2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)

Human Activity Recognition using Triaxial Acceleration Data from Smartphone and Ensemble Learning

Narit Hnoohom
Department of Computer Engineering,
Faculty of Engineering, Mahidol University,
Nakhon Pathom, Thailand
narit.hno@mahidol.ac.th

Sakorn Mekruksavanich
Department of Computer Engineering,
School of Information and Communication Technology,
University of Phayao, Phayao, Thailand
sakorn.me@up.ac.th

Anuchit Jitpattanakul
Department of Mathematics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok,
Bangkok, Thailand
anuchit.j@sci.kmutnb.ac.th

***Abstract*— In recent years, the use of smartphone sensors in human activity recognition (HAR) has been well studied. Mostly, a smartphone accelerometer has played the main role to solve the problem of HAR. However, the role of a gyroscope is to be explored, both when used alone as well as in combination with an accelerometer. For this purpose, the researchers investigated the role of these two smartphone sensors in human activity recognition. Two ensemble learning based approaches, i.e., majority voting and stacking, to improve recognition performance were presented. Also, the researchers evaluated the roles of the approaches on two body positions using the two ensemble classifiers while recognizing six physical activities, i.e., standing, sitting, laying, walking, walking upstairs, and walking downstairs. From the experimental results, it was shown that in general an accelerometer and a gyroscope complement each other, thereby making the recognition performance higher. Moreover, the ensemble learning based approaches could improve the recognition performance in terms of accuracy to 91.1667 percent.**

Keywords—activity recognition; smartphone; ensemble learning

I. INTRODUCTION

Physical activities play an important role in people's physical and mental wellbeing. The lack of physical activities can badly affect their well-being in general. Physical activity recognition is mainly divided into two categories: Vision-based [1] and non-vision based approaches. The vison-based activity recognition has limitations to its sensitivity of light in the environment and low detection range. At present, the activity recognition based on the non-vision approach is mainly based on various sensors on wearable devices.

Human activity recognition (HAR) is a research area that aims to capture the state of the human and its environment by exploiting heterogeneous sensors in order to provide resources for computing in various real-world applications. The sensors are attached to a human's body; such as, the waist, wrist, and other body parts. These sensors permit continuous monitoring of numerous physiological signals. This has appealing use in healthcare applications, e.g. daily activity monitoring for children, elder people, and people with disabilities. For example, medical staff can monitor health conditions from the information of their activity [2]

Nowadays, smartphones mostly incorporate many diverse and powerful sensors; such as, light sensors, temperature sensors, magnetic compasses, gyroscopes, and accelerometers, [3]. One of the most recent, challenging and appealing applications in the HAR consists of sensing human body motion using smartphones to gather information about the motion of people activities. The smartphones are generating new research in HAR, as the devices come with embedded built-in sensors. The use of smartphones with inertial sensors is an alternative solution for the HAR problem. Consequently, in the last few years, some works aiming to understand human behavior using smartphones have been proposed; for instance, [4] one of the first approaches to exploit an Android smartphone for HAR employing its embedded tri-axial accelerometers. The extended research has also been presented [5,6]. Improvements are still expected in topics; such as, in multi-sensor fusion for better HAR recognition performance with standardizing performance evaluation metrics [7]. In recent years, various machine learning algorithms have been used to build a HAR system to apply in automatic healthcare systems [8]; such as, support vector machine [9], back propagation neural network [10], decision tree [11], etc. In recent years, the combination of multiple learning, which is called ensemble learning, has become an alternative machine learning method. Many research studies have investigated and shown that ensemble learning can achieve a much better recognition performance than a single learning model [12, 13].

978-1-5386-4283-2/17 $31.00 © 2017 IEEE
DOI 10.1109/SITIS.2017.73

Therefore, the researchers assume that ensemble learning will improve the recognition performance of the HAR problem.

In this paper, the researchers focus on human activity recognition of basic activities that people perform on six daily activities as standing, sitting, laying, walking, walking upstairs, and walking downstairs. Two ensemble based approaches as majority voting and stacking were comparatively studied for human activity recognition from accelerometer data and gyroscope data obtained from a smartphone at the arm position and waist position.

