Software Projects

Julia Kreutzer, Stefan Riezler

SS 2019

Projects Overview

Simple seq2seq learning using joeynmt

- JoeyNMT is a a minimalist framework for neural machine translation, developed especially for educational purposes
- Your task is to extend JoeyNMT for seq2seq tasks beyond supervised machine translation
- Each project is a re-implementation of an interesting task, based on a single paper
- Your goal is to become proficient in s2s learning, not scientific innovation

Projects Overview

Sample projects

- Neural machine translation without parallel data
- Task-oriented dialogue without database execution
- Text summmarization without simple baselines

Project 1: NMT without parallel data

Neural machine translation

- Standard approaches to NMT such as RNNs with attention [Bahdanau et al., 2015] are part of the core implementation of JoeyNMT
- Standard training is performed by supervised learning on parallel sentences in source and target language

Challenge

- How far can one get by learning to translate from monolingual corpora only?
- Which extensions to JoeyNMT are needed in order to achieve this?

Project 1: NMT without parallel data

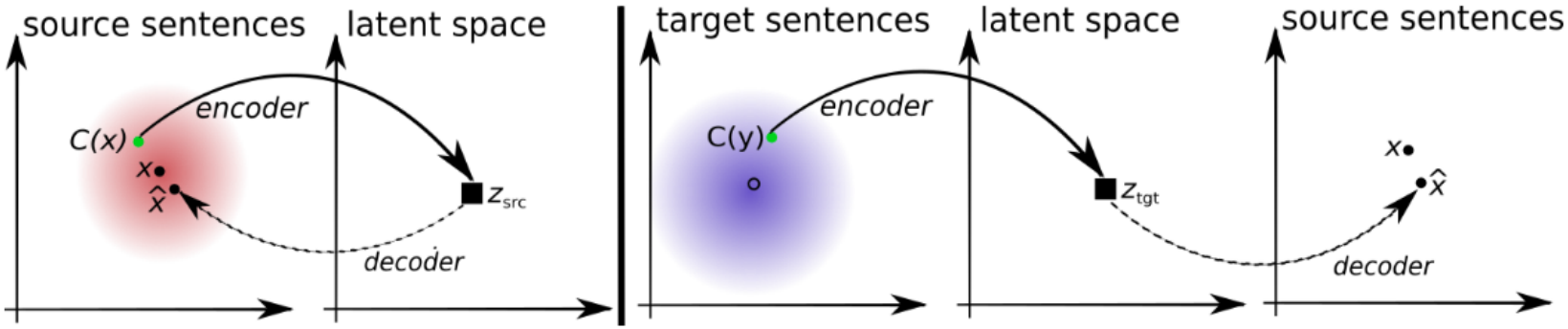
Neural machine translation

- Standard approaches to NMT such as RNNs with attention [Bahdanau et al., 2015] are part of the core implementation of JoeyNMT
- Standard training is performed by supervised learning on parallel sentences in source and target language

Challenge

- How far can one get by learning to translate from monolingual corpora only?
- Which extensions to JoeyNMT are needed in order to achieve this?

Approach



- Reconstruction [Lample et al., 2018, Artetxe et al., 2018]
  - Map sentences from monolingual corpora in two different languages into the same latent space
  - Learn to translate by learning to reconstruct in both languages
    - Reconstruct source from noise version of it, using autoencoder, in order to learn about structure
    - Reconstruct source from self-generated translation, in order to learn about translation

Project 2: Dialogue without database execution

Task-oriented dialogue systems

- Standard approaches apply seq2seq models similar to those used in NMT to “parallel” data consisting of sequence of turns of a dialogue [Vinyals and Le, 2015]
- Standard training is performed by supervised learning to predict the next dialogue turn

Challenge

- As soon as a knowledge base needs to be accessed in the dialogue, end-to-end differentiability is broken. What about using attention mechanism for soft incorporation of KB while still preserving trainability?
- Which extensions to JoeyNMT are needed in order to perform task-oriented dialogue-turn prediction?

Project 2: Dialogue without database execution

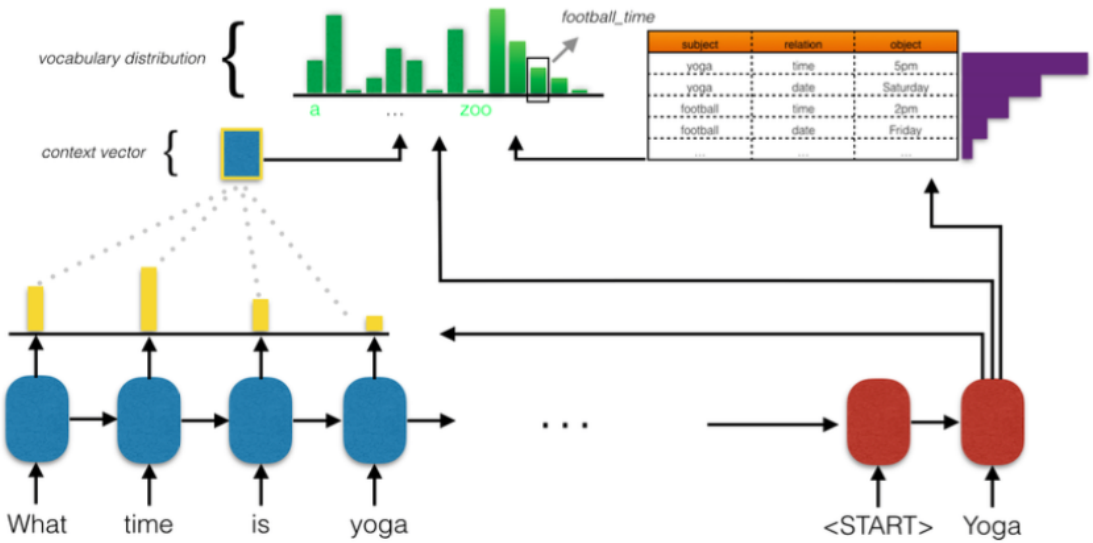
Task-oriented dialogue systems

- Standard approaches apply seq2seq models similar to those used in NMT to “parallel” data consisting of sequence of turns of a dialogue [Vinyals and Le, 2015]
- Standard training is performed by supervised learning to predict the next dialogue turn

