International Journal of Computer Applications (0975 - 8887)
Volume 124 - No.12, August 2015

Statistical based Neural Network in Human Activity Recognition

Mohammad Pivezhandi
Department of Electrical and Computer Engineering
Shahid Beheshti University
Tehran, Iran

Baback Mazloom-Nezhad Maybodi
Department of Electrical and Computer Engineering
Shahid Beheshti University
Tehran, Iran

ABSTRACT

This research proposed an efficient method in classification of Human Activity Recognition tasks. The evaluated tuned models show higher than 99 percent mean accuracy and gain more training and testing accuracy in comparison to previous studies. Dimensionally reduction have been introduced based on P-value evaluation in feature space. Finally a hybrid model that compressed statistically in feature space alongside with Neural Network architecture have been proposed. The final model could be used as best architecture of hardware implementation in gesture recognition applications.

Keywords

Dimensionality reduction, Human Activity Recognition, Neural network, P-value extension, Statistical Analysis

1. INTRODUCTION

Human Activity Recognition (HAR) has been a key research area in last 8 years and has broad range of applications in smart human activity recognition. Automatically detection of human activities is interesting field of science. Various sensory methods like image recognition, GPS, accelerometer, gyroscope and etc have been proposed for detection of human daily activity recognition. Checking of Athletes incorrect activity techniques in weight lifting exercises, control of elder people status, energy consumption in weight loss programs and etc are some applications of HAR.
Yuting Zhang et.all in [15] proposed a continuous functional activity monitoring device based on decision trees that identifies daily life different types of activities in application of detection strokes of Parkinson's disease. In [13] Decision Tree have been used along side with Naive Bayes classifier, as preprocessing stage in human activity recognition problems. Jun Nishimure et.all proposed a novel Haar-like method that have been used for frequency contents extraction out of temporal signal in classifying the speech/nonspeech human activity recognition [11]. Chernbumroong et.all made a comparison between Neural Network and decision tree method with different feature sets [1]. Finally in [5] a detector of daily activities based on decision trees, multilayer perceptron and a combination of statistical and machine learning called hybrid model have been proposed.
There are several ways to earn primary data sets in HAR, like sensors or using image processing skills in camera recording. Video processing needs professional and costly prerequisite equipment and sufficient post processing skills. Sensor covered clothes causes simultaneously position recognition with high coverage of feature space in a low costly way.
So well trained model could be a good classifier of the desired position in human activities, the main concern in this research is to propose a best classifier model based on dataset of [14].
Here different Tree based statistical and multi layered neural network machine learning model and a combination of this two learning methods have been used to classify HAR. Finally the output of training and testing data in different models compared with each other. the feature selection procedure have been illustrated on best proposed model in the case of removing the unnecessary variables. By applying statistical procedure like P-value significant feature extraction important features have been applied to the multi layered neural network model.
All parts of this research organized as below: section 2 is about features in dataset. Section 3 introduces all of models. Section 4 introduces all of models. Section 5 concludes with the results.

2. HUMAN ACTIVITY RECOGNITION CASE STUDY

In the data set of this research for increasing the quality of study, test constructed with 6 number of patients and 4 sensors that sensed 4 different positions (classes). Position of sensors are shown in Figure 1. Class A shows exactly correct execution; class B shows how throwing elbows to the front; class C is about lifting dumbbell halfway; lowering dumbbell halfway shows class D; and finally class E shows throwing the hips to the front [6]. Class A corresponded to specified execution of exercise and other 4 classes showed mistakes in execution of exercise. All patients were between the ages of 20-28 and surveillance had been done by experienced weight lifter[9].
The first step in statistical analysis of pattern recognition tasks is cleaning the dataset from unnecessary and not assigned data. This procedure causes elimination of unnecessary features in model and effective training process. all of steps have been done in R and finally the tidy data consists of 52 features, the frequency of data is shown in Figure 2.

3. MODEL SELECTION

The goal of this study is to suggest classification models with best train and test accuracy.

