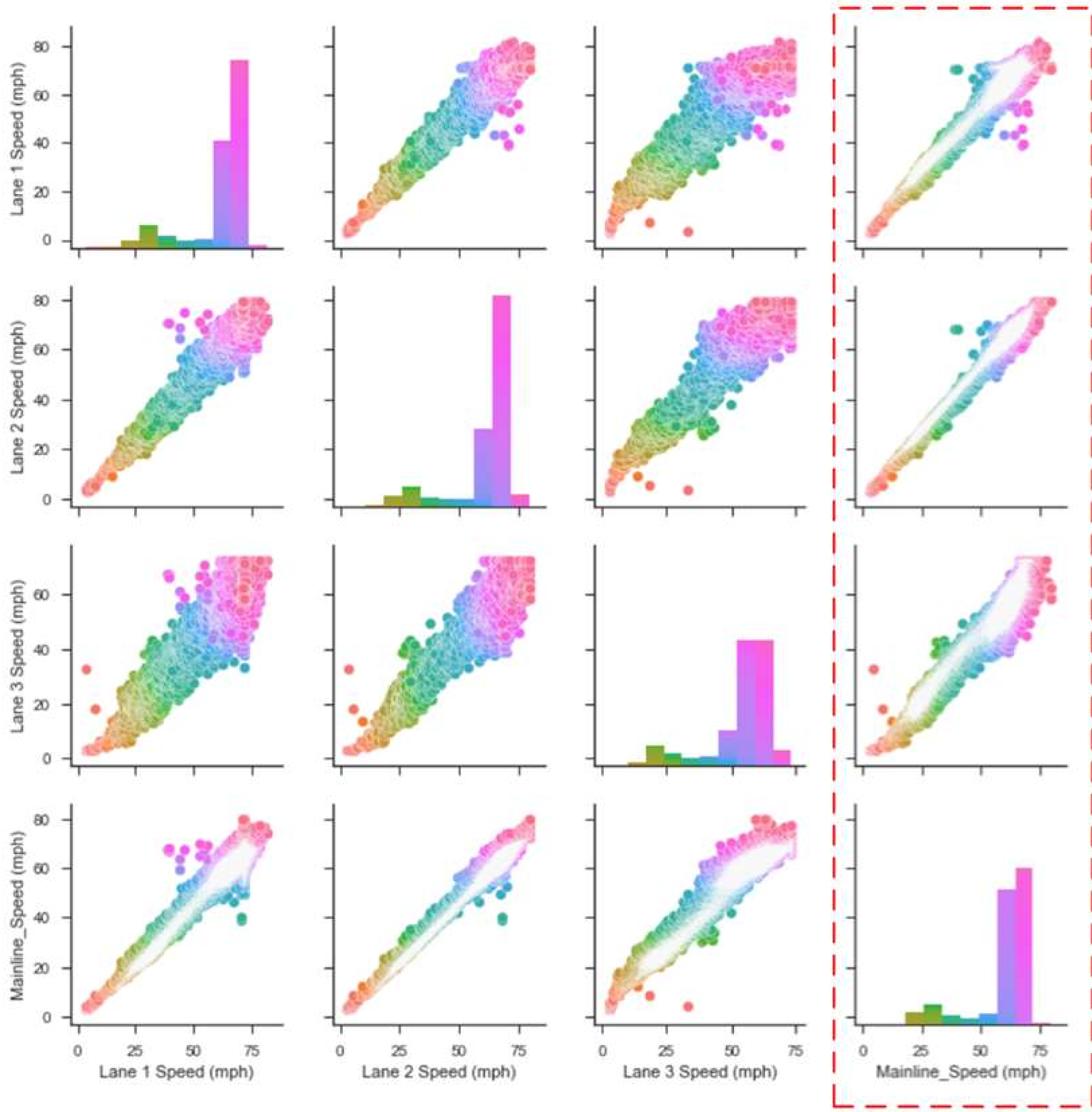

Figure 12: Scatterplot of Speed of Each Lane of Mainline

From the red dotted box which represents speed of each lane of mainline determine the horizontal axis of average mainline speed, we can find that the correlations between speed of each lane of mainline and average mainline speed are similar, and the relations are linear.

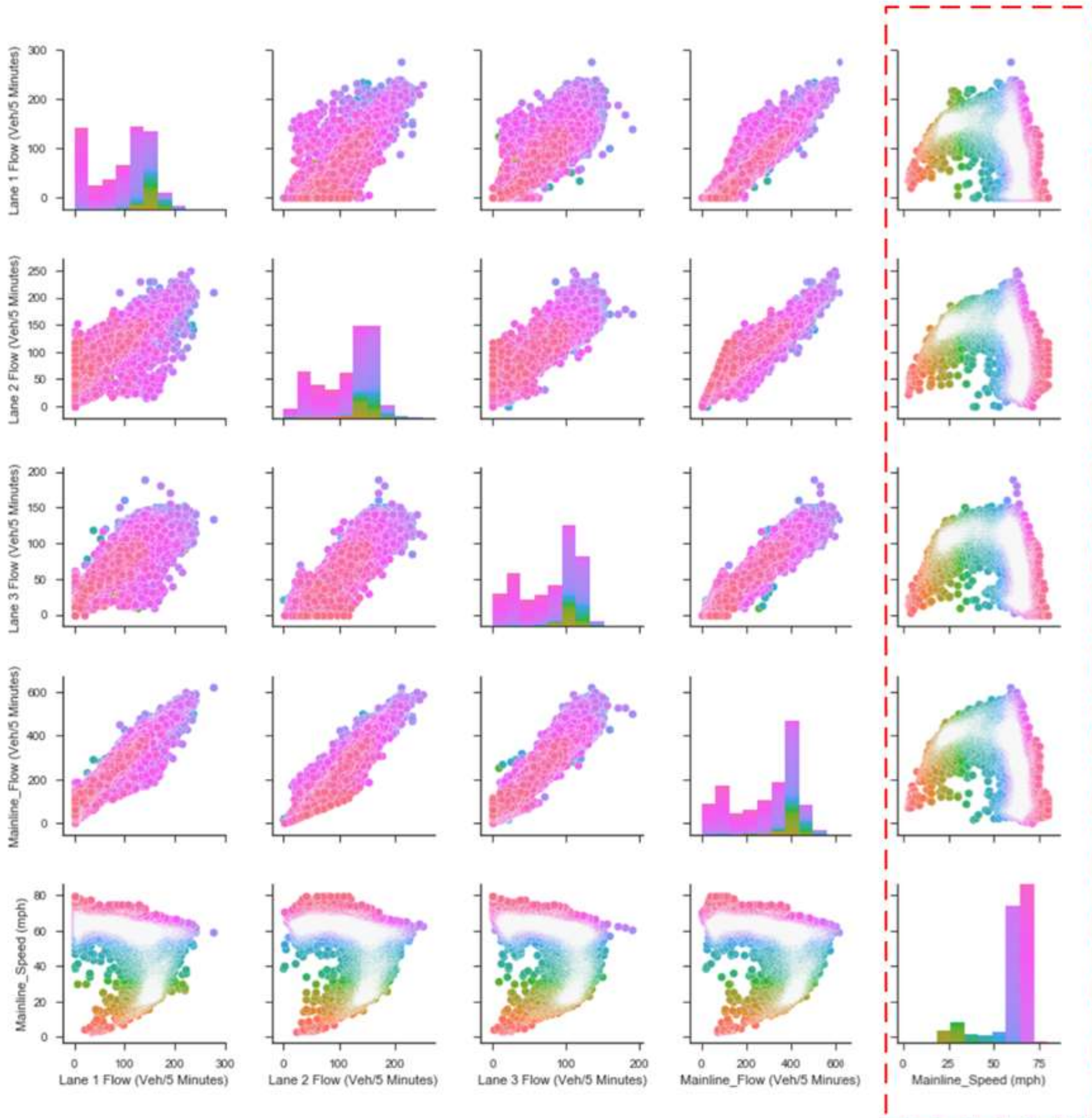

Figure 13: Scatterplot of Flow of Each Lane of Mainline

and Total Flow of Mainline

From the red dotted box which represents flow of each lane of mainline and total flow of mainline determine the average mainline speed, we can find that the correlations between flow of each lane of mainline, the total flow of mainline and average mainline speed are similar, and the relations are nonlinear.

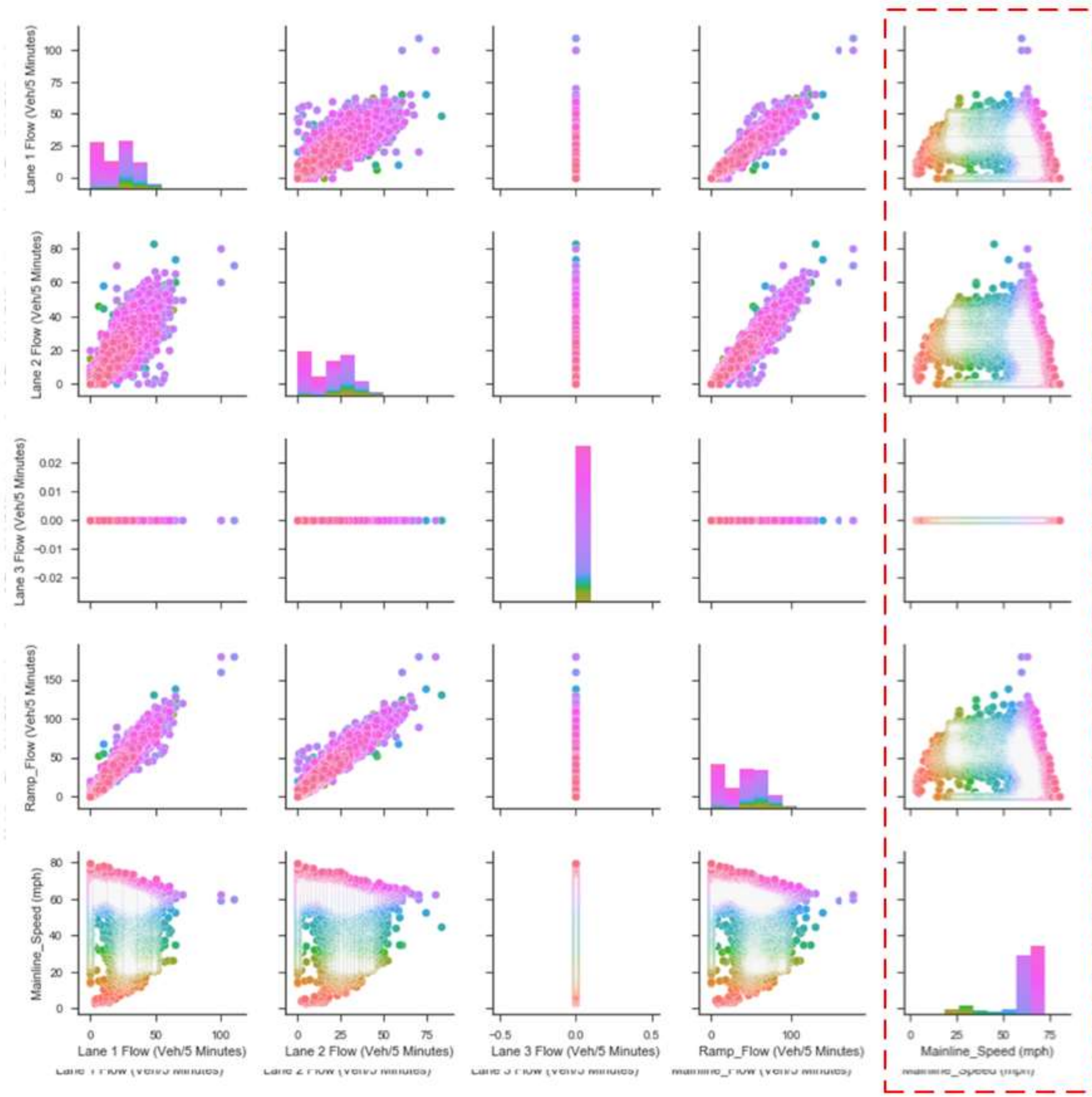

Figure 14: Scatterplot of Flow of Each Lane of Ramp

Total Flow of Ramp

From the red dotted box which represent flow of each lane of ramp and total flow of ramp determine the average mainline speed, we can find that the correlations between flow of each lane of ramp, total flow of ramp and average mainline speed are similar, and the relations are nonlinear.

5. EXPERIMENT AND RESULT ANALYSIS

5.1 Evaluation Measurements

To rigorously evaluate the performance of LSTM, root-mean-square error (RMSE) and Mean absolute percentage error (MAPE) are proposed to evaluate the performance of short-term average mainline traffic speed prediction. Based on the 5-fold cross validation, 5 pairs of training and test datasets are proposed, the average values of 5 pairs RMSE and MAPE are shown in the section.

RMSE and MAPE are frequently used measure of the differences between values predicted by a mode and the values actually observed, the formula is as follows,

$$RMSE = \frac{\sqrt{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{n}$$

$$MAPE = \frac{100}{n} \times \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right|$$

Where $\hat{y}_i$ denotes the predicted value of average mainline traffic speed, y_i denotes the real value of the average mainline traffic speed, n is the number of instances in test dataset. The results of the 4 situations are shown as follows,

a. Average mainline speed

Table 2: RMSE and MAPE of Average Mainline Speed Prediction

with Average Mainline Speed Feature

	Decision Tree		Random Forest		Linear Regression		Bagging		Adaboosting		LSTM	
	RMSE	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**
5min	2.86	1.71	2.87	1.72	2.79	1.68	2.86	1.71	3.88	2.64	2.81	1.56
10min	4.51	2.56	4.51	2.57	4.41	2.50	4.51	2.56	5.28	3.83	4.41	2.57
15min	5.75	3.25	3.25	3.25	5.65	3.16	5.75	3.25	6.48	4.70	5.65	3.13
20min	6.72	3.82	6.73	3.82	6.62	3.70	6.72	3.82	7.58	5.65	6.72	4.26
25min	7.49	4.32	7.50	4.32	7.42	4.20	7.49	4.32	8.45	6.46	7.4	4.31
30min	8.13	4.76	8.14	4.76	8.09	4.64	8.13	4.76	8.92	6.85	8.07	4.92

From the RMSE and MAPE of average mainline speed prediction with average mainline speed feature, the performance of random forest for 15 min average mainline speed prediction is significantly good, AdaBoosting shows the worst performance for 15 min average mainline speed prediction, the predicted performance of linear regression and LSTM are the best for over the 5 to 30 min average mainline speed prediction, besides, LSTM shows the best performance for 30 min average mainline speed prediction.

b. Average mainline speed and total mainline flow

Table 3: RMSE and MAPE of Average Mainline Speed Prediction

with Average Mainline Speed and Total Mainline Flow

	Decision Tree		Random Forest		Linear Regression		Bagging		Adaboosting		LSTM	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
5min	3.84	2.24	3.18	1.94	2.78	1.69	3.84	2.24	2.72	2.56	3.08	2.01
10min	6.00	3.33	4.98	2.88	4.37	2.53	6.00	3.33	4.27	3.79	4.29	2.44
15min	7.52	4.12	6.33	3.62	5.58	3.22	7.52	4.12	5.42	4.82	5,54	3.16
20min	8.64	4.74	7.29	4.16	6.50	3.78	8.64	4.74	6.31	5.47	6.35	3.62
25min	9.57	5.25	8.10	4.64	7.25	4.28	9.57	5.25	6.99	6.10	7.21	4.77
30min	10.24	5.70	8.72	5.05	7.87	4.74	10.24	5.70	7.56	6.64	7.68	4.64

Based on the RMSE and MAPE of average mainline speed prediction with average mainline speed and total mainline flow features, the predicted performance of Decision Tree, Random Forest and Bagging get worse, the predicted performance of Linear Regression, Adaboosting and LSTM has been improved except 5 min average mainline speed with LSTM.

c. Speed and flow of each lane of mainline

Table 4: RMSE and MAPE of Average Mainline Speed Prediction

with Speed and Flow of Each Lane of Mainline

	Decision Tree		Random Forest		Linear Regression		Bagging		Adaboosting		LSTM	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
5min	3.96	2.38	3.06	1.88	2.78	1.70	3.96	2.38	3.66	2.52	2.91	1.85
10min	6.15	3.49	4.72	2.76	4.38	2.54	6.15	3.49	5.15	3.73	4.31	2.48
15min	7.69	4.30	5.95	3.45	5.58	3.23	7.69	4.30	6.32	4.54	5.61	3.23
20min	8.73	4.87	6.86	3.93	6.50	3.79	8.73	4.87	7.48	5.50	6.37	3.65
25min	9.62	5.62	7.59	4.41	7.25	4.29	9.62	5.38	8.21	6.11	7.17	4.78
30min	10.37	5.82	8.16	4.79	7.86	4.73	10.37	5.82	8.74	6.52	7.64	4.61

Based on Table 2, Table 4 illustrates the predicted performance of Decision Tree, Random Forest and Bagging become worst, the predicted performance of Linear Regression, Adaboosting and LSTM has been improved, besides, the predicted performance of LSTM for 20 to 30 min average mainline speed is better than Linear Regression.

d. Speed and flow of each lane of mainline and ramp

Table 5: RMSE and MAPE of Average Mainline Speed Prediction

with Speed and Flow of Each Lane of Mainline and Ramp

	Decision Tree		Random Forest		Linear Regression		Bagging		Adaboosting		LSTM	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
5min	3.85	2.33	2.88	1.77	2.75	1.67	3.85	2.33	3.69	2.68	2.8	1.75
10min	6.10	3.47	4.52	2.64	4.32	2.52	6.10	3.47	5.33	3.89	4.24	2.43
15min	7.72	4.29	5.71	3.30	5.51	3.21	7.72	4.29	6.46	4.66	5.48	3.14
20min	8.90	4.92	6.64	3.84	6.41	3.77	8.90	4.92	7.38	5.38	6.29	3.56
25min	9.84	5.46	7.31	4.27	7.13	4.27	9.84	5.46	8.17	6.01	7.1	4.74
30min	10.50	5.84	7.92	4.66	7.71	4.71	10.50	5.84	8.57	6.30	7.5	4.56

Based on Table 2, Table 3, Table 4 and Table 5, the overall performance of prediction with speed and flow of each lane of mainline and ramp features is the best, and LSTM achieves the best performance compared with the other machine learning models. In addition, LSTM performs the best robustness of long time prediction.

5.2 Traffic Breakdown Prediction

Based on Section 5.1, short-term traffic speed predictions (STTSP) with speed and flow of each lane of mainline and ramp achieves the best results, in order to evaluate the prediction performance, 30-min average mainline speed prediction is used for consideration. Based on RMSE and MAPE results of 7.5 and 4.56, the overall performance of STTSP is satisfiable, however, most part of the average mainline traffic speed is at the situation of free flow, the STTSP for free flow is not difficult, the STTSP for the congestion area plays a significant role for evaluating the research. The status of average mainline traffic speed is defined as free flow, breakdown, congestion and recovery. The status of free flow represents the fluctuations of average mainline traffic speed between 55 to 75 mph, the status of congestion is defined as the average mainline traffic speed is around or less than 35 mph and last for some time, breakdown is the transition process from free flow to congestion, the average mainline traffic speed decreases quickly within short time and recovery is the transition process from congestion to free flow. The evaluation focus on detecting the breakdown starting point of the first congestion of a day, based on the predicted breakdown starting point and real breakdown starting point of a day, people can get to know when the traffic congestion will happen.

