Sochiatrist – Using Conversational and Biometric Data to Predict Mood

Brown HCI Lab Research Project Report

By: Jessica Fu

INTRODUCTION

During my graduate studies at Brown, I worked with Professor Jeff Huang, Sachin Pendse, Polina Tamarina and Palak Goel in the Brown HCI Lab on a research project called Sochiatrist. We work with partner research teams at hospitals in RI and use social media, conversational and biometric data to model changes in emotion in different populations. One of the goals of the past year is to conduct our own study with Brown undergraduate students and see if we could use machine learning models to predict their mood. I worked on submitting the IRB proposal and conducting the study with the team. I also built an automated system to send and track participant EMAs and developed preliminary predictive models of participant PANAS scores.

PREVIOUS WORK

In the fields of public health and clinical psychology, researchers often want to obtain a measure of participant behavior and emotion over time. One standard approach to collecting such information is through EMAs - ecological momentary assessments [5]. Participants are asked to provide self-reports a set number of times during the day, where each prompt time is randomly chosen, for an ongoing period. The assessments used can differ for each study but participants can be asked to respond to numerous questions many times a day and for months on end. As a result, researchers often find it difficult to recruit participants and must compensate heavily to incentivize the consistent completion of these self-reports.

A promising alternative to using the traditional EMA methodology is to supplement participant self-reporting with predictions of affect measures using social media, conversational and biometric data that is available to us with the ubiquitous use of mobile phones and wearable technologies. Current research in the area only utilizes either public social media data or mobile phone metadata to ascertain some form of affect [1-4]. Considering the common problems associated with public posts: a misleading public persona and a lack of samples, we suggest turning to private conversational content.

Our research acknowledges that even though there is an untapped abundance of private user data, there is currently no clean and quick solution to extracting it. The Sochiatrist team developed a system that is not only able to extract public social media data but also conversational texts sent through mobile messaging applications. We are hoping that by combining sources of more "private" data, we will be able to develop better prediction models than previously realized.

COLLECTING PARTICIPANT DATA

We conducted an IRB approved study with Brown University undergraduate students in the fall semester where we collected messaging and biometric data from each participant and asked them to fill out EMAs for the 2-week study period. Textual data from participants was anonymized because of the private nature of the collected data. Each participant was asked to wear a Microsoft Band for the duration of the study and respond to online questionnaires when prompted through text message or email. At the end of the study period, they were asked to come in, return the Microsoft Band and allow us to extract conversational textual data from their mobile devices.

Worksheet 3.1 The Positive and Negative Affect Schedule (PANAS; Watson et al., 1988)

PANAS Questionnaire

This scale consists of a number of words that describe different feelings and emotions. Read each item and then list the number from the scale below next to each word. **Indicate to what extent you feel this way right now, that is, at the present moment *OR* indicate the extent you have felt this way over the past week (circle the instructions you followed when taking this measure)**

1	2	3	4	5
Very Slightly or Not at All	A Little	Moderately	Quite a Bit	Extremely

________ 1. Interested	________ 11. Irritable
________ 2. Distressed	________ 12. Alert
________ 3. Excited	________ 13. Ashamed
________ 4. Upset	________ 14. Inspired
________ 5. Strong	________ 15. Nervous
________ 6. Guilty	________ 16. Determined
________ 7. Scared	________ 17. Attentive
________ 8. Hostile	________ 18. Jittery
________ 9. Enthusiastic	________ 19. Active
________ 10. Proud	________ 20. Afraid

Scoring Instructions:

Positive Affect Score: Add the scores on items 1, 3, 5, 9, 10, 12, 14, 16, 17, and 19. Scores can range from 10 – 50, with higher scores representing higher levels of positive affect. Mean Scores: Momentary = 29.7 (*SD* = 7.9); Weekly = 33.3 (*SD = 7.2)*

Negative Affect Score: Add the scores on items 2, 4, 6, 7, 8, 11, 13, 15, 18, and 20. Scores can range from 10 – 50, with lower scores representing lower levels of negative affect. Mean Score: Momentary = 14.8 (*SD* = 5.4); Weekly = 17.4 (*SD* = 6.2)

Copyright © 1988 by the American Psychological Association. Reproduced with permission. The official citation that should be used in referencing this material is Watson, D., Clark, L. A., & Tellegan, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54(6), 1063–1070.

Figure 1. PANAS questionnaire administered as EMA

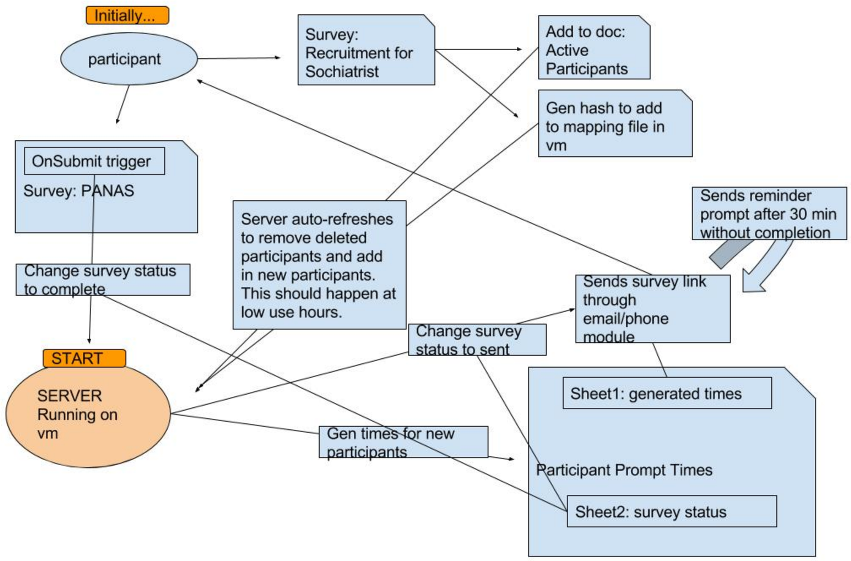

Figure 2. Survey prompting service high-level design

We administered the PANAS as the participant EMA for this study. The PANAS (Positive and Negative Affect Schedule) is a questionnaire that accurately scores a person's mood by providing a positive and negative affect score (see figure 1) [6]. Participants received prompts through text messages or emails randomly, three times a day for the 2-week period. To automate this process, I developed a survey prompting service that generated random times to send the prompts and tracked participant completion of the PANAS.

Extracting Microsoft Band Data

Using sensors embedded in the Microsoft Bands, we could collect biometric data such as participants' heart rate data, skin temperature data, heart rate variability interval data, ambient light data, and acceleration data. Initially, I replicated Professor Huang's band extraction experiments (https://tinyurl.com/mrknnjw) in an effort to obtain more granular minute-by-minute data. The Band's phone app could only export a summary of the sensor data while other helper apps that were designed to listened in on requests being sent to the Microsoft servers were unreliable.

After some more research, we decided to use the Microsoft Band's public API to obtain sensor data from an account to account basis. To get this to work, participants were asked to toggle on a "run activity" on the Band throughout the day so that it would be able to continuously record granular biometric data. Although the data collected only had around a ten-minute granularity level, it was the simplest way to get something that worked no matter which kind of phone the participant had synced the Band to.

Automating Participant EMAs

I developed a survey prompting service that allowed researchers to follow the EMA methodology by prompting participants a set amount each day for a fixed amount of time. In our case, our participants were prompted at 3 random times during the day for the 2-week study period. This service allowed participants to start their EMA period retroactively as well as be removed if they chose to withdraw from the study for any reason.

I decided to use Google Sheets as the main means of data storage by interfacing with the Google Sheets API. The reason we decided to forgo more traditional database systems is because gsheets is easier to use for people with little experience with databases. Even though there exists software that interfaces with databases well, the learning curve is substantial for someone who is not familiar with database administration. By using ghseets as the "database", users can easily and clearly see participant information, survey prompt statuses, and generated random schedules for each EMA. Any changes that need to be made can be made by editing the appropriate cell, deleting/adding a row, etc, as the system retrieves updated data every night. The built-in features of gsheets mirrors many of the perks of most modern database systems, for example, retrieving revision history and exporting of data to different formats. Since Google Drive and its applications are so familiar, the intuitive interface and editing tools make it easier for anyone to pick up and use.

The survey prompting service architecture is illustrated in figure 2. For each new participant added to the list of active participants, the service generates a random schedule for the next two weeks, accounting for the participant's self-reported waking and sleeping time during the week and the weekend. The service follows the generated times to prompt each participant with their preferred method of communication (text or email). We put together communication modules that interfaced with Twilio and Google's Gmail API to send the messages. Once a participant completes the Google Form PANAS questionnaire, the onPrompt trigger set with the Google Script engine sets off a call to the service. If the service doesn't receive the completion request in 30 minutes, it sends a reminder prompt to the participant.

