

Training ELMo from scratch on Saga



Andrey Kutuzov
Language Technology Group
University of Oslo

Are transformers **always** better?

Task: **classification of 7 000 documents into two classes (positive and negative movie reviews, SST-2 dataset).**

Approach: naively averaging all token embeddings from the document, logistic regression, 10-fold cross-validation.

	BERT-base uncased	ELMo trained on the News on Web corpus
Number of parameters	110M	57M
Time to classify	43 seconds	32 seconds
Model size	440 MBytes	223 MBytes
Macro F1	0.835	0.843

Useful links

- Pre-trained ELMo models:
 - <http://vectors.nlpl.eu/repository/>
- How to train your model on Saga:
 - <http://wiki.nlpl.eu/index.php/Vectors/elmo/tutorial>
- Code for TensorFlow < 2.0:
 - https://github.com/lmgoslo/simple_elmo
 - https://github.com/lmgoslo/simple_elmo_training
- Takes about 24 hours to train ELMo on 1 billion words using two NVIDIA P100 GPUs.
- TF 2.0 code is coming.