Variational Nonparametric Bayesian Hidden Markov Model

Nan Ding, Zhijian Ou
ozj@tsinghua.edu.cn
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

Objective: discover the structure of the HMM state space.
Method: propose a nonparametric Bayesian HMM (NBHMM) based on Dirichlet Process
Advantage: theoretically sound, efficient computation with variational inference

Differences from other existing nonparametric Bayesian HMM


1. iHMM and HDP-HMM employ sampling based inference. We apply the efficient variational inference for the NBHMM.
2. iHMM deals only with discrete observations. NBHMM supports continuous observations via (infinite) Gaussian mixtures.
3. The transition distribution in iHMM and HDP-HMM is generated from HDP. In the NBHMM, directly created from a stickbreaking construction, simpler

Variational Inference on NBHMM

Basic Idea: minimize the Kullback-Leibler distance

\[ KL(q||p) = \sum_{x} q(x) \log \frac{q(x)}{p(x)} = \sum_{x} q(x) \log q(x) - \sum_{x} q(x) \log p(x) \]

\[ q(x) = \frac{q(x)}{\sum_{x} q(x)} \]

\[ p(x) = \frac{p(x)}{\sum_{x} p(x)} \]

Two variational assumptions:

Assume \((\pi, A, \mu, \Sigma)\) and \((s, h)\) are mutually independent.
We only compute the posterior probabilities for \(L\) states of the infinite large state-space.
Only the states corresponding to "large" posteriors are effective in explaining the observed data.

Experiment Results

(1) NBHMM vs. Classical HMM

(2) Triphone model

(3) Chinese isolated syllable recognition

- 1254 syllables, whole-syllable HMMs
- (14 MFCCs+ E) \(3 = 45\)-dim feature
- 50 males data, leave-one-out test
- classic HMM: 6-state 73.4%, 16-state 80.1%
- NBHMM: discover 14-18 effective states for the syllables, 78.9% accuracy