IDENTIFICATION OF TRAFFIC PREDICTION PARAMETERS

Anuchit Ratanaparadorn
Department of Industrial Engineering, Kasetsart University, Thailand
b4705091@gmail.com

Sasivimol Meeampol
Faculty of Business, Kasetsart University, Thailand
fbussas@ku.ac.th

Thaneerat Siripachana
Office of the National Boardcasting and Telecommunication Commission

Pornthep Anussornnitisarn
Department of Industrial Engineering, Kasetsart University, Thailand
fengpta@ku.ac.th

Abstract:

Most traffic information we have are real-time, but it is not so useful because we use this information when we decide which way we should go. Imagine we are about to go to work, we look at traffic information and choose the least traffic route. Eventually, traffic might be jam already when we arrive at the road. The problem is real-time information becomes outdated very fast. So how could we fix this? Using Forecasting? It might be good but what factors affects the traffic condition? We should identify it first. There are many factors that could affect the traffic condition. The current and historical traffic condition could be the factors that could predict traffic condition. The ideas are very simple, if the traffic is bad right now, it is more likely that the next five, ten or fifteen minutes later the traffic condition might be the same. Historical traffic condition, we mention is the traffic condition at the same time and same day, for example Traffic condition on this Monday at 8 o'clock might look like the last Monday at the 8 o'clock.

Different day such as weekday, weekend or holiday might have different traffic volume, this could also affects the traffic condition. Time period in the day, for example, rush hours or congestion period might be easier to forecast because the driving speed is limited by the vehicle in front of us. And uncongested period might be harder to forecast because it depends on the drivers to decide how fast they want.

Many researchers found that the driving speed distribution is not only influenced by weather conditions but also by the degree of the roads and the time of the day. However, Bangkok does not have the weather data available on the road; there are only fifty-two stations (plus two airport weather condition) around it. No thorough study has been conducted to determine if the weather data near the road and time of the day combining with traffic volume and traffic flow speed in multiple periods: weekday, weekend, rush hours could increase prediction result. Even though, the prediction accuracy result is not high as expected but it is possible that current and history speed can be used to predict the future traffic speed. We also proved that other parameters might be time period. When we include time period into the model, the first two links and the last link have better correlation but the rest are the same. It might depend on the link, some link may have different traffic condition in different time and some link may have almost the same traffic condition all day. The weekday, weekend, and holiday could help prediction accuracy higher, but their significant statistical correlations are minimal.

Keywords: traffic prediction, traffic condition, traffic prediction accuracy

1. INTRODUCTION

Traffic jam is a well-known problem in all major cities. There have been continuous efforts to improve and mitigate traffic condition; still there are rooms for improvement. Appropriated vehicle routing system in logistics may contribute to reduction in traffic jam and transportation cost. With rising gasoline prices, the need for an effective routing system to help reduce traffic congestion is becoming increasingly important.

In Bangkok, there are two sources of real-time traffic information: Camera Detectors (Closed Circuit Television: CCTV) and Radio-frequency identification (RFID) Detectors. The first one is from Bangkok Metropolitan Administration (BMA) and provided online via National Electronics and Computer Technology Center (NECTEC). The system is based on camera detectors which are all over Bangkok for more than 10,000 stations; however, the only traffic information recorded by these detectors is occupancy level in a form 'H' (traffic jam), 'M' (moderate congestion) and 'L' (low congestion). And they are updated in every five minutes. The occupancy level is calculated by processing image, estimates the traffic density, judges the occupancy level and store in information in 'L', 'M', or 'H'. The accuracy of information is still in doubt because there is no process that can validate it; and the information is too rough for predicting traffic. Normally, every single camera detector needs to be calibrated in order to detect accurate occupancy ratio, traffic volume and traffic flow speed.

Another source of online traffic information is from Burapha Advanced Logistics Labs (BAL-Labs). BAL-Labs are providing the traffic flow speed in five minutes interval. This detector system is initiated by Department of Land Transport cooperate with BAL-Labs for detecting public bus and van speed. Method of gathering information here is the RFID Sensor Network; consisting of four elements: Base Stations, RFID tags, Central Control Center, and Software. RFID readers are installed in the Base Stations. RFID tags are installed in the buses, commuter van, and trucks. Concerns should be taken when analyzing the data close to intersection. To illustrate, if the gather the information around the intersection, the flow speed might be very fluctuate because when it might calculate the speed at time 0 when the light is green and then calculate again at time 5 when the light might be red. The advantage of the bus and van are the steady of information; however, the disadvantage is that they have to stop at every bus stop and is likely to drive slower than private cars.

