Smartwatch-based Activity Recognition: A Machine Learning Approach*

Gary M. Weiss, Jessica L. Timko, Catherine M. Gallagher, Kenichi Yoneda, and Andrew J. Schreiber

***Abstract*— Smartwatches and smartphones contain accelerometers and gyroscopes that sense a user's movements, and can help identify the activity a user is performing. Research into smartphone-based activity recognition has exploded over the past few years, but research into smartwatch-based activity recognition is still in its infancy. In this paper we compare smartwatch and smartphone-based activity recognition, and smartwatches are shown to be capable of identifying specialized hand-based activities, such as eating activities, which cannot be effectively recognized using a smartphone (e.g., smartwatches can identify the "drinking" activity with 93.3% accuracy while smartphones achieve an accuracy of only 77.3%). Smartwatch-based activity recognition can form the basis of new biomedical and health applications, including applications that automatically track a user's eating habits.**

I. Introduction

Activity recognition research originally utilized specially engineered devices distributed across a subject's body [2, 6, 11] to identify the subject's physical activities, but in recent years much of this research began utilizing commercial smartphones [7, 9, 11, 14], which include the requisite sensors (i.e., accelerometers and gyroscopes). The use of these ubiquitous commercial devices greatly expanded the applications of activity recognition, but also introduced limitations due to their placement on a user's body and inconsistent orientation. For example, the smartphone could shift in a person's pocket and the pocket position (near the upper thigh) is not ideal for tracking hand-based activities. Smartphone-based activity recognition is especially limited for women, since they typically do not keep the phone in their pocket. Most of these limitations are addressed by smartwatches, which are worn in a consistent position and which are ideally situated for tracking hand-based activities. Furthermore, since virtually all smartwatches work in tandem with smartphones, the sensor information from both devices can be utilized for activity recognition.

This paper examines the use of smartphones and smartwatches for activity recognition. The performance of smartphone-based activity recognition is compared with the performance of smartwatch-based activity recognition—although we recognize that ultimately a combination of both devices may work best. The efficacy of the smartwatch accelerometer is also compared with the efficacy of the smartwatch gyroscope sensor for performing activity recognition. Our prior research demonstrated that smartphone-based personal activity recognition models—built with training data from the intended user—vastly outperform impersonal models, and this study shows that this advantage extends to smartwatch-based activity recognition models. Finally, this study also extends our prior work by recognizing many more activities, including hand-based activities (e.g., typing and writing). Of special note, this study includes many eating-related activities (e.g., eating soup, eating a sandwich, drinking), which opens up the possibility for new health-related activity recognition applications. The results in this paper demonstrate that smartwatches have the potential to accurately identify a large variety of activities, including hand-based and eating-based activities that cannot be effectively recognized by smartphones. Consistent with our prior smartphone work [7, 9], the classification models are induced from labeled training data using standard machine learning algorithms.

Smartwatch and smartphone-based activity recognition has many applications [8]. Generally, these devices can operate more intelligently if they are aware of what their user is doing (e.g., forwarding a call to voicemail during a meal). However, the main application of our research has been to improve people's health and wellbeing. Physical inactivity and unhealthy eating are two of the most powerful, modifiable risk factors for disease. Performing a sufficient amount of physical activity is important because physical inactivity dramatically increases the health risks for cardiovascular disease [4], colon cancer [5], and many other diseases—while a healthy amount of physical activity reduces the risk of all-cause mortality [1] and could prevent two million deaths per year [12]. Similarly, there are significant health risks associated with excessive caloric intake [10]. While many types of interventions seek to reduce the tendency toward overeating, long-term dietary adherence remains a major challenge [3]. Activity monitoring can help combat both inactivity and overeating by providing accurate, real-time information about sedentary behavior, exercise, and eating behavior. The smartwatch is perfectly poised to convey this information since it is always readily and unobtrusively accessible—which is one reason why smartwatch manufacturers tout its potential to improve health. While there are some basic activity recognition applications for smartwatches, our work involves much more specific activities (including a variety of eating activities). Smartwatch-based apps capable of tracking eating activities could ultimately replace (or augment) the manually intensive methods for maintaining a food diary.

