Posudek diplomové práce
Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Abhishek Agrawal
Název práce	Eye-tracking features in syntactic parsing
Rok odevzdání	2020
Studijní program	Informatika **Studijní obor** Matematická lingvistika

Autor posudku	Rudolf Rosa **Role** vedoucí
Pracoviště	Ústav formální a aplikované lingvistiky

Text posudku:

In his thesis, Abhishek Agrawal studies the potential utility of eye tracking features for syntactic parsing. He uses the Dundee corpus which contains sentences manually annotated with dependency syntax and accompanied by eye tracker measurements of 10 people reading the sentences. He uses two existing dependency parsers, which he slightly modifies to enable inclusion of the eye tracking features. He then performs a range of experiments to optimize several hyperparameters of the parsing setup (most importantly the subset of the eye tracking features to use and the method to normalize their values) on a development set. The best-performing setups are then evaluated on a test set, comparing the parser enriched with eye tracking features to a vanilla baseline. The author experiments with two settings of the parser, either using a standard lexicalized parser, or a delexicalized parser which may be more appropriate especially in low-resource settings.

The results show that for a standard lexicalized parser, no clear benefit of including the eye-tracking features can be demonstrated, as the small observed increase in accuracy is not statistically significant. However, for the delexicalized parser, inclusion of eye tracking features leads to smallish but statistically significant improvements in accuracy, showing that eye tracking features have some relevance towards syntactic parsing. Eye tracking features thus might be used to improve parsing accuracies especially in low-resource scenarios; as the author notes, eye trackers are becoming quite cheap recently and even mobile phones can be used to provide such measurements.

I believe that the author completed an interesting piece of work, investigating an interesting idea and finding trustworthy mildly positive results. I believe his approach is solid, is described in sufficient detail in the thesis, with adequate references to relevant literature, and the experiments are properly evaluated. The approach is also well motivated, as eye tracking features had previously been found to have some relevance to other NLP tasks, such as part of speech tagging and named entity recognition; previous works had also indicated potential utility of eye tracking features for syntactic parsing, but this had not been studied extensively. The presented work thus tries to partially fill this gap and provide some more reliable answers in this area, as well as an investigation of the possible setups with a suggestion of optimal hyperparameters backed by experiments on development data.

I have one serious concern though. After carefully investigating the detailed evaluation in chapter 5, I realized that the part-of-speech tagset used in the Dundee corpus is the actually the Google universal POS tagset, not the Universal Dependencies tagset (this is not stated accurately in the thesis). While these tagsets are very similar, they differ in the UD tagset distinguishing several categories which the Google tagset groups under one label, such as nouns and proper nouns (NOUN, PROPN), full verbs and auxiliary verbs (VERB, AUX), or coordinating and subordinating conjunctions (CONJ, SCONJ). While this may not seem to be of a high importance, it is worth noting that these distinctions were made in UD partially due to some of these distinctions (AUX/VERB, CONJ/SCONJ) having clear syntactic relevance. In case of a delexicalized parser, which only has access to the POS label, the distinction between an auxiliary verb and a full verb can actually help the parser a lot. I thus propose an alternative hypothesis to the one presented by the author: I hypothesize that maybe the eye tracking features are not by themselves significantly syntactically relevant, but they only help disambiguate some of the part of speech categories into

more fine-grained ones, such as distinguishing full verbs from auxiliary verbs and distinguishing coordinating conjunctions from subordinating conjunctions, which then in turn helps the parser; however, this is more of a morphological disctinction rather than a syntactic distinction (even though the boundary may not be absolutely clear here). This hypothesis is corroborated by the detailed analyses in tables 5.5, 5.6, 5.7 and 5.8, showing that the incorporation of eye tracking features into parsing brings particularly high improvements in accuracy for parsing verbs and conjunctions (VERB, CONJ) and for distinguishing the verbal relations (root, aux, cop) and conjuncts (conj). I feel that had the dataset used the more modern UD POS labels which feature more fine-grained distinctions in these categories, the observed improvements brought by eye tracking features might considerably diminuish. Thus, while I do not question the obtained results, I do question the conclusions drawn from them, as I believe that this alternative explanation of the results is also possible.

To conclude, I still believe that the thesis is solid and I do recommend it for defence, but I worry that its main findings need to be taken with a grain of salt, and should withstand further investigation before their possible publishing in a research paper.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 2. 9. 2020

Podpis: