Machine Learning for Signal Processing
Human Activity Recognition

Shridivya Sharma
School of Informatics and Computing
Indiana University
Email: shrishar@iu.edu

Venkatesh Raizaday
School of Informatics and Computing
Indiana University
Email: vraizada@umail.iu.edu

Jongwon Lee
College of Arts and Sciences
Department of Statistics
Indiana University
Email: lee647@umail.iu.edu

***Abstract*—Human Activity Recognition (HAR) is the most important task in medical science. A model for HAR providing accurate and precise prediction is the main topic in medical computing. With wearable devices, inertial signals for human activity is collected. In basis of classes for human activity, five classification models are provided in this article. Multi-class Support Vector Machine (SVM) without feature selection is the initial model and SVM with L1 based feature selection and tree based feature selection are followed in order to compare the SVMs. Then, Deep Neural Network (DNN) and Recurrent Neural Network (RNN) are suggested as the additional models. Consequently, we present the accuracy for the five models and propose the pros and cons for each model.**

1. Introduction

HAR dataset is from University of California Irvine Machine Learning Repository. 30 volunteers within in an age bracket of 19 to 48 years joined the HAR data collection. They performed six activities wearing a smart phone, Samsung Galaxy S II, on their waist. The activities are walking, walking upstairs, walking downstairs, sitting, standing and laying. With the embedded accelerometer and gyroscope, 3-axial linear acceleration and angular velocity at a constant rate of 50 Hz were captured by the smart phone. The accelerometer and gyroscope (sensor signals) were preprocessed by applying noise filters and then sampled in fixed width sliding windows of 2.56 seconds and 50

The main topic for this article is to compare prediction accuracy for each models. The initial model is SVM without feature selection, so the model includes 561 features. Then, SVM with L1 based feature selection and tree based feature selection are built to avoid over-fitting problem. L1 regularization, or lasso, is widely used regularization method. L1 regularization is the loss function with a penalty,

$$\alpha \sum_{i=1}^{n} |w_i|$$

. Each non-zero coefficient adds to the penalty. So, the weak features have zero coefficient. Tree based feature selection is from a decision tree. DNN is a neural network method with more than a single hidden layer and RNN is the method with loops allowing information to persist. Finally, the accuracy for test dataset are compared and we propose the pros and cons for each models.

2. Related Work

The problem of human activity recognition(HAR) using accelerometers was established way back in the 90s when Foerster, Smeja, & Fahrenberg, 1999 came out with their paper but the first major breakthrough came from Bao and Intille (2004) who placed multiple sensors on different parts of the body and applied multiple data mining techniques to produce satisfactory results for the problem statement. This paper concluded that the sensor placed on the thigh was the most effective in recognizing different activities, using the assumption Kwapisz et al., 2010 performed human activity recognition using a single smartphone. They used hand crafted features from the data on multilayer perceptron and j48 decision trees. Their experiments showed that machine learning based classification techniques achieve higher performance in terms of accuracy compared to other data mining techniques. However, none of the classifiers were able to efficiently distinguish between similar activities like moving upstairs and downstairs.

Some more machine learning techniques were employed for solving the problem such as Sharma, Lee,& Chung, 2008 applied an artificial neural network, Khan (2013) used decision trees and the Wii Remote to classify basic activities. Wu, Dasgupta, Ramirez, Peterson, & Norman, 2012 declared k-nearest neighbors (kNN) as the best classifier, but still failed to effectively classify very similar activities. The use of gyroscope along with an accelerometer for classification was introduced in Shoaib, Bosch, Incel, Scholten, & Havinga, 2014. Anguita, Ghio, Oneto, Parra, & Reyes-Ortiz, 2012 used 561 hand-designed features to classify six different activities using a multiclass support vector machine, the feature set that we have used in part of our experiments.

Deep learning algorithms have also been used for HAR, Convolutional neural networks were used together with accelerometer and gyroscope data in the gesture recognition work by Duffner, Berlemont, Lefebvre, and Garcia (2014).

