Supervised Learning on Activity Recognition Problem, and Study of Dimensionality Reduction on Different Classification Models

Mohsen Nabian
Northeastern University
360 Huntington Avenue, Boston ,MA 02445
nabian.m@husky.neu.edu

Abstract

Activity Recognition (AR) algorithms are machine learning algorithms developed for cellphones and smart wearables to recognize real-time human activities such as walking, standing, siting, running and biking. This paper applied several well-known supervised and unsupervised learning models including 'Logistic Regression', 'Support Vector Machine', 'K-nearest neighbors' 'Artificial Neural Network', 'Naive Base', Decision Tree, 'Random Forest' and 'Principal Component Analysis' on an AR dataset[1] with 12 discrete classes (labels). The first four algorithms are developed from scratch by the author to achieve deeper understanding of those ML models and to examine the effect of hyper parameters on the accuracy and running time of these algorithms. Moreover, the performance of these different classifiers are compared on the AR dataset. Although non-linear classifiers such as Neural Network, KNN and Random Forest have shown to provide very accurate predictions, considering the accuracy and running time, it is shown that 'Random Forest' outperforms other models, therefore it is a very suitable to be used as a real time activity recognition software. Also the effectiveness of PCA preprocessing has been investigated on the mentioned learning algorithms. Results of PCA investigation provided the following conclusions: Dim. reduction using PCA must be done for Neural Network since it improves the prediction accuracy of Neural Network models. It is also highly recommended to use PCA for KNN and SVM. However, PCA should not be implemented for Random Forest or Decision Tree since PCA increases their running time and decreases the accuracy.

1. Introduction

Wearable-based activity recognition (AR) systems are typically built to recognize a predefined set of common activities such as sitting, walking, and running [1]. AR has captured the attraction of computer science communities due to its capabilities on providing supporting personalized information and its many applications in Human-Computer interface, Medicine, Insurance Businesses, Sports and Sociology. AR is currently being used by some Companies like Nike for a limited number of Activities. There are wide variety of studies on AR systems, from Data provision research [4] to introducing semi-supervised nonlinear models for AR[3].For instance in a study by Nguyen [2], he has added semantic relationships between activity definitions to compensate unbalanced training data for activity recognition. However, there is no literature found to comprehensively investigate the performance of well-known classification models on the Activity Recognition problem. In this paper, Logistic Regression, K-Nearest Neighbor, Support Vector Machine and Neural Network have been developed by the author to achieve deep and detailed understanding of the implementation of these algorithm as well as to investigate the effect of hyper parameters on the performance of these models for AR dataset. Other well-known classifiers including Random Forest and Decision Tree were implemented to provide a comprehensive comparison study on the prediction accuracy of wide variety of available Supervised Machine Learning algorithms. Also dimensionality reduction using PCA is done as a preprocessing step for all these classifiers and the effect of PCA is investigated on all those models.

Figure.1 Activity Recognition Process Pipeline

2. Problem Statement

Research studies in Activity recognitions as mentioned above, have been focused on developing novel models for rare activities and for unbalanced training data and novel methods for data collections [1-5]. However, there is gap for comprehensive study of the performance of the conventional Machine Learning Classifiers on Activity Recognition. In this rese