In this study, we review, select source of real-time traffic, find the relationship between traffic speeds with/without weather data in multiple time periods in order to identify the parameters that could affect the traffic condition.

2. LITERATURE REVIEW

Many forecasting algorithms for traffic forecasting model have been proposed. The simplest one is Historical Average, which is the easiest one in term of implementation but lack of accuracy. ARIMA time series is the popular parametric modeling approach that was used to predict traffic in many researches and return good accuracy. Nonparametric Regression approach is the group identification of past cases that has been developing over the last 30 years. Neural networks, a mathematical modeling approach developed in the field of artificial intelligence, have recently gained significant attention from the transportation research community.

Smith & Demetsky, (1997) proposed four short-term traffic volume forecasting models for the freeway traffic flow forecasting problem at two sites on Northern Virginia's Capital Beltway. The models based on Historical Average, ARIMA Time-Series, Back-Propagation Neural Network, and Nearest Neighbor Nonparametric Regression models. The historical average model simply uses an average of past traffic volumes to forecast future traffic volume. The ARIMA model was developed with five consecutive days of data, with no missing values. By using a statistical software package, an ARIMA(2,1,0) process was identified and then the parameters were estimated. The Back-Propagation Neural Network model was trained with the development database collected at the Telegraph Road location. The model contained 10 hidden processing elements on one hidden layer. The Nearest Neighbor Nonparametric Regression model attempts to identify groups of past cases whose input values, or states, are similar to the state of the system at prediction time with the least Euclidean distance to the input state. Finally, they used Wilcoxon signed-rank test to test the forecasting accuracy for each of models and the result revealed that the Nonparametric Regression model significantly outperformed the other models. Moreover, the Nonparametric Regression model was easy to implement, and proved to be portable, performing well at two distinct sites.

Yamamoto and Morikawa (2004) proposed that the rainfall has effect on the travel speed in a signalized urban surface road network by using probe vehicle data and rainfall data. The data sets were obtained from the central area of Nagoya City, Japan in May and June 2004. The results show that as the rainfall becomes heavier, the average travel speed becomes lower significantly. In the case of heavy rain, the average travel speed decreases 6.03 km/h when compared to the case of no rain. However, the size of the effects varies across the hierarchy of the road, the number of lanes, and the travel speed at no rain. The model also provides more accurate travel time forecasts at signalized urban surface road networks by considering these factors when the weather forecasts are obtained.

Bertini (2009) described an analysis of hourly rain and traffic parameters (speed and flow) at several locations on northbound Interstate 5 in Portland, Oregon using data collected over three years and based on a platform of archived freeway speed, flow, incident and weather data. The study aims to quantify a possible measurable effect of rainfall on traffic speed and flow. First, the connection between rainfall conditions and incidents is studied. It is shown that the presence of incidents slightly influences the analysis of the effects of rain on measured traffic flow. After removing periods affected by incidents, traffic data are examined under different rainfall conditions. During uncongested hourly periods, at the locations analyzed, a significant difference was noted between speed and flow under different rainfall conditions. During congested periods, the flow and speed differences observed were not significant. Further analysis using an hour-by-hour probabilistic approach was performed in order to analyze weather and traffic data during congested time periods. It was found that the probability of having higher speeds and flows in hours with rain was lower than during hours with no rain for congested time periods. Finally, a macroscopic analysis is used to display possible effects of rainfall on the traffic flow fundamental diagram. In the context of intelligent transportation systems, analyses such as these may lead to improved weather-responsive applications in traffic management and information.

3. TRAFFIC DATA SELECTION

CCTV use image processing to calculate occupancy ratio, it is the ratio that tell us how much the road is occupied by the vehicle. The ratio is from 0 to 1; 0 means no vehicle on the road. Unfortunately, the traffic information from this source is only in the form of “H”, “M” and “L” which are very rough information. And it is also discredited by (Jarusakwong, 2009): the information is about 60% inconsistency with people perspective.

