IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-8, Aug 2019 E-ISSN: 2347-2693

Performance Evaluation of Various Machine Learning Techniques for Human Activity Recognition using Smartphone

Anju S.S.[1*], Kavitha K.V.[2]

[1,2] Dept. of Computer Science and Engineering, Sree Chitra Thirunal College of Engineering, Thiruvananthapuram (Kerala) 695018, India

Corresponding Author: anjuss1997@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i8.316319 | Available online at: www.ijcseonline.org

Accepted: 25/Aug/2019, Published: 31/Aug/2019

***Abstract*—** The process of Human Activity recognition nowadays had found a wide variety of applications in healthcare and security surveillance. The commonly used smartphones are now available with inbuilt accelerometer and gyroscope sensors. The data collected using these sensors are used for recognizing the activity performed by the person who carries the smartphone. The sensor data collected from these sensors are fed to activity classifiers to train them. In this paper, the performance of various machine learning techniques are trained and evaluated for finding the better classification technique. In particular, examines the use of Decision tree, Naive bayes, K-nearest neighbour, Support Vector Machine and Random forest. The evaluation metrics used are accuracy, sensitivity, specificity and precision. During evaluation the results showed that the SVM showed better accuracy with the smartphone data**.**

***Keywords*—**Activity Recognition, Smart phone, Accelerometer, Machine Learning, Support Vector Machines.

I. INTRODUCTION

Introduction In the field of healthcare, elderly rehabilitation is an important area where Human Activity Recognition (HAR) is the problem of classifying sequences of accelerometer data recorded by smart phones into known well-defined activities like walking, sitting, standing etc. Motivation for Human activity recognition is being one of the important and challenging research area with many applications including, smart environments, surveillance and security, human survey, study human daily activities, medical care, elderly rehabilitation, home behavior analysis, video surveillance, gait analysis and gesture recognition. There are two types of HAR: Video-based HAR and Sensor-based HAR. In Video-based HAR videos or images containing human motions from the camera are analyzed. In sensor-based HAR focuses on the motion data from smart sensors such as an Accelerometer, Gyroscope, Bluetooth, or sound sensors. Sensor-based HAR is becoming more popular, due to the development of sensor technology and pervasive computing. Raw data collected from the sensors enables the classification of human activities with machine learning algorithms. In studies focusing on activity recognition, usually the motion sensors, such as accelerometer and gyroscope are used. Machine learning methods used includes decision trees, naive bayes, random forest classifier and support vector machines. In this study a dataset consist of signals from accelerometer and gyroscope of a smartphone carried by different man and women volunteers while doing different activities are classified using different machine learning approaches. Performance of different approaches are analysed and compared in terms of accuracy and precision.

The paper is organized as follows, Section I contains the introduction of Human Activity Recognition, Section II describes the related works, Section III explains the methodology of the system, Section IV describes results and discussions, and Section V concludes research work with future directions.

II. RELATED WORK

The problem of Human Activity Recognition using accelerometer signals from a smartphone can be realized as a classification problem where from the signals for each of the activity a pattern is searched by the classifier. In recent years many machine learning approaches were used for activity recognition but the feature extraction process end up with many features. For the selection of appropriate features a variety of methods was followed by researchers.

Tuan Dinh Le et.al. [1] proposed a robust system for human activity recognition by smartphone. Different from other work, investigated the use and combination feature selection and instance selection to reduce dimensionality of dataset in order to enhance the performance. They implemented the system on Android and the experimental results show that the system achieves better accuracy of up to 15% and the response time is 3 to 5 times faster when compared to the original system.

Erhan Bulbul et.al. [2] in this work focuses on recognition of human activities using smartphone sensors using different machine learning classification approaches. Data retrieved from smartphone accelerometer and gyroscope sensors are classified in order to recognise human activity. Results of the approaches used are compared in terms of efficiency and precision.

A.M. Khan et.al. [3] presented a method to address this problem. The proposed method is validated using five daily physical activities. Activity data is collected from five body positions using a smartphone with a built-in triaxial accelerometer. Features including autoregressive coefficients and signal magnitude area are calculated. Kernel Discriminant Analysis is then employed to extract the significant nonlinear discriminating features which maximize the between class variance and minimize the within-class variance.

Jindong Wang et. al. [4] presented a survey on different approaches of activity recognition using traditional classification algorithms and advances to deep learning. It studies the different methods in smartphone based activity recognition and also provides the overall idea of the current strategies in activity recognition.