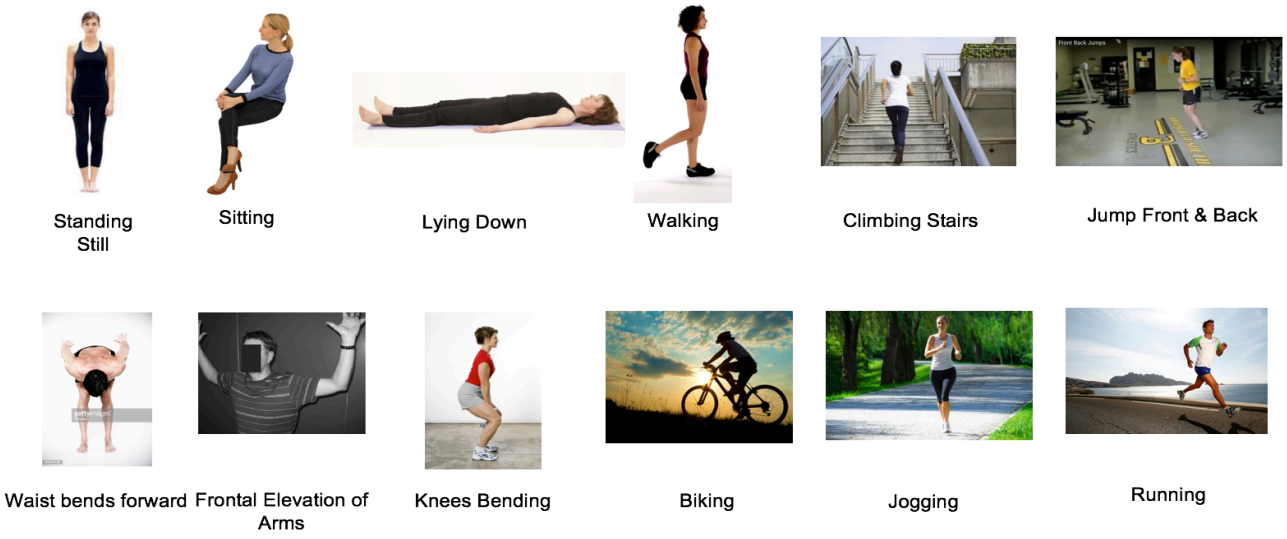

Figure.2 Type of Activities (Labels) of the Dataset

arch a comprehensive study is performed over several well-known supervised learning algorithms. Moreover, a study is done on the dimensionality reduction using PCA as preprocessing of the learning process.

3. Methodology

A large AR dataset provided by UCI is retrieved and then cleaned. The raw data contains 1.2 million instances and after the cleaning, Dataset had 346 K instances almost uniformly over 12 activities. The data is then normalized by each column and mapped to [0,1]. Next process is to shuffle the data to increase accuracy in the model. Then 80% of the data points is selected for training and 20% is selected for testing. 4 Models (Logistic Regression, KNN, SVM and Neural Network) were developed from scratch and their parameters were discussed. Moreover, other conventional Classifiers were used for training and their accuracy were tested using the test data. It was shown that Naive Base was the fastest in overall and Random Forest

4. Theory

In this paper, the following classifiers and algorithms were developed by the author: 1) Ridge Regression, 2) Support Vector Machine 3) K-nearest neighbor 4) Artificial Neural Network and 5) Principal Component Analysis. Table1 summarizes the hypothesis, approach and cost functions of all these methods. Since the problem is multi-class classification, softmax algorithm is used for Ridge-Regression and softmax activation function is used for the last layer of neural network.

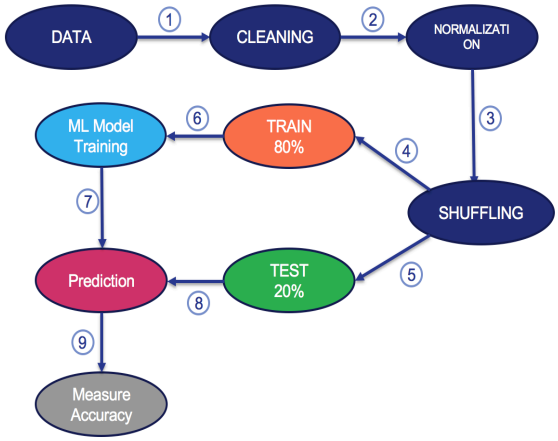

Figure.3 Machine Learning Process road map

Furthermore, using the method of 'One-vs-all' SVM coefficients were classified 12 times to address the multiclass classification problem. Notice that Ridge Regression is a linear classifier, while SVM, KNN and ANN are nonlinear and since the activity measures such as velocity and accelerations are non-linear with respect to the provided the highest accuracy. Figure 2 provides the road map of this machine learning research.

activities, it is expected to have better prediction for non-linear classifiers. The hyper parameter for each classifier should be set to a value that maximizes the accuracy of the classifiers over the test data. In the experiment section, hyper parameters for each classification will be examined.

Algorithm	Hypothesis	Approach	Cost Function	hyper Param.
Ridge Regression	$p(y=1\|x) = \frac{1}{1+e^{-w^T x}}$	model should maximize likelihood of data and penalize magnitudes of w	$J = \sum \log\left(1+e^{w^T x}\right) - yxw + \frac{\lambda}{2}\|w\|_2^2$	λ
SVM	$y = sgn(w^T x + b)$	model should maximize geometric margin	$J = \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{N}\xi_i$ s.t. $y^{(i)}\left(w^T x^{(i)} + b\right) > 1 - \xi_i$	*C*
KNN	majority vote of its k-neighbors	model is majority vote of its k-neighbors	No cost function	*K*
ANN	$y = h(x) = \frac{1}{1+e^{-z}}$ where z is vector of values of the last layer of the neural network.	model should minimize error and penalize magnitudes of w	$J = \frac{1}{N}\sum_{i=1}^{N}\left\|h\left(x^{(i)}\right) - y^{(i)}\right\|_2^2 + \frac{\lambda}{2}\sum_{l}^{n-1}\left\|w^{(l)}\right\|_2^2$	*H*

Table.1 Summary of the mathematical background of classification models

PCA convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Eliminating components that contain low variation, will decrease the dimension of the observation without losing significant amount of information. Dimensionality reduction may cause significant computational expenses for learning algorithms.

Dimensionally reduced data is obtained by the following:

$$y^{(i)} = u^T(x^{(i)} - \bar{x})$$

where $\bar{x}$ and u are calculated as:

$$\bar{x} = \frac{1}{N}\sum_{i}^{N} x^{(i)}$$

$$x = u_x \sum_x V_x^T$$

$$u = u_x(:, 1:d)$$

and d is the reduced dimension of the data.

5. Experiment

5.1 Data Description

The dataset comprises body motion and vital signs recordings for ten volunteers of diverse profile while performing several physical activities. Sensors placed on the subject's chest, right wrist and left ankle are used to measure the motion experienced by diverse body parts, namely, acceleration, rate of turn and magnetic field orientation. The sensor positioned on the chest also provides 2-lead ECG measurements, which can be potentially used for basic heart monitoring, checking for various arrhythmias or looking at the effects of exercise on the ECG. Figure 4 indicates where sensors are located and which information they convey:

The activities selected, are the most common in people's daily life. The full list of activities is depicted in Figure 2.

in average, there are about 30000 data points allocated for each activity except the last activity (jump front & back) which has 10000 data point, which is still a very reasonable amount of data. In overall, the data set is quite well prepared for model fitting.

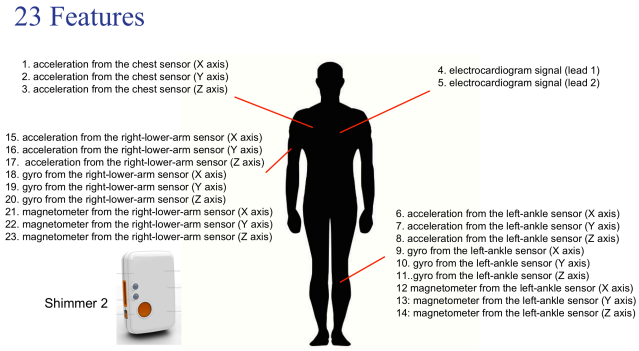

Figure.4 Dataset Features for the Activity Recognition (AR) problem

5.2 Experimental Results

The performance of different learning algorithms on the same AR dataset is explored and compared. Moreover, the effect of hyper parameters on the accuracy of classifiers are also investigated. Finally, it is shown whether PCA as preprocessing step can decrease the running time of each algorithm while maintaining the accuracy.

5.2.1 Performance Measurement of Supervised Algorithms on AR Dataset:

Here, 7 supervised learning classifiers are trained for the same AR dataset and their prediction accuracy for the Test data were measured. Figure 6 demonstrates a comparison of accuracies of the classifiers.

