On Study of Vehicular Movement Data Analysis for Generating Driving Behavior and Road Situation Information

Yamashita, Rei-Jo and Yao, Hsiu-Hsen

Abstract**—**In our study, we use tachograph to collect *vehicular movement data*, such as GPS data, velocity, acceleration, crosswise-acceleration, angular-speed, and distance data. These data are all deposited in a data warehouse. Then, the *information of driving behavior*, such as "stepping-on-accelerators", "stepping-on-brakes", and "transferring-the-steering-wheel" is generated. Using data mining and data analysis approach, we set up a pattern rule base to recognize various road situation, driving behavior, and vehicle types. Finally, a *VDR(vehicle, driver, and road) quality knowledge*, including driving smoothness, driving stability, driving safety degree, driving comfort level, and road situation as well as driver condition, is generated.

Index Terms**—**GPS, tachograph, Driving Behavior, Driving Pattern Recognition

I. Introduction

Researchers have found that vehicular movement data is important for driving safety improvement, traffic road situation understanding, and driver management [11,14]. Various efforts try hard using IT techniques to improve traffic environment safety and traffic flow efficiency to provide high-quality service for passengers. The goal of our study is trying to develop an integrated information system to help manager as well as traffic policy-makers in promotion of road transportation quality, improvement of driving safety management, and enhancement of driving behavior better. Furthermore, the motivation of our study is try to build up a high quality, convenient and digital traffic environment for road traffic decision making support.

The recently development of intelligent transportation system (ITS), using IT techniques, offers high quality road service and well traffic decision making by automatic data collection ways, which can continuously improve traffic service quality and traffic safety level. This research uses a new approach to help traffic manager to reduce driving management cost, to find more driving behavior dominant factors, to increase the passenger satisfaction, to improve efficiency in traffic safety management, and to discover more demands in road traffic.

In this study, the basic information comes from vehicular movement data. We use tachograph to collect the vehicular movement data, and then transform them into driving behavior information. The collected data is integrated with some necessary information (driver name, vehicle ID, etc) together. The program converts the data to the information that is deposited in a data warehouse. One may find that useful knowledge for road traffic decision or driver management can be generated through analysis of these raw vehicular movement data.

Yamashita, Rei-Jo is with ShiGun Systems.
Yao, Hsiu-Hsen is with the Department of Computer Engineering, Yuan-Ze University.

In our study, we use tachograph to collect *vehicular movemen data*, such as GPS data, velocity, acceleration, crosswise-acceleration, angular-speed, and distance data. These data are*t* all deposited in a data warehouse. Then, the *information of driving behavior*, such as "stepping-on- accelerators", "stepping-on-brakes", and "transferring-the- steering-wheel" is generated. Using data mining and data analysis approach, we set up a pattern rule base to recognize various road situation, driving behavior, and vehicle types. Finally, a *VDR(vehicle, driver, and road) quality knowledge*, including driving smoothness, driving stability, driving safety degree, driving comfort level, and road situation as well as driver condition, is generated.

Kuhler and Karstens[1] define ten kinds of drive behavior parameter threshold value in 1978. They are average speed or rate, average drive speed of rate(amputate the motionless time), eration, retardation and stop in single drive period, acceleration time, stable speed or rate time, retardation time, average acceleration, average retardation, average drive period time, and the average numbers of acceleration.

Ericsson[2] concludes 26 types of driving behavior parameters to describe the quality of driving behavior. Ericsson uses SPSS package to calculate the parameters' statistics, in order to get the threshold of the values.

Van der Voort[3] in 2001 concludes the threshold values of steering wheel turning angle, average speed, average acceleration, and use logical procedures to determine kinds of drive behavior, then give suggestion for the improper drive behavior.

In Hauer's research(1996), the researchers discuss to sort dangerous level by the numbers according to the traffic accident ratio or accident-vehicle / numbers-of-vehicle-in-the-area. Some of them use the index of frequency, and some use both of them.

