DUBLIN CITY UNIVERSITY

AUGUST/RESIT EXAMINATIONS 2017/2018

MODULE: CA4012 - Statistical Machine Translation

PROGRAMME(S):

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

YEAR OF STUDY: 4,O

EXAMINER(S):

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

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.

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) **[9 Marks]**

Any statistical approach to MT requires the availability of aligned bilingual corpora which are (i) large, (ii) good-quality, and (iii) representative. Explain why all three requirements are important. What are the potential problems if the training corpus is (iv) small, (v) noisy, and (vi) unrepresentative compared to the intended test data?

Q 1(b) **[7 Marks]**

Provide the fundamental equations of (i) the noisy channel model of SMT, and (ii) the log-linear model of SMT. With reference to these equations, name the different components in both models, and describe their basic function.

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

Give **two** reasons why customised engines are likely to produce better output than a freely available web-based system such as Google Translate.

[End of Question 1]

QUESTION 2 ***[TOTAL MARKS: 20]***

Q 2(a) **[5 Marks]**
Until the 1980s, all MT systems were rule-based. Now pretty much all MT systems are corpus-based. Why did corpus-based MT replace rule-based MT? Provide **two** examples where you think linguistic rules may still have a place in today's state-of-the-art MT systems.

Q 2(b) **[6 Marks]**
The number of translation use-cases where there is no place for a human-in-the-loop are increasing. Provide **three** use-cases where MT is the only solution, i.e. that there is no place for human intervention in the translation pipeline for such use-cases.

Q 2(c) **[5 Marks]**
Recently, Microsoft have claimed to have achieved "human parity" for Chinese-to-English neural MT (NMT), while SDL have claimed to have "cracked Russian-to-English NMT". What do you think of such claims? How would you propose to test their validity?

Q 2(d) **[4 Marks]**
Given claims such as those in 2(c) above, some human translators are fearful of the impact of MT on their profession. Give **two** reasons why human translators remain essential cogs in the MT pipeline.

[End of Question 2]

END OF SECTION A

SECTION B

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

Q 3(a) **[6 Marks]**

What are the **three** main types of MT evaluation? Briefly describe how you would conduct each of these types of evaluation.

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

When deciding whether to adopt MT or not, apart from translation quality, what other criteria need to be taken into account?

Q 3(c) **[6 Marks]**

Why was the “brevity penalty” introduced in the BLEU automatic evaluation metric? Is fluency accounted for in BLEU via “*n*-gram precision” or “word precision”? How can the correlation between BLEU and human quality scores be improved?

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

How can the MT quality on new text be predicted? Provide **three** features which could be useful in this regard.

[End of Question 3]

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

Q 4(a) **[8 Marks]**
Describe in your own words the "*n*-gram language model". Show how a bigram language model would decompose the following sentence to calculate its probability, both with and without sentence boundaries:

"They did n't evaluate their systems ."

Q 4(b) **[4 Marks]**
How can the quality of a language model be evaluated? Why is this useful?

Q 4(c) **[5 Marks]**
Why do we need "smoothing" in language modelling? Describe **two** smoothing methods.

Q 4(d) **[3 Marks]**
Describe **three** techniques that can be used to manage very large language models.

[End of Question 4]

END OF SECTION B

SECTION C

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

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

Assume the following Chinese—English sentence pairs:

S_1	S_2
diannao *computer*	*xin diannao* *new computer*

The source side is Chinese, and the target side is English. In this question, the *NULL* token is ignored. Assuming that only one-to-one alignment is allowed, list all possible word alignments for the two sentence pairs.

Q 5(b) **[10 Marks]**

Considering all the word alignments you computed in (a), (i) state what the following translation probabilities will be after **two** iterations of the Expectation Maximisation algorithm, and (ii) show all the steps followed to arrive at these values:

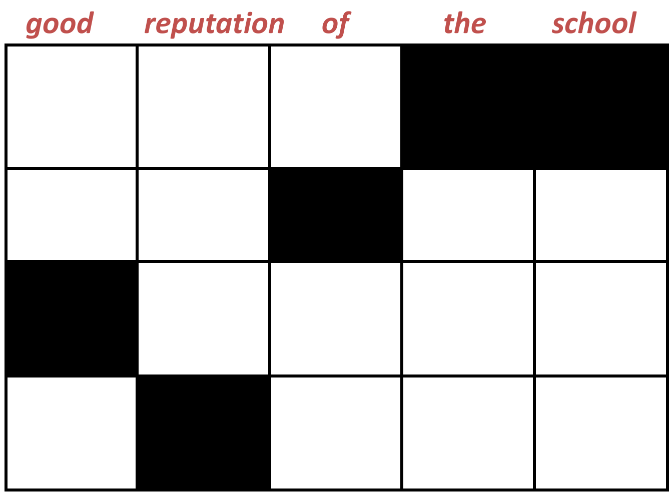
$t(computer|diannao)$
$t(new|diannao)$
$t(new|xin)$
$t(computer|xin)$

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

Assume the following English—Chinese parallel sentence pair:

good reputation of the school
xuexiao de hao shengyu

List all phrase pairs that are consistent with the following word alignment:



[End of Question 5]

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

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

In the context of phrase-based SMT: (i) what is the definition of the term "phrase"? (ii) List **three** advantages of using phrases as atomic units in phrase-based SMT compared to word-based SMT.

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

In order to build a phrase-based SMT model from a parallel corpus: (i) what are the **three** basic steps that need to be followed? (ii) What kinds of methods can we use to obtain a symmetrised word alignment from two unidirectional word alignments? (iii) In the process of extracting parallel phrases, what rules need to be followed?

Q 6(c) **[8 Marks]**

After phrase pairs are extracted from the symmetrised word alignment, (i) in your own words, describe the basic idea of how phrase pairs are scored.

(ii) The following table provides four different Chinese translations for the English phrase "Good morning".

Translation	**Counts**
zaoshang hao	50
zaoshang	5
zhongwu hao	15
nihao	30

Using the method you described in (i), calculate the probability of each phrase pair.

[End of Question 6]

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

Q 7(a) **[4 Marks]**
Why do we typically use “recurrent” neural networks instead of “feed-forward” neural networks in neural machine translation (NMT)? Describe the **single** fundamental difference between these two architectures.

Q 7(b) **[4 Marks]**
Explain what the “encoder” and the “decoder” do in an encoder-decoder NMT system.

Q 7(c) **[4 Marks]**
A baseline encoder-decoder NMT system can be improved by means of an attention mechanism.

(i) what is an “attention mechanism”?
(ii) how can the attention mechanism improve translation quality compared to the baseline encoder-decoder NMT system with no attention?

Q 7(d) **[4 Marks]**
In NMT, briefly describe the stochastic gradient descent algorithm. List **two** commonly used optimisers for NMT, and briefly describe how they work.

Q 7(e) **[4 Marks]**
In NMT, briefly describe the term “activation function”. List **two** commonly used activation functions in neural networks, and briefly describe how they differ.

[End of Question 7]

END OF SECTION C

[END OF EXAM]