ISSN No. 0976-5697

Volume 8, No. 3, March – April 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Human Activity Recognition from Sensor data using Random Forest Algorithm

Sijo Antony N J
Department of Computer Science
Christ University
Bengaluru, India

Kavitha R
Department of Computer Science
Christ University
Bengaluru, India

Abstract: The advancement of technologies have facilitated the monitoring of human activities through the embedded sensors in a smartphone. Since the smartphone is very common with the public, it has paved way for the researchers to work on activity recognition systems. In this work, we propose a system to recognize the basic activities of an individual. For this publicly available dataset was used. The data collected from the sensors are used to predict the activities. The proposed system obtained the highest accuracy of 99.6%. This system is implemented on R programming using Random Forest algorithm.

Keywords: Human Activity Recognition, Sensor, Random Forest

I. INTRODUCTION

The essential motivation behind the activity recognition is to observe the actions of an individual and their surrounding environment. The rapid growth in the technology has made a huge forward leap in the field of human activity recognition. This has made a significant changes in the way the elderly people, athletes, patients etc. are monitored. The human sensor based activity recognition is a combination of sensor networks hand-in-hand with the data mining and machine learning techniques. The smartphones provide enormous amount of sensor data for one to understand the daily activity patterns of an individual. The sensor data is also available from different sources such as wearable devices like smartwatches, fitness tracking bands, E-textiles etc.

By monitoring and analyzing these data from the sensor it is easy to recognize the day to day activities of each individuals. For e.g. a person staying away from his/her elderly parent can monitor their daily activities. Activities like eating, moving to different rooms, medication, sleep etc. are monitored. They can raise an alarm if there is a change in the pattern or an early alarm of healthcare emergency. In this paper, our goal is to recognize six different human activities using the sensor data.

II. LITERATURE REVIEW

Human activity recognition has been a vital topic for the researchers. In the era of cloud computing and Internet of Things, the activity recognition has made a tremendous change in the lives of the people. The researchers have found multiple ways to collect data. They also have introduced the world to new algorithms in the field of activity recognition. Each time researchers comes up with better ways of using the algorithms and setting new benchmarks on the accuracy rates of the new models. The use of smartphones and fitness bands are feasible ways of gathering the sensor data of an individual's everyday activity. Thanks to the energy efficient sensors such as accelerometer, gyroscope etc. used in the devices. TABLE 1 represents the studies carried out by various researchers using the sensor data.

III. IMPLEMENTATION

A. *Human Activity Recognition Dataset*

In this work, we experimented the publicly available dataset [1] for Human Activity Recognition. The dataset comprises of sensor reading of 6 basic activities. The activities were performed by wearing a smartphone (Samsung Galaxy S II) on the waist. The embedded accelerometer and gyroscope helped in capturing tri-axial linear acceleration and tri-axial angular velocity at a constant rate of 50Hz. Dataset was divided into 2 sets where the 70% of the volunteers were selected to generate the training dataset and the rest 30% for the test data. The sensor signals 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 gravitational and body motion components present in the sensor acceleration signal was separated using a Butterworth low pass filter. Assuming the gravitational force contains only low frequency components, a filter with 0.3 Hz cutoff frequency was used. From each window, by calculating variables from the time and frequency domain, a new vector of features is obtained.

B. *Activities*

A group of 30 volunteers within an age bracket of 19-48 years were selected to perform this experiment. The smartphones attached to their waist captured the signals of the activities. The six activities are Sitting, Walking, Laying, Standing, Walking upstairs and walking downstairs.

Running: A high vitality, tedious physical activity with three stages in particular i.e. upward motion, descending motion and high vitality sway.

Walking: A low vitality tedious activity. Unlike running, walking is a typical day to day activity of an individual.

Walking upstairs: It is not an everyday activity of an elderly person. On walking upstairs an individual is conflicting with the gravitational force and hence obliges expanded blood stream to the outskirts, which consequently expands the amount of work on the cardiovascular yield.

Walking downstairs: It is the opposite activity of walking upstairs. When walking downstairs the individual works with the gravitational force. This results in using less energy [3].

