Physical Activity Recognition in Free-living from Body-worn Sensors

Katherine Ellis†
kellis@ucsd.edu

Suneeta Godbole‡
sgodbole@ucsd.edu

Jacqueline Chen‡
jac018@ucsd.edu

Simon Marshall‡
sjmarshall@ucsd.edu

Gert Lanckriet†
gert@ece.ucsd.edu

Jacqueline Kerr‡
jkerr@ucsd.edu

†Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92023-0407

‡Department of Family and Preventive Medicine
University of California, San Diego
La Jolla, CA 92093-0628

ABSTRACT

Machine learning techniques are used to improve accelerometer-based measures of physical activity. Most studies have used laboratory-collected data to develop algorithms to classify behaviors, but studies of free-living activity are needed to improve the ecological validity of these methods. With this aim, we collected a novel free-living dataset that uses SenseCams to obtain ground-truth annotations of physical activities. We trained a classifier on free-living data and compare it to a classifier trained on prescribed activities. The classifier predicts five activity classes: bicycling, riding in a vehicle, sitting, standing, and walking/running. When testing on free-living data, classifiers trained on free-living data significantly outperform those trained on a controlled dataset (89.2% vs. 70.9% accuracy).

Author Keywords

Activity Recognition; Free-living; Accelerometer; GPS; SenseCam; Wearable Cameras

ACM Classification Keywords

I.5.4 Pattern Recognition: Applications

INTRODUCTION

Recently in the public health community there has been rising interest in using machine learning to improve objective measurement of physical activity from body-worn sensor data. Researchers use supervised machine learning techniques to train classifiers that map patterns of sensor signals to activity types [1, 3, 4, 6].

Machine learning techniques require a training dataset, *i.e.*, known examples of sensor data that correspond to the activities one is interested in predicting. Most prior studies have used laboratory simulations of behavior to train classifiers, with the hope of applying these algorithms to free-living populations (*i.e.*, people going about their normal daily life). However, researchers have noted that classifiers that show high performance on very controlled datasets often do not translate to high performance in free-living. For example, Gyllensten and Bonomi [2] observed that while cross-validation of classifiers on a lab dataset achieved high accuracy, the accuracy was significantly lower when these trained classifiers were applied to free-living data.

Bicycling Vehicle

Figure 1. Examples of SenseCam images and annotations

Few previous studies have used free-living data for training activity classifiers, in part because it is difficult to obtain reliable ground truth annotations describing what activities the participants were doing at a particular time, which are essential for the training stage of machine learning algorithms. Full-time direct observation is expensive and infeasible for large populations [5]. However, an option that allows for accurate time-stamped annotations at relatively low cost is equipping participants with wearable cameras.

In this study, we collected a free-living dataset in which participants wore the SenseCam along with accelerometer and GPS sensors for 3-4 days, going about their normal lives. Researchers then used the SenseCam images to annotate data with a timestamped activity label. The annotated data was used to train a machine learning classifier to predict five different activities from accelerometer and GPS data. We compared the performance of this classifier trained on free-living data to a classifier trained on another dataset that was collected in controlled conditions.

DATA COLLECTION

We collected two datasets with comparable outcomes. One dataset was collected in controlled conditions; the other was collected in free-living conditions. Both datasets used a hip-worn Actigraph GT3X+ accelerometer and Qstarz BT1000X GPS device. Accelerometer data was collected along three axes at 30 Hz and GPS data was collected every 15 seconds.

SenseCam 2013 Extended Abstracts, San Diego, USA.
ACM 978-1-4503-2247-8.

Train Dataset	Test Dataset	Algorithm	Sitting	Standing	Walking	Bicycling	Vehicle	**Average**
Controlled	Controlled	RF	88.5	91.4	97.0	96.8	96.2	**94.0**
	Free-living	RF	76.2	87.9	81.6	78.6	81.9	**70.9**
Free-living	Free-living	RF	95.7	91.6	91.2	83.0	86.7	**89.2**
		RF + HMM	97.0	93.7	93.9	87.9	87.8	**91.3**

Table 1. Percent accuracy observed in each experiment. Experiments tested on the controlled dataset use leave-one-day-out cross-validation, experiments tested on the free-living dataset use leave-one-user-out cross-validation. Accuracy for each activity label is reported as normalized accuracy, *i.e.*, the average of the true positive rate and the true negative rate.

Data was labeled with five activities: sitting, standing, walking/running, bicycling, and vehicle.

