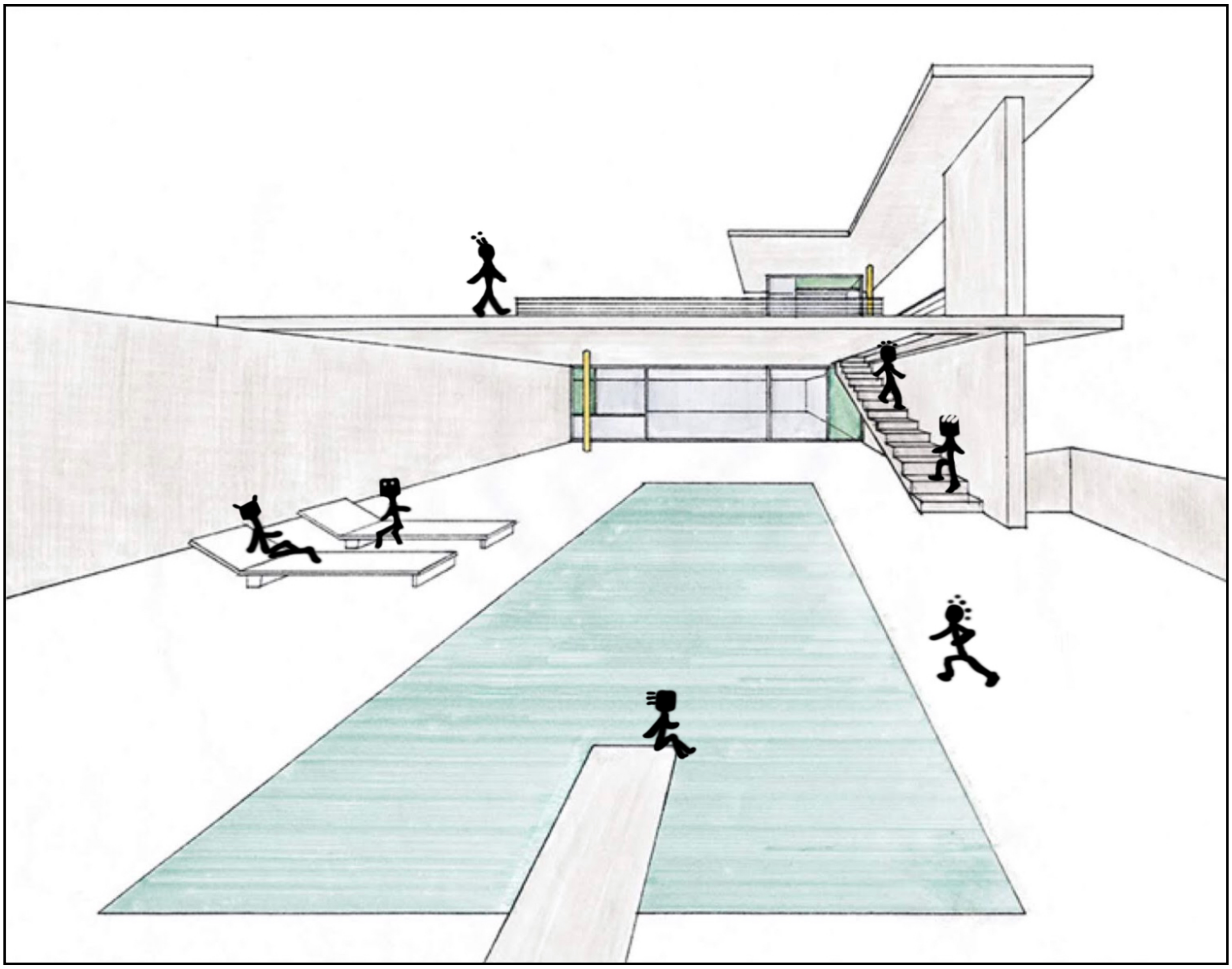
HUMAN ACTIVITY RECOGNITION USING SMARTPHONES DATA SET



A Capstone Project

Prepared for: Udacity Machine Learning Engineer Nanodegree Program
Prepared by: Shan Sun, London, UK
4 June 2018

PROJECT PROPOSAL

Domain Background

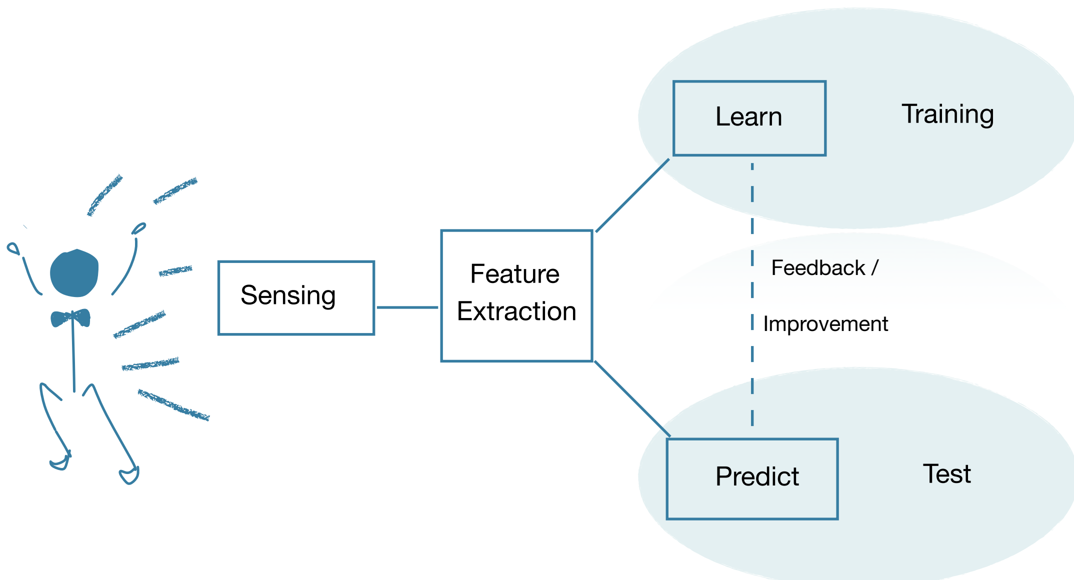
The availability of affordable wearable equipments and portable computing devices results in behemoth amount of data being collected including motion, location, physiological signals and environmental information. The human activity recognition (HAR) is an active research field to understand how human behaviours are developed by interpreting attributes derived from this data. "*This field is the first component (sensing) of the sequence for achieving Smarter Interactive Cognitive Environments together with data analysis, decision making and taking action, and our subject of research.*"[1] HAR is especially appealing to healthcare applications, such as caring and monitoring for elderly people. For example, physiological signals could be an indicator of a health condition via a change in heart rate and body temperature.

I become interested in wearable devices since I live far away from my parents which means I do not have the opportunity to watch over their health on a regular basis. It is my personal interest to explore this area. Additionally, for similar reasons, I also started virtual reality (VR) development. Although not yet apparent, I believe HAR can provide some indirect assistance to solve the motion sickness problem in VR by classifying human movements accurately and assisting real time re-calibration adjustments in VR visions.

Problem Statement

This paper addresses the HAR problem. The availability of affordable wearable devices, such as smart phones, enables mass data collection from users thanks to the embedded inertial sensors. Sensor signals from accelerometer, heart rate monitors, thermometers and gyroscopes are physiological signals data collected from these devices attached to human bodies. With the availability of data, analyses can be done to better recognise, classify, cluster and predict what human activities are carried out (e.g. stand, sit, lie, walk) for further decision making. Several papers already exploited this area, using supervised and semi-supervised learning methods such as support vector machine, ensemble methods with boosting and, deep learning techniques such as artificial neural network[1][2]. The diagram below summarises the process pipeline for HAR.