III. METHODOLOGY

The objective is to evaluate the performance of different machine learning algorithms in smartphone based human activity recognition. For this the main steps included are data collection, feature extraction, feature reduction, model training and evaluation of models.

In data collection phase, the 3-axial data of accelerometer is collected using smart phone. For feature extraction the data is subjected to time and frequency domain analysis resulting in 561 features. The extracted features are further reduced to its 10% i.e. Reduced to 56 features using the Principal Component Analysis. The feature reduction step is used to decrease the processing cost by removing irrelevant and redundant features, but at the same time ensuring the accuracy of recognition. The removed features are those which provide very less or no information to the recognition process. In other words, the features which are irrelevant to the current scenario of activity recognition are not present in the selected features set. These results in improvements in model interpretability reduced training time and also enhance the generalization by reducing over fitting.

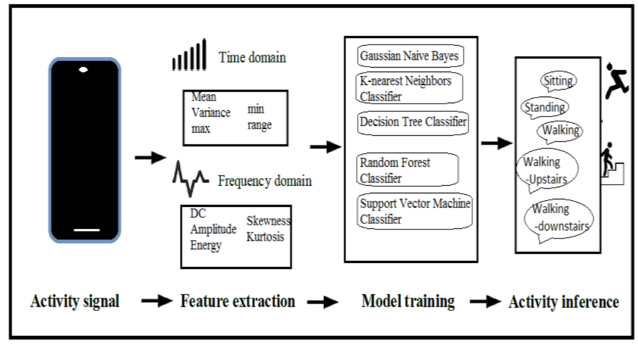

Figure 1: Activity recognition system

The data is normalized and using this data various machine learning classification models are trained. Figure 1 shows the workflow of the human activity recognition system.

The machine learning models evaluated here are Gaussian Naive Bayes, K-nearest Neighbours Classifier, Decision Tree Classifier, Random Forest Classifier and Support Vector Machine Classifier.
For evaluation of trained classification models, 5 fold cross validation is used to achieve more precise results.

IV. RESULTS AND DISCUSSION

In this section, to evaluate the machine learning classification algorithms accuracy, confusion matrix and classification report are produced. The dataset used is obtained from UCI Machine Learning Repository. The dataset is created by conducting the experiment with 30 individuals with each one is carrying a Samsung Galaxy S2 smartphone.

The signals from the embedded accelerometer within the smartphone are collected while the individuals are doing various activities. The experiment was done for 6 activities namely sitting, standing, laying, walking, walking upstairs and walking downstairs. The data of 3-axis of accelerometer with a constant rate of 50 Hz was recorded and performed time domain and frequency domain analysis for obtaining 561 features. The dataset then divided into 2 parts one with 70% of data tuples for training the modals and 30% for testing the modals. For each machine learning approach, the 5 fold cross validation is performed.

A. Accuracy

Figure 3 shows the accuracy measure of various machine learning algorithms during the testing and training phases. The six classifiers Gaussian naive bayes, decision tree, random forest, k-nearest neighbour and SVM classifier are evaluated.

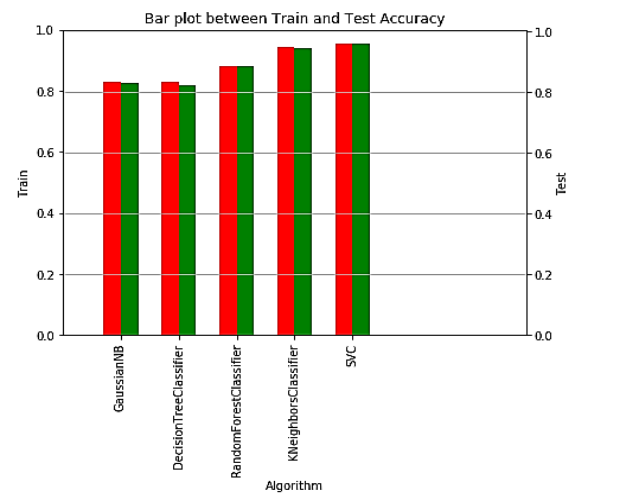

Figure 3. Bar plot of training and testing accuracy

During evaluation of the results, SVM showed maximum accuracy at training and testing phase. Among 5 classifiers evaluated, SVM with linear kernel gave the maximum accuracy 95.9%.

