The authors propose a technique novel technique to segment words into "subword units" based on the Byte Pair Encoding (BPE) algorithm in order to handle open vocabularies in Neural Machine Translation (NMT). Using this technique the authors achieve significant improvements over baseline systems without needing to resort to tricks such as UNK replacement and backoff dictionary alignments.

KeyPoints:

\* Translation of rare words is an open problem. Typically, NMT vocab is limited to fixed 30k-50k words.

\* Intution: Many words are translatable using smaller units than words (e.g. word stems, suffixes, etc).

\* Technique(byte pair encoding) is purely a data pre- and post-processing step. Nothing in the model needs to change.

\* Basic idea behind algorithm: Start with a vocabulary of all character in the text, then iteratively merge the most frequent pair of characters or character sequences. Thus, the final vocabulary size is num\_chars + num\_merge\_operations where num\_merge\_operations is a hyperparameter of the method.

\* Can learn BPE for source and target separately, or joint. Joint BPE has the advantage that words appearing in both source and target are split exactly the same way, making it easier for the model to learn the alignments.

\*Training details:Trained using Groundhog

Model based on Bahdanau's alignment paper

Beam size 12; Adadelta with batch size 80;

Authors try 60k separate and 90k joint BPE vocab for English-German