AN ARTIFICIAL INTELLIGENCE BASED PLATFORM FOR DRIVER BEHAVIOUR EVALUATION AND PREDICTION USING MINIMAL SENSING INFORMATION

[1]MOHAMED JAAFAR, [2]FLORIAN JOUY, [3]MAXENCE GALOPIN, [4]JESUS ZEGARRA FLORES

[1,2,3,4]V3R, Département de recherche d'Altran, 6, AV des Usines, 90000, Belfort
E-mail: [1] mohamed.jaafar@altran.com

Abstract - With the advent of artificial intelligence and data, intelligent vehicles have become a central matter both in academia and industry for the promises they exhibit. Through this paper, we explore the possibility of building a system that allows a personalized driving experience. After collecting non-invasive minimal sensing information, the driver can be identified through his profile where his own driving style is characterized and evaluated using both supervised and unsupervised machine learning approaches. Results are assessed in terms of intelligible driving behavior related features. Based on the extracted information, individual-tailored eco-driving recommendations can be displayed online or after the drive.

Keywords - Artificial Intelligence, Driving Analysis, Driving Style Evaluation, Driving Prediction, Intelligent Vehicles

I. INTRODUCTION

A. Context

Designing and implementing data-driven approaches to personalize and enhance one's driving experience is an increasingly explored concept. While data traditionally stems directly from on-board instrumentation, in the past few years, smartphones have been used as well to collect car sensor data directly through the vehicle [1,2]. Alternatively, data collection can be performed using smartphones' sensing elements themselves. Both approaches have been combined by Jachimczyk *et al* [3]. Such data are used for several purposes, including driver identification. Thus, Van Ly *et al.* [4] used inertial sensors to perform a driver recognition task whereas Gahr *et al.* proposed an approach based on steering wheel angle measurements [5]. Several other methodologies have been explored in this perspective [6]-[8]. Researches on driving style recognition and analysis were also performed in order to characterize an individual's way of controlling a vehicle [9]-[12]. Driving style assessment is another task that can be achieved. Some researches use the vehicle's data in order to improve one's driving behavior, [3]. This idea has been explored by Johnson and Trivedi, they proposed a smartphone application "MIROAD", to make the driver aware of its own aggressiveness in real time, [13]. Rui *et al.* designed another application helping to reduce an individual's fuel consumption by providing direct feedback [14]. Such approaches can be linked to researches led in [15] where results suggested that feedback provided by a relevant smartphone application helps reducing driver's fuel consumption. Another extensively research topic consists in personalized driving cycle generation, which are supposed to deliver more authentic consumption results than the standard ones. Realistic driving cycle generation has been investigated through several angles, including statistical [16, 17], clustering [18] and optimization based approaches [19].

B. Contribution

Through this work, we aim to build a system gathering most of the task quoted previously in a unique software platform using minimal sensing information. A smartphone application designed by Altran Technologies is used to collect data in a non-invasive way. Such information is then used to identify the driver, characterize and evaluate his own driving style regarding easily understandable features while providing feedback to adapt its behaviour to reach a determined objective (for instance, maximal eco-driving). These tasks are performed using both supervised and unsupervised learning techniques. Data acquisition and processing is exposed together with our general methodology concerning each task in the second paragraph, the third paragraph displays our experiments and results. The fourth paragraph states our overall conclusions while the last one allows us to cast a glance into future improvements and perspectives.

II. METHODOLOGY

A. Data acquisition and processing

Data stemming from two sources is gathered. The former source is an android smartphone application developed by Altran Technologies. This application is collecting time, GPS data and vehicle's instantaneous speed measurements at a 1 Hz rate during the trip. This application was used with different types of smartphone (Samsung and Xiaomi) under Android 9. The second source consists in a video stream displaying time, GPS, and vehicle's instantaneous speed information recorded by a board camera facing the road (which can be discarded from the study to make the system non-invasive). Using image

processing techniques and optical character recognition, data can be extracted at a sampling time of 1 Hz from the recordings. All data is then formatted into JSON files, where each file accounts for a trip. As we had to deal with strongly-varying trips' lengths, a slicing algorithm has been designed to perform data homogenization and augmentation in terms of length. The resulting slices are determined by the slicing window's length d_μ and the overlapping time with the previous slice d_r. For the sake of our experiments, we chose d_μ = 90 s and d_r = 15 s as a trade-off between length of the slice and quantity of available data. Afterwards, hand-designed features noted $\{C_i\}$ relative to driving are extracted from each sliced trip. Those features consist of:

- Statistical values such as standard deviation of speed and extrema ($km.h^{-1}$), average and standard deviation for acceleration ($km.h^{-1}.s^{-1}$) and jerk ($km.h^{-1}.s^{-2}$).
- Counts, mean and standard deviation of magnitude relating to braking followed by acceleration (BTA: Brake To Accelerate) or accelerations followed by braking (ATB: Accelerate To Brake). Except for the counts which have no unit, those features are in $km.h^{-1}.s^{-1}$.
- Relative Positive Acceleration (RPA, in $km.h^{-2}.s^{-1}$), Acceleration and Deceleration shifts (ADS, which consists in counting how many times the magnitude of acceleration exceeds $3.6\ km.h^{-1}.s^{-1}$ in absolute value, there is no unit, [20,21]), Positive Kinetic Energy (PKE, in $km.h^{-2}$).

We then form data sets using each slice as an observation and the features extracted from this slice as the columns. A label is added to know who the driver was. We collected approximately two hours of driving (≈150 km) from three drivers on the same type of road (urban, extra-urban roads and highways). Data from drivers 1 and 3 has been collected from the application, whereas driver 2's data stems from the board camera.

B. Driver identification & recognition system

1) Support Vector Classification

In this section, we discuss the driver identification and recognition problematic. A supervised classification algorithm is trained on a dataset comprising the labels of all known drivers. Specifically, after some preliminary tests, we decided to use a Support Vector Classifier (SVC) considering its relatively good performances on small datasets (which is our case). This algorithm attempts to find an optimal separating hyperplane between multiple classes such that margin between samples and boundary is maximised. This method is initially designed for a two-class problem but can be extended to more classes using one-vs-one or one-vs-rest approaches. Finding the optimal hyperplane is equivalent to solving the following constrained quadratic optimization problem:

$$\min_{w,b,\{\xi_i\}} \frac{1}{2}\|w\|^2 + \frac{C}{2}\sum_{i=1}^{m}\xi_i^2 \quad (1)$$

With constraints: $\forall\, i, (w^T\phi(x_i) + b)y_i \geq 1 - \xi_i$

$\forall i, \xi_i \geq 0$

Where we considered: a labelled dataset $\{x_i, y_i\}$, parameters w, b, a margin of $\frac{2}{\|w\|}$, a L2-regularization coefficient C, slack variables $\{\xi_i\}$ and ϕ, a mapping to treat non-linearly separable data. Thanks to the kernel trick, it can be shown that the calculations involving ϕ are replaced by computing the output of a known kernel function K which is a SVC's parameter. Such a problem can be solved in its dual form, which can be obtained using Lagrange's multipliers method.

2) Driver recognition system

Using the trained support vector classifier, known drivers can be classified provided features extracted from a sliced trip of length d_μ. This classifying block can be inserted in the configuration of Fig. 1 shown below in order to build a dynamical driver recognition system:

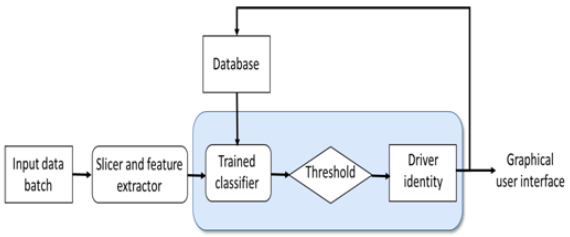


Fig. 1: Driver recognition's system global architecture.

The system works as follows: when driving data enters the system, it is pre-processed and fed to the train classifier. Afterwards, for each processed batch's slice, a vector describing the probability of belonging to one of the classes (including the unknown class, which depends on a threshold as shown on Fig. 1) is computed. The highest probability is chosen for each observation and compared to the threshold. A majority voting is then performed across all observations to determine the driver for the current data batch. If the driver is known, the current batch is used to upload the corresponding profile in the database, otherwise, a new location is created for the unknown driver. Providing the right implementation, this system could be used in real time.

C. Driving style identification and evaluation

In this section, we seek to use available data and features in order to characterize an individual driving style on a specified trip (or over all data) in terms of concrete and easily understandable aspects. For this purpose, on the one hand, a dataset comprising one or more real drivers is selected. On the other hand, we used another application being developed by Altran

Technologies to synthetize an approximately 2 hours-long trip comprising several types of road while taking into account specifications to make it meet specific aspects in terms of driving attitude. More accurately, this synthetic profile is considered as a reference in terms of calmness, consistency and ecological driving for this study. This behavior consists in [22]:

- Keeping to the speed limits.
- Drive as steadily as possible.
- Performing smooth accelerations (we chose $1.1\ m.s^{-2}$) and brakings (we chose $-1.1\ m.s^{-2}$).
- Using engine and gears at maximal efficiency.
- Taking advantage of the road gradients.

