Neural Network Language Generation (CL3.407)

Spring 2023, IIIT Hyderabad
Assignment 1

Abhinav S Menon
(2020114001)

Part 1

While reading a research paper, I usually make use of the following procedure. First, I make a preliminary pass over the paper. During this pass, I focus on the abstract and the major part of the introduction, and partially on the conclusion; some parts of the introduction may be omitted if they are not immediately relevant. The aim at this stage is to get a high-level view of where the paper stands – which field and subfield is it in? does it aim to solve a specific problem? if so, which? if not, what are the directions in which it goes? is it descriptive or theoretic? and so on.

Next, I use the second pass of the paper to refine my idea of where the paper is located, and understand more of the details. At this point, I carefully look at the intervening sections, like the methodology, theoretical results/proofs, experimental evidence, and so on.

Finally, I look for gaps in my understanding and find further resources to fill myself in. These may be cited or related works, or videos or blog posts connected to the paper itself, or textbooks on similar topics. If necessary, at this point, I make a third pass and repeat this step.

Part 2

Consider the paper *Get to the Point – Summarisation with Pointer-Generator Networks* (See et al., 2017). The following criticisms can be made:

- The method is described sometimes as abstractive, and sometimes as partially abstractive and partially extractive. This is inconsistent.
- Using min to find the coverage loss is ill-motivated. It may well be necessary for the model to attend to the same location more than once. This is not countered.

- No justification is given for the choice to not use pre-trained embeddings, like Nallapati et al. (2016).
- The claim that "in the early phase of training, the coverage objective interfered with the main objective" is not substantiated with any data, and not presented as a hypothesis. It is only empirically observed that adding coverage loss after some epochs helps.
- Potential future directions are not well-summarised, but scattered throughout the paper.
- There is a lack of examples – especially of error cases. Although the supplementary materials include them, some more should have accompanied the claims in the main paper as well.
- The ROUGE metric rewards truncated summaries. This issue is not addressed.