Michael Auli - Research Statement

Overview

My research is motivated by the prospect of computers understanding language. This requires machines to be able to analyse the syntactic and semantic structure of a sentence. The predominant approach to natural language processing (NLP) is the use of probabilistic models for which the parameters can be learned from annotated training examples. A major challenge presents itself in the desire to capture important but complex relationships in language. This results in ever more complicated models which makes inference and optimisation less tractable.

My work is about making complex models in NLP practical via using approximation methods such as loopy belief propagation which computes approximate marginals that can be used for both training and inference. I believe that being aware of the effectiveness of approximations relative to the exact solution can help us to reason better about our models and will ultimately lead to the design of better models. I used these insights in improving a parser to achieve the best reported results in CCG parsing to date (increasing labelled F-measure by 1.6%; reducing error-rate by 13%) while maintaining high efficiency (17ms per sentence).

Combined Models made Practical

My thesis deals with the shortcomings of a pipeline for Combinatory Categorial Grammar (CCG) parsing: In a first step, individual words are labelled with syntactic types such as determiners and nouns using a sequence model. A second step builds syntactic structures such as noun phrases or verb phrases on top of these types using a parsing model. This increases speed since the first step restricts the second step. A downside is that errors are propagated up the pipeline, making it impossible for the second step to recover from errors made in the first step.

I demonstrate that the components of the CCG parsing pipeline can be integrated into a combined model. The resulting complexity can be is tackled with either of two methods: Dual decomposition repeatedly solves individual sub-problems of the original problem and combines them into a global solution that is exact under certain circumstances. Second, loopy belief propagation computes approximate marginals via message passing which can be used for parsing but which lead to entirely approximate solutions.

I have shown that a combined model can be made practical using these methods while achieving significantly better accuracy. Furthermore, I verified that the accuracy of approximate inference is on par with exact inference for combined CCG parsing; this has very favourable implications on practical parsing since the approximation does not significantly change speed (Auli and Lopez, 2011a).

Efficient Search

Natural language is inherently ambiguous which typically results in a vast number of possible structures for every sentence. Computers are quickly overwhelmed by this ambiguity despite that most of these structures are highly unlikely, whereas humans can tackle this task effortlessly.

The statistical approach allows us to search for the most likely structure but doing this efficiently is crucial for speed. The most popular parsing algorithm proceeds in a breadth-first fashion. I adopted A*, a heuristic depth-first algorithm, to CCG parsing. A* processes sub-solutions based on their probability and a heuristic estimate of the entire solution instead of blindly searching the entire space. I found that it is not only useful for exact search but it can also improve the efficiency of the approximate pipeline approach which prunes part of the search space (Auli and Lopez, 2011b).

Task-specific Optimisation

Optimisation of NLP models towards their evaluation metric is very successful in machine translation. I adopted the same approach for statistical parsing using the softmax-margin objective which minimizes a bound on expected risk and allows integration of task-specific loss functions that can encode metrics. I used a novel algorithm to optimise the evaluation criterion in CCG parsing exactly as well as an approximation. This leads to substantial gains in parsing accuracy at no loss in efficiency (Auli and Lopez, 2011c).

Surprisingly, a comparison between the methods reveals that the approximation achieves comparable accuracy despite its simplicity. This allows the training of highly accurate models at negligible additional computational cost. Furthermore, it presents another example of a problem in which we can be sure that it is not an approximation that hampers accuracy.

Model Analysis in Machine Translation

My early PhD work addressed the following question: How do the translations of the two predominant translation models differ? Phrase-based models translate continuous chunks of text with rules such as “der Baum → the tree”. Hierarchical models translate discontinuous chunks of text to capture constructs such as the French “ne X pas → not X”. I asked the question of how the search spaces of the models differ from each other and found that they are highly overlapping despite well known structural differences (Auli et al., 2009).

Future Directions: Tackling More Complex Models

The recent introduction of dual decomposition into the NLP community has sparked a plethora of applications to various problems ranging from parsing to word alignment. However, the well-known trade-off between optimality and efficiency still applies: exactness comes at the price of efficiency. My work has shown that exactness is not necessarily required to achieve high accuracy. I am therefore very interested in applying approximation methods to other areas, including problems where exact methods have been successful, in order to improve efficiency. More concretely, I would like to apply approximate inference methods to machine translation which has recently been tackled with exact inference using dual decomposition.

References

Auli, M. and Lopez, A. (2011a). A Comparison of Loopy Belief Propagation and Dual Decomposition for Integrated CCG Supertagging and Parsing. In *Proc. of ACL.*

Auli, M. and Lopez, A. (2011b). Efficient CCG Parsing: A* versus Supertagging. In *Proc. of ACL.*

Auli, M. and Lopez, A. (2011c). Training a Log-Linear Parser with Loss Functions via Softmax-Margin. In *Proc. of EMNLP.*

Auli, M., Lopez, A., Koehn, P., and Hoang, H. (2009). A systematic analysis of translation model search spaces. In *Proc. of WMT.*