The rest of this paper has been organized as follows: Section II where the framework of human activity recognition is introduced. Section III describes ensemble learning for HAR. The researchers describe an experimental design and show the experimental results in Section IV. Finally, the conclusion of the paper is provided in Section V.

II. THE FRAMEWORK OF HUMAN ACTIVITY RECOGNITION

Typically, activity recognition consists of four stages: preprocessing, feature extraction, model training, and classification. When putting these basic steps into a recognition system, a generic model can be used to explain the system's flow while highlighting important steps; such as, data acquisition, input data labels, feature extraction, where the selected machine learning algorithms and classification can be shown in Figure 1.

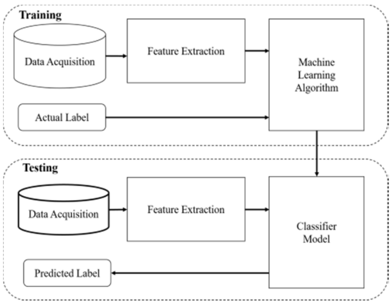

Fig. 1 Flow diagram of the machine learning for HAR

The following subsections present a brief summary of each process of human activity recognition proposed in this work.

A. Data Acquisition

In this work, the raw data was acquired from 10 voluntary subjects aged between 18 to 25 years. The majority of these subjects were females. Each subject wore one monitoring device at the waist position and one monitoring device at the arm position while performing six different types of activities; i.e., walking, walking upstairs, walking downstairs, standing, sitting, and laying. The device used for data collection was the iPhone7 running on an iOS version 11.0.1. The smartphone applications utilized in this study were SensorLog that can be downloaded from the App Store. This application provides data from the smartphone's build-in sensors, which are utilized to collect the triaxial accelerometer data and gyroscope data, with a sample rate of 50 Hz. For every activity of each subject, 10 samples were generated for each Cartesian axis, i.e., acceleration in the x, y, z-direction and gyroscope in the x, y, z-direction. Thus, the dataset contained a total of 600 human activity samples. Examples of the accelerometer data and gyroscope data are shown in Figures 2 and 3.

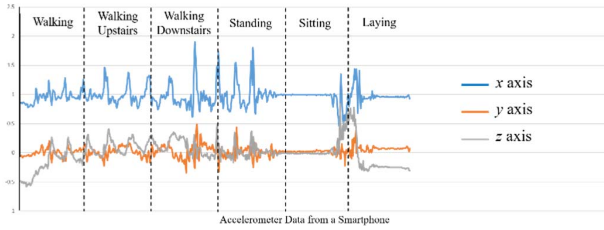

Fig. 2 An example of the accelerometer data of the six activities

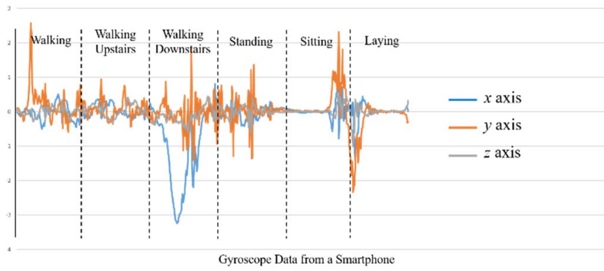

Fig. 3 An example of the gyroscope data of the six activities

B. Feature Extraction

The accelerometer data and gyroscope data from the sensor could not be directly used for classification. The feature extraction could be seen as a data pre-processing step where different types of features were extracted from the raw data. A proper feature set made the machine learning model more effective.

Suppose that S_t represents the data fragment in a data window, S_i represents the ith data obtained by the accelerometer or gyroscope sensor during the time. The nine features function used in this paper were calculated by formulas shown in Table1.

C. Machine Learning and Classification

Human activity recognition is usually treated as a supervised learning problem. Some of the most common supervised learning models include decision tree classifier, support vector machine, and multilayer perceptron. Certain classifiers were more accurate when determining specific activities;, for example, multilayer perceptron models were better at recognizing slightly more complex activities; such as, walking upstairs and walking downstairs [14]. For the simpler activities; such as, sitting and standing, a decision tree was found to achieve better accuracy [15].

III. ENSEMBLE LEARNING FOR HAR

In the present study, the researchers performed the experiments on the various machine learning methods; both the basic machine learning model and ensemble machine learning model were used to evaluate the recognition performance of the HAR problem. Three basic machine learning models, decision tree, multilayer perceptron, and support vector machine, were built for human activity recognition. The three basic learning models with two ensemble based approaches, i.e., majority voting and stacking were then combined.

Table 1. Types of features used for the HAR in this work

Domain	Feature	Formula
Time	Mean value	$mean = \frac{1}{N}\sum_{i=1}^{N} S_i$
Time	Standard deviation	$std = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(S_i - mean\right)^2}$
Time	Median absolute deviation	$mad = median\left(\lvert S_i - median(S_i)\rvert\right)$
Time	Maximum value	$max = \mathrm{MAX}(S_i)$
Time	Minimum value	$min = \mathrm{MIN}(S_i)$
Time	Signal magnitude area	$sma_{xyz} = \frac{1}{3}\left(\sum_{j=1}^{N}\lvert S_{xi}\rvert + \sum_{j=1}^{N}\lvert S_{yi}\rvert + \sum_{j=1}^{N}\lvert S_{zi}\rvert\right)$
Time	Energy measure	$iqt = Q_3(S_t) - Q_1(S_t)$
Time	Interquartile range	$energy = \frac{1}{N}\sum_{i=1}^{N} S_i^2$
Other	Correlation between axis	$c_{xy} = \frac{\sum_{i=1}^{N}\left(S_{xi} - mean(S_x)\right)\left(S_{yi} - mean(S_y)\right)}{\sqrt{\sum_{i=1}^{N}\left(S_{xi} - mean(S_x)\right)^2 \cdot \sum_{i=1}^{N}\left(S_{yi} - mean(S_y)\right)^2}}$

1) Decision Tree (DT): Decision trees are one of the most utilized classification techniques in practical applications of machine learning [16]. They are often used to obtain underlying patterns established with an attribute as the root and others as extended branches whose leaf nodes can be regarded as categories to express a concept; namely, intuitive cognition of things. The rationality of the attributes is the key to the comprehensibility of decision trees.

2) Multilayer Perceptron (MLP): A neural network is a virtual intelligence tool that can be used to simulate the human brain to perform analysis and generate results [17]. This machine learning algorithm has excellent abilities in non-linear mapping, generalization, and self-organization. MP is a multilayer neural network to give reasonable answers when presented with inputs that have never been seen by the neural network. This generalization property makes it possible to train a network on a representative set of input-output pairs and obtain good results training the network on all possible input-output pairs.

3) Support Vector Machine (SVM): The support vector machine (SVM) was originally introduced to solve the two-class pattern recognition problem [18]. The main idea is to build a hyperplane that separates the positive and negative examples while maximizing the smallest margin. SVM has two importance advantages: (1) feature selection is often not needed, and (2) no effort in parameter tuning is needed.

To increase the accuracy, combining the basic machine learning models is an alternative approach for the HAR problem. The alternative approach is well-known as ensemble learning. Many studies have demonstrated that the combination of the multiple base learning model significantly improves accuracy than any one of the base learning models [18]. In this work, the researchers solved the HAR problem with two ensemble learning algorithms, i.e., majority voting and stacking. Each ensemble method was a combination of the three basic machine learning models as DT, MP, and SVM.

4) Majority Voting (MV): Majority voting is the most common ensemble method, which does not require any parameter turing once the base classifiers have been constructed [20]. Assuming $d(y)$ is the domain of the class label y, $y_k(x)$ is the class label of the fundus image x assigned by kth base model, and $v(y_k(x), c_i)$ is the indicator function:

$$v(y_k(x), c_i) = \begin{cases} 1, & y_k(x) = c_i \\ 0, & y_k(x) \neq c_i \end{cases} \tag{1}$$

The formula to compute $c(x)$ assigned to an unlabeled fundus image x is given as below:

$$c(x) = \arg\max_{c_i \in d(y)} \left(\sum_k v(y_k(x), c_i) \right) \tag{2}$$

5) Stacking (ST): The stacking ensemble method employs a meta classifier that consumes the classification results of multiple base classifiers to generate the final classification [21]. The implementation of stacking uses a two-layer frame structure, where a number of base classifiers is generated from the training dataset in the level-0 layer. These individual classifiers are then combined by the meta classifier in the level-1 layer that can be a decision tree (called ST-DT), multilayer perceptron (ST-MLP), or support vector machine (ST-SVM). Figure 4 shows the two-layer stacking diagram of the ensemble learning method used in this work for the HAR problem.

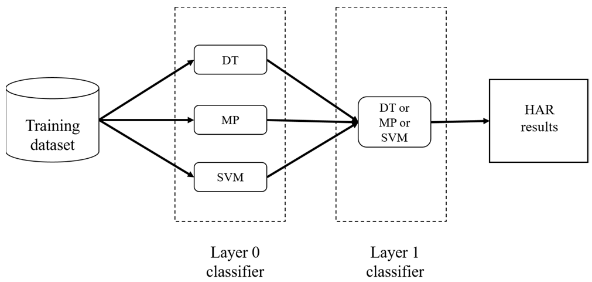

Fig. 4 Two-layer stacking diagram of the HAR problem

IV. EXPERIMENTAL DESIGN AND RESULTS

In order to evaluate ensemble learning methods to the collected data, the researchers used a WEKA machine learning tool (Waikato Environment for Knowledge Analysis). A personal computer used for the experiment had an i7-4500 CPU, 8 GB memory, and Microsoft Windows 10 as the operation system. Then seven classifiers as DT, MLP, SVM, MV, ST-DT, ST-MLP, and ST-SVM were applied on the activity dataset collected from 10 different subjects as described in Section II to see their performance. The researchers used a 10-fold cross validation technique to evaluate these classifiers. In cross validation, each fold or part of data contains all classes in equal proportion to ensure fairness. Seven commonly used classifiers were evaluated as listed in Table 2. Then short notations were provided in this table for these classifiers in all the next sections.

Table 2. Seven classifiers used in this work

Type of Classifiers	WEKA-version	Notation
Decision trees	J48	DT
Neural networks	MulitlayerPerceptron	MLP
Support vector machines	LibSVM	SVM
Majority voting	Vote	MV
Stacking	Stacking with J48	ST-DT
Stacking	Stacking with MLP	ST-MLP
Stacking	Stacking with SVM	ST-SVM

In the experiment, the accuracy value was calculated to measure the performance of the human activity recognition. These values were calculated from the value of true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*) defined by the formula given below:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}. \tag{3}$$

The researchers used the accelerometer data and gyroscope data of a smartphone at two different positions (arm and waist) to represent each activity. The six human activities were evaluated in combination with the seven classifiers in Table 2, reasonable results were obtained and shown in Tables 3 and 4.

Table 3. Experimental results of the recognition performance at the arm position

Type	Classifers	Recognition Performance (% Accuracy) at Arm Position		
		Acc. data	Gyro. data	Acc. and Gyro. data
Basic	DT	83.6667	58.0000	88.1667
	MLP	86.5000	58.8333	88.5000
	SVM	69.3333	55.6667	81.8333
Ensemble	MV	**88.0000**	61.5000	**90.6667**
	ST-DT	83.0000	59.1667	89.1667
	ST-MLP	86.0000	59.6667	88.3333
	ST-SVM	86.6667	**62.1667**	89.6667

Table 3 provides a summary of the overall recognition performance at the arm position in terms of accuracy. From this table, it can be seen that MV has the highest recognition performance for using both the accelerometer data and gyroscope data with an accuracy of 90.6667 percent.

Table 4 shows a summary of the improved performance of each classifier using an activity dataset obtained from the sensors on a smartphone at the waist position. Based on the experimental results, all seven classifiers had a higher recognition performance. From this table, it can be seen that MV had the highest recognition performance for using both the accelerometer data and gyroscope data with an accuracy of 91.1667 percent.

Table 4. Experimental results of recognition performance at the waist position

Type	Classifers	Recognition Performance (% Accuracy) at Waist Position		
		Acc. data	Gyro. data	Acc. and Gyro data
Basic	DT	82.6667	61.6667	84.6667
	MLP	87.6667	62.5000	88.1667
	SVM	75.3333	52.1667	79.3333
Ensemble	MV	86.8333	62.6667	**91.1667**
	ST-DT	83.6667	58.3333	88.6667
	ST-MLP	87.8333	61.3333	91.0000
	ST-SVM	**88.1667**	**68.0000**	90.5000

Table 5. Best classifier at the arm position of each activity

Type	Classifers	Best Classifier at Arm Position of each Activity					
		Walking	Walking upstairs	Walking downstairs	Sitting	Standing	Laying
Basic	DT	88.00	83.00	83.00	87.00	75.00	95.00
	MLP	92.00	90.00	90.00	**91.00**	72.00	**96.00**
	SVM	75.00	80.00	84.00	77.00	82.00	93.00
Ensemble	MV	93.00	**93.00**	**93.00**	**91.00**	**85.00**	95.00
	ST-DT	**94.00**	87.00	90.00	90.00	79.00	95.00
	ST-MLP	92.00	87.00	90.00	88.00	78.00	95.00
	ST-SVM	**94.00**	88.00	81.00	85.00	**85.00**	95.00

Table 5 shows a summary of how well each classifier performed with respect to the activity collected from a smartphone at the arm position. Based on the results, the majority voting model performed with the highest accuracy on four of the six activities. It was found that when classifying the laying, the multilayer perceptron yielded the highest accuracy with 96.000 percent.

Table 6. Best classifier at the waist position of each activity

Type	Classifers	Best Classifier at Waist Position of each Activity					
		Walking	Walking upstairs	Walking downstairs	Sitting	Standing	Laying
Basic	DT	90.00	69.00	80.00	85.00	83.00	100.00
	MLP	95.00	86.00	90.00	81.00	77.00	**100.00**
	SVM	88.00	67.00	85.00	61.00	76.00	99.00
Ensemble	MV	96.00	84.00	**96.00**	83.00	**88.00**	**100.00**
	ST-DT	**97.00**	**88.00**	92.00	83.00	73.00	99.00
	ST-MLP	96.00	87.00	93.00	84.00	84.00	99.00
	ST-SVM	95.00	**88.00**	95.00	**86.00**	80.00	99.00

Table 6 shows a summary of how well each classifier performed with respect to the activity collected from a smartphone at the waist position. Based on the results, the majority voting model performed with the highest accuracy on four of the six activities. It was found that when classifying layiny, the multilayer perceptron yielded the highest accuracy with 100.000 percent.

V. CONCLUSION

This paper regarding the ensemble learning algorithms used in human activity recognition using a smartphone would be very important for researchers to get a clearer picture from the experimental results. The researchers investigated the role of the traiaxial accelerometer and gyroscope in smartphones to human activity recognition. From the experimental results, The researchers showed that in general an accelerometer and a gyroscope complement each other, thereby making the recognition process more reliable. Moreover the ensemble learning based approaches could improve the recognition performance in term of accuracy to 91.1667 percent.

ACKNOWLEDGMENTS

This paper was supported by the Department of Computer Engineering, Faculty of Engineering, Mahidol University and the Faculty of Applied Science, King Mongkut's University of Technology North Bangkok under Grant No. 6142101.

REFERENCES

[1] R. Poppe, "A survey on vision-based human action recognition", Image and Vision Computing, vol. 28, no. 6, pp. 976-990, 2010.

[2] H. Ghasemzadeh and R. Jafari., "Physical movement using body sensor networks: A phonological approach to construct spatial decision trees", IEEE Transactions on Industrial Informations, vol. 7, no. 1, pp. 66-70, 2011.

[3] J. Kwapisz, G. Weiss, and S. Moore, "Activity recognition using cell phone accelerometers," SIGKDD Explorations, vol. 12, no. 2, pp. 74-82, 2010.

[4] J.R. Kwapisz, G.M. Weiss, and S.A. Moore, "Activity recognition using cell phone accelerometers", *SIGKDD Explorations Newsletter*, vol. 12, no.2, 2011,pp. 74–82.

[5] T. Brezmes, J.L. Gorricho, and J. Cotrina, "Activity recognition from accelerometer data on a mobile phone", *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, 2009, pp. 796–799.

[6] W. Wu, S. Dasgupta, E.E. Ramirez, C. Peterson, and G.J. Norman, "Classification accuracies of physical activities using smartphone motion sensors", *Journal of Medical Internet Research*, vol. 14, no.5, 2012.

[7] M.F.A. bin Abdullah, A.F.P. Negara, M.S. Sayeed, D.J. Choi, and K.S.Muthu, "Classificationalgorithms in human activity recognition using smartphones", *International Journal of Computer and Information Engineering*, vol. 6, 2012, pp. 77–84.

[8] L. Bao and S.S. Intille, "Activity recognition from user-annotated acceleration data", In T. Kanade, J. Kittler, J.M. Kleinberg, F. Mattern, J.C. Mitchell, O. Nierstrasz, C. Pandu Rangan, B. Steffen, D. Terzopoulos, D. Tygar, M.Y. Vardi, and A. Ferscha, editors, *Pervasive Computing*, 2004, pp. 1–17. 2004.

[9] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J.L. Reyes-Ortiz, "Human activity recognition on smart-phones using a multiclass hardware-friendly support vector machine", In *Proceedings of the Interna- tional Workshop of Ambient Assited Living*, 2012.

[10] J.Y. Yang, J.S. Wang, and Y.P. Chen, "Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers", *Pattern recognition letters*, vol.29, no.16, 2008, pp.2213– 2220.

[11] D.M. Karantonis, M.R. Narayanan, M. Mathie, N.H. Lovell, and B.G. Celler, "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring", *IEEE Transactions on Information Technology in Biomedicine*, vol.10, no.1, 2006, pp.156–167.

[12] G. Brown, J.L. Wyatt, and P. Tino, "Managing diversity in regression ensembles", Journal of Machine Learning Research, Vol.6 pp.1621-1650, 2005.

[13] E.K. Tang, P.N. Suganthan, and X. Yao, "An analysis diversity in measures", Machine Learning, Vol. 65. No.1, pp. 247-271, 2006.

[14] J. R. Kwapisz, G.M. Weiss, and S.A. Moore, "Activity recognition using cell phone accelerometers", ACM SIGKDD Explor. Newskett., vol. 12, no. 2, pp. 74-82, 2010.

[15] P. Siirtola and J. Roning, "Ready-to-use activity recognition for smarphones", in Proceeding of IEEE Symposium of Computing Intelligence and Data Mining (CIDM), Singapore, pp. 59-64.

[16] P.E. Utgoff, "Incremental induction of decision trees' *Machine learning*, vol.4, no.2, 1989, pp.161-186.

[17] L. V. Fausett, Fundamentals of Neural Networks: Architectures, Algorithms, and Applications, Prentice-Hall, 1994

[18] V.N. Vapnik, "The nature of statistical learning theory" Springer-Verlag New York Inc., USA, 1995.

[19] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, and A. Ferscha, "Collecting complex activity data sets in highly rich networked sensor envi- ronments", In *Proceedings of the 7th International Conference on Networked Sensing Systems 2010*, 2010.

[20] L. Rokach, "Ensemble-based classifiers", Artificial Intelligence Review, vol.33(1-2), 2010, pp. 1-39.

[21] [R. Polikar, "Ensemble based systems in decision making", IEEE Circ. Syst. Mag vol. 6, no.3, 2006, pp. 21-45.