Limitations with Activity Recognition Methodology & Data Sets

Jeffrey W. Lockhart
Fordham University
441 E. Fordham Rd.
Bronx, NY, 10458 USA
lockhart@cis.fordham.edu

Gary M. Weiss
Fordham University
441 E. Fordham Rd.
Bronx, NY, 10458 USA
gweiss@cis.fordham.edu

UbiComp '14, September 13 - 17 2014, Seattle, WA, USA

ACM 978-1-4503-3047-3/14/09…$15.00.
http://dx.doi.org/10.1145/2638728.2641306

Abstract

Human activity recognition (AR) has begun to mature as a field, but for AR research to thrive, large, diverse, high quality, AR data sets must be publically available and AR methodology must be clearly documented and standardized. In the process of comparing our AR research to other efforts, however, we found that most AR data sets are sufficiently limited as to impact the reliability of existing research results, and that many AR research papers do not clearly document their experimental methodology and often make unrealistic assumptions. In this paper we outline problems and limitations with AR data sets and describe the methodology problems we noticed, in the hope that this will lead to the creation of improved and better documented data sets and improved AR experimental methodology. Although we cover a broad array of methodological issues, our primary focus is on an often overlooked factor, *model type*, which determines how AR training and test data are partitioned—and how AR models are evaluated. Our prior research indicates that personal, hybrid, and impersonal/universal models yield dramatically different performance [30], yet many research studies do not highlight or even identify this factor. We make concrete recommendations to address these issues and also describe our own publically available AR data sets.

Model Type	Personal Hybrid Impersonal
Collect Method	Fully natural Semi-natural Laboratory
Data	
Subjects	1, 2, 3 …
Population	Students, elderly,..
Traits	Height, weight, …
Activities	Run, Jog, …
Duration	1 hour, …
Sensors	
Type	GPS, accel, gyro
Sample rate	20Hz, 50Hz, …
Number	1, 2, 3, …
Location	Pocket, belt, …
Orientation	Up, down, changing
Features	
Raw	m/s^2, …
Xformed	statistical, FFT
Window	10 sec.
Results	
Accuracy	94%, 98%, ..
Consistency	Mean, variance

Table 1. AR data set characteristics.

Author Keywords
Activity Recognition; Data Mining; Methodology

ACM Classification Keywords
J.3. Computer Applications: Life and Medical Sciences-Health.

Introduction
Activity recognition research requires high quality and diverse activity data. For research in this area to thrive, such data should be publically available. The activity data should match the intended range of activities and settings that is being studied. But even more fundamentally, AR data sets should *clearly describe the characteristics* of the data so that researchers can determine if the data meets their needs—and how to qualify any conclusions drawn from that data. Thus, at a minimum, AR data sets should specify the values for the various data set characteristics listed in Table 1. However, they should go beyond that and ensure that the values of these characteristics are as realistic as possible, to increase the generality of inferences drawn from the data. As our analysis of data sets in related AR research shows, most research studies are based on data that have many limitations (e.g., few subjects, collected under laboratory settings, etc.) and sometimes these limitations are not even discussed. Such limitations are sufficiently important that they should ideally be described within the research papers, not just in metadata associated with the data sets.

A key concern of this paper is not just the underlying raw AR sensor data, but the methodology used to build and evaluate AR systems. In this paper we focus on the fact that many existing research studies do not explicitly discuss the type of model they are building (personal, hybrid, or impersonal), even though our research results show that this has a huge impact on AR performance [30]. Model type is related to AR data in that it determines how the data is collected, organized, and partitioned into training and test data.

The observations and conclusions that we present in this paper are based on the Master's thesis [31] of one of the authors, which analyzed 34 published AR research papers. While this sample is not exhaustive, it is reasonably large and focuses on AR research similar to our smartphone-based AR research [25, 30]. A few of the 34 papers cover multiple datasets, and hence 38 data sets are analyzed; furthermore, since several of them generate multiple types of models, a total of 47 distinct models are analyzed. A table that provides much of the information listed in Table 1 for each of the individual 38 data sets is available [31]. Information from that table is provided throughout this paper.

