TopKG: Target-oriented Dialog via Global Planning on Knowledge Graph

Zhitong Yang[1,3,2], **Bo Wang**[2,1]*, **Jinfeng Zhou**[2,1], **Yue Tan**[2,1], **Dongming Zhao**[4], **Kun Huang**[4], **Ruifang He**[2,1], **Yuexian Hou**[2]

[1]State Key Laboratory of Communication Content Cognition, People's Daily Online, Beijing, China

[2]College of Intelligence and Computing, Tianjin University, Tianjin, China

[3]School of New Media and Communication, Tianjin University, Tianjin, China

[4]AI Lab, China Mobile Communication Group Tianjin Co., Ltd.

{yyyyyyzt, bo_wang, jfzhou, tanyue_098}@tju.edu.cn

Abstract

Target-oriented dialog aims to reach a global target through multi-turn conversation. The key to the task is the global planning towards the target, which flexibly guides the dialog concerning the context. However, existing target-oriented dialog works take a local and greedy strategy for response generation, where global planning is absent. In this work, we propose global planning for target-oriented dialog on a commonsense knowledge graph (KG). We design a global reinforcement learning with the planned paths to flexibly adjust the local response generation model towards the global target. We also propose a KG-based method to collect target-oriented samples automatically from the chit-chat corpus for model training. Experiments show that our method can reach the target with a higher success rate, fewer turns, and more coherent responses.

1 Introduction

Human-like dialog agents have three types of approaches: open-domain (Zhang et al., 2019a; Huang et al., 2020), task-oriented (Budzianowski et al., 2018; Rastogi et al., 2020; Yang et al., 2020), and target-oriented dialog (Tang et al., 2019; Qin et al., 2020; Zhong et al., 2021). The open-domain dialog only requires the dialog generation to be fluent and context coherent. In contrast, typical task-oriented dialog further completes a specific task by understanding users' intention and collecting the required information of predefined sub-tasks of the intention. However, as a more challenging task, target-oriented dialog aims to achieve a global target that often can not be clearly defined as sub-tasks. The dialog agents are required to lead the conversation to the target flexibly, and the process is excepted to be coherent, effective, and successful. Due to its purpose and flexibility, target-oriented dialog agents have a broad-based demand, e.g., conversational recommendation (Li et al., 2018; Kang et al., 2019), psychotherapy (Sharma et al., 2020), and education (Clarizia et al., 2018). In these fields, a typical expectation of target-oriented dialog is to actively lead the conversation by smoothly changing the dialog topic to a designated one, e.g., a product, a stimulus of mind, and a knowledge point.

To reach a target topic effectively and coherently in dialog, existing approaches primarily represent the topic as keywords and adopt a two-stage architecture, i.e., predicting a next-turn keyword and keyword-augmented response retrieval (Tang et al., 2019). In this direction, Xu et al. (2020b) further introduces reinforcement learning with "target similarity" rewards to target-oriented dialog learning. However, the target-oriented dialog is a typical knowledge-rich task. Although dialog context can support the semantic concern of dialog generation, it is not quite effective to model the knowledge-driven process in the target-oriented dialog. To involve global knowledge, Qin et al. (2020) and Xu et al. (2020a) incorporate a dialog graph into the target-oriented dialog and Zhong et al. (2021) uses the external commonsense KG (ConceptNet (Speer et al., 2017)) to improve the performance.

Although existing target-oriented dialog works have demonstrated practical approaches in self-simulation test, there is still some open issues: (1) Lack of multi-turn target-oriented dialog corpus for training and benchmarks. Most existing target-oriented corpus are prepared for next-turn local target (e.g., OTTers(Sevegnani et al., 2021)), or adopt chit-chat corpora and randomly select a keyword in the next-turn utterance as the local target, (2) Lack of global planning of dialog process. Although the latest works use a global target to guide every turn of response generation, they adopt a short-sighted and greedy strategy instead of global planning to optimize the process towards the global target.

To this end, we propose **T**arget-**O**riented dialog with global **P**lanning on **K**nowledge **G**raph (**TopKG**), which effectively supports the target-

*Corresponding author.

Proceedings of the 29th International Conference on Computational Linguistics, pages 745–755
October 12–17, 2022.

oriented process by global reasoning on KG concerning the dialog context. Specifically, to address the first data issue, we automatically select a new dataset named Target-Guided ConvAI (**TGConv**) from the chit-chat corpus ConvAI2 (Dinan et al., 2020). We select target-oriented samples from ConvAI2 by identifying the dialog utterances containing a go-through entity sequence that aligns with the KG path. Furthermore, we distinguish the selected dialog samples according to whether the global target is easy to reach or not to verify the performance of TopKG in dealing with hard global target-oriented cases. For instance, the sample in the left part of Figure1 is target-oriented because the keywords in this dialog are connected (direct or low-order connected) in a commonsense KG, which embodies a smooth transition towards global target words. To address the second issue, we first improve the existing one-turn target-oriented response generation, trained in a supervised fashion to predict a next-turn keyword and generate a fluent and coherent response with the predicted keyword. Using the improved one-turn model as local-model, we further introduce a reinforcement learning based global-model to effectively guide the local-model towards a global target with global planning on KG. Specially, the global-model adjusts the next-turn keyword selection of the local-model to follow the global planning path on KG and reward the keyword-based response generation with success in reaching the global target.

