Augmenting Deep Learning with Syntactic and Semantic Annotation

Shalom Lappin*

University of Gothenburg, King's College London, and Queen Mary University of London

*Joint work with Adam Ek and Jean-Philippe Bernardy, University of Gothenburg

In recent years deep neural networks (DNNs) of various architectures have achieved impressive results across a large set of natural language processing tasks. These include, among others, machine translation, paraphrase identification, graphic image description, question-answering, sentiment analysis, and natural language inference. Attempts to enrich DNNs with explicit syntactic and semantic representations, either in the training data, or through biasing the learning process in the internal design of the network, have yielded mixed results that vary with the task. Gupta and Zhang (2018) provide evidence that Tree-LSTMs with the addition of progressive attention improve the performance of Long Short Term Memory (LSTM) networks on the recognition of semantic relatedness. Williams et al. (2018) show that latent tree learning Recurrent Neural Networks (RNNs) outperform Tree-LSTMs on several semantic relatedness applications and the Stanford Natural Language inference task. However, the parse trees that these RNNs construct are not consistent across sentences, and they do not resemble the structures posited in formal syntactic or semantic theories. Bernardy and Lappin (2017) find that substituting part of speech sequences for rich lexical embeddings in training data decreases the accuracy of an LSTM in identifying syntactic subject-verb agreement across sequences of intervening subject candidates.

In this talk I present joint work on the effect of adding syntactic and semantic markers to LSTM training data for the prediction of sentence acceptability task, discussed in Lau et al. (2017). We compare the performance of an LSTM language model (LM) trained on raw Wikipedia text with that of LSTM LMs trained on text annotated with universal semantic category markers (Abzianidze et al., 2017), with universal syntactic dependency roles (Nivre et al., 2016), and with syntactic dependency tree depth features (Nivre et al., 2016), respectively. We test the LMs for perplexity, and we assess their performance in predicting the mean human gradient acceptability judgments that Lau et al. (2017) collected with crowd sourcing for 2500 British National Corpus sentences. Some of these sentences were subjected to round trip machine translation through other languages in order to introduce infelicities of different kinds into the test suite.

We found that syntactic role and depth markers reduce LM perplexity, but semantic features increase it. More significantly, the LSTM LM trained on non-annotated text outperformed the other LMs in the prediction of human acceptability judgments. This result supports the view proposed in Bernardy and Lappin (2017), and Gulordava et al. (2018) that, for at least some NLP tasks, LSTMs learn a considerable amount of syntactic structure, and that learning is achieved through distributed lexical representations. Our results suggest that the addition of explicit syntactic and semantic representations to training data can interfere with this learning. An important question for future research is to determine the extent to which the way in which humans acquire knowledge of the syntactic and semantic properties of their language resembles the processes that we are discovering in deep learning.