What is your driving identity?
Some empirical findings using large-scale smartphone sensors' data

Eleni G. Mantouka[1], Eleni I. Vlahogianni[1], John C. Golias[1]

[1]Department of Transportation Planning and Engineering, School of Civil Engineering, National Technical University of Athens, 5 Iroon Polytechniou Str., 15773, Zografou Campus, Athens, Greece
E-mail: elmant@central.ntua.gr

Abstract

Driving behavior is of great importance when road safety and improved road network conditions are being considered. Each driver has been seen to move between different driving styles and do not maintain a stable driving behavior. In this paper, we take advantage of a big dataset of real driving data collected through a smartphone application, that include among others harsh events (accelerating, braking and cornering), speeding, mobile usage etc. The aim of this paper is to identify universal driving profiles which consistently appear in the dataset and categorize drivers based on the frequency and the type of unsafe driving behavior that they perform during their trips. Findings revealed that there are two main groups of drivers in the dataset, those who seem to travel in an aggressive manner and those who perform a number of unsafe behaviors, such as risk taking and distraction from the driving task.

Keywords: *Driving behavior, driving profiles, clustering techniques, smartphone sensors, self-organizing maps*

1. Introduction

Driving behavior is of great importance in maintaining safe and sustainable traffic conditions. Furthermore, in the anticipated new era of mixed traffic conditions in relation to the various level of vehicle automation, the understanding of driving behavior and more specifically the analysis of unsafe driving styles seems vital. Advanced driving assistance and recommendation systems should be able to provide those recommendations that improve driving behavior and ensure road safety and therefore, detection and even prediction of driving behavior is crucial.
Current technological advances and especially smartphones, offer the opportunity to collect massive amount of high-quality data regarding human mobility and behavior, in a cost-effective manner in contrast to the inefficient and expensive solutions of OBDs and GPS devices (Montini et al., 2015, Ashbrook & Starner, 2002). Due to their available sensors, such as accelerometer, GPS, Bluetooth, microphone etc., smartphones can gather a variety of data. More specifically, in the case of driving behavior analytics, researchers take advantage of smartphones' sensors to capture driving behavior characteristics (Kanarachos et al., 2018), such as acceleration, deceleration, speed, etc. Driving behavior data could be collected without requiring any effort from the user and of course without distracting them from the driving task. Data gathered through smartphones, first, can be used to extract interesting features of the driving task and then, to further communicate them to the drivers to help them adopt safer and

more effective driving habits and increase their awareness regarding road safety (Papadimitriou et al., 2019). In addition, the identification of driving profiles can be used for the development of usage-based insurance schemes, such as Pay-as-you-drive and Pay-how-you-drive and assist insurance companies at offering services which are dedicated to each users' specific requirements and needs (Tselentis et al., 2017).

The various driving behaviors have been investigated by researchers using several machine learning techniques, such as neural networks (Meseguer et al., 2017), classification methods (Quek & Ng, 2013), Dynamic Time Warping (Johnson & Trivedi, 2011, Saiprasert & Pattara-Atikom, 2013) and many more. Studies have shown that driving behavior may vary from aggressive to inattentive and from risky to more safe behaviors. Drivers behave differently while performing the driving task in relation to the manner they alter their longitudinal (accelerate, decelerate) and lateral position (steering), the distance they choose to keep from the leading car as well as the time they choose to drive with an excessive speed (speeding) (Miyajima et al., 2007). Most of the relevant research has focused on the detection of unsafe driving events such as harsh braking and acceleration, rapid turning and rapid lane change (Saiprasert et al., 2013, Mitrović, 2005, Aljaafreh et al., 2012). Moreover, recent evidence emphasizes on a desirable driving style the so-called eco-driving, which refers to a rational way of driving and consumption of fuel (Sivak & Schoettle, 2012).

Each driver has been seen to move between different driving styles between successive trips without necessarily follow a specific pattern (Mantouka et al., 2019). But how easy is to identify the single driving identity of each user?

In this paper, we take advantage of an always increasing dataset of trip data which are gathered through a smartphone application. The dataset includes trip attributes, driving behavior characteristics, parking circulation data and many more, which come from the monitoring of more than 200000 trips performed by more than 3000 drivers in large road networks. The aim of this paper is to identify unsafe driving behaviors and then group drivers based on the stability of their unsafe driving behavior. Specifically, we aim to understand whether drivers maintain a stable driving behavior or they move between different unsafe driving profiles.

