DSGA 1011 Natural Language Processing: Project Proposal

Sang Jun Yum and Yurui Mu

Sang Jun Yum
Sang Jun Yum, New York University, e-mail: sjy269@nyu.edu

Yurui Mu
Yurui Mu, New York University e-mail: ym1495@nyu.edu

Abstract For the final project, we will implement a Neural Translation Model using RNN Encoder-Decoder and Soft Attention Mechanism. Overall structure of the model is well explained in NVIDIA's blog post[1, 2, 3], Cho et al., 2014[4] and Bahdanau et al., 2016[5].

1 Encoder-Decoder

Encoder-Decoder is a 'concatenated' RNN model that is used frequently in Neural Machine Translation tasks. Suppose a source sentence $X = \{x_1, x_2, ..., x_T\}$ and a target sentence $Y = \{y_1, y_2, ..., y_T\}$ are present in the given corpus.
'Encoding' is analogous to human's reading a source sentence from which translation is to be made. Embedded representations of words from source sentence are fed into the RNN

$$h_{\mathrm{T}} = \phi_\theta(h_{T-1}, s_T) \tag{1}$$

'Decoding' is analogous to human's translating the source sentence to the target sentence in target language. Decoding step involves another RNN that computes hidden states, z's.

$$z_{\mathrm{i}} = \phi_{\theta'}(h_T, u_{i-1} z_{i-1}) \tag{2}$$

where u's are one hot representations of the target words. After computing z's, the model scores each target word based on how likely it is to follow all the preceding translated words given the source sentence. Each score is computed as

$$e(k) = u_k^T zi + b_k \tag{3}$$

This process is done for every hidden state z's, and probability distribution of these scores are computed using *Softmax*.
By training the Encoder-Decoder model, conditional log-likelihood is obtained using SGD.

$$\max_\theta \frac{1}{N} \sum_{n=1}^{N} p_\theta(Y_n | X_n) \tag{4}$$

2 Attention Mechanism

A single layered Feed-Forward Neural Network is implemented inside decoder; this network takes as input the decoder's previous state z_i and source word representation h_j to return the probability of decoder selecting h_j out of T source words. Simply put, this small network decides out of a translated word, which source word is the word translated from. And such decision is made again by *Softmax*.

3 Dataset and Evaluation

In this project, we plan to use WIT3(Web Inventory of Transcribed and Translated Talks) training and evaluation datasets[6], to translate from English to French.
We will use BLEU, an automatic machine translation evaluation metric that is quick, inexpensive, and language-independent.[7]

References

1. Kyunghyun Cho. Introduction to neural machine translation with gpus (part 1), May 2015.
2. Kyunghyun Cho. Introduction to neural machine translation with gpus (part 2), May 2015.
3. Kyunghyun Cho. Introduction to neural machine translation with gpus (part 3), May 2015.
4. Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *CoRR*, abs/1406.1078, 2014.
5. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2014.
6. Mauro Cettolo, Christian Girardi, and Marcello Federico. Wit^3: Web inventory of transcribed and translated talks. In *Proceedings of the 16th Conference of the European Association for Machine Translation (EAMT)*, pages 261–268, Trento, Italy, May 2012.
7. Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia, PA, USA.*, pages 311–318, 2002.