Master Degree in Computer Engineering

Natural Language Processing
Final Exam

February 21st, 2024

1. **[2 points]** Introduce the notions of word type and word token, and provide a simple example. Do you know any general relation between their numbers in a corpus?

2. **[6 points]** With reference to the feedforward neural language model (FNLM), answer the following questions.

   (a) How does FNLM approximate the probability $P(w_t \mid w_{1:t-1})$?

   (b) Provide and explain the mathematical equations defining FNLM.

   (c) Discuss the training process for FNLM.

3. **[2 points]** When sampling text from a neural language model, what is usually meant by trade-off between coherency and diversity? What is a popular method for modifying these two model behaviors and how does it work?

4. **[6 points]** Consider the task of POS tagging and the following hidden Markov model, consisting of transition and emission matrices specified in terms of costs (negative log probabilities). As an example, the transition from PRON to VERB has score 19, and the transition from VERB to PRON has score 16.

	ADJ	PRON	VERB	PROPN	⟨EOS⟩
⟨BOS⟩	23	13	19	8	14
ADJ	21	24	23	11	21
PRON	22	24	11	23	13
VERB	21	14	13	16	11
PROPN	23	21	12	12	14

	sie	besuchen	Berlin
ADJ	16	18	7
PRON	5	17	16
VERB	18	6	15
PROPN	19	18	5

Consider the German sentence fragment 'sie besuchen Berlin' (literal translation: they visit Berlin). In the following table each entry represents the lowest cost of reaching the associate word/POS pair, starting from ⟨BOS⟩. Use the Viterbi algorithm to fill in the table and to calculate the lowest cost POS tag sequence.

	⟨BOS⟩	sie	besuchen	Berlin	⟨EOS⟩
⟨BOS⟩	0	–	–	–	–
ADJ	–				–
PRON	–				–
VERB	–				–
PROPN	–				–
⟨EOS⟩	–	–	–	–	

(see next page)

5. **[5 points]** With reference to contextualized language models, also called pre-trained language models, answer the following questions.

   (a) Briefly explain the notion of fine tuning, providing a simple example.

   (b) Introduce and motivate the use of adapters in the transformer, outlining the basic architecture of these components.

6. **[5 points]** In the context of transition-based dependency parsing, consider the French sentence 'Au début de la réunion Pierre a présenté sa démission' (literal translation: at the start of the meeting Pierre has presented his resignation) along with the projective dependency tree consisting of the following unlabeled dependency relations.

head	début	présenté	réunion	réunion	début	présenté	présenté	⟨ROOT⟩	démission	présenté
dependent	Au	début	de	la	réunion	Pierre	a	présenté	sa	démission

   Answer the following questions.

   (a) Draw a graphical representation of the dependency tree above, with arcs directed from the head to the dependent.

   (b) Apply to the above tree the oracle presented in class to construct a sequence of training instances for the arc-standard parser.

7. **[7 points]** In the context of neural machine translation, answer the following questions.

   (a) Introduce the notion of word alignment and discuss the possible typologies of this relation.

   (b) How do we model the probability $P(y \mid x)$, with $x = x_1 \cdots x_n$ a sentence in the source language and $y = y_1 \cdots y_m$ a sentence in the target language?

   (c) Introduce the encoder-decoder architecture based on recurrent neural network, dynamic context vector and dot-product attention, that we have illustrated in the class lectures. Present and discuss the model equations.