HAR Process Pipeline



Inspired by: Reyes-Ortiz et al (2013)

Datasets and Inputs

The dataset I used for this paper is from the UCI Machine Learning Repository titled "Smartphone-Based Recognition of Human Activities and Postural Transitions Data Set" (SBHAR). The data was collected from experiments with a group of 30 volunteers aged between 19 to 48 years wearing a Samsun Galaxy SII on the waist. Volunteers performed a protocol of activities including six basic postures: three static - standing, sitting, lying and three dynamic - walking, walking downstairs and walking upstairs. The experiment also included transitional postures between static postures. These are: stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie and lie-to-stand. The sensor signals were collected from the accelerometer and gyroscope in the smartphones including 3-axial linear acceleration and 3-axial angular velocity at a constant refresh rate of 50Hz. Manually labelled data was also added to the video footage of this experiment [1]. SBHAR was partitioned randomly into test and training sets at the ratio of 70:30.

The SBHAR data generated around 5-hours of experimental data [3], and was also pre-processed with noise filters sampled by fixed-width sliding windows (2.56 sec with 50% overlap, i.e.128 readings/window). Other data transformations were also applied including the calculation of Jerk signals from time, body linear acceleration and angular velocity information; magnitude using Euclidean norm and, the frequency domain signals using Fast Fourier Transform ("FFT"). The resulting dataset has 561-feature vector per example derived from 17 action patterns and 17 functions over 12 activities labels. The table below summarise the SBHAR dataset structure :

SBHAR DATA SUMMARY

Index	Signal Variables	Functions	Labels	Activity Labels
1	tBodyAcc-XYZ	mean()	1	WALKING
2	tGravityAcc-XYZ	std()	2	WALKING_UPSTAIRS
3	tBodyAccJerk-XYZ	mad()	3	WALKING_DOWNSTAIRS
4	tBodyGyro-XYZ	max()	4	SITTING
5	tBodyGyroJerk-XYZ	min()	5	STANDING
6	fBodyAcc-XYZ	sma()	6	LAYING
7	fBodyAccJerk-XYZ	energy()	7	STAND_TO_SIT
8	fBodyGyro-XYZ	iqr()	8	SIT_TO_STAND
9	tBodyAccMag	entropy()	9	SIT_TO_LIE
10	tGravityAccMag	arCoeff()	10	LIE_TO_SIT
11	tBodyAccJerkMag	correlation()	11	STAND_TO_LIE
12	tBodyGyroMag	maxInds()	12	LIE_TO_STAND
13	tBodyGyroJerkMag	meanFreq()		
14	fBodyAccMag	skewness()		
15	fBodyAccJerkMag	kurtosis()		
16	fBodyGyroMag	bandEnergy()		
17	fBodyGyroJerkMag	angle()		

Features are selected from eight 3-axial raw signal variables and nine single component variables passing through 17 functions resulting in 8*3*17 +(17-8)*17 = 561 features for each example.

I believe the pre-processing (albeit with some arbitrary decisions such as the selection of filters) and the construction of various components within the dataset is reasonable and sufficient to train and test for the HAR problem from the aforementioned experiments. Further, the dataset can be divided into two separate parts for independent analyses:

1.) Inertial sensor data with raw sensor signals (accelerometer and gyroscope) with labelled activities;

2.) Activity windows (each window is one example) each has a 561-feature vector with time and frequency domain variables and labels for activities and user identifier.

Solution Statement

With the dataset ready, the HAR problem essentially becomes a classification problem with labelled data under supervised learning. This paper addresses the HAR problem by applying two types of machine learning techniques: supervised learning using several classifiers (e.g. Random Forest, Support Vector Machines, and Gaussian Naive Bayes). Subject to time constraint and a deeper exploration of the dataset, I also propose to use a Convoluted Neural Network (CNN) to classify activities. For each of the proposed algorithm, there are some common metrics to use to compare performance. This is discussed in more detail in the "Evaluation Metrics" section below.