Our main contributions are as follows:

(1) We propose a simple yet effective way to automatically extract multi-turn global target-oriented dialog from the chit-chat corpus to develop global target-oriented dialog agent. We also distinguish the selected dialog into easy-to-reach target and hard-to-reach target.

(2) We make the first step towards global planning in global target-oriented dialog. A two-stage learning framework is designed to guide a next-turn local model with a reinforcement learning based global model which is guided by global planning in commonsense KG.

(3) With automatic and human metrics, we verify that TopKG exceeds baselines on reaching global target with more coherent semantics, fewer turns, and a higher success rate in reaching targets.

The dataset can be downloaded in data folders from https://github.com/yyyyyyzt/topkgchat

2 Related Work

Target-oriented dialogue systems. Current target-oriented dialog studies can generally be divided into local-target oriented and global-target oriented methods. Local-target oriented methods (Wang et al., 2021) pays attention to the next-turn target. For example, Xu et al. (2020b,a) proposes a hierarchical policy model to plan and generate responses of different levels where the high-level policy plans a topic. However, the low-level policy plans responses that are coherent to this topic instead of approaching it. Global-target oriented methods (Qin et al., 2020; Zhong et al., 2021) uses global target to guide every turn of response generation. These methods propose a keyword predictor to determine the next-turn keyword to talk about and produce a response relevant to the determined keyword. However, they adopt a short-sighted and greedy strategy instead of explicit planning to optimize the process towards the global target.

KG-grounded dialogue systems. Leveraging background information for dialogue system improvement is a well-researched topic, especially in target-oriented settings. Some work uses structured knowledge, DKRN (Qin et al., 2020) incorporates a dialog graph, and CKC (Zhong et al., 2021) uses the ConceptNet to improve the performance. For how to utilize KG, classical methods are divided into using full path (Ma et al., 2021) and using flexible path fragments (Zhou et al., 2021). These models enjoy rich knowledge augmentation since short KG paths relating to the context are encoded, but they lack the ability to plan on KG. Another set of works focuses on grounds in unstructured knowledge (Zhao et al., 2020; Wu et al., 2020), which can also be divided into independent sentences and documents. This unstructured knowledge is more challenging to use than KG.

3 Our Approach

Task Definition Formally, $C = \{c_1, \cdots, c_i\}$ is the current dialog context involving latest i utterances. A knowledge graph $G_{KG} = V_{KG} \times E_{KG}$ is composed of the commonsense entities V_{KG} and relations E_{KG}. Given C, G_{KG} and a global target keyword K_{target}, the global target-oriented dialog is firstly required to figure out a next-turn keyword z from the G_{KG}, and generate a response r related to z. Furthermore, with multi-turn response generation, the global target-oriented dialog need to successfully mentioned a global target keyword

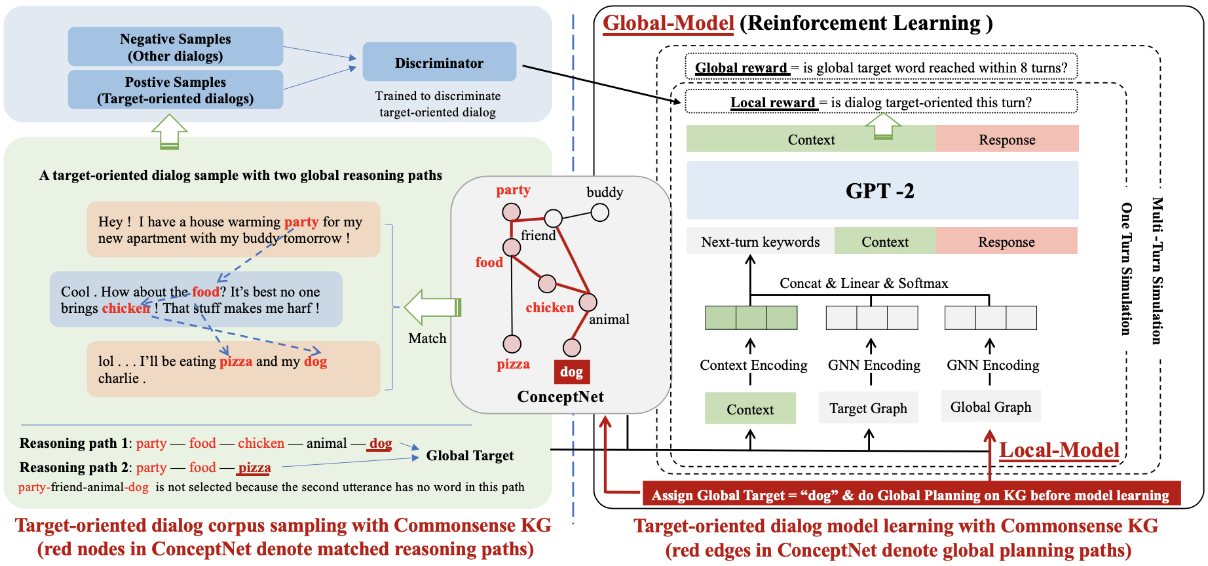

Figure 1: The left part illustrates how to select target-oriented dialogs for model learning by matching the word sequence across utterances with reasoning paths in KG. The right part illustrates how the global-model uses reinforcement learning to guide a GPT-2 based local-model to follow the global planning in KG. Global Planning on KG is pre-performed before the learning of Local-Model and Global-Model, and planning paths are essential to guide the multi-turn responses generation.

