Computational Physiology — Papers from the AAAI 2011 Spring Symposium (SS-11-04)

Activity Recognition with Time-Delay Embeddings

Jordan Frank
McGill University, Montreal, Canada

Shie Mannor
Technion, Israel

Doina Precup
McGill University, Montreal, Canada

Abstract

We outline an approach that uses time-delay embedding models and machine learning in order to recognize physical activities, based on data coming from cell phone accelerometers. The approach is very robust, cheap in terms of the amount of data and computation required, and easy to deploy.

As the world's population ages, automated monitoring of physical activities is becoming an increasingly important task. Applications range from the detection of potential problems (such as an elderly person who has fallen down in their home) to general monitoring of disease progression (e.g. in Parkinson's disease), or simply tracking the amount of exercise and physical activity that a person gets. Ideally, such activities should be monitored as precisely as possible, but using cheap or easily available devices, and in a way that does not interfere with daily life.

In this extended abstract we outline our approach to this problem, which relies on machine learning to build automated activity recognition systems. The data we use is acquired from accelerometers embedded in standard smart phones; the monitoring system is just another application, which makes it very easy to deploy and use in the general population.

Time-Delay Embeddings for Feature Construction

The main building block in our approach is the use of *time-delay embedding* models for time series (e.g. Kantz & Schreiber, 2004). We assume measurements coming in a univariate time series, which is generated from a set of non-linear, periodic dynamical systems. The underlying system may consist of unobserved latent variables, governed by complex laws; for example, in the case of human motion, the measured acceleration is the result of the non-linear dynamics of interacting muscles and joints. Each activity (walking, running, skating etc.) generates different dynamics, which are reflected in observed acceleration magnitude. However, in order to classify the activity, it is not necessary to identify and model all hidden variables; instead, one can use a data-driven approach, in which models are only based on observable quantities. Note that similar issues arise in modeling other types of physiological data, e.g. EEG (which captures brain activity) and ECG (measuring the electrical activity of the heart). Indeed, time-delay embeddings have proven useful for classifying both types of signals.

Specifically, given a sequence of observations, at time-step t, the system is represented through a vector $\langle o_t, o_{t+\tau}, \ldots o_{t+m\tau}\rangle$ where τ is a delay parameter and m is the chosen dimension. A sequence of observations $o_1, \ldots o_T$ will then be represented by a collection of such vectors for all times $t = 1, \ldots T - (m-1)\tau$. We refer to such a sequence as a *model* of the system. Note that these models are non-parametric. Theoretically, under some smoothness assumptions (Takens, 1981), if m is big enough, and τ is not a multiple of the period of the system, such a model captures all the relevant dynamics. However, real data is noisy, so non-parametric models of the same activity can have high variability. Hence, it is beneficial to extract several models from data segments of a given activity. We typically use a small number (5-10) of segments picked at random from a given activity in order to construct models.

When new data is acquired, in order to label it, we need to compute its similarity with the existing activity models. To do this, we generate a time-delay reconstruction of the data, using the same parameters (m and τ) that were used to construct the models. Afterwards, the similarity of the data to the existing models can be computed using the *geometric template matching* algorithm (Frank et al., 2010a). Briefly, the algorithm computes a normalized cosine similarity between the data reconstruction and the model.

The similarity scores can then be used as input to other machine learning algorithms. For example, labeling the activity for new data can be achieved by nearest-neighbor classification based on the model similarity scores. More generally, the similarity scores can be provided as inputs for any supervised learning algorithm. In our work so far we have experimented with support vector machines, random forests and boosting as classifiers, with similar results.

If no labelled data for activities is available, similarity scores between different segments of data can be used as inputs to a clustering algorithm, in order to identify which segments may be grouped together, and which correspond to distinct activities. Hierarchical clustering is particularly well suited for this task (Frank et al., 2010b), because it can determine both the number of clusters, and the partitioning of the data segments into clusters.

Empirical Results

We implemented our approach on the Google Android platform, and built an application based on it for the Nexus One mobile phone (Frank et al., 2010a, 2010b). We have experimented with several data sets from the phone, including walking data from 40 different subjects (in which the goal is to identify the subject), walking data gathered from 25 different subjects on two different days (in which again the goal is to identify the subject) and exercise data including 4 subjects and 6 activities (in which the goal is to perform exploratory data analysis and see how well both the subjects and the activities are separated).

