The 61st Annual Meeting of the Association for Computational Linguistics

ACL 2023

Author Response

Title: Incomplete Utterance Rewriting by A Two-Phase Locate-and-Fill Regime
Authors: Zitong Li, Jiawei Li, Haifeng Tang, Kenny Q. Zhu and Ruolan Yang

Instructions

The author response period has begun. The reviews for your submission are displayed on this page, and you can respond to the points raised in the reviews in the boxes provided below.

The main purpose of author response is to help the reviewers and meta-reviewer to identify and resolve any significant misunderstandings of the submission. If you see any, please respond to correct the record and say what would be clarified and how if the paper is accepted. You may **not** include new results in the response, or link to any new materials. The response should follow the same anonymity requirements as the submission, i.e. the identity of the authors should not be disclosed (including through non-anonymized URLs).

NEW You will notice that at ACL'23 the "overall recommendation" has been split into two scores: soundness and excitement. The former is more objective and is a strict pre-requisite for acceptance. Thus, we recommend focusing the response on correcting inaccuracies or misunderstandings that could change the soundness rating. If a reviewer spotted any unclarity, framing issues, etc -- please be explicit about how you would address the issue in camera-ready if the paper is accepted. The reviews may contain questions to the authors. We asked the reviewers to use the letter numbering, which you can use to refer to the questions in your response (e.g., Question 2C is Question C from Reviewer 2).

While you are not obligated to respond, the reviewers are volunteer researchers and it is a professional courtesy to at least thank them for their time. Even if you disagree with their position, and neither of you manages to convince the other, both sides still learn something from the exchange about the norms of the field and the current state of knowledge about a contentious topic.

Unfortunately, despite our best efforts, a few reviews are still missing. We apologize and ask for your understanding. These papers have been assigned emergency reviewers, but we decided to release existing reviews on time to let the authors start working. If some of your reviews are missing, please check back later, hopefully they will arrive soon. Please use the "response to chairs" box to flag missing and late reviews, so that the meta-reviewers can take this into account.

Conversely, some authors may have received more than three reviews because of emergency over-assignments. While more feedback is generally a good thing, such authors may feel stressed to respond to more reviews within the same space limit. In such cases, please focus on the most serious issues, and the meta-reviewers will be asked to take the space limitations into consideration. NEW The "response to chairs" mechanism should also be used in the rare cases when the reviewers have not done due diligence and the reviews are of unacceptably low quality, violating the conference review policy. A new element at ACL'23 is that we ask you to specify both the type of issue and the rationale for flagging it. Please see the instructions below.

Last but not least: if a paper is flagged for ethical issues, it means that it may (but not necessarily will) be sent for a separate ethics review after the author response period. The chairs might also recommend the ethics review, irrespective of what you see in regular reviews. Please see EACL ethics FAQ, our process is similar.

For reference, you may see the review form that reviewers used to evaluate your submission. If you do not see some of the filled-in fields in the reviews below, it means that they were intended to be seen only by the committee. See the review form HERE.

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

Submit Response to Reviewers

Use the following boxes to enter your response to the reviews. Please limit the total amount of words in your comments to 900 words (longer responses will not be accepted by the system).

Response to Review #1:

Response to Review #2:

Response to Review #3:

General Response to Reviewers:

Response to Chairs

In rare cases reviews may be of unacceptably low quality, which violates the conference peer review policy. If this happened to you, you can use the box below to report the type of the issue and explain your rationale to the chairs. This mechanism should only be used for serious issues. It is not in the authors' interest to make their meta-reviewers investigate cases where the authors disagree with the reviewers, but the reviewers have done due diligence and provide their arguments/evidence/references.

The following types of issues are known from past conferences:

A. The review is not specific enough, e.g. missing references are not specified
B. The review exhibits one of the heuristics discussed in the ACL23 review policy blog post, such as "not novel", "not surprising", "too simple", "not SOTA". Note that these criticisms may be legitimate, if the reviewer explains their reasoning, and backs up the criticism with arguments/evidence/references. Please flag only the cases where you believe that the reviewer has not done due diligence.
C. The scores do not match the review text. Note that in ACL23, the "soundness" score is meant to reflect the technical merit of the submission, and low soundness should be backed up with serious objections to the work. The "excitement" score is more subjective, and its justification may not be reflected in the text.
D. The review is rude/unprofessional
E. The reviewer does not seem to have the expertise to review this paper
F. The review does not match the paper type (e.g. short paper expected to produce more experiments than is necessary to support the stated claim)

G. The review does not match the type of contribution (e.g. experimental work expected of a paper stating a different kind of contribution)
H. The review is missing or too short and uninformative
I. The review was late and could not be addressed in the author response
J. Other (please explain)

If you feel that you have such a problem, please use the following format to report it in the text box below (without the #comment lines, 250 words max). In this example, Reviewer 1 had issue A (unspecific review) and Reviewer 2 had issues C and D (rude review, scores don't match the text).

```
# review problem type(s), as a capital letter corresponding to the issue type in the above list of possible issues. If there is more than one, list them comma-separated (e.g. A, I)
R1: A

# explanation
R1 states [reviewer statement], which we believe corresponds to the review issue type A. It is unreasonable in this case because [rationale].

R2: C,D

R2 states [reviewer statement]...
```

Submit

START Conference Manager (V2.61.0 - Rev. 6823)