Measuring the symptoms of Parkinson's Disease with a Smartphone

The Michael J. Fox Foundation for Parkinson's research proposed as a data challenge on Kaggle to develop and propose efficient algorithms to detect and m quantify symptoms of Parkinson disease from data monitored in a smartphone. We think that this challenge present both a genuine interest and present major clinical applications. First, the theoretical problem of detecting Parkinson's symptom is challenging. On a clinical perspective, the tracking of theses symptoms with a simple device, such has a smartphone, in one individual presents a real clinical interest. Not only such algorithm can be used to track the evolution of the disease in time for one patient, but can also be used to maximize the effect of medication for this single patient.

The available data consisted of acceleration, audio, battery, compass and GPS information collected from an Android smartphone place in the subject's pocket or worn around the neck. Though the data is large the most important data is the accelerometer data.

For the capstone project we will focus on the algorithm presented by the two leading teams trying to use commonly available python tool to solve the problem. The winner team (the LIONgroup)[1] was using the LION software (available online free of charge for non-profit user)[2]. The second group[3] (Michelle Wang a single graduate student from MIT) used a toolbox in Matlab for Support Vector Machine Learning (SVM)[4]. Matlab is a really expansive software for industry and therefore should not be used if a python equivalent exist. Sklearn is a free and widely use python library seems to propose a free alternative for SVM. Therefore, we will use it here. We therefore propose to perform similar analysis as what they did, but using python.

The data cleaning here seems to be a key issue for both group and the strategy they proposed is different. The Lion group isolates the time windows where the patient was at rest using available GPS and compass data. Michelle Wang uses a strategy I prefer. She reduced the three variables to one by taking the root mean square. She also reduced the granularity by average over hourly window rather than the 2sec proposed. This might be big and may explain that her success rate with SVM algorithm was a bit lower than for the other group. Both group than filter the data on different frequency values. Here, I propose to use the approach of Wang, but also

---

[1] https://www.michaeljfox.org/foundation/news-detail.php?winning-data-

[2] lionso.org

[3] https://www.kaggle.com/c/predicting-parkinson-s-disease-progression-with-smartphone-data/visualization/1143

[4] http://www.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html?searchHighlight=support%20vector%20machine%20learning

test the reliability of different time windows. The LION group also proposed a linear regression model[5] to track the strength of the symptom and compare them to severity assessment questionnaire. Here, we will want to reproduce it using the linear regression model tool available in sklearn.

What we propose to do is to use the data cleaning strategy proposed by Michelle Wang and than with python base SKlearn library doing both a SVM analysis to separate the patient from the control group and a linear regression model to try to quantify the severity of the disease. We will deliver python code and a report on our finding.

[5] http://www.mathworks.com/help/stats/what-is-linear-regression.html