Learning state space models from time series data

Jordan Frank JORDAN.FRANK@CS.MCGILL.CA
Reasoning and Learning Lab, McGill University, Montréal, Québec, Canada

1. Overview

We consider the problem of reconstructing a state space and state transition model from a set of observations given in the form of a time series. Our approach builds a non-parametric generative model from data that can be used to both ground the state of an agent and make predictions about future states. We demonstrate our method by constructing models of human activities from data collected by an accelerometer mounted in a mobile device and showing that activities can be accurately classified by training an SVM on these state space models. Previous methods for activity recognition use hundreds or thousands of complex features of the data, whereas our method only requires a small number of features corresponding to samples from the raw signal.

2. Time-delay embeddings

The proposed approach is based on the method of time-delay embedding. We assume that the underlying hidden state of the system at time t is a point $x_t \in \mathbb{R}^k$ which lies on an attractor $A \subset \mathbb{R}^k$ in an unknown phase space of dimension k. We are given access to a time series $\{s_t = s(x_t)\}$ generated by a measurement function $s : \mathbb{R}^k \to \mathbb{R}$ which is a smooth map from the phase space to a scalar value. Throughout this abstract, we use smooth to mean continuously differentiable. If we take a set of m of these scalar measurements to be a point in $\mathbb{R}^m$, then this constitutes a map from $\mathbb{R}^k$ to $\mathbb{R}^m$. We call this space $\mathbb{R}^m$ the *reconstruction space*. An *embedding* is a map from the attractor A into reconstruction space $\mathbb{R}^m$ that is one-to-one and preserves local differential structure. In the language of differential topology, such a map is a local diffeomorphism on A.

The *time-delay reconstruction* in m dimensions with time delay τ is formed by the vectors $\mathbf{s}_t \in \mathbb{R}^m$, given as

$$\mathbf{s}_t = (s_t, s_{t-\tau}, s_{t-2\tau}, \ldots, s_{t-(m-2)\tau}, s_{t-(m-1)\tau}). \quad (1)$$

Takens (1981) showed that if A is a d-dimensional smooth compact manifold, then provided $m > 2d$, $\tau > 0$, and the attractor contains no periodic orbits of length τ or 2τ and only finitely many periodic orbits of length $3\tau, 4\tau, \ldots, m\tau$, then for almost every smooth function s, the map from $\mathbb{R}^k$ to the time-delay reconstruction 1 is an embedding. This tells us that we can find a projection that accurately preserves the dynamics of the underlying dynamical system from a time-delay embedding of the data, provided m is large enough and τ doesn't conflict with any periodic orbits in the system.

Takens' theorem only holds when the measurement function s is deterministic, but in practice all measurements of a system are corrupted by noise, whether it be noise in the measurement apparatus, or simply the error introduced by representing a real value on a finite . When noise is present in our observations, we need to be more careful about our choices of m and τ. In the next section we discuss a method for estimating good values for the time-delay parameters m and τ from the data.

3. Spectral embeddings

From a qualitative standpoint, a good embedding is one that makes optimal use of the reconstruction space. Since an embedding should be a one-to-one map from state space to reconstruction space, a desireable property of an embedding would be that trajectories that pass through different parts of the state space would also remain well separated in the reconstruction space. Therefore we prefer embeddings that spread the data out and thus obtain a maximal separation of the trajectories. One way to approximate how the data is spread out is via Principal Component Analysis (PCA) (Buzug & Pfister, 1992). PCA can be used to reveal the internal structure of the data in a way that best explains the variance in the data. We pick a very large embedding dimension m, and then for a range of values for τ, we compute the time-delay reconstruction of our data, perform PCA, and then inspect the eigenvalues of the covariance matrix. Figure 1 shows an example of such a plot, computed using accelerometer data collected while the subject was riding a stationary bicycle.