Benchmark Model

Most literature around the HAR problem based on inertial sensor signals are rather recent due to the increased adoption and popularity of smart devices. Much research was dedicated on applying machine learning techniques to this problem, which are often some form of nonlinear multivariate regressions and pattern recognition [4]. The table below summarises some relevant papers and results. Early studies did not combine 3-D accelerometer and gyroscope data but was included as long as at least one type of sensors were used. Note the metric used is accuracy in almost all papers which is discussed in more detail in the section below.

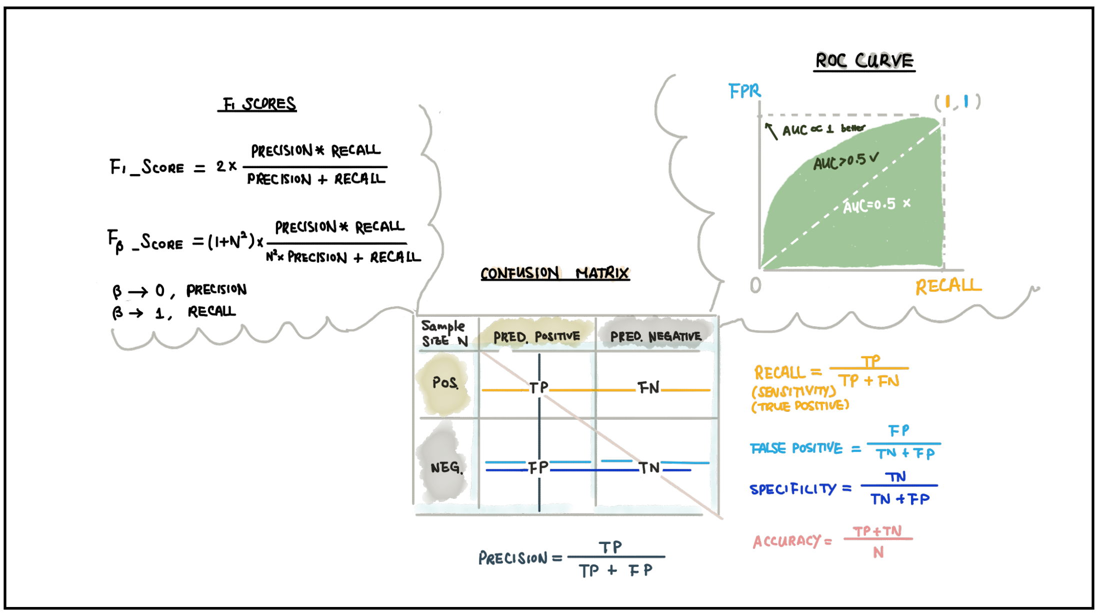
SUMMARY OF BENCHMARK MODELS

Sensors used	Features	Classifiers	Activities	Accuracy (%)	Reference
1 2D acc 1 gyro	Wavelet coefficients	Threshold-based	5	> 90	[5]
1 2D acc 1 gyro 1 compass	Raw data, Standard Deviation, Derivative	Threshold-based	5	92.9 to 95.9	[6]
1 3D acc	Raw data, Delta Coefficients, DC Component	GMM	8	91.3	[7]
2 3D acc	Wavelet coefficients	ANN	4	83 - 90	[8]
2 3D acc 1 gyro	Constructed features from raw data and signal frequency components	Multiclass SVM Multiclass Hardware Friendly SVM	12	Precision: 89.3 Recall: 89	[3]

This paper classifies 12 activities including three static, three dynamic and six transitional postures using data from the two 3-D accelerometers and the gyroscope built in within smartphones. The bare minimum accuracy I expect from a model is therefore 1/12 or 8.33%, i.e. equiprobability of a random guess. From the summary result above for similar studies, I consider an accuracy above 90% an appropriate benchmark.