Controlled Dataset
The controlled dataset was collected by two trained research assistants in San Diego performing a variety of prescribed activities, in a series of over 500 prescribed trips.

Free-Living Dataset
We recruited 40 adult cyclists through a university based cycle-to-work network. Eligible participants were aged 18 - 70 years, were university employees and routinely bicycled for transportation. Participants wore the SenseCam on a lanyard around their neck, and a group of researchers and assistants annotated the SenseCam images with activity labels according to a protocol. The full annotation protocol is available from the first author on request.

DATA PROCESSING
We used a sliding window of width 1-minute and step size 30 seconds to process the sensor data. From each window we extracted a 47-dimensional feature vector consisting of 41 acceleration features and 6 GPS features.

MACHINE LEARNING ALGORITHMS
Our activity classification method is based on a random forest (RF) classifier. The random forest classifier takes as input a feature vector corresponding to one minute of data, and outputs a predicted activity label. On the free-living dataset, we added an additional layer to the system — a Hidden Markov model (HMM) that models the probabilities of transitioning between activities throughout a day and smooths the first-level predictions made by the random forest classifier.

EXPERIMENTS AND RESULTS
We first test the performance of our random forest classifier on the controlled dataset, training the random forest classifier on held-out data from the same dataset. We test the RF classifier with leave-one-day-out cross-validation to prevent training on data immediately adjacent to data in the test set. Next we test the performance of a RF classifier trained on data in the controlled dataset the free-living dataset. We next train and test the RF classifier on the free-living dataset, using leave-one-user-out cross validation. Finally, we add the HMM as a second-layer classifier that is applied after the RF classifier that was learned on free-living data.

Table 1 summarizes the results of these experiments. Accuracy reported is the result of leave-one-user-out cross validation, which aligns with a real-world scenario, in which we would like to apply our learned classifier to previously unseen subjects. In each experiment the RF classifier uses 50 trees, 25 features and 10,000 training examples per tree. These parameters were chosen based on a held-out validation set of data from the free-living study.

DISCUSSION AND CONCLUSIONS
We first notice that our results corroborate those made by previous work [2], that prediction performance is high when classifiers are cross-validated on controlled or lab data, but underperform when applied to free-living data. However, the accuracy significantly improves when the classifier is trained on free-living data. This cross-validated free living accuracy is still lower than the cross-validated controlled dataset accuracy — not surprisingly, as we expect that free-living data is inherently more variable, and in fact even human annotators may not always agree on activity labels. Our results demonstrate that if the goal is to classify human behavior in the wild, it is important to train classifiers on free-living data. The SenseCam proved to be a feasible mechanism for collecting ground-truth activity annotations in free-living.

Addtional Authors
John Staudenmayer, Department of Mathematics and Statistics, University of Massachusets, Amherst, Amherst MA 01003-9305, jstauden@math.umass.edu.

Acknowledgements
The analysis in this paper was supported in part through a collaboration with the PALMS Project (UCSDPalmsProject.wikispaces.com) and TREC project 4 at the University of California, San Diego. Funded by NIH/NCI Grant 1 U01 CA130771 and NCI Grant U53 CA1555435-01.

REFERENCES
1. Bonomi, A. G., Goris, A., Yin, B., and Westerterp, K. R. Detection of type, duration, and intensity of physical activity using an accelerometer. *Med Sci Sports Exerc 41*, 9 (2009), 1770–1777.
2. Gyllensten, I. C., and Bonomi, A. G. Identifying types of physical activity with a single accelerometer: evaluating laboratory-trained algorithms in daily life. *Biomedical Engineering, IEEE Transactions on 58*, 9 (2011), 2656–2663.
3. Pober, D. M., Staudenmayer, J., Raphael, C., Freedson, P. S., et al. Development of novel techniques to classify physical activity mode using accelerometers. *Medicine and science in sports and exercise 38*, 9 (2006), 1626.
4. Preece, S. J., Goulermas, J. Y., Kenney, L. P., and Howard, D. A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *Biomedical Engineering, IEEE Transactions on 56*, 3 (2009), 871–879.
5. Sallis, J., and Saelens, B. Assessment of physical activity by self-report: status, limitations, andfuture direction. *Res Q Exerc Sport 71*, 2 (2000), 114.
6. Staudenmayer, J., Pober, D., Crouter, S., Bassett, D., and Freedson, P. An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of Applied Physiology 107*, 4 (2009), 1300–1307.