Schema Guided Zero-Shot End-to-End Task Oriented Dialog System using Context Summarization

Anonymous Authors
Affiliation
Address Line 1

Abstract

Creating Task Oriented Dialog (TOD) Systems that generalize well to unseen domains has been a challenging research area in dialog. Most systems use the dialog history as context, and as the number of turns in a dialog increases, the context becomes too long and in many cases contain repetitive and unnecessary information. Also, the metrics used to evaluate dialog systems mainly focus on Dialog State Tracking (DST) and response generation, and seem to neglect system actions. In this paper, we propose a novel TOD system that uses a compressed context consisting of the latest DST and the last user utterance. Using a compressed context allows us to feed additional information like the schema, list of system actions, user actions and service results to the model, and still use language models like GPT-2. We propose a two step training process, where in the first pass we calculate the cross entropy loss on the context and target, and in the second pass we calculate the loss only on the target. To get a better understanding of the system actions, we propose two new metrics called Average Action Accuracy (AAA) and Joint Action Accuracy (JAA). Experimental results on the Schema Guided Dialogue (SGD) dataset show that our model outperforms the state of the art models in terms of zero shot generalization.

1 Introductions

- Basic Tod intro
- Challenge of zero-shot TOD
- Existing methods
- Our Method
- Pros of our methodology
- Summary of results

TOD systems interact with users in the form of dialog using natural language, to accomplish user tasks. The system needs to understand user needs and provide the best possible response to the user. The task of extracting user intent and goals from conversations by filling belief slots is called Dialog State Tracking (DST) (Wang, Liu, and Zhao 2016). Using the DST and dialog history, the system needs to decide what actions to take and then convey that action in the form of natural language to the user.

Traditional TOD systems were built using a pipeline approach, where each component was created separately and then integrated together. However, with the adaptation of large pretrained language models (Devlin et al. 2019; Radford et al. 2019), researchers have moved towards end-to-end systems (Hosseini-Asl et al. 2020; Peng et al. 2021; Lee et al. 2020; Yang, Li, and Quan 2020; Jeon and Lee 2021; Sun et al. 2022a; Yang et al. 2022), In these systems, the dialog history is fed to the model as input and the output is a cascaded generation (Su et al. 2021) of the DST, System Actions and System Response.

A major drawback of most of these systems is that they fail to generalize to unseen domains. In real-world setting, ideally a model should have the capablity to adjust to new domains. Some work has been done to address this issue (Feng, Wang, and Li 2020; Lee, Cheng, and Ostendorf 2021; Noroozi et al. 2020; Mosig, Mehri, and Kober 2020; Mehri and Eskénazi 2021), but the focus has been on DST and next action prediction, not end-to-end systems. Another issue is that in dialogs with many turns, the dialog history becomes very long, repetitive, and slot values could be updated multiple times in different turns depending on the needs of the user, thus making it difficult for systems to correctly model long-range semantic dependencies (Sun et al. 2022b).

To address the aforementioned challenges, we propose a novel Schema Guided Zero-Shot Generalizable End-to-End TOD system using Context Summarization that outperforms existing state-of-the-art systems. We use a summarized context consisting of the latest DST, the last user utterance, related domain schemas and system action names. We also propose a two step training process, where in the first pass we calculate the cross entropy loss on the context and target, and in the second pass we calculate the loss only on the target. We conduct experiments on the Schema-Guided Dialog (SGD) dataset and perform a thorough analysis and an ablation study to show the effectiveness of our approach. To the best of our knowledge, this is the first Zero-Shot End-to-End TOD system designed for the SGD dataset.

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

- Summary and importance of approach
- Base model (GPT-2)
- Prompt (context, additional info)
- Training (pretraining, training)

2.1 Problem Formulation

A dialog session is composed of multiple turns, which consists of interactions between the user and the system in natural language utterance. The SGD dataset provides a list of Schemas, $S = (s_1, ..., s_n)$ and each dialog contains a list of service names, which can be used to extract the relevent schema for that dialog, $S_r \in S$.

For a turn t, the inputs to the model are the following: user utterance U_t, DST from the previous turn D_{t-1}, relevant schemas S_r, database search results Db_t and a list of system action names $Action_{all}$. The model autoregressively generates the dialog state D_t, user actions UA_t, system actions SA_t and system response R_t. Figure 1 shows a visual representation of the overall approach.

2.2 Training

The model is trained in two steps, where in the first step we calculate the cross entropy loss on the context and target, and in the second step we calculate the loss only on the target. The intuition behind this is that in the first step the model understand the general structure of the text and in the second step the model fine tunes the target text.

3 Experimental Setup

3.1 Datasets

The Schema Guided Dialogue (SGD) dataset is a large scale dataset for task oriented dialogue that consists of over 16K multi domain dialogs between a human and a virtual assistant covering 16 domains. The dataset also provides a schema for each domain that provides a textual description of the domain, list of slots and list of intents. A slot contains a name, textual description, and possible values for categorical slots and an intent contains a name, textual description, optional slots and result slots.

