TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol. 12, No. 9, September 2014, pp. 7041 ~ 7045
DOI: 10.11591/telkomnika.v12i9.5055

Human Activity Recognition on Smartphone: A Classification Analysis

Shianghau Wu*, Yanan Song
Faculty of Management and Administration, Macau University of Science and Technology,
Avenida Wai Long, Taipa, Macau, China, telp/fax: (853)88972399
*Corresponding author, e-mail: shwu@must.edu.mo

Abstract

The study hinged on the human activity recognition on smartphones by using the random forests model and Ada Boost model to make the classification. The study compared the classification results of two models and found the AdaBoost model had the better classification results. The study also found the Ada Boost model had the advantage of less calculation time.

Keywords*: activity recognition, random forests model, AdaBoost model, classification, machine learning*

1. Introduction

Smartphones are the indispensable gadgets for nowadays life. Many people many people use smartphones with built-in accelerometers of smartphones to recognize daily activities [1]. The recognition of human activities is essential for smartphone developers [2].

The study attempted to analyze the data from Anguita et. al. (2012) to analyze the features of human activity recognition. The goal of human activity recognition is to identify the human actions given a set of observations of itself and the surrounding environment. Recognition can be accomplished by analyzing the information retrieved from inertial sensors such as accelerometers [3].

The rest of the paper was organized as follows. First, the study began with the introduction to the database. Second, the overall research design was outlined. Third, in order to comprehend the features of human activity recognition, the study applied the random forests and AdaBoost Model classification analysis. The paper concluded with implications and future research avenues.

2. Research Method

2.1. Data

The study analyzed the data from the study and experiments of Anguita et. al. (2012). The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed the six activities previously mentioned wearing the smartphone on the waist. The experiments have been video-recorded to facilitate the data labeling. A Samsung Galaxy S2 smartphone has been exploited for the experiments, as it contains an accelerometer and a gyroscope for measuring 3-axial linear acceleration and angular velocity respectively at a constant rate of 50Hz, which is sufficient for capturing human body motion. The recognition process starts with the acquisition of the sensor signals, which are subsequently pre-processed by applying noise filters and then sampled indexed-width sliding windows of 2.56 sec and 50% overlap [4]. The database was originally randomly partitioned into two sets, where 70% of the patterns were used for training purposes and 30% as test data: the training set is then used to train random forests model and AdaBoost model for classification which were described in the following section.

Received November 4, 2013; Revised June 27, 2014; Accepted July 20, 2014

2.2. Variables Description

The data included the training data and test data. The amount of training data was 7352 and 2947 for the test data. The relabeled variables of the database analyzed in the study were introduced as follows:

(1) X1: training or test set of human activity features

(2) x: the acceleration signal from the smartphone accelerometer X axis in standard gravity units ’g’. Every row shows a 128 element vector.

(3) y: the acceleration signal from the smartphone accelerometer Y axis in standard gravity units ’g’.

(4) z: the acceleration signal from the smartphone accelerometer Z axis in standard gravity units ’g’.

In the study, the training set was categorized as “Type 0”, and the test set was categorized as “Type 1” for classification.

2.3. Random Forests Model

The random forests classification included the following steps [5-7],

(1) Draw the n_{tree} bootstrap samples from the original data.

(2) For each of the bootstrap samples grow an unpruned classification or regression tree, with the following modification: at each node, rather than choosing the best split among all predictors, randomly sample mtry of the predictors and choose the best split from among those variables.

(3) Predict new data by aggregating the predictions of the n_{tree} trees (i.e., majority votes for classification, average for regression).

2.4. AdaBoost Model

AdaBoost model is a machine learning algorithm which builds a strong classifier from a small set of efficient but weak classifiers. The idea is to choose the weak classifiers in such a way that when combined they perform much better. In the result, the final strong classifier builds a model that is able to predict the class of a new observation given a data set [8-9]. Viola and Jones (2001) also developed the AdaBoost algorithm further to boost the classification performance by combining collections of weak classifiers to form a stronger classifier. In the beginning, a set of weak classifiers are chosen with the lowest classification error. Then the sequence of machine learning problems is solved and the final strong classifier which takes a weighted combination of the weak classifiers is determined. The final strong classifier determines the optimal threshold classification function for each feature [10-11].

The general procedure of AdaBoost algorithm is shown as Figure 1 [12].

Input: Data set $D = \{(x_1, y_1), (x_2, y_2), \cdots, (x_m, y_m)\}$;
Base learning algorithm $\mathcal{L}$;
Number of learning rounds T.

