Natural Language Processing

Assignment- 1

TYPE OF QUESTION: MCQ

Number of questions: 10 **Total mark: 10 X 1 = 10**

Question 1: In a corpus, you found that the word with rank 4th has a frequency of 500. What can be the best guess for the rank of a word with frequency 250?

1. 2
2. 4
3. 8
4. 6

Answer: 3

Solution:
frequency * rank =k [by Zipfs law]
500*4 = 250*r
r = 8

Question 2: In the sentence, "In Mumbai I took my hat off. But I can't put it back on.", total number of word tokens and word types are:

1. 14, 13
2. 13, 14
3. 15, 14
4. 14, 15

Answer: a) 14, 13.
Solution: Here, the word "I" is repeated two times so type count is one less than token count.

Question 3: Let the rank of two words, w1 and w2, in a corpus be 400 and 100, respectively. Let m1 and m2 represent the number of meanings of w1 and w2 respectively. The ratio m1 : m2 would tentatively be

1. 1:4
2. 4:1
3. 1:2
4. 2:1

Answer: 3

Solution:
m1/m2 = sqrt(rank2)/sqrt(rank1) = sqrt(100)/sqrt(400) = 1:2

Question 4: What is the valid range of type-token ratio of any text corpus?

1. TTR$\in$(0,1] (excluding zero)
2. TTR$\in$[0,1]
3. TTR$\in$[−1,1]
4. TTR$\in$[0,+∞] (any non-negative number)

Answer: 1.
Solution: Number of unique words or type ≤ Total number of tokens in text, and both are greater than 1

Question 5: If first corpus has $TTR_1 = 0.075$ and second corpus has $TTR_2 = 0.15$, where TTR_1 and TTR_2 represents type/token ratio in first and second corpus respectively, then

1. First corpus has more tendency to use different words.
2. Second corpus has more tendency to use different words.
3. Both a and b
4. None of these

Answer: b
Solution: Second corpus has more tendency to use different words. If TTR scores are higher then there is more tendency to use different words.

Question 6: Which of the following is/are true for the English Language?

1. Lemmatization works only on inflectional morphemes and Stemming works only on derivational morphemes.
2. The outputs of lemmatization and stemming for the same word might differ.
3. Output of lemmatization are always real words
4. Output of stemming are always real words

Answer: 2, 3
Solution: *Stemming* usually refers to a crude heuristic process that chops off the ends of words in the hope of achieving this goal correctly most of the time, and often includes the removal of derivational affixes. *Lemmatization* usually refers to doing things properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the *lemma* .

Question 7: An advantage of Porter stemmer over a full morphological parser?

1. The stemmer is better justified from a theoretical point of view
2. The output of a stemmer is always a valid word
3. The stemmer does not require a detailed lexicon to implement
4. None of the above

Answer: 3
Solution: The Porter stemming algorithm is a process for removing suffixes from words in English. The Porter stemming algorithm was made on the assumption that we don't have a stem dictionary (lexicon) and that the purpose of the task is to improve Information Retrieval performance. Stemming algorithms are typically rule-based. You can view them as a heuristic process that sort-of lops off the ends of words.

Question 8: Which of the following are instances of stemming? (as per Porter Stemmer)

1. are -> be
2. plays -> play
3. saw -> s
4. university -> univers

Answer: 2,4
Solution: Stemming cannot convert are->be as it can only convert or chop off word suffixes. Also Porter Stemmer wouldn't chop off if the final outcome is of length 1 as in saw -> s.

Question 9: What is natural language processing good for?

1. Summarize blocks of text
2. Automatically generate keywords
3. Identifying the type of entity extracted
4. All of the above

Answer: 4

Solution:
For all the above-mentioned task, NLP can be used

Question 10: What is the size of unique words in a document where total number of words = 12000. K = 3.71 Beta = 0.69?

1. 2421
2. 3367
3. 5123
4. 1529

Answer: 1

Solution: 3.71 x 12000^0.69 = **2421** unique words. Heap's Law

---

**********END*******

Natural Language Processing

Assignment- 2

TYPE OF QUESTION: MCQ

Number of questions: 10 **Total mark: 10 X 1 = 10**

QUESTION 1:

According to Zipf's law which statement(s) is/are correct?

(i) A small number of words occur with high frequency.

(ii) A large number of words occur with low frequency.

a. Both (i) and (ii) are correct

b. Only (ii) is correct

c. Only (i) is correct

d. Neither (i) nor (ii) is correct

Correct Answer: a

Solution:

QUESTION 2:

Consider the following corpus C_1 of 4 sentences. What is the total count of unique bi-grams for which the likelihood will be estimated? Assume we do not perform any pre-processing.

today is Sneha's birthday
she likes ice cream
she is also fond of cream cake
we will celebrate her birthday with ice cream cake

a. 24
b. 28
c. 27
d. 23

Correct Answer: a

Detailed Solution:

Unique bi-grams are:

<s> today	today is	is Sneha's	Sneha's birthday	birthday <\s>
<s> she	she likes	likes ice	ice cream	cream <\s>
She is	is also	also fond	fond of	of cream
cake <\s>	<s> we	we will	will celebrate	celebrate her
her birthday	birthday with	with ice	cream cake	

QUESTION 3:

A 3-gram model is a _____ order Markov Model.

a. Two
b. Five
c. Four
d. Three

Correct Answer: a

Detailed Solution:

QUESTION 4:

Which of these is/are - valid Markov assumption?

a. The probability of a word depends only on the current word.

b. The probability of a word depends only on the previous word.

c. The probability of a word depends only on the next word.

d. The probability of a word depends only on the current and the previous word.

Correct Answer: a, c, d

Solution:

QUESTION 5:

For the string **'mash'**, identify which of the following set of strings have a Levenshtein distance of 1.

a. smash, mas, lash, mushy, hash
b. bash, stash, lush, flash, dash
c. smash, mas, lash, mush, ash
d. None of the above

Correct Answer: c

Detailed Solution:

QUESTION 6:

Assume that we modify the costs incurred for operations in calculating Levenshtein distance, such that both the insertion and deletion operations incur a cost of 1 each, while substitution incurs a cost of 2. Now, for the string **'lash'** which of the following set of strings will have an edit distance of 1?

a. ash, slash, clash, flush
b. flash, stash, lush, blush,
c. slash, last, bash, ash
d. None of the above

Correct Answer: d

Detailed Solution:

QUESTION 7:

Given a corpus C_2, the Maximum Likelihood Estimation (MLE) for the bigram "dried berries" is 0.4 and the count of occurrence of the word "dried" is 680. for the same corpus C_2, the likelihood of "dried berries" after applying add-one smoothing is 0.05. What is the vocabulary size of C_2?

a. 4780
b. 3795
c. 4955
d. 3995

Correct Answer: a

Detailed Solution:

$$P_{MLE}(berries \mid dried) = \frac{C(dried, berries)}{C(dried)}$$

0.4 = C(dried, berries) / 680
C(dried, berries) = 680*0.4 = 272

$$P_{Add-1}(berries|\ dried) = \frac{C(dried, berries)+1}{C(dried)+\ V}$$

0.05 = (272+1) / (680+V)
V=4780

For Question 8 to 10, consider the following corpus C_3 of 3 sentences.

there is a big garden
children play in a garden
they play inside beautiful garden

QUESTION 8:

Calculate **P(they play in a big garden)** assuming a bi-gram language model.

a. 1/8
b. 1/12
c. 1/24
d. None of the above

Correct Answer: b

Detailed Solution:

P(they | <s>) = 1/3
P(play | they) = 1/1
P(in | play) = 1/2
P(a | in) = 1/1
P(big | a) = 1/2
P(garden | big) = 1/1
P(<\s>|garden) = 3/3
P(they play in a big garden) = 1/3 x 1/1 x 1/2 x 1/1 x 1/2 x 1/1 x 3/3 = 1/12

QUESTION 9:

Considering the same model as in Question 7, calculate the perplexity of **<s> they play in a big garden <\s>.**

a. 2.289
b. 1.426
c. 1.574

d. 2.178

Correct Answer: b

Detailed Solution:

$perplexity = \sqrt[7]{12} = 1.426$

QUESTION 10:

Assume that you are using a bi-gram language model with add one smoothing. Calculate **P(they play in a beautiful garden).**

a. 4.472 x 10^-6
b. 2.236 x 10^-6
c. 3.135 x 10^-6
d. None of the above

Correct Answer: b

Detailed Solution:

|V|=11
P(they | <s>) = (1+1)/(3+11)
P(play | they) = (1+1)/(1+11)
P(in | play) = (1+1)/(2+11)
P(a | in) = (1+1)/(1+11)
P(beautiful | a) = (0+1)/(2+11)
P(garden | beautiful) = (1+1)/(1+11)
P(<\s>|garden) = (3+1)/(3+11)
P(they play in a beautiful garden) = 2/14 x 2/12 x 2/13 x 2/12 x 1/13 x 2/12 x 4/14
= 2.236 x 10^-6

**********END*******

Natural Language Processing

Assignment- 3

TYPE OF QUESTION: MCQ

Number of questions: 10 **Total mark: 10 X 1 = 10**

Question 1: Let's assume the probability of rolling three (3), two times in a row of a uniform dice is p. Consider a sentence consisting of N random digits. A model assigns probability to each of the digit with the probability p. Find the perplexity of the sentence.

1. 10
2. 6
3. 36
4. 3

Answer - 3

Solution: Probability of rolling 3 two times in a row is(⅙)*(⅙) = (1/36) . Then perplexity is ((1/36)^N)^-1/N = 36

Question 2: Which of the following is false?

1. Derivational morphology creates new words by changing part-of-speech
2. Inflectional morphology creates new forms of the same word
3. Reduplication is not a morphological process
4. Suppletion is a morphological process

Answer - 3

Solution: Reduplication is also one of the morphological process

---

Question 3: Assume that “x” represents the input and “y” represents the tag/label. Which of the following mappings are correct?

1. Generative Models - learn Joint Probability p(x, y)
2. Discriminative Models - learn Joint Probability p(x, y)
3. Generative Models - learn Posterior Probability p(y | x) directly
4. Discriminative Models - learn Posterior Probability p(y | x) directly

Answer: 1, 4
Solution: Generative classifiers learn a model of the joint probability p(x, y) and make their predictions by using Bayes rules to calculate p(y | x). Discriminative classifiers model the posterior p(y | x) directly, or learn a direct map from inputs x to the class labels y.

Question 4: Which one of the following is an example of the discriminative model?

1. Naive Bayes
2. Bayesian Networks
3. Hidden Markov models
4. Logistic Regression

Answer - 4
Solution: Others model in the option are generative model

Question 5. Natural language processing is essentially the study of the meaning of the words a human says or writes. Natural language processing is all around us all the time, but it also happens to be a way to improve the chatbot or product we interact with on a regular basis. Natural language processing is all about mimicking our own language patterns. Natural language processing can also improve the efficiency of business transactions and customer care. Natural language processing is the application of computer technology.

Suppose we want to check the probabilities of the *final words* that succeed the *string* language processing in the above paragraph. Assume d= 0; it is also given that no of unigrams = 78, no of bigrams = 122, no of trigrams = 130,, Question 6 and Question 7 are related to Question 5 corpus.

Solve the question with the help of **Kneser-Ney backoff technique.**

What is the continuation probability of *"is"* ?

1. 0.0078
2. 0.0076
3. 0.0307
4. 0.0081

Answer: 2

Solution: Refer week 3 lecture 12

Continuation probability of is = 1/130 = 0.0076

The numerator means the number of different string types preceding the final word, (here only 1 type– language processing is) and the denominator means the number of different possible n-gram types , in this case trigram = 130

Question 6: What will be the value of P(is| language processing) using Kneser-Ney backoff technique and choose the correct answer below. . Please follow the paragraph in Question .

1. 0.5
2. 0.6
3. 0.8
4. 0.7

Answer: 3

Solution: Refer week 3 lecture 12

P(is| language processing) = ⅘ + 0*0.0076 = 0.8 [as d=0 so lambda = 0]

In this example is equal to the frequency of language processing *: the frequency of language processing is (here it occurs 4 times) plus the frequency of language processing can (occurs only once). Therefore, for word is, firstTerm(is) = 4/(4+1) = 0.8

Question 7. What is the value of P(can| language processing)? Please follow the paragraph in Question 5

1. 0.1
2. 0.02
3. 0.3
4. 0.2

Answer: 4

Solution: Refer week 3 lecture 12

Similarly P(can| language processing) = ⅕+ 0*Continuation probability = 0.2

Language processing * occurs 5 times, language processing can occurs only once.

Question 8: Which of the following morphological process is true for motor+hotel → motel?

1. Suppletion
2. Compounding
3. Blending
4. Clipping

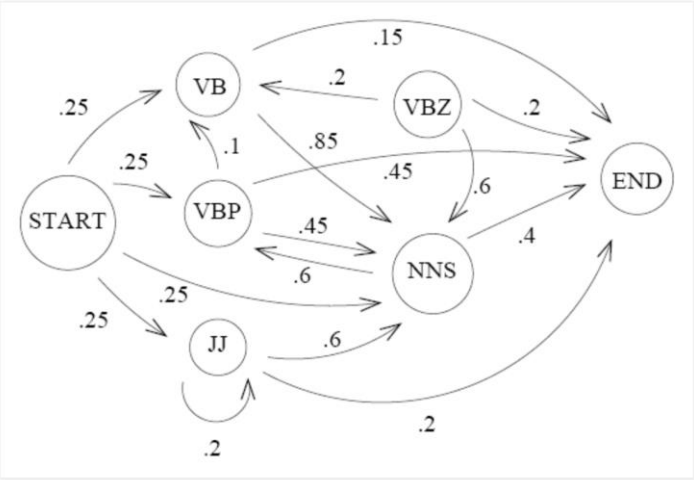
Answer: 3

Solution:
Blending combines parts of two different words. Refer Lecture 13

Question 9: Consider the HMM given below to solve the sequence labeling problem of POS tagging. With that HMM, calculate the probability that the sequence of words "free workers" will be assigned the following parts of speech;

VB NNS

	free	*workers*
JJ	**0.00158**	**0**



NNS	**0**	**0.000475**
VB	**0.0012**	**0**
VBP	**0.00081**	**0**
VBZ	**0**	**0.00005**

The above table contains emission probability and the figure contains transition probability

1. $4.95 * 10^{-8}$
2. $9.80 * 10^{-8}$
3. $4.84 * 10^{-8}$
4. $3.94 * 10^{-7}$

Answer: 3

Solution:

P(free workers, VB NNS)

= P(VB|*start*) * P(*free*|VB) * P(NNS|VB) * P(*workers*|NNS)

* P(end|NNS)

= 0.25 * 0.0012 * 0.85 * 0.000475 * 0.4

= $4.84 * 10^{-8}$

Question 10: Which of the following is/are true?

