Clustering Of Drivers' State Before Takeover Situations Based On Physiological Features Using Unsupervised Machine Learning

Emmanuel de Salis[1], Quentin Meteier[2], Colin Pelletier[1], Marine Capallera[2], Leonardo Angelini[2], Andreas Sonderegger[3], Omar Abou Khaled[2], Elena Mugellini[2], Marino Widmer[4] and Stefano Carrino[1]

[1] Haute Ecole Arc Ingénierie, HES-SO, Saint-Imier, Switzerland
{emmanuel.desalis, colin.pelletier, stefano.carrino}@he-arc.ch
[2] HumanTech Institute, HES-SO//University of Applied Sciences and Arts Western Switzerland, Fribourg, Switzerland
{quentin.meteier, marine.capallera, leonardo.angelini, omar.aboukhaled, elena.mugellini}@hes-so.ch
[3] Bern University of Applied Sciences, Business School, Institute for New Work, Bern, Switzerland
andreas.sonderegger@bfh.ch
[4] DIUF, University of Fribourg, Fribourg, Switzerland
marino.widmer@unifr.ch

Abstract. Conditionally automated cars share the driving task with the driver. When the control switches from one to another, accidents can occur, especially when the car emits a takeover request (TOR) to warn the driver that they must take the control back immediately. The driver's physiological state prior to the TOR may impact takeover performance and as such was extensively studied experimentally. However, little was done about using Machine Learning (ML) to cluster natural states of the driver. In this study, four unsupervised ML algorithms were trained and optimized using a dataset collected in a driving simulator. Their performances for generating clusters of physiological states prior to takeover were compared. Some algorithms provide interesting insights regarding the number of clusters, but most of the results were not statistically significant. As such, we advise researchers to focus on supervised ML using ground truth labels after experimental manipulation of drivers' states.

Keywords: Automated Vehicles · Clustering · Machine Learning · Physiological state · Takeover · TOR

1 Introduction

Research on conditionally automated driving has been extensively conducted in the last few years. One of the main challenges is to optimize the interaction between the car and the driver, especially in situations of handover of control. When a takeover request (TOR) is issued, drivers are expected to take over the control of the car in a short period of time and with the adequate behavior (e.g. braking and/or turning the wheel). The physiological state of the driver prior the TOR is critical and was shown to impact the takeover performance [1], [2], motivating the need to know the physiological state of the driver before triggering a TOR. This information could then be used to improve takeover performance. In previous studies, this has been addressed by experimentally influencing drivers' physiological states with the use of secondary tasks which was then used as ground truth information. To the extent of our knowledge, little was done on studying natural clusters of driver physiological states. Knowing the general state of the driver, meaning in which cluster they are, could be used to improve the takeover quality by adjusting the TOR to their state. As such the goal of this study is to find out if Machine Learning (ML) can be used as a tool to distinguish relevant clusters of drivers based on their physiological state before a TOR.

2 Related Work

2.1 Psychophysiological Data Clustering

The ML algorithms need to focus on the most relevant features and on coherent data only [3]. Furthermore, the Bayesian Information Criterion (BIC) score cannot be used with a dataset in which the number of samples are not considerably higher than the number of features extracted [4], which is the case in this study (108 feature for 304 samples).

As such, the mean and the standard deviation provide information about data distribution into a feature [3]. This can be exploited in the feature selection with a variance threshold.

Models of the driver's state are being studied, but they usually aim at classifying driver's condition using states that are manipulated experimentally [5], [6]. Other sources of data such as driving behavior [7] were shown to be valuable to classify the driver's state. Yet, this source of data cannot be used anymore in high levels of automation. Previous studies classified various drivers' states using physiological data, such as stress [8], [9], fatigue [10], or cognitive workload [6]. However, only few studies tried to perform unsupervised ML tasks using physiological indicators in such context [11], [12].

3 Methodology

3.1 Data collection

Electrodermal activity (EDA), Electrocardiogram (ECG) and respiration (RESP) of 80 subjects were collected during an hour of conditionally automated driving in a driving simulator. Participants spent the first 5 minutes in autonomous mode monitoring the environment only. This phase was used as a baseline for the physiological state of each driver. Then, participants had to drive manually for 5 minutes, while the takeover process was explained and showed to them (three takeovers). The TOR was indicating by displaying a red icon on the dashboard and an audio chime. In the main driving session, participants had to perform a succession of non-driving related, cognitive demanding tasks (N-back task) on a tablet held in hands while the car was driving. Participants were instructed to take over control of the car when requested. All participants had to react to 5 takeover situations due to an automation limitation. The cause of each takeover request was different but in the same order for all participants: steep road ahead and no visibility behind, vanishing and then fully erased lane markings, a rock on the driving lane, heavy rain, and a deer standing on the right side of the road and then crossing. There were 5 different tasks: visual low cognitive, visual high cognitive, auditory low cognitive, auditory high cognitive or no task (e.g. monitoring the environment). Each participant was completing all the five non-driving task in one driving session. The order of non-driving tasks presented to each participant was controlled using a Latin Square Design.

3.2 Physiological features

The Neurokit [13] library in Python was used to process the 3 signals and compute physiological indicators. Different time windows before takeovers were used to calculate them: 30, 60, 90, 120 and 150 seconds. Each indicator calculated with Neurokit served as a feature for the clustering task. Also, an additional feature was created from each indicator: a baseline correction was applied to each indicator (e.g. subtracting the value of that indicator during baseline) in order to remove the individual differences of drivers at rest. Overall, a total of 120 features (10 for EDA, 27 for ECG, 19 for RESP, 4 for Respiratory Sinus Arrhythmia) were calculated using each time window before each takeover for the 80 participants.

3.3 Machine Learning

Outliers detection The Z-Score is used to detect outliers based on a Gaussian probabilistic distribution. For each feature, every samples with an absolute Z-Score strictly greater than 3 (|Z-Score| > 3), which corresponds to 3 standard deviations from the mean, are considered as outliers and are ignored for the following processes.

Features selection Since the dataset contains a large amount of features (108), it is necessary to select the most important ones. The Pearson correlation coefficient gives information about the linear relation between two features. Firstly, relations with a coefficient of correlation greater than 0.8 are removed. Removing a relation corresponds to remove one of the two features which are part of the relation. Finally, the Pearson correlation coefficient is used again to remove relations with a correlation coefficient greater than 0.8. The feature selection is made in two steps to keep as many features as possible.

Another avenue of research is using a variance threshold to select the features. However, this method has not provided better results than the one with Pearson correlation coefficient and was thus abandoned.

The total number of features kept for each time window is presented in Table 1.

Table 1. Total number of features kept in each time window after the feature selection process

Time window	30s	60s	90s	120s	150s
Final number of features	13	13	14	15	13

Clustering algorithms The selected machine learning algorithms K-Means, EM-GMM, DBSCAN and Mean-Shift have been tested on this dataset, once outliers detection and features selection have been processed. All those algorithms hyperparameters were optimized using the methodology presented in the following section. The resulting clusters from these algorithms are then evaluated and compared in order to choose final hyperparameters and also the most efficient algorithm for this specific task.

Parameters definition With K-Means, the elbow method was used to estimate the optimal number of clusters. Then, to further refine the exact number of clusters, it was combined with the Silhouette score which informs about clustering quality. It was also further confirmed with the Mutual Information (MI) score, which informs about clusters' stability. Silhouette score and MI score range typically from 0 to 1, with 1 being the best score possible.

With EM-GMM, the BIC score is usually a good indicator to select the optimal number of clusters and the optimal covariance type. Since the number of features has been sufficiently reduced, it can be used on this dataset. Nevertheless, these supposedly optimal parameters do not provide the best results for the MI score or the Silhouette score. As every indicator provided different optimal parameters, multiple combinations have been compared with K-Means clustering.

With Mean-Shift, the parameter to define is the bandwidth. A finite field of bandwidths values (from 2 to 20) were tested until the algorithm clustered every sample into a cluster. For each bandwidth of the finite field, the population of every cluster as well as the number of clusters that it involves were checked.

With DBSCAN, the parameters to fine-tune are the maximum distance between two samples for one to be considered as in the neighborhood the other (eps) and the minimum of samples in a neighborhood to consider the neighborhood as a cluster (minPts). Two finite fields of values were created for both parameters. All the combinations between the values of these two fields have been tested until every sample was contained into a cluster or until most of the samples were considered as noise. For each combination, the population of every cluster as well as the number of clusters created were analyzed.

4 Results

Overall, Mean-Shift and DBSCAN were not able to achieve significant results, despite trying several configurations of parameters. For K-Means, a Silhouette score of 0.10 was attained (std: 0.00) and a MI of 0.8 (std: 0.1) using the 30s time window and 3 clusters. EM-GMM (spherical) presented two interesting clustering results, with a Silhouette score and MI of respectively 0.25 (std: 0.01) and 0.87 (std: 0.27) for the 60s time window, and 0.38 (std: 0.08) and 0.85 (std: 0.33) for the 120s time window, with

2 clusters each time. Both situations presented one cluster significantly smaller than the other one 70 vs 255 (for the 60s) and 41 vs 291 (for the 120s). Also, we analysed if those two clusters match the 2 physiological states that were manipulated experimentally: no task (20% of the dataset) vs. task (80% of the dataset). After a comparison of the samples in each cluster, it appears that this is not the case.

In both situations, the models seem to gather more extreme values in a cluster (the small one), and keep the other ones in another cluster (the large one). As such, analysis of the features revealed extreme standard deviations in the smaller cluster, rendering further analysis unnecessary.

5 Conclusion

The current results show that there are no clear clusters of physiological data prior a TOR. Unsupervised ML algorithms were not successful in creating relevant clusters, despite testing several time windows. Bigger time windows and more refined features can still be tested, but our first results seem to hint that this is a difficult task.

As such, we advise researchers to focus on supervised ML using ground truth labels after experimental manipulation of drivers' states.

Acknowledgments. This work is part of the AdVitam project funded by the Hasler Foundation. We would also like to thank our colleagues who helped us during this project.

References

1. Merat, N., Jamson, A.H., Lai, F., Carsten, O.: Highly Automated Driving, Secondary Task Performance, and Driver State. Human Factors: The Journal of Human Factors and Ergonomics Society, 54, 762 - 771 (2012)
2. Zeeb, K., Buchner, A., Schrauf, M.: What determines the take-over time? An integrated model approach of driver take-over after automated driving. Accident; analysis and prevention, 78, 212-221 (2015)
3. Johannes, B., Gaillard, A. W.: A methodology to compensate for individual differences in psychophysiological assessment. Biological Psychology, 96, 77-85. doi:10.1016/j.biopsycho.2013.11.004 (2014)
4. Giraud, C.: Introduction to High-Dimensional Statistics. doi:10.1201/b17895 (2014)
5. Darzi, A., Gaweesh, S. M., Ahmed, M. M., Novak, D.: Identifying the Causes of Drivers' Hazardous States Using Driver Characteristics, Vehicle Kinematics, and Physiological Measurements. Frontiers in Neuroscience, 12. doi:10.3389/fnins.2018.00568 (2018)
6. Meteier Q., Capallera M., Ruffieux S., Angelini L., Abou Khaled O., Mugellini E., Widmer M. and Sonderegger A.: Classification of Drivers' Workload Using Physiological Signals in Conditional Automation. Front. Psychol. 12:596038. doi: 10.3389/fpsyg.2021.596038 (2021)
7. Darzi, A., Gaweesh, S. M., Ahmed, M. M., Novak, D.: Identifying the Causes of Drivers' Hazardous States Using Driver Characteristics, Vehicle Kinematics, and Physiological Measurements. Frontiers in Neuroscience, 12. doi:10.3389/fnins.2018.00568 (2018)
8. Healey, J., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems, 6, 156-166 (2005)

9. Chen, L., Zhao, Y., Ye, P., Zhang, J., Zou, J.: Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. Expert Syst. Appl., 85, 279-291. (2017)
10. Awais, M., Badruddin, N., Drieberg, M.: A Hybrid Approach to Detect Driver Drowsiness Utilizing Physiological Signals to Improve System Performance and Wearability. Sensors (Basel, Switzerland), 17. (2017)
11. Li, N., Misu, T., Miranda, A.: Driver behavior event detection for manual annotation by clustering of the driver physiological signals. 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), 2583-2588. (2016)
12. Noh, Y., Kim, S., Jang, Y.J., Yoon, Y.: Modeling Individual Differences in Driver Workload Inference Using Physiological Data. International Journal of Automotive Technology, 22, 201-212. (2021)
13. Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & S H Chen, A.: NeuroKit2: A Python Toolbox for Neurophysiological Signal Processing. Retrieved March 28, 2020, from https://github.com/neuropsychology/NeuroKit (2020)