

Hide menu

Text Summarization

Lecture Notes (Optional)

Practice Quiz

Practice Quiz: Text Summarization

Started

Assignment

Congratulations! You passed!

Grade received 100% To pass 80% or higher

Go to next item

Text Summarization

1. Select all the correct answers.

Submit your assignment

1 / 1 point

Try again

With transformers, the vanishing gradient problem isn't related with length of the sequences because we have access to all word positions at all times.

Correct

Correct.

Receive grade

To Pass 80% or higher

Your grade

100%

View Feedback

We keep your highest score

Transformers are able to take more advantage from parallel computing than other RNN architectures previously covered in the course.

Correct

Correct.

Like

Dislike

Report an issue

Transformers are models that use both recurrent units and attention mechanisms.

Correct

Correct.

Even RNN architectures like GRUs and LSTMs don't work as well as transformers for really long sequences.

Correct

Correct.

2. Which of the following are applications of transformers?

1 / 1 point

Text summarization.

Translation

Question Answering

Chatbots

All of the above.

Correct

Correct. There are others too.

3. What is one of the biggest techniques that the T5 model brings about?

1 / 1 point

It's attention mechanism is far more superior than the one used in other models.

It makes use of transfer learning and the same model could be used for several applications. This implies that other tasks could be used to learn information that would benefit us on different tasks.

T5 model is very cheap to train from scratch.

It allows for interpretability.

Correct

Correct.

4. When it comes to translating french to english using dot product attention:

1 / 1 point

You find the distribution by multiplying the queries by the keys (you might need to scale), take the softmax and then multiply it by the values.

Correct

Correct.

A CPU is more than enough to train this type of model.

The intuition is that each query  $q_i$ , picks most similar key  $k_j$ . This allows the attention model to focus on the right words at each time step.

Correct

Correct.

The queries are the english words and the keys and values are the french words.

Correct

Correct.

5. Which of the following corresponds to the causal (self) attention mechanism?

1 / 1 point

One sentence (decoder) looks at another one (encoder)

In one sentence, words look at previous words (used for generation). They can not look ahead.

In one sentence, in this attention mechanism, words look at both previous and future words.

In causal attention, queries and keys come from different sentences and queries search among words before only

Correct

Correct.

6. Let's explore multi-headed attention in this problem. Select all that apply.

1 / 1 point

Each head learns a different linear transformation to represent words.

Correct

Correct.

Those linear transformations are combined and run through a linear layer to give you the final representation of words.

Correct

Correct.

Multi-Headed models attend to information from different representations at different positions

Correct

Correct.

Multi-Headed attention allows you to capture less information than single headed attention.

Correct

Correct.

7. Which of the following is true about about bi-directional attention?

1 / 1 point

It only attends to words before.

It used an encoder and decodes it using a decoder.

It could attend to words before and after the target word.

It is less powerful than regular uni-directional attention.

Correct

Correct.

8. Why is there a residual connection around each attention layer followed by a layer normalization step in the in the decoder network?

1 / 1 point

To speed up the training, and significantly reduce the overall processing time.

To break the symmetry in the back-prop.

To help with the interpretability.

To help with the parallel computing component during the training.

Correct

Correct.

9. In the lecture, the way summarization is generated is using:

1 / 1 point

Next sentence prediction.

Next character generation.

Next word generation.

By extracting key sentences from the original article.

Correct

Correct.