Homework 11 – Language

Arthur J. Redfern
arthur.redfern@utdallas.edu
Apr 17, 2019

0 Outline

1 Logistics

Assigned: Mon Apr 17, 2019
Due: Mon Apr 24, 2019
Format: PDF uploaded to eLearning

2 Reading

1. Exceedingly strong pre trained embeddings for ~ sentence length inputs are a significant development in natural language processing. Read the BERT blog post and paper illustrating 1 method for architecting a network and creating unsupervised language modeling tasks for training the network

>   Open sourcing BERT: state-of-the-art pre-training for natural language processing
>   https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html
>
>   BERT: pre-training of deep bidirectional transformers for language understanding
>   https://arxiv.org/abs/1810.04805

2. Next, read the GPT / GPT 2 blog posts and papers on the same topic; GPT 2 has been in the news a lot recently

Improving language understanding with unsupervised learning
https://openai.com/blog/language-unsupervised/

Improving language understanding by generative pre-training
https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

Better language models and their implications
https://openai.com/blog/better-language-models/

Language models are unsupervised multitask learners
https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

3 Theory

None

4 Practice

3. Word embedding. Work through the TensorFlow tutorial on Word2Vec

Vector Representations of Words
https://www.tensorflow.org/tutorials/representation/word2vec

4. Language modeling. Work through the TensorFlow tutorial on language modeling that trains a character predicting RNN using the text of Shakespeare's plays

Text generation using a RNN with eager execution
https://www.tensorflow.org/tutorials/sequences/text_generation
https://colab.research.google.com/github/tensorflow/docs/blob/master/site/en/tutorials/sequences/text_generation.ipynb
https://github.com/tensorflow/docs/blob/master/site/en/tutorials/sequences/text_generation.ipynb

5. Language translation. Work through the TensorFlow tutorial on language translation using attention

Neural machine translation with attention

https://colab.research.google.com/github/tensorflow/tensorflow/blob/master/tensorflow/contrib/eager/python/examples/nmt_with_attention/nmt_with_attention.ipynb

6. Sentiment prediction. Work through the TensorFlow tutorial on sentiment prediction for movie reviews using BERT

TensorFlow code and pre-trained models for BERT
https://github.com/google-research/bert

Predicting movie review sentiment with BERT on TF Hub
https://colab.research.google.com/github/google-research/bert/blob/master/predicting_movie_reviews_with_bert_on_tf_hub.ipynb