Human Activity Recognition using data collected by Smartphone sensors

November, 2019

Prepared by:

Abhishek Gupta (2017120015)

Rupal Handoo (2017120017)

Siddhant Bhagat (2017120006)

Problem Introduction

Smartphones have become an integral part of our lives today. We carry it with ourselves everywhere we go. Hence, there also comes a question about what all activities humans perform with smartphones in their possession. The other side of the problem highlights the capability of a smartphone to deliver useful human activity insights, by means of the data collected by the sensors of the smartphone (the gyrometer and the accelerometer being the most prominent). It is in this aspect that we use our knowledge of EDA and data classification to make sense of the expansive data collected.

Dataset

The dataset was created by 30 participants while they performed activities of daily living while carrying a waist-mounted smartphone. The phone was configured to record two implemented sensors (accelerometer and gyroscope). For these time series the directors of the underlying study performed feature generation and generated the dataset by moving a fixed-width window of 2.56s over the series. Since the windows had 50% overlap the resulting points are equally spaced (1.28s).

Features and Processing

An overwhelmingly elaborate feature set consisting of 564 columns. Our job as a data scientist is to make sure that we select only the most relevant predictors for our desired output. There are numerous statistical parameters to ensure we select the data predictors, which ensure maximum efficiency of our trained model. Various techniques have been employed to ensure the same. All the data predictors are numerical in nature. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers were selected for generating the training data and 30% the test data.

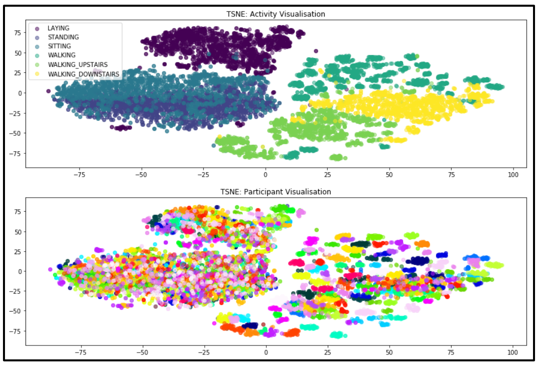
Models and Techniques

PCA: Principal components analysis (PCA) has been extensively used in multivariate statistical process control. We use it here to find a unique walking style for each participant. We used it to categorize and analyze interrelationships among a large number of variables. It finds a sequence of linear combinations of the variables that have maximal variance and are mutually uncorrelated. The objective of using PCA was to reduce the number of variables and to cluster them into more close and manageable groups. PCA also serves as a tool for data visualization, which could help us, track the path of walking structure of each individual.

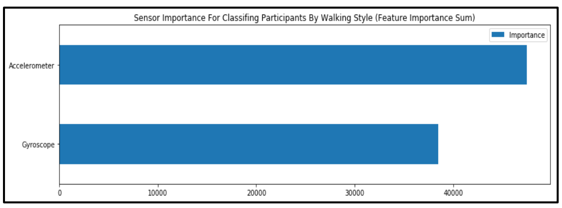
LGBM Classifier: LightGBM is a gradient boosting framework that uses tree based learning algorithms. A GBM is a model that combines the efforts of multiple weak models to create a strong model.

t-SNE: t-Distributed Stochastic Neighbor Embedding (t-SNE) is a recently discovered dimensionality reduction technique. It is a non-linear technique for dimensionality reduction that is particularly well suited for the visualization of high-dimensional datasets. It is extensively applied in image processing, NLP, genomic data and speech processing. t-SNE maps the multi-dimensional data to a lower dimensional space and attempts to find patterns in the data by identifying observed clusters based on similarity of data points with multiple features. To sum up, it is a very powerful EDA technique.

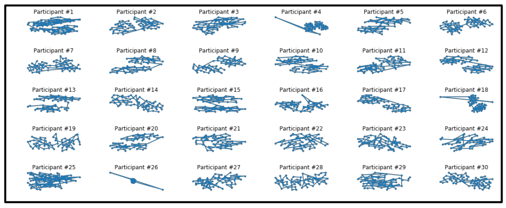
Results and Discussion



The top plot shows that the different activities are separable in most cases. The bottom plot shows that there exist a number of walking styles across individuals. With a basic untuned model the activity of the smartphone user can be predicted with an accuracy of 95.57%.



The accelerometer supplies slightly more information. Both sensors are important for classification and refraining from using both sensors will be a drawback for the quality of the model.



Each cluster represents a single walking cluster, and also consists of physically recognizable outliers.

The other notable observations have been attached with the code notebook as a part of this project.

Conclusion

Within a short time (1-1.5 min) the smartphone has enough data to determine what its user is doing (95%: 6 activities) or who the user is (Walking 94%: 30 participants) and even the basics of a person's specific walking style (slow steps per second). By linking this insights to more personal data of the participants extensive options open up.

In addition these insights have been extracted from only two smartphone sensors, which probably could be accessed, by most of our mobile applications.

References

Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine (IWAAL 2012, Spain) & *Human Activity and Motion Disorder Recognition: Towards Smarter Interactive Cognitive Environments (ESANN 2013, Belgium)*