

Human Activity Recognition

CSCI-P556: Applied Machine Learning

Spring 2019

Bivas Maiti (bmaiti)
Virendra Wali (vwali)
Darshan Shinde(dshinde)

February 28,2019

1 Problem Statement

Human Activity Recognition is the task of classifying different commonly carried out activities(like lying, standing, running, cycling, ascending stairs, etc.) based on sensory data from devices like accelerometer and gyroscope. It is an important task in the healthcare domain. By classifying activities of patients and monitoring their trends, valuable insights can be obtained. Most of today's smartphones come with IMUs (inertial measurement units), which have in-built accelerometer and gyroscope sensors. We will be using the [PAMAP2 Physical Activity Monitoring Data Set](#) from the UCI Machine Learning Repository. This is the dataset which hold the entries for readings from different sensors. Using this dataset, we will try to answer a set of questions based on the data, and will also try to tackle the issue of user-dependency on the sensory data for different activities.

2 Data

We are using PAMAP2 Dataset (Physical Activity Monitoring Data Set) from the UCI machine learning repository. This dataset is collected by setting 3 sensors at wrist, chest and ankle. 9 subjects (1 female and 8 male) are participated in the data collection. 18 different activities are performed by every subject and sensor readings are collected and mapped with activity ground truths. Here are some points to mention about the dataset:

- Data feature includes 9-axis IMU data streams for sensors on each of hand, chest, and ankle and subject heart rate.
- 1.9 million data points of 52 features each, spread over nine subjects.
- 18 different activity IDs, including sitting, walking, running, folding laundry, and cycling..9 million data points of 52 features each, spread over nine subjects.
- Total size of dataset: 1.61 GB

3 Questions

General perception is that there may exist a correlation between a subject and his/her activities. There can be two scenarios when we want to recognize the activity.

- We have information about the subject.
- We don't have information about the subject.

We are looking towards human activity recognition as a multi-class classification problem. In the first scenario, we are trying to answer two different questions.

1. For the given subject, what is the activity?
2. For the given activity, who is the subject?

For the second scenario, we will try to build a user-independent model for classification problem. We will try to generalize the model so that our model can be used for any user, whilst still achieving decent performance. So, our question is as follows

3. Predict what is the activity without any prior information about the user.

4 Evaluation Criteria

There are a lot of existing projects and papers which deal with HAR. We also have a publication for this very dataset, [“Creating and benchmarking a new dataset for physical activity monitoring”](#) , which can be used as a standard benchmark for the evaluation of our models.

5 Timeline and Roles

- Week 1:
 - Feature Extraction
 - Data Pre-Processing
- Week 2-4:
 - Fitting and building Different Models for all questions
- Week 5:
 - Comparison of different Models and evaluating performance
- Week 6-7:
 - Try to optimize the model that works best for the dataset for different questions, including user-independent one.
 - Generation of Reports
- Week 8:
 - If time permits, we want to check how we can achieve decent accuracy with minimal data, i.e. only heart rate monitor and 1 IMU.

bmaiti:

1. Data Preprocessing
2. Build Model for Question 1
3. Model Optimization

dshinde:

1. Data Preprocessing
2. Build Model for Question 2
3. Model Optimization

vwali:

1. Data Preprocessing
2. Build Model for Question 3
3. Model Optimization