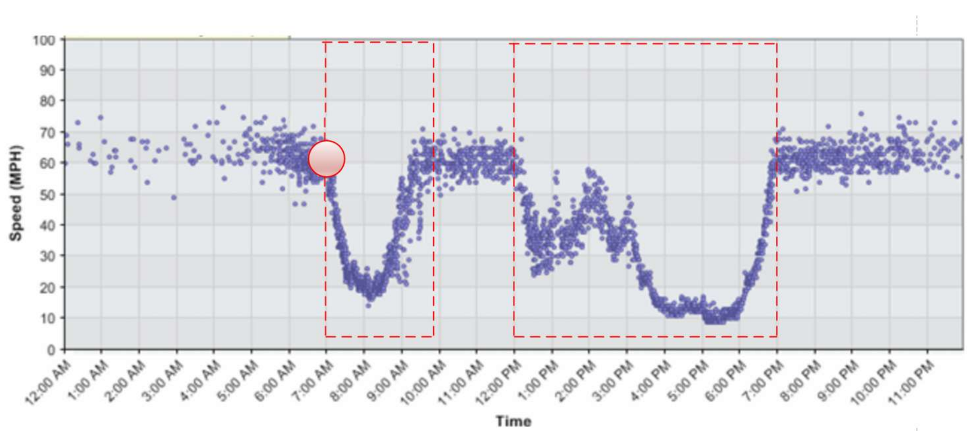

Figure 15: Average Mainline Speed of a day

Figure 14 illustrates the average mainline traffic speed of freeway of a day, the 2 red dotted windows shows the traffic congestions duration of the day, this research focus on the breakdown starting point of the first traffic congestion, the read circle on the first red dotted window illustrates the evaluation area of this research.

There are two typical cases of STTSP for average freeway mainline speed, one is the predicted breakdown starting point shifted right or left to the real breakdown starting point of the first congestion of a day, the other is the predicted breakdown starting point matches the real breakdown starting point of a day. The capabilities Continuous Wavelet Transform (CWT) can be used for analyzing important features related to bottleneck activations and traffic oscillations in congested traffic in a system manner [62], thus CWT is proposed to detect breakdown starting point of predicted and real average mainline speed. Figure 16 illustrates first case of STTSP for average mainline speed, Figure 17 illustrates the second case of STTSP for average freeway mainline speed.

Figure 16: Predicted Average Mainline Speed of the First Case

Figure 16 illustrates the first case of STTSP for average freeway mainline speed, the two red dotted line illustrates the breakdown starting points of predicted and real average mainline speed, we can find the time difference of breakdown starting points for the 2 diagram can be around 25 to 30 min.

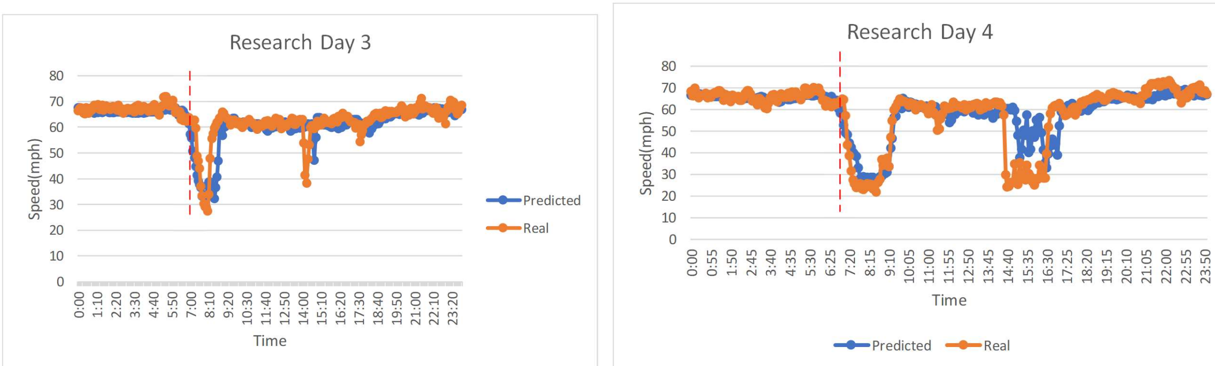

Figure 17: Predicted Average Mainline speed of the Second Case

Figure 17 illustrates the second case of STTSP for average mainline speed, the red dotted line illustrates the breakdown starting points of predicted and real average mainline speed, which illustrates the breakdown starting points of predicted and real average freeway mainline speed match exactly, with the very well breakdown starting point, travelers can plan traveling route and time accurately.

59 days of 30-min predicted and real average mainline speed data set is generated, 55 days of that contain congestions which are proposed for evaluation, the time difference of breakdown starting points for the 55 days is recorded, Table 6 and Figure 18 illustrate the number of days for breakdown starting time difference of predicted and real average mainline speed.

