Human activities classification Project Proposal

CMPE255 - Data Mining Prof David Anastasiu San Jose State University

Team: The Revivalists Akshay Jaiswal (013774137)

Ayaz Khan (013749710) Vishwanath Manvi (013776204)

Project Description:  
Fitness as a lifestyle is trending and gaining increasing importance with the advent of fitness trackers including smart watches. [Fitness as a lifestyle is trending. Its not only fashionable to be fit but people are also embracing the health benefits that comes with leading an active lifestyle. Chronic conditions like obesity, diabetes, blood pressure, cholesterol are on the rise, and a sendentarty lifestyle makes it that much harder for people to effectively combat these conditions. The availability of devices like fitness trackers and smart watches is making it easier than ever for people to monitor and set fitness and wellness goals for themselves. Many of us lead a sedentary lifestlye owning to the dictates of modern working environment and these devices give us the ability to take charge of our wellbeing into our own hands]. It’s extremely relevant in current times given the sedentary lifestyles of most of the urban population and the direct correlation of a person’s health of a person given chronic diseases like obesity, diabetes, blood pressure, cholesterol, etc are on the rise. [However the fitness devices are still evolving and one of the main challenges is the ability to the accurately identity and track human activities.]One of the main challenge a fitness tracker faces is be able to accurately identify and track daily various activities of a person and set appropriate fitness goals.

Our project goal is to apply various machine learning algorithms on the data generated from IMA and Heart beat sensors and be able to accurately identify a person’s activity beyond the standard categories like walking, running and resting. We plan to classify 18 categories like playing soccer, driving car, house cleaning or watching TV. Once these activities are accurately identified, people can come up with fitness and health routines that is customized for them.

Methodologies:  
The dataset contains sensor data from 9 subjects, each with 376K records with information about activity being performed, heart rate and readings from 3 separate Synchronized sensors (IMUs). As one of the steps of preprocessing, we will remove the last 3 fields of each sensors’ data which belongs to ‘Orientation’ readings. These readings are not valid in this data collection and need to be ignored. We’’ll also combine demographic and physical attributes of subjects with the sensor data. After thorough preprocessing, we will apply various classification and regression algorithms like, Logistic regression, Decision trees, Random forest, Neural networks and more, to find out the best possible classifier for the problem. We will also try to determine if the same level of results can be obtained with reduced number of features, like using readings from only hand IMU and compare the results with original results involving all 3 IMUs.

Technologies:  
Python libraries (SciPy, NumPy, Pandas, NLTK, Scrapy, Scikits-Learn, TensorFlow) and Jupyter notebook

Algorithms​:  
Logistic regression, Decision trees, Random forest, SVM, kNN, Naive Bayes, Convoluted Neural Networks

Dataset:  
3.8 million instances of 52 attributes spanning 18 activities of 9 subjects.

This includes IMA sensory data along with Heartbeat information the select demographic data of the subjects http://archive.ics.uci.edu/ml/datasets/pamap2+physical+activity+monitoring

References(s):

https://ieeexplore.ieee.org/document/8286571

page3image60242560page3image60231424