By ordering the eigenvalues, we can determine how

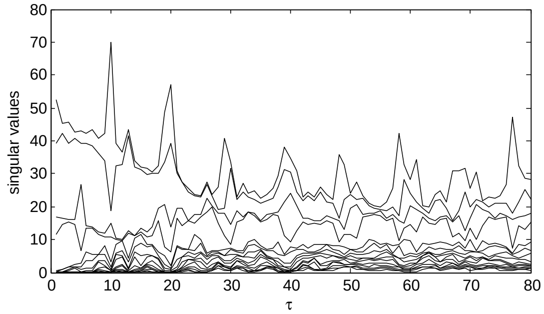

Figure 1. Spectral embedding of biking data with $m = 15$.

much of the variance in the data is accounted for by each of the principal components, and thus through manual inspection, or by setting some threshold, we can pick a number of principal axes that best represent the structure in the data, and assume that the remaining eigenvalues are only accounting for the noise. This provides a method for choosing m. In Figure 1, we see that $m = 4$ is a good choice, as the four largest eigenvalues seem to stand out from the others.

To choose τ, we consider the behaviour of the eigenvalues as we vary τ. We notice that for certain choices of τ, such as $\tau = 10$ in Figure 1, the reconstruction collapses in certain dimensions. To choose τ, it is best to look for local maxima in all of the eigenvalues, though in practice this does not typically occur at any one value of τ. In Figure 1, no value of τ stands out as an obvious choice, although at $\tau = 7$, there is a local maxima in the first eigenvalue, and there are no local minima in the other eigenvalues.

4. Experimental Results

In order to demonstrate one use of these time-delay embedding models, we collected data from the accelerometer built into the HTC Dream© cellular telephone while the device was in the pocket of a subject performing a number of activities. The activies were walking, running, riding a stationary bicycle, and standing still. For each of the activities, we constructed a time-delay embedding from training data using the parameters $\tau = 3$ and $m = 7$. We trained a set of binary support vector machines (SVMs) with radial basis kernels to discriminate between each pair of activities using 10-fold cross-validation. The inputs to the SVMs are the points in the reconstruction space and the class labels are the activities. No tuning was done on the parameters of the SVM kernels. We then had the subject perform a sequence of activities and classified each data point in the test set using the SVMs and a majority voting scheme. The results on the test set are shown in Figure 2 and the classification accuracy is approximately 97.4%.

Previous work in activity recognition achieves similar

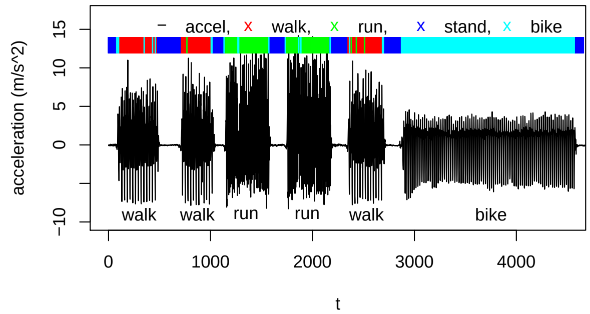

Figure 2. Classification results. The labels below the data show the true activities and the coloured bar above the data shows the labels assigned by the classifier.

classification accuracy, but these approaches typically employ on the order of hundreds of features, most of which require a nontrivial cost to compute (Mahdaviani & Choudhury, 2007). Our method uses 7 features, which consist of samples within a half-second time window, and thus require a minimal cost in both computation and memory requirements. This makes such a solution far more attractive for performing real-time activity recognition on a low-powered mobile device.

5. Concluding Remarks

We present a method for reconstructing the state space and dynamics of some underlying system given only a sequence of scalar measurements. The model that we build preserves the uniqueness of states in the underlying system as well as the differential structure, and thus we can use such a model for prediction and planning. In many applications, constructing a generative model of a nonlinear dynamical system from first principles is difficult. However, if the system is partially observable then our method can learn a nonparametric model that can be used in place of the MDP model typically used in reinforcement learning.

Future research directions include further investigation into the use of these models for reinforcement learning as well as looking at ways to incorporate multivariate time series data. We are also currently validating our methods on a larger data set consisting of more activities.

References

Buzug, T., & Pfister, G. (1992). Optimal delay time and embedding dimension for delay-time coordinates by analysis of the global static and local dynamical behavior of strange attractors. *Phys. Rev. A*, *45*, 7073–7084.

Mahdaviani, M., & Choudhury, T. (2007). Fast and scalable training of semi-supervised CRFs with application to activity recognition. *Proceedings of NIPS 2007*.

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