Other algorithms also have a significant accuracy scores as follows, Gaussian Naive Bayes: 83.4%, Decision Tree Classifier: 83.5%, Random Forest Classifier: 88.6% and K-Neighbors Classifier: 95%. In figures 4-8 accuracy values and confusion matrix for each classifier is shown.

B. Classification Report

The classification report contains the precision, recall, support and f1-score for each of the iterations in the evaluation process with the 5-fold cross validation. The classification report generated for each of the evaluated machine learning algorithms is shown in below figures 4-8.

Support vector machine shows the maximum precision, recall, f1-score and support during the testing of the trained machine learning classifiers. K-nearest neighbours classifier also showed better precision than others. The naive bayes showed comparatively poor performance with least precision value. Even though the random forest classifier is an ensemble classifier and expected a maximum performance but it showed less precision than SVM classifier.

On the basis of f1-scores also, the SVM classifier was better. The f-1 scores of the classifiers are the following, Naive Bayes: 0.83, Decision Tree: 0.84, Random Forest Classifier: 0.89 and K-Neighbors Classifier: 0.95 and Support Vector Machine Classifier: 0.96.

```
============= SVC ==================

Model Accuracy: 95.92233009708738
Confusion Matrix:
[[501   1   0   0   0   0]
 [  2 392  48   0   0   1]
 [  0  31 426   0   0   0]
 [  0   0   0 440   1   5]
 [  0   0   0   4 337   2]
 [  0   0   0   3   7 374]]
classificationReport:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       502
           1       0.92      0.88      0.90       443
           2       0.90      0.93      0.92       457
           3       0.98      0.99      0.99       446
           4       0.98      0.98      0.98       343
           5       0.98      0.97      0.98       384
```

Figure 4. Confusion matrix and classification report of Support vector machine classifier

```
============= DecisionTreeClassifier ==================

Model Accuracy: 83.57281553398059
Confusion Matrix:
[[480  16   5   0   1   0]
 [ 21 321 100   0   0   1]
 [  2 105 350   0   0   0]
 [  0   0   0 393  32  21]
 [  0   0   0  28 289  26]
 [  0   0   0  23  42 319]]
classificationReport:
              precision    recall  f1-score   support

           0       0.95      0.96      0.96       502
           1       0.73      0.72      0.73       443
           2       0.77      0.77      0.77       457
           3       0.89      0.88      0.88       446
           4       0.79      0.84      0.82       343
           5       0.87      0.83      0.85       384
```

Figure 5. Confusion matrix and classification report of Decision Tree classifier

```
============= GaussianNB ==================

Model Accuracy: 83.49514563106796
Confusion Matrix:
[[468   3   2   0  28   1]
 [  5 309 107   1  18   3]
 [  1  34 397   4  15   6]
 [  0   0   0 385  42  19]
 [  0   0   0  38 244  61]
 [  0   0   0   6  31 347]]
classificationReport:
              precision    recall  f1-score   support

           0       0.99      0.93      0.96       502
           1       0.89      0.70      0.78       443
           2       0.78      0.87      0.82       457
           3       0.89      0.86      0.88       446
           4       0.65      0.71      0.68       343
           5       0.79      0.90      0.85       384
```

Figure 6 Confusion matrix and classification report of Gaussian Naive Bayes classifier

```
============= KNeighborsClassifier ==================

Model Accuracy: 95.02912621359224
Confusion Matrix:
[[494   3   4   1   0   0]
 [  5 373  64   0   0   1]
 [  0  33 424   0   0   0]
 [  0   0   0 442   3   1]
 [  0   0   0   5 334   4]
 [  0   0   0   3   1 380]]
classificationReport:
              precision    recall  f1-score   support

           0       0.99      0.98      0.99       502
           1       0.91      0.84      0.88       443
           2       0.86      0.93      0.89       457
           3       0.98      0.99      0.99       446
           4       0.99      0.97      0.98       343
           5       0.98      0.99      0.99       384
```

Figure 7. Confusion matrix and classification report of K-Nearest Neighbors classifier

```
============= RandomForestClassifier ==================

Model Accuracy: 88.62135922330097
Confusion Matrix:
[[494   5   1   1   1   0]
 [ 18 340  84   0   0   1]
 [  1  82 374   0   0   0]
 [  0   0   0 420  14  12]
 [  0   0   0  25 302  16]
 [  0   0   0  12  20 352]]
classificationReport:
             precision    recall  f1-score   support

          0       0.96      0.98      0.97       502
          1       0.80      0.77      0.78       443
          2       0.81      0.82      0.82       457
          3       0.92      0.94      0.93       446
          4       0.90      0.88      0.89       343
          5       0.92      0.92      0.92       384
```