TABLE 1: BRIEF SUMMARY OF PREVIOUS WORK

Reference	Sensors used	Activities	Features	ML	Accuracy
[2]	Accelerometer	A1/A2 on treadmill, A3, A10, A4, A5, A6, A7, A8, A10, A11	Max, Min, AM, Std, MCR, (10th, 25th, 50th, 75th)percentile	RF, k-NN, SVM, Decision Tree (DT), Multi-Layer Perceptron (MLP), Naive Bayes (NB)	83%-90%
[3]	Accelerometer, Gyroscope	Serve, Return, Smash, Backhand, Forehand	Mean, median, Std, Variance, RMS, Highest peak, Lowest peak, Energy	k-NN, SVM	88.89%
[4]	Accelerometer	Falling forward, Falling to left, Falling to right, A2, A3, Stand to sit, Lying to sit	Average, Std	-	95%
[5]	Accelerometer, Gyroscope, Heart Rate monitor, Barometer, Skin Temperature, Compass GPS, Ambient light, GRS, UV	A12, A13, A14, A15, A16, A11, A17, A18, A19	Mean, Std	SVM	88%-100%
[6]	Accelerometer, Gyroscope, GPS	A2, A3, A1, A6, Non-uniform motion, Unknown	-	Viterbi programming	-
[7]	Accelerometer, Gyroscope	A2, A4, A5, A6, Lying down	-	SVM, MLP, Deep CNN	96%-97%
[8]	Accelerometer	A3, Serves, Forehand, Backhand	Mean, Probability, Std, Average	Naïve Bayes, MLP, Decision Tree(J48), SVM	96.25%
[9]	Accelerometer, Gyroscope	A2, A3, A6, A4, A5	Mean, Std	Naïve Bayes, C4.5 decision tree, SVM, RF, Multilayer perception	-
[10]	Accelerometer, Gyroscope, Light, Proximity, Barometer, GPS	A2, A4, A5, A23, and jumping	Mean, Std, Max, Min, IQR, Wavelength energy, FFT, DTW distance, Correlation	RF, SVM, k-NN	-
[11]	Accelerometer	A2, A4, A5, A6, A1, A20, A21, A22, A23, A24, A25, Laying,	Mean	RF	100%

A1-standing;A2- Walking ;A3-Running;A4-Walking Upstairs;A5- Walking Downstairs;A6- Sitting;A7- Uphill;A8- Downhill; A9- Driving;A10- folding/stacking;A11 Brushing teeth;A12- Writing;A13- Typing;A14- Stirring;A15- Eating; A16- Cutting with knife;A17- Drinking;A18- Washing;A19- Resting;A20- Stand;A21- Sit;A22- Lie;A23-Jogging; AM- Arithmetic mean; Std- Standard deviation; Mcr- Median crossing rate; RMS- Root Mean Square; RF- Random Forest; SVM-Support Vector Machine; k-NN-K-Nearest Neighbor;

Sitting: This activity is the continuous period of being seated. The sitting posture of the individuals brings a change in the sensor readings. The sitting posture of each individual is not the same.

Standing: Maintaining an upward position with the support of one's feet is called standing. The gravitational force acting on the body puts pressure on the body parts which helps in maintain the upright position.

Laying: It is the position of rest. The body is more or less horizontal to the surface underneath.

C. *Feature Extraction*

The features captured for this database came from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ ('t' represents time domain signals). These signals were captured at a constant rate of 50 Hertz. These signals were filtered using a medium filter and a 3rd order low pass Butterworth filter to remove noise with a corner frequency of 20Hz. Another low pass Butterworth filter at a corner frequency of 0.3Hz was used to separate the acceleration signal into body and gravity acceleration signal. The angular velocity and the body linear acceleration were derived in time to obtain Jerk signals. Euclidean norm was used to calculate the magnitude of these 3 dimensional signals. Some of these signals were subjected to Fast Fourier Transform (FFT) to produce fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag ('f' denoted frequency domain signals) [13].

