Natural Language Processing

Assignment 5

Type of Question: MCQ

Number of Questions: 9 **Total Marks: (8×1)+(1×2)=10**

Question 1: Which of the following is/are false? (1 mark)

a. PCFG is a better language model for English than an n-gram model.
b. The probability of a smaller tree is greater than a larger tree.
c. The Inside-Outside algorithm is faster compared to HMMs.
d. In CKY algorithm, grammar must be converted to Chomsky normal form (CNF).

Answer: a, c
Solution:

Question 2: Consider the CFG given below:
$S \rightarrow aSb|D$
$D \rightarrow Dc|\epsilon$
How many non-terminals should be added to convert the CFG into CNF? (1 mark)

a. 3
b. 2
c. 4
d. 5

Answer: d
Solution:
The final CNF is:
$S' \rightarrow AE|AB|DC|c$
$S \rightarrow AE|AB|DC|c$
$E \rightarrow SB \quad D \rightarrow DC|c$
$A \rightarrow a \quad B \rightarrow b \quad C \rightarrow c$

For question 3 to 8 consider the following PCFG fragment:

$S \rightarrow$ NN VP	0.50	$S \rightarrow$ VP NN	0.50
$NP \rightarrow$ NN PB	0.40	$PB \rightarrow$ PP NN	0.30
$VP \rightarrow$ VB NN	0.30	$VP \rightarrow$ VB NP	0.20
$VP \rightarrow$ NN VB	0.25	$VP \rightarrow$ NN PB	0.15
$PP \rightarrow$ with	0.10	$PP \rightarrow$ without	0.10
$VB \rightarrow$ play	0.30	$VB \rightarrow$ enjoy	0.20
$VB \rightarrow$ watch	0.25	$NN \rightarrow$ childern	0.15
$NN \rightarrow$ cricket	0.15	$NN \rightarrow$ friends	0.20
$NN \rightarrow$ football	0.10	$NN \rightarrow$ music	0.12

For a sentence $S = w_1w_2w_3w_4$, assume that the cells in the table are indexed as follows:

	1	2	3	4	
w_1	11	12	13	14	1
	w_2	22	23	24	2
		w_3	33	34	3
			w_4	44	4

Question 3: Using CKY algorithm, find the probability score for the most probable tree for the sentence S_1 = **"children play cricket with friends"**. (1 mark)

a. 5.06×10^{-4}
b. 2.73×10^{-3}
c. 1.62×10^{-6}
d. None of the above

Answer: c
Solution:
Calculate the probability using the Bottom-Up method as explained in the lecture.

Question 4: How many entries are generated while filling the parsing table with CKY algorithm in Question 3? Note: There can be multiple entries in one cell.
(1 mark)

a. 12
b. 13
c. 14
d. 15

Answer: c
Solution:

Question 5: Using CKY algorithm, find the number of parse trees for the sentence S_2 = **children enjoy music** and the probability score for the most probable tree.
(2 marks)

a. 1, 4.95×10^{-3}
b. 2, 0.36×10^{-3}
c. 3, 0.99×10^{-3}
d. 2, 0.54×10^{-3}

Answer: d
Solution:
There are two parse trees.
$S \rightarrow NN_{11}\ VP_{23} = 0.5 \times 0.15 \times (0.3 \times 0.2 \times 0.12) = 0.54 \times 10^{-3}$
$S \rightarrow VP_{12}\ NN_{33} = 0.5 \times (0.25 \times 0.15 \times 0.2) \times 0.12 = 0.45 \times 10^{-3}$

We get the above probabilities with CKY algorithm.

Question 6: Using the Inside Algorithm, find the probability for generating the sentence S_2 = **children enjoy music**. (1 mark)

a. 0.99×10^{-3}
b. 1.10×10^{-3}
c. 0.55×10^{-3}
d. 0.78×10^{-3}

Answer: a
Solution: Refer to solution 5. Add the probabilities of the two trees.

Question 7: Consider the expression below:
P("children watch football enjoy music", $N_{34}|G$) = $\sum_j$ P("children watch football enjoy music" $|N_{34}^j, G$)
What does the L.H.S. represent? (1 mark)

a. Probability of the sentence "children watch football enjoy music", given a grammar G.
b. Probability of the sentence "children watch football enjoy music", given a grammar G and some rule which derives the segment "football enjoy".
c. Probability of the sentence "children watch football enjoy music", given a grammar G and that there is some consistent spanning of the segment "football enjoy", i.e. from word 3 to 4.
d. None of the above

Answer: c
Solution: Refer to Inside-Outside Probabilities.

Question 8: Suppose after parsing the sentence S_2 = **children enjoy music** with CKY algorithm, the non-terminals that appear in position 12 and 23 are NT_1 and NT_2 respectively. Compute the outside probabilities for $\alpha_{NT_1}(12)$ and $\alpha_{NT_2}(23)$. (1 mark)

a. 0, 0.075
b. 0.25, 0
c. 0.30, 0.06
d. None of the above

Answer: d
Solution:
$\alpha_{VP}(12) = 0.5 \times 1 \times 0.12$ =0.060
$\alpha_{VP}(23) = 0.5 \times 1 \times 0.15$ =0.075

Question 9: Which of the following grammars are valid CNF? (1 mark)

1. A → B
   B → CD
   C → c
   D → d

2. A → BCD
   B → b
   C → c
   D → d

3. A → BC
   B → ϵ
   C → c

a. 1.
b. 2.
c. 3.
d. None of the above

Answer: d
Solution: Valid CNF form is as follows:
$A \rightarrow BC$
$A \rightarrow a$
