Computational Linguistics

Quarterly (March, June, September, December)
160pp. per issue
6 3/4 x 10
Founded: 1974
2018 Impact Factor: 1.319
2018 Google Scholar h5-index: 32
ISSN: 0891-2017
E-ISSN: 1530-9312

Finite-State Chart Constraints for Reduced Complexity Context-Free Parsing Pipelines

Brian Roark, Kristy Hollingshead and Nathan Bodenstab

Posted Online November 15, 2012
https://doi.org/10.1162/COLI_a_00109

Computational Linguistics
Volume 38 | Issue 4 | December 2012
p.719-753

Abstract Full Text Authors

We present methods for reducing the worst-case and typical-case complexity of a context-free [illegible]eline via hard constraints derived from finite-state pre-processing. We perform

Author Reprints

Reader Resources

Rights and Permissions
Most Read
Most Cited

More About Computational Linguistics

Metrics

Open Access

O(n) predictions to determine if each word in the input sentence may begin or end a multi-word constituent in chart cells spanning two or more words, or allow single-word constituents in chart cells spanning the word itself. These pre-processing constraints prune the search space for any chart-based parsing algorithm and significantly decrease decoding time. In many cases cell population is reduced to zero, which we term chart cell “closing.” We present methods for closing a sufficient number of chart cells to ensure provably quadratic or even linear worst-case complexity of context-free inference. In addition, we apply high precision constraints to achieve large typical-case speedups and combine both high precision and worst-case bound constraints to achieve superior performance on both short and long strings. These bounds on processing are achieved without reducing the parsing accuracy, and in some cases accuracy improves. We demonstrate that our method generalizes across multiple grammars and is complementary to other pruning techniques by presenting empirical results for both exact and approximate inference using the exhaustive CKY algorithm, the Charniak parser, and the Berkeley parser. We also report results parsing Chinese, where we achieve the best reported results for an individual model on the commonly reported data set.

Forthcoming

Most Read

See More

Lexicon-Based Methods for Sentiment Analysis (14019 times)
Maite Taboada et al.
Computational Linguistics
Volume: 37, Issue: 2, pp. 267-307

Computational Linguistics and Deep Learning (10513 times)
Christopher D. Manning
Computational Linguistics
Volume: 41, Issue: 4, pp. 701-707

Near-Synonymy and Lexical Choice (3658 times)
Philip Edmonds et al.
Computational Linguistics
Volume: 28, Issue: 2, pp. 105-144

(Note that the Most Read numbers are based on the number of full text downloads over the last 12 months.)

Most Cited

See More

Lexicon-Based Methods for Sentiment Analysis (436 times)
Maite Taboada et al.
Computational Linguistics
Volume: 37, Issue: 2, pp. 267-307

A Systematic Comparison of Various Statistical Alignment Models (174 times)
Franz Josef Och et al.
Computational Linguistics
Volume: 29, Issue: 1, pp. 19-51

Opinion Word Expansion and Target Extraction through Double Propagation (147 times)
Guang Qiu et al.
Computational Linguistics
Volume: 37, Issue: 1, pp. 9-27

(Note that the Most Cited numbers are based on Crossref's Cited-by service and reflect citation information for the past 24 months.)