Indraprastha Institute of Information Technology, Delhi — **Monsoon Semester 2018**

End Semester Examination in CSE556: Natural Language Processing — **Time: 3:00 - 5:00**

26th November, 2018 — **Full Marks: 50**

Subhankar

Q1. [4 marks]

System generated summary of a text document: the cat was found under the bed
Reference summary for that text document: the cat was under the bed
Calculate the values of ROUGE-3 precision and ROUGE-3 recall. (No need to consider start and end sentence tags)

Subhankar

Q2. [4 marks]

In a question answering system, 50 questions are asked. The system returned a ranked answer list of 50 answers for each question asked. It was found that 1st answer of 1st question, 2nd answer of 2nd question, 3rd answer of 3rd question, 4th answer of 4th question,..., 40th answer of the 40th question are correct; and none of the answers of last 10 questions are correct. Calculate the "Mean Reciprocal Rank" for the system.

Q3. [4 marks]

Shubhrata

	Count(w,context)				
	computer	data	pinch	result	sugar
apricot	0	0	1	0	1
pineapple	0	0	1	0	1
digital	2	1	0	1	0
information	1	6	0	4	0

What is the value of PMI(information,data)? Show each of the components to measure the PMI value.

Subhankar

Q4. [2+1+1=4 marks]

Like NLP, IR will also be taught by Tanmoy. NLP was interesting course. IR will also be interesting. IR may have some concepts in common with NLP. Basic math is prerequisite for these subjects.

(i) Create a term-context matrix based on the above text, with context size of 10 words window (non-overlapping, new window starts from the next word to the last window. Ignore " ." & "," for window length). Consider only the 4-dimensional vectors of these words {NLP, IR, Math, interesting} as dimension for the vectors.

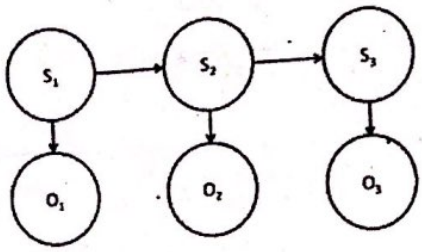
(ii) Calculate the cosine similarity between (IR and NLP) & between (IR and Maths).

(iii) Calculate the Jaccard similarity between (IR and NLP) & between (IR and Maths).

Ramesh

Q5. [6 marks]

Use Viterbi algorithm to find the most likely sequence for the given setting:



State	$P(S_1)$
A	0.99
B	0.01

(a) Initial probs.

S_1	S_2	$P(S_2\|S_1)$
A	A	0.99
A	B	0.01
B	A	0.01
B	B	0.99

(b) Transition probs.

S	O	$P(O\|S)$
A	0	0.8
A	1	0.2
B	0	0.1
B	1	0.9

(c) Emission probs.

Vishal

Q6. [5 marks]

Let $S = \{a, b, c\}$ (the sample space), and p be the joint distribution on a sequence of two events (i.e., on $S \times S$, ordered). If you know that $p(a,a)$ [a followed by a] = 0.25, $p(c,c)$ [c followed by c] = 0.25, $p(b,a)$ [b followed by a] = 0.125, $p(b,b)$ [b followed by b] = 0, $p(a,c)$ [a followed by c] = 0.25, $pL(a)$ [unigram probability of a as a left-hand bigram member] = 0.5, and $pR(b)$ [unigram probability of b as the right-hand bigram member] = 0.125, is it enough to compute $p(b|c)$ (i.e., the probability of seeing b if we already know that the preceding event generated c)? If yes, how?

Ramesh

Q7. [2+2=4 marks]

Model A: Set of possible tags {Tag1, Tag2, Tag3, Tag4}

Tag1	Tag2	Tag3	Tag4	Tag5	Tag6	Tag7	Tag8
1/5	1/5	2/5	1/5	0	0	0	0

Model B: Set of possible tags {Tag1, Tag2, Tag3, Tag4, Tag5, Tag6}

Tag1	Tag2	Tag3	Tag4	Tag5	Tag6	Tag7	Tag8
1/6	1/6	1/6	1/6	1/6	1/6	0	0

Which of the two models would you use for an MEMM (no calculations needed)? Why is the model with maximum entropy used in MEMM?

Vishal

Q8. [2+2=4 marks]

Given the following short movie reviews, each labelled with a genre, either COMEDY or ACTION:

1. fun, couple, love, love => COMEDY

2. fast, furious, shoot => ACTION
3. couple, fly, fast, fun, fun => COMEDY
4. furious, shoot, shoot, fun => ACTION
5. fly, fast, shoot, love => ACTION

and a new document D: fast, couple, shoot, fly

Compute the most likely class for D. Assume a naïve Bayes classifier and use add-1 smoothing for the likelihoods.

Shubhrata

Q9: [3+3=6 marks]

Convert the following CFGs to CNF. Here, all non-terminals are in uppercase (e.g X,Y) and all terminals are in lowercase (e.g. a, b)

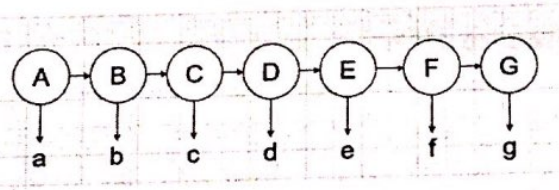
(i) $S \rightarrow aXbX, X \rightarrow aY \mid bY \mid \varepsilon, Y \rightarrow X \mid c$

(ii) $S \rightarrow ASA \mid aB, A \rightarrow B \mid S, B \rightarrow b \mid \varepsilon$

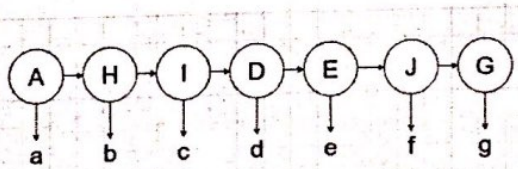
Mohit

Q10: [9×1=9 marks]

(i) In HMM the probability of a tag depends on its previous tag. So the probability of tag A depends on tag B's probability which in turn depends on tag C and so on. How many probabilities are there that differ between (a) and (b) mentioned below:



(a)



(b)

(ii) Define Meronymy with an example.

(iii) What is the time complexity of CYK algorithm?

(iv) What is the major advantage of MEMM over HMM?

(v) You have created a document term-matrix of the data, treating every tweet as one document. Which of the following is correct, in regards to document term matrix?

1. Removal of stopwords from the data will affect the dimensionality of data
2. Normalization of words in the data will reduce the dimensionality of data
3. Converting all the words in lowercase will not affect the dimensionality of the data

(vi) When do we avoid using Word2Vec model?

(vii) Give an example where you find syntactic ambiguity.

(viii) If you were to design a trigram language model, how would the final smoothed distribution be defined if you use the linear interpolation smoothing method?

(ix) What was your attendance in the NLP'18 course at IIITD (choose one from the following)

(a) >=95%, (b) >=85% and <95%, (c) >=75% and <85%, (d) >=65% and <75%, (e) <65%