Challenge

- As soon as a knowledge base needs to be accessed in the dialogue, end-to-end differentiability is broken. What about using attention mechanism for soft incorporation of KB while still preserving trainability?
- Which extensions to JoeyNMT are needed in order to perform task-oriented dialogue-turn prediction?

Approach



- Key-value retrieval network [Eric et al., 2017]
  - Separate attention over encoder states and KB keys
  - Combine both to get distribution over target vocabulary

Project 3: Summarization without first-sentence baseline

Extractive summarization

- Standard approaches apply seq2seq models similar to those used in NMT to “parallel” data consisting of sentences and keep/drop decision [Nallapati et al., 2017]
- Standard training is performed by supervised learning to predict which sentence to keep in summary

Challenge

- News articles can be summarized by simply extracting the first three sentences. What about using non-news data like WikiHow [Koupaee and Wang, 2018]?
- Which extensions to JoeyNMT are needed in order to perform extractive summarization?

Project 3: Summarization without first-sentence baseline

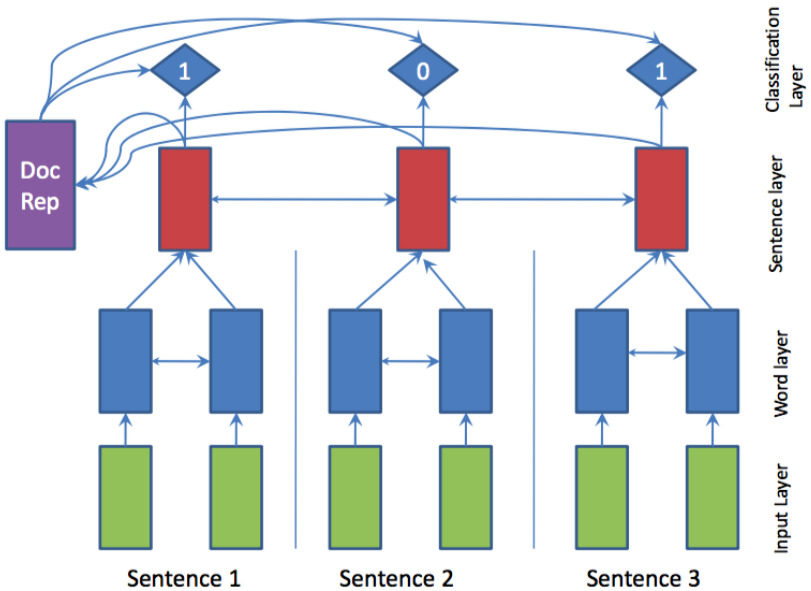
Extractive summarization

- Standard approaches apply seq2seq models similar to those used in NMT to “parallel” data consisting of sentences and keep/drop decision [Nallapati et al., 2017]
- Standard training is performed by supervised learning to predict which sentence to keep in summary

Challenge

- News articles can be summarized by simply extracting the first three sentences. What about using non-news data like WikiHow [Koupaee and Wang, 2018]?
- Which extensions to JoeyNMT are needed in order to perform extractive summarization?

Approach



- RNN-based summarization [Nallapati et al., 2017]
  - Bottom layer operates on words, top layer on sentences
  - Classification layer decides whether or not sentence belongs into summary

References


Artetxe, M., Labaka, G., Agirre, E., and Cho, K. (2018).
Unsupervised neural machine translation.
In *Proceedings of the 6th Conference on Learning Representations (ICLR)*, Vancouver, BC, Canada.


Bahdanau, D., Cho, K., and Bengio, Y. (2015).
Neural machine translation by jointly learning to align and translate.
In *Proceedings of the International Conference on Learning Representations (ICLR)*, San Diego, CA.


Eric, M., Krishnan, L., Charette, F., and Manning, C. D. (2017).
Key-value retrieval networks for task-oriented dialogue.
In *Proceedings of the SIGDIAL Conference*, Saarbrücken, Germany.


Koupaee, M. and Wang, W. Y. (2018).
Wikihow: A large scale text summarization dataset.
CoRR, abs/1810.09305.


Lample, G., Denoyer, L., and Ranzato, M. A. (2018).
Unsupervised machine translation using monolingual corpora only.
In *Proceedings of the 6th Conference on Learning Representations (ICLR)*, Vancouver, BC, Canada.


Nallapati, R., Zhai, F., and Zhou, B. (2017).
Summarunner: A recurrent neural network based sequence model for extractive summarization of documents.

In *Proceedings of the 31st Conference on Artificial Intelligence (AAAI)*, San Francisco, CA, USA.

Vinyals, O. and Le, Q. V. (2015).
A neural conversational model.
In *Proceedings of the ICML Deep Learning Workshop*, Lille, France.