Background on Model Type
All activity recognition research must decide on which type of AR model(s) to analyze. The model types are determined by the way in which the generated model is *used* and dictate how the training and test data are partitioned. Although there may be many variations, we have identified three basic model types [30]:

- <u>Impersonal models</u>, or universal models, use training data from a panel of users who will not subsequently use the model (i.e., won't be present in the test set). These models can be built once and used on new users without requiring labeled training data from those users.
- <u>Personal models</u> use training data only from the user who will utilize the model, so the training and test data come from the same person; all users must provide labeled training data.

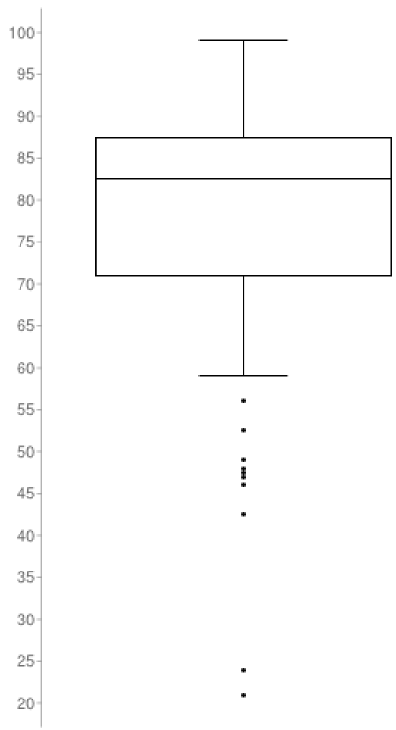

Figure 1. Distribution of impersonal model performance across 59 users in prior work [30].

- Hybrid models are a combination of personal and impersonal models. Users of the model will have labeled data in the training set, but the training set will also include data from other users.

Hybrid models make the most straightforward use of the labeled data—the labeled data can be partitioned into training and test data, randomly or via cross validation, without regard to who the data came from. But while this model type is used in many research studies, it *almost never* applies to real-world situations: commercial AR systems will use universal models or personal models. Our research also shows that it is unlikely that data from other users will improve AR performance if one has personal training data—even if only small amounts of such data is available [30].

Our own work [30] is the only exhaustive study of AR model type and training set composition (two other studies [10, 40] compare universal and personal models but provide very little discussion or analysis). Virtually all other work discusses only a single model type and thus cannot provide a comparative analysis of this factor; some work does generate two types of models but provides very little in the way of comparison or discussion. The main conclusions from our study can be summarized as follows: for the best performing algorithm (Random Forest) personal models perform extremely well (98% accurate), universal models perform much worse (76% accurate), and hybrid models perform in between the two (95% accurate). Furthermore, personal models perform well with very little training data (e.g., 1-2 minutes per activity) and the performance of personal models is very consistent over users, whereas the performance of impersonal models is very inconsistent (distribution shown in Figure 1). Thus, the number of subjects is of particular interest when considering model type since 1) performance results for impersonal models are unreliable if only a few users are analyzed and 2) one cannot evaluate the distribution of performance results with few users. This highlights the need for obtaining AR data from more than a few users.

It is also very important to note that the performance of the hybrid models is much closer to that of the personal models than the impersonal models. This was a surprise to us, since in our case the vast majority of the training data was from “other” users. This result throws enormous doubt on the results of other studies that use hybrid models, since hybrid usage scenarios will really not be found in actual deployed systems. Most prior work assumes (incorrectly and often implicitly) that hybrid models approximate the performance of impersonal models. Hybrid models are most likely utilized in existing research because they simplify the partitioning of the training and test data.

In the next few sections we examine various AR data set and methodological factors and see how they are handled in prior work. We point out any problems and limitations and provide some recommendations.

Factor 1: Model Type

Model type has a huge impact of AR performance and hence it is important to consider what types of models were used in prior AR research. The breakdown, based on our analysis of 47 models (built from 38 data sets in 34 published papers) is shown in Table 2. The table demonstrates that a variety of model types were used and that the hybrid model [4-5, 11-14, 17, 20-21, 25, 28, 32, 34, 37-38, 42] is the most popular (40%). These authors often claim that their results can be

Model Type	Count	%
Personal	12	26%
Impersonal	10	21%
Hybrid	19	40%
Unknown	6	13%

Table 2. Model type distribution across large sample of AR papers.

