School of
Engineering and Computing Sciences

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

REU fellows: Bruno Correa[1] and Yang Peng[2], Faculty mentor: Dr. Huanying Gu[3],
Affiliation: 1. University of West Florida, 2. Cornell University, 3. School of Engineering and Computing Sciences, NYIT

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 (MJFF), 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, tremoring, 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, MJFF 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][6].

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 MJFF 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, like the ones shown, 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.

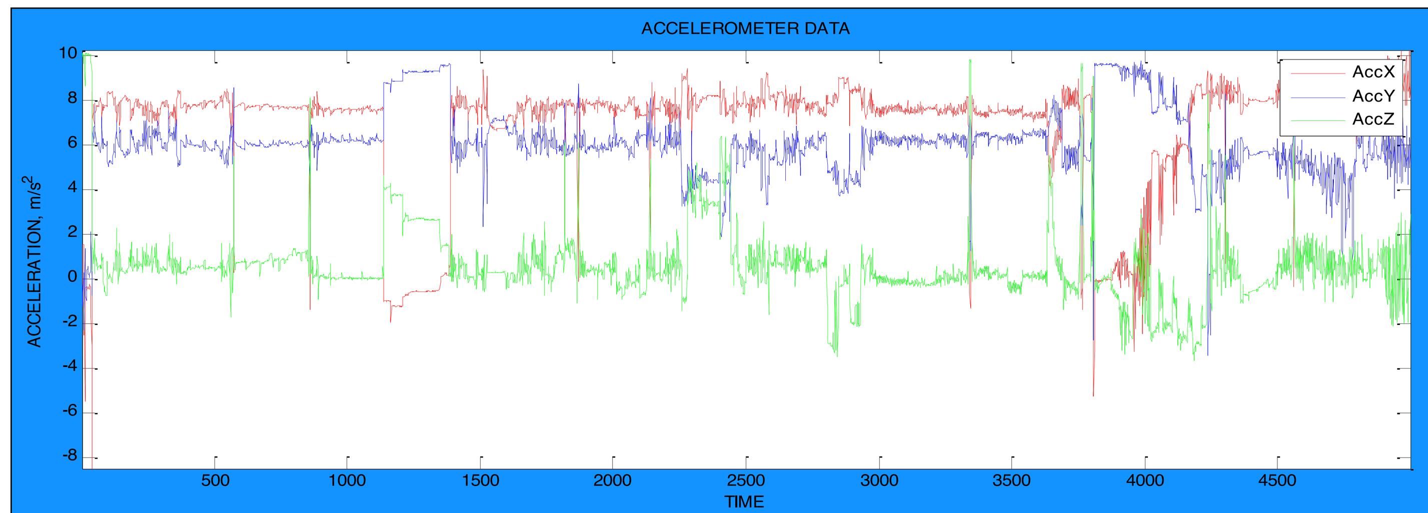

Figure A. MJFF accelerometer data recorded by phone, at 50 Hz sampling frequency, for an individual

