Capstone Proposal

Lucas Lofaro

November 7, 2017

1 Domain Background

What if we could use our phones or smart watches to track our fitness regiment and suggest improvements and optimizations to our daily routines? Many modern devices are beginning to implement these features, but there is already so much more we could be doing. At this point, almost everyone walks around with devices equipped with a accelerometers and gyroscopes. We carry these devices around all day long collecting data that could provide key insights into our behaviour and give us real-time feedback about our health. That data could be used to identify the activities that make up our daily routine, keeping track of how long we spend sitting at a computer, walking and talking with a friend, or playing our favourite sport. Ultimately, our health and well-being could improve dramatically if our devices were just a little more conscious of how we spend our day.

2 Problem Statement

Before we can do the fancy things—like make a recommendation for an exercise routine or plan a meal designed to perfectly compensate for our caloric loss—we need to address a more fundamental question. Specifically, our goal is to translate raw time series data into a recognizable human activity. Our data should consist of streaming data collected from a sensor attached to an individual performing a particular activity. Features can then be extracted from that data based on patterns within the data itself or statistical indicators such as mean and variance.

3 Datasets and Input

The data used in this project comes from the UCI Machine Learning Repository [1]. The data set consists of fifteen participants, each of whom performed seven different activities:

1. Working at Computer
2. Standing Up, Walking and Going up/down stairs
3. Standing
4. Walking

	x	y	z	activity
0.0	1984	1998	1645	1
1.0	1984	2003	1653	1
2.0	1987	2002	1648	1
3.0	1990	2001	1650	1
4.0	1992	2000	1657	1

Figure 1: First few entries in the dataset

5. Going Up/Down Stairs
6. Walking and Talking with Someone
7. Talking while Standing

The data is formatted as a time series where each sample provides the x, y, and z components of acceleration, as well as the corresponding activity label. Each participant performed all seven activities for a total of 105 participant and activity combinations. Of those pairs, 84 will be used for training while the remaining 21 will be used for testing.

This data is ideal for this project because it aims to identify everyday activities. Our goal is to determine what someone is doing at any given time simply based on a series of consecutive samples taken from sensors on the person's body. In this case, a simple accelerometer was mounted to the participant's chest and recorded a continuous stream of data while the participant performed the activity. In other words, the participants went about their normal activities as they would on any other day, simulating a real-world environment where these activities would normally take place. If we are able to gleam enough information from the accelerometer data, it will serve as proof of concept and allow further exploration using more readily available sensors—like those in modern phones or smart watches.

4 Solution Statement

In order to solve this problem, we need to look for trends in the time series data that will serve as features in a classification algorithm. We want to extract useful patters that characterise the type of motion associated with a given activity. Much like a fingerprint, each activity has distinctive qualities that uniquely identify it. Certain activities will be rhythmic in nature (e.g. walking or climbing up and down stairs) while others will be characteristically arrhythmic (e.g. talking). In some cases we might also expect the magnitude of one component to be larger than another (at least on average). The presence—or absence—of any of these traits will give clues to the identity of an unlabelled activity.

After the feature engineering phase, we need to classify each activity based on its similarity to other (known) examples. This particular set-up lends itself well to the canonical K-means approach since we would expect the same activity to present similarly when performed by all people. In addition, there could be some empirical guidelines that we use to quickly classify different activities. Implementing something like a decision tree would allow us to see those logical steps and get a better sense of the traits that best characterize different activities. After testing a few possible models, the one that performs best according to our evaluation metrics will ultimately be selected.

5 Benchmark Model

There is a lot of fascinating research going on in this field, particularly in the area of on-body accelerometers. While the implementation of these models can differ significantly from the solution outlined here, the conceptual model is nearly identical (see Figure 2). The benchmark model from [3]

Figure 2: Conceptual scheme of a generic classification system with supervised learning [3]

identifies seven key activities, though many differ from those in our datasets. Ultimately they achieved extremely high classification accuracy—over 99% after rejecting spurious data. Overall, the biggest difference is that the benchmark model trains classifiers for each individual in an attempt to capture unique mannerisms. Since our goal is to detect the generic motion involved in certain tasks, that level of modelling is not required here. If nothing else, this benchmark model demonstrates that these techniques can be used to effectively classify human activities from nothing more than a body-mounted accelerometer.

6 Evaluation Metrics

Classification accuracy will serve as the primary evaluation metric for this project. This is a standard metric for classification problems and will allow for comparison of results with the benchmark model. In addition, since our gaol is to distinguish different forms of activity, correct classification is a chief concern. Classification accuracy (a_c) is defined as the ratio of correctly classified samples (n) to the sample size (n_0).

$$a_c = \frac{n}{n_0}$$

However, classification accuracy may not be the most appropriate metric because some of these activities consist of combinations of other activities (i.e. walking and talking with someone—activity 6—involves walking—activity 4). For that reason, false positives may be more likely to occur, and therefore an adjusted f-score might serve as a better metric. Another possibility is the Fowlkes Mallows Index [2], which is the geometric mean of precision and recall.

$$FM = \sqrt{\frac{TP}{TP+FP} \cdot \frac{TP}{TP+FN}}$$

7 Project Design

Feature Extraction from Time Series

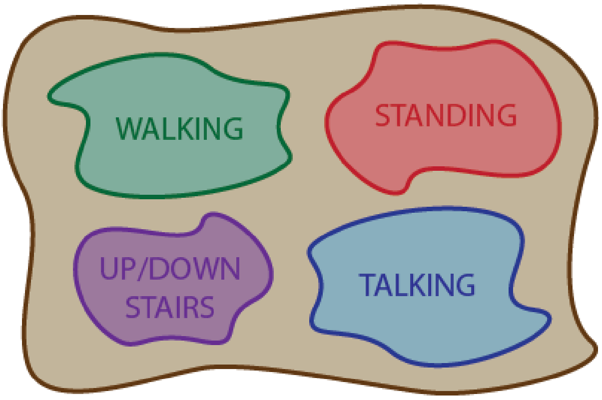
The data is provided as incremental measurements of the x, y, and z components of acceleration. This data cannot be input into our classification algorithm directly, which means we must derive a feature vector from the data. For the purposes of this project we will consider a four-dimensional feature vector. This vector will contain the mean and standard deviation of the data as well both the amplitude and phase of the dominant frequency.

$$\vec{v} = \begin{bmatrix} \mu \\ \sigma \\ f_0 \\ \phi \end{bmatrix}$$

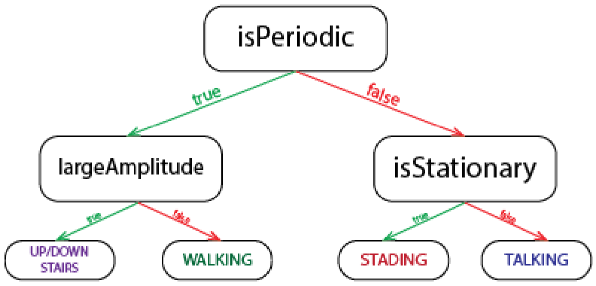
Classification Algorithm

One we have extracted the necessary feature vectors, we can input them into a classification algorithm to group them together based upon similarity. Of the possible choices we will consider the k-means algorithm as well as decision trees. We expect that each activity will posses a unique acceleration pattern that will remain consistent amongst all participants. If this assumption is true, then k-means should be able to group the data points appropriately. However, because we are only using a four-dimensional feature vector, we may see unexpected groupings.

Intuitively, we would imagine that there are some rules of thumb that human might use to distinguish these activities. Decision trees could show that dividing the data according to a number of conditions is a more effective approach. Not only that, but if the algorithm proves successful we could actually see what criteria the algorithm used to separate the data.



(a) Representation of clustering different activities.



(b) Simple decision tree model of classifying activities.

Evaluation

Finally, we will evaluate the performance of the algorithms on the data. We will use classification accuracy to determine how well the algorithm predicts the correct activity from the derived feature vectors. We will compare the accuracy of each model to determine which algorithm performed better on the test set.

References

[1] O. Casale, P. Pujol and P. Radeva. Personalization and user verification in wearable systems using biometric walking patterns. *Personal and Ubiquitous Computing*, 2012.

[2] E. B. Fowlkes and C. L. Mallows. Machine learning methods for automatic classification of human physical activity from on-body accelerometers. *Journal of the American Statistical Association*, 1983.

[3] A. Mannini and A. M. Sabatini. Machine learning methods for automatic classification of human physical activity from on-body accelerometers. *sensors*, 2010.