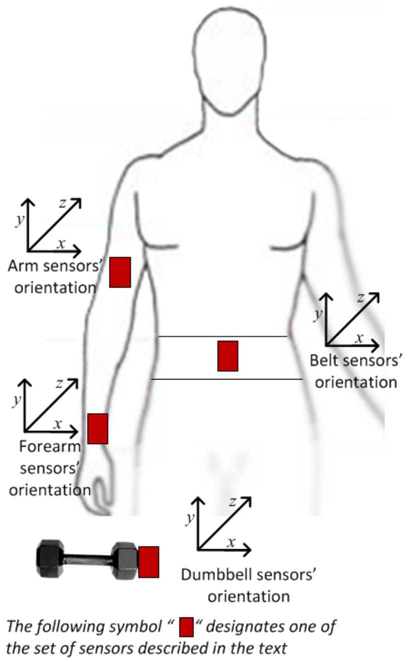

Fig. 1. Situation of evaluated sensors[14]

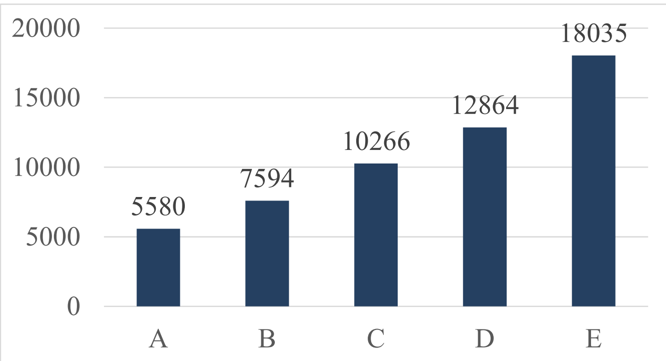

Fig. 2. Frequency of classes.

3.1 Using Decision Trees for Pattern Recognition Problems

Decision trees are hierarchical models that separate input space into K independent class and is drawn as upside down, because the place of leaves are in the bottom of the trees. Points along the tree where separation in predictor space occurred called internal nodes, so decision trees uses top-down greedy approach and have been known as recursive binary splitting, this is greedy because best split is at particular step of tree building process.
In classification trees, prediction comes by most commonly occurring classes of training observation in the region that it has been related. This kind of problems should use a mechanism for making binary splits and one of this measurements is Gini index that defined by Equation 1.
The Gini index takes a small value if all values of $\hat{p}_{mk}$ get close to the zero or one, for this reason the Gini index referred to measurement of node purity[9].

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}). \quad (1)$$

Another alternative to the Gini index is cross entropy that have been shown as Equation 2.

$$D = -\sum_{k=1}^{K} \hat{p}_{mk} \, log \, \hat{p}_{mk} \quad (2)$$

Decision Trees are so interpretable but have lower accuracy in some of applications versus traditional methods. Most of the people believes decision trees are closer to human body decision making procedure rather than other methods.
First Decision tree based model is bootstrap aggregating that simply called Bagging; and is the basic structure of other decision tree methods. Using Bootstrap by repeatedly sampling from training set causes reduction in variance of statistical learning methods, this mechanism have been invented by Bradley Efron in 1979[4]. Bootstrap is a flexible and powerful statistical tools to quantify the uncertainty associated with a known estimator or in statistical learning method. In this kind of classification trees the class will be predict by taking the majority vote so the most commonly occurring class among the B prediction shows the final prediction.
Random forest provides an improvement over bagged trees, each time a split has been considered on a tree, a random selection of m predictors are chosen as split candidates from the full set of p predictors[7]. Typically best number of evaluated features at each split could be assigned by the value $m \approx \sqrt{p}$. Random forest could be the best and fastest decision tree model in big classification problems like pattern recognition by assigning appropriate tuning parameters[10].
Another decision tree model is Boosting that is an approach of bagging model. As illustrated in bagging approach, multiple trees that come from bootstrap mechanism are independent, Boosting combination is implemented by adding trees sequentially[12].
Out of Bag error is a way of error estimations of test results in bagged models. The key idea here is that in bootstrap, sampling occurs on two-thirds of observations and another one-thirds that not used in fitting, could be referred as out of bag samples. The prediction on ith observation using each of trees yield B/3 of predictions, which B implies on number of bootstrapped trees[3].

