

Project Title

Using Mobile Data to Manage Parkinson's Disease

Business Understanding

Parkinson's Disease is a progressive neurological disorder affecting at least one million Americans, with 60,000 new cases diagnosed in the US every year - and there are strong reasons to believe that the majority of Parkinson's sufferers go undiagnosed. Nonetheless even at the current rate of diagnosis, Parkinson's Disease is estimated to cost the US economy over 25 billion dollars annually. While Parkinson's is not considered a terminal disease, it is incurable and progressive and so the identification of the most effective therapies for each patient would have an immediate effect on quality of life.

Parkinson's disease is characterised by three principal symptoms - Bradykinesia or slowness of movement, rigidity of the body evident in many ways including reduction in facial expression, and resting tremor. Tremor is present in about 75% of patients and is often the first symptom leading to diagnosis, but it is significant in a number of other ways; the tremor-dominant Parkinson's subtype is a milder disorder than the non-tremor subtype and many other minor differences between subtypes have been noted; also in cases where symptoms occur predominantly on one side of the body, tremor is sometimes found on the opposite side to the other symptoms.

The focus of this project will be a detailed characterization of patient tremor and the extent to which tremor characteristics are related to, and predictive of, severity, symptoms, and treatment response. The goal of the project is to determine the extent to which a smartphone accelerometer is an effective tool for the collection of tremor data, and to investigate the extent to which medication can control tremor in a patient with Parkinson's Disease.

References:

- Parkinson's Disease Foundation: Parkinson's Statistics
http://www.pdff.org/en/parkinson_statistics
- Cerebral causes and consequences of parkinsonian resting tremor: a tale of two circuits? <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3501966/>
- Parkinson's disease tremor: pathophysiology
<https://www.ncbi.nlm.nih.gov/pubmed/22166464>

Data Understanding

In March 2015, Apple launched ResearchKit, a toolkit for the rapid construction of research-oriented mobile applications, and local Seattle non-profit Sage Bionetworks used this

technology to create a range of applications intended to further healthcare research. One of the first was the Parkinson's Study M-Power App (<https://itunes.apple.com/us/app/parkinson-mpower-study-app/id972191200?mt=8>) which tracks the health of patients through questionnaires and the collection of data using phone sensors and an optional wearable device. The data are then made available for authorized research purposes via the Synapse data portal, with copious metadata and documentation available in a public area: (<https://www.synapse.org/#!/Synapse:syn4993293/wiki/247859>). Currently, three patient surveys are available along with four Activities - a set of Activity data comprises the output of a range of phone sensors while the patient conducts a specified set of tasks.

I will be assisted in my understanding of the data by the copious documentation available through the Synapse mPower Portal, as well as through the kind assistance of Dr. Larsson Omberg, Chief Data Scientist at Sage Bionetworks, who has offered me the opportunity to sit at Sage if needed during the project. I am also grateful to Dr. J. Witt, head of the Movement Disorders clinic at Swedish Neurology Institute for her agreement to provide assistance in better understanding the clinical aspects of Parkinson's Disease.

The data I will be using are not public and I will be the only authorized user; while I may share summarized data and results as a part of my project, I will be the sole accessor of the underlying dataset.

Data Preparation

In the case of resting tremor, the data have not yet been cleaned and documented for public release, but consist of 19,000 users (approximately 1200 with a diagnosis of Parkinson's Disease, the rest controls) performing a set of six movements corresponding to a standard clinical tremor test, with the phone accelerometer sampling tremor at 100 Hz (tremor is typically in the 3-5 Hz range). Approximately 500,000 tests have been recorded, with approximately 15,000 related to tremor.

A significant aspect of the project will be the preliminary analysis and characterization of data, although some code will be available to assist me from a previous student project. Data will need to be de-noised and the time-series analyzed, and research will need to be done to determine the state of the art for the features used to decompose, recognize and characterize movements from accelerometer data. Preliminary analysis will be followed by data cleaning, normalization if required, and feature extraction. One project deliverable will be sufficient information about the data to generate a page similar to that available for the walking Activity: <https://www.synapse.org/#!/Synapse:syn4993293/wiki/376006>.

Due to issues of privacy and patient confidentiality, this data is available to authorized users only through the Synapse data platform (www.synapse.org). Currently I am registered as a user of Synapse and have completed the preliminary stages of validation and certification - this

proposal will form the basis of my Statement of Use, required to document access and usage of the data.

Modeling

The intent is to classify patients using tremor characteristics using a model which clearly conveys the contributions of each factor to the final prediction. My initial approach would therefore be logistic regression. A decision on the factors to model will require closer evaluation of the data, but as currently envisaged this project will seek to identify features in the mobile data that are predictive of the nature and severity of the disease, in particular response to medication. The data do contain information on medications and whether a patient has a confirmed diagnosis. A first step will be to a minimal analysis pipeline from raw data to prediction, and to determine whether all six movements can be distinguished. This base model will then be elaborated to include other factors such as medication, with a goal of predicting whether a task was performed by a participant with Parkinson's or a control, and whether the participant was medicated or not.

Evaluation

While collection of mPower data is ongoing, the available data is a six-month snapshot and so effectively static; it is however quite large, and so I should be able to use cross-validation.

Deployment

Availability of the mPower data is restricted, and while I hope this project finds an association between tremor and response to therapy, it would not be advisable to make any public release of a predictive tool. Deployment will therefore be restricted to the fitted model.