Geurts, Wets, Brijs, and Vanhoof[6], they used Flemish's index of the number of history traffic accidents which happen in 3 years or longer time, then one location danger degree is calculated by the formula as follows(Ministry of Flemish Community, 2001): $S = 1*X + 3*Y + 5*Z$, where

X is light injury number, Y is heavy injury number and Z is dead number in traffic accidents.

Sec 2 discusses a global view of our vehicular movement data collection and the derivation of driving behavior information. In sec 3, we discuss on the driving behavior and patterns recognition, especially on the unsafe driving patterns. Section 4 describes data analysis of the driving information. In this study, we provide two kinds of data analysis, **Driver Behavior Analysis, and Road Features and Situations Analysis**. Section 5 addresses how to develop a list of **Driving Quality Knowledge indices.**

II. DATA COLLECTION AND INFORMATION DERIVATION

In our study, digital tachographs are installed in vehicles to collect vehicular movement data, consisting of GPS data and vehicular movement data. Besides the vehicular movement data, some “external” related data, including driver ID, vehicle ID, and video data are also collected.

The tachograph used in this study collects 4 kinds of data using the following components,

1. GPS receiver, collecting the data of date, time, latitude, longitude, velocity, and height of the trip;
2. Gyro and accelerators, collecting the data of vehicle forward/backward acceleration, crosswise acceleration, and angular-speed;
3. Memory Stick for depositing data;
4. RS232c link, and Modem link.

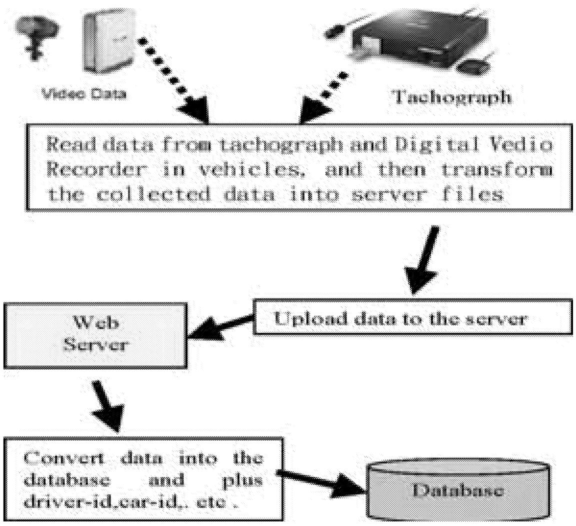
Data Collection

Tachographs may be installed in vehicles for collecting GPS data, and vehicular movement data (i.e., Back-and-Forth Acceleration, Crosswise-Acceleration, and Angular-Speed, Velocity, and Distance). The driving distance can be derived directly from GPS data. These data are all deposited in the data warehouse.

Information of Driving Behavior

One may transform the collected data from the tachograph and their related data to derive “stepping-on-accelerators”, “stepping-on-brakes”, and “transferring-the-steering-wheel” immediately.

The information system architecture and data flow chart in this study are designed as shown below,



Data Warehouse

The data, information, and knowledge in this study are all stored in an OLAP data warehouse, which provides various views for users to check traffic and driving data by time, by vehicles, or by road zones. The data stored in this data warehouse consists of 4 parts, the related image material, the driving data of vehicular movements, the GPS data, and the map data. The data materials present the way to be possible as the statistical graph and with GIS map. The data materials include aggregated data and the detailed information. The system may then provide a user interface using the OLAP operations via the multiple views to see the traffic and driving data.

The OLAP data warehouse provides users to view data via various perspectives. Some basic operations as follows, (a) Slice -A single dimension operation to see the fact data; (b) Dice- multidimensional operation to see the fact data; (c) Roll-up-A higher level of aggregation; (d) Drill-down-A detailed view to see the fact data; (e) Rotation-Viewing fact data from a new perspective

In our study, we support some useful decision resource to help the user to make effective traffic safety decisions. During the decision process, the programs can output some statistical chart and some geographic information. For instance, the hazardous road location is more appropriate to be viewed on the map. The data analysis based on some derived road traffic data: (1) driving behavior, (2) road traffic situations. These reports provide measurement of whether the roads are traffic jams, the stability, the smoothness condition. The driving managers and traffic management officers may check each index for each road section or for each driver to analyze traffic situations or driving behaviors.

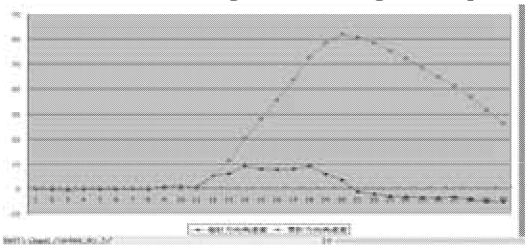
III. PATTERN RULE BASE

This study uses some pattern recognition methods to find some driving patterns, or to check whether the driving behavior is dangerous. We define some thresholds for these checking. We then aggregate these raw data and reorganize them to be road traffic index. Some basic data are GPS data, velocity, acceleration, crosswise-acceleration, angular-speed, and distance data. These data are all deposited in the **data warehouse**. Then, the *information of driving behavior*, such as “stepping-on-accelerators”, “stepping-on-brakes”, and “transferring-the- steering-wheel” can be generated. The Pattern Rule Type is a list of movement conditions in several steps. Some instances from the report in [11] can be referenced.

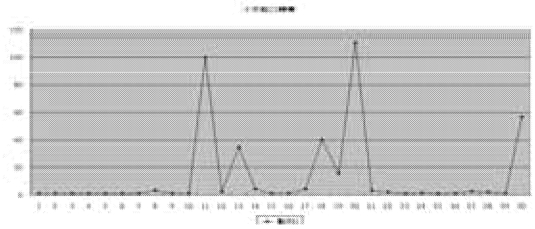
The rules of some basic driving patterns are as below,

- L/R/U-turn
- Road type
- Vehicle type
- Driver type

Below is the integrals of angular-speed in L/R-turn



Below is the (pseudo) curvature in L/R-turn.



The dangerous driving behavior rules are like the form as below, e.g., W-type traffic lane changing

step1 : crosswise-acceleration≦0.23g, and
step2 : crosswise-acceleration≦－ 0.2g, and
step3 : crosswise-acceleration≦0.15g, and
All these actions happen in 5 seconds.

The codes of this rule is as followings,

```
//step1
if er[I].side > 0.23 then
   begin
//step2
      for k :=0 to 5 do
      begin
      if er[I+k].side < -0.2 then
           break;
      end;
//step3
      for k := k to 5 do
      begin
      if er[I+k].side > 0.15 then
         er[I+k].flag :=1;
         iErr := I+k;
      end;
   end;
```

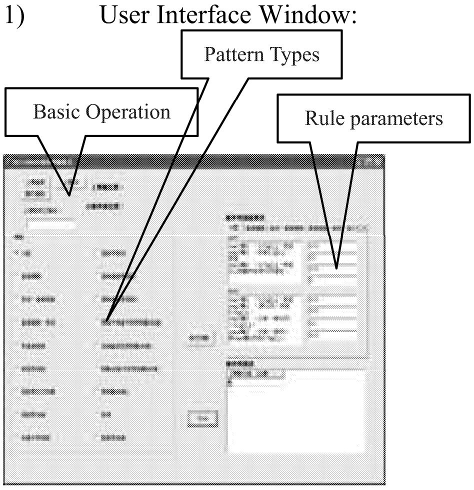
We implemented the other Pattern rules from the report provided by [11], as follows,

Below are Dangerous Driving Patterns in the rule base

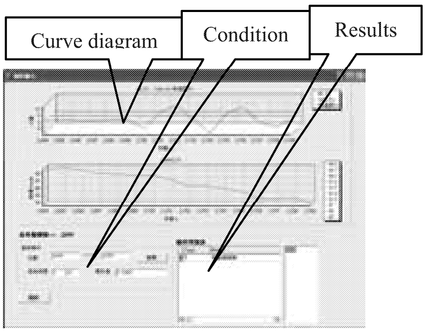
- w-type traffic lane changing
- suddenly passing another vehicle
- suddenly turning around
- traffic lane changing twice
- traffic lane changing frequently
- traffic lane changing with emergently breaking
- urgently breaking with traffic lane changing
- urgently traffic lane changing
- urgently turning around during moving off
- urgently transfers the steering wheel during turning around
- urgently stemming on accelerator during turning around
- urgently turning around during braking
- urgently breaking during turning round
- high-speed driving with urgently turning around
- urgently braking during moving off or back
- urgently braking while stemming accelerator
- extremely braking
- turning around and then moving in reverse
- taking a long term rest without flameout
- moving back
- turning around back
- urgent stemming on accelerator at the start

Our pattern recognition program’s user interface is as following,

1) User Interface Window:



Row Data Diagram:



Pattern Rule Query and List:

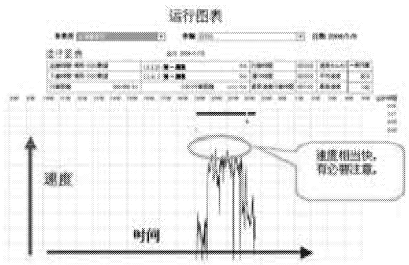
IV. Data Analysis

Two kinds of data analysis are provided. They are Driver behavior analysis, and Road features and situations analysis.

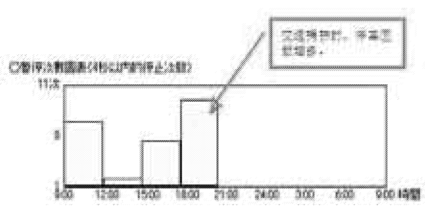
4.1 Driver Behavior Analysis

In this study division, we use "Andokun" web software, developed by Shigun Systems from Japan[14], to analyze the driving behavior for each driver.

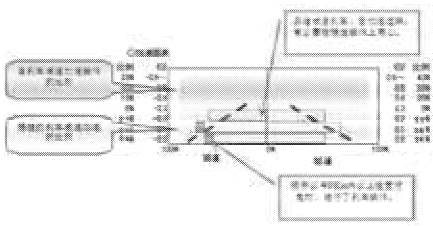
The following diagram provides the basic information of some driving records.



Some diagram shows the driving custom of one driver.

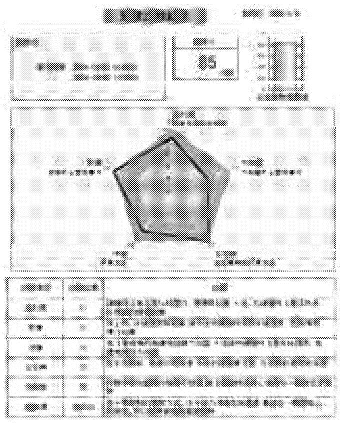


Some diagram shows the driving trends of one driver.

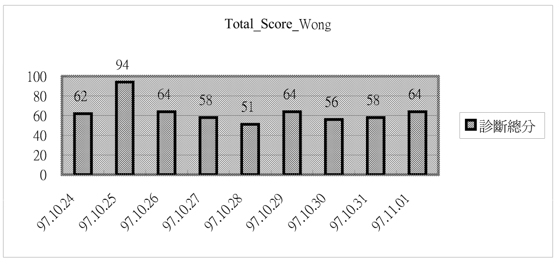


Below is Dangerous Driving Event List for each driver

The following diagram shows the driving safety evaluation scores for each driver.



The diagram below is the comparison of various driver's scores.



4.2 Road Features and Situations Analysis

In this section, we discuss some features of the road segment.

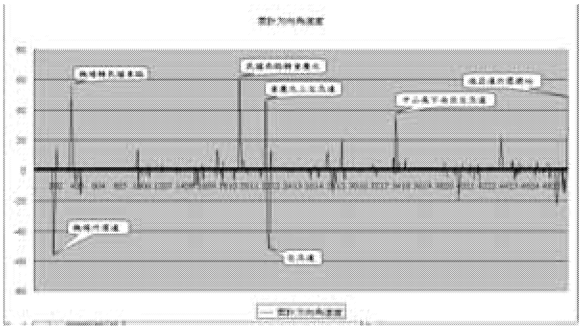
Curvature of a road

Using parameters Px_Angular and Px_Speed, one may generate a pseudo "curvature", denoted K. Please see the following formula of K.