3.2 Learning in MLP Neural Network Machine Learning

Artificial Neural Networks (ANN) are imitation of natural neural system in living things body and are powerful, common method in pattern recognition problems.
Traditional training procedure of ANN's consists of three parts: first the weights initialization step, then feedforward step that constructs product of multiplication in inputs to the weights in each of layer with algebraic mechanism, finally error correction step takes optimizing methods for reinitializing of each weight in proposed network. Three layer architecture of proposed model have been shown in Figure 3.
In the feedforward step all of the values in the nodes should be multiplied with weights, Equation 3 have been called regularized

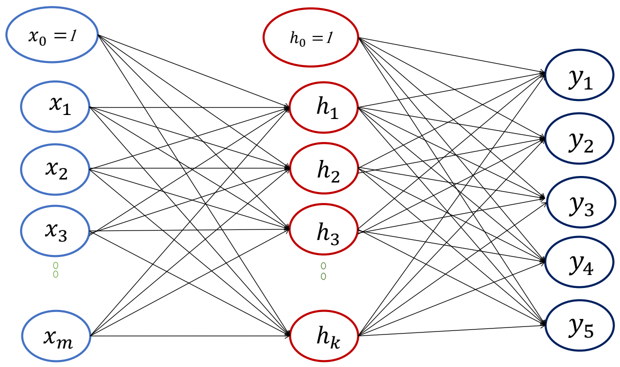

Fig. 3. Architecture of Proposed Artificial Neural Network.

logistic regression cost function. The regularization in cost function causes shrinkage of weights, by avoiding overfitting in Neural Network learning procedure.

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{k=1}^{K} y_k^{(i)} log\left(h_\theta(x)\right)_k + (1-y_k^{(i)}) log\left(1-h_\theta(x)\right)_k\right] + \frac{\lambda}{2m}\sum_{l=1}^{L-1}\sum_{i=1}^{S_i}\sum_{j=1}^{S_i+1}\left(\theta_{ij}^{(l)}\right) \quad (3)$$

$$h_\theta(x) \in R^k, \left(h_\theta(x)\right)_i = i^{th} output$$

In the Equation 3, K implies on number of output patterns and m is number of observations. $\theta_{ij}^{(l)}$ shows the weights that connected input nodes(i) to the output(j) in the layer(l) of the proposed feed-forward architecture.

4. SIMULATION AND RESULTS

4.1 Statistical model selection

Bagging classification model have been implemented via adabag package in R [2]. Another model is Random forest and its equivalent package in R is a powerful package in classification problems however there should be some consideration in specifying tuning parameters like number of variables that sampled as candidate at each split[10]. The typical value in number of splits for this study is square root of total evaluated features and as shown in Figure 4 optimum value is near the 7th splits.
Generalized Boosted Model (gbm) is another package in R that implements Freund and Schapire's adaboost algorithm[12]. In boosting models three tuning parameters have to be assigned: number of trees, shrinkage value and interaction values. In contrary to bagging and Random Forest models the big number of trees in Boosting model could cause overfitting[8]. learning rate in Boosting model known as shrinkage value(λ), this mechanism controls the rate that model could learn, this value depends on the case study. Number of splits(d) in Boosting model controls the complexity of the boosted ensemble and generally d shows the interaction depth. Figure 5 shows the comparison between three introduced classification trees. All the evaluated trees are tuned via tuning parameters with minimum test error.
As shown in Figure 5 two Random Forest and Boosting model have best response according to test error evaluation in 500 number of trees, in Figure 6 the Random Forest algorithm results, shows fastest testing Error convergence in small number of trees.

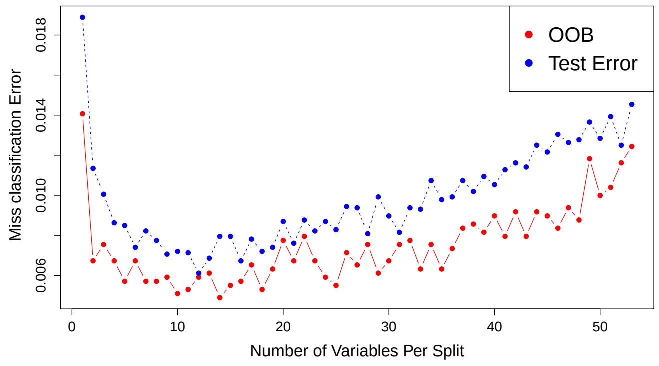

Fig. 4. Random Forest number of variable per each split base on Misclassification Error.

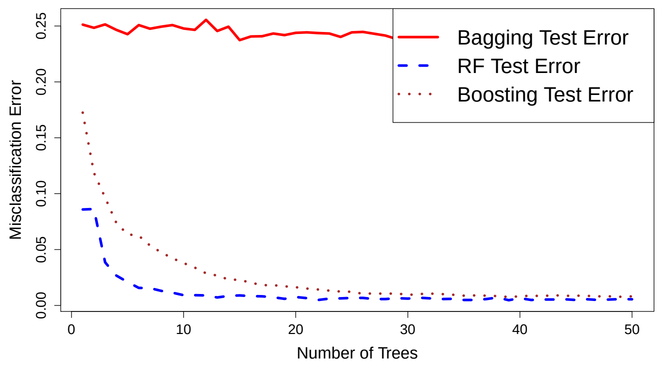

Fig. 5. Testing Misclassification Error of decision tree models.

