Approximating the Permanent with Belief Propagation

Bert Huang, Tony Jebara
Computer Science Department
Columbia University
New York, NY 10027
{bert,jebara}@cs.columbia.edu

This work describes a method of approximating matrix permanents efficiently using belief propagation. We formulate a probability distribution whose partition function is exactly the permanent, then use Bethe free energy to approximate this partition function. After deriving some speedups to standard belief propagation, the resulting algorithm requires $\mathcal{O}(n^3)$ time per iteration and seems empirically to take a constant number of iterations.

The permanent is a scalar quantity computed from a matrix and has been an active topic of research for well over a century. It plays a role in cryptography and statistical physics where it is fundamental to Ising and dimer models. While the determinant of an $n \times n$ matrix can be evaluated exactly in sub-cubic time, efficient methods for computing the permanent have remained elusive. Since the permanent is #P-complete, efficient exact evaluations cannot be found in general. Recently, promising fully-polynomial randomized approximate schemes (FPRAS) have emerged which provide arbitrarily close approximations. Significant progress has produced an FPRAS that can handle arbitrary $n \times n$ matrices with non-negative entries [5]. The method uses Markov chain Monte Carlo and only requires a polynomial order of samples.

It remains to be seen if other approximate inference methods can be brought to bear on the permanent. For instance, loopy belief propagation has also recently gained prominence in the machine learning community. The method is exact for singly-connected networks such as trees. In certain special loopy graph cases, including graphs with a single loop, bipartite matching graphs [1] and bipartite multi-matching graphs [4], the convergence of BP has been proven. In more general loopy graphs, loopy BP still maintains some surprising empirical success. Theoretical understanding of the convergence of loopy BP has recently been improved by noting certain general conditions for its fixed points and relating them to minima of Bethe free energy. This article proposes belief propagation for computing the permanent and investigates some theoretical and experimental properties. In particular, we provide evidence that BP produces a bound on the permanent for any non-negative matrix while most other known bounds require assumptions on the structure of the input matrix.

We construct a bipartite factor graph similar to [1] that represents a distribution over permutations parameterized by a weight matrix. Observing that the partition function of this distribution is the permanent of the weight matrix, we approximate the permanent using Bethe free energy, which approximates the negative log-partition function. We compute the Bethe free energy of the converged pseudo-marginal sum-product beliefs and exponentiate its negation to obtain our approximation.

Empirically, we have found that using the Bethe approximation appears to always underestimate the permanent, but the actual value output by the approximation is more monotonic with respect to the true permanent than sampling methods. This monotonicity implies that our approximation, though less accurate with respect to actual distance from the true permanent, is more useful as a measure, or as a kernel. We show this monotonicity in experiments including using the Bethe approximation as a permutational invariant kernel.

The kernels we computed using our approximation were valid Mercer kernels, which was a surprise, and we wish to explore the theoretical implications of this. It is now known that the permanent can be used as a valid kernel [2], but it is not trivial that our approximation of the permanent is also valid.

Finally, one important open question about this work is whether or not we can guarantee convergence. Though the sum-product algorithm always converged in practice, so far only the max-product algorithm is guaranteed to converge to the correct maximum matching [1, 4] via arguments on the unwrapped computation tree of belief propagation. We attempted convexity analysis of the Bethe free energy of our distribution, but found that our formulation did not meet the sufficient conditions provided in [3] nor does our distribution fit the criteria for non-convex convergence provided in [6]. Since all our empirical evidence implies that BP always converges, we suspect that we have not yet correctly analyzed the true space traversed during optimization. One promising avenue we are exploring is to exploit the deterministic nature of the matching constraints and how they force certain conditional entropies to be zero.

References

[1] M. Bayati, D. Shah, and M. Sharma. Maximum weight matching via max-product belief propagation. In *Proc. of the IEEE International Symposium on Information Theory*, 2005.

[2] M. Cuturi. Permanents, transportation polytopes and positive definite kernels on histograms. In *International Joint Conference on Artificial Intelligence, IJCAI*, 2007.

[3] T. Heskes. Convexity arguments for efficient minimization of the bethe and kikuchi free energies. *Journal of Artificial Intelligence Research*, 26, 2006.

[4] B. Huang and T. Jebara. Loopy belief propagation for bipartite maximum weight b-matching. In *Artificial Intelligence and Statistics (AISTATS)*, 2007.

[5] M. Jerrum, A. Sinclair, and E. Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, 2004.

[6] S. Tatikonda and M. Jordan. Loopy belief propagation and Gibbs measures. In *Proc. Uncertainty in Artificial Intell., vol. 18*, 2002.