(WIP Title) Predicting Sobriety Levels Using Gait and Other Biometrics

Kevin Martin
dept. name of organization (of Aff.)
name of organization (of Aff.)
Los Angeles, California
kmarti44@syr.edu

Ravjot Sachdev
dept. name of organization (of Aff.)
name of organization (of Aff.)
Syracuse, New York
rssachde@syr.edu

***Abstract*—We combine modern telemetry recording techniques and novel alcohol content recording methods with advanced machine learning techniques to detect and identify whether or not an individual is inebriated. As the quantity of alcohol increases in the body, so too do the changes in movement and walking patterns for the individual. We supplement our analysis with similar work, ranging from other movement patterns to entirely different biometric analysis. The comparison between methods shows which direction may hold the most promise for future, real-time inebriation detection.**

***Index Terms*—BAC, biometrics, gait, sobriety, TAC**

I. Introduction

There are many ways to qualitatively and quantitatively measure a subject's blood alcohol content (BAC). Standard Field Sobriety Tests (SFSTs) might be administered by a trained individual, such as a police officer, to determine whether it is safe for a driver to continue to operate a vehicle. Breathalyzers are also used to give a precise measurement of one's BAC. In recent years, more automatic means of measuring BAC have emerged including, but not limited to, thermal eye imaging, thermal facial imaging, and gait. It is well-known that alcohol affects one's ability to walk, which is why SFSTs, such as the walk-and-turn test, measure. Therefore, it should be equally possible to extract features from accelerometer data to analyze a subject's gait and estimate their BAC.

II. Previous Works

A. Thermal Eye Imaging

For a sober person, the iris and the sclera in the eye are similar temperatures. Therefore, thermal imaging of the eye will show little to no differences for these areas. As an individual's BAC increases, so does the temperature of their sclera. This has the effect of the iris appearing *darker* in a thermal image of an inebriated person than in a sober person, as indicated by **Figure 1**. This data gathered from cameras can be used to classify an individual as either sober or drunk [1].

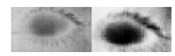

Fig. 1. Thermal images of an eye. Sober person on the left, drunk person on the right
Source: Drunk Person Screening using Eye Thermal Signatures

B. Thermal Facial Imaging

Alcohol has a visible affect on the color of one's skin, making areas, particularly around the nose, eyes, and forehead, more distinguishable. This is due to the blood vessels becoming more active as a result of the alcohol. To other people, these areas of the face appear more *red*. These temperature differences can be measured, and a classification can be made on whether the subject is drunk or sober [2].

C. Gait and Walking Patterns

A person's ability to walk degrades as their BAC increases. This fact is what makes SFSTs, such as the walk-and-turn, so effective. "However, SFSTs were designed to make intoxication apparent to a trained law enforcement office who manually proctors them" [On Smartphone Sensability of Bi-Phasic...]. Manual administration of these tests introduces room for error. 3-axis accelerometer data, gathered from mobile devices, makes way for a biometric system to automatically classify a user has either drunk or sober. Studies have attempted to classify participants when their BAC was both increasing and decreasing. Regression techniques used include linear regression, a Bayesian regularized neural network, and a support vector machine [Using phone sensors and...]. By comparing the results from previous studies, we hope to better determine the usefulness of our own experiment.

III. Preparation

The dataset is comprised of multiple CSV files, each related to 12 individuals from a prior study. There is a separate master file which contains all the telemetry data (the x, y, and z-axis readings) measured each nanosecond, classified by individual [Learning to Detect Heavy Drinking Episodes Using Smartphone Accelerometer Data]. This file contains approximately 14 million telemetry readings. There are also 12 files (one for each participant) which hold the alcohol readings throughout each time frame. The alcohol is measured in TAC (Transdermal Alcohol Content) which allows for less invasive data collection than something like a traditional breathalyzer. Additionally, the telemetry readings are taken every nanosecond, whereas the TAC readings are taken every thirty minutes.

As such, there are significantly less TAC reading records than there are telemetry readings.

A. Feature Extraction

The master file, which holds the telemetry readings, was the starting point. We imported the file using the Python library Pandas and created a dataframe. This allowed us to easily manipulate and match up the participants and their readings. We created a new column that was a copy of the original timestamp column, only this time it was formatted in seconds. This allowed for us to match the TAC readings (which are formatted in seconds) to the accelerometer readings (which were originally formatted in milliseconds).

