Comparative Analysis of Human Activity Classification Techniques

Raghuveer Krishnamurthy Kanchibail, rkanchib@iu.edu

Supreeth Keragodu Suryaprakash, skeragod@indiana.edu

Shree Harsha Sridharamurthy, sridhash@iu.edu

I. Objectives and Significance

Today's needs of the consumer technology market are higher than ever before influencing positive breakthroughs in innovation, solving real world problems or simply provide means to upgrade to faster, robust and better user friendly devices. Specialized electronics such as wearable devices are one among the many technological advances gaining popularity.

Human activity recognition is one of the few fields of Computer Science which aims at providing a customized interfaces to the users for their interaction with technology. We observe significant difference in personalization with wearable devices which offer new features in addition to widely used smartphones, such as motion detection and activity monitoring. In this project, we focus our attention on the data science part of HAR (Human Activity Monitoring) in determining the most appropriate algorithm for classification by building a system based on approaches of machine learning, data mining and artificial intelligence.

In this paper, we present our analysis after investigating four such approaches highlighting their key concepts and evaluating them based on their performance in different situations. Per our analysis however, we found each of the algorithmic approaches themselves has its own strengths and drawbacks when it comes to support the mentioned factors. First, we build programs to create models out of the input dataset. Then we classify the test data to obtain the predicted output. Later, we compare the expected classes of test data with the predicted class values to obtain the accuracy for each of the programs, redoing the process for each dimensionality reduction approach.

II. Background

71% of 16 to 24-year-olds want wearable technology[1]. Twenty percent of American adults already own a wearable device and the adoption rate on par with tablets in 2012 is quickly expected to rise, according to PwC's Consumer Intelligence Series The Wearable Future report an extensive U.S. research project that surveyed 1,000 consumers, wearable technology influencers and business executives, as well as monitored social media chatter, to explore the technology's impact on society and business[2]. Human activity recognition - a field related to Human Computer Interaction aims at personalizing the way people interact with technology. Although mobile devices have been closely coupled with people, wearable devices take a significant step forward in detecting motion and monitoring human activity. Our objective is to determine the most appropriate algorithm for classification on the HAR dataset by building a classification system based on the approaches of machine learning, data mining and artificial intelligence.

Wearable devices, also called fashion electronics, is integration of electronic technologies with clothing. We use wearable devices due to their many uses such as for fitness or activity tracking, home automation, remote monitoring systems, assisted living technologies, and athletic tracking activities. They also form the internet of things as they connect devices with each other over multiple networks. Activity recognition is the process of recognizing the actions and goals of one or more agents based on observed data of the agents activities. 1 Applications of activity recognition include assisting the sick and disabled by homebased rehabilitation, security-related applications, plan recognition, goal recognition, intent recognition, behavior recognition, location estimation, and personalized support. Furthermore, activity recognition also associates with other domains such as medicine, human-computer interaction and sociology. There has been considerable amount of research accomplished through computer vision which has its applications in the field of surveillance by Professor Michael S. Ryoo of Indiana University, where mainly a set of motion features extracted from the streaming video signal is used to predict the ongoing activities in the video stream[3] . Real-time identification of an activity having a video as the input has been a problem oriented towards computer vision. However, we do see data mining approaches in classification of human activity as discussed below. Gu et al. have proposed an approach in recognizing human activity with pattern-based classification. They have used discriminative pattern recognition technique which describes significant changes between any two activity classes of data which further help recognize sequential, interleaved and concurrent activities in a unified solution. Gilbert et al. use 2D corners in both space and time. These are grouped spatially and temporally using a hierarchical process, with an increasing search area. At each stage of the hierarchy, the most distinctive and descriptive features are learned efficiently through data mining[4] J.C. Nascimento M. A. T. Figueiredo J. S. Marques have proposed an algorithm for seg- 2 menting and classifying human activities from video sequences of a shopping center[5] . Zhanqing Wu Xianping Tao and Hung Keng Pung propose an emerging patterns based approach to sequential, interleaved and concurrent activity recognition (epSICAR) where they exploit emerging patterns as powerful discriminators to differentiate activities[6] . Our project mainly focuses on building four classifiers and analyzing their efficiency. We will classify the data according to the four classification algorithms which in turn will be made less error prone. We then evaluate the algorithms against each other based on their correctness to infer the one which we will find to be the most appropriate in classifying the data.

