Monte Carlo inference for probabilistic context-free grammars

Wilker Aziz and Milos Stanojevic

April 24, 2015

Description

Probabilistic context-free grammars (PCFGs) are vastly used in natural language processing with applications such as parsing and machine translation. Inference under PCFGs typically requires running a cubic time algorithm (the CKY algorithm) either to compute expectations (necessary in learning) or to make decisions (e.g. finding a good parse). For large PCFGs, a complete run of CKY is prohibitively slow. The simplest approximate inference technique involves a pruned run of CKY which discards parses deemed unlikely early on in the search. However, arbitrary pruning introduces arbitrary bias harming learning (biased expectations) and inference (local optimum). In this project, you will deal with a different class of approximate inference methods, namely, Monte Carlo (MC) and Markov chain Monte Carlo (MCMC) techniques. The principle is to simulate a complex target distribution and reason over a tractable statistical sample of the space of solutions.

Organisation

The project starts with a crash course on important algorithms for PCFGs.

- parsing algorithms such as CKY (Younger, 1966), and EARLEY parsing (Earley, 1970), particularly, their generalisations to weighted intersection;

- semiring parsing (Goodman, 1999) and INSIDE-OUTSIDE algorithm (Lari and Young, 1990);

- decision rules (e.g. Viterbi, MBR).

The crash course is followed by a lecture on statistical sampling where special attention will be given to the techniques we will investigate in the rest of the project.

The next phase of the project has the format of a reading group. First, we will discuss exact sampling from a PCFG (Chappelier and Rajman, 2000) which can be done trivially provided an unpruned chart can be produced. We will then discuss two MCMC techniques that allow samples to be obtained without ever instantiating a complete chart (Bouchard-côté et al., 2009; Blunsom and Cohn, 2010).

The final part of the project (which amounts to half the program) involves coding in hackathon-style sessions. We will implement at least one of the MCMC techniques discussed in the reading group (depending on the number of participants we might implement both techniques) and move towards a novel importance sampler (Neal, 2001) which addresses some of the issues with previous work.

Prerequisites

Interest in statistical parsing and machine translation, optimisation and sampling for large and complex discrete distributions. The ideal student will be comfortable with context-free grammars, dynamic programming and statistics. This project requires good Python programming skills.

Assessment

In order to pass, participants must attend all meetings (lectures, reading group, and hackaton) and contribute to a fruitful research experience. The final grade will be the geometric average of the following three components (uniformly weighted). Crash course: exercises and programming assignment. Reading group: presentation in which the student discusses one of the papers. Hackaton: quality and performance of code.

References

Blunsom, P. and Cohn, T. (2010). Inducing synchronous grammars with slice sampling. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 238–241, Los Angeles, California. Association for Computational Linguistics.

Bouchard-côté, A., Petrov, S., and Klein, D. (2009). Randomized pruning: Efficiently calculating expectations in large dynamic programs. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., and Culotta, A., editors, *Advances in Neural Information Processing Systems 22*, pages 144–152. Curran Associates, Inc.

Chappelier, J.-C. and Rajman, M. (2000). Monte-Carlo sampling for NP-hard maximization problems in the framework of weighted parsing. In Christodoulakis, D., editor, *Natural Language Processing — NLP 2000*, volume 1835 of *Lecture Notes in Computer Science*, pages 106–117. Springer Berlin Heidelberg.

Earley, J. (1970). An efficient context-free parsing algorithm. *Commun. ACM*, 13(2):94–102.

Goodman, J. (1999). Semiring parsing. *Computational Linguistics*, 25(4):573–605.

Lari, K. and Young, S. J. (1990). The estimation of stochastic context-free grammars using the inside–outside algorithm. *Computer Speech and Language*, 4:35–36.

Neal, R. M. (2001). Annealed importance sampling. *Statistics and Computing*, 11(2):125–139.

Younger, D. H. (1966). Context-free language processing in time n^3. In *Proceedings of the 7th Annual Symposium on Switching and Automata Theory (swat 1966)*, SWAT '66, pages 7–20, Washington, DC, USA. IEEE Computer Society.