JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Human Activity Recognition using Machine Learning Techniques

Harpreet kaur lohia, Simran kaur Dari , Shravanti Kalwal, Roshni Singh
Dept. of computer science and engineering, RCERT Chandrapur
Asst. Prof. Anand D.G. Donald

Abstract-
Activity recognition is one of the leading applications of machine learning algorithms nowadays. It is being used in the field of biomedical engineering, game development, developing better stats for sports training etc. Data from the sensors attached to a person can be utilised to train supervised machine learning models in order to predict the activity being carried out by the person. In this paper we will be using Data available at UCI machine learning Repository. It contains data generated from accelerometer, gyroscope and other sensors of Smartphone to train supervised predictive models using machine learning techniques like SVM, Random Forest and decision tree to generate a model. which can be used to predict the kind of movement being carried out by the person which is divided into six categories walking, walking upstairs, walking downstairs, sitting, standing and laying.

Keywords-
Human Activity Recognition, Random Forest, SVM, classifier, Logistic Regression

Introduction-

In the last decade there has been an exponential increase in the number of smartphones that are available in the market. In 2014, in India alone the smartphone users are expected to increase by more than double from existing 156 million users to 364 million.[1]. With the growing number of smartphones, the amount of data that can be generated from the sensor of the smartphone is also growing. The smartphone comes equipped with various inertial sensors such as gyroscope, accelerometer. The reading of these sensors changes according to the movement of the smartphone.

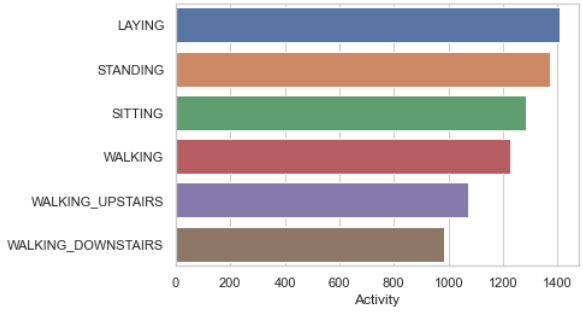
These smart phones have integrated seamlessly into everyday life of the people around us. People carry mobile devices around for most part of the day. This enables us to track activity of the person carrying the smart phone using the data of the sensors of their smartphone. Activity recognition takes into account this data and uses it to produce models which can be used to predict the activity being carried out by the person.

In this paper we use the data available at the UCI machine learning repository which has been modified for our research. The data will be processed through supervised machine learning algorithms to produce predictive classification models that will be used to classify physical activities of the person into six categories namely Sitting, Standing, Laying, Walking, walking upstairs, Walking Downstairs. Various models thus created will be analysed for accuracy using Random Forest, SVM, Logistic Regression.

DATA SET-

The dataset used in this paper comes from the UCI machine learning repository. The original Dataset was derived by carrying experiments on 30 Different Individuals volunteers within the age of 19-40 years. In the experiments each volunteer was made to perform six activities namely Sitting, Standing, Laying, Walking, walking upstairs, Walking Downstairs. During the experiment the volunteer was wearing a Smartphone on the waist. Using the embedded Accelerometer and Gyroscope on the given Smartphone 3-axial linear acceleration and 3-axial angular velocity were recorded at a constant rate of 50 Hz.

The Dataset is converted to sub data that contains 563 columns and 7325 rows. The feature values contained in these columns were the values from the gyroscope and accelerometer. Each such row of observation also contained the Id number of the volunteer ranging from 1 to 30. Also, each row contained the cell named Activity which contained the value of the activity being carried out by the volunteer. This cell is required for creating supervised ML models.



Methodology-

The problem we are dealing with in this paper falls into the category of Classification modelling. In this process we will be creating ML predictive models, which will then

be used to decide which category a particular observation belongs to, predictive models are formulated using the training data. The observations are needed to Classify under six categories namely sitting, standing, laying, walking, walking upstairs, walking downstairs. We will be using supervised learning methods to create the models. Various models thus created will be analysed for accuracy using Random Forest, SVM, Logistic Regression.

1.Decision Tree

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure. Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand. The logic behind the decision tree can be easily understood because it shows a tree-like structure.

2.Random Forest

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

3.Kneighbors classifier

The k-nearest neighbours algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another.

4. Logistic Regression

Logistic regression is a linear model for classification. In this model, the probabilities describing the possible outcomes of a single trial are modelled using a logistic function. The logistic function is a sigmoid function, which takes any real input and outputs a value between 0 and 1, and hence is ideal for classification. When a model learns the training data too closely, it fails to fit new data or predict unseen observations reliably. This condition is called overfitting and is countered, in one of many ways, with ridge (L2) regularisation. Ridge regularisation penalises model predictors if they are too big, thus enforcing them to be small. This reduces model variance and avoids overfitting.

5.SVC

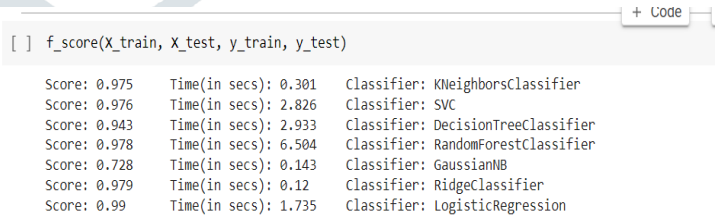
The objective of a Linear SVC (Support Vector Classifier) is to fit to the data you provide, returning a "best fit" hyperplane that divides, or categorises, your data. From there, after getting the hyperplane, you can then feed some features to your classifier to see what the "predicted" class is.

6. ensemble learning bagging

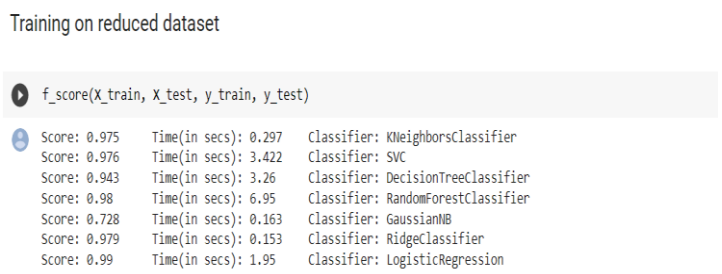
Bagging, also known as bootstrap aggregation, is the ensemble learning method that is commonly used to reduce variance within a noisy dataset. In bagging, a random sample of data in a training set is selected with replacement—meaning that the individual data points can be chosen more than once. After several data samples are generated, these weak models are then trained independently, and depending on the type of task—regression or classification, for example—the average or majority of those predictions yield a more accurate estimate. As a note, the random forest algorithm is considered an extension of the bagging method, using both bagging and feature randomness to create an uncorrelated forest of decision trees.

Observations

1. Accuracy of algorithms before reducing the dataset.



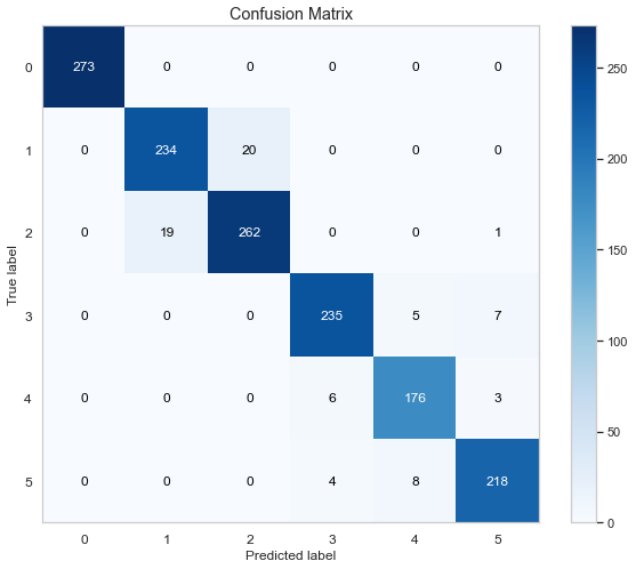
2. Accuracy of algorithms after reducing the dataset.



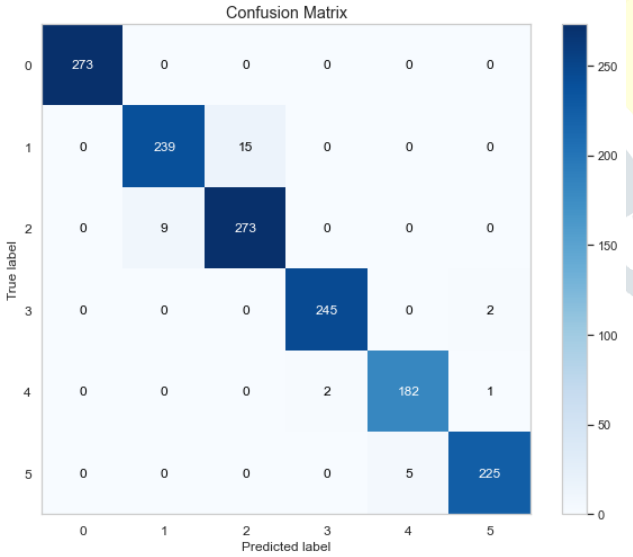
Confusion Matrix

Confusion Matrix are used to measure the performance of an algorithm in Machine learning process, it is mostly used in supervised learning. Each row of the matrix represents the actual value of class to be predicted and column represents the value of the predicted class. In our research we used Confusion Matrix to measure the rate of misclassification of the different models created.