RFID provides traffic speed in km/h which is very useful because it is continuous variable. With this source, we could build the models that predict traffic speed. However, the number of vehicle in the system is questionable. A lack of vehicle in the system could lead to less continuity and reliability of the information. Considering that we need at least one vehicle to pass every RFID node within five minutes. The system has 716 nodes, which means that the system needs 716 vehicles in every five minutes. This could lead to great amount of vehicles required by the system, in order to sustain the system continuity. If the system needs reliability, it needs at least twice the continuity volume. Various type of vehicle in the system is another issue. Different vehicle type can lead to more variation of the speed that a driver drive, for example, bus and commuter van cannot drive exceed the speed limit of Land Transport Authority (for safety of commuter) and truck usually drive slower than other type of vehicle.

These two sources of traffic flow speed have its own advantage and disadvantage. The only advantage of the CCTV could be the continuity and reliability. The major disadvantage is the information is categorized data which is too rough to use in the model. The consequence is the result would end up in the form of “H”, “M” and “L” which is stated that it is about 60% inconsistency with people perspective. On the other hand, the RFID has advantage over CCTV in the detailed traffic speed in km/h which is much more useful. Even though, it could have issues on continuity and reliability of the information. Therefore, we decided not to use traffic flow speed from CCTV.

4. DATA ANALYSIS

4.1 Current traffic condition (Hypothesis 1)

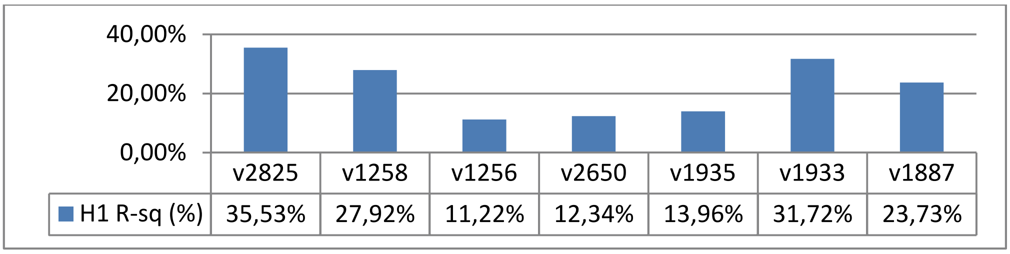
The expected traffic flow speed may affect by the last 5, 10 or 15 minutes traffic speed.

$$v' = a + b_i v_i$$

$$\begin{aligned} v' &= expected\ velocity(km/h) \\ v_i &= velocity(km/h) of\ i\ minutes\ before\ v' \\ a &= constant\ value \\ b_i &= coefficient \\ i &= minutes\ lag; from\ 5,\ 10,\ and\ 15\ minutes \end{aligned}$$

In **Napaka! Vira sklicevanja ni bilo mogoče najti.**1, The R-sq is not high but it is statistically significant in all links. Therefore, we can conclude that current traffic condition affects the expected traffic speed.

Picture 1: Regression Analysis of current traffic condition and expected traffic speed



4.2 Historical traffic condition (Hypothesis 2)

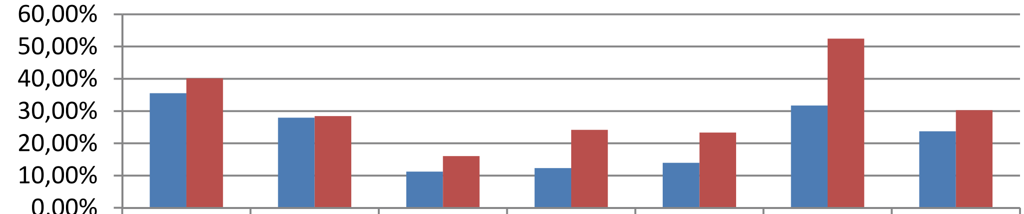
The expected traffic flow speed may affect by the last 5, 10 or 15 minutes and historical traffic flow speed.

$$v' = a + b_i v_i + c_j v_j$$

$$\begin{aligned} v' &= expected\ velocity(km/h) \\ v_i &= velocity(km/h) of\ i\ minutes\ before\ v' \\ v_j &= historical\ v' in\ same\ day\ and\ time(km/h) \\ a &= constant\ value \\ b_i, c_j &= coefficient \\ i &= minutes\ lag; from\ 5,\ 10,\ and\ 15\ minutes \\ j &= week\ lag; from\ 1\ to\ 5\ week \end{aligned}$$

In picture 2, the R-sq of all link are significantly higher than the first hypothesis. Therefore, we can conclude that historical traffic condition significantly affects the expected traffic speed.

