An Event-Based Framework for Process Inference

Michael Joya
Department of Computing Science
University of Alberta
Edmonton, Canada
joya@cs.ualberta.ca

Abstract

We focus on a class of models used for representing the dynamics between a discrete set of probabilistic events in a continuous-time setting. The proposed framework offers tractable learning and inference procedures and provides compact state representations for processes which exhibit variable delays between events. The approach is applied to a heart sound labeling task that exhibits long-range dependencies on previous events, and in which explicit modeling of the rhythm timings is justifiable by cardiological principles.

Introduction

Consider the problem of identifying patterns of sequential events within multivariate data observed at a very high and possibly nonuniform rate. In this general setting one might include the task of analyzing speech and audio signals, the analysis of sensor data collected at different clock rates, and the simulation of asynchronous events that occur very sporadically. We take the approach that the stream of data is the result of a set of instantaneous events occuring sparsely in the time domain. The goal is to model the process dynamics using observed sequences of data and to explain the observations of sequences with similar dynamics. (E.g.: Given that event of type A has just occured, what is the chance of witnessing event of type B? If such an instance of B occurs, what is the distribution of its occurrence over time?) An important feature of these queries is the need for probabilistic modeling of the interval times between the events. Although the processes in these applications may exhibit stochasticity in the space of their observations, the sequences of discrete events are often significantly less noisy or sometimes completely deterministic. In many situations, it is the omission of certain events in a chain that incurs variability in the process, even though interval times are predictable. We argue that this type of "rhythm" variability is difficult to capture with available methods.

Background

Classical approaches to this problem arise from the Markov models (Rabiner and Juang 1986; Yu 2010; Fine, Singer, and Tishby 1998; Beal, Ghahramani, and Rasmussen 2002). Here, the state variable $\mathbf{s}_t \in \mathbb{S}$ is used to summarize the history of the process prior to t and to provide a context for beliefs about the current state.

Variants of the existing Markov models do address some of the aspects of our problem. For example, hidden Markov models (HMMs) (Rabiner and Juang 1986) can be used to represent the times between states of interest, although there is a difficulty in modeling intervals that are known to span many time steps; the memory requirement for the belief state grows in $O(t)$ where t is the number of steps between the onset of a new state.

A similar approach is the continuous-time Bayes' net framework (Nodelman, Shelton, and Koller 2002), which proposes a state space consisting of discrete-space variables and whose transition distributions factor according to a set of independence relationships to produce a directed graph structure. The major difference between this approach and the one presented here is that the current approach does not presume a modal "state", rather the process is described in terms of instantaneous occurrences and the observations they produce. In lieu of a state variable, our implementation retains a memory of all previous events that are known to predict at least one future event with some positive probability.

An Event-Based Modeling Framework

We propose a novel formulation of the Markovian state. Consider a countable, finite set of instantaneous event classes $\mathbb{E}$ of size $|\mathbb{E}| = N$. Given an event a of type $A \in \mathbb{E}$ at time $T_a = t$, we state a distribution over the future occurences of an event b of type B at time T_b in terms of the tuple $\langle S_{ab}, D_{ab} \rangle$ where S_{ab} is a Bernoulli random variable indicating whether or not event b occurs, and D_{ab} is a positive-valued random variable describing the interval time between event a and event b. Thus, D_{ab} is meaningful iff $S_{ab} = 1$ and is undefined otherwise. We refer to this type of distribution simply as an *occurrence*, and it forms the basis of an alternative approach to composing a DBN. This model produces a conditional distribution of the form:

$$P(\ D_{ab} \mid S_{ab} = 1, T_i\ ) \quad \text{where} \quad T_a = T_a + D_{ab}$$

We define one of the event types as a "start event" that serves as an analogue to the initial state distribution of a Markovian model. Choices for the above distributional model can include any density function defined over the positive real numbers. In our present application, we use the log-normal distribution with parameters $\langle\, \mu_{D_{ij}},\, \sigma_{D_{ij}} \,\rangle$. (However, the option to use a nonparametric density estimate is available, see (Parzen 1962)). The events provide a generational model that produce the observed data $X(t) \in \mathbb{R}^M$ according to the distribution:

$$P(\ X(t) \mid T_a = t, \theta_A)$$

Here, θ_i are the parameters of a linear observational model $X(t{+}k) = \theta_i(k) \cdot \mathbb{I}_{\{T_i = t\}}$ where k ranges over some positive and measurable time span and describes the progression of the event in the observation space.