Plotz et al., 2011) made use of restricted Boltzmann machines (RBM), and (Bhattacharya *et al.*, 2014; Li *et al.*, 2014 ; Vollmer *et al.*, 2013), which both made use of slightly different sparse-coding techniques to use deep learning as an automatic feature extraction mechanism. Hammerla, Halloran, Ploetz (2016) explore deep, convolutional, and recurrent approaches across three representative datasets providing a guideline on how to apply deep learning on HAR. Edel and Koppe (2016) have proposed Binarized Long Short-Term Memory Network to minimize the memory footprint associated with using a RNN in a wearable. [1], [2], [3], [4], [5], [6], [7], [8]

3. HAR Dataset and Feature Extraction

A set of experiments were carried out to obtain the HAR dataset. A group of 30 volunteers with ages ranging from 19 to 48 years were selected for this task. Each person were asked to follow certain set of activities while wearing a waist-mounted Samsung Galaxy S II smartphone. The six selected activities were *standing, sitting, laying down, walking, walking downstairs and upstairs.* Each user were asked to perform the activity twice : on the first attempt, the smartphone was fixed on the left side of the belt of the user and second attempt smartphone can placed by the user himself. Also, there was a 5 seconds break between each activities where user were asked to rest.

3.1. Exploratory Data Analysis on Inertial Signals

We selected 6 signals from each 3-axial accelerometer and 3-axial gyroscope, and plot them in order to see the pattern that each activity has. The signals were chosen because the plots show patterns clearly comparing to other plots. The figure 1, figure 2 and figure 3 are the body accelerometer plots. In case of walking plot, x, y and z axis shows similar pattern. Also, walking upstairs and walking downstairs have similar pattern like walking plot shows. However, sitting plot shows curved shape. In x-axis, the curve shows upward. On the contrary, in y-axis and z-axis, the curves are downward. In case of standing, there is a large fluctuation in between 40 and 60. We expect that the large fluctuation of accelerometer is the point of standing. Laying graphs show a decreasing shape in x-axis and y-axis and show an increasing shape in z-axis.

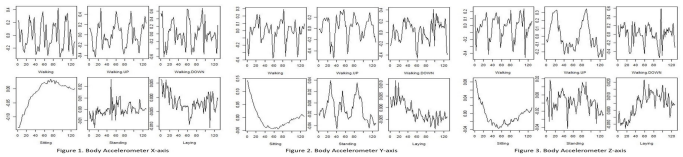


Figure 1. Body Accelerometer X-axis

Figure 2. Body Accelerometer Y-axis

Figure 3. Body Accelerometer Z-axis

Now, the following graphs are for 3-axial gyroscope. The gyroscope graphs show different aspect from accelerometer graphs. Especially, walking, walking upstaris and walking downstairs show different patterns. In case of walking upstais, there is an obvious change between 20 and for all 3-axial. But, only walking downstairs in y-axis shows the rising off between 80 and 100. Gyroscope sitting plot shows patterns, but it is different from accelerometer sitting plot. In x-axis, it is relatively consistent after 20. And, it is increasing in y-axis and decreasing in z-axis. But, the common thing is that 20 is the starting point to show pattern. In case of standing, the large dropping between 20 and 40 in x-axis is an action right before standing and the peak between 80 and 100 in z-axis is an action right after standing. And, the rising off between 40 and 60 in y-axis is the standing point. Laying plots show many of fluctuation, but the plot in x-axis shows contantly increase and the plot y-axis shows constantly decrease after 20. And, the plot in z-axis shows constant horizontal shape between 20 and 100, and then, it drops.

