Ollscoil Chathair
Bhaile Átha Cliath
Dublin City University

AUGUST/RESIT EXAMINATIONS 2018/2019

MODULE: CA4012 - Statistical Machine Translation

PROGRAMME(S):

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

YEAR OF STUDY: 4,O,X

EXAMINER(S):

Prof Andrew Way	(Internal)	(Ext: 5074)
Dr Mohammed Hasanuzzaman	(Internal)	(Ext: 6719)
Dr Dimitar Shterionov	(Internal)	(Ext: 6719)
Dr Hitesh Tewari	(External)	
Prof Brendan Tangney	(External)	

TIME ALLOWED: 3 Hours

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

PLEASE DO NOT TURN OVER THIS PAGE UNTIL YOU ARE INSTRUCTED TO DO SO.
The use of programmable or text storing calculators is expressly forbidden.
Please note that where a candidate answers more than the required number of questions, the examiner will mark all questions attempted and then select the highest scoring ones.

There are no additional requirements for this paper.

SECTION A

QUESTION 1 ***[TOTAL MARKS: 20]***

Q 1(a) **[6 Marks]**

For a language pair of your choice, provide **three** examples of translational phenomena which demonstrate why MT is a difficult problem, no matter what type of system might be built.

Q 1(b) **[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(c) **[6 Marks]**

Until recently, the freely available web-based system Google Translate used entirely phrase-based SMT models, but now its engines are built using Neural MT (NMT). From the perspective of MT, what factors would Google have taken into account in coming to this decision? What would the main differences be for Google in providing this new service compared to their previous offering?

Q 1(d) **[4 Marks]**

With the advent of AI approaches, some researchers consider MT to be a 'solved problem'. Do you agree with that claim? Would you expect NMT to beat SMT on all language pairs, in all domains?

[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]**

Why did developers of SMT systems switch from word-based to phrase-based models?

Q 2(c) **[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(d) **[4 Marks]**

MT has never been used as much as it is today, but 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, which could prove useful in their work.

[End of Question 2]

[END OF SECTION A]

SECTION B

QUESTION 3 ***[TOTAL MARKS: 20]***

Q 3(a) **[8 Marks]**

Explain the *n*-gram language model. Show how a bigram language model would decompose the following sentence in order to calculate its probability, both with and without sentence boundaries.

"I strongly believe that Australia is going to win ICCWC2019."

Q 3(b) **[3 Marks]**

How can we measure the quality of a language model? Why is this useful?

Q 3(c) **[4 Marks]**

Describe **four** techniques to manage very large language models.

Q 3(d) **[5 Marks]**

Assume three separate language models based on unigrams, bigrams and trigrams trained on standard English. Based on each of these language models, how would you expect the probability of the sentence 'the doughnut Homer ate' to compare to the probability of 'Homer ate the doughnut'?

[End of Question 3]

QUESTION 4 ***[TOTAL MARKS: 20]***

Q 4(a) **[5 Marks]**

What is the main task of decoding? Explain **two** types of errors encountered in decoding. How can the quality of the decoding process be evaluated?

Q 4(b) **[9 Marks]**

In decoding, what is hypothesis recombination? Why it is important? Explain **two** types of pruning strategies, using examples of your choice.

Q 4(c) **[6 Marks]**

Assume the following partial phrase table:

ta	she	0.4
shanchang	*likes*	0.3
shanchang	*is good at*	0.7
paobu	*running*	0.6
paobu	*run*	0.4
ta shanchang	*she likes*	0.2
ta shanchang	*she is good at*	0.8
shanchang paobu	*likes running*	0.3
shanchang paobu	*is good at running*	0.7

Assume that:

i) only monotone word order is permitted;

ii) the language model is ignored.

Draw the search diagram for the following input sentence:

ta shanchang paobu

[End of Question 4]

[END OF SECTION B]

SECTION C

QUESTION 5 ***[TOTAL MARKS: 20]***

Q 5(a) **[4 Marks]**

Explain **two** approaches to human evaluation of MT output. Identify **two** drawbacks to human evaluation.

