

Activity Classification with Smartphone Data

Matt Brown, Trey Deitch, and Lucas O’Conor

Abstract—We present analysis of several methods for classifying activities, such as walking up stairs or standing, using data from a gyroscope and accelerometer. Results are shown for both a processed dataset and raw data. Analysis is informed by a visualization of the data. We analyze the error due to differences between users and present a potential solution to this problem.

1 INTRODUCTION

OVER 60% of the U.S. population owns a smartphone [1], which increasingly include accelerometers and gyroscopes. We hope that a successful activity classifier will enable a new class of user-mobile interactions. For health reasons, some people are interested in tracking the amount of time they spend sitting down or the number of stairs they climb each day. For home automation, walking up stairs or sitting on a couch could trigger the lights or television to turn on. Phones could also react to their environments, for example by switching into “car mode” when they detect that a user is driving or “airplane mode” when on an airplane. On a larger scale, understanding the activity patterns of populations has implications for transportation and urban design.

Previous experiments required too much time and data [2] or used complex dedicated sensors [3–8]. Further, they frequently relied only on accelerometer data, while today many mobile phones contain gyroscopes as well. Therefore, our research focuses on developing a fast classifier that operates on accelerometer and gyroscope data from mobile phones.

In this paper, we present our efforts to develop an effective activity classifier. We first exposit our dataset, which contains both raw and preprocessed versions of the same data. Next, we present several classifiers that operate on the processed data and analyze their effectiveness. We then provide motivation for working with raw rather than processed data, and develop and analyze several classifiers that

operate on the raw data. Finally, we present further analyses of our classifiers, which in turn motivate our suggestion for future work.

2 DATASET

We drew our data from a prepared dataset [9], which includes labeled data collected from 30 subjects who engaged in six different activities—walking on flat ground and up and down stairs, sitting, standing, and lying down—all while wearing a smartphone that recorded accelerometer and gyroscope readings at a rate of 50 hz. One portion of the dataset contains the raw gyroscope and accelerometer readings, which we call the “raw data.” The other portion consists of vectors that each contain 561 features and represent 2.56 seconds of time. Each vector encodes characteristics such as the triaxial average, maximum, and minimum acceleration and angular velocity over the given interval, as well as more complex properties such as the Fourier transform and autoregressive coefficients. We call this portion of the data set the “processed data” or “preprocessed data.” We used this data to train and test all of our classifiers.

3 APPROACHES

3.1 Preprocessed Data

As a starting point for our project, we trained and tested Naive Bayes and Gaussian Discriminant Analysis classifiers on the preprocessed data. We then applied a Hidden Markov Model

classifier to our GDA classifier, using the activities as the states and the outputs of our GDA as the emissions. Our goal in adding an HMM to our GDA classifier was to capture time-based relationships in the data which the static GDA classifier failed to use, a technique which has been used successfully in the past [10, 11].

3.1.1 Results

Algorithm	Accuracy
Naive Bayes	80%
GDA	96%
GDA + HMM	98%

3.1.2 Analysis

The accuracy of the GDA model was much higher than that of the Naive Bayes model because Naive Bayes makes the assumption that the features are independent of one another, which is far from the truth in the case of our data. For example, average acceleration is dependent on minimum and maximum acceleration for a time window, but our data lists all of these properties separately. Therefore Naive Bayes is a bad fit for our data set. The suitability of GDA to the preprocessed data can be inferred from the PCA projection in Figure 2, in which there are clearly visible clusters corresponding to each activity. Adding the HMM improved performance further, which is not surprising given the heavily time-dependent nature of our data.

3.1.3 PCA

In order to better understand the processed data set, we performed principle component analysis. Figure 1 shows the training and testing accuracy of GDA as a function of the dimensionality of the feature space. Note that many dimensions can be removed without drastically affecting performance. However, these dimensions are linear combinations of the original features, so this analysis is not particularly useful for reducing the number of features used in the model. Because the data only varies greatly in a few dimensions, we can use PCA to project the data onto the plane of maximal variance and visualize the data. Figure 2 shows this projection, with activities distinguished by color. Note that the activities

are fairly well-grouped, even in two dimensions. This helps explain why GDA works so well on the processed data.

