Accident Analysis and Prevention 121 (2018) 118–128

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Adapting artificial neural networks to a specific driver enhances detection and prediction of drowsiness

Charlotte Jacobé de Naurois[a,b,*], Christophe Bourdin[a], Clément Bougard[b], Jean-Louis Vercher[a]

[a] Aix Marseille Univ, CNRS, ISM, Marseille, France
[b] Groupe PSA, Centre Technique de Vélizy, Vélizy-Villacoublay, Cedex, France

ARTICLE INFO

Keywords:
Monitoring
ANN
Adaptive learning
Inter-individual variability
Drowsiness

ABSTRACT

Monitoring car drivers for drowsiness is crucial but challenging. The high inter-individual variability observed in measurements raises questions about the accuracy of the drowsiness detection process. In this study, we sought to enhance the performance of machine learning models (Artificial Neural Networks: ANNs) by training a model with a group of drivers and then adapting it to a new individual. Twenty-one participants drove a car simulator for 110 min in a monotonous environment. We measured physiological and behavioral indicators and recorded driving behavior. These measurements, in addition to driving time and personal information, served as the ANN inputs. Two ANN-based models were used, one to detect the level of drowsiness every minute, and the other to predict, every minute, how long it would take the driver to reach a specific drowsiness level (moderately drowsy). The ANNs were trained with 20 participants and subsequently adapted using the earliest part of the data recorded from a 21st participant. Then the adapted ANNs were tested with the remaining data from this 21st participant. The same procedure was run for all 21 participants. Varying amounts of data were used to adapt the ANNs, from 1 to 30 min, Model performance was enhanced for each participant. The overall drowsiness monitoring performance of the models was enhanced by roughly 40% for prediction and 80% for detection.

1. Introduction

Driving while drowsy is a safety issue, and a major cause of accidents. Numerous fundamental and applicative studies focus on detection of drowsiness as a way to improve accident prevention. However, simply detecting drowsiness is not enough: once the driver is drowsy, it is probably already too late to prevent the accident. The key challenge is to predict how and when drowsiness will occur, how often it will occur and who might become drowsy under which conditions. Prediction refers here to the timely identification of when a given event will occur within a given range of future states, in our case a given level of drowsiness. Watson and Zhou (2016) detected the occurrence of micro-sleep episodes with 96% accuracy and were able to predict the next micro-sleep between 15 s and 5 min in advance, although obviously not the time of occurrence of the first micro-sleep. A recent study (Jacobé de Naurois et al., 2017) showed that an Artificial Neural Network (ANN) can not only detect the level of drowsiness but can also predict, in advance, the time at which this impaired driver's state will occur.

Various sources and types of information can be used to estimate the operator's functional state. For car driving, measurements must be easily recordable, not invasive, and reliable. The literature contains a variety of sources of information (Dong et al., 2011), mainly based on ocular and eyelid movements (Chen and Ji, 2012; Liu et al., 2009). For instance, PERCLOS (PERcentage of eye CLOSure, the percentage of time, generally during one minute, when eyes are closed more than 80%) indicates how long on average the eyes are closed. Physiological measurements are also often used to assess the driver's state through the central and the neuro-vegetative systems, offering the advantage of being continuously available, objective and fairly direct indicators of the functional state. The most commonly used physiological signal is the electroencephalogram (EEG). However, EEG recording during driving is rather intrusive and constraining (despite continuous technological advances), which can be a real disadvantage. Electrocardiogram (EKG) and respiration measurements are also often used. Yet it remains difficult to define a direct relationship between physiological features and a given cognitive state, since these physiological features vary with other states like stress, emotions, workload, physical effort and fatigue, or with the context.

Finally, driving behavior and performance, such as the standard deviation of car position relative to lane midline (also termed standard deviation of lane position (SDLP)) or steering wheel movements,

* Corresponding author at: Aix Marseille Univ, CNRS, ISM, Marseille, France.
E-mail addresses: charlotte.jacobe-de-naurois@etu.univ-amu.fr (C. Jacobé de Naurois), jean-louis.vercher@univ-amu.fr (J.-L. Vercher).

https://doi.org/10.1016/j.aap.2018.08.017
Received 5 April 2018; Received in revised form 18 July 2018; Accepted 14 August 2018

(Arnedt et al., 2001; De Valck et al., 2003; Liu et al., 2009; Philip et al., 2004) are also common measures used to detect the driver's state. However, here again, driving performance and activity are not specific indicators of drowsiness.

To deal with the above limitations, recent research has sought to improve prediction through complex approaches combining multivariate, heterogeneous information via data fusion (Dong et al., 2011; Samiee et al., 2014). Findings from these studies show that this hybrid approach can provide better accuracy (Awais et al., 2014).

However, current models need to deal with yet another challenge to their prediction power. It is now widely recognized that neurobehavioral and cognitive performance vary considerably from one individual to another (Van Dongen et al., 2004aa; b). In car driving tasks, according to (Ingre et al., 2006), there is extensive inter-individual variability in driving behavior and eye behavior. Under similar conditions, individuals' patterns of drowsiness evolution over time can differ, and for a given self-declared drowsiness level, markers such as eye blink duration also vary considerably. Van Dongen et al. (2003) showed that individuals probably also differ in their vulnerability to sleep deprivation, and that this is partially predictable from individual cognitive performance without deprivation, i.e. from the individual cognitive profile. In driving simulator studies, drowsiness is often observed to develop in differing ways (Thiffault and Bergeron, 2003). Situational and personality factors, sleeping habits and driving history help explain why some people fall asleep at the wheel while others do not. This confirms the need to consider drivers' traits or profiles to calibrate systems for the detection and prediction of drowsiness (Jacobé de Naurois et al., 2017).

Such large inter-individual variability makes creating algorithms that will perform well for all individuals a challenge. As most studies use machine-learning algorithms, the difficulty is finding a general model trained with a limited number of drivers which can then be applied to the majority of individual drivers (Karrer et al., 2004). One of the main issues with machine learning is uncertainty about the generalization of a given model to a new participant. To ascertain whether an algorithm generalizes well, the dataset is segregated into either two (training and testing) or three (training, validating and testing) datasets (in most cases, the segregation is randomly performed on the full set of recorded data). Thus, it is impossible to be sure that the algorithm will perform well for another participant whose data is unknown to the model.

This problem can be approached in different ways. One is to train the model with as large a population as possible: the more data, the better the model. However, this method is based on the assumption that for each new individual, the model has previously encountered a similar individual. This makes it difficult to determine the number of participants required to deal with the large inter-individual variability. Furthermore, the level of similarity between two individuals is hard to quantify. This method would thus be extremely time-consuming, not only in terms of training the model but also in terms of data collection. A second solution is to have a specific model for each driver, but this obviously involves collecting and labeling sufficient data from each driver as well as training the specific model with these data, another time-consuming option. A third way is to use methods such as transfer learning or adaptive learning, which combine the advantages of the two preceding methods by permitting capitalization on a group of individuals and personalization for each new individual. In particular, these methods are applied on Brain Computer Interface systems (Wang et al., 2015). To detect driver drowsiness, studies applied such techniques on EEG signals (Wei et al., 2015; Wu et al., 2015, 2016) and found that transfer learning applied to EEG significantly enhances model performance. Our aim here was to test a similar method based on adaptive learning but using non-intrusive measurements including eyelid movements, head movements, EKG, respiration rate, driving activity and performance, as in our previous study (Jacobé de Naurois et al., 2017).

The goal of the present study is to enhance the performance of machine-learning models both in detecting the level of driver drowsiness and in predicting when a given impaired state will be reached, by first training a model and then adapting it to each new individual. The model uses Artificial Neural Networks. We hypothesize that training an ANN with a group of individuals and then personalizing the ANN for a new individual (whose data were not encountered by the model during training) will improve the performance of the model for this specific individual. We also assess the amount of data required to enhance the generalization performance of the model.

2. Materials and methods

The participants and the protocol, including data collection and preprocessing, were the same as used for our previous study (Jacobé de Naurois et al., 2017). Data modeling methods were specifically developed for the present study.

2.1. Participants

Twenty-one participants were included in the study (mean age 24.09 ± 3.41 years; 11 men, 10 women). Inclusion criteria were: valid driver's license for at least 6 months, no visual correction needed to drive, not susceptible to simulator sickness, as assessed by the Motion Sickness Susceptibility Questionnaire, Short-form (MSSQ-Short, Golding, 1998), and an Epworth scale score (assessing susceptibility to drowsiness) below 14 (Johns, 1991) (for more detail, see (Jacobé de Naurois et al., 2017)). The following participant information was collected: Epworth scale score (assessing susceptibility to drowsiness (Johns, 1991)), quality of the previous night's sleep (on a scale from 1 to 10), caffeine consumption (never, rarely, one or two cups per day, more than two cups per day), driving frequency (occasionally, several times a month/a week/a day), distance (kilometers) driven per year and score on the Horne and Östberg morning/evening questionnaire (Horne and Ostberg, 1975).

2.2. Protocol

The participants drove for 100 to 110 min in a static driving simulator in an air-conditioned room with temperature control set at 24 °Celsius. They drove just after lunchtime, a time considered as risky in terms of drowsiness (Horne and Reyner, 1999). The road and traffic were generated with SCANeR Studio®. A webcam located on top of the central screen of the simulator video-recorded the participants during the session to establish the ground truth (see below). The (static) simulator, provided by Oktal® and powered with SCANeR Studio® software, is made of a real car seat, 3 video screens (24″ in format 16/9 each, forming a tryptic), a steering wheel, pedals and a small screen (10″) for the dashboard, located just behind the wheel. The driving environment was displayed at a resolution of 1280 × 1024 pixels onto the three forward screens providing a 210° horizontal forward field of view. A rear screen provided a 60° rear field of view, corresponding to the normal use of the central rearview and two side mirrors. A stereo sound system provided simulated engine, road, and traffic sounds. An example of the field of view is presented on the figure below, which has been added to Fig. 1. The simulated car had an automatic gearbox, so the driver had only access to the steering wheel, gas and brake pedals. At the beginning of the session, the participants drove along a highway for roughly 90 min, then turned off the highway and drove for around 5 min to reach a city. Finally, they drove in an urban environment for roughly 5 min. There was no traffic during most of the highway stretch. The very monotonous environment (without event or traffic) was selected in order to induce drowsiness. Somewhere 2/3 of the way along, 22 cars appeared from the right of the highway, disappearing a few kilometers later (Fig. 2). This sudden addition of traffic was intended to change the driver's level of drowsiness. Rossi et al. (2011)