INTELLIGENT AUTOMATION & SOFT COMPUTING, 2017
https://doi.org/10.1080/10798587.2017.1342400

Feature Selection for Activity Recognition from Smartphone Accelerometer Data

Juan C. Quiroz[a], Amit Banerjee[b], Sergiu M. Dascalu[c] and Sian Lun Lau[a]

[a]Department of Computing and Information Systems, Sunway University, Bandar Sunway, Malaysia; [b]School of Science, Engineering and Technology, Penn State Harrisburg, Middletown, USA; [c]Department of Computer Science and Engineering, University of Nevada, Reno, USA

ABSTRACT

We use the public Human Activity Recognition Using Smartphones (HARUS) data-set to investigate and identify the most informative features for determining the physical activity performed by a user based on smartphone accelerometer and gyroscope data. The HARUS data-set includes 561 time domain and frequency domain features extracted from sensor readings collected from a smartphone carried by 30 users while performing specific activities. We compare the performance of a decision tree, support vector machines, Naive Bayes, multilayer perceptron, and bagging. We report the various classification performances of these algorithms for subject independent cases. Our results show that bagging and the multilayer perceptron achieve the highest classification accuracies across all feature sets. In addition, the signal from gravity contains the most information for classification of activities in the HARUS data-set.

KEYWORDS

Activity recognition; accelerometer; angular velocity; smartphone; supervised learning

1. Introduction

Sensor data from smartphones, such as accelerometer and gyroscope data, has been used to determine the physical activity performed by a user (Fahim, Fatima, Lee, & Park, 2013; Reyes Ortiz, 2015; Reyes-Ortiz, Oneto, Samà, Parra, & Anguita, 2016; Shoaib, Bosch, Incel, Scholten, & Havinga, 2015; Weiss & Lockhart, 2012). Activity recognition is not limited to fitness tracking applications. Instead, activity recognition has applications in telemedicine services where monitoring of physical activity is part of treatment or therapy (Lau et al., 2010), remote patient monitoring for disabled and elderly patients, healthcare and assisted living technologies, and smart environments (Anguita, Ghio, Oneto, Parra, & Reyes-Ortiz, 2013a; Zhang & Sawchuk, 2011).

The Human Activity Recognition Using Smartphones (HARUS) data-set (Anguita, Ghio, Oneto, Parra, & Reyes-Ortiz, 2013b; Anguita et al., 2013a) is a public data-set built from the recordings of 30 users. Data was collected from 30 users, with a smartphone mounted on their waist, while they performed six physical activities: walking, walking upstairs, walking downstairs, sitting, standing, and lying down. The data recorded included the sensor readings from the accelerometer and gyroscope embedded in the smartphone. From the sensor readings, 561 features were extracted and used to build models for the classification of physical activities (Anguita et al., 2013b; Reyes-Ortiz et al., 2016).

The HARUS data-set is unique in (1) deriving the acceleration and gyroscope readings over time to generate jerk and angular velocity jerk, and (2) extracting a large set of features–561–from body acceleration, gravity (low frequency acceleration readings), angular velocity, and the jerk signals. Jerk is the rate of change of acceleration, as acceleration is the rate of change of velocity over time. The 561 features provide a large set of features for classifiers to differentiate between all the physical activities at high accuracies. However, the large feature set also makes the classifiers prone to overfitting, increases the calculations needed to extract the features, and increases the training and testing time of the classifiers. Most importantly, the extraction of 561 features makes it unfeasible to do online activity recognition. As the research in physical activity recognition progresses to identifying more complex activities and every day activities, finding the most informative features becomes even more important. In addition, while accurate results (nearly 100% accuracy) have been achieved in prior work (Fahim et al., 2013; Reyes Ortiz, 2015; Reyes-Ortiz et al., 2016, p; Shoaib et al., 2015; Weiss & Lockhart, 2012), these results have been obtained when the physical activities are simple, well separated, and performed under controlled conditions according to researchers' instructions (Lockhart & Weiss, 2014). Studies also suffer from data collection from a small group of participants (typically less than 10), obscure and poorly documented experimental and data processing techniques, and having the researchers being one of the participants for the data collection (Lockhart & Weiss, 2014). Our goal is to advance research in physical activity recognition by identifying the most informative features from the 561 features of the HARUS set.

In Reyes Ortiz (2015), the performance of the following groups of features was compared using a linear SVM model: gyroscope time domain features, gyroscope time and frequency domain features, acceleration time domain features, acceleration time and frequency domain features, and all 561 features. In this paper we use the HARUS data-set and build on the analysis of different feature sets from Reyes Ortiz (2015). Over the 561 features, and specifically over the time-domain features, we answer the following questions: (1) which feature sets yield the highest accuracies? (2) how do these sets compare with feature sets tested in prior work and (3) how much does

CONTACT Juan C. Quiroz juanq@sunway.edu.my

accuracy improve when using large feature sets? Our goal is to identify feature sets, which derived from the accelerometer and angular velocity readings of a smartphone, maximize the classification performance of activities. The six activities we classify are: walking, walking upstairs, walking downstairs, sitting, standing, and lying down (Anguita, Ghio, Oneto, Parra, & Reyes-Ortiz, 2012). Our results show that bagging and the multilayer perceptron achieved the highest classification accuracies across all our feature sets tested. In addition, the signal from gravity contained the most information for the accurate classification of activities in the HARUS data-set.

The rest of the paper is organized as follows: Section 2 describes prior work on activity recognition based on accelerometer readings. Section 3 presents our methodology for activity recognition. Section 4 discusses the results obtained. Finally, we conclude and present directions for future work in Section 5.