K_{target} with fewer turns and keep the response be coherent to the context in each turn.

3.1 Method Overview

Our approach consists of two main contributions: an automatic method for target-oriented conversation dataset collection and a two-stage learning model for global target-oriented dialog generation.

Target-oriented Conversation dataset As existing multi-turn dialog corpora are not specially created for target-oriented tasks, we firstly propose automatically selecting the target-oriented dialog session from the general dialog corpora. A dialog session was selected from the general chit-chat corpus by examining whether a KG-explainable entity path is running through a dialog. In addition, we indicate the entity path and specify the easy target and the hard target. The example shown in Table1.

Two-stage learning model We divide the task into two progressive stages in Fig1: local-model of next-turn strategy learning (stage 1) and global-model of multi-turn strategy learning (stage 2). Specifically, at stage 1, the local-model is supervised trained to predict next-turn keywords and generate a response related to the keywords. In stage 2, we design a reinforcement learning to adjust the local-model to explore all potential paths by global planning in a commonsense KG towards the global target word, where a bidirectional heuristic reasoning obtains the paths. We also reward the generated response in each turn by whether the dialog till this turn is target-oriented and whether the dialog finally reaches the global target word.

Dialog	**A:** I spend a lot of time **outside**. **B:** I like the outdoors as well, especially **gardening** . **A:** Wow! I used to have a **garden** too. **B:** I love sipping coffee while enjoying **flowers** in my garden. **A:** Flowers are always beautiful and **colorful** ! **B:** I like anything with **art**, especially colorful things.
Entity Path	**Outside-Garden-Flower-Color-Art**
Target	**Art**

Table 1: A target-oriented example dialog in TGConv

3.2 Target-oriented dialog corpus sampling

In this section, we construct a target-oriented dialog corpus (named **TGConv**) from chit-chat corpus ConvAI2 (Dinan et al., 2020).

3.2.1 Identify target-oriented dialog

We suppose a dialog is a positive example of target-oriented dialog if there is a consistent reasoning path of words linking all the utterances in their order in the dialog. A reasoning path of words is $p = \{w_1 \rightarrow w_2 \rightarrow \cdots \rightarrow w_n\}$, where w_i is a word, e.g., "*Outside-Garden-Flower-Color-Art*" in Table 1. To be logical, each neighbor word pair, i.e.,

w_i and w_{i+1}, should match the names of the two nodes of an edge in the ConceptNet, respectively. To link all the utterances in dialog, each utterance in the dialog should provide at least one word to p. To keep the order in dialogue, w_i should be in the same order in p as they appear in the dialog. Except for positive samples, other samples in the corpus are identified as negative examples.

3.2.2 Global target assignment

For each positive example dialog associated with a reasoning path p, we select the last word w_n in p as the global target K_{target}. Furthermore, to better evaluate the model's ability to guide the dialog to the target of different difficulties, we distinguish target words into "easy-to-reach" and "hard-to-reach". Specifically, target words with low frequency in the corpus are classified as "hard-to-reach" target, because there are fewer cases to learn the semantic transition to low-frequency target words (less than 800) in local-model and global-model.

3.3 Global Planning

Global planning is the key to successfully accomplishing target-oriented task. We finally obtain a graph consisting of a set of potential paths through global planning, which embodies the keyword transition from the initial context to the global target word. Building a connected graph G_{global} from the starting to target allows us to learn a better graph representation and facilitate our model to explore better paths. Specifically, we identify the noun and verb concepts in the dialogue context and then use a bidirectional reasoning method to find KG paths over ConceptNet effectively. Bidirectional reasoning is a graph search algorithm that finds smallest path from the initial to the target entity. It runs two simultaneous search: 1) Forward search from source/initial entity toward goal entity and 2) Backward search from goal/target entity toward source entity. This algorithm is very suitable for target-oriented task scenarios, and the detailed process is shown in Algorithm 1.

Algorithm 1: Global Planning by Bidirectional Reasoning over $ConceptNet$

Input: ConceptNet, G_{KG}; Target, K_{target};
The set of concepts in start:
$V_{start} = \{v_1, v_2 \cdots v_m\}$;
Output: A graph consists of all potential paths from source to target, G_{global}

```
Initialize graph G_global;
foreach node v_i of the V_start do
    Initialize a concept stack S contain v_i;
    for h from 1 to maximum hops H do
        while S is not emtpy do
            Pop a head entity v_h from S;
            N_i: the neighbouring concepts of v_h in
              ConceptNet;
            Select the top K concepts most similar
              to the head entity v_h from N_i;
            Select the top K concepts most similar
              to the target entity K_target from N_i;
            Add them to an empty temporal triple
              list T;
            foreach (v_h, r, v_t) in T do
                Add v_h, v_t and r into G;
                if v_t not in G_global then
                    Push v_t in S
                end
            end
        end
    end
end
Repeat the above process from K_target to V_start;
```

3.4 Supervised Learning of Local-Model

We let the local-model learn next-turn target-oriented policy in a supervised fashion. The local-model architecture is shown in the right part of Fig 1. In order to predict the next turn keywords z, we need to model the candidate words, the context, and the target, respectively. Firstly, we get the target entity and its neighbors on the ConceptNet to build a subgraph G_{target} and use the method in the previous section to get a global graph G_{global}. Then we apply a multi-layer GCN encoder to model the graphs. Besides, we use a typical transformer encoder for context understanding. Finally, we predict a keyword and generate a coherent response by generator for approaching the target.

