Cloud Computing Project Proposal

Latent-structure conditional random fields for machine translation

Michael Subotin
Department of Linguistics, University of Maryland
3126 AV Williams Building

1. Problem

Many training approaches for statistical machine translation have been proposed. However, a tractable formulation of parameter estimation as a single optimization problem has been elusive. This project, described in more detail in [1], builds on practical insights underlying state-of-the-art translation methods and theoretical insights of machine learning literature to formulate latent-structure conditional random fields, a novel model that makes it possible to generalize core algorithms of word alignment and hierarchical phrase-based translation [2] and subsume them into a unified and rigorous probabilistic framework. Under the proposed framework parameter estimation for alignment and translation is performed by optimizing weights for large numbers of features, including both indicator function for rules in a grammar and millions of other features that can model arbitrarily rich dependencies on source sentences and their annotation. Language model scores, estimated from large monolingual corpora by standard algorithms, are incorporated as features into the model via a novel generalization of the inside-outside algorithm, which leads to the trade-off between the language model and translation model components being optimized at the level of millions individual weights rather than by a handful of interpolation coefficients as it is done in the mainstream paradigm. The proposal specifies several methods of approximation and dimensionality reduction to control computational complexity at various stages of parameter estimation. Pilot experiments with a piecewise estimate of the translation model have yielded promising results.

2. Resources

The proposed approach is attractive in part because its implementation can capitalize extensively on existing state-of-the-art software. Thus, translation with a trained model can be performed via a simple modification of Chiang's decoder [2]. Training, likewise, can make both conceptual and physical use of existing code, including efficient optimization algorithms [4-6].

3. The MapReduce Perspective

In common with several other learning algorithms [7] parameter estimation in latent-structure conditional random fields can efficiently utilize parallel computation. For all standard optimization algorithms the cost of training is dominated by evaluations of the objective function and its gradient, both of which are additive with respect to training examples. Thus, the Map function should evaluate these quantities while the Reduce function adds the results. For the special case of the piecewise estimate without unary factors [1,3] the problem decomposes into independent optimization sub-problems, so that training can be fully parallelized. For online and sampling-based algorithms, which update the parameter vector based on values computed from portions of the training set, Map-Reduce can be applied to batches of training examples.

4. Interesting Extensions

Although the proposed algorithms synthesize the insights of machine translation and machine learning to make parameter estimation for the proposed model tractable, the uncommon computational challenges of translation for resource-rich language pairs make global training impractical without intelligent use of parallel computing resources. The opportunities provided by the Cloud Computing initiative and the proposed framework, which establishes close connections between parameter estimation for machine translation and optimization problems extensively studied in many other fields, open the road to a broad range of exciting developments that would remain unexplored otherwise.

5. REFERENCES

[1] Subotin, M. Conditional random fields for machine translation. Generals paper draft, 2007.

[2] Chiang, D. Hierarchical phrase-based translation. 2007. Computational Linguistics 33(2):201–228.

[3] Sutton, C. and Minka, T. Local Training and Belief Propagation. Microsoft Research Technical Report TR-2006-121. 2006.

[4] Jones. E, Oliphant. T, Peterson P. and others. SciPy: Open Source Scientific Tools for Python, 2001.

[5] Andrew G. and Gao J. Scalable training of L1-regularized log-linear models. ICML, 2007.

[6] Collins M., et al. Exponentiated Gradient Algorithms for Conditional Random Fields and Max-Margin Markov Networks. Journal paper in submission.

[7] Chu C.-T. et al. Map-Reduce for Machine Learning on Multicore. In NIPS 19, 2007.