II. The Activity recognition Task

The activity recognition task involves mapping time series sensor data from a smartwatch and/or smartphone to a single activity. In our approach the time series data is

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

All authors are with the Department of Computer and Information Science, Fordham University, Bronx, NY 10458 USA. (Corresponding author: Phone 718-817-0785; email: gaweiss@fordham.edu).

aggregated into examples based on non-overlapping 10-second intervals of data; an activity is recognized correctly if the single activity that occurred during the 10-second interval is correctly classified. Table I lists the eighteen activities included in this study. The activities are grouped into three general categories. The details associated with the data collection process are described in the next section.

TABLE I. EIGHTEEN ACTIVITIES UTILIZED IN THIS STUDY

- General activities (not hand-oriented)
  - Walking
  - Jogging
  - Climbing Stairs
  - Sitting
  - Standing
  - Kicking Soccer Ball
- General activities (hand-oriented)
  - Dribbling Basketball
  - Playing Catch with Tennis Ball (two people)
  - Typing
  - Handwriting
  - Clapping
  - Brushing Teeth
  - Folding Clothes
- Eating activities (hand-oriented)
  - Eating Pasta
  - Eating Soup
  - Eating Sandwich
  - Eating Chips
  - Drinking from a Cup

III. EXPERIMENT METHODOLOGY

The methodology associated with generating and evaluating the activity recognition models is discussed in this section. This includes the data collection procedure, the method for transforming the low-level time-series sensor data into examples, and the model induction process. Some of this material has been presented in our prior smartphone-based work [7, 9], so in these cases some details may be omitted and replaced by appropriate citations.

A. *Data Collection*

The data for this study was collected from 17 test subjects, each of whom were asked to perform the 18 activities listed in Table I. Each activity was performed for two minutes while the subject wore a smartwatch on their dominant hand and a paired smartphone in their front-right pocket with the phone oriented upright with the screen facing outward. All data was collected using the LG G Watch smartwatch and the Samsung Galaxy S4 smartphone, both of which were running the Android Wear mobile operating system. The data collection process took approximately one hour per person. All test subjects provided written informed consent prior to participating in this study, which was approved by our university's Institutional Review Board.

The data collection process utilizes our custom designed smartwatch/smartphone app. The app collects data from the accelerometer and gyroscope on both the phone and watch at a rate of 20Hz, with each sensor providing data for the three spatial dimensions. After two minutes of data is collected for an activity, the smartphone automatically sends the data, in an email message, to a server for storage and later retrieval. Unfortunately, unresolved and sporadic issues with the phone's gyroscope prevented sufficient phone-gyroscope data from being collected and hence this sensor is not incorporated into our analysis. The data received by the server was later trimmed to ensure that the collected data corresponded to an activity (sometimes test subjects delayed in starting the activity or ended early). The trimming resulted in having approximately 100 seconds of data for each activity rather than the full two minutes.

B. *Data Transformation*

Conventional classification algorithms do not operate on time-series data, so the raw time series sensor data was transformed into examples by taking 10-second chunks of data and generating higher level features that describe the behavior over this time period. The 43 higher level features involve computing averages, standard deviations, and other measures from the x, y, and z axis values for the sensor. The details of the data transformation process, including a description of the higher level features, is provided in our prior work on smartphone-based activity recognition [7, 9]. Note that in this study we generate examples from only one sensor at a time; thus when we present the results in Section IV, the activity recognition performance of each sensor is provided separately (i.e., the models utilize a single sensor).

C. *Model Induction*

The activity recognition models are induced from the labeled training examples using the WEKA [13] Random Forest (RF) algorithm, J48 decision tree algorithm, IB3 instance-based learning algorithm, Naïve Bayes (NB) algorithm, and the multi-layer perceptron (MLP) algorithm. The default parameter settings for each algorithm are used, except for NB where kernel estimation is enabled, and the instance-based learning algorithm where 3 nearest neighbors are used (there are dozens of default parameters for the algorithms employed in this paper so please refer to WEKA online documentation for more details).

Two types of models are induced. Personal models utilize data from only the intended user while impersonal models are built using training data from everyone but the intended user. Our prior research has shown that personal models vastly outperform impersonal models for activity recognition [9], but at the cost of requiring each user to provide labeled training data. In order to build and evaluate the personal models, data from each of the 17 subjects is partitioned into training and test sets using 10-fold cross validation. Therefore the results for the personal models in the next section are based on results averaged over $17 \times 10 = 170$ models. The impersonal models are generated by taking training data from 16 users, building a model, and then evaluating that model on the remaining user; this is repeated 17 times so that all subjects are evaluated. In this case 10-fold cross validation is not used since we have plenty of training data. The results for the impersonal models are averages over the 17 generated models.

IV. RESULTS

The activity recognition results are presented and analyzed in this section. The results for the personal and impersonal activity recognition models are presented in Table II and Table III, respectively. Results are measured using classification accuracy, which in this case corresponds to the percentage of the classifications that correctly identify the activity the user is performing. Each model utilizes a single sensor—models using the watch-accelerometer, phone-accelerometer, and watch-gyroscope are all evaluated. A model is also generated using each of the five learning algorithms described earlier. All results in Table II and III are based on 17 test subjects who performed the 18 activities that were listed in Table I.

TABLE II. OVERALL ACCURACY FOR PERSONAL MODELS

Algorithm	Phone accel (%)	Watch accel (%)	Watch gyroscope (%)
RF	75.5	93.3	79.0
J48	65.5	86.1	73.0
IB3	67.7	93.3	60.1
NB	77.1	92.7	80.2
MLP	77.0	94.2	70.0
Ave	72.6	91.9	72.4

TABLE III. OVERALL ACCURACY FOR IMPERSONAL MODELS

Algorithm	Phone accel (%)	Watch accel (%)	Watch gyroscope (%)
RF	35.1	70.3	57.5
J48	24.1	59.3	49.6
IB3	22.5	62.0	49.3
NB	26.2	63.8	53.5
MLP	18.9	64.6	57.7
Ave	25.3	64.0	53.5

The results show that, consistent with our previous smartphone-based activity recognition results [9], personal models outperform impersonal models, even though they are built using much less training data. The results also show that the watch accelerometer provides much better results than the phone accelerometer, with an average accuracy of 91.9% versus 72.6% for personal models and 64.0% versus 25.3% for impersonal models. These differences are largely due to the hand-based activities included in this study. The watch gyroscope performs much more poorly than the watch accelerometer, with an average accuracy of 72.4% versus 91.9% for the personal models and 53.5% versus 64.0% for the impersonal models (as mentioned earlier there were technical difficulties in capturing the phone gyroscope data). While it makes sense to generate models using the fusion of multiple sensors, this is not done because of issues synchronizing the data from different sensors—possibly running on different devices. Since the watch accelerometer yields the best models for both personal and impersonal models, we focus most of our attention on this sensor. It is worth emphasizing that the results for this sensor are quite impressive, given the diversity of activities that are tracked, the granularity of the activities (i.e., eating activities are not all grouped together), and the fact that a strategy of guessing the most common activity would yield an accuracy of only about 5% (i.e., about 1 in 18).

