Ollscoil Chathair
Bhaile Átha Cliath
Dublin City University

SEMESTER 2 EXAMINATIONS 2021/2022

MODULE: CA4012 - Machine Translation

PROGRAMME(S):

CASE BSc in Computer Applications (Sft.Eng.)

YEAR OF STUDY: 4

EXAMINER(S): Prof. Andy Way (ext. 5074)
Dr. Maja Popović (ext. 6736)
Dr. Guodong Xie
Dr. Sheila Castilho
Dr. Pintu Lohar

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) **[6 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.

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]**

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(d) **[3 Marks]**

Do you agree with the claim by some researchers that with the advent of AI approaches, MT should now be considered to be a ‘solved problem’?

[End of Question 1]

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

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

Machine translation is being used by millions of people on a daily basis. Provide *two* use-cases which freely available services such as Google Translate are good for, and provide reasons behind your selection.

Q 2(b) **[6 Marks]**

What is the market need for MT companies which build customised engines for their clients? Why can those clients expect better translation quality compared to using (say) Google Translate?

Q 2(c) **[6 Marks]**

Give *three* ways in which professional human translators are of critical importance in the development and testing of MT systems.

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

Despite the obvious advances in quality seen from the move to neural models, why do you think some human translators are still reluctant to embrace MT technology? How might you try to persuade them that MT can be a useful tool in the translator's armoury?

[End of Question 2]

[END OF SECTION A]

SECTION B

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

Q 3(a) **[9 Marks]**

Calculate the WER score for the source sentence (SRC), reference sentence (REF) and the MT output hypothesis (HYP) below:

SRC: Kuopion kaupunginvaltuusto hyväksyi liitoksen yksimielisesti maanantaina
REF: The union was unanimously approved by the Kuopio City Council on Monday
HYP: The city council unanimously approved the joint Niiralan on Monday

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

Explain the concepts of “adequacy”, “fluency” and “comprehensibility” and state why each of them is important for the evaluation of MT output.

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

What are advantages and disadvantages of automatic evaluation of MT systems?

[End of Question 3]

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

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

Explain the Markov assumption. Why do we need to take it into account when building *n*-gram language models?

Q 4(b) **[7 Marks]**

Consider the following sentences:

- *I love the cat .*
- *I love the dog .*
- *They see the dog .*
- *They talk to the girl .*
- *They talk to the dog .*
- *I talk to the cat .*

Using a unigram language model, calculate the probability of the sentence "*I see the cat .*"

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

Explain the advantages and disadvantages of lower and higher order *n*-gram models.

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

How can the problem of unseen *n*-grams be overcome in count-based *n*-gram language modelling? Describe one possible method which can be applied to resolve this issue.

[End of Question 4]

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

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

Explain the differences between the IBM models of word alignment. How does the "Hidden Markov model (HMM)" differ from "IBM model 2"?

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

What is the main disadvantage of using sentence-level translation probabilities in SMT? Derive the fundamental equation of translation probability when word alignment is introduced.

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

Name and define *four* types of word alignments based on mapping cardinality between source and target words. Which of those alignments are supported by the IBM-models?

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

Calculate IBM-1 lexical probabilities for the following parallel training corpus:

English	**French**
the house	la maison
the dog	le chien
the cat	le chat

[End of Question 5]

[End of SECTION B]

SECTION C

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

Q 6(a) **[10 Marks]**

Select the following statements which represent the characteristics of neural MT. (2 marks for each correct choice, -2 marks for each incorrect choice but the total score cannot be negative).

a) it can be regarded as using only local context
b) it can be regarded as using global context
c) it can be regarded as being guided by neural language model
d) the neural language model is one of its many components
e) it includes a coverage constraint mechanism that enables avoiding repetitions and omissions in translation
f) sometimes a part of a sentence can be translated more than once
g) there are no obvious word alignments
h) it is possible to provide a model introspection and understand well the source of any translation errors
i) because of the sub-word technology, it sometimes generates misspelled or non-existing words

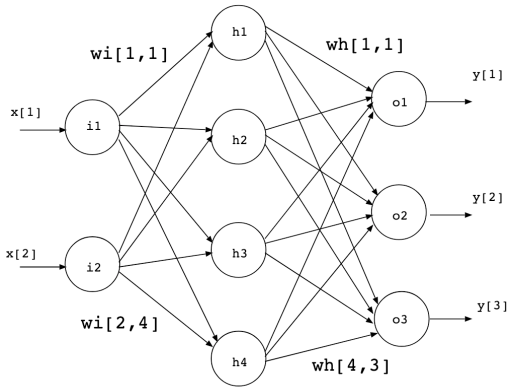
Q 6(b) **[5 Marks]**

If a neural network with three artificial neurons in the output layer is used for classifying images of animals, and the outputs of the network represent the labels y_1 = “cat”, y_2 = “dog” and y_3 = “mouse”, which activation function would you choose for the neurons in the output layer? Explain your reasoning.

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

For the neural network on the right, the following weights connecting the input layer with the hidden layer are given:

$w^i_{11} = 0.7$
$w^i_{12} = 1.4$
$w^i_{13} = 3.2$
$w^i_{14} = -0.3$
$w^i_{21} = -2.3$
$w^i_{22} = 1.2$
$w^i_{23} = -4.7$
$w^i_{24} = 2.3$



If the input values coming into the network are x_1=3 and x_2=2, what will be the input to the activation function of the fourth neuron in the hidden layer? (h_4)

[End of Question 6]

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

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

Explain why feed-forward neural networks are not appropriate for MT, and why recurrent neural networks are.

Q 7(b) **[8 Marks]**

The following Spanish sentence:

Me gustan los gatos

should be translated into English as:

I like cats

If we have an NMT system which decodes using greedy search, what would be the translation of the entire sentence if the following probabilities of partial hypotheses are generated by the model:

<u>first target word:</u>

they 0.28
I 0.16
we 0.10
like 0.02
it 0.01

<u>first two target words:</u>

they like 0.15
they are 0.10
I like 0.11
I want 0.09
I think 0.03
we like 0.05
we will 0.02
like me 0.002
like you 0.001
it will 0.001

<u>first three target words (complete hypothesis):</u>

they are nice 0.002
they are here 0.001

they are mine 0.003
they like cats 0.08
they like dogs 0.06
they like me 0.01
I like cats 0.09
I like dogs 0.05
I think we 0.002
I think that 0.0001
we like cats 0.01
we like mice 0.005
we will go 0.0012
we will see 0.0001
like me , 0.0012
like you , 0.0007
it will be 0.0002
it will go 0.0001

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

Explain what is meant by the term ‘’vanishing gradient”. What network architectures are prone to it, under which circumstances?

[End of Question 7]

QUESTION 8 ***[TOTAL MARKS: 20]***

Q 8(a) **[5 Marks]**
Name and describe *five* differences between RNN and transformer architectures used for NMT.

Q 8(b) **[5 Marks]**
Explain the term ‘’attention” in the context of neural networks. Which types of attention are used in the transformer architecture?

Q 8(c) **[5 Marks]**
What are ‘’word representations”? Where do they come from?

Q 8(d) **[5 Marks]**
What is the dimension of the input layer of a recurrent neural language model? Explain your answer.

[End of Question 8]

[End of SECTION C]

[END OF EXAM]