

CS 4650/7650 Fall 2020: Homework 5

November 3, 2021

Instructions

1. This homework has two parts: Q1 is a theory question, and Q2 is a programming assignment with some written components within a Jupyter Notebook.
We will be using Gradescope to collect your assignments. Please read the following instructions for submitting to Gradescope carefully!
 - (a) Each subproblem must be submitted on a separate page. When submitting to Gradescope (under **HW5 Writing**), make sure to mark which page(s) correspond to each problem or subproblem. For instance, Q1 has four subproblems, so the solution to each must start on a new page.
 - (b) For the coding problem (Q5), please upload 'MyQA.py' and 'HW5.ipynb' under the **HW5 programming** assignment on Gradescope. Write your solutions for Q5 (a) and (b) in your writeup, and Attach a pdf export of 'HW5.ipynb', including outputs, to your writeup.
 - (c) Note: This is a large class and Gradescope's assignment segmentation features are essential. Failure to follow these instructions may result in parts of your assignment not being graded. We will not entertain regrading requests for failure to follow instructions.
2. \LaTeX solutions are strongly encouraged (a solution template is available on the class website), but scanned handwritten copies are also acceptable. Hard copies are not accepted.
3. We generally encourage collaboration with other students. You may discuss the questions and potential directions for solving them with another student. However, you need to write your own solutions and code separately, and not as a group activity. Please list the students you collaborated with on the submission site.

Questions

1. (a) The **BiLingual Evaluation Understudy (BLEU)** is a method for automatic evaluation of machine translation that is quick, inexpensive and language-independent, and correlates highly with human evaluation.

The formal definition of BLEU is

$$\text{BLEU} = \text{BP} \cdot \exp\left(\sum_{n=1}^N w_n \log p_n\right),$$

where

$$\text{BP} = \begin{cases} 1 & \text{if } c > r, \\ \exp(1 - r/c) & \text{otherwise.} \end{cases}$$

p_n represents the n -gram precision, associated with its positive weight w_n .

c is the length of the candidate translation and r is the length of the reference translation.

Suppose there are a reference translation and a candidate translation:

- i. reference: “Tom likes to study natural language processing at night”;
- ii. candidate: “Tom loves to study language processing at night”.

Assume $N = 4$ and $w_1 = w_2 = w_3 = w_4 = \frac{1}{4}$.

Compute p_1, p_2, p_3, p_4 , and then compute BP and BLEU. [6 points]

- (b) When humans evaluate translations, they consider many aspects, including adequacy, fidelity, and fluency.
- **Adequacy:** Translated text should contain the same linguistic contents as the source text.
 - **Fidelity:** Translated text should carry the same meaning as the source text.
 - **Fluency:** Translated text should be grammatical/natural/fluent in the target language.

Automated evaluation metrics typically try to cover some of those criteria. Which of these three criteria does the BLEU metric consider? And how does this metric account for them?

- (c) If the translation model produces shorter translations, then it is easy to achieve high n -gram precision scores. How does BLEU solve this issue?
- (d) What are the weaknesses of the BLEU metric? (List and explain two).

2. In this assignment, you will implement a simple Seq2Seq model for machine translation. You will train an encoder-decoder neural network to translate sentences from French to English. You will work with three files: the main notebook **HW5_MT.ipynb**, **dataset.py**, and **model.py**. Please start with **HW5_MT.ipynb** and follow instructions there.

All three files should be submitted to the Gradescope assignment **HW5 Coding**. For your write-up, please answer any questions in the following parts and attach a PDF for **HW5_MT.ipynb** at the end of your write-up.

The assignment zip can be downloaded here:

https://www.cc.gatech.edu/classes/AY2022/cs4650_fall/programming/h5_mt.zip

- (a) Implement `dataloader` in 'dataset.py' to retrieve data pairs for French and English sentences. No write-up is required for this part. [2 points]
- (b) Implement `EncoderRNN` and `DecoderRNN` in 'models.py'. No writeup required. [10 points]
- (c) Implement `train` and `evaluate` functions in the notebook and train the model. **In your write-up**, include plots (loss and bleu scores) and final performance numbers as well as ten sample sentences generated (obtained by running `evaluateRandomly` in the notebook. [15 points implementation + 5 points write-up]
- (d) **Where** is a potential representational bottleneck of the seq2seq architecture you implemented? **Explain. What** are some alternatives to overcome such a bottleneck? [3 points]
- (e) **[BONUS]** Implement `AttnDecoderRNN` and its corresponding `train` and `evaluate` functions in the notebook. [15 points]
Train the attention model. Include plots, final performance numbers and ten sample sentences generated by this model **in your write-up** [5 points].