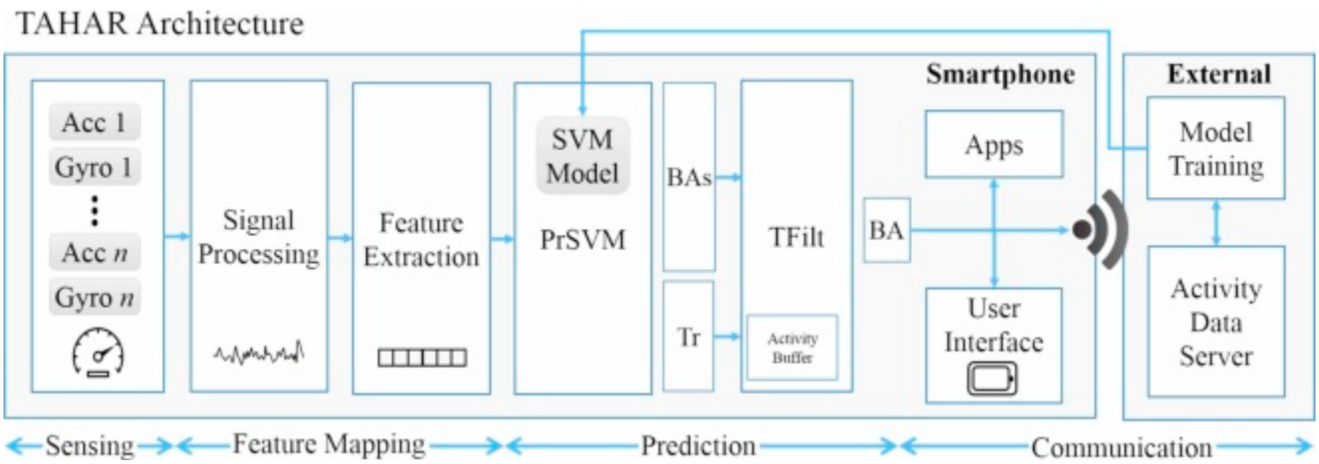
Evaluation Metrics

As the table above indicated, the most commonly used evaluation metrics for HAR problem is accuracy. Some research also looked at the confusion matrix which allows clear algorithm representation whilst identifying error types. Since the HAR problem can be considered as a classification problem in machine learning, a confusion matrix is particularly helpful to evaluate the models used. Specifically, precision represents the number of correctly classified activities out of all the activities being classified; whilst recall represents each type of activity in truth how many the machine correctly classified. The relationship between precision and recall can also be calculated by F1 Scores and represented by ROC curves if a single measurement is desirable. The relationships of confusion matrix are summarised in the graph below.



Project Design

This paper loosely follows the transition-aware human activity recognition (TAHAR) architecture for workflow purpose [9] with some changes to the sections of feature selection and machine learning classification techniques applied. The TAHAR system architecture is summarised in the figure below:



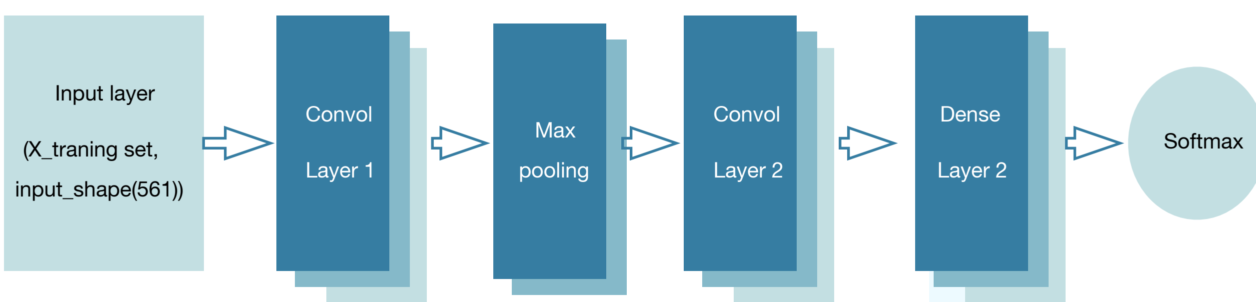
Source: TAHAR System Architecture (Reyes-Ortiz, 2016)[9]