The remainder of the paper is organized as follows: Section 2 provides an overview of the methodological approach followed and a comprehensive presentation of the data and the methods that are used. The most significant findings of this paper are presented in Section 3, followed by some discussion and future research steps in Section 4.

2. Methodological approach

Each driver drives differently from trip to trip but also in relation to the way other drivers drive. Therefore, driving behavior analysis is a complex process extensively researched over the past decades. In this paper, in order to disentangle this process, the methodological steps shown in Figure 1 were followed.

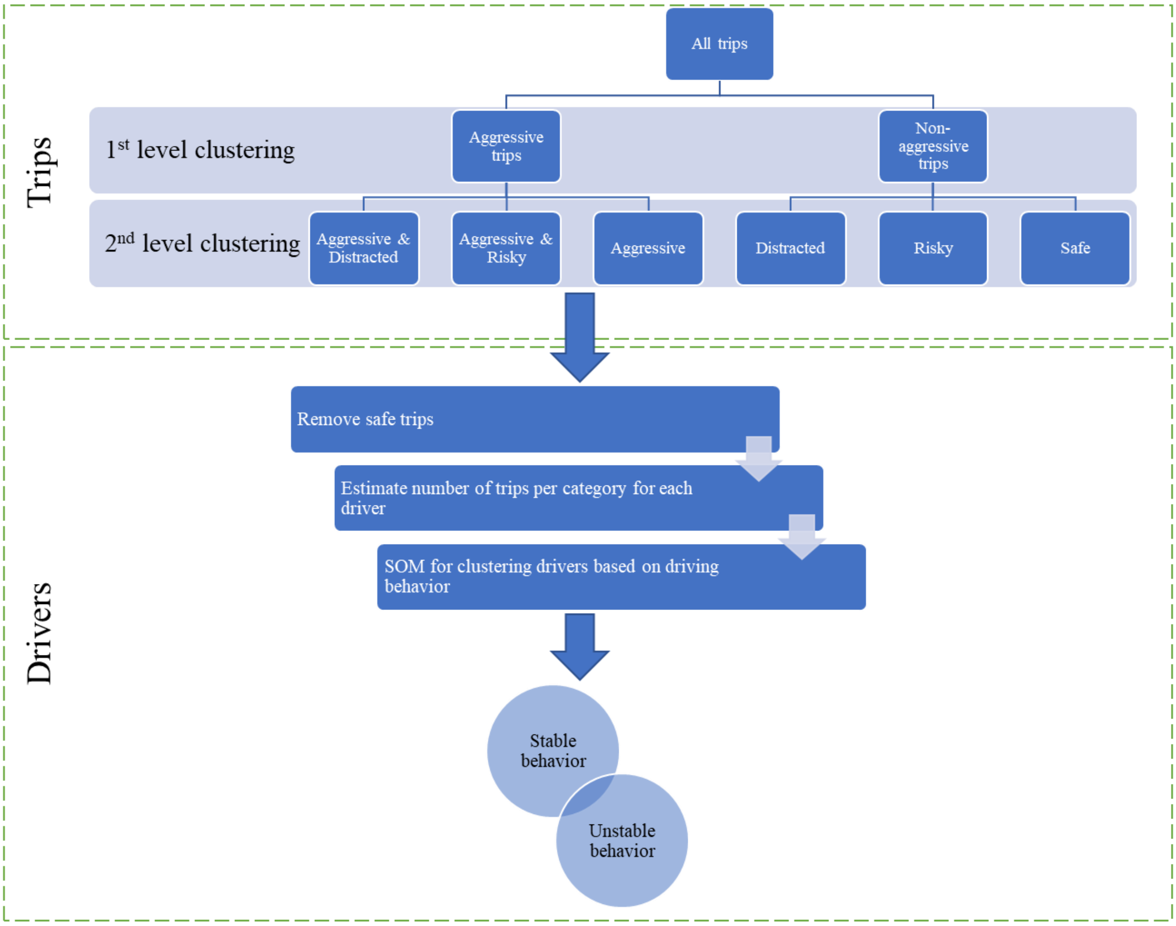


***Figure 1:** Methodological approach for grouping drivers based on unsafe driving behavior*

First, K-means clustering algorithm was applied in order to identify driving profiles with regards to driving safety. Unsafe driving profiles include driving aggressively, using a mobile phone and thus, being distracted from the driving task, as well as having risk taking behaviors such as driving with an excessive speed. Then, the analysis focuses on the unsafe driving behaviors, and therefore safe trips were removed from the dataset. Finally, in order to identify group of drivers who have similar driving behavior over time, Self-Organizing maps were developed and a k-means clustering algorithm was applied in order to separate drivers in two main groups.

2.1 K-means Clustering

Clustering is a well-known task of dividing a set of observations into a number of groups so that the observations within the same group are similar. The most widely used clustering technique is K-means clustering, where a cluster can be thought as a group of data points whose interpoint distances are small compared with the distances of points outside of the cluster. For each data point X_n, a corresponding set of binary indicator variables $r_{nk} \in \{0.1\}$ are introduced, where $k = 1, \ldots\ldots, K$ describing which of K clusters the data point X_n is assigned to, so that if a data point is assigned to cluster k then $r_{nk} = 1$, and $r_{nj} = 0$ for $j \neq k$. Then, an objective function is defined, given by:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n - \mu_k\|^2 \qquad (1)$$

Which represents the sum of squares of the distances of each data point to its assigned vector μ_k, where μ_k represents the center of the k^{th} cluster. The goal is to find values for $\{r_{nk}\}$ and the $\{\mu_k\}$ so as to minimize J (Bishop, 2006).

2.2 Self-Organizing Maps

Self-Organizing Maps (SOMs) are an unsupervised data visualization method usually used to visualize high-dimensional data in 2- dimensional maps. They were first introduced by Finnish professor Teuvo Kohonen in 1980 and is basically a method to do dimensionality reduction, and thus, sometimes are called the Kohonen maps (Oja & Kaski, 1999).
