I J C T A, 9(15), 2016, pp. 7273-7281

Analysis of Car-Following Behavior Using Weighted Cluster Model

Kakde Swati Sudhakar* and C. Malathy**

ABSTRACT

Nowadays people's preference for the need of transportation by car is more as it provides more comfort than any other vehicle. Every time the driver hit the road they face many pressures and it has become necessary to provide the secured environment. The focus was made on traffic safety. Analyzing driver behavior has become vital so to provide comfortable driving and safer journey. In this paper, it is proposed to analyze speed, trip identifier, time, heading, gas pedal position, light intensity; break on off, turn signal state, acceleration, etc. that will help to improve the efficiency of the car-following behavior and driving pattern of car drivers. Weighted Cluster Model is used here as it supports for nonlinear prediction of outputs from inputs.

Index Terms: Clustering methods, data mining, ITS (Intelligent Transportation System),WCM (Weighted Cluster Model).

1. INTRODUCTION

A driver's behavior is the way a driver reacts to his/her current systems (e.g. speed, distance, accelerate or steer, lane-marking). His/her behavior can be formally defined as the function that maps traffic states to a driver's actions. Typically, driver behavior studies collect data that incorporate both states and actions and attempts to develop a more accurate mapping between them. These efforts do not usually include many influencing factors of human behavior, such as emotion, personality, hunger, age, gender [18].

The driver's internal factors are Essential in impacting traffic safety in addition to vehicle road environment. The driver's internal factors include the driver psychological characteristics. The differences of driver's age, gender, driving experience and individuality result in different psychological characteristics and the different psychological characteristics are reflected as driving tendencies. The system of driving inclination recognition plays an important role in improving the applicability and accuracy of vehicle active safety systems (collision avoidance as warning system) drivers choose their desired level of task difficulty [16]. The proposed system is structured such that a known state can link to multiple actions thus accounting for the effects of developing a new driving state. A sequential clustering algorithm is used for the clustering of driving patterns. It is proposed to use sequential clustering algorithm which will segment and cluster a car following behaviors based on different variables such as speed, trip identifier, time, heading, gas pedal position, light intensity; break on off, turn signal state, etc. Driver behavior is what the driver chooses to do with these attributes. The objective of this system is to find a perspective set of driving state. Driving is a complex task and requires people to see and hear clearly, pay close attention to other cars, traffic signs, signals and react quickly to events. Better driving style and encourage self-management, active driver feedback promotes safety by presenting live feedback and drivers behavior. The Purpose of this analysis is to recognize the optimal cluster size so that they do not overfit/underfit their model to the dataset.

* MTech Computer Science & Engineering Srm University, Kattankulathur, 603203, Tamilnadu, India, *Email: swatikakde@hotmail.com*

** Professor, Department of Computer Science and Engineering Srm University, Kattankulathur, 603203, Tamilnadu, India, *Email: malathyc03@gmail.com*

2. CAR-FOLLOWING BEHAVIOR

In recent years, modeling and recognizing driver behavior have become critical to understanding intelligent transport system, human vehicle systems, and smart vehicle system [5]. The Intelligent transportation systems (ITS) are innovative applications which, without expressing intelligence as such, aim to provide advanced services involving to different modes of transport and traffic management and enable several users to be better informed and make them safer, more coordinated, and 'smarter' use of transportation networks. Smart transport systems technologies applied from basic management systems such as car navigation, traffic signal control systems, container management systems, variable message signs, automatic number plate recognition or speed cameras to monitor applications, such as security CCTV systems; and to more advanced applications that integrate live data and feedback from a number of other sources, such as parking guidance and information systems; weather information.