Specifically, this paper focuses on the feature mapping and prediction phases of the TAHAR architecture and is organised in the following steps:

A. data exploration: in this step, the paper will explore the shape of the SBHAR data and consider if there is any outliers, abnormalities or class imbalance thereof.

B. data preprocessing: based on the result from the previous step in data exploration, this step may see some preprocessing methodologies being implemented. One apparent characteristics of the data is high dimensionality of feature vectors (561 columns), the common technique of dimension reduction such as the Principle Component Analysis (PCA) will be deployed to address this issue.

C. machine learning algorithms selection: as discussed above, there are two types of machine learning algorithms this paper will aim to adopt, supervised learning classifiers and a simple CNN. In the supervised learning, this paper intends to try out three common classifiers first - Random Forest, Support Vector Machines, and Gaussian Naive Bayes. From the literature review in the Benchmark Model section above, some ensemble method such as Adaboosting may also be explored. Additionally, a simple CNN may also shed some light on addressing the classification problem, the structure of which is included below:



D. **implementation:** the selected machine learning methodologies discussed above are implemented in this step with the first round of evaluation metrics calculated.

E. **refinement:** based the evaluation metrics results calculated above, further considerations are given in this step to fine tune the hyper-parameters in relevant algorithms and refine the models to better performance.

F. **results:** final results are presented in this step with the best performing algorithms. This paper intends to calculate the confusion matrix with some free style visualisation to represent the best performer.

G. **conclusion and future work:** this step provides more discussions on what the results mean in a contextual environment and what could be done for future work.

References:

[1] Jorge Luis Reyes-Ortiz, Alessandro Ghio, Xavier Parra-Llanas, Davide Anguita, Joan Cabestany, Andreu Català. Human Activity and Motion Disorder Recognition: Towards Smarter Interactive Cognitive Environments. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.

[2] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A Public Domain Dataset for Human Activity Recognition Using Smartphones. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.

[3] Jorge-Luis Reyes-Ortiz, Luca Oneto, Alessandro Ghio, Albert SamÃ¡, Davide Anguita and Xavier Parra. Human Activity Recognition on Smartphones With Awareness of Basic Activities and Postural Transitions. Artificial Neural Networks and Machine Learning â€“ ICANN 2014. Lecture Notes in Computer Science. Springer. 2014.

[4] Andrea Mannini and Angelo Maria Sabatini. Machine learning methods for classifying human physical activity from on-body accelerometers. Sensors, 10(2):1154–1175, 2010

[5] Najafi B., Aminian K., Paraschiv-Ionescu A., Loew F., Büla C., Robert P. Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. IEEE Trans. Biomed. Eng. 2003;50:711–723.

[6] Lee S.W., Mase K. Activity and location recognition using wearable sensors. IEEE Perv. Comput. 2002;1:24–32.

[7] Allen F.R., Ambikairajah E., Lovell N.H., Celler B.G. Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models. Physiol. Meas. 2006;27:935–951.

[8] Mantyjarvi J., Himberg J., Seppanen T. Recognizing human motion with multiple acceleration sensors. Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics; Tucson, AZ, USA. October 7–10, 2001; pp. 747–752.

[9] Reyes-Ortiz, Jorge Luis et al. "Transition-Aware Human Activity Recognition Using Smartphones." Neurocomputing 171 (2016): 754-767.

REPORT

(to be completed)