3.4.1 Graph-based Encoder

We use a graph-based encoder to model graph node representations for predicting keywords. Here we use two graphs G_{global} and G_{target}, The G_{global} is a large graph that contains all potential paths from start context to target, and G_{target} only contains target entity and its neighbor nodes to enhance the target representation.

Therefore, to obtain the representation of concepts and relations, we apply multi-layer GCN (Kipf and Welling, 2016) encoders to encode the G_{global} and G_{target}. Moreover, following the idea of the TransE model (Bordes et al., 2013), we update a concept embedding with the subtraction between each neighbor concept embedding and the corresponding relation embedding to obtain the relation representation. The concepts V

in two graphs are initialized by pretrained word embeddings[1], and the relations R in graph are initialized with randomly embeddings. For each concept v_i, we update its embedding at the $(l+1)^{th}$ layer by aggregating its neighbours N_i including pairs of the concept and the relation liking to v_i:

$$h_i^{(l+1)} = \sigma \left(W_s^{(l)} h_i^{(l)} + \sum_{(j,r) \in N_i} \frac{1}{|\mathcal{N}_i|} W_n^{(l)} \left(h_j^{(l)} - h_r^{(l)} \right) \right) \quad (1)$$

where h_i^l, h_j^l and h_r^l are the embeddings of node v_i, node v_j, and the relation between v_i and v_j at layer $(l)^{th}$; $W_s^{(l)}$ and $W_n^{(l)}$ are the two trainable parameter matrices specific to the layer $(l)^{th}$; and σ is a non-linear active function. The relation embedding is also updated at the $(l+1)^{th}$ layer via a linear active function: $h_r^{(l+1)} = W_R^{(l)} h_r^{(l)}$. After L layers, we are able to obtain a set of concept representations $\{h_{v_1}^{(L)}, \ldots, h_{v_{|V|}}^{(L)}\}$.

3.4.2 Conversation Context Encoder

We utilize a transformer encoder for conversation context understanding. Same as previous works, we flatten conversation context in C, and then add a special token $[CLS]$ at the beginning of the input. $\bar{C} = [CLS; C]$ is fed into Transformer Encoder, then output representation of $[CLS]$ token denoting the global memory of the whole sequence.

3.4.3 Classification

Now we have the context representation, G_{global} concepts representation, and G_{target} concepts representation for predicting words. Finally, we concatenated these vectors and fed to a linear transformation layer, followed by a softmax layer. We limited the candidates to two-hop entities based on context. The entire model is optimized by minimizing the cross-entropy loss.

3.4.4 Keyword Augmented Generator

After we get the next-turn keywords word z, we employ a keyword-augmented GPT (Radford et al., 2019) to generate a response to approaching the target. The generator takes keywords z and context C as the input, and the following text r as the target reference. Specifically, the z and C are first concatenated by a special separator token. The training objective follows a standard language model (LM) loss(Zhang et al., 2019b):

$$p_\Theta(r \mid C, z) = \prod_{t=0}^{|r|} p\left(r_t \mid x, z, r_{0:t-1}\right) \quad (2)$$

where r_t is the t-th token in r.

3.5 Reinforcement Learning of Global-Model

As our main contribution, we propose a global-model to explore better dialog strategies toward the global target through reinforcement learning. Although the local-model performs well on next turn response generation, it tends to be short-sighted and ineffective in reaching the global target in the multi-turn dialog. Therefore, we design a simulation-based environment to guide the local-model toward the global target through reinforcement learning. To this end, we let the model talk to itself. At the start of the dialog, we explicitly search a set of planning paths (described in 3.3) in ConceptNet from the initial context to the global target word. Then we use searched planning paths to adjust the next-turn keyword prediction to obey the planning paths and generate a response with the keyword. Furthermore, the generated response is rewarded by its target-oriented coherence to the context and the success of the global target. Global-model consists of the following components.

3.5.1 State/Action

At each time step t, the **state** S_t is a tuple of $[G_{global}; G_{target}; C]$, where G_{global} is a graph of planning paths obtained at the start of the dialog, and G_{target} is the predefined global target word and its neighbors, and C is the current context. Given the current dialog state, an **action** is the next-turn keyword z, and the action space is the potential paths obtained by global planning.

3.5.2 Reward

We use Local Reward and Global Reward to encourage the dialog to be contextual and coherent and explore global target-oriented strategy.

Local Reward encourages the contextual consistency at each turn of dialog, which is the discriminator score of the utterances sequence containing the current context and generated response, the detail are as below 3.5.3.

Global Reward encourages the global target-oriented response by giving a positive reward of "1" if the global target word finally appears in the last turn or a negative reward of "-1" otherwise.

