Utilizing Smartphone Sensors for Daily Physical Activity Recognition

Rajeev Piyare[1]*, Min A Jung[2] and Seong Ro Lee[3], *Member, IEEE*
[1,3]Department of Electronics Engineering, [2]Department of Computer Engineering
Mokpo National University
Mokpo, Korea South
rajeev.piyare@hotmail.com*; majung@mokpo.ac.kr; srlee@mokpo.ac.kr

***Abstract*—Smartphones with built-in sensors promise a conducive, objective way to quantify everyday body movements and classify those movements into activities. Utilizing smartphone accelerometer data we estimate the following daily activities performed by the user: walking, jogging, using stairs, sitting, standing and lying down. The proposal is tested experimentally via evaluations on real data obtained from 50 test users. The evaluation indicates that the J48 classifier using a window size of 512 samples with 50% overlapping yields the highest accuracy (i.e., up to 96.02%).**

***Keywords*—daily activity recognition; feature extraction; WEKA; machine learning**

I. Introduction

Context- awareness is one of the important components of ubiquitous computing and HAR (human activity recognition) has originated as an essential way to classify the user's context for automatic service delivery in ubiquitous application. In ubiquitous healthcare applications, recognition and pattern analysis of everyday activities could enable such systems to monitor and predict any changes in daily behavior of an elderly that might be the indicators of developing physical or mental medical conditions. In addition to, patients with heart disease, diabetes, or obesity are often required to follow a well-defined exercise routine as part of their treatment. Therefore, recognizing activities such as resting, jogging, walking or running becomes quite useful to provide feedback to the caregiver about the patient's behavior.

Mainly two approaches have been employed to achieve the goal of recognizing activities of daily living. These are use of external sensors (sensors attached to predetermined points) and wearable sensors (sensors attached to the users). Smart homes [1-3] equipped with sensors embedded in everyday objects are a typical example of external sensors. Such systems are able to recognize finer grain activities (e.g. sitting eating, sitting reading, fall detection, washing dishes, etc.) because they rely on data from numerous sensors placed in target objects which people are supposed to interact with. The foremost problem with using the external approach is its lack of pervasiveness, i.e. it forces the user to stay within a perimeter defined by the position and the capabilities of the sensors. Additionally, the installation and maintenance of the sensors lead to increase the cost of the system. The majority of approaches in human activity recognition have relied on multiple wearable sensors attached at different locations of the body to detect everyday tasks such as sitting, walking, and running, ascending and descending stairs and jogging [4, 5]. To integrate activity recognition into the daily life of the intended users, unobtrusiveness is an important characteristic that needs to be considered. Particularly, the acquisition of sensor data and context information, using multiple body worn sensors are often obtrusive and can cause inconvenience and discomfort to the users. This may likely lead to the rejection of the newly introduced technology by the users. Therefore, there is a need to investigate new approaches that are unobtrusive and highly to be accepted by the intended users without altering his or her normal habits and lifestyle. Thus, the issue to be addressed in this work is the use of a smart phone as a suitable sensor and computing device for real-time human activity and context recognition.

Motivated by lack of comprehensive approach in online smartphone based activity recognition research, we propose a use of smartphone accelerometer for recognizing daily activities. We define the HAR problem as following:

Given a set $W=\{W_0,\ldots,W_{m-1}\}$ of m equally sized time windows, totally of partially labeled and such that each W_i contains a set of time series $Si=\{S_{i,0,\ldots,}S_{i,k-1}\}$ from each of the k measured attributes, and a set $A=\{a_0,\ldots,a_{n-1}\}$ of activity labels, the goal is to find a mapping function $f{:}S_i \rightarrow A$ that can be evaluated for all possible values of S_i, such that $f(S_i)$ is as similar as possible to the actual activity performed during W_i.

II. Materials and Methods

A. Proposed Architecture for Activity Recognition using Smartphone

We propose architecture (Fig.1) that can be used to design a user-centric context aware system. The proposed activity recognition system consists of three components: to collect sensor data using smartphone, preprocess the data and recognize a user's activity. The first step is modeling process (feature computation, selection and extraction –indicated by blue lines). The data from the smartphone is prepared and processed by a selected learning algorithm. We define these data as training data. The learning algorithms discover useful information in the data. This information is then used to build models. These models are defined as classifiers. In the second process (indicated by red line) the acceleration data are measured and then pre-processed to produce test data. The test

This work was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2009-0093828) and the MSIP (Ministry of Science, ICT and Future Planning) (No. 2011-0029321), Korea, under the C-ITRC (Convergence Information Technology Research Center) support program (NIPA-2014-H0401-14-1009) supervised by the NIPA (National IT Industry Promotion Agency).

ISSN 1975-4736 ©MITA 2014

data are then sent to the designated classifier to produce the recognized activity.

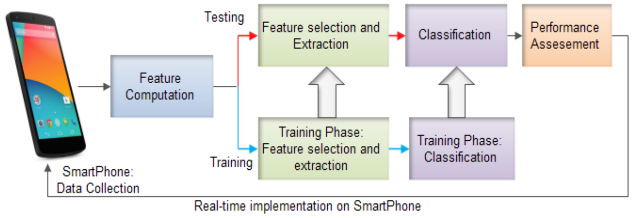

Fig. 1. Activity Recognition system overview.

B. Device and Data Collection

For this research work, activity data was collected from 50 healthy subjects (30 males and 20 females) between the ages of 21 and 35 years old, with an average height of 172 cm and average weight of 67 kg. Six common daily activities were selected: walking, jogging, using stairs, sitting, standing and lying down. A custom build Android application was used for data collection and annotation. Subjects carried the Android phone in their front pants leg pocket and were trained on the use of this application prior to data collection. The acceleration signals were sampled at 20Hz from which features were extracted and stored on a SD card in the smartphone. This sampling frequency has shown to be sufficient to capture most of the everyday activities [6]. Moreover, lower sampling frequency allows reducing the computational load of the smartphone in terms of data collection and storage. The feature extracted data stored on the SD card has been used for building classification models using WEKA (Waikato Environment for Knowledge analysis) [7]. The classifiers were trained offline, and the best classifier was transferred to the smartphone after cross validation and non-parametric statistical test.

III. Data Processing and Classification

A. Feature Extraction

Feature extraction is an important step in any accelerometer based HAR system. Mobile phones are energy constrained devices therefore, it is essential to prolong the device battery life, utilizing features that are energy efficient (light weight) and accurate to preserve battery life and ensure high accuracy. Vast number of time and frequency domain features has been reported in the previous studies with varying success rate. The most widely used time domain features include: mean [4, 8, 9], variance [4, 8] or standard deviation [4, 8], energy, entropy correlation between axes [4, 8], signal magnitude area and so on. The popular frequency domain features are FFT (Fast Fourier Transform) and DCT (Discrete Cosine Transform) coefficients [10].

Frequency domain features require longer time windows which increases the computational cost, thus are not suitable for real-time applications. While time domain features can be easily extracted in real-time. Due to these advantages, for this study, simple time domain features were chosen. Time domain statistical features were extracted from acceleration data using a window frame of 512 samples with 256 samples overlapping between consecutive windows. Feature extraction on sliding windows with a 50% overlap has demonstrated reasonable results in previous works [4, 8, 11]. Five features have been selected to be evaluated. These are: mean, standard deviation, MAD (mean absolute deviation), time between peaks and the resultant magnitude. These features were extracted from each window, giving a total of 13 attributes. The following Equations (1)-(4) denote the features that were calculated from the datasets. These features are then used as an input for WEKA data mining software to train and build the classifiers.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$\sigma_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2}$$

$$MAD_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}|x_i - \bar{x}|} \tag{3}$$

$$mag = \sqrt{x^2 + y^2 + z^2} \tag{4}$$

B. Classification Models

Various classification models have been applied to problem of activity recognition. However, there is no universally accepted method of detecting a particular range of activities and all these techniques have associated limitations and benefits. In order to identify which machine learning algorithm provided the most accurate activity detection, J48 (C4.5 Decision Tree), BN (Bayesian Network), MLP (Multilayer perceptron), NB (Naïve Bayes), RT (Random tree), RBFNet (Radial Basis Function), SMO (Sequential Minimal Optimization) and Logistic Regression were applied to data. However, in order to choose the most accurate one by comparing the two classifiers, a 5x2-fold cross validation along with a paired *t*-test as recommended in [12] was subsequently performed. A paired *t*-test was performed on the results to identify if the percentage of correctly classified instances was significantly different using the J48 as the baseline scheme, with the other seven algorithms being compared to it. A value of $p < 0.05$ was considered statistically significant. The *5x2cv* $\tilde{t}$ statistic is defined as follows:

$$\tilde{t} = \frac{p_1^{(1)}}{\sqrt{\frac{1}{5}\sum_{i=1}^{5} s_i^2}} \tag{5}$$

Where $p_i^{(j)}$ is the difference between the accuracies of both classifiers in the *j-th* iteration for $1 \le j \le 2$ and the *i-th* replication; $s_i^2 = (p_i^{(1)} - \bar{p})^2 + (p_i^{(2)} - \bar{p})^2$ is the estimated variance from the *i-th* replication for $1 \le i \le 5$; and $\bar{p} = (p^{(1)} + p^{(2)})/2$. Under the null hypothesis, $\tilde{t}$ follows the Student's t distribution with 5 degrees of freedom. It has been proven that the 5x2-fold cross validation with the paired *t*-test is more powerful than the non-parametric McNemar's test and provides a better measure of the variations due to the choice of the training set [12].

IV. RESULTS AND DISCUSSION

Two types of analyses have been proposed to evaluate activity recognition systems: subject-independent and subject-dependent evaluations [13]. In the former, only one classifier is built for all individuals and the evaluation is carried out by cross validation analysis. In the second one, a classifier is trained and tested for each individual. For this study, we have carried out both the evaluations.

A. Subject-Independent Analysis

There were approximately 6,000 samples for each activity with a total number of 1,800,000 samples. The resulting examples contain 5656 instances, 13 attributes and cover 50 users. In total eight classifiers were evaluated with five different random seeds $s_i \in \{1, 128, 255, 1023, 4095\}$. Table I shows the results of the 5x2-fold cross validation for the accelerometer dataset. Note that the highest overall accuracy was achieved by the J48 classifier (i.e., 95.156%). In order to find the best classifier for the dataset, a paired two-tailed *t*-test was performed between the J48 and all other classifiers with a significance level $\alpha = 0.05$. The null hypothesis is that each pair of classifiers achieved the same mean accuracy. As a result of the tests, the *p*-values are included in the last column of Table I. Since all the *p*-values are below the significance level, there is strong statistical evidence that J48 is more accurate than all other classifiers in the tested dataset.

Since, 2-fold cross validation takes into account only half of the dataset during training, it is essential to point out that a 5x2-fold cross validation is not performed to measure the classification accuracy but to rather find differences in the overall accuracy of the classifiers. To measure the actual classification accuracy, a 5x10-fold cross validation was performed on the dataset. After evaluating the best classifiers in each dataset for all five random seeds, the overall accuracy for J48 classifier reached 96.02%. A more detailed analysis was carried out for each activity by calculating a number of performance metrics: *precision, recall, F-measure, False Positive Rate, and False Negative Rate.* Fig. 2 compiles the results of the evaluation metrics.

TABLE I. CLASSIFICATION ACCURACY OF 5X2-FOLD CROSS VALIDATION (UNITS: %)

	s_1	s_2	s_3	s_4	s_5	Avg.	*p*-value
BN	76.82	77.81	77.19	77.38	77.29	77.30	<0.001
MLP	93.90	94.44	93.86	93.86	94.14	94.04	0.001
NB	58.06	57.60	57.42	57.70	56.48	57.45	<0.001
J48	**94.97**	**95.15**	**95.03**	**95.40**	**95.20**	**95.16**	-
RT	93.67	94.40	94.48	94.67	94.51	94.35	0.004
RBFNet	72.02	71.79	71.03	73.03	72.77	72.13	<0.001
SMO	89.48	89.78	90.11	90.27	89.71	89.87	<0.001
Logistic	91.90	92.60	92.45	92.71	91.77	92.29	<0.001

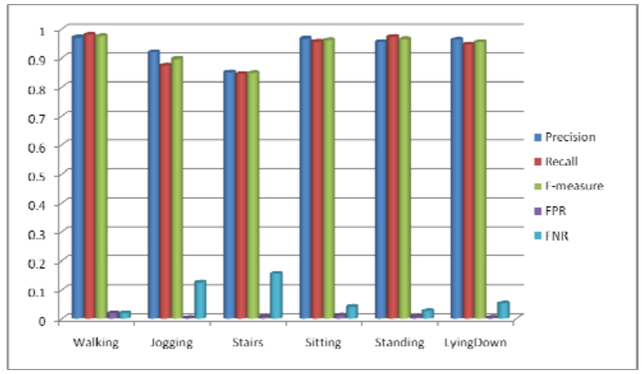

Fig. 2. Classification Accuracy: Precision, Recall, FPR, FNR, and F-measure.

The confusion matrix for the best classifier, J48, is shown in Table II after five iterations using five different random seeds. As we took a closer look at the classification results, not all activities had equally high recognition accuracy. The recognition accuracy for activity, walking, sitting, standing and lying down using J48 classifier was higher than 94%. The confusion matrix could be used to understand how well a certain classifier had performed. For example, by comparing the ground truth and the classification outcomes for each movement, 2204 instances of walking were correctly classified, while others were misclassified as jogging (2 instances), stairs (20 instances), sitting (14 instances), standing (6 instances) and lying down (4 instances).

B. Subject-dependent Analysis

Given that the training phase is generally more expensive than the evaluation; the classifier implemented on the smartphone was trained a priori in WEKA using J48 classifier and then transferred to the smartphone. The results presented in this section were accomplished online where feature extraction and performance metrics were computed on the smartphone. Thus, the results presented here are more realistic.

For this experiment, two individuals *A* and *B* who were new to the system were employed. Each individual performed all the six activities in a naturalistic environment. In training mode, each user set the real activity as the ground truth and performed said activity for a certain amount of time. Meanwhile, the smartphone application computes the classified activity using J48 algorithm and displayed it on the phone. Table III (A) and (B) display the confusion matrices for J48 classifier for both individuals.

TABLE II. CONFUSION MATRIX FOR THE BEST CLASSIFIER *J48*.

	Predicted Class (Overall Accuracy: 96.02%)						
	Walk	Jog	Stairs	Sit	Stand	Lay Down	% Correct
Walk	2204	2	20	14	6	4	97.96
Jog	6	161	9	3	2	3	87.50
Stairs	31	5	212	1	2	0	84.46
Sit	18	4	7	1425	22	11	95.83
Stand	5	1	1	12	842	4	97.34
LayDown	5	2	0	19	6	587	94.83

ISSN 1975-4736 ©MITA 2014

TABLE III. CONFUSION MATRIX FOR INDIVIDUALS *A* AND *B*

(A)

	Predicted Class (Overall Accuracy: 92.36%)					
	Walk	Jog	Stairs	Sit	Stand	LayDown
Walk	30	0	1	0	0	0
Jog	0	19	0	0	0	0
Stairs	0	1	39	0	0	0
Sit	1	0	0	7	5	0
Stand	0	0	0	3	62	0
LayDown	0	0	0	2	0	0

(B)

	Predicted Class (Overall Accuracy: 97.30%)					
	Walk	Jog	Stairs	Sit	Stand	LayDown
Walk	60	0	0	0	0	1
Jog	0	12	0	1	0	0
Stairs	0	0	4	0	0	0
Sit	0	0	0	18	0	0
Stand	0	0	0	0	0	0
LayDown	0	0	0	1	0	14

The confusion matrices show that the overall accuracy for individual *A* is quite lower (92.36%) than individual *B* (97.30%). However, the accuracies achieved show encouraging results even though these two individuals were not part of the training phase. The discrepancies in the results can be inferred that the gait and the intensity at which activities are performed are individual specific. Also, individuals *A* and *B* were of different physical characteristics than the participants who collected training data for offline analyses.

V. CONCLUSION

This work proposes the idea of online Human Activity Recognition system that focuses on using unobtrusive devices and services for the designated context aware features. The evaluation indicates that the J48 classifier using window size of 512 samples with 50% overlapping yields the highest accuracy (i.e., up to 96.02%). The system does not require server for feature extraction and processing thus, reducing the energy expenditures and making it more robust and responsive. The mobile HAR system is scalable, as performing the feature extraction and classification computations locally on the mobile device alleviates the server load.

REFERENCES

[1] F. J. Ordóñez, P. de Toledo, and A. Sanchis, "Activity Recognition Using Hybrid Generative/Discriminative Models on Home Environments Using Binary Sensors," *Sensors,* vol. 13, pp. 5460-5477, 2013.

[2] J. Vandewynckel, M. Otis, B. Bouchard, B.-A.-J. Ménélas, and A. Bouzouane, "Towards a real-time error detection within a smart home by using activity recognition with a shoe-mounted accelerometer," *Procedia Computer Science,* vol. 19, pp. 516-523, 2013.

[3] C. Belley, S. Gaboury, B. Bouchard, and A. Bouzouane, "An efficient and inexpensive method for activity recognition within a smart home based on load signatures of appliances," *Pervasive and Mobile Computing,* 2013.

[4] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive Computing*, ed: Springer, 2004, pp. 1-17.

[5] W. He, Y. Guo, C. Gao, and X. Li, "Recognition of human activities with wearable sensors," *EURASIP Journal on Advances in Signal Processing,* vol. 2012, pp. 1-13, 2012.

[6] C. V. Bouten, K. T. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen, "A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity," *Biomedical Engineering, IEEE Transactions on,* vol. 44, pp. 136-147, 1997.

[7] *Waikato environment for knowledge analysis (WEKA)*. Available: http://www.cs.waikato.ac.nz/ml/weka

[8] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *AAAI*, 2005, pp. 1541-1546.

[9] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explorations Newsletter,* vol. 12, pp. 74-82, 2011.

[10] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors,* vol. 10, pp. 1154-1175, 2010.

[11] I. Cleland, B. Kikhia, C. Nugent, A. Boytsov, J. Hallberg, K. Synnes*, et al.*, "Optimal Placement of Accelerometers for the Detection of Everyday Activities," *Sensors,* vol. 13, pp. 9183-9200, 2013.

[12] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural computation,* vol. 10, pp. 1895-1923, 1998.

[13] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King*, et al.*, "Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor," in *Wearable Computers, 2007 11th IEEE International Symposium on*, 2007, pp. 37-40.