

START OF QUIZ

Student ID:

30821250,Huang,Chloe

Question 1

Topic: Lecture 6

Source: Lecture 6

Think back to week 1 of this block when we were doing word sense disambiguation. Do you think there would be benefits or disadvantages to disambiguating all words before running word2vec? Explain. (2)

Question 2

Topic: Lecture 5

Source: Lecture 5

Which is likely to have the highest PMI? A rare word and a frequent word that appear together frequently, or two frequent words that appear together frequently? (1)

Question 3

Topic: Lecture 7

Source: Lecture 7

We took a look at 2 different ways of implementing the TextTiling algorithm - one with vector overlap, and one with BERT. Can you think of how we might modify the algorithm further to strengthen up its weaknesses? (No is not a valid answer.) (2)

Question 4

Topic: Lecture 7

Source: Lecture 7

Explain salience with respect to entities in a sentence (ie, when identifying Cf). (1)

Question 5

Topic: Lecture 6

Source: Lecture 6

Why do we say that the analogy task is an "intrinsic" evaluation of our word embeddings? (1)

Question 6

Topic: Lecture 8

Source: Lecture 8

What is the purpose of an antecedent in anaphoric resolution? (1)

Question 7

Topic: Lecture 8

Source: Lecture 8

Do you think we could use word embeddings for coreference resolution? What kind of assumptions would we be making, and why do you think it might still be a very difficult task?
(2)

Question 8

Topic: Lecture 5

Source: Lecture 5

We've seen co-occurrence matrices weighted by TF-IDF; would it make sense to weight them by PMI? Briefly explain. (1)

Question 9

Topic: Long

Source: Lecture 6

Imagine we were trying to find a word that is the best prototype of its synonyms. Write a short function that grabs the lemmas of each synset in wordnet, and calculates which lemma is the best prototype (ie, which lemma is the closest to the centroid of the synset) by using the word embeddings. Ignore words that do not have embeddings in gensim. (3)

END OF QUIZ