UNIVERSITY OF CALIFORNIA,
IRVINE

Investigation of A New Method for Improving Individual Vehicle Speeds Estimation with Advanced Loop Data

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Civil Engineering

by

Yiqiao Li

Thesis Committee:
Professor Stephen G. Ritchie, Chair
Professor Will Recker
Professor R. Jayakrishnan

2017

DEDICATION

To my mother and my friends,

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Firstly, I would like to thank Professor Ritchie who accepts me as a MS/PhD student and gives me suggestions to improve this thesis.

Secondly, I would also acknowledge Dr. Andre Tok who gives me valuable technique advice and writing tips for this thesis.

Thirdly, I would like to thank my committee members, Professor Recker and Professor Jay. They encourage me to develop this thesis and give me supports.

Then, I would also thank Professor Saphores, Professor McNally, and Professor Jin who enhance my professional knowledge and skills.

Finally, I want to thank all my friends in ITS. Without their supports and loves, I may not finish my thesis.

ABSTRACT OF THE THESIS

Investigation of A New Method for Improving Individual Vehicle Speeds Estimation with Advanced Loop Data

By

Yiqiao Li

Master of Science in Civil Engineering

University of California, Irvine, 2017

Professor Stephen G. Ritchie, Chair

In this study, a previous speed estimation model supported by advanced loop data was evaluated in a new dataset with a high volume of large vehicles. With 23% large vehicles in the traffic stream, the performance of the previous model got worse. Duration of a vehicle traversing the loop sensor, which was the feature selected to develop the previous model was easily influenced by the variation of vehicle lengths. Leading edge duration which indicates the time period from when the front bumper of the vehicle first gets onto the loop to the vehicle fully covering the loop was investigated. A new speed estimation model with a new feature (leading edge duration) was developed. The leading edge duration was justified as a more reliable feature than duration to be used in speed estimation. Two models based upon the corresponding features were compared. Finally, results of the model comparison and suggestions for the further model improvements were provided.

Chapter 1

Introduction

1.1 Background

With the development of traffic detection technology, essential information can be captured on the roadway network. In order to improve the travel efficiency for the roadway users, accurate network information is required.

In recent years, researchers have been trying to improve the accuracy of traffic parameter estimation by using advanced traffic detection sensors. Non-intrusive detection technologies can provide relatively accurate measurements of aggregated vehicle speed and travel time. However, the spatial coverage, as well as penetration rate are limited by those methods. For example, the penetration rate of Bluetooth devices has been investigated in different research and it varies from 3% to 10%[1][4][10]. The penetration rate is also spatially dependent[23]. The Bluetooth method will not be applicable in the area where the traffic stream contains undiscoverable Bluetooth devices.

Alternatively, information collected from conventional loop detectors provides excellent temporal coverage. Meanwhile, single-loop detectors are widely installed in the US, which are also solved the spatial coverage issues. Currently, the conventional loop detector uses the effective vehicle length for vehicles traversing the detector which can only provide the aggregated speed for certain time period. With advanced detector cards, the inductive loop detector system is capable of getting much more detailed information from the waveform generated by the individual vehicle. Taking the advantages of this technology, researchers have started to use features extracted from the inductive vehicle signature to get accurate individual vehicle speed[19]. Subsequently, researchers explored more speed-related features from vehicle signatures, in order to increase the accuracy of the speed models[16][17][22].

This study evaluates the latest individual speed estimation model as well as develops a new model to enhance the estimation of individual vehicle speed.

1.2 Motivation

Traffic speed is an essential input of traffic flow models. It is a good indicator of traffic condition on the roadway network. An accurate speed estimation can give travelers an appropriate guidance that helps them to avoid congestion on road segments.

The inductive loop detector is one of the most commonly used data collection devices in the US and many other countries. They are widely spread along highways and mostly deployed as single-loop detectors per lane. Compared to other advanced sensors, the inductive loop sensor is considered as one of the most reliable and cost-effective methods in the vehicle detection field[14].

In recent years, with the limitation of conventional loop detectors, studies have focused on aggregated speed estimation by using single-loop detector data. However, individual vehicle speed estimation is much more important. Different vehicle types have different physical properties. For example, on highways, trucks typically travel at a lower speed than passenger cars. From the safety perspective, individual speed differences have a significant influence on accidents[13]. Also, the upstream and downstream speed can be used to identify the moving bottleneck which can cause congestion condition on the roadway segment[15]. In this case, individual speed would be a good indicator to diagnose congestion and incidents. With accurate speed estimation on a link of the network, an accurate travel time estimation which can reflect traffic conditions can be provided. The precise traffic information will help travelers to avoid congestion conditions and will improve the traffic performance on the roadway network.

This research investigates a previous speed estimation model and develops a new speed estimation model. Two models are compared with each other, and suggestions for improving the speed estimation model is given at the end of this thesis.

1.3 Thesis Outline

The rest of this thesis is organized as follows:

Chapter 2, Literature Review: This chapter reviews the literature on speed estimation. First, the relationship between vehicle speed, effective vehicle length, and duration is described. Then, the methods using conventional loops are summarized. Subsequently, speed estimation with advanced sensors is introduced. At the last part of this chapter, an investigation of prior inductive signature method is presented.

Chapter 3, Preliminary Data Analysis: This chapter presents an introduction of the dataset that is used for this research. It mainly includes the statistical summary of the dataset. Also, the geographical information and the layout of road segment are given.

Chapter 4, Model Development: This chapter is the main part of this research. In this chapter, a detailed feature extraction process is given. The prior inductive loop signature model is re-applied in this chapter and a new speed model is developed.

Chapter 5, Results Analysis: The old and new speed estimation models are implemented with a test dataset and the accuracy of both models is calculated. At the end of this chapter, a comparison between the two speed models is given.

Chapter 6, Conclusion: The last chapter summarizes this research and suggests the directions for future studies.

Chapter 2

Literature Review

2.1 Speed, Vehicle Length, and Duration

In order to better understand the performance of the transportation supply system, many researchers have attempted to figure out the relationship between traffic variables in mathematical and statistical terms. In 1952, Wardrop[25] derived the fundamental law of traffic flow which illustrated the relationship between three traffic parameters. In a traffic stream, with flow q, speed v, and spacing $1/q$, the density of this stream should be:

$$k = q/v \tag{2.1}$$

When a vehicle passed the single-loop detector, it triggers the inductive changes in the loop sensor. The vehicle is recorded during the time period when the front bumper of the vehicle gets onto the leading edge of the loop, and the rear bumper of the vehicle leaves the following edge of the loop. The signal for one vehicle traversing a conventional loop detector is shown below:

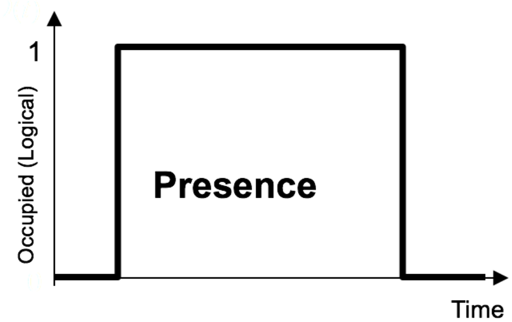

Figure 2.1: The Signal from the Conventional Loop Detector

The time period that the vehicle is on the loop is called on-time[12]. It is also called duration in some research[17]. For consistency, in this study, it is called duration and denoted by d.

With the occupancy (occ) and aggregated flow rate (q) given by the presence detectors, space mean speed can be derived by using the equation shown below:

$$v \approx \frac{q \cdot l_e^m}{occ} \tag{2.2}$$

v stands for space mean speed. q is the aggregated flow given by the presence detectors. occ stands for the ratio in time for a loop detector to be covered by vehicles.l_e^m represents the mean of effective vehicle length. The effective vehicle length is the summation of the vehicle length and the loop length. In most of the speed estimation studies using loop detector data, speed estimated by double-loop detectors is considered as the true speed, denoted by $\dot{x}_{true}$. Then the effective vehicle length can be estimated by using the equation shown below:

$$l_e = l_v + l_l = d \cdot \dot{x}_{true} \tag{2.3}$$

The graph shown below illustrates the relationship between effective vehicle length and individual vehicle length.

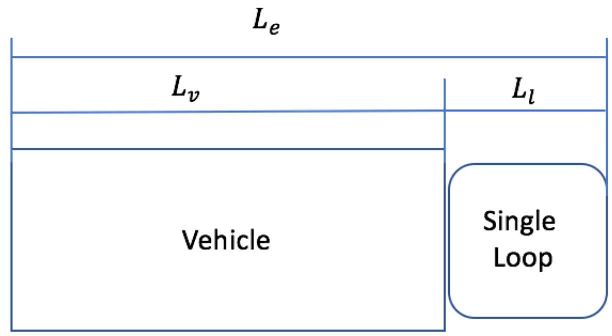

Figure 2.2: Effective Vehicle Length

2.2 Conventional Loop Methods

Basically, most of the conventional loop methods are based on the speed, vehicle length and duration relationship shown in the previous section. The first speed estimation method by using single loop detectors was proposed by Athol[2]. In 1965, Athol[2] used the definition of occupancy and introduced a g factor into speed estimation. The equation used by Athol[2] is shown below:

$$v = \frac{N}{g \cdot T \cdot occ} \tag{2.4}$$

Here, v represents space mean speed. The g factor is an estimator which incorporates site characteristics of effective vehicle length. In this approach, g only varies site by site. However, in reality, effective vehicle lengths have large variations, although they are obtained in one site. Also, it is hard to observe uniform traffic on freeways as Athol[2] assumed in his method. In this case, many researchers have sought better ways to reduce the bias caused by vehicle length and congestion conditions. Pushkar[18] used a cusp catastrophe theory model to estimate the average vehicle length, and a new speed model was developed based on the estimated average vehicle lengths. The new speed model showed a better performance than just using constant average vehicle length. Rather than using static effective vehicle length for different time periods within one day, Coifman[6] used an exponential filter to dynamically update the average vehicle length, which makes the speed model much more reliable. After that, Coifman[7] found that instead of using mean speed, median speed is much less sensitive to outliers and performs better than the previous model. In 2003, Kwon et al.[11] classified vehicles into "Passenger Vehicle" and "Truck" by using the vehicle length distribution and estimated the effective vehicle length based upon that. Similarly, Wang and Nihan[24] used the distribution of vehicle length to estimate the truck volume. However, these two methods operate under two similar assumptions: First, each study period contains more than 2 intervals and vehicle speed can be considered as constant. Second, no large trucks are present in each period of at least two intervals. In reality, it is hard for vehicles to maintain a constant speed within these two intervals. Also, the large truck volume is an essential element which can not be ignored. Such a model can only therefore be applied in that specific scenario. The performance of these methods with respect to the temporal and spatial transformability are undesirable. Then, Coifman[8] integrated the sequence method that uses a sequence of vehicles to estimate space mean speed in a small time interval and moving median method into the on time distribution method to improve the length-based vehicle classification and enhance the vehicle speed estimation. Three years later, Lao et al.[12] analyzed the on time distribution pattern of vehicles and used a Gaussian Mixture Model (GMM) to classify different types of vehicles by length. The performance of the speed model was improved by using GMM to cluster vehicle by length, and the model accuracy was increased under high truck flow. Since the vehicle length distributions are different between each loop site, the model cannot avoid a low transferability issue which means the model may not work well in other loop sites. In 2014, Coifman[9] introduced hybrid methodology into speed estimation to cope with issues that were brought into previous studies by high truck flows. The new sequence method proposed by Coifman[9] can provide good speed estimation results under 40% of long vehicles within a day.

2.3 Speed Estimation with Other Sensors

In recent years, various sensors have been used in the traffic surveillance field. Based on the installation method, detection technologies can be broadly classified as intrusive detection and non-intrusive detection. Pneumatic road tube, inductive loop detectors, piezoelectric sensors, and Weigh-in-Motion can be categorized as intrusive detection. On the other hand, video image processing, microwave radar, infrared sensors, GPS (Global Position System), Bluetooth sensors, and probe vehicles that do not involve in-pavement installations can be called non-intrusive detection. Since some of the non-intrusive detectors have relatively good spatial flexibility, non-intrusive detection technologies such as GPS, Bluetooth sensor and probe vehicle are used for performance measurement on roadway networks. Since vehicles equipped with a GPS device can report the individual vehicle speed, Cheu et al.[5] calculated the optimal sample size of probe vehicles equipped with GPS devices and got relatively accurate speed estimation for the link of an arterial road. In 2013, by using a data fusion algorithm, Bachmann[3] integrated Bluetooth data and conventional loop detector data and improved the aggregated speed estimation on freeways.

2.4 Inductive Signature Method

Generally, the conventional loop detector card can only collect bivalent data. The output of the detector card is either "0" or "1", which can directly produce occupancy and vehicle counts. "1" represents the presence of a vehicle. Usually, the occupancy can be calculated by using the presence of a vehicle passing over the loop detector as shown in Figure 2.3. The advanced detector card can capture the inductance change caused by the vehicle passing over the loop. The inductance change produces a waveform called a "vehicle signature". By using the analog signal, the advanced detector is able to obtain the vehicle signature of an individual vehicle. The two curves shown in figure 2.3 are two vehicle signatures generated by double-loop detectors for a single vehicle traversing the sensors. The x axis represents time. In this study, the sampling rate of the detector card is 1000 points per second. So the time gap between each point is 0.001s. The vertical axis refers to the inductance changes and is usually described as magnitude. Vehicle signatures can be used not only for getting volume and occupancy but for getting individual vehicle speed and vehicle types[20].

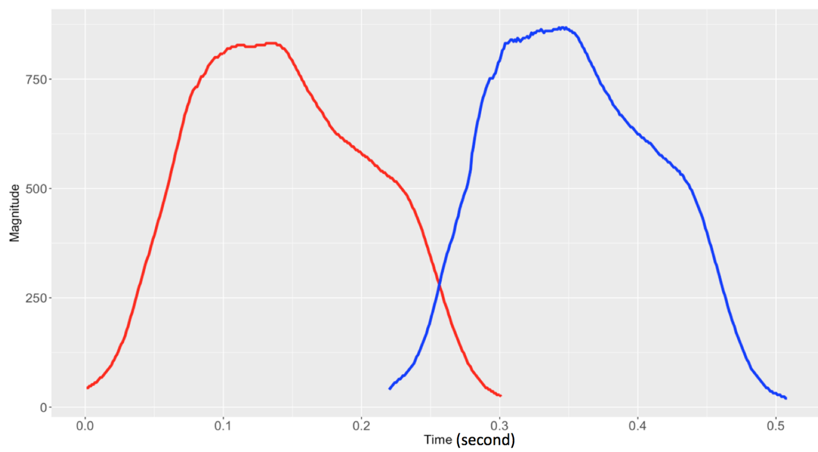

Figure 2.3: Vehicle Signature from the Double-Loop Detector

Using inductive signature data collected from the single loop detector is also a major approach for speed estimation. Some of the features extracted from the vehicle signature can represent a certain characteristic of the vehicle. In 1999, Sun and Ritchie[19] found that the slew rate of individual vehicles had a strong correlation with vehicle speed and they are linearly related. They developed a linear model between individual vehicle speed and slew rate. After conducting several statistical tests of the model, they found that the model is statistically significant and the data can be well explained by the model. However, the linear relationship between different vehicles would be different. In order to cope with this problem, Oh[16] used a probabilistic neural network (PNN) to group vehicles by vehicle length and applied a linear regression model for each vehicle group to increase the accuracy of the model. Later, Park [17] introduced a shape parameter into a vehicle grouping model in order to get more specific vehicle types within each group in the model. Since two different vehicles traversing an inductive loop sensor with the same velocity may have different slew rates and duration, Tok[22] adopted k-means clustering and artificial neural network classification methods to group vehicles into homogeneous groups based on natural slew and signature length characteristics. The linear model within each group was improved. However, the regression of the speed model corresponding to the cluster representing long vehicles was significantly weaker than others.

2.5 Summary

Generally, compared with the conventional loop approach, models developed by using inductive signature approach have a better accuracy and transferability. Compared to the non-intrusive technologies, the inductive signature approach provides a better temporal and spatial coverage. However, the performance of these models may be reduced with the presence of large vehicles. In this case, more specific vehicle classes, as well as a better feature selection, may be helpful to improve the performance of the model under high truck flow conditions.

Chapter 3

Preliminary Data Analysis

The data used in this study were obtained from the Castro Valley double-loop station which covers both directions of I-580 in the city of Castro Valley, California. This loop station is located in the middle section of the Sergeant Daiel Sakai Memorial Highway. There are four general purpose lanes for each direction. Each lane is equipped with double round loops. The spacing between the leading edges in a speed trap of each pair of double loop is 6m (19.685ft). The length of each single round loop is 1.8m (5.906ft). In this site, the configuration of the double loop speed trap provides the ability to obtain accurate individual vehicle speed for developing and evaluating single inductive loop speed estimation models.

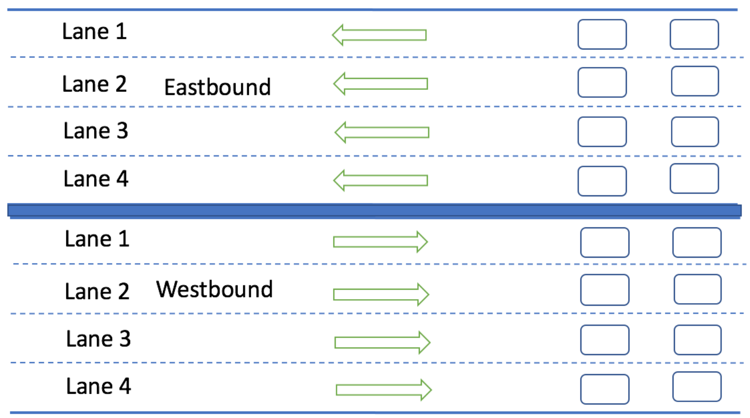

Figure 3.1: Study Site Configuration

The data were collected on January 31st, 2017 between 13:50 and 16:10. Within this time period, 12,008 vehicle signatures were recorded. The total vehicle count within the collection time period was 6004. Excluding bad signatures (signature length less than 10ft and larger than 80ft) and vehicles with lane changing behavior (vehicles just show up in one of the loop sensors), 5980 vehicles were used for developing and testing the speed model.

In order to develop a reliable speed estimation model, the data set was separated into a training set (80%) and a testing set (20%). The training set was used to develop a vehicle grouping model. The training set was then separated into a sub-training set, and a validation set which are correspondingly used for training and validating the neural network model. The remaining 20% testing set was used to evaluate the performance of the new speed estimation model.

Chapter 4

Model Development

4.1 Feature Extraction

In the model development process, features which can represent the vehicle characteristics and are correlated with vehicle speed are required. Based on previous studies, the slew rate located at the leading edge of the inductive waveforms and the inverse duration has been found to be highly correlated with individual vehicle speed[17][19][22]. A detailed description of the features which are used in the model development is provided in this section. Furthermore, the reliability of a new feature for a new speed estimation model is investigated.

4.1.1 Signature Preprocessing

Prior to the feature extraction procedure, the raw signatures provided by high-resolution inductive loop detector card need to be pre-processed. There are several steps to pre-process the raw signatures. First, when vehicles pass over the loop magnetic field, it causes the reduction in inductance. For the convenience of analysis, the inductive signatures were flipped to the positive side of the y-axis. Second, based on the double loop data, the vehicle lengths can be calculated. In this study, vehicle length less than 10 ft or larger than 80 ft are considered to be outliers and excluded in the pre-processing step. Some of the raw signatures are relatively rough, and the "bad" slew rate caused by noise may be captured. Then, the moving average filtering method should be conducted to get "smooth" signatures before feature extraction. Next, loop sensors deployed on different lanes have different sensitivities. In order to cope with the sensitivity issue, signatures are normalized. Finally, the oscillations near the baseline inductance should be eliminated. An arbitrary threshold which can exclude all the oscillations as well as preserve as much of the vehicle signature information as possible is used. According to previous studies, a value between 10%-20% of the signature peak magnitude was found to be a reliable threshold from an empirical experiment with various datasets[19]. In this study, the magnitude values below 0.2 were excluded.

