❐ 333

Bus Arrival Prediction – to Ensure Users not to Miss the Bus

Lutfi Fanani*[#], Achmad Basuki*, Deron Liang[#]
* Departement of Electrical Engineering, University of Brawijaya
[#] Department of Computer Science and Information Engineering, National Central University

Article Info	ABSTRACT
	Predicting arrival times of buses is a key challenge in the context of building intelligent public transportation systems. The bus arrival time is the primary information for providing passengers with an accurate information system that can reduce passenger waiting times. In this paper, we used the normal distribution method to the random of travel times data in a bus line number 243 in Taipei area. In developing the models, data were collected from Taipei Bus Company. A normal distribution method used for predicting the bus arrival time in bus stop to ensure users not to miss the bus, and compare the result with the existing application. The result of our experiment showed that our proposed method has a better prediction than existing application, with the probability user not to miss the bus in peak time is 93% and in normal time is 85%, greater than from the existing application with the 65% probability in peak time, and 70% in normal time.

Corresponding Author:

Lutfi Fanani,
Department of Electrical Engineering,
University of Brawijaya,
Jl. Veteran, Malang 65145, East Java, Indonesia.
Department of Computer Science and Information Engineering,
National Central University,
No. 300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan, ROC.
Email: lutfifanani@yahoo.com

1. INTRODUCTION

A good public transport is of increasingly to maintain and improve quality of life by providing mobility and accessibility. Any traveler who wants to travel between two places will have to take a number of decisions such as mode of transportation, route, travel time and so on [4].

When travelling with buses, the travelers usually want to know the accurate travel times of the bus. Excessive long waiting times at the bus stops may drive away the anxious travelers and make them reluctant to use the bus. Nowadays, most bus operating companies have been providing their timetable for the travelers [3]. To estimate the bus arrival times at stops is a challenging task, because bus travel times (from a real-time observation position) to a specified stop depend on a number of factors (e.g., delays of intersections, dwell times at stops, travel times on links, etc.), and fluctuate spatially and temporally [1].

Travel time information is the most preferred information by travelers [14], [15]. However, this information can not be measured directly. As has been explained on [16], the provision of timely and accurate transit travel time information is important because it attracts additional ridership and increases the satisfaction of transit users, which will ultimately result in a decrease in congestion.

In the previous work, the Taipei Bus Company Route Planner Application has a feature of the bus arrival prediction time. This application feature used a prediction time data from Taipei Bus Company API (Application Programming Interface). Google application also has a prediction feature, using data that they are provided. There are some conditions when the predicted result is too fast than actual bus arrival time, in this case users should wait a few minutes before the bus arrived. It also has the condition when prediction

result is too slow than actual bus arrival time, which makes users missed the bus and they should wait for the next bus that stops about 15 – 20 minutes in peak hour and 30 – 40 minutes in normal hour [6]. This condition can make users loss of a lot of their precious time and expense. The proposed system aiming to predict the bus arrival time to ensure users not to miss the bus. This system proposed the Normal Distribution method using random travel time variable, to make the arrival time prediction.

A variety of the prediction model to predict the bus arrival time has been studied by many in recent years. The tree most widely used models include historical data based model, regression model, and time series model.

Historical data based models predict travel time for a given time period using the average travel time for the same time period obtained from a historical database. These models assume that traffic patterns are cyclical and the ratio of the historical travel time on a specific link to the current travel time reported in real-time will remain constant [1]. The procedure requires an extensive set of historical data and it is difficult to install the system in a new setting [1]. Real-time models assume that the most recently observed transit travel times will stay consistent in the future.

The regression model is conventional approaches for predicting the arrival time. These models predict and explain a dependent variable with a mathematical function formed by a set independent variables [2]. To establish the regression models, the dependent variables need to be an independent. Mishalani, et.al. [5] proposed a multi linear regression to predict the bus arrival times using the data collected by Automatic Passenger Counter (APC). They used distance, number of stops, dwell times, boarding and alighting passengers and weather descriptors as independent variables. However, this approach is reliable when the dependent variable has a correlation with the independent variable. In this case we have tried to implement this method in our experiment, but the data that we have can't be processed using this method because of no correlation between the dependent variable (time travel) with the independent variable (distance). The farther distance there should have been the longer travel time, but the existing data are different. We have the different distances with same travel time, for example, in 500 and 1500 meters it has around 50 second to 100 second of travel time. In this experiment distance can't be used to be an independent variable. This is caused by the traffic condition, speed of bus, and other parameters that included in this line.