Table 6: Number of Days for Breakdown Starting Time Difference

Time Difference	0 min	5 min	10 min	15 min	20 min	25 min	30 min	35 min
Number of Days	30	14	2	3	1	2	2	1
Proportion	0.54545	0.2545	0.0364	0.05455	0.01818	0.03636	0.03636	0.0182

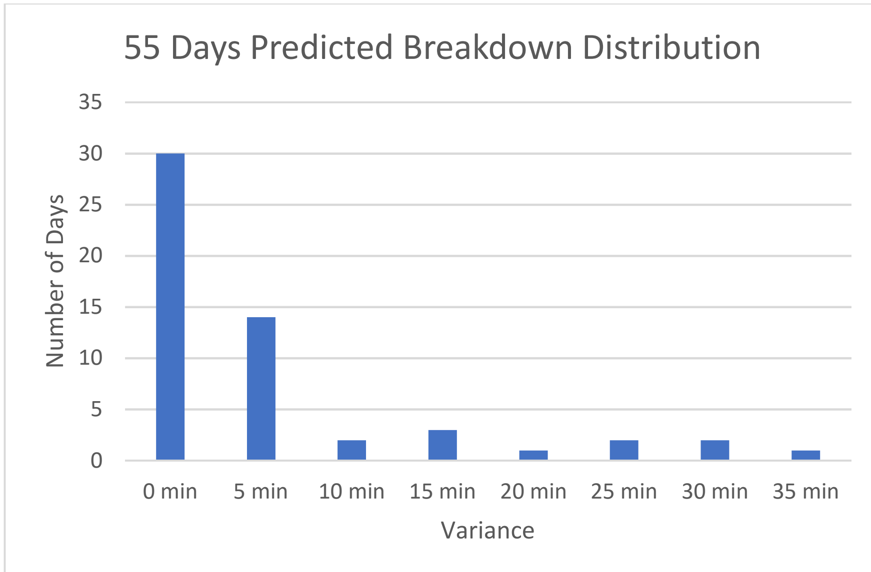

Figure 18: 55 Days Predicted Breakdown Distribution

From the results of breakdown prediction above, we can find that there are 30 days which breakdown starting points of predicted and real average mainline speed match exactly, 14 days have 5 min time difference of breakdown starting points of predicted and real average mainline speed.

6. CONCLUSION

The main contribution of this thesis is the development of an architecture of Long Short-term Memory of Recurrent Neural Network (LSTM) of Short-term traffic speed prediction (STTSP) for freeway mainline and traffic congestion breakdown predictions associated with reliable traffic data processing. Traffic speed prediction is difficult mainly on whether the model or methodology can capture the continuity or discontinuity of the speed, and whether the breakdown of predicted speed matches that of the real speed. From the experiments, the overall accuracy of speed prediction is outstanding due to the small value of RMSE and MAPE, but due to the characteristics of traffic speed of a day, if there is no congestion, the traffic speed will fluctuate around a certain value, which is around 60 to 65 mph, but when the congestion will happen, the traffic speed will decrease a lot within a very short time, the detection of the breakdown is difficult to control, because after generating the predicted speed, then we can compare the predicted and real traffic speed then, in this case, the current results of current experiments is satisfiable. Based on the current work, I conclude that the data processing step is significant to generate the results, because the past 90 min traffic data was involved into one data instance to predict the future 30 min traffic speed, the model can learn more information from the past, and the model can help to predict more reliable traffic speed. In addition, LSTM is good for temporal data prediction, the special recurrent connections between hidden layer to hidden layer is essential for model to remove the unnecessary information. The current architecture of LSTM has two hidden layers, each of the hidden layer has four hidden units, based on the literature review, more complex architecture of LSTM can involve more computation, and drop out operation can improve the ability of computation.

6.1 Future Research/Work

The current research can help predict average freeway mainline speed accurately, in order to predict the breakdown starting points of the first congestion more accurately, two methodologies can be considered for further research. The current work involves the past 90 min traffic data for the future 5 to 30 min average freeway mainline speed prediction, in order to improve the ability of prediction, the traffic data of more than past 90 min will be involved to generate data instance, in addition, the traffic data of upstream and downstream freeway segments will be considered and involved into prediction for the research.

7. REFERENCE

[1] Adeli, H., 2001. Neural networks in civil engineering: 1989-2000. Computer-Aided Civil and Infrastructure Engineering 16, 126-142.

[2] Vlahogianni, E.I., Golias, J.C., Karlaftis, M.G., 2004. Short-term traffic forecasing: overview of objectives and methods. Transportation Reviews 24 (5), 533-557.

[3] Van Lint, J.W.C., Van Hinsbergen, C.P.I.J., 2012. Short term traffic and travel time prediction models, in artificial intelligence applications to critical transportation issues. In: Chowdhury, R., Sadek, S. (Eds.), Transportation Research Circular. National Academies Press, Washington DC, Number E-C168, November.

[4] Chen, G., Govindan, K. & Yang, Z. Z. (2012), Managing truck arrivals with time windows to alleviate gate congestion at container terminals, *International Journal of Production Economics,* **141**(1), 179-88.

[5] Chen, G., Govindan, K. & Golias, M. M. (2013), Reducing truck emissions at container terminals in a low carbon economy: proposal of a queueing-based bi-objective model for optimizing truck arrival pattern, *Transportation Research Part E: Logistics and Transportation Review,* **55,** 3-22.