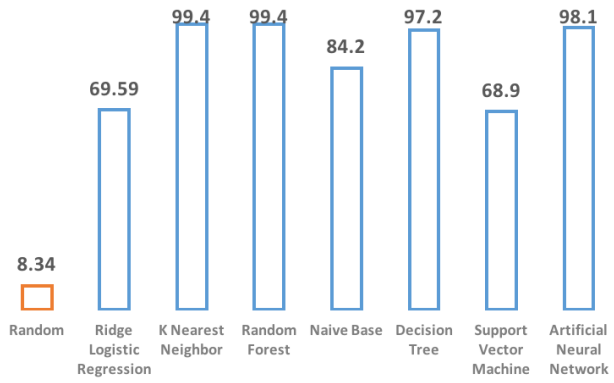

Figure.5 Prediction Accuracy (%) of classification models over AR dataset

It is show in Figure 6 that KNN and Random Forest provide very accurate prediction (>99%). Artificial Neural Network and Decision Tree performed very well. (>98%) However, Naive Base and Ridge Logistic Regression as well as SVM demonstrate relatively poor performance. Naive Base assumes independency between features of the observations. This assumption is not descent in the case of Activity Recognition Dataset since there are some significant correlations between the values of the sensors attached on the body. (features). As a result, Naive Base does not provide very high accuracy. Moreover, due to the non-linearity of the data for different activities, it is expected to have poor result for linear classifiers. Which is the case for Ridge Logistic Regression and SVM with no kernel which are both linear classifiers.

Besides the accuracy, running time of classifiers are important factor for the performance of the machine learning classifiers. The importance of the running time will become tangible once the dimension of the data becomes large. The running time of classifiers are provided in the table 2.

Classification Model	Time (sec)
Ridge Logistic Regression	64.3
K-nearest neighbor	149.2
Random Forest	19
Naive Base	0.6
Decision Tree	15.2
Support Vector Machine	131.2
Artificial Neural Network	2052

Table.2 Running Time for training and testing of classification models over AR dataset

Although Artificial Neural Network (ANN) provided very accurate prediction, it also has the highest running time, while Random Forest and Decision Tree are very fast. In summary, considering both accuracy and running time, Random Forest could be selected as the most suitable classifier for the AR problem.

5.2.2 Effect of hyper-parameters on the classifiers:

Hyper parameters as opposed to learning parameters are parameters that are defined in the algorithm and doesn't change during the learning process. Thus, it is expected to have different classification results with different prior of hyper parameters. Therefore, it is essential to investigate the optimum hyper parameter for each classification problem for the given dataset.

Here, Ridge Logistic Regression, KNN, SVM and Artificial Neural Network are investigated for the optimum corresponding hyper parameters.

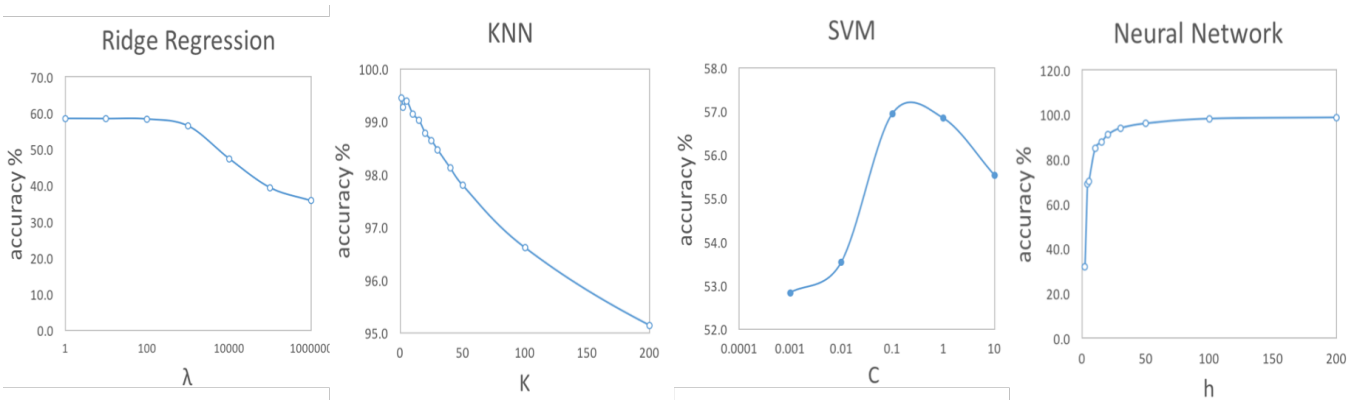

Figure.6 Effect of hyper parameter value on the prediction accuracy of classification models; from left to right: Ridge Logistic Regression, KNN, SVM, Neural Network

\r Ridge Regression, λ is the hyper parameters which is the measure on how much we are willing to penalize the magnitude of the model coefficients. large λ may prevent over-fitting of the model to the training data. in figure 6, by increasing λ until around 100 there is no effect found on the prediction accuracy, however increasing λ above 100 gradually decreases the model accuracy.