Time series models assume that the historical traffic patterns will remain the same in the future. The accuracy of time series models is a function of the similarity between the real-time and historical traffic patterns [2]. Variation in historical data or changes in the relationship between historical data and real-time data could significantly cause inaccuracy in the prediction results [1]. They used a non-linear time series model to predict a corridor travel time on a highway [1]. He compared two cases: the first model used only speed data as a variable, while the second model used speed, occupancy, and volume data to predict travel time. It was found that the single variable model using speed was better than the multivariable prediction model. We have the bus speed data from Taipei Bus Company API, but we could not use this data because is not valid, the value of bus data speed always below 10 km/h. And we don't have a volume traffic data on the road to be a variable in this time series model.

2. RESEARCH METHOD

The experiment is divided by two time condition, in peak hour and normal hour. We choose one bus stop in one line to make an experiment and testing, based on stops that have the most error prediction, in this case is the longest waiting time. Normal distribution method with random variable travel time has been chosen to make a better prediction/ estimation of travel time between each bus stop, and compare the result with the existing application.

2.1. Normal Distribution Method

In probability theory, the normal (or Gaussian) distribution is a very commonly occurring continuous probability distribution – a function that tells the probability that any real observation will fall between any two real limits or real numbers, as the curve approaches zero on either side [8]. In the case of a single real-valued variable x, the Gaussian distribution is defined by [11]:

$$N(x; \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \tag{1}$$

Which is governed by two parameters: μ, called the mean, and σ^2, called the variance. The square root of the variance, given by σ, is called the standard deviation and the reciprocal of the variance, written as $\beta=1/\sigma^2$, is called the precision [11]. Figure 1 shows a plot of normal (Gaussian) distribution.

The probability of the normal as given above is difficult to work with in determining areas under the curve, and each set of X values generates another curve as long as the means and standard deviations are translated to a new axis, a Z -axis, with the translation defined as:

$$Z = \frac{X - \mu}{\sigma}$$ (2)

The Gaussian distribution arises in many different contexts and can be motivated from a variety of different perspectives. For example, we have already seen that for a single real variable, the distribution that maximizes the entropy is the Gaussian [11]. This property applies also to the multivariate Gaussian. The normal distribution has an infinite support for many applications. In reality, we can not get really an infinite support. Some authors used and implement a truncated normal distribution to their research. [A. Kuang-17] develop a bi-level programming model based on the OD pair travel time reliability to study the road network capacity reliability with the assumption that the link capacity follows a truncated normal distribution. For next experiment, researchers can use the truncated normal distribution to solve this travel time problem.

2.2. Experiment Design

Figure 1 shows the experiment design in this research. We construct the model by using data from Taipei Bus Company. Collecting training data needed for this experiment to make the dataset that we use for make the prediction system.
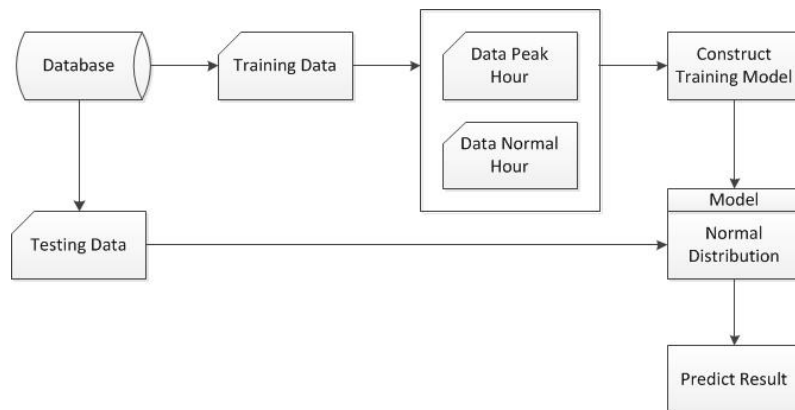

Figure 1. Experiment system design

2.3. Data Collecting

The data used for this study were collected from bus route number 243 in Taipei area. The route length is approximately 14 km and 53 minutes of travel time between the source and destination, and spanning 25 bus stops in each direction. The route starts at Zhonghe Station located in Jinhe Road and stop in Zhonghua Road near Taipei Museum Ximen. The schedule of bus operation start at 5.30 AM until 9.00 PM on the weekday and start at 5.30 AM until 6.00 PM at the weekend, and the time interval between each bus divided by two conditions, in the peak hour: 15 – 20 minute and in the normal hour: 30 – 40 minute [6]. Figure 2 shows the map of bus route number 243 in Taipei with the blue color in all bus stops.