[6] Chen, G. & Yang, Z. Z. (2010), Optimizing time windows for managing arrivals of export container in Chinese Container terminals, *Maritime Economics & Logistics,* **12**(1), 111-26.

[7] Cacciabue, C. (2007) *Modelling Driver Behavior in Automotive Environments: Critical Issues in Driver Interactions with Intelligent Transport Systems.* London: London: Springer.

[8] European Commission (ed.) (2001) *Commission of European Communities.* WHITE PAPER - European Transport Policy for 2010: Time.to Decide. Held 12/09/2001 at Brussels. Brussels: European Commission

[9] Vlahogianni, E,I., 2009. Enhancing predictions in signalized arterials with information on short-term traffic flow dynamics. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 13(2), 73-84.

[10] Kamarianakis, Y., Shen, W., 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.

[11] van Lint, J.W.C., Van Zuylen, H.J., 2005. Monitoring and predicting freeway travel time reliability. Transportation Research Record 1917, 54-62.

[12] Castro-Neto, M., Jeong, Y.-S., Jeong, M.-K., Han, L.,D., 2009. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. Expert Systems with Applications 36(3 PART 2), 6164-6173.

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

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

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

[16] Chen, C., Wang. Y., Li, L. Hu, J., Zhang, Z., 2012. The retrieval of intra-day trend and its influence on traffic prediction. Transportation Research Part C: Emerging Technologies 22, 103-118.
[17] Haworth, J., Cheng, T., 2012. Non-parametric regression for space-time forecasting under missing data. Computers. Environment and Urban Systems 36 (6), 538-550.
[18] Sun, S., Huang, R. , Gao, Y., 2012. Network-scale traffic modeling and forecasting with graphical lasso and neural networks. Journal of Transportation Engineering 138 (11), 1358-1367.
[19] Innamaa, S., 2009. Self-adapting traffic flow status forecasts using clustering. IET Intelligent Transport Systems 3(1), 7-76..
[20] Oh, C., Park, S., 2011. Investigating the effects of daily travel time patterns on short-term prediction. KSCE Journal of Civil Engineering 15 (7), 1263-1272.
[21] Li, C.-S., Chen, M.-C, 2013. Identifying important variables for predicting travel time, of freeway with non-recurrent congestion with neural networks. Neural Computing & Applications 23 (6), 1611-1629.
[22] Jiang, X., Adeli, H., 2004. Wavelet packet-autocorrelation function method for traffic flow pattern analysis. Computer-Aided Civil and Infrastructure Engineering 19)5_, 324-337.
[23] Boto-Giralda, D., D., Diaz-Pernas, F.J., Gonzalez-Ortega, D., Diez-Higurera, J.F., Anton-Rodriguez, M., Martinez-Zarzuela, M., Torre-Diez, 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.
[24] Stathopoulos, A., Karlaftis, M.G., Dimitriou, L., 2010. Fuzzy rule-based system approach to combining traffic count forecasts. Journal Transportation Research Record: Journal of the Transportation Research Board 2183, 120-128.
[25] Cheng, T., Haworth, J., Wang, J., 2012. Spatio-temporal autocorrelation of road network data. Journal of Geographical System 14, 389-413.
[26] Dunne. S., Ghosh,. B., 2012. Regime-based short-term multivariate traffic condition forecasting algorithm. Journal of Transportation Engineering 138 (4), 455-466.
[27] Xia, J., Huang, W., Guo. I., 2012. A clustering approach to online freeway traffic state identification using ITS data. KSCE Journal of Civil Engineering 16 (3), 426-432.
[28] Sun, S., Huang, R. Gao, Y., 2012. Network-scale traffic modeling and forecasting with graphical lasso and neural networks. Journal of Transportation Engineering 138 (11), 1358-1367.
[29] Miles, J.C., Walker, A.J., 2006. The potential application of artificial intelligence in transport. IEEE Proceedings = Intelligent Transport Systems 153 (3), 183.
[30] Sadek, A.W., (Ed.), 2007. Artificial Intelligence in Transportation: Information for Application. Transportation Research Board Circular (E-C113), TRB, National Research Council, Washington, D.C.
[31] Van Lint, J.W.C., Van Hinsbergen, C.P.I.J., 2012. Short term traffic and travel time prediction models, in artificial intelligence applications to critical transportation Issues. In: Chowdhury, R., Sadek, S. (Eds.), Transportation Research Circular. National Academies Press, Washington DC, Number E-C168, November.
[32] Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2014. Short-term traffic forecasting: where we are and where we’re going. Transport. Res. Part C: Emerg. Technol. 43, 3-19.
[33] Karlaftis, M.G., Vlahogianni, E.I., 2011. Statistics versus neural networks in