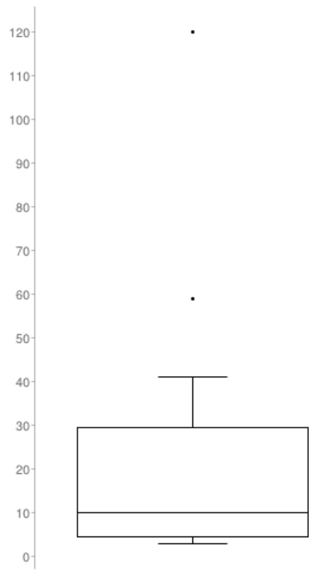

Figure 2. Data set sizes in related literature (HASC omitted due to scale but subset of HASC comparable to these studies contains 414 users)

generalized to new users, but as our prior work has shown [30] and as we discussed in the prior section, the results of evaluating hybrid models are dramatically better than the results from previously unseen users (i.e., for impersonal models). Furthermore, in 11 out of the 19 hybrid models that we analyzed, there were 10 or fewer users in the data set—a situation that even more closely resembles the personal model situation than in our case since we utilized 59 users [30]. Our analysis also found that personal models accounted for 26% of the models [2, 8, 10, 15-16, 18, 26-27, 33, 43, 45] and impersonal (universal) models accounted for 21% of the models [2-3, 6, 10, 19, 24, 39, 44]. We were unable to determine the model type in several studies [9, 23], which accounted for 13% of the models. This is significant given that model type dramatically impacts AR performance and is an enormous methodological oversight that impacts the utility of these paper's conclusions. Overall, these results are quite disturbing since in more than half of all cases (53%) the methodology used is either not proper (40%) or not even described (13%).

It is not easy to compare the relative performance for the three different types of models from our analysis of related work, because in most cases the models were built from different data sets and different AR tasks. Based only on the *averages* over the models in the related work analysis [31], hybrid models performed best (90% accurate), impersonal models second best (87% accurate), and personal models worst (84% accurate). In our AR research study [30] personal models did best, hybrid models second best, and impersonal models worst; thus the only consistent relationship between the two sets of results is that in both cases hybrid models outperformed the impersonal models. However, if we analyze only the cases from the related work where models are generated on the same datasets (the only fair comparison), we find that in 2 of 2 cases [2, 10] the personal model outperforms the impersonal model and in 2 out of 3 cases [26-27, 45] the personal model outperforms the hybrid model.

Factor 2: Number of Subjects & Diversity

Many studies use very limited datasets, often with fewer than 5 subjects [2-3, 12, 15-16, 18, 28-29, 33, 38, 44] or 10 subjects [11, 13-14, 17, 20, 23, 32-33, 37, 39, 45]. Compounding the issue, the most widely used AR datasets, COSAR and OPPORTUNITY, have data from only 4 and 12 subjects, respectively [35-36]. Data sets with more users, such as HASC 2010 and HASC 2011 [22], contain relatively small amounts of data per person. This relative lack of AR data motivated us to release our AR dataset [41] with 59 subjects. A summary box plot of data set sizes from our related work analysis [31] is presented in Figure 2.

The number of subjects present in the data set does not just impact the quality and robustness of the induced AR model, but also the ability to evaluate the consistency of results across subjects. This is important since our prior results (summarized in Figure 1) show that AR performance of impersonal models is *extremely* inconsistent across users [30]. Thus it is critically important to have a substantial number of subjects for evaluation in order to obtain reliable results. In our analysis of related work we found 4 studies that utilized universal models with fewer than 8 subjects [2, 3, 39, 44] and, aside from our study, only 2 studies of universal models included at least 30 subjects [6, 24].

