Pre-trained Language Models

* ELMo: Deep contextualized word representations
  + LSTM; Word-level Language Model; Fixed
  + <http://aclweb.org/anthology/N18-1202>
  + <https://github.com/allenai/bilm-tf>
  + <https://allennlp.org/elmo>
* OpenAI GPT
  + Transformer(decoder); fine-tuning;
  + <https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf>
  + <https://blog.openai.com/language-unsupervised/>
* Flair: Contextual String Embeddings for Sequence Labeling
  + LSTM; Character-level Language Model; Fixed;
  + <http://aclweb.org/anthology/C18-1139>
  + <https://research.zalando.com/welcome/mission/research-projects/flair-nlp/>
  + <https://github.com/zalandoresearch/flair>
* ULMFiT: Universal Language Model Fine-tuning for Text Classification
  + LSTM; two steps fine-tuning;
  + <http://aclweb.org/anthology/P18-1031>
  + <http://nlp.fast.ai/category/classification.html>
* BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding
  + Transformer(encoder); fine-tuning; masked-lm; next sentence prediction
  + <https://arxiv.org/pdf/1810.04805.pdf>
  + <https://github.com/google-research/language/bert>