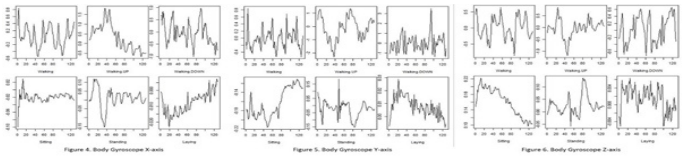


Figure 4. Body Gyroscope X-axis

Figure 5. Body Gyroscope Y-axis

Figure 6. Body Gyroscope Z-axis

3.2. Data Collection

Triaxial linear acceleration and velocity data was collected from the smartphone built in accelerometer and gyroscope which includes acceleration in x-axis , y-axis and z-axis. The signals were then sampled at 50 Hz. The signals were then passed to median filter and 3rd order low pass butter worth filter with a cut off frequency of 20 Hz to remove noise from collected data. The reason behind selecting 20 Hz as a cut off frequency because 99 % of human body motion energy is below 15 Hz.
The second preprocessing step was to separate gravitational and body components from acceleration signal. Another butter worth low-pass filter for this purpose. The corner frequency was selected as 0.3 Hz for a constant gravity signal since gravitational force is assumed to have low frequency components.

Additional two time signals i.e. jerks and acceleration by calculating Euclidean magnitude and time derivatives of triaxial signals.
[9]

Total features obtained after preprocessing steps :

tBodyAcc-XYZ
tGravityAcc-XYZ
tBodyAccJerk-XYZ
tBodyGyro-XYZ
tBodyGyroJerk-XYZ
tBodyAccMag
tGravityAccMag
tBodyAccJerkMag
tBodyGyroMag

tBodyGyroJerkMag
fBodyAcc-XYZ
fBodyAccJerk-XYZ
fBodyGyro-XYZ
fBodyAccMag
fBodyAccJerkMag
fBodyGyroMag
fBodyGyroJerkMag

3.3. Feature Mapping

To extract descriptive features for the four time series obtained in the previous section, window overlapping technique was applied.We subdivide the data set into smaller subsets and window them individually. We choose a window of 128 samples which corresponds to 1.28 seconds of accelerometer data for feature extraction.Also,to reduce information loss at the edges of the window, individual sets may overlap in time. For this purpose, features are computed with 64 samples overlapping, corresponding to 50% of overlap between consecutive windows. The reason behind selecting 1.28 seconds and 50% because of the following reasons:[1]

1) The cadence of an average person walking is within [90, 130] steps/min [2], i.e. a minimum of 1.5 steps/sec
2) At least a full walking cycle (two steps) is preferred on each window sample
3) People with slower cadence such as elderly and disabled should also benefit from this method. We supposed a minimum speed equal to 50% of average human cadence
4) Signals are also mapped in the frequency domain through a Fast Fourier Transform (FFT), optimized for power of two vectors (2.56sec 50Hz = 128cycles)

Thus, a total of 17 signals were obtained with this method,which are as follows:
mean(): Mean value
std(): Standard deviation
mad(): Median absolute deviation
max(): Largest value in array
min(): Smallest value in array
sma(): Signal magnitude area
energy(): Energy measure. Sum of the squares divided by the number of values.
iqr(): Interquartile range
entropy(): Signal entropy
arCoeff(): Autorregresion coefficients with Burg order equal to 4
correlation(): correlation coefficient between two signals
maxInds(): index of the frequency component with largest magnitude
meanFreq(): Weighted average of the frequency components to obtain a mean frequency
skewness(): skewness of the frequency domain signal
kurtosis(): kurtosis of the frequency domain signal
bandsEnergy(): Energy of a frequency interval within the 64 bins of the FFT of each window.
angle(): Angle between to vectors.

These 17 signals are important as they can improve the learning performance,including energy of different frequency bands, frequency skewness,and angle between vectors (e.g. mean body acceleration and y vector).

In nutshell, total **561** attributes were extracted to describe each activity window. [9], [10]

4. Experiment

Since we had two data sets: raw data i.e. inertial signals which consists of body-acc,gyro -acc and total-acc x,y,z signals with corresponding labels and other data set which has 561 features with corresponding labels. We conducted some experiments on Human Activity Recognition data set to compare prediction accuracy between classification and neural network approach using 561 features and raw inertial signal data.

