

Review of Thesis Proposal

Faculty of Mathematics and Physics, Charles University

Reviewer David Mareček
ÚFAL MFF UK

Candidate Sourabrata Mukherjee
Thesis Title Text Style Transfer
Supervisor Mgr. et Mgr. Ondřej Dušek, Ph.D.
ÚFAL MFF UK

Topic

The thesis proposal by Sourabrata Mukherjee focuses on a very important and up-to-date problem of controlling and transferring the style of automatically generated texts. It turns out that even the evaluation of this task is very complicated, which makes the topic very challenging.

Content

The proposal is clearly written and well structured. The experiments described and proposed are presented in a reasonably self-contained manner, allowing for easy comprehension. Since I am not an expert in Natural Language Generation, I cannot evaluate the completeness of the related work, however, the detailed analysis confirms Sourabrata's expertise in the research topic.

After the problem's introduction, the main challenges are outlined, including the lack of usable parallel data and the absence of standardized evaluation metrics. The key issue here is the distinction between content and style. While the content of the text must be preserved, its style is to be changed.

The third section contains Sourabrata's work done so far. He describes two experiments with novel strategies: The Sentiment Transfer model and the Politeness transfer model, which is applied to a Chatbot. Both the models were evaluated as competitive or even better (in some aspects) than previously existing approaches. However, since there are many evaluation aspects, it is hard to decide whether a system with a significantly better style but a bit worse in content preservation is really better.

The fourth section proposes future experiments. The first one is building a multilingual biased-to-neutral 5k sentence pairs with the aim of fixing the biased text containing offensive, prejudiced or hurtful words, which I find very useful and without any higher risks. The second one is evaluating the content preservation: to select the style-related words and run the evaluation on the rest, and, besides the word-based metrics as BLUE, to compare also some kind of meaning representations

of the source and target sentences. This I find much harder with more potential risks (see the comments).

Questions and Comments

1. In your sentiment transfer experiment: What is the main motivation for using German as a transfer language? Would it be possible to do it in English only without the translation to/from German?
2. Have you tried ChatGPT or GPT-3 model with a few-shot setting? What would be their answers for style transfer tasks? Are they comparable to your systems?
3. I find the Positive \rightarrow Negative style change very debatable. In your example “The food is tasteless \rightarrow The food is delicious” I think it has really a different (opposite) meaning instead of a style change. Here I would say “The food could be much better” to preserve the meaning and say it much more positively at the same time.
4. The previous example would be also a problematic case for decoupling the content from the style. The words “tasteless” and “delicious” are both the main content and style words here.
5. Have you directly compared your MaskBLEU and MaskSim metrics with the human evaluation of content preservation? Would it be possible on the level of sentences?
6. From the automatic and human evaluations, we still do not know what should be the optimal trade-off between style, content, and fluency. We would like to have only one evaluation score. Do you think it would be useful to make an application-based human evaluation (e.g. the Polite Chatbot) with only one general score?

Conclusions

I find the thesis proposal of sufficient quality. The plan is clear, well-structured, and sufficiently extensive. The goals are achievable in the next couple of years. I suggest focusing on the evaluation because that seems to be the main challenge and without any clear metric, the systems could be never easily compared.

Prague, 15. 3. 2023

Signature: