Using Smartphones for Behavior Study of Individuals with Parkinson’s Disease

NYIT Research Experience for Undergraduates (REU) May 26 – July 30, 2015

Abstract
The purpose of this study is to use smartphone sensors in pocket usage in order to identify changes in Parkinson’s disease (PD) patients’ actions at different times of the day. We collected data from control subjects to create a model to classify actions in an unlabeled data set from the Michael J. Fox Foundation (MUFF), and then compare morning and afternoon sequences. Action classification with the model on our own data achieved maximum of 97.9% accuracy with 3 action labels and 88% with 5 labels, using random forest classifier. The preliminary results show a concentration of activities towards the mean value in the late-day from the PD data set. However, more data is desired for further validation of the machine learning algorithms.

Background
Parkinson’s disease (PD) is a neurodegenerative disease that affects movement, such that actions become slow, trembling, and in severe cases, freezing completely [1]. Currently, it is diagnosed by having the patient meet face-to-face with medical experts, but recent studies have utilized motion sensors (dedicated or on smartphones) to identify features of the disease, typically by securing the sensors to areas like the waist or legs [2][3][7]. Because PD patients were not available for this study, MUFF data from a 2013 smartphone challenge, which includes 9 PD subjects and 7 control subjects, were used even though it did not contain ground truth [4]. A few papers that use this data were able to achieve 70-90% action classification or PD identification, though such studies did not study changes in behavior of individuals with PD during the course of a day or longer [5][8].

Approach
The study is composed of a series of steps:
1. Collect our own smartphone data and manually label walking, sitting, and standing actions (Fig. B, Fig. C) for accelerometer and gyroscope.
2. Use various classifiers or cluster analysis on the collected data to identify actions in the unlabeled MUFF data (Fig. A).
3. Compare the behavior of PD patients between morning (before 11 am) and afternoon (after 1 pm) with symbolic aggregate approximation (SAX), dynamic time warping (DTW), and symbol frequency count (Fig. D, Fig. E.)

Experimental Results
Table 1 and Table 2. show action classification results using 4 different classifiers in a machine learning software called Weka. The collected control subject data was first labeled with the 3 main actions (sit, stand, walk) and applied to itself to achieve a maximum of 97.9% accuracy with random forest classifier and 10-fold cross-validation. When labels were expanded to include transition actions (sit to stand, stand to sit), as expected, the accuracy decreased to 88% because transition actions are difficult to distinguish. However, the MJFF data was not able to be identified by actions. Figure C and Figure D. show samples of symbol frequency count for two individuals, in which time series data is partitioned and changed into symbols, and the fraction of each symbol appearing is calculated. Originally this was to be done with action labels to assess the extent that certain actions appeared, but without labeled data, this could not be done. For some instances in the small sample size, control subjects were more consistent and PD patients sometimes appeared to have symbols concentrated at the mean or other places the afternoon (less movement or smaller movement), but more distributed in mornings.

Discussion
The experimental results exhibited maximum of 97.9% accuracy in action classification with 3 actions (sit, stand, walk) and 88% with 5 actions between sitting and standing (induced), using the random forest algorithm. However, even though classifying actions was demonstrated to be a way to classify the MUFF data, since we could not collect our own PD data with ground truth, the model created with control subjects was not suitable enough to classify the MJFF data. Cluster analysis provided a way to separate parts of the MJFF into similar groups. When similar clusters for one individual were compared at different times of the day for similarity with SAX and DTW, these data have thus far provided no significant results. Clusters may signify certain parts of the data appear to do the same general actions, but perhaps not necessarily accurately or correspond to the specific actions we are looking for. SAX and DTW also seem more useful for instances of comparing series that are less dependent on the specific actions that may just happen to occur at certain times, and so work better for long-term trends or well-matched actions. The main limitations of this study were small sample size and lack of real PD patients for creating the model. Some of the MJFF data also lacked long, continuous data ranging from morning to evening, which makes meaningful comparisons difficult.

Future Work
Since recording our own PD patient data with ground truth was not achieved, it would be useful to do so and have a more direct comparison with the knowledge of actions. This way, SAX and DTW could be used for small-scale comparison (changes in a certain action), and frequency count (by symbol or labeled actions) could be done for long-scale comparison for changing behavior. With more data, even if not recorded by this study, clustering and action classification would be more accurate as well. Furthermore, long-scale data spanning months would be useful in observing not just small cycles in actions, but also the progression of symptoms. In addition to more data, future endeavors would include involving security in the implementation of this study, such as limiting access only to patient and medical professionals.

Acknowledgement
The project is funded by National Science Foundation Grant number CNS-1263283 and New York Institute of Technology.

References