4.1. Classification

Support Vector Machine:
A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. [11]

For classification purpose, we applied SVM with 10-fold Cross Validation procedure and Gaussian kernels on transformed feature data set.Along with SVM, we experimented with random forest, decision tree and SGD classifier.

4.2. Neural Network Approach

Neural networks have been a recent rage and most state art of the systems employ it.The dataset gives us an opportunity to compare feature extraction vs. raw data. Used 561 dimensional pre-processed data on a 3 hidden layer deep neural network.We used raw data for a 512 LSTM cell recursive neural network.Both networks are made using tensor flow

Deep Neural Network:
Neural Networks with more than a single hidden layer are usually called deep networks.In our current network we have used 3 hidden layers with 500, 200, 200 units respectively.The accuracy of the given network with 10 epochs after shuffling has been 94.27%.For optimizing this value we tried varying number of layers, number of hidden units and epochs.

Recurrent Neural Network:
As discussed before, the raw data is a 2.56s (128 frames when sampled at 50 Hz).The time dependency presents

an opportunity to fit a model with temporal dependency.Recurrent Neural Networks are networks with loops in them, allowing information to persist.It can be thought of as multiple copies of the same network, each passing a message to a successor.We are using LSTM cells to tackle fading gradient problem. We also implemented dropout to improve the performance.

5. Results

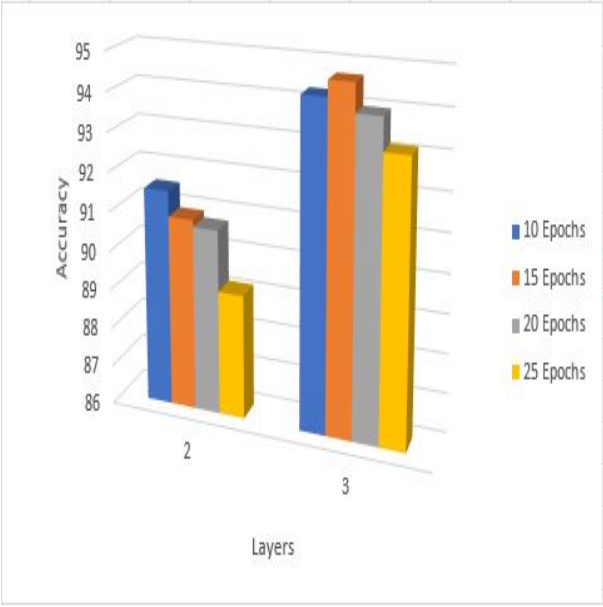
Using supervised classification method like SVM , we could see recognition accuracy of up to 93% without feature selection. We tried using dimension by selecting best features with Tree based classifier and L1 based classifier , but somehow accuracy went down. Results with SVM are as follows:

Accuracy w/o feature Selection	Time Taken w/o feature selection	Feature Selection	Accuracy w/ feature Selection	Time taken w/ feature selection
93 %	17.07 sec	L1 based classifier	92.96 %	3.43 sec
		Tree Based Classifier	89.2 %	2.46 sec

We evaluated the performance of the following neural networks - DNN and RNN using tensor flow. The summary results for our activity recognition experiments are as follows:

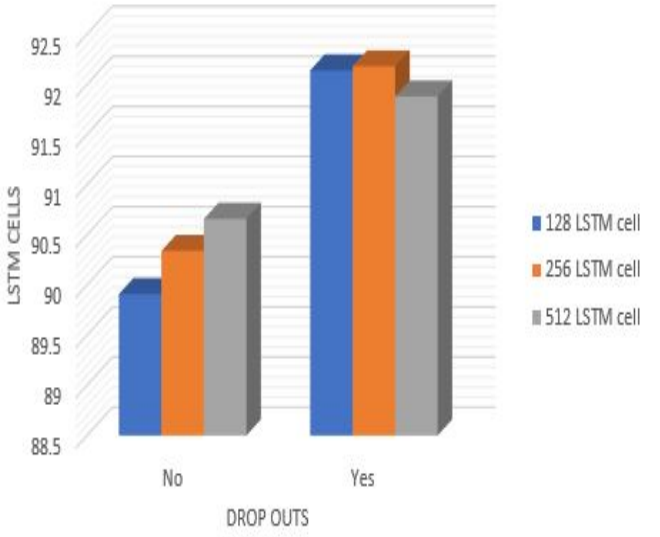
DNN:As we can see for 2 layers, accuracy decreased as we increased epochs. But for 3 layers, accuracy remained almost same.

Layers/Epochs	10	15	20	25
2	91.52	90.88	90.71	89.16
3	94.27	94.68	93.96	93.14



RNN: The results are as follows

Dropout/LSTM cells	128	256	512
No	89.91	90.34	90.66
Yes	92.14	92.18	91.88



6. Conclusion and Future Work

We ran experiments on the UCI human activity recognition dataset to compare performance evaluation between feature extraction with a deep neural network and raw data with a recurrent neural network. To our surprise the DNN performed better than the RNN on data that had time associated with it. The RNN consistently works better with dropout and DNN are prone to overfitting when trained for more than 15 epochs. Experiments showed that SVMs do not need shuffling of training data but neural networks do. Shuffling the data prior to training improved the performance by about 2%. Some activities like sitting and standing were consistently misclassified by all classifiers. Also with tree based feature selection and SVM with kernel='rbf', random-state='8' and decision-function-shape='ovo', we could see an accuracy of 95.14%
The current signal was only 2.56 seconds long and we believe the accuracy might improve with RNNs with longer signal length. Also, we have not used bidirectional recurrent neural networks which tend to perform better. As for feature extraction, comparison between hand crafted features and feature representation by a convolutional neural network might provide us more insight. Activity recognition in wearables can be seen as real time problem hence reducing the memory footprint and classification in real time also a future area to work on.

References

[1] Detection of posture and motion by accelerometry: A validation study in ambulatory monitoring, F. Foerster, M. Smeja, J. Fahrenberg.

Computers in Human Behavior, 15 (5) (1999), pp. 571583

[2] Activity recognition from user-annotated acceleration data, L. Bao, S. Intille. Pervasive computing, Lecture notes in computer scienceVol. 3001, , Springer (2004), pp. 117

[3] Activity recognition using cell phone accelerometers, J. Kwapisz, G. Weiss, S. Moore. SIGKDD Explorations, 12 (2) (2010), pp. 7482

[4] High accuracy human activity monitoring using neural network, A. Sharma, Y.-D. Lee, W.-Y. Chung. Proceedings of international conference on convergence and hybrid information technology (2008), pp. 430435

[5] Fusion of smartphone motion sensors for physical activity recognition, M. Shoaib, S. Bosch, O.D. Incel, H. Scholten, P.J.M. Havinga. Sensors, 14 (6) (2014), pp. 1014610176

[6] Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine, D. Anguita, A. Ghio, L. Oneto, X. Parra, J.L. Reyes-Ortiz. Proceedings of international conference on ambient assisted living and home care (IWAAL) (2012), pp. 216223.

[7] 3D gesture classification with convolutional neural networks, S. Duffner, S. Berlemont, G. Lefebvre, C. Garcia. Proceedings of international conference on acoustic, speech, and signal processing (ICASSP) (2014), pp. 54325436

[8] Binarized-BLSTM-RNN based Human Activity Recognition, M. Edel, E. Koppe. Indoor Positioning and Indoor Navigation (IPIN), 2016 International Conference.

[9] Human Activity Recognization using UCI Data:pdfs.semanticscholar.org-HAR.pdf

[10] A Study on Human Activity Recognition Using Accelerometer Data from Smartphones: www.sciencedirect.com

[11] SVM documents :doc.opencv website