DCU

DUBLIN CITY UNIVERSITY

SEMESTER 2 EXAMINATIONS 2017/2018

MODULE: CA4012 - Statistical Machine Translation

PROGRAMME(S):
  CASE   BSc in Computer Applications (Sft.Eng.)
  ECSAO   Study Abroad (Engineering & Computing)
  CPSSD   BSc in ComputationalProblem Solv&SW Dev.

YEAR OF STUDY: 4,O

EXAMINER(S):

          Prof. Andrew Way   (Ext: 5074)
          Dr. Jinhua Du   (Ext: 6716)
          Dr. Mohammed Hassanuzzaman (Ext: 6912)
          Prof. Brendan Tangney
          Dr. Hitesh Tewari

TIME ALLOWED: 3 Hours

INSTRUCTIONS: **Answer 5 questions. You must attempt at least one question from each of Sections A, B and C.**
All questions carry equal marks.

SECTION A

QUESTION 1 *[TOTAL MARKS: 20]*

Q 1(a) **[4 Marks]**
What are the main advantages of the log-linear model of SMT compared to the noisy channel model of SMT? How might you argue that in some cases, the move from the noisy channel model to the log-linear model could be interpreted as a disadvantage?

Q 1(b) **[6 Marks]**
Provide the fundamental equations of these two SMT models. Name the different components in each, and describe their basic function. Demonstrate how the two equations might be equivalent.

Q 1(c) **[6 Marks]**
Until quite recently, the freely available web-based system Google Translate used entirely phrase-based SMT models, but now, certain language pairs are serviced via Neural MT (NMT). How would you build a case for switching over from SMT to NMT? What would the main differences be for Google in providing this new service compared to their previous offering?

Q 1(d) **[4 Marks]**
Machine-learning researchers have had remarkable success in building specific models to beat the very best humans in certain fields, e.g. Go, Chess, Jeopardy. Why does MT continue to perform at a level lower than human translators, in general?

[End of Question 1]

QUESTION 2 *[TOTAL MARKS: 20]*

Q 2(a) **[8 Marks]**

SMT learns from data. What two types of data do we need to build SMT systems? What recommendations regarding training data would you provide to someone intending to build an SMT system so that the best possible performance was achieved?

Q 2(b) **[4 Marks]**

Ideally, the translations output by MT systems should be both adequate and fluent. Explain why. Which components of an SMT system are primarily responsible for ensuring that these two constraints are met?

Q 2(c) **[3 Marks]**

Assume the reference translation for a particular source sentence is "Fruit flies like bananas". Provide examples of (i) an inadequate but fluent translation; (ii) an adequate disfluent translation; and (iii) an inadequate disfluent translation.

Q 2(d) **[5 Marks]**

Despite the fact that MT has never been used as much as it is today, some translators continue to argue that MT will never be useful to them. Why do you think this is? Give **two** reasons that you believe would persuade them to try MT, and find it useful in their work.

[End of Question 2]

[END OF SECTION A]

QUESTION 3   *[TOTAL MARKS: 20]*

Q 3(a)   **[10 Marks]**
Given the following:

1. Source sentence:
   Nous avons besoin d'évaluer nos systèmes statistiques de traduction.

2. Machine translation outputs:
   We need to evaluate our statistical systems translation.
   We evaluate our statistical translation systems.

3. Translation reference:
   We need to evaluate our statistical translation systems.

Calculate the BLEU-3 scores of the two candidate translations.

Q 3(b)   **[5 Marks]**
Why is standard BLEU not suitable to evaluate MT output at sentence level
(especially when sentences are short)? How would you modify BLEU to make it
suitable to be used at sentence level?

Q 3(c)   **[5 Marks]**
What are the main benefits of performing automatic evaluation of MT output
compared to human evaluation? At the same time, explain why it remains important
to conduct human evaluations from time to time.

[End of Question 3]

QUESTION 4   *[TOTAL MARKS: 20]*

Q 4(a)   **[10 Marks]**
Describe Word Error Rate (WER) in your own words. How is it calculated? What is it used for? What are the benefits/disadvantages associated with WER compared to the BLEU MT evaluation metric?

Q 4(b)   **[5 Marks]**
Given the following strings:

        Reference: They are responsible for the airport security.
        MT output: They responsible security are airport.

Calculate the WER Score.

Q 4(c)   **[5 Marks]**
Explain the differences between WER and Translation Edit Rate (TER).

[End of Question 4]

[END OF SECTION B]

SECTION C

QUESTION 5   *[TOTAL MARKS: 20]*

Q 5(a)   **[6 Marks]**
What is the EM algorithm? How is EM used to compute the word alignment and lexical probabilities in IBM Model 1? Provide **one** specific example of your own choosing which demonstrates how EM is calculated.

