

Lagrangian Relaxation for Inference in Natural Language Processing

Michael Collins

Department of Computer Science
Columbia University
mcollins@cs.columbia.edu

Abstract

There has been a long history in combinatorial optimization of methods that exploit structure in complex problems, using methods such as dual decomposition or Lagrangian relaxation. These methods leverage the observation that complex inference problems can often be decomposed into efficiently solvable sub-problems. Thus far, however, these methods are not widely used in NLP.

In this talk I will describe recent work on inference algorithms for NLP based on Lagrangian relaxation. In the first part of the talk I will describe work on non-projective parsing. In the second part of the talk I will describe an exact decoding algorithm for syntax-based statistical translation. If time permits, I will also briefly describe algorithms for dynamic programming intersections (e.g., the intersection of a PCFG and an HMM), and for phrase-based translation.

For all of the problems that we consider, the resulting algorithms produce exact solutions, with certificates of optimality, on the vast majority of examples; the algorithms are efficient for problems that are either NP-hard (as is the case for non-projective parsing, or for phrase-based translation), or for problems that are solvable in polynomial time using dynamic programming, but where the traditional exact algorithms are far too expensive to be practical.

While the focus of this talk is on NLP problems, there are close connections to inference methods, in particular belief propagation, for graphical models. Our work was inspired by recent work that has used dual decomposition as an alternative to belief propagation in Markov random fields.

This is joint work with Yin-Wen Chang, Tommi Jaakkola, Terry Koo, Sasha Rush, and David Sontag.