Given an observed training data sequence $X(t)$ and a set of labeled event instances at sparse points in the time sequence, we can obtain a maximum likelihood estimator for the parameters of the observation model and for the interval times D_{ab} via convex optimization.

Additionally, we can obtain population- and single-sample estimators for the interval times for situations involving multiple sequences of data.

With known parameters $\hat{\theta}_{\text{ML}}$ for the observational model, it is possible to attain a density model for event classes over all points in time. Obtaining precise estimates of the times of event instances can be achieved using dynamic programming approaches.

Application: Heart Sound Identification

Our framework is applied to the task of identifying the two primary heartbeat sounds, known in cardiological literature as $S1$ and $S2$, or coloquially as the "lub-dub" sound. The data consist of 20-second long samples of patient data collected from 25 patients. Only 10 of these patients are labeled with known $S1$ and $S2$ beats. The task is to identify the precise instants at which the $S1$ and $S2$ beats occur in the remaining 15 patients, and to label these points in the sequence.

The waveform signal data for each patient are converted to their *pseudo Wigner-Ville distribution* (Kudriavtsev, Polyshchuk, and Roy 2007) to produce a spectrogram in the joint time and frequency space. The interpolated spectrogram is then used as the multivariate observations $X(t)$.

Using the maximum likelihood estimation procedures described above, we obtain both population and subject parameters for the observational model and for the transition time distributions between the two event classes, i.e.: for $S1 \to S2$ and for $S2 \to S1$.

It is notable that although both the $S1$ and the $S2$ beats produce very similar "signatures" in the observational space, they are relatively easier to identify using interval distributions unique to each patient. These rhythms are represented by subject parameters of the log-normal distribution $\langle\, \mu_{D_{ij}},\, \sigma_{D_{ij}} \,\rangle$ that is used to model the transition times between the beats. For example, it is understood that the complete cardiac cycle event, once initiated, will proceed to completion reliably using a timing mechanism that is intrinsic to the nerve networks surrounding the heart. Within a single patient, the $S1 \to S2$ transition time exhibits a low variance and a low mean time while the alternate $S2 \to S1$ transition time is less predictable.

Rhythmic properties of the problem domain can thus be used to improve inference about the exact timing of $S1/S2$ events in new patients once population parameters for these transitions have been learned. Such properties would have been quite difficult to model explicitly with conventional Markovian models.

Future Work

We would like to extend our framework by showing that short, reliable sequences of events can be treated as a "macro-event" and that these larger events can be treated using the same framework to produce hierarchical models of sequence data. The motivation for this work is to determine whether it is possible to use the distributional parameters of the transition times to aid in the learning of multiscale model structure. We will therefore examine other application domains for which the input sequence exhibits clear rhythmic patterns at several time scales. Such applications may include speech recognition, weather data, stock market data, and the analysis and generation of musical scores.

References

Beal, M.; Ghahramani, Z.; and Rasmussen, C. 2002. The infinite hidden markov model. *Advances in Neural Information Processing Systems* 1:577–584.

Fine, S.; Singer, Y.; and Tishby, N. 1998. The hierarchical hidden markov model: Analysis and applications. *Machine learning* 32(1):41–62.

Kudriavtsev, V.; Polyshchuk, V.; and Roy, D. 2007. Heart energy signature spectrogram for cardiovascular diagnosis. *Biomedical engineering online* 6(1):16.

Nodelman, U.; Shelton, C.; and Koller, D. 2002. Continuous time bayesian networks. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, 378–387. Citeseer.

Parzen, E. 1962. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* 33(3):1065–1076.

Rabiner, L., and Juang, B. 1986. An introduction to hidden markov models. *IEEE ASSp Mag.* 3(1):4–16.

Yu, S. 2010. Hidden semi-markov models. *Artificial Intelligence* 174(2):215–243.