On flat cues and deep learning: the benefits of Convolutional Neural Networks for (historical) linguistic research

The past decade has witnessed a massive take-over by Deep Learning algorithms in virtually all domains of Machine Learning, including Natural Language Processing. As far as language is concerned, the bloom of Deep Learning has culminated in the emergence of neural language models that yield impressive results on a variety of applied linguistic tasks (e.g. GPT, BERT). While these models are powerful tools, they are not always applicable in settings where computational power and training data are limited. In such settings, tailored neural approaches might be more desirable than generic ones. The present paper showcases one such tailored model architecture. The approach revolves around Convolutional Neural Networks (CNNs), a type of neural network that recently made its transition from image recognition to NLP (Vanni et al. 2018).

The benefits of CNNs for linguistic research will be shown by means of two case studies on Early Modern English, each one highlighting a different application area of the algorithms. The first case study is practical in nature and illustrates how CNNs can be tweaked to retrieve (formally ill-defined) cleft constructions from raw corpus data in a semi-automatic fashion. The second case study shows how CNNs, as radical implementations of analogical processing, can contribute to the study of linguistic alternations. By generating token-level sentence embeddings, the networks allow to retrieve not only the contexts a competitor actually occurred in, but also the contexts the competitor could have occurred in, but did not. This, in turn, provides insight in the context-sensitive dynamics between the competitors in a fully bottom-up way.

The versatility of CNNs for linguistic research stems from their algorithmic design. CNNs hold the middle ground between full-fledged neural language models and multinomial classifiers commonly used in the study of linguistic alternations. Just like neural language models, CNNs know how all words in a language interact. Just like multinomial classifiers, their goal is to model just one alternation rather than a language as a whole. Their specific focus makes them considerably lighter than generic neural language models, both in terms of computational complexity as in terms of data greediness. In addition, the possibility to open up the networks and visualise what they have learned enables users to monitor and debug the models during training and at the same time provides direct insight in the factors that govern the classification problem under scrutiny.

References

Vanni, Laurent, Mélanie Ducoffe, Damon Mayaffre, Frédéric Precioso, Dominique Longrée, Veeresh Elango, Nazly Santos, Juan Gonzalez, Luis Galdo & Carlos Aguilar. 2018. Text Deconvolution Saliency (TDS): a deep tool box for linguistic analysis. *56th Annual Meeting of the Association for Computational Linguistics*, July 2018, Melbourne.

Model Papers

GPT: Radford, Alec, Karthik Narasimhan, Tim Salimans & Ilya Sutskever. 2018. Improving language understanding by generative pre-training. *Preprint at OpenAI.*

Bert: Devlin, Jacob, Ming-Wei Chang, Kenton Lee & Kristina Toutanova. 2018. Bert: pre-training of deep bidirectional transformers for language understanding. *Preprint at ArXiv 810.04805*.