

Efficient incremental beam-search parsing with generative and discriminative models

Brian Roark

Center for Spoken Language Understanding
OGI School of Science & Engineering, Oregon Health & Science University

Extended Abstract:

This talk will present several issues related to incremental (left-to-right) beam-search parsing of natural language using generative or discriminative models, either individually or in combination. The first part of the talk will provide background in incremental top-down and (selective) left-corner beam-search parsing algorithms, and in stochastic models for such derivation strategies. Next, the relative benefits and drawbacks of generative and discriminative models with respect to heuristic pruning and search will be discussed. A range of methods for using multiple models during incremental parsing will be detailed. Finally, we will discuss the potential for effective use of fast, finite-state processing, e.g. part-of-speech tagging, to reduce the parsing search space without accuracy loss. POS-tagging is shown to improve efficiency by as much as 20-25 percent with the same accuracy, largely due to the treatment of unknown words. In contrast, an ‘islands-of-certainty’ approach, which quickly annotates labeled bracketing over low-ambiguity word sequences, is shown to provide little or no efficiency gain over the existing beam-search.

The basic parsing approach that will be described in this talk is stochastic incremental top-down parsing, using a beam-search to prune the search space. Grammar induction occurs from an annotated treebank, and non-local features are extracted from each derivation to enrich the stochastic model. Left-corner grammar and tree transforms can be applied to the treebank or the induced grammar, either fully or selectively, to change the derivation order while retaining the same underlying parsing algorithm. This approach has been shown to be accurate, relatively efficient, and robust using both generative and discriminative models (Roark, 2001; Roark, 2004; Collins and Roark, 2004).

The key to effective beam-search parsing is comparability of analyses when the pruning is done. If two competing parses are at different points in their respective derivations, e.g. one is near the end of the derivation and another is near the beginning, then it

will be difficult to evaluate which of the two is likely to result in a better parse. With a generative model, comparability can be accomplished by the use of a look-ahead statistic, which estimates the amount of probability mass required to extend a given derivation to include the word(s) in the look-ahead. Every step in the derivation decreases the probability of the derivation, but also takes the derivation one step closer to attaching to the look-ahead. For good parses, the look-ahead statistic should increase with each step of the derivation, ensuring a certain degree of comparability among competing parses with the same look-ahead.

Beam-search parsing using an unnormalized discriminative model, as in Collins and Roark (2004), requires a slightly different search strategy than the original generative model described in Roark (2001; 2004). This alternate search strategy is closer to the approach taken in Costa et al. (2001; 2003), in that it enumerates a set of possible ways of attaching the next word before evaluating with the model. This ensures comparability for models that do not have the sort of behavior described above for the generative models, rendering look-ahead statistics difficult to estimate. This approach is effective, although somewhat less so than when a look-ahead statistic is used.

A generative parsing model can be used on its own, and it was shown in Collins and Roark (2004) that a discriminative parsing model can be used on its own. Most discriminative parsing approaches, e.g. (Johnson et al., 1999; Collins, 2000; Collins and Duffy, 2002), are re-ranking approaches, in which another model (typically a generative model) presents a relatively small set of candidates, which are then re-scored using a second, discriminatively trained model. There are other ways to combine a generative and discriminative model apart from waiting for the former to provide a set of completed candidates to the latter. For example, the scores can be used simultaneously; or the generative model can present candidates to the discriminative model at intermediate points in the string, rather than simply at the end. We discuss these options and their potential

benefits.

Finally, we discuss and present a preliminary evaluation of the use of rapid finite-state tagging to reduce the parsing search space, as was done in (Ratnaparkhi, 1997; Ratnaparkhi, 1999). When the parsing algorithm is integrated with model training, such efficiency improvements can be particularly important. POS-tagging using a simple bi-tag model improved parsing efficiency by nearly 25 percent without a loss in accuracy, when 1.2 tags per word were produced on average by the tagger. Producing a single tag sequence for each string resulted in further speedups, but at the loss of 1-2 points of accuracy. We show that much, but not all, of the speedup from POS-tagging is due to more constrained tagging of unknown words.

In a second set of trials, we make use of what we are calling ‘syntactic collocations’, i.e. collocations that are (nearly) unambiguously associated with a particular syntactic configuration. For example, a chain of auxiliaries in English will always combine in a particular syntactic configuration, modulo noise in the annotation. In our approach, the labeled bracketing spanning the sub-string is treated as a tag for the sequence. A simple, finite-state method for finding such collocations, and an efficient longest match algorithm for labeling strings will be presented. The labeled-bracketing ‘tags’ are integrated with the parse search as follows: when a derivation reaches the first word of such a collocation, the remaining words are attached in the given configuration. This has the effect of extending the look-ahead beyond the collocation, as well as potentially reducing the amount of search required to extend the derivations to include the words in the collocation. However, while POS-tagging improved efficiency, we find that using syntactic collocations does not, indicating that ‘islands-of-certainty’ approaches are not what is needed from shallow processing; rather genuine dis-ambiguation of the sort provided by the POS-tagger.

Acknowledgments

Most of this work was done while the author was at AT&T Labs - Research. Some of it was in collaboration with Michael Collins.

References

- Michael Collins and Nigel Duffy. 2002. New ranking algorithms for parsing and tagging: Kernels over discrete structures and the voted perceptron. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 263–270.
- Michael Collins and Brian Roark. 2004. Incremental parsing with the perceptron algorithm. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*. To appear.
- Michael J. Collins. 2000. Discriminative reranking for natural language parsing. In *The Proceedings of the 17th International Conference on Machine Learning*.
- Fabrizio Costa, Vincenzo Lombardo, Paolo Frasconi, and Giovanni Soda. 2001. Wide coverage incremental parsing by learning attachment preferences. In *Conference of the Italian Association for Artificial Intelligence (AIIA)*, pages 297–307.
- Fabrizio Costa, Paolo Frasconi, Vincenzo Lombardo, and Giovanni Soda. 2003. Towards incremental parsing of natural language using recursive neural networks. *Applied Intelligence*. to appear.
- Mark Johnson, Stuart Geman, Steven Canon, Zhiyi Chi, and Stefan Riezler. 1999. Estimators for stochastic “unification-based” grammars. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 535–541.
- Adwait Ratnaparkhi. 1997. A linear observed time statistical parser based on maximum entropy models. In *Proceedings of the Second Conference on Empirical Methods in Natural Language Processing (EMNLP-97)*, pages 1–10.
- Adwait Ratnaparkhi. 1999. Learning to parse natural language with maximum entropy models. *Machine Learning*, 34:151–175.
- Brian Roark. 2001. Probabilistic top-down parsing and language modeling. *Computational Linguistics*, 27(2):249–276.
- Brian Roark. 2004. Robust garden path parsing. *Natural Language Engineering*, 10(1):1–24.