[1] We use GloVe embedding of size 300 (Pennington et al., 2014)

3.5.3 Discriminator for local reward evaluation

To reward the dialog (context+response) which are more likely to be target-oriented, we train a discriminator to tell whether an utterance sequence is semantically target-oriented. To this end, the discriminator is trained to classify the positive and negative samples collected in section 3.2. Specially, the positive and negative samples with 1/0 label: $\mathbf{X} = [CLS; c; SEP; r]$ or $\mathbf{X} = [CLS; c; SEP; \overline{r}]$ is fed into pre-trained language model (BERT) (Devlin et al., 2018), then output representation of [CLS] token is used for classification. The classification score is formulated as

$$f_{score}(X) = \sigma(\mathbf{w}^\top \mathbf{x}_{[\text{CLS}]} + b) \tag{3}$$

where w and b are trainable parameters. We use binary cross-entropy loss to optimize the models.

3.5.4 Training

We apply Proximal Policy Optimization (Schulman et al., 2017), a stable policy based RL algorithm using a constant clipping mechanism as the soft constraint, for dialog policy optimization:

$$J_\pi(\theta) = E_{s,a\sim\pi}\left[\min\left\{\beta_t \hat{A}_t, \mathit{clip}\left(\beta_t, 1-\epsilon, 1+\epsilon\right)\hat{A}_t\right\}\right] \tag{4}$$

$\hat{A}_t = R_t - \hat{V}_\phi(s_t)$ is the estimated advantage, where $R_t = \sum_{\tau=t}^{T}$ is the local reward adding global reward, $\hat{V}_\phi$ is the estimated value function of state S_t with parameters ϕ, $\beta_t = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$ is the ratio of the probability under the new and old policies, δ is TD residual, λ and ε are hyper-parameters.

4 Experiments and Results

4.1 Datasets

We evaluate TopKG and baselines on two datasets. To verify the ability to guide the user to the target topic in multi-turn of dialogue, we use our proposed dataset TGConv, which is extracted from ConvAI2 (Dinan et al., 2020) and is distinguished into "easy-to-reach/hard-to-reach" targets with the method in section 3.2. ConvAI2 is a chit-chat dataset based on the PersonaChat for NIPS 2018 competition, which contains high-quality open-domain dialogues, including diverse topics. In addition, one-turn dialogue is a special case of multi-turn, therefore we also conduct our evaluation on a next-turn target-oriented dataset OTTers(ood)[2] (Sevegnani et al., 2021). OTTers requires the agent pro-actively generate an "bridging" utterance to approach the target, which is consistent with the input and output of the task on TGConv. The statistics of the two datasets are presented in Table2.

Dataset	Split	#Conv.	Avg. #Utter.	Avg. #Word.	Avg. #Entity.	Avg. #Coh.
OTTers	Train	2034	3.0	9.47	2.86	0.45
	Valid	1152	3.0	9.56	2.95	0.45
	Test	1130	3.0	9.19	2.80	0.44
TGConv	Train	15197	8.35	12.60	2.89	0.32
	Valid	2681	7.96	12.29	2.85	0.31
	Test	1000	8.97	12.47	2.91	0.32

Table 2: Dataset statistics. Avg.#Utter., #Word., #Entity., #Coh. denotes the average number of utterances, words, entities, semantic similarity per dialogue, utterance, utterance, utterance.

4.2 Baselines

We select four baselines in end-to-end (GPT-2, MultiGen) and pipeline style (DKRN, CKC), respectively. The first baseline is **GPT-2** (Radford et al., 2019). Next, we test the recent **Multi-Gen** (Ji et al., 2020), extends GPT-2 with multi-hop reasoning on commonsense knowledge graphs. The third baseline is **DKRN** (Qin et al., 2020), which builds a dialog graph from the corpus for topic transition. The last baseline is **CKC** (Zhong et al., 2021), the state-of-the-art approach using ConceptNet for this task. In addition, DKRN and CKC are retrieval models. Here we replace the retriever with the generator in our paper.

4.3 Metrics

Local-Evaluation To evaluate models' performance in generating next-turn response, we firstly perform automatic evaluation using commonly adopted text generation metrics, including CIDEr (Vedantam et al., 2015), ROUGE-L (Lin, 2004) and METEOR (Banerjee and Lavie, 2005). However, we report the full BLEU score[3] (Papineni et al., 2002) that accounts for the overlap across 1-4 ngrams instead of only 4-grams (BLEU-4). In addition, we use hits@K ratio to measure the number of relevant entities correctly predicted by the

[2]OTTers have different train-dev-test (in-domain and out-of-domain) splits, we choose out-of-domain(ood) split. The ood split resembles a zero-shot scenario, where the model has to generate a shift between two topics it has never been fine-tuned on.

[3]SacreBLEU (Post, 2018) provides hassle-free computation of shareable, comparable, and reproducible BLEU scores.