1. Only a few non-deterministic automation can be transformed into a deterministic one
2. Recognizing problem can be solved in quadratic time in worst case
3. Deterministic FSA might contain empty (ε) transition
4. There exist an algorithm to transform each automation into a unique equivalent automation with the least no of states

Answer: 4

Solution: Refer Lecture 13

Natural Language Processing

Assignment 4

Type of Question: MCQ

Number of Questions: 7 Total Marks:(4×1)+(3×2)=10

==========

1. Baum-Welch algorithm is an example of - **[Marks 1]**
A) Forward-backward algorithm
B) Special case of the Expectation-maximisation algorithm
C) Both A and B
D) None

Answer: C

Solution: Theory.

==========

2. Once a day (e.g. at noon), the weather is observed as one of state 1: rainy state 2: cloudy state 3: sunny The state transition probabilities are :

0.4	0.3	0.3
0.2	0.6	0.2
0.1	0.1	0.8

Given that the weather on day 1 (t = 1) is sunny (state 3), what is the probability that the weather for the next 7 days will be “sun-sun-rain-rain-sun-cloudy-sun”?
[Marks 2]

A) $1.54 * 10^{-4}$
B) $8.9 * 10^{-2}$
C) $7.1 * 10^{-7}$
D) $2.5 * 10^{-10}$

Answer: A

Solution:

O = {S3, S3, S3, S1, S1, S3, S2, S3}

P(O | Model)

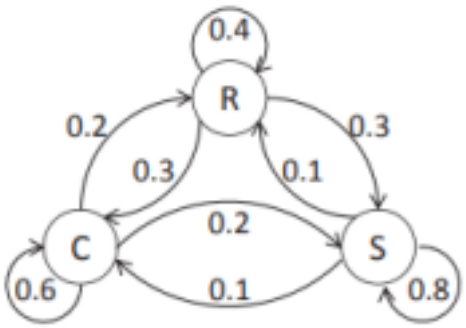
= P(S3, S3, S3, S1, S1, S3, S2, S3 | Model)

= P(S3) P(S3|S3) P(S3|S3) P(S1|S3) P(S1|S1) P(S3|S1) P(S2| S3)

P(S3|S2) = Q3 · a33 · a33 · a31 · a11 · a13 · a32 · a23

= (1)(0.8)(0.8)(0.1)(0.4)(0.3)(0.1)(0.2)

= 1.536 × 10-4



==========

3. In the question 2, the expected number of consecutive days of sunny weather is:

A) 2
B) 3
C) 4
D) 5 **[Marks 1]**

Answer: D

Solution:

Exp(i) = $1/(1-p_{ii})$ So for sunny the exp = 1/(1-0.8) = 5

==========

4. You are building a model distribution for an infinite stream of word tokens. You know that the source of this stream has a vocabulary of size 1200. Out of these 1200 words you know of 200 words to be stop words each of which has a probability of 0.001. With only this knowledge what is the maximum possible entropy of the modelled distribution. (Use log base 10 for entropy calculation) **[Marks 2]**

A) 2.079
B) 4.5084
C) 2.984
D) 3.0775

Answer: D

Solution: There are 200 stopwords with each having an occurrence probability of 0.001. Hence,
P(Stopwords) = 200 ∗ 0.001 = 0.2
P(non − stopwords) = 1 − 0.2 = 0.8

For maximum entropy, the remaining probability should be uniformly distributed. For every non-stopword w, P(w) = 0.8/(1200 − 200) = 0.8/1000 = 0.0008. Finally, the value of the entropy would be,
H = E(log(1/p))
= −200(0.001 ∗ log(0.001)) − 1000(0.0008 log(0.0008))
= −200(0.001 ∗ (-3)) − 1000(0.0008 * (-3.0969))
= 0.6 + 2.4775
= 3.0775

==========

5. Suppose you have the input sentence “Sachin Tendulkar is a great player”. And you know the possible tags each of the words in the sentence can take.
• Sachin: NN, NNS, NNP, NNPS
• Tendulkar: NN, NNS, NNP, NNPS
• is: VB
• a: DT
• great: ADJ
• player: NN, NNS, NNP
How many possible hidden state sequences are possible for the above sentence and States? **[Marks 1]**

A) 4 × 3 × 3
B) $4^{3\wedge 3}$
C) $2^4 \times 2^3 \times 2^3$
D) 3×4^2

Answer: D

Solution: Each possible hidden sequence can take only one POS tag for each of the words. Hence the total possibility will be a product of the number of candidates for each word.

==========

6. What are the space and time complexity order of the Viterbi algorithm? K is the

number of states and N number of time steps.

[Marks 1]

A) KN, K^2N
B) K^2N, KN
C) K^2N, K^2N
D) KN, KN

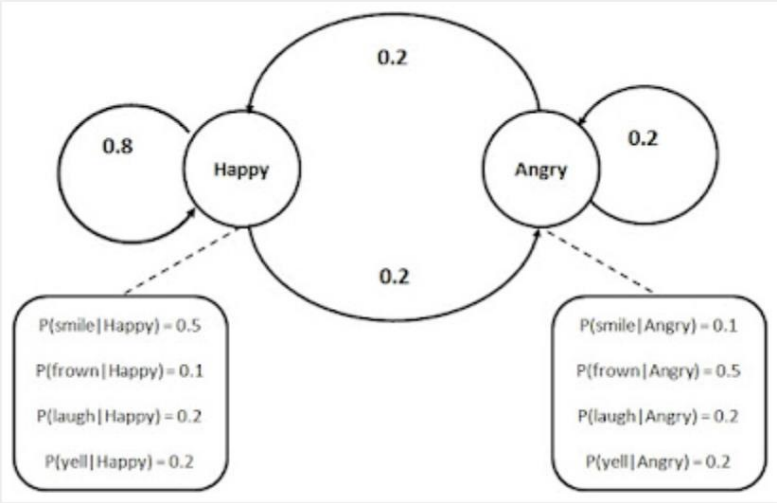
Answer: A

Solution: The sum-product algorithm is polynomial. The time complexity is $O(K^2N)$, the space complexity is O(KN), where K is the number of states and N number of time steps.

==========

7. Mr. X is happy someday and angry on other days. We can only observe when he smiles, frowns, laughs, or yells but not his actual emotional state. Let us start on day 1 in a happy state. There can be only one state transition per day. It can be either a happy state or an angry state. The HMM is shown below-

Assume that q_t is the state on day t and o_t is the observation on day t. Answer the following questions;



What is $P(o_2 = frown)$?

[Marks 2]

A) 0.56
B) 0.18
C) 0.03
D) 0.78

Answer: B

Solution: We need to find the probability of observation ***frown*** on day 2. But we don't know whether he is happy or not on day 2 (we know he was happy on day 1). Hence, the probability of the observation is the sum of products of observation probabilities and all possible hidden state transitions.

$P(o_2 = \text{frown}) = P(o_2 = \text{frown} \mid q_2 = \text{Happy}) + P(o_2 = \text{frown} \mid q_2 = \text{Angry})$

$= P(\text{Happy} \mid \text{Happy}) * P(\text{frown} \mid \text{Happy}) + P(\text{Angry} \mid \text{Happy}) * P(\text{frown} \mid \text{Angry})$

$= (0.8 * 0.1) + (0.2 * 0.5) = 0.08 + 0.1 = 0.18$

X

(https://swayam.gov.in)

(https://swayam.gov.in/nc_details/NPTEL)

pp0783@

NPTEL (https://swayam.gov.in/explorer?ncCode=NPTEL) » Natural Language Processing (course)

If already registered, click to check your payment status

Course outline

About NPTEL ()

How does an NPTEL online course work? ()

Week 0 ()

Week 5 : Assignment 5

The due date for submitting this assignment has passed.