Process:

1. $\mathcal{D}_1(i) = 1/m$. % Initialize the weight distribution
2. **for** $t = 1, \cdots, T$:
3. $h_t = \mathcal{L}(D, \mathcal{D}_t)$; % Train a learner h_t from D using distribution $\mathcal{D}_t$
4. $\epsilon_t = \Pr_{x \sim \mathcal{D}_t, y} \mathbf{I}[h_t(x) \neq y]$; % Measure the error of h_t
5. **if** $\epsilon_t > 0.5$ **then break**
6. $\alpha_t = \frac{1}{2} \ln \left(\frac{1-\epsilon_t}{\epsilon_t} \right)$; % Determine the weight of h_t
7. $\mathcal{D}_{t+1}(i) = \frac{\mathcal{D}_t(i)}{Z_t} \times \begin{cases} \exp(-\alpha_t) & \text{if } h_t(x_i) = y_i \\ \exp(\alpha_t) & \text{if } h_t(x_i) \neq y_i \end{cases}$
$= \frac{\mathcal{D}_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$ % Update the distribution, where
% Z_t is a normalization factor which
% enables $\mathcal{D}_{t+1}$ to be a distribution
8. **end**

Output: $H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x) \right)$

Figure 1. The AdaBoost Algorithm

3. Results and Analysis
3.1. Random Forests Model

The study applied the random forests classification analysis to make the classification of the train data and test data. The study set the category of train data as type 0 and test data as type 1. The “rattle” package of the R software randomly chose 7209 data as the test data (7090 Type 0 data, 119 Type 1 data). The number of trees was set as 700, and the number of variables tried at each split was set as 2. The duration of the calculation was 9.29 minute. The error matrix of the random forests model for test data is as Table 1.

Table 1. The Error Matrix of Random Forests Model

Observed	*Predicted* Type No.0	Type No.1	Percentage error
Type No. 0	5001	106	2.07%
Type No. 1	13	2089	0.61%
Overall Error Percentage	1.65%		

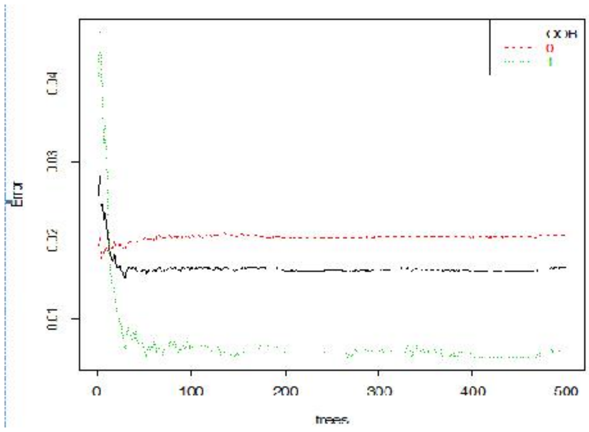

Figure 2. Error Rates of Random Forests

According to Table 1 and Figure 2, the overall error rate of random forests model was 1.65%, and the error rate of Type 1 data was lower than Type 1 data when the number of classification tree increased.

The random forests model also calculated the variable importance, mean decrease accuracy and mean decrease gini of keywords which were listed as Table 2 and Figure 3.

Table 2. Valuable Importance, Mean Decrease Accuracy and Mean Decrease Gini

Variables	Valuable Importance (Type 0 Data)	Valuable Importance (Type 1 Data)	Mean Decrease Accuracy	Mean Decrease Gini
X1	107.58	488.22	466.00	552.30
x	112.62	484.25	453.36	546.13
y	58.75	405.76	377.03	399.75
z	44.62	398.80	373.50	384.73

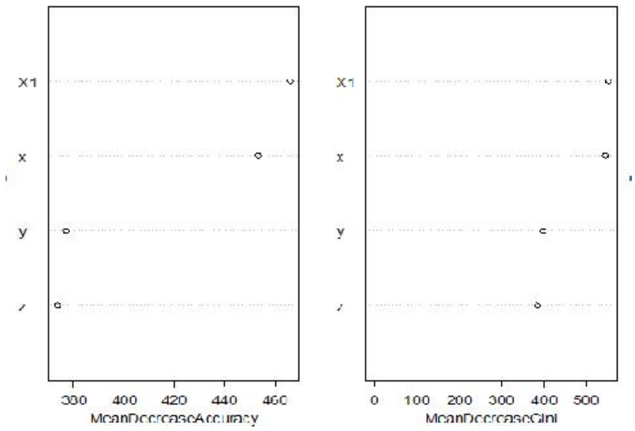

Figure 3. Mean Decrease Accuracy and Mean Decrease Gini of Random Forests Model

As part of the random forests construction, several measures of variable importance can be defined. Variable importance is a measurement of how much influence an attribute has on the prediction accuracy [13]. At every split of random forests, one of the randomly chosen variables was used to form the split and there is a resulting decrease in the Gini index. The Gini based variable importance measure was defined as the sum of all decreases in the forest due to the given variable, normalized by the number of trees. According to Table 2 and Figure 3, the importance ranking of human action features of smartphones usage was X1 followed by x, y,z.

