Runtime Error: Human Activity Recognition using Smartphones

Tejas Srinivasan
PES1201800110
Department of CSE, PES University
tejas.srinivasan007@gmail.com

Joseph Dominic Cherukara
PES1201800328
Department of CSE, PES University
jd.cherukara@gmail.com

Anirudh S Ayya
PES1201800338
Department of CSE, PES University
anirudhshrikanth65@gmail.com

GBS Akhil
PES1201800188
Department of CSE, PES University
akhilgbs123@gmail.com

***Abstract*—Human activity recognition (HAR) is the automatic identification of physical activities performed by human subjects and is one of the most essential and active areas in the fields of computer vision and machine learning. It has various applications in the fields of assistive healthcare, security and surveillance and human-computer interaction (HCR) to mention a few of the many applications. In this paper we have attempted to get a clear understanding of the sensor data provided to us in the dataset and its source and the conditions that were provided during its collection. We have done a thorough comparison of the models that have been implemented in previous papers through our literature survey and a summary of the results obtained by these papers and the limitations in their approach have been mentioned in the appropriate section as well. We propose a neural network based model to predict the activity undertaken by a subject and plan to compare this model with the previous models for results and tune it to see how it performs under different conditions on different metrics. We also aim to explore Hidden Markov Models (HMMs) to see how they perform on transitional activities. Extensive experiments with a publicly available dataset of human activity with smart phone inertial sensors can be used to show that the proposed approach/approaches can indeed lead to development of intelligent and automatic real time human activity monitoring for eHealth application scenarios for elderly, disabled and people with special needs.**

***Index Terms*—Human activity recognition and prediction, Gyrometer and Accelerometer (sensor) data, Deep learning models**

I. Introduction

The first commercial hand-held mobile phones appeared in 1979, and since then there has been an unprecedented growth in the adoption of mobile phone technology, reaching to more than 80% of the world population by 2011. Lately, smartphones, which are a new generation of mobile phones, are equipped with many powerful features including multitasking and a variety of sensors, in addition to the basic telephony. The integration of these mobile devices in our daily life is growing rapidly, and it is envisaged that such devices can seamlessly monitor and keep track of our activities, learn from them and assist us in making decisions. Such assistive technologies can be of immense use for remote health care, for the elderly, the disabled and those with special needs, if there are autonomous and intelligent. However, currently, though there is good capacity for collecting the data with such smart devices, there is limited capability in terms of automatic decision support capability and making sense out of this large data repository. There is an urgent need for new data mining and machine learning techniques to be developed to this end.

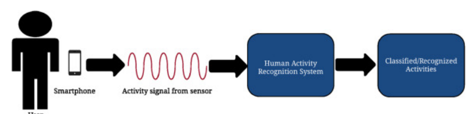

Fig. 1. A depiction of a human activity recognition system

Mobile devices are becoming increasingly sophisticated and the latest generation of smart cell phones now incorporates many diverse and powerful sensors. These sensors include GPS sensors, vision sensors (i.e., cameras), audio sensors (i.e., microphones), light sensors, temperature sensors, direction sensors (i.e., magnetic compasses), and acceleration sensors (i.e., accelerometers). The availability of these sensors in mass-marketed communication devices creates exciting new opportunities for data mining and data mining applications. This can be immensely useful in healthcare applications, for automatic and intelligent daily activity monitoring for elderly.

If one would be able to analyze complex acceleration data, it would be possible to develop an Activity Recognition System, using a smartphone. The reason for using an acceleration sensor for activity recognition is to obtain information without depending on the environment, such as being indoors. If the situation and activity of a user is obtained from acceleration data of the smartphone, an ambient network service is made possible people. The dataset has been created using inertial data from smartphone accelerometers and gyroscopes, targeting the recognition of six different human activities, (standing, sitting, walking, laying, walking upstairs, walking downstairs).

II. Related Work

In this section we summarize the existing work that has done using the dataset for various purposes, especially the aspect that is relevant to our work. The assumptions made, a summary of the approach used and the results obtained have been mentioned. Furthermore any lacuna inferred from the following approaches along with any limitations have also been appropriately mentioned.