The SOMs are a special class of neural networks which are trained using competitive learning, where the output neurons compete among themselves to become active (winner neuron). In competitive learning only one neuron can be active as opposed to Hebbian learning rule. This feature helps at discovering statistically salient features that may be used for the classification of the inputs. SOMs differ from other competitive neural networks since weights can be corrected not only for the winning neuron but also for the neighboring neurons (Principe et al., 2000). This fact is critical because it allows the regeneration of the features of the input in a topological order which means that data with similar characteristics are placed in the same area of the self-organizing map (Vlahogianni et al., 2005).

2.3 The dataset

The data used in this work has been provided by OSeven Telematics, a company that works in the fields of insurance telematics and driving behaviour analysis. OSeven has developed a complete system for the recording, evaluation, storage, presentation of driving data, including machine learning algorithms, driving scoring models and gamification schemes. The data recording is carried out through the OSeven smartphone application for both iOS and Android operational systems. The application exploits smartphone's embedded sensors in order to collect valuable data concerning among others, trip characteristics, driving behaviour, eco behavior and parking circulation. The app is constantly running in the background of the smartphone and thus data is collected without requiring any user input. The raw data gathered from smartphone's sensors such as accelerometer, gyroscope and GPS sensors, are uploaded to the server for storage after being anonymized and then several techniques are applied in order to remove data noise (Vlahogianni & Barmpounakis, 2017). All data was provided by OSeven Telematics in a fully anonymized format.
The application tracks trips performed by all means of transport and therefore those trips which have been identified using another mode rather than car have been removed. The dataset initially included 245000 trips performed by more than 3000 drivers. For the purpose of this work 200 drivers were chosen from the whole dataset with the criterion to have performed more than 300 trips each.
For each trip, statistical measurements of acceleration and deceleration are estimated that describe how smoothly the driver changes their longitudinal position. In addition, speeding

measurements are collected, namely the percentage of trip duration when the driver drives over the speed limit. Finally, mobile usage indicators are estimated that describe how cautious the driver is.

