Cluster Analysis for Driver Aggressiveness Identification

Fabio Martinelli[1], Francesco Mercaldo[1], Vittoria Nardone[2], Albina Orlando[3] and Antonella Santone[4]
[1]*Istituto di Informatica e Telematica, Consiglio Nazionale delle Ricerche (CNR), Pisa, Italy*
[2]*Department of Engineering, University of Sannio, Benevento, Italy*
[3]*Istituto per le Applicazioni del Calcolo "M. Picone", Consiglio Nazionale delle Ricerche (CNR), Napoli, Italy*
[4]*Department of Bioscience and Territory, University of Molise, Pesche (IS), Italy*

Keywords: CAN, OBD, Automotive, Safety, Cluster Analysis, Machine Learning.

Abstract: In the last years, several safety automotive concepts have been proposed, for instance the cruise control and the automatic brakes systems. The proposed systems are able to take the control of the vehicle when a dangerous situation is detected. Less effort was produced in driver aggressiveness in order to mitigate the dangerous situation. In this paper we propose an approach in order to identify the driver aggressiveness exploring the usage of unsupervised machine learning techniques. A real world case study is performed to evaluate the effectiveness of the proposed method.

1 INTRODUCTION

As of 2015 there were over 263 million registered vehicles on the roads in the United States. Of those millions of registered vehicles, each year there are also millions of vehicle crashes. In 2015, there were 32,166 fatalities, 1,715,000 injuries and 4,548,000 car crashes which involved property damage. Of these fatalities, there are far more driver deaths, than passenger, pedestrian or motorcyclist deaths[1]. So while many of us feel secure in vehicles, the statistics indicate the importance of automobile insurance and in most cases, auto insurance is required by law. This is the reason why car insurance is really important because it not only covers any physical damage that may occur in an accident, but also any damage or injury that might be caused because of a vehicular accident or which may be done upon oneself or ones vehicle by another vehicle or accident, as a falling tree for example (Marotta et al., 2017).

The insurance industry is a key component of the economy by virtue of the amount of premiums it collects, the scale of its investment and, more fundamentally, the essential social and economic role it plays by covering personal and business risks.

Usage-based insurance (UBI) also known as "pay as you drive" (PAYD) and "pay how you drive" (PHYD) and mile-based car insurance is a type of vehicle insurance whereby the costs are dependent upon type of vehicle used, measured against time, distance, behavior and place (Desyllas and Sako, 2013; Tselentis et al., 2016; Kantor and Stárek, 2014) and they represent the emerging trend in the insurance area.

This represents a different approach with respect to traditional insurance, which attempts to differentiate and reward "safe" drivers, giving them lower premiums and/or a no-claims bonus. However, conventional differentiation is a reflection of history rather than present patterns of behaviour. This means that it may take a long time before safer (or more reckless) patterns of driving and changes in lifestyle feed through into premiums.

UBI programs offer many advantages to insurers, consumers and society. Linking insurance premiums more closely to actual individual vehicle or fleet performance allows insurers to more accurately price premiums (Boquete et al., 2010). This increases affordability for lower-risk drivers, many of whom are also lower-income drivers. It also gives consumers the ability to control their premium costs by encouraging them to reduce miles driven and adopt safer driving habits. Fewer miles and safer driving also aid in reducing accidents, congestion, and vehicle emissions,

[1]https://www.statista.com/topics/3087/car-insurance-in-the-united-states/

Martinelli F., Mercaldo F., Nardone V., Orlando A. and Santone A.
Cluster Analysis for Driver Aggressiveness Identification.
DOI: 10.5220/0006755205620569
In *Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP 2018)*, pages 562-569
ISBN: 978-989-758-282-0

which benefits society[2].

Starting for these considerations, in this paper we propose an approach able to characterize the driver behaviour using a set of feature gathered from the vehicle CAN bus. The proposed method considers the unsupervised machine learning i.e., the machine learning task of inferring a function to describe hidden structure from unlabeled data to discriminate between urban and highway roads. In order to perform this task, we consider cluster analysis in order to group the feature extracted from the driver under analysis: the main assumption that will be verified in the experiment is that CAN bus features gathered from the highway path exhibits different values from the ones gathered from urban road (and for this reason grouped in different clusters). As a matter of fact, as demonstrated in current literature, drivers typically exhibit different driving style in different kind of roads (Mehar et al., 2013; Wang et al., 2004). Furthermore, on the basis on the cluster analysis results, we compute an aggressiveness index of the driver under analysis in order to boost the research into "pay how you drive" possible risk assessment calculation.

