DS-GA 1011: Natural Language Processing with Representation Learning
Project Proposal

Yurui Mu (ym1495) Sang Jun Yum(sjy269)

November 9, 2017

1 Introduction

In this project, we aim to implement and compare two bilingual machine translation systems:

1. RNN Encoder-Decoder
2. Encoder-Recurrent Memory Network

We will take RNN Search[1] as baseline model. And we propose to build a machine translation system: Encoder-RMN, where we instead of using RNN decoder, use Recurrent Memory Network[2] to generate target sentences.

2 RNN Encoder-Decoder

RNN Encoder-Decoder[3] is a classical machine translation model proposed by Cho et al. It encodes the input sentence into a sequence of vectors and chooses a subset of these vectors adaptively while decoding the translation.

3 Encoder - RMN

According to literature [4], it is plausible to integrate language models into machine translation systems. And we plan to substitute RNN Decoder in RNN Search with an RMN Decoder, with a Memory Block stacking on top of an LSTM. Recurrent Memory Network[2] consists of two components: an LSTM and a Memory Block. The Memory Block takes the hidden state of the LSTM and compares it to the most recent inputs using an attention mechanism[2].
A Memory Block is a variant of single-layer Memory Network[5][6]. Different from Memory Network, instead of using $g(s_t, h_t) = s_t + h_t$, RMN uses a gating unit that decides how much it should trust the hidden state h_t and context s_t at time step t. And such gating unit comes in form of Gated Recurrent Unit[7]:

$$z_t = sigm(W_{sz}s_t + U_{hz}h_t)$$

$$r_t = sigm(W_{sr}s_t + U_{hr}h_t)$$

$$\tilde{h}_t = tanh(W_{s_t} + U(r_t \odot h_t))$$

$$h_t^m = (1 - z_t) \odot h_t + z_t \odot \tilde{h}_t$$

RMN has worked well in language modeling in German, English and Italian datasets, outperforming LSTM baselines in terms of perplexity. It has also worked well in sentence completion challenge with 69.2% accuracy[2]. Thus hopefully we could see an increase of BLEU score in our Encoder-RMN model.

4 Dataset

In this project, we plan to use WIT3(Web Inventory of Transcribed and Translated Talks) training and evaluation datasets[8], to translate from English to French.

5 Evaluation

We will use BLEU, an automatic machine translation evaluation metric that is quick, inexpensive, and language-independent.[9]

References

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, abs/1409.0473, 2014. 1

[2] Ke M. Tran, Arianna Bisazza, and Christof Monz. Recurrent memory network for language modeling. *CoRR*, abs/1601.01272, 2016. 1, 3

[3] 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. 2

[4] Caglar Gulcehre, Orhan Firat, Kelvin Xu, Kyunghyun Cho, and Yoshua Bengio. On integrating a language model into neural machine translation. *Computer Speech Language*, 45(Supplement C):137 – 148, 2017. 3

[5] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. Weakly supervised memory networks. *CoRR*, abs/1503.08895, 2015. 3

[6] Junyoung Chung, Çaglar Gülçehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *CoRR*, abs/1412.3555, 2014. 3

[7] KyungHyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *CoRR*, abs/1409.1259, 2014. 3

[8] 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. 4

[9] 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. 5