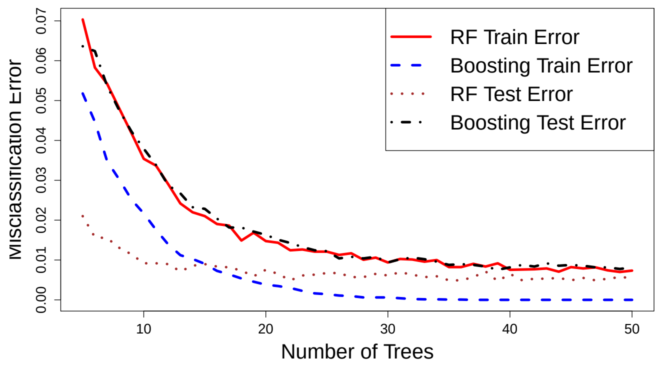

Fig. 6. Random forest Versus Boosting Test and Train Misclassification Error.

Table 1. Top and bottom three important predictors based on mean decrease in accuracy.

Priority	Predictor	Accuracy Mean Decrease	Gini Mean Decrease
1	magnet_dumbbell_y	0.1232	546.3155
2	roll_belt	0.1210	924.2964
3	magnet_dumbbell_z	0.1160	625.1139
50	gyros_dumbbell_z	0.0054	64.4434
51	gyros_forearm_x	0.0053	52.6771
52	gyros_arm_z	0.0043	38.9661

This values are based on Random Forest Permute function that Earn P-values in R console.

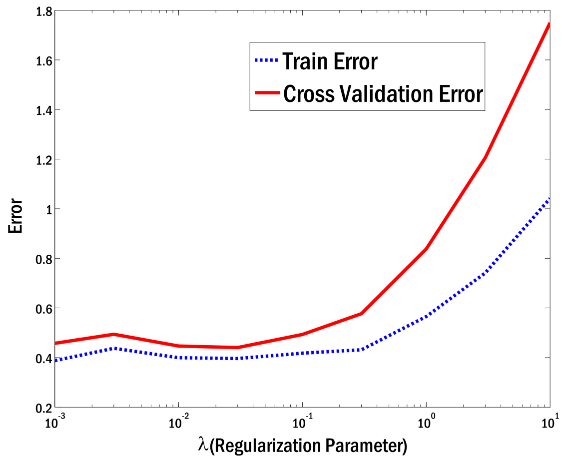

Fig. 7. Selection of Regularization Parameter in NN model.

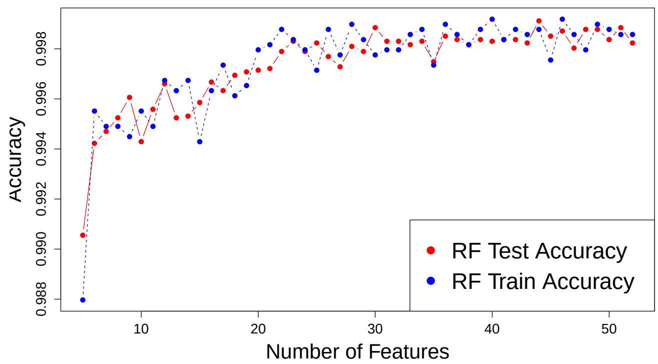

Fig. 8. Random forest important feature extraction base on Mean Decrease in Accuracy.

4.2 Machine learning model selection

In Figure 3 multi layered Neural Network with 3 number of layers and 26 neurons in hidden layer proposed. For prevention of overfitting problem in neural network there used regularization parameter that shrinkages value of weights.
In this research the dataset divided into three categories, 60 percent for training, 20 percent for cross validation and other 20 percent are for final evaluation in test set. Figure 7 shows validation and train error by earning appropriate regularization parameter.

