Drunk User Interfaces: Determining Blood Alcohol through everyday smartphone tasks

Introduction: breathyluzers are standar method to assess inebriation but the DUI app can user ML modesl to determine inebriation, specifically Bal. Modern smartphones can evaluate human performance.

Necessity: situations where DUI can be used to do stuff and a person can use it to evaluate their own performance. DUI measures side effects of blood alcohol, not blood alcohol level. Using human performance metrics allows ml model to account for a lot of variables. Current data only shows sober and inebriated.

Discussion: DUI vs breathylyzer: different approaches. Limitations of study: make sit less practical but fatigue may not be accounted for well in a irl situation. Its very hard to extend this from a controlled setting to a real setting. Lots of comment on impracticality of process. Many cases where without alcohol consumption, cognitive performance might still be bad(someone working a 20 hour shift). Perhaos study can account for other variables as well? Alcohol can be a motivating factor but given that cognitive performance can be impaired from whatever else, what should we do? While designing tech, you prob shouldn’t re-create tech nd you should probably check your assumptions. Are you accessing a broader space than what you are focusing on?

Previous attemots tomeasure alcohol consumption:

Breathlyzers, LCD watch, SCRam, saliva color analysis, user self reports drunk behavior, geotagged tweets, nystagmus: using eye movement, various attempts without using ML models

Typing task:

User’s mototr abilities and coordination, type given phrase as quickly and accurately, autocorrect and cursor are disabled, trade off – fix mistakes with backspace or leave mistakes as is. Baseline is established when you are sober.

Discussion: How do you hack this system? Establish baseline when drunk, give it to a friend etc. Generally a good idea to keep in mind that people will hack a given design.

Swiping task:

Fine motor control thru gesturing. Swipe correctly thru a 3 by 3 lock screen. Compares trajectory of user’s fingers to the ideal 3-segment chase that connects 4 circles.

Discussion: very similar to Hevelius.

Balancing + heart rate task:

Measures user’s ability to keep smartphone flat. PPG – light that see how much blood goes thru a person. With alcohol, it might be different.

Discussion: At CMU, a camera device that tracks hepatitis and jaundice across people.

Simple reaction task:

User’s alertness and motor speed. Perform touch down and finger lift gesture in response to a randomly timed stimulus. DUI records time difference between square’s color change and the expected action. Gets data from touchscreen, gyroscope etc.

Discussion:

Choice reaction task:

Measures user’s alertness and motor speed. Compliacated version of fourth task

Discussion: Why the need for fourth task? Could be motor vs cognitive task- fourth vs fifth.

Machine learning:

Generates human perf metrics used for training a regression model that estimates BAL. DUI uses random forest regression models, which are known to be accurate.

Discussion: paper is an artifact(new app), methodological(method since authors made this new method and evaluated it afterwards), survey?(since they do a good job of existing work in the area), empirical. Is this theoretical? Based on concepts and materials that have been used already and there is no new concept that we are learning.

User Study for DUI:

Satisfy NIAAA guidelines, proof of being over 21, no family history of alcoholism, not taking emds interacting with alcohol, female participants take pregnancy test, required to stay with research team until sober.

Participants:

14 participants – 9 male, 5 female

21 to 35 years old

Caucasian, South Asian and Asian Students

Probably grad students? Given racial makeup and age

Schedule:

Control for fatigue: Sessions had to have same start time +- 1 hour. Conflict pushed session back full day. Max 48 hour gap between sessions,

Procedure:

One shot of 80-proof vodka every 10 mins.

Continue shots until target BAL

5 trials of all 5 tasks take about 5 mins.

Discussion: Placebo control may be a good study because people’s conginitives and external behaviors might be affected. It may not be a good study because at that point, you would be evaluating the placebo as opposed to evaluating the app.

Two competing models:

First baseline does not account for learning and may stop users from gaming the app.

Second baseline may be good in accounting for learning but from a usability perspective, it may not be good in IRL situations.

Training:

Leave one out cross validation

Measures: Absolute mean error of BAL. pearson Correlation coefficient of BAL. Model Sensitivity and Specificity rates.

Results:

Single vs multiple calibration sessions

Number of trials of same interface

Efficacy of combining interfaces

Experiment 1:

Baseline: biased towards saying people are drunk

Learning curve:

Experiment 2: Single task

Choice reaction task on its own very effective.

Experiment 3: Multiple tasks:

The good performing tasks always have CR. Having more tasks may prevent users from gaming the interface(for instance, if we only have CR tasks, users may game the interface) but that is sort of weak reasoning. Why not just write the paper on the CR task? Additionally, were the tasks even randomized in order at all? Was CR always the 5th task.

Utility:

CR task on tis won achieves best results.

Not valid if user is adversarial.

Limitations:

Users under placebo may act intoxicated and get false positive. But that risk is very minor to not effectuate discussion. Measures behavioral manifestations of BAL, does not account for sleeping.

Conclusion, future work:

Comparable eror to breathlyzer

Tasks mimic common smartphone behavior.