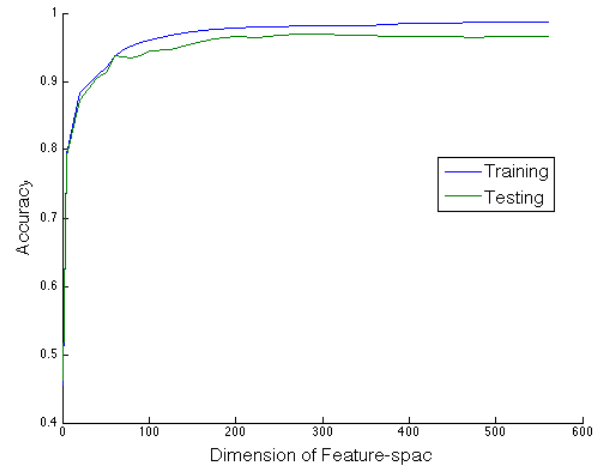


Fig. 1. Accuracy vs feature-space dimension

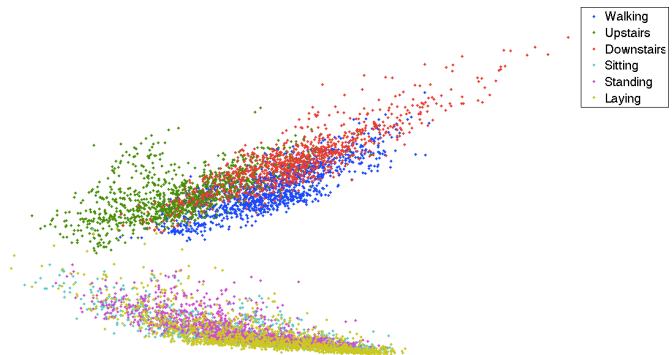


Fig. 2. Processed data projected onto plane

3.2 Raw Data

The next problem we looked at was classifying the raw data. While we achieved very good performance using the processed data, there are a few advantages to using raw data. Each processed data vector is created using 2.56 seconds of raw data, which means the maximum precision of the classifier is limited to one estimate per 2.56 seconds. Additionally, some of the features in the preprocessed data might be difficult to generate in real-world settings, especially given the memory, processor, and battery constraints of multitasking mobile

devices. In such cases, a classifier based only on raw data would be a good alternative.

To establish a baseline, we first applied the same GDA model that we used for the preprocessed data to the raw data. We then decided to apply a Hidden Markov Model approach where our states were the activities that the subject was performing at a given time step, using the Baum-Welch algorithm to estimate the probabilities of transitioning between states and the probabilities of emissions at each state, and the Viterbi algorithm to predict the series of underlying states for a given test sequence of emissions. The observables in our data were the continuous raw data readings, but we used a discrete emissions model HMM for simplicity. In order to convert our data points into discrete values we tried three approaches:

- 1) Using the GDA classification of the data point as the emission.
- 2) Finding the centroids of the data points corresponding to each activity in the training set and using the nearest centroid to each data point as our emission.
- 3) running k-means with different numbers of clusters on the entire training set and using the nearest cluster centroid to each data point as the emission.

3.2.1 Raw Data Results

Algorithm	Accuracy
GDA	52%
HMM + GDA	54%
HMM + Activity Centroids	38%
HMM + K-means (75 clusters)	87%

3.2.2 Analysis

The GDA approach did poorly because its method of fitting a single Gaussian to each activity was unsuited to the raw data set. This is visible in the PCA projection in Figure 4, which shows that the data points for each example were either evenly distributed in space or arranged into multiple clusters. There was a significant amount of overlap for the distributions of data points for each user and activity, making the prediction task extremely difficult.

The GDA approach was also weak because it only looked at a single time step rather than

changes over time. Time-dependent information is very valuable in the multiple-activity context (consider how a transition from standing to walking is much more likely than one from laying down to walking). The HMM approach allowed us to take into account changes in the readings over time by treating the data points as a sequence rather than just individual examples and estimating the transition probabilities between the hidden states.

HMM + GDA and HMM + Activity Centroids did much worse than HMM + K-means because the former two involve fitting a single cluster or distribution in order to characterize an emission while the latter involves using multiple clusters to characterize the emissions. As can be seen in Figure 7, a given activity can have data points clustered around several distinct centroids, so using K-means clustering and allowing multiple centroids per activity is better suited to the nature of the data than using one centroid per activity. The accuracy of HMM+ K-means improved as we increased the number of clusters until we encountered a memory error after 75 clusters.

To further analyze the HMM + K-means model, we computed the testing and training errors as a function of training set size. Figure 3 shows little gap between training accuracy and testing accuracy, both of which are lower than the baseline accuracy established by running HMM+GDA on the processed data. This indicates that our classifier has high bias [12] and confirms our suspicion that moving from the processed data to the raw data reduces the complexity of the model, which isn't surprising since the processed data has 561 dimensions and the raw data has only 9.

