Master Degree in Computer Engineering

Natural Language Processing
Final Exam

June 27th, 2024

1. **[2 points]** Introduce the algorithm BPE token segmenter (encoder) and the algorithm BPE token merger (decoder), explaining their functions and providing very simple examples of applications.

2. **[6 points]** In the context of POS tagging for Spanish, consider the hidden Markov model consisting of the transition and emission matrices reported below. These matrices are specified in terms of **costs** (negative log probabilities). As an example, the transition from NOUN to ADJ has score 16, and the transition from ADJ to NOUN has score 5.

	NOUN	VERB	ADJ	⟨EOS⟩
⟨BOS⟩	5	11	8	17
NOUN	8	5	16	8
VERB	8	8	15	11
ADJ	5	17	10	16

	Carmen	cuadros	dos	compró
NOUN	5	5	17	17
VERB	18	19	18	5
ADJ	13	11	7	18

Consider the Spanish sentence fragment 'Carmen compró dos cuadros' (literal translation: Carmen bought two paintings). In the following table each entry represents the **lowest cost** of reaching the associated 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⟩	Carmen	compró	dos	cuadros	⟨EOS⟩
⟨BOS⟩	0	–	–	–	–	–
NOUN	–					–
VERB	–					–
ADJ	–					–
⟨EOS⟩	–	–	–	–	–	

3. **[6 points]** Specify the algorithm called skip-gram with negative sampling (SGNS), for the construction of static word embeddings using logistic regression. Introduce and discuss the objective function used by this algorithm.

4. **[2 points]** Briefly outline the four historical periods of the field of natural language processing. Discuss the dominant research paradigm associated to each of these periods, as discussed in lecture 1.

(see next page)

5. **[2 points]** When sampling text using a neural language model, what is usually meant by the trade-off between coherency and diversity? What is a popular method for modifying these two model behaviors and how does it work?

6. **[5 points]** Recall that for a transition-based parser, a configuration consists of a stack, an input buffer, and a set of dependencies constructed so far.

   Let $w_1 w_2 \cdots w_8$ be an input sentence, where each w_i is a word token. We indicate as (w_i, w_j) a dependency arc with head w_i and dependent w_j. Consider the following set of dependencies

   $$A \;=\; \{(w_1, w_2), (w_2, w_3), (w_3, w_4), (w_4, w_6), (w_6, w_5)\}\ .$$

   For the arc-standard parser, show a sequence of actions that takes the parser from the initial configuration with stack = $[\langle \text{ROOT} \rangle]$, buffer = $[w_1, w_2, \ldots, w_8]$, arcs = $\emptyset$, to the intermediate configuration with stack = $[\langle \text{ROOT} \rangle, w_1, w_7]$, buffer = $[w_8]$, arcs = A.

7. **[8 points]** In the context of the task of machine reading, answer the following questions.

   (a) Introduce the notions of query, passage and span and the associated notation.

   (b) Define the main equations of the neural model known as Stanford attentive reader, and explain how the model implements the attention mechanism.

   (c) Make a comparison between the attention implementation in point (b) and the standard attention mechanism as defined for the transformer neural architecture.

8. **[2 points]** With reference to semantic parsing, define the notions of semantic role and thematic grid. Discuss some linguistic examples.