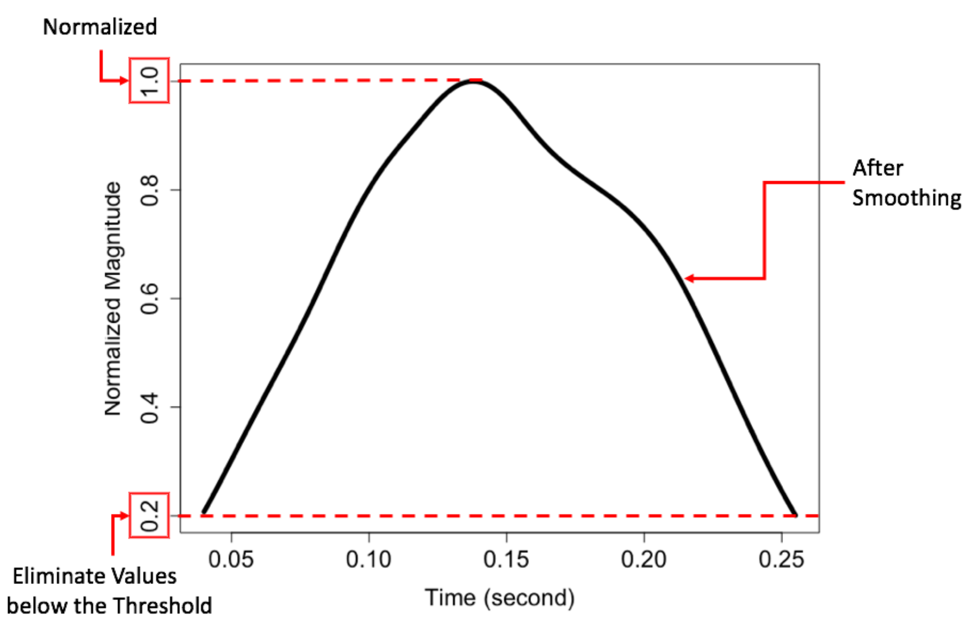

Figure 4.1: The Signature After Pre-processing

4.1.2 Slew Rate and Duration

Slew rate is a feature located at the leading edge of a vehicle inductive waveform signature. This feature represents the rate of the metallic mass of the vehicle traversing the magnetic field of inductive loop sensor. It is the steepest slope of the leading edge curve of a signature. To extract the slew rate of a signature, the leading edge should be identified. In this study, the leading edge is defined as the part of the signature before the first point which has a derivative value equal to 0. The leading edge of a signature was tagged in Figure 4.2. Then, the largest derivative value of the leading edge is picked and considered as the slew rate of the signature.

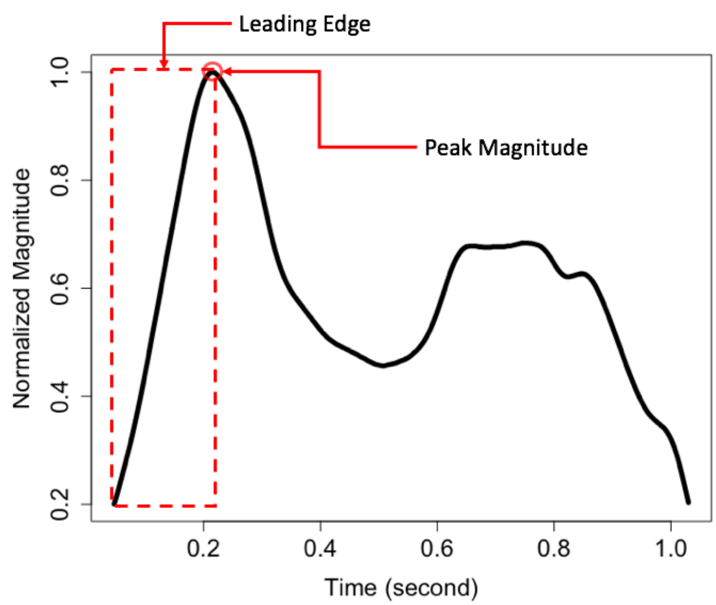

Figure 4.2: Leading Edge of a Signature

Duration is an important feature which can be extracted from the vehicle signature. It records the time from when vehicle's front bumper gets onto the leading edge of the front loop sensor and triggers the inductance change above the loop to the rear bumper of the vehicle and leaves the trailing edge of the loop sensor. The slew rate and duration extraction were illustrated in Figure 4.3. The duration value in Figure 4.3 has eliminated the oscillation part of the signature. Here, mag represents the magnitude value. t stands for time. Slew rate is denoted by SR. Duration is denoted by DUR.

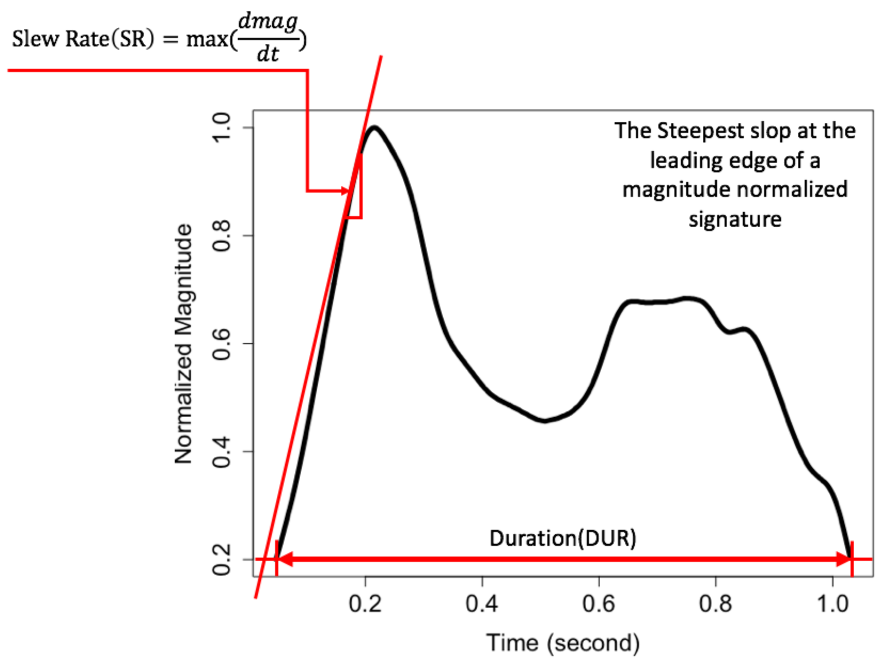

Figure 4.3: Slew Rate and Duration of a Signature

4.1.3 Natural Slew and Signature Length

Various lengths of vehicles traversing the same inductive loop sensor with the same speed may have different slew rates and duration measures. Therefore, vehicle length should be considered the speed model to reduce the error caused by various vehicle lengths. In this case, natural slew and signature length which can capture the vehicle length information are introduced[22].

Natural slew (NS) is a feature belonging to the magnitude normalized and spatially transformed signature. It is the maximum slope on the leading edge of the magnitude normalized and spatially transformed signature. The signature length is calculated by using the duration of the signature and the true vehicle speed:

$$L_{sig} = DUR \cdot \dot{x}_{true} \tag{4.1}$$

The figure shown below depicts the method for extraction of natural slew and signature length of a signature. Here, NS represents natural slew. l indicates the vehicle passing time multiplied by the vehicle true speed. L_{sig} stands for the signature length.

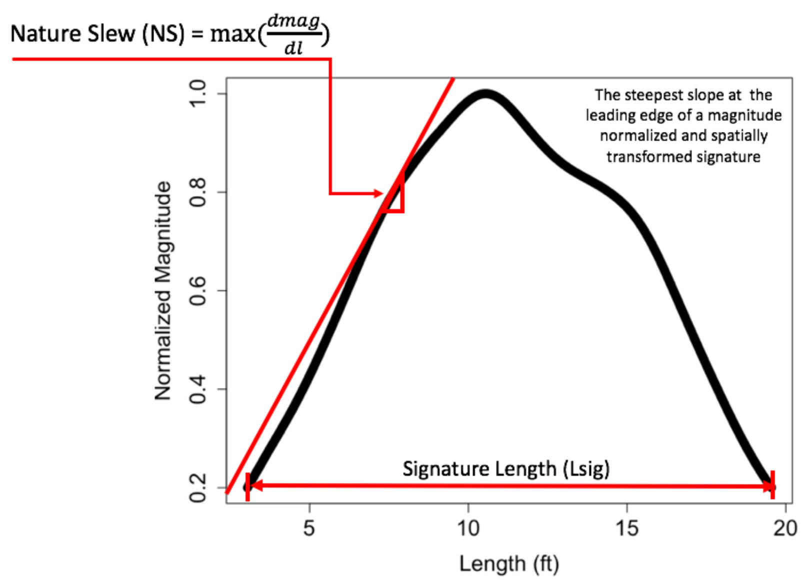

Figure 4.4: Vehicle with Same Signature Length

4.1.4 Investigation of Leading Edge Duration and Length

Generally, when a vehicle's front bumper travels from the leading edge of the inductive loop sensor to the trailing edge of the sensor, the inductance may present a decreasing trend until the vehicle covers the whole loop (Note: The vehicle waveforms have been inverted). In a sense, the leading edge of a signature approximately indicates the duration that the vehicle's front bumper traverses from the leading edge of a single loop to its trailing edge. The process is illustrated in Figure 4.5. Here, the leading edge duration is denoted by $LDUR$. The single loop length is approximatly 6ft.

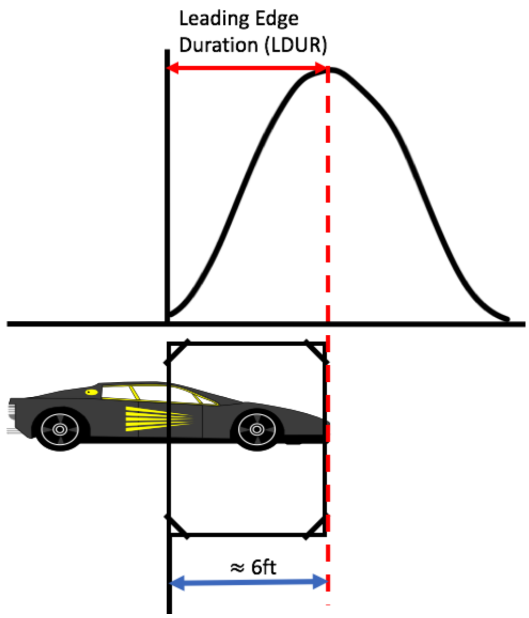

Figure 4.5: Leading Edge Duration

Similarly, after spatial transformation of the signature, the leading edge length of the signature can be obtained. Figure 4.6 describes the method of getting the leading edge length.

