

Structured Belief Propagation for NLP

Matthew R. Gormley **Jason Eisner**
Department of Computer Science
Johns Hopkins University, Baltimore, MD
{mrg, jason}@cs.jhu.edu

1 Tutorial Overview

Statistical natural language processing relies on probabilistic models of linguistic structure. More complex models can help capture our intuitions about language, by adding linguistically meaningful interactions and latent variables. However, inference and learning in the models we *want* often poses a serious computational challenge.

Belief propagation (BP) and its variants provide an attractive approximate solution, especially using recent training methods. These approaches can handle joint models of interacting components, are computationally efficient, and have extended the state-of-the-art on a number of common NLP tasks, including dependency parsing, modeling of morphological paradigms, CCG parsing, phrase extraction, semantic role labeling, and information extraction (Smith and Eisner, 2008; Dreyer and Eisner, 2009; Auli and Lopez, 2011; Burkett and Klein, 2012; Naradowsky et al., 2012; Stoyanov and Eisner, 2012).

This tutorial delves into BP with an emphasis on recent advances that enable state-of-the-art performance in a variety of tasks. Our goal is to elucidate how these approaches can easily be applied to new problems. We also cover the theory underlying them. Our target audience is researchers in human language technologies; we do not assume familiarity with BP.

In the first three sections, we discuss applications of BP to NLP problems, the basics of modeling with factor graphs and message passing, and the theoretical underpinnings of “what BP is doing” and how it relates to other inference techniques. In the second three sections, we cover key extensions to the standard BP algorithm to enable modeling of linguistic structure, efficient inference, and approximation-aware training. We survey a variety of software tools and introduce a new software framework that incorporates many of the modern approaches covered in this tutorial.

2 Outline

1. Probabilistic Modeling [15 min., Eisner]
 - Intro: Modeling with factor graphs
 - Constituency and dependency parsing
 - Joint CCG Parsing and supertagging
 - Transliteration; Morphology
 - Alignment; Phrase extraction
 - Joint models for NLP; Semantic role labeling; Targeted sentiment
 - Variable-centric view of the world
2. Belief Propagation Basics [40 min., Eisner]
 - Messages and beliefs
 - Sum-product algorithm
 - Relation to the forward-backward and Viterbi algorithms
 - BP as dynamic programming
 - Acyclic vs. loopy graphs
3. Theory [25 min., Gormley]
 - From sum-product to max-product
 - From arc consistency to BP
 - From Gibbs sampling to particle BP to BP
 - Convergence properties
 - Bethe free energy
4. Incorporating Structure into Factors and Variables [30 min., Gormley]
 - Embedding dynamic programs (e.g. inside-outside) within factors
 - String-valued variables and finite state machines
5. Message approximation and scheduling [20 min., Eisner]
 - Computing fewer messages
 - Pruning messages
 - Expectation Propagation and Penalized EP
6. Approximation-aware Training [30 min., Gormley]
 - Empirical risk minimization under approximations (ERMA)
 - BP as a computational expression graph
 - Automatic differentiation (AD)
7. Software [10 min., Gormley]

3 Instructors

Matt Gormley is a PhD student at Johns Hopkins University working with Mark Dredze and Jason Eisner. His current research focuses on joint modeling of multiple linguistic strata in learning settings where supervised resources are scarce. He has authored papers in a variety of areas including topic modeling, global optimization, semantic role labeling, relation extraction, and grammar induction.

Jason Eisner is a Professor in Computer Science and Cognitive Science at Johns Hopkins University, where he has received two school-wide awards for excellence in teaching. His 90+ papers have presented many models and algorithms spanning numerous areas of NLP. His goal is to develop the probabilistic modeling, inference, and learning techniques needed for a unified model of all kinds of linguistic structure. In particular, he and his students introduced structured belief propagation (which incorporates classical NLP models and their associated dynamic programming algorithms), as well as loss-calibrated training for use with belief propagation.

References

- Michael Auli and Adam Lopez. 2011. A comparison of loopy belief propagation and dual decomposition for integrated CCG supertagging and parsing. In *Proceedings of ACL*.
- David Burkett and Dan Klein. 2012. Fast inference in phrase extraction models with belief propagation. In *Proceedings of NAACL*.
- Markus Dreyer and Jason Eisner. 2009. Graphical models over multiple strings. In *Proceedings of EMNLP*.
- Jason Naradowsky, Sebastian Riedel, and David Smith. 2012. Improving NLP through marginalization of hidden syntactic structure. In *Proceedings of EMNLP 2012*.
- David A. Smith and Jason Eisner. 2008. Dependency parsing by belief propagation. In *Proceedings of EMNLP*.
- Veselin Stoyanov and Jason Eisner. 2012. Minimum-risk training of approximate CRF-Based NLP systems. In *Proceedings of NAACL-HLT*.