D. *Machine learning Algorithmic Description*

Random Forest is one of the most accurate algorithms available. For multiple datasets, this algorithm produces a highly accurate classifier. It can also run on database of larger sizes. The best attribute of this algorithm is that it can handle input variables of size ranging from hundreds to thousands without variable deletion and estimate the important variables required in the classification. It effectively estimates the missing data and maintains accuracy when a large proposition of the data are missing.

Random Forest starts by selecting multiple bootstrap samples from the original dataset. The original training instances might contain out of the bootstrap samples which is called out-of-bag (OOB) instances. Once the bootstrap samples are selected, the decision tree learns each bootstrap samples. At every node only a few number of randomly selected variables are available for binary partitioning. Largest possible trees are grown using the random forest. The tree grown with the bootstrap sample predicts the OOB examples at each bootstrap iteration. The Aggregated OOB predictions helps in calculating the error rate [10].

E. *Implementation Setup*

The available datasets were thoroughly understood using R tool. R is a programming language and environment designed and developed for the statisticians and researcher for statistical analysis. It is the most comprehensible statistical analysis package available. It provides a better way of managing and manipulating data along with the standard statistical tests, models and analyses. It also possess outstanding graphical capabilities surpassing other graphical and statistical packages. Various packages are used for conducting the experiment and implementation of Random Forest Algorithms. 'readxl', 'XLConnect' 'ggplot', 'randomForest' etc. are few of the important packages used in this experiment. The dataset was divided in to two sets where 70% of the dataset was used as training dataset and the rest 30% as test dataset. The train data consisting of sensor reading of x, y, z axes are combined to form a single data frame. This data frame is used by the Random Forest algorithm to create decision tree. The algorithm predicts the activities of the test data with an accuracy of 99%.

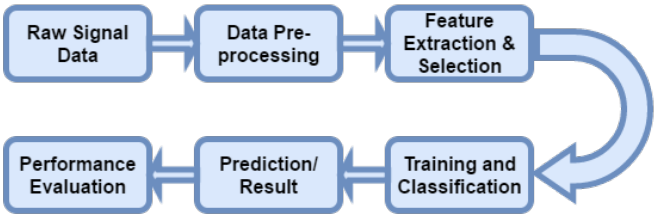

Fig. 1. Basic Activity Recognition steps

The figure 1 shows the steps involved in the activity recognition process. The main steps are i) Collection of Signal data ii) Data pre-processing iii) Feature extraction and selection iv) Training and classification v) Prediction vi) Performance evaluation. The initial step is to gather sensor data from the smart devices like smartphone, fitness bands etc. The raw data collected is preprocessed to remove noise which prevents us from achieving quality results. Further the pre-processed data is fed to the next stage to extract features. The feature extraction converts the data to informative and non-redundant data. This facilitates better recognition of human activities. This data is split into two sets namely train data and test data. The data is trained using various machine learning algorithms. The prediction algorithm uses the test data to predict the activities. The better the dataset, the better the accuracy of the result.

IV. RESULT AND DISCUSSION

To re-examine the effectiveness of the experiment, driven by Random Forest Algorithm, validation was carried out on the existing model using the validation subset of the studied dataset. The cross validation model was tested after applying the 5 fold cross validation approach upon training subset. Different random seeds were used when the experiment was run for multiple times. Confusion matrix were produced to demonstrate the performance of the model. The highest values in the result percentage of the confusion matrix indicate a higher success in recognition performance i.e. the error percentage is minimal. The results are displayed in Table 2.

The experiment was setup for six basic activities. We applied random forests on the raw features. On the Raw features we set the size to 500. The Table 2 shows the recognition performance using Random. The features were fed to Random Forest to build the final HAR recognition system. The graph (Fig.2) clearly indicates the higher precision of the Random Forest algorithm in the activity recognition experiment.