In the field of transportation and engineering many research and field experiments to study car-following behavior have been conducted on test tracks and roadways, and then modeled to represent driver's behavior [17].A car-following model was first proposed by Reachel (1950) and Pipes (1953) and extended by Herman(1959~1967) and has been continuously refined up to present with various approaches to describing relationships between the leader and the following vehicles. Most of the researches into clustering have occurred outside of the field transportation engineering. Largely, there have been two kinds of data collection methods for car-following experiments, using either video recording (or aerial film) to capture many anonymous or wire-linked test vehicles on a test track or a roadway [19]. In the first case, it should note that this type of data collection is really tedious work. One advantage is that real drivers are observed; who are unaware that the experiment is taking place, thus, their behavior is as "natural" as can be expected.

2.1. Data Collection Method

The car-following behavior typically collects vehicle trajectory data through various factors, including naturalistic, simulator and video data collection methods. It also considers the different lane, weather factor, road construction, types of vehicle and traffic flow to analyze it. Data collection mainly observed the drivers physical actions as well as steering movement. The following Table I gives the description of the different data type and their advantages and flaws including driver's reaction into a different environment. It captures and analyze. The 100 car naturalistic driving study was an instrumented vehicle study conducted in the Northern Virginia/Washington, D.C. area over a two-year period. The primary purpose of the analysis was to collect large-scale naturalistic driving data to this end. The instrumentation was designed to be discreet, study participants were given no special instructions, and experimenters were not present. Around 100 vehicles were instrumented with a suite of sensors including forward and rearward radar, lateral and longitudinal accelerometers, gyro, GPS, access to the vehicle CAN, lane marking, and five channels of compressed digital video. The collection rates for the various sensors ranged from 1Hz to 10Hz. This collection effort resulted in approximately 2,000,000 vehicles miles and 43,000 hours of driving data

2.2. Data Clustering

Cluster analysis is a method of unsupervised learning and a statistical methodology used to categories individual objects into groups with similar meanings (homogeneous) there are several clustering techniques

Table 1

Data Type	*Description*	*Strengths*	*Weaknesses*
Naturalistic	An instrumental vehicle is drive in normal driving routines.	- The driver is in the natural environment. - Multiple trajectories observed for each driver.	Drivers know that they will be observed.

are available within each group, but for the application car-following behavior, the time-series clustering is more suitable. Time series clustering is to partition time series data into groups based on similarity or distance, so that time series in the same cluster are similar. For time series clustering with R, the first part is to work out an appropriate distance/similarity metric, and then, at the second step, use existing clustering techniques, such as K-means, hierarchical clustering, and density-based clustering or subspace clustering, to find clustering structures. Clustering the segments shows how certain behaviors repeat throughout the data within and between drivers.

3. IMPLEMENTATION

Sequential clustering technique is used to capture the full range of state–action clusters for car drivers. First, observed the naturalistic database, followed by extract car-following periods from the huge amount of data. Next, the data is divided into different states to finding a cluster. A car-following period will be defined as a period during which one vehicle is reacting to a leading vehicle in the same direction of the car. The behavior is the link between the data set and the action variables that the driver expresses (acceleration, deceleration, lane change, turning). The segmentation process divides each car-following period into multiple segments, and each segment will be defined as a new data set. The sequential clustering process group's similar parts into a single cluster for the new dataset will generate. Thus, a cluster gives the grouping of a certain behavior. Fig 1 shows coordinate system. To compare the different patterns associated with multiple drivers and observing similarities and differences between different drivers. The components used in the analysis are Feature Reduction using Principal Component Analysis. The Computing Root Mean Squared Error Rate deviation cluster size n versus the size of 1 cluster taking into account and possible reduction in RMSE.