2. Related Work

In the past 15 years, human activity recognition using wearable sensors has been an active research area. A typical example is provided by accelerometer sensors. Their application includes healthcare and medicine (Jehn et al., 2009; Lau et al., 2010), daily activities (Bao & Intille, 2004; Bieber, Voskamp, & Urban, 2009) and sports (Mladenov & Mock, 2009). The earlier work investigated recognition techniques using a number of selected sensors, placed on different parts of the body. For example, Bao and Intille (2004) used five sensor boards on different parts of the body such as arm, wrist, knee, ankle and waist. Each sensor board consisted of a biaxial accelerometer, four AAA batteries, and a memory card for storage. Another investigation, carried out by Kern et al., used 12 body worn tri-axial accelerometers to perform activity recognition (Kern, Schiele, & Schmidt, 2007). Both investigations demonstrated how one can achieve recognition accuracy up to around 90%. These investigations proposed the use of multiple sensors placed at fixed strategic positions depending on the targeted activities (Kern et al., 2007).

Over the last 7 years, there was a shift towards using built-in sensors on smartphones instead of multiple dedicated sensor devices. Mladenov et al. presented a step counter application using the accelerometer of a Nokia N95 (Mladenov & Mock, 2009). The results showed that such smartphones can provide accurate step-counts, comparable to some of the commercial and dedicated step counter products. However, users needed to ensure that the phone was firmly attached to the body during data collection. The DiaTrace project (Bieber et al., 2009) used a mobile phone with accelerometers for physical activity monitoring. The prototype obtained accuracy of more than 95% for activities like resting, walking, running, cycling, and driving a car. Lau et al. used the Nokia N95 smartphones to record movements such as walking, going upstairs and downstairs, standing, and sitting, using only the accelerometer sensor (Lau et al., 2010). With a small number of features (only three to five features), they applied classification algorithms to build movement recognition models and achieved accuracy up to 94%. Even with low sampling rates (lowest was 8 Hz), it was possible to have recognition accuracy around 90% if meta-classifiers were applied. All these investigations have shown the potential of activity recognition using only a single smartphone placed at a specific position, such as the front pants pocket.

Most of the recent research now utilizes Android or iOS-based smartphones. In (Kwapisz, Weiss, & Moore, 2011), Kwapisz et al. had users carry an Android smartphone in the front pants leg pocket as they performed various activities, including walking, jogging, ascending stairs, descending stairs, sitting, and standing. They enlisted 29 subjects and used a 10 s window for the data analysis. They used Weka to compare decision trees (J48), logistic regression, and multilayer neural networks, using ten-fold cross-validation for their testing averaged over 10 runs. The smartphone included a tri-axial accelerometer, thus each reading included an x, y, and z value. In contrast to our work, they generated only a total of 43 features, as opposed to the 561 total features of the HARUS data-set. The features included average, standard deviation, average absolute difference, average resultant acceleration, time between peaks, and binned distribution.

Bayat et al. collected data from four users (Bayat, Pomplun, & Tran, 2014). Their experiments included collecting data with the smartphone in hand and with the smartphone in the pocket. They used Weka to train various classifiers, with the multilayer perceptron performing the best. They separated the raw accelerometer signal into body acceleration and gravity by using a low pass and a high pass filter.

Much of the previous work focused on feasibility and exploratory work that applied sensors on smartphones for human activity recognition, with the majority limited to accelerometers. In our work, we focus on two aspects. First, we investigate how feature selection can be optimized for activity recognition using features extracted from both the accelerometer and the gyroscope embedded in a smartphone. Second, we assess how the signals from accelerometer and gyroscope can be maximized by deriving jerk signals, and how the features from all these signal types can be combined to optimize accuracy of activity recognition.

3. Methodology

To conduct our experiments we used the Human Activity Recognition Using Smartphones (HARUS) Data-Set (Anguita et al., 2013b). Our motivation was to compare and analyze how various feature sets, in combination with various classifiers, affect the accuracy of activity recognition. In addition, we use an open and public data-set to enable comparison of feature extraction techniques and choice of classifier systems with prior work.

3.1. Data-Set

The HARUS data-set includes the recordings of 30 subjects. The data-set is divided into a training set, with the data from 21 subjects, and a testing set, with the data from the remaining 9 subjects. The subjects were between the ages of 19–48, though the data does not identify the age of each subject. The subjects performed six activities while wearing a smartphone mounted on their waist: walking, walking upstairs, walking downstairs, sitting, standing, and lying down. As opposed to prior work where several dedicated sensors are placed at strategic locations on the users body (Bao & Intille, 2004; Preece, Goulermas, Kenney, & Howard, 2009), the HARUS data-set is generated from a triaxial accelerometer and a triaxial gyroscope embedded on a smartphone, the Samsung Galaxy S II. By relying on data from a smartphone, the data collection is more natural and unobtrusive.

