

CS 5316 – Natural Language Processing

Quiz 8

(Time limit: 12 minutes)

1. ((4 points) Given the TF-IDF vectors of four sentences in a document, create a two sentence summary for it. Show / justify your approach.

S1: [0, 0.5, 0, 0.25]

S2: [0.5, 0.5, 0, 0.5]

S3: [0, 0, 0.5, 0]

S4: [0.5, 0.5, 0.25, 0.5]

We will rank the sentences based on their information content. The information content of a sentence can be defined as: sum of its tf-idf values

S1: 0.75, S2: 1.5, S3: 0.5, S4: 1.75

Two sentence summary: S4, S2

2. (2 points) Describe how coherence can be used to order sentences in a summary. How can coherence be measured?

A greedy approach based on coherence uses the following steps:

Compute the relevance of each sentence (generic or query-focused)

Add the most relevant sentence that is most coherent with the existing summary (e.g., score can be product of relevance and coherence)

Coherence of a sentence to be added to a summary can be estimated by the cosine similarity of the sentence with the existing summary.

3. (2 points) List the two classes of conversational agents commonly used in practice

Chatbots: Used for fun, psychological purposes

Goal-based agents: Used for tasks like making reservations, technical support, etc

4. (1 point) What type of summary is generated by a seq2seq model? (a) extractive, (b) abstractive.

Abstractive