

CS5803
Natural Language Processing
MTech in Data Sciences
27-Apr-2024

Instructions:

- Answers should be crisp, and to-the-point. Vague, ambiguous, unnecessarily verbose answers will not be considered.
- Use of electronic items (e.g. mobiles, tablets, laptops, headphones, earphones, smartwatches, etc.) is strictly prohibited.

Question 1: [(2+2)+2+2 = 8 Marks]

- a) What is constituency parsing? Explain in a generic manner keeping in mind its input and the output, without referring to any example. Pick any application where you think constituency parsing could be helpful, and explain how it will help in the application.
- b) How does dependency parsing differ from semantic parsing? Answer clearly. (For example, you should not write the definitions for these parsings and expect the evaluator to find the difference.)
- c) Cite two differences between the resources PropBank and FrameNet.

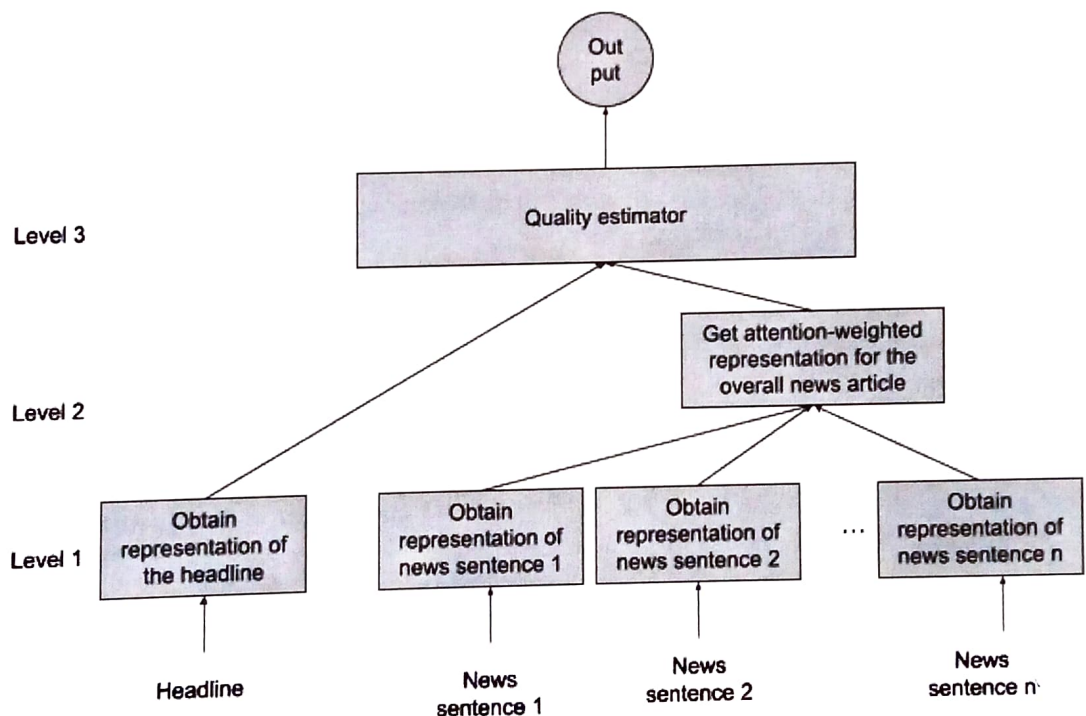
Question 2: [2+3+6+3 = 14 marks]

- a) What is a Context Free Grammar (CFG)? *CNF*
- b) Convert the following grammar into a ~~CFG~~ *CFG*. Assume that the capital letters represent non-terminals/variables, and small letters are for terminals.
 - 1) $S \rightarrow AbC$
 - 2) $A \rightarrow CCd$
 - 3) $C \rightarrow BD \rightarrow C \rightarrow DD$
 - 4) $D \rightarrow aE$
- c) Parse the sentence "avoid drinks with sugar" using CKY parsing. Show the status of the table/matrix after each processing each diagonal. Draw the complete parse tree obtained by you. The grammar rules are as follows:
 - 1) $S \rightarrow NP VP$
 - 2) $VP \rightarrow V NP \mid VP PP$
 - 3) $V \rightarrow \text{drinks} / \text{avoid}$
 - 4) $NP \rightarrow NP PP \mid \text{we} \mid \text{drinks} \mid \text{sugar}$
 - 5) $PP \rightarrow P NP$
 - 6) $P \rightarrow \text{with}$

✓ What is the time complexity of CYK parsing, and why?

