**Addressing the Rare Words Problem in Neural Machine translation.**

**Summary:**

In the above mentioned paper, the researches proposed and implemented an effective technique to overcome the inability of existing Neural Machine Translation to correctly translate the very rare words. They used English to French for the test data. The basic area that they tried to cover was Out-Of-Vocabulary problem. The researchers used the deep LSTM (Long short-term Memory) encoder decoder neural network. The treated the NMT model as a black box and train it in 2 ways: i) alignments are produced with an unsupervised aligner. ii) Used alignment links to construct a word dictionary which was used in post-processing step. The Rare word model which was implemented consist of following models:

* Copyable Model--multiple tokens to represent the various unknown words.
* Positional All Model (PosAll)—complete alignments between source and target sentences.
* Positional Unknown Model (PosUnk)—unkposd token to denote a) words are unknown b) relative position d w.r.t aligned source.

**Strengths:**

* This method provides a substantial improvement of **2.8 BLEU** points over an equivalent NMT that does not use their technique.
* With **37.5 BLEU** points, there model was the first to surpass the best result achieved in **WMT’14** contest task.
* First one to address the rare words problem and to some extent handled it well.
* If the word is not known, it gets replaced with UNK(unknown) but also the relative position of the word is also showed in the base like **unkposd**.
* It can be used for any NMT system besides the deep LSTM model they used as their base.

**Observations:**

* The observations made in their model w.r.t to some changes are shown in below table:

|  |  |  |  |
| --- | --- | --- | --- |
| *end-to-end Neural Machine Translation* | | | |
| **System** | **Vocab** | **Corpus** | **BLEU** |
| Single LSTM with 4 layers | 40K | 12M | 29.5 |
| Single LSTM with 4 layers + PosUnk | 40K | 12M | **31.8 (+2.3)** |
| Single LSTM with 6 layers | 40K | 12M | 30.4 |
| Single LSTM with 6 layers + PosUnk | 40K | 12M | **32.7 (+2.3)** |
| Ensemble of 8 LSTMs | 40K | 12M | 34.1 |
| Ensemble of 8 LSTMs + PosUnk | 40K | 12M | **36.9 (+2.8)** |
| Single LSTM with 6 layers | 80K | 36M | 31.5 |
| Single LSTM with 6 layers + PosUnk | 80K | 36M | **33.1 (+1.6)** |
| Ensemble of 8 LSTMs | 80K | 36M | 35.6 |
| Ensemble of 8 LSTMs + PosUnk | 80K | 36M | **37.5 (+1.9)** |

**Weakness:**

* The main weakness of PosAll model was that it doubles the length of the target sentence, which makes learning more difficult and slows the speed with the factor of 2.
* The main weakness in the Copyable model was its inability to translate the unknown words which are aligned to the known words.
* For English to French translation, this model gives an improvement of 2.8 BLEU points if and only if vocabulary is relatively small.