Comparing the Performance of Neural Machine Translation and Skip-Thought Vectors in Neural Paraphrase Generation

Jacob Gursky, St. Olaf College '19 | Advisor: Professor Matthew Richey

Research Question

What encoding architectures yield thought vectors from which diverse and semantically-similar paraphrases can be produced?

Introduction

Neural paraphrase generation is a nontrivial task in the subfield of natural language processing regarding the generation of novel paraphrases using neural networks. A common architecture used in generating a sequential output given a sequential input is the Seq2Seq model [1], which uses a encoder that models the sequence of tokens as a fixed-length vector, or thought vector, and a decoder that takes the thought vector as input and produces a series of tokens.

Of interest to our work is the quality of the thought vector, which represents the meaning of a sentence as a point in space. An important property of thought vectorization in reconstructing paraphrases is that semantically-similar sentences should have vectors that are closer to one another in space. We analyze which encoding architectures produce thought vectors with this property using two methods: NMT [2] and Skip-Thoughts [3].

About the Data

The corpus for this study comes from 3,500 texts scraped from the Project Gutenberg project, a free online repository of classic literature.

To prepare the data for model training, only the 15,000 most common words were kept, sentences were restricted to 22 tokens in length, and a subsample of 300,000 sentences was taken from the resulting corpus with an 80/20 train-test split.

Methods: NMT vs. Skip-Thoughts

NMT: Neural Machine Translation, or NMT, uses an encoder-decoder architecture to translate one sequence of tokens to another [2]. We use a given sentence as both the input and output so that the encoder learns to thought-vectorize in a way that the source sentence can be reconstructed.

Skip-Thoughts: Skip-thoughts is a variation on NMT that attempts to reconstruct the preceding and succeeding sequences given a sentence, rather than itself [3]. It has been shown that skip-thought vectors better model the semantic properties of a sentence than other architectures.

Results

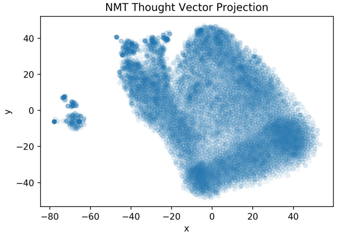
Table 1: Paraphrases generated with varying levels of noise (shown in parentheses)

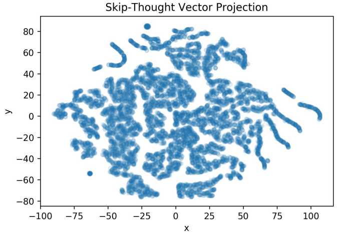
1. He is a good man.
NMT(0): He is a good man.
NMT(0.1): He is a good man.
NMT(0.2): He was a good man?
Skip-Thought(0): He is a very good.
Skip-Thought(0.1): He was a little man too.
Skip-Thought(0.2): "that was."

2. What did you do yesterday?
NMT(0): What did you do yesterday?
NMT(0.1): What did you happened yesterday?
NMT(0.2): What sang not forget.
Skip-Thought(0): What have you been doing?
Skip-Thought(0.1): What have I done?
Skip-Thought(0.2): "Well, it is a good deal for you."

3. I asked her about you!
NMT(0): I asked her about you!
NMT(0.1): I asked her about you!
NMT(0.2): I telling her that you Anne!
Skip-Thought(0): "What is it?"
Skip-Thought(0.1): "What's you," he said, "that's what I want to do."
Skip-Thought(0.2): "You're right."

Figures 1&2: Visualizations of 20,000 thought-vectors using both NMT and Skip-Thoughts projected using t-SNE





Implications

As shown in Table 1, NMT is able to reconstruct the source sequence quite accurately, but does not produce diverse paraphrases, and excessive noise expresses a completely separate thought. It seems that NMT thought vectors are insufficient for paraphrasing tasks.

It seems that a larger training set is required to effectively train a Skip-Thoughts model, as performance is quite poor for sequences with rare tokens. However, we do see initial signs of paraphrasing, with noised reconstructions relating to the source sentence.

Future Study

- Train on a larger corpus and vocabulary
- Compare the performance of GRU and LSTM cells
- Use bidirectional encoders rather than unidirectional
- Use adversarial training for the final decoders [4]
- Use classifier rather than sequential output
- Incorporate attention mechanism
- Utilize beam search for decoding

Acknowledgements

We would like to thank Professor Richey of St. Olaf College for his advising on this project

References

[1]Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." *Advances in neural information processing systems.* 2014.
[2]Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." *arXiv preprint arXiv:1409.0473* (2014).
[3]Kiros, Ryan, et al. "Skip-thought vectors." *Advances in neural information processing systems.* 2015.
[4]Papernot, Nicolas, et al. "Crafting adversarial input sequences for recurrent neural networks." *MILCOM 2016-2016 IEEE Military Communications Conference.* IEEE, 2016.