Exploring the Copying Mechanism in Sequence-to-Sequence Learning

Anonymous ACL submission

Abstract

The copying mechanism, defined as selectively repeating a certain segment of previously text, is common in human language processing. In this paper, we explore the copying mechanism in the context of sequence-to-sequence learning and devise a neural network model, called COPYNET. Our empirical study on both synthetic data sets and real world application demonstrates the efficacy of COPYNET on mixing these two generation modes for fulfilling the tasks.

1 Introduction

Recently the Sequence-to-sequence (SEQ2SEQ) learning framework achieves great success in various basic natural language processing (NLP) applications but not limited to Machine Translation (Cho et al., 2014; Bahdanau et al., 2014), Syntactic Parsing (Vinyals et al., 2015), Text Summarization (Rush et al., 2015) and Dialogue Systems (Shang et al., 2015; Vinyals and Le, 2015). The SEQ2SEQ is typically achieved with the encoder-decoder ...

The introducing of attention mechanism, which was first proposed for machine translation tasks as ”automatic” alignment, has greatly improved the performance on Seq2Seq tasks such as XXXX and YYYY. Different from the canonical encoder-decoder architecture, attention-based Seq2Seq model revisits the source in its relatively raw form (array of word-level representations) to dynamically fetch the relevant piece of information based mostly on content-based addressing.

In this paper, we explore another mechanism important to the natural language communication of human being, called ”copying mechanism”. Basically, copying mechanism locates a certain segment of source sentence and puts this segment into the target. For example, in the following three dialogue turns we observe different patterns that some sub-sequence in the response (R) copied from the the utterance (O) (colored blue)

O: `Hello Jack, my name is Chandralekha.`
R: `Nice to meet you, Chandralekha.`

O: `I watched The Big Lebonski last night. Absolutely hilarious!`
R: `The movie's name is The Big Lebonski?`

O: `This new guy doesn't perform exactly as we expected.`
R: `What do you mean by "doesn't perform exactly as we expected"?`

Unlike the canonical encoder-decoder and attention model, both of which rely heavily on the representation of ”meaning”, copy mechanism is more related to the rote memorization in the natural language processing of human being. This divergence calls for different strategies in modeling the two mechanism in neural network-based models, and probably more sensibly, an elegant unified model to accommodate (mix) both understanding and memorization. Towards this goal, we propose COPYNET, which can nicely combine the regular generation process and the operation of copying appropriate segment of the input sequence. Our empirical study on both synthetic data sets and real world application demonstrates the efficacy of COPYNET on mixing these two generation modes for fulfilling the tasks.

RoadMap

2 COPYNET

2.1 Hybrid Addressing

We hypothesize that COPYNET uses a hybrid addressing strategy in fetching the content in the short-term memory (STM) of source, which nicely mixes content-based and location-based addressing for COPYNET's hybrid style of generation. Unlike the explicit design for hybrid addressing in Neural Turing Machine(**?**), COPYNET is more subtle: it provides the architecture that can facilitate some particular location-based addressing without any explicit notion of location, and let the model to figure out the details from the data (task).

Assume there is absolution location information in h_i, the information flow (where $\tilde{h}_t$ stands for the location-input at time t (a convex combination of $\{h_i\}$) and $\hat{h}_t$ stands for the probability of source words)

$$\cdots \to \tilde{\mathbf{h}}_t \xrightarrow{\text{update}} \mathbf{s}_t \xrightarrow{\text{predict}} \hat{h}_t \xrightarrow{\text{average}} \tilde{\mathbf{h}}_{t+1} \to \cdots$$

provides a simple way to "moving one word to right" on source. More specifically, the state-update operation $\tilde{\mathbf{h}}_t \xrightarrow{\text{update}} \mathbf{s}_t$ potentially acts as "move right" , leading to the prediction of the location $\mathbf{s}_t \xrightarrow{\text{predict}} \hat{h}_t$ right to that, which is then again transformed into the location-input ($\hat{h}_t \xrightarrow{\text{average}} \tilde{\mathbf{h}}_{t+1}$) fed to the state update for next round. This operation is learned through training (for both encoder and decoder) and has been controlled by the content ...

In contrast, this move-right operation is a built-in in NTM, and

It is fairly hard to verify the addressing strategy as above directly, but Observations and suggestive evidence

- toy data
- UNKs

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Alexander M Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685*.

Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural responding machine for short-text conversation. *arXiv preprint arXiv:1503.02364*.

Oriol Vinyals and Quoc Le. 2015. A neural conversational model. *arXiv preprint arXiv:1506.05869*.

Oriol Vinyals, Łukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey Hinton. 2015. Grammar as a foreign language. In *Advances in Neural Information Processing Systems*, pages 2755–2763.