|  |
| --- |
| Review #1 |

**What is this paper about and what contributions does it make?**

This paper describes a two-phase model for the task of rewriting to improve the context understanding of dialogue models. The main contribution of the paper is that the two-phase approach improves the current SOTA. Specifically, The approach consists of locating positions by unsupervised and supervised methods and filling blanks by LMs.

**Reasons to accept**

The paper is well organized to identify the research problem and an effective approach to tackle the problem. The approach is well-described and experimented with multiple datasets to prove its validity. Results are shown to be the state of the art compared to the previous research.

**Reasons to reject**

The two-phase approach is effective. However, as the author mentioned in the abstract, this task is a step towards the understanding of dialogues by LM models. The rule-based method in the first phase takes risks of affecting the understanding task if there are mistakes or missing rules in the step as mentioned in the paper.

|  |  |
| --- | --- |
| **Soundness**: | 3 |
| **Excitement (Long paper)**: | 3.5 |
| **Reviewer Confidence**: | 4 |
| **Recommendation for Best Paper Award**: | No |
| **Reproducibility**: | 5 |
| **Ethical Concerns**: | No |

|  |
| --- |
| Review #2 |

**What is this paper about and what contributions does it make?**

The authors argued that rewriting incomplete and ambiguous utterances can help the dialogue model generate better results. They proposed a two-phase rewriting framework to locate the empty slots and filled in those positions using T5 or BART. Their contributions include proposing a two-phase framework for rewriting incomplete utterances, aligning the two sentences before and after rewriting based on the longest common subsequences, and carrying out experiments on 5 datasets (MuDoCo, CQR, REWRITE, RES, and CANARD).

**Reasons to accept**

Extensive experiments have been conducted on MuDoCo, CQR, REWRITE, RES, and CANARD, though their evaluation methods are biased.

**Reasons to reject**

In the context of ChatGPT, it is doubted that if the rewriting of incomplete utterances is a valid problem or not. Rewriting those utterances using T5 or BART cannot guarantee not to distort the original utterance. First, it is not convincing if rewriting is a valid problem to solve. Second, they did not compare with ChatGPT or GPT-3 in the experiments. Their automatic metrics and human evaluation cannot demonstrate that their approach can help the dialogue model generate better results.

|  |  |
| --- | --- |
| **Soundness**: | 2 |
| **Excitement (Long paper)**: | 1.5 |
| **Reviewer Confidence**: | 4 |
| **Recommendation for Best Paper Award**: | No |
| **Reproducibility**: | 3 |
| **Ethical Concerns**: | No |

|  |
| --- |
| Review #3 |

**What is this paper about and what contributions does it make?**

This work proposes a two-stage method for incomplete utterance rewriting: the first stage points out which tokens need to be replaced, and the second stage just aims to replace those tokens with the corresponding tokens in the context for resolving co-references and ellipses.

**Reasons to accept**

1. The results are quite comprehensive, which has used 5 datasets in two languages for benchmark. And the baselines have included the latest SOTA models. Meanwhile, the improvements also look significant.

**Reasons to reject**

1. The method is limited in novelty. The two-stage framework for utterance rewriting has been proposed by many previous works several years ago. And the first stage is almost the same as previous methods: first using LCR to get those co-reference and ellipses out and then fine-tuned a NER model on it. And the second stage has something new: it uses a pre-trained seq-2-seq mdoel while previous methods mainly used LSTM or pointer network. And instead of using the T5 model directly, it also proposes to divide the sentence with multiple replacements into multiple segments. So only this part is novel.

**Questions for the Author(s)**

Question A: In line 295, I think you used T5-small and BART-Base.

Question B:

Question C:

**Missing References**

1. Di Jin, Sijia Liu, Yang Liu, and Dilek Hakkani-Tur. 2022. Improving Bot Response Contradiction Detection via Utterance Rewriting. In Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue, pages 605–614, Edinburgh, UK. Association for Computational Linguistics.

|  |  |
| --- | --- |
| **Soundness**: | 3 |
| **Excitement (Long paper)**: | 3.5 |
| **Reviewer Confidence**: | 4 |
| **Recommendation for Best Paper Award**: | No |
| **Reproducibility**: | 3 |
| **Ethical Concerns**: | No |