Bern, 01.04.2021

Bern Data Science Day 2021 - April 23

UNIVERSITÄT BERN

Contribution – Visualizing Language Models

Keywords: Digital Humanities, Machine Learning, Natural Language Processing

Sebastian Flick, Tobias Hodel, Reinhard Priber, Christa Schneider, Jonas Widmer
Digital Humanities/Walter Benjamin Kolleg, Mittelstrasse 43, 3012 Bern
Mail: givennamen.familyname@wbkolleg.unibe.ch

Abstract

Language models (e.g. character embeddings) are essential to succeed in NLP tasks. Especially when it comes to Part-of-Speech and Named Entity Recognition, sequence taggers result in more precise models when supported by adequate language models already. Since the advent of word2vec and large transformer-based language models (such as BERT [Devlin et al. 2019] or GPT-3 [Brown et al. 2020]) a variety of specialized and fine-tuned language models is currently available. Despite the widespread use and the necessity when it comes to specific model training (e.g. for language entities with only sparse data), our understanding of the models themselves is limited at best. In order to strengthen our knowledge about language models and to start the process of reflecting them, we propose to analyze language models based on calculated vectors. In order to make language models also visually approachable we furthermore reduced dimension by applying PCA (t-SNE is also foreseen) to plot the results in three dimensions. Thus, we start to understand how embeddings are being constructed and what they tell us about the underlying (grammatical) system of a certain language.

Through the use of current frameworks [Akbik et al. 2019] which embed tokens (“words”) as parts of sentences, semantic and grammatical context plays an important role. Based on these contexts, homographs can for example be compared, when visualized even physically. This leads us to new discussions and first very simple explanations of the inner functions of language models. As such models are currently implemented in tools (like search engines) and machines (like smartphones) that we use on a daily basis, we can expect to deal with other, even more complex questions very soon.

First results based on a custom-made character language model for 14th-16th century German charters can be found online: https://nlp-hack-4.fdn-dev.iwi.unibe.ch/.

Bibliography:

Akbik, Alan; Bergermann, Tanja; Blythe, Duncan et al.: FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, Minneapolis 2019. DOI: 10.18653/v1/N19-4010.

Brown, Tom B.; Mann, Benjamin; Ryder, Nick et al.: Language Models are Few-Shot Learners, in: arXiv:2005.14165 [cs], 22.07.2020. Online: <http://arxiv.org/abs/2005.14165>.

Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton u. a.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in: arXiv:1810.04805 [cs], 24.05.2019. Online: <http://arxiv.org/abs/1810.04805>.