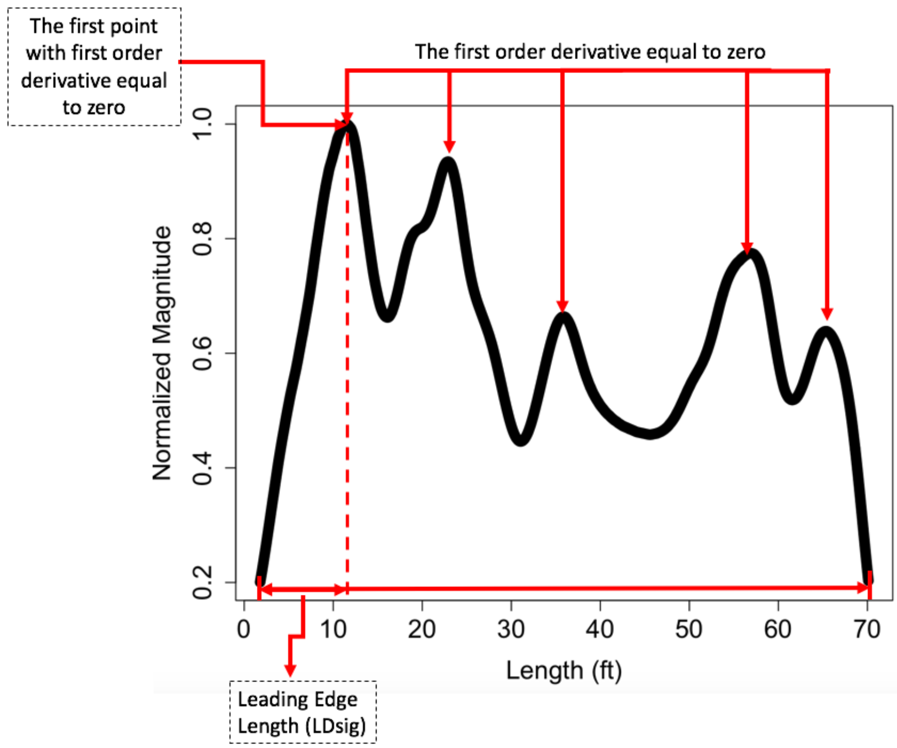

Figure 4.6: Vehicle with Same Signature Length

Based on the explanation shown in the graph, the new feature, leading edge signature length, indicates the loop length and certain vehicle characteristics. One concern of speed models that use the full duration measure is that the high length variance in large commercial vehicles will adversely compromise the linear relationship assumption between duration and speed, and consequently impact the accuracy of speed estimates of such vehicles. In this case, the leading edge signature length has lower variance than the full signature length. It is possible to be a more reliable variable for developing an accurate speed estimation model. A statistical test is conducted to verify the reliability of this variable for the speed estimation model in the next section.

4.1.5 Feature Comparsion

In this section, the length-based vehicle classification model is used to distinguish different vehicle types. Based on different research purposes, vehicles are classified into different groups by their lengths. According to the Analysis of Variance (ANOVA) test results, the length-based vehicle classification model using the Gaussian Mixture Model (GMM) performs significantly better than other length-based models[9].

The GMM vehicle classification model classified vehicles into three classes[12]. Vehicle length within class 1 is less than 22 ft. For class 2, vehicle length is in the range of 22 ft to 40 ft. Class 3 includes vehicles which have lengths larger than 40 ft. Then, based on the double loop signature, the "true" vehicle length can be calculated. The whole dataset is separated into 3 groups and vehicles within one group are considered to be one class. Subsequently, the signature length and leading edge signature length is plotted for each vehicle class. The results are shown below:

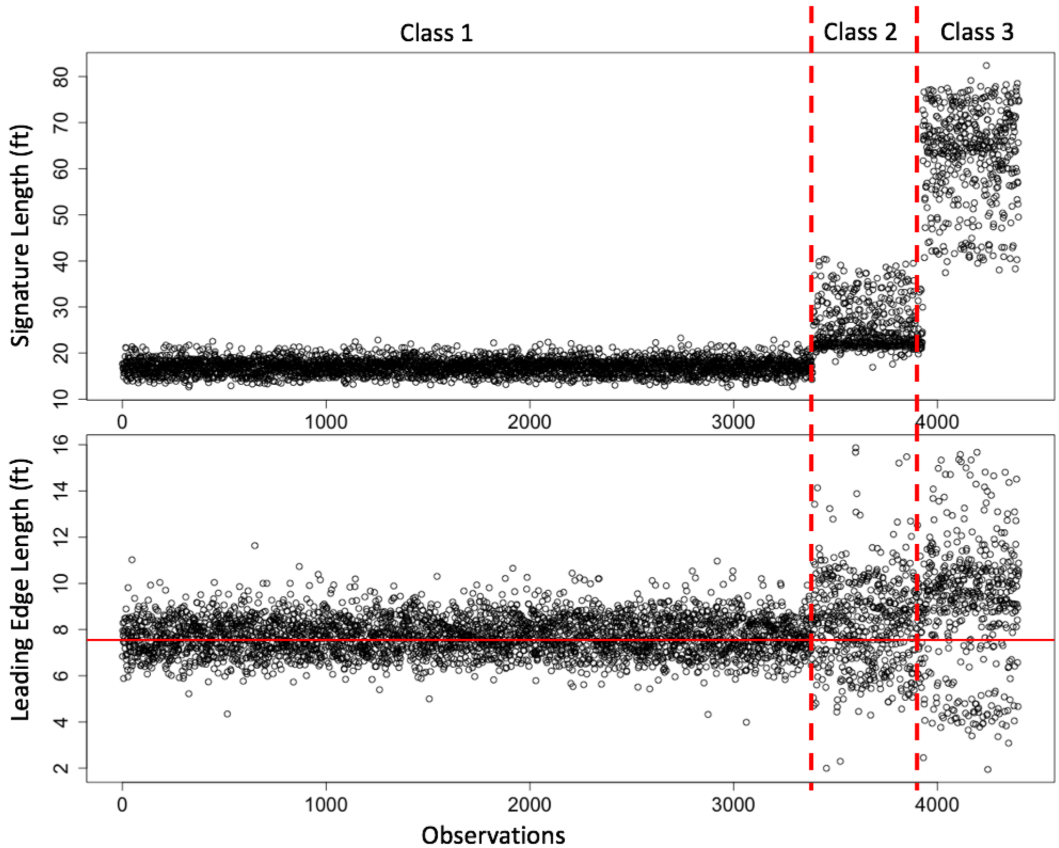

Figure 4.7: Signature Length and Leading Edge Length within Each Class

Signature length for different vehicle classes is obviously centered to a different length. On the contrary, leading edge length for different classes is centered to approximately 7 ft. As expected, the leading edge length captures the loop sensor's length and certain characteristics of the vehicle's physical attributes. To further statistically verify whether the variance of signature length is larger than the variance of the leading edge length within a specific class, a F-test has been conducted. The main purpose of the F-test is to verify whether the variance of signature length for a certain class is significantly larger than the variance of the leading edge length in the same class. The null hypothesis H_0 is that the ratio of variances $(\frac{Var(L_{sig,i})}{Var(LD_{sig,i})})$ is not greater than 1. Here, i indicates the class number. $i = 1, 2, 3$. The F-test results are shown in Table 4.1.

In all three classes, under 0.05 significant level, H_0 is rejected. The final decision of the

Table 4.1: F-test Results

Class (i)	Vehicle Length	F-Statistic	p - value	Decision
1	< 22ft	5.653	2.20×10^{-16}	Reject H_0
2	22ft - 40ft	12.802	2.20×10^{-16}	Reject H_0
3	> 40ft	70.927	2.20×10^{-16}	Reject H_0

F-test suggests:

$$Var(L_{sig,i}) > Var(LD_{sig,i}), i = 1, 2, 3 \tag{4.2}$$

Generally, the vehicle length's variety causes the inaccuracy of the speed model, especially under high truck volume traffic conditions. In this case, introducing the leading edge length to the speed estimation model may provide much more accurate estimation results.

In the next section, the signature length-based approach is applied, and a new vehicle grouping model is developed by using the leading edge length of the signature. A comparison between these two models is provided in the next section. The outcome of the model which has better performance is subsequently used as the input to the vehicle classification model for the next step.

4.2 Vehicle Grouping Model

Since two different vehicles with the same speed passing over an inductive loop sensor may have different slew rates and duration measures, vehicles have to be grouped by their features which are related to individual vehicle speed. In the first part of this section, the old method which used natural slew and signature length as grouping variables are implemented. In the second part of this section, a new variable is introduced and a new vehicle grouping model is developed. To differentiate the two models, in this study, the old model is referred to as signature length-based approach and the new model is refered to as leading edge length-based approach. The comparison between two models is provided at the end of this section.

In vehicle grouping process, K-means clustering algorithm is used. K-means clustering algorithm is an unsupervised learning method. It tries to group the unlabeled data based on their attributes or features into K number of groups[21]. The grouping process is ended by minimizing the sum of square distances within each group and the corresponding clusters' centroids. The variables' attributes would be relatively homogeneousness within one cluster which is the objective of the vehicle grouping model in this study. Therefore, K-means clustering algorithm is introduced in the speed estimation model.

Before conducting the K-means clustering method, the outliers and bad signatures are excluded, and the training dataset is scaled by the mean and standard deviation. In this case, the clustering model would not be biased by unequal variances. The equation shown below explains the scaling process.

$$Scaled.NS_i = \frac{NS_i - Mean(NS_{tr})}{Std(NS_{tr})} \tag{4.3}$$

$$Scaled.L_{sig,i} = \frac{L_{sig} - Mean(L_{sig,tr})}{Std(L_{sig,tr})} \tag{4.4}$$

$$Scaled.LD_{sig,i} = \frac{LD_{sig} - Mean(LD_{sig,tr})}{Std(LD_{sig,tr})} \tag{4.5}$$

Here, NS_i represents the natural slew rate of the i^{th} vehicle. L_{sig} represents the signature length of the i^{th} vehicle. Similarly, LD_{sig} represents the signature length of the i^{th} vehicle. The subscript "tr" represents the fact that the feature is obtained from the training dataset.

4.2.1 Signature Length-Based Approach

Optimal Vehicle Grouping

Prior to the K-means clustering analysis, the number of clusters has to be declared. The final speed estimation regression model may be affected by the configuration of each cluster. In order to get the optimal number of clusters, the sum of square distances should be calculated when introducing one more cluster into the model. The number of data points would influence the goodness of fit of the regression model in each cluster. Then, the optimal group numbers should have limited numbers and a relatively small sum of square distance. After 5 iterations, the RMSE between each iteration was less than 15%. Obviously, when the cluster number reaches 5, the sum of the square distance would not decrease significantly as the cluster size increases. The optimal cluster size is determined as five. Here, A_1 is used to denote signature length-based approach. A_2 is used to denote leading edge based approach.

