Carnegie Mellon University

THE ROBOTICS INSTITUTE

THESIS DEFENSE

Personalized and weakly supervised learning for Parkinson's disease symptom detection

Monday, October 21, 2019

8102 Gates Hillman Center

12:30 p.m.

Thesis Committee:

Fernando De la Torre
Co-chair

Jessica Hodgins
Co-chair

Artur Dubrawski

Anthony Jarc
Intuitive Surgical Inc.

Ada Zhang

Abstract

Parkinson's Disease (PD) is a neurodegenerative disorder that affects approximately one million Americans. Medications exist to manage the symptoms, but doctors must periodically adjust dosage level and frequency as a patient's disease progresses. These adjustments are typically based on observations made during short clinic visits, which provide an incomplete picture of a patient's daily quality of life. To provide a more accurate assessment of a patient's at-home experience, this thesis explores the use of wearable accelerometers for monitoring PD in the wild. A central challenge of this application is human variability, which makes it difficult for algorithms to generalize to "unseen" subjects who were not in the training set. This thesis investigates two major approaches to address this issue: (1) stratified algorithms for learning from in-the-wild data, and (2) personalization algorithms to actively adjust for human variability.

A major challenge with in-the-wild data collection is obtaining labels of tremor symptoms. We developed a data collection protocol in which subjects used a cellphone app to log the *approximate* ([0-33%], [33-66%], or [66-100%]) amount of tremor experienced within short time segments and collected in-the-wild data from six PD patients. We call these approximate percentages *stratified* labels, and we present a novel stratified weakly supervised learning technique that benefits from labels with this structure. Experiments demonstrated that this new technique leads to higher performance on this dataset compared to standard weakly supervised learning algorithms. Furthermore, stratified weakly supervised algorithms performed better on in-the-wild data than fully supervised algorithms trained on accurately labeled laboratory data. These findings suggest that collecting stratified labels from each subject in the wild is more effective than collecting accurate labels from other subjects in the laboratory. In summary, stratified learning provides a practical and efficient method for collecting weak labels in-the-wild that greatly improves accuracy of person-specific PD tremor detection.

This thesis also explores the use of several unsupervised personalization algorithms, which either reweight the training data to align its distribution with the test data or learn a regression from distributions to classifiers. We extended these personalization algorithms to the multiclass scenario. In our dataset, the personalization algorithms demonstrated modest improvement over non-personalized algorithms. We further analyzed the behavior of these algorithms with synthetic data, gaining insights into their limitations when learning appropriate weights. We addressed this issue by developing new constraints to the cost function, which improved performance on synthetic data, and experimented with global optimization.