Assignment 1

February 13, 2024

1 BLEU Score[Papineni et al., 2002]

Measuring similarity between texts is important in NLP because it helps us to evaluate quality of machine translation. One of the most common method for this is to calculate the BLEU Score between the texts. In this question, we'll explore the BLEU Score in a simplified setting.

Let x and y be two sentences that we want to compare. Then we define the *Modified n-gram precision* as

$$p_n = \frac{\sum_{n\text{-}gram \in x \cap y} min(count_x(n\text{-}gram), count_y(n\text{-}gram))}{\sum_{n\text{-}gram \in x} count_x(n\text{-}gram)} \tag{1}$$

Note that here y is akin to the ground truth and x is akin to machine translated text. Now we define the BLEU Score as

$$BLEU = BP \cdot \exp \sum_{n=1}^{N} w_n \log(p_n) \tag{2}$$

where $N = 4$, $w_n = \frac{1}{N}$ and BP is called the Brevity Penalty. For our purposes, we'll take $BP = 1$.

1. Implement BLEU Score metric. Pre-process the text by lower-casing the text and removing punctuation.(3 marks)
2. Use this implementation to find BLEU Score when x = "The boys were playing happily on the ground." and y = "The boys were playing football on the field.". (2 mark)
3. Can you explain why we are taking minimum in numerator in equation 1? (2 marks) Hint: You may refer the paper.
4. Use your implementation to find BLEU Score between any 5 sentence pairs and explain what are potential disadvantages of using the BLEU Score. (1+2 marks)

References

[Papineni et al., 2002] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. *ACL Anthology*.