Assignment submitted on 2024-02-28, 06:36 IST

1)

Which of the following are true?

A) Given a CFG and its corresponding CNF, they produce d
B) It requires '2n-1' productions or steps in CNF to generate 'n'.
C) For a given grammar, there can be more than one CNF.
D) None of the above

- [ ] A)
- [x] B)
- [x] C)
- [ ] D)

Yes, the answer is correct.
Score: 1

Accepted Answers:
B)
C)

2) Consider the CFG given below:

$S \rightarrow xSy|V$
$V \rightarrow Vz|\epsilon$

How many non-terminals should be added to convert the Cl

A) 2
B) 4
C) 5
D) 3

- ( ) A)
- ( ) B)

C)

D)

Yes, the answer is correct.
Score: 2
Accepted Answers:
C)

3) In the above Q. 2) How many different numbers of Null proc the CFG to CNF converted form?

A) 0
B) 1
C) 2
D) 3

A)

B)

C)

D)

No, the answer is incorrect.
Score: 0
Accepted Answers:
A)

4) In the above Q. 2) How many different numbers of productic rules/steps in the CFG to CNF converted form??

A) 2
B) 4
C) 7
D) 10

- (○) A)
- (○) B)
- (●) C)
- (○) D)

Yes, the answer is correct.
Score: 1

Accepted Answers:
C)

For **Question 5 to 7** consider the following PCFG fragment:

$S \rightarrow$ NN VP	0.50	$S \rightarrow$ VP NN	0.50
$NP \rightarrow$ NN PB	0.40	$P\,B \rightarrow$ PP NN	0.30
$V\,P \rightarrow$ VB NN	0.30	$V\,P \rightarrow$ VB NP	0.20
$V\,P \rightarrow$ NN VB	0.25	$V\,P \rightarrow$ NN PB	0.15
$P\,P \rightarrow$ with	0.10	$P\,P \rightarrow$ without	0.10
$V\,B \rightarrow$ play	0.30	$V\,B \rightarrow$ *enjoy*/like	0.1
$V\,B \rightarrow$ watch/enjoy	0.25	$NN \rightarrow$ children/student	
$NN \rightarrow$ cricket/football	0.15	$NN \rightarrow$ friends	0.2
$NN \rightarrow$ football/cricket	0.10	$NN \rightarrow$ music/painting	

For a sentence S = w1w2w3w4, assume that the cells in the tab 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

5) Using CKY algorithm, find the probability score for the most tree for the sentence S_1 = "**students play football with frie**

A) 6.06×10^{-4}
B) 1.62×10^{-6}
C) 2.73×10^{-3}
D) 4.33×10^{-6}

- A)
- (•) B)
- C)
- D)

Yes, the answer is correct.
Score: 1

Accepted Answers:
B)

6) Using CKY algorithm, find the number of parse trees for the sen **like painting** and the probability score for at least one of the proba

A) 1, 4.95×10^{-3}
B) 3, 0.36×10^{-3}
C) 2, 0.99×10^{-3}
D) 2, 0.54×10^{-3}

- A)
- B)
- C)
- (•) D)

No, the answer is incorrect.
Score: 0

Accepted Answers:
C)

7) Consider the expression below:

P("students enjoy cricket like painting", $N_{34}|G$) = P_j P("studer painting" |N^j_{34}, G)
What does the L.H.S. represent?

A) Probability of the sentence "students enjoy cricket lik grammar G.
B) Probability of the sentence "students enjoy cricket lik grammar G and that there is some consistent spanni "cricket like", i.e. from word 3 to 4.
C) Probability of the sentence "students enjoy cricket lik grammar G and some rule which derives the segmer
D) None of the above

- A)
- B)
- C)
- D)

Yes, the answer is correct.
Score: 1

Accepted Answers:
B)

8)

Which of the following grammars are valid CNF?

a) $A \rightarrow B$ 2. $A \rightarrow BCD$ 3. $A \rightarrow BC$
b) $B \rightarrow CD$ $B \rightarrow b$ $B \rightarrow \epsilon$
c) $C \rightarrow c$ $C \rightarrow c$ $C \rightarrow c$
d) $A \rightarrow BC$ $A \rightarrow a$

A) a
B) b
C) c
D) d

- A)
- B)
- C)
- D)

Yes, the answer is correct.
Score: 1

Accepted Answers:
D)

Natural Language Processing

Assignment- 6

TYPE OF QUESTION: MCQ

Number of questions: 8 **Total mark: 6X 1 + 2*2 = 10**

Question 1 : Which of the following is/are False about the criteria of head (H) and dependants (D) in a construction (C) in a dependency graph?

1. H is obligatory; D is mandatory
2. H selects D and determines whether D is obligatory
3. The form of D depends on H
4. H specifies D

Answer: 1,4

Solution: H is obligatory; D may be optional. D specifies H

Question 2. Which of the following is/are True about formal conditions on Dependency graph G?

1. G can be cyclic
2. G is projective
3. G is disconnected
4. G obeys the rule: if i→j then not k→j, for any k not equal to i

Answer: 2, 4

Solution: G is connected and acyclic. G is projective and follows single head constraint (option 4)

Question 3: Which of the following is false?

1. Deterministic parsing requires an oracle and an oracle can be approximated by a classifier
2. Each vertex in the graph greedily selects the incoming edge with the highest weight in the Chu-Liu-Edmonds Algorithm
3. During the iteration of Chu-Liu-Edmonds Algorithm it never produces cycle
4. A multi-digraph is a digraph where multiple arcs between vertices are possible

Answer: 3
Solution: During the iteration of the Chu-Liu-Edmonds Algorithm it might produce cycle also. Refer lecture 30

Question 4: Suppose you write down the sequence of actions that generate the parse tree of the sentence "I love AI course" using Arc-Eager Parsing. The number of times you have to use Right Arc, Left Arc, Reduce, Shift is:

Format of the answer is [a, b, c, d] corresponding to the 4 values in the order specified in the query.

1. [3, 0, 2, 1]
2. [1, 2, 1, 3]
3. [1, 2, 0, 2]
4. [1, 2, 0, 3]

Answer: 4
Solution: Please refer lecture

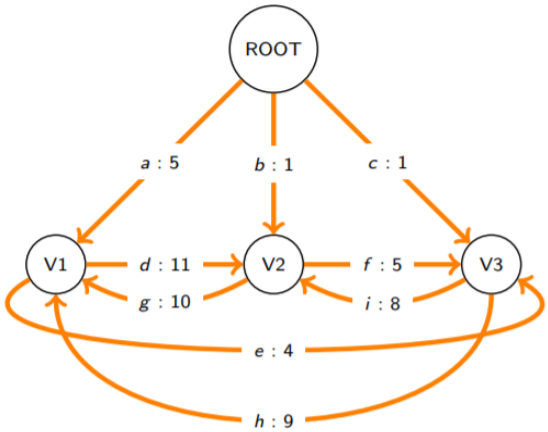
Question 5: Correct sequence of actions that generates the parse tree of the sentence "I love AI course" using Arc-Eager Parsing is:
Note: Right Arc (RA), Left Arc(LA), Reduce(RE), Shift(SH)

1. SH->LA->SH->RE->LA->RA
2. SH->LA->SH->SH->LA->RA
3. SH->LA->SH->RE->LA->RA
4. SH->LA->SH->SH->RA->LA

Answer: 2

Solution: Solve by arc-eager parsing, Refer lecture 29

Question 6: Consider the following graph with a root node and 3 other vertices. I had drawn and put initial edge weights between all the pair of modes at night as shown below. In the morning, my little brother added one (1) to each of the edge weight and said now the edge weight looks good. Suppose you use Chu-Liu-Edmonds algorithm to find the MST for the modified edge weighted graph by my brother. Which pair of nodes will have to be contracted in the modified edge weighted graph to form a single vertex during the algorithm in the 1st iteration?



