# NATIONAL INSTITUTE OF TECHNOLOGY ROURKELA

**End Semester Examination, Spring 2022**

**Subject: Natural Language Processing** **Subject Code: CS6314** **Full Marks: 50**

**Time: 3 hours** **Number of pages: 2**

**Questions 1 – 7 are mandatory. Any marks fetched in Question 8 will be added to the secured marks and scaled to 50. Mere answer without justification will not fetch any mark.**

1. Consider the following two matrices $A$ and $B$ for transition probabilities and observation probabilies respectively. $A[start, verb] = 0.1$ denotes that the probability of transition from start state to verb state is 0.1. Similarly, $B[noun, gave] = 0.001$ denotes that the probability of observation of the word gave from noun state is 0.001. 8 marks

**A** =

| | *verb* | *noun* | *propnoun* | *det* | *end* |
|---|---|---|---|---|---|
| *start* | 0.1 | 0 | 0.4 | 0.5 | 0 |
| *verb* | 0 | 0 | 0.25 | 0.25 | 0.5 |
| *noun* | 0.9 | 0 | 0 | 0 | 0.1 |
| *propnoun* | 0.8 | 0 | 0 | 0.1 | 0.1 |
| *det* | 0.05 | 0.95 | 0 | 0 | 0 |

$B$ =

| | *Ram* | *gave* | *an* | *umbrella* |
|---|---|---|---|---|
| *verb* | 0.01 | 0.1 | 0.001 | 0.001 |
| *noun* | 0.01 | 0.001 | 0.001 | 0.1 |
| *propnoun* | 0.1 | 0.001 | 0.001 | 0.01 |
| *det* | 0.01 | 0.001 | 0.1 | 0.01 |

Use the viterbi algorithm to find the most likely sequence for the sentence “Ram gave umbrella”.

2. Use the CKY algorithm to generate the parse tree for the sentence “fish people fish tanks”. 8 marks
The probabilistic context free grammar in CNF with unarys is given in the figure below. When the start position is 0 and end position is 1, the sentence is “fish”. Looking at the grammar, we see that there are two rules that generate “fish” directly: $N \rightarrow fish$ with probability 0.2, and $V \rightarrow fish$ with probability 0.6. We place these two transitions in the location 0 to 1 corresponding to fish. There is a rule $NP \rightarrow N$ with probability 0.7 that can combine with $N \rightarrow fish$ with probability 0.2 to generate fish from $NP$ with probability $0.7 \times 0.2 = 0.14$. Similarly, all the single word entries are pre-filled in the table. Fill the remaining entries in the table and generate the desired parse tree if possible from the grammar.

| Grammar Rules | Corresponding Probabilites |
|---|---|
| S → NP VP | 0.9 |
| S → VP | 0.1 |
| VP → V NP | 0.5 |
| VP → V | 0.1 |
| VP → V @VP_V | 0.3 |
| VP → V PP | 0.1 |
| @VP_V → NP PP | 1.0 |
| NP → NP NP | 0.1 |
| NP → NP PP | 0.2 |
| NP → N | 0.7 |
| PP → P NP | 1.0 |
| N → *people* | 0.5 |
| N → *fish* | 0.2 |
| N → *tanks* | 0.2 |
| N → *rods* | 0.1 |
| V → *people* | 0.1 |
| V → *fish* | 0.6 |
| V → *tanks* | 0.3 |
| P → *with* | 1.0 |

| | fish (0–1) | people (1–2) | fish (2–3) | tanks (3–4) |
|---|---|---|---|---|
| 0 | N → fish 0.2<br>V → fish 0.6<br>NP → N 0.14<br>VP → V 0.06<br>S → VP 0.006 | | | |
| 1 | | N → people 0.5<br>V → people 0.1<br>NP → N 0.35<br>VP → V 0.01<br>S → VP 0.001 | | |
| 2 | | | N → fish 0.2<br>V → fish 0.6<br>NP → N 0.14<br>VP → V 0.06<br>S → VP 0.006 | |
| 3 | | | | N → tanks 0.2<br>V → tanks 0.1<br>NP → N 0.14<br>VP → V 0.03<br>S → VP 0.003 |

3. In the graph based dependency parsing, we discussed the Chu-Liu-Edmonds algorithm of constructing the dependency tree from a directed graph. Simulate the algorithm in the directed graph given below as a matrix $C$ and obtain the desired dependency tree. 8 marks

$$C = \begin{array}{cccccc} & book & that & flight & to & boston \\ root & 15 & 5 & 6 & 7 & 8 \\ book & - & 50 & 100 & 120 & 10 \\ that & 8 & - & 200 & 81 & 15 \\ flight & 11 & 11 & - & 150 & 16 \\ to & 14 & 100 & 50 & - & 14 \\ boston & 14 & 22 & 51 & 79 & - \end{array}$$