Canova et.al [1] mainly focus on building classifiers to accurately predict the activities performed by individuals using their smartphones sensor data. The original approach detailed a support vector machine (SVM) adapted for multiclass classification, using computational efficiencies that exploit fixed-point arithmetic. Further studies also showed multiple sensors aid in recognition because conjunctions in acceleration feature values can help to identify many activities. A section on data visualization is also there in the paper, with an emphasis on two main techniques (PCA and t-SNE) to clearly understand the distinction in the categories of data, an approach that we plan to take as well. Both algorithms effectively distinguish between activities of motion (walking, walking upstairs, walking downstairs) and static activities (sitting, laying,standing), and each of the activities are well represented by a cluster.

Different algorithms such as a baseline multinomial model (given the high dimensionality of the data), linear, radial-based and polynomial kernel SVMs and in comparison to these, gradient boosted trees. At the end of the paper, the result of the models were given using the overall misclassification rate as the primary performance metric. Each of the models displayed a similar performance in terms of the train and test error, which indicates that increasing the complexity of the model does not significantly improve its overall performance. Since the linear kernel SVM was found to have a low misclassification rate and is computationally efficient to train, it was further diagnosed and a confusion matrix was constructed for it using the entire train and test data. Further, activity specific observations were made (such as sitting being the one with the highest misclassification rate) etc. and t-SNE plots were constructed to explore the variance.

In conclusion the paper confirmed the hypothesis that some activities are classified more accurately with subject independent training data, while others require subject specific training data. It was also suggested to use hidden markov models (HMMs) to capture the dependency between activities that involve transitions, an approach that we would love to explore as well. All references used to write the paper were cited in the appropriate section at the end of the paper.

All the latest smartphones capable of detecting the orientation of the device with the help of tri-axial accelerometers that measure acceleration in all three spatial dimensions. The data is divided into 10-second segments and then generated features that were based on the readings contained within each 10-second segment. A 10-second duration was chosen because it provided sufficient time to capture several repetitions of the (repetitive) motions involved.

Jennifer et al. [2] observed the repetitive activities like walking, generate repetitive waves for each axis and this feature tries to measure the time between successive peaks. To estimate this value, for each example we first identify all of the peaks in the wave using a heuristic method and then identify the highest peak for each axis. Then the threshold is set based on a percentage of this value and find the other peaks that met or exceed this threshold; if no peaks meet this criterion then the threshold is lowered until we find at least three peaks. We then measure the time between successive peaks and calculate the average.

The z-axis captures the forward movement of the leg and the y-axis captures the upward and downward motion. The x-axis captures horizontal movement of the user's leg. Sitting and standing do not exhibit periodic behavior. The four other activities, which involve repetitive motions, do exhibit periodic behavior. It was quite easy to identify activities such as walking, laying, sitting and standing since they involve more extreme changes in acceleration.

Three learning algorithms we used - The J48 decision tree algorithm, logistic regression for multi-class classification and the multi layer neural network. it was observed that none of the algorithms consistently performs well but the multi layer perceptron performed comparatively better. The climbing up and down activities were difficult to recognize. Many of the prediction errors are due to confusion between these two activities. The most common incorrect classification occurs when we predict "downstairs," which occurs 107 times out of 545 times and accounts for a decrease in accuracy of 19.6%.

Girija et al. [3] aimed at developing the best model or scheme to classify various activities performed by users with the help of inertial sensors present in the smartphones.

For the validation of the approach the dataset used consists of labelled data of 30 subjects from age group of (19-48) years. The activity was performed wearing smartphone containing accelerometer and gyroscope for measuring 3-axial linear acceleration and angular velocity at constant rate of 50Hz, it was placed around the waist. The dataset consists of total 561 features.

