Deep Learning for NLP 2020
Exercise 07 Solution

June 2, 2020

1 Pingo

- Which of the following statements about recurrent neural networks are true?
  - ☐ Every iteration has its own weight matrices.
  - ☑ Information about previous inputs can be stored via “feedback loops”.
  - ☑ They can process input sequences of arbitrary length.
- What is the main reason for CBOW and skip-gram not being applicable on full sentences?
  - ☐ The training process would be too computationally expensive
  - ☐ There are no sufficiently large corpora with full sentences
  - ☑ The probability for out-of-vocabulary sentences when applying the embeddings would be too high

2 Sentence Embeddings

1. List two reasons why the naïve approach of creating a sentence embedding by averaging word embeddings might cause problems.

   Answer: See lecture slides:

   - averaging does not take word order into account
   - stopwords (articles, pronouns, interjections) need special treatment such as filtering or weighting

2. List four sentence representation techniques covered in the lecture and explain their main principle in one sentence.

   Answer:

   a) Stacked Denoising Auto-Encoder (SDAE) by Hill et al. 2015
   - Use an auto-encoder as an auxiliary task to learn a representation of one sentence without context.

   b) Skip-Thought Vectors by Kiros et al. 2015
   - Use an encoder-decoder to learn a sentence representation based on the preceding and following sentence.

   c) InferSent by Conneau et al. 2017
   - Use a BiLSTM sentence encoder trained on the SNLI corpus (as an auxiliary task).

   d) Paragraph Vectors by Le et al. 2014
   - Learn a paragraph representation which predicts the words occurring in the paragraph.

e) Siamese CBOW by Kenter et al. 2016

- Learn word embeddings so that the average embeddings of nearby sentences are close.

f) Sent2Vec by Pagliardini et al. 2017

- Similarly to Siamese CBOW, learn sentence embeddings which average nicely, but also include bigrams and trigrams.

g) Concatenated Power Mean Embeddings

- Compute the power mean between pretrained word embeddings.

h) BERT by Devlin et al. 2018

- Learn a transformer model in unsupervised fashion using cloze-style auxiliary tasks on a large body of text.