The number of subjects is not all that matters for building good AR models. The subject population should be as diverse as possible (for specific applications one can always restrict the subjects that are used). Diversity takes many forms but it should consider age, gender, health, height, weight, and other demographic factors. While we do not present all of the details here about the diversity of the observed data sets—and the information is not available for many data sets—the majority of studies do mention the general population from which they subjects are drawn. Based on our analysis of related work [31], a large fraction of the studies focus on college students—most likely because most research occurs on college campuses. Data sets intended for widespread use and varied applications should try to sample a much more diverse population and should also document the characteristics of the subjects and link them to the underlying data (some datasets provide this information but many do not). We do this for one of our activity recognition data sets [41], which contains detailed information about each subject (e.g., gender, height, weight, age).

Factor 3: Collection Methodology

The way the data is collected is important and should be noted. While many distinctions can be made, at a high level there are three types of AR collections:

- Fully natural: subjects go about their normal daily activities without altering their behavior.
- Semi-natural: subjects operate in their normal environment but modify their behavior in modest ways, such as ensuring that they walk or perform other specific activities.
- Laboratory: subjects perform specific structured activities in a laboratory environment.

Activity data sets should document the methodology used and try to also map the collection methodology into one of these categories. It is important to document this because performance will be superior in laboratory settings and hence for fairness this information should be disclosed. Some studies do provide this information, but some do not. We have released two datasets [41]: an "activity recognition" data set that uses semi-natural AR data collection and our "Actitracker" data set that relies on fully natural data collection via our Actitracker app [1].

Factor 4: Sensors

All aspects of the sensors that impact the AR data should be described. Some of these, such as the type of sensor (accelerometer, gyroscope, GPS) and number of sensors, are always appropriately described. However, the precise location of each sensor is not always specified and the orientation is almost never specified. For example, if a smartphone is placed in one's pants pocket, which pocket is it placed in? How is the phone oriented? While this information may not seem important, our experience has demonstrated that these factors impact AR performance. Thus, in our experimental protocol, we often do specify exact location and orientation; however, we have also, like many others, failed to provide these details in the methodology section of our research papers. For personal models these details may not be quite as important since individualized models are generated—but even in these cases the models may have problems if the location and orientation changes over time. If protocols constrain the location and orientation but are not documented, this will impact the ability of others to reproduce the work and draw appropriate conclusions.

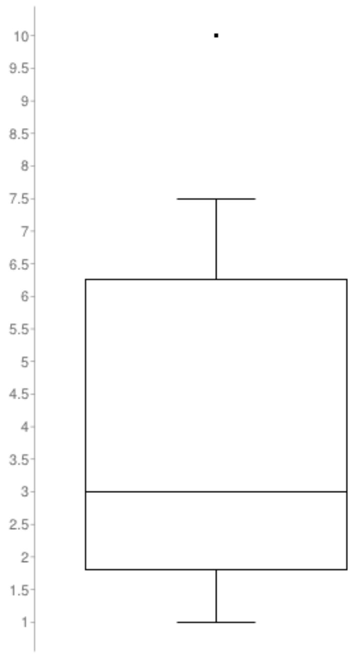

Figure 3. Distribution of window sizes for data transformation for the 52% of related work that reports this information.

Factor 5: Features and Feature Generation

The features used to describe the data are of critical importance. Generally there is little choice in the representation of the raw time-series data since this is determined by the sensor. Tri-axial accelerometers in smartphones yield three real numbers for the x, y, and z directions, with the values on Android phones generally varying between $-2g$ and $+2g$ ($\pm 19.6 m/s^2$). However, the sampling rate used to record the data in the dataset does involve some choice. In our work we use a sampling rate of 20Hz for the accelerometer [25, 30], which generally provides sufficient resolution for activity recognition.

The raw time-series sensor data is typically transformed into a higher level multivariate record format, using a sliding window technique, so that conventional classification algorithms can be used. It is important for AR researchers to describe this transformation process and make either the transformed data—or ideally the code to generate the transformed data—available so that others can reproduce their experimental results and analyze the efficacy of alternate encoding schemes. In the transformed data each generated variable represents some aspect of the data (e.g., average *x* acceleration) over the time period associated with the sliding window. Window lengths vary from 1 to 10 seconds and the values used in practice, based on our analysis of the 38 research datasets from related work, are detailed in Figure 3. Almost half of the research papers do not mention the window size that is utilized, further indicating omissions in providing relevant methodological details. Our research [25, 30] utilizes a window size of 10 seconds and currently each window slides the full 10 seconds so each sensor reading appears in only one example (i.e., we use non-overlapping windows).

