Master Degree in Computer Engineering

Natural Language Processing
Final Exam

June 30th, 2023

1. **[2 points]** Define the general problem of structured prediction and express it in a mathematical way as an optimization problem. Introduce and discuss an example of a natural language processing task that can be formulated as a structured prediction problem.

2. **[7 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 PN to VB has score 7, and the transition from VB to PN has score 9.

	DT	PN	PP	VB	⟨EOS⟩
⟨BOS⟩	7	7	13	17	22
DT	22	11	19	19	19
PN	11	7	9	7	9
PP	7	9	17	14	19
VB	7	9	9	8	9

	ils	sont	allés
DT	15	17	21
PN	5	19	18
PP	19	17	18
VB	17	6	6

   (a) Draw a finite state representation of the model, including the special states ⟨BOS⟩ and ⟨EOS⟩.
   (b) Consider the French sentence fragment 'ils sont allés'. 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⟩	ils	sont	allés	⟨EOS⟩
⟨BOS⟩	0	–	–	–	–
DT	–				–
PN	–				–
PP	–				–
VB	–				–
⟨EOS⟩	–	–	–	–	

3. **[7 points]** Considering contextualized word embeddings, answer the following questions.
   (a) Introduce the general architecture of the model known as ELMo (embeddings from language model).
   (b) With reference to the char-CNN layer, specify and discuss the so-called highway block.
   (c) Specify how the output embeddings at each layer are combined.

4. **[1 points]** Introduce the linguistic resource called WordNet. How is the notion of synset defined in WordNet?

(see next page)

5. **[6 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 $w_1 w_2 \cdots w_9$ 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

   $$A \;=\; \{(w_4, w_3), (w_2, w_4), (w_5, w_2), (w_1, w_5), (w_7, w_6)\} \,.$$

   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 = $[w_1, w_2, \ldots, w_9]$, arcs = $\emptyset$, to the intermediate configuration with stack = $[\langle \text{ROOT} \rangle, w_1, w_7]$, buffer = $[w_8, w_9]$, arcs = A.

6. **[8 points]** In the context of neural machine translation, answer the following questions.

   (a) Introduce the encoder-decoder neural architecture using RNN and static context vector $\mathbf{c}$, and draw the model recurrent equations.

   (b) Discuss the so-called bottleneck problem for the model in (a).

   (c) Augment the model in (a) with the use of dynamic context vector $\mathbf{c}_t$ and dot-product attention.

7. **[1 points]** Introduce and discuss the two evaluation metrics that are most often used for machine reading systems.

8. **[1 points]** In the context of human-machine dialogue, introduce the technique usually called grounding, and provide some example.