Data were collected using Adam's Automatic Data Collector (ADC) System [12]. Data collecting process showed in Figure 3. In the system we insert data from Taipei Bus Company API to our system. The Taipei Bus Company provided the Bus Location data with a GPS technology for each bus on the route. The estimation time and each time the bus stopped, the bus location was recorded using the GPS receiver and sends the data to the database that we collected for this study.
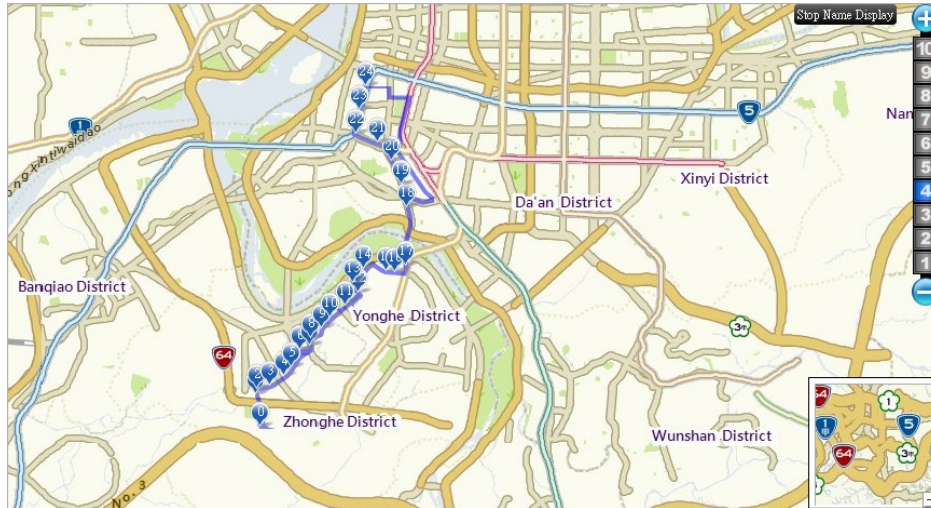
Figure 2. Bus line number 243 in Taipei [6]

Arrival and departure time records at each stop are the most important ones for predicting the travel time. The data were collected in June 2014 from bus route number 243 in Taipei area. The data collection scheme is presented in Figure 3 below. Bus route number 243 has been chosen for the case study because it has a longest time interval between each bus in Taipei area (the time interval has been explained before). With the longest time interval, the waiting time at the stop will be longer.

So, our focus is to improve the prediction time in this line to ensure users that they will not miss the bus, because if they missed the bus, waiting time for the next bus is very long, about 20 – 40 minutes in different time condition.
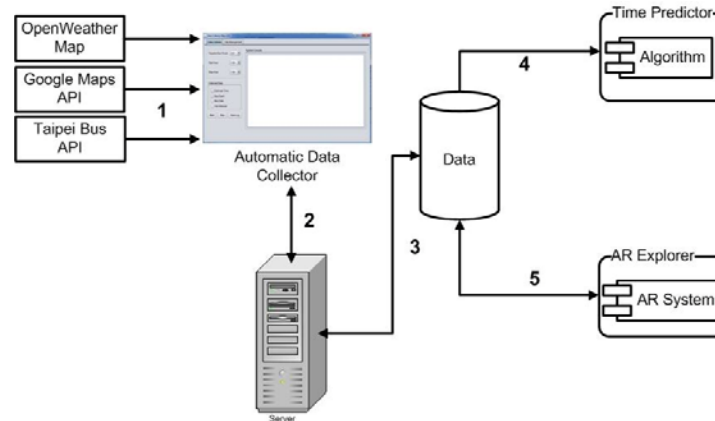

Figure 3. Data collecting scheme [12]

In this data collection we have separated based on time condition, in the peak hour and normal hour. Because the travel time on both conditions is different. The data were collected in June 2, 2014 to June 9, 2014 in weekday.

