Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation

https://arxiv.org/abs/1406.1078

EuiYul Song (20214426)

Files

- main.py: a main class that preprocess, train, and evaluate the model
- vocab.py: a class that tokenizes and builds vocabulary dictionaries of input and output phrases
- encoder.py: a RNN Encoder with GRU that encodes a sequence of symbols into a fixed-length vector representation
- decoder.py: a RNN Decoder with GRU that decodes the representation into another sequence of symbols
- app.py: a Flask website that enables users to run French to English translation against this model
- split.py: splits train and test set
- Makefile: a command line automation

Data

- data/eng-fra.txt: an English to French dictionary (separated by "
  \t") that was downloaded from [link (http://www.manythings.org/anki/fra-eng.zip)](http://www.manythings.org/anki/fra-eng.zip) and refined
- data/eng-fra-train.txt: literally training data
- data/eng-fra-train.txt: literally test data

Command

- make run: preprocess, train, and evaluate the model
- make app: run website that translates inputted French phrase to English using this model at localhost:5000 or 127.0.0.1:5000

Analysis

- BLEU score on the trained dictionary

  - BLEU1: 0.925128729444704
  - BLEU2: 0.747157671833037
  - BLEU3: 0.6159862239384737
  - BLEU4: 0.4493687308739829
  - Desc: Macroscopically looking at the BLEU score, I maintain that my model trained French to English dictionary decently without high underfitting. However, I saw some grammatical mistakes and bad synonym choice after translation.

- BLEU score on the test dictionary

  - BLEU1: 0.9003110619801925

- BLEU2: 0.704739339739333
- BLEU3: 0.5752640535287554
- BLEU4: 0.41597006978070844
- Desc: I maintain that BLEU score and evaluation on the test dictionary proves that my model can generalize pretty decently and can train dictionary decently. However, my model has punctuational, grammatical, and word-choice issues when I manually tested my model against the test dictionary.

- Prediction on unseen French word

  - Looking at the website I implemented, it was obvious that this model cannot predict unseen French phrases very well. Thus, this model has overfitting problem besides the problems I mentioned above.

Visualization

- Website at localhost:5000 or http://127.0.0.1:5000 if the localhost does not work (refer to Command section)

Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation (Link)

| Type in French to retrieve English tranlsation... | Search |

French: **je n en ai pas fini avec toi .**

English: **i m not finished with you .**

Limitation

- Due to time-constraint to complete this project, I was not able to solve overfitting problem. I will add more data, add dropout, or add regularization in order to reduce overfitting. Additionally, I will add t-SNE to visualize my model's embeddings in the future.