These behaviors are explained with more details in paragraph III. C. A dataset including real drivers and the synthetic driver is then created. All the following steps are applied to this dataset.

1) Principal Component Analysis

Principal Component Analysis (PCA) [23, 24] can be defined as a transformation which aims to embed data in a new coordinate system whose axes are orthogonal and decorrelated from each others. It can be seen as a form of data compression that can be used to reduce dimensionality and explore data more easily. Mathematically, it is equivalent to performing eigenvalue decomposition on the data correlation matrix. As the first step, the data correlation matrix is computed and eigenvalue decomposition is performed. To perform dimensionality reduction, a specific criterion is applied to the eigenvalue matrix, for example, it can be kept a fixed number of dimensions or to choose the number allowing to keep 95% of data variance. Such variance is represented by the magnitude of the decreasingly sorted eigenvalues.

2) Clustering

The next step is to apply a clustering process on PCA-transformed data. Without knowing who the driver for each instance is, we seek to find a partition such that each point belongs to a specific group called a cluster. The idea consists in interpreting each cluster as a driving style. It is then possible to highlight an individual's sliced trips repartition through different driving styles. Indeed, an individual's driving style is not necessarily consistent in time due to contextual or emotional factors. For this study, after preliminary tests, we chose to use spectral clustering to perform driving style separation. Another reason for choosing spectral clustering over other methods is that this approach does not make any *a priori* assumption about the clusters shape and withstands relatively well overlapping clusters [25]. The cluster number has to be picked manually. In our case, as there is no available ground truth, the cluster number is selected such that the mean silhouette score is maximized.

3) Mapping principal components to driving scores

During the first step, hand-designed driving features were mapped to principal components through a transformation matrix. Considering x as a vector representing the original instance, M the PCA-transformation matrix and x_{PCA} the transformed vector, the transformation is linear and can be written as:

$$x_{PCA} = xM \tag{2}$$

For each principal component, the highest magnitudes of the original features are considered following a criterion while the others are neglected. An example is given below in Table I, the original features are noted $\{C_i\}$ and the principal components $\{PC_i\}$:

	PC4	PC5
C1	-0,07	0,36
C2	-0,03	0,01
…	…	…
C11	-0,08	-0,03
C12	0,72	-0,32
C13	-0,01	-0,04
C14	0,11	-0,04
C15	0,04	0,03
C16	-0,56	-0,53
C17	-0,14	0,06
C18	0,07	-0,28
C19	-0,03	-0,60

Table I: Displaying highest magnitude features choice for two principal components in part of the transformation matrix.

Following Table I, PC4 can be approximated as C12+C16 and PC5 as C16+C19. We create another matrix F which is the filtered version of M, where all the selected coefficient (Table I) are kept while the other terms are set to zero. More formally:

$$F_{i,j} = \begin{cases} M_{i,j} \text{ if } (i,j)_{th} \text{ is selected} \\ 0 \quad \text{otherwise} \end{cases} \tag{3}$$

Furthermore, each original driving feature C_i is associated with a positive or a negative way with a unique adjective among the following list: aggressiveness, consistency, consumption, brakings for acceleration and accelerations for braking). Those adjectives will now be called driving domains. It is therefore possible to find the mapping between each principal component and each domain through the original components. Considering m original features, for a specific column j of F, we have:

$$PC_j = \sum_{i=1}^{m} F_{i,j}\, C_i \tag{4}$$

See Fig. 2 below for an illustration of the process:

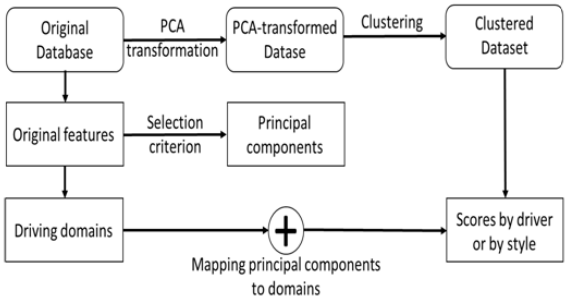


Fig. 2: Displaying a synthetic view of driving style analysis and evaluation system

To summarize, the original features are transformed using PCA in order to compress and clarify data structures. Afterwards, the link between original features and driving domains is used to link principal components to the domains themselves.