DATA ANALYSIS AND MACHINE LEARNING WORK

I have experimented with generating some preliminary predictive models based on the data we've collected from study participants. We also experimented with the scikit-learn library that was then used to streamline some of the machine learning code. We worked with 25 participants who have completed PANAS surveys, worn the Microsoft Band and have given us conversational data from applications such as Facebook Messenger, Instagram, WhatsApp, Kik, and Twitter. I have outlined some of the general statistics surrounding this dataset in figures 3-6.

The PANAS scores range from 10-50 because there are 10 positive moods and 10 negative moods where the lowest rating for each question is 1 and the highest is 5. We can see from the distribution of positive and negative scores (figures 3-4) that most participants felt more positive moods on average while many simultaneously reporting fewer negative moods. It's interesting to see how negative and positive scores interact with each other over time through figures 5-7. We can see that some participants (figures 5-6) see their positive and negative lines cross while others look completely disjoint (figure 7).

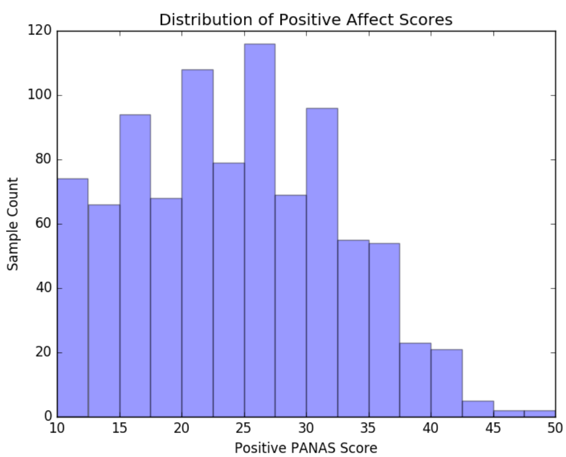

Figure 3. All positive PANAS affect scores from all participants over the 2-week study period

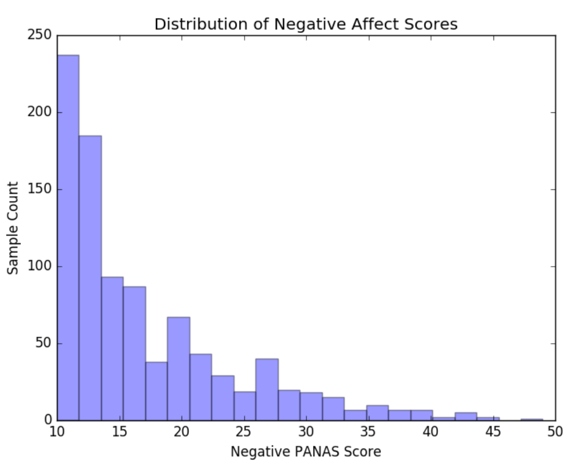

Figure 4. All negative PANAS affect scores from all participants over the 2-week study period

Featurizing Participant Data

We are currently working on featurizing conversational data and biometric data by combining natural language processing techniques such as using dictionaries like LIWC, ANEW and LabMT. We are also looking at the temporal aspects of the data and comparing sequence-based predictive models to traditionally non-temporal models.

Personalized vs Generalized Models

Currently, personalized models for each participant perform slightly better than generalized models of all participant data on average. In addition, the range of accuracies per participant model differs as well. This can be explained by the differences in expression of mood in individual participants. For example, between two participants who report the same higher than average negative affect score,

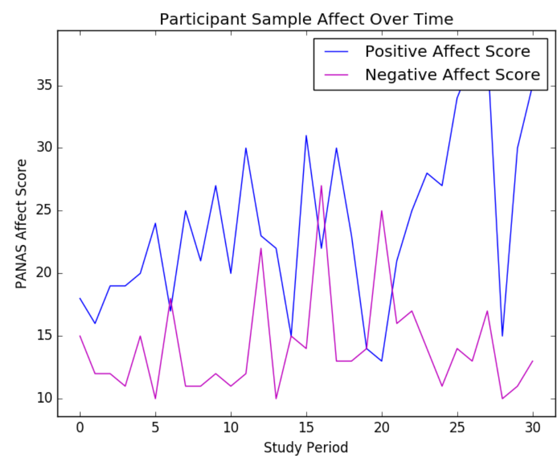

Figure 5. PANAS affect scores over the 2-week study period for sample participant A

