Summary of changes

In this revision, we have responded to one of the core flaws: lack of recent baselines. We have added 3 new baselines: SyncToD(2024), InstructTods(2023), QTOD(2022).

We performed a neuron activation analysis to explore how the activation patterns of a model changes with the schema augmentation mechanism. This analysis gives more insight in how models learn to generalize across layers in a model. We show that augmentation enables models to recognize semantic similarities across domains in lower layers while maintaining domain-specific distinctions in higher layers.

We have organized the paper around our core ideas, which are no dependency on turn annotations, evaluating API calls, schema augmentation mechanism to improve task completion performance and neuron analysis of how schema augmentation allows better out of domain generalization.

Response to previous reviews

The meta review asked for the following baselines: SyncTOD (Saley et al., 2024), DialoKG (Rony et al., 2022), GraphMemDialog (Wu et al., 2021a), COMET (Gou et al., 2021), MAKER (Wan et al., 2023), and CDNet (Raghu et al., 2021).

Of these we have implemented SyncTOD as it is related to our approach. The other approaches are very different from what we are doing. The other baselines mainly build knowledge graphs, and model external knowledge. We are focusing on out-of-domain generalization of ToD systems, so these baselines are not relevant. We have added 3 recent baselines that are more related to our work for comparison.

Review 1

The rationale behind the selection of the baselines is not clear. SOLOIST and SimpleToD are relatively older works. AutoTOD, although recently published, is used out of the box in a zero-shot setup. Comparison with more recent baselines like [1] and [2] would help to understand the gap between the proposed method and the existing works. Please note that [1] also proposes a method that does not depend upon the annotations and hence shares a similar motivation as ZeroToD.

[1] King, Brendan and Jeffrey Flanigan. “Unsupervised End-to-End Task-Oriented Dialogue with LLMs: The Power of the Noisy Channel.” Conference on Empirical Methods in Natural Language Processing (2024). [2] Du, Huifang et al. “Rewarding What Matters: Step-by-Step Reinforcement Learning for Task-Oriented Dialogue.” Conference on Empirical Methods in Natural Language Processing (2024).

In this revision, we have added 3 recent baselines that would help to understand the gap between our method and the baselines.

Of the methods suggested, [1] lacks a lot of implementation details and so it is difficult to reproduce and [2] has not provided any source code.

Results on MultiWoz

Multiwoz is indeed one of the popular datasets in supervised TOD settings. Nonetheless, the focus of our work is on out-of-domain generalization, for which Multiwoz dataset is not well-suited. The multi-domain nature of dialog in this dataset makes it impossible to split dialogs into seen/unseen sets. Additionally, Multiwoz dataset **does not provide API calls**, which we use as a proxy for task completion, i.e., reserving a restaurant via an API call including method name, parameters, and corresponding values. For these reasons, we have decided not to include Multiwoz. We want to highlight that one of our datasets, SGD, is the most comprehensive one available in the public domain (with 20 domains) with clear seen/unseen domains for experimentation.

Reviewer 2

However, I still hold some reservations regarding the originality and overall excitement of the contribution, as the core ideas build upon existing work and the methodological novelty is relatively incremental.

We would like to mention that a core part of our methodology is data augmentation and how it assists in out of domain generalization. Additionally, in this work we have also performed neuron activation analysis to explore how the activation patterns of a model changes with augmentation. This analysis gives more insight in how models learn to generalize across layers in a model.

Reviewer 3

We have already provided a response to the questions that were raised by this reviewer. Unfortunately, we did not get any response to our comments.