
Your ACL 2023 Submission (Number 1887)

发件人 : ACL 2023 Program Chairs
<papers_acl2023@softconf.com>

2023年5月02日, 周二, 下午 10:16

发件人 : start@z.softconf.com

主题 : Your ACL 2023 Submission (Number 1887)

收件人 : blmoistawinde@sjtu.edu.cn

回复 : : papers_acl2023 <papers_acl2023@softconf.com>

Dear Zhiling Zhang:

On behalf of the ACL 2023 Program Committee, we are sorry to inform you that the following submission was not selected by the program committee to appear at ACL 2023: Semantic Space Grounded Weighted Decoding for Multi-Attribute Controllable Dialog Generation

All rejection notifications always say that the process was very competitive, and some authors—especially the more junior authors—may interpret us saying that as a formality. Let us assure you that this is actually true with respect to ACL. The final decisions were based on a range of factors, including the editorial priorities of the area chairs, clarity of presentation, reviewer confidence (taking into account any author-flagged review issues). We hope that the perspective of ACL reviewers and meta-reviewers will still be helpful revising your work.

The good news is that the experiment with excitement/soundness scores this year allowed us to collect more detailed statistics of the distribution of these scores, which might provide a clearer picture of the possible future acceptance rates for Findings, to be considered by the community and the future program chairs. We will make our analysis public, stay tuned. We hope that you can still attend the conference. If you plan to attend ACL on-site, but require a visa to visit Canada, and have not obtained it yet - please have a look at this post: <https://2023.aclweb.org/blog/visa-info/>

We have enclosed the reviewer and meta-reviewer comments for your perusal.

Best Regards,
ACL 2023 Program Chairs
ACL 2023

=====
=====
ACL 2023 Reviews for Submission #1887
=====

Title: Semantic Space Grounded Weighted Decoding for Multi-Attribute Controllable Dialog Generation

Authors: Zhiling Zhang, Mengyue Wu and Kenny Q. Zhu

=====
=====
META-REVIEW
=====

Comments: This paper studies multi-attribute controllable dialogue generation. The author proposes to use semantic space similarity for decoding, and the experimental results show that the model can perform better than the baseline model.
This paper's writing and presentation are far from satisfactory and require significant modification before publication.

=====

=====

REVIEWER #1

=====

=====

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

The author proposes to use semantic space similarity to decode for multi-attribute controlled tasks to generate controllable text. Experimental results show that the model can achieve better performance than the baseline model. The author also shows some potential applications.

Reasons to accept

-
- (1)The motivation of this paper is interesting and clear. Because there are many multi-attribute control scenarios in real applications.
 - (2)The writing of background Weighted Decoding for Controllable Generation is clear and reasonable.
-

Reasons to reject

-
- (1) The writing of their model Dialogue Attribute Space Controller is not clear. In line 267, how to obtain h_n^k for a_k ? I suggest the author use standard mathematical formula to explain.
 - (2) The loss function of classification is also clear. In line 307, we didn't know how to train these two important loss functions.
-

Questions for the Author(s)

-
- (1) The writing of their model Dialogue Attribute Space Controller is not clear. In line 267, how to obtain h_n^k for a_k ? I suggest the author use standard mathematical formula to explain.
 - (2) The loss function of classification is also clear. In line 307, we didn't know how to train these two important loss functions.
-

Reviewer's Scores

Soundness: 4

Excitement (Long paper): 2.5
Reviewer Confidence: 2
Recommendation for Best Paper Award: No
Reproducibility: 2
Ethical Concerns: No

=====

=====

REVIEWER #2

=====

=====

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

This paper proposes a novel controllable generation framework called DASC possesses strong controllability with the weighted decoding paradigm while improving generation quality with the grounding in an attribute semantics space. The main contribution of this paper is to propose a multi-attribute controllable generation framework that achieves high control accuracy as well as high response quality. The experimental result shows that DASC outperforms strong baselines such as Director with a large margin

Reasons to accept

- 1) The idea is novel, makes sense, and effective. To the best of my knowledge, this is the first time that learning a latent attribute space is the guidance of controllability. Obviously, this new paradigm is more smoothy than previous classifier-based controllable generation method
 - 2) The experimental result, especially the robustness, shows a very good balance between controllability and response quality. As I know, all controllable text generation model fails to generate stylistic and high-quality response simultaneously. The experimental result in table 3 shows that the authors did it via the semantic space.
 - 3) The paper is easy to follow
-

Reasons to reject

- 1) The experimental setting is unclear. Which backbone model did you use in the experiments? Since some of your datasets are in Chinese, did you use a multi-lingual BART?
 - 2) The Human evaluation is weak. Where did you find the annotators? Why would they volunteer for your experiment? How do you ensure they're eligible for this judgment?
-

Questions for the Author(s)

All questions are listed in the section 'reasons to reject'

Reviewer's Scores

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

=====

=====

REVIEWER #3

=====

=====

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

This paper proposes a new framework for controlled response generation based on multiple attributes (e.g., emotions, gender style). Comparison results with certain related previous works are given and various analyses are provided.

The main contribution of this paper is that it extends the previous work that mainly considers one attribute to a framework that can handle multiple attributes in controlled generation. The framework also introduces additional attribute-sensitive embedding mapping in order to reduce computational load and improve performance.

Reasons to accept

The framework proposed is new and the analyses have been conducted to analyze various aspects, including extensibility, of the proposed algorithm.

Reasons to reject

The paper is problematic in multiple aspects, listed below. It should be significantly improved before publication.

1. The paper has poor clarity with a lot of grammar/format errors (e.g., a^k should be a_k in Formula 3/4/6), and is not well organized. For example, the introduction of the related previous approach Director in Section 3.2 should be moved to Section 2, while Section 3 should focus on the proposed methods only. Another example is that Figure 2 should be moved to a earlier position in Section3, making the explanations of the framework easier to understand for the readers.
2. The reported improvements (e.g., as in Table 1/2/3) are unconvincing. BScore achieved by the proposed method is worse than the CTRL baseline and Bscore is not reported in Table 3.
3. The proposed method fails to outperform baseline(s) on Sensible in both Table 2 and Table 3. The ACC_x results are evaluated based on classifier-generated labels instead of ground-truth labels. And it is unclear whether the ACC_E reported in Table 3 is automatic or

human evaluated.

In addition, regarding the human evaluation, only 2 volunteers are involved in labeling 100 contexts. It is unclear how the 100 contexts were selected and what procedure has been taken to ensure the quality of human evaluation.

Questions for the Author(s)

Question A: Were the 100 contexts used in human evaluation (see Lin 397) randomly selected or selected with certain strategy?

Question B: Is there any procedure adopted to ensure the evaluation quality of the human evaluation?

Question C: What are the BScore results for the comparison experiments shown in Table 3?

Missing References

More references should be added for the weighted decoding models mentioned in Line 206.

Typos, Grammar, Style, and Presentation Improvements

As mentioned in "Reasons to reject" section above, the paper need to be thoroughly modified to improve clarity and content organization. There are also a lot of spelling/grammar/format errors (e.g., "state-off-the-art" in Line013) in the paper. Please thoroughly proofread the paper to correct all the writing errors.

Reviewer's Scores

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

--

ACL 2023 - <https://softconf.com/acl2023/papers>