3.2 Evaluation Metrics

To have a fair comparison with other methods, we use the evaluation script from SGD dataset and report the Active Intent Accuracy, Requested Slot F1, Average Goal Accuracy and Joint Goal Accuracy.

However, the metrics in the SGD dataset are geared towards DST, and do not take into account the system actions and response. To evaluate task completion we report Inform, Success, Average Action Accuracy (AAA) and Joint Action Accuracy (JAA). For evaluating 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, Srinivasan, and Eskenazi 2019) with (Inform + Success) × 0.5 + GLEU.

The metric Inform measures whether a system has provided a correct entity and Success measures whether it has answered all the requested information. To calculate inform, from the ground truth system actions, we filter actions by action type inform (Inform, Inform Count) and check if they are predicted correctly. To calculate success, we filter actions by slot names that are in the requested slots and check if the action slot values are predicted correctly. AAA and JAA are similar to the goal metrics in SGD, but are calculated from system actions.

4 Results

- Best Results
- Comparison Table with other methods
- Ablation Study
- Ablation table of our model to show importance of each component

We compare our approach with other popular approaches that have been proposed for task oriented dialog and report the results in Table 1.

4.1 Ablation Study

To better understand the effect of different components in our approach, we present an ablation study in Table **??**.

Describe effect of different components.

5 Related Works

5.1 Supervised End to End Models

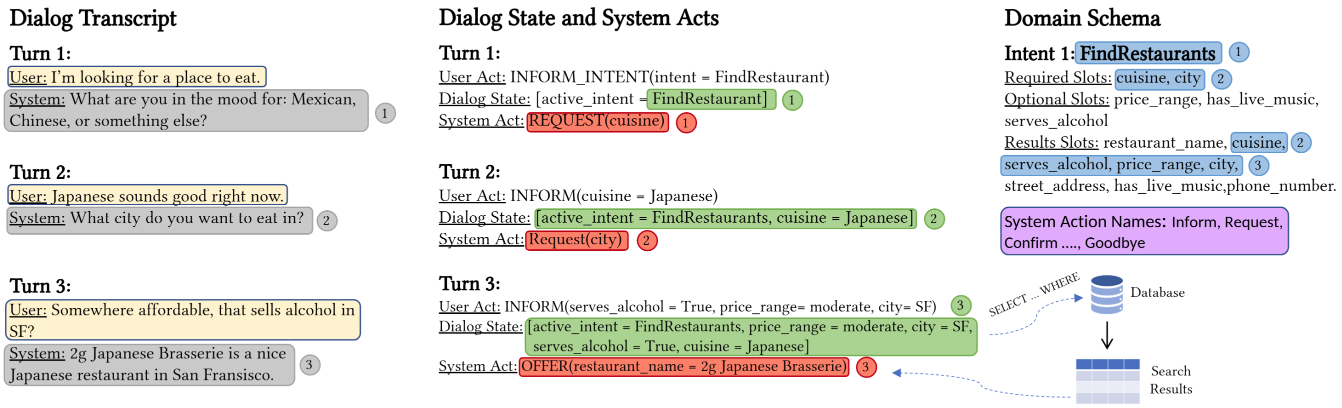
Pretrained language models like BERT (Devlin et al. 2019), GPT-2 (Radford et al. 2019) and T5 (Raffel et al. 2019) have been used extensively in the literature for End to End models for TOD systems (Hosseini-Asl et al. 2020; Peng et al. 2021; Lee et al. 2020; Yang, Li, and Quan 2020; Jeon and Lee 2021; Sun et al. 2022a; Yang et al. 2022; Noroozi et al. 2020). In these models, the context consists of the dialog history, whereas our approach uses 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 generalize well to unseen domains.

5.2 Zero Shot Dialog Models

Recently, some work has been done on Zero Shot generalizability by incorporating schema on certain components of TOD systems, such as for DST (Feng, Wang, and Li 2020; Lee, Cheng, and Ostendorf 2021; Noroozi et al. 2020) and next action prediction and response generation (Mosig, Mehri, and Kober 2020; Mehri and Eskénazi 2021). However, unlike ours, these systems are not end-to-end.

References

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv* abs/1810.04805.



Dialog Transcript

Turn 1:

User: I'm looking for a place to eat.

System: What are you in the mood for: Mexican, Chinese, or something else?

Turn 2:

User: Japanese sounds good right now.

System: What city do you want to eat in?

Turn 3:

User: Somewhere affordable, that sells alcohol in SF?

System: 2g Japanese Brasserie is a nice Japanese restaurant in San Fransisco.

Dialog State and System Acts

Turn 1:

User Act: INFORM_INTENT(intent = FindRestaurant)

Dialog State: [active_intent = FindRestaurant]

System Act: REQUEST(cuisine)

Turn 2:

User Act: INFORM(cuisine = Japanese)

Dialog State: [active_intent = FindRestaurants, cuisine = Japanese]

System Act: Request(city)

Turn 3:

User Act: INFORM(serves_alcohol = True, price_range= moderate, city= SF)

Dialog State: [active_intent = FindRestaurants, price_range = moderate, city = SF, serves_alcohol = True, cuisine = Japanese]

System Act: OFFER(restaurant_name = 2g Japanese Brasserie)

Domain Schema

Intent 1: FindRestaurants

Required Slots: cuisine, city

Optional Slots: price_range, has_live_music, serves_alcohol

Results Slots: restaurant_name, cuisine, serves_alcohol, price_range, city, street_address, has_live_music,phone_number.