Here $C$ denotes the adjacency matrix of the underlying directed graph with vertices $\{root, book, that, flight, to, boston\}$. The entry in the matrix $C[x, y]$ denotes the weight of the edge starting from $x$ and ending at $y$. If $C[x, y]$ is unspecified for any pair $(x, y)$, then edge $(x, y)$ is not present in the graph. During expansion of tree, assume that dependency tree is constructed by adding one zero weight edge at a time starting from the $root$ vertex.

4. Consider the english sentence $e$="he was eating" and a sentence in foreign language $f$="woh kha raha tha". The translation probabilities are shown below. 6 marks

$$D = \begin{array}{ccccc} & null & he & was & eating \\ main & 0.1 & 0.1 & 0.01 & 0.01 \\ woh & 0.1 & 0.4 & 0.01 & 0.01 \\ kha & 0.1 & 0.1 & 0.01 & 0.3 \\ raha & 0.1 & 0.1 & 0.01 & 0.1 \\ tha & 0.1 & 0.1 & 0.1 & 0.05 \end{array}$$

Here $D[main, was] = 0.01$ denotes that the translation probability $t(main|was) = 0.01$.

We are using IBM Model 1 to generate $f$ from $e$. Assuming length 4 for $f$ was chosen with probabilty 0.1, pick an arbitrary alignment $a$ and compute the probabilities $p(f, a|e)$ and $p(f|a, e)$.

5. Consider the english sentence $e$="my house" and a sentence in foreign language $f$="apna ghar". The translation probabilities are shown below. 7 marks
$p(apna|my) = 0.7$, $p(ghar|my) = 0.05$, $p(apna|house) = 0.1$, $p(ghar|house) = 0.8$.
Use expectation maximization algorithm for two steps to determine the correct alignments (you need to consider all 4 possible alignments; do not use null in $e$).

6. Consider the the problem of word sense disambiguation of the word "play". There are 3 documents in the corpus and they are either labelled s or m. 6 marks

| *Document* | *label* |
|---|---|
| play like sachin | $s$ |
| the play was based on sachin | $m$ |
| play cricket | $s$ |

Choose the correct label for documents (1) sachin was play and (2) sachin was the. Use Näive Bayes with add-one smoothing.

7. A self attention layer maps the vectors $x_1, x_2, x_3 \in \mathbb{R}^6$ to $y_1, y_2, y_3 \in \mathbb{R}^3$. 7 marks
Compute $y_1$ using the data provided below.
$x_1 = [0, 0.9, 0.8, 0.2, 0.5, 0.8]$, $x_2 = [0.3, 0.8, 0.5, 1, 0.8, 0.2]$, $x_3 = [0.2, 0.7, 0.2, 0.8, 0, 0.3]$.

$$W^Q = \begin{bmatrix} 0.8 & 0.6 & 0.5 & 0 & 0.2 & 0.3 \\ 0.5 & 0.6 & 0.5 & 0.3 & 0.3 & 0 \\ 0.2 & 0.5 & 0 & 0.1 & 0.7 & 0.2 \\ 1 & 0.4 & 0.6 & 0.9 & 0.1 & 0.6 \\ 0.5 & 1 & 0.8 & 0.1 & 0.5 & 0 \end{bmatrix} \quad W^K = \begin{bmatrix} 0.2 & 0.2 & 0.6 & 0.5 & 0.3 & 0.2 \\ 0.4 & 0.8 & 0.7 & 0.5 & 0.6 & 0.7 \\ 0.4 & 0.7 & 0 & 0.9 & 0.1 & 0.9 \\ 0.8 & 0.3 & 0.8 & 0.2 & 0.6 & 1 \\ 0.8 & 0.6 & 0.6 & 0.9 & 0.2 & 0.6 \end{bmatrix} \quad W^V = \begin{bmatrix} 1 & 0.3 & 0.5 & 0.7 & 0.7 & 0 \\ 0.7 & 0.5 & 0.2 & 0.3 & 0.6 & 0.2 \\ 0.6 & 0.2 & 0.6 & 0.8 & 0.9 & 0.1 \end{bmatrix}$$

8. Explain the term project you have submitted this semester. Provide the problem definition, literature survey, the architecture proposed in the paper, and results. If it was a survey paper, outline the main papers on the problem, the architectures proposed and a comparison of the results. 6 marks

$- - \bigstar - -$