1. (V1, V2)
2. (V2, V3)
3. (V1, V3)
4. (ROOT, V1)

Answer: 1
Solution: Solve by applying Chu-Liu-Edmonds Algorithm

Question 7: Suppose you are training MST Parser for dependency and the sentence, "I like offline exam" occurs in the training set. The POS tags for these words are Pronoun, Verb, PropNoun and Noun, respectively. Also, for simplicity, assume that there is only one dependency relation, "rel". Thus, for every arc from word wi to wj, your features may be simplified to depend only on words wi and wj and not on the relation label.

Below is the set of features

f1: pos(wi) = Verb and pos(wj) = Noun|Pronoun

f2: wi = Root | wi occurs before wj in the sentence

f3: wi = Root and pos(wj) = Verb

f4: wj occurs before wi in the sentence

The feature weights before the start of the iteration are: [6,19,15,12]

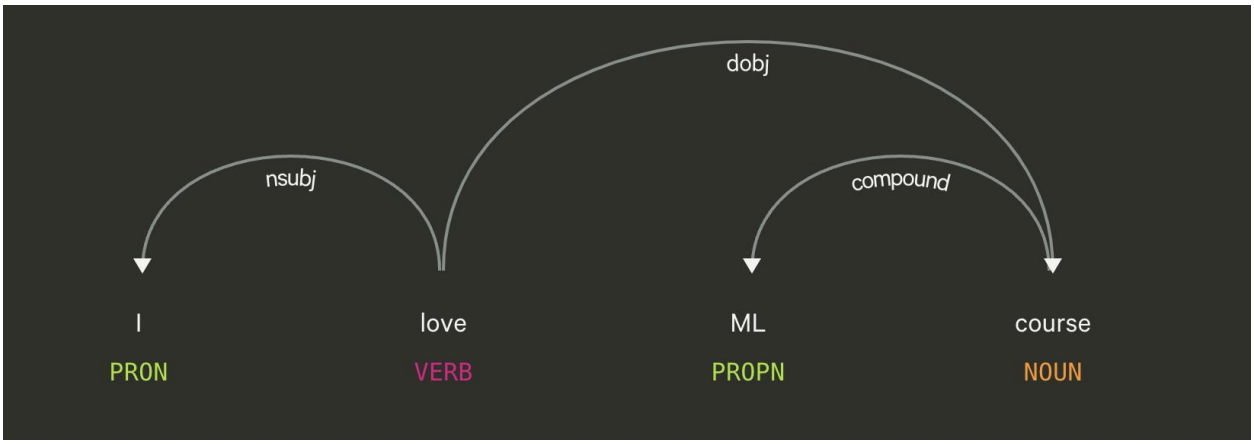
Suppose you are also given that after applying the Chu-Liu Edmonds, you get the following parse tree {Root → like, like → I, I → offline, offline → exam}

What would be the weights after this iteration?

1. [7, 19, 14, 13]
2. [7, 18, 15, 13]
3. [6, 19, 13, 13]
4. [6, 19, 15, 12]

Answer: 2
Solution: Please refer lecture 30

Question 8: Assume that you are learning a classifier for the data-driven deterministic parsing and the sentence 'I love ML course' is a gold-standard parse in your training data. You are also given that ML and 'course' are 'Nouns', 'I' is a 'Pronoun' while the POS tag of 'love' is 'Verb'. Find the dependency graph for this sentence below. Assume that your features correspond to the following conditions:



1. **The stack is empty.**
2. **Top of stack is Noun and Top of buffer is Verb.**
3. **Top of stack is Pronoun and Top of buffer is Verb.**
4. **The word at the top of stack occurs before word at the top of the buffer in the sentence**

The initial weights of your features are
[2,2,2,2 | 3,3,3,2| 2,2,2,2 | 2,2,2,2] where the first four features correspond to LA, and then to RA, SH and RE, respectively

Use this gold standard parse during online learning. What will be the weights after completing two iteration of Arc-Eager parsing over this sentence:

1. [2,2,2,2 | 3,3,3,2| 2,2,2,2 | 2,2,2,2]
2. [2,2,3,2 | 2,3,2,1| 3,2,2,2 | 2,2,2,2]
3. [2,2,3,3 | 3,3,2,1| 3,2,2,2 | 2,2,2,2]
4. [2,2,3,3 | 2,3,2,1| 3,2,2,2 | 2,2,2,2]

Answer: 4
Solution: Refer lecture 29 of week 6

Natural Language Processing

Assignment 7

Type of Question: MCQ

Number of Questions: 7 **Total Marks: (5×1)+(3×2)=10**

==========

Question 1: Suppose you have a raw text corpus and you compute word co occurrence matrix from there. Which of the following algorithm(s) can you utilize to learn word representations? (Choose all that apply) **[1 mark]**

a. CBOW
b. SVM
c. PCA
d. Bagging

Answer: a, c
Solution:

==========

Question 2: What is the method for solving word analogy questions like, given A, B and D, find C such that A:B::C:D, using word vectors? **[1 mark]**

a. $v_c = v_a + (v_b - v_d)$, then use cosine similarity to find the closest word of v_c.
b. $v_c = v_a + (v_d - v_b)$ then do dictionary lookup for v_c
c. $v_c = v_d + (v_a - v_b)$ then use cosine similarity to find the closest word of v_c.
d. $v_c = v_d + (v_a - v_b)$ then do dictionary lookup for v_c.
e. None of the above

Answer: c
Solution: $v_d - v_c = v_b - v_a$
$v_c = v_d + v_a - v_b$ then use cosine similarity to find the closest word of v_c.

==========
=

Question 3: What is the value of $PMI(w_1, w_2)$ for $C(w_1) = 250$, $C(w_2) = 1000$, $C(w_1, w_2) = 160$, $N = 100000$? N: Total number of documents.
$C(w_i)$: Number of documents, w_i has appeared in.
$C(w_i, w_j)$: Number of documents where both the words have appeared in.
Note: Use base 2 in logarithm. **[1 mark]**

a. 4
b. 5
c. 6
d. 5.64

Answer: c

Solution:

PMI = log2 [(160*100000) / (250*1000)] = log2(64) = 6

==========
=

Question 4: Given two binary word vectors w_1 and w_2 as follows:
w_1 = [1010101010]
w_2 = [0011111100]
Compute the Dice and Jaccard similarity between them. **[2 marks]**

a. 6/11, 3/8
b. 10/11, 5/6
c. 4/9, 2/7
d. 5/9, 5/8

Answer: a

$$\text{Dice coefficient} = \frac{2 \times 3}{5+6} = \frac{6}{11}$$

$$\text{Jaccard coefficient} = \frac{3}{8}$$

Solution:

==========

Question 5: Consider two probability distributions for two words be *p* and *q*. Compute their similarity scores with KL-divergence. **[2 marks]**
p = [0.20, 0.75, 0.50]
q = [0.90, 0.10, 0.25]
Note: Use base 2 in logarithm.

a. 4.704, 1,720
b. 1.692, 0.553
c. 2.246, 1.412
d. 3.213, 2.426

Answer: c
Solution:

$$\begin{aligned}
\text{KL-div}(p,q) &= \sum_i p_i \log_2 \frac{p_i}{q_i} \\
&= 0.2 \log \frac{0.2}{0.9} + 0.75 \log \frac{0.75}{0.1} + 0.5 \log \frac{0.5}{0.25} \\
&\approx 2.246 \\
\text{KL-div}(q,p) &= 0.9 \log \frac{0.9}{0.2} + 0.1 \log \frac{0.1}{0.75} + 0.25 \log \frac{0.25}{0.5} \\
&\approx 1.412
\end{aligned}$$