TABLE IV. PER-ACTIVITY ACCURACY FOR RF MODELS

Activity	Impersonal (%)			Personal (%)		
	Watch accel	Phone accel	Watch gyro	Watch accel	Phone accel	Watch gyro
Walking	79.8	60.7	87.0	94.2	88.5	93.5
Jogging	97.7	93.8	48.6	99.2	68.8	98.1
Stairs	58.5	66.7	43.1	88.9	66.7	80.0
Sitting	84.9	26.9	70.5	97.5	87.0	82.2
Standing	96.3	65.9	57.9	98.1	73.1	68.6
Kicking	71.3	72.5	41.4	88.7	91.7	67.9
Dribbling	89.3	26.1	86.0	98.7	84.8	96.9
Catch	66.0	26.1	68.9	93.3	78.3	94.6
Typing	80.4	76.9	60.8	99.4	72.0	88.6
Handwriting	85.2	12.9	63.1	100.0	75.9	80.5
Clapping	76.3	40.9	67.9	96.9	77.3	95.6
Brush Teeth	84.5	19.2	66.2	97.3	96.2	89.6
Fold Clothes	80.8	8.3	37.8	95.0	79.2	73.1
Eat Pasta	47.1	0.0	57.9	88.6	40.0	72.9
Eat Soup	52.7	0.0	47.7	90.7	82.4	69.8
Eat Sandwich	29.0	7.1	31.1	68.9	63.0	44.2
Eat Chips	65.0	16.0	50.6	83.4	76.0	52.5
Drink	62.7	31.8	61.1	93.3	77.3	78.5
Overall	70.3	35.1	57.5	93.3	75.5	79.0

It is informative to examine the accuracies associated with individual activities, in order to determine which activities are easy to recognize and which ones are difficult to recognize. Due to space limitations, we focus our analysis on models induced using the Random Forest (RF) algorithm, since this algorithm performs well over the 6 configurations (3 sensors and 2 types of models). The accuracies of the RF models, over the 18 activities for both personal and impersonal models, are shown in Table IV. These results indicate that the accuracy for the activities varies widely. The last group of activities (i.e., eating activities) seems to have the lowest accuracy when compared to the results for the other two groupings. This suggests perhaps that the eating activities may be similar and may be getting confused with one another. We also see that the phone sensor performs much worse than the two watch sensors for the eating activity, for both the impersonal and personal models—but especially for the impersonal models. This is most likely due to that fact that many of the hand-based activities in the second grouping also involve significant body motion (e.g., dribbling and folding clothes). It is interesting that a phone in one's pocket can be effective at identifying many hand-based activities, but is ineffective at identifying eating activities. We had previously seen that personal models outperform impersonal models, but the results in Table IV show that personal models are particularly effective at boosting the performance of the eating activities—perhaps suggesting that there is wide variance in how people eat.

The results in Table IV do not tell us how there errors are distributed or—more to the point—which activities are confused with one another. A model's confusion matrix can answer this, but since each confusion matrix has dimensions 18×18, it is not practical to display them here. But we can analyze these matrices for the most problematic cases. For example, if we look at the results for the personal RF models built from the watch-accelerometer sensor, we see that the hardest activity to recognize is "eating a sandwich," which has a 68.9% accuracy. The confusion matrix indicates that this activity is misclassified as "eating soup" 10.9% of the time, "eating chips" 9.4% of the time, and "eating pasta" 5.8% of the time (other misclassifications occur less frequently). This shows that the common mistakes involve other eating activities, suggesting that much higher accuracy could be achieved by grouping all of the eating activities together. Similarly, the impersonal model using the phone accelerometer has a 0% accuracy for "eating soup" and the underlying confusion matrix shows that the single largest source of errors is due to misclassifying this activity as "drinking." It seems reasonable that these two activities would appear similar based on the motion measured at a person's upper thigh (i.e., by the pants pocket).

V. Conclusion and Future work

This paper demonstrates that smartwatch-based activity recognition is capable of recognizing a wide variety of activities—including some activities that a smartphone cannot effectively recognize. This paper also includes what is perhaps the first research study for tracking eating activity with a smartwatch. Overall, the results in this paper suggest that commercial smartwatches can recognize a wide variety of activities with relatively good accuracy. The results further show that personal models perform best—especially for the eating activities. Thus, our observations about the superiority of personal models, previously demonstrated for smartphones [9], is now shown to hold for smartwatches and for a much more diverse set of activities. This paper also compares the performance of the accelerometer and gyroscope for activity recognition, and the watch accelerometer is shown to significantly outperform the watch gyroscope.

