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 a smartwatch in order to predict drunkenness in real-time, discreetly, and non-invasively via 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 determine personal intoxication level without the use of intrusive breathalyzers or guess work. Using smartwatch data collected from several volunteers, we trained a machine learning model that may work with a smartphone application to predict the user's intoxication level 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" [8]. 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 [9]. 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 [13].

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 [4]. 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 [11].

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 [14]. This system can also be used to warn friends and family, or prevent the operation of the user's car.

In this paper, we investigate the prediction of intoxication level from smartwatch sensor data via machine learning. We also briefly discuss a general Android-based gateway system which can collect data from any type of physical or virtual sensor accessible by the host smartphone.

2 Related Work

There are a few ways of approaching the problem of determining a person's blood alcohol content. One approach is to devise a mathematical model of 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 [10]. 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 number of drinks taken by the user to accurately predict it [15].

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 [3]. 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 [7]. 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 [6], 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 [12], 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

A huge end-goal of ours is to have a public system where participating users' phones transmit sensor information to a central database through a REST-based web service. The service stores this information in the database and another data processing system. Other users who are not contributing information can connect to our system to access other related services, like a choropleth map of the sensor data; BAC prediction data in our case, but the system would not be limited to it. The particular protocols and setup used should be designed to protect the privacy of the sensor data contributors.

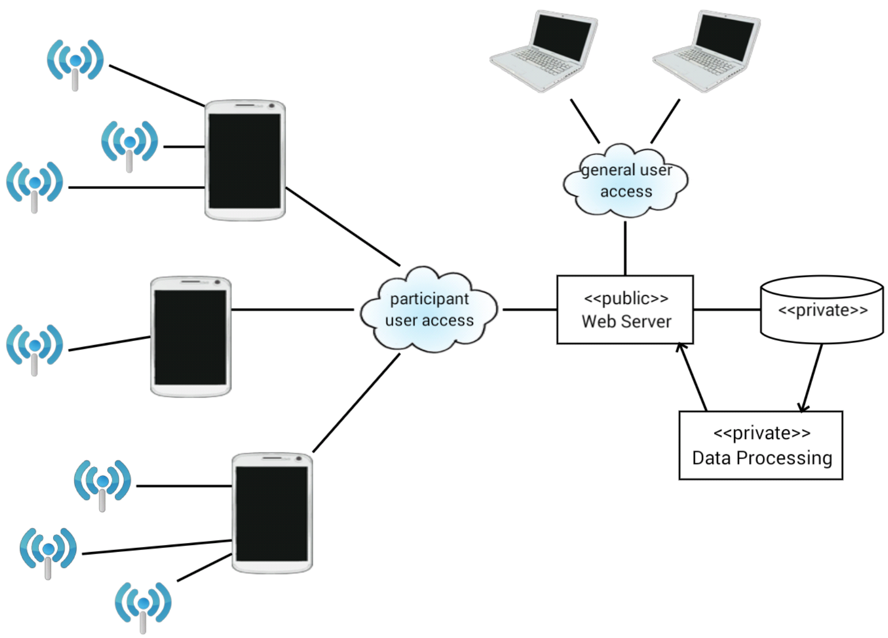

Fig. 1. Example general system for large-scale sensor data collection

3.1 Smartphone Sensor Gateway

Each of the smartphones in this system are behaving as a type of sensor gateway. In order to collect our data, we built such a system on the Android platform. Though we used primarily a Microsoft Band, our system allows the easy addition of any sensor that can be connected to from the Android smartphone. In Figure 2, we show an overview of the classes used in this system.

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.