==========

Question 6: Consider the following word co-occurrence matrix given below. Compute the cosine similarity between
(i) w1 and w2, and (ii) w1 and w3. **[2 mark]**

	w4	w5	w6
w1	2	8	5
w2	4	9	7
w3	1	2	3

a. 0.773, 0.412
b. 0.881, 0.764
c. 0.987, 0.914
d. 0.897, 0.315

Answer: c
Solution:

$$\text{cosine-sim}\ (\overrightarrow{p}, \overrightarrow{q}) = \frac{\overrightarrow{p} \cdot \overrightarrow{q}}{\|\overrightarrow{p}\| \cdot \|\overrightarrow{q}\|}$$

Cosine-sim (w1, w2) = (2*4 + 8*9 + 5*7) / (√(2*2 + 8*8 + 5*5) * √(4*4 + 9*9 + 7*7)) = 0.987
Cosine-sim (w1, w3) = (2*1 + 8*2 + 5*3) / (√(2*2 + 8*8 + 5*5) * √(1*1 + 2*2 + 3*3)) = 0.914

==========

Question 7: Which of the following type of relations can be captured by word2vec (CBOW or Skipgram)? **[1 mark]**

1. Analogy (A:B::C:?)
2. Antonymy
3. Polysemy
4. All of the above

Answer: 1
Solution: Word vectors learnt using CBOW or Skipgram models can't disambiguate between Antonyms or Polysemous words.

==========

Natural Language Processing

Assignment- 8

TYPE OF QUESTION: MCQ

Number of Questions: 8 [Question 4, 7 carries 2 marks] **Total Marks: 10**

Question 1: Consider the following sentences:

1. After a hectic day, she found **peace** in finishing the last **piece** of her puzzle.
2. He used a **bat** to hit the ball out of the park, then later that evening, he saw a **bat** flying in the night sky.

The lexical relation between the highlighted words in sentences 1, 2 are

a. Homograph, Synonym
b. Homonymy, Homophones
c. Synonym, Hyponym
d. Homophones, Homonymy

Answer: d
Solution:

Question 2: Consider the following sentences. Which of the following is/are False?

a. Fruit is a hypernym of mango.
b. Animal is hyponym of cat.
c. Homographs are the words with the same pronunciation but different spelling
d. Bank (financial organization) vs Bank (riverside) is an example of homonym

Answer: b, c

Question 3:

Which of the following is /are True?

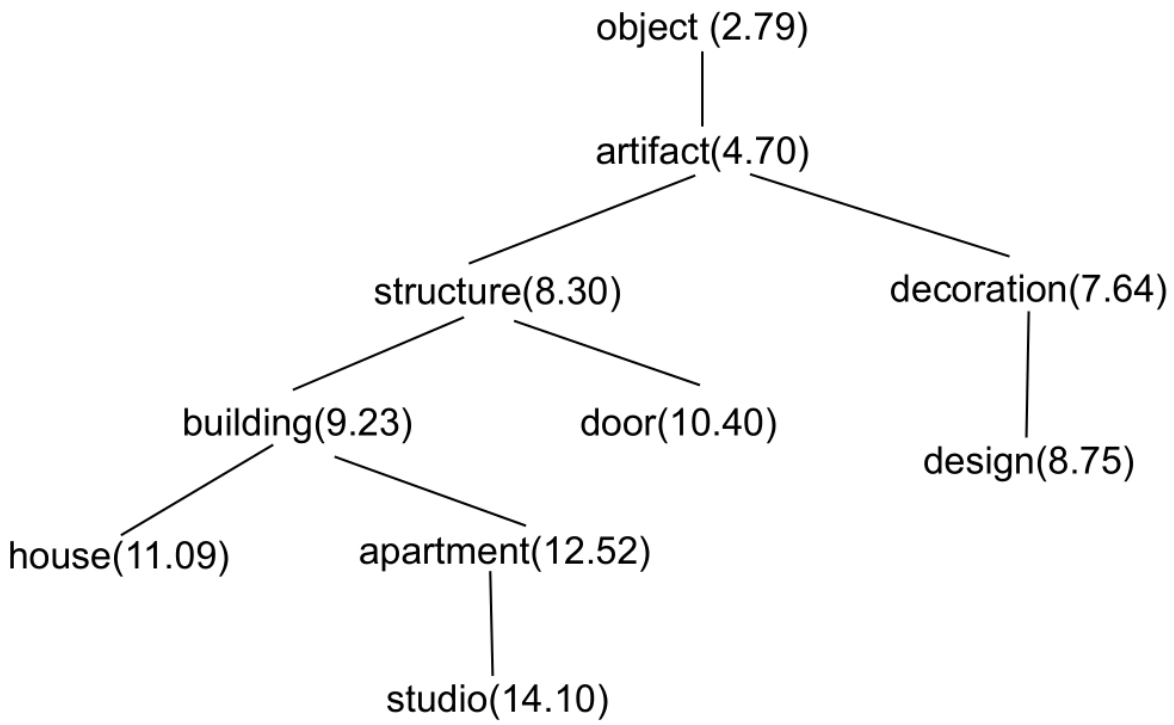
1. Synonymy is a binary relation
2. Thesaurus-based algorithms are based on whether words are nearby in Wordnet
3. Synset is a set of antonyms representing sense
4. Troponym is from a verb to a specific manner elaboration of that verb

Answer: 1,2,4

Solution: Refer lecture 3

For Question 4 to 6, consider a hypothetical wordnet noun taxonomy with their information content as shown in Figure 1. Question 4 carries 2 marks

Note: Use base 10 in logarithmic calculations



Question 4: What is the Lin similarity between **house** and **decoration**?

a. 0.564
b. 0.433
c. 0.501
d. 0.473

Answer: c
Solution: (2×4.7)/(11.09+7.64) ≈ 0.501

Question 5: What is the Resnic similarity between **building** and **door**?

a. 11.09
b. 8.30
c. 9.23
d. 4.70

Answer: b
Solution:

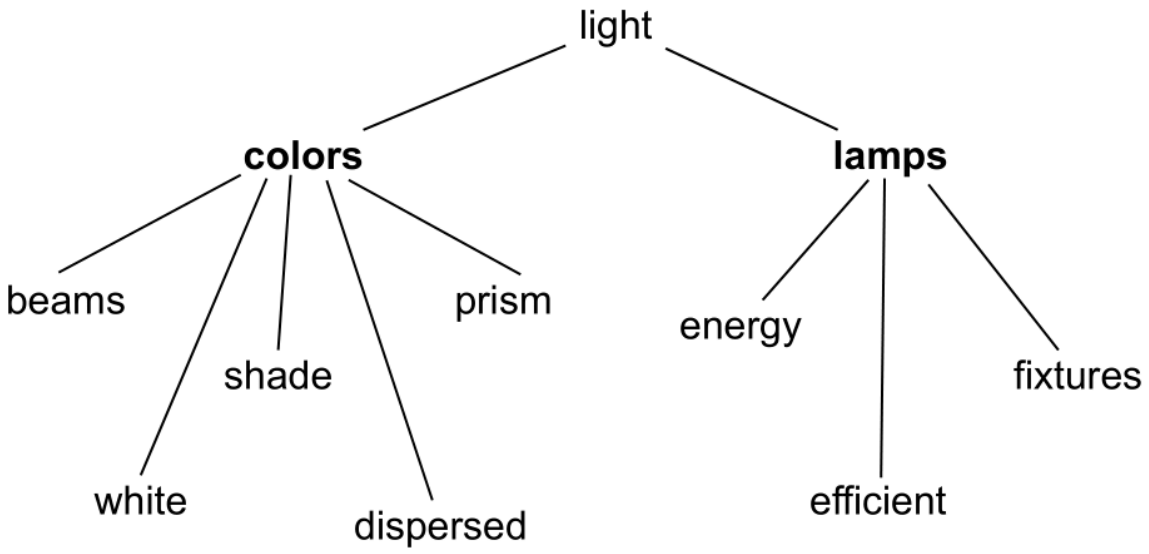
Question 6: What is the Leacock–Chodorow similarity between **house** and **design**?

a. 0.694
b. 0.398
c. 0.097
d. None of the above

Answer: b

Solution:
LC similarity = −log pathlen(c1,c2)/2d = −log 4/(2×5) ≈ 0.398

For Question 7 consider the network of words for disambiguation of the word "light" as shown in Figure 3. The hubs are "colors" and "lamps". Note: Take the distance between two words as the path length between them.