	BLEU$^{1-4}$	METEOR	ROUGE-L	CIDEr	hits@1	hits@3
GPT2	11.58	10.26	17.67	13.75	4.39	15.79
MultiGen	13.57	12.51	26.27	15.48	6.58	20.51
DKRN	12.86	11.90	21.52	14.33	4.91	17.72
CKC	13.34	11.65	24.77	14.46	6.87	21.89
TopKG	**15.35****	**13.41***	27.16	**17.18***	7.78	**22.06***
w/o global plan	14.89	12.89	26.99	16.22	7.45	21.14
w/ small graph(K=5,H=3)	13.24	10.65	25.53	15.62	6.77	21.22
w/ large graph(K=20,H=6)	15.24	11.65	**27.53**	16.62	**7.79**	21.63

Table 3: Automatic evaluation of next-turn response generation on OTTers. Numbers marked with * indicate that the improvement is statistically significant compared with the best baseline(t-test with p-value < 0.05).

	Easy Target			Hard Target		
	Succ.(%)	Turns	Coh.	Succ.(%)	Turns	Coh.
GPT2	22.3	2.86	0.23	17.3	**2.94**	0.21
MultiGen	26.7	**2.55**	0.21	19.6	7.31	0.24
DKRN	38.6	4.24	0.33	21.7	7.19	0.31
CKC	41.9	4.08	**0.35**	24.8	6.88	0.33
TopKG	**48.9***	3.95	0.31	**27.3***	4.96	**0.33**
w/o global plan	35.4	4.51	0.32	21.3	7.18	0.32

Table 4: Automatic evaluation of global guiding on TGConv. Note that our task requirement is to reach the target smoothly and fast. “Coh.” and “Turns” not the higher / lower the better.

	Easy Target		Hard Target	
	G-Coh.	Effect.	G-Coh.	Effect.
GPT2	1.13	1.20	1.13	0.86
MultiGen	1.24	1.29	1.17	1.13
DKRN	1.26	1.23	1.19	1.18
CKC	**1.53**	1.31	1.23	1.16
TopKG	1.51	**1.67**	**1.37**	**1.48**
w/o global plan	1.42	1.34	1.24	1.13
kappa	0.45	0.55	0.51	0.58

Table 5: Comparison of human evaluation metric Coherence and Effectiveness results on self-chat dialogues among our model and baselines. The agreement among the annotators is measured by the Fleiss’s kappa. The agreement ratio kappa in [0.41, 0.6] denotes the moderate agreement.

local-model, out of the K most important entities identified in the target references.

Global Evaluation To evaluate models’ performance in guiding the dialog to global target, as existing works (Qin et al., 2020; Zhong et al., 2021) do, we use a simulator to simulate multi-turn dialog. "Succ." automatically measures the success rate of generating the global target word within 8 turns. "Turns" indicates the average turns of all dialogs which successfully generate the global target word. "Coherence (Coh.)" automatically measures the contextual semantic similarity between the last utterance in the context and generated response.

As automatic metrics are widely criticized due to their lack of correlation with human judgments (Novikova et al., 2017; Reiter, 2018), we also employ two dialogue-level metrics by human evaluation: "Global-Coherence (G-Coh.)" manually examine whether the whole dialog is logical, fluent, and target-oriented. "Effectiveness (Effect.)" measures whether the target is efficiently reached. 100 dialogs are generated by simulation for each model. Three annotators score the generated dialogs rating in $\{0, 1, 2\}$, higher is better.

4.4 Results

Local Evaluation Table 3 shows the next-turn results on OTTers. GPT2 performs poorly on the one-turn transition. The multi-hop reasoning component added by the MultiGen and our TopKG leads to substantial improvements. TopKG achieves substantially better performance than all baselines on OTTers. This advantage indicates that TopKG successfully improves the performance of the existing local-model, which aims at the next-turn target.

Global Evaluation As shown in Table 4, on cases with "easy-to-reach" targets, end-to-end models such as GPT and MultiGen have fewer turns.

Start: music → dance	Easy Target: Travel
Context	**A:** My favorite type of music is from the south. **B:** Good ! I really like dancing, especially in the moonlight.
CKC	**A:** I watch sci fi. mainly, I read supernatural romances. **B:** Do you read ? My sister says fantasy novels are the best. **A:** I have a library in my house , I like to read books. **B:** I prefer historical story to romance. **A:** Cool! I love to wander and travel the world. **Path: romance → novel → book → story → travel**
TOPKG	**A:** I watch sci fi. mainly, I read supernatural romances. **B:** Reading fantasy novel and poetry are my true loves ! **A:** Awesome! I like adventures , been in sales all my life . **B:** I love to travel and see different places. Do you? **Path: romance → novel → adventure → travel**

Start: game → fun	Hard Target: Revolutionary
Context	**A:** Nice . i'm playing some card games with my family. **B:** That sounds like fun. i like wine and dancing too!
CKC	**A:** You sound like a fun person! **B:** What music do you like? **A:** Classical. how about you, what kind do you like? **B:** I like all music. ....(get trapped)
TOPKG	**A:** And what games are you into? **B:** Action games, do you? **A:** Not much into those. i like the rebel. **B:** Revolutionary. is that your favorite? **Path: game → action → rebel → revolutionary**

Table 6: Case study from self-play simulations on TGConv. In the left easy target case, TopKG generates responses similar in quality to CKC but plans a diverse path to the target. In the right hard target case, CKC gets trapped, but TopKG successfully reaches the target.

We notice that they tend to directly generate an utterance containing the target, despite that the utterances are of low quality in human evaluation. This may be due to that they are designed without global view. However, our TopKG has a higher success rate and higher efficiency in manual evaluation benefiting from the global planning.

