

FastInf: An Efficient Approximate Inference Library

Ariel Jaimovich

Ofer Meshi

*School of Computer Science and Engineering
Hebrew University of Jerusalem
Jerusalem, Israel 91904*

ARIELJ@CS.HUJI.AC.IL

MESHI@CS.HUJI.AC.IL

Ian McGraw

*Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139 USA*

IMCGRAW@MIT.EDU

Gal Elidan

*Department of Statistics
Hebrew University of Jerusalem
Jerusalem, Israel 91905*

GALEL@HUJI.AC.IL

Editor: Cheng Soon Ong

Abstract

The FastInf C++ library is designed to perform memory and time efficient approximate inference in large-scale discrete undirected graphical models. The focus of the library is propagation based approximate inference methods, ranging from the basic loopy belief propagation algorithm to propagation based on convex free energies. Various message scheduling schemes that improve on the standard synchronous or asynchronous approaches are included. Also implemented are a clique tree based exact inference, Gibbs sampling, and the mean field algorithm. In addition to inference, FastInf provides parameter estimation capabilities as well as representation and learning of shared parameters. It offers a rich interface that facilitates extension of the basic classes to other inference and learning methods.

Keywords: graphical models, Markov random field, loopy belief propagation, approximate inference

1. Introduction

Probabilistic graphical models (Pearl, 1988) are a framework for representing a complex joint distribution over a set of n random variables $\mathcal{X} = \{X_1 \dots X_n\}$. A qualitative graph encodes probabilistic independencies between the variables and implies a decomposition of the joint distribution into a product of local terms:

$$P(\mathcal{X}) = \frac{1}{Z} \prod_i \psi_i(C_i),$$

where C_i are subsets of $\mathcal{X}$ defined by the cliques of the graph structure and $\psi_i(C_i)$ are the quantitative parameters (potential functions) that define the distribution. Computing marginal probabilities and likelihood in graphical models are critical tasks needed both for making predictions and to facilitate learning. Obtaining exact answers to these inference queries is often infeasible even for relatively

modest problems. Thus, there is a growing need for inference methods that are both efficient and can provide reasonable approximate computations. Despite few theoretical guarantees, the Loopy Belief Propagation (LBP, Pearl, 1988) algorithm has gained significant popularity in the last two decades due to impressive empirical success, and is now being used in a wide range of applications ranging from transmission decoding to image segmentation (Murphy and Weiss, 1999; McEliece et al., 1998; Shental et al., 2003). Recently there has been an explosion in practical and theoretical interest in propagation based inference methods, and a range of improvements to the convergence behavior and approximation quality of the basic algorithms have been suggested (Wainwright et al., 2003; Wiegnerinck and Heskes, 2003; Elidan et al., 2006; Meshi et al., 2009).

We present the *FastInf* library for efficient approximate inference in large scale discrete probabilistic graphical models. While the library’s focus is propagation based inference techniques, implementations of other popular inference algorithms such as mean field (Jordan et al., 1998) and Gibbs sampling are also included. To facilitate inference for a wide range of models, *FastInf*’s representation is flexible allowing the encoding of standard Markov random fields as well as template-based probabilistic relational models (Friedman et al., 1999; Getoor et al., 2001), through the use of shared parameters. In addition, *FastInf* also supports learning capabilities by providing parameter estimation based on the Maximum-Likelihood (ML) principle, with standard regularization. Missing data is handled via the Expectation Maximization (EM) algorithm (Dempster et al., 1977).

FastInf has been used successfully in a number of challenging applications, ranging from inference in protein-protein networks with tens of thousands of variables and small cycles (Jaimovich et al., 2005), through protein design (Fromer and Yanover, 2008) to object localization in cluttered images (Elidan et al., 2006).

2. Features

The *FastInf* library was designed while focusing on generality and flexibility. Accordingly, a rich interface enables implementation of a wide range of probabilistic graphical models to which all inference and learning methods can be applied. A basic general-purpose propagation algorithm is at the base of all propagation variants and allows straightforward extensions.

A model is defined via a graph interface that requires the specification of a set of cliques $C_1 \dots C_k$, and a corresponding set of tables that quantify the parametrization $\psi_i(C_i)$ for each joint assignment of the variables in the clique C_i . This general setting can be used to perform inference both for the directed Bayesian network representation and the undirected Markov one.

