CS 74 Project Proposal: Predicting Stress from Android Sensor Data

Nishant Kumar, Minsoo Kim
January 22, 2015

1. PROBLEM:

Stress is an indicator for a wide variety of both physical and mental health disorders. Although stress is relatively easily self-assessed and addressed, it often goes unnoticed as a natural part of everyday life. Unaddressed in this way, especially over a prolonged period of time, stress may lead to more serious mental health disorders such as depression or increase the likelihood of physical disorders such as heart attack.

An automatic stress level prediction system that can infer a user's level of stress from sensory input will be helpful in addressing the issues caused by stress. Because sensory data is gathered continuously, such a system can monitor a user's stress level with little interruption and alert/remind the user of his or her current stress level or make recommendations based on it.

We propose an application that can predict the perceived stress scale(PSS) of a user based on the data collected from his/her Android device usage.

2. METHOD:

Neural Network in MATLAB

We plan to use a multilayer neural network with backpropagation. The motivation is as follows. Logistic regression is often used in looking for correlations between factors such as physical activity, social activity and mental health. These results have been validated in multiple papers. However, we hypothesize that due to the complex and interdependent nature of brain processes, there may be some important relationships between features that are not captured in a logistic regression model. To address this issue, we will apply neural networks, which are specialized for inferring such complex nonlinear functions from data, and compare the accuracy of our Perceived Stress Scale predictions with those from a logistic regression or other similar methods.

Java & MySQL

Because our data consists of separate sets of sensory, GPS, and survey data, it makes sense to integrate these into a single efficient database. We also plan to write few Java classes that will process the data and infer higher order features from the raw sensor data, eg: infer sleep duration from sensor data such as light, audio input, phone lock etc.

3. DATA:

The data used to learn this model is from the Dartmouth StudentLife study. This is available on the CS department's website. This data was collected from student volunteers from the CS65-Smartphone programming class offered last year. Student volunteers were given android phones which had an in built app which recorded raw sensor data such as audio inputs, activity, location, light sensor data etc and uploaded to a cloud.

3.1 Data collection monitoring

Students were sent reminder e-mails if there were any gaps in the data collection or the data was not getting uploaded at all.

3.2 Incentive

The student volunteers were offered incentives such as free T-shirts, Google Nexus smartphones etc.

3.3 Privacy

The student volunteers' identities were hidden by using random ids. The call logs and sms logs were hashed.

3.4 Sensor Data

There are 10 different sensor data: physical activity, audio inferences, conversation inferences, Bluetooth scan, light sensor, GPS, phone charge, phone lock, WiFi, WiFi location. All sensor data are stored in csv files. Additionally, we are able to make inferences from the raw data to fit our model, such as sleep and co-location.

3.5 Survey data

Survey data were collected via EMA (Ecological Momentary Assement) surveys. EMAs' were administered to students via their smartphones. On average, 3-13 EMA questions were administered per day. The content of the survey was selected out of a possible 27 survey types (Stress, Sleep, Classes, etc.).

4. MILESTONE GOAL

The milestone goal consists of three parts:

1) Code completion of the neural network in MATLAB. This consists of implementing the multilayer neural network architecture and the backpropagation algorithm.
2) Code completion of the preprocessing classes in Java. This consists of processing raw sensor data such as light sensor data, GPS data, microphone, accelerometer data and the building, from raw data, of some necessary inferred features such as sleep and conversation.
3) Storage of StudentLife data in MySQL. This will make the process of pulling data based on certain criteria much faster and more robust.

5. REFERENCES

1. *Wang, Rui, Fanglin Chen, Zhenyu Chen, Tianxing Li, Gabriella Harari, Stefanie Tignor, Xia Zhou, Dror Ben-Zeev, and Andrew T. Campbell. "StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students using Smartphones." In Proceedings of the ACM Conference on Ubiquitous Computing. 2014.*

2. *S. E. Taylor, W. T. Welch, H. S. Kim, and D. K. Sherman. Cultural differences in the impact of social support on psychological and biological stress responses. Psychological Science, 18(9):831–837, 2007*

3. *N. D. Lane, M. Mohammod, M. Lin, X. Yang, H. Lu, S. Ali, A. Doryab, E. Berke, T. Choudhury, and A. Campbell. Bewell: A smartphone application to monitor, model and promote wellbeing. In Proc. of PervasiveHealth , 2011*

4. *CS65 Smartphone Programming. http://www.cs.dartmouth.edu/~campbell/cs65/cs65.html*

5. *Depression. http://www.nimh.nih.gov/health/topics/depression/index.shtml*

6. *StudentLife Dataset 2014. http://studentlife.cs.dartmouth.edu/*

7. *C. M. Aldwin.Stress, coping, and development: An integrative perspective. Guilford Press, 2007.*

8. *Eftekhar B, Mohammad K, Ardebili HE, Ghodsi M, Ketabchi E. Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. BMC Medical Informatics and Decision Making 2005;5:3. doi:10.1186/1472-6947-5-3.*