ECS7001 NN & NLP
Assignment 2: Neural Machine Translation and Neural Dialogue Systems

March 31, 2020

In this assignment, you will gain practice with two of the most popular applications of NN in NLP: Neural Machine Translation and Dialogue Systems.

The assignment consists of two parts:

- In Part A of the assignment, to be carried out in Lab 8, week 8, and Lab 9, week 9, you will first implement a version of the seq2seq NMT model, and then you will enrich that with attention.
- In Part B, to be carried out in weeks 10-12, you will implement a neural dialogue act tagger, comparing different methods, and then optionally extend that either by incorporating state-of-the art transfer learning (BERT) or by creating a generative seq2seq dialogue agent.

When all parts of the assignment are completed, you will have to submit two things:

- A PDF report describing what you did (instructions below);
- Your completed Python code.

The deadline for returning all completed parts of the assignment (Parts A and B) is **10:00:00 Thursday 23rd April 2020**.

Part A: Neural Machine Translation [50 marks]

For this part of the assignment, worth 50 marks in total, you will have to carry out the steps specified by the Lab script ("Lab 8 - Neural Machine Translation" - separately provided). These marks are broken down among the several parts of the assignment as follows (please refer to the lab script):

1. Task 1: Implementing the encoder [15 marks].
   Your report must include the code and an explanation.

2. Task 2: Implementing the decoder [15 marks].
   Again, you must include the code, an explanation, the BLEU score, and a sample of the output

3. Adding attention [20 marks]. *Again, you must include the code, an explanation, the BLEU score, and a sample of the output*

Part B: Dialogue [50 marks]

For this part of the assignment, worth 50 marks in total, you must carry out the steps specified by the Lab script ("Labs 10-12 - Dialogue Act Tagging" - separately provided). These marks are broken down among the several parts of the assignment as follows (please refer to the lab script):

1. Task 1: Implementing an utterance-based tagger, using standard text classification methods from lectures [20 marks].
   Your report must include the code, accuracy figures and explanation as specified in the script.

2. Task 2: Implementing a hierarchical utterance+DA-context-based tagger [20 marks].
   Your report must include the code, accuracy figures and explanation as specified in the script.

3. Task 3: EITHER (following the lab script) Using a pre-trained BERT model to see if it can help improve DA tagging performance, OR (as in original assignment specification) adding a decoder to make a seq2seq dialogue system [10 marks]. If you don't have time for this, remember it's only 10 marks out of 50! You can do very well on this assignment without tackling this part. If you do have time, though, these are useful techniques to learn. *Your report should include the code and some explanation of what you found out, with some examples of output.*

Submission

Please submit one zip file with all your answers to Parts A and B together.

As well as code, you should include text explanations, descriptions and answers to specific questions as necessary and as specified above. Code should be in Python; explanatory text can be either as a separate report in PDF format (not Word, please), or included together with the code as a Jupyter/Colab notebook.

For each section, marks will be awarded for correctness of code and classifier performance, but also for clarity of explanations and justifications.