3.2. AdaBoost Model

The study also applied the AdaBoost classification analysis to make the classification of the train data and test data. The study set the category of train data as type 0 and test data as type 1. The "rattle" package of the R software randomly chose 7209 data as the test data (5107 Type 0 data, 2102 Type 1 data). The maximized depth was set as 30, the minimum split was set as 20 and the iterations were set as 50. The duration of calculation was 13.43 seconds. The error matrix of the random forests model for test data is as Table 3. The error during 50 iterations training process was shown in Figure 4.

Table 3. The Error Matrix of AdaBoost Model

Observed	Predicted		
	Type No.0	***Type No.1***	***Percentage error***
Type No. 0	5025	82	1.60%
Type No. 1	0	2102	0
Overall Error Percentage	1.10%		

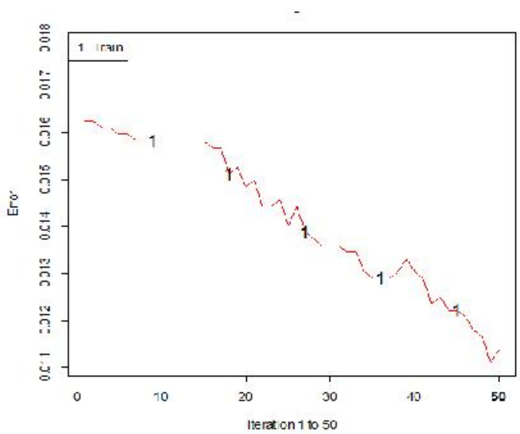

Figure 4. Training Error during 50 iterations

The error rate of AdaBoost model classification was 1.10%. According to Figure 4, the training error was decreased as the iterations time increased. It showed the machine learning process of the AdaBoost Model.

4. Discussion

The study compared the classification results of two models and found the AdaBoost model had better classification performance for human action recognition on smartphones. The error rate of AdaBoost model was 1.10% as compared with 1.65% of random forests model. The AdaBoost model had another advantage for calculation time. It only took 13.43 seconds for AdaBoost model calculation as compared with 9.29 minute of random forests model. The study also found the importance ranking of classification from random forests model. The variable X1 ranked as the first followed by three other variables (x, y, z). .

5. Conclusion

The contributions of the study were as follows. First, the study used two different models for the human actions recognition on smartphones by classifying the training and test data with related features. From the study, the study found the AdaBoost model with better classification results and less calculation time. The study offered more insights for related researches.

Acknowledgements

The study was partially funded by the Faculty Research Grant of Macau University of Science and Technology.

References

[1] Khan AM, Lee YK, Lee SY, Kim TS, The Fifth International Conference on Future Information Technology, Busan, Korea. 2010; 1-6.
[2] Landwehr N, Gutmann B, Thon I, Philipose M, De Raedt L, Proceedings of the International Workshop on Knowledge Discovery from Ubiquitous Data Streams, Warsaw, Poland. 2007; 83-94.
[3] Allen FR, Ambikairajah E, Lovell NH, Celler BG, Classification of a Known Sequence of Motions and Postures from Accelerometry Data using Adaptive Gaussian Mixture Models, Physiological Models. 2006; 27(10): 935.
[4] Anguita D, Ghio A, Oneto L, Parra X, Reyes-Ortiz JL, Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine, Lecture Notes in Computer Science. 2012, 7657: 216-223.
[5] Liaw A, Wiener M. Classification and Regression from Random Forest. 2002; 2(3): 18-22.
[6] Breiman L. Random Forests, Machine Learning. 2001; 45(1): 5-32.
[7] Zhou LF, Wang H, Loan Default Prediction on Large Imbalanced Data Using Random Forests, *TELKOMNIKA Indonesian Journal of Electrical Engineering.* 2012; 10(6): 1519-1525.
[8] Freund Y, Schapire RE, A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence*. 1999; 14(5): 771-780.
[9] Shin Y, Kim DW, Yang SW, Cho HH, Kang KI. *Decision Support Model using the AdaBoost Algorithm to Select Formwork Systems in High-Rise Building Construction.* The 25th International Symposium on Automation and Robotics in Construction, Vilnius, Lithuania. 2008: 644-649.
[10] Viola P, Jones M. *Rapid Object Detection using a Boosted Cascade of Simple Features.* Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. 2001: 511-518.
[11] Chen Y, Target Tracking Feature Selection Algorithm Based on Adaboost. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(12): 7373-7378.
[12] Wu X, Kumar V (eds.). The Top Ten Algorithms in Data Mining, Taylor & Francis. 2009.
[13] Treeratpituk P, Giles CL, Disambiguating Authors in Academic Publications using Random Forests, Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries. 2009: 39-48.