The ability of the BP-AR-HMM to find common behaviors among a collection of time series was demonstrated on data from the CMU motion capture database [34]. As an illustrative example, a set of six exercise routines were examined, where each of these routines used some combination of the following motion categories: running in place, jumping jacks, arm circles, side twists, knee raises, squats, punching, up and down, two variants of toe touches, arch over, and a reach-out stretch. The overall performance of the BP-AR-HMM showed a clear ability to find common motions and provided more accurate movie frame labels than previously considered approaches [35]. Most significantly, the BP-AR-HMM provided a superior ability to discover the shared feature structure, while allowing objects to exhibit unique features.

OUR FOCUS IN THIS ARTICLE HAS BEEN THE ADVANTAGES OF VARIOUS HIERARCHICAL, NONPARAMETRIC BAYESIAN MODELS.

CONCLUSIONS

In this article, we explored a Bayesian nonparametric approach to learning Markov switching processes. This framework requires one to make fewer assumptions about the underlying dynamics, and thereby allows the data to drive the complexity of the inferred model. We began by examining a Bayesian nonparametric HMM, the sticky HDP-HMM, that uses a hierarchical DP prior to regularize an unbounded mode space. We then considered extensions to Markov switching processes with richer, conditionally linear dynamics, including the HDP-AR-HMM and HDP-SLDS. We concluded by considering methods for transferring knowledge among multiple related time series. We argued that a featural representation is more appropriate than a rigid global clustering, as it encourages sharing of behaviors among objects while still allowing sequence-specific variability. In this context, the beta process provides an appealing alternative to the DP.

The models presented herein, while representing a flexible alternative to their parametric counterparts in terms of defining the set of dynamical modes, still maintain a number of limitations. First, the models assume Markovian dynamics with observations on a discrete, evenly-spaced temporal grid. Extensions to semi-Markov formulations and nonuniform grids are interesting directions for future research. Second, there is still the question of which dynamical model is appropriate for a given data set: HMM, AR-HMM, SLDS? The fact that the models are nested (i.e., HMM $\subset$ AR-HMM $\subset$ SLDS) aids in this decision process—choose the simplest formulation that does not egregiously break the model assumptions. For example, the honey bee observations are clearly not independent given the dance mode, so choosing an HMM is likely not going to provide desirable performance. Typically, it is useful to have domain-specific knowledge of at least one example of a time-series segment that can be used to design the structure of individual modes in a model. Overall, however, this issue of model selection in the Bayesian nonparametric setting is an open area of research. Finally, given the Bayesian framework, the models that we have presented necessitate a choice of prior. We have found in practice that the models are relatively robust to the hyperprior settings for the concentration parameters. On the other hand, the choice of base measure tends to affect results significantly, which is typical of simpler Bayesian nonparametric models such as DP mixtures. We have found that quasi-empirical Bayes' approaches for setting the base measure tend to help push the mass of the distribution into reasonable ranges (see [36] for details).

Our focus in this article has been the advantages of various hierarchical, nonparametric Bayesian models; detailed algorithms for learning and inference were omitted. One major advantage of the particular Bayesian nonparametric approaches explored in this article is that they lead to computationally efficient methods for learning Markov switching models of unknown order. We point the interested reader to [13], [14], [19], and [28] for detailed presentations of Markov chain Monte Carlo algorithms for inference and learning.

AUTHORS

Emily B. Fox (fox@stat.duke.edu) received the S.B., M.Eng., E.E., and Ph.D. degrees from the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology in 2004, 2005, 2008, and 2009, respectively. She is currently a postdoctoral scholar in the Department of Statistical Science at Duke University. She received both the National Defense Science and Engineering Graduate (NDSEG) and National Science Foundation (NSF) Graduate Research fellowships and currently holds an NSF Mathematical Sciences Postdoctoral Research Fellowship. She was awarded the 2009 Leonard J. Savage Thesis Award in Applied Methodology, the 2009 MIT EECS Jin-Au Kong Outstanding Doctoral Thesis Prize, the 2005 Chorafas Award for superior contributions in research, and the 2005 MIT EECS David Adler Memorial Second Place Master's Thesis Prize. Her research interests are in Bayesian and Bayesian nonparametric approaches to multivariate time series analysis.

Erik B. Sudderth (sudderth@cs.brown.edu) is an assistant professor in the Department of Computer Science at Brown University. He received the bachelor's degree (summa cum laude) in electrical engineering from the University of California, San Diego, and the master's and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology. His research interests include probabilistic graphical models; nonparametric Bayesian analysis; and applications of statistical machine learning in computer vision, signal processing, artificial intelligence, and geoscience. He was awarded a National Defense Science and Engineering Graduate Fellowship (1999), an Intel Foundation Doctoral Fellowship