In cases with "hard-to-reach" targets, GPT, which does not rely on KG, can also directly generate responses, and its performance is similar to that of "easy-to-reach" cases. For all KG-based methods, the performance significantly degrades on "hard-to-reach" targets, but our TopKG still exceeds all baselines. The ablation discussion below demonstrates the contribution of our global planning. Furthermore, our generated responses' average contextual "Semantic Similarity(Coh.)" is similar to the golden similarity in Table 2, which shows that our TopKG effectively learns the semantic patterns in the corpora. We also found that KG methods (CKC and TopKG) outperform the other models, which verifies the benefits of using KG in global target-oriented dialog.

4.5 Ablation Studies

We perform ablation studies for TopKG to better analyze the main components' relative contributions. The results are shown in Tables 3, 4, 5.

Does the global planning work? To prove the contribution of proposed global planning, we replace the global planning (w/o global plan) with a 2-hop neighbors graph (based on context entities), which results in the most significant performance drop in multi-turn evaluation. In contrast, the drop in the next-turn evaluation is not noticeable. The main reason is that the target often can be found in two-hop neighbors on the graph in a next-turn dialogue. This verifies the contribution of global KG planning to global target-oriented dialog.

How much graph information we need? We also explore the number of neighbors needed for initializing the G_{global} graph's nodes in two aspects (refer in Algorithm1): the maximum number of hops H, and the number of neighboring nodes in the h_{th} hop (denoted as K). Contrary to our expectations, expanding the average size of the knowledge graphs from 1000 nodes to 2000 did not improve the $hits@K$ ratio, as shown in the last row of Table 3. Therefore, the final version of TopKG adopts the global planning with K = 10, H = 3.

4.6 Case study

In the case study, we compare our TopKG with CKC, the most competitive baseline. In the left case of "Easy Target" in Table 6, TopKG and CKC followed different KG paths. In the first path followed by CKC, the novel indicates books, and the following two keywords are the topics of the books. In the second path followed by TopKG, the novel is an adjective, adventure is novel, and travel is one kind of adventure. In such easy cases, although the best existing method works well, TopKG can further explore diverse paths based on reinforcement learning. In the right case of "Hard Target" in Table 6, CKC gets trapped and fail to reach the goal. However, TopKG still successfully guides the dialog to the goal with effective global planning.

5 Conclusion and Future Work

We propose effectively guiding the target-oriented dialog towards a global target with global planning on KG. We first design a novel method to automatically select target-oriented samples from the chit-chat corpus by identifying KG reasoning paths throughout the dialog. We train a reinforcement learning model with a selected high-quality corpus that can guide a GPT-2 based response generation model to reach a global target word by global planning on ConceptNet. Automatic and human evaluations show that our method exceeds the baselines from both local and global views, and global planning provides a significant contribution. We will explore to balance the coherence and number of turns in global planning in future work.

6 Acknowledgements

This work was supported by National Natural Science Foundation of China (62272340, 61876128, 61876129, 61976154), State Key Laboratory of Communication Content Cognition (Grant No.A32003).

References

Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gasic. 2018. Multiwoz-a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5016–5026.

Fabio Clarizia, Francesco Colace, Marco Lombardi, Francesco Pascale, and Domenico Santaniello. 2018. Chatbot: An education support system for student. In *International Symposium on Cyberspace Safety and Security*, pages 291–302. Springer.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Emily Dinan, Varvara Logacheva, Valentin Malykh, Alexander Miller, Kurt Shuster, Jack Urbanek, Douwe Kiela, Arthur Szlam, Iulian Serban, Ryan Lowe, et al. 2020. The second conversational intelligence challenge (convai2). In *The NeurIPS'18 Competition*, pages 187–208. Springer.

Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in building intelligent open-domain dialog systems. *ACM Transactions on Information Systems (TOIS)*, 38(3):1–32.

Haozhe Ji, Pei Ke, Shaohan Huang, Furu Wei, Xiaoyan Zhu, and Minlie Huang. 2020. Language generation with multi-hop reasoning on commonsense knowledge graph. *arXiv preprint arXiv:2009.11692*.

Dongyeop Kang, Anusha Balakrishnan, Pararth Shah, Paul Crook, Y-Lan Boureau, and Jason Weston. 2019. Recommendation as a communication game: Self-supervised bot-play for goal-oriented dialogue. *arXiv preprint arXiv:1909.03922*.

Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Raymond Li, Samira Ebrahimi Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. Towards deep conversational recommendations. *Advances in neural information processing systems*, 31.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Wenchang Ma, Ryuichi Takanobu, and Minlie Huang. 2021. Cr-walker: Tree-structured graph reasoning and dialog acts for conversational recommendation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1839–1851. ACL.

Jekaterina Novikova, Ondřej Dušek, Amanda Cercas Curry, and Verena Rieser. 2017. Why we need new evaluation metrics for nlg. *arXiv preprint arXiv:1707.06875*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Matt Post. 2018. A call for clarity in reporting bleu scores. *arXiv preprint arXiv:1804.08771*.

Jinghui Qin, Zheng Ye, Jianheng Tang, and Xiaodan Liang. 2020. Dynamic knowledge routing network for target-guided open-domain conversation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34 Issue 05, pages 8657–8664.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

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*, pages 8689–8696.