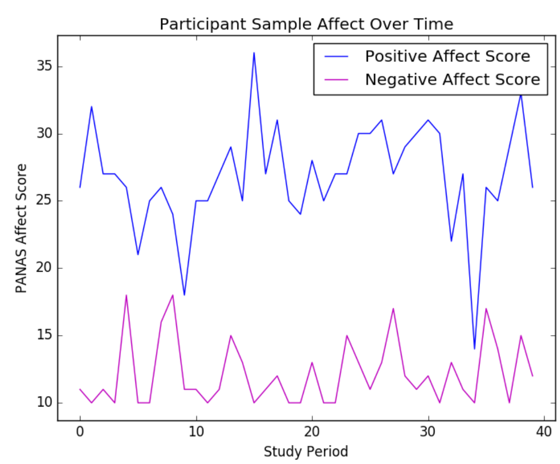

Figure 6. PANAS affect scores over the 2-week study period for sample participant B

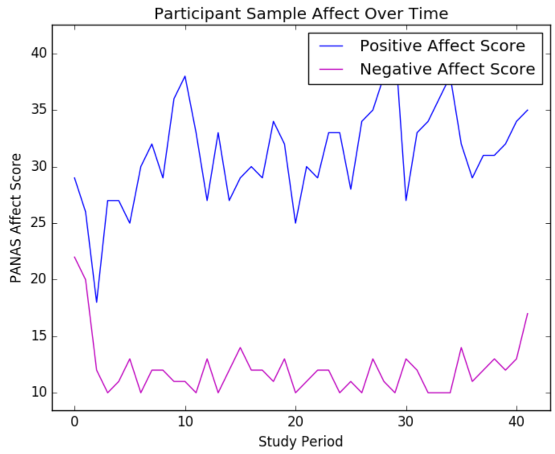

Figure 7. PANAS affect scores over the 2-week study period for sample participant C

participant A may use more negative affect words in their messages while participant B may cease to send messages altogether. Because of the lack of sample data per participant, supplementing personalized models with data from the general pool of participants also gives a noticeable boost to the accuracy of predicting mood for some participants.

*The concrete analysis of machine learning techniques and models will be in the ongoing paper on this research.

CONCLUSION AND FUTURE WORK

The continuous use of mobile devices for communication and wearables with sophisticated biometric sensors provide the unique opportunity to leverage the user data generated to predict changes in mood and quite possibly provide correctly timed intervention in at-risk populations. As illustrated in this report, our collection of participant information requires participants to hand us their devices for data extraction. This provides its own set of limitations: the complexity of the extraction software for the general population and the infeasibility of live data collection. The live streaming of personal messaging data can allow researchers to develop methods to regulate mood or provide intervention on the fly. However, most messaging applications are proprietary and privacy issues push back against opening any doors for messaging data to be accessible to 3rd party software.

I will continue to develop the predictive models for mood and apply our models onto datasets for different target populations from our research partners. The Sochiatrist team and I will also be working on submitting a research paper on our findings in the near future.

ACKNOWLEDGMENTS

This work was done in collaboration with the Sochiatrist team: Sachin Pendse, Polina Tamarina and Palak Goel, and Professor Jeff Huang. Special thanks to all the participants of our study and our research partners and sponsors for all the hard work and dedication put into this impactful research.

REFERENCES

1. Asselbergs, J., Ruwaard, J., Ejdys, M., Schrader, N., Sijbrandij, M. and Riper, H. Mobile Phone-Based Unobtrusive Ecological Momentary Assessment of Day-to-Day Mood: An Explorative Study. *Journal of Medical Internet Research 18*, 3 (2016), e72.
2. De Choudhury, M., Gamon, M., Counts, S. and Horvitz, E. Predicting Depression via Social Media. AAAI (2013).
3. Reece, A., Reagan, A., Lix, K., Dodds, P., Danforth, C. and Langer, E. Forecasting the onset and course of mental illness with Twitter data. *Scientific Reports 7*, 1 (2017).
4. Saha, K., Chan, L., De Barbaro, K., Abowd, G. and De Choudhury, M. Inferring Mood Instability on Social Media by Leveraging Ecological Momentary Assessments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 1*, 3 (2017), 1-27.
5. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 4 (2008), 1–32.
6. Watson, D., Clark, L. and Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology 54*, 6 (1988), 1063-1070.