**Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates**

**Summary**

The research paper presents a simple regularization method, **subword regularization** for NMT (nueral machine translation), with no change to the network architecture. The central idea is to virtually augment training data with on-the-fly subword sampling, which helps to improve the accuracy as well as robustness of NMT models. In addition, for better subword sampling, a new subword segmentation algorithm based on the unigram language model is proposed in this paper. Experiments on multiple corpora with different sizes and languages show that subword regularization leads to significant improvements especially on low resource and open-domain settings. The segmentation algorithm is compared with different techniques i.e BPE (byte pair encoding).

**Strengths**

a new subword segmentation algorithm based on a unigram language model is used, which is capable of outputing multiple subword segmentations with probabilities and is more Flexible then BPE(byte pair encoding) and gives better results than other word segmentation algorithms.

**Observations**

* Subword segmentation with the unigram language model can be seen as a probabilsitic mixture of characters, subwords and word segmentations.
* The unigram language model is more flexible than BPE(byte pair encoding) as it is based on a probabilistic language model and can output multiple segmentations with their probabilities, which is an essential requirement for subword regularization.
* BPE and unigram language model without subword regularization (l = 1) show almost comparable BLEU scores. This is not surprising, given that both BPE and the unigram language model are based on data compression algorithms.
* Subword regularization (l > 1) boosted BLEU scores quite impressively (+1 to 2points) in almost all language pairs. The gains are larger especially in lower resource settings.