Real-Time Prediction of Blood Alcohol Content using Smartwatch Sensor Data

Mario A. Gutierrez, Michelle L. Fast, Anne H. Ngu, and Byron J. Gao

Department of Computer Science, Texas State University, San Marcos, Texas, USA
{mag262|mlf96|angu|bgao}@txstate.edu

Abstract. This paper proposes an application that collects sensor data from smartwatches and smartphones in order to predict blood alcohol content (BAC) or drunkenness in real-time, discreetly, and non-invasively through a machine learning approach. This system could prevent drunk driving or other dangers related to the consumption of alcohol by giving users a way to measure personal BAC without the use of intrusive breathalyzers or guess work. Using data collected by a custom application from many volunteers, we trained a machine learning model that will work with an Android application to predict the user's BAC in real-time.

1 Introduction

Internet of Things (IoT) is a domain that represents the next most exciting technological revolution since the Internet. IoT will bring endless opportunities and impact every corner of our planet. In the healthcare domain, IoT promises to bring personalized health tracking and monitoring ever closer to the consumers. This phenomena is discussed in a recent Wall Street Journal article, "Why Connected Medicine Is Becoming Vital to Health Care" [?]. Modern smartphones and smartwatches now contain a more diverse collection of sensors than ever before, and people are warming up to them. In January 2014, approximately 46 million US smartphone owners were reported to have used health and fitness applications [?]. Currently, sports and fitness are the predominant foci of IoT-based health applications. However, applications in disease management and health care are becoming increasingly prevalent. For example, detecting falling of elderly patients [?].

Drunk driving is a dangerous, worldwide problem. This problem is not only a hazard to the drunk drivers, but also to pedestrians and other drivers. It is reported by the Bureau of Transportation Statistics that in 2010, 47.2% of pedestrian fatalities and 39.9% of vehicle occupant fatalities were caused by drunk driving [?]. The Centers for Disease Control and Prevention (CDC) reported that between the years 2008 and 2010, roughly two-thirds of adults were drinkers, with adults between the ages 18 and 24 having the greatest association with heavy drinking [?].

At dangerous levels of intoxication, it can be difficult to judge ones own drunkenness. Instead it would be better to get a definitive measurement of the

BAC, or simply a binary response: ”drunk” or ”not drunk.” Compact breathalyzers are probably the best option at the moment, but these are not discreet and require deliberate action by the user. The other option is to use a smartphone application to manually calculate BAC, but these demand a greater deal of involvement from the user. To be practical, it would be useful to have some sort of non-invasive and accurate monitoring system that will warn its user if they become too intoxicated. It has been shown that electronic intervention programs are more successful at reducing college student drinking than a general alcohol awareness program [**?**]. This system can also be used to warn friends and family, or prevent the operation of the user's car.

The main contributions of our paper are:

- A machine learning model for the prediction of intoxication level from smartwatch sensor data.
- A general Android-based gateway system that can collect data from any type of physical or virtual sensor accessible by the host smartphone.
- An overall, sensor middleware system, including a geographic visualization service, that may be useful for answering many other important research questions.

2 Related Work

There are a few ways of approaching the problem of determining a person's blood alcohol content. One approach is to model mathematically the elimination of ethanol in the human body. In this case, the Widmark equation, published in 1932 by E.M.P. Widmark, is a very popular one,

$$C = \frac{A}{rW} - (\beta t) \tag{1}$$

where C is the BAC, r and β are empirically determined constants, A is the mass of the consumed alcohol, and W is the body weight of the person. These days, there have been several improvements and variants. Douglas Posey and Ashraf Mozayani published an excellent article comparing this model using parameters determined by different researchers and discussing different models [**?**]. They found that the Widmark equation tends to overestimate, and that there can be significant discrepancy between the results of the different models. Despite that, they do provide a rough estimate. The problem is that these models also require a good deal of information that prohibit their use in a non-intrusive, drunkenness warning system.

Another approach to the problem is simply to measure the BAC directly. Transdermal ethanol sensors have been a recent option for this approach. These can provide a discreet way to measure intoxication, but they are accompanied by the problem of a significant time lag between the sensed alcohol concentration and actual blood alcohol concentration. Gregory D. Webster and Hampton C. Gabler closely investigated this problem. They found that the lag is predictable, but not constant, and requires additional information about the drinks taken to accurately predict [?].

Similar to our project, James A. Baldwin has a patent on a system involving a wearable transdermal ethanol sensor and a mobile device to capture the information [?]. Baldwin describes his system as using a mathematical model to predict the user's BAC given the transdermal sensor data and information about the drinks the user plans to consume. A benefit of our system is that it involves no input by the user about the drinks taken, and the user need not buy a special sensor dedicated to this task alone.