Picture 2: Regression Analysis of current traffic condition and historical traffic condition



	v2825	v1258	v1256	v2650	v1935	v1933	v1887
H1 R-sq (%)	35,53%	27,92%	11,22%	12,34%	13,96%	31,72%	23,73%
H2 R-sq (%)	40,15%	28,45%	16,02%	24,18%	23,35%	52,47%	30,26%

4.3 Period of Time: Weekday, Weekend, Holiday (Hypothesis 3)

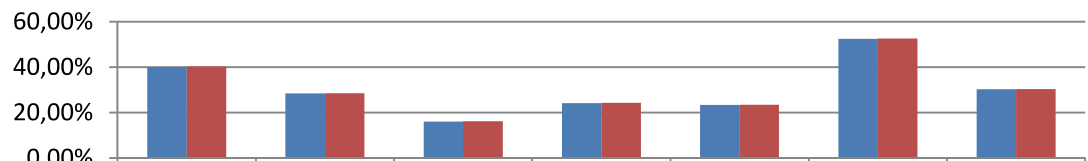
The expected traffic flow speed may affect by the last 5, 10 or 15 minutes and historical traffic flow speed in different day such as weekday, weekend, and holiday.

$$v' = a + b_i v_i + c_j v_j \begin{cases} catetory\ 0: weekday \\ catetory\ 1: weekend \\ catetory\ 2: holiday \end{cases}$$

$$\begin{aligned} v' &= expected\ velocity (km/h) \\ v_i &= velocity (km/h) of\ i\ minutes\ before\ v' \\ v_j &= historical\ v' in\ same\ day\ and\ time (km/h) \\ a &= constant\ value \\ b_i, c_j &= coefficient \\ i &= minutes\ lag; from\ 5,\ 10,\ and\ 15\ minutes \\ j &= week\ lag; from\ 1\ to\ 5\ week \end{aligned}$$

In picture 3, the R-sq of all link are very little higher than the second hypothesis. We can conclude that type of day factor affects the expected traffic speed. However, if we want to implement a vehicle routing system, we could cut off this factor in order to simplify the model.

Picture 3: Regression Analysis of current traffic condition and specific time interval



	v2825	v1258	v1256	v2650	v1935	v1933	v1887
H2 R-sq (%)	40,15%	28,45%	16,02%	24,18%	23,35%	52,47%	30,26%
H3 R-sq (%)	40,35%	28,51%	16,18%	24,28%	23,44%	52,60%	30,38%

4.4 Time in Advance (Hypothesis #4)

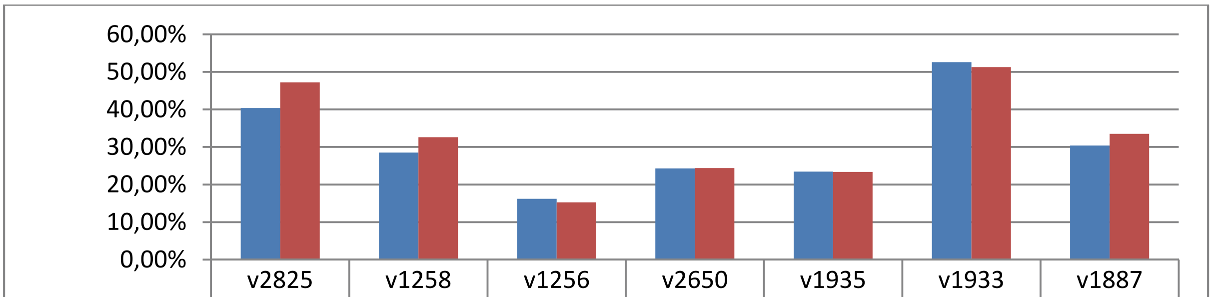
The expected traffic flow speed may affect by the last 5, 10 or 15 minutes and historical traffic flow speed in different time period.

$$v' = a + b_i v_i + c_j v_j : hourly\ catetory\ from\ 6\ am\ to\ 21\ pm$$

$$v' = expected\ velocity(km/h)$$
$$v_i = velocity(km/h) of\ i\ minutes\ before\ v'$$
$$v_j = historical\ v' in\ same\ day\ and\ time(km/h)$$
$$a = constant\ value$$
$$b_i, c_j = coefficient$$
$$i = minutes\ lag; from\ 5,\ 10,\ and\ 15\ minutes$$
$$j = week\ lag; from\ 1\ to\ 5\ week$$

In **Napaka! Vira sklicevanja ni bilo mogoče najti.**, the R-sq of some links is higher and some links is lower than the third hypothesis. We conclude that time period also affect the traffic condition but it depends on the link. When we include time period into the model, the first 2 links and the last link have better r-sq but the rest are the same. It depends on the link, some link may have different traffic condition in different time and some link may have almost the same traffic condition in different time.