transportation research: differences, similarities and some insights. Transportation Research Part C Emerging Technologies 19 (3), 387–399.
[34] Zheng, Z., Su, D., 2014. Short-term traffic volume forecasting: a k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm. Transport. Res. Part C: Emerg. Technol, 43(Part 1), 143-157..
[35] Crosby, H., Davis, P., Jarvis, S.A., 2016. Spatially-intensive decision tree prediction of traffic flow across the entire UK road network. Distributed Simulation and Real Time Applications (DS-RT), 2016 IEEE/ACM 20th International Symposium On.
[36] Hamner, B., 2010. Predicting travel times with context-dependent random forest by Modelling local and aggregate traffic flow. Data Mining Workshops (ICDMW), 2010 IEEE International Conference on.
[37] Smith, B.L., Demetsky, M.J., 1994. Short-term traffic flow prediction models-a comparison of neural network and nonparametric regression approaches. IEEE International Conference on. 1707-1709..
[38] Yin, H., Wong, S.C., Xu, J., Wong, C.K., 2002. Urban traffic flow prediction using a fuzzy-neural approach. Transport. Res. Part C: Emerg. Technol. Vol 10, Issue 2, 85-98.
[39] Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C., 2005. Optimized and meta-optimized Neural networks for short-term traffic flow prediction: a genetic approach. Transport. Res. Part C: Emerg. Technol. Vol 13, Issue 3, 211-234.
[40] Zheng, W., Lee, D., ASCE, M., Shi, Q., 2006. Short-term freeway traffic flow prediction: Bayesian combined neural network approach. Journal of Transportation Engineering. Vol. 132, Issue 2.
[41] Manoel, C.N., Jeong, Y.S., Jeong, M.K., Han, L.D., 2009. Online-SVR for short-term Traffic flow prediction under typical and atypical traffic conditions. Transport. Res. Expert Systems with Applications. Vol 36, Issue 3, Part 2, 6164-6173.
[42] Tan, M.C., Yong, S.C., Xu, J.M., Guan, Z.R., Zhang, P., 2009. An aggregation approach to short-term traffic flow prediction. IEEE Intelligent Transportation Systems Society, Vol 10, Issue 1, 60-69.
[43] Polson, N.G., Sokolov, V.O., 2017. Deep learning for short-term traffic flow prediction. Transport. Res. Part C: Emerg. Technol. Vol 79, 1-17.
[44] Anacleto, O., Queen, C., Albers, C.J., 2013. Multivariate forecasting of road traffic flows in the presence of heteroscedasticity and measurement errors. J. R. Statist. Soc.: Ser. C (Appl. Statist.) 62 (2), 251–270.
[45] Blandin, S., Couque, A., Bayen, A., Work, D., 2012. On sequential data assimilation for scalar macroscopic traffic flow models. Phys. D: Nonlin. Phenom. 241(17), 1421–1440.
[46] Chiou, Y.-C., Lan, L.W., Tseng, C.-M., 2014. A novel method to predict traffic features based on rolling self-structured traffic patterns. J. Intell. Transport. Syst.18 (4), 352–366.
[47] Work, D.B., Blandin, S., Tossavainen, O.-P., Piccoli, B., Bayen, A.M., 2010. A traffic model for velocity data assimilation. Appl. Math. Res. eXp. 2010 (1), 1–35.
[48] "Cross-Validation (Statistics)." *Wikipedia*, Wikimedia Foundation, 20 Mar. 2018, en.wikipedia.org/wiki/Cross-validation_(statistics).
[49] "Decision Tree." *Wikipedia*, Wikimedia Foundation, 14 Mar. 2018,

en.wikipedia.org/wiki/Decision_tree.
[50] "Random Forest." *Wikipedia,* Wikimedia Foundation, 14 Mar. 2018, en.wikipedia.org/wiki/Random_forest.
[51] "Linear Regression." *Wikipedia,* Wikimedia Foundation, 14 Mar. 2018, en.wikipedia.org/wiki/Linear_regression.
[52] "Booststrap aggregating." *Wikipedia,* Wikimedia Foundation, 14 Mar. 2018, en.wikipedia.org/wiki/Bootstrap_aggregating.
[53] "Boosting (machine learning)." *Wikipedia,* Wikimedia Foundation, 14 Mar. 2018, en.wikipedia.org/wiki/Boosting_(machine_learning).
[54] Deep Learning; Ian Goodfellow, Yoshua Bengio, Aaaron Courville; MIT Press; 2016, p 196.
[55] Michael I Jordan. Serial order: A parallel distributed processing approach. *Advances in Psychology,* 121:471-495, 1986.
[56] Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14:179–211.
[57] AJ Robinson and Frank Fallside. *The utility driven dynamic error propagation network.* University of Cambridge Department of Engineering, 1987.
[58] Paul J Werbos. Generalization of backpropagation with application to a recurrent gas Market model. *Neural Networks,* 1(4):339-356, 1988.
[59] Michael C Mozer. A focused back-propagation algorithm for temporal pattern recognition. *Complex systems,* 3(4):349=381, 1989.
[60] A. Graves. Generating sequences with recurrent neural networks. *arXiv preprint arXiv: 1308.0850, 2013.*
[61] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation,* 9(8):1 735-1780, 1997.
[62] Zheng et al, 2011. Applications of wavelet transform for analysis of freeway traffic: Bottlenecks, transient traffic, and traffic oscillations. Transport. Res. Part B, 45 (2011) 372-384.