Question 7: Compute the scores for (i) the hub "colors" and the component "white" and (ii) the hub "colors" and the component "fixtures".

a. 0.2, 0.25

b. 1.0, 0.0

c. 0.5, 0.25

d. None of the above

Answer: d

Solution:

(i) 1/(1+1) = 0.5

(ii) 0 as "colors" is not an ancestor of "fixtures"

Question 8: Which of the following is/are false?

1. Hyponym is also called superordinate
2. Hypernym is also called subordinate
3. Snore→ Sleep is an example of entails
4. Meal→ lunch is a example of hyponym

Answer: 1,2

Solution: Refer lecture 37

Natural Language Processing

Assignment- 9

TYPE OF QUESTION: MCQ

Number of questions: 8

[Question 7 and 8 carries 2 marks each] **Total mark: 6 X 1 + 2 X 2 = 10**

Question1. Which of the following is/are true?

1. Topic modelling discovers the hidden themes that pervade the collection
2. Topic modelling is a generative model
3. Dirichlet hyperparameter Beta used to represent document-topic Density?
4. None of the above

Answer: 1,2

Question2. Which of the following is/are true?

1. The Dirichlet is an exponential family distribution on the simplex positive and negative vectors sum to one
2. Correlated Topic Model (CTM) predicts better via correlated topics
3. LDA provides better fit than CTM
4. CTM draws topic distributions from a logistic normal

Answer: 2, 4

Solution: Refer Lecture 44

Question 3: You have a topic model with the parameters $\alpha = 0.89$ and $\beta = 0.04$. Now, if you want to have sparser distribution over words and denser distribution over topics, what should be the values for α and β?

1. Both α and β values should be decreased
2. Both α and β values should be increased
3. α should be decreased, but β should be increased

4. α should be increased, but β should be decreased

Answer: 4

Solution:
α : topic distribution
β : word distribution

Question4: Which of the following is/are false about LDA assumption?

1. LDA assumes that the order of documents matter
2. LDA is not appropriate for corpora that spans hundreds of years
3. LDA assumes that documents are a mixture of topics and topics are a mixture of words
4. LDA can decide on the number of topics by itself.

Answer: 1,4

Solution: Refer Lecture 44

Question 5: Which of the following is/are True about Relational Topic Model (RTM) ?

1. RTM formulation ensures that the same latent topic assignments used to generate the content of the documents
2. In RTM, link function models each per-pair binary variable as linear regression
3. In RTM, covariates are constructed by the Hadamard product
4. Link probability function is dependant on the topic assignments that generated their words

Answer: 1,3,4

Solution: Refer Lecture 45

Question 6:
Classically, topic models are introduced in the text analysis community for_____ topic discovery in a corpus of documents.

1. Unsupervised.

2. Supervised.
3. Semi-automated.
4. None of the above.

Answer - 1. Unsupervised

Question 7: Which of the following is/are False about Gibbs Sampling?

1. Gibbs sampling is a form of Markov chain Monte Carlo (MCMC)
2. Sampling is done sequentially and proceeds until the sampled values approximate the target distribution
3. It can not estimate the posterior distribution directly
4. Gibbs sampling falls under the category of variational methods

Answer: 3,4
Solution: Refer Gibbs Sampling slide

For question 8 use the following information.
Suppose you are using Gibbs sampling to estimate the distributions, θ and β for topic models. The underlying corpus has 3 documents and 5 words, **{machine, learning, language, nature, vision}** and the number of topics is 2. At certain point, the structure of the documents looks like the following

Doc1: nature(1) language(1) vision(1) language(1) nature(1) nature(1) language(1) vision(1)
Doc2: nature(1) language(1) language(2) machine(2) vision(1) learning(2) language(1) nature(1)
Doc3: machine(2) language(2) learning(2) language(2) machine(2) machine(2) learning(2) language(2)

(number) –number inside the brackets denote the topic no. 1 and 2 denote whether the word is currently assigned to topics t1 and t2 respectively. $\eta = 0.3$ and $\alpha = 0.3$

For question 8 calculate the value upto 4 decimal points and choose your answer

Question 8 : Using the above structure the estimated value of $\beta(2)_{\text{nature}}$ at this point is

1. 0.0240
2. 0.02459
3. 0.0260
4. 0.0234

Answer: 1

Solution:

	t1	t2
machine	0	4
nature	5	0
language	5	4
vision	3	0
learning	0	3

$\beta(2)_{nature}$ = (0+0.3)/(11+5*0.3) = 0.3/12.5 = 0.024

Natural Language Processing

Assignment 10

Type of Question: MCQ

Number of Questions: 10 **Total Marks:(10×1)= 10**

==========

Question 1. Common Steps of entity linking are -

A) Reference Disambiguation -> Candidate Selection
B) Reference Disambiguation -> Candidate Selection -> Mention Identify
C) Mention Identify -> Candidate Selection -> Reference Disambiguation
D) All of the above

[Marks 1]

Answer: C

Solution: Theory. Slide 7; Lecture 1, Week 10 (Entity Linking 1).

==========

Question 2. The text span s="world" occurs in 764 different Wikipedia articles.

c1 189
c2 273
c3 87
c4 53
No Link 162

Calculate the keyphraseness of "world". **[Marks 1]**

Answer:

A) 0.232
B) 0.788
C) 0.688
D) 0.976

Solution: B)

$CF(s_i) / CF(s) = 189 + 273 + 87 + 53 / 764 = 602 / 764 = 0.788$

==========

Question 3. What is the commonness of (s, c2) in the above question?
[Marks 1]

Answer:

A) 0.765
B) 0.389
C) 0.453
D) 0.910

Solution: C) 273/(189 + 273 + 87 + 53) = 273/602 = 0.453

==========

Question 4. Higher value of keyphraseness represents higher probability of:

A) an article to get selected as linkable candidate
B) a phrase to get detected as a mention
C) an article to get disambiguated from other candidates
D) None of the above

Solution: B) a phrase to get detected as a mention. Lecture Slide.

==========

Question 5. Which of the following problem exists in bootstrapping technique for Information extraction are: **[Marks 1]**

A) Sensitiveness towards the seed set
B) High precision
C) Less manual intervention
D) All of the above

Answer: A

Solution: Theory

==========

Question 6. Which of the following is an advantage of unsupervised relation extraction: **[Marks 1]**

A) Can work efficiently with small amount of hand-labeled data
B) Not easily generalizable to different relations
C) Need training data.
D) Always perform better than supervised techniques.
E) None of the above

Answer: E

Solution: Theory

==========

Question 7. Which of the following is/are a Hearst's Lexico Syntactic Patterns for automatic acquisition of hyponyms -

A) X or other Y
B) X and other Y
C) Y including X
D) X but not Y

[Marks 1]

Answer: A, B, C

Solution: Theory. Lecture Video 48 : Information Extraction - Introduction, Week 10, Lecture 3 - Slide 15/18.