We evaluate the proposed approach on a real-world dataset gathered from a vehicle running through several (urban and highway) roads.

The reminder of the paper is organized as follows: Section 2 discusses the current literature, Section 3 introduces the method, Section 4 illustrates the results of the cluster analysis based experiment. Finally, conclusions and future works are given in Section 5.

2 RELATED WORK

In the following we section review current literature related to the driving style and aggressiveness recognition.

In the past, the automotive real-world data retrieving was limited due to the difficulty to equip the sensors in cars, since the introduction of CAN this limit is overcome.

Authors in (Wakita et al., 2006) propose a driver identification method that is based on the driving behavior signals that are observed while the driver is following another vehicle. They analyze signals, as accelerator pedal, brake pedal, vehicle velocity, and distance from the vehicle in front, were measured using a driving simulator. The identification rates were 81% for twelve drivers using a driving simulator and 73% for thirty drivers.

[2]http://www.naic.org/cipr_topics/topic_usage_based_insurance.htm

Data from the accelerator and the steering wheel were analyzed by researchers in (Zhang et al., 2014). Observing the considered features, they employ hidden Markov model (HMM) to model the driver characteristics. They build two models for each driver, one trained from accelerator data and one learned from steering wheel angel data. The models can be used to identify different drivers with an accuracy equal to 85%.

Researchers in (Kedar-Dongarkar and Das, 2012) classify a set of features extracted from the powertrain signals of the vehicle, showing that their classifier is able to classify the human driving style based on the power demands placed on the vehicle powertrain with an overall accuracy of 77%.

Van Ly et alius (Van Ly et al., 2013) explore the possibility of using the inertial sensors of the vehicle from the CAN bus to build a profile of the driver observing braking and turning events to characterize an individual compared to acceleration events.

Researchers in (Miyajima et al., 2007; Nishiwaki et al., 2007) model gas and brake pedal operation patterns with Gaussian mixture model (GMM). They achieve an identification rate of 89.6% for a driving simulator and 76.8% for a field test with 276 drivers, resulting in 61% and 55% error reduction, respectively, over a driver model based on raw pedal operation signals without spectral analysis.

Driver behavior is described and modeled in (Choi et al., 2007) using data from steering wheel angle, brake status, acceleration status, and vehicle speed through Hidden Markov Models (HMMs) and GMMs employed to capture the sequence of driving characteristics acquired from the CAN bus information. They obtain 69% accuracy for action classification, and 25% accuracy for driver identification.

In reference (Meng et al., 2006) the features extracted from the accelerator and brake pedal pressure are used as inputs to a fuzzy neural network (FNN) system to ascertain the identity of the driver. Two fuzzy neural networks, namely, the evolving fuzzy neural network (EFuNN) and the adaptive network-based fuzzy inference system (ANFIS), are used to demonstrate the viability of the two proposed feature extraction techniques.

A hidden-Markov-model-(HMM)-based similarity measure is proposed in (Enev et al., 2016) in order to model driver human behavior. They employ a simulated driving environment to test the effectiveness of the proposed solution.

Authors in (Kwak et al., 2016) propose a method based on driving pattern of the car. They consider mechanical feature from the CAN vehicle evaluating them with four different classification algo-

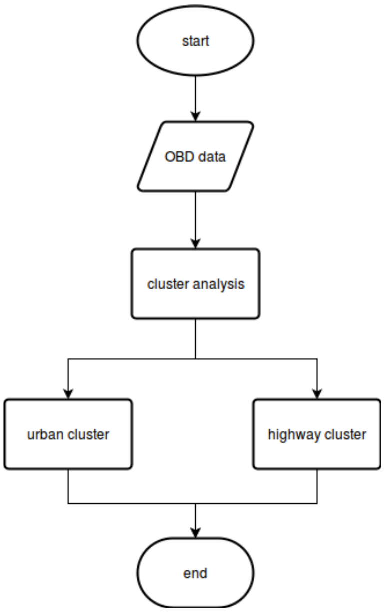

Figure 1: Flow diagram of the proposed approach for risk index computation.

rithms, obtaining respectively an accuracy equal to 0.939 with Decision Tree, equal to 0.844 with KNN, equal to 0.961 with RandomForest and equal to 0.747 using MLP algorithm.

