# Advanced Traffic Prediction System by Socio-Technical Sensor Fusion using Machine Learning

Swe Swe Aung[1], Shiro Tamaki[2] and Itaru Nagayama[3]
[1] Department of Information Engineering, University of the Ryukyu
1 Senabru, Nishihara, Okinawa 903-0213, Japan
[2] Department of Information Engineering, University of the Ryukyu
1 Senabru, Nishihara, Okinawa 903-0213, Japan
[3] Department of Information Engineering, University of the Ryukyu
E-mail: [1] sweswe@ie.u-ryukyu.ac.jp , [2] shiro@ie.u-ryukyu.ac.jp , [3] nagayama@ie.u-ryukyu.ac.jp

**Abstract:** Nowadays, as traffic jam is an everyday facing problem in the developed and developing countries, monitoring, predicting and detection current traffic condition systems are playing an important role in research fields. For this case, many researchers have been trying to solve this problem by using many kinds of technologies, especially road side sensors. However, these sensors have already injected with the good and bad things altogether. Therefore, it may not be a better way to use it in stand alone for traffic prediction system. For this issue, the paper mainly focus on predicting traffic condition based on multiple points of view such as the data from road side camera, weather condition, weekday or weekend, rush hour time and special day. The system has three parts: Resources Side (RS), Traffic Prediction Server (TPS) and Display Side (DS).

*Keywords*-- **Display Side (DS), Traffic Prediction Server (TPS), Resources Side (RS)**

## 1. Introduction

Traffic congestion is not an unusual problem however it is a big everyday facing trouble to everyone in their daily life functions as they rely on the road network transportation. In consequently, it can be seen that human has been destroyed gradually to be the one who is impatient. From the perspective of healthcare, everyday facing traffic jams is very similar to a kind of diseases which is eroding human's healthiness. For that reason, many researchers have been seeking many different kinds of issues and correspondences brilliant solutions aiming at reducing road traffic jams sooner or later.

For the sake of road congestion issues, it is obvious that monitoring road is critical to city life throughout the world. There are many kinds of sensors designed to collect different types of traffic data. These traffic sensor types are loop detectors, radar, camera, license plate readers, global positioning system (GPS) and so on.

Inductive-loop detectors are widely used at intersections with traffic-actuated signals, freeway entrance with automatic ramp metering, highway segments monitored by traffic counting programs, and entrances of gated parking facilities. They are built into the roadway so that they can detect each vehicle that passes over them. A Loop detector is able to provide high accuracy data.

When placed directly above a lane of traffic, license plate readers are capable of automatically extracting the numbers and letters from passing vehicles. When multiple readers are setup at two points along the road, it is possible to extract travel time information for vehicles passing both locations [4].

The fundamental challenge in using cell tower information for estimating position and motion of vehicles is the inherent inaccuracy of the method, which poses significant difficulties to the computation of speed. Several solutions have been implemented to circumvent this difficulty, in particular by the company Airsage, which historically developed its traffic monitoring infrastructure based on cell tower information [5].

The use of video cameras (many of which are already installed to survey road networks), coupled with computer vision techniques, offers an attractive alternative to other sensors before. Video-based camera was mainly used to detect moving object, including moving vehicle in Traffic Monitoring System. Video-based camera systems are more sophisticated and powerful because the information content associated with image sequences allows precise vehicle tracking and classification. [6]

GPS enable mobile phone sensors are enabling new applications across a wide variety of domains, such as healthcare, social networks, safety environmental monitoring, and transportation, traffic detection system, and give arise to a new area of research called mobile phone sensing. However, this sensor introduces many accuracy problems [3].

These sensors play an essential role in Intelligent Transportation System, especially traffic prediction system. This paper mainly focused on utilizing road side camera in the side of getting traffic data.

The rest of this paper is organized as follows. Section 2 describes related works. In section 3 presents traffic prediction system architecture. In section 4 describes analytical model and experimental environment presented in section 3. Finally, section 4 is the conclusion and future work.

## 2. Related Works

The video stream processing for traffic congestion is essential one in controlling the city traffic. [1] P.Niksaz proposed a system that estimated the size of traffic in highways by identifying the number of cars utilized one of image processing techniques, background subtraction. Before this task, each frame was compared with the first frame and then if the number of cars was more than a threshold, it assumed that there was a traffic. However, the system predicted the traffic congestion by considering only does not consider any other affection that can cause traffic. D. Rosenbaum et al. [2] designed a system which extracted traffic data from road side camera by a priori knowledge from a road database of the approximation location of road axes in the georeferenced and orthorectified images. Firstly, vehicle detection was executed and then velocity was obtained by applying vehicle tracking technique.

## 3. Naïve Bayes Classifier based Traffic Prediction System

Figure 1 illustrates the overall architecture of the system. The system is composed of three main parts, Resources Side, Traffic Prediction Server (TPS) and Display Side. RS has the responsibility of tracking traffic data, filtering the raw data into usable data and then transmitting it to the prediction section, TPS. Resource Side has two types of resources that are road side camera and weather station. Traffic Prediction Server plays a vital role in predicting the current traffic condition. What its responsibility is to the integrate history data and the current input data from Resource Side by applying Naive Bayes Classifier, using m-estimate of probabilities. After that, it sends the final result to Display Side as well as stores it in the traffic database for future prediction.

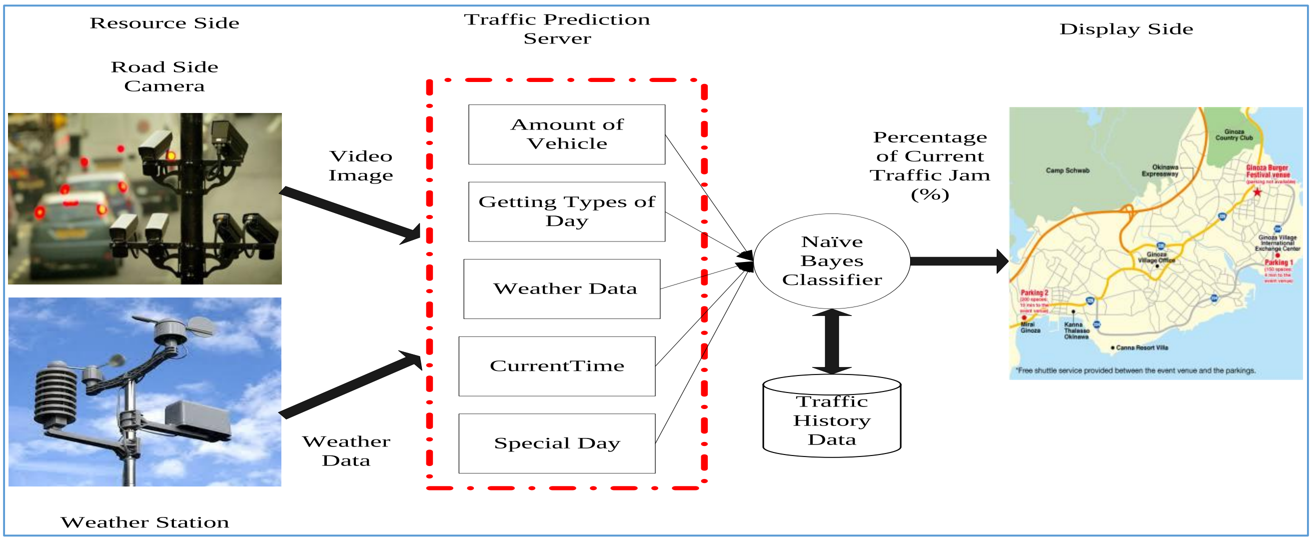


**Figure 1. Architecture of Traffic Prediction System**

This paper assumes that it has already receipted the data from the road side camera and the weather station, and does not emphasize on the detail calculation of image processing. In Figure 1, it can be seen that Server Side has to get the data from road side camera and weather station, and to analyze a day whether it is weekday or weekend or a special day. There are many situations that traffic jam can happen. Among them, the bad weather condition is most likely to occur traffic jam. Besides, it needs to consider about the special days. On these day, traffic jam can be occurred due to a large number of people going outside. For example Naha Festival on that day the roads related with this area are almost blocked by cars, which makes traffic jam terribly. From my experience, traffic congestion is more likely to be faced in weekday than in weekend. These traffic attributes described above badly affect having almost non-stop traffic jam in downtown area in different way.

Table 1 illustrates the attributes and their related values for traffic prediction system. The input attributes are ***AmountofVehicle***, ***Specialday***, ***WeatherCondition***, ***CurrentTime*** and ***TypesofDay.*** The output attribute is ***Traffic***. ***AmountofVehicle*** means the total amount of vehicle running on a road and it has three values, High, Medium and Low. Moreover, it also has to get the live weather data from the weather station(***WeatherCondition***) with the values, 1, 2 and 3. In this case, 1 means sunny, 2 cloudy and 3 raining. Besides, it analyses the current time too if it is rush hour time or not (***CurrentTime***). It has ***Hight, Medium and Low.*** Furthermore, it looks back about the current day whether it is a weekend or weekday(***TypesofDay***). In the rush hour times, the traffic jam is usually high or medium. The rest time is low.

**Table 1. Attribute Value Table**

| No | | Attribute Name | Value |
|---|---|---|---|
| 1 | ***Inputs*** | ***AmountofVehicle*** | High / Medium / Low |
| 2 | | ***SpecialDay*** | Yes / No |
| 3 | | ***WeatherCondition*** | 1 / 2 / 3 |
| 4 | | ***CurrentTime*** | High / Medium / Low |
| 5 | | ***TypesofDay*** | Weekday / Weekend |
| 6 | ***Output*** | ***Traffic*** | Red/Yellow/Green |

Another important thing for traffic prediction system is **traffic history data** as well as how much fast it can notify the current traffic condition to the end user. It is obvious that the amount of history data is getting bigger and bigger each year. In consequently, the Traffic Prediction System is going to face the problem of performance degrading according to the heuristic processes on an enormous amount of history data. For these issues, the parallel process using Multithreading technique is applied. Therefore, roughly half of processing time is reduced and consequently the system will not make the end user be waiting for ages. In this work, the database is divided into five parts, ***AmountofVehicle, SpecailDay, WeatherCondition, TypesofDay*** and ***CurrentTime.*** The main thread assigns each part (attribute) to each thread and then collects the final result from each thread. After that, it predicts the current traffic status and then alerts the end user and stores the final result in database as a training data for future use.

Table 2 describes traffic prediction algorithm. The task is to learn to predict the percentage of current traffic condition, ***Traffic,*** for an arbitrary time, based on the values of other attributes. In this work, hypothesis, X, is a vector of six constraints, specifying the values of the six attributes ***CurrentTime, KingofDay, SpecialDay, AmountofCar,*** and ***WeatherData***. V is the set of target values, Red, Yellow and Green. $P(v_j)$ terms the class prior probability and $P(a_k| v_j)$ represents the likelihood.

**Table 2. Traffic Prediction Algorithm**

**Algorithm: Traffic Prediction System**

**$v_j \in V$ = {Red, Yellow, Green} where (j=1,2,3)**

**Input :** ***current time, kind of day, special day, amount of car, weatherdata***

**Output :** ***percentage of current traffic condition*** **(Red, Yellow and Green)**

1. Instant <---- Receipt input data
2. Hypothesis <---- Read traffic history data
3. Calculate the required $P(v_j)$ and $P(a_k | v_j)$ probability
    1. For each target value $v_j$ in V do
    2. n <---- the number of training examples for which v = $v_j$
    3. nc <---- the total number of training examples for which v= vj and a = aj
    4. p = a prior estimate for $P(a_j | v_j)$
    5. m= the equivalent sample size
    6. $P(v_j)$ <----- (|n|) / (|Hypothesis|)
    7. For each attribute aj in Input Instance
       $P(a_k | v_j)$ <---- (nc + (m*p)) / (n+m)
4. VNB <---- $\operatorname{argmax}_{v_j \in V} P(v_j) \prod_{i \in attributes} (a_k | v_j)$

## 4. Analytical Performance Model of Naïve Bayes based Traffic Prediction System

This section will discuss about the performance of the system and data collection used as history data. Now the system is focusing on the Route 58 Ojana junction in Okinawa, Japan. The history data collection has been doing for the last two months for this junction from Google Map as well as video camera. As the system very depends on history data, the accuracy of hypothesis is also very important for estimating future result well.