Our motivation to use the HAR dataset is its dimensionality. It is one of the few datasets available at UCI machine learning repository which have more than 500 features. Our work also shows some of the techniques which can be applied to the algorithms to reduce the time required in building and classification of the models.

III. Methods

The HAR (Human Activity Recognition) dataset is available at the UCI machine learning repository. We felt the dataset "Human Activity Recognition Using Smartphones" at UCI Machine Learning repository to best match our interests due to its dimensionality and continuousness. It has been obtained with the help of 30 volunteers of unknown age. Data of six basic activities have been collected in this dataset and labelled as following classes:

1. Walking 2. Walking Upstairs

3. Walking Downstairs 4. Sitting

5. Standing 6. Sleeping

Volunteers performed the above activities with wearable devices. As per the description of the dataset, 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz were obtained using the accelerometer and gyroscope in the cellphone. The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain[7].

Each record in the main dataset has the following information:

- Total Acceleration from the accelerometer and the estimated body acceleration.
- Total Angular velocity from the gyroscope.
- A vector containing 561 features of time and frequency domain variables.
- Activity Label.

An identifier of the subject who carried out the experiment. All the 561 attributes can be seen in the features.txt file present in the dataset. Some more key points about the dataset are as below,

- Normalization is performed on features and each feature is bounded within [-1,1].
- Each row in the text file will be a feature vector.
- The units used for the accelerations for both total and body are 'g's which is nothing but the gravity of the earth 9.80665 m/seg2
- The gyroscope units are rad/seg. We have divided the whole dataset into training and testing set where 70% being the training data and the rest 30% being the test data

We used our knowledge gained through the curriculum in developing various classification models in order to evaluate our programs against the HAR(Human Activity Recognition) dataset. Additionally, we also measure the Support Vector Machine model based on the standard package scipy to compare against the others. Following are the classifiers evaluated:

- Naive Bayes classifier
- KNN classifier
- Decision Tree classifier
- SVM

Further, we also process the dataset with our dimensionality reduction algorithms and record the accuracy of these classifiers for these different datasets. These indicators for accuracy enable us in choosing better algorithm.

Before we move into the details of the above classifiers, we first take a look at the dimensionality reduction approaches. To deal with large number of features during classification, we reduce the features logically using either of the two types of approaches - Feature selection and feature extraction. Feature selection is the process of selecting only a subset of the actual dataset, ignoring those which may not influence the result by a great extent. On the other hand, feature extraction is the process of using the whole of data and then transforming it into reduced number of dimensions. We have used correlation filter and variance filters as the feature selection and PCA (Principal Component Analysis) as the strategy for feature extraction.

We now discuss the various methods of classification we have built and used in our project. The first model specified in the above list is the Naive Bayes classifier. Naive Bayes is a classic probabilistic classifier in machine learning. Our dataset not only has many features but is also continuous. Therefore, we use Gaussian Naive Bayes classifier which assumes that the continuous values are in a Gaussian distribution. The following equation depicts the Gaussian distribution:

$$p(x = v|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}}\, e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}}$$

where: x is the continuous variable indicating the distribution of the feature

μ_c is the mean of values in x

c is the class associated with x

σ_c^2 is the variance of values in x

v is the observed value

The Naive Bayes classification algorithm assumes that the joint probability of the features given the class is the same as the probability of the features given class, individually. Also, we assume $p(x = v|c)$ follows a normal (Gaussian) distribution.

The second model we analyze is the KNN. The k-nearest neighbors classification method is different from the others we have compared in the aspect of model building. This is a lazy (instance-based) learning algorithm - in other words, this does not build the classification model before test data is input to the system. Every data point input to this classifier triggers a distance calculation and subsequent top k similar features listing based on which the class is assigned. Although it is non parametric and no assumptions are made for this method of classification, we do observe a hit to the performance as the distance is calculated every time a new data point is input to the classifier. We use the most common euclidean distance formula to find the nearest neighbors.

$$\begin{aligned} \mathrm{d}(\mathbf{p}, \mathbf{q}) = \mathrm{d}(\mathbf{q}, \mathbf{p}) &= \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} \\ &= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}. \end{aligned}$$

where p = (p1, p2,..., pn) and q = (q1, q2,..., qn) are two points in Euclidean n-space distance (d) from p to q, or from q to p is given by the Pythagorean formula.

