

Deep Learning for NLP 2020

Exercise 06

May 26, 2020

1 Pingo

Try to find the right answer(s) to each question on your own or in a group with your colleagues. The interactive survey will be conducted near the end of the practice class.

- Which of the following statements about sense-disambiguated word embeddings are correct?
 - ☐ They differentiate better between two similarly written words with different meanings.
 - ☐ They differentiate better between two differently written words with the same meaning.
 - ☐ Their use is limited because they require more effort to set up.
- Which of the following statements apply on bilingual embeddings?
 - ☐ They are helpful for training neural networks for languages where few resources exist.
 - ☐ In order to train bilingual embeddings, one always needs parallel texts.
 - ☐ BiVCD relies on correlation analysis.
- Which of the following statements on syntactic embeddings are correct?
 - ☐ Contexts in word2vec do not include information about word order.
 - ☐ The structured skip-gram relies on word contexts of variable length.
 - ☐ ~~Dependency-based embeddings are an attempt at solving the problem of long dependencies.~~

2 Using Word Embeddings

2.1 Contextualized Embeddings

Multilingual BERT is trained by running standard BERT on the concatenation of multilingual text from the largest 100 languages in Wikipedia. Does this yield multilingual contextualized embeddings? Which problems do you see? Explain in up to 5 sentences.

2.2 Handling out-of-vocabulary Words

When using pre-trained word embeddings, one can run into the issue that certain words/tokens are missing from the embedding vocabulary (so called out-of-vocabulary words), i.e. the embedding vector is unknown for these words.

Think of at least three possible solutions and state their advantages and disadvantages.