The proposed method, differently from the ones explained in this section, consider the road identification issue in order to compute two different aggressiveness indexes (related to the urban and to highway roads). In addition, relating to road identification task, we highlight that the classification task does not require previous knowledge about the type of road since a cluster analysis algorithm is considered.

3 THE METHOD

In following section we describe the considered approach in order to evaluate the driving style (in terms of aggressiveness) from a set of features extracted from the in-vehicle CAN data and from GPS sensor.

Figure 1 depicts the flow diagram of the proposed approach for the road identification.

As Figure 1 shows, the cluster analysis process is concerned to OBD data (Martinelli et al., 2018) gathered by the CAN bus in order to label the data as belonging to the urban or to the highway roads.

In our analysis we considered the feature set (belonging to OBD and to GPS) shown in Table 1.

We considered features gathered from different sources: the first source is represented by the OBD (i.e. F1, F2, F3, F4, F5, F6 and F7) while the second one is computed by the user device GPS sensor (i.e. F8 and F9).

The GPS sensor features are considered to add meta information useful to have the confirmation about the kind of road (i.e., urban or highway) identified by the cluster analysis using the F8 and F9 features.

As stated into the introduction in order to characterize the driver style in terms of aggressiveness we resort to an unsupervised machine learning approach i.e., cluster analysis that, differently from the supervised machine learning does not require a labelled dataset to perform the classification task (Cimitile et al., 2017; Canfora et al., 2014; Canfora et al., 2016; Battista et al., 2016; Mercaldo et al., 2016). The cluster analysis itself is not one specific algorithm, but the general task to be solved: it can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them (Kaufman and Rousseeuw, 2009). Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions.

In this paper we consider the k-means algorithm (MacQueen et al., 1967), one of the simplest unsupervised learning algorithms that solve the well known clustering problem (Har-Peled and Kushal, 2007).

The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters, with particular regards to the designed approach we consider k=2) fixed a priori. The main idea is to define 2 centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result (Jain, 2010). So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done (Arthur et al., 2009). At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid (MacQueen et al., 1967). A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more.

We consider the k-means implementation in the Weka data mining toolkit[3] i.e., SimpleKMeans. This implementation can use either the Euclidean distance (default) or the Manhattan distance. Euclidean, in

[3]https://www.cs.waikato.ac.nz/ml/weka/

Table 1: Features involved in the study.

Feature	Description	Info	OBD	GPS
F1	Engine RPM	Revolutions Per Minute	X	
F2	Mass Air Flow	expressed in g/s	X	
F3	Instantaneous Fuel Consumption	expressed in liters/100 km	X	
F4	Boost pressure estimation	expressed in KPa/Bar/Kg	X	
F5	Acceleration	expressed as g (gravity)	X	
F6	Engine power	expressed in KW	X	
F7	Engine torque	expressed in NM/Kg	X	
F8	Altitude	expressed in degree		X
F9	Longitude	expressed in degree		X

this study we set the SimpleKMeans algorithm with the Euclidean distance, maximum iterations number equal to 500 and maximum of generated clusters equal to 2. Since the features given to the learner are unlabeled, there is no evaluation of the accuracy of the structure that is output by the relevant algorithm (this is one way of distinguishing unsupervised learning from supervised learning): for this reason we consider the incorrectly clustered instances number and percentage in order to evaluated the goodness of the proposed method (i.e., to evaluate whether the first cluster contains the majority of urban while the second one contains the majority of highway ones).

Once evaluated the k-means algorithm in order to distinguish between feature gathered while the driver is traveling on urban roads and features gathered while the driver is traveling on highway ones, we discuss an approach to apply this information to provide an aggressiveness index for the PHYD car insurance.

4 EXPERIMENTAL EVALUATION

In this section we discuss the experiment we performed related to cluster analysis in order to classify between urban and highway paths.

The evaluation consists of two stages: (i) a comparison of descriptive statistics of the populations of feature and (ii) an unsupervised classification analysis aimed at assessing whether the urban and highway features are grouped in different clusters.

We realize a real-world dataset, gathering data from the in-vehicle CAN bus. The vehicle involved in the experiment is a Fiat Punto Evo 1.3 Diesel with 75 horsepowers with one driver.