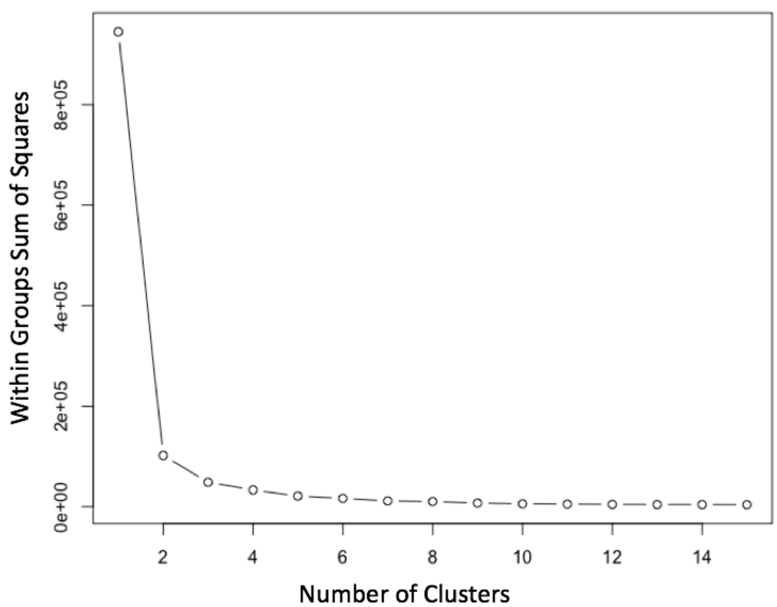

Figure 4.8: K-means Clustering Results (A1)

After determining the optimal group numbers, the K-means clustering algorithm is applied on NS and L_{sig} in the training dataset. The clusters' statistical summary is shown below. Clusters' size, mean of cluster variable, and standard deviation (Std) is calculated.

Table 4.2: Statisitical Summary of K-means Clustering (A1)

Cluster Number	Count	Natural Slew		Signature Length	
		Mean	Std	Mean	Std
1	1954	0.1605	0.005	17.558	2.995
2	1408	0.142	0.008	19.075	4.183
3	843	0.180	0.011	18.620	5.228
4	310	0.112	0.018	62.686	10.642
5	173	0.157	0.019	63.136	9.138

The K-means clustering result is shown below. According to natural slew and signature length, vehicles are classified into five groups. Those five groups are labeled by five distinct colors.

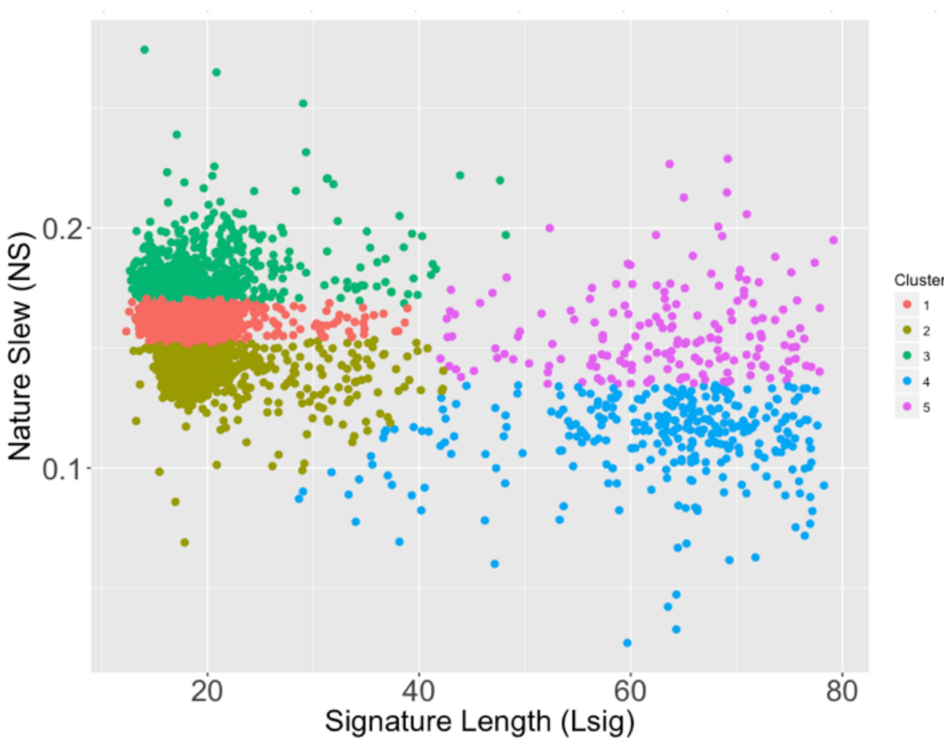

Figure 4.9: K-means Clustering Results (A1)

Speed Estimation Regression Models

In this model, vehicles are categorized into five groups based on their natural slew and signature length. Within each vehicle group, linear regression models are developed. Slew rate and inverse of the duration (DUR^{-1}) extracted from single loop are considered to be the explanatory variables. The formulation is shown below:

$$\hat{x} = \beta_0 + \beta_1 \cdot SR + \beta_2 \cdot DUR^{-1} \tag{4.6}$$

Where, $\hat{x}$ stands for the estimated speed from the linear model. β_i, $i = 0, 1, 2$ is the coefficient of the linear model.

Table 4.3: Summary of Linear Regression Models (A1)

Cluster Number	Constant	Slew Rate Coefficient	Inverse DUR Coefficient	Adjusted R^2	F Statistic	Degree of Freedom
1	1.503 (7.748)	5.922 (142.999)	0.502 (5.150)	0.989	86600.0	1951
2	2.854 (6.362)	6.407 (83.361)	0.972 (5.655)	0.958	15910.0	1405
3	1.225 (3.612)	4.761 (93.019)	2.208 (17.760)	0.986	28890.0	840
4	16.434 (11.050)	4.540 (21.080)	15.837 (13.990)	0.841	815.9	307
5	2.930 (2.285)	4.146 (18.746)	18.897 (10.041)	0.940	1354	170

The statistical summary of five regression models is shown above. Estimated parameters of each variable and their t-statistics (in parentheses) are listed in Table 4.3. Under 0.05 significance level, based on the values of t-statistics, all parameters are statistically significant to their corresponding models. Additionally, F-statistics of each model suggest that all five models are significant. According to adjusted R^2, cluster 1, 2, and 3 which represent three types of short vehicles have a relatively excellent goodness of fit. Cluster 4 and 5 which represent longer vehicles have lower adjusted R^2, but those two models are still look pretty good. However, this problem can be solved by using the leading edge based approach.

4.2.2 Leading Edge Length-Based Approach

Optimal Vehicle Grouping

The K-means clustering method was re-applied to the leading edge length-based approach. The clustering results are shown in Figure 4.10 and Figure 4.11. Based on the sum of squared error and the number of observations within one cluster, the optimal number of clusters is determined as five.

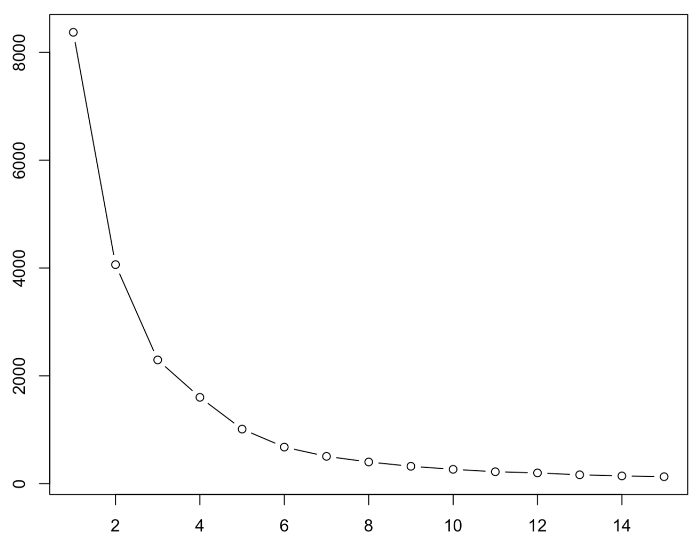

Figure 4.10: K-means Clustering Results (A2)

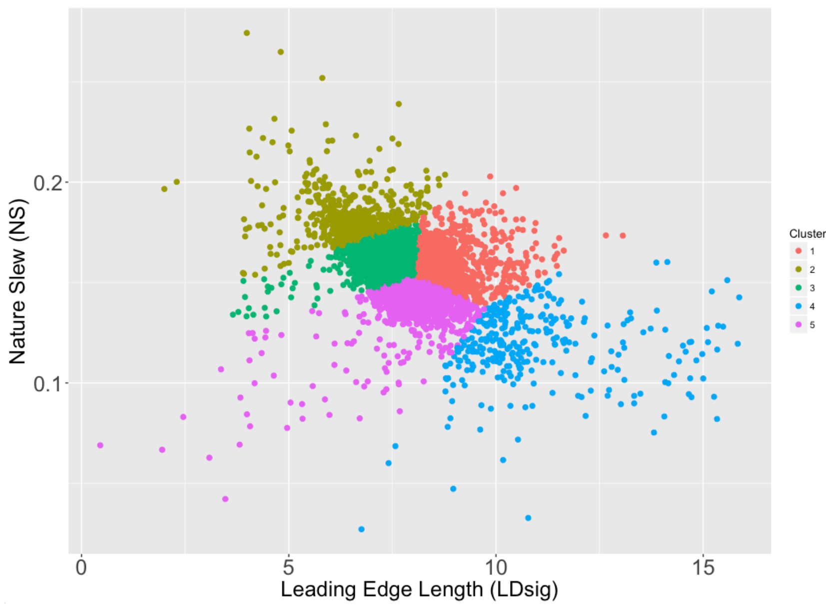

Figure 4.11: K-means Clustering Results with Five Clusters (A2)

Table 4.4: Leading Edge Length-Based Approach with Five Clusters

Cluster	Count	NS Mean	LDsig Mean	Adjusted R^2	F statistic
1	988	0.160	8.891	0.983	28710
2	602	0.183	6.519	0.976	11950
3	1585	0.161	7.258	0.986	53920
4	318	0.118	10.925	0.878	315
5	1195	0.139	7.979	0.941	1192

The result of the leading edge length-based method with five clusters is shown above. Apparently, the adjusted R^2 of the linear relationship between independent variable ($\hat{x}$) and dependent variable (NS and LD_{sig}) are better than the adjusted R^2 in Equation 4.6, even in the cluster which indicates vehicles with larger vehicle length. In this case, if the vehicle with certain characteristics is assigned into the right cluster, the estimated vehicle speed will be extremely accurate. However, during the model development process, one drawback of this approach was found. Since the variance of different vehicle leading edge lengths for each groups is low, the vehicle characteristics obtained from a single loop detector are not sufficient to better describe each cluster. The similarity between each cluster will confuse the vehicle classification model. Vehicle signatures obtained from single loop sensors cannot be correctly assigned to the right clusters. Therefore, the number of clusters has to be reduced.

