DL Project Proposal: Shakespeare to English

Morris Kraicer
Johns Hopkins University
mkraice1@jhu.edu

Riley Scott
Johns Hopkins University
rscott39@jhu.edu

William Watson
Johns Hopkins University
billwatson@jhu.edu

Abstract

Add abstract to be fancy

1 Introduction

Abstract

Inspiration

Data web scrap label via aligner

Preprocessing Tokenize!!!!! NUM PROPER NOUNS PUNCTUATION USE SPACY FOR POS TAGS

Models Baseline Seq2Seq Baseline No attention, no Bidirectional encoder RNN-¿ RNN GRU-¿GRU bidirectional GRU -¿ Bidirectional GRU With Attentions Teacher Forcing

Use Aligner to measure quality of data, and learnability. Not learnable -¿ still train, but when report, do backwards, suggest data is difficult Learnable -¿ do aligner do well, then models

Final Tricks: Num -¿ Num PN -¿ PN Punct -¿ Punct

Expectation: Hopefully provides good translations Hopefully can do a pseudo style transfer on English Funny translations? 50 Shades of Grey

2 Data Procurement

[1]

3 Preprocessing

[2]

[1]`https://github.com/cocoxu/Shakespeare`
[2]`https://spacy.io`

4 Architectures

We seek to develop several models to improve our translations, incorporating context, attention, and novel training methods.

4.1 Baseline RNN Sequence to Sequence

Our baseline model will be a simple RNN sequence to sequence model. It will accept a source sentence s and will decode it to an output sequence t. A simple RNN will only incorporate previous states, i.e. source words, during prediction, and might be prone to the vanishing gradient problem. It will incorporate an encoder-decoder style model [2] [5].

4.2 GRU Sequence to Sequence

We will improve our baseline thorugh the use of the GRU layer, to hopefully offset any vanishing gradients. In addition, we use a GRU over an LSTM to help bound the number of parameters used in the model for training purposes, as the GRU and LSTM fix issues regarding long term dependencies within sequences.

4.3 Bidirectional Model

In order to consider the context of words before an after, we will alter the model to use a bidirectional GRU [1], and hopefully see better translations.

4.4 Attention Mechanisms

Attention mechanisms have been shown to improve sequence to sequence translations from Bahdanau et al [1], and further work from Luong et al [4] examines global vs local approached to attention-based encoder-decoders.

Roles	Members
Data Procurement	Morris, Riley, Bill
Preprocessing Scripts	Riley, Bill
Baseline Model	Riley
GRU Extension	Morris
Bidirectional Extension	Morris
Attention Mechanisms	Bill
Train and support code	Bill
Training experiments	Morris, Riley
Writeup and analysis	Morris, Riley, Bill

Figure 1: Role Distribution

4.5 Teacher Forcing

In terms of training, an encoder-decoder system can either accept the target token or the model's prediction as input during the decoding step. When we use the target token, this is known as teacher forcing, and is shown to be favored during initial trainging iterations, but should be backed off to use the model's own predicitons, as it will exhibit instability in the translations otherwise [3].

5 Roles and Responsibilities

We intend to split the project into the following subsections according to Figure 1.

6 Expectations

References

[1] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[2] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[3] A. M. Lamb, A. G. A. P. GOYAL, Y. Zhang, S. Zhang, A. C. Courville, and Y. Bengio. Professor forcing: A new algorithm for training recurrent networks. In *Advances In Neural Information Processing Systems*, pages 4601–4609, 2016.

[4] M.-T. Luong, H. Pham, and C. D. Manning. Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*, 2015.

[5] I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.