Evaluation of D. Marecek: 'Unsupervised dependency parsing'

David Marecek presents an interesting new dependency model for unsupervised dependency parsing and evaluates it against a large collection of syntactically annotated corpora from different languages. The model consists of four sub-models: an edge model, a fertility model, a distance model and a reducibility model. The first sub-model is completely standard, but while some of the intuitions behind the three other sub-models are also shared by many authors (e.g. Spitkovsky et al., 2009 and Headden III et al., 2009), their implementations and combination are novel and point in an interesting direction. Parameter estimation is done using Gibbs sampling, and evaluation is done on a large collection of manually annotated corpora, including the widely used CoNLL 2006 and CoNLL 2007 datasets. The descriptions of the dependency models, the sampling technique and the empirical evaluation are adequate. In sum, the thesis presents a very interesting contribution to research on unsupervised parsing and certainly fulfills the requirements for a PhD thesis in computation linguistics.

Critical remarks

Assumptions

a) D. Marecek states that "in unsupervised parsing, we do not allow ourselves to use any kind of linguistic rules" (p. 7). This is not completely true. In recent years several authors have proposed models that rely on universal dependency rules (Naseem et al., 2010; McDonald et al., 2011). Moreover, it is not clear (to me) why assuming that frequency is important for fertility (allowed; p. 41) is more innocent than assuming that "roots of dependency trees are often verbs" (forbidden; p. 7) (if you assume information about parts of speech). In the real-world scenario where we want to parse a language X for which no syntactically annotated corpora exist, one would think that anything goes.

b) In the same vein, D. Marecek follows Gillenwater et al. (2011) in tuning his model on English development data. Why not tune the model on all-but-one languages - which would be the obvious choice in a real world scenario?

c) Finally, D. Marecek mentions the possibility of relying on annotated corpora for other languages X' to induce a model of a language X (Sgaard, 2011; McDonald et al., 2011; Naseem et al., 2012). He dismisses this option as being "burdened by linguistic conventions" (p. 17). However, intrinsic evaluation of unsupervised parsers is already burdened by linguistic conventions, and most work on cross-language adaptation uses several source languages and thereby reduces the influence of unconventional annotation.

Method

a) D. Marecek states that his contribution is three-fold: an application of a sampling technique, a dependency model, and an empirical evaluation. The empirical evaluation is thorough, but relatively standard, and Gibbs sampling has been used in Johnson et al. (2007).
b) The choice of Gibbs sampling could be motivated in more detail. D. Marecek states that EM is easily trapped in local optima when applied to complex models (p. 20), but EM with random restarts often performs as well as Gibbs sampling and is usually faster.
c) Finally, D. Marecek mentions three decoding algorithms (pp. 58-61). Two of them are said to have been implemented (p. 68) - but which one was chosen? It is not clear why a non-projective parsing algorithm was chosen for the second method.

Empirical evaluation I have no comments on the empirical evaluation; except maybe that results could have been compared to more results reported in the literature, and that no results are reported for the Turkish CoNLL 2006 data in Table 7.8 (Gillenwater et al., 2011, include this dataset).

Language The thesis is generally well-written, but some technical terms are used incorrectly in a few places, e.g. "bad convergence to a global optimum" (p. 12).

In sum, the thesis presents a very interesting contribution to research on unsupervised parsing and certainly fulfills the requirements for a PhD thesis in computation linguistics.

Anders Søgaard, Assoc.Prof., PhD