After considering the accuracy of the vehicle classification model, the optimal number of clusters is determined as three. The clustering results are shown in Figure 4.12 and Table 4.5.

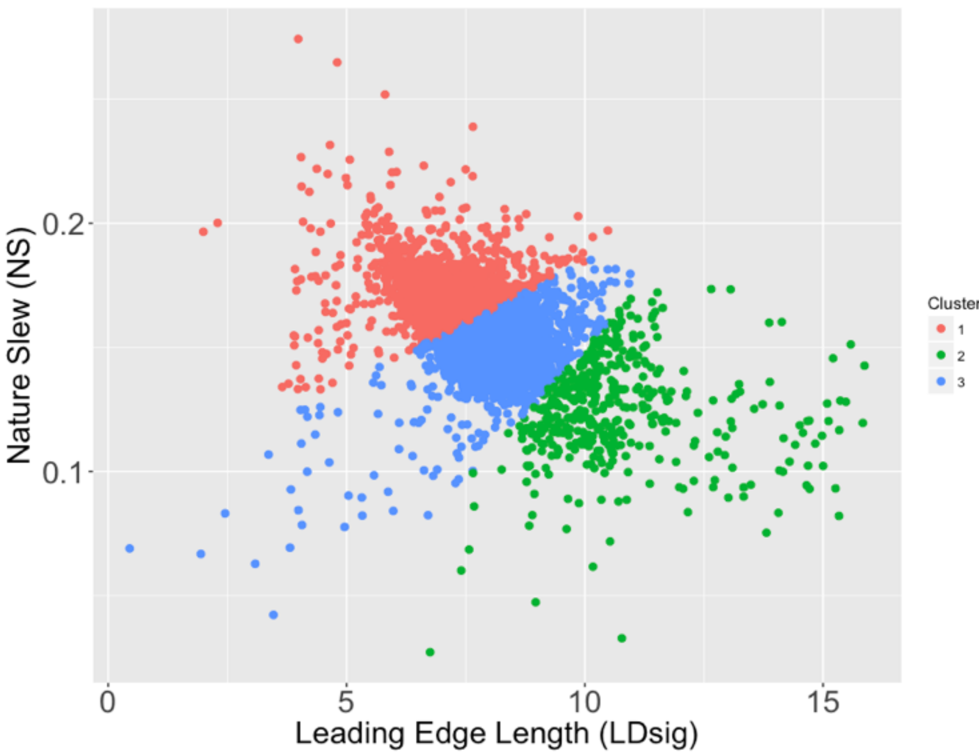

Figure 4.12: K-means Clustering Results with Three Clusters (A2)

Table 4.5: Statisitical Summary of K-means Clustering (A2)

Cluster Number	Count	Natural Slew		Leading Edge Length	
		Mean	Std	Mean	Std
1	1726	0.172	0.012	7.006	0.882
2	427	0.125	0.021	10.690	1.570
3	2535	0.149	0.012	8.113	0.812

Speed Estimation Regression Models

Similarly, the linear regression model is developed in each cluster. Under 0.05 significant level, all parameters are significant to their corresponding linear model. F statistics of each model indicate that all models are statistically significant. The adjusted R^2 of each model shows that all the regression models can better explain the data in their corresponding cluster.

Table 4.6: Summary of Linear Regression Models (A2)

Cluster Number	Constant	Slew Rate Coefficient	Inverse LDUR Coefficient	Adjusted R^2	F Statistic	Degree of Freedom
1	1.141 (3.260)	4.755 (71.980)	1.182 (16.400)	0.969	27190.0	1723
2	9.582 (8.205)	2.868 (17.300)	5.367 (24.621)	0.885	1634.0	424
3	1.951 (5.955)	5.155 (119.101)	1.658 (37.744)	0.957	27980.0	2532

However, natural slew and vehicle length cannot be captured by using a single loop sensor.

Subsequently, vehicle classification models are developed. The vehicle classification model is used for connecting the vehicle feature obtained from single loop and the pre-determined clusters. Detailed information is provided in the next section.

4.3 Speed Estimation Vehicle Classification Model

In the previous section, vehicles were grouped based upon their natural slew and signature length. These two features can only be extracted from single loop signatures. Thus, in these two approaches, vehicle signatures obtained from single loop sensors cannot be directly assigned into the pre-determined clusters. In this step, supervised learning has to be introduced into the vehicle classification model, since the input and output of the model are well defined. For this classification problem, artificial neural network (ANN) is considered to be used as a classifier.

4.3.1 Extraction of Selected feature

Certain features should be extracted from single loop signatures and are used as input of the vehicle classification model. The pre-determined clusters for each approach are considered as the output of the model. The selected input features include nine first order magnitude differences and ten magnitudes from a magnitude normalized signature. First, signatures obtained from single loop sensors are normalized based upon their peak magnitude and their duration. Then, on the x-axis, the signature is evenly divided into ten equal segments. Across the ten equal segments, the ten corresponding magnitudes and the nine magnitude differences are extracted. These nineteen total features are decided to be the input features of the neural network model.

4.3.2 Train and validate the Neural Network

The neural network model is developed with a single hidden layer and 38 neurons which ensure the accuracy of the model both in the training and the testing set. The sigmoid activation function is used as the activation function which helps the artificial neuron process incoming information and pass it throughout the network. The sigmoid function is shown below:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{4.7}$$

The output value of the sigmoid function can fall anywhere in the range from 0 to 1 which helps the model make a decision.

The neural network is trained in the training dataset, and it is validated in the validation set to avoid overfitting problems. The neural network model used for signature length-based approach provides 69% correct classification rate in the training set and 62% in the validation set. In the leading edge length-based approach, after reducing the number of clusters to three, the neural network model has 77% accuracy in the training set and 75.3% accuracy in the validation set.

Chapter 5

Model Results

5.1 Measures of Effectiveness

In this chapter, two quantities are used to measure the performance of the signature length-based model and the leading edge length-based model:

Mean absolute error (MAE) is used to measure how close the estimated values are to the true values. In this study, it represents the deviation of estimated speed from the true vehicle speed. n denotes the number of vehicles. Other notations are the same as the notations used in previous sections. The equation of MAE is shown below:

$$MAE = \frac{1}{n} \sum_{n=1}^{n} \left| \hat{\dot{x}} - \dot{x}_{true} \right| \tag{5.1}$$

The mean absolute percentage error (MAPE) is a measure of estimation accuracy of a predictive model in statistics. In this study, it is used to measure the percentage that estimated speed deviated from the true speed. The representation of MAPE is shown below:

$$MAPE = \frac{1}{n}\sum_{n=1}^{n}\left|\frac{\hat{\dot{x}} - \dot{x}_{true}}{\dot{x}_{true}}\right| \times \frac{100}{1} \tag{5.2}$$

The final results of the two models are shown in Table 5.1. Individual vehicle speeds are estimated in the training set as well as in the testing set. Classes 1, 2, and 3 indicate the length-based vehicle classification. Although the overall performance of the new model is slightly worse than the old model, it has a better performance with respect to short vehicles.

Table 5.1: Measure of Effectiveness Results for Two Speed Model

Dataset	Vehicle Length	Signature Length Based		Leading Edge Based	
		MAE(mph)	MAPE(%)	MAE(mph)	MAPE(%)
Training	Total	3.19	6.87%	3.43	7.54%
	Class 1	2.81	5.89%	2.89	6.09%
	Class 2	4.42	9.37%	4.87	11.28%
	Class 3	4.56	11.27%	5.78	13.95%
Testing	Total	3.51	7.64%	3.56	8.00%
	Class 1	2.94	6.15%	2.78	5.71%
	Class 2	4.50	9.48%	4.88	10.15%
	Class 3	5.92	14.75%	7.01	19.86%

