Duke Master of Interdisciplinary Data Science (MIDS) Capstone Project
Project Description

Project Title: Decoding hidden information in activity-tracking data: Using neural, behavioral, and physiological data to identify cognitive and emotional states

Team Composition and Involvement

This is an IGE Capstone, thus the student team will include MIDS students and Mikella Green, a Duke PhD student in Psychology and Neuroscience (the IGE Fellow). All student team members will be expected to contribute equal magnitudes of time and take on equivalent levels of responsibility.

Faculty mentors for the project will include Gregory Samanez-Larkin, Assistant Professor of Psychology and Neuroscience.

The outside partner for this project will be Fitbit, pending legal approval. At present, Fitbit is expected to give advice and feedback, but not proprietary data (beyond data that was collected by Gregory Samanez-Larkin's group through IRB-approved studies).

Background

As the popularity and capability of wearable activity trackers has increased over the last 20 years, so has an interest in what their internal sensors can tell us about our overall health. In particular, there is much interest in whether activity trackers can be used to decode emotional states. If so, interventions could potentially be created to help people when their activity trackers suggest they are feeling depressed or anxious. This project aims to explore whether wearable activity tracker data can be used to classify cognitive and emotional states through leveraging relationships with brain imaging and behavioral data.

Neural responses have been used as "biomarkers" for emotional states in the past. For example, multi-voxel pattern analysis of functional brain activity can successfully predict what type of emotional stimulus a person saw or heard during an experiment (Kragel & Labar, 2015). Physiological responses can be used as "biomarkers" for emotional states as well. Multivariate pattern classification of electrodermal, cardiac, respiratory, and gastric activity predicts what emotion a person reports feeling (Kragel & Labar, 2013). While informative, both neural data and physiological data are costly and time-intensive to collect. Activity tracker data, on the other hand, is non-invasive and easily collected while people live their daily lives. Ideally, activity tracker data could be used as a biomarker for emotional states the same way neural and physiological data has been used in the past.

Approach

The educational team will design a modeling strategy for using activity tracker data to predict negative emotional states. To do this, they will use data sets that have Fitbit data (which yields numbers of steps, heart rate, and sleep measures), neuroimaging data, behavioral data, neuropsychological data, and experience sampling data from the same participants. They will design models to predict emotional states that use only variables that can be collected from people outside of the lab. However, the weights and parameters of these models will be trained using the neuroimaging data and neuropsychological data, which has been shown

previously to correlate with self-reported emotional states. Once a successful model is generated and thoroughly tested, the team will design an app that can integrate with Fitbit trackers to identify when a wearer is anxious. The goal of the app will be to offer personalized strategies for relieving anxiety at exactly the moment the wearer needs them.

Deliverables

- Documentation of data cleaning steps and modeling details
- Model that successfully predicts emotional states from Fitbit data
- Characterization of when the model performs successfully or unsuccessfully
- Commented code to implement modeling and any related data visualizations
- Slide presentation describing the project to a non-technical audience

Description of Protected Data Set and other Data

The dataset is comprised of physical activity data, neuroimaging data, behavioral and neuropsychological measure data, and experience sampling data. All of the data in the dataset were collected from previous or on-going studies conducted under the guidance of Dr. Samanez-Larkin. The data has been deidentified and contains no personal identifiers. Most of the previously collected datasets are available publicly online. Data not available online is stored on server space provided by the Duke University Brain Imaging and Analysis Center. All participants in the studies have consented to the use of their deidentified data for research and academic purposes.

Activity tracking data. The dataset currently contains activity monitoring data, measured with a FitBit Charge, on approximately 110 participants. This data contains steps, sleep, and heart rate. On an additional 67 participants we have daily step data recorded on a pedometer. All physical activity data is stored in a tabular format.

Experience Sampling. We have conducted 4 previous studies using experience sampling. Experience sampling involves asking participants to answer a survey about their thoughts, feelings, environment and/or desires on multiple occasions across a time period. All experience sampling data is stored in a tabular format.