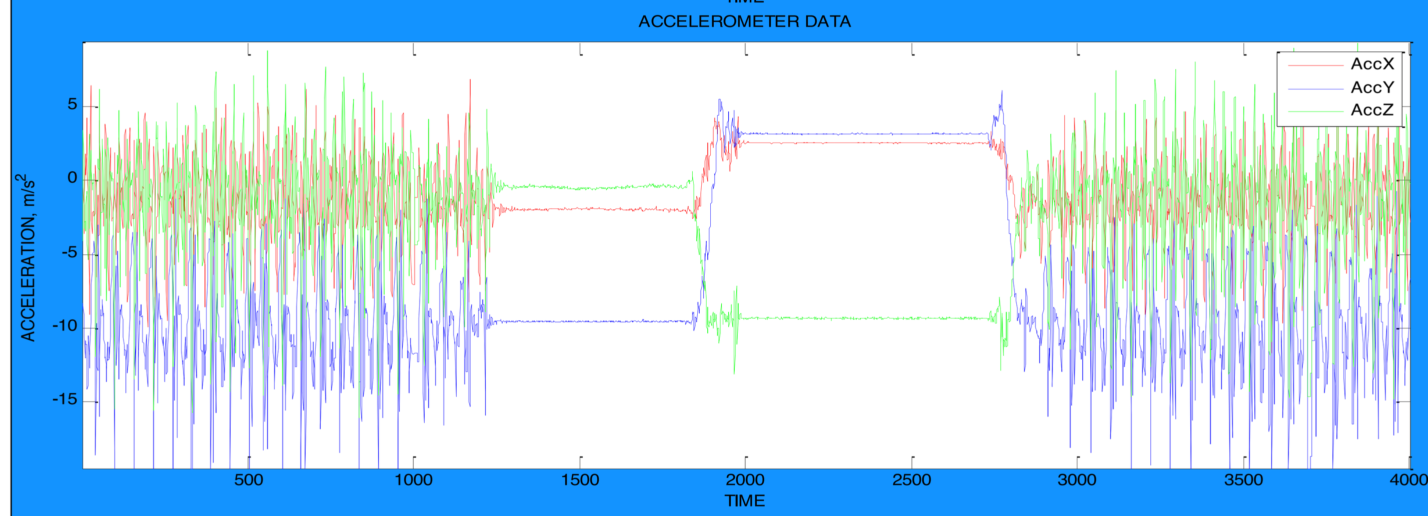

Figure B. Our own sample accelerometer data recorded by phone, at 50 Hz sampling frequency, for an individual

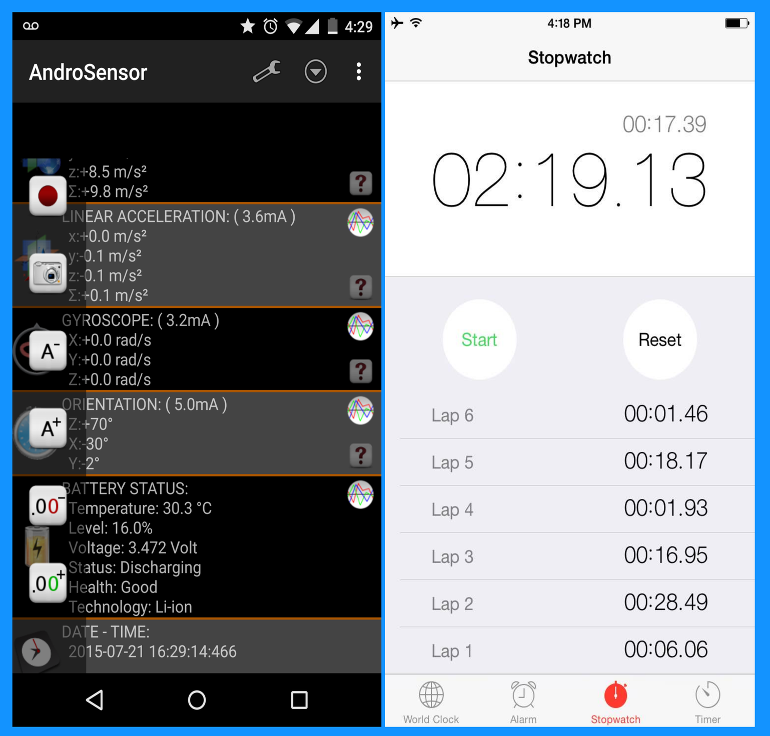

Figure C. An app called AndroSensor and digital stopwatch were used. Recording was done on a Nexus 5 placed in clothing pockets at 50 Hz sampling frequency.

Table 1. Classification for 3 actions: sit, stand, walk

	% Precision Average (Cross-Validation 10 folds)	% Precision Average (Percentage Split 30%)
Random Forest	97.9	97.5
Decision Tree (J48)	95	94
Naïve Bayes	79.4	88.9
Decision Table	93.6	91.1

Table 2. Classification for 5 actions: sit, stand, walk, stand-to-sit, sit-to-stand (transition actions)

	% Precision Average (Cross-Validation 10 folds)	% Precision Average (Percentage Split 30%)
Random Forest	88	85.9
Decision Tree (J48)	86.8	82.2
Naïve Bayes	81.9	82.9
Decision Table	82.6	77.6

