

## **Daniël de Kok and Tobias Pütz: Task-specific pre-training for dependency parsing**

Context-sensitive word representations (ELMo, BERT, XLNet) have provided large improvements over co-occurrence based word embeddings across various tasks. Context-sensitive word representations use deep models that are trained on an auxiliary objective, such as language modeling or masked word prediction. An important criticism of context-sensitive word representations is the excessive amount of computation time that running both training and inference with such models require.

In this talk, we discuss an alternative to such word representations in the form of task-specific pretraining. Task-specific pretraining uses the objective of a specific task and a corpus that was automatically annotated using a (weaker) model for that task. We show that in dependency parsing, task-specific pretraining results in models that achieve labeled attachment scores similar to models that use context-sensitive word representations. However, it achieves such performance with models that are smaller, faster, and require far less compute to run training and inference.

We will also attempt to address the question what task-specific pre-trained models actually learn and how that differs from models that do not use pretraining through model introspection in the form of probing tasks and gradient analysis.