Q 5(b) **[6 Marks]**

Name **three** issues that the METEOR evaluation metric addresses which BLEU, WER and TER do not. Provide examples to support your answer.

Q 5(c) **[6 Marks]**

Given the following two translations (from MT system 1 and MT system 2), compute their BLEU score (max. *n*-gram size 3) with respect to the reference.

Candidate MT system 1: near the shore the ship sank
Candidate MT system 2: the big ship sank close to the shore
Reference: the large ship sank near the shore

Q 5(d) **[4 Marks]**

Identify **two** drawbacks of automatic evaluation metrics, and **two** reasons why they are useful.

[End of Question 5]

QUESTION 6 ***[TOTAL MARKS: 20]***

Q 6(a) **[6 Marks]**

Training an MT system is typically based on a corpus of parallel sentences. Why do we need to compute word alignments? What alignment patterns are you aware of, and why is word alignment a challenging task?

Q 6(b) **[7 Marks]**

Explain the four steps of the expectation maximisation algorithm (as defined for translation modelling). Focus on what counts or probabilities are computed at each step and how are they connected.

Q 6(c) **[7 Marks]**

What are the differences between the Normal and the Simplified IBM model 1? For the Simplified IBM model 1, write down the formula and explain what it takes into account in terms of word alignment.

[End of Question 6]

QUESTION 7 ***[TOTAL MARKS: 20]***

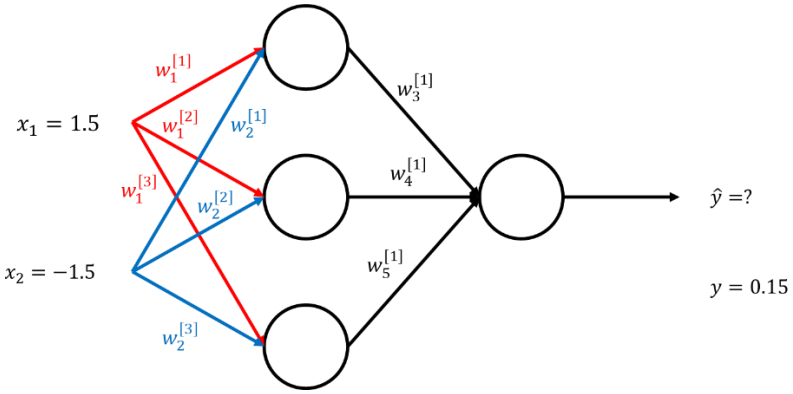
Q 7(a) **[7 Marks]**

What are the differences between a multi-layer perceptron and a Recurrent Neural Network? Which one would you use in an encoder-decoder NMT architecture, and why?

Q 7(b) **[6 Marks]**

Below we provide a simple neural network with two inputs, one output layer, and one hidden layer. All biases are equal to 1. All initial weights are uniformly distributed. Their values are noted in the table below.

Given the inputs x1 = 1.5, x2 = -1.5 and the expected output of y = 0.15, choose two activation functions and for each perform one forward pass through this network. Explain the differences between the chosen activation functions. Fill in the blanks in the table below the network, corresponding to the weights, network output and error after the forward and the backward steps. Use the quadratic cost function in your calculations.



			Weights from input 1 to hidden layer			Weights from input 2 to hidden layer			Weights from hidden layer to output layer					
	Input 1	Input 2	$w_1^{[1]}$	$w_1^{[2]}$	$w_1^{[3]}$	$w_2^{[1]}$	$w_2^{[2]}$	$w_2^{[3]}$	$w_3^{[1]}$	$w_4^{[1]}$	$w_5^{[1]}$	Network Output ($\hat{y}$)	Expected Output (y)	Error
Forward (activation function 1)	$x_1 = 1.5$	$x_2 = -1.5$	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3		y = 0.15	
Forward (activation function 2)	$x_1 = 1.5$	$x_2 = -1.5$											y = 0.15	

Q 7(c) **[7 Marks]**

Explain the gradient descent algorithm for training neural networks. What is the difference between gradient descent and stochastic gradient descent? What is the role of the learning rate?

[End of Question 7]

[END OF SECTION C]

[END OF EXAM]