20233_csci_544_30249: Applied Natural Language Processing   Assignments   Review Test Submission: Quiz 8

20233_csci_544_30249: Applied Natural Language Processing

- Announcements
- Syllabus
- Content
- Assignments

- Tools
- USC Code of Ethics

Review Test Submission: Quiz 8

User	Kayvan Shah
Course	20233_csci_544_30249: Applied Natural Language Processing
Test	Quiz 8
Started	11/7/23 5:35 PM
Submitted	11/7/23 5:45 PM
Due Date	11/7/23 5:50 PM
Status	Completed
Attempt Score	65 out of 100 points
Time Elapsed	9 minutes out of 10 minutes
Results Displayed	All Answers, Submitted Answers, Correct Answers

Question 1   10 out of 10 points

Automatic metrics are low-cost to compute while being unable to capture high-level grammatical correctness and meaningfulness.

Selected Answer: ✔ True

Answers: ✔ True

False

Question 2   10 out of 10 points

Consider the sentence $x_1, x_2, \ldots, x_n$ where each x_i is a word in the sentence. Which of the following probability decompositions correctly describes the RNN approach in modeling the sequence?

Selected Answer:

✔ $p(x_1, x_2, \ldots, x_n) = \prod_i p(x_i | x_{i-1}, \ldots, x_1)$

Answers:

✔ $p(x_1, x_2, \ldots, x_n) = \prod_i p(x_i | x_{i-1}, \ldots, x_1)$

$p(x_1, x_2, \ldots, x_n) = \prod_i \sum_{x_{i-1}} p(x_i, x_{i-1})$

$p(x_1, x_2, \ldots, x_n) = \sum_i p(x_i)$

$p(x_1, x_2, \ldots, x_n) = \prod_i p(x_i | x_{i-1})$

Question 3   10 out of 10 points

Consider a sequence of length 2, with the following vectors:

$h_1 = [1, 0, -1]$

$h_2 = [2, 1, 0]$

In fact, h_1 and h_2 are the vectorized representations of words 1 and 2 respectively. What is the new representation for the first word, if we apply a self-attention mechanism on this sequence? (Consider the Key, Query, and Value attributes to be the same).

Selected Answer:

✔ $[1.5, 0.5, -0.5]$

Answers:

✔ $[1.5, 0.5, -0.5]$

$[3, 1, -1]$

$[2, 1, 0]$

$[1, 0, -1]$

Question 4   5 out of 10 points

Select all that is correct:

Select all that is correct.

Selected Answers:
❌ It is computationally harder to model positional information in RNNs compared to Transformers.

✓ In an encoder-decoderTransformer architecture for translation, the decoder takes the target language sentence as input during training.

✓ Both RNN and Transformer architectures can theoretically process arbitrary-length input sequences.

Answers:
It is computationally harder to model positional information in RNNs compared to Transformers.

✓ In an encoder-decoderTransformer architecture for translation, the decoder takes the target language sentence as input during training.

✓ Both RNN and Transformer architectures can theoretically process arbitrary-length input sequences.

In a Seq2Seq RNN architecture, Bi-LSTM architecture can be used in the decoder to produce higher-quality outputs.

Question 5   0 out of 10 points

Consider an example of machine translation where the reference sentence and hypothesis sentence both have the same length. Given the following metrics, calculate the F2 score.

precision (2-gram) = 3/7

precision (3-gram) = 2/6

precision (4-gram) = 1/5

BLEU = 0.382

Selected Answer: ❌ 0.457
Correct Answer: ✓ 0.75
Answer range +/- 0.1 (0.65 - 0.85)

Question 6   0 out of 10 points

Select the incorrect items:

Selected Answers:
✓ In modern architectures, RNNs are more popular compared to Transformers because they are more effective in coding positional information.

❌ Self-attention is too expensive to compute and therefore Transformers face serious challenges dealing with long sequences.

Answers:
✓ In modern architectures, RNNs are more popular compared to Transformers because they are more effective in coding positional information.

Transformers can capture word dependencies more easily by design.

Self-attention is too expensive to compute and therefore Transformers face serious challenges dealing with long sequences.

✓ Similar to RNNs, Transformers can be parallelized for computational speed up.

Question 7   0 out of 10 points

Consider the following two sentences:

Reference sentence: "she ate three apples while she was walking in the garden."

MT Model's output: "garden she she"

What is the precision metric for the model's output?

Selected Answer: ❌ 0.333
Correct Answer: ✓ 1
Answer range +/- 0.01 (0.99 - 1.01)

Question 8   10 out of 10 points

Consider the following two sentences:

Reference sentence: "the dark knight rises"

Hypothesis: "the dark night"

What is the Word Error Rate metric evaluated for this hypothesis?

Selected Answer: ✓ 0.5
Correct Answer: ✓ 0.5

Answer range +/- 0 (0.5 - 0.5)

Question 9

Which of the following is true about the IBM machine translation model?

Selected Answer: ✅ It has problems dealing with the pronoun ambiguity problem.

Answers: ✅ It has problems dealing with the pronoun ambiguity problem.

It is flexible to new phrases.

It is capable of handling sentences with varying lengths.

It is capable of understanding the context.

Question 10

Max-pooling layer is a non-differentiable function and therefore it can not be used in neural network architectures.

Selected Answer: ✅ False

Answers: True

✅ False

Saturday, January 20, 2024 11:42:42 PM PST

← OK