Information based ranking of features was used to discard some features. Random Forests model was used to reduce the issues with high bias and variance by computing an average and balancing the two extremes, due to this it is used to yield an efficient and fast model to solve the problem at hand. Lazy learners such as IBK was also used which are similar to k-nearest neighbour classifier.The models such as Naive bayes and k-mean clustering produced poor accuracies compared to the lazy IBK learner, random forests and ensemble based approaches accuracies.

III. Problem Statement

The dataset contains various physical values as well as labels of what activity the subject was performing when the readings were taken.The dataset has already been divided into testing and training sections. The problem we intend to solve is to be able to predict what the subject is doing based on the given physical values. As the activity labels are manually observed, they are not subject to errors and hence can be used as a scale to test the accuracy of our models.
Each activity has a markedly different accelerometer and gyroscope read-out as is explained in [3]. In Fig. 2., the frequency of acceleration values on each axis is visibly different. Thus it is evident that by analysis of the physical values, we can calculate what activity is being performed by the subject. We intend to use various models to perform these predictions and compare the accuracy of the same.

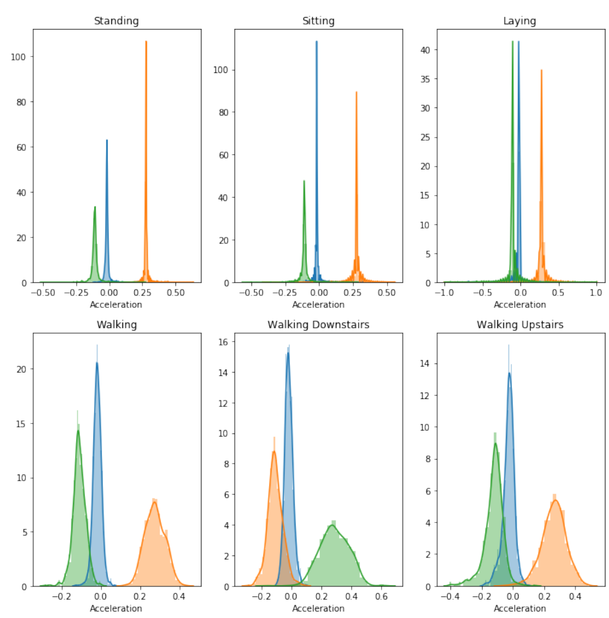

Fig. 2. Frequency of acceleration on different axes for each activity

A. Dataset

The Dataset [4] consists of 561 features.The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities in total (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using the built-in accelerometer and gyroscope, 3-axial linear acceleration and 3-axial angular velocity were measured at a constant rate of 50Hz. The obtained dataset has been partitioned into two sets , 70% training set and 30% test data. The sensor signals were pre-processed by applying noise filters and then sampled in fixed-width sliding windows (128 readings/window).

B. Exploratory Data Analysis and Visualizations done

The dataset has 563 columns in total [5]. Out of the 563 entries it was found that 561 were of type float64, 1 was of type int and 1 object which was the column containing activity label.
On performing basic data preprocessing it was found that there were no missing values or duplicate values in the dataset. Apart from the activity label associated with each attribute, there was also the triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration, triaxial Angular velocity from the gyroscope and an identifier of the subject that carried out the experiment.
The frequency of each activity was found to be around the same so none of the activities are under-represented or over-represented in the dataset.

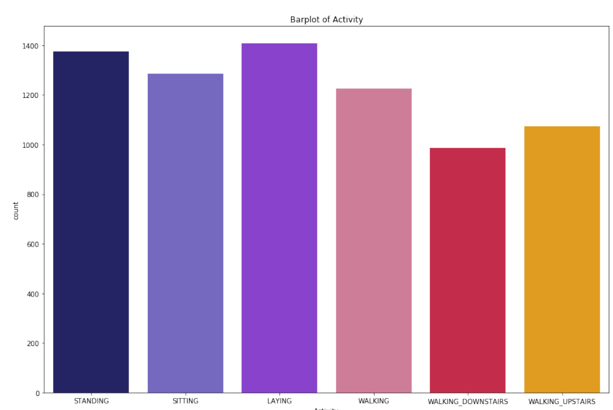

Fig. 3. Activity frequency

Based on the nature of the given activities we can divide the activities as static (SITTING, STANDING, LAYING) and dynamic activities (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS). We can classify an activity one variable at a time and hence construct a probability density function.

