

T5: Variational Inference for Structured NLP Models

David Burkett, Dan Klein

ABSTRACT

Historically, key breakthroughs in structured NLP models, such as chain CRFs or PCFGs, have relied on imposing careful constraints on the locality of features in order to permit efficient dynamic programming for computing expectations or finding the highest-scoring structures. However, as modern structured models become more complex and seek to incorporate longer-range features, it is more and more often the case that performing exact inference is impossible (or at least impractical) and it is necessary to resort to some sort of approximation technique, such as beam search, pruning, or sampling. In the NLP community, one increasingly popular approach is the use of variational methods for computing approximate distributions.

The goal of the tutorial is to provide an introduction to variational methods for approximate inference, particularly mean field approximation and belief propagation. The intuition behind the mathematical derivation of variational methods is fairly simple: instead of trying to directly compute the distribution of interest, first consider some efficiently computable approximation of the original inference problem, then find the solution of the approximate inference problem that minimizes the distance to the true distribution. Though the full derivations can be somewhat tedious, the resulting procedures are quite straightforward, and typically consist of an iterative process of individually updating specific components of the model, conditioned on the rest. Although we will provide some theoretical background, the main goal of the tutorial is to provide a concrete procedural guide to using these approximate inference techniques, illustrated with detailed walkthroughs of examples from recent NLP literature.

Once both variational inference procedures have been described in detail, we'll provide a summary comparison of the two, along with some intuition about which approach is appropriate when. We'll also provide a guide to further exploration of the topic, briefly discussing other variational techniques, such as expectation propagation and convex relaxations, but concentrating mainly on providing pointers to additional resources for those who wish to learn more.

OUTLINE

1. Introduction

1. Approximate inference background
2. Definition of variational inference
3. Structured NLP problem setting, loglinear models
4. Graphical model notation, feature locality

2. Mean Field Approximation

1. General description and theoretical background
2. Derivation of updates for simple two-variable model
3. Structured mean field: extension to joint CRFs
4. Joint parsing and word alignment
5. High level description of other models (Coref, Nonparametric Bayes)

3. Belief Propagation

1. General description and theoretical background
2. Factor graph notation
3. Formulas for messages and beliefs, with joint CRF example
4. Dependency parsing
5. Word alignment

4. Wrap-up

1. Mean Field vs Belief Propagation (i.e. what to use when)
2. Other variational methods
3. Additional resources

BIOS

David Burkett

University of California, Berkeley
dburkett--AT--cs.berkeley.edu

David Burkett is a Ph.D. candidate in the Computer Science Division at the University of California, Berkeley. The main focus of his research is on modeling syntactic agreement in bilingual corpora. His interests are diverse, though, and he has worked on parsing, phrase alignment, language evolution, coreference resolution, and even video game AI. He has worked as an instructional assistant for multiple AI courses at Berkeley and won the Outstanding Graduate Student Instructor award in 2009.

Dan Klein

University of California, Berkeley
klein--AT--cs.berkeley.edu

Dan Klein is an Associate Professor of Computer Science at the University of California, Berkeley. His research includes many areas of statistical natural language processing,

including grammar induction, parsing, machine translation, information extraction, document summarization, historical linguistics, and speech recognition. His academic awards include a Sloan Fellowship, a Microsoft Faculty Fellowship, an NSF CAREER Award, the ACM Grace Murray Hopper Award, Best Paper Awards at ACL, EMNLP and NAACL, and the UC Berkeley Distinguished Teaching Award.