Expectation propagation as a way of life

Yingzhen Li

Department of Engineering

Feb. 2014

Reference

This talk mainly references the paper
Gelman et al. Expectation propagation as a way of life. arXiv:1412.4869.
Also this paper
Xu et al. Distributed Bayesian Posterior Sampling via Moment Sharing. NIPS '14
has very similar idea.

Sketch of the problem

Given a model $p(\mathbf{x}|\theta)$ and a prior distribution $p_0(\theta)$

- observed the dataset $D = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$
- **Goal**: Evaluate the computationally intractable posterior

$$p(\theta|D) \propto p(\theta, D) = p_0(\theta) \prod_{i=1}^{N} p(\mathbf{x}_i|\theta)$$

- assume we have a "black-box" like algorithm: given an input f

$$approximate(f) \approx f$$

  measured by some meaningful distance/divergence/similarity
- **One possible solution**: use the algorithm to form approximation

$$q(\theta) := p_0(\theta) \prod_{i=1}^{N} q_i(\theta) \approx p(\theta|D)$$

Smart partitioning

- Obviously that the factorisation of the posterior is non-unique.
- ...so instead we can partition the dataset

$$D = \cup_{k=1}^{K} D_k, \quad D_i \cap D_j = \emptyset, \forall i \neq j$$

- ...and change to a better approximation

$$\prod_{k=1}^{K} q_k(\theta) p_0(\theta) \approx \prod_{k=1}^{K} p(D_k|\theta) p_0(\theta)$$

Why EP-like approximations

- mean-field: forming q_k where

$$q_k(\theta) = approximate(p(D_k|\theta))$$

  - Pro: easy to implement, natural to go parallel
  - Con: less inaccurate since computing updates based on local info only

Why EP-like approximations

- Assumed density filtering (ADF): forming q_k by

$$q_k(\theta) \prod_{i<k} q_i(\theta) p_0(\theta) = approximate \left(p(D_k|\theta) \prod_{i<k} q_i(\theta) p_0(\theta) \right)$$

  - Pro: more accurate approximations, can be done by one-pass
  - Con: highly dependent on the update scheduling, not seeing the whole structure

Why EP-like approximations

- Expectation propagation (EP): looping ADF for more iterations

$$q_k(\theta)\prod_{i\neq k} q_i(\theta)p_0(\theta) = \textit{approximate}\left(p(D_k|\theta)\prod_{i\neq k} q_i(\theta)p_0(\theta)\right)$$

  - Pro: less dependent on scheduling, "correct itself", updates based on the info from other sites
  - Con: no convergence guarantee

The EP-like algorithm

Algorithm 1 the EP-like algorithm

1: **while** not converged **do**
2: choose a factor q_k to refine:
3: exclusion: $q^{\backslash k}(\theta) = p_0(\theta) \prod_{i \neq k} q_i(\theta)$
4: update: $q_k(\theta) \leftarrow approximate(q^{\backslash k}(\theta) p(D_k|\theta)) / q^{\backslash k}(\theta)$
5: inclusion: $q(\theta) \leftarrow q_k(\theta) q^{\backslash k}(\theta)$
6: **end while**

Paralleled EP-like algorithm

Assume we have $K + 1$ processing units (one central processor and K worker units)

1. Partition dataset into K subsets
2. Initialise local approximations q_k
3. Run EP:
   - compute cavities: worker units receive cavities from the central processor
   - compute updates: run *approximate*() function locally at each worker unit site
   - update global approximation q: worker units send back the current updated local functions q_k to the central processor

Some examples of the "black-box" algorithm

- moment matching: minimise $KL(q^{\backslash k}(\theta)p(D_k|\theta)||q(\theta))$
  - often have analytical solutions when using Gaussians
  - can use numerical approximations, e.g. sampling, Gaussian quadrature
  - use nested EP to form a second-level approximation
  - Relaxed EP: introducing a relaxation factor into the KL term
- power EP: minimise α-divergence
- mode matching
- Laplace propagation
- EP-ABC
- ...

More references

Minka, T. Expectation propagation for approximate Bayesian inference. UAI '01

Minka, T. Power EP. Technical report, Microsoft Research, Cambridge. 2005

Riihimaki et al. Nested expectation propagation for Gaussian process classification with a multinomial probit likelihood. JMLR '14

Smola et al. Laplace propagation. NIPS '04

Qi, Y. and Guo, Y. Message passing with l_1 penalized KL minimization. ICML '13

Zoeter, O. and Heskes, T. Gaussian quadrature based expectation propagation. AISTATS '05

Barthelme, S. and Chopin, N. ABC-EP: Expectation propagation for likelihood-free Bayesian computation. ICML '11