Figure 1 shows the process of gathering raw sensor readings from the accelerometer and gyroscope embedded in the

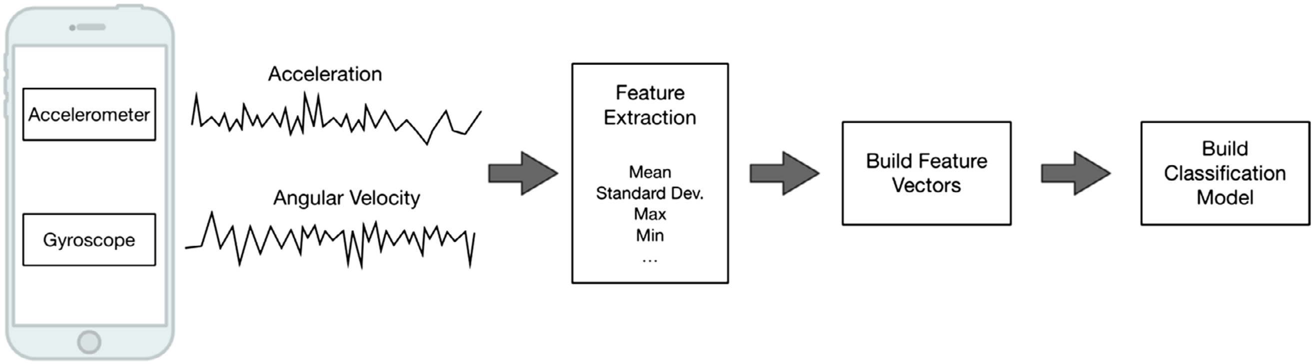

Figure 1. The Process of Sampling Acceleration Sensor Readings and Angular Velocity Sensor Readings from the Accelerometer and the Gyroscope Embedded in a Smartphone, Extracting Features, and using the Features to Train a Model for Activity Recognition.

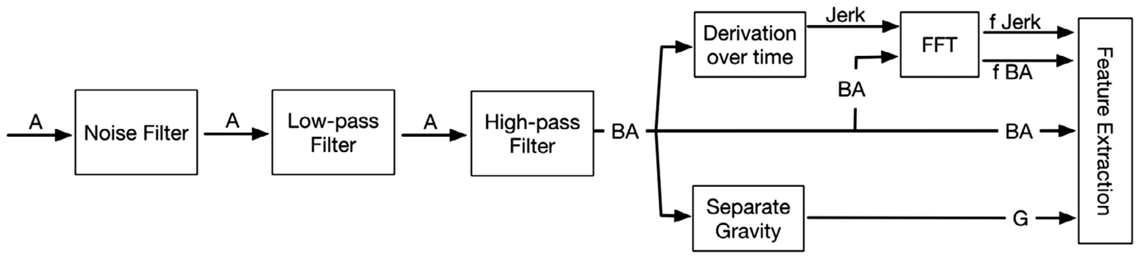

Figure 2. Processing of Raw Accelerometer Sensor Readings from Embedded Accelerometer of Smartphone to Generate Body Acceleration Data (BA), Gravity Data (G), Body Acceleration Jerk Data (Jerk), and Body Acceleration Frequency Data (f Jerk and f BA).

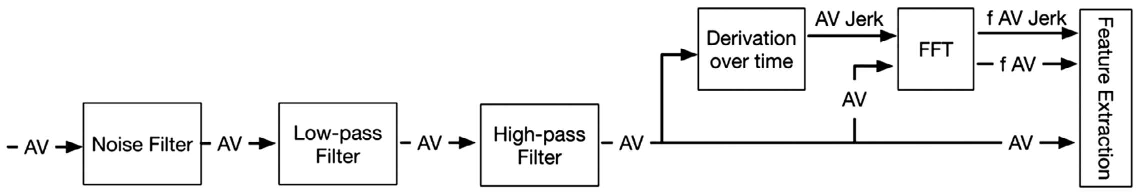

Figure 3. Processing of Sensor Readings from Embedded Gyroscope of Smartphone to Generate Angular Velocity Data (AV), Angular Velocity Jerk Data (AV Jerk), and Angular Velocity Frequency Data (f AV and f AV Jerk).

smartphone, and using it to build a model, which can then recognize activities. The accelerometer and gyroscope generate time series data, where at each point in time there is an accelerometer reading and an angular velocity reading, each having an x, y, and z component. The HARUS data-set was gathered from data sampled at 50 Hz.

The HARUS data-set maximizes the acceleration and angular velocity data from the smartphone by deriving additional features. Figures 2 and 3 show how additional features are derived from the raw acceleration and angular velocity readings. The sensor signals, for the accelerometer and gyroscope, were processed with noise filters: a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz (Anguita et al., 2012). Body acceleration (BA) was obtained by passing the acceleration signal through a high-pass filter with a corner frequency of 0.3 Hz. The gravity signal (G) was obtained by subtracting body acceleration (BA) from the acceleration signal (A). Deriving the body acceleration signals over time resulted in acceleration jerk signals. A similar process, detailed in Figure 3, was used to derive various signals from the raw angular velocity (AV) readings. Time-domain features were extracted from body acceleration, gravity, the acceleration jerk, angular velocity, and angular velocity jerk. This process of signal processing and feature extraction was documented in (Reyes Ortiz, 2015; Reyes-Ortiz et al., 2016).

The HARUS data-set includes features extracted from frequency domain signals, generated by applying the Fast Fourier Transform (FFT) to the body acceleration signals and the angular velocity readings. Out of the 561 features from the HARUS data-set, 289 features are frequency-domain features. In this paper, we limit our experiments and analysis to the time-domain features, and leave the analysis of frequency domain-features and their combination with time-domain features for future work.

In the data-set, the acceleration, angular velocity, and jerk signals are segmented into 2.56 s sliding windows with 50% overlap. The features extracted from each of these windows formed a feature vector of size 561, with each feature normalized and bounded within [−1, 1]. The training set includes 7352 feature vectors, generated from the data from 21 subjects (out of the total 30 subjects). We randomly split the training set into 70/30 for training and testing of the classifier systems. To test for statistical significance, we repeat the random 70/30 splitting 10 times. The original test data from the 9 subjects is left as validation data. This has been shown to be a reliable method of training, testing, and building models (Bishop, 1996).