==========

Question 8. Advantage of Distant supervision over bootstrapping method

A) Need more data
B) Less human effort
C) Can handle noisy data better
D) No Advantage

[Marks 1]

Answer: C

Solution: Theory

==========

Question 9. Consider a dataset with a very low number of relations - all of which are very important. For a relation extraction task on that dataset, which of the following is the most useful metric

A) Precision
B) Recall
C) Accuracy
D) F1-Score

[Marks 1]

Answer: B
Solution: Theory

==========

Question 10. Bootstrapping can be considered as

A) Supervised Approach
B) Unsupervised Approach
C) Semi-supervised Approach
D) All of the above
E) None of the above

[Marks 1]

Answer: C

Solution: Theory. Slide => Relation Extraction, Slide - 2/17, Week4, Lecture 10.

==========

Natural Language Processing
Assignment 11
Type of Question: MCQ

Number of Questions: 8 [Question 4,5 carries two marks] Total Marks: 6*1+2*2=10

Question 1: Your teacher recommended you to read the book 'Deep Learning with Python'. After reading the book, you want to summarize it. What kind of summarization method would you use for this purpose?

1. Abstractive single document summarization
2. Abstractive multi document summarization
3. Extractive single document summarization
4. Extractive multi document summarization

a. 1, 2
b. 3, 4
c. 1, 3
d. 2, 4

Answer: c
Solution:

Question2: Which of the following is/are True?

1. PageRank based algorithm is used to compute the sentence centrality vector
2. Query-focused summarization can be thought of as a simple question answering system
3. The underlying hypothesis of LexRank algorithm is sentences that convey the theme

of the document are more similar to each other

4. All of the above

Answer: 1,3

Solution: Refer Lecture 51

Question 3 : Which of the following is/are True?

1. ROUGE metric is as good as human evaluation
2. LexRank can be applied for multi-document summarization
3. In optimization based approach for summarization, the inference problem is to select a subset S of textual units found such that summary score is maximized
4. Maximum Marginal Relevance strives to reduce redundancy while maintaining query relevance.

Answer: 2,3,4

Question 4: It is estimated that 20% of GPT-4 generated texts are fake. Google built some AI system to filter these fake contents. An AI system claims that it can detect 99% of fake contents, and the probability for a false positive (a real content detected as fake) is 3%. Now if a content is detected as fake, then what is the probability that it is in fact a real content?

a. 0.084
b. 0.118
c. 0.108
d. None of the above

Answer: c

Solution:

Let, A = Event that a content is detected as fake
B = Event that a generated text is fake
$P(B) = 0.2$

$P(B') = 0.8$
$P(A|B) = 0.98$
$P(A|B') = 0.03$

$$P(B'|A) = P(A|B')P(B')/P(A)$$

$$= P(A|B')P(B')/(P(A|B)P(B) + P(A|B')P(B'))$$

$$= (0.03 \times 0.8)/(0.99 \times 0.2 + 0.03 \times 0.8)$$

$$\approx 0.108$$

For question 5-8 follow the below Table . One classifier predicts the following. The correct prediction is shown by the tick mark under Match column.

No	Actual	Predicted	Match
1	Airplane	Airplane	✓
2	Car	Boat	
3	Car	Car	✓
4	Car	Car	✓
5	Car	Boat	
6	Airplane	Boat	
7	Boat	Boat	✓
8	Car	Airplane	
9	Airplane	Airplane	✓
10	Car	Car	✓

Question 5: What is the macro-averaged f1 score?

a. 0.54
b. 0.56
c. 0.58
d. 0.64

Answer: c

Question 6: What is the micro averaged precision?

a. 0.58
b. 0.64
c. 0.50
d. 0.60

Answer: d

Question 7: What is the f1 score of boat class?

a. 0.40
b. 0.30
c. 0.58
d. 0.67

Answer: a

Question 8: What is the accuracy of the classifier?

a. 0.40
b. 0.50
c. 0.60
d. 0.90

Answer: c

Solution:

			Predicted	
	Label	**Airplane**	**Boat**	**Car**
	Airplane	**2**	1	0
Actual	**Boat**	0	1	0
	Car	1	2	**3**

Label	True Positive (TP)	False Positive (FP)	False Negative (FN)
Airplane	2	1	1
Boat	1	3	0
Car	3	0	3

Airplane :

Precision: 0.67, Recall: 0.67

Boat:

Precision: 0.25, Recall: 1.00

Car:

Precision: 1.00, Recall: 0.50

macro-f1: 0.58

micro-f1: 0.60

Accuracy: 6/10 = 0.60

Natural Language Processing

Assignment- 12

TYPE OF QUESTION: MCQ

Number of questions: **8** [Question 4,6 carries two marks] **Total mark: 6 X 1+2*2 = 10**

Question 1: The word ‘nervous’ and ‘jealous’ falls under which affective state typology?

1. Emotion
2. Mood
3. Personality traits
4. Interpersonal stances

Answer: 3

Solution: Refer lecture 56

Question2: Which of the following is/are tokenization issues for Sentiment Analysis?

1. Handling emoticons
2. Handling negation
3. Word lengthening
4. None

Answer: 1,2,3

Question 3: Which of the following is/are false?

1. The picture quality of this camera is great - ‘picture quality’ is implicit aspect
2. This laptop is expensive - ‘expensive’ is an implicit aspect expression
3. This phone will not fit in a pocket - ‘fit in a pocket’ is an explicit aspect expression indicates the aspect ‘size’
4. All of the above

Answer: 1,3

Solution: Refer lecture 60

Question 4: Consider the sentence: "The display was cool; but, the network quality and camera were very dull". Which of the following are true?

1. Aspect: "display", Sentiment: Positive, Opinion Phrase: "cool".
2. Aspect: "camera", Sentiment: Negative, Opinion Phrase: "very dull".
3. Aspect: "were", Sentiment: Negative, Opinion Phrase: "very dull".
4. Only (1) and (3)

Answer: 1,2

Solution:
'were' is not a aspect term

Question 5: Let P(w|c) represent the probability of a word given its rating. Further, let P(w) represent the probability of the word in the overall corpus. While analyzing the polarity of a word in a given corpus, what is the intuition behind dividing P(w|c) by P(w)?

1. To make it comparable across different ratings.
2. To make it comparable across different words.
3. Both (1) and (2).
4. None of the above.

Answer: 2
Solution: Refer to Lecture 60 "Computing with Affective Lexicons".

Question 6: The word 'loving' and 'buoyant' falls under which affective state typology respectively?

1. Attitudes, Mood
2. Emotion, Attitudes
3. Emotion, Mood
4. Mood, Personality traits

Answer: 1

Solution: Refer to Lecture 56

Question 7: Which of the following is/are True regarding indicators of Irrealis mood?

1. Conditional markers(if)
2. Negative polarity items like ‘any’ and ‘anything’
3. Words enclosed in quotes and necessarily reflective of the authors opinion
4. None

Answer: 1, 2

Solution: Refer Lecture 59

Question 8: Which of the following is/are True about the Turney algorithm?

1. It extracts a phrasal lexicon from reviews
2. Learns polarity of each phrase
3. Rate a review by the average polarity of its phrases
4. All of the above

Answer: 4
Solution: Refer Lecture 58