Question 3: [5+5 = 10 marks]

13. a) Suppose you have a document collection D , with n documents. Assume that there are two in-built functions $\text{nsent}(d)$, and $\text{sent}(d,i)$, which return the number of sentences in document d , and the i -th sentence in document d respectively. You are creating a training dataset for coherence quality estimator, using self-supervision. What components will be there in each training data instance? Using the above information, write a pseudo-code for creating such training data.
- b) Suppose you are creating an application to determine the goodness of a headline for a given short news. The simple algorithm considers that there are n sentences in the news, and contextual representations are used throughout the algorithm. A block diagram of the algorithm is given below. Write down the mathematical functions that could describe the operations inside the shaded blocks at each level. Essentially, you will have to write the mathematical operations that happen in Level 1, Level 2, and Level 3.



Question 4: [3+5+2 = 10 Marks]

14. a) Amar and Bhaskar have created word2vec embeddings for the same vocabulary of words. For any word w from the vocabulary, the center and context vector

representations obtained by A are given by $u_{w,A}$ and $v_{w,A}$ respectively. Similarly, for the word w the center vector and context vector representations obtained by B are $u_{w,B}$ and $v_{w,B}$. Also, it has been observed that, for each word w , the dot products of the center and context vectors are the same in both A and B's models. Which means, $u_{w,A} \cdot v_{w,A} = u_{w,B} \cdot v_{w,B}$. Does it mean that for each word, A and B have obtained the same representations, i.e. $u_{w,A} = u_{w,B}$ and $v_{w,A} = v_{w,B}$? Answer with justification.

- b) Rahul has obtained the SVD decomposition of a matrix A, to obtain the matrices U, Σ , and V, and had noted down the same on a paper. However, a couple of entries of the V matrix got smudged. The entries are marked as a, b and c in the following matrix. Can you predict what the values of a, b and c were? The only thing that he remembers is that c was a positive entry.

a	$-2/\sqrt{6}$	$-1/\sqrt{6}$
$2/\sqrt{5}$	$-1/\sqrt{5}$	$1/\sqrt{5}$
$1/\sqrt{30}$	b	c

- ✓ LDA and SVD both can give vector representations for different words from a collection. How are these LDA and SVD vectors characteristically different from each other?

Question 5 [2+2+2+2 = 8 Marks]:

- 11
- Name at least two types of features that would be helpful to provide as input to a neural dependency parser and explain why.
 - A news article has an image associated with it. You are developing an algorithm to generate a caption for the image. Do you see co-reference chains can be used for this task? If yes, briefly explain how.
 - ✓ Suppose you have to create data for training a system for identifying elementary discourse units, using BIOES technique. Create the training instances for the input: "Tim said Bill was eating while Anne was baking the cake."
 - ✓ How can entity grid based mechanism be used to identify valid discourses?

Question 6 : [8+6+8 = 22 Marks]:

PS. Consider the following sentences in Hindi, with each word annotated using a named entity tag {Per, Loc, Org, Other}

Roger/Per Hyderabad/Loc daure/Other per/Other hein/other "
Money/Other to/Other Swiss/Org bank/Org mein/Other the/Other"
Roger/Per Swiss/Loc se/Other hei/Other

- ① [Marks : 3+3+2 = 8] Consider you want to develop a NER based on the Hidden Markov Model. Provide the transition probabilities and emission probabilities in the form of a Table. What is the probability of assigning the tag Roger/Per Swiss/Org the/Other vs Roger/Per Swiss/Loc the/Other based on the HMM model. Provide clearly the probability computations.
- ✓ [Marks : 3+3 = 6] Assume you want to develop a Conditional Random Field model for solving the problem. Provide the expression for predicting the output sequence Y from input sequence X. How is this different from Logistic regression ? Provide some set of global features that will be useful for this problem.
- ✓ [Marks : 5 + 3 = 8] Assume you want to develop a deep learning model to solve the problem of NER. More specifically, you want to use a encoder-decoder model with CNN encoder and LSTM decoder. Provide the architecture details and explain with a diagram (you may choose number of layers, width etc. appropriately to fit the diagram). Provide the training loss you will use to train the model, expressing in terms of the sequence of output labels y_1, y_2, \dots, y_L .

Question 7 : [3+5 = 8 Marks]:

- ✓ Explain with an example the functionality of the Dialogue state tracking Module in the Task Oriented Dialogue systems. Why do you think CopyNet is an appropriate deep learning model to solve the DST task based on the example you provide.

Provide the architecture and functioning of the CopyNet model and explain how it achieves Generate and Copy operations.