3.2. Classifier Systems

We used WEKA, a machine learning workbench written in Java, for our experiments (Hall et al., 2009). We compare the following base-level classifiers from the WEKA toolkit: decision trees (J48), support vector machines (libSVM), radial basis function (RBF) network (RBFNetwork), Naive Bayes classifier (NaiveBayes), and multilayer perceptron (MultilayerPerceptron). In addition, we test the adaptive boosting (AdaBoostM1) and bagging (Bagging) meta-classifiers. We use these classifiers with their default settings in Weka.

3.3. Feature Selection

Figures 2 and 3 identify the data sets that are derived from the raw acceleration and angular velocity sensor readings. From the acceleration sensor readings and the angular velocity sensor readings, the following sensor signals are generated: body acceleration (BA), gravity (G), body acceleration jerk (AJ), acceleration frequency data, angular velocity (AV), angular velocity jerk (AVJ), and angular velocity frequency data. We discuss the use of the acceleration and angular velocity frequency signals as future work in Section 5.

Our goal is to analyze how features extracted from the various acceleration and angular velocity signals affect the accuracy of activity recognition, and which features may be combined to maximize classification performance. The data-set is structured such that there is a total of 40 time-domain features calculated for each of the sensor signals: mean (3), standard deviation (3), median absolute deviation (3), max (3), min (3), signal magnitude area (1), energy (3), interquartile range (3), entropy (3), auto regression coefficients with Burg order equal to 4 (12), and correlation between axes (3). All of these features are computed for each of the x, y, and z components of acceleration and angular velocity, with signal magnitude area and the auto regression coefficients with Burg order equal to 4, four features per axis, being the exception.

For our experiments, first we selected all of the time-domain features for body acceleration (feature vector of size 40), and compared the classification accuracy against angular velocity, gravity, and using the entire set of time-domain features (feature vector of size 256). Next, we compared body acceleration to body acceleration jerk, and angular velocity to angular velocity jerk, to determine whether the jerk signals resulted in higher accuracies. Table 1 lists the sets of time-domain features extracted from the available signal types.

When using all of the time-domain features for one signal, such as body acceleration, make up a feature vector with 40 dimensions. However, evaluating every feature reduces the speed of training and execution of the model. High dimensional data can also make models prone to overfitting and reduce their predictive power due to the curse of dimensionality (Donoho, 2000). Thus, we selected subsets of features from the feature vector of size 40, in order to identify high performing feature subsets for activity recognition on the HARUS data-set. Table 2 shows ten different sets of time-domain features selected from the 40 time-domain features listed above. The selection of these ten sets was driven by our experimentation and by the use of particular time-domain features in prior work (Bao & Intille, 2004; Khan, Siddiqi, & Lee, 2013; Lara & Labrador, 2013; Shoaib et al., 2015).

For the first set, we use the following 13 time-domain features: mean (for each axis), standard deviation (for each axis),

Table 1. Original Sets of Time-domain Features in the HARUS data-set, the Features Extracted from Body Acceleration, Angular Velocity, Gravity, Body Acceleration Jerk, and Angular Velocity Jerk Signals. These Sets all have a Total of 40 Features.

Set	No Features	Signal Type	Features Extracted
Set BA	40	BA	mean (3), standard deviation (3), median absolute deviation (3), max (3), min (3), signal magnitude area (1), energy (3), interquartile range (3), entropy (3), auto regression coefficients with Burg order equal to 4 (12), correlation between axes (3)
Set AJ	40	BA jerk	
Set AV	40	AV	
Set AVJ	40	AV Jerk	
Set G	40	G	
Set All	265	BA, AV, G, BA jerk, AV jerk, magnitude	

Table 2. Smaller Sets of Time-domain Features Generated by Selecting a Subset of Features from the Original 40 Time-domain Features Listed in Table 1 for Various Signal Types.

Set	No Features	Signal Type	Features Extracted
Set 1	13	BA	mean (3), standard deviation (3), median absolute deviation (3), signal magnitude area (1), correlation (3)
Set 2	22	BA	mean (3), standard deviation (3), median absolute deviation (3), signal magnitude area (1), correlation (3), energy (3), interquartile range (3), entropy (3)
Set 3	13	AV	mean (3), standard deviation (3), median absolute deviation (3), signal magnitude area (1), correlation (3)
Set 4	22	AV	mean (3), standard deviation (3), median absolute deviation (3), signal magnitude area (1), correlation (3), energy (3), interquartile range (3), entropy (3)
Set 5	26	BA and AV	Set 1 + Set 3
Set 6	44	BA and AV	Set 2 + Set 4
Set 7	13	G	mean (3), standard deviation (3), median absolute deviation (3), signal magnitude area (1), correlation (3)
Set 8	22	G	mean (3), standard deviation (3), median absolute deviation (3), signal magnitude area (1), correlation (3), energy (3), interquartile range (3), entropy (3)
Set 9	6	BA	mean (3), standard deviation (3)
Set 10	6	BA	median absolute deviation (3), correlation (3)

median absolute deviation (for each axis), the correlation between the X and Y axis, the correlation between the X and Z axis, and the correlation between the Y and Z axis. For the second set, we add nine 9 features, for a total of 22 features. Third, can we achieve high accuracy activity recognition with only angular velocity data? That is, in the absence of acceleration data, do the features extracted from angular velocity provide accurate results for activity recognition? The third and fourth sets are used to test this hypothesis.