3. Findings

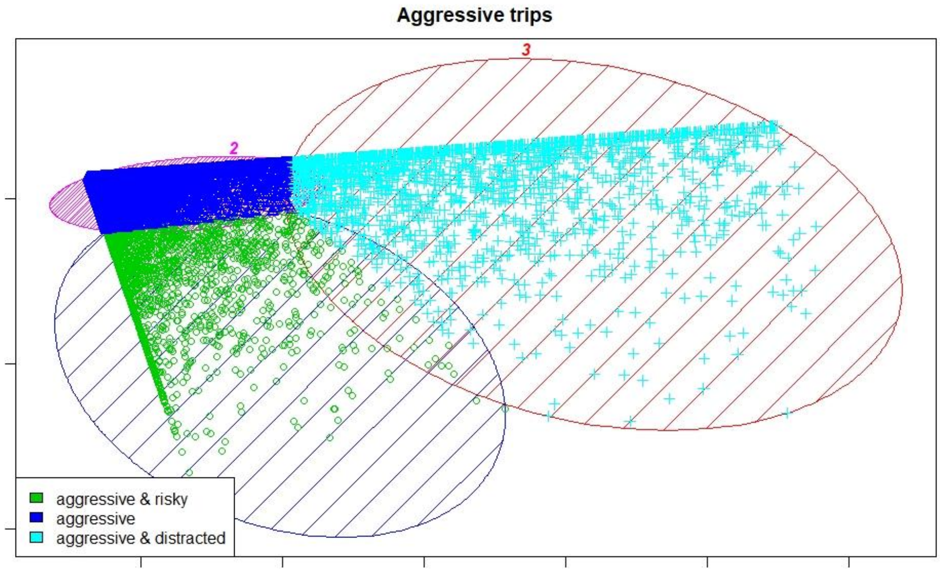
The identification of driving safety profiles has been already done before, and the methodological approach has been presented thoroughly in Mantouka et al. (2019). For the identification of different driving profiles, a two-level clustering approach was used. In the first level, trips are categorized based on the number of harsh alterations of the longitudinal position of the vehicle (acceleration and braking), while the rest of them are essentially measures of the average acceleration of the trip. The first level of clustering resulted in separating trips in two main categories: Aggressive and Non-aggressive trips. Subsequently, a second level of clustering is performed in order to identify additional unsafe driving behaviors, namely distraction and speeding. This process resulted in identifying 6 driving profiles:

- Safe trips
- Aggressive trips
- Risky trips
- Distracted trips
- Aggressive & Risky trips
- Aggressive & Distracted trips

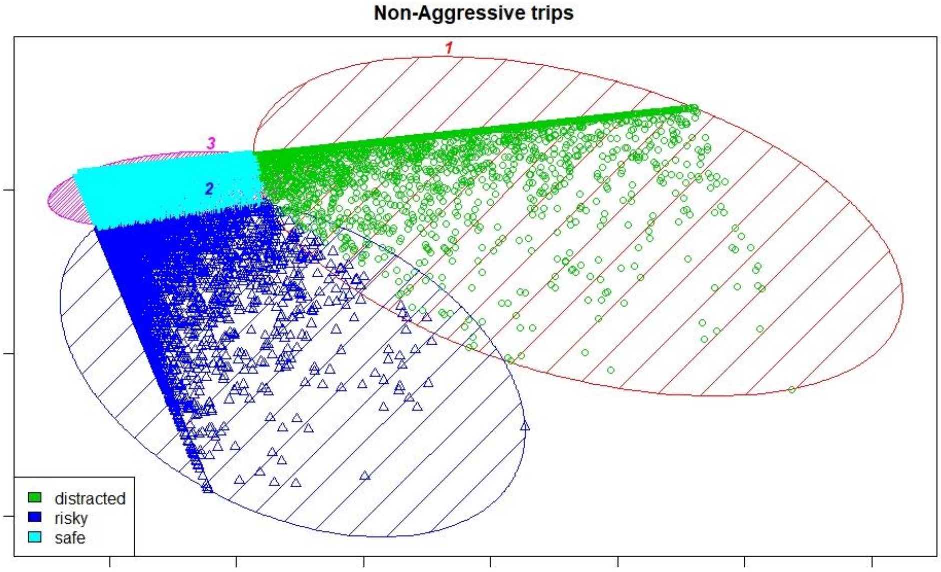
The variables used for the clustering process as well as the centers of each cluster are presented in Table 1. The results of clustering implementation are depicted in Figure 2.

