

Text Classification with Recurrent Neural Network Based Language Models

Jakab Buda

Using different machine learning models in almost any applied statistical field is mainstream today – and this holds for natural language processing (NLP), too. But this field is changing so fast, and new methods are constantly popping up that it is hard to keep track or find a systematic review on the advantages and disadvantages of the tools used. The presentation will discuss whether using multiple language models (LMs) to classify texts has advantages over using usual classifier architectures. After a brief overview on LMs and on why recurrent neural networks, especially those with long-short term memory cells, are particularly appropriate tools for NLP tasks, I will present some experimental results on classifying texts by the gender of their author with different classifier setups, and I will discuss the limitations of such models.