**CS 5316–Natural Language Processing**

Midterm ExamSolution

March 10, 2019

Duration: 2 hours (10.30 to 12.30)

1. Briefly describe and/or distinguish between pragmatic and syntactic knowledge in linguistics.

Pragmatic knowledge refers to information about the context surrounding or behind a passage. This information may not be readily available from the passage itself, and therefore, can be called as background knowledge. For example, the acronym LUMS in a passage can only be related to the university LUMS if this background information is known.

Syntactic knowledge refers to information about how words and sentences are formed in a given language. Grammar is a key component of syntactic knowledge.

1. Given the dictionary below, segment the sequence a3- bba- 5kaa- bb - 3 using max-match algorithm.

a, bb, a3,bba,5kaa

a3,bba,5kaa,bb,3

1. Write a regex for dates using one or two digits for day and month, 2 or four digits for year (last component), separator can be dash (-) or slash (/) (e.g., 9/5/18, 09/05/2018, 9-5-18).

[0-9]{1,2}[-/][0-9]{1,2}[-/]([0-9]{2}) | ([0-9]{4})

1. Consider the following sentences written in informal language. (a) List the normalization steps you will apply before text classification, (b) List the tokens and types after your normalization steps.

The phone looks GREAT!!!!

I like the fon,,, it has gr8 looks, goooood color

Me liking n loving the phone ALOT

Preprocessing steps:

1. Remove punctuations
2. Tokenize with space delimitation
3. Lowercase all words (assuming that we do not want to distinguish between all caps and lowercase words in the analysis)
4. Use an informal word dictionary to map informal words to their corresponding standard forms, e.g., fone🡪 phone
5. Remove stop words
6. Lemmatize
7. Build vocabulary

Tokens – 14

phone look great

like phone has great look good color

like love phone lot

Types =9

phone look great like has good color love lot

1. Explain briefly (with an example) why MLE is not recommended for language modeling.

MLE tends to over-fit to data and not generalize well. For language modeling, the possibilities are numerous while all such possibilities may not be observed in finite-sized data. For example,if the bigram virus infection is not observed in a corpus, MLE will give it a probability of zero even though this is a legitimate bigram.

1. In Good-Turing estimation, if and , calculate the discounted count of an item that is observed 5 times.

c\* = 6\*8/10 = 4.8

1. Sentiment analysis extracts and aggregates affective states expressed in natural language. In addition to attitude analysis, give examples of two other types of sentiment analyses.

* Emotion analysis
* Mood analysis
* Inter-personal behavior analysis
* Personality trait analysis

1. Consider the following confusion matrix obtained during the evaluation of a POS tagging algorithm. The rows indicate true tags and the columns indicate predicted tags. Compute the macro averaged precision, recall, and F1-value of the algorithm.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Noun | Verb | Adjective |
| Noun | 550 | 40 | 60 |
| Verb | 100 | 700 | 200 |
| Adjective | 20 | 50 | 300 |

For Noun

PreN = 550/(550+100+20) = 0.8209

RecN = 550/(550+40+60) = 0.8462

F1N = 2 \* 0.8209 \* 0.8462/ (0.8209 + 0.8462) = 0.8334

For Verb

PreV = 700/(700+40+5) = 0.9396

RecV = 700/(700+100+200) = 0.7

F1V = 2 \* 0.9396 \* 0.7/ 0.9396 + 0.7 = 0.8023

For Adjective

PreA = 300/(300+60+200) = 0.5357

RecA = 300/(300+20+50) = 0.8108

F1A = 2 \* 0.5357 \* 0.8108/ 0.5357 + 0.8108 = 0.6451

macro averaged precision = (PrecN + PrecV + PrecA) / 3 = 0.7654

macro averaged recall = (RecN + RecV + RecA) / 3 = 0.7857

macro averaged F1 = (F1N + F1V + F1A) / 3 = 0.7603

Macro average (weighted by class size)

Noun = 650, Verb = 1000, Adjective = 370, total = 2020

1. Using the IOB encoding, label the following sentence with NER tags: Prime Minister Imran Khan took oath in Islamabad.

Prime/B/Per Minister/I/Per Imran/I/Per Khan/I/Per took/O oath/O in/O Islamaba/B/Loc

1. Write the computational time complexity of solving the decoding problem using (a) Vitterbialgorithm, and (b) brute-force approach. Define the parameters in the expression.

Vitterbi algorithm:

Brute force approach:

no. of states/tags

sequence length

1. 15 points) Determine the minimum edit distance and the corresponding alignment between BARLEY and BEARY. Do this using (a) unit costs for insertion, deletion, and substitution, and (b) insertion has cost 1, deletion has cost 2, and substitution has cost 3.

let source = beary

target = barley

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Y | 5 | 4 | 3 | 2 | 2 | 3 | 3 |  |  | Y | 10 | 8 | 6 | 4 | 5 | 6 | 4 |
| R | 4 | 3 | 2 | 1 | 2 | 3 | 4 |  |  | R | 8 | 6 | 4 | 2 | 3 | 4 | 5 |
| A | 3 | 2 | 1 | 2 | 3 | 4 | 4 |  |  | A | 6 | 4 | 2 | 3 | 4 | 5 | 6 |
| E | 2 | 1 | 1 | 2 | 3 | 3 | 4 |  |  | E | 4 | 2 | 3 | 4 | 5 | 3 | 4 |
| B | 1 | 0 | 1 | 2 | 3 | 4 | 5 |  |  | B | 2 | 0 | 1 | 2 | 3 | 4 | 5 |
| # | 0 | 1 | 2 | 3 | 4 | 5 | 6 |  |  | # | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|  | # | B | A | R | L | E | Y |  |  |  | # | B | A | R | L | E | Y |