TABLE 2 : CONFUSION MATRIX OF RECOGNIZED ACTIVITIES

HAR	Walking	Walking upstairs	Walking downstairs	Sitting	Standing	Laying	Result %
Walking	494	1	1	0	0	0	99.60%
Walking upstairs	0	471	0	0	0	0	100.00%
Walking downstairs	4	7	409	0	0	0	97.38%
Sitting	0	0	0	476	15	0	96.95%
Standing	0	0	0	10	522	0	98.12%
Laying	0	0	0	0	0	537	100.00%
Precision %	99.19%	98.30%	99.76%	97.90%	97.13%	100.00%	99.96%

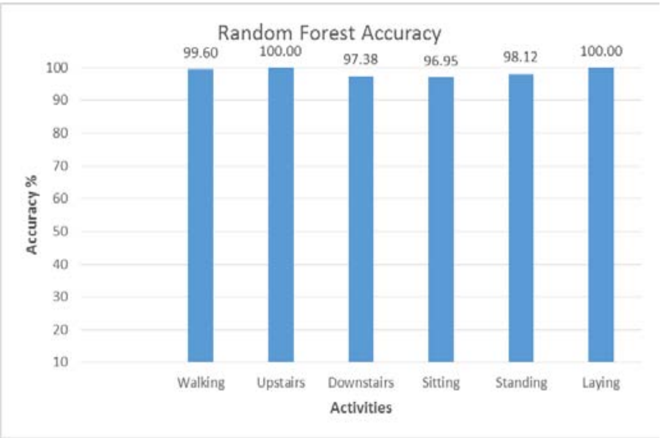

Fig. 2. Classification Accuracy of Random Forest

V. CONCLUSION

With this paper, we propose and implement the basic activity recognition model with the help of Random Forest Algorithm from the inertial sensors (accelerometer and gyroscope sensors) of smartphones. The future enhancements to this system include the extension of the number of activities performed. Also bring up new techniques to deal with the sensor data of specially abled people’s motion patterns. This would also include the gathering of data from multiple devices from the user to design more robust HAR systems.

VI. REFERENCES

[1] 2017. [Online]. Available:] https://archive.ics.uci.edu/ml/machine-learning-databases/00240/

[2] D. Yazdansepas et al., "A Multi-featured Approach for Wearable Sensor-Based Human Activity Recognition," 2016 IEEE International Conference on Healthcare Informatics (ICHI), Chicago, IL, 2016, pp. 423-431.

[3] M. A. I. Anik, M. Hassan, H. Mahmud and M. K. Hasan, "Activity recognition of a badminton game through accelerometer and gyroscope," 2016 19th International Conference on Computer and Information Technology (ICCIT), Dhaka, 2016, pp. 213-217.

[4] A. Babu, K. Dube, S. Mukhopadhyay, H. Ghayvat and Jithin Kumar M. V, "Accelerometer based human activities and posture recognition," 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE), Ernakulam, 2016, pp. 367-373.

[5] A. Filippoupolitis, B. Takand and G. Loukas, "Activity Recognition in a Home Setting Using Off the Shelf Smart Watch Technology," 2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS), Granada, 2016, pp. 39-44.

[6] G. Sebestyen, I. Stoica and A. Hangan, "Human activity recognition and monitoring for elderly people," 2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, 2016, pp. 341-347.

[7] T. Zebin, P. J. Scully and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," 2016 IEEE SENSORS, Orlando, FL, 2016, pp. 1-3.

[8] M. Dangu Elu Beily, M. D. Badjowawo, D. O. Bekak and S. Dana, "A sensor based on recognition activities using smartphone," 2016 International Seminar on Intelligent Technology and Its Applications (ISITIA), Lombok, 2016, pp. 393-398.

[9] Xizhe Yin, Gary Shen, Xianbin Wang and Weiming Shen, "Mitigating sensor differences for phone-based human activity recognition," 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, 2016, pp. 003550-003555.

[10] Y. Chen; C. Shen, "Performance Analysis of Smartphone-Sensor Behavior for Human Activity Recognition," in IEEE Access , vol.PP, no.99, pp.1-1.

[11] M. T. Uddin, M. M. Billah and M. F. Hossain, "Random forests based recognition of human activities and postural transitions on smartphone," 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, 2016, pp. 250-255.

[13] 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 24-26 April 2013.