



Application number/Title: 32961: Learning a Lexicon of Human Movements: Automated, data-driven detection of elementary movement units for objective, accurate and interpretable analysis of physical activities and human behavior.

Applicant PI: Professor Thomas Ploetz

Application Institution: Georgia Institute of Technology, USA

Keywords provided by the Applicant PI to describe the research project:

activities, behavioural-phenotype, machine-learning

Application Lay Summary:

We develop methods that enable automated, quantitative assessments of behavioral phenotypes of conditions such as Parkinson's, Dementia, or Autism. Assessments are based on the analysis of movement data recorded using wearable sensing platforms. The main research question of our fundamental machine learning research focuses on how effective analysis methods can be derived from large scale unlabeled movement datasets. We aim at automatically identifying elementary building blocks of human movements. Once such a lexicon of human movements has been learned we are able to develop effective recognition and assessment systems for fine grained, and accurate clinical analysis of human behaviors. A multitude of health conditions are characterized by their behavioral phenotypes. Being able to automatically assess human behavior has the potential for improved diagnosis as well as care and treatment through objective and meaningful outcome measures derived from automated behavior assessments. Automation as enabled through our research thereby promises large scale, possibly population-scale uptake as well as overcoming subjective and potentially biased clinical behavior assessments that not necessarily represent real-world conditions. As such our research has the potential for improving the health of future generations and thus aligns directly with the UK Biobank's stated purpose. We will use the UK Biobank accelerometry dataset for automatically finding generic representations of human behavior from wearable movement sensors. We will develop machine learning algorithms that allow us to analyze what is considered a representative sample of human movements. Our algorithms will extract/learn sets of building blocks that form a lexicon of human movements with which subsequently any

behavior can be modeled. Only raw accelerometer data but no ground truth annotation will be used. The resulting generic representation of movements will then be validated on existing, annotated behavior datasets from assessment domains such as Parkinson's and Autism. We are interested in using as much raw movement data as possible with the greatest variability of participants who wore the recording devices. The reason for this is that we aim at extracting generic representations of human movements that will be used for behavior modeling and assessment. As such we plan to use the full cohort included in the UK Biobank physical activity measurement dataset. Having access to generic movement and physical activity data of a large sample of more than 100k participants will facilitate model robustness through automatically learning from what is considered a representative sample of human movements.