Zero Shot Task Oriented Dialog

Anonymous ACL submission

Abstract

This document is a supplement to the general instructions for *ACL authors. It contains instructions for using the LaTeX style file for ACL 2023. The document itself conforms to its own specifications, and is, therefore, an example of what your manuscript should look like. These instructions should be used both for papers submitted for review and for final versions of accepted papers.

1 Introduction

2 Methodology

Steps: Pretraining(CE Loss), Training (loss on target only)

Context: Dialog History like SimpleTOD DST instead of dialog history

Additional info Schema List of system actions User actions Service Results

Contrastive Loss User action, System action System Action, NLG

3 Experimental Setup

3.1 Datasets

We use the SGD (Rastogi et al., 2020) and SGD-X (Lee et al., 2022) datasets to conduct experiments. Seen domains are obtained from the train schema file and unseen domains from the eval and test schema files. The models are trained on seen domains and inference is performed on all, seen and unseen domains.

To evaluate the performance of the models, we use metrics from SGD, MultiWoz 2.0 benchmark (Ramadan et al., 2018) and introduce 2 novel metrics: Average Action Accuracy, Joint Action Accuracy. These new metrics are similar to the goal metrics in SGD, but are performed on the actions. For response generation, we report the ROUGE-2 (Lin and Och, 2004) score and GLEU (Wu et al., 2016) instead of BLEU as it performs better on individual sentence pairs. The combined score is calculated as suggested in (Mehri et al., 2019) with (Inform + Success) $\times$ 0.5 + GLEU.

4 Related Works

4.1 Supervised End to End Models

Pretrained language models like BERT (Devlin et al., 2019), GPT-2 (Radford et al., 2019) and T5 (Raffel et al., 2019) have been used extensively in the literature for End to End models for TOD systems (Hosseini-Asl et al., 2020), (Peng et al., 2021),(Lee et al., 2020), (Yang et al., 2020), (Jeon and Lee, 2021), (Sun et al., 2022), (Yang et al., 2022), (Noroozi et al., 2020). In these models, the context consists of user and system utterance, whereas in our model we use the last user utterance and the previous state DST as context. Moreover, most of these models have the best performance in supervised settings and do not have the primary focus on zero-shot generalization.

4.2 Zero Shot End to End Models

Zero Shot DST models (Feng et al., 2020), (Zhao et al., 2022) incorporate schema as part of the context and generalize well for DST, however these models do not focus on system actions and response generation.

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv*, abs/1810.04805.

Yue Feng, Yang Wang, and Hang Li. 2020. A sequence-to-sequence approach to dialogue state tracking. In *Annual Meeting of the Association for Computational Linguistics*.

Ehsan Hosseini-Asl, Bryan McCann, Chien-Sheng Wu, Semih Yavuz, and Richard Socher. 2020. A simple language model for task-oriented dialogue. *ArXiv*, abs/2005.00796.

Hyunmin Jeon and Gary Geunbae Lee. 2021. Dora: Toward policy optimization for task-oriented dialogue system with efficient context. *Comput. Speech Lang.*, 72:101310.

Harrison Lee, Raghav Gupta, Abhinav Rastogi, Yuan Cao, Bin Zhang, and Yonghui Wu. 2022. Sgd-x: A benchmark for robust generalization in schema-guided dialogue systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 10938–10946.

Hwaran Lee, Seokhwan Jo, Hyungjun Kim, Sangkeun Jung, and Tae-Yoon Kim. 2020. Sumbt+larl: End-to-end neural task-oriented dialog system with reinforcement learning. *ArXiv*, abs/2009.10447.

Chin-Yew Lin and FJ Och. 2004. Looking for a few good metrics: Rouge and its evaluation. In *Ntcir workshop*.

Shikib Mehri, Tejas Srinivasan, and Maxine Eskenazi. 2019. Structured fusion networks for dialog. *arXiv preprint arXiv:1907.10016*.

Vahid Noroozi, Yang Zhang, Evelina Bakhturina, and Tomasz Kornuta. 2020. A fast and robust bert-based dialogue state tracker for schema guided dialogue dataset. *ArXiv*, abs/2008.12335.

Baolin Peng, Chunyuan Li, Jinchao Li, Shahin Shayandeh, Lars Lidén, and Jianfeng Gao. 2021. Soloist: Building task bots at scale with transfer learning and machine teaching. *Transactions of the Association for Computational Linguistics*, 9:807–824.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Colin Raffel, Noam M. Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *ArXiv*, abs/1910.10683.

Osman Ramadan, Paweł Budzianowski, and Milica Gasic. 2018. Large-scale multi-domain belief tracking with knowledge sharing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, volume 2, pages 432–437.

Abhinav Rastogi, Xiaoxue Zang, Srinivas Sunkara, Raghav Gupta, and Pranav Khaitan. 2020. Towards scalable multi-domain conversational agents: The schema-guided dialogue dataset. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8689–8696.

Haipeng Sun, Junwei Bao, Youzheng Wu, and Xiaodong He. 2022. Bort: Back and denoising reconstruction for end-to-end task-oriented dialog. In *NAACL-HLT*.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation.

Yunyi Yang, Hong Ding, Qing Liu, and Xiaojun Quan. 2022. Ubarv2: Towards mitigating exposure bias in task-oriented dialogs. *ArXiv*, abs/2209.07239.

Yunyi Yang, Yunhao Li, and Xiaojun Quan. 2020. Ubar: Towards fully end-to-end task-oriented dialog systems with gpt-2. In *AAAI Conference on Artificial Intelligence*.

Jeffrey Zhao, Raghav Gupta, Yuanbin Cao, Dian Yu, Mingqiu Wang, Harrison Lee, Abhinav Rastogi, Izhak Shafran, and Yonghui Wu. 2022. Description-driven task-oriented dialog modeling. *ArXiv*, abs/2201.08904.

A Example Appendix

This is a section in the appendix.