In order to collect data, the DashCommand (OBD ELM App)[4] application and Mini Bluetooth ELM327 OBD 2 Scanner were used.

OBD is available on modern car to produce the self-diagnostic report by monitoring vehicle system in terms of measurement and vehicle failure (Martinelli et al., 2017). The data are recorded every 1 second during driving using the DashCommand application by an Android smartphone (i.e., a Huawei p8 lite 2017 with Android 7.0 Nougat onboard) fixed in the car by a support.

In order to label the track using the “urban” or the “highway” label we developed a Java script able to generates an address from a latitude and longitude through the reverse geocoding Java wrapper [5] able to query the Nominatim search engine for OpenStreetMap data[6].

We collected data from the vehicle in an urban an a highway area in Italy, in Figure 2 the urban path considered: it consists of 22 Km from the Istituto di Informatica e Telematica in Pisa to Cascina, in the center of Italy. The highway path 3 is related to the main Italian highway (the A1, i.e. Autostrada del Sole) between the Center and the South of Italy and it consists of 234 Km. In order to balance to traveled kilometers between the urban and the highway paths, we have considered 10 urban paths (i.e., ten different travels of the urban path of 22 Km) and one highway path: in this way we have a dataset composed of 220 Km of urban path and 234 Km of highway path.

We represent two scatterplots with the aim to give statistical evidence that considered feature population exhibit different trend between the urban and the highway populations. Similar consideration can addressed for the other features.

Figure 4 shows the scatterplot related to the Engine RPM (i.e., the F1 feature) and Boost pressure estimation (i.e. the F4 feature): Engine RPM feature is represented on the X axis while the Boost pressure estimation one on the Y axis.

The red distribution is related to the urban path, while the blue one is related to the highway path: from the scatterplot it is clear the division between the red points, mostly allocated on the center-low left

[4]https://play.google.com/store/apps/details?id=com.palmerperformance.DashCommand&hl=it

[5]https://www.daniel-braun.com/technik/reverse-geocoding-library-for-java/

[6]http://nominatim.openstreetmap.org/

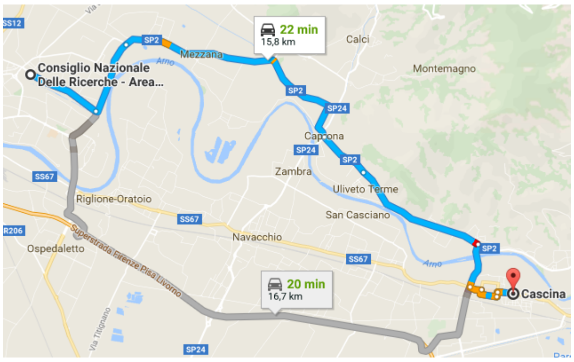

Figure 2: The urban path considered in the study highlighted in blue: it consists of 22 Km from the Istituto di Informatica e Telematica in Pisa to Cascina, in the center of Italy.

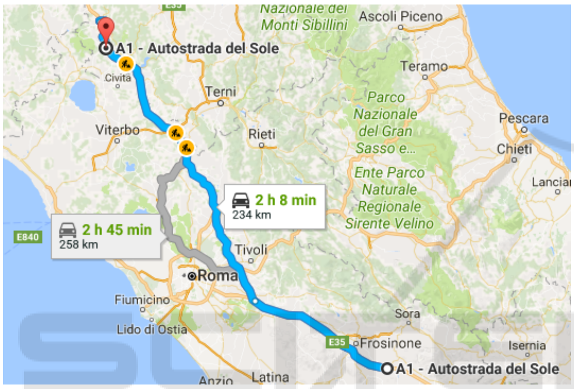

Figure 3: The highway path considered in the study highlighted in blue: it is related to the main Italian highway between the Center and the South of Italy and it consists of 234 Km.

side of the scatterplot, and the blue one, mostly allocated on the high and low right side of the scatterplot.

Figure 5 shows the scatterplot related to the F1 feature and the F7 one i.e., the Engine torque: the F1 feature is represented on the X axis while the F7 one on the Y axis.

