Sum-Product Problem

Tran The Truyen
thetruyen.tran@postgrad.curtin.edu.au

April 27, 2008

Contents

1 Problem Statement

Given N discrete variables $x_i \in \{1, 2, ..., S\}$ for $i = 1, 2, ..., N$. Let $x = (x_1, x_2, ..., x_N)$ be the joint variable. Denote by c the index of the subset of variables x_c, and $\psi_c(x_c)$ be positive real functions. Let $\mathcal{C}$ be the set of all those indices. We want to compute the following sum

$$Z = \sum_{x_1, x_2, ..., x_N} \prod_{c \in \mathcal{C}} \psi_c(x_c) \tag{1}$$

Z is often referred to as *partition function*, a term borrowed from statistical mechanics and thermal dynamics. This is a very hard problem because the number of all possible assignments of x is S^N. In practice, N can be quite large, e.g. $N \in [10^2, 10^{12}]$ so brute-force computation is intractable.

2 Applications

The partition function has wide applications in spatial stochastic processes. Here we present one of the most popular formulations - the Markov Random Fields.

2.1 Markov Random Fields

Markov Random Fields are popular tool for modelling spatial stochastic processes (e.g. see [2, 6]). The joint variable $x = (x_1, x_2, ..., x_N)$ is modelled as a random variable with the following distribution

$$\Pr(x) = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(x_c) \tag{2}$$

where $\psi_c(x_c)$ encodes interaction between variables in x_c and Z is the normalisation term, i.e. $\sum_x \Pr(x) = 1$.

3 State-of-the-Arts

The most well-known methods are:

- Markov chains Monte-Carlo methods [3, 1] can theoretically estimate Z but they can be very slow in practice.
- Loopy Belief Propagation, also known as Sum-Product algorithm: this is originated from Artificial Intelligence [7], often yields good solutions but may not converge. Later, this algorithm was proven to be consistent to the state-of-the-art decoding algorithm in Information Theory [5]. Recently, it was shown that the algorithm can be derived by minimising the Bethe free energy [10] - a well-known concept in statistical physics.
    - Let us represent each variable as a node and each pair of dependent variables as an edge. If the result graph is a tree, then Belief Propagation is guaranteed to compute Z exactly in linear time. In this case, Belief Propagation reduces to dynamic programming.
- In [8] the authors exploit the convexity of Z and propose a method to compute the upper bound of Z. The final algorithm is quite similar to the Sum-Product.
- Variational methods which include Mean Field as a special case [4, 9]. This class of algorithms can compute the lower bound of Z.

References

[1] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1-2):5–43, 2003.

[2] Julian Besag. Spatial interaction and the statistical analysis of lattice systems (with discussions). *Journal of the Royal Statistical Society Series B*, 36:192–236, 1974.

[3] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[4] M.I. Jordan, T. Ghahramani, Z.and Jaakkola, and L.K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, 1999.

[5] F. R. Kschischang, B. J. Frey, and H. A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, February 2001.

[6] S.L. Lauritzen. *Graphical Models*. Oxford Science Publications, 1996.

[7] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, CA, 1988.

[8] M. J. Wainwright, T. Jaakkola, and A. S. Willsky. A new class of upper bounds on the log partition function. *IEEE Transactions on on Information Theory*, 51:2313–2335, Jul 2005.

[9] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. Technical Report 649, Department of Statistics, University of California, Berkeley, 2003.

[10] Jonathan Yedidia, Willia Freeman, and Yair Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312, 2005.