The large majority of smartphone AR research uses only basic statistics (e.g. mean, standard deviation, binned distribution) derived from accelerometer data in each window as the input features for classification [4, 10-14, 19, 20, 25, 29, 32-33, 35, 39, 44-45]. Research using these features has been extremely successful with personal and hybrid model types, often achieving accuracies in the high 90's [25, 30]. Universal models using these features have often performed generally in line with work using more advanced features [30]. The key advantage to using only basic statistics is that they are computationally lightweight to calculate [25], which enables them to be calculated on smartphones, which have limited CPU and battery resources. Features which are quick to calculate are also advantageous in real-time scenarios and in aggregate central-processing schemes, where computation time is a key concern.

A growing trend is to use more advanced signal processing techniques [4, 6, 17, 24, 26-27, 34]. While the 'basic statistics' approach extracts features from the time domain, this approach uses Fourier transforms (FFT) to transform the time-series data into frequency domain data. New features are then extracted from the FFT output by an additional processing step according to either domain expertise (e.g. knowledge about the rate of human gait), or to signal processing standards from other fields (e.g. audio analysis). Studies using FFT information sometimes [34], but not always [26], perform well, and overall there is not conclusive evidence as to whether these features improve activity recognition performance.

WISDM Activity Recog. Data (Semi-Natural)

Raw labeled time-series data

# examples:	1,098,207
# attributes:	6

Transformed labeled data

# examples:	5,424
# attributes:	46

Approximate Class Distribution

Walking:	39%	Standing:	4%
Jogging:	31%	Upstairs:	11%
Sitting:	6%	Downstairs:	9%

WISDM Actitracker Data Set (Fully Natural Labeled)

Raw labeled time-series data

# examples:	2,980,765
# attributes:	6

Transformed labeled data

# examples:	5435
# attributes:	46

Approximate Class Distribution

Walking:	42%	Standing:	10%
Jogging:	15%	Lie down:	9%
Sitting:	22%	Stairs:	2%

WISDM Actitracker Data Set (Fully Natural Unlabeled)

Raw unlabeled time-series data

# examples:	751,004,153
# attributes:	6

Transformed unlabeled data

# examples:	1,369,349
# attributes:	46

Table 3. WISDM AR data sets.

Researchers have also occasionally tried alternative approaches. Several papers [2, 8, 43] have used the raw time series accelerometer directly as input to classification algorithms (usually nearest neighbor). While this approach shows promise, it presents researchers with new challenges. First, the problem of segmenting and aligning the data becomes much more important, since misaligned segments will classify incorrectly. Second, when each 3-dimensional accelerometer reading (sampled at 20-100Hz) is used as a feature in nearest-neighbor algorithms, the computation time grows rapidly in comparison to statistical feature approaches, which typically only use a few dozen features. Other research teams have constructed features from more esoteric statistics [5, 21, 23-24], or used data transformations other than FFT such as discrete cosine transforms [19, 24]. While each method shows promise, there is little consistency across these studies when compared with studies using more common methodologies. It is worth noting that these methods generally require more computational resources than basic statistics.

It is quite important that each released AR data set include the raw sensor data and the transformed data or, alternatively, a script to generate the transformed data from the raw data. Specifications of the transformation process, perhaps documented in a research paper, are often not sufficient given the potentially large time effort for implementing the transformation code, as well as potential ambiguity in the specification. In our released data sets, we provide both the raw and transformed data and reference papers which describe the transformation process [25]. Interestingly, several researchers have identified minor unexpected differences in our actual transformed data, leading us to further enhance our documentation. This lesson reinforces the need for complete and detailed information about AR methodology.

WISDM Activity Recognition Data Sets

In this section we briefly describe our publically available data sets [41]. While we do believe that we provide a more complete description than many other researchers, we in no way believe our data set, or associated documentation, is ideal. In the remainder of this section we specify the values for the information listed earlier in Table 1 for our WISDM data sets. Some of the most basic statistics are provided in Table 3.

