SEQUENCE TO SEQUENCE MACHINE TRANSLATION AND GENDER BIAS ANALYSIS

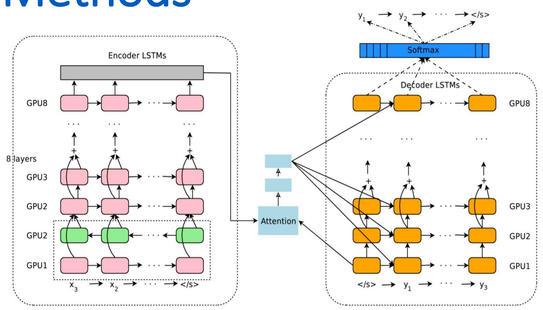
CSE 702 - Applied Natural Language Processing and Computational Social Science

Team: Payraw Salih, Parth Shah

Introduction

The Project is on translation from German to English using seq2seq and then analysis on basis of gender bias. Gender bias is the tendency to translate a sentence into either the masculine form and feminine form when both forms do exist or the form being used is not the best.

Methods



Reference to Model

Model

4 layer model

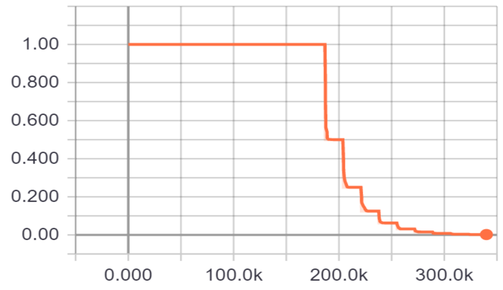
Encoder: 1 bidirectional, 3 unidirectional;
Decoder: 4 unidirectional layers

2 layer model

Encoder: 1 bidirectional, 1 unidirectional
Decoder: 2 unidirectional layers

Units in layer: 1024 units (LSTM)
Optimizer : SGD
Batch size: 128
Search parameter: Beam search (width=10), bleu score metric
Learning rate: Start with 1 then change following t2t (warmup_steps=0):

Graphs and Charts

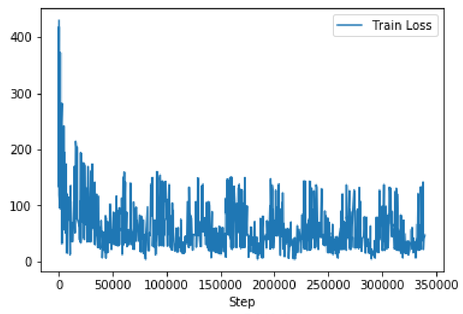


Learning Rate

$$lrate = d_{model}^{0.5}.min(step\_num^{0.5}, step\_num.warmup\_steps^{1.5})$$

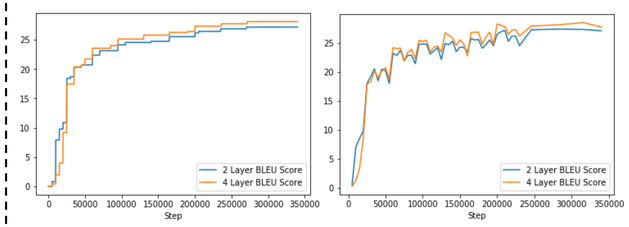
Training Loss

2 Layer GNMT



4 Layer GNMT

BLEU SCORE VS STEP



Training

Test Set

Results

1. We took a sample of 10000 random sentence translation pairs
2. Of these 10,000 translation pairs, ~9,000 were translated with a bleu score less than 0.9.
3. Of the 1000 that had a bleu score higher than 0.9, only ~100 received the perfect 1.0 score.
4. We focus on higher bleu scores because we want to focus mainly on gender bias
5. From the 1000 remaining translations, 600 contained words that expressed gender. These words are: he, she, her, him, miss, mister, madame, sir, sister, brother, mother, father, daughter, son, girl, boy, women, woman, men, man, female, male, hers, his.
   a. We found that of the 600 sentences that contained gender, approximately 180 sentences had translated gender subjectively.
   b. An example of a subjective translation include translating the German word for nurse into “she”. Example output that came from an input which didn’t express gender “the she nurse went to work”. This is, in our opinion, due to the fact that “krankenschwester” the word for nurse, is genitive and female. The correct translation would have been “the nurse went to work”
6. If we extrapolate this data then out of 600 sentences that contained gender words; 180 had incorrect gender in the translation. This accounts for 30% of sentences that have an incorrect reference to gender.

Performance Comparison

Model	BLEU Score
NMT (greedy)	27.6
NMT + GNMT attention (beam=10)	29.9
Our GNMT 2 layer	27.6
Our GNMT 4 layer	**28.7**

Conclusion

1. LSTM DNN was found to be a good step towards a more accurate machine translation
2. Mimicked the LSTM network which returned a bleu score of 28.7.
3. Did gender analysis and found that 36% of sentences containing gender from those sentences with bleu score higher than 0.9 (or 90 if going by 100 scale), will have error with respect to gender.

References

1. Attention Is All You Need. arXiv:1706.03762 [cs.CL]
2. Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. arXiv:1609.08144 [cs.CL]
3. On Using Very Large Target Vocabulary for Neural Machine Translation. arXiv:1412.2007 [cs.CL]
4. Sequence to Sequence Learning with Neural Networks (rep.). Sequence to Sequence Learning with Neural Networks. [NIPS]
5. BLEU : a Method for Automatic Evaluation of Machine Translation. [ACL]
6. Comparing Machine Translation and Human Translation: A Case Study [Ahrenberg, L]
7. Shared Task: Machine Translation. [ACL]
8. A Large-Scale Test Set for the Evaluation of Context-Aware Pronoun Translation in Neural Machine Translation arXiv:1810.02268 [cs.CL]

University at Buffalo The State University of New York

Department of Computer Science and Engineering
University at Buffalo, SUNY
cse.buffalo.edu