System Action Names: Inform, Request, Confirm ...., Goodbye

Figure 1: Using the user utterance (brown), domain schema (blue), system action names (purple) and database results, our model autoregressively generates the dialog state (green), user actions (yellow) system action (red) and system response (gray)

Model	Setting	Avg act Acc	Joint act Acc	Avg goal Acc	Joint goal Acc	Inform	Intent Acc	Req Slots F1	Response BLEU	Response ROUGE	Success	Avg u_act Acc	Joint u_act Acc	Combined
Our	all	56.25	46.07	73.53	45.59	66.46	88.31	95.61	16.22	18.52	58.63	74.08	65.63	78.76
	seen	58.41	48.64	75.7	49.87	68.7	89.09	95.97	23.12	18.52	61.05	75.97	67.99	88
	unseen	55.56	45.31	73.15	44.85	65.14	88.19	95.51	14.29	18.52	58.07	73.5	64.94	75.89

Table 1: Main Results

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

Hosseini-Asl, E.; McCann, B.; Wu, C.-S.; Yavuz, S.; and Socher, R. 2020. A simple language model for task-oriented dialogue. *ArXiv* abs/2005.00796.

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

Lee, H.; Jo, S.; Kim, H.; Jung, S.; and Kim, T.-Y. 2020. Sumbt+larl: End-to-end neural task-oriented dialog system with reinforcement learning. *ArXiv* abs/2009.10447.

Lee, C.-H.; Cheng, H.; and Ostendorf, M. 2021. Dialogue state tracking with a language model using schema-driven prompting. In *Conference on Empirical Methods in Natural Language Processing*.

Lin, C.-Y., and Och, F. 2004. Looking for a few good metrics: Rouge and its evaluation. In *Ntcir workshop*.

Mehri, S., and Eskénazi, M. 2021. Schema-guided paradigm for zero-shot dialog. In *SIGDIAL Conferences*.

Mehri, S.; Srinivasan, T.; and Eskenazi, M. 2019. Structured fusion networks for dialog. *arXiv preprint arXiv:1907.10016*.

Mosig, J. E. M.; Mehri, S.; and Kober, T. 2020. Star: A schema-guided dialog dataset for transfer learning. *ArXiv* abs/2010.11853.

Noroozi, V.; Zhang, Y.; Bakhturina, E.; and Kornuta, T. 2020. A fast and robust bert-based dialogue state tracker for schema guided dialogue dataset. *ArXiv* abs/2008.12335.

Peng, B.; Li, C.; Li, J.; Shayandeh, S.; Lidén, L.; and Gao, J. 2021. Soloist: Building task bots at scale with transfer learning and machine teaching. *Transactions of the Association for Computational Linguistics* 9:807–824.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2019. Language models are unsupervised multitask learners.

Raffel, C.; Shazeer, N. M.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; and Liu, P. J. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *ArXiv* abs/1910.10683.

Su, Y.; Shu, L.; Mansimov, E.; Gupta, A.; Cai, D.; Lai, Y.-A.; and Zhang, Y. 2021. Multi-task pre-training for plug-and-play task-oriented dialogue system. *arXiv preprint arXiv:2109.14739*.

Sun, H.; Bao, J.; Wu, Y.; and He, X. 2022a. Bort: Back and denoising reconstruction for end-to-end task-oriented dialog. In *NAACL-HLT*.

Sun, H.; Bao, J.; Wu, Y.; and He, X. 2022b. Mars: Semantic-aware contrastive learning for end-to-end task-oriented dialog. *arXiv preprint arXiv:2210.08917*.

Wang, B.; Liu, K.; and Zhao, J. 2016. Inner attention based recurrent neural networks for answer selection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1288–1297. Berlin, Germany: Association for Computational Linguistics.

Wu, Y.; Schuster, M.; Chen, Z.; Le, Q. V.; Norouzi, M.; Macherey, W.; Krikun, M.; Cao, Y.; Gao, Q.; Macherey, K.; Klingner, J.; Shah, A.; Johnson, M.; Liu, X.; Łukasz Kaiser; Gouws, S.; Kato, Y.; Kudo, T.; Kazawa, H.; Stevens, K.; Kurian, G.; Patil, N.; Wang, W.; Young, C.; Smith, J.; Riesa, J.; Rudnick, A.; Vinyals, O.; Corrado, G.; Hughes, M.; and Dean, J. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation.

Yang, Y.; Ding, H.; Liu, Q.; and Quan, X. 2022. Ubarv2:

Towards mitigating exposure bias in task-oriented dialogs. *ArXiv* abs/2209.07239.

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