CS224n Assignment5 Solution

shawnye

August 2020

1 Character based convolutional encoder for NMT

(a)

No. For CNNs, the input length of all examples need to be the same, otherwise the output shape of CNN layer would be different, if there are following linear neural networks layer in the model, the model cannot be trained.

(b)

The padding size should be 3 if we hope that there exists at least one window for all possible values of m_{word}. The smallest possible value that m_{word} can take is 1. Because we have padded the $start\_of\_word$ token and end_of_word token to the word character embedding, then for $k = 5$, we need to pad another $k - 2$ zeros to both sides.

(c)

Highway layer enables the model to adaptively decide to project some of the input convolutional features into another space, and keep some features unchanged.

The b_gate should be better to initialize to be negative. In this case, the sigmoid function value will be biased to 0 during initial stage, then the highway just pass the identity x_{conv_out}, it should be better for the training warmup.

(d)

(i) Transformer is able to do parallel computation, so it is faster to train and inference compared with the LSTM.

(ii) LSTM-with-attention model only make each step hidden state of decoder attend to the encoder hidden states to produce the prediction, the hidden state representations of encoder doesn't take the advantage of attention. While the

self-attention in transformer encoder generate better representation of the source sentence.

2 Character-based LSTM decoder for NMT

The final test BLEU score is 35.91

3 Analyzing NMT system

(a)

"traducir", "traduzco", and "traduce" exsist in the word-vocabulary, "traduces", "traduzca" and "traduzcas" are not included in the word-vocabulary.

For the word-based NMT from Spanish to English, different forms of "traducir" which did not included in the training word-vocabulary cannot be translated because they are represented with "<unk>". Even though the meaning of different forms are closely similar to each other. Character-aware NMT is able to represent any input words with appropriate embeddings. Besides, different forms of words which with similar meaning should be easily inferenced by the character-aware NMT.

(b)

(i) The nearest neighbor of "financial" is "economic". The nearest neighbor of "neuron" is "neurons". The nearest neighbor of "Francisco" is "san". The nearest neighbor of "naturally" is "occurring". The nearest neighbor of "expectation" is "operator".

(ii) Similar words of character-based embeddings.

The nearest neighbor of "financial" is "vertical". The nearest neighbor of "neuron" is "Newton". The nearest neighbor of "Francisco" is "France". The nearest neighbor of "naturally" is "practically". The nearest neighbor of "expectation" is "exception".

Compared with Word2Vec, CharCNN models the character similarity of words, words composed by similar characters is more close to each other. Such as "neuron" and "Newton", "expectation" and "exception". There is no words meaning in the embedding. On the contrary, Word2Vec models the meaning of words, words with similar meanings or usually appear together close to each other. such as "financial" and "economic", "Francisco" and "san".

(c)

Test example 1:

Source Sentence: Bien, al da siguiente estbamos en Cleveland.
Reference: Well, the next day we were in Cleveland.
Assignment4 output: Well, the next day we were in <unk>.
Character-based output: Well, the next day we were in Cleveland.

This is a good example, character-based model successfully keep the name of a location, a word which is not included in the word-based model training vocabulary.

Test example 2:
Source sentence: Sobresale del agua unos 120 pies o 40 metros.
Reference: It's about 120 feet above the water, or 40 meters.
Assignment4 output: <unk> of the water about 120 feet or 40 meters.
Character-based output: It's about water about 120 feet or 40 meters.

This is an unacceptable translation for character-based NMT. Even though it doesn't generate "<unk>" as word-based NMT, but we cannot understand the meaning of translated sentence. The character-based model failed to translate "Sobresale" to "above", it generate "about" instead. The reason is that embeddings of "about" and "above" are close to each other in character-based model.