We have two main data sets. The first “activity recognition” data set was generated by having 59 test subjects perform a specific sequence of activities while outside, which is a semi-natural form of data collection [25]. The second data set was generated from our Actitracker activity recognition app [1], which is available for free from the Google Play store, and which is continuously collecting new data. The data from this app is stored on our server and corresponds to fully natural data collection since we have no control over the user. There are two subsets of data associated with this Actitracker data set: the first is labeled data that is generated when the user optionally executes the app’s training mode and the second is the unlabeled data that is captured otherwise. Since some users run this app throughout the day, the app has captured thousands of hours of unlabelled data.

Our research-oriented activity recognition data set has been utilized to generate and evaluate personal, impersonal, and hybrid models. Others may generate such models from our data by properly partitioning the data file. The publically available Actitracker app utilizes an impersonal model by default for users; however, once the training mode is executed to generate labeled training data, the system automatically generates and deploys a personal model.

We currently only collect tri-axial accelerometer data, although we intend to start collecting gyroscope and barometer data (for phones with these sensors). Our accelerometer sampling rate is 20Hz. For the more highly controlled activity recognition data set, we ask subjects to place the phone in their right front pants pocket facing up and out. For the fully natural Actitracker data set, subjects put the phone in whatever location and orientation they desire and this may vary over time. We use a variety of basic statistical features and a window size of 10 seconds to generate the transformed data [25].

Conclusions and Recommendations

In this paper we described information that should be provided for every AR data set and recommend that each research study provide all of the information listed in Table 1. We further recommend that AR data sets provide as much diversity as possible. Our analysis of dozens of research studies demonstrated that many research studies do not document important methodological information and also that the studies and data sets are often quite limited by the diversity of the data set and/or methodological limitations. We also focused on model type, which in prior work we showed to have a dramatic impact on AR performance. We showed that more than half of the models from related work use hybrid models or unspecified models—and in these cases any conclusions drawn about AR performance are suspect.

In summary, we strongly recommend that each research study provide detailed information about their AR data sets and the methodology used to create the data sets and build predictive models. The data sets should be as diverse as possible and the model types that are employed should match the real-world setting in which they would be used. Ideally the results for all three types of models would be reported. Ultimately AR problems differ sufficiently so that one "ideal" format and configuration may not be possible, but by generating diverse data sets the data at least could prove useful in a greater variety of situations. Hopefully this paper will encourage researchers to consider methodological issues much more carefully.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1116124.

References

[1] actitracker.com

[2] Abdallah, Z.S., Gaber, M.M., Srinivasan, B., and Krishnaswamy, S. CBARS: Cluster based classification for activity recognition systems. In *Proc. 1st Int. Conference on Advances Machine Learning Technologies* (2012).

[3] Abdallah, Z.S., Gaber, M.M., Srinivasan, B., and Krishnaswamy, S., StreamAR: incremental and active learning with evolving sensory data for activity recognition. In *Proc. 24th IEEE Int. Conf. on Tools with AI* (2012).

[4] Anguita, D., Ghio, A., Oneto, L., Parra, X., and Reyes-Ortiz, J. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *Proc. Int. Workshop of Ambient Assisted Living* (2012).

[5] Anguita, D., Ghio, A., Oneto, L., Parra, X., and Reyes-Ortiz, J.L. A public domain dataset for human activity recognition using smartphones. In *Proc. European Symposium on Artificial Neural Networks, Comp. Intelligence, and Machine Learning* (2013), 437-442.

[6] Anguita, D., Ghio, A., Oneto, L., Parra, X., and Reyes-Ortiz, J.L. Energy efficient smartphone-based activity recognition using fixed-point arithmatic, *Journal of Universal Comp. Science*, 19, 9 (2013), 1295-1314.

[7] Bao, L., and Intille, S. Activity recognition from user-annotated acceleration data. *Lecture Notes Computer Science 3001* (2004), 1-17.

[8] Brezmes, T., Gorricho, J.L., and Cotrina, J. Activity recognition from accelerometer data on mobile phones. In *Proc. 10th International Work-Conference on Artificial Neural Networks* (2009), 796-799.

[9] Cho, Y., Nam, Y., Choi, Y-J, and Cho, W-D. SmartBuckle: human activity recognition using a 3-axis accelerometer and wearable camera. *HealthNet* (2008).

[10] Derawi, M., and Bours, P. Gait and activity recognition using commercial phones. *Computers and Security*, 39 (2013): 137-144.

[11] Dernbach, S., Das, B., Krishnan, N.C., Thomas, B.L., and Cook, D.J. Simple and complex activity recognition through smart phones. In *Proc. 8th Int. Conference on Intelligent Environments* (2012), 214-221.

[12] Do, T.M., Loke, S.W., and Liu, F. HealthyLife: An activity recognition system with smartphone using logic-based stream reasoning. *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, 120 (2013), 188-199.

[13] Fahim, M. Evolutionary learning models for indoor and outdoor human activity recognition. *PhD Thesis*, Kyung Hee University (2014).

[14] Fahim, M., Fatima, I., and Lee, S. EFM: Evolutionary fuzzy model for dynamic activities recognition using a smartphone accelerometer. *Applied Intelligence*, 39, 3, (2013), 475-488.

[15] Gomes, J.B., Krishnaswamy S., Gaber, M.M., Sousa, A.C., and Menasalvas, E. Mobile activity recognition using ubiquitous data stream mining. In *Proc. 14th Int. Conf. on Data Warehousing and Knowledge Discovery* (2012).

[16] Gyorbiro, N., Fabian, A., and Homanyi, G. An activity recognition system for mobile phones. *Mobile Networks and Applications* 14, 1 (2008), 82-91.

[17] M. Han, An intergativ Human Activity Recognition Framework based on Smartphone Multimodal Sensors, *Doctoral Thesis*, Kyung Hee University, 2013.

[18] Han, M., Bang, J-H, Nugent, C., McClean, S., and Lee, S. HARF: A hierarchical activity recognition framework using smartphone sensors, *Ubiquitous Computing and Ambient Intelligence, Lecture Notes in Computer Science,* 8276 (2013), 159-166.

[19] He, Z., and Jin, L. Activity recognition from acceleration data based on discrete cosine transform and SVM. In *Proc. IEEE International Conference on Systems, Man, and Cybernetics* (2009).

[20] He, Y., and Li, Y. Physical activity recognition utilizing the built-in kinematic sensors of a smartphone*. Int. Journal of Distributed Sensor Networks* (2013).

[21] Kastner, M., Strickert, M., and Villmann, T. A sparse kernelized matrix learning vector quantization model for human activity recognition. In *Proc. European Symposium on Artificial Neural Networks, Comp. Intelligence, and Machine Learning* (2013).

[22] Kawaguchi, N., et al. HASC Challenge: gathering large scale human activity corpus for the real-world activity understandings. In *Proc. 2nd Augmented Human International Conference* (2011).

[23] Khan, A.M., Lee, Y.K., Lee, S.Y., and Kim, T.S. Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis. In *Proc. 5th Int. Conf. on Future Information Technology (2010),* 1-6.

[24] Khan, A.M., Siddiqi, M.H., and Lee, S-W. Exploratory data analysis of acceleration signals to select light-weight and accurate features for real-time activity recognition on smartphones. *Sensors* 13 (2013), 13099-13122.

[25] Kwapisz, J.R., Weiss, G.M., and Moore, S.A. Activity recognition using cell phone accelerometers, *SIGKDD Explorations*, 12 2 (2010), 74-82.

[26] Lane, N.D., Xu, Y., Lu, H., Hu, S., Choudhury, T., Campbell, A.T., and Zhao, F., Enabling large-scale human activity inference on smartphones using community similarity networks (CSN). In *Proc. 13th International Conference on Ubiquitous Computing* (2011), 355-364.

[27] Lane, N.D., Xu, Y., Lu, H. Hu, S., Choudhury, T. Campbell, A.T., and Zhao, F. Community similarity networks. *Personal Ubiquitous Computing*, 18, 2 (2014), 355-358.

