Behavioural Context Detection

Recognition and Analysis from Mobile Sensors

TA Guide:
Garvit Chugh

Course Instructor:
Suchetana Chakraborty

Presented by:
Abhijit S Iyer (M21CS001)
Shruti Sureshan(M21CS015)

Brief Introduction

In this project, we look forward to analyse data that has been obtained from various sensors present in smartphones and smart watches, based on which we will try and find how a unique combination of these data points can be used to identify a user uniquely.

To facilitate our purpose, we have taken data from the ‘Extrasensory Dataset’ that contains sensor data taken from various smartphone sensors of different users at different points of time.

The Extrasensory Dataset

The extrasensory dataset contains sensory data and labels collected from over 60 smartphone users in a span of 7 days.

There are about a thousand instances of data for each user, and every instance contains data collected from sensors such as accelerometer, gyroscope, magnetometer, phone state, etc.

Experiments so far

- Activity detection (walking) based on features - phone accelerometer and smartwatch accelerometer
- Activity detection (walking) based on features - phone accelerometer, watch accelerometer, gyroscope and location sensor.
- Activity detection (walking) based on features - location sensor and audio
- Clustering data based on every feature, i.e sensor available using K-Means Clustering
- Clustering data based on all features, using GMM
- Dimensionality reduction using PCA and checking for prediction accuracy
- Clustering data based on accelerometer and gyroscope features only (using K-means clustering and GMM)

Activity Detection based on Accelerometers

In this experiment, we have predicted the label of whether the user using the smartphone is *walking*, based on data collected from smartphone accelerometers and smartwatch accelerometers only. Three different classification algorithms were used to solve this purpose. The results are as follows:

	Features selected	Classifier	Accuracy
0	Phone accelerometer, Watch accelerometer	Logistic reg	0.947598
1	Phone accelerometer, Watch accelerometer	KNN	0.954876
2	Phone accelerometer, Watch accelerometer	Random Forest	0.959243

Activity Detection based on Accelerometers, Gyroscope and Location Sensor

A similar experiment was conducted as the previous experiment, but the gyroscope sensor data and location data was included to predict if a user was walking. The results are as shown in the figure below:

	Features selected	Classifier	Accuracy
0	Phone accelerometer, Watch accelerometer	Logistic reg	0.947598
1	Phone accelerometer, Watch accelerometer	KNN	0.954876
2	Phone accelerometer, Watch accelerometer	Random Forest	0.959243
3	Phone acc, Watch acc, Gyroscope, location sensor	Logistic regression	0.946143
4	Phone acc, Watch acc, Gyroscope, location sensor	KNN	0.946143
5	Phone acc, Watch acc, Gyroscope, location sensor	Random forest	0.950509

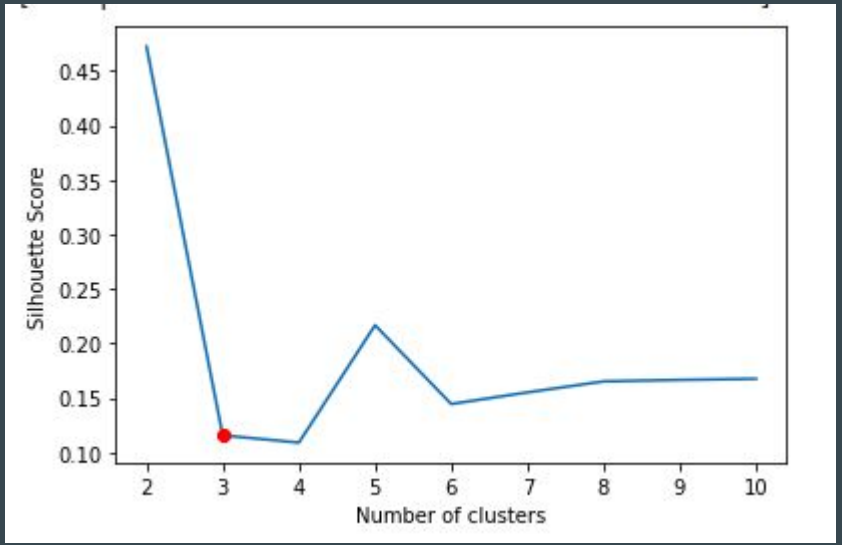
Activity Detection based on Location Data and Audio Data

This is again another similar experiment that was performed, using location sensory data and audio sensory data. Data from these points were used to predict if a person is walking or not, according to the labels assigned in the Extrasensory dataset. In this case too, the experiment was performed on three different classifiers.