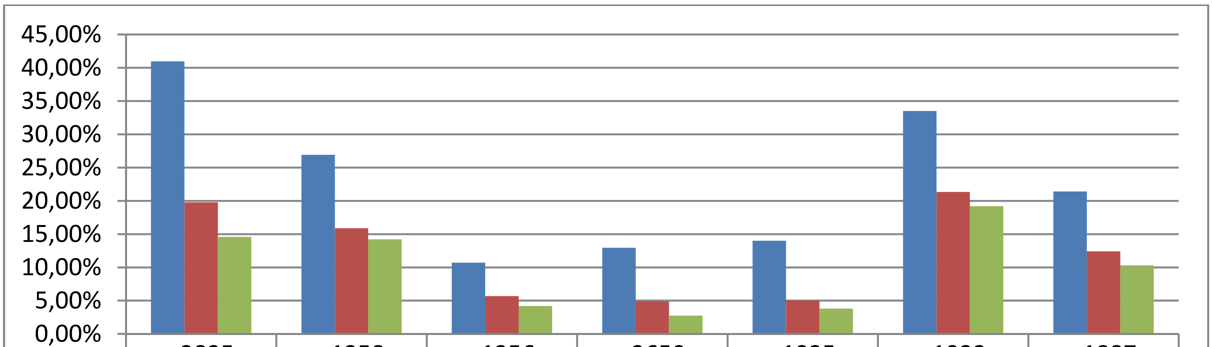
Picture 4: Regression Analysis of current traffic condition and time in advance



	v2825	v1258	v1256	v2650	v1935	v1933	v1887
H3 R-sq (%)	40,35%	28,51%	16,18%	24,28%	23,44%	52,60%	30,38%
H4 R-sq (%)	47,21%	32,60%	15,27%	24,37%	23,34%	51,27%	33,51%

We further analyze the mean of traffic speed by comparing current traffic speed to the last 5, 10, and 15 minutes by using ANOVA. The result is as expected. The all traffic speed at the last 5 minutes has highest R-sq then lower at 10 minutes and lowest at 15 minutes. This means that forecasting traffic by using traffic speed at the last 5 minute is more accurate than using 15 minute.

Picture 6: ANOVA of current traffic speed with speed from the last 5, 10, 15 minutes

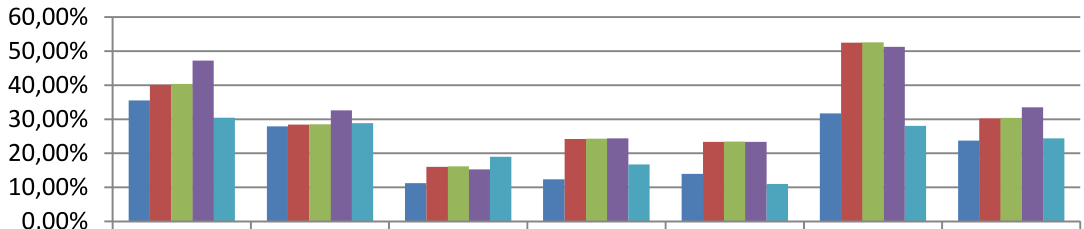


	v2825	v1258	v1256	v2650	v1935	v1933	v1887
x5	40,97%	26,92%	10,70%	12,94%	14,01%	33,50%	21,41%
x10	19,77%	15,89%	5,69%	4,88%	5,04%	21,35%	12,41%
x15	14,57%	14,22%	4,19%	2,76%	3,82%	19,19%	10,31%

5. CONCLUSION AND RECOMMENDATION

In **Napaka! Vira sklicevanja ni bilo mogoče najti.**7, the parameters that can certainly predict the traffic condition are the current and historical traffic condition. Although, the weekday, weekend, and holiday could help prediction accuracy higher, still they are just a little bit. When we include time period into the model, the first 2 links and the last link have better r-sq but the rest are the same. This may depends on the link, some link may have different traffic condition in different time and some link may have almost the same traffic condition in different time. We conclude that the weather data near the link is not the factor that helps increase the prediction accuracy.

Picture 7:



	v2825	v1258	v1256	v2650	v1935	v1933	v1887
H1 R-sq (%)	35,53%	27,92%	11,22%	12,34%	13,96%	31,72%	23,73%
H2 R-sq (%)	40,15%	28,45%	16,02%	24,18%	23,35%	52,47%	30,26%
H3 R-sq (%)	40,35%	28,51%	16,18%	24,28%	23,44%	52,60%	30,38%
H4 R-sq (%)	47,21%	32,60%	15,27%	24,37%	23,34%	51,27%	33,51%
H5 R-sq (%)	30,47%	28,87%	18,99%	16,73%	10,98%	28,06%	24,41%

Even though, the prediction accuracy is not high as expected but it is possible that current and history speed can be used to predict the future traffic speed. We also proved that other parameters might be time period. When we include time period into the model, the first 2 links and the last link have better r-sq but the rest are the same. It might depend on the link, some link may have different traffic condition in different time and some link may have almost the same traffic condition all day. The weekday, weekend, and holiday could help prediction accuracy higher, still they are just a little bit.

During development of the model, we encountered many difficulties when we try to include traffic information collected from BAL-Labs into the model. The information contains heavily traffic information outlier from 10 – 50% and missing from 6 – 11%. (See section **Napaka! Vira sklicevanja ni bilo mogoče najti.** for more detail). Bangkok Metropolitan Administration should collect more traffic information. Only 'H', 'M', and 'L' is inadequate. CCTV could be powerful source of traffic variable such as traffic speed and traffic volume (Kwon & Park). Many authors conclude that the rainfall condition affect the vehicle speed (Wang, Yamamoto, Miwa, & Morikawa, 2004), (Do, Oh, & Kim, 2003) and (Bertini & K. , 2009),. Weather probe on highway could be another key that help increase traffic prediction. And if we have more reliable traffic information, we could build an effective vehicle routing system that could reduce transportation cost.

REFERENCE LIST

1. Bertini, R. L., & K. , M. S. (2009). Empirical Analysis of the Effects of Rain on Measured Freeway Traffic Parameters. Transportation Research Board.
2. Boto-Giralda, D., Diaz-Pernas, F. J., Gonzalez-Ortega, D., Diez-Higuera, J. F., Anton-Rodriguez, 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, (pp. 530-545).
3. Charusakwong, N. (2010). Evaluation of traffic indicators for intelligent traffic signs from motorists' perspectives.
4. Do, M., Oh, J., & Kim, M. (2003). Effects of Weather Conditions on Free-flow Speeds. Eastern Asia Society for Transportation Studies.
5. Jarusakwong, N. (2009). Evaluation of traffic indicators for intelligent traffic signs from motorists' perspectives.
6. Kamarianakis, Y., & Prastacos, P. (2003). Forecasting Traffic Flow Conditions in an Urban Network: Comparison of Multivariate and Univariate Approaches. Transportation Research Record: Journal of the Transportation Research Board, 74-84.
7. Lee, S.-H., Lee, B.-W., & Yang, Y.-K. (2006). Estimation of Link Speed Using Pattern Classification of GPS Probe Car Data. Computational Science and Its Applications, 495-504.
8. Li, Z.-P., YU, H., Liu Y.-C., & Liu, F.-Q. (2008). An Improved Adaptive Exponential Smoothing Model for Short-term Travel Time Forecasting of Urban Arterial Street. Acta Automatica Sinica, 1404-1409.
9. Liu, B.-s., Li, Y.-j., Yang, H.-t., Sui, X.-s., & Niu, D.-f. (2006). Research on Forecasting Model in Short Term Traffic Flow Based on Data Mining Technology. International Conference on Intelligent Systems Design and Applications. IEEE.
10. Smith, B. L., Williams, B. M., & Oswald, R. K. (1999). Comparision of parametric and nonparametric models for traffic flow forecasting. Transportation Research, 303-321.
11. Wang, L., Yamamoto, T., Miwa, T., & Morikawa, T. (2004). An Analysis of Effects of Rainfall on Travel Speed at Signalized Surface Road Network Based on Probe Vehicle Data