3.1. Naturalistic Driving Data

Naturalistic driving data refer to the data collected from drivers in their natural environment. It is taken by different vehicles with specialized sensors and "vehicle network" recording equipment's, then allowing the drivers to drive the vehicles as they feel comfortable. The tools records a different number of variables (e.g., speed, acceleration, and steering wheel positions, lane markings), and of particular interest to car following, a radar system positioned at the front of the vehicle records the differences in the position of the car and the speed between the target and leading vehicles. The equipment also includes cameras that record what the driver sees from the front as well as from the two side mirrors, the driver's face, and what the

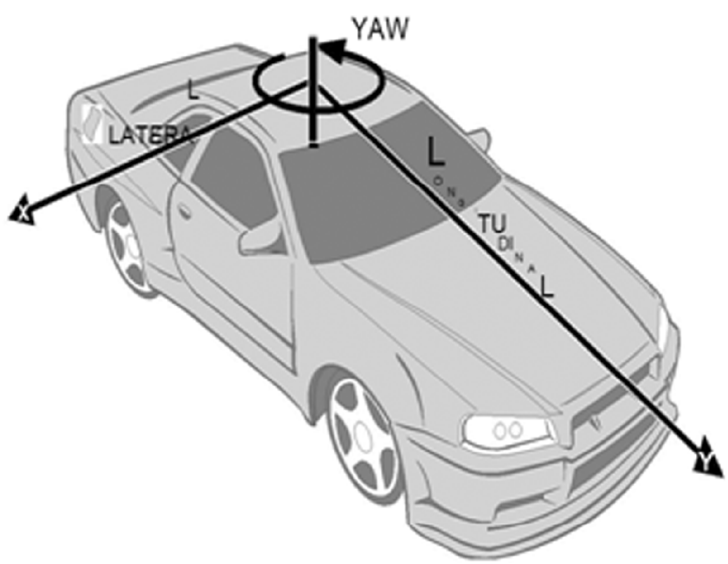

Figure 1: Coordinate System [19]

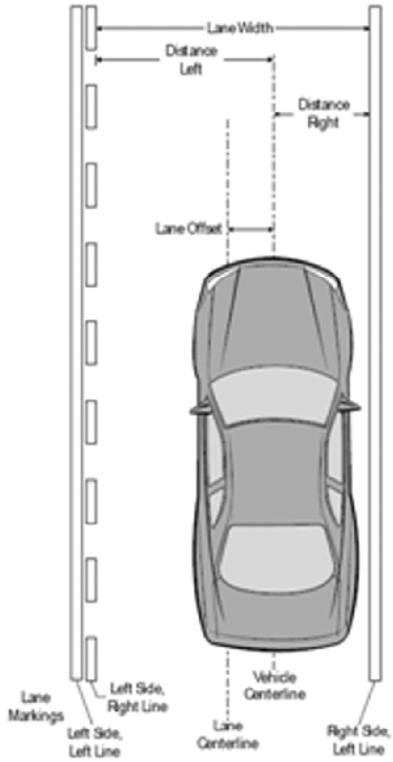

Figure 2: Machine-Vision-Based Lane Tracker Variable Names and Dimension [19]

driver is doing inside the vehicle. The naturalistic data used in this system were collected from VTTI. These data are collected from 100 cars that were used by multiple drivers. The 100 Car Naturalistic Driving Study was an instrumented vehicle study conducted in the Northern Virginia / Washington. These data describe crashes or near crashes from the 100 Car Naturalistic Driving Study (Dingus et al., 2006). The data include 68 crashes and 760 near crashes. Each file contains the time series data spanning 30s before an event and 10s after an event. The data in the libraries stored as comma delimited text, with each column representing a variable, and each row representing a time sample.

3.2. Data Dictionaries

There are different variables used for deriving driver pattern. The variables record will be taken for each driver and analyze it. The Table shows the different variables used this variable will help to observe a record of driver reacting with those variables.

- List of Dictionary Fields – A description of the components or fields described in the dictionary for each variable entry.
- List of Variables – A list of the entries (variables) in the dictionary which can be used as a table of contents to locate specific variables in the document.
- Conversions, Coordinate System, and Formulas – A catalog of unit conversions, sign conventions, and formulas which may be of value to researchers working with these data. Table II
- Data Dictionary Entries – The dictionary entries themselves, one for each variable included in the data set.