4) Evaluating styles and individuals

Considering an instance transformed through PCA, two analyses are possible. First, clustering makes it possible for all instances to be separated into different groups. Therefore, each driver can have one or more driving style that can be identified through the clustering. Afterwards, if an instance v of a particular driving style or driver is considered, it can be written as a linear combination of principal components, considering p principal components and a scalar set of real numbers $\{\lambda_t(v)\}$:

$$v = \sum_{t=1}^{p} \lambda_t(v) PC_t \quad (5)$$

For each driving domain, the score in a specific domain S_D of the instance represented by v is a real number and can be computed using the following equation:

$$S_D(v) = \sum_{t=1}^{m} \sum_{j=1}^{p} \lambda_j(v)\, F_{t,j}\, \Pi_D(C_t) \quad (6)$$

Where $\Pi_D(u)$ is a scalar projection of u vector on domain D:

$$\Pi_D(u) = \begin{cases} 1 \text{ if } u \text{ is associated with } D \\ 0 \text{ otherwise} \end{cases} \quad (7)$$

Therefore, each instance (each slice of trip) of a driver or a style can be rated using the constructed scores. As the synthetic driver is considered as a reference and is rated as well, scores for real drivers can be compared to it.

III. EXPERIMENTS AND RESULTS

A. Driver identification

Let us consider a dataset consisting of drivers 1, 2 and 3 (paragraph II.A.). A SVC is trained on this dataset using 2/3 of data quantity for training and the rest for testing. The classifier's hyper-parameters are fine-tuned using a grid searching with cross-validation process. Using this method yields a radial basis function (RBF) kernel with a gamma (γ) coefficient of 0.25 and a regularization coefficient C of 2. Concerning performance evaluation, we used F1-score and micro-averaged it over the classes to take into account small imbalances in the dataset. The evaluation has been performed 1000 times, regenerating data at each new iteration. The results of these 1000 experiments are displayed in Table II below:

Dataset	Driver 1: 87 instances Driver 2: 57 instances Driver 3: 54 instances
Classifier	RBF-kernel SVC, $\gamma = 0.25$, $C = 2$
Mean F1-score (micro)	0.9501
F1-score (micro) std	0.0274

Table II: Summarizing performances for the three drivers classification over 1000 experiments.

We can observe that mean F1-score is near 95%. The standard deviation of these scores is about 2.75%. To further analyze these results, the averaged confusion matrix in frequencies is given in Table III below:

	Driver 1	**Driver 2**	**Driver 3**
Driver 1	97.1%	0%	2.9%
Driver 2	0%	98.3%	1.7%
Driver 3	10.4%	1.1%	88.5%

Table III: Displaying the mean confusion matrix in terms of percentages for classification of dataset 3 over 1000 experiments.

Observing results from Table III indicates that most of the classification error was made by mistaking Driver 3 for Driver 1. This highlights the fact that almost perfect performances for classifying driver 1 and 2 are possibly due to a bias stemming from heterogeneous data distributions: Driver 1's data were obtained through the smartphone application, whereas Driver's 2 data were extracted from a video stream. Table III of this experiment underlines that performances are lower considering drivers whose data come from the same source (drivers 1 and 3). In this case, the maximal classification error between two drivers is about 10.4% (Table III).

B. Driver recognition system

In this section, we implemented the architecture given in Fig. 1. In order to test the performances, two tasks were selected. The first task is recognition: for an input data batch stemming from a driver known by the database, the system has to assign this batch to the correct driver. The other task is identification: for an input data batch stemming from an unknown driver,

the system has to detect the driver is unknown. To tune the system, we first chose an empirical probability threshold, though other alternatives can be considered, as we shall see in the conclusion section. The experiments we made were conducted (approximately 100 procedure runs while shuffling data each time) using the dataset presented in paragraph III.A. Considering such data, results are displayed in Fig. 3 and 4 below:

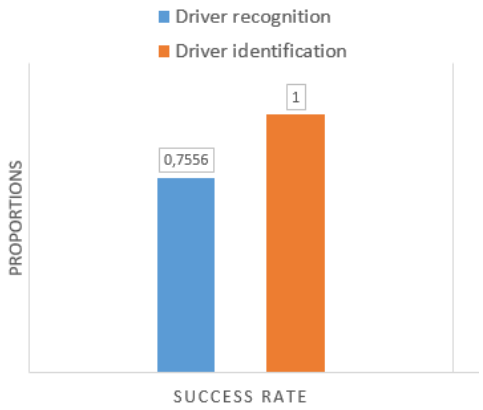


Fig. 3: Displaying the success rate for each task for experiments we led on our dataset.