2.1 Inference Methods

FastInf includes implementations of the following inference methods:

- Exact inference by the Junction-Tree algorithm (Lauritzen and Spiegelhalter, 1988)
- Loopy Belief Propagation (Pearl, 1988)
- Generalized Belief Propagation (Yedidia et al., 2005)
- Tree Re-weighted Belief Propagation (Wainwright et al., 2005)
- Propagation based on convexification of the Bethe free energy (Meshi et al., 2009).
- Mean field (Jordan et al., 1998)
- Gibbs sampling (Geman and Geman, 1984)

By default, all methods are used with standard asynchronous message scheduling. We also implemented two alternative scheduling approaches that can lead to better convergence properties (Wainwright et al., 2002; Elidan et al., 2006). All methods can be applied to both sum and max product propagation schemes, with or without damping of messages.

2.2 Relational Representation

In many domains, a specific local interaction pattern can recur many times. To represent such domains, it is useful to allow multiple cliques to share the same parametrization. In this case a set of template table parametrizations $\psi_1, \dots, \psi_T$ are used to parametrize all cliques using

$$P(\mathcal{X}) = \frac{1}{Z} \prod_t \prod_{i \in I(t)} \psi_t(C_i),$$

where $I(t)$ is the set of cliques that are mapped to the t 'th potential. This template based representation allows the definition of large-scale models using a relatively small number of parameters.

2.3 Parameter Estimation

FastInf can also be used for learning the parameters of the model from evidence. This is done by using gradient-based methods with the Maximum-Likelihood (ML) objective. The library also handles partial evidence by applying the EM algorithm (Dempster et al., 1977). Moreover, FastInf supports L_1 and L_2 regularization that is added as a penalty term to the ML objective.

3. Documentation

For detailed instructions on how to install and use the library, examples for usage and documentation on the main classes of the library visit FastInf home page at: <http://compbio.cs.huji.ac.il/FastInf>.

Acknowledgments

We would like to acknowledge Menachem Fromer, Haidong Wang, John Duchi and Varun Ganapathi for evaluating the library and contributing implementations of various functions. This library was initiated in Nir Friedman's lab at the Hebrew University and developed in cooperation with Daphne Koller's lab at Stanford. AJ was supported by the Eshkol fellowship from the Israeli Ministry of Science. IM and GE were supported by the DARPA transfer learning program under contract FA8750-05-2-0249

References

- A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B 39:1–39, 1977.
- G. Elidan, I. McGraw, and D. Koller. Residual belief propagation: Informed scheduling for asynchronous message passing. In *UAI 2006*.

- G. Elidan, G. Heitz, and D. Koller. Learning object shape: From drawings to images. In *CVPR* 2006.
- N. Friedman, L. Getoor, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In *IJCAI* 1999.
- M. Fromer and C. Yanover. A computational framework to empower probabilistic protein design. In *Bioinformatics*, pages 214–222, 2008.
- S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 721–741, 1984.
- L. Getoor, N. Friedman, D. Koller, and B. Taskar. Learning probabilistic models of relational structure. In *ICML* 2001.
- A. Jaimovich, G. Elidan, H. Margalit, and N. Friedman. Towards an integrated protein-protein interaction network. In *RECOMB*, 2005.
- M. I. Jordan, Z. Ghahramani, T. Jaakkola, and L. K. Saul. An introduction to variational approximations methods for graphical models. In M. I. Jordan, editor, *Learning in Graphical Models*. Kluwer, Dordrecht, Netherlands, 1998.
- S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*, 1988.
- R. McEliece, D. McKay, and J. Cheng. Turbo decoding as an instance of pearl’s belief propagation algorithm. *IEEE Journal on Selected Areas in Communication*, 16:140–152, 1998.
- O. Meshi, A. Jaimovich, A. Globerson, and N. Friedman. Convexifying the bethe free energy. In *UAI* 2009.
- K. Murphy and Y. Weiss. Loopy belief propagation for approximate inference: An empirical study. In *UAI* 1999.
- J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Francisco, California, 1988.
- N. Shental, A. Zomet, T. Hertz, and Y. Weiss. Learning and inferring image segmentations with the GBP typical cut algorithm. In *ICCV* 2003.
- M. J. Wainwright, T. Jaakkola, and A. S. Willsky. Tree-based reparameterization for approximate estimation on loopy graphs. In *NIPS* 2002.
- M. J. Wainwright, T. Jaakkola, and A. S. Willsky. Exact map estimates by (hyper)tree agreement. In *NIPS* 2002.
- M. J. Wainwright, T.S. Jaakkola, and A. S. Willsky. A new class of upper bounds on the log partition function. *IEEE Transactions on Information Theory*, 51(7):2313–2335, 2005.
- W. Wiegerinck and T. Heskes. Fractional belief propagation. In *NIPS* 2002.
- J.S. Yedidia, W.T. Freeman, and Y. Weiss. Constructing free energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51:2282–2312, 2005.