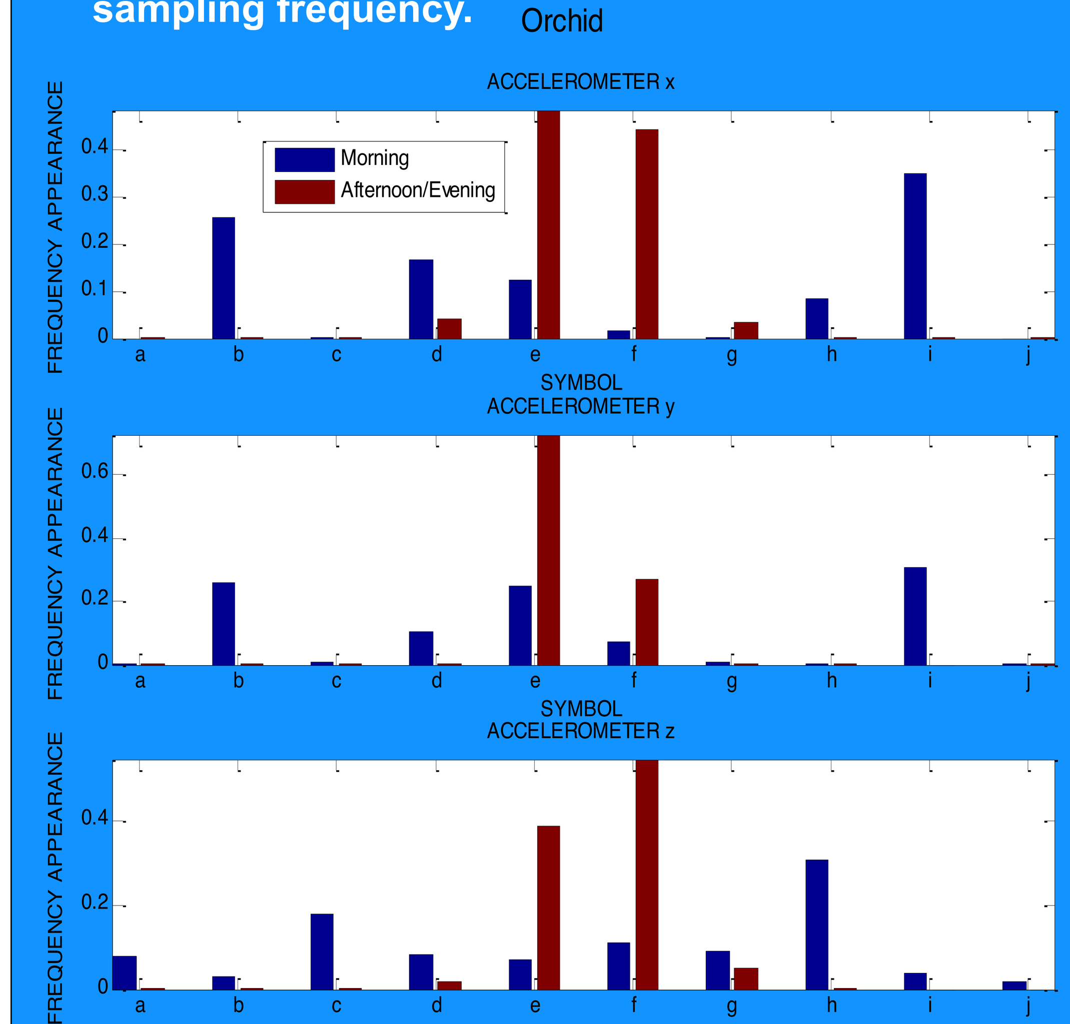

Figure D. Symbol frequency in a PD subject. It is sparse and more variable in time.