For (a) alignment is

B E A R \* \* Y

B \* A R L E Y

del ins ins

MED =3

For (b) alignment is

B E A R \* \* Y

B \* A R L E Y

del ins ins

MED =4

Another solution could be:

let source = barley

target = beary

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Y | 6 | 5 | 4 | 4 | 4 | 3 |  |  | Y | 12 | 10 | 8 | 9 | 7 | 5 |
| E | 5 | 4 | 3 | 4 | 3 | 3 |  |  | E | 10 | 8 | 6 | 7 | 5 | 6 |
| L | 4 | 3 | 3 | 3 | 2 | 2 |  |  | L | 8 | 6 | 7 | 5 | 3 | 4 |
| R | 3 | 2 | 2 | 2 | 1 | 2 |  |  | R | 6 | 4 | 5 | 3 | 1 | 2 |
| A | 2 | 1 | 1 | 1 | 2 | 3 |  |  | A | 4 | 2 | 3 | 1 | 2 | 3 |
| B | 1 | 0 | 1 | 2 | 3 | 4 |  |  | B | 2 | 0 | 1 | 2 | 3 | 4 |
| # | 0 | 1 | 2 | 3 | 4 | 5 |  |  | # | 0 | 1 | 2 | 3 | 4 | 5 |
|  | # | B | E | A | R | Y |  |  |  | # | B | E | A | R | Y |

For (a) alignment is

B \* A R L E Y

B E A R \* \* Y

in del del

MED =3

For (b) alignment is

B \* A R L E Y

B E A R \* \* Y

in del del

MED =5

1. ((15points) Consider the following five documents belonging to three classes (d1, d2 is class 1, d3, d4 is class 2, d5 is class 3).For each document, the words and their counts are given.

d1:plot/1, absurd/1, okay/1, bad/2

d2:screenplay/1, terrible/1, novel/1, absurd/1

d3:plot/1, engaging/1, good/2

d4:screenplay/1, excellent/2, okay/1,

d5:plot/1, okay/2, novel/1

* 1. (8 points) Develop a multinomial naïve Bayes classifier for this problem. Show the model and the estimated parameters using add-one smoothing.
  2. (7 points) Determine the best class for a new document: good screenplay, good plot.

1. P(c1) = 2/5

P(c2) = 2/5

P(c3) = 1/5

multinomial naïve bayes model is given as:

--------------- (A)

vocabulary ={plot, absurd, okay, bad, screenplay, terrible, novel, engaging, good, excellent}

|V| = 10

P(plot| c1) =

P(absurd| c1) =

P(okay| c1) =

P(bad| c1) =

P(screenplay| c1) =

P(terrible| c1) =

P(novel| c1) =

P(engaging| c1) =

P(good| c1) =

P(excellent| c1) =

P(plot| c2) =

P(absurd| c2) =

P(okay| c2) =

P(bad| c2) =

P(screenplay| c2) =

P(terrible| c2) =

P(novel| c2) =

P(engaging| c2) =

P(good| c2) =

P(excellent| c2) =

P(good| c3) = P(absurd|c3) = P(bad|c3) = P(screenplay|c3) = P(terrible|c3) = P(engaging|c3) = P(excellent|c3) =

P(okay| c3) =

P(plot| c3) = P(novel|c3) =

1. let d6 be good screenplay, good plot

then using multinomial naïve bayes model in previous part:

P(c1|d6) = P(c1)P(good|c1)P(screenplay|c1)P(good|c1)P(plot|c1)

= P(c1)[P(good|c1)]2 P(screenplay|c1) P(plot|c1)

= 2 \* \* = 0.00001228

similarly

P(c2|d6) = 2 \* \* = 0.000137

P(c3|d6) = 2 \* \* = 0.0000104

since

Hence d6 belongs to class 2.

1. (15 points) Given the following passage, estimate using(a) add-one smoothing, and (b) Kneser-Ney smoothing. Take discount equal to 0.75.

New Delhi's narrative on the India-Pakistan standoff appeared to crumble further on Wednesday as foreign journalists unearthed new details about the events of the preceding weeks.

|V| = 24

Add-one smoothing estimate

(1+1)/(2+24) = 0.08

Kneser-Ney smoothing estimate

(1-0.75)/(2) + 0.75 \* 1/25 = 0.155

Lambda = 0.75/2\*2, P\_cont = 1/25 (25 is the no. of unique non-zero bigrams)

1. (15 points) Consideran HMM over 2 states and 3 events, and the probability matrices as given below. Compute the likelihood of the event sequence .

0.8 0.2

0.7 0.3

0.4 0.2

0.3 0.2

0.3 0.6

Assume all probabilities from initial state and to final state are uniform.

Using Forward algorithm:

|  |  |  |  |
| --- | --- | --- | --- |
|  | r | p | q |
| start state |  |  |  |
| m | 0.5 \* 0.3 = 0.15 | 0.132 | 0.03672 |
| n | 0.5 \* 0.6 = 0.3 | 0.024 | 0.00672 |
| end state |  |  |  |

Required likelihood = (f[m,2] \* 0.5) + (f[n,2] \* 0.5)

= (0.03672 \* 0.5) + (0.00672 \* 0.5)

= 0.01836 + 0.00336

= 0.02172

* students are expected to show full work for all questions