The vehicle speed, the lane offset, yaw angle, the range rate represent state variables. The lateral acceleration and the yaw rate represent the driver's action of steering. The longitudinal acceleration represents the driver's actions of accelerating or breaking.

Table 2
(List of the Variables)

List of variables in existing system		List of Variables in proposed system	
The variables included in the existing text files are,		The variables included in the text files are	
Variable	**Definition**	Sr.	Name
Radar Target ID	Unique identification number assigned to each lead vehicle.	1	Trip Identifier
		2	Sync
Longitudinal Acceleration(g)	Acceleration of the subject vehicle along the roadway.	3	Time
		4	Gas pedal position
Lateral Acceleration(g)	Acceleration of the subject vehicle across the roadway.	5	Speed, vehicle composite
		6	Speed, GPS horizontal
Yaw Angel(rad)	The angle between the vehicle heading and the center of the roadway.	7	Yaw rate
		8	Heading, GPS
Yaw Rate(rad/s)	The rate of the change in the yaw angle.	9	Lateral acceleration
Vehicle Speed (km/h)	The speed of the subject vehicle.	10	Longitudinal acceleration
		11	Lane Markings, Continuity, Left Side Left Line
Lane Offset(in)	The distance between the center of the vehicle and the center of the lane travel.	12	Lane Markings, Continuity, Left Side Right Line
		13	Lane Markings, Continuity, Right Side, Left Line
Range (ft)	The distance from the front of the subject vehicle to the back of the lead vehicle.	14	Lane Markings, Continuity, Right Side, Right Line
		15	Lane Markings, distance left
		16	Lane Markings, distance right
Range Rate(ft/s)	The rate of change in the range or the differences in speed between the subject vehicle and the lead vehicle.	17	Lane Markings, type left
		18	Lane Markings, type right
		19	Lane markings, probability left
		20	Lane markings, probability right
		21	Radar, forward, ID
		22	Radar, rearward, ID
		23	Radar, forward, range
		24	Radar, rearward, range
		25	Radar, forward, range rate
		26	Radar, rearward, range rate
		27	Radar, forward azimuth
		28	Radar, rearward azimuth
		29	Light intensity
		30	Brake on off
		31	Turn signal state

3.3. Sequential Clustering

The second part of this methodology clusters the segments, which were found by the segmentation algorithm, to find similar segments or behaviors in the car-following data. Clustering the segments shows how certain behaviors repeat throughout the data within and between drivers. Sequential clustering algorithm performs easy computations and processes pattern samples sequentially without special storage requirements. Its behavior is biased by the first patterns passed to the algorithms.

3.4. Sequential Algorithm

Let $d(x, C)$ denote the distance between data vector x and a cluster C. Furthermore, q is the allowed threshold of dissimilarity. [9]

Let $m = 1$

$Cm = X1$

for $i = 2$ to N

Ck; $d(xiCk) = \min i \le j \le md\ (XiCj)$

If$(d(xiXCk) > \phi) AND (m < q)$

$m = m + 1$

$Cm = Xi$

Else

$Ck = Ck \cup xi$

Where necessary, update representatives

End

End

The algorithm begins choosing the first cluster center among all the pattern samples randomly. Then, it processes the remaining patterns serially. It computes the distance from the actual pattern to its nearest cluster center. If it is smaller than or equal to *x*, the pattern is assigned to its nearer cluster. If not, a new cluster is formed with the definite form. Every m patterns, clusters are merged using a distance criterion (two groups are combined into one if the distance among their centroids is below C). If there are still more than k clusters, clusters are lumped together using a size criterion (clusters with less than patterns are merged with their next groups). If we still have too many clusters, the adjacent pairs of clusters are merged until there are exactly k clusters left.

3.5. WCM (Weighted Cluster Model)

The proposed WCM cluster model is an algorithm based on the non-linear prediction of outputs (dependent variables) from inputs (independent variables) with density estimation using a set of models(clusters).The procedure for cluster-weighted modeling of an input-output problem as follows. [20].

1. To build the predicted values for an output variable *y* from an input variable *x*, the modeling and calibration technique reaches a joint probability density function, $p(y, x)$. The “variables” might be uni-variate, multivariate or time-series.
2. The required expected values are obtained by creating the conditional probability density $p(y|x)$ from which the estimate using the conditional expected value can be obtained.
3. The important step of the modeling is that $p(y|x)$ is assumed to take the following form, as a mixture model:

$$(y,x) = \sum_{1}^{n} \omega j pj(y,x)\, p,$$

Here n is the number of clusters and w_j are weights that sum to one. The functions $p_j(y,x)$ are joint probability density functions that relate to each of the n clusters.

$$pj(y,x) = pj(y \mid x)\, pj(x)$$

Where,

- *Pj* ($y|x$) is a model for predicting *y* assumed *x*, and given that the input-output pair should be associated with cluster *j* by the value of *x*.
- *Pj(x)* is formally a density for values of *x*, given that the input-output pair should be associated with cluster *j*. The similar method used for regression analysis.

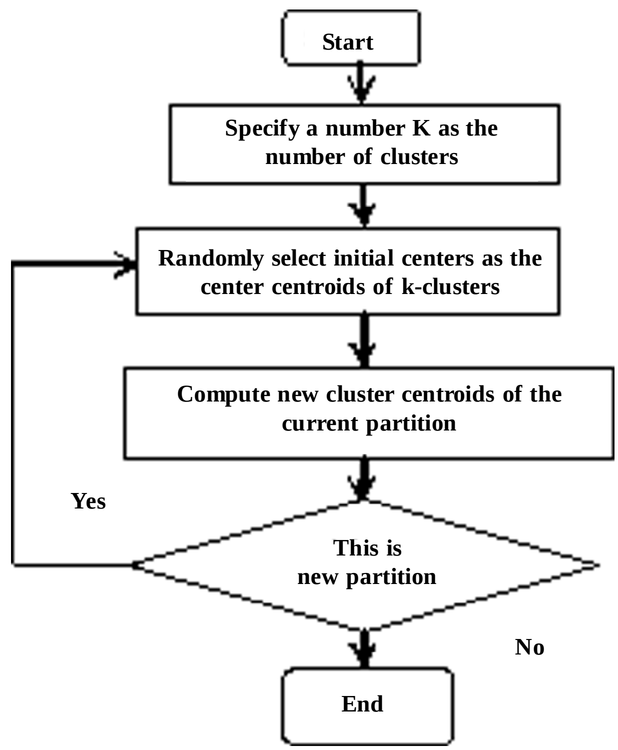

Figure 3: Flow chart for Sequential Clustering Algorithm.

4. RESULTS

Table III shows some clusters that occur very often. These clusters are investigated further for analyzing the behavior. The findings were that the clustering method produced clusters of the different data set. Some of the clusters have data errors where one sensor produces data that conflict with another sensor. The clusters containing data errors were removed for further analysis.

Here principal component analysis is used for analysis of multiple components and Root Mean Squared Error is reduced. The value obtained shows the deviation from cluster Size n Versus Cluster Size1. For each level, the WCM model was optimized to find the minimum RMSE for the difference in speed between the model and the data. For the last level, each cluster was optimized in order to find parameters that apply to each cluster individually.

The clustering method appears to be successful not only in grouping car-following behaviors but also in extracting and grouping data errors that occur over all 100 car drivers.

4.1. Car Driver

The results of clustering for car drivers are exposed in Fig. 4 which represents the cluster distributions or the “driving patterns” of each driver. Each division on the rectangle represents a cluster (1–10). The Graph represents the number of occurrences of each cluster by the car user.

