Encoder-decoder architectures

Quiz, 8 questions

Congratulations! You passed!

Next Item

1 / 1 points

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: I, J, x_i, y_j, h_i, v_j, s_j

- [x] length of source, length of target, source word, target word, encoder state, context vector, decoder state

Correct

- [ ] 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

---

1 / 1 points

2.
How do we compute the context (thought) vector v for the decoder position j in a seq2seq model without attention?

- [ ] h_j, where h_j is the j-th state of the encoder
- [x] h_I, where h_I is the last encoder state

Correct
Correct!

$\sum_i \alpha^j h_i$, where α^j are some weights

Encoder-decoder architectures

Quiz, 8 questions

1 / 1 points

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

Correct

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

- No new parameters

1 / 1 points

4.

Which of the following formulas stand for the **additive attention**? Note that h_i is the i-th encoder state, s_j is the j-th decoder state, and we are interested in the similarity between them.

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

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.

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

1 / 1 points

5.

Encoder-decoder architectures

Quiz, 8 questions

Let us denote encoder states by h_i with i going from 1 to I. Lets us denote by a_i^j the similarities computed using the additive attention formula from the previous question. How should we compute the context vector v_j for the decoder position j?

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

Correct

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

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

---

1 / 1 points

6.

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

- [ ] h_{i-1} - the previous encoder state

Un-selected is correct

- [x] v_j - the context vector for position j, calculated using attention

Correct

- [ ] s_j - the j-th decoder state

Un-selected is correct

- [ ] x_{i-1} - the previous word in the source sequence

Un-selected is correct

- [ ] y_j - the j-th word in the target sequence

Encoder-decoder architectures

Quiz, 8 questions

Un-selected is correct

- [x] x_i - the i-th word in the source sequence

Un-selected is correct

- [x] y_{j-1} - the previous word in the target sequence

Correct

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

Correct

- [x] h_i - the i-th encoder state

Un-selected is correct

---

1 / 1 points

7.

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

- [ ] Hierarchical softmax

Un-selected is correct

- [x] Byte-pair encoding

Correct

- [ ] Negative sampling

Un-selected is correct

Encoder-decoder architectures

Quiz, 8 questions

Sub-word modeling

Correct

Copy mechanism

Correct

1 / 1 points

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?

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!

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

Un-selected is correct

The bot makes lots of spelling mistakes

Un-selected is correct

The bot doesn't remember what has already been decided in your dialogue

Correct

No memory of the context / history!

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

Encoder–decoder architectures

Quiz, 8 questions

Correct
No personality problem!