***Table 1:** Cluster Centers of Aggressive and Non–aggressive Trips after 2-level clustering*

	1st level of Clustering			
Variable/ cluster	Harsh Acceleration/km	Harsh Brake/km	Smoothness Indicator	Standard Deviation of Acceleration
Aggressive Trips	0.281	1.801	0.455	0.509
Non-Aggressive Trips	0.038	1.169	0.299	0.093
	2nd level of Clustering NON-AGGRESSIVE TRIPS			
	Percent of mobile usage		Percent of speeding	
Distracted trips	0.540		0.065	
Risky trips	0.029		0.289	
Safe trips	0.013		0.024	
	AGGRESSIVE TRIPS			
Aggressive & Risky trips	0.038		0.292	
Aggressive trips	0.20		0.032	
Aggressive & Distracted trips	0.547		0.100	



(a)



(b)

Figure 2: *Two-level clustering of trips*

After assigning each trip to a single driving profile, the number of trips in each category per driver were calculated. In order to visualize unsafe driving behavior per driver (namely the number of trips in each cluster), a 2x2 SOM was created. Each node of the SOM includes more than 40 drivers. The distance between each node and its neighbors is depicted in the so-called U-Matrix which is shown in Figure 3. In this kind of plots, areas of low neighbor distance indicate groups of nodes that are similar, while areas with large distances indicate that nodes are much more dissimilar. This can also be seen as the natural boundaries between node clusters. It is clear that 1 node (yellow one) is significantly different from the rest of the nodes which seem to be really close to each other (red nodes).

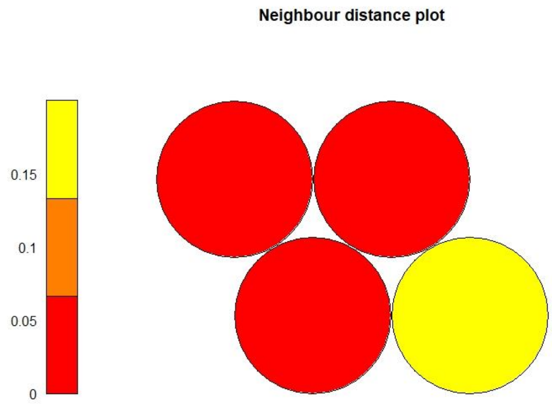


Figure 3: *Distance between nodes of SOM*

SOM provide the opportunity to visualize the normalized values of the original attributes which were used to generate the SOM in first place, the so-called "weight vectors". Weight vectors are representative of the observations included in the corresponding node. In this case as well, the visualization of the weight vectors across a plot may be indicative of the nodes' clustering. The corresponding plot is depicted in Figure 4 and the values of the weight vectors are presented in Table 2.

Table 2: *Variables in each SOM node*

	Driving profiles				
SOM component	Distracted	Risky	Aggressive-Risky	Aggressive	Aggressive-Distracted
V1	0.044	0.107	0.059	0.759	0.030
V2	0.168	0.397	0.110	0.267	0.057
V3	0.045	0.099	0.059	0.770	0.027
V4	0.042	0.114	0.058	0.757	0.029

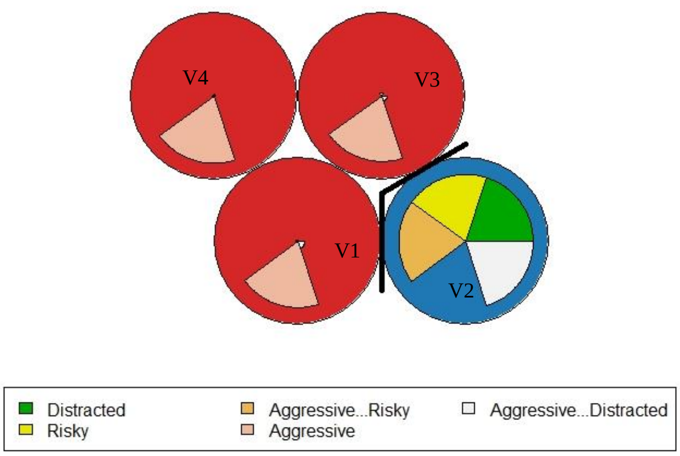


Figure 4: *Share of driving profiles within SOM clusters*

Table 2 presents the average distribution between driving profiles of the drivers represented in the corresponding node. More specifically, SOM node V2 holds all the drivers whose 17% of trips were categorized as distracted, 40% of them were categorized as risky and only 27% of them were clustered as aggressive trips. Contrary, the three other components of the SOM hold drivers who drive in an aggressive manner during the majority of their trips (around 76%) while at the same time do not seem to have any other unsafe driving behavior.