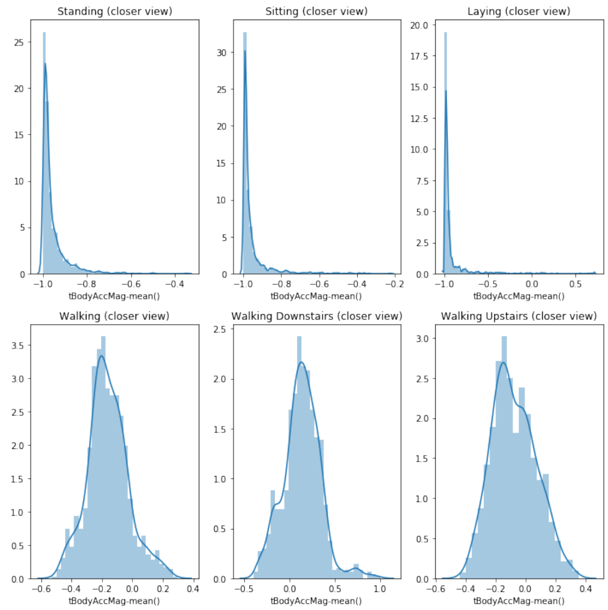

Fig. 4. Probability densities for each activity

Since the data had 561 features and it is hard to visualize the importance of all of it, the tSNE technique was used to reduce the dimensions from a higher dimensional space to a 2D space while still retaining a lot of the information. We can use a seaborn scatterplot to distinguish between each of the activities by dividing them into clusters in a 2D space.

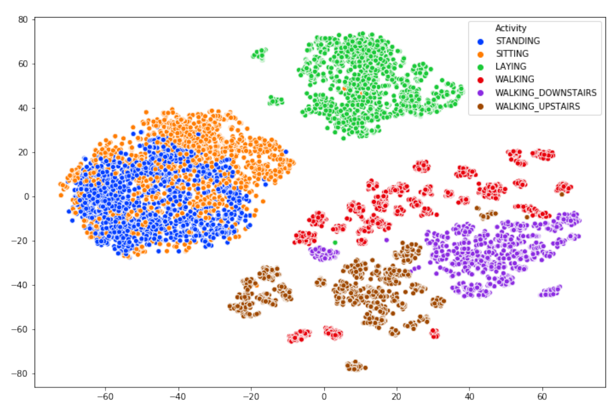

Fig. 5. tSNE plot to reduce dimensionality

IV. Proposed Methodology

In order to use the values to predict what activity the subject is performing, we intend to use a number of different models. We intend to use logistic regression to predict the type of activity the user is performing (i.e. static and dynamic activities). We also plan to try k-means clustering to achieve the same and will compare the outputs.

We then intend to use Support Vector Machines (SVM) to classify the data based on the given accelerometer and gyroscope readings.

We intend to be use neural network architectures such as regular neural networks, and convolution neural networks to predict the activity the subject is performing given the accelerometer and gyroscope readings.

We will then compare the results of the various models to understand which model produces the best results for the dataset.

References

[1] F. S. D. . Nicholas Canova, “Human activity recognition using smartphone sensor data.” http://cs229.stanford.edu/proj2016/report/CanovaShemaj-HumanActivityRecognitionUsingSmartphoneData-report.pdf.

[2] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.

[3] G. Chetty, M. White, and F. Akther, “Smart phone based data mining for human activity recognition,” *Procedia Computer Science*, vol. 46, pp. 1181–1187, 2015.

[4] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, “A public domain dataset for human activity recognition using smartphones.,” in *Esann*, vol. 3, p. 3, 2013.

[5] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, “Energy efficient smartphone-based activity recognition using fixed-point arithmetic.,” *J. UCS*, vol. 19, no. 9, pp. 1295–1314, 2013.