In this work, the accuracy of hypothesis is measured by using Standard Deviation for errors(h). Where errors(h) means errors of hypothesis h with respect to target function f and data sample S

$$\boldsymbol{error_s(h) = r/n} \qquad \text{equation (1)}$$

Where *r* means the number of instances from *S* misclassified by *h* and *n* means the number of instances in the sample *S.*

$$\delta_{errors(h)} \approx \sqrt{\frac{errors(h)(1-errors(h))}{n}} \qquad \text{equation (2)}$$

In this case, error for hypothesis was observed to be 0.03 (3%).Figure 2 shows the traffic condition for rush hour time on 13 April in 2016. It can be seen that the heavy traffic jam is in rush hour time, in the morning and in the evening. The morning rush hour time is from the number time 1 to time 13 and the evening rush hour time is from time 35 to time 53. The other numbers are the normal time. During the normal time, the traffic condition is almost in Yellow and Green state. It is obvious that the heavy traffic condition remains stable in rush hour time from time 1 to 7 and 42 to 58. Although there is traffic in Red condition in fluctuation from time 8 to 13, there is no Green condition.

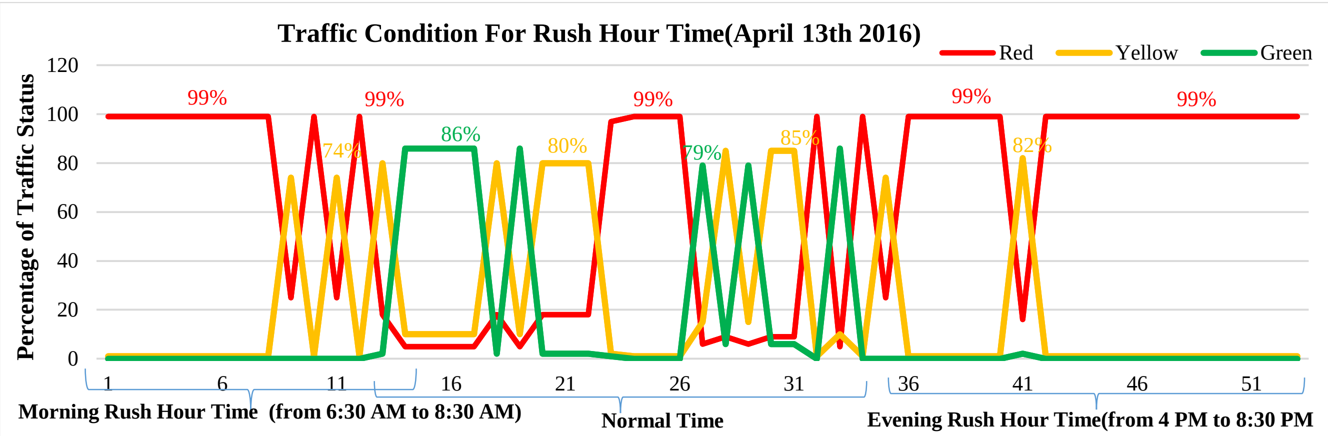


**Figure 2. Traffic Prediction for Rush Hour Time**

Figure 3 describes the fluctuation in the traffic jam in weekend day of Route 58, Ojana on April 9th – 10th, 2016. The time number 1 to 20 is normal time, 21 to 40 rush hour time, 41 to 65 normal time and 66 to 69 rush hour time. There is a sharp increment between 1 and 6, with 100 % being in Red condition. However, that duration is normal time. After this time, the percentage drops quickly to less than 80 %. Even though, the percentage of red condition takes more than other condition. Another peak condition is after 59. Normally, there is seldom traffic jam in weekend days, but on April 9 to 10, there was a traffic jam even during normal time because these days were raining days.