Q 5(b)   **[5 Marks]**
Describe the basic principles of word-based SMT and phrase-based SMT. Name **three** advantages and disadvantages of (i) word-based SMT and (ii) phrase-based SMT models.

Q 5(c)   **[5 Marks]**
In phrase-based SMT, what rules need to be followed in order to extract parallel phrases from a word-aligned parallel corpus?

Q 5(d)   **[4 Marks]**
Explain the main differences between higher IBM Models 2—5 and IBM Model 1.

[End of Question 5]

QUESTION 6 [TOTAL MARKS: 20]

Q 6(a) [10 Marks]

Assume the following partial phrase table:

ta	he	0.6

ti	plays	0.6
ti	is playing	0.4

zuqiu	football	0.6
zuqiu	soccer	0.4

ta ti	he plays	0.7
ta ti	he is playing	0.3

ti zuqiu	plays football	0.5
ti zuqiu	plays soccer	0.5

Consider the following input sentence: *ta ti zuqiu*

Assume that (i) only monotone word order is permitted; and (ii) the language model is ignored. Then we have the following search diagram (partial search space):

(1)	he	0.6

(2)	plays	0.6
(3)	is playing	0.4
(4)	he plays	0.7
(5)	he is playing	0.3

(6)	football	0.6
(7)	soccer	0.4
(8)	football	0.6
(9)	soccer	0.4
(10)	football	0.6
(11)	soccer	0.4
(12)	football	0.6
(13)	soccer	0.4
(14)	plays football	0.5
(15)	plays soccer	0.5
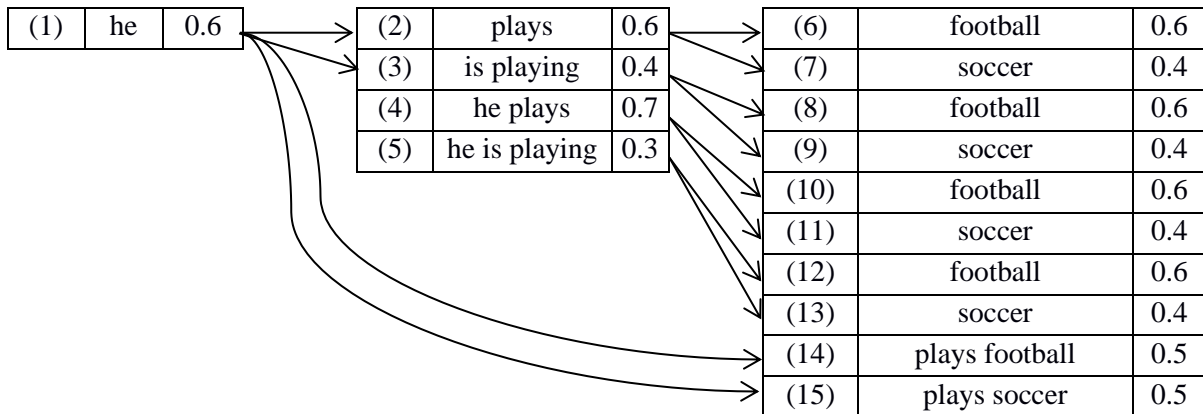
Given the search diagram, calculate the probabilities for all possible hypotheses (search paths). Indicate also which hypothesis provides the optimal translation for the input sentence.

Q 6(b) [6 Marks]

Given the search diagram above, indicate which groups of hypotheses can be recombined and indicate which hypothesis should be selected to represent each group.

Q 6(c) [4 Marks]

Assuming threshold pruning after recombination, where the threshold is 0.5, indicate which hypotheses will be pruned in the last stack.

[End of Question 6]

QUESTION 7 *[TOTAL MARKS: 20]*

Q 7(a) **[5 Marks]**
What is the activation function in neural machine translation (NMT)? Provide the formulae of **two** commonly used activation functions and specify the output range of these functions.

Q 7(b) **[5 Marks]**
What do you understand by the term "gradient"? In your own words, describe the gradient vanishing problem. Provide **one** method to alleviate this problem in neural networks.

Q 7(c) **[5 Marks]**
What is the overfitting problem in neural networks? Describe **two** methods to alleviate this problem.

Q 7(d) **[5 Marks]**
Draw the basic architecture of a recurrent neural network (RNN), and explain how the context of the input sequence is memorised in an RNN.

[End of Question 7]

[END OF SECTION C]

[END OF EXAM]