Aside from measuring BAC directly, or developing a biologically-based mathematical model, machine learning is another good approach. Georgia Koukiou and Vassilis Anastassopoulos published research this year in using a neural network to identify drunkenness from thermal infrared images of peoples' faces [?]. Neural networks were trained on different parts of the face in order to determine which areas can be used to classify drunkenness. They found the forehead was the most significant facial location to observe for determining the drunkenness of a person. Their study takes advantage of the effect of alcohol making blood vessels dilate allowing warm blood to come closer to the skin; which is also an important effect for our research. Such a system may be good for ignition interlock systems, or drunk surveillance.

Outside of BAC studies, there has been plenty of research into detecting other activities using smartphone and smartwatch sensor data. In [?], John J. Guiry, Pepijin van de Ven, John Nelson, attempt using the sensor data to identify various daily activities, such as: walking, running, cycling, and sitting. In their study, they use several machine learning algorithms for their approach: C4.5, CART, Naïve Bayes, ANN, and SVM. Their results showed some promise for better future models, with their model for classifying whether a user is indoors or outdoors being the most impressive. Successful models for predicting daily activities will certainly be important in a practical implementation of our system. This is because the body's response to alcohol consumption may share significant similarity to exercise, dance, or other activities.

Using smartphones and smartwatches, there is an active desire to create monitoring applications for serious health problems. Such as in [?], where Vinod Sharma, Kunal Mankodiya, Fernando De La Torre, *et. al.*, developed a smartwatch-smartphone system for the monitoring and analysis of data from patients with Parkinson Disease. This system, named SPARK, includes the analysis of speech and detection of: facial tremors, dyskinesia, and freezing of gait. Their system is intended to provide useful recommendations to physicians based on the collected information. They concluded noting some potential problems of a full implementation of their system, the most relevant problem being misplacement of the sensors. This may also be a problem for us considering the potential importance of the motion-based accelerometer and gyroscope data.

3 System Architecture

3.1 Android-based Sensor Gateway

4 Methodology

In this section, we will describe how we collected the data, we will present an analysis of the data, and then lead into the discussion of the machine learning models used.

4.1 Collection

Our collection began with the development of an Android application that connected to a Microsoft Band smartwatch and collected data from all of its available sensors. This development led to the general Android-based sensor gateway discussed in the previous section. Next, we developed a general procedure for our volunteers to follow during the collection of data. Our volunteers were eager to freely contribute the anonymous data used in this paper. The system collected the data into a .csv file on the Android smartphone and was also transmitted to a central MySQL server.

The Android platform was version 5.0 (kernel 3.4.0-4432708) running on a Samsung S5 smartphone. The Microsoft Band was Build Version 10.2.2818.0 09 R. Samples of the sensor data were collected every three seconds, based on the update speed of the Band's heart rate sensor. For every sample, the most recent sensor value was used for every sensor if available, else it would use the last updated value; or in the case of the accelerometer and gyroscope sensors, the last three values were averaged with linear weighting (the most recent having the most importance). There may be a better weighting, but this weighting was suitable for our purposes.

We designed a simple, two-hour procedure for the collection of the data. First, we established some necessary information about the subject to estimate the amount of alcohol necessary to reach 0.08 BAC in a 1.5 hour period using the Widmark equation (1). The particular formulation of it we used is the following:

$$SD = BW \cdot Wt \cdot (EBAC + (MR \cdot DP)) \cdot 0.4690 \tag{2}$$

where SD is the number of standard drinks (10 grams ethanol), BW is the body water contant (0.58 for men and 0.49 for women), Wt is the body weight in lbs, EBAC is the estimated BAC, MR is the metabolism rate (0.17 for women and 0.18 for men), DP is the drinking period in hours, and $0.4690 = 0.4536 \div (0.806 \cdot 1.2)$, a combination of two constants from the equation and a converstion from kg to lbs [?][?]. This amount was used to estimate the number of standard drinks to be consumed over the set time period, distributed over equal intervals. During this process, at every 25 minutes we took a measurement of the BAC using a BACtrack Trace$^{\text{TM}}$ Pro breathalyzer. This measurement interval was determined by the cooldown rate of the breathalyzer. The activity chosen for the volunteers to engage in was a card or board game of their choice. Drinking stopped before 1.5 hours while collection and BAC measurement continued for another 30-45 minutes.

4.2 Data Analysis

4.3 Feature Selection

4.4 Model Building

5 Evaluation

The evaluation.

6 Conclusions

Conclude and stuff.

Acknowledgement

We thank the National Science Foundation for funding the research under the Research Experiences for Undergraduates Program (CNS-1358939) at Texas State University to perform this piece of work and the infrastructure provided by a NSF-CRI 1305302 award.