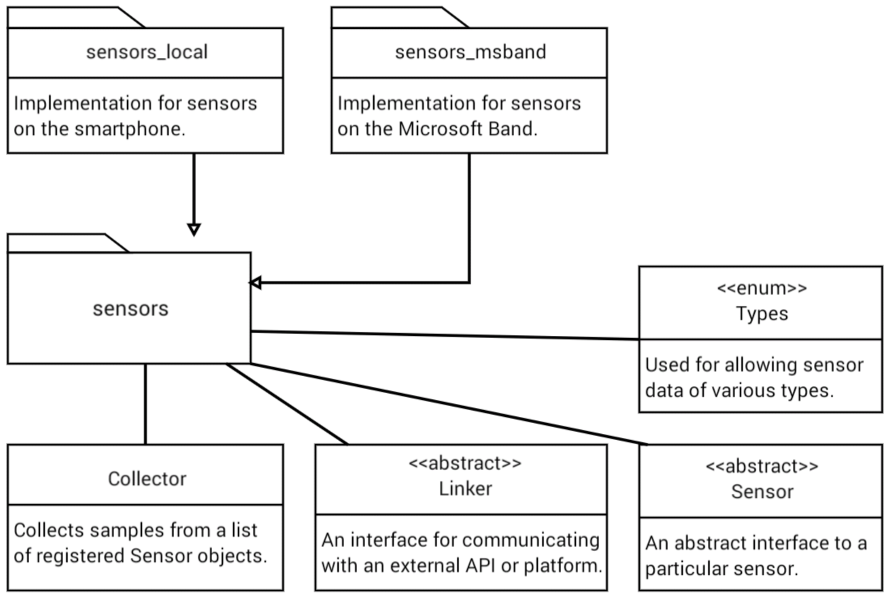

Fig. 2. Gateway implementation for our data collection application

4.1 Collection

Our collection began with the development of an Android application that can connect to and store data from available sensors on the Microsoft Band smartwatch. 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 data collected by the system was stored into a .csv file on the Android smartphone and also transmitted to a central database server.

We used Android platform version 5.0 (kernel 3.4.0-4432708) running on a Samsung S5 smartphone. The Microsoft Band had 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. Each sample used the most recent sensor value if available, otherwise it would use the last value; or in the case of the accelerometer and gyroscope sensors, the three most recent values were averaged with linear weighting (the most recent having highest importance). There may be a better weighting, but this weighting was suitable for our purposes.

The procedure followed by the volunteers during data collection lasted two-hours. First, we established some necessary information about the subject in order 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 [1][16]. 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

The Microsoft Band we used has an assortment of interesting physical and virtual sensors, including: accelerometer, gyroscope, distance, heart rate, pedometer, skin temperature, and ultraviolet level. With our sample size of only five volunteers and a controlled setting for the experiments, some of these sensors will not be very useful for this study, such as the: distance, pedometer, and ultraviolet level, sensors. In fact their usefulness may be limited even with larger datasets.

We begin our analysis with the heart rate (HR) data by normalizing the heart rates and BACs per subject. This way we can plot and compare the HR values over the BAC and see if there are any obvious patterns. Doing this we get the plots shown in Figure 3.

The heart rate increases over time for three of the subjects (0, 2, 3) at different rates, decreases for one (4), and stays level for one (1). These observations are consistent with data in [2]; a study where the authors relate low and high HR responses to behavioral traits. Interestingly, most of the subjects seem to show two patterns of HR activity. One is the baseline HR activity, and another is an excitied HR activity that seems to have better correlation with BAC. This is most clearly seen in the plot for subject 0, shown in Figure 4. It's not a consistent pattern, however. Subject 1 had a very level heart rate with spikes at the regular drinking intervals, and subject 2 had relatively little variance throughout. Overall, it seems that experiencing excitement while intoxicated results in a more exaggerated HR response than while sober. This may be useful information for determining drunkenness.