Location and Audio features	Logistic regression	0.928675
Location and Audio features	KNN	0.947598
Location and Audio features	Random forest	0.947598

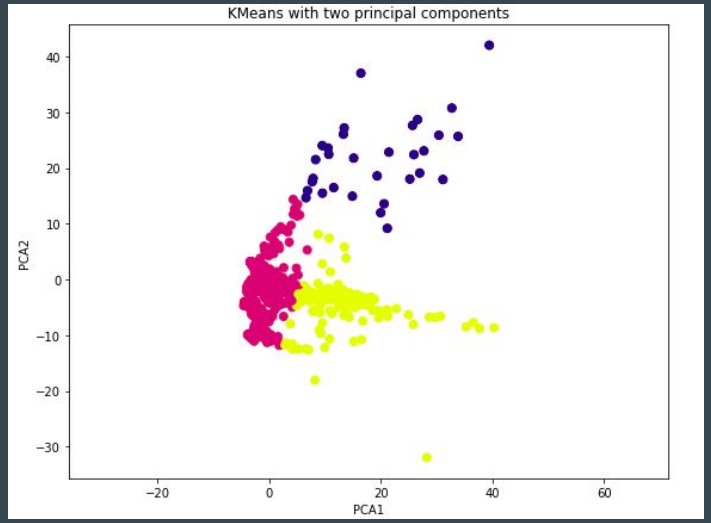
Clustering all instances, based on all features available in dataset

In this experiment, we applied K-Means clustering and GMM on the dataset, without filtering out on any feature. The silhouette scores for different number of clusters were observed as follows:



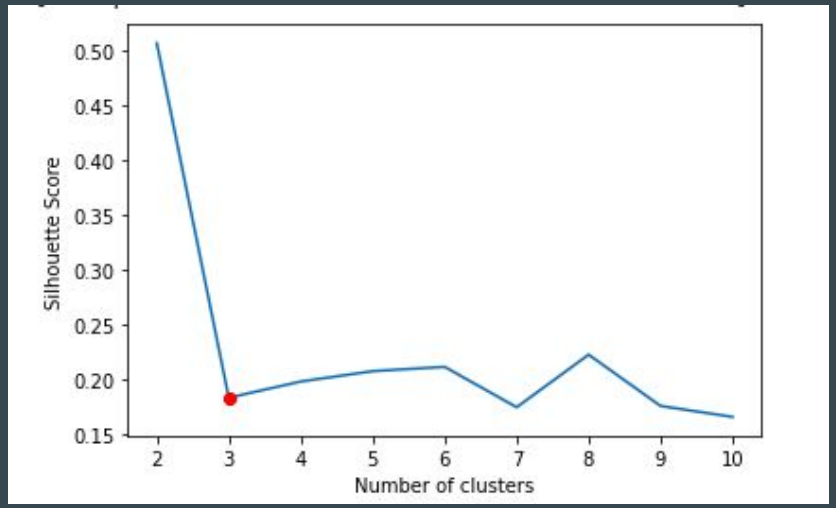
Clustering all instances after applying PCA

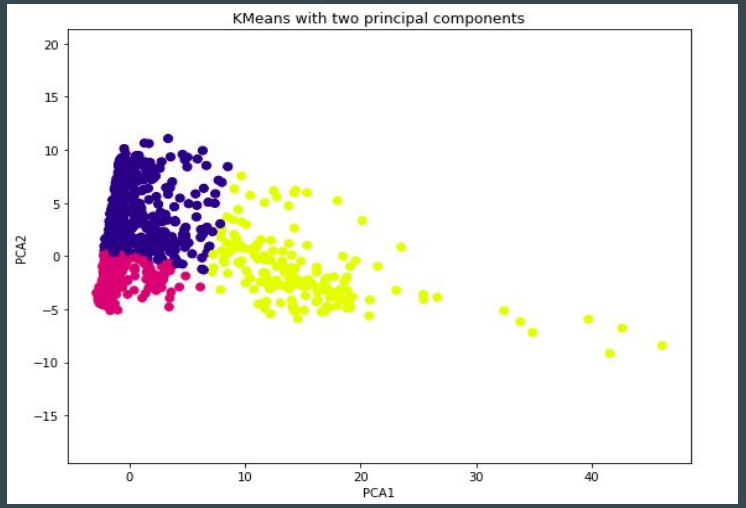
Since the number of features are quite high, we used dimensionality reduction to remove highly correlated features to obtain better clarity on the data. Post dimensionality reduction using PCA, we applied clustering mechanisms to check on silhouette score.