Next we created our new features. Using a single point in time (x, y, and z positioning) could not possibly help identify a person's inebriation level. We are focusing on motion. In order to capture motion, we used three rolling averages for each of the three axes. We used a rolling previous 500, 1,000, and 5,000 mean to approximate motion volatility. Our thought was that the higher the rolling averages were, the more likely the individual was to be over the legal limit. See **Figure 2** for an example of how rolling averages and TAC display similar movements throughout the relevant time period. Additionally, we kept all three axes separate, but used all as features in our classification algorithms. This allowed us to test three different features for each TAC reading across three different time frames, for a total of nine unique features.

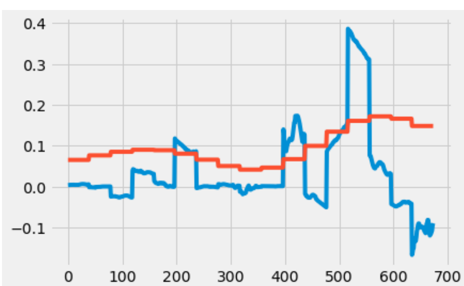

Fig. 2. Sample rolling average of the x axis vs. TAC reading

B. Data Manipulation

Next we imported each user TAC reading data, which were saved as separate CSV files. Again we used Pandas to create a dataframe for each, only this time we added a merge function to also pull in the accelerometer readings. Because the TAC readings were done every 30 minutes, we were only able to match a few thousand user timestamps. However, this also meant we could include all of the nine rolling averages as well. In effect we were pulling in the absolute x, y, and z reading, but also the previous 500, 1,000, and 5,000 averages. Our rationale was that since the TAC readings were so spread out, the accelerometer readings between those measurements were largely irrelevant. Thus we focus on only the critical times.

After merging the accelerometer readings to the user data, we had twelve new dataframes. We kept these preserved, but found it most useful to combine all into one summary dataframe using the simple concat function built into pandas. This left us with over 5,000 instances of TAC readings matched to precise accelerometer readings as well as our nine rolling averages. From there we could extract both user-level data as well as aggregate information.

IV. Methodology

We had two main goals for the actual testing. First we wanted to see if we could, based on our rolling averages, predict the TAC level of a participant. This takes the form of a regression problem as the target variable is continuous. We look to use multiple methods and evaluate the results of each: simple linear regression (as a baseline), a decision tree regressor, and a random forest regressor.

Our second approach, and the one that we will focus on more for our biometric system, is a binary classification system. Our goal is to predict whether or not someone is over the legal limit, so while the regression results would be interesting, they would not prove as useful for our overall goal. Additionally, we believe that it is an easier problem to solve in that the classifier can be more imprecise with its predictions.

In order to train classifier algorithms, we added a new column to the summary dataframe. The column would be populated with a "1" for any TAC greater than or equal to .08, and a "0" for any reading less than .08. This new column will be the target column for which we will have the classification algorithms focus. This gives us a decent distribution, with about twice as many sober readings as inebrieated, which can be seen in **Figure 3** below:

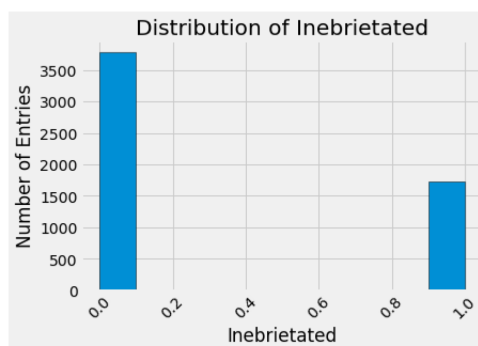

Fig. 3. Binary classification results

Again, we utilize several different methods and compare the results. First we use a Naive Bayes as a general baseline. Next we try a decision tree classifier without any hyperparameter adjustments. This also serves as a baseline, but more for the tuning aspect. Next we try a grid search and use cross-validation to help find the optimal hypterparameters for the decision tree. Finally, we use a "pre-tuned" model, XGBoost. XGBoost is a variant on gradient boosting, and generally performs well despite not requiring much tuning.