In KNN, K is the hyper parameter which specifies how many observations to be considered as neighbors for taking the majority vote. Figure 6 shows that increasing K in the case of AR dataset will decrease the accuracy.

In SVM, C penalizes the observations that are not in the functional margin of 1. In other words, small Cs allows for some misclassification of data while large C strictly penalize misclassifications. C is particularly useful when the data is not linearly separable. According to figure 6, the optimum C is between 0.1 to 1.That is certainly a hint that the data is not linearly separable and requires relaxation. However, high relaxation may negatively affect the accuracy.

In Artificial Neural Network, H is the hyper parameters which is the number of neurons in the middle layer of the network. As H increases, the neural network model will have more parameters to learn thus useful for complicated models. Based on figure 6, low values of H (1 : 50) is not enough to encompass the complexity of the data. However, for H between 50 to 200 there is insignificant change in the accuracy.

5.2.3 Effect of Dimensionality Reduction on the performance of classification models:

PCA is the technique used for the dimensionality reduction of the dataset. This section intends to investigate how dimensionality reduction will improve the running time and to explore whether the classification accuracy will be affected or not. In PCA, we add principal directions until we capture ρ fraction of total variance in the original dataset and ρ is typically among these values: ρ={0.99,0.98,0.95}

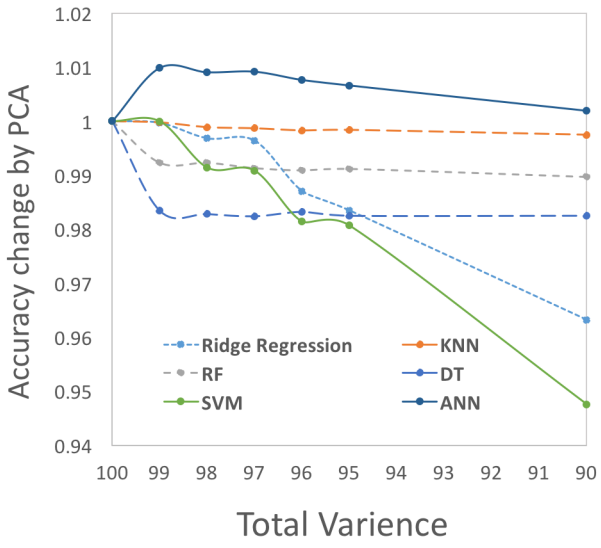

Figure.7 Effect of Dimensionality reduction on the prediction accuracy of classifiers. the values of prediction accuracies are normalized with values in figure 6.

Here, the running time and accuracy of the classification models that were discussed earlier, were examined under the range of ρ=[0.99,0.90].figure 7 demonstrates how classification accuracy may alter by reducing dimensions of the data using PCA and Figure 8 projects the effect of dimensionality reduction on the running time of the learning and prediction process of supervised learning models.

It is very interesting that **neural network accuracy improves** around 1% although we have reduced the dimensions of the data using PCA. This improvement still exists for ρ=0.90. This phenomenon was discovered by 'Junita Mohammad-Saleh'[8] at 2008. Apparently, Neural Network works more efficiently on dataset that have their features orthogonal or independent. Also, by looking at figure 8, the running time of Neural Network drastically decreases by PCA to 20% of time required using the original data. The result should put PCA a must

preprocessing step for Neural Network classification. It is also important to mention that Neural Network is the **only model** that has its prediction accuracy improved by dimensionality reduction using PCA.