Fig. 3 shows that the system was able to identify the correct driver from an input data batch stemming from a known individual 75.56% of the time. This may be explained by the fact that the results are depending on a lot of parameters such as: input batch size, database size, probability threshold, classifier's hyper-parameters, etc. Moreover, the classifier used in this system is not perfect, which may result in classification errors. Therefore, these results should be taken as an order of magnitude showing the possibilities of this approach. On the other hand, concerning the driver identification task, the system was always able to detect an unknown driver, which result in a 100.00% success rate during our experiments.

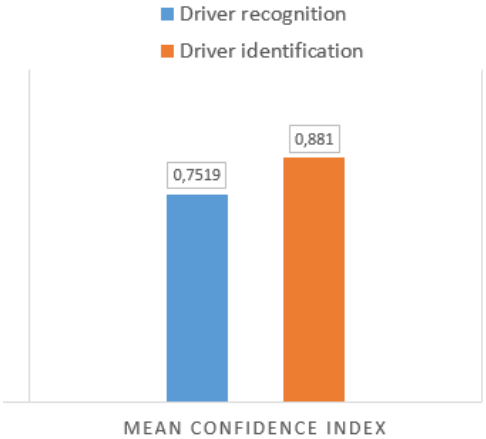


Fig. 4: Displaying the mean confidence index for each task when the system succeeded.

Fig. 4 displays the mean confidence index for each task. The mean confidence index can be defined as the proportion of correct votes during the majority vote for all instances of the input data batch when the system completes its task successfully. For the experiments we conducted, the mean confidence index was 75.19% and 88.10% for driver recognition and driver identification respectively (Fig. 4). This means that for the recognition task, when the system recognized the driver, 75.19% of the votes were in favor of the correct answer. Regarding the identification task, when the classifier was successful, 88.10% of the votes were in a favor of the right answer. Concerning the probabilistic threshold, it is highly dependent on training data and classifier's hyper-parameters, therefore, an automatic and adaptive threshold tuning procedure should be designed in order to ensure consistent performances through dynamic regime.

C. Driving style evaluation

Once a driver has been recognized or identified as new, driving style can be analyzed. The same input batch as before is used as a design matrix consisting of several instances and fed to the driving style analysis module. The current user's data is concatenated with the reference data batch stemming from the synthetic ideal driver's trip using the same route or a representative one. For these experiments, we synthetized an ideal driving profile which is used as a reference for ecological driving, by respecting as much as possible the following aspects:

- Keeping to the limit speeds.
- Accelerating and braking as smoothly and gently as possible (we used a constant acceleration of 1.1 $m.s^{-2}$) and drive as steadily as possible.
- Use optimal engine speed and gear and vehicle's inertia on slopes. Even if this aspect is not directly visible on the speed profiles, the application we use to make the computations is tuned to respect these behaviors as much as possible.

Fig. 5 and 6 shown below display the generated speed profile and its road type repartition respectively.

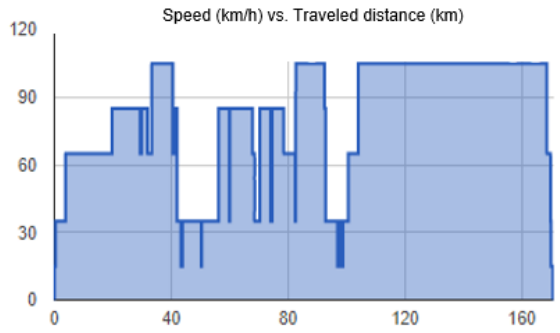


Fig. 5: Displaying the reference speed profile synthesized by HEVES.

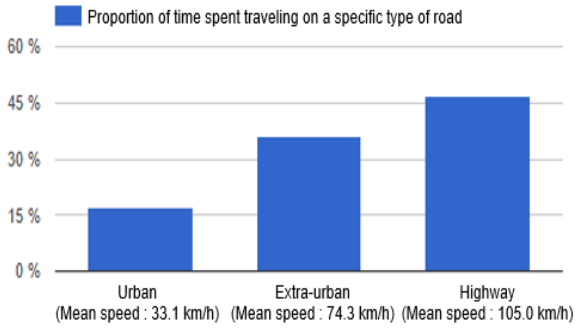


Fig. 6: Displaying the road type proportions for the trip used to synthetize the reference profile.

PCA was then applied to consisting of four drivers (three drivers from the dataset of paragraph III. A. and the reference driver which is indexed 4). Fig. 7 shows the initial data cloud where sliced trips have been projected from a 19-dimensional space (corresponding to the 19 original features) to a 3-dimensional space.

