Collaborative Distributed Inference with Minimal Online Communication

O. Patrick Kreidl, Müjdat Çetin and Alan S. Willsky
MIT Laboratory for Information and Decision Systems
77 Massachusetts Ave, 32-D572
Cambridge, MA 02139
{opk,mcetin,willsky}@mit.edu
http://ssg.mit.edu

Given a probabilistic graphical model with both hidden and observation nodes, common inference problems include solving for the maximum a-posteriori (MAP) and maximum-posterior-marginal (MPM) estimates (i.e., the modes of, respectively, the joint distribution of all hidden nodes or the marginal distributions of every hidden node upon conditioning on any realization of all observation nodes). Belief propagation (BP) and related algorithms [6], [11] are popular methods for solving such inference problems. BP originates as an exact solution on tree-structured graphs, converging to the optimal solution after a finite number of iterations. Though convergence is not guaranteed on more general graphs, there is empirical (e.g., [5]) and theoretical (e.g., [10]) evidence that BP can converge to acceptable approximate solutions.

Conceptually, the message-passing formulation of BP extends naturally to a distributed network setting: associating to each node and edge in the graph a distinct processor and communication link, respectively, the algorithm is equivalent to a sequence of purely-local computations interleaved with only nearest-neighbor communications. Specifically, each computation event corresponds to all nodes simultaneously evaluating a local *message function*, or a mapping from all messages received in the preceding communication event to the messages that must be transmitted in the next communication event. Practically, the viability of BP in a distributed network setting appears to rest upon an implicit assumption that online communication resources are abundant. In a general network, because mere termination of the algorithm is in question, the required communication resources are a-priori unbounded. Even in a tree-structured network, iterative reliable transmission of the exact messages presumes communication channels with infinite capacity (in bits per observation), or at least of sufficiently high bandwidth such that the resulting finite message precision is essentially error-free.

In some distributed settings (e.g., power-constrained sensor networks with voluminous local observations), it may be prohibitively costly or simply infeasible to justify such idealized online communication assumptions. Fortunately, there is evidence to suggest ideal communication is not essential for BP-like algorithms to achieve acceptable performance. A message function by which nodes can decide to discard certain messages is examined in [2], where empirical results indicate such a "communication-sensitive" message-passing algorithm (if it converges) achieves comparable performance given low-to-moderately frequent occurrences of discarded messages. The analysis in [3] studies the effect of multiplicative errors in BP message computations, establishing bounds on the accumulation of these errors as the algorithm proceeds and, in turn, implying "small-enough" message errors will not alter the behavior of BP. A converse of the evidence in [2] and [3], however, is that BP-like algorithms may perform poorly when communication resources become severely constrained.

Assuming communication constraints are severe, we examine the extent to which alternative message functions can serve to minimize the unavoidable loss in performance. Our analysis begins with a very specific inference objective and severely constrained communication scheme: we wish to solve for MAP estimates in a discrete-variable tree-structured graphical model, where (i) all edges correspond to low-capacity communication links (e.g., one bit per observation) and (ii) the message schedule is restricted to exactly one directed sweep through the tree network. We cast the problem within a variational inference formulation [4], [12], viewing the message functions as the degrees-of-freedom subject to the constrained information flow implied by the stipulated link capacities and restricted message schedule.

The resulting variational problem turns out to be equivalent to the optimization problem underlying the well-studied decentralized detection paradigm [7], [8], [9], [1]. Known necessary optimality conditions in decentralized detection immediately provide a finite parameterization for the message functions, clearly distinct from the traditional BP counterpart, as well as an iterative algorithm to be executed *offline* (i.e., before observations are realized). This offline procedure serves to couple the parameters of all local message functions, in a manner that depends non-trivially on global problem statistics, in order to (at least partially) mitigate the loss in MAP performance due to the stipulated online communication constraints. The same variational analysis applies for the MPM inference objective, where we discover an added advantage with respect to the distributed network setting: the global offline procedure can itself be expressed as an iterative message-passing algorithm with favorable convergence properties.

The proposed variational approach for distributed inference under severe communication constraints illuminates upon a number of design principles that merit further exploration. Firstly, we articulate a distinction between the graph that defines the probabilistic model and the graph that defines the online communication constraints—do they need to be commensurate as assumed above? Secondly, mitigating the performance loss that results from an imposed communication graph requires some type of offline preprocessing to globally couple the local message functions. Indeed, such an offline procedure is itself a tax on distributed network resources—under what circumstances does the potential for improved online inference justify this offline resource expenditure? Finally, analogous to the success demonstrated by traditional BP when applied to inference problems beyond those for which it was originally derived, we conjecture that the offline message-passing algorithm derived for MPM inference on a tree-structured probabilistic/communication graph will find success in more general communication-constrained inference problems.

REFERENCES

[1] R. S. Blum, S. A. Kassam, and H. V. Poor. Distributed detection with multiple sensors: Part ii–advanced topics. *Proceedings of the IEEE*, 85(1):64–79, January 1997.

[2] L. Chen, M. J. Wainwright, M. Çetin, and A. S. Willsky. Data association based on optimization in graphical models with application to sensor networks. *Mathematical and Computer Modeling (Special Issue on Optimization and Control for Military Applications)*, 2005. To appear.

[3] A. T. Ihler, J. W. Fisher III, and A. S. Willsky. Message errors in belief propagation. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems 17*. MIT Press, Cambridge, MA, 2005.

[4] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 105–161. The MIT Press, Cambridge, MA, 1999.

[5] K. P. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief propagation for approximate inference: an empirical study. *Uncertainty in Artificial Intelligence*, 15:467–475, 1999.

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

[7] R. R. Tenney and N. R. Sandell, Jr. Detection with distributed sensors. *IEEE Transactions on Aerospace Electronic Systems*, 17(4):501–510, September 1981.

[8] J. N. Tsitsiklis. Decentralized detection. In H. V. Poor and J. B. Thomas, editors, *Advances in Statistical Signal Processing*, volume 2, pages 297–344. JAI Press, Greenwich, CT, 1993.

[9] R. Viswanathan and P. K. Varshney. Distributed detection with multiple sensors: Part i–fundamentals. *Proceedings of the IEEE*, 85(1):54–63, January 1997.

[10] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. Tree-based reparameterization framework for analysis of sum-product and related algorithms. *IEEE Transactions on Information Theory*, 49(5):1120–1146, May 2003.

[11] A. S. Willsky. Multiresolution markov models for signal and image processing. *Proceedings of the IEEE*, 90(8):1396–1458, August 2002.

[12] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free energy approximations and generalized belief propagation algorithms. Technical Report MERL TR-2004-040, Mitsubishi Electric Research Laboratories, May 2004. [Online] Available: http://www.merl.com/reports/docs/TR2004-040.pdf.

Topic: graphical models/controls
Preference: oral