3.2.3 PCA

We performed principle component analysis on the raw data to provide insight into its structure. Because the raw data comes from independent sensors, we expected the variation in all dimensions to be fairly significant, and indeed that is the case. Shown in Figure 4 is a two dimensional projection which is useful for understanding the general nature of the data, but there may be structure lost in the projection. There are several visible clusters of



Fig. 3. Training and testing error vs. size of training set

points corresponding to activities, namely lying, sitting, and standing. The cluster of standing points appears in the middle of a cloud of walking upstairs, downstairs, and walking. This seems consistent with an intuitive view that standing is somehow “in the middle of” walking, both of which are fairly distinct from lying and sitting.

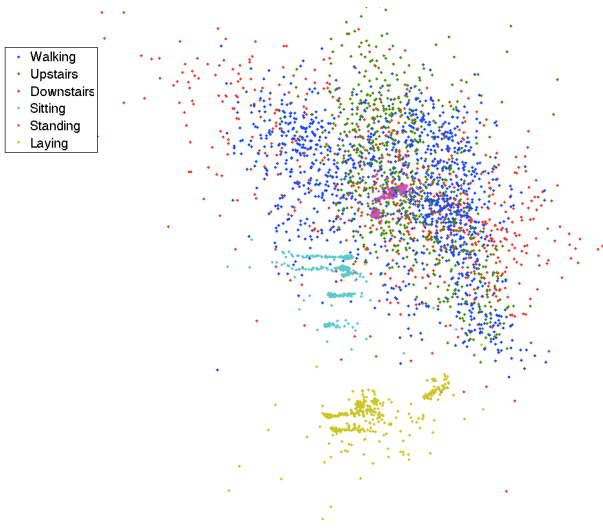


Fig. 4. Raw data projected onto plane

4 NEW USER CROSS-VALIDATION

If any of these classifiers are to be used in practice, it is very likely they will be used by someone on whom the classifier was not trained. Therefore, we are interested in the testing error where the test set consists of data from a

single subject and the training set consists of data from other subjects. We performed this “new user cross-validation” on the HMM+k-means classifier (our most successful raw data algorithm). As shown in Figure 6, the average cross validation accuracy was relatively low (75%) and the variation in error between users was high. This tells us that it may be necessary to incorporate each user’s individual data into the model to achieve high accuracy. In contrast, Figure 5 shows the results of running new user cross validation on the processed data. Here, the cross validation errors are lower and are more consistent from user to user. We can conclude that the processed data forms a more general model for each of the activities that new users will tend to fit.

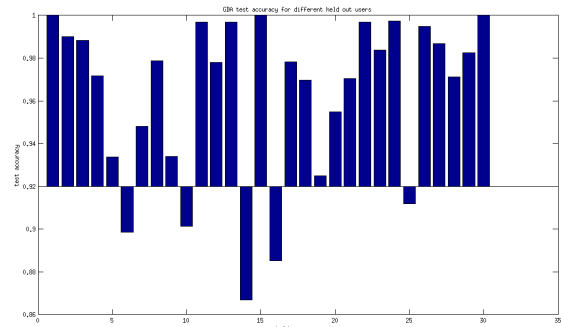


Fig. 5. Accuracy of GDA on preprocessed data for each new user

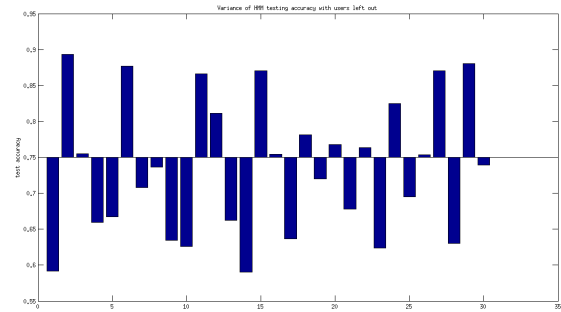


Fig. 6. Accuracy of HMM with k-means on raw data for each new user

5 FUTURE WORK

5.1 Real Time K-Means Adaptation

To improve accuracy by adapting the model to individual users after training, we propose a

modification to k-means as the emission model in an HMM. Previously, we used k-means to cluster the training data and recorded the cluster centroids. Then we treated the closest centroid to each data point as an emission of an HMM. We propose that once the algorithm is deployed, the cluster centroids should be moved slightly in the direction of each new point to which they are closest.