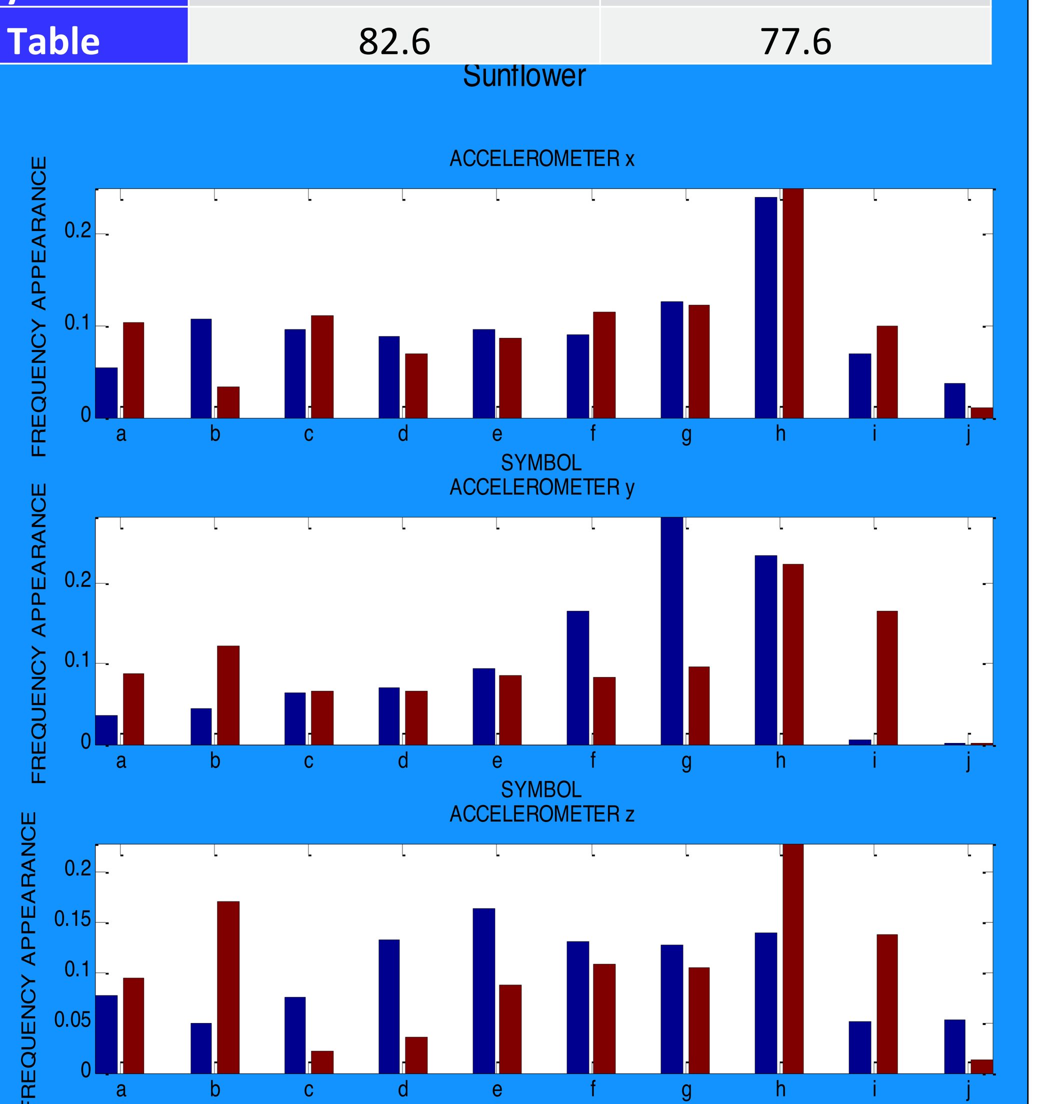

Figure E. Symbol frequency in a control subject. It seems more consistent throughout the day.

Discussion

The experimental results exhibited maximum of 97.9% accuracy in action classification with 3 actions (sit, stand, walk) and 88% with 5 actions (transitions between sitting and standing included), using the random forest algorithm. However, even though classifying actions was demonstrated to be a way to classify the MJFF 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 signifying 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 by 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, longer-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 to only patient and medical professionals.

Acknowledgement

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

References

[1] Parkinson's Disease Foundation (2015). *What is Parkinson's Disease?* [Online]. Available: http://www.pdf.org
[2] Steven T. Moore; MacDougall, Hamish; Ondo, William, "Ambulatory monitoring of freezing of gait in Parkinson's disease," *Journal of Neuroscience Methods*, Volume 167, Issue 2, 30 January 2008, Pages 340-348, ISSN 0165-0270
[3] Hundza, S.R.; Hook, W.R.; Harris, C.R.; Mahajan, S.V.; Leslie, P.A.; Spani, C.A.; Spalteholz, L.G.; Birch, B.J.; Commandeur, D.T.; Livingston, N.J., "Accurate and Reliable Gait Cycle Detection in Parkinson's Disease," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* , vol.22, no.1, pp.127,137, Jan. 2014
[4] Michael J. Fox Foundation, "Predicting Parkinson's Disease Progression with Smarphone Data," *Access Data and Biospecimens*, 2013. Dataset. https://www.kaggle.com/c/predicting-parkinson-s-disease-progression-with-smartphone-data/data
[5] C. Boussios, Greenbaum, J., Ieong, B., Kokkotos, F., Kokkotos, S., Zalesak, M., "The Construction of a Novel Statistical Algorithm to Objectively Diagnose Parkinson's Disease Use Smartphone Data," Michael J. Fox Foundation: Predicting Parkinson's Disease Progression with Smartphone Data, Unpublished.
[6] P. Anantharam, Tirunarayan, K., Taslimi, V., Sheth, A., "Predicting Parkinson's Disease Progression with Smartphone Data," Michael J. Fox Foundation: Predicting Parkinson's Disease Progression with Smartphone Data. Unpublished.
[7] Kun-Chan Lan, Wen-Yuah Shih, "Early Diagnosis of Parkinson's Disease Using a Smartphone," *Procedia Computer Science*, Volume 34, 2014, Pages 305-312, ISSN 1877-0509