International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 3 Issue 6, October 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Daily Human Activity Recognition using Adaboost Classifiers on Wisdm Dataset

Khin Khin Oo

Lecturer, University of Computer Studies, Lashio, Myanmar

ABSTRACT

Human activity recognition is an important area of machine learning research as it has much utilization in different areas such as sports training, security, entertainment, ambient-assisted living, and health monitoring and management. Studying human activity recognition shows that researchers are interested mostly in the daily activities of the human. Nowadays mobile phone is well equipped with advanced processor, more memory, powerful battery and built-in sensors. This provides an opportunity to open up new areas of data mining for activity recognition of human's daily living. In the paper, the benchmark dataset is considered for this work is acquired from the WISDM laboratory, which is available in public domain. We tested experiment using AdaBoost.M1 algorithm with Decision Stump, Hoeffding Tree, Random Tree, J48, Random Forest and REP Tree to classify six activities of daily life by using Weka tool. Then we also see the test output from weka experimenter for these six classifiers. We found the using Adaboost,M1 with Random Forest, J.48 and REP Tree improves overall accuracy. We showed that the difference in accuracy for Random Forest, REP Tree and J48 algorithms compared to Decision Stump, and Hoeffding Tree is statistically significant. We also show that the accuracy of these algorithms compared to Decision Stump, and Hoeffding Tree is high, so we can say that these two algorithms achieved a statistically significantly better result than the Decision Stump, and Hoeffding Tree and Random Tree baseline.

How to cite this paper: Khin Khin Oo "Daily Human Activity Recognition using Adaboost Classifiers on Wisdm Dataset" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-6, October 2019, pp.205-209, URL: https://www.ijtsrd.com/papers/ijtsrd28073.pdf

IJTSRD28073

KEYWORDS: *Human activity recognition, machine learning, data mining, classifier, accuracy*

1. INTRODUCTION

Human Activity recognition (HAR) is the root of many applications, such as those which deal with personal biometric signature, advanced computing, health and fitness monitoring, and elder-care, etc. The input of HAR models is the reading of the raw sensor data and the output is the prediction of the user's motion activities. The HAR system becomes an emerging discipline in the area of pervasive computing in the intelligent computing applications. According to the World Health Organization (WHO), the number of diabetic patients among the world population drastically increases from time to time (WHO, 2016). In the world, the first time it is happening that the proportion of older persons (60 years or older) increases in the proportion of young (below 15). For the first time in history, the number of older persons in the world will exceed the number of young by year 2050. Such ageing population need care. Activity recognition is a significant research area can provide a solution to such problem. This area has many applications in healthcare, elder care, user interfaces, smart environments, and security. Image and video based human activity recognition has been studied since a long time but they have limitation of mostly require infrastructure support, for example, the installation of video cameras in the monitoring areas. There are alternative approaches are available such as a body worn sensors or a smart phone which have built-in sensors to recognize the human activity of daily living. But a normal human can't wear so many sensors on the body excluding a patient. Today's smartphone is well equipped with powerful sensors and long lasting battery with small in size provides an opportunity for data mining research and applications in human activity recognition using mobile phones. Some existing works have explored human activity recognition using data from accelerometer sensors. Many researches received very good accuracy by using tri-axial accelerometer for activity recognition the daily.

2. Sensor approaches

There are two types of sensors to recognize the human activities; using external or wearable sensors. In the past, the sensors were settled in predetermined points of interest, therefore the detecting of activities is essentially based on the interaction of the users with the sensors. One of the examples of external sensors applications is the intelligent home, which has a capability to identify the complicated activities, eating, taking a shower, washing dishes, etc., because they depend on data that is collected from various sensors which are placed in specific objects. Those objects are supported by peoples' interaction with them (e.g., stove, faucet, washing machine, etc.). However, there is no useful response if the user is out of the sensor area or the activities of the user do not need to interact with those objects. Moreover, the composition and servicing of sensors require high costs.

Also, some of the extensive researches have been focused on the recognition of activities and gestures from video sequences. This is most appropriate for security and interactive applications. Microsoft developed the Kinect game console that let the user interact with the game using the gestures without any controller devices. However, there are some issues in video sequences of HAR such as:

- The privacy, as no one wants to be always monitored and recorded by cameras.
- The pervasiveness, it is difficult to attach the video recording devices to the target of individuals in order to collect the images of their entire body during daily living activities.
- Video processing techniques are comparatively costly and consuming time.

