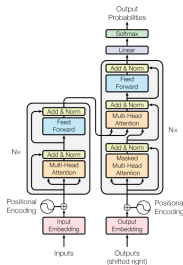


Transformer

BytePair encoding (BPE)



Learning goals

- Understand inner workings of BPE
- Being able to compare BPE to other tokenization approaches

BYTEPAIR ENCODING (BPE)

Data compression algorithm ► Gage (1994)

- Considering data on a *byte*-level
- Looking at pairs of bytes:
 - ❶ Count the occurrences of all byte pairs
 - ❷ Find the most frequent byte pair
 - ❸ Replace it with an unused byte
- Repeat this process until no further compression is possible

BYTEPAIR ENCODING (BPE)

Open-vocabulary neural machine translation ► Sennrich et al. (2016)

- Translation as an open-vocabulary problem
- Word-level NMT models:
 - Handling out-of-vocabulary word by using back-off dictionaries
 - Unable to translate or generate previously unseen words
- Subword-level models alleviate this problem

BYTEPAIR ENCODING (BPE)

Adapt BPE for word segmentation ▸ Sennrich et al. (2016)

- *Goal:* Represent an open vocabulary by a vocabulary of fixed size
→ Use variable-length character sequences
- Looking at pairs of characters:
 - ➊ Initialize the the vocabulary with all characters plus end-of-word token
 - ➋ Count occurrences and find the most frequent character pair, e.g. "A" and "B" (⚠ Word boundaries are **not** crossed)
 - ➌ Replace it with the new token "AB"
- Only one hyperparameter: Vocabulary size
(Initial vocabulary + Specified no. of merge operations)
→ Repeat this process until given $|V|$ is reached

BYTEPAIR ENCODING (BPE)

test

BYTEPAIR ENCODING (BPE)