Human Activity Recognition from Continuous Ambient Sensor Data Set

- Vasudevan, Vignesh, Vasudevan.v@northeastern.edu
- Palanisamy, Rajaragunanthan, p.r@northeastern.edu
- Rajan, Sujai, rajan.suj@northeastern.edu

Objective:

Human activity recognition is a challenging time series classification problem. It involves predicting the movement of a subject purely from data collected from sensors and traditionally requires the deep domain expertise and involvement from signal processing to understand raw data and fit a predictive model. In latest developments, deep learning techniques like CNN and RNN bring out state-of-art results by self-learning features from the available raw sensor data. The objective is to recognize the activities of people who volunteered for the collection of data from various sensors placed at different spots in their homes. So, using the input sensorial data - Multivariate, Sequential, Time-Series, we need to classify them into one of the below labels:

Current state-of-art:

The current research works include: Single layer LSTMs and Deep convolutional LSTM architectures to accurately classify a given input activity data point into one of the correct classes. Contrary to popular belief, an LSTM model with relatively lesser number of layers outperforms the other complex LSTMs. Modelling 3D data is also a technique that has been extensively researched in recent times. The human body consists of limbs connected by joints, and stronger features captured from depth cameras is known to provide a higher accuracy on predicting human activity than a model built to predict on 2D data.

Approach:

Since the dataset essentially is a timeseries data and the objective is a classification problem, we are planning to equip the multi-layered perceptron architecture as the baseline model for this setting. So, we will first train and test the perceptron model, and examine the goodness of the fit based on accuracy, F1 score, and other metrics, and based on the results, we will then move on to the more flexible models such as CNN (convolutional Neural Networks), RNN (Recursive Neural Networks), and LSTM models which are proven to give best results for timeseries classification problems. In the end, we will compare and contrast the results from all the models to determine which would best fit this classification task.

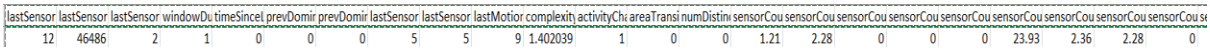
Dataset:

For the purpose of this classification task, we will be making use of a dataset which has 35 classes present in the training data which include 321429 rows, containing data from all sensors and the predicted variable column with activity labels such as cook_breakfast, cook_lunch, cook_dinner, entertain_guests, groom, wake_up, etc. We will use a subset of the test data available which contains 1048576 rows of data containing all features except the output variable – ‘activity’, as

validation data and another subset as the test data. The dataset is publicly available in the UCI repository.

Link - https://archive.ics.uci.edu/ml/machine-learning-databases/00506/

A snapshot of the dataset is included below:



lastSensor	lastSensor	lastSensor	windowDu	timeSincel	prevDomir	prevDomir	lastSensor	lastSensor	lastMotior	complexity	activityCh	areaTransi	numDistin	sensorCou	sensorCou	sensorCou	sensorCou	sensorCou	sensorCou	sensorCou	sensorCou	sensorCou	se
12	46486	2	1	0	0	0	5	5	9	1.402039	1	0	0	1.21	2.28	0	0	0	23.93	2.36	2.28	0	

Project Timeline:

Before November 10st – Studying the existing research and understanding the problem description and the dataset

Before November 17th – Deciding and formulating the models, finding the right tools and techniques to be equipped to solve the problem.

Before December 1st – Performing the validations, evaluating the performances of the models, and Fine tuning based on accuracy, F1 score, tec.

Before December 7th – Writing conclusions and the final paper

Deliverables:

- **Research –** Detailed project execution plan
- **Data exploration –** Report with the results
- **Modeling –** Model prototype, Model evaluation report
- **Result analysis –** Report with the results
- **Final –** Code reviewed model