The above-mentioned limitations motivate to use a wearable sensor in HAR. Where the measured attributes almost depend on the following: environmental variables (such as temperature and humidity), movement of the user (such as using GPS or accelerometers), or physiological signals (such as heart rate or electrocardiogram). These data are indexed over the time dimension.

Accelerometer sensors sense the acceleration event from mobile phone, WII remote or wearable sensors. The raw data stream from the accelerometer is the acceleration of each axis in the units of g-force. The raw data is represented in a set of 3D space vectors of acceleration. A time stamp can also be returned together with the three axes readings. Most of the existing accelerometers provide a user interface to configure the sampling frequency so that the user have to choose the best sampling rate which match his needs. There are many causes that encourage developing new techniques for enhancing the accuracy under more factual conditions. However, the first works on HAR date back to the late 90's.

3. Challenges face HAR system designers

Any HAR system design relies on the activities to be recognized. The activities kinds and complexity are able to affect the quality of the recognition. Some of challenges which face researches are (1) how to select the attributes to be measured, (2) how constructing the system with portable, unobtrusive, and inexpensive data acquisition, (3) how extracting the features and designing the inference methods, (4) how collecting the data in the real environment, (5) how recognizing activities of the new users without the need of re-training the system, and (6) how can be implemented in the mobile devices which meeting energy and processing limitations.

4. Offline versus online HAR systems

The recognition of human activity could be done using offline or online techniques. Whenever online processing is not necessary for the application, the offline processing can always be used. For example, if the tracking of person's daily routine is the goal such as in, the data was collected during the day by using the sensors and then it could be uploaded to a server at the end of the day. The data can be processed offline for classification purposes only.

However, some of the applications such as fitness coach where the user applies the given program which contains on a set of activities with sequence and duration. It is widely required to identify what the user is currently doing; therefore it requires using online technique.

Another application can be the recruitment for participatory sensing applications. For instance, the application aimed to collect the information from users during walking in a specific location in the city. Thus, online recognition of activities becomes significant. Some researches on human activities, which work on offline recognition, are using machine learning tools such as WEKA. Nowadays, some of clouding systems are being used for online recognition.

5. Data collection

In this paper, we have uses a standard HAR dataset which is publicly available from the WISDM group. Android smart phone based application was used to collect data. Each user was asked to take the smart phone in a front leg pocket and performed five different activities in supervised condition which were walking, jogging, walking upstairs, walking downstairs, sitting, and standing. While performing these activities, the sampling rate for accelerometer sensor was kept of 20Hz. WISDM HAR dataset consists the accelerometer's raw time series data and detail descriptions are shown in the Table 1.

Description	Nos. of Record	% of Records
Total Nos. of Samples	10,98,207	100%
Nos. of Attributes	6	
Any missing value	None	
Ativity wise distribution	**Total nos. of Samples**	**Percentage**
Walk	4,24,400	38.6%
Jog	3,42,177	31.2%
Up-stairs	1,22,869	11.2%
Down-stairs	1,00,427	9.1%
Sit	59,939	5.5%
Stand	48,395	4.4%
Transformed Examples		
Total Nos. of samples	5,424	
Nos. of attributes	46	
Any missing value	None	
Activity wise distribution	**Total nos. of samples**	**Percentage**
Walk	2,082	38.4%
Jog	1,626	30.0%
Up-stairs	633	11.7%
Down-stairs	529	9.8%
Sit	307	5.7%
Stand	247	4.6%

5.1. Feature generation

Before applying the classifier algorithm, it is necessary to transform the raw sensor's data. The raw accelerometer's signal consists of a value related each of the three axes. To accomplish this J. R. Kwapisz et al has segmented into 10-second data without overlapping. This is because he considered that 10seconds data consist of sufficient recreations that consist of 200 readings. Then they have generated features that were based each segment data of 200 raw accelerometer readings. A total 43 features are generated. All these are variants are based on six extraction methods. Average, Standard Deviation, Average Absolute Difference and Time between Peaks for each axis are extracted. Apart from these Average Resultant Acceleration and Binned Distribution is also extracted.

5.2. Classification

In this paper for classification of human activity of daily living, we have used the classifiers available in the Weka tool. In this paper, we have presented selected classifier algorithms like Decision Stump, Hoeffding Tree, Random

Tree, REP Tree, J48 and RAndom Forest, decision tree algorithms along with Adaptive Boosting available in Weka Adaboost.M1 with default setting.

5.3. Performance measure

During this experimentation following performance measures has been used. The Overall accuracy is used to summarize the overall classification performance for all classes. It is defined as follows:

- Overall accuracy=TP/ (TP+FP+FN+TN)
- Precision=TP/ (TP+FP)
- Recall=TP/ (TP+FN)
- Specificity=TN/ (TN+FP)

6. Experimental results

The experiments are performed by the following steps.

- Acquisition of standard WISDM HAR Dataset for Human Activity Recognition through a mobile device which is available in public domain.
- Partitioning dataset into training, testing and cross validation by using 10-fold cross-validation.
- A Selection of Meta Adaboost.M1 classifier for classification with selected decision tree classifier with default parameters.
- Examination of each classification model on 10-fold cross validation.
- Comparative analysis on the basis of performance measures such as, classification accuracy, TP rate, FP rate, minimum
- RMSE, F-measure, precision, recall and ROC.
- We used experiment environment from weka in determining mean and standard deviation performance of a classification algorithm on a WISDM dataset.
- We choose decision tree classifiers, experiment type has been chosen as 10-fold cross-validation in which WISDM dataset is divided into 10 parts (folds) and compare their results with meta classifier Adaptive Boosting. The confidence kept at 0.05.

Finally, we used weka experimenter to evaluate the performance of the classifiers mentioned in an earlier section on standard WISDM dataset. Each classifier is trained and tested using 10-fold cross validation with 10 times' repetition.

6.1. Confusion matrix for classifiers

The Confusion Matrix for Decision Stump, Hoeffding Tree, Random Tree, REP Tree, J48 and Random Forest are shown in the Table 2 to Table 7.

Table2. Confusion Matrix for Adaboost.M1 Meta Classifier with Decision Stump

classified as	a	b	c	d	e	f
a=Walking	2014	67	0	0	0	0
b=Jogging	185	1440	0	0	0	0
c=Upstairs	588	44	0	0	0	0
d=Downstairs	519	9	0	0	0	0
e=Sitting	306	0	0	0	0	0
f=Standing	246	0	0	0	0	0

Table3. Confusion Matrix for Adaboost.M1 Meta Classifier with Hoeffding Tree

classified as	a	b	c	d	e	f
a=Walking	1863	89	67	42	5	15
b=Jogging	53	1520	38	4	0	10
c=Upstairs	346	46	115	94	3	28
d=Downstairs	327	11	61	109	1	19
e=Sitting	0	0	1	0	288	17
f=Standing	0	0	19	3	25	199

Table4. Confusion Matrix for Adaboost.M1 Meta Classifier with Random Tree

classified as	a	b	c	d	e	f
a=Walking	2042	5	18	14	0	2
b=Jogging	10	1601	6	6	1	1
c=Upstairs	27	19	501	80	4	1
d=Downstairs	36	9	119	360	2	2
e=Sitting	1	0	2	1	299	3
f=Standing	1	2	5	2	6	230

Table5. Confusion Matrix for Adaboost.M1 Meta Classifier with REP Tree

classified as	a	b	c	d	e	f
a=Walking	2065	9	3	3	1	0
b=Jogging	32	1575	10	8	0	0
c=Upstairs	4	8	500	120	0	0
d=Downstairs	9	6	112	401	0	0
e=Sitting	2	1	6	3	292	2
f=Standing	5	2	6	9	2	222

Table6. Confusion Matrix for Adaboost.M1 Meta Classifier with J48

classified as	a	b	c	d	e	f
a=Walking	2019	11	23	26	2	0
b=Jogging	10	1585	16	14	0	0
c=Upstairs	39	30	445	115	2	1
d=Downstairs	34	16	83	390	4	1
e=Sitting	2	0	3	1	297	3
f=Standing	0	2	6	1	2	235

Table7. Confusion Matrix for Adaboost.M1 Meta Classifier with Random Forest

classified as	a	b	c	d	e	f
a=Walking	2048	2	15	14	0	2
b=Jogging	2	1605	11	6	1	0
c=Upstairs	15	13	516	85	2	1
d=Downstairs	29	7	95	393	3	1
e=Sitting	0	0	2	0	302	2
f=Standing	1	0	4	1	0	240

As shown a confusion matrix in the Table– 2 and performance criteria in table 8 for Decision Stump, the classifier found confused over the Jogging stairs standing and Laying Down. It is found that there is common misclassification of the stairs and sitting with walking has been observed. But still the performance of the REP Tree, J49 and Random Forest is much better compared with others.

6.2. Performance criteria for classifiers
The performance criteria for classifiers are as shown in Table 8 to Table 13.

Table8. Performance Criteria for Adaboost.M1 Meta Classifier with Decision Stump

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Walking	0.968	0.553	0.522	0.968	0.678	0.446	0.708	0.519
Jogging	0.886	0.032	0.923	0.886	0.904	0.865	0.93	0.876
Upstairs	0	0	?	0	?	?	0.624	0.152
Downstairs	0	0	?	0	?	?	0.651	0.135
Sitting	0	0	?	0	?	?	0.654	0.08
Standing	0	0	?	0	?	?	0.651	0.064
Weighted Avg.	0.638	0.222	?	0.638	?	?	0.753	0.5

Table9. Performance Criteria for Adaboost.M1 Meta Classifier with Hoeffding Tree

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Walking	0.895	0.218	0.72	0.895	0.798	0.66	0.91	0.851
Jogging	0.935	0.038	0.912	0.935	0.924	0.891	0.979	0.971
Upstairs	0.182	0.039	0.382	0.182	0.247	0.201	0.832	0.327
Downstairs	0.206	0.029	0.433	0.206	0.279	0.25	0.81	0.316
Sitting	0.941	0.007	0.894	0.941	0.917	0.912	0.998	0.979
Standing	0.809	0.017	0.691	0.809	0.745	0.735	0.991	0.783
Weighted Avg.	0.756	0.104	0.719	0.756	0.725	0.653	0.92	0.778

Table10. Performance Criteria for Adaboost.M1 Meta Classifier with Random Tree

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Walking	0.981	0.022	0.965	0.981	0.973	0.956	0.997	0.996
Jogging	0.985	0.009	0.979	0.985	0.982	0.974	0.999	0.998
Upstairs	0.793	0.031	0.77	0.793	0.781	0.752	0.976	0.857
Downstairs	0.682	0.021	0.778	0.682	0.727	0.701	0.975	0.813
Sitting	0.977	0.003	0.958	0.977	0.968	0.966	1	0.995
Standing	0.935	0.002	0.962	0.935	0.948	0.946	0.997	0.981
Weighted Avg.	0.929	0.017	0.927	0.929	0.928	0.913	0.993	0.962

Table11. Performance Criteria for Adaboost.M1 Meta Classifier with REP Tree

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Walking	0.992	0.016	0.975	0.992	0.984	0.974	0.997	0.992
Jogging	0.969	0.007	0.984	0.969	0.976	0.967	0.989	0.988
Upstairs	0.791	0.029	0.785	0.791	0.788	0.76	0.982	0.872
Downstairs	0.759	0.029	0.737	0.759	0.748	0.721	0.978	0.801
Sitting	0.954	0.001	0.99	0.954	0.972	0.97	0.991	0.981
Standing	0.902	0	0.991	0.902	0.945	0.943	0.946	0.915
Weighted Avg.	0.933	0.014	0.934	0.933	0.933	0.92	0.988	0.954

Table12. Performance Criteria for Adaboost.M1 Meta Classifier with J48

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Walking	0.97	0.025	0.96	0.97	0.965	0.943	0.995	0.993
Jogging	0.975	0.016	0.964	0.975	0.97	0.957	0.998	0.994
Upstairs	0.704	0.027	0.773	0.704	0.737	0.705	0.964	0.829
Downstairs	0.739	0.032	0.713	0.739	0.726	0.696	0.967	0.794
Sitting	0.971	0.002	0.967	0.971	0.969	0.967	0.994	0.972
Standing	0.955	0.001	0.979	0.955	0.967	0.966	0.996	0.988
Weighted Avg.	0.917	0.021	0.916	0.917	0.917	0.898	0.989	0.953

Table13. Performance Criteria for Adaboost.M1 Meta Classifier with Random Forest

Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area
Walking	0.984	0.014	0.978	0.984	0.981	0.969	0.998	0.997
Jogging	0.988	0.006	0.986	0.988	0.987	0.982	1	0.999
Upstairs	0.816	0.027	0.802	0.816	0.809	0.784	0.986	0.908
Downstairs	0.744	0.022	0.788	0.744	0.765	0.741	0.984	0.883
Sitting	0.987	0.001	0.981	0.987	0.984	0.983	1	0.999
Standing	0.976	0.001	0.976	0.976	0.976	0.974	1	0.995
Weighted Avg.	0.942	0.012	0.941	0.942	0.942	0.93	0.996	0.976

7. Conclusion
This paper surveys the state-of-the-art in human activity recognition based on measured acceleration components.
It can be concluded that the Random Forest, REP Tree and J48 algorithms which have a little "v" next to their results means that the difference in the accuracy of these algorithms compared to Decision Stump, and Hoeffding Tree is statistically significant. This paper also shows the accuracy of these algorithms compared to Decision Stump, and Hoeffding Tree is high, so it can be said that these two algorithms achieved a statistically significantly better result than the Decision Stump, and Hoeffding Tree and Random Tree baseline.

8. References

[1] Khan, Adil Mehmood and Lee, Young-Koo and Lee, Sungyoung Y and Kim, Tae-Seong [2010] A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer,
Information Technology in Biomedicine, *IEEE Transactions*;14:5–1166.

[2] Casale, Pierluigi, Pujol Oriol, Radeva Petia. [2011] Human activity recognition from accelerometer data using a wearable device, Pattern Recognition and Image Analysis, 289, Springer.

[3] A. G. Renehan, M. Tyson, M. Egger, R. F. Heller, M. Zwahlen, Bodymass index and incidence of cancer: a systematic review and metaanalysis of prospective observational studies, The Lancet 371 (2008) 569-578.

[4] A. P. C. S. Collaboration, et al., Body mass index and cardiovascular disease in the asia-pacific region: an overview of 33 cohorts involving 310 000 participants, International journal of epidemiology 33 (2004) 751-758.

[5] M. Javan Roshtkhari and M. D. Levine, "Human activity recognition in videos using a single example," Image Vis. Comput., vol. 31, no. 11, pp. 864–876, Nov. 2013.

[6] Redmon, J., Farhadi, A.: Yolo9000: Better, faster, stronger. In: Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, IEEE (2017) 6517–6525.

[7] Slim S. O., Atia A., Mostafa MS.M. (2016) An Experimental Comparison Between Seven Classification Algorithms for Activity Recognition. In The 1st International Conference on Advanced Intelligent System and Informatics (AISI2015), November 28-30, 2015, Beni Suef, Egypt. Advances in

Intelligent Systems and Computing, vol 407. Springer, Cham.

[8] T. HAYASHI, M. NISHIDA, N. KITAOKA, T. TODA and K. TAKEDA, "Daily Activity Recognition with Large-Scaled Real-Life Recording Datasets Based on Deep Neural Network Using Multi-Modal Signals", IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. 101, no. 1, pp. 199-210, 2018.

[9] T. Hayashi, M. Nishida, N. Kitaoka, and K. Takeda, "Daily activity recognition based on dnn using environmental sound and acceleration signals," European Signal Processing Conference, pp.2306–2310, IEEE, 2015.

[10] M. Saeed, J. Pietilä, and I. Korhonen. "An Activity Recognition Framework Deploying the Random Forest Classifier and A Single Optical Heart Rate Monitoring and Triaxial Accelerometer WristBand." Sensors (Basel, Switzerland), 18(2), 2018.

[11] Shirahama, K.; Grzegorzek, M. On the Generality of Codebook Approach for Sensor-based Human Activity Recognition. Electronics 2017, 6, 44.

[12] Kose, Mustafa.; Incel, O.D.; Ersoy, C. Online Human Activity Recognition on Smart Phones. In Proceeedings of the Workshop on Mobile Sensing: From Smartphones and Wearables to Big Data, Beijing, China, 16 April 2012; pp. 11–15.

[13] Dean M. Karantonis, Michael R. Narayanan, Merryn Mathie, Nigel H. Lovell, Branko G. Celler [2006] Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring, IEEE Transactions on Information Technology in Biomedicine, 10(1): 156–167

[14] N Ravi, N Dandekar, P Mysore, and ML Littman. [2005] Activity recognition from accelerometer data. IAAI-05: American Association for Artificial Intelligence.