Interestingly, as shown in Figure 4, drivers can be grouped in two main categories regarding the frequency of the appearance of each unsafe driving profile over the total number of their trips. More specifically, some of the drivers seem to drive in an aggressive manner without performing any other unsafe behavior. Aggressive drivers tend to perform harsh accelerations and braking and in addition make high accelerations during their trips. On the other hand, there is another group of drivers whose behavior is unstable and they behave differently in each trip. Those drivers perform several unsafe behaviors and do not maintain a certain driving style. Specifically, they may sometimes be distracted from the driving task, drive with excessive speed for a long period of time, or even drive in an aggressive manner and simultaneously use their mobile phone and perform speeding. This group of drivers is particularly risky and needs special attention when it comes to providing recommendations to improve their behavior. Moreover, the presence of this category of drivers constitutes the development of a driving behavior prediction scheme a very complicated process which requires further driving analytics techniques and deep learning approaches.

Figure 5 depicts the 2 clusters of SOM nodes as well as the number of observations in each node. Observations of the blue cluster seem to be very similar, since they are all placed in the center of the node. On the contrary, as far as concerns the red cluster although it is clear that the observations are correctly placed in the cluster, there are some observations that move towards the burden of the nodes.

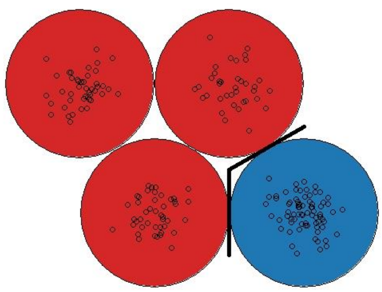


***Figure 5:** Number of drivers in each SOM-node*

4. Conclusions and future research

This paper takes advantage of an always increasing trip database which monitors hundreds of drivers, in order to identify driving profiles. The data are collected through a smartphone application using smartphones’ embedded sensors. For the identification of driving profiles, a two-level k-means clustering algorithm was applied and 6 distinct driving profiles were identified. Subsequently, Self-Organizing maps were used to investigate similarities of unsafe driving behaviors between drivers. K-Means clustering was applied, after the representation of all drivers to the SOM, and results revealed that there are two main groups of drivers as far as concerns the adoption of unsafe driving styles. On one hand, there are a large number of drivers who drive aggressively without adopting any other unsafe driving behavior. On the other hand, there are those drivers who have an unstable behavior and perform several road safety violations, such as speeding, mobile phone usage, large accelerations etc.

These findings can be really useful especially for insurance companies who are interested in identifying driving behavior of their customers in order to offer them personalized and efficient services. In the era of autonomous vehicles and mixed traffic conditions, there results can be exploited for the development of real time recommendation systems which aim to improve driving behavior in relation to safe, efficient and sustainable driving as well as prevent collisions and road accidents between the several types of vehicles existing simultaneously on the road.

Future research will focus on discovering the factors that lead to each risky behavior and understand interrelations between driving behavior and other trip's characteristics. In addition, the identification of driving profiles can be enriched with additional variables which also describe unsafe driving behavior, such as lane changing, overtaking, abnormal steering and other causes of distraction (conversations with passengers, listening to music, out of vehicle incidents etc.).

Acknowledgements

This research is part of a Doctoral Dissertation which is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project "Strengthening Human Resources Research Potential via Doctorate Research – 2nd Cycle" (MIS-5000432), implemented by the State Scholarships Foundation (ΙΚY).

The authors would like to thank Oseven Telematics for providing the data for this analysis.