The next model under consideration is the Decision Tree classifier. This classifier is a classic predictive modeling technique which may be visually represented as a n-ary tree with a class tagged to every value of the corresponding feature. The tree is built by first zero-one normalizing the data and then discretizing it into 10 values. Each data point from the test dataset is then classified by traversing through the tree nodes. Information gain, which uses the entropy to assess the importance of the features themselves in classification. The formula for calculating the information gain is as below:

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

where S is the main set and A is the attribute under consideration

$$Entropy(S) \equiv \sum_{i=1}^{c} -p_i \log_2 p_i$$

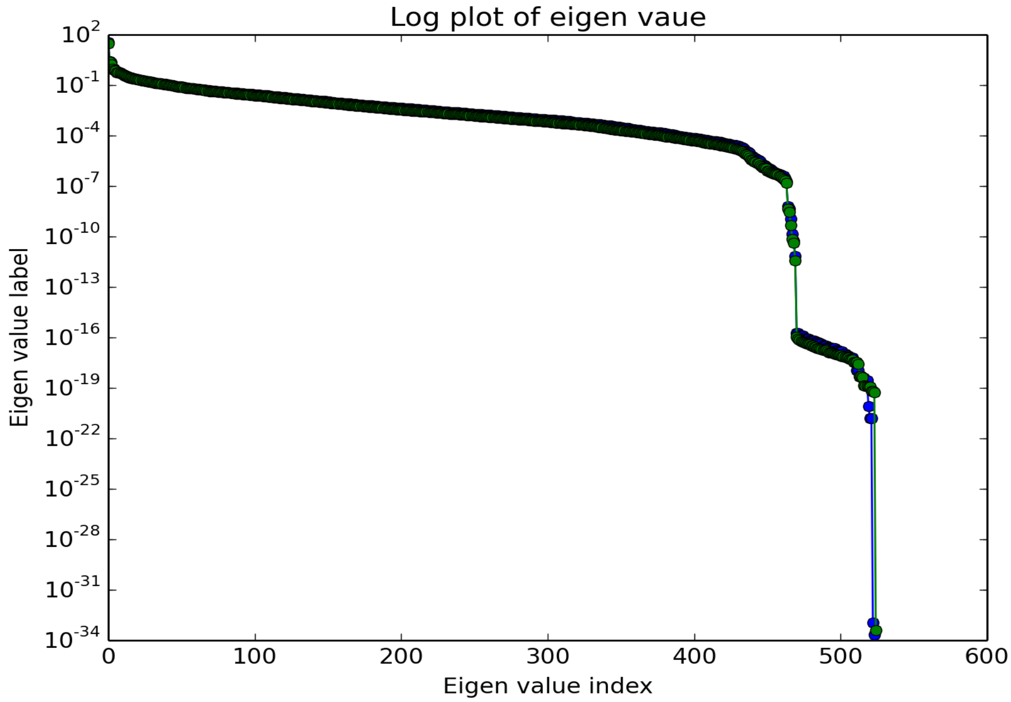
where p is the probability of the class c

Lastly, we analyze the SVM (Support Vector Machine) model. Support Vector Machines are non probabilistic linear classifiers. We made use of the the classifier built by Cornell University to classify the data. It usually classifies the data by making use of a discriminative hyperplane that distinguishes data into different classes. In layman terms, if a labeled training set is provided, the classifier outputs an optimal hyperplane which will categorize test examples.

IV. Results

All the analyses were performed for the below limits corresponding to each of the dimensionality reduction methods:

PCA: When we plotted the Eigen values against the Eigen vectors and observed a dip at 400 hence we reduced the number of features to 400 from the initial 516.



Correlation filter: The correlated features were eliminated based on the threshold of 0.98 - i.e., any features positively or negatively correlated above 0.98 were considered redundant and were eliminated.

Variance filter: The features having less than 5% variance were removed assuming little or no effect on the dataset classification.