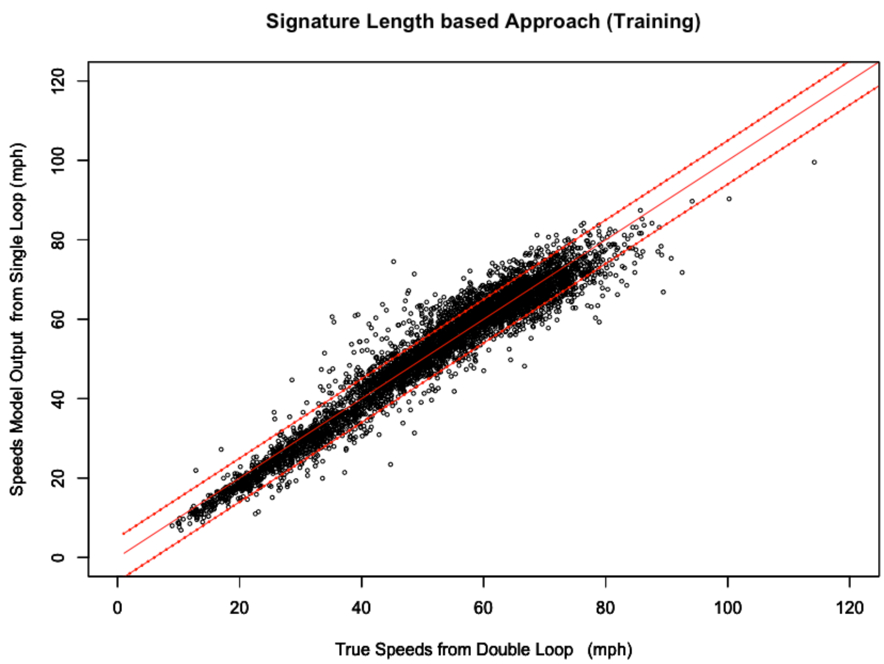

Figure 5.1: Signature Length based Approach Results on Training Set

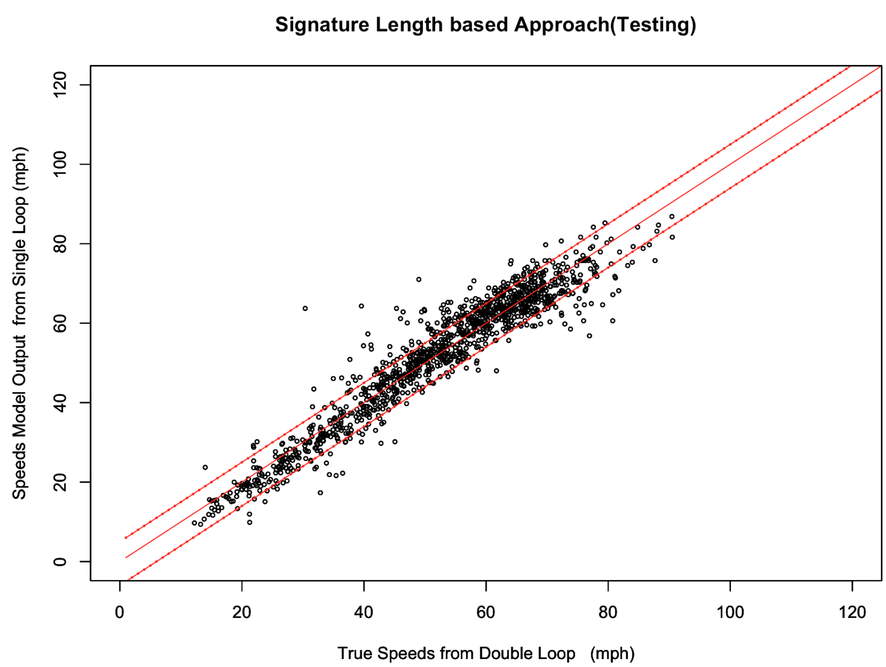

Figure 5.2: Signature Length based Approach's Results on Testing Set

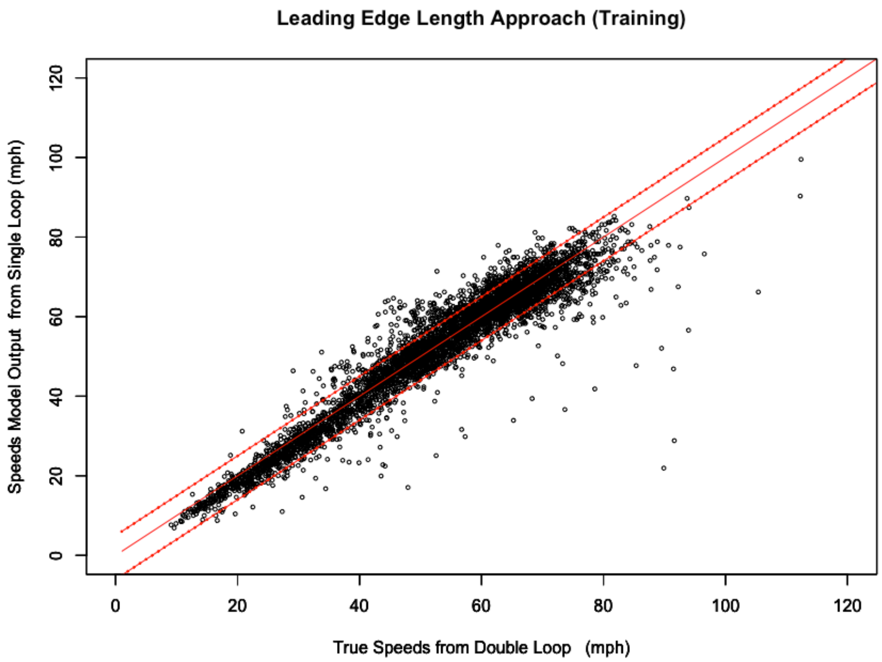

Figure 5.3: Leading Edge Length based Approach's Results on Training Set

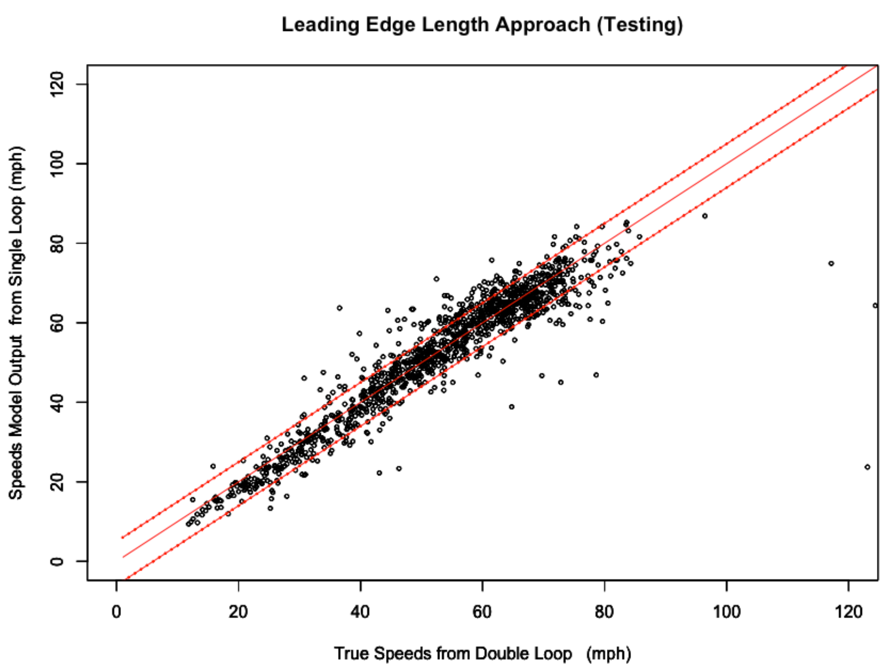

Figure 5.4: Leading Edge Length based Approach's Results on Testing Set

Figure 5.1 and Figure 5.2 present the signature length based model outcome from single loop sensors versus the true speed obtained from the double-loop speed traps on the training set and the testing set. The two red dash lines around the solid line are the ±5 mph error. Similarly, Figure 5.3 and Figure 5.4 show the leading edge length-based model estimation results from single loop sensors versus the true speed extracted from the double-loop speed traps on the training and the testing set. When the leading edge length-based model applies on both the training and the testing dataset, there is a line formed by data points which is parallel to the true speed trend. It is probably caused by the inaccurate assignment of single loop signature to the pre-determined clusters. Besides, models can capture the true speed accurately.

5.2 Model Comparsion

According to the grouping results, it seems that the leading edge length-based approach increases the goodness of fit of the linear model within each cluster.

However, because of the low variance of the leading edge length among different kinds of vehicles, it is hard for the neural network model to classify single loop signatures into the right groups. Therefore, the optimal group numbers have to be reduced based on the performance of the neural network model. In this case, the final result of the new approach is not as good as expected. Obviously, there is a trade-off between the vehicle grouping model and vehicle classification model. However, with distinct vehicle length between each group, the trade-off problem is automatically solved. On the other hand, the signature lengths have large variation between each group. Compared to the leading edge length, they also have relatively large variations within one group. Therefore, the signature length-based model cannot guarantee a promising estimation result for the traffic flow with large diversity of trucks.

Compared to the signature length, the leading edge length has low variance within one vehicle group. Theoretically, the leading edge method can provide a better result. However, based on the value of MAE and MAPE, overall, the signature length model has better performance than the leading edge length method. Interestingly, the leading edge length method can accurately estimate the speed of the vehicles with lower length.

5.3 Result Analysis

The results indicate that the leading edge lengths of the long vehicles are probably not being correctly captured. For example, in large vehicles, the first peak magnitude occurs in different places which may be caused by the different configurations of tractor parts of large vehicles.

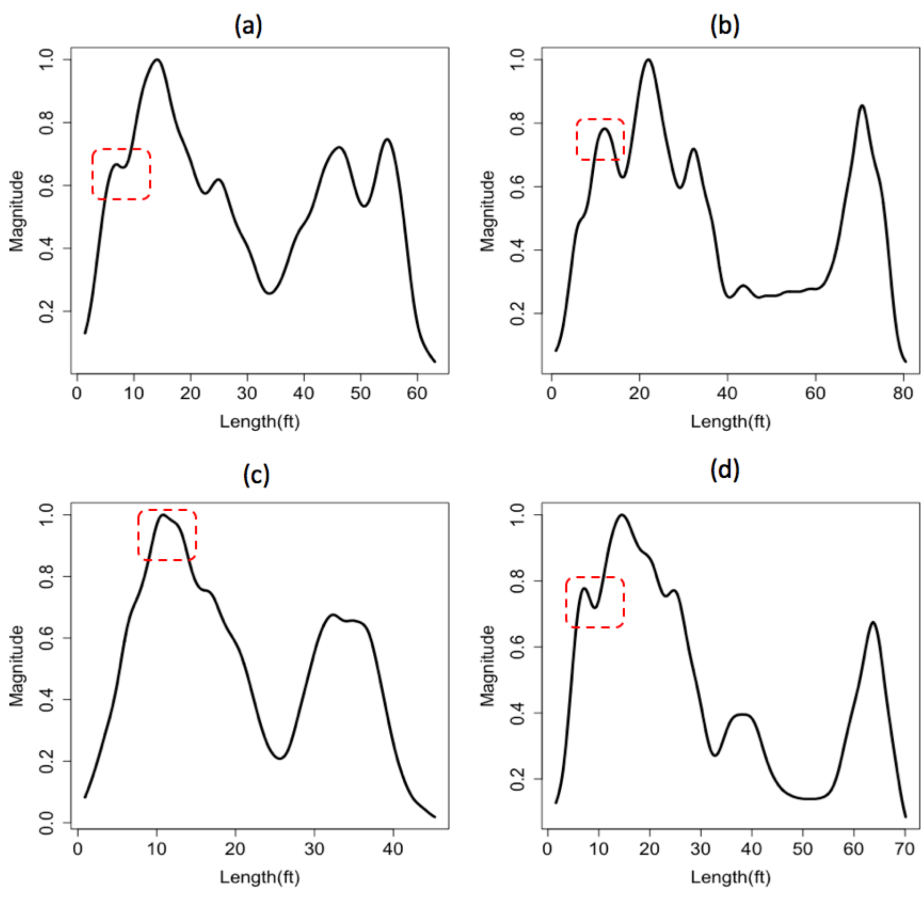

Figure 5.5: Leading Edge Length in Large Vehicles' Signatures

As the figures show, the length related to the first peak magnitude of a signature may be an inaccurate representation of the leading edge length. Taking a physical perspective, when the vehicle first gets onto the loop, the inductive change reduces. After the loop sensor is fully covered by the vehicle, the reduction rate decreases. Therefore, the first peak magnitude of the first derivative of the signature may be a better feature to represent the leading edge length of a signature. (a)(b)(c) in Figure 5.6 represents the first derivative of the signatures from passenger cars. (d)(e)(f) stand for the first derivative of the signatures from large vehicles.

