Encoder-decoder architectures [Solution by Karthikeyan.S]

LATEST SUBMISSION GRADE

100%

1.Question 1

In the lecture as well as in this test we will have lots of formulas. Let us first make sure that we remember the used notation.

Please, name the following objects: II, JJ, x_ix_i, y_jy_j, h_ih_i, v_jv_j, s_js_j

length of target, length of source, encoder state, decoder state, source word, target word, context vector

length of source, length of target, source word, target word, encoder state, decoder state, context vector

length of source, length of target, source word, target word, encoder state, context vector, decoder state

Correct

1 / 1 point

2.Question 2

How do we compute the context (thought) vector vv for the decoder position jj in a seq2seq model without attention?

h_Ih_I, where h_Ih_I is the last encoder state

\sum_i \alpha_i^j h_i$\sum_i \alpha_{ij} h_i$, where \alpha_i^jα_{ij} are some weights

h_jh_j, where h_jh_j is the jj-th state of the encoder

Correct

Correct!

1 / 1 point

3.Question 3

How many new parameters for the network are introduced to calculate **multiplicative attention** weights? (Just to calculate, we are not yet looking into how we use them afterwards).

The length of the source, multiplied by the length of the target

The dimension of an encoder state, multiplied by the dimension of a decoder state

No new parameters

Correct

Exactly! This is the number of parameters for **multiplicative** attention.

1 / 1 point

4.Question 4

Which of the following formulas stand for the **additive attention**? Note that h_ih_i is the ii-th encoder state, s_js_j is the jj-th decoder state, and we are interested in the similarity between them.

w^T tanh (W [h_i, s_j])$w_T tanh(W[h_i,s_j])$, where the brackets denote concatenation of the vectors, and ww and WW are a vector and a matrix of parameters respectively.

h_i^T W s_j$h_i^T W s_j$, where W is a matrix of parameters

h_i^T s_j$h_i^T s_j$

Correct

Exactly! Take a moment to see, that this is just a different form of the same additive attention formula that was introduced at the lecture.

1 / 1 point

5.Question 5

Let us denote encoder states by h_ih_i with ii going from 1 to II. Lets us denote by a_i^ja_{ij} the similarities computed using the additive attention formula from the previous question. How should we compute the context vector v_jv_j for the decoder position jj?

\sum_{i=1}^I a_i^j h_i$\sum_{i=1}^I a_{ij} h_i$

\sum_{i=1}^I \frac{\exp a_i^j}{\sum_{i'} \exp a_{i'}^j} h_i$\sum_{i=1}^{I} \frac{\exp a_i^j}{\sum_{i'} \exp a_{i'}^j} h_i$

\sum_{i=1}^I \frac{\exp a_i^j}{\sum_{j'} \exp a_i^{j'}} h_i$\sum_{i=1}^{I} \frac{\exp a_i^j}{\sum_{j'} \exp a_i^{j'}} h_i$

Correct

Correct! We apply *softmax* to transform the weights into probabilities and compute the average of the encoder states.

1 / 1 point

6.Question 6

Which three vectors should be passed to a decoder state s_js_j in a seq2seq with attention model from the lecture?

- [ ] x_ix_i - the ii-th word in the source sequence
- [ ] h_{i-1}h_{i-1} - the previous encoder state
- [x] y_{j-1}y_{j-1} - the previous word in the target sequence

Correct

- [ ] x_{i-1}x_{i-1} - the previous word in the source sequence
- [ ] y_jy_j - the jj-th word in the target sequence
- [ ] h_ih_i - the ii-th encoder state
- [x] v_jv_j - the context vector for position jj, calculated using attention

Correct

- [ ] s_js_j - the jj-th decoder state
- [x] s_{j-1}s_{j-1} - the previous decoder state

Correct

1 / 1 point

7.Question 7

Which techniques would help if the data has rich morphology, informal spelling, and other sources of OOV tokens?

- [x] Byte-pair encoding

Correct

- [ ] Negative sampling
- [x] Sub-word modeling

Correct

- [ ] Hierarchical softmax
- [x] Copy mechanism

Correct

1 / 1 point

8.Question 8

Let us imagine we have trained a conversational chat-bot as a seq2seq model on Harry Potter movies subtitles. What problems could we expect?

- [x] The bot suggests to use a time-turner or probably some spell if you say you do not have enough time for your Coursera studies

Correct

Dramatic/unrealistic topics, influenced by the training corpus!

- [ ] The bot makes lots of spelling mistakes
- [x] The bot doesn't remember what has already been decided in your dialogue

Correct

No memory of the context / history!

- [x] When asked "What's your name?", the bot is not sure and says Harry, or Ron, or Hermione from time to time.

Correct

No personality problem!

If asked in English, the bot replies in French or some other language

1 / 1 p