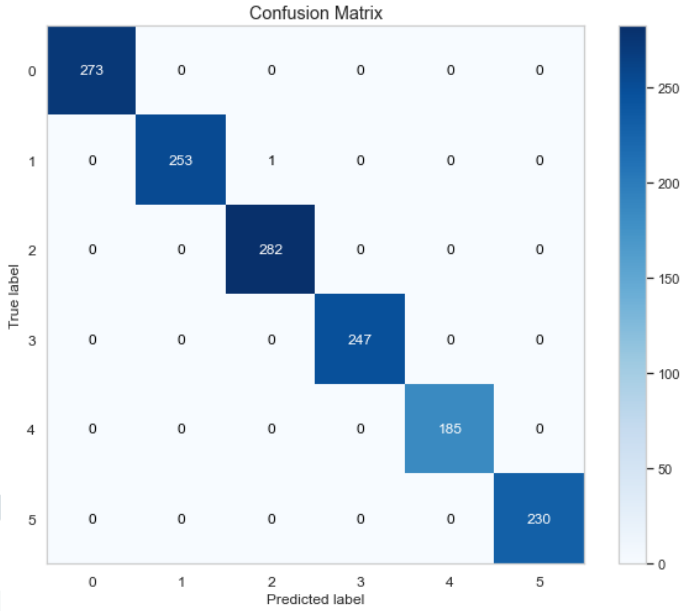
1. Confusion matrix for decision tree



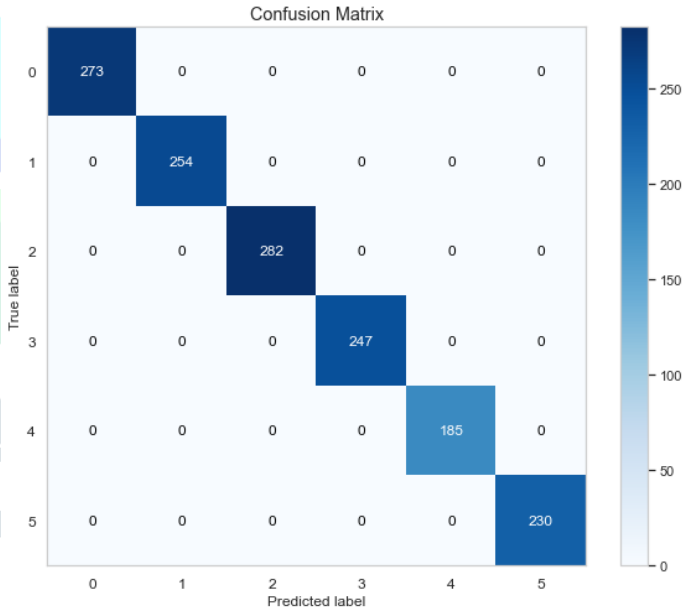
2. Confusion matrix for random forest.



3. Confusion matrix for ensemble learning bagging.



4. Confusion matrix for random forest classifier after applying bagging.



Conclusion

From the analysis conducted in this paper out of different algorithms used the accurate result that we have got is from random after applying bagging. Also, the Confusion Matrix analysis concluded that the misclassification error in the Random Forest Models was zero than other models. other models have given fewer effective results while the results provided by the random forest were accurate. It can be clearly said that the Random Forest model can be effectively utilized when used for activity recognition using the data provided by the accelerometer and gyroscope of a Smartphone.

References

[1] Charles Arthur. Smartphone explosion in 2014 will see ownership in India pass US. Report: 13th Jan 2014. http://www.theguardian.com/technology/2014/jan/13/smartph one-explosion-2014-india-us-china-firefoxos-android

[2] - Jorge L. Reyes-Ortiz, Davide Anguita, Alessandro Ghio, Luca Oneto. Human Activity Recognition Using Smartphones Dataset. UCI Machine Learning Repository. 10 Dec 2012. http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recog nition+Using+Smartphones

[3] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a

[4] Multiclass Hardware-Friendly Support Vector Machine. International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012.

[5] Jun Liu, Amir Shahroudy, Dong Xu, Gang Wang(2016). Spatio-Temporal LSTM with Trust Gates for 3D

[6] Academic Performance. International Journal of Computer Science and Information Security,7(1).292-295

[7] JuhaVesanto(1999). SOM-Based Data Visualization Methods. Intelligent Data Analysis, Laboratory of Computer and Information Science, Helsinki University of Technology, P. O. Box 5400, FIN-02015 HUT, Finland Vol. 3(2), pp. 111-126.