Natural Language Processing with Small Feed-Forward Networks

We show that small and shallow feedforward neural networks can achieve near state-of-the-art results on a range of unstructured and structured language processing tasks while being considerably cheaper in memory and computational requirements than deep recurrent models. Motivated by resource-constrained environments like mobile phones, we showcase simple techniques for obtaining such small neural network models, and investigate different tradeoffs when deciding how to allocate a small memory budget.

Authors: Mr JAN, A. Botha (Google Inc); Mr ANTON, Bakalov (Google Inc.) Presenters: Mr JAN, A. Botha (Google Inc); Mr ANTON, Bakalov (Google Inc.)