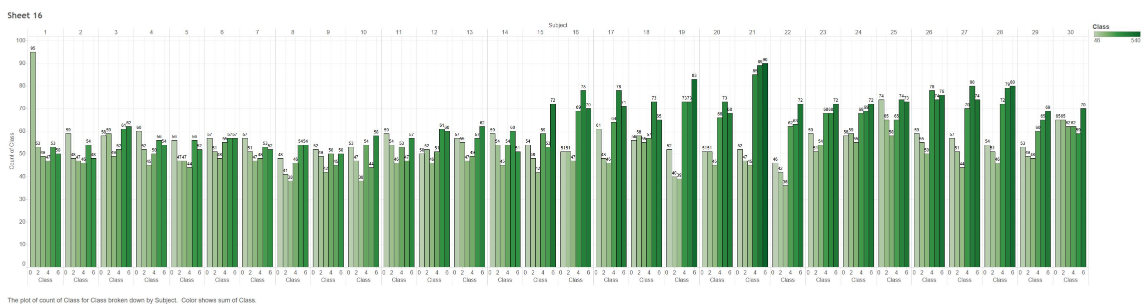
We compared and contrasted the above algorithms and found the accuracy of each of them as tabulated below:

Algorithm	PCA	Correlation Filter	Variance Filter	Without Reduction
KNN	0.16	90.21	90.43	90.73
Naive Bayes	25.16	63.42	78.76	57.86
Decision Tree	18.05	78.21	78.75	79.23
SVM	30.35	95.63	95.76	94.33

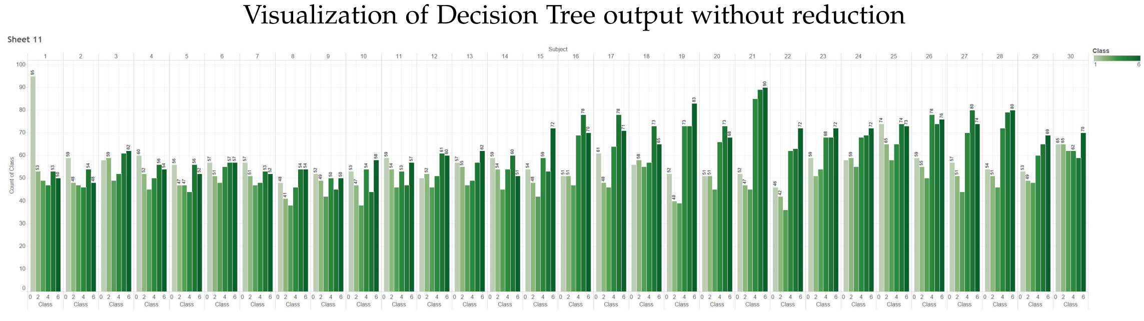
The predicted class values and the specific subject classes were combined to produce a comprehensive list of all the 10,000 datapoints and then we plotted a graph of subject v/s class where each subject's class counts were visualized. The higher the count of the class, more are the chances the subject performed the activity denoted by the class numbers. Visualizations are self-explanatory and can be found at the following box location: https://iu.app.box.com/files/0/f/7710213701/Images

The Dataset is in the below location: https://iu.app.box.com/files/0/f/7710170793/Code_%2B_All_Data_Files

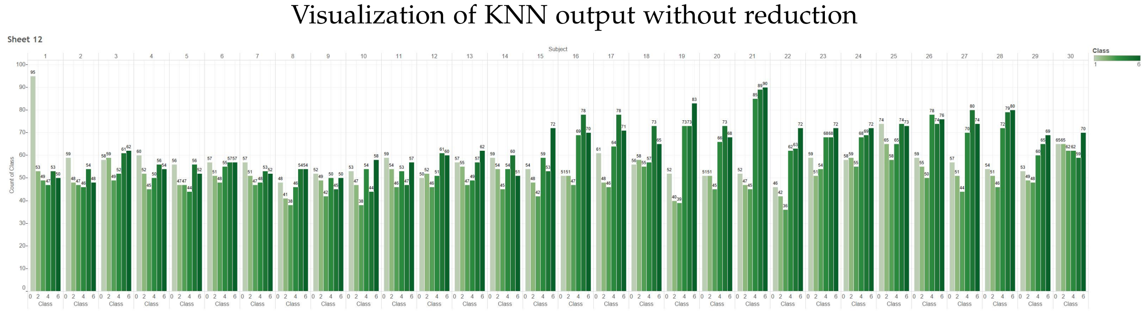
Visualized output for three algorithms are as follows:



Visualization of Decision Tree output without reduction



Visualization of KNN output without reduction



Visualization of Naive Bayes output without reduction

V. Conclusions

Overall, our project was successfully completed. With various techniques and algorithms applied such as the Naive Bayes, KNN, Decision Tree for classification and PCA, Correlation and variance based dimensionality reduction. However, PCA didn't work properly may be because of the fact that it completely created new dimensions. Overall it was a great experience working on this project. We would like to thank Dr. Predrag Radivojac and all the Associate Instrcutors of CSCI-B565 for their support during our tough times.

VI. Individual Tasks

Supreeth Keragodu Suryaprakash

This was an amazing project where I learnt many Data mining methodologies and techniques. To implement some of the major Data Mining algorithms we had studied theoretically was indeed amazing. I implemented Naive-Bayes, Correlation Filter to reduce dimensions and developed code to match the Subjects with Test and Train Data to get an overall figure on what the data says. I was bestowed with some amazing project mates who made this work possible.

Raghuveer Krishnamurthy Kanchibail

I really loved the project from day 1 because of its complexity. I learnt lot of Data Mining Techniques and Tools during the course of this project. I developed KNN, Covariance filter to reduce the dimensionality and visually analyzed the data using Tableau. My team mates had also put in similar amount of effort to make this project a success. It was amazing to have them in the same team.

Shree Harsha Sridharmurthy

I developed Decision Tree. The program I coded was way too complex for anyone's liking. I discretized the entire data to form a proper tree structure and to get the accuracy we had expected.

I also coded the PCA dimensionality reduction technique to reduce the complexity of the dataset. I had very good teammates who worked tirelessly to achieve the expected end product. We split the entire work into three parts, which didn't make this complex project too burdensome upon us.

1. By Victor Lipman. Forbes "71% Of 16-To-24-Year-Olds Want 'Wearable Tech.' Why Don't I Even Want To Wear A Watch?". September 22, 2014. Retrieved September 22, 2014.

2. Wearable Technology Future is Ripe for Growth - Most Notably among Millennials, Says PwC US. (n.d.). Retrieved March 03, 2016, from http://www.pwc.com/us/en/pressreleases/2014/wearable-technology-future.html

3. M. S. Ryoo,"Human Activity Prediction: Early Recognition of Ongoing Activities from Streaming Videos", International Conference on Computer Vision (ICCV), Barcelona, Spain, November 2011.

4. Gilbert A, Illingworth J, Bowden R. Action Recognition using Mined Hierarchical Compound Features. IEEE Trans Pattern Analysis and Machine Learning

5. Nascimento, J. C., Figueiredo, M. A., and Marques, J. S. (n.d.). On-Line Classification of Human Activities. Pattern Recognition and Image Analysis Lecture Notes in Computer Science, 444-451.

6. Tao Gu, Zhanqing Wu, Xianping Tao, Hung Keng Pung, and Jian Lu. epSICAR: An Emerging Patterns based Approach to Sequential, Interleaved and Concurrent Activity Recognition. In Proc. of the 7th Annual IEEE International Conference on Pervasive Computing and Communications (Percom '09), Galveston, Texas, March 913, 2009.

7. UCI website for dataset: http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphone

8. Dimensionality reduction https://en.wikipedia.org/wiki/Dimensionality_reduction

9. https://en.wikipedia.org/wiki/Naive_Bayes_classifier

10. https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

11. http://machinelearningmastery.com/

12. https://en.wikipedia.org/wiki/Euclidean_distance

13. https://www.cs.cornell.edu/people/tj/svm_light/ svm_multiclass.html

14. http://personal.disco.unimib.it/Vanneschi/McGrawHill_-_Machine_Learning_-Tom_Mitchell.pdf

15. http://connor-johnson.com/2014/04/02/computing-principal-components-in-python/