Improvement on Knowledge-backed Generation Model Using Post-Modifier Dataset (PoMo)

Priyanka Nath / SBU ID: 112715634
Shanuj Shekhar / SBU ID: 112670784
Trisha Kanji / SBU ID: 112686934

Abstract

Post-modifier is a short descriptive phrase that comes after a word (usually a noun) in a sentence which gives more detailed contextual information about that word. Post-modifier generation is a contextual data-to-text generation problem. A good post-modifier generation model must select relevant facts about the word, given the text and the output post-modifier should cover these facts in such a way that it maintains the context with respect to the original text.

1 Introduction

The main objective is to improve the already existing knowledge-backed generation model which uses the PoMo data-set[1]. One can always look up Wikipedia for information concerning a particular topic but it is always convenient to get relevant and concise information instead of being bombarded with tons of irrelevant information. It is an interesting problem to train a model that understands the context pertaining to .

The paper concludes that given relevant claims, sequence-to-sequence models can generate good post-modifiers but the performance dwindles if the models have to do claim selection by themselves.

It is mentioned that F1 score for most-common baseline model is lesser than neural-baseline model for claim selection.

We will try to improve claim selection and add third party facts to improve PoMo generation.

We intend to use accuracy metrics like precision, recall, F1 score etc. to evaluate the performance of our improved PoMo Generation model.

2 Current Progress

Currently, we are in discussion with Jun S. Kang who worked on the paper "PoMo: Generating Entity-Specific Post-Modifiers in Context"

3 Expected Results

We expect to improve the performance of the existing model. We are yet to quantify our expectations.

4 Three Questions we have

Should we look into modifying PoMo dataset or should we use the PoMo dataset as it is and improve the model?

Is the existing model doing claim selection or is a set of pre-extracted claims given as an input to the model?

We want to extract facts from an external source other than wiki data graphs to improve performance. News articles seem like a reliable and somewhat accurate source because Post Modifiers are extensively used in them. What other textual sources should we consider?

5 References

[1] Kang, Jun Seok, I. V. Logan, L. Robert, Zewei Chu, Yang Chen, Dheeru Dua, Kevin Gimpel, Sameer Singh, and Niranjan Balasubramanian. "PoMo: Generating Entity-Specific Post-Modifiers in Context." arXiv preprint arXiv:1904.03111 (2019)