2.4. Preliminary Analysis

Taipei Bus Company application is Taiwan public transit system that has been developed to help users for checking bus line, times, and routes. The bus system operated under cooperation between 15 private agencies, so translation is not always consistent. It is recommended to always keep a Chinese written version of your destination for comparison [7].

In this paper we used some terms to explain some condition in predicting a bus arrival time. First is Travel time, is defined as pure running time in sections, does not contain short delayed time because of traffic

signal control, the time for passengers getting on and off at each stop station and the stop time for vehicle technical problems [10]. The travel time on sections from the bus's current running location of the prediction stop station is calculated from arrival time (*AT*) data in stop S_{i+1} minus stop S_i.

$$T_i = (AT_i + 1) - AT \qquad (3)$$

Where Ti denotes the travel time from Si to stop S; AT denotes the arrival time in Stop i.

Second is waiting time occurs when waiting for a bus at a bus stop, when the duration of the wait may exceed the time needed to arrive at a destination by another means [5]. It will be worst if waiting time in bus stop is more than 15 minutes. In bus arrival prediction system, if prediction showed the wrong value, for example prediction is too slower than the bus actual arrived 1 minute or 5 minutes it means users must to wait for the next bus.

The next is rush hour or peak hour, it is a part of the day during which traffic congestion on roads and crowding on public transport is at its highest. Normally, this happens twice a day-once in the morning and once in the evening, the times during when the most people commute. The term is very broad but often refers to specifically private automobile transportation traffic, even when there is a large volume of cars on a road, but not a large number of people, or if the volume is normal but there is some disruption of speed [9]. In this experiment we decided into peak hour and normal hour condition. In normal condition the probability user missed the bus is greater than in peak hour. We can determine the peak hour and normal hour based on data that we have collected for this experiment. From the data the condition in peak hour is from 7.00 AM – 10.00 AM and from 3.00 PM – 6.00 PM. Normal hour is from 11.00 AM – 2.00 PM and from 7.00 PM – 9.00 PM

3.   RESULTS AND ANALYSIS

This is a preliminary study, and bus line 243 has been chosen to be an experiment testing because that has a longest waiting time. In this experiment testing bus stop number 14 has a longest waiting time, with the average waiting time is 10.2 minutes. We make a testing and comparison with the existing application from stop 7 to 14 with 7 bus stops (S_7,S_{14}), stop 8 to 14 with 6 bus stops (S_8,S_{14}), stop 9 to 14 with 5 bus stops (S_9,S_{14}), stop 10 to 14 with 4 bus stops (S_{10},S_{14}), stop 11 to 14 with 3 bus stops (S_{11},S_{14}), stop 12 to 14 with 2 bus stops (S_{12},S_{14}) in peak hour and normal hour. Testing was conducted 10 times in different time condition at each stop to know how many probabilities that user do not miss the bus, and compare the result with the existing application. Figure 4 shows the average waiting time in all bus stops in bus line 243. In this part, we would like to see the comparison result of existing application and our proposed method from experiment above. The comparison result of user average waiting time are presented in Figure 5. And the result of probability users not to miss the bus during peak hour and normal hour are presented in Figure 6.
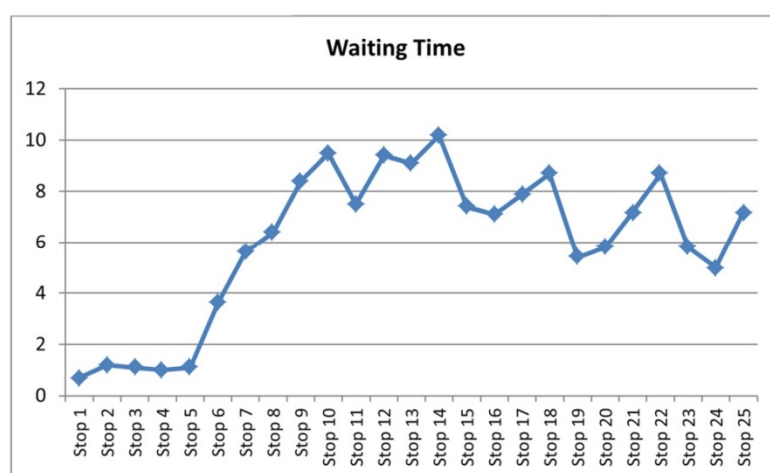