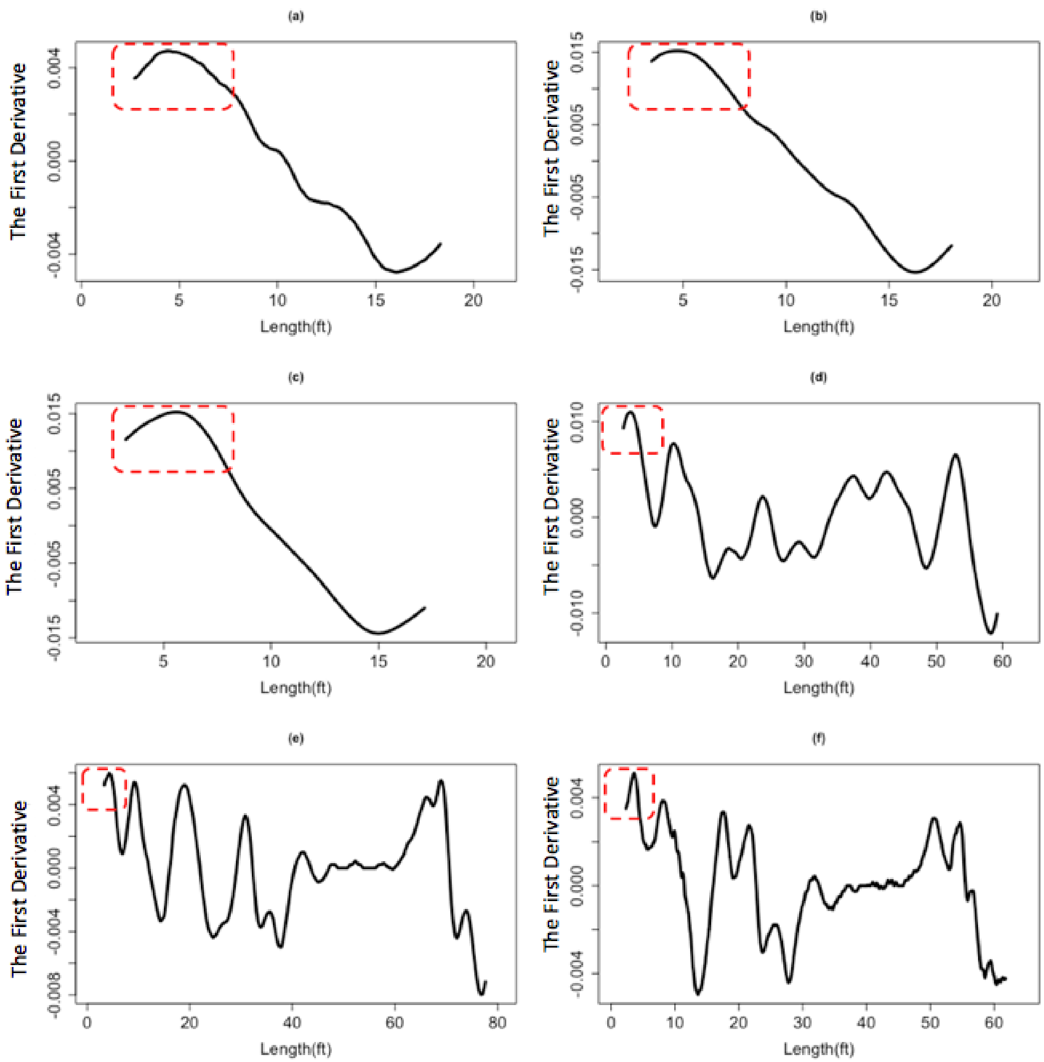

Figure 5.6: Suggested Leading Edge Length

If the leading edge length is correctly captured, the result of the leading edge model will be encouraging.

Chapter 6

Conclusion

6.1 Thesis Summary

This thesis extracted a new feature, leading edge length, from vehicle signatures and developed a new speed estimation model based on the new feature. F-tests were conducted between the new feature and the vehicle signature length within a certain length based vehicle class. The new feature was proved to be a more realiable feature than the signature length for a speed estimation model.

A prior vehicle signature-based model was applied to the new dataset. And the results of these two models were compared. However, the final result of the new speed model was not as good as expected. Compared to the old speed model, the new feature performed well in the vehicle grouping model. Due to the similarity between each cluster in the new model, the number of clusters had to be reduced to increase the accuracy of the new speed estimation model. Unfortunately, based on Figure 5.3 and Figure 5.4, it seems that around 10% of the vehicle have relatively large deviation from the true speed trend.

The slightly weak results of the new model do not mean that the new feature is not a good feature for a speed estimation model. The new speed model can better capture the speed of vehicles which have vehicle length less than 22 ft. This result indicates the potential for the future improvement of the new speed model.

6.2 Future Study

In summary, there are two possible ways to improve the accuracy of the model:

1. Leading edge length indicates that the vehicle first gets onto the loop sensor until it fully covers the loop sensor. Before the vehicle fully covers the inductive loop sensor, the reduction rate of the inductive change will increase. After the loop sensor is fully covered by the vehicle, the reduction rate of the inductive change will reduce. The first peak magnitude is not necessarily the point which can capture the reduction rate of the inductive change. So, if the leading edge length can be captured correctly, the estimated speed provided by the model will be extremely accurate.

2. Since a simple structure of the neural network can not handle the vehicle classification model in the new approach, a multi-layer neural network can be introduced and the deep learning algorithm can be used to cope with the problem.

Bibliography

[1] M. Asudegi. *Optimal number and location of Bluetooth sensors for travel time data collection in networks*. PhD thesis, 2009.

[2] P. Athol. Interdependence of certain operational characteristics within a moving traffic stream. Technical report, 1965.

[3] C. Bachmann, M. J. Roorda, B. Abdulhai, and B. Moshiri. Fusing a bluetooth traffic monitoring system with loop detector data for improved freeway traffic speed estimation. *Journal of Intelligent Transportation Systems*, 17(2):152–164, 2013.

[4] D. Bullock, A. Hainen, J. Wasson, S. Hubbard, S. Remias, and G. Farnsworth. Estimating route choice and travel time reliability using field observations of bluetooth probe vehicles. In *Transportation Research Board 90th Annual Meeting*, 2011.

[5] R. L. Cheu, D.-H. Lee, and C. Xie. An arterial speed estimation model fusing data from stationary and mobile sensors. In *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*, pages 573–578. IEEE, 2001.

[6] B. Coifman. Improved velocity estimation using single loop detectors. *Transportation Research Part A: Policy and Practice*, 35(10):863–880, 2001.

[7] B. Coifman, S. Dhoorjaty, and Z.-H. Lee. Estimating median velocity instead of mean velocity at single loop detectors. *Transportation Research Part C: Emerging Technologies*, 11(3):211–222, 2003.

[8] B. Coifman and S. Kim. Speed estimation and length based vehicle classification from freeway single-loop detectors. *Transportation research part C: emerging technologies*, 17(4):349–364, 2009.

[9] B. Coifman and S. Neelisetty. Improved speed estimation from single-loop detectors with high truck flow. *Journal of Intelligent Transportation Systems*, 18(2):138–148, 2014.

[10] A. Haghani and S. Young. Bluetooth traffic monitoring-technical attributes and applications. In *North American Travel Monitoring Exposition and Conference, ed*, 2010.

[11] J. Kwon, P. Varaiya, and A. Skabardonis. Estimation of truck traffic volume from single loop detectors with lane-to-lane speed correlation. *Transportation Research Record: Journal of the Transportation Research Board*, (1856):106–117, 2003.

[12] Y. Lao, G. Zhang, J. Corey, and Y. Wang. Gaussian mixture model-based speed estimation and vehicle classification using single-loop measurements. *Journal of Intelligent Transportation Systems*, 16(4):184–196, 2012.

[13] C. A. Lave. Speeding, coordination, and the 55 mph limit. *The American Economic Review*, 75(5):1159–1164, 1985.

[14] X.-Y. Lu, P. Varaiya, R. Horowitz, and J. Palen. Faulty loop data analysis/correction and loop fault detection. In *15th World Congress on Intelligent Transport Systems and ITS America's 2008 Annual Meeting*, 2008.

[15] J. C. Muñoz 1 and C. F. Daganzo. Moving bottlenecks: a theory grounded on experimental observation. In *Transportation and Traffic Theory in the 21st Century: Proceedings of the 15th International Symposium on Transportation and Traffic Theory, Adelaide, Australia, 16-18 July 2002*, pages 441–461. Emerald Group Publishing Limited, 2002.

[16] S. Oh, S. Ritchie, and C. Oh. Real-time traffic measurement from single loop inductive signatures. *Transportation Research Record: Journal of the Transportation Research Board*, (1804):98–106, 2002.

[17] S. Park and S. G. Ritchie. Innovative single-loop speed estimation model with advanced loop data. *IET intelligent transport systems*, 4(4):232–243, 2010.

[18] A. Pushkar, F. L. Hall, and J. A. Acha-Daza. Estimation of speeds from single-loop freeway flow and occupancy data using cusp catastrophe theory model. *Transportation research record*, 1457:149, 1994.

[19] S. Ritchie and C. Sun. Individual vehicle speed estimation using single loop inductive waveforms, 1999.

[20] S. G. Ritchie, S. Park, C. Oh, and C. Sun. Field investigation of advanced vehicle reidentification techniques and detector technologies-phase 1. *California Partners for Advanced Transit and Highways (PATH)*, 2002.

[21] K. Teknomo. K-means clustering tutorials. *Medicine*, 100(4):3, 2006.

[22] A. Tok, S. V. Hernandez, and S. G. Ritchie. Accurate individual vehicle speeds from single inductive loop signatures. *Transportation Research Board*, 2009.

[23] T. Vo. *An investigation of bluetooth technology for measuring travel times on arterial roads: a case study on spring street*. PhD thesis, Georgia Institute of Technology, 2011.

[24] Y. Wang and N. L. Nihan. Dynamic estimation of freeway large-truck volumes based on single-loop measurements. In *Intelligent Transportation Systems*, volume 8, pages 133–141. Taylor & Francis, 2004.

[25] J. Wardrop. Some theoretical aspects of road traffic research. In *Inst Civil Engineers Proc London/UK/*, 1900.