Exam
Natural Language Processing (201600074)
29 October 2020 (13:45 - 16:45)

READ ME

This exam consists of 18 multiple choice questions (each 5 points) and 1 open question (10 points). The grade is computed as total points/10.

You may use your own notes as well as any materials from the book "Speech and Language Processing" by Jurafsky & Martin, the slides, homework assignments, in printed form and/or in PDF format via the university laptop. Don't use any other electronic devices; there is a calculator on the laptop.

We recommend that you read through the entire exam first, and then start with those questions you think are easy or quick to answer.

Regarding the multiple choice questions we recommend that you first mark your answer on draft paper, check all the answers and then fill in your final choice on the multiple choice answer form. For all MC questions only 1 of the 4 options is correct. We will post the correct answers (but not the questions!) on Canvas after the exam.

For asking questions about the exam, please use the chat widget on the web page with the exam materials. This will put you directly in touch with the teachers. The other students will not see your chat.

When you are finished please **raise your hand** and we will come and collect your answers. You should hand in the following papers:

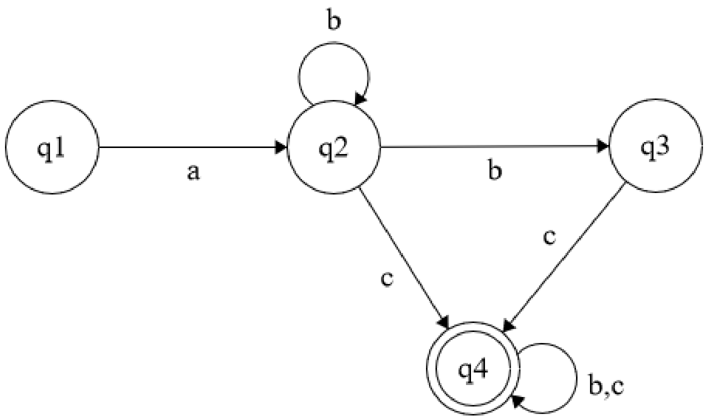
1. The filled-in MC answer form
2. A separate sheet with your answer to the open question

Don't forget to put your name and student number on everything you hand in.

Good luck!

Multiple choice questions

1. The figure below shows a finite state automaton M with as alphabet $\{a, b, c\}$. The initial state of M is state 1; the final state is state 4.

   Which of the following languages is the one defined by M?

   (a) ab+c(bc)*
   (b) ab*bc+b*
   (c) a(bc+|c+)
   (d) ab*c(b|c)*

2. Below you see a list of words and their stemmed version, as produced by the English Porter stemmer.

Word	Stemmed version
grate	*grate*
grated	*grate*
grateful	*grate*
grates	*grate*

   In the context of this list, which of the following statements is true?

   (a) That *grateful* was stemmed to *grate* is an error of omission
   (b) That *grateful* was stemmed to *grate* is an error of commission
   (c) That *grateful* was stemmed to *grate* is not an error
   (d) It's impossible to tell from this list if the stemming of *grateful* is an error or not

3. U.S. President Donald Trump is known for sometimes making unusual word choices. The following is a quote from his announcement speech on June 16, 2015:

   > *Think of it. Iran is taking over Iraq, and they're taking it over big league.*

   Many people thought he said *bigly* instead of *big league*, and argued over whether this was even a real word.

   If Trump had really used the word *bigly* here, what kind of word formation process would he have applied?

   (a) Inflection
   (b) Derivation
   (c) Compounding
   (d) Cliticisation

4. Here is our corpus:
   $\langle s \rangle$ *let it be* $\langle /s \rangle$
   $\langle s \rangle$ *to be or not* $\langle /s \rangle$
   $\langle s \rangle$ *let it not be let yet* $\langle /s \rangle$

   If we train a bigram model where we model the start and end of the sentence, using "add-1" smoothing, what is the probability of this sentence: $\langle s \rangle$ *not to be* $\langle /s \rangle$?

   (a) $\frac{1}{3960} \approx 0.000253$
   (b) $\frac{1}{50820} \approx 1.97 \times 10^{-5}$
   (c) $\frac{1}{12005784} \approx 8.33 \times 10^{-8}$
   (d) None of above

5. Which of the following statements about Naïve Bayes is **false**?

   (a) A Naïve Bayes classifier can use multiple types of features, including some not directly based on words
   (b) Stop word removal often does not significantly improve the performance of a Naïve Bayes classifier
   (c) The assumption of Naïve Bayes is that the probabilities of words can be independently multiplied, regardless of their class
   (d) Naïve Bayes classifiers can be "made aware" of negations with some corpus pre-processing

6. Perplexity is a way to evaluate the quality of a probabilistic model of text sequences.

   Which of the following statements about perplexity is true?

   (a) Perplexity can be evaluated on different test sets, for different applications and of different text lengths. The results can be compared directly and quantify how good the model is for each particular application.

   (b) Perplexity can be evaluated on a single test set to compare different models, of different vocabulary sizes, for that particular application

   (c) Perplexity must be evaluated on the test set, because otherwise we can have zero probabilities in the denominator and the result is undefined

   (d) Models that consider a larger vocabulary tend to have smaller perplexities, because they overfit the data

7. When applying Recurrent Neural Networks (RNN) to machine translation, the initial unrolled network is used to encode the sentence in the source language, and the subsequent part is used to generate the corresponding sentence in the target language. At test time, the target sentence is generated by "Beam Search". This means that:

   (a) A greedy search is performed on the most probable target (output) sentences, where the most probable generated tokens with index i are used as input for the generation of the most probable tokens at index $i+1$, including the "end-of-sentence" token.

   (b) An exhaustive search of target (output) sentences is performed, where all generated tokens with index i are used as input for the computation of the probability of the tokens at index $i+1$ and the most probable sentence is selected at the end.

   (c) A greedy search is performed on the most probable target (output) sentence length by computing the probability of the "end-of-sentence" token. An exhaustive search of all possible sentences of that length is then performed.

   (d) A greedy search is performed on the most probable target (output) sentence length by computing the probability of the "end-of-sentence" token. A greedy search of the most probable sentences of that length is then performed where the most probable generated tokens with index i are used as input for the generation of the most probable tokens at index $i+1$.

8. Consider the formula for estimating the probability of word i given the positive class:

   $P(f_i|+) = \frac{C(w_i,+)}{\sum\limits_{w \in V} C(w,+)}$

   Which of the following statements is true?

   (a) $C(w_i,+)$ is the number of times word w_i appears in the *positive* class, out of the times w_i appears in the whole corpus

   (b) $C(w_i,+)$ is the number of times word w_i appears in the *positive* corpus, multiplied by the prior probability of the positive class

   (c) V is the set of unique words that appear in the *whole* corpus

   (d) $P(f_i|+)$ is added to all the other word probabilities $P(f_1|+)...P(f_n|+)$ to obtain the likelihood of the document $P(d|+)$, which is then multiplied by the prior probability of the *positive* class $P(+)$ to estimate the correct class $\hat{c}$

9. Compared to human-labelled sentiment lexicons, supervised and semi-supervised sentiment lexicons have

   (a) smaller dictionary size (lower "recall") and lower quality of the annotated values (lower "precision")

   (b) larger dictionary size (higher "recall") and higher quality of the annotated values (higher "precision")

   (c) smaller dictionary size (lower "recall") but higher quality of the annotated values (higher "precision")

   (d) larger dictionary size (higher "recall") but lower quality of the annotated values (lower "precision")

10. Which of these statements about Hidden Markov Models (HMMs) is true?

    (a) HMMs can be used to model both for part-of-speech tagging and named entity recognition (NER). For NER, however, they are limited by the fact that they can only recognize individual words as an entity (and not multiple tokens)

    (b) Despite their limitations, they are a popular choice for part-of-speech tagging since they are an unsupervised method

    (c) Retrieving the most likely states sequence does not necessarily mean obtaining the correct sequence; however, at least for part of speech tagging, it works remarkably well

    (d) The Viterbi algorithm does not guarantee retrieval of the most likely states sequence; however, it works remarkably well and it is the only way to simplify the problem to a computationally-tractable state

11. Student A and student B have independently applied the Word2Vec algorithm to obtain word embeddings using the same corpus. With the same vocabulary of words V, student A has the *context* vector c_w^A and the *target* vector t_w^A for every word $w \in V$, while student B has the *context* vector c_w^B and the *target* vector t_w^B for every word $w \in V$.

    Which of the following statements is true?

    I. For a word $w \in V$, $t_w^A \approx t_w^B$

    II. For a pair of words, u and $w \in V$, $(c_u^A)^T t_w^A \approx (c_u^B)^T t_w^B$

    (a) Statement I is true
    (b) Statement II is true
    (c) Both are true
    (d) Neither is true

12. We have learned that pre-trained word vectors work very well in many practical tasks. However, for some special applications, such as medical information retrieval, we might need to retrain the model to cover the domain-specific vocabulary and semantic relationships.

    What might happen if the data set for the specific task is too small?

    I With careful tuning of hyper-parameters and controlled training, the re-trained model can capture well domain-specific semantics while avoiding over-fitting.

    II The structure of the semantic space in the pre-trained model might get destroyed, because the words occurring in the new training set get updated while those that are not in the new training set do not. This can lead to over-fitting.

    (a) Statement I is correct
    (b) Statement II is correct
    (c) Both are correct
    (d) Both are incorrect

13. Which of the following tasks can be used to extrinsically evaluate word embeddings?

I Word analogy test

II Compute word similarities and their correlation with human judgments

III Question answering

IV Sentiment analysis

(a) Tasks I and II

(b) Tasks III and IV

(c) Tasks I, III and IV

(d) All of the above

14. Consider the following grammar.

Grammar rules			**Lexicon**		
S	$\rightarrow$	NP VP	Verb	$\rightarrow$	like | likes | fly | flies
S	$\rightarrow$	VP			| time | times
NP	$\rightarrow$	Det Noun	Noun	$\rightarrow$	time | times | fly | flies
NP	$\rightarrow$	Noun Noun			| fruit | banana | bananas
VP	$\rightarrow$	Verb NP			| arrow | arrows | like | likes
VP	$\rightarrow$	Verb PP	Prep	$\rightarrow$	like
VP	$\rightarrow$	Verb NP PP	Det	$\rightarrow$	the | a | an
PP	$\rightarrow$	Prep NP			

Which of the following sentences can NOT be parsed by this grammar?

(a) *time flies like an arrow*

(b) *fruit flies like a banana*

(c) *arrows fly*

(d) *like time like a fly*

15. Consider the following grammar.

Grammar rules		
S	$\rightarrow$	NP VP
NP	$\rightarrow$	I | you | we
NP	$\rightarrow$	morning | flight
NP	$\rightarrow$	Det Nominal
Nominal	$\rightarrow$	morning | flight
Nominal	$\rightarrow$	Nominal Noun
VP	$\rightarrow$	VP NP
VP	$\rightarrow$	Verb NP

Lexicon		
Noun	$\rightarrow$	morning | flight
Pronoun	$\rightarrow$	I | you | we
Det	$\rightarrow$	a | the
Verb	$\rightarrow$	book | reserve

We use this grammar to fill in a CKY parse table for the sentence *we book the morning flight*.

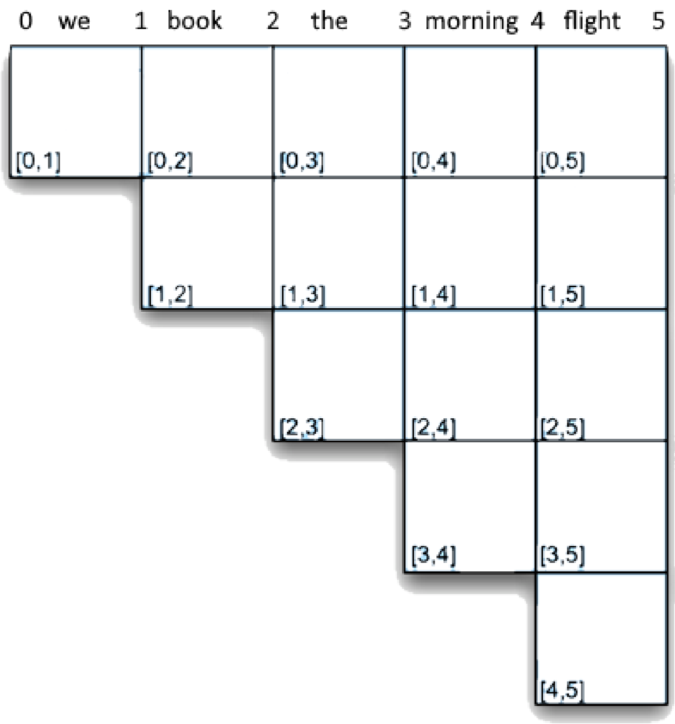
What will the contents of cell [1, 5] be when the entire table is filled?

(a) S

(b) NP

(c) VP

(d) VP, VP



16. Which of the following statements is true?

    For a sentence that is syntactically ambiguous:

    I Dependency parsing normally returns multiple parses

    II Constituency parsing normally returns multiple parses

    III Probabilistic constituency parsing normally returns multiple parses

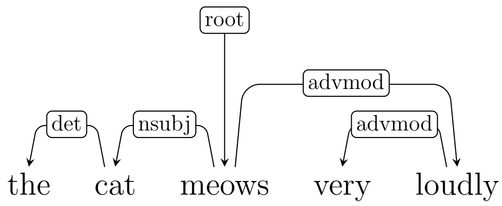
    (a) Statement I is true

    (b) Statement II is true

    (c) Statement III is true

    (d) All of the statements are true

17. Consider the following dependency structure for the sentence *the cat meows very loudly.*



    We use this structure as a reference parse, and simulate the dependency parsing of the sentence using an arc standard transition-based approach. What is the sequence of operators in the trace of this simulated parse?

    (a) SHIFT, SHIFT, LEFTARC, SHIFT, LEFTARC, SHIFT, SHIFT, LEFTARC, RIGHTARC, RIGHTARC

    (b) SHIFT, SHIFT, LEFTARC, SHIFT, LEFTARC, RIGHTARC, SHIFT, RIGHTARC, SHIFT, LEFTARC

    (c) SHIFT, LEFTARC, RIGHTARC, LEFTARC, RIGHTARC, SHIFT, LEFTARC, RIGHTARC, REDUCE, REDUCE

    (d) None of the sequences given above

18. Consider this newspaper article with a funny ambiguous headline:

Boy paralysed after tumour fights back to gain a black belt

A BOY left paralysed after a brain tumour has baffled doctors by making an unexpected recovery – and gaining a black belt in Taekwondo.

Daniel Kimmins, now 14, underwent emergency surgery days before his seventh birthday when doctors discovered the tumour.

After the operation at Bristol's Frenchay Hospital, the youngster was left paralysed down the left side of his body and doctors warned him that it could be permanent.

But now Daniel, from Bath, Somerset, has fought against the odds to make a full recovery.

He first took up the martial art when he was six – and within a year had earned his yellow belt.

By **Mark Reynolds**

ability to speak. We were never given any hope and got the impression he would not recover from the tumour.

"But he has learnt to walk and talk all over again and he is able to do all of the things he did before – which makes me unbelievably proud of him."

At first, Daniel found the Taekwondo a lot harder than he used to.

So he was over the moon to be rewarded with the black belt a few weeks ago.

He said: "It was just absolutely amazing."

Think of what a correct dependency parse of the headline would look like. By a correct dependency parse we mean one that corresponds to the non-funny reading of the headline, reflecting the actual content of the article.

Which of the following statements about the correct dependency parse would be true?

(a) There is an arrow labeled NSUBJ from *fights* to *tumour*

(b) There is an arrow labeled NSUBJ from *boy* to *paralysed*

(c) *Paralysed* is the head of the sentence

(d) *Fights* is the head of the sentence

Open question

The context-free grammar below can be used to assign a syntactic structure to sentences such as the following:

The three happy women believe the men sleep
The two sad women want to swim
The women want to believe the men sleep
The four men want to greet the happy women

Question: provide the CNF version of this grammar.

Grammar rules

S	→	NP VP
NP	→	Det Noun
NP	→	Det Num Noun
NP	→	Det Adj Noun
NP	→	Det Num Adj Noun
VP	→	believe S
VP	→	want INF-VP
INF-VP	→	to VP
VP	→	Verb
VP	→	Verb NP

Lexicon

Verb	→	sleep | swim | greet
Noun	→	men | women
Adj	→	happy | sad
Num	→	two | three | four
Det	→	the