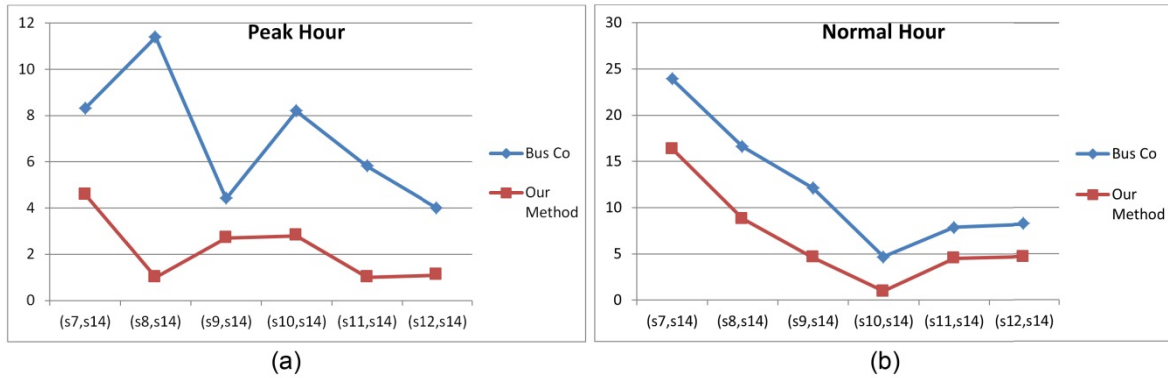Figure 4. Average waiting time in bus line 243

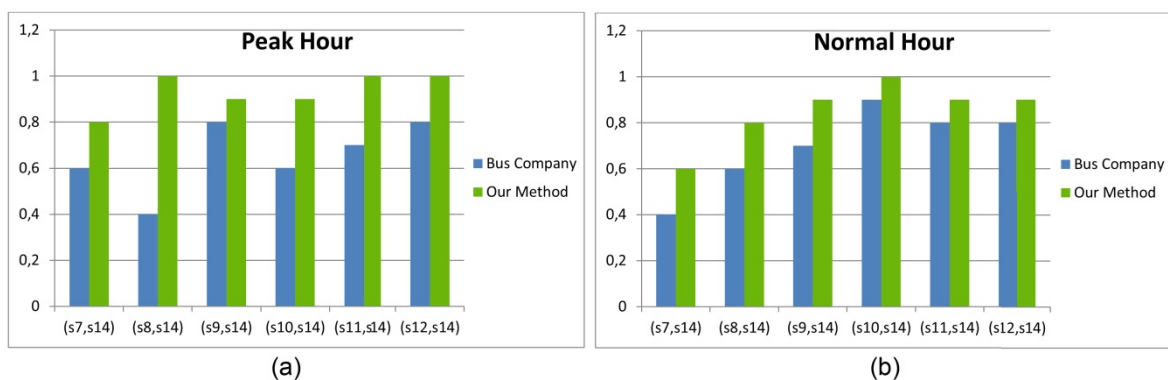Figure 5. Average users waiting time

Figure 6. Comparison of probability users not to miss the bus

From our experiment result, we observe that our proposed method is better than the existing application (Taipei Bus Company) in predicting the arrival time to ensure users not to miss the bus. We have some different arrival prediction time in peak hour and normal hour because the number of passenger and traffic condition in peak hour (a) and normal hour (b) are different. There are two conditions on average users waiting time and the probability users not to miss the bus, peak hour (a) and normal hour (b) condition. The users waiting time from our proposed method have a less waiting time than bus company (existing application), and the probability of users not to miss the bus is greater than existing application.

3.1. Discussion

Our experiment design is divided by time condition, peak hour and normal hour. Peak hour and normal hour condition obtained from data that provided by Taipei Bus API. The aims of divided the time condition to make a prediction model is to get a more accurate prediction result, because the condition of road in peak hour and normal hour is different.

3.1.1. Weekend Condition

To obtain more accurate results, we can add to the conditions of the time in weekday and weekend. Because the condition in 7.00 – 8.00 PM at the weekday is different in weekend, in weekday we considered that time is normal time, but at the weekend it can be a peak time situation. In this experiment we used weekday time condition.