As in Figure 4, also in the scatterplot in Figure 5 the red distribution is related to the urban path, while

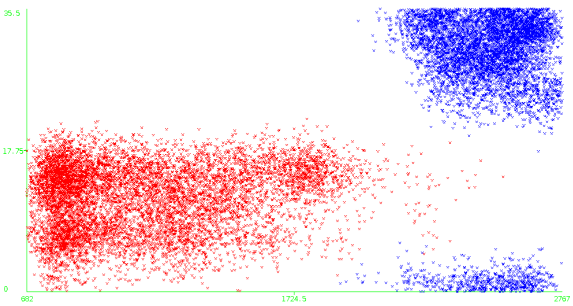
Figure 4: Scatterplot related to the F1 feature and the F4 feature (the red distribution is related to the urban path, while the blue distribution in related to the highway path).

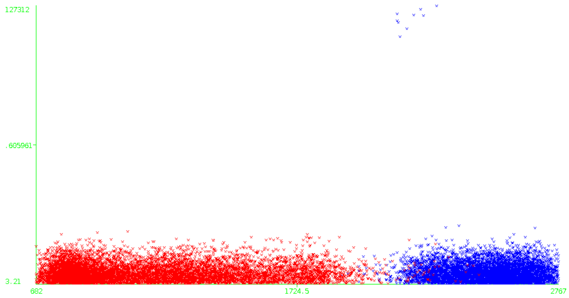
Figure 5: Scatterplot related to the F1 feature and the F7 feature (the red distribution is related to the urban path, while the blue distribution in related to the highway path).

the blue one is related to the highway path. In this case both the red and the blue distributions are allocated in the down side of the graph, but also in this case their division between is clear: the red points are in the left and middle part of the scatterplot, while the blue points are most allocated in the left side.

From the considerations related to Figures 4 and 5 we state that the features under analysis can be useful to discriminate between urban and highway paths and, consequently, good candidates for the cluster analysis phase.

Relating the unsupervised classification, we compute the incorrectly clustered instances number and percentage in three different scenarios (i.e., we perform three different clustering experiments) with following instances:

- C1: instances related only to the urban path;
- C2: instances related only to the highway path;
- C3: instances related to the urban and highway path (i.e., the full dataset);

The aim on this experiment is to demonstrate that the cluster analysis is useful to discriminate between urban and highway path (the C3 scenario) producing two different clusters, while the performances obtained in the unsupervised classification related to urban (i.e., C1) and highway (i.e., C2) paths are not able to obtain a good clustering (i.e., if we fix the number of clusters to generate as two, when we consider only urban/highway instances the cluster algorithm is not able to create the first cluster with all the urban/highway instances and to generate the second cluster with few instances).

We consider three difference instance set (i.e., C1, C2 and C3) with the aim to demonstrate that the more appropriate clusters are obtained using the C3 instances (related to the urban and highway path).

Table 2 shows the results of the C1, C2 and C3 unsupervised classifications.

As shown in Table 2, the C1 experiment (with only urban path instances) obtains an Incorrectly cluste-

Table 2: Results of the C1, C2 and C3 experiments.

Exp.	ICI	%	time
C1	5551	63.4545%	0.06
C2	8735	83.5437%	0.09
C3	444	2.4636%	0.06

red instances value equals to 5551 (i.e., 63% of the instances considered); the C2 experiment (with only highway path instances) obtains an Incorrectly clustered instances (ICI) value equal to 8735 (i.e., 83% of the instances considered) while the C3 experiment (with both urban and highway paths instances) obtain an Incorrectly clustered instances value equals to 444 with a percentage of incorrectly clustered instances equal to 2%.

These results demonstrate that the adoption of the unsupervised machine learning techniques is promising: as a matter of fact, considering the different driving styles that should be adopted in urban and highway roads, we can consider the Incorrectly clustered instances value as an estimator of the driving style. In case this value is low, the driver exhibits a different driving style between urban and highway paths and this is the result of the different driving style that should be adopted in different roads. From the other side, whether the Incorrectly clustered instances value exhibits an high value (for instance, in the C1 and C2 experiment), as we demonstrated cluster analysis is not able to correctly define the clusters (C1 and C2 experiment), and this is symptomatic that the driver under analysis exhibits a driving style pretty similar in urban and highway roads and the feature set considered is representative of the kind of traveled roads.