For every new data point x and cluster centroid c_j with weight w_j

$$c_j^{new} = w_j \cdot c_j^{old} + \alpha \cdot x$$

$$w_j^{new} = w_j^{old} + \alpha$$

This method was motivated after inspecting PCA projections of multiple users, and noticing similar but distinct clusters for a given activity. Figure 7 shows data taken for two different subjects while sitting. Note the similar, but approximately translated groups of points. If the cluster centroids were trained using the green subject and then run on the blue subject, this algorithm would slowly move the centroids toward the centroids of the blue clusters, creating a better fit for that subject.

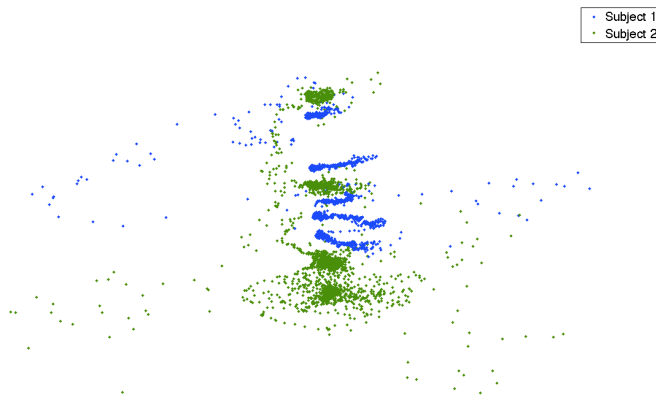


Fig. 7. *Sitting data for two different users*

6 CONCLUSION

We created a classifier that is very accurate on processed data and relatively insensitive to new users. However, the feature vectors are computationally expensive to generate and only give an activity estimate every 2.56 seconds. We transitioned to raw data to make

a faster, more efficient classifier. A hidden Markov model proved useful in capturing information in the transitions between activities and the time history of the estimate. K-means clustering worked well as the emission model for the HMM because it allowed multiple clusters per activity. More work is needed so that this method generalizes well to new users, and shifting the centroids to better fit each user is one promising way to achieve that.

REFERENCES

- [1] *Mobile Majority: U.S. Smartphone Ownership Tops 60%*. Nielsen, June 6 2013, <http://www.nielsen.com/us/en/newswire/2013/mobile-majority--u-s--smartphone-ownership-tops-60-.html>.
- [2] Kwapisz, Jennifer R., Gary M. Weiss, and Samuel A. Moore. "Activity recognition using cell phone accelerometers". *SIGKDD Explorations*: 12 (2011), 74-82.
- [3] Bao, Ling and Stephen S. Intille. *Activity recognition from user-annotated acceleration data*, 2004.
- [4] Casale, Pierluigi, Oriol Pujol, and Petia Radeva. *Human Activity Recognition from Accelerometer Data Using a Wearable Device*, 2011.
- [5] Cho, Yongwon, Yunyoung Nam, Yoo-Joo Choi, and We-Duke Cho. "SmartBuckle: Human Activity Recognition Using a 3-axis Accelerometer and a Wearable Camera." *Proceedings of the 2nd International Workshop on Systems and Networking Support for Health Care and Assisted Living Environments*, 2008.
- [6] Lee, Seon-Woo and Kenji Mase. "Activity and Location Recognition Using Wearable Sensors." *Pervasive Computing*: 2 (2005), 24-32.
- [7] Pärkkä, Juha, Miikka Ermes, Panu Korpipää, Jani Mäntyjärvi, Johannes Peltola, and Ilkka Korhonen. *Activity Classification Using Realistic Data From Wearable Sensors*, 2006.
- [8] Ravi, Nishkam, Nikhil Dandekar, Preetham Mysore, and Michael L. Littman. "Activity recognition from accelerometer data." *In Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence (IAAI)*, 2005.
- [9] Anguita, Davide, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L. Reyes-Ortiz. *Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine*. Vitoria-Gasteiz, Spain: International Workshop of Ambient Assisted Living (IWAAL 2012), Dec 2012.
- [10] Lester, Jonathan, Tanzeem Choudhury, Nicky Kern, Gaetano Borriello, and Blake Hannaford. "A hybrid discriminative/generative approach for modeling human activities." *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2005.
- [11] Yin, Pei, Irfan Essa, Thad Starner, and James M. Rehg. *Discriminative Feature Selection for Hidden Markov Models Using Segmental Boosting*, 2004.
- [12] Ng, Andrew. *Advice for applying Machine Learning*, <http://cs229.stanford.edu/materials/ML-advice.pdf>.