Table 3
Cluster Occurrences Optimal Cluster size for Car Analysis

Number of Clusters	*Scaled PCA RMSE*	*Scaled Reduction in RMSE*
2	15.6414139865	
3	15.5829255075	5.8488478996
4	14.9880409012	59.4884606366
5	14.4934933817	49.4547519445
6	14.3038752334	18.9618148302
7	14.2270632625	7.6811970916
8	14.1256759492	10.1387313278
9	14.030479838	9.5196111197
10	13.9222513182	10.8228519828

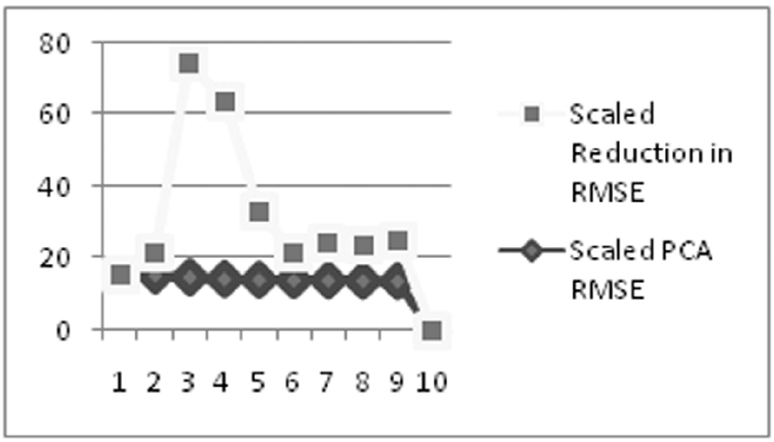

Figure 4:Graph for Cluster Occurances

5. DISCUSSION AND IMPLICATION

The method built-in in this paper exposes some characteristics of driver behavior. The results show heterogeneity among car drivers. The implications behind those declarations are that car drivers should be independently revealed. The heterogeneity among the car drivers can be due to different vehicle characteristics, but these drivers will carry on to drive their own vehicles, which mean that the varied behaviors still will be present in their driving under normal conditions. Although a behavior or a cluster might appear infrequently, that cluster can represent cases where the driver poses a risk to others or interrupts the flow of traffic. Car drivers show much more complex behavior that consists of a scattering of multiple clusters for each driver. This paper analyzes data from a variety of driving environments in order to capture the full range of behavior of drivers.

The number of drivers to contain should be based on the need to study additional patterns in more detail. For cars, studying more drivers could potentially result in finding more unique driving patterns. Drivers are humans, and human behavior has multiple impelling factors that are tough to consider such as emotions, personality, medical conditions, hunger or thirst [18]. These influencing factors can cause a driver to in a different way behave when the similar situation repeats.

The method offered in this paper would theoretically account for unobserved inducing factors. Finding of the characteristics of the impelling factors will require data from each cluster that include additional variables.

We have considered 31 variables in our analysis so as to get more accuracy.

6. CONCLUSION

It was proposed to find a prescriptive set of state-action clusters that will be used to analyze possible driving patterns of car driver as well as the safety of the driver. The comparison was made to analyze different patterns associated with multiple drivers and observation was made for the similarities and differences between different drivers driving pattern. The analyzed patterns can be used to support car traffic control system that will improve the driver behavior.

REFERENCES

[1] A. Kesting and M. Treiber, "Calibrating car-following models using trajectory data: Methodological study," Transp. Res. Rec., no. 2088, pp. 148–156, 2008.

