Prof. Dr. Andreas Hotho, Albin Zehe, Jonas Kaiser
Data Science Chair

03./04.07.2024

9. Assignment in "Machine Learning for Natural Language Processing"

Summer Term 2024

1 General Questions

1. A convolution transforms the input by weighting the input values with a learned filter. A self-attention layer also combines the values that are computed from the input using weights. What are the differences between the two?

2. Explain why we need positional encodings when training a transformer but not for simple multi layer perceptrons or recurrent neural networks!

3. The basic transformer architecture uses attention in two different ways per layer. Explain how they differ and what function each one serves intuitively!

2 Byte-Pair Encoding

Given a (very small) dataset consisting of the words "hello", "hell", "love", and "lovely".

Perform byte pair encoding using these words for 3 steps to create a new vocabulary and rules to encode new words. If there are multiple subtoken pairs with the same count, use the alphabetically first pair.