V. Results

A. Regression Testing

Initially, we used three separate regression models to analyze the correlation between the moving averages from the telemetry data and the TAC readings. The three regression models we chose were linear regression, decision tree regression, and random forest. For each regression model, we calculated the root mean square error(RMSE) and the mean absolute error. Of the three models, the random forest model performed the best, with a RMSE of approximately 0.0085. In comparison, the RMSE of the linear regression model for the same dataset was approximately 0.049, and 0.011 for the decision tree.

B. Binary Classification

To examine the data through a more concrete lens, we added a field to group the TAC readings into one of two classes: inebrieated or sober. The dividing line was a TAC reading of greater than or equal .08, the legal limit for operating a motor vehicle in most states. This allowed us to use a different set of algorithms, and the results were very encouraging. Based on the three sets of rolling averages for each axis reading, three of our algorithms are able to achieve accuracy ratings of over 97%. Our most successful method, XGBoost, which had an accuracy of over 98%, was able to beat not only the decision tree, but also the grid search with optimized hyperparameters. See **Figure 4** below for a summary of these results:

Method	Hyperparameters	Accuracy	Precision	Recall	F1
Gradient Boosting - XGB	N/A	98.97%	97.84%	98.81%	98.32%
Decision Tree - Grid Search	critereon, max_depth	97.76%	95.34%	97.42%	96.37%
Decision Tree	None	97.70%	95.33%	97.22%	96.27%
Naive Bayes	None	35.69%	30.65%	87.89%	45.46%

Fig. 4. Binary classification results

VI. TAC vs BAC

VII. How Our System Compares

Classifying an individual as sober or inebriated using a biometrics system is hardly a novel idea. In fact, there have even been studies that used gait. Comparing the results of our experiment with previous studies helps to determine which methodology may prove more accurate or more beneficial for future automatic detection of inebriated individuals.

A. Facial Thermal Imaging Techniques

Normally, the termperature of a person's skin is around 33.5°C, but after consuming alcohol "there is a temperature increase in capillary density, such as around the nose, forehead, and eyes" [**?**]. This provides measurable features that could be used by a biometric system to determine whether a subject is inebriated.

One study collected thermal images of participants and attempted to determine how many beers they had consumed, ranging from zero to four beers. As seen in **Figure 5**, a 22-point grid was overlaid over each face to obtain a 22 dimension feature vector. A Gaussian Mixture Model (GMM), a supervised learning algorithm which utilizes Gaussian distributions to form a probability density function, was selected to classify the images into the five different classes.

The test data selected for the study was from the class of subjects who drank 4 beers. In the end, the biometric system classified approximately 13% of these subjects as *sober*, meaning it had an accuracy rate of about 87%. One of the limitations with this study is in the test data, as the test data was limited to non-sober participants only.

Another study followed a very similar methodology: a 22 dimensional vector used as input to a GMM algorithm. In this study, instead of trying to predict the number of beers one had consumed, they simplified it to a binary classification problem of either drunk or sober. Although the problem is simplified, it is important to note that the results for this study seem misleading: "When the classification was done on drunk drivers approximately 13% were classified as drunk and 87% as sober. Hence the accuracy of the classification stage is 87% approximately" [**?**]. Even if the accuracy was 87%, as the paper states, our system outperforms both of these systems that use thermal facial imaging.

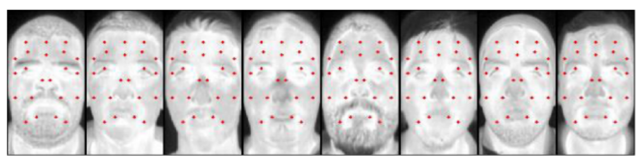

Fig. 5. 22-point grid overlaid over thermal images

B. Eye Thermal Imaging Techniques

VIII. Future work

IX. Conclusion

References

[1] Koukiou, G.; Anastassopoulos, V. Drunk person screening using eye thermal signatures. J. Forensic Sci. 2016, 61, 259–264.

[2] M. K. Bhuyan, S. Dhawle, P. Sasmal and G. Koukiou, "Intoxicated Person Identification Using Thermal Infrared Images and Gait," 2018 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, 2018, pp. 1-3, doi: 10.1109/WiSPNET.2018.8538761.