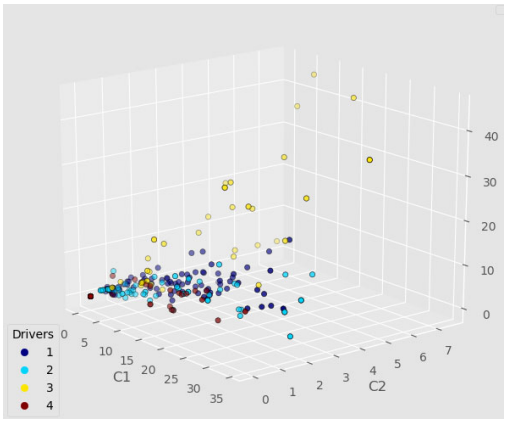


Fig. 7: Original data cloud embedded in a 3-dimensional initial features' space.

After applying PCA to the original data cloud, approximately 90% of the explained variance was kept. Fig. 8 shows the projection of PCA-transformed data into a 3-dimensional space. Fig. 9 illustrates the effect of the clustering procedure. For the dataset we used, we found two clusters that can be interpreted as two driving styles (see Fig. 9).

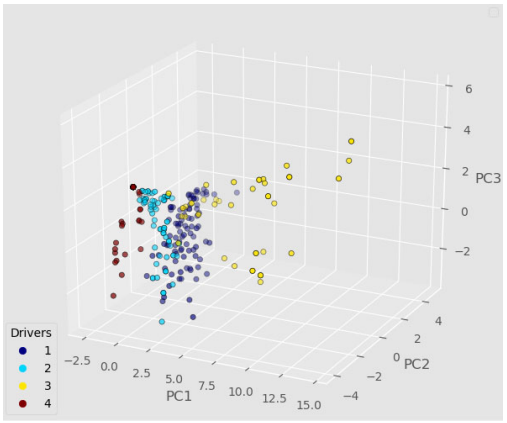


Fig. 8: PCA-transformed data cloud projected in the space of the three first principal components.

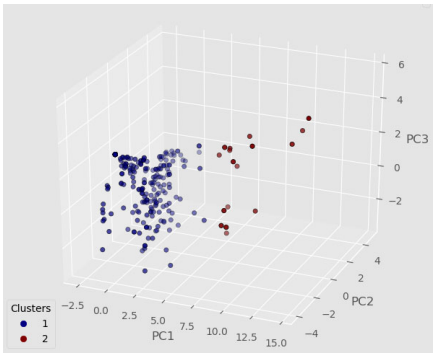


Fig. 9: Clustered data after applying our procedure projected in the three first principal components space.

Using the now clustered data, several kinds of analyses can be carried out. First, observing the created clusters, it can be said that two predominant driving styles can be distinguished in this dataset. The clusters' compositions in terms of drivers are given below in table IV:

	Style 1	Style 2
Driver 1	100%	0%
Driver 2	100%	0%
Driver 3	46.8%	53.2%
Driver ref.	100%	0%

Table IV: Displaying the driving style's compositions in terms of driving instances for each individual.

The results displayed in Table IV show that drivers 1, 2 and reference have a consistent driving style through the studied trips. Driver 3 is mostly responsible for the existence of style 2 as roughly half of his time while driving accounts for 100% of instances in this style. Drivers are analyzed before style in order to understand more individual behaviors, which is one of the key aspects of this study. Through the mappings we created (paragraph II.C.3) the average score for each domain can be computed for each driver by averaging the relevant data instances' features. Scores can be shifted with respect to the reference driver by applying the following transformation:

$$S_D^{driver} = 100 - \mathrm{K}\frac{(S_D^{ref} - S_D^{driver})}{S_D^{ref}} \quad (8)$$

Where S_D^{driver} is the mean shifted score in domain D and S_D^{driver} is the mean score of a driver in a particular domain. K is a visualization parameter that rules the gap between tested drivers. Using this transformation, the objective score to achieve should be 100%, which corresponds to the reference driver's score. Plotting a radar chart using the average scores for each driver leads to Fig. 10 shown below. Table V presented after displays the scores obtained using our method.

	Driver 1	Driver 2	Driver 3
ATB	0.6815	0.8770	0.6805
Aggressivity	0.8223	0.8645	0.4266
BTA	0.6536	0.8982	0.7028
Consistency	0.6172	0.8072	0.9089
Consumption	0.5709	0.9406	0.7999

Table V: Summarizing the scores for all tested drivers with a visualization factor K=20.

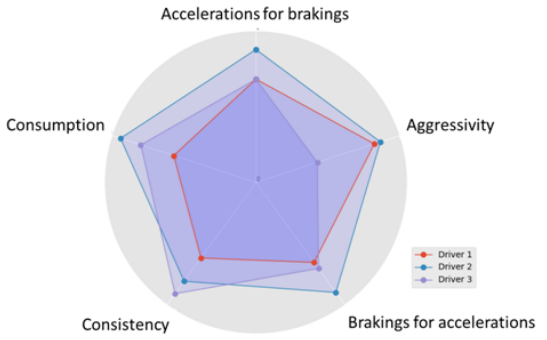


Fig. 10: Radar chart summarizing scores for the three tested drivers.

