

COMP 550: READING ASSIGNMENT 2

JAN TIEGGES

OCTOBER 24, 2023

The paper *Grammar as a Foreign Language* by Vinyals et al. (2015) presents a domain agnostic attention enhanced sequence-to-sequence model for syntactic constituent parsing. The authors experimented with two different sequence-to-sequence models, one of which had an attentional mechanism and the other did not. Trained on a small human-annotated dataset, the attention model matches the performance of standard parsers on standard syntactic constituent parsing datasets and achieves state-of-the-art performance when trained on a large synthetic dataset that the authors created using existing parsers. In doing so, the model is extremely data efficient and fast. The paper contains a detailed analysis of the model performance compared to other state-of-the-art parsers while also providing insight into various aspects relevant to training these models (e.g. dropout, pre-training). In addition, the analysis of the attention matrix gives a visual insight into the functioning of the model.

In contrast to their approach, the CYK algorithm is a traditional approach to constituency parsing based on probabilistic context-free grammars (CFGs). The paper clearly demonstrates the advantages of the neural approach, as it achieves comparable or even better results than these traditional approaches, while being significantly more data efficient and faster. The computational cost of traditional parsers is typically cubic in sentence length, and even recent advances improve this only slightly. However, the neural approach is less interpretable as it does not generate parse trees that can be easily analyzed. On the other hand, this allows for easier transfer to other tasks. Lastly, the experiments show that these models only reach their true potential when they are trained on large data sets.

The use of a large text corpus allows the model to learn from a variety of examples, which helps it to better understand the context in which words and expressions are used. This is related to concepts from our lecture where little to no domain-specific knowledge is given and instead structures are learned from the data, such as in the case of bootstrapping. We also talked about the effect of context on word meaning, which also matters here, where a large corpus allows for learning of meaning in different contexts, allowing the model to generalize better to unseen data.

The paper is well structured and clearly demonstrates the advantages of neural approaches over traditional approaches. Moreover, new state-of-the-art results are obtained on common syntactic constituent parsing dataset. The detailed description of model architecture and training procedures as well as the creation of the synthetic dataset enable reproducibility and further experiments.

However, it would have been interesting to perform a detailed error analysis to identify the areas of improvement of the method. In addition, discussing the limitations of the method and providing directions for future research would have been valuable. Also, the method was only used for syntactic constituent analysis, and other NLP tasks were not considered. Finally, no experiment was conducted with standard parsers trained on the large data sets, which is a missed opportunity for a fair comparison regardless of resource consumption.