Natural Language Processing: Exam information

Miryam de Lhoneux

Autumn 2024

This document contains information about the exam for the Natural Language Processing course. The exam is divided into two parts: theoretical and practical. Each part has two questions (possibly with subquestions) and each question is worth 5 points. The practical part includes questions that can be of two types: *algorithm application* and *analysis*. I discuss each type of question in turn below, and include grading criteria for each.

The expected length of each answer is between half a page and a page, with a max limit of 1.5 pages, to give you freedom to extend your answer and/or include illustrations (e.g. trees).

The exam is **open book**, you are allowed to bring any paper material you want (we will not check its content) as well as a calculator (any calculator that is not connected to the internet). Prepare the exam with this in mind: memorizing the course material is not what is expected here and writing down answers that contain the course material verbatim does not demonstrate that you understand what you write since you can just copy it directly. Rather, what is expected is that you explain the material in your own words.

1 Theoretical question

A theoretical question tests your understanding of the core concepts, models, and algorithms of the course. It may require you to write pseudocode to illustrate the explanation of an algorithm, or to understand python or pseudocode of an algorithm. Given that the exam is open book, of special importance is to write out full sentences explaining the different concepts in your own words. A list of bullet points is not sufficient to demonstrate your understanding of the material. Dumping a lot of information without relating the different pieces of information together shows that you can identify what is relevant (surface-level understanding of the material) but not more than that. To do well in this question (grade above 3), you need to demonstrate that you can connect the knowledge from the different lectures. To get above 4, you need to demonstrate that you have reflected on the material (in-depth knowledge). Here are some example questions:

- What are regular expressions and how can they be used in NLP systems? How useful are they in current state-of-the-art LLMs?
- In n-gram language modelling, what are the advantages and disadvantages of using a higher n?
- Given pseudocode (or code in python) for BPE training which contains mistakes, can you identify and correct the mistakes? Explain the algorithm, using the correct pseudocode as illustration.
- Explain HMM models for sequence labelling and how to do viterbi decoding for an HMM model (illustrate viterbi with pseudocode). What are advantages and disadvantages of using viterbi?

Grading criteria are in table 1.

Grade	Criteria
1	The answer contains some elements but there are gaps (parts unanswered or answers are not clear, underspecific or inaccurate).
2	The answer contains most of the expected elements but some parts are missing, unclear or inaccurate.
3	The answer contains most relevant parts and is clearly presented. It demonstrates at least surface-level understanding of the material.
4	The answer contains most if not all relevant parts, connects them together and is generally clearly presented. It demonstrates more than surface-level knowledge of the material.
5	The answer contains most if not all relevant parts, connects them together and is generally clearly presented. It demonstrates in-depth knowledge of the material.

Table 1: Grading criteria for theoretical questions.

2 Algorithm application question

In this type of question, you apply an algorithm or a probabilistic model to new data (similar to the exercises in class or the paper exercises in the labs). You briefly analyse the results. The analysis does not need to be extensive here, the aim is to verify that you are not just blindly applying a method but understand why you are doing it. Make sure this is not just a general description of the algorithm but instead, try to clearly illustrate this description with the result of your exercise. Note that we will be forgiving of mistakes, as long as the reasoning is correct: minor mistakes cost you a maximum of 0.25 points (we will do our best to trace back the errors that may propagate through the exercise, so an early mistake will not cost you more than a late mistake). Example questions:

- Given a small corpus, carry out 3 iterations of BPE and give the resulting vocabulary as well as merged text. Do inference on new text and analyse the result.
- Given a small corpus, build a bigram probability table. Compute the probability of a new sentence, given this table, and discuss the result.
- Given an emission and a transition probability table, apply the viterbi algorithm to POS tag a short sentence. Discuss the result.

Grading criteria are in table 2.

Grade	Criteria
1	Some of the exercise has been applied correctly.
2	The general reasoning of how to do the exercise is correct but there is one or more reasoning mistake in parts of the exercise. The analysis is lacking or does not adequately describe the result.
3	The exercise has been applied correctly (possibly with minor mistakes), the analysis is lacking or does not adequately describe the result.
4	The exercise has been applied correctly (possibly with minor mistakes) and the analysis adequately describes the result.
5	The exercise has been applied correctly and the analysis is insightful, it is well illustrated.

Table 2: Grading criteria for algorithm application questions.

3 Analysis question

In this type of question, you will be asked to analyse the output of an NLP tool. Typically, this corresponds to output of code you ran in the notebooks in the labs. If you have gone through the labs and reflected on them, you should do well in this question. Try to not only describe the parts of the course material that explain the output but to also illustrate your answer by referring to specific parts of the output.

Example questions:

- Analyse the output of a tokenizer, e.g.:

  ```
  ['The', 'Solar', 'System', 'is', 'about', '4.568', 'billion', 'years', 'old.[1]']
  ```

  Can you identify what system was used for tokenization? What issues does it present?

- Given a table of emission and transmission probabilities, a sentence, and two different tag sequences for the sentence outputted by two different systems, e.g:

  - `Book|NOUN your|PRON flight|NOUN`
  - `Book|VERB your|PRON flight|NOUN`

  Can you identify which decoding strategy was used by each system? What are stengths and weaknesses of each decoding strategy?

Grading criteria are in Table 3.

Grade	Criteria
1	The answer contains some description of the output but lacks analysis, or the analysis does not really make sense.
2	The answer describes the output adequately but lacks analysis, or the analysis does not really make sense.
3	The answer describes the output adequately and includes some surface-level analysis, without much detail.
4	The answer describes the output adequately and includes more than surface-level analysis with some details.
5	The answer describes the output adequately and includes detailed/deep analysis.

Table 3: Grading criteria for analysis questions.