Next, we plotted the normalized skin temperature over time on top of the normalized BAC values; this is shown in Figure 5. Visually, we see that the correlation of skin temperature and BAC is highly significant. This is because of

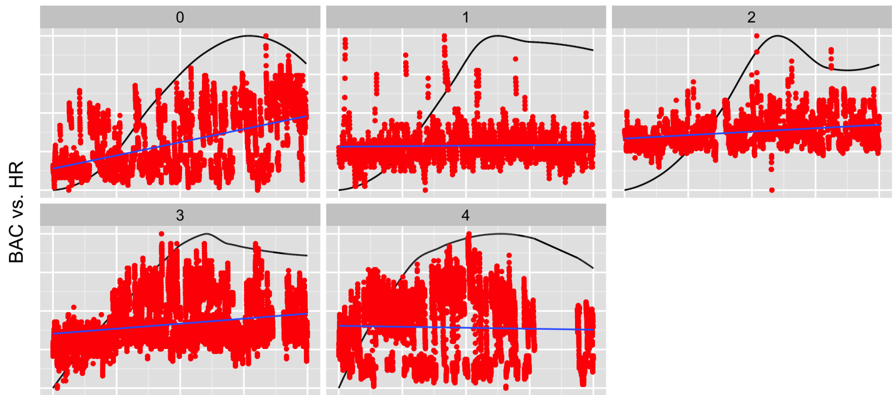

Fig. 3. Heart Rate and BAC per subject (normalized)

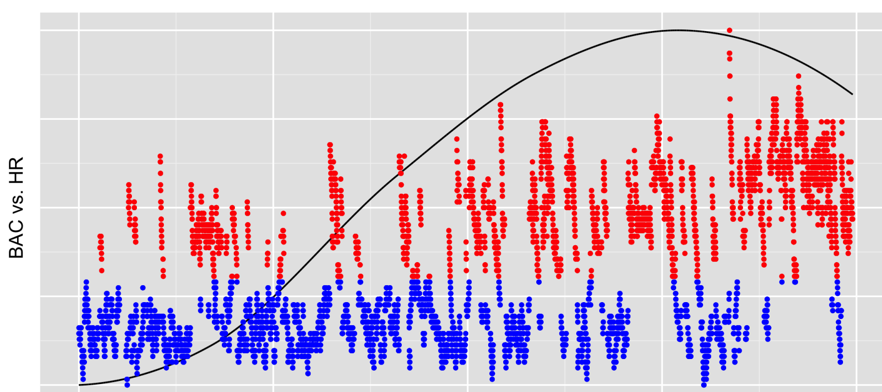

Fig. 4. Heart rate split into baseline and excited states

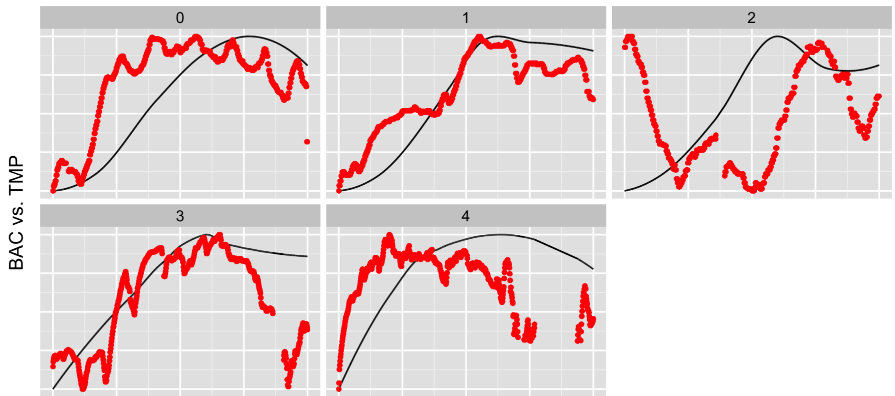

Fig. 5. Skin Temperature and BAC per subject (normalized)

the well known vasodilation effect of alcohol causing warm blood to come near the surface of the skin [5]. It is also known that with a high enough level of intoxication it behaves as a vasoconstrictor and drops skin temperature, which we see with subjects 3 and 4 who had reached the highest BAC of the group. Subject 2 had a rapid decrease in temperature at the beginning (we may have started collection too soon before the smartwatch sensor adjusted to his skin temperature), it may be appropriate to discard this portion.

The movement sensors (accelerometer and gyroscope) were not too interesting visually or statistically. They might have some hidden patterns that contribute to the performance of a modeling technique. We think its usefulness will be increased if we can use it first to predict whether a user is taking a drink, then provide a time-based aggregate of the estimated number of drinks as a feature in our model to determine drunkenness.

4.3 Features Setup

Skin temperature and heart rate will have different ranges of values unique to each subject, so per-subject normalization must be performed to put them on the same scale. A simple unity-based normalization (min-max) is used to put the feature values in the range 0 to 1. No transformations are applied to the movement data.

4.4 Models and Tools

We will be taking a look at BAC prediction as both a classification and regression problem. As a classification problem, we will threshold the BAC values at a point where we may want to warn the user. This allows the data to be used as a binary classification problem. For this we will use logistic regression and SVM.

As a regression problem, we will be attempting to predict the observed BAC directly. To do this, we will use linear regression and artificial neural networks (ANN). Using a regression approach, this will allow a user to select their own threshold rather than the threshold determined in a classification model. We may however want the threshold to be fixed.

All work will be performed in R version 3.2.1 using the additional 'nnet' and 'kernlab' packages. Any necessary additional information will be provided in the evaluation section.

5 Evaluation

Overall, our data set contains 233,538 samples that were collected from five volunteers. Each sample was collected every three seconds from a Microsoft Band smartwatch by our Android data collection system. In this section, we evaluate the performance of a few machine learning models on our dataset. We use some standard performance measures for our evaluation: precision, recall, and F1-score, for the classification models; RMSE and R^2 for the regression models. All reported performance values are determined via 5-fold cross-validation.

5.1 Classification

We want to warn a person if they are close to reaching the legal limit of 0.08 BAC. A good time to warn a user is at about 0.065 BAC. Using this threshold, the classes are split into 64% is *DRUNK*, and 36% *SOBER*.

To get a baseline, we first trained a logistic regression model. This model outputs values from 0.0 to 1.0, so we need to determine where to best split this output into each class. In Figure 6, we show a plot of the predictions of the model on the test data, using the actual labels to distinguish the output. In this plot, we see that the best threshold value for the logistic model predictions is around 0.32; above that we classify *SOBER* and below that *DRUNK*. Using this, we achieved a precision of 0.855 ± 0.002, recall of 0.730 ± 0.004, and F1-score of 0.787 ± 0.003.

Moving forward, we trained a SVM model using the Gaussian Radial Basis Function (RBF) kernel. For time constraints, the dataset was reduced by half using a uniform subsample. Even so, we found the SVM was able to achieve great performance with a precision of 0.886 ± 0.002, recall of 0.930 ± 0.002, and overall F1-score of 0.907 ± 0.001. Ideally we do not want to warn our users that they are drunk when they are not actually drunk, so we want to try and optimize the model to be as precise as possible. By modifying the error weighting to train against false positive errors, the SVM model achieved a precision of 0.970 ± 0.002, with recall of 0.729 ± 0.003, and F1-score of 0.832 ± 0.002. Our recall dropped for a higher precision, which is a well-known tradeoff for this kind of tuning, but this is fine. We can set a threshold on our smartphone warning system that if at least a recall fraction of the samples from our smartwatch are classified as *DRUNK*, then there is a precision chance that the user is actually at 0.065 BAC.

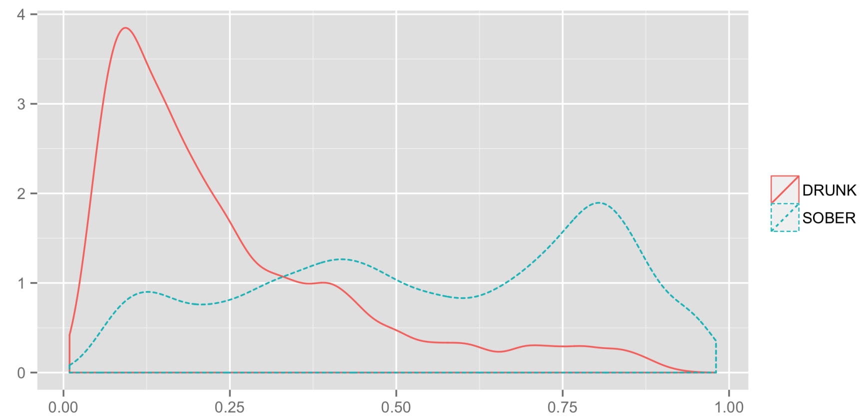

Fig. 6. Logistic regression output frequency with actual labels

5.2 Regression

So how does this do as a regression problem? We first considered the most basic: a linear regression model. This did not perform well at all. We next trained a neural network (ANN) model on the data. The best performing ANN had

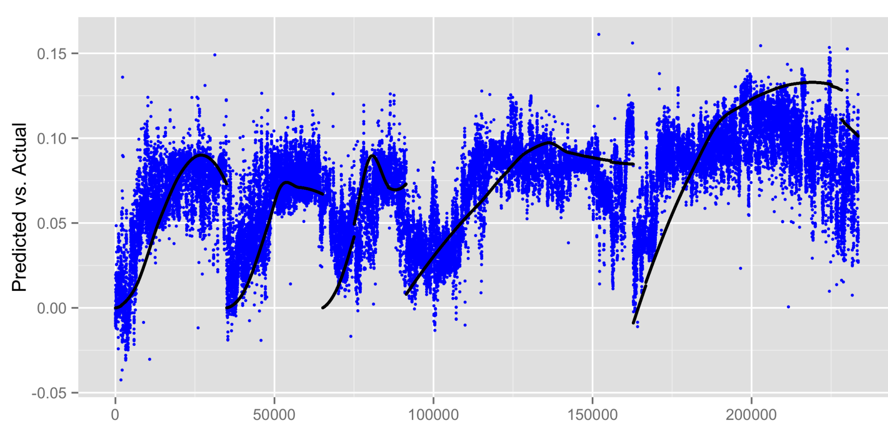

Fig. 7. ANN predictions (blue) with actual BAC data (black) on test partition

a structure with twenty nodes in one hidden layer. This was determined by doubling and then reducing node count to find the best performance. Additional layers did not show any improvement (the 'neuralnet' package was used to test multiple layers). Figure 7 plots the BAC predictions of the ANN along with the actual values using data from a test partition. The ANN model achieved a R^2 value of 0.524 ± 0.015, with RMSE of 0.026 ± 0.000. It performed much better than the linear regression model, but not nearly as good as the classification models.

6 Conclusions

We investigated the design of a general system that could be used for many applications, whether it be a BAC warning system, or a geographical mapping service that displays the smartwatch data of users on a choropleth map. Part of this system was a general Android-based gateway that can be used to collect data from a variety of sensors connected to an Android smartphone.

We then designed an experiment to collect labeled data consistently from volunteers. After this, we analyzed the data to discover if there were any interesting patterns that were immediately obvious. We found that skin temperature was a good indicator of drunkenness (in our controlled setting). We also discovered that excited heart rate looks to also be a good indicator of intoxication level.

Following the overview analysis, we dove into training some regression and classification models. Achieving good performance as a regression problem was difficult. We found that the problem was much better tackled as a classification problem. This worked better because our classification models could ignore a good deal of the variance in the straight BAC predictions. In the end, we found SVM to perform the best on our data.

