

Dialogue Planning via Brownian Bridge Stochastic Process for Goal-directed Proactive Dialogue

Jian Wang*, Dongding Lin*, Wenjie Li

Department of Computing, The Hong Kong Polytechnic University
{jian-dylan.wang, dongding88.lin}@connect.polyu.hk
cswjli@comp.polyu.edu.hk

Abstract

Goal-directed dialogue systems aim to proactively reach a pre-determined target through multi-turn conversations. The key to achieving this task lies in planning dialogue paths that smoothly and coherently direct conversations towards the target. However, this is a challenging and under-explored task. In this work, we propose a coherent dialogue planning approach that uses a stochastic process to model the temporal dynamics of dialogue paths. We define a latent space that captures the coherence of goal-directed behavior using a Brownian bridge process, which allows us to incorporate user feedback flexibly in dialogue planning. Based on the derived latent trajectories, we generate dialogue paths explicitly using pre-trained language models. We finally employ these paths as natural language prompts to guide dialogue generation. Our experiments show that our approach generates more coherent utterances and achieves the goal with a higher success rate¹.

1 Introduction

Dialogue systems have made significant progress in generating high-quality responses for open-domain chitchat (Zhang et al., 2020; Roller et al., 2021) and assisting users in completing specific tasks (Madotto et al., 2018; Wu et al., 2019a). Instead of passively responding to users, dialogue systems can also take a proactive role to direct a conversation towards specific goals, such as introducing new and interesting topics (Wu et al., 2019b) or providing sociable recommendations on target items (Wang et al., 2022a). Such a proactive target-oriented or goal-directed dialogue system can guide conversations towards topics that the system knows how to discuss, making it promising to build autonomous conversational AI.

*Equal contribution.

¹Our code and data are available at <https://github.com/iwangjian/Color4Dial>.