Once obtained the clusters with regards to the urban and to the highway path, in order to compute the driver aggressiveness index in urban and highway paths we consider the acceleration feature (i.e., F5) variation: this is the reason why we resort to the *standard deviation* statistical dispersion index i.e., an estimate of the variability of a data population or a random variable (in this case the variable is represented by the F5 feature).

Considering u_i the value of the i-th urban path occurrence of the F5 feature, N_u the total number of urban path occurrences of the F5 feature (with $1 \leq i \leq N_u$) we defined the driver aggressiveness index σ_{urban} in urban path as follows:

$$\sigma_{urban} = \sqrt{\frac{\sum_{i=1}^{N_u}(u_i - \bar{x}_{urban})^2}{N_u}}$$

where $\bar{x}_{urban}$ represents the arithmetic mean of F5 feature urban path distribution and it is defined as:

$$\bar{x}_{urban} = \frac{1}{N_u}\sum_{i=1}^{N_u} u_i$$

Relating to the driver aggressiveness index $\sigma_{highway}$ in highway path, considering h_k the value of the k-th highway path occurrence of the F5 feature, N_h the total number of highway path occurrences of the F5 feature (with $1 \leq i \leq N_h$), we define the $\sigma_{highway}$ index as follows:

$$\sigma_{highway} = \sqrt{\frac{\sum_{i=1}^{N_h}(h_k - \bar{x}_{highway})^2}{N_k}}$$

where $\bar{x}_{highway}$ represents the arithmetic mean of F5 feature highway path distribution and it is defined as:

$$\bar{x}_{highway} = \frac{1}{N_h}\sum_{i=1}^{N_k} u_k$$

We compute the driver aggressiveness indexes: with regards to the urban path σ_{urban} we obtained following value:

$$\sigma_{urban} = 6.4734$$

while relating to the driver aggressiveness index in highway path we obtain following value:

$$\sigma_{highway} = 2.4519$$

From these results we deduce that the driver under analysis exhibit a drive style more aggressive in the urban path (with $\sigma_{urban} = 6.4734$) than in the highway one (i.e., $\sigma_{urban} = 2.4519$).

This behaviour can be considered as normal, as a matter of fact, typically urban roads require more accelerations and decelerations if compared to the highway ones.

The opposite behavior would be considered highly aggressive.

5 CONCLUSION AND FUTURE WORK

In this paper, starting from the consideration that depending on the type of road run (i.e., urban and highway) drivers adopt different driving style, we propose an approach to compute the driver aggressiveness. The aim of the proposed approach is to identify the kind of road traveled through unsupervised machine learning in order to identify the driver aggressiveness index is urban and highway paths. In order to evaluated the cluster analysis method to discern between urban and highway data, we use a set of features extracted from the CAN bus of real-world car while traveling in different roads (i.e., urban and highway) in the center and south of Italy. As future work we plan to adopt formal verification techniques aimed to identify whether a driver can be classified in several predefined categories (for instance: the young driver, the ruthless driver, the cautious driver) in order to propose a risk index considering the category to which a driver belongs.

ACKNOWLEDGMENTS

This work has been partially supported by H2020 EU-funded projects NeCS and C3ISP and EIT-Digital Project HII and PRIN “Governing Adaptive and Unplanned Systems of Systems” and the EU project CyberSure 734815.

REFERENCES

Arthur, D., Manthey, B., and Röglin, H. (2009). k-means has polynomial smoothed complexity. In *Foundations of Computer Science, 2009. FOCS’09. 50th Annual IEEE Symposium on*, pages 405–414. IEEE.

Battista, P., Mercaldo, F., Nardone, V., Santone, A., and Visaggio, C. A. (2016). Identification of android malware families with model checking. In *ICISSP*, pages 542–547.

Boquete, L., Rodríguez-Ascariz, J. M., Barea, R., Cantos, J., Miguel-Jiménez, J. M., and Ortega, S. (2010). Data acquisition, analysis and transmission platform for a pay-as-you-drive system. *Sensors*, 10(6):5395–5408.

Canfora, G., Medvet, E., Mercaldo, F., and Visaggio, C. A. (2016). Acquiring and analyzing app metrics for effective mobile malware detection. In *Proceedings of the 2016 ACM on International Workshop on Security And Privacy Analytics*, pages 50–57. ACM.

Canfora, G., Mercaldo, F., Visaggio, C. A., and Di Notte, P. (2014). Metamorphic malware detection using code metrics. *Information Security Journal: A Global Perspective*, 23(3):57–67.