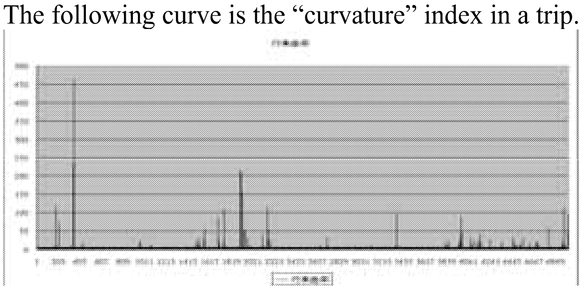
$$K = lim_{P \to Q} \frac{\theta}{\overline{PQ}}$$

Θ is the turning angule, and P is the previous point, and Q is the next point.
The pseudo curvature is a basic parameter for road feature references.

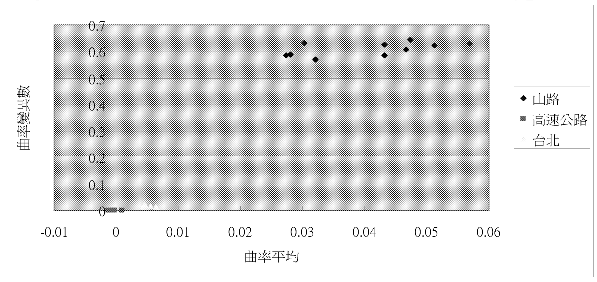
The following curve is the "integral" of angular-speed index in a trip, which shows the L/R-turn place.



The following curve is the "curvature" index in a trip.



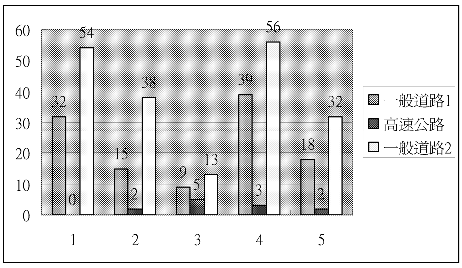
Below the diagram is the comparison of Var(curvature) vs. Avg(curvature) in three different kinds of road types.

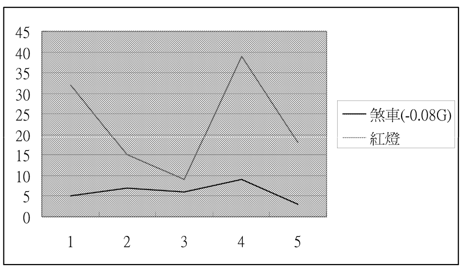


One may find that the average and variation of the curvature in a mountain road is major different from the curvature of the other kinds of roads. Note that we use resample method to normalize various runs of driving data.

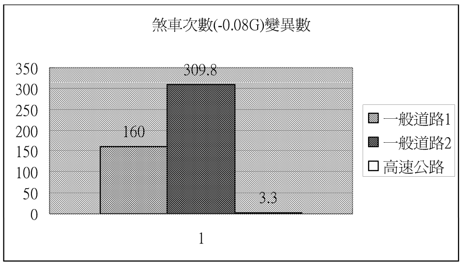
Traffic Jam Degree

While the threshold of braking number is -0.08g, we get the following diagrams.
Note that there are 5 run data.





It shows positive correlation, where the threshold of braking number is -0.08g.



As we see in the diagram, the variances of braking in different kinds of roads are quite different from the others.

V. DRIVING QUALITY KNOWLEDGE GENERATION

In our study, various driving quality knowledge are generated based on the vehicular movement data. The following four basic types of the driving quality indices are used to evaluate some driving quality degrees. Note that these indices can be used to evaluate various drivers', vehicular, or roads' driving quality degrees.