Additionally, for activity recognition, we used a large, noisy data set comprising over 50 hours of data from six different subjects, including activities such as running, walking, and going up or down stairs (Subramanya et al., 2006). We focus on this data set here, as it is the most challenging. The data set was gathered using the Intel Mobile Sensing Platform (Lester et al., 2006). For the experiment, we split the data set into six parts, each containing the data from a specific participant. The accelerometer data is sampled at 512Hz, which we decimate to 32Hz. The accelerometer data consists of three measurements at each time step, corresponding to the acceleration along each of the three axes, x, y, and z. We combine these three measurements to form a single measure of the magnitude of the acceleration vector $a = \sqrt{x^2 + y^2 + z^2} - g$, where $g = 9.8m/s^2$ is the Earth's gravity. Subtracting g causes the acceleration when the device is at rest or moving at a constant velocity to be 0. We projected all of the accelerometer data into a time-delay reconstruction space with parameters $\tau = 3$ and $m = 16$. For each user, we constructed a training set by selecting randomly 25% of these embedded data points. This corresponds to approximately 2 hours of data, or 230,000 samples for each participant. For this data set, we also used the gradient of the barometric pressure in the classification task, as it is necessary to separate the activities of walking up or down the stairs.

The classifier performs well across all users, regardless of the user on which it was trained. The average accuracy for the experiments using the time-delay embedding on the entire data set is 85.5$\pm$0.30% (see Frank et al., 2010a, for details). This result demonstrates that these features are useful for activity recognition devices, because the system can be calibrated on one user, then deployed to other users, and the performance is very similar. Note that some of the misclassifications are very short segments in which the class label briefly flips, which could be removed by further post-processing. Using the raw data yields much worse results, with an accuracy of 77.1%$\pm$0.3; this supports the intuition that time-delay embedding features are particularly well-suited for this type of data. These results were generated using support vector machine classifiers, but using random forests yields very similar accuracy.

Our approach also compares favorably with that of Lester et al. (2006), which computes 650 features of the time series, composed of cepstral coefficients, FFT frequency coefficients, spectral entropy, band-pass filter coefficients, correlations, and a number of other features that require a nontrivial amount of computation. In their work, a modified version of AdaBoost is used to select the top 50 features for classification, but other classifiers could be used as well. Preliminary comparisons show their approach has 3%-5% lower accuracy; but what is most important is that our approach is comparatively "light-weight", computing much fewer features and requiring a lot less memory and battery power. This is important, as we do not want this application to interfere with the normal use of the phone. Also, our approach can provide a class label very quickly, after roughly 2 seconds of observed data, and has very short lags in the correct identification when the activity changes. This is important in the context of real-time health monitoring.

Conclusions and Future Work

Our approach is well suited for real-time applications on low-powered portable devices, such as off-the-shelf smartphones. The model quality is robust to noise in the data, as well as to different parameter settings. The results are also consistent when using different supervised learning algorithms. We are currently evaluating more in-depth a semi-supervised approach, where small amounts of labelled data are used to seed a classifier. We are also examining active learning approaches for identifying new activities. Specifically, if similarity scores for new data against all existing models are low, the user is assumed to be performing a "new" activity and we would like to prompt them for a label. However, this part of the project requires addressing some interface design issues as well.

Acknowledgements

This work has been supported by NSERC. We thank Dieter Fox for providing access to his activity recognition data, and Tanzeem Choudhury for useful discussions.

References

Kantz, H., and Schreiber, T. 2004. *Nonlinear time series analysis*. Cambridge University Press.

Frank, J.; Mannor, S.; and Precup, D. 2010a. Activity and Gait Recognition with Time-Delay Embeddings. In *Proceedings of AAAI*.

Frank, J.; Mannor, S.; and Precup, D. 2010b. A novel similarity measure for time series data with applications to gait and activity recognition. In *UbiComp (Adjunct Papers)*

Lester, J.; Choudhury, T.; and Borriello, G. 2006. A practical approach to recognizing physical activities. *Lecture Notes in Computer Science* 3968:1–16.

Subramanya, A.; Raj, A.; Bilmes, J.; and Fox, D. 2006. Recognizing activities and spatial context using wearable sensors. In *Proceedings of Uncertainty in Artificial Intelligence*.

Takens, F. 1981. Detecting strange attractors in turbulence. *Dynamical systems and turbulence* 898(1):365–381.