There are still many other factors to consider in forming a better models. From this research, we found that the accurate prediction of drunkenness in a real application looks possible in theory. There are still many obstacles surrounding the collection of a larger and better data set. Ideally, we would want a data set from a thousand volunteers in candid situations over several days. The two biggest problems are, how do we label the data with the alcohol levels in this scenario, and how can we model this enormous amount of data in reasonable time? Also, will we need to first determine the activity of the user, or will the other sensors provide sufficient information? And if the former, how can we consistently tag these activities in the data? The answers to this may be simple, or infeasible. In any case, it would be worth it to try and find out.

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.

References

1. Andersson, A., Wiréhn, A.B., Ölvander, C., Ekman, D.S., Bendtsen, P.: Alcohol use among university students in sweden measured by an electronic screening instrument. BMC Public Health 9, 229 (2009)
2. Assaad, J.M., Pihl, R.O., Séguin, J.R., Nagin, D.S., Vitaro, F., Tremblay, R.E.: Intoxicated behavioral disinhibition and the heart rate response to alcohol. Experimental and clinical psychopharmacology 14(3), 377 (2006)
3. Baldwin, J.: Wearable blood alcohol measuring device (2014), http://www.google.com/patents/US20140365142, uS Patent App. 14/298,847
4. Chambers, M., Liu, M., Moore, C.: Drunk driving by the numbers
5. Dekker D.O., A.: What are the effects of alcohol on the brain? (July 1999), http://www.scientificamerican.com/article/what-are-the-effects-of-a/, [Online; accessed 16-Aug-2015]
6. Guiry, J.J., van de Ven, P., Nelson, J.: Multi-sensor fusion for enhanced contextual awareness of everyday activities with ubiquitous devices. Sensors 14(3), 5687–5701 (2014)
7. Koukiou, G., Anastassopoulos, V.: Neural networks for identifying drunk persons using thermal infrared imagery. Forensic Science International "252", 69 – 76 (2015), http://www.sciencedirect.com/science/article/pii/S0379073815001681
8. Landro, L.: Why connected medicine is becoming vital to health care (June 2015), http://www.wsj.com/articles/why-connected-medicine-is-becoming-vital-to-health-care-1435243250
9. Hacking health: How consumers use smartphones and wearable tech to track their health (April 2014), http://www.nielsen.com/us/en/insights/news/2014/hacking-health-how-consumers-use-smartphones-and-wearable-tech-to-track-their-health.html
10. Posey, D., Mozayani, A.: The estimation of blood alcohol concentration. Forensic Science, Medicine, and Pathology 3(1), 33–39 (2007)
11. Schoenborn, C.A., Adams, P.F., Peregoy, J.A.: Health behaviors of adults: United states, 2008-2010. Vital and health statistics. Series 10, Data from the National Health Survey (257), 14–29 (2013)
12. Sharma, V., Mankodiya, K., De La Torre, F., Zhang, A., Ryan, N., Ton, T., Gandhi, R., Jain, S.: Spark: Personalized parkinson disease interventions through synergy between a smartphone and a smartwatch. In: Marcus, A. (ed.) Design, User Experience, and Usability. User Experience Design for Everyday Life Applications and Services, Lecture Notes in Computer Science, vol. 8519, pp. 103–114. Springer International Publishing (2014)
13. Tacconi, C., Mellone, S., Chiari, L.: Smartphone-based applications for investigating falls and mobility. In: Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on. pp. 258–261 (May 2011)
14. Ward, L.: How to cut student drinking. The Wall Street Journal p. R5 (2015)
15. Webster, G.D., Gabler, H.C.: Feasibility of transdermal ethanol sensing for the detection of intoxicated drivers. Associating for the Advancement of Automotive Medicine 51, 449–464 (2007)
16. Wikipedia: Blood alcohol content — wikipedia, the free encyclopedia (2015), https://en.wikipedia.org/w/index.php?title=Blood_alcohol_content&oldid=674051485, [Online; accessed 13-July-2015]