VDR(vehicle, driver, and road) Quality Knowledge

the indices in this study consist of

1.Driving smoothness (*D-Smooth*),
2.Driving Stability(*D-Stable*),
3.Driving Safety Degree(*DSD*), and
4.Driving Comfort Level(*DCL*).

These indices can be used to weight the qualities of driving by various views.

Driving smoothness(*D-Smooth*)

The formula of *D-Smooth* = [count($|acc_t - acc_{t+\delta}| \leqq \varepsilon$, for all t +δ) + count($|brakes_t - brakes_{t+\delta}| \leqq \varepsilon$, for all t +δ) + count($|c\text{-}acc_t - c\text{-}acc_{t+\delta}| \leqq \varepsilon$, for all t +δ) + count($|wheel_t - wheel_{t+\delta}| \leqq \varepsilon$, for all t +δ)] / (3 * total-data-quantity) for each time t (1 sec),

Where δ and ε are two given small number, acc_t is the quantity of stepping- on- accelerators at time t, $c\text{-}acc_t$ is the quantity of crosswise- accelerations at time t, $brakes_t$ is the quantity of stepping- on- brakes at time t, $wheel_t$ is the quantity of angular-speed of transferring- the- steering- wheel at time t, εandδare two small given numbers, total-data-quantity is the data quantity per driver, per road section, or per vehicle during a period of times.

Driving Stability(*D-Stable*)

The formula of *D-Stable* = [count($acc \leqq \sigma threshold_{acc}$) + count($brake \leqq \sigma threshold_{brakes}$) + count($|c\text{-}acc| \leqq \sigma threshold_{c\text{-}acc}$) + count($|wheel| \leqq \sigma threshold_{wheel}$)] / (3* total-data-quantity) for each time t (1 sec),

where *acc* is the quantity of stepping- on- accelerators per second, *c-acc* is the quantity of crosswise- accelerations per second, *brakes* is the quantity of stepping- on- brakes per second, *wheel* is the quantity of angular-speed of transferring- the- steering- wheel per second, $\sigma threshold_{acc}$ is 1.5 or 2 times the standard deviation of the total-data of stepping- on- accelerators per second in the data warehouse, $\sigma threshold_{brakes}$ is 1.5 or 2 times the standard deviation of the total-data of stepping- on- brakes per second in the data warehouse, $\sigma threshold_{c\text{-}acc}$ is 1.5 or 2 times the standard deviation of the total-data of crosswise- accelerations per second in the data warehouse, $\sigma threshold_{wheel}$ is 1.5 or 2 times the standard deviation of the total-data of angular-speed of transferring- the- steering- wheel per second in the data warehouse, and total-data-quantity is the data quantity per driver, per road section, or per vehicle during a period of times.

Driving Safety Degree(*DSD*)

The formula of *DSD* = [count($acc \leqq Threshold_{acc}$) + count($brake \leqq Threshold_{brakes}$) + count($|c\text{-}acc| \leqq Threshold_{c\text{-}acc}$) + count($|wheel| \leqq Threshold_{wheel}$)] / (3 * total- data- quantity) for each time t (1 sec),

where *acc* is the quantity of stepping- on- accelerators per second, *c-acc* is the quantity of crosswise- accelerations per second, *brakes* is the quantity of stepping- on- brakes per second, *wheel* is the quantity of angular-speed of transferring- the- steering- wheel per second, $Threshold_{acc}$ is the safety threshold quantity of stepping- on- accelerators per second, $Threshold_{brakes}$ is the safety threshold quantity of stepping- on- brakes per second, $Threshold_{c\text{-}acc}$ is the safety threshold quantity of crosswise- accelerations per second, $Threshold_{wheel}$ is the safety threshold quantity of angular-speed of transferring- the- steering- wheel per second, and total-data-quantity is the data quantity per driver, per road section, or per vehicle during a period of times.

The index of *DSD* is an important quality weighting level, which can be used to weight the quality of the driving safety. Please note that this index can provide a way to weight the potential driving safety factor.