Clustering Based on Accelerometer, Gyroscope, Location and Audio

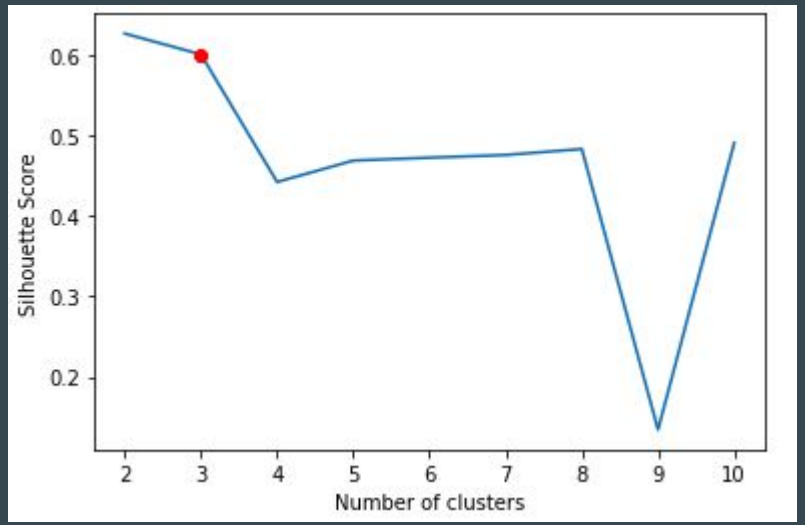
Instead of choosing all features available in the input dataset, the clustering experiment was carried out only on the above mentioned features of data points. The silhouette score graphs for different values of number of clusters can be seen as follows:

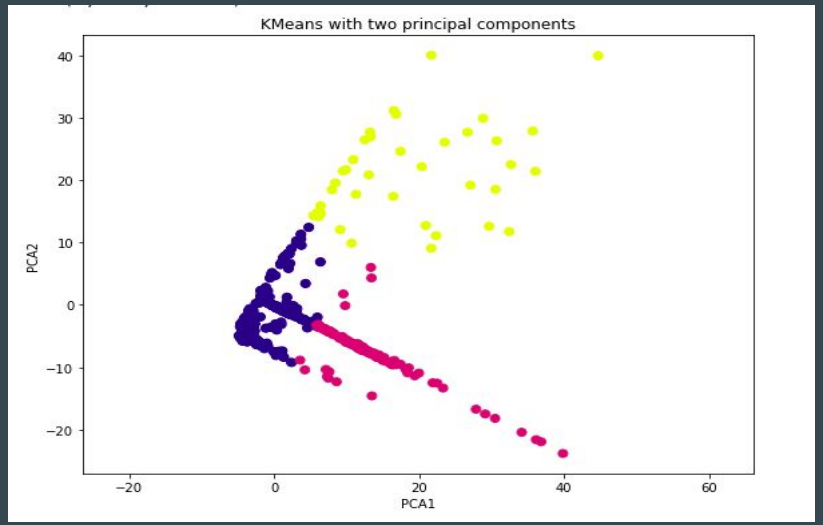




Clustering Based on Accelerometer and Gyroscope

Clustering was done only on accelerometer data and gyroscope data. The silhouette score chart can be shown as follows:





Future Directions

- Identification of the right set of features that will help in identifying each user uniquely
- Performing more experiments, i.e., make use of more algorithms and try and find out of goals can be achieved
- Multiclass classification for activity labelling, i.e walking, lying down, etc
- Subcategorization of activities
- Thorough analysis of experiments performed so far so as to check if worthwhile information can be concluded from them

Conclusion

So far, we have experimented with different features available to check if they can be used for identifying a person uniquely. We will analyse the information obtained thoroughly to see where we can proceed further. We are looking to find out the best set of features that can help us identify users uniquely.

Once the right set of features can be found out, we can apply clustering/ classification techniques to label users uniquely based on behavioural patterns.

Thank You