The fifth and sixth sets test combinations of sets 1–4. Sets 7 and 8 test the accuracy achieved by using only gravity data. Set 9 consists of the mean and standard deviation extracted from accelerometer readings, with features widely used for activity recognition in prior work (Bao & Intille, 2004; Preece et al., 2009). Set 10 includes the median standard deviation and the correlation between axes. For this set, we found empirically that the median absolute deviation and correlation between axes to be features that provided high classification performance. Correlation between axes has been used in prior work for activity recognition (Bao & Intille, 2004; Fahim et al., 2013; Ravi, Dandekar, Mysore, & Littman, 2005; Zheng, Wong, Guan, & Trost, 2013). The median absolute deviation has not been used in prior work.

4. Results and Discussion

For all of our results, we performed the Friedman test to test the differences between the classifiers and the Nemenyi post hoc test to determine whether the performance of two classifiers is significantly different ($p < 0.05$) (Demšar, 2006). The Friedman test is a non-parametric equivalent to the repeated measures ANOVA.

4.1. Acceleration vs Angular Velocity vs Gravity

Table 3 compares the classification performance using a subset of time-domain features for acceleration, acceleration jerk, gravity, angular velocity, angular velocity jerk, and all time-domain features. When classifying activities using a feature vector of size 40 from accelerometer readings, the multilayer perceptron resulted in the highest accuracy (83.9%). The multilayer perceptron has an input layer with size equal to the size of the input vector, in this case being size 40. The single hidden layer includes I + 1 nodes where I is the size of the input layer. Finally, the output layer has 6 nodes, with each node representing one of the six physical activities.

For body acceleration, the perceptron accuracy of 83.9% was statistically significant over the performance of the other classifiers. For angular velocity, the multilayer perceptron performed best (82.2%), with this difference being statistically significant over the performance of the other classifiers, with the exception of bagging. For gravity, bagging performed best with an accuracy of 91.5%, with this performance being statistically better than all classifiers except for the perceptron. The last column shows the classification performance using all statistical features, consisting of a feature vector of size 265. The multilayer perceptron achieved the highest classification accuracy of 97.2%, which was statistically significant over all the classifiers.

The gravity features resulted in the highest classification accuracies. This was surprising, as we expected acceleration and angular velocity readings to provide more useful information for classification. We believe that this may be due to gravity providing useful information for discerning activities that would typically require less movement and energy, such as sitting, standing, and lying down.

Table 3 shows that angular velocity is not as effective as body acceleration for the classification of physical activities. The performance of all the classifiers drops slightly when using angular velocity features instead of body acceleration features, with the largest drop being for the SVM. The performance of the RBF network and Naive Bayes actually increases from 64.6% to 67.3% and from 53.9% to 58.2% respectively. For the multilayer perceptron, the drop in performance is 1.6%, yet the accuracy remains over 80%. Thus, in the absence of accelerometer readings, or in the absence of reliable accelerometer readings, angular velocity readings provide suitable data for classification of physical activities, with the multilayer perceptron being a good choice.

Table 4 shows the confusion matrix for one run of a decision tree (J48) on the body acceleration set BA(40). The overall accuracy was 75%, which was consistent with the result on Table 3. The highest misclassification is in rows 4–6, which corresponds to the static activities of sitting, standing, and lying down. Table 5 shows the confusion matrix for one run of a decision tree on the gravity set G(40). In this case the overall accuracy was 90.3%. The difference is that features extracted from the gravity signal result in the static activities (rows 4–6) being classified more accurately. However, when using the gravity signal, the prediction accuracy decreases slightly for the dynamic activities of walking, walking upstairs, and walking downstairs, as evident from rows 1–3 of Table 5.

4.2. Jerk Signals

Table 6 compares the accuracy results between acceleration, angular velocity, and jerk. Results show that classification performance drops when using features from jerk as opposed to acceleration. For jerk, the multilayer perceptron outperforms all the other classifiers ($p < 0.05$). For angular velocity jerk, the performance of the multiplayer perceptron and bagging is statistically significant better over the other classifiers, though

Table 3. Activity Recognition Accuracy of Classifiers using Time-domain Features Extracted from Acceleration, Angular Velocity, Gravity, and all 265 Time-domain Features from the HARUS Data-set.

Classifier	BA (40)	AV (40)	G (40)	All (265)
J48	75.6	73.5	89.7	94.5
AdaBoost	35.7	35.7	37.8	35.8
Bagging	81.0	79.5	**91.5**	95.9
SVM	79.2	72.9	84.2	94.8
RBF	64.6	67.3	79.6	88.7
NaiveBayes	53.9	58.2	75.0	85.4
Perceptron	**83.9**	**82.2**	89.8	**97.2**

Note: Statistically significant results in bold.

Table 4. Confusion Matrix for Decision Tree (J48) with Feature Set A(40), with an Accuracy of 75.0% (1655 Correctly Classified Instances and 551 Incorrectly Classified Instances). Activities: (1) Walking , (2) Walking Upstairs (3) Walking Downstairs (4) Sitting (5) Standing (6) Lying Down.

		Prediction					
		1	2	3	4	5	6
Truth	1	330	34	4	0	0	0
	2	34	262	37	0	0	0
	3	12	29	269	0	0	0
	4	0	0	0	202	112	56
	5	0	0	0	119	267	38
	6	0	0	0	56	20	325

Table 5. Confusion Matrix for RF with Feature Set G(40), with an Accuracy of 90.3% (1992 Correctly Classified Instances and 214 Incorrectly Classified Instances). Activities: (1) Walking, (2) Walking Upstairs (3) Walking Downstairs (4) Sitting (5) Standing (6) Lying Down.

		Prediction					
		1	2	3	4	5	6
Truth	1	304	12	40	2	10	0
	2	13	300	17	0	3	0
	3	42	25	240	2	1	0
	4	1	0	0	350	19	0
	5	5	8	5	9	397	0
	6	0	0	0	0	0	401

Table 6. Activity Recognition Accuracy of Classifiers using Time-domain Features Extracted from Acceleration, Acceleration Jerk, Angular Velocity, and Angular Velocity Jerk.