Driving Comfort Level(*DCL*)

The formula of *DCL* = [count($acc \leqq threshold_{acc}$) + count($brake \leqq threshold_{brakes}$) + count($|c\text{-}acc| \leqq threshold_{c\text{-}acc}$) + count($|wheel| \leqq threshold_{wheel}$)]/(3 * total-data-quantity) for each time t (1 sec),

where *acc* is the quantity of stepping- on- accelerators per second, *c-acc* is the quantity of crosswise- accelerations per second, *brakes* is the quantity of stepping- on- brakes per second, *wheel* is the quantity of angular-speed of transferring- the- steering- wheel per second, $threshold_{acc}$ is the comfort threshold quantity of stepping- on- accelerators per second, $threshold_{brakes}$ is the comfort threshold quantity of stepping- on- brakes per second, $threshold_{c\text{-}acc}$ is the comfort threshold quantity of crosswise- accelerations per second, $threshold_{wheel}$ is the comfort threshold quantity of angular-speed of transferring- the- steering- wheel per second, and total-data-quantity is the data quantity per driver, per road section, or per vehicle during a period of times.

The index of *DCL* is an important quality evaluation, being used to weight the quality of passengers riding a driving vehicle.

Some enhanced index may be developed to add count($joltiness \leqq threshold_{joltiness}$) into *DCL*, where*joltiness* is the quantity of driving joltiness degree per second, and $threshold_{joltiness}$ is the threshold quantity of driving joltiness degree per second.

VI. Conclusion

One of the goal of this study is for developing a traffic safety information system and its applications. The main idea is to build a traffic safety analyze database and combine the geographic information system, find out the dangerous road segments and driving behaviors, know the traffic situations and conditions in different road sections and in various time zone, according these to build traffic safety decision making system, help us to refine transportation facilities, to promote traffic safety and to make good transportation decisions.

References

[1] Kuhler, M. and D. Karstens, 1978, “Improved driving cycle for testing automotive exhaust emissions,” SAE Technical Paper Series 780650.
[2] Ericsson, E., 2001, “Independent driving pattern factors and their influence on fuel-use and exhaust emission factors,” Transportation Research Part D, Vol. 6, pp. 325-345.
[3] Vander Voort, M., M. S. Dougherty, and M. van Maarseveen, 2001, “A prototype fuel-efficiency support tool,” Transportation Research Part C, Vol. 9, pp. 279-296.
[4] Flahaut, B. “Impact of infrastructure and local environment on road unsafety. Logistic modeling with spatial auto correlation. Crash Analysis and Prevention,” 36(6), 1055- 1066., 2004
[5] Flahaut, B., Mouchart, M., San Martin, E., & Thomas, I. (2003). “The local spatial autocorrelation and the kernel method for identifying blackzones. A comparative approach. Crash Analysis and Prevention, 35(6),991- 1004.
[6] Geurts, K., Wets, G., Brijs, T., & Vanhoof, K. (2004). “Identification and ranking of black spots: Sensitivity analysis. Electronic proceedings (CD-ROM) of the 83th annual meeting of the transportation research board. Washington, DC' National Research Council.
[7] Ministry of Flemish Community. (2001). Ontwerp-Mobiliteitsplan Vlaanderen, Brussels, Belgium, http://www.viwc.lin.vlaanderen.be/mobiliteit,Accessed July 2003. (in Dutch).
[8] “Identification of Human Driver Models in Car Following, IEEE, Burnham et al.,1974
[9] “Mining Sequential Patterns by Pattern-Growth: The Prefix Span Approach, Jian Pei et al.，2004
[10] Ericsson, Eva, “Independent Driving Pattern Factors and Their Influence on Fuel-Use and Exhaust Emission Factors, Transportation Research, Part D, Vol.6, No.5, pp.325-345，2000
[11] www.datatec.co.jp
[12] Case Study：An Intelligent Decision Support System, IEEE, Zbigniew Michalewicz, 2005
[13] http://www.moi.gov.tw/Scripts/tornado
[14] http://andokun.com/shiguntw/