

UNIVERSITY OF EDINBURGH  
COLLEGE OF SCIENCE AND ENGINEERING  
SCHOOL OF INFORMATICS

**INFR11157 NATURAL LANGUAGE UNDERSTANDING,  
GENERATION, AND MACHINE TRANSLATION**

**Tuesday 14<sup>th</sup> May 2019**

**14:30 to 16:30**

**INSTRUCTIONS TO CANDIDATES**

- 1. Note that ALL QUESTIONS ARE COMPULSORY.**
- 2. DIFFERENT QUESTIONS MAY HAVE DIFFERENT NUMBERS OF TOTAL MARKS. Take note of this in allocating time to questions.**
- 3. CALCULATORS MAY NOT BE USED IN THIS EXAMINATION.**

MSc Courses

Convener: M.Mistry

External Examiners: W. Knottenbelt, M. Dunlop, M. Niranjan, E. Vasilaki

**THIS EXAMINATION WILL BE MARKED ANONYMOUSLY**

1. (a) EXCLUSIVE OR is a logical function that returns TRUE only if one of its two inputs is TRUE and the other is FALSE. Written as a truth table with 1 as TRUE and 0 as FALSE, it looks like this:

first input $x_1$	second input $x_2$	output $y$
0	0	0
0	1	1
1	0	1
1	1	0

Now recall that a perceptron with two inputs computes a function:

$$y = \begin{cases} 1 & \text{if } w_1x_1 + w_2x_2 > \theta \\ 0 & \text{otherwise} \end{cases}$$

You can set  $w_1$ ,  $w_2$  and  $\theta$  in different ways to implement different functions.

- Could you implement EXCLUSIVE OR with a single perceptron? Why or why not?
- Illustrate how you would implement EXCLUSIVE OR using a multilayer perceptron—which can be a single perceptron, if that is sufficient. You may assume that there are perceptrons to compute AND, OR, and NOT, you do not need to show the weights of these perceptrons.

[4 marks]

- (b) A language model defines the probability of a sequence  $x = x_1 \dots x_{|x|}$  as:

$$P(x) = \prod_{i=1}^{|x|+1} P(x_i \mid x_1, \dots, x_{i-1})$$

Now consider three types of language model: (1)  $n$ -gram, (2) feedforward, and (3) recurrent. For each model, explain the independence assumptions that it makes in terms of the above equation.

[3 marks]

- (c) Are the following statements true or false? Give a short (1-sentence) justification of each answer.

- A machine translation model cannot learn how to translate words that were never observed at training time.
- Self-attentional models rely on positional encodings in their input, but recurrent neural network models do not.

[4 marks]

- (d) In an RNN attentional sequence-to-sequence model, the encoder is typically a bidirectional RNN. Why is the decoder unidirectional?

[2 marks]

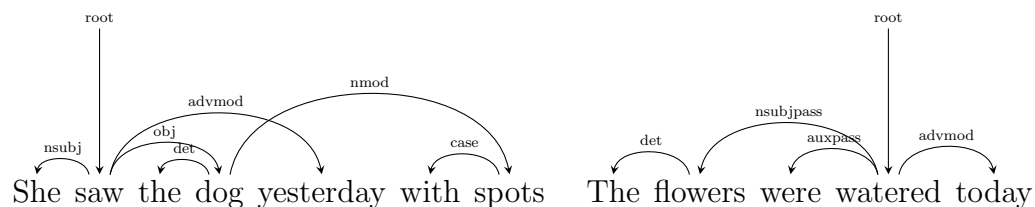
*QUESTION CONTINUES ON NEXT PAGE*

*QUESTION CONTINUED FROM PREVIOUS PAGE*

- (e) Most automatic MT evaluation metrics, including BLEU, compare the translation output to a human reference translation. We have learned that many other metrics have been proposed in the last decade, and that their quality is typically judged based on correlation with human judgments at MT competitions (such as the WMT shared tasks). One exception is the metric MTeRater, which uses an automated essay scoring system to measure the fluency of the translation. MTeRater has the advantage of not requiring reference translations, and the authors found that the sentence-level correlation with human judgments of MTeRater is equivalent to that of BLEU.

Would you recommend optimizing an MT system to maximize a fluency-based MT metric like MTeRater instead of BLEU? Why or why not? [4 marks]

- (f) Consider the two sentences and their dependency parses, below.



Which tree is projective, and which is not? Justify your answer. [3 marks]

2. You join a company as an NLP engineer. The company uses a standard sequence-to-sequence model as used in NMT (an attentional encoder-decoder with self-attention, i.e. the Transformer) to model many different tasks. Discuss how the modelling assumptions in NMT may be unsuitable for these different tasks, and how you could change the model to improve the quality, reliability or simplicity of the model. Try to point out one inappropriate assumption, and one suggestion to improve the model for each task. Your answers should be short: each assumption or improvement should be described in a sentence or two. For all models, you can assume that the company has labelled datasets:

- (a) A part-of-speech tagger. Given a sequence of words, produce a sequence of POS tags.

**Input** That parrot is definitely deceased .

**Output** DT NN VBZ RB VBN .

[5 marks]

- (b) A sentiment analysis system. Given a short text, determine whether the text expresses a positive or negative sentiment.

**Input** This movie is so bad it took me 3 watches until I could finish it without falling asleep.

**Output** negative

[5 marks]

- (c) A chat bot. Given an utterance in a conversation, produce a response. The model is trained on logs of online conversations.

**Input** Alexa, what book would you recommend?

**Output** 1984 is a fantastic book about omnipresent surveillance.

[5 marks]

3. A startup wants to hire you to develop an NLP system that fully automates **medical coding** from doctor's reports. One type of medical code identifies a diagnosis indicated in the report, while another identifies a prescribed treatment. Each code comes from a finite set. Here is a simplified example:

**Input** (doctor's report): *Patient is a 27-year-old white male. Height is 74 inches, weight 220 lbs. Patient states he is allergic to penicillin, but has no other outstanding medical history. Does not smoke, exercises moderately.*

*Patient presents with chills, headache, cough, fever (101 degrees), difficulty breathing. Examination via stethoscope yields heavy rales. Percussion test on thorax suggests buildup in lungs. Streptococcal pneumoniae suspected.*

*Prescribed patient two weeks of 500mg azithromycin (Zithromax), and scheduled follow-up for next week.*

**Output 1** (medical diagnosis code): **pneumonia**

**Output 2** (medical treatment code): **azithromycin-500mg**

Medical codes determine medical bills and authorize patients to receive certain treatments like prescription medicines, so they are a critical part of the modern healthcare system. They are also costly because they are produced by trained professionals. Many companies are now aiming to reduce this cost by using AI to automate medical coding, and a successful solution could earn your company a lot of money.

- (a) Design a simple model to predict both types of medical codes from reports. You may explain with words, diagrams, and/ or maths, but it should be clear how the architecture of your model treats input and output. [3 marks]
- (b) Explain how you would train and test your model, and what you think its weaknesses might be. Your answer should state any assumptions you make about the data, and justify your technical decisions. [4 marks]
- (c) Who would benefit from this system? Who could be harmed, and how? [4 marks]
- (d) What measures could you take to minimize these harms? [4 marks]

Your answer to this question might reasonably require two pages of the exam booklet. Longer than that is unnecessary. A concise answer that is clear and specific will earn more points than a long answer that is vague and imprecise.