3.1.2. Dwell Time

In order to predict travel time, in an accurate and timely manner, the consideration of traffic condition, including traffic congestion, dwell time at stops, etc. The bus dwell time at a bus stop is defined as the time spent by a bus at the bus stop for passenger alighting and boarding, including time of opening and closing bus doors [13]. The bus dwell time is of great importance to estimate the capacity of a bus station, and it is also a major component of bus travel time. An Automatic Passenger Data needed to predict the dwell

time at a bus stop. Taipei Bus API did not provide us an APC data. We can use APC data at bus stops to predict bus arrival time with dwell time to make more accurate prediction result.

4.   CONCLUSION

In this paper, we dealt with the problem of bus arrival time prediction in the existing applications. Sometimes the prediction of bus arrival time in the existing application is too late and sometimes is too early. This condition can cause users to miss the bus and waiting for a long time for the next bus. We proposed a normal distribution method to predict the best prediction of bus arrival time to ensure users not to miss the bus. This proposed method is based on random variable travel time data between bus stops in a preliminary study based on bus line number 243 in Taipei. We performed experiments on two types of time condition, in peak hours and normal hours. Because the condition in peak hours and normal hours are different, the travel times are different too. In this experiment, we compared the probability user not to miss the bus from our proposed method with the existing application. From our experiment, we concluded that both of time condition showed the better result than the existing application to give a probability to ensure users not to miss the bus. With the probability average in peak hour is 93% than 65%, and in normal hour is 85% than 70% of an existing application.

REFERENCES

[1]   Cheng, Shaowu. Liu, Baoyi. "Bus Arrival Time Prediction Model Based on APC Data". *The sixth advantage forum on Transportation of China. China*. 2010.
[2]   Chien, S.I.J., Ding, Y., and Wei, C. "Dynamic Bus Arrival Time Prediction with Artificial Neural Networks". *Journal of Transportation Engineering*. 2002.
[3]   Zhou, Pengfei, Zheng, Yuanqing. "How Long to Wait?: Predicting Bus Arrival Time with Mobile Phone Based Participatory Sensing". Nanyang Technological University Publications. Singapore. 2012.
[4]   Litman, Todd. "Evaluating Public Transit as an Energy Conservation and Emission Reduction Strategy". *Victoria Transport Policy Institute Publication*. 2012.
[5]   Mishalani, Rabi, McCord, Mark. "Passenger Waiting Time Perception at Bus Stop: Empirical Result and Impact on Evaluating Real Time Bus Arrival Information". *Journal of Publict Transportation*. 2006.
[6]   Taipei City Public Transportation Office. "Taipei E-Bus System". *Available at: http://www.e-bus.taipei.gov.tw/new/english/en_index_6_1.aspx*
[7]   165284 Taipei Bus Information and Transit System. *Available at: http://www.5284.com.tw/Dybus.aspx?Lang=En*
[8]   Wikipedia. "Normal Distribution". *Available at: http://en.wikipedia.org/wiki/Normal_distribution*
[9]   Wikipedia. "Rush Hour". *Available at: http://en.wikipedia.org/wiki/Rush_hour*
[10] Zang, Jian, Yan, Ling. "Study of the Prediction Model of Bus Arrival Time". *IEEE Management and Service Science Conference*. 2009.
[11] Bishop, Christoper M. "Patern Recognition and Machine Learning". *Springer Science and Business Media. Singapore*. 2006.
[12] Hendra Brata, Adam. "Thesis: Software Development of Automatic Data Collector". *National Central University. Taiwan*. 2014.
[13] Meng, Qiang, Qu, Xiaobo. "Bus Dwell Time Estimation at Bus Bays: A Probabilistic Approach". *National University of Singapore*. 2001.
[14] L. Vanajakshi, S.C. Subramanian, and R. Sivanandan. "Travel time prediction under heterogeneous traffic conditions using global positioning system data from buses", *IET Intell. Transp. Syst. 3,*. 2009
[15] Pu W., J. Lin "Urban Street Travel Time Prediction Using Real Time Bus Tracker Data". *Transport Chicago 2008, Chicago, IL*, 2008.
[16] Jeong R.H. "The Prediction of Bus Arrival time Using Automatic Vehicle LocationSystems Data", *A Ph.D. Dissertation at Texas A&M University*. 2004
[17] A. Kuang, Z. Tang, L. Shan. "Road network capacity reliability considering travel time reliability". *Procedia Social and Behavioral Science*. 2013