The study described in this paper is at an early stage, and we expect to extend the study in several key ways. First, we plan to extend the study to include a minimum of 50-100 test subjects (this study included 17 subjects). The amount of data per activity will be increased to 3 minutes since after trimming the data we fell below our goal of having 2 minutes of data per activity (prior research [9] indicates that performance plateaus at 2 minutes of personal training data per activity). We also expect to overcome the technical difficulties with collecting the phone gyroscope data so we can better compare the utility of the various sensors for activity recognition. Fusing the sensor data is likely to yield improved activity recognition results and we plan to evaluate such fusion strategies in the future, after developing methods to better synchronize the sensor data. Finally, our activity recognition models rely on the fact that a clear pattern will occur within 10 seconds of data. This was valid for the clearly repetitive activities that we initially studied—like walking, jogging, and stair climbing—but this may not be appropriate for eating activities that may not repeat in such a short time frame. We plan to explore alternative strategies for handling such activities.

References

[1] S. N. Blair, H. W. Kohl, R. S. Paffenbarger, D. G. Clark, K. H. Cooper, and L. W. Gibbons, "Physical fitness and all-cause mortality: A prospective study of healthy men and women," *Journal of the American Medical Association, vol.* 262, pp. 2395-2401, 1989.

[2] Y. Cho, Y. Nam, Y-J Choi, and W-D Cho, "SmartBuckle: human activity recognition using a 3-axis accelerometer and a wearable camera," in *HealthNet '08 Proceedings of the 2nd Int. Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*.

[3] K. R. Fontaine, and L. J. Cheskin, "Self-efficacy, attendance, and weight loss in obesity treatment," *Addict Behav*, vol. 22, pp. 567–70, 1997.

[4] S. M. Fox, J. P. Naughton, and W. L. Haskell, W. L., "Physical activity and the prevention of coronary heart disease," *Annals of Clinical Research, vol.* 3, pp. 404-432, 1971.

[5] M. Gerhardsson, S.E. Norell, H. Kiviranta, N. L. Pedersen, and A. Ahlbom, A, "Sedentary jobs and colon cancer," *American Journal of Epidemiology, vol.* 123, pp. 775-780, 1986.

[6] N. Gyorbiro, A. Fabian, and G. Homanyi, "An activity recognition system for mobile phones," in *Mobile Networks and Applications*, vol. 14, no. 1, pp. 82-91, 2008.

[7] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SIGKDD Explorations*, vol. 12, no. 2, pp. 74–82, December 2010.

[8] J. W. Lockhart, T. Pulickal, and G. M. Weiss, "Applications of mobile activity recognition," in *Proc. of the 2012 ACM UbiComp International Workshop on Situation, Activity, and Goal Awareness*, Pittsburgh, PA.

[9] J. W. Lockhart, and G. M. Weiss, "The benefits of personalized models for smartphone-based activity recognition," in *Proc, of the 2014 SIAM International Conference on Data Mining*, Philadelphia, PA, pp. 614–622.

[10] J. E. Manson, P. L. Skerrett, and W. C. Willett, "Epidemiology of health risks associated with obesity", in *Eating disorders and Obesity*, C.G Fairburn, and k\K.D. Brownell, Ed., New York, Guilford Publications, 2001.

[11] U. Maurer, A. Smailagic, D. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in 2006 *IEEE Proc. on the International Workshop on Wearable and Implantable Sensor Networks*, 3(5).

[12] Physical inactivity a leading cause of disease and disability, warns WHO. World Health Organization. [http://www.who.int/mediacentre/news/releases/release23/en/index.html], 2002.

[13] I. H. Witten, and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed., Morgan Kaufmann, 2005.

[14] J. Yang, "Toward physical activity diary: Motion recognition using simple acceleration features with mobile phones," In *First Int. Workshop on Interactive Multimedia for Consumer Electronics at ACM Multimedia*, pp. 1-10, 2009.