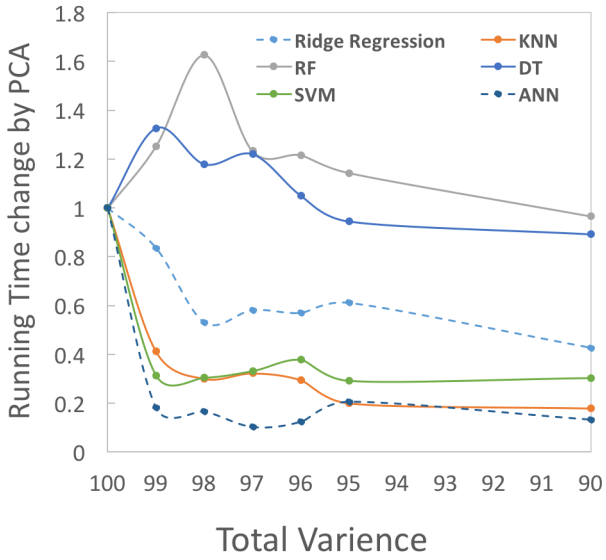

Figure.8 Effect of dimensionality reduction on the running time of classifiers. the values of running time are normalized with values in table 2.

The prediction accuracy of KNN and Random Forest drop only less than 1% for a wide range of ρ. According to Figure 8, there is around 60% running time drop for KNN. That implies that dimensionality reduction using PCA is **highly recommended for KNN**. However, PCA is apparently **increasing the running time** of Random Forest which makes the PCA preprocessing useless and not recommended for Random Forest.
For SVM ρ=0.99 is highly recommended since the prediction accuracy doesn't change significantly while the running time may drop over 60%.
For decision tree, PCA is not recommended since it has increased the running time.

6.Summary and Discussion:

In this paper Activity Recognition problem was investigated using several well-known classification models in machine learning. Due to non-linearity nature of the AR dataset, non-linear classifiers outperforms linear models such as Ridge Regression and linear SVM. Models including KNN, Neural Network and Random Forest provided more than 99% prediction accuracy for this problem, however, by considering running time, random forest indicated the highest performance over other models. Moreover, the effect of hyper parameters on the prediction accuracy of different models were investigated. and Lastly, dimensionality reduction were applied as a preprocessing step to these mention models. By investigation of the running time and prediction accuracy of different models after PCA the following recommendation were provided:

a) PCA must be done before Neural Network problem which will improve the accuracy while immensely decrease the running time.

b) PCA with ρ=0.99 is highly recommended for KNN and SVM with linear kernel.

c) PCA should NOT be used for Random Forest or Decision Tree as shown in the paper. Since, PCA will increase the running time as well as decrease the accuracy of the model.

7. References:

1. Lara, O. D., and Labrador, M. A. A survey on human activity recognition using wearable sensors. Communications Surveys & Tutorials, IEEE 15, 3 (2013), 1192–1209.
2. Nguyen, L. T., Zeng, M., Tague, P., Zhang, J. (2015). Recognizing New Activities with Limited Training Data. In IEEE International Symposium on Wearable Computers (ISWC).
3. Stikic, M., Van Laerhoven, K., and Schiele, B. Exploring semi-supervised and active learning for activity recognition. In IEEE International Symposium on Wearable Computers, IEEE (2008), 81
4. Stikic, M., and Schiele, B. Activity recognition from sparsely labeled data using multi-instance learning. In Location and Context Awareness. Springer, 2009
5. Nguyen, L. T., Zeng, M., Tague, P.: Supervised activity discovery. In International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, ACM (2015).
6. Banos, O., Garcia, R., Holgado, J. A., Damas, M., Pomares, H., Rojas, I., Saez, A., Villalonga, C. mHealthDroid: a novel framework for agile development of mobile health applications. Proceedings of the 6th International Work-conference on Ambient Assisted Living an Active Ageing (IWAAL 2014), Belfast, Northern Ireland, December 2-5, (2014).
7. Banos, O., Villalonga, C., Garcia, R., Saez, A., Damas, M., Holgado, J. A., Lee, S., Pomares, H., Rojas, I. Design, implementation and validation of a novel open framework for agile development of mobile health applications. BioMedical Engineering OnLine, vol. 14, no. S2:S6, pp. 1-20 (2015).
8. Mohamad-Saleh, Junita, and Brian S. Hoyle. "Improved neural network performance using principal component analysis on Matlab." *International journal of the computer, the internet and Management* 16.2 (2008): 1-8.

9. DATA Source: The following link provides access to the AR dataset.

https://www.dropbox.com/sh/zfv7qjndjkqc7kn/AADpRnHJ7OiWo6Y2DuPDL0G4a?dl=0