[28] Lee, Y., and Cho, S. Activity recognition using hierarchical hidden Markov models on a smartphone with 3D accelerometer. In *Proc. 6th Int. Conf. on Hybrid Artificial Intelligent Systems* (2011), 460-467.

[29] Lee, Y., and Cho, S. Activity recognition with Android phone using mixture-of-experts co-trained with labeled and unlabeled data. *Neurocomputing, 126* (2014), 106-115.

[30] Lockhart, J.W., and Weiss, G.M. The benefits of personalized smartphone-based activity recognition models. In *Proc. SIAM International Conference on Data Mining (SDM)* (2014), 614-622.

[31] Lockhart, J.W. The Benefits of Personalized Data Mining Approaches to Human Activity Recognition with Smartphone Sensor Data. *Master's Thesis*, Dept. of Computer & Info. Science, Fordham University (2014). [http://www.cis.fordham.edu/wisdm/Lockhart-MS-thesis.pdf]

[32] Miluzzo, E., Lane, N.D., Fodor, K., Peterson, R., Lu, H., Musolesi, M., Eisenman, S.B., Zheng, X., and Campbell, A.T. Sensing meets mobile social networks: the design, implementation and evaluation of the CenceMe application. In *Proc. 6th ACM Conference on Embedded Networked Sensor Systems* (2008), 337-350.

[33] Qi, X. Keally, M., Zhou, G., Li, Y., Ren, Z. AdaSense: adapting sampling rates for activity recognition in body sensor networks. In *Proc. IEEE 19th Real-Time and Embedded Technology and Applications Symposium (*2013), 163-172.

[34] Reiss, A., Hendeby, G., and Stricker, D. A competitive approach for human activity recognition on smartphones. In *Proc. European Symposium on Artificial Neural Networks, Computational Intelligence*, *and Machine Learning* (2013), 455-460.

[35] Riboni, D., and Bettini, C. COSAR: hybrid reasoning for context-aware activity recognition. *Personal and Ubiquitous Computing*, 15, 3 (2011), 271-289.

[36] Roggen, D. et al. Opportunity: towards opportunistic activity and context recognition systems. In *Proc. 3rd IEEE Workshop on Autonomic and Opportunistic Communications* (2009).

[37] Roy, N., Misra, A., and Cook, D. Infrastructure-assisted smartphone-based ADL recognition in multi-Inhabitant smart environments. *In Proc. IEEE Int. Conf. on Pervasive Computing and Communications* (2013), 38-46.

[38] Shoaib, M., Scholten, H., and Havinga, P.J.M. Towards physical activity recognition using smartphone sensors. In *Proc. 10th International Conference on Ubiquitous Intelligence & Computing* (2013), 80-87.

[39] Siirtola, P., and Roning, J. Ready-to-use activity recognition for smartphones, In *Proc. IEEE Symposium on Comp. Intelligence and Data Mining (*2013), 59-64.

[40] Tapia, E., Intille, S., Haskell, W., Larson, K., Wright, J., King, A., and Friedman, R. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In *Proc. 11th IEEE Int. Symp. on Wearable Computers* (2007), 37-40.

[41] www.cis.fordham.edu/wisdm/dataset.php

[42] Wu, W., Dasgupta, S., Ramirez, E.E., Peterson, C., and Norman, G.J. Classification accuracies of physical activities using smartphone motion sensors. *Journal of Medical Internet Research,* 14, 5 (2012).

[43] Yan, Z., Chakraborty, D., Mittal, S., Misra, A., and Aberer, K. An exploration with online complex activity recognition using cellphone accelerometer. *UbiComp Adjunct* (2013), 199-202.

[44] Yang, J. Toward physical activity diary: motion recognition using simple acceleration features with mobile phones. I*n Proc. 1st Int. Workshop on Interactive Multimedia for Consumer Electronics* (2009).

[45] Yuan, B., Herbert, J., and Emamian, Y. Smartphone-based activity recognition using hybrid classifier utilizing cloud infrastructure for data analysis. In *Proc. of the 4th Int. Conference on Pervasive and Embedded Computing & Comm. Systems*. 2014.