Behavioral tasks and neuropsychological measures. Previously collected task measures include decision-making and reinforcement learning tasks. We also have data from neuropsychological batteries that measure various cognitive abilities and working memory. Additionally, we have self-report data measuring affect, risk-taking, depression, and personal financial information (e.g., investments, credit usage).

Neuroimaging data. Pre-existing brain imaging data contains diffusion-weighted images, structural MRI, and functional MRI. All imaging data has been pre-processed and is ready for analysis.

Future data collection. Two new studies are currently underway in Dr. Samanez-Larkin's lab. Both studies are collecting activity tracker data and neuroimaging data.

Allowable Reference to Project/Data Provider

The members of the capstone team will be allowed to reference the project and data provider and the project in job interviews, resumes, or other contexts during or subsequent to their participation in this project.

Data Hosting and Computational Environment

Data will be stored and analyzed in a compute environment approved by the relevant IRB and information security office.

Publication

If a publication is desired from the work performed in this project, the publication should be pursued according to the most recent version of the "Guidelines for Authorship and Authorship Dispute Resolution" in the Duke Faculty Handbook. The IGE Fellow's doctoral advisor should generally be considered the "Lead Author" who serves as the corresponding author and takes responsibility for the integrity of the work as a whole, unless the IGE Fellow's doctoral advisor delegates that responsibility to another party agreed upon by the Capstone team (for example, if the IGE Fellow's doctoral advisor is not going to be an author on the paper, the "Lead Author" could be the IGE Fellow or another faculty member who made a significant intellectual contribution to the work). It is the Lead Author's responsibility to ensure that questions of authorship are discussed and clarified at the earliest practical stage of a research project, and to anticipate possible disagreements concerning authorship, especially if authors' relative contributions change over the course of the work that will be included in the publication.

Since an IGE Fellow and MIDS students are equal partners on Capstone projects, if an IGE Fellow meets the criteria described in the Faculty Handbook for being included as an author on a publication, it is probable that the participating MIDS students will meet the criteria as well. Thus all students involved in the Capstone team should be informed if a publication related to the project work is desired, and should be given the opportunity to participate in the drafting, reviewing, and/or revision of related publications, and should also be given the opportunity to approve the final version of related publications. As a general guideline, if a member of the Capstone team chooses not to participate as an author in a publication related to Capstone project work, the authors of the publication should still send the member a copy of the publication at least 30 days prior to publication in order to allow the member an opportunity to protect proprietary information or intellectual property that might be disclosed. The authors should also notify the MIDS Capstone Director about which members of the Capstone team have chosen not to participate in the publication.

If a publication conflict arises that involves a MIDS student, a resolution should first be attempted by the Capstone team and co-authors of the paper. If desired, the MIDS Director may be asked to consult with the team to help mediate a resolution. If a resolution cannot be reached, the matter can be referred to the Duke Authorship Dispute Board according to the guidelines in the Duke Faculty Handbook.

If and when a non-Duke partner becomes involved in the Capstone project, this publication policy will be replaced by the legal agreement made between Duke and the non-Duke partner.

Work Plan Expectations

The project described above must receive appropriate IRB approval and have all team members added to the approved IRB protocol before any work is done, and will be pursued in stages. Study data that has already been collected will be analyzed in Stage 1. The app that uses analyses of these data to offer personalized strategies for relieving anxiety will be analyzed in Stage 2. Stage 1 will be approved through DUHS IRB Pro00103312, which has already been submitted. Stage 2 requires a separate protocol submission that must go through full review. This protocol should be submitted as soon as possible so that work is not delayed.

Legal agreements with Fitbit will be pursued in parallel with iRB submissions. The following contacts are the ones MIDS has on file:

Sylvester Mbonu (smbonu@fitbit.com)
Nova Hammerquist (nhammerquist@fitbit.com)