Hilbert Space Embeddings of Predictive State Representations

Byron Boots BBOOTS@CS.WASHINGTON.EDU
Arthur Gretton ARTHUR.GRETTON@GMAIL.COM
Geoffrey J. Gordon GGORDON@CS.CMU.EDU

Many problems in machine learning and artificial intelligence involve discrete-time partially observable nonlinear dynamical systems. If the observations are discrete, then *Hidden Markov Models* (HMMs) (Rabiner, 1989) or, in the control setting, *Input-Output HMMs* (IO-HMMs) (Bengio & Frasconi, 1995), can be used to represent belief as a discrete distribution over latent states. *Predictive State Representations (PSRs)* (Littman et al., 2002) are generalizations of IO-HMMs that have attracted interest because they can have greater representational capacity for a fixed model dimension. In contrast to latent-variable representations like HMMs, PSRs represent the state of a dynamical system by tracking occurrence probabilities of future observable events (called *tests*) conditioned on past observable events (called *histories*). One of the prime motivations for modeling dynamical systems with PSRs was that, because tests and histories are observable quantities, learning PSRs should be easier than learning IO-HMMs by heuristics like Expectation Maximization (EM), which suffer from bad local optima and slow convergence rates. For example, Boots et al. (2010) proposed a statistically consistent *spectral* algorithm for learning PSRs with discrete observations and actions.

Despite their positive properties, many algorithms for PSRs are restricted to sets of actions and observations with only moderate cardinality. For continuous actions and observations, learning algorithms for PSRs often run into trouble: we cannot hope to see each action or observation more than a small number of times, so we cannot gather enough data to estimate the PSR parameters accurately without additional assumptions. Previous approaches attempt to learn continuous PSRs by leveraging kernel density estimation (Boots et al., 2010) or modeling PSR distributions with exponential families (Wingate & Singh, 2007a;b); each of these methods must contend with drawbacks such as slow rates of statistical convergence and difficult numerical integration.

In this paper, we fully generalize PSRs to continuous observations and actions using a recent concept called Hilbert space embeddings of distributions (Smola et al., 2007; Sriperumbudur et al., 2008). The essence of our method is to represent distributions of tests, histories, observations, and actions as points in (possibly) infinite-dimensional reproducing kernel Hilbert spaces. During filtering we update these embedded distributions using a kernel version of Bayes' rule (Fukumizu et al., 2011). The advantage of this approach is that embeddings of distributions can be estimated without having to contend with problems such as density estimation and numerical integration.

Our approach is similar to recent work that applies kernel methods to dynamical system modeling and reinforcement learning, which we summarize here. Song et al. (2010) proposed a nonparametric approach to learning HMM representations in RKHSs. The resulting dynamical system model, called Hilbert Space Embeddings of Hidden Markov Models (HSE-HMMs), proved to be more accurate compared to competing models on several experimental benchmarks. Despite these successes, HSE-HMMs have two major limitations: first, the update rule for the HMM relies on *density estimation* instead of Bayesian inference in Hilbert space, which results in an awkward model with poor theoretical guarantees. Second, the model lacks the capacity to reason about actions, which limits the scope of the algorithm. Our model can be viewed as an extension of HSE-HMMs that adds inputs and updates state using a kernelized version of Bayes' rule.

Grunewalder et al. (2012) proposed a nonparametric approach to learning transition dynamics in Markov decision processes (MDPs) by representing the stochastic transitions as conditional distributions in RKHS. This work was extended to POMDPs by Nishiyama et al. (2012). The resulting Hilbert space embedding of POMDPs represents distributions over the states, observations, and actions as embeddings in RKHS and uses kernel Bayes rule to update these distribution embeddings. However, the algorithm requires training data that includes labels for the true latent states. This is a serious limitation: it precludes learning dynamical systems directly from sensory data. By contrast, our algorithm only requires access to an unlabeled sequence of actions and observations, and learns the more expressive PSR model, which includes POMDPs as a special case.

References

Bengio, Yoshua and Frasconi, Paolo. An Input Output HMM Architecture. In *Advances in Neural Information Processing Systems*, 1995.

Boots, Byron, Siddiqi, Sajid M., and Gordon, Geoffrey J. Closing the learning-planning loop with predictive state representations. In *Proceedings of Robotics: Science and Systems VI*, 2010.

Fukumizu, Kenji, Song, Le, and Gretton, Arthur. Kernel bayes' rule. In Shawe-Taylor, J., Zemel, R.S., Bartlett, P., Pereira, F.C.N., and Weinberger, K.Q. (eds.), *Advances in Neural Information Processing Systems 24*, pp. 1737–1745. 2011.

Grunewalder, Steffen, Lever, Guy, Baldassarre, Luca, Pontil, Massimiliano, and Gretton, Arthur. Modelling transition dynamics in mdps with rkhs embeddings. *CoRR*, abs/1206.4655, 2012.

Littman, Michael, Sutton, Richard, and Singh, Satinder. Predictive representations of state. In *Advances in Neural Information Processing Systems (NIPS)*, 2002.

Nishiyama, Y, Boularias, A, Gretton, A, and Fukumizu, K. Hilbert space embeddings of pomdps. 2012.

Rabiner, L. R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77(2):257–285, 1989.

Smola, A.J., Gretton, A., Song, L., and Schölkopf, B. A Hilbert space embedding for distributions. In Takimoto, E. (ed.), *Algorithmic Learning Theory*, Lecture Notes on Computer Science. Springer, 2007.

Song, L., Boots, B., Siddiqi, S. M., Gordon, G. J., and Smola, A. J. Hilbert space embeddings of hidden Markov models. In *Proc. 27th Intl. Conf. on Machine Learning (ICML)*, 2010.

Sriperumbudur, B., Gretton, A., Fukumizu, K., Lanckriet, G., and Schölkopf, B. Injective Hilbert space embeddings of probability measures. 2008.

Wingate, David and Singh, Satinder. Exponential family predictive representations of state. In *Proc. NIPS*, 2007a.

Wingate, David and Singh, Satinder. On discovery and learning of models with predictive representations of state for agents with continuous actions and observations. In *Proc. AAMAS*, 2007b.