5. References

Aljaafreh, A., Alshabatat, N. & Najim Al-Din, M. S. (2012). Driving style recognition using fuzzy logic, *2012 IEEE International Conference on Vehicular Electronics and Safety*, ICVES 2012, pp. 460–463

Ashbrook, D. & Starner, T. (2002). Learning significant locations and predicting user movement with GPS, *Proceedings - International Symposium on Wearable Computers, ISWC, 2002– Janua*, pp. 101–108

Bishop, C. (2006). Pattern Recognition and Machine Learning. *Springer Science+ Business Media*, New York, pp. 78-90

Johnson, D. A. & Trivedi, M. M. (2011). Driving style recognition using a smartphone as a sensor platform, *IEEE Conference on Intelligent Transportation Systems*, Proceedings, ITSC, 1609–1615

Kanarachos, S., Christopoulos & S. R. G., Chroneos, A. (2018), "Smartphones as an integrated platform for monitoring driver behaviour: The role of sensor fusion and connectivity," *Transportation Research Part C: Emerging Technologies*, March.

Mantouka, E. G., Barmpounakis, E. N., & Vlahogianni, E. I. (2019). Identification of driving safety profiles from smartphone data using machine learning techniques. *Safety Science*. https://doi.org/10.1016/j.ssci.2019.01.025

Meseguer, J. E., Toh, C. K., Calafate, C. T., Cano, J. C. & Manzoni, P. (2017). Drivingstyles: A mobile platform for driving styles and fuel consumption characterization, *Journal of Communications and Networks*, vol.19, 2, pp.162–168

Mitrović, D. (2005). Reliable method for driving events recognition, *IEEE Transactions on Intelligent Transportation Systems*, vol.6, 2, pp. 198–205

Miyajima, C., Nishiwaki, Y., Ozawa, K., Wakita, T., Itou, K., Takeda, K., & Itakura, F. (2007). Driver modeling based on driving behavior and its evaluation in driver identification. *Proceedings of the IEEE*, vol. 95, 2, pp.427-437

Montini, L., Prost, S., Schrammel, J., Rieser-Schüssler, N. & Axhausen, K. W. (2015). Comparison of travel diaries generated from smartphone data and dedicated GPS devices, *Transportation Research Procedia*, vol. 11, pp. 227–241

Oja, E., & Kaski, S. (Eds.). (1999). Kohonen maps. *Elsevier*

Papadimitriou, E., Argyropoulou, A., Tselentis, D. I. & Yannis, G. (2019). Analysis of driver behaviour through smartphone data: The case of mobile phone use while driving, *Safety Science*, January 2018

Principe J.C., Euliano N. R., & Lefebvre C. W., (2000). "Neural and Adaptive Systems: Fundamentals Through Simulations". *John Wiley and Sons*, Inc

Quek, Z. F. & Ng, E. (2013). Driver Identification by Driving Style. Technical Report, CS 229 Project, *Stanford University*

Saiprasert, C., Pholprasit, T. & Pattara-Atikom, W. (2013). Detecting Driving Events using Smartphone, *20th ITS World Congress*, October, pp. 1–12.
Saiprasert, C. & Pattara-Atikom, W. (2013). Smartphone enabled dangerous driving report system, *Proceedings of the Annual Hawaii International Conference on System Sciences*, pp. 1231–1237

Sivak, M., & Schoettle, B. (2012). Eco-driving: Strategic, tactical, and operational decisions of the driver that influence vehicle fuel economy. *Transport Policy*, Vol. 22. https://doi.org/10.1016/j.tranpol.2012.05.010

Tselentis, D. I., Yannis, G. & Vlahogianni, E. I (2017), Innovative motor insurance schemes: A review of current practices and emerging challenges, *Accident Analysis and Prevention*, vol.98, pp. 139–148

Vlahogianni, E. I., Karlaftis, M. G., & Stathopoulos, A. (2005). An extreme value based neural clustering approach for identifying traffic states. In Proceedings. *2005 IEEE Intelligent Transportation Systems*, pp. 320-325, IEEE

Vlahogianni, E. I., and E. N. Barmpounakis. Driving Analytics Using Smartphones: Algorithms, Comparisons and Challenges. (2017), *Transportation Research Part C: Emerging Technologies*, Vol. 79, pp. 196–206. https://doi.org/10.1016/j.trc.2017.03.014.