4.3 Feature selection in statistical and machine learning model

There are different methods In the case of removing correlated features in statistical and machine learning based training procedure. In the statistical method the P-value shows the number of null hypothesis that could effect on sample datasets. Occurrence of null hypothesis for the P-values of lower than 0.05 implies on rare events that happened otherwise it is a usual event. Variables with P-values lower than 0.05 are more uncorrelated and independent so are more important in comparison to other predictors.
Table 1 has been illustrated via RFpermute package in R where predictors sorting based on mean decrease in accuracy. Gini mean decrease shows roll belt predictors have higher impact on prediction status while magnet dumbel in y direction has higher impact on mean accuracy decrease. In Figure 8 as number of evaluated features become more than 20, testing and training error become more stable. Another method for gaining compression in feature set is using methods like Principle Component Analysis(PCA). In

Table 2. Training Set and Testing Set accuracy of different models

Fitting Model	Train Accuracy	Test Accuracy
Bagging(R)	78.3734	77.2634
Random Forest(R)	100	99.4290
Boosting(R)	100	99.4086
Neural Network(Matlab)	98.5934	96.8501
Neural Network using PCA(Matlab)	96.5386	93.9755
Hybrid(Top 32 features)	95.8072	94.7615

PCA compression comes by taking important eigenvalues in evaluated feature set. The advantage of PCA method versus statistical one that removed unimportant features is impact of unessential predictors in new feature space. In this problem new features space in PCA comes by 32 top eigenvalues.
In the Hybrid model top 32 predictors in statistical evaluation used in Neural Network architecture and accuracy on both training and test sets gained as shown in Table 2

5. CONCLUSION

According to Table 2 Boosting and Random Forest shows best response in training and test set evaluation, by approximately 99.5 percent accuracy in testing and 100 percent in training set accuracy. The hybrid model with Equivalent number of features to the PCA shows better response in testing accuracy. Better compression method has been introduced in this model with comparison to common dimensionally reduction methods. Accuracy of classification in tuned decision trees are about 2 percent higher than equivalent model that introduced in [14].
New mechanism in Hybrid model proposed, here mean decrease accuracy per feature in random forest algorithm used as input variables in neural network architecture. The proposed Hybrid model

could be used as substitute to previous algorithms in dimensionally reduction procedures.
One of the main benefits of proposed hybrid model is important feature extraction without any impact on features in contrary to PCA algorithm where features are mapped in new dimenssionally space.

6. REFERENCES

[1] Saisakul a, Anthony S Atkins, and Hongnian Yu. Activity classification using a single wrist-worn accelerometer. In *Software, Knowledge Information, Industrial Management and Applications (SKIMA), 2011 5th International Conference on*, pages 1–6. IEEE, 2011.

[2] Esteban Alfaro, Matıas Gámez, and Noelia Garcia. Adabag: An r package for classification with boosting and bagging. *Journal of Statistical Software*, 54(2):1–35, 2013.

[3] Leo Breiman. Out-of-bag estimation. Technical report, Citeseer, 1996.

[4] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[5] Miikka Ermes, Juha Parkka, Jani Mantyjarvi, and Ilkka Korhonen. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *Information Technology in Biomedicine, IEEE Transactions on*, 12(1):20–26, 2008.

[6] HAR. Human activity recognition. http://groupware.les.inf.puc-rio.br/har.

[7] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The elements of statistical learnin, 2009.

[8] Benjamin Hofner, Andreas Mayr, Nikolay Robinzonov, and Matthias Schmid. Model-based boosting in r: a hands-on tutorial using the r package mboost. *Computational Statistics*, 29(1-2):3–35, 2014.

[9] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*. Springer, 2013.

[10] Andy Liaw and Matthew Wiener. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[11] Jun Nishimura and Tadahiro Kuroda. Multiaxial haar-like feature and compact cascaded classifier for versatile recognition. *Sensors Journal, IEEE*, 10(11):1786–1795, 2010.

[12] Greg Ridgeway. Generalized boosted models: A guide to the gbm package. *Update*, 1(1):2007, 2007.

[13] Liu Rong and Liu Ming. Recognizing human activities based on multi-sensors fusion. In *Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference on*, pages 1–4. IEEE, 2010.

[14] Eduardo Velloso, Andreas Bulling, Hans Gellersen, Wallace Ugulino, and Hugo Fuks. Qualitative activity recognition of weight lifting exercises. In *Proceedings of the 4th Augmented Human International Conference*, pages 116–123. ACM, 2013.

[15] Yuting Zhang, Stacey Markovic, Inbal Sapir, Robert C Wagenaar, and Thomas DC Little. Continuous functional activity monitoring based on wearable tri-axial accelerometer and gyroscope. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on*, pages 370–373. IEEE, 2011.