[2] Alejandra Medina Flintsch, Bryan Higgs, Linsen Chong, Montasir M. Abbas, "A rule-based neural network approach to model driver naturalistic behavior in traffic" Transportation Research C. Sep-2011

[3] Andrew Liu, NobuyukiKuge, Tomohiro and Osamu Shimoyama "A Driver Behavior Recognition Method Based on a Driver Model Framework" Society of Automotive Engineers, 2013

[4] Alex P. Pentland "Driver Behavior Recognition and Prediction in a Smart Car" Media Laboratory Massachusetts Institute of Technology (MIT) 20 Ames St E15-385 Cambridge MA02139

[5] B. Higgs, M. Abbas, and A. Medina, "Analysis of the Wiedemann car following model over different speeds using naturalistic data," presented at the Road Safety Simulation, Indianapolis, Indiana, USA, 2011.

[6] Bryan Higgs and Montasir Abbas, Member, IEEE "Segmentation and clustering of Car-Following Behavior: Recognition of Driving Patterns" Ieee Transactions on Intelligent Transportation Systems, VOL. 16, NO. 1, FEBRUARY 2015

[7] Biagio Ciuffo, Carlos Lima Azevedo, JoãoLourenço Cardoso, Moshe E. Ben-Akiva, "Dealing with uncertainty in detailed calibration of traffic simulation models for safety assessment" Transportation Research C 2015

[8] Bryan Higgs and Montasir Abbas Ieee transactions on intelligent transportation systems, "Segmentation and Clustering of Car-Following Behavior: Recognition of Driving Patterns", Member, IEEE.

[9] C. C. Aggarwal and C. K. Reddy, Data Clustering: Algorithms and Applications. Boca Raton, FL, USA: CRC Press, 2013.

[10] ChiyomiMiyajima, FumitadaItakura, Koji Ozawa, Kazuya Takeda, Toshihiro Wakita, Katsunobu Itou, and Yoshihiro Nishiwaki "Driver Modeling Based on Driving Behavior and Its Evaluation in Driver Identification" IEEE 2007.

[11] Christine Buisson**,** LudovicLeclercq, Nicolas Chiabaut. "From heterogeneous drivers to macroscopic patterns in congestion" International Journal of Transportation Engineering 2014.

[12] David J. Lovell, David J. Lovell, Gang-Len Chang, Kent L. Norman, Taehyung Kim, William S. Levine, Yongjin Park, "Analysis of variability in car-following Behavior over long-term driving maneuvers" Transportation & Environmental Research 2005

[13] David J. Lovell, Martin Hall, Taehyung Kim, Yongjin Park "Limitations of Previous Models on Car-Following Behavior and Research Needs" Submitted on the 82th Transportation Research Board Annual Meeting Washington.January -2013.

[14] Ingmar Andréass on and Xiaoliang Ma "Driver reaction time estimation from a real car following data and application in GM-type model evaluation" Center for Traffic Simulation Research (CTR) Royal Institute of Technology (KTH), Stockholm 10044, Sweden. Jinglei Zhang, Xiaoyuan Wang "Extraction and Recognition Methods of VehicleDriving Tendency Feature Based on Driver vehicle-environment.

[15] Car following "International Journal of Computational Intelligence Systems", Dec2011

[16] Jiri Macur, Michal Radimsky, TomasApeltauer "Validation of microscopic Model" http://www.researchgate.net/publication/257465232,April 2013.

[17] M. M. Abbas, B. Higgs, and A. Medina Flintsch, "Naturalistic car driver behavior: Hybrid Wiedemann and Gazis-Herman-Rothery model cali-bration," presented at the Transportation Research Board 91st Annual Meeting, Washington, DC, USA, 2012.

[18]] N. A. Shenge, S. E. Oladipoand, S. K. Balogun, "Psychosocial factors influencing aggressive driving among commercial and private automobile drivers in Lagos metropolis," Social Sci. J., vol. 49, no. 1, pp. 83–89, Mar. 2012

[19] T. A. Dingus et al "The 100-car naturalistic driving study; Phase II— Results of the 100-car field experiment" Virginia Tech Transp. Inst., Blacksburg, VA, USA, VA DOT HS 810 593, 2006

[20] https://en.wikipedia.org/wiki/Cluster-weighted_modeling.

.