Fig. 10 displays a representative vision of each drivers' performances between each other and compared to the reference (which is represented by the unit circle). In an ideal perspective, the perfect driver should be able to get all scores to 1 (or 100%). Displaying such a plot to users can help to quickly visualize which aspects are prominent and which ones could be enhanced. For example, Driver 3 seems to have the lowest score in "Aggressivity" which means that his driving is the most aggressive among all the tested drivers. On the other hand, driver 1 seems to have the lowest consistency in his driving and to be the most consuming one (scores of 100% have to be seen as an ideal to achieve, so a low score indicates a low performance). These results can be extensively analyzed using Table V. Driver 1's weaknesses seem to reside in consumption, consistency (which are linked) and Braking For Accelerations. Driver's 3's lowest score is due to aggressivity. Driver 2 seems to have relatively good scores in each domain. In order to communicate the results to the users, a Global Ecodriving Score is computed by averaging all the scores of Table V. Afterwards, the lowest score of each driver is selected and key advices can be automatically generated (using the mappings between domains and features) in order to improve one's driving behavior on the road. To improve, driver 1 should try to anticipate more to avoid unnecessary brakings and acceleration. Driver 2 should try to be more consistent when possible. In order to reduce aggressivity, Driver 3 should lower acceleration and braking magnitudes. These results can be found in Table VI displayed right after:

Driver	Global Ecodriving Score	Lowest score	Key advices
1	66.92%	Consumption	Try to anticipate traffic and stops. Try avoiding unnecessary accelerations and brakings.
2	87.75%	Consistency	Try to maintain your speed when possible. Try using a speed regulator.
3	70.38%	Aggressivity	Try to accelerate and brake as smoothly as possible.

Table VI: Displaying Global Ecodriving Score and key advices to each driver.

An analogous analysis may be carried out by examining and averaging data of each driving style, however, considering that this work emphasizes with individual driving behavior, this will not be covered in this paper. The most significant results are illustrated through the drivers analysis.

IV. CONCLUSION

Through this work, we presented several aspects of an artificial intelligence based platform for driver behavior analysis, assessment and prediction. Three different drivers were identified and analyzed. Despite the little data used for this study (2 hours of driving per driver), we were able to identify each driver with a high percentage of accuracy with a mean F1-score of 0.95 (97% for driver 1, 98% for driver 2 and 88% for driver 3). It can be noticed that the data acquisition technology has a slight influence on the identification of the driver and this amounts to the error induced during the acquisition linked for example to the GPS position or the acceleration. Though this platform is still under development, this paper highlighted some key concepts. Driver identification using driving related features and a supervised learning procedure has been performed. Such information is used to create an individual profile in the database that can be later used to recognize an existing driver.

Using database's information for one or several drivers, individual performances regarding Ecodriving can be assessed and personalized advices displayed. The evaluation of the driving performances is carried out by comparing the characteristics of an optimal economic driving with the criteria of the different drivers grouped in 5 performances (ATB, aggressiveness, BTA, consistency and consumption) with the PCA method. These characteristics are represented in a radar chart making it possible to visualize the behavior of drivers regarding economic driving. Driver 2 is the driver closest to economical driving except that he has to improve his speed maintenance to come closer to economical driving.
Finally, personalized advice is given to each driver by analyzing the difference between his driving and economical driving. Indeed, the notion of Global Ecodriving Score may be perceived as a sort challenge

of self-improvement for the user, which may encourage positive behaviors. Overall, we have introduced and proposed a collection of analytic and predictive tools based on statistics and artificial intelligence which aim to improve ecological behaviors while providing a customized experience tailored on the individual's attitude.

PERSPECTIVES

Considering the concepts we proposed, in the future, we intend to explore two directions. First, a rigorous study should be led in order to observe the influence of our driving evaluation system on drivers' behaviour, particularly regarding ecology. Second, we carried out our experiments and applied the presented methodology on a small pool of drivers and a relatively small data quantity. The concepts covered in this study should be tested and adapted (if necessary) to a greater scale in terms of individuals and collected data. Moreover, we intend to deploy this platform in a vehicle-embedded system in real time in order to deliver the solutions we proposed to the drivers.