Classifier	BA (40)	AJ (40)	AV (40)	AVJ (40)
J48	75.6	64.4	73.5	65.8
AdaBoost	35.7	35.8	35.7	35.8
Bagging	81.0	71.2	79.5	**74.4**
SVM	79.2	73.2	72.9	68.2
RBF	64.6	64.4	67.3	55.7
NaiveBayes	53.9	48.9	58.2	43.6
Perceptron	**83.9**	**76.2**	**82.2**	**76.5**

Note: Statistically significant results in bold.

there was no statistically significance between the perceptron and bagging.

Overall the results show that the statistical features extracted from acceleration readings yield higher classification performances compared to classification with features extracted from acceleration jerk readings. Similarly, features from angular velocity readings are more useful than features from angular velocity jerk readings.

4.3. Performance of Smaller Feature Sets

We experimented with feature combinations to find effective, yet smaller feature sets. Table 7 shows sets of features by taking the 40 time-domain features listed in Table 1, and selecting only a subset of those features for a particular signal type. The feature sets are listed in Table 2. Set 1 includes 13 time-domain features extracted from body acceleration. Set 2 includes the same features from set 1, plus nine additional body acceleration features. Sets 3 and 4 are the same as set 1 and 2, except that they are extracted from angular velocity. Set 5 combines sets 1 and 3, while set 6 combines sets 2 and 4.

Over all the sets, bagging yielded the highest accuracy, with the multilayer perceptron performing second-best. When using all 40 time-domain features for body acceleration, BA(40), the multilayer perceptron achieved the highest accuracy of 83.9%, whereas using only 13 time-domain features, BA(13), with bagging yields 80% accuracy. Thus, classifiers can be trained on smaller feature spaces, leading to faster training and testing times for classifiers, while achieving accuracies comparable to larger feature sets. Set 5, which combines 13 time-domain features of body acceleration and 13 time-domain features of angular velocity, performed better than each of the larger sets BA(40) and AV(40).

The median standard deviation and correlation between axes (set 10) outperforms the accuracy of the feature set including the mean and standard deviation (set 9). For set 10, bagging achieved the highest accuracy of 78.9%, statistically significant over all classifiers except for the decision tree with accuracy of 75.6%. The accuracy achieved compares favorably with the accuracy of sets 1 and 2, with the difference being less than 2%, even though sets 9 and 10 use half the features of sets 1 and 2. The use of smaller feature sets results in faster training times, faster modeling times, and simpler models. When considering online training and classification, the use of small feature sets is key to an effective system implementation.

Table 8 shows time-domain feature subsets extracted from gravity. For most of the classifiers, using 22 features instead of 13 results in a modest increase in classification accuracy. For the decision tree and bagging, the increase is less than 1%. The biggest increase is in the SVM accuracy, increasing from 67.9% to 72.3%. For SVM and the multilayer perceptron, the increase is more modest; the highest accuracy achieved 87.8%

Table 7. Activity Recognition Accuracy of Classifiers using the Feature Sets from Table 2.

	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 9	Set 10		
Classifier	BA(13)	BA(22)	AV(13)	AV(22)	BA+AV(26)	BA+AV(44)	BA(6)	BA(6)	BA(40)	AV(40)
J48	75.9	76.6	73.3	73.9	82.7	82.5	65.8	75.6	75.6	73.5
AdaBoost	35.5	35.5	35.7	35.7	35.6	35.7	35.5	35.4	35.7	35.7
Bagging	**80.0**	**80.8**	**78.1**	**78.5**	**87.1**	**87.5**	**69.7**	**78.9**	81.0	79.5
SVM	69.5	73.7	60.3	66.5	74.7	80.0	50.1	69.5	79.2	72.9
RBF	62.0	63.2	60.1	61.3	64.2	65.5	55.3	67.0	64.6	67.3
NaiveBayes	53.3	52.4	53.2	51.8	57.9	56.6	49.3	61.2	53.9	58.2
Perceptron	76.1	79.3	70.2	73.2	82.3	85.5	60.9	73.5	**83.9**	**82.2**

Note: Statistically significant results in bold.

Table 8. Classification Accuracy of Activities using Features Extracted from Gravity Signals.

	Set 7	Set 8	
Classifier	G (13)	G (22)	G (40)
J48	85.0	85.2	89.7
AdaBoost	37.8	37.8	37.8
Bagging	**87.8**	**87.9**	**91.5**
SVM	67.9	72.3	84.2
RBF	72.3	70.9	79.6
NaiveBayes	59.9	59.1	75.0
Perceptron	78.9	80.9	89.8

Note: Statistically significant results in bold.

with 13 features. By increasing the number of features to 40, the accuracy of bagging increases about 4% to 91.5%. Thus, the set of 13 time-domain features extracted from the gravity readings yields high accuracies, without having to explore or build models on higher dimensional feature spaces. Out of sets 1–10, sets 7 and 8 perform the best.

In this paper we do not test and explore the performance of frequency domain features. We expect that the adding frequency-domain features to our feature sets would show the accuracy of all the classifiers consistent with the results in (Reyes Ortiz, 2015). Using the frequency-domain features would almost double the total number of features available for classification. In practice, the use of the frequency-domain features for online activity recognition would be more costly due to the Fast Fourier Transform (FFT) needing to be calculated to transform the signals to the frequency domain. Our current approach is computationally simpler by not using the frequency domain signals.

5. Conclusions and Future Work

We presented a comparison of the accuracy of classifiers for activity recognition. We used the HARUS data-set to compare the accuracy achieved with features extracted from body acceleration, angular velocity, gravity, body acceleration jerk, and angular velocity jerk. We found that for this particular data-set, the gravity signals provide high classification accuracy, especially for static activities of sitting, standing, and lying down. This is important since gravity signals are usually filtered and discarded in favor of linear acceleration. Jerk signals do not provide as high accuracy as simply using acceleration and angular velocity. Thus, jerk signals alone do not merit the additional derivation step. Lastly, we found that features from angular velocity are not as helpful for classification as body acceleration, yet a few features extracted from angular velocity can considerably improve the accuracy of body acceleration feature sets. Our contributions not only benefit traditional physical activity recognition, but also applications that rely on sensor data extracted from smartphones for the purpose of remote patient monitoring and smart environments.

For future work we plan to analyze how the frequency-domain features from various signals affect activity recognition accuracy. A limitation from our work is that we did not use algorithmic feature selection. The HARUS data-set, with its large number of features, provides an ideal test case for testing various feature selection algorithms to identify small feature sets that yield high classification accuracies in this data-set. We plan to use the feature importance from random forests for feature selection. Other algorithmic methods to be tested include subset correlation-based feature selection, principal component analysis, and genetic algorithms. Finally, having identified the features that are the most informative, these features need to be tested on additional data sets to assess how well they can generalize?

Disclosure Statement

No potential conflict of interest was reported by the authors.

Notes on Contributors

Juan Quiroz is a senior lecturer in the Department of Computing and Information Systems at Sunway University, Malaysia. He received a Ph.D. in Computer Science and Engineering from the University of Nevada, Reno, USA (2010). His research interests are in machine learning, pattern recognition, and evolutionary computation.

Amit Banerjee, Ph.D. is an associate professor in the Mechanical Engineering Department at Pennsylvania State University Harrisburg. His research interests include robotics and automation, computational intelligence in manufacturing, evolutionary computation, machine learning and design creativity. He teaches courses in robotics, manufacturing, design and mechanics. He has an M.Des in Product Design from the Indian Institute of Science, Bangalore and a Ph.D. in Mechanical

Engineering from the New Jersey Institute of Technology. He has also been a postdoctoral research fellow at the Evolutionary Computational Systems Lab in the Department of Computer Science and Engineering at the University of Nevada Reno. His work has been published in ASME Journal of Risk and Uncertainty in Engineering Systems, Chaos and Solitons, and Swarm and Evolutionary Computation, among others.

Sergiu Dascalu is a professor in the Computer Science and Engineering Department at the University of Nevada, Reno (UNR). He received a Master's degree in Automated Control from the Polytechnic University of Bucharest, Romania and a Ph.D. in Computer Science from Dalhousie University, Canada. His research interests are in software engineering, human-computer interaction, and data science, in particular in software for scientific projects, interaction design, and simulation environments. Sergiu is the Director of the Software Engineering Laboratory at UNR and the leader of the Nevada cyberinfrastructure team in a 5-year $24-million NSF-funded project. He has served as investigator or co-investigator on several other projects funded by NSF, NASA, and ONR. Since joining UNR in 2002 he was the main advisor of over 50 graduate students.

Sian Lun Lau received his MSc and Dr.-Ing. at the University of Kassel, Germany, where he worked as a full-time researcher since year 2004. He has contributed and managed various German national as well as EU funded projects, such as EU FP6 MobiLife, ITEA S4ALL, BMBF MATRIX and EU FP7 SEAM4US. Since February 2013, he has returned to his home country and took up the position of the head of Department at the Department of Computing and Information Systems, Sunway University. He is currently an associate professor at the department, and also holds the position of the Associate Dean for the School of Science and Technology. He has published over 50 publications in conferences, workshops, book chapters, and journals. He is currently a senior member of the Institute of Electrical and Electronics Engineers (IEEE). His research interests include context-awareness, crowd-sourced mobile systems, sustainable smart city and applied machine learning.

References

Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J.L. (2012). Human Activity Recognition on Smartphones using a Multiclass Hardware-friendly Support Vector Machine. In David Hutchison, Takeo Kanade, & Josef Kittler (Eds.), *Ambient Assisted Living and Home Ccare, Lecture Notes in Computer Science* (pp. 216–223). Berlin Heidelberg: Springer.

Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J.L. (2013a). Energy Efficient Smartphone-based Activity Recognition using Fixed-point Arithmetic. *J. Univers. Comput. Sci., 19*, 1295–1314.

Anguita, D., Ghio, A., Oneto, L., Parra, X., & Reyes-Ortiz, J.L. (2013b). A public domain dataset for human activity recognition using smartphones. In: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN.

Bao, L., & Intille, S.S. (2004). Activity Recognition from User-annotated Acceleration Data. In A. Ferscha & F. Mattern (Eds.), *Pervasive Computing, Lecture Notes in Computer Science* (pp. 1–17). Berlin Heidelberg: Springer.

Bayat, A., Pomplun, M., & Tran, D.A. (2014). A Study on Human Activity Recognition using Accelerometer Data from Smartphones. Procedia Comput. Sci., The 9th International Conference on Future Networks and Communications (FNC'14)/The 11th International Conference on Mobile Systems and Pervasive Computing (MobiSPC'14)/Affiliated Workshops 34, 450–457. DOI:10.1016/j.procs.2014.07.009

Bieber, G., Voskamp, J., & Urban, B. (2009). Activity Recognition for Everyday Life on Mobile Phones. In: Proceedings of the 5th International on Conference Universal Access in Human-Computer Interaction. Part II: Intelligent and Ubiquitous Interaction Environments, UAHCI '09. Springer-Verlag, Berlin, Heidelberg, pp. 289–296. DOI:10.1007/978-3-642-02710-9_32

Bishop, C.M. (1996). *Neural Networks for Pattern Recognition* (1 ed.). Oxford: New York: Clarendon Press.

Demšar, J. (2006). Statistical Comparisons of Classifiers over Multiple Data-Sets. *J. Mach. Learn. Res., 7*, 1–30.

Donoho, D.L. (2000). High-dimensional data analysis: The curses and blessings of dimensionality. In: Ams Conference on Math Challenges of the 21st Century.

Fahim, M., Fatima, I., Lee, S., & Park, Y.-T. (2013). EFM: Evolutionary fuzzy model for dynamic activities recognition using a smartphone accelerometer. *Applied Intelligence, 39*, 475–488. DOI:10.1007/s10489-013-0427-7

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I.H. (2009). The WEKA Data Mining Software: An update. *ACM SIGKDD Explorations Newsletter, 11*, 10–18. DOI:10.1145/1656274.1656278

Jehn, M., Schmidt-Trucksäess, A., Schuster, T., Hanssen, H., Weis, M., Halle, M., & Koehler, F. (2009). Accelerometer-based Quantification of 6-minute Walk Test Performance in Patients with Chronic Heart Failure: Applicability in Telemedicine. *Journal of Cardiac Failure, 15*, 334–340. DOI:10.1016/j.cardfail.2008.11.011

Kern, N., Schiele, B., & Schmidt, A. (2007). Recognizing Context for Annotating a Live Life Recording. *Personal and Ubiquitous Computing, 11*, 251–263. DOI:10.1007/s00779-006-0086-3

Khan, A.M., Siddiqi, M.H., & Lee, S.-W. (2013). Exploratory Data Analysis of Acceleration Signals to Select Light-weight and Accurate Features for Real-time Activity Recognition on Smartphones. *Sensors, 13*, 13099–13122. DOI:10.3390/s131013099

Kwapisz, J.R., Weiss, G.M., & Moore, S.A. (2011). Activity Recognition using Cell Phone Accelerometers. *ACM SIGKDD Explorations Newsletter, 12*, 74–82. DOI:10.1145/1964897.1964918

Lara, O.D., & Labrador, M.A. (2013). A Survey on Human Activity Recognition using Wearable Sensors. *IEEE Communications Surveys & Tutorials, 15*, 1192–1209. DOI:10.1109/SURV.2012.110112.00192

Lau, S.L., König, I., David, K., Parandian, B., Carius-Düssel, C., & Schultz, M. (2010). Supporting patient monitoring using activity recognition with a smartphone. In: 2010 7th International Symposium on Wireless Communication Systems (ISWCS). Presented at the 2010 7th International Symposium on Wireless Communication Systems (ISWCS), pp. 810–814. DOI:10.1109/ISWCS.2010.5624490

Lockhart, J.W., & Weiss, G.M. (2014). Limitations with Activity Recognition Methodology & #38; Data-Sets. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, UbiComp '14 Adjunct. ACM, New York, NY, USA, pp. 747–756. DOI:10.1145/2638728.2641306

Mladenov, M., & Mock, M. (2009). A Step Counter Service for Java-enabled Devices using a Built-in Accelerometer. In: Proceedings of the 1st International Workshop on Context-Aware Middleware and Services: Affiliated with the 4th International Conference on Communication System Software and Middleware (COMSWARE 2009), CAMS '09. ACM, New York, NY, USA, pp. 1–5. DOI:10.1145/1554233.1554235

Preece, S.J., Goulermas, J.Y., Kenney, L.P.J., & Howard, D. (2009). A Comparison of Feature Extraction Methods for the Classification of Dynamic Activities from Accelerometer Data. *IEEE Transactions on Biomedical Engineering, 56*, 871–879. DOI:10.1109/TBME.2008.2006190

Ravi, N., Dandekar, N., Mysore, P., & Littman, M.L. (2005). Activity Recognition from Accelerometer Data. In: Proceedings of the 17th Conference on Innovative Applications of Artificial Intelligence - Volume 3, IAAI'05. AAAI Press, Pittsburgh, Pennsylvania, pp. 1541–1546.

Reyes Ortiz, J.L. (2015). *Smartphone-based Human Activity Recognition* (Springer Theses). Springer International Publishing, Cham.

Reyes-Ortiz, J.-L., Oneto, L., Samà, A., Parra, X., & Anguita, D. (2016). Transition-aware Human Activity Recognition using Smartphones. *Neurocomputing, 171*, 754–767. DOI:10.1016/j.neucom.2015.07.085

Shoaib, M., Bosch, S., Incel, O.D., Scholten, H., & Havinga, P.J.M. (2015). A Survey of Online Activity Recognition using Mobile Phones. *Sensors, 15*, 2059–2085. DOI:10.3390/s150102059

Weiss, G.M., & Lockhart, J. (2012). The Impact of Personalization on Smartphone-based Activity Recognition. In: Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence.

Zhang, M., & Sawchuk, A.A. (2011). A Feature Selection-based Framework for Human Activity Recognition using Wearable Multimodal Sensors. In: Proceedings of the 6th International Conference on Body Area Networks, BodyNets '11. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, pp. 92–98).

Zheng, Y., Wong, W.-K., Guan, X., & Trost, S. (2013). Physical Activity Recognition from Accelerometer Data using a Multi-scale Ensemble Method. In: Twenty-Fifth IAAI Conference. Presented at the Twenty-Fifth IAAI Conference.