Ehud Reiter. 2018. A structured review of the validity of bleu. *Computational Linguistics*, 44(3):393–401.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Karin Sevegnani, David M Howcroft, Ioannis Konstas, and Verena Rieser. 2021. Otters: One-turn topic transitions for open-domain dialogue. *arXiv preprint arXiv:2105.13710*.

Ashish Sharma, Adam S Miner, David C Atkins, and Tim Althoff. 2020. A computational approach to understanding empathy expressed in text-based mental health support. *arXiv preprint arXiv:2009.08441*.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Thirty-first AAAI conference on artificial intelligence*.

Jianheng Tang, Tiancheng Zhao, Chenyan Xiong, Xiaodan Liang, Eric Xing, and Zhiting Hu. 2019. Target-guided open-domain conversation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5624–5634.

Ramakrishna Vedantam, C Lawrence Zitnick, and Devi Parikh. 2015. Cider: Consensus-based image description evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4566–4575.

PeiFeng Wang, Jonathan Zamora, Junfeng Liu, Filip Ilievski, Muhao Chen, and Xiang Ren. 2021. Contextualized scene imagination for generative commonsense reasoning. *arXiv preprint arXiv:2112.06318*.

Zeqiu Wu, Michel Galley, Chris Brockett, Yizhe Zhang, Xiang Gao, Chris Quirk, Rik Koncel-Kedziorski, Jianfeng Gao, Hannaneh Hajishirzi, Mari Ostendorf, et al. 2020. A controllable model of grounded response generation. *arXiv preprint arXiv:2005.00613*.

Jun Xu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, Wanxiang Che, and Ting Liu. 2020a. Conversational graph grounded policy learning for open-domain conversation generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1835–1845.

Jun Xu, Haifeng Wang, Zhengyu Niu, Hua Wu, and Wanxiang Che. 2020b. Knowledge graph grounded goal planning for open-domain conversation generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34 Issue 05, pages 9338–9345.

Yunyi Yang, Yunhao Li, and Xiaojun Quan. 2020. Ubar: Towards fully end-to-end task-oriented dialog systems with gpt-2. *arXiv preprint arXiv:2012.03539*.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2019a. Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv preprint arXiv:1911.00536*.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2019b. Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv preprint arXiv:1911.00536*.

Xueliang Zhao, Wei Wu, Can Xu, Chongyang Tao, Dongyan Zhao, and Rui Yan. 2020. Knowledge-grounded dialogue generation with pre-trained language models. *arXiv preprint arXiv:2010.08824*.

Peixiang Zhong, Yong Liu, Hao Wang, and Chunyan Miao. 2021. Keyword-guided neural conversational model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35 Issue 16, pages 14568–14576.

Jinfeng Zhou, Bo Wang, Ruifang He, and Yuexian Hou. 2021. CRFR: Improving conversational recommender systems via flexible fragments reasoning on knowledge graphs. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4324–4334.

A Appendix

A.1 Discriminator

Discrimination measures the different between the positive/negative examples we extracted. In experiments, we verify the significant difference between the conversational behaviour of selected positive and negative examples by accurately classifying positive and negative examples by a discriminator.

	Accuracy
BERT-small	76.8%
BERT-base	78.1%

Table 7: The accuracy of the discriminator classification

According to the result of the discriminator and data analysis, we found that some negative samples may contain positive segments (e.g., the first few utterances in a negative sample can generate a short reasoning path), so false negative may be a problem in training the discriminator.

A.2 Fail cases

Here are some cases of TopKG failures. It can be seen that the failure can be roughly divided into two situations. One is that in the face of some uncommon words, even if there is a correct path, the generative model can only generate sentences with similar semantics. The other is that the path is very long and it is difficult to generate replies that contain relevant words.

Start: dye Hard Target: volley	
Context	**A:** hi i dye my hair 4 times a year. **B:** do you live in the sky ? i do . in the clouds.
TOPKG	**A:** i dye it dark brown **B:** do you also like the martant? it is my favorite. **A:** yes i like it because it is on the ground **B:** do you have a favorite field? ....(can not reached) **Path: dye-mordant-ground-volley**

Start: mowing Hard Target: ethicist	
Context	**A:** i also enjoy lawn mowing on sunny days . **B:** i enjoy green living and just put solar panels on roof
TOPKG	**A:** that's awesome. solar roofs are awesome. **B:** yes they're. are you good at math **A:** i am excellent at it. mornings mornings are like night. **B:** i like the night **A:** it is my favorite part **B:** what is your favorite part about it? ....(can not mention related words and can not reached) **Path: mowing-morning-forenoon--part-particular-ethicist**

Table 8: Some failure cases

A.3 Implementation Details

The default parameter settings across all experiments are as follows. We select conceptnet-5.7.0 and glove.6B.300d for initialization. The discount factor γ is 0.99 and the advantage discount factor is 0.95. Our policy reasoning model is trained for 200 epochs using Adam optimization with the learning rate of 5e-4 for actor network and 1e-3 for critic network with the batch size of 64. 256 action-state pairs to rollout for trajectory collection per epoch, 4 steps of gradient descent to perform on each batch, capacity of the replay buffer is 8.