REFERENCES

[1] Johan Wideberg, Pablo Luque, Daniel Mantaras. "A smartphone application to extract safety and environmental related information from the OBD-II interface of a car". International Journal of Vehicle Systems Modelling and Testing 7(1):1 - 11 · February 2012.

[2] Michael Reininger and al., "A First Look at Vehicle Data Collection via Smartphone Sensors" 2015 IEEE Sensors Applications Symposium (SAS).

[3] Bartosz Jachimczyk and al., "IoT On-Board System for Driving Style Assessment", Sensors 2018, 18, 1233; doi:10.3390/s18041233.

[4] Minh Van Ly et al., "Driver Classification and Driving Style Recognition using Inertial Sensors". 2013 IEEE Intelligent Vehicles Symposium (IV).

[5] B. Gahr et al., "Driver Identification via the Steering Wheel" , arXiv:1909.03953 [cs, eess, stat], September 2019.

[6] Hallac et al., "Driver Identification Using Automobile Sensor Data from a Single Turn", arXiv:1708.04636 [cs.HC].

[7] Thitaree Tanprasert et al., "Combining Unsupervised Anomaly Detection and Neural Networks for Driver Identification", Hindawi Journal of Advanced Transportation Volume 2017, Article ID 6057830, 13 pages https://doi.org/10.1155/2017/6057830.

[8] Ruhallah Ahmadian, Mehdi Ghatee, "Driver Identification by Neural Network on Extracted Statistical Features from Smartphone Data", arXiv:2002.00764 [cs.HC].

[9] Dominik Dörr et al., "Online Driving Style Recognition using Fuzzy Logic", 17th International IEEE Conference on Intelligent Transportation Systems (ITSC).

[10] Zoran Constantinescu et al., "Driving Style Analysis Using Data Mining Techniques", International Journal of Computers, Communications & Control (IJCCC) V:654-663 · December 2010.

[11] Weishan Dong et al., "Characterizing Driving Styles with Deep Learning", arXiv:1607.03611v2 [cs.AI] 8 Oct 2016.

[12] Yuxiang Feng et al., "A Support Vector Clustering Based Approach for Driving Style Classification", International Journal of Machine Learning and Computing, Vol. 9, No. 3, June 2019.

[13] Derick A. Johnson, Mohan M. Trivedi, "Driving Style Recognition Using a Smartphone as a Sensor Platform", 2011 14th International IEEE Conference on Intelligent Transportation Systems.

[14] Rui Araújo et al., "Driving Coach: a Smartphone Application to Evaluate Driving Efficient Patterns", 2012 Intelligent Vehicles Symposium.

[15] Johannes Tulusan et al., "Providing eco-driving feedback to corporate car drivers: what impact does a smartphone application have on their fuel efficiency?", UbiComp '12: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, September 2012, Pages 212–215 https://doi.org/10.1145/2370216.2370250.

[16] Qin Shi et al., "The study of a new method of driving cycles construction", International Workshop on Automobile, Power and Energy Engineering, Procedia Engineering 16 (2011) 79 – 87.

[17] Ying Yang et al., "Markov chain-based approach of the driving cycle development for electric vehicle application", CUE2018-Applied Energy Symposium and Forum 2018: Low carbon cities and urban energy systems, 5–7 June 2018, Shanghai, China.

[18] A. Fotouhi, M. Montazeri-Gh, "Tehran driving cycle development using the k-means clustering method", Scientia Iranica A (2013) 20 (2), 286–293

[19] I.N. Anida et al., "Characterisation and development of driving cycle for work route in Kuala Terengganu", International Journal of Automotive and Mechanical Engineering Volume 14, Issue 3 pp. 4508-4517 September 2017.

[20] Bosetti P, Da Lio M, Saroldi A. On the human control of vehicles: an experimental study of acceleration. European Transport Research Review. 2014 Jun 1;6(2):157-70.

[21] Karjanto, J., Yusof, N.M., Terken, J., Delbressine, F., Hassan, M.Z. and Rauterberg, M., 2017. Simulating autonomous driving styles: Accelerations for three road profiles. In MATEC web of conferences (Vol. 90, p. 01005). EDP Sciences.

[22] SenterNovem, Utrecht for the EC TREATISE project, "ECODRIVING The smart driving style", September 2005.

[23] J. Shlens, "A Tutorial on Principal Component Analysis", arXiv:1404.1100 [cs.LG], April 2014.

[24] Jolliffe IT, Cadima J., "Principal component analysis: a review and recent developments", Phil. Trans. R. Soc. A 374: 20150202. http://dx.doi.org/10.1098/rsta.2015.0202, January 2016.

[25] M. Meila, "Spectral Clustering: a Tutorial for the 2010's", September 2016.

★★★