Figure 8. Confusion matrix and classification report of Random Forest classifier

V. CONCLUSION AND FUTURE SCOPE

During evaluation of the results, SVM showed maximum accuracy at training and testing phase. The K-nearest neighbors classifier also showed significant accuracy on testing. Least accuracy was given by the naive bayes classifier. The random forest classifier, which is an ensemble of decision trees, was expected to give more accuracy. It showed greater accuracy than the accuracy obtained from using a single decision tree, but gave accuracy which is less than that given by the SVM classifier. The K-nearest neighbour's classifier gave accuracy closer to the obtained maximum accuracy of the SVM Classifier. On considering the f1-scores also the SVM classifier was found better. As the result of evaluation of performance, it is clear that with feature reduction, SVM with linear kernel showed better accuracy among the above classifiers.

As a future work, complex neural network models like CNN, RNN, LSTM etc can be cross verified and evaluated for better performance. Also more complex activities can be included in the dataset for recognition.

REFERENCES

[1] T.Dinh Le and C. Van Nguyen, "*Human activity recognition by smartphone*", 2015 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS), **Ho Chi Minh City**, pp. **219-224**, **2015**.

[2] J. E. Bulbul, A. Cetin and I. A. Dogru, "*Human Activity Recognition Using Smartphones*", 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), **Ankara**, pp. **1-6**, **2018.**

[3] A. M. Khan, Y. K. Lee and S. Y. Lee, "*Human activity Recognition via An Accelerometer Enabled Smartphone Using kernel discriminent Analysis*", 2010 5th International Conference on Future Information Technology, **Busan**, pp. **1-6**, **2010.**

[4] Jindong Wang, Yiqiang Chen, Shuji Hao, Xiaohui Peng, Lisha Hu," *Deep Learning for Sensor-based Activity Recognition: A Survey*", Pattern Recognition Letters, pp. **1-9**, **2018.**

[5] M. C. Sorkun, A. E. Danisman and O. D. Incel, "*Human activity recognition with mobile phone sensors: Impact of sensors and window size*", 2018 26th Signal Processing and Communications Applications Conference (SIU), **Izmir**, pp. **1-4, 2018**.

[6] K. Altun, B. Barshan, Tuncel," *Comparative study on classifying human activities with miniature inertial and magnetic sensors*", Pattern Recognition Letters, vol. **43**, no.**10**, pp. **3605–3620**, **2010**.

[7] A. Wang, G. Chen, J. Yang, S. Zhao and C. Chang, "*A Comparative Study on Human Activity Recognition Using Inertial Sensors in a Smartphone*", IEEE Sensors Journal, vol. **16**, no. **11**, pp. **4566-4578, 2016.**

[8] Meenu Shukla, Sanjiv Sharma, "Analysis of Efficient Classification Algorithm for Detection of Phishing Site", International Journal of Scientific Research in Computer Science and Engineering, Vol.**5**, Issue**.3**, pp.**136-141**, **2017.**

[9] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. "*A Public Domain Dataset for Human Activity Recognition Using Smartphones", 21th European Symposium on Artificial Neural Networks*, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, **Belgium**, pp. **24-26**, **2013**.

[10] Gagandeep Kaur, Harmanpreet Kaur, "*Ensemble based J48 and random forest based C6H6 air pollution detection*", International Journal of Scientific Research in Computer Science and Engineering, Vol.**6**, Issue.**2**, pp.**41-50, 2018**.

Authors Profile

Anju S S pursed Bachelor of Technology from Kerala University, India in 2018. She is currently pursuing Master of Technology in Sree Chitra Thirunal College of Engineering, Thiruvananthapuram, India. Her main research work focuses on Machine Learning, Data Analytics, Data Mining, Deep Learning, Computer Vision, and Computational Intelligence.

Kavitha K V pursed Bachelor of Technology and Master of Technology from Kerala University, India. She is currently working as an Assistant Professor in CSE at SCT College of Engineering, Pappanamcode, Thiruvananthapuram. Her areas of interest are Machine Learning, Network Security, IOT and Cloud computing.