

Joint Inference for Natural Language Processing

Andrew McCallum

Department of Computer Science
University of Massachusetts Amherst
Amherst, MA 01002
mccallum@cs.umass.edu

Abstract of the Invited Talk

In recent decades, researchers in natural language processing have made great progress on well-defined subproblems such as part-of-speech tagging, phrase chunking, syntactic parsing, named-entity recognition, coreference and semantic-role labeling. Better models, features, and learning algorithms have allowed systems to perform many of these tasks with 90% accuracy or better. However, success in integrated, end-to-end natural language understanding remains elusive.

I contend that the chief reason for this failure is that errors cascade and accumulate through a pipeline of naively chained components. For example, if we naively use the single most likely output of a part-of-speech tagger as the input to a syntactic parser, and those parse trees as the input to a coreference system, and so on, errors in each step will propagate to later ones: each component's 90% accuracy multiplied through six components becomes only 53%.

Consider, for instance, the sentence "I know you like your mother." If a part-of-speech tagger deterministically labels "like" as a verb, then certain later syntactic and semantic analysis will be blocked from alternative interpretations, such as "I know you like your mother (does)." The part-of-speech tagger needs more syntactic and semantic information to make this choice. Consider also the classic example "The boy saw the man with the telescope." No single correct syntactic parse of this sentence is possible in isolation. Correct interpretation requires the integration of these syntactic decisions with semantics and context.

Humans manage and resolve ambiguity by unified, simultaneous consideration of morphology, syntax, semantics, pragmatics and other contextual information. In statistical modeling such unified

consideration is known as joint inference. The need for joint inference appears not only in natural language processing, but also in information integration, computer vision, robotics and elsewhere. All of these applications require integrating evidence from multiple sources, at multiple levels of abstraction. I believe that joint inference is one of the most fundamentally central issues in all of artificial intelligence.

In this talk I will describe work in probabilistic models that perform joint inference across multiple components of an information processing pipeline in order to avoid the brittle accumulation of errors. I will survey work in exact inference, variational inference and Markov-chain Monte Carlo methods. We will discuss various approaches that have been applied to natural language processing, and hypothesize about why joint inference has helped in some cases, and not in others.

I will then focus on our recent work at University of Massachusetts in large-scale conditional random fields with complex relational structure. In a single factor graph we seamlessly integrate multiple subproblems, using our new probabilistic programming language to compactly express complex, mutable variable-factor structure both in first-order logic as well as in more expressive Turing-complete imperative procedures. We avoid unrolling this graphical model by using Markov-chain Monte Carlo for inference, and make inference more efficient with learned proposal distributions. Parameter estimation is performed by SampleRank, which avoids complete inference as a subroutine by learning simply to correctly rank successive states of the Markov-chain.

Joint work with Aron Culotta, Michael Wick, Rob Hall, Khashayar Rohanimanesh, Karl Schultz, Sameer Singh, Charles Sutton and David Smith.