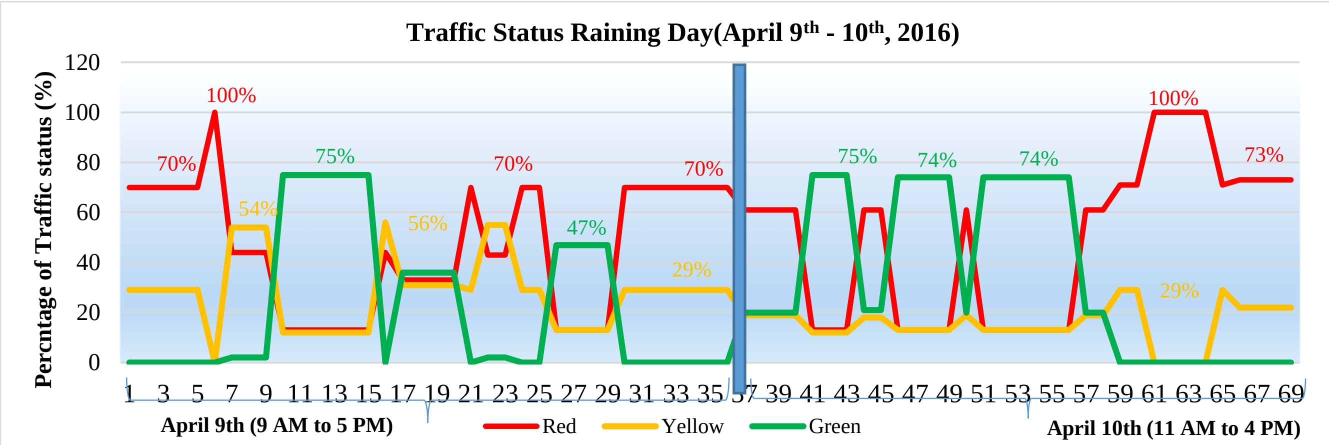


**Figure 3. Traffic Condition During Rush Hour Time**

According to Figure 3, road conjunction for raining day in Red condition was 50%, in Yellow condition 8.5 % and in Green condition 38 % even in weekend day.

**Table 2. Metric For Performance Evaluation**

| | | PREDICTED CLASS | | |
|---|---|---|---|---|
| ACTUAL CLASS | | RED | YELLOW | GREEN |
| | RED | A | B | C |
| | YELLOW | D | E | F |
| | GREEN | G | H | I |

Table 2 illustrates the predicted class and actual class for computing system's accuracy. The accuracy of the system is obtained by dividing the number of truth outputs by the total number of truth and wrong outputs.

$$\text{Accuracy} = \frac{A+E+I}{A+B+C+D+E+F+G+H+I} \qquad \text{equation (3)}$$

The system could predict the result as accurate as 93% by using the equation (3). The Purity for system's prediction is 93% by utilizing the following equation.

$$\text{Purity} = \sum_{k=1}^{n} \frac{n_i}{n} \, purity_i = \frac{1}{n} \sum_{k=1}^{n} max^{k}_{j=1} \{n_{ij}\} \qquad \text{equation (4)}$$

## 5. Conclusion

This paper describes the traffic prediction system which has been implemented for Ojana junction Root 58 Okinawa, Japan. Such tasks, weather condition and data from camera, analyzing day and time, are considered. The system shows that 93% purely predicted using Naïve Bayes Theorem using m-estimate of probability. According to the accuracy of hypothesis measurement using standard deviation, it has 3% error still left. The system performance is also maintained by using Multithreading which decreases 50% system processing time. Now the paper only emphasizes on only junction. As a future work, more than two junctions are going to focus on.

## References


[1] D. Rosenbaum, J. Leitloff, F. Kurz, O. Meynberg, and T. Reize, "Real-Time Image Processing for Road Traffic Data Extraction from Aerial Images", Vienna, Austria, July 5-7, 2010, IAPRS, Vol.XXXVIII, Part 7B.

[2] P.Niksaz, "Automatic Traffic Estimation Using Image Processing", International Journal of Signal Processing, Image Processing and Pattern Recognition, vol.5, No. 4, December, 2012.

[3] S.S.Aung and T.T.Naing, "Mining Data For Traffic Detection System Using GPS-enabled Mobile Phone on Mobile Cloud Infrastructure", in Proceedings of International Journal on Cloud Computing: Services and Architecture (IJCCSA),Vol.4, June 2014.

[4] R . J . Herring, "Real-Time Traffic Modeling and Estimation with Streaming Probe Data using Machine Learning", University of California, Berkeley,2010.

[5] M.Wang, T. Mulinazzi and S. D. Schrock , University of Kansas, "Feasibility of Using Cellular Telephone Data to Determine the Truckshed of Intermodal Facilities", Proceedings of the 2009 Mid-Continent Transportation Research Symposium, Ames, Iowa, August 2009. © 2009 by Iowa State University.

[6] T. Wang and Z. Zhu, "Real Time Moving Vehicle Detection and construction for Improving Classification ", IEEE, ISBN: 978-1-4673-0233-3, Jan 9, 2012.