Kernel & Spectral Algorithms for PGMs

Hilbert Space Embeddings of Conditional Distribution with Applications of Dynamical Systems

In this paper the author uses the concept of Hilbert space embedding approach to handle conditional distributions. They start by introducing Hilbert spaces as spaces in which the distribution can in theory be completely reconstructed just by using the summary statistics from a lower dimensional plane leveraging the fact that empirical estimates of points in lower dimensions are consistent. But these estimates are good in case of discrete but what happens in case of continuous where not all points are observed? This is the key idea of author to generalize to unobserved points using other points that are close to it via reproducing kernel Hilbert space of kernel estimation functions. They extend on the ability of conditional embeddings and cross-covariance operations to manipulate these RKHS embeddings which makes it useful for a wide number of applications in dynamic systems.

Authors demonstrate the ability of RKHS by creating a simple non-parametric algorithm for learning models and doing probabilistic inference for filtering using this approximate approach performing prediction and conditioning operations in parallel in dynamical systems creating a sample use case for famous Kalman filters being exact method performs the operations sequentially making a case for effective comparison between the two. There has been work in the past related to above mentioned and the author's improve upon those approaches in the paper. For experiments, they simulate a particle rotating around origin for comparing performance with Kalman filters where the Kalman filters had constant error the RKHS first had to learn after which it performed equal and better in complex datasets without the need for exact dynamics of the environment. The second experiment involved estimating camera rotation matrices with Kalman and Random where also the RKHS performed significantly better but also note the performance degrades for increasing noise level the comparison for which is not reported.

A Spectral Algorithm for Learning Hidden Markov Models

Authors start by introducing Hidden Markov Models (HMM) as the prime choice for modelling sequential data as random variable in probabilistic framework with numerous applications like finding the most likely sequence of hidden variables hence unsupervised learning or sampling from joint posterior, monitor belief state of dynamical systems, infer latent variables from time series, density estimation, etc. The parameter of a HMM are its hidden states, observations, conditional independencies, initial state distribution, transition and observation matrix. For learning discrete HMMs popular hueristic algorithms like Expectation-Maximization a.k.a. Baum-Welch algorithm exist which are shown to be computationally hard under cryptographic conditions (same distribution of observation from a state or mixture of other states), the author addresses this issue by offering an efficient algorithm with learning guarantees for invertible HMMs by borrowing subspace identification techniques from linear systems theory and applying it to realize observation operators. Work on identification for general linear models has been done by using Singular Value Decomposition (SVD) to discover subspace containing relevant states and then learning effective transition and observation matrices assuming additive noise (independent of state) models such as Gaussian noise applicable to Kalman filters but not to HMMs, authors combines above mentioned approach with observation operators obtained by doing an SVD on correlation matrix between past and future observations resulting in a efficient learnable representation of HMM.

The key points of the algorithms only maintain low dimensional estimates of conditional and sequence prediction of observed variables only which avoid estimating transition and observation matrix, then finding a set of linear transformations that iteratively updates these estimates with SVD that captures the dynamics of observable representation and then use co-occurrence statistics and spectral decomposition to learn this parametrization. The authors provide a theoretical proof of correct approximation of sequence and conditional probability where the computation cost attached are that of computing SVD $O(n^3)$ and calculation joint probability of transition variables same as in original HMM. Authors leave on the note of potential for further improvements since current algorithm cannot compute probabilities that involve hidden states, and also the approximation of joint distribution over observation sequences degrades polynomially.