Choi, S., Kim, J., Kwak, D., Angkititrakul, P., and Hansen, J. H. (2007). Analysis and classification of driver behavior using in-vehicle can-bus information. In *Biennial Workshop on DSP for In-Vehicle and Mobile Systems*, pages 17–19.

Cimitile, A., Martinelli, F., and Mercaldo, F. (2017). Machine learning meets ios malware: Identifying malicious applications on apple environment. In *ICISSP*, pages 487–492.

Desyllas, P. and Sako, M. (2013). Profiting from business model innovation: Evidence from pay-as-you-drive auto insurance. *Research Policy*, 42(1):101–116.

Enev, M., Takakuwa, A., Koscher, K., and Kohno, T. (2016). Automobile driver fingerprinting. *Proceedings on Privacy Enhancing Technologies*, 2016(1):34–50.

Har-Peled, S. and Kushal, A. (2007). Smaller coresets for k-median and k-means clustering. *Discrete & Computational Geometry*, 37(1):3–19.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666.

Kantor, S. and Stárek, T. (2014). Design of algorithms for payment telematics systems evaluating driver’s driving style. *Transactions on Transport Sciences*, 7(1):9.

Kaufman, L. and Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons.

Kedar-Dongarkar, G. and Das, M. (2012). Driver classification for optimization of energy usage in a vehicle. *Procedia Computer Science*, 8:388–393.

Kwak, B. I., Woo, J., and Kim, H. K. (2016). Know your master: Driver profiling-based anti-theft method. In *PST 2016*.

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA.

Marotta, A., Martinelli, F., Nanni, S., Orlando, A., and Yautsiukhin, A. (2017). Cyber-insurance survey. *Computer Science Review*.

Martinelli, F., Mercaldo, F., Nardone, V., Orlando, A., and Santone, A. (2018). Who’s driving my car? a machine learning based approach to driver identification. In *ICISSP*.

Martinelli, F., Mercaldo, F., Nardone, V., and Santone, A. (2017). Car hacking identification through fuzzy logic algorithms. In *Fuzzy Systems (FUZZ-IEEE), IEEE International Conference on. IEEE*. IEEE.

Mehar, A., Chandra, S., and Velmurugan, S. (2013). Speed and acceleration characteristics of different types of vehicles on multi-lane highways. *European Transport*, 55:1825–3997.

Meng, X., Lee, K. K., and Xu, Y. (2006). Human driving behavior recognition based on hidden markov models. In *Robotics and Biomimetics, 2006. ROBIO’06. IEEE International Conference on*, pages 274–279. IEEE.

Mercaldo, F., Nardone, V., Santone, A., and Visaggio, C. A. (2016). Download malware? no, thanks. how formal methods can block update attacks. In *Formal Methods in Software Engineering (FormaliSE), 2016 IEEE/ACM 4th FME Workshop on*, pages 22–28. IEEE.

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

Nishiwaki, Y., Ozawa, K., Wakita, T., Miyajima, C., Itou, K., and Takeda, K. (2007). Driver identification based on spectral analysis of driving behavioral signals. In *Advances for In-Vehicle and Mobile Systems*, pages 25–34. Springer.

Tselentis, D. I., Yannis, G., and Vlahogianni, E. I. (2016). Innovative insurance schemes: pay as/how you drive. *Transportation Research Procedia*, 14:362–371.

Van Ly, M., Martin, S., and Trivedi, M. M. (2013). Driver classification and driving style recognition using inertial sensors. In *Intelligent Vehicles Symposium (IV), 2013 IEEE*, pages 1040–1045. IEEE.

Wakita, T., Ozawa, K., Miyajima, C., Igarashi, K., Katunobu, I., Takeda, K., and Itakura, F. (2006). Driver identification using driving behavior signals.

IEICE TRANSACTIONS on Information and Systems, 89(3):1188–1194.

Wang, J., Dixon, K., Li, H., and Ogle, J. (2004). Normal acceleration behavior of passenger vehicles starting from rest at all-way stop-controlled intersections. *Transportation Research Record: Journal of the Transportation Research Board*, (1883):158–166.

Zhang, X., Zhao, X., and Rong, J. (2014). A study of individual characteristics of driving behavior based on hidden markov model. *Sensors & Transducers*, 167(3):194.