Expert Systems With Applications 52 (2016) 99–107

Contents lists available at ScienceDirect

Expert Systems With Applications

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

The application of data mining techniques to model visual distraction of bicyclists

Cristofer Englund*, Maria Nilsson, Alexey Voronov

Viktoria Swedish ICT, Lindholmspiren 3A, SE 417 56 Göteborg, Sweden

ARTICLE INFO

Keywords:
Bicycle simulator
Data mining
Bicyclist distraction
Random forest
Support vector machine
Automated driving systems

ABSTRACT

This paper presents a novel approach to modelling visual distraction of bicyclists. A unique bicycle simulator equipped with sensors capable of capturing the behaviour of the bicyclist is presented. While cycling two similar scenario routes, once while simultaneously interacting with an electronic device and once without any electronic device, statistics of the measured speed, head movements, steering angle and bicycle road position along with questionnaire data are captured. These variables are used to model the self-assessed distraction level of the bicyclist. Data mining techniques based on random forests, support vector machines and neural networks are evaluated for the modelling task. Out of the total 71 measured variables a variable selection procedure based on random forests is able to select a fraction of those and consequently improving the modelling performance. By combining the random forest-based variable selection and support vector machine-based modelling technique the best overall performance is achieved. The method shows that with a few observable variables it is possible to use machine learning to model, and thus predict, the distraction level of a bicyclist.

1. Introduction

Electronic devices become bigger and bigger part of our life. It is important to understand the effect of their usage, especially in a high-risk environment like traffic, since the distraction these devices can cause may lead to severe consequences. Significant research has been done on the usage of electronic devices such as in-vehicle systems and mobile phones (e.g., Bach, Jæger, Skov, & Thomassen, 2009; Fitch et al., 2013; Gelau et al., 2003; Horberry, Anderson, Regan, Triggs, & Brown, 2006; Strayer, Watson, & Drews, 2011). For example, Horberry et al. (2006) highlight the fact that the usage of electronic devices while driving increases the perceptual and cognitive demands. The results of Hurwitz and Wheatley (2002) showed that driving performance was impaired when performing a visual task. Engström, Johansson, and Östlund (2005) showed that visual demand increases variation in lane-keeping. Visual distraction has been defined and examined for vehicle drivers (e.g., Engström, 2010; Engström et al., 2005; Liang & Lee, 2010; Young & Regan, 2007) and methods for how to measure and analyse distraction have been identified by Kircher (2007). However, as the vehicle safety increases, the forefront of safety research has to move towards less investigated and less protected Vulnerable Road Users (VRU) like bicyclists.

Previous studies on distraction of bicyclists indicate, on the one hand, that only 0.5% of the respondents were using a phone at the time of the accident (de Waard, Schepers, Ormel, & Brookhuis, 2010), and that electronic devices do not contribute to a large extent to severe bicyclist accidents (Niska & Eriksson, 2013). However, Goldenbeld, Houtenbos, Ehlers, and de Waard (2012) concluded that "the odds of being involved in a bicycle crash were estimated to be higher for teen bicyclists and young adult bicyclists who used electronic devices on every trip compared to the same age groups bicyclists who never used these devices". In de Waard et al. (2010) it was shown that using mobile phone "coincided with reduced speed, reduced peripheral vision performance and increased risk and mental effort ratings. Text messaging had the largest negative impact on cycling performance. Higher mental workload and lower speed may account for the relatively low number of people calling involved in accidents". In de Waard, Lewis-Evans, and Jelijs (2014) it was found that sending text messages from mobile phones reduces bicycle speed and increases variation in lateral position. Moreover, the authors found that the use of a touch screen has a larger negative effect on bicycling performance compared to a keypad-based phone.

* Corresponding author. Tel.: +46 708560227.
E-mail addresses: Cristofer.Englund@viktoria.se (C. Englund), Maria.Nilsson@viktoria.se (M. Nilsson), Alexey.Voronov@viktoria.se (A. Voronov).

http://dx.doi.org/10.1016/j.eswa.2016.01.006
0957-4174/

Beside the qualitative research reported above that study the behaviour of bicyclists is the work from Dozza and Fernandez (2014); Gustafsson and Archer (2013) and Jahangiri, Rakha, and Dingus (2015a) that report on naturalistic studies where large amount of data are stored during normal bicycling conditions enabling behaviour studies during realistic traffic situations. Jahangiri, Rakha, and Dingus (2015b) also developed machine learning methods to detect driver behaviour while approaching a traffic light controlled intersection. Consequently, there are three approaches to study the behaviour of VRUs (i) Qualitative studies where questionnaires and interviews are used to capture the self reported distraction (ii) Naturalistic studies where e.g. video data of both the view of the bicyclist and the face of the bicyclist are captured during realistic traffic situations and (iii) Bicycle simulator studies where a combination of qualitative and quantitative data are captured. The first approach may give valuable insights of what the VRUs think about different traffic situations however it is typically not feasible to train machine learning algorithms based on the data. The second method is capable of capturing the behaviour of the bicyclist however from an ethical point of view it is not recommended to create hazardous situations in real traffic to study the behaviour of the bicyclist. The third approach, as proposed in this study, can be used to obtain both qualitative data and quantitative data from the simulator during experiments that may involve hazardous events that would be difficult and dangerous to perform in a real traffic environment.

The work in this study is motivated by the rapid introduction of automated vehicles. General knowledge about automatic detection of bicyclist behaviour may improve the safety of automated vehicles and thus improve their acceptance and consequently speed up the introduction. Specifically, characterising bicyclist distraction is important for both the bicyclists themselves and for other road users. Applications where this technology may be of high value are for example, a wearable device that can detect if the bicyclist wearing it is distracted and should take a pause in the activity. Another potential application is an in-vehicle warning system (e.g. Coelingh, Eidehall, & Bengtsson, 2010) that can warn a driver about potentially dangerous distracted bicyclist nearby. Another potential application is in autonomous driving systems, to automatically adjust the distance to the bicyclist based on the estimated distraction level of the bicyclist.

A number of studies exist that use data mining for analysing vehicle driver distraction. In Liang, Lee, and Reyes (2007a, 2007b) cognitive distractions was detected from visual behaviour of the driver and driving performance with the help of dynamic Bayesian networks and support vector machines (SVM). Recently, Liang and Lee (2014) developed a layered algorithm that combined two data mining methods, Dynamic Bayesian Network (DBN) and supervised clustering, to detect cognitive distraction by observing eye movement and driving performance measures. In Doshi and Trivedi (2009) vehicle driver distraction was predicted from vehicle state variables (gas pedal position, brake pedal depression, longitudinal acceleration, vehicle speed, steering angle, yaw rate, and lateral acceleration), environmental variables (road curvature metric, heading, lateral lane position, lateral lane position 10 m ahead, and lateral lane position 20 m ahead) and driver state variables (head dynamics, and eye-gaze measurements). The method used a DIIS (discriminative classifier) that in turn involves a RVM classifier that is based on sparse Bayesian learning. The research by Tango, Minin, Tesauri, and Montanari (2010) also modelled driver distraction using machine learning methods e.g. adaptive neuro fuzzy inference systems and artificial neural networks. The variables in the study was: traffic density, road visibility, standard deviation of speed, standard deviation of steering angle, standard deviation of lateral position, standard deviation of lateral acceleration, deceleration jerk, mean time to line crossing, steering reversal rate and the reaction time of the driver completing a secondary task. Several of these measures are also used in this study. In Liang, Lee, and Yekhshatyan (2012) data from the 100-Car Naturalistic Driving Study was used and showed promising driver distraction classification results using measurements from an eye tracker, the methods used are based on statistics on eye glance history. In our study the experimental design includes a head mounted display (HMD), that completely coves the eyes, and instead of measuring eye glances with an eye tracker, the head movements are used to obtain the gaze direction.

This study addresses the estimation of visual distraction for bicyclists and the aim of this study is twofold:

- To model the self-assessed distraction level based on the experimental data using machine learning.
- To determine the most descriptive variables that affect the model performance.

Experimental data using a bicycle simulator developed at Viktoria Swedish ICT is collected along with a questionnaire accompanying each test regarding distraction, situation awareness and risk assessment and the self-assessed distraction levels.

The methodology in this paper is inspired from both the research fields of cognitive science and machine learning and the main contribution of this paper is the introduction of a bicycle simulator and the data mining methodology where qualitative and quantitative data from simulator experiments are combined to model the self-assessed distraction as a function of the obtained data. In the paper we investigate the effect of combining quantitative data measured from the bicycle simulator and qualitative questionnaire data obtained from the participants. The overall goal is to to model the self-assessed distraction level and we explore the method with the two different data sets (qualitative and quantitative). The combined data is used to investigate what modelling performance can be achieved with the available data. However in a real-life situation it is more likely that only the quantitative data is available and therefore this data is used for comparison.

The remainder of the paper is organised as follows. Section 2 describes the methodology, the bicycle simulator, the experimental design, the participants, the experimental procedure, the data collection and the suggested modelling approach. Section 3 presents the results of the paper including manipulating visual distraction, identification of distraction level, variable importance and distraction level regression results. The findings of the paper are discussed in Section 4 and the paper is concluded in Section 5.

2. Methodology

The methodology is based on a cyber-physical bicycle simulator developed for the experiment. Once the simulator and its virtual environment was established experiments was performed and data collected. The data is used to model the self-assessed distraction level and a variable selection procedure is used to find the distinguishing variables. The following Sections describe the methodology in detail.

2.1. Bicycle simulator

The study was conducted using an interactive fixed-base bicycle simulator built at Viktoria Swedish ICT, see Fig. 1. The simulator has a city bicycle mounted on a bicycle trainer Kurt Kinetic Road Machine. Instead of using projectors for displaying the environment, the simulator is equipped with a head mounted display (HMD) Oculus Rift. The HMD has 110° diagonal field of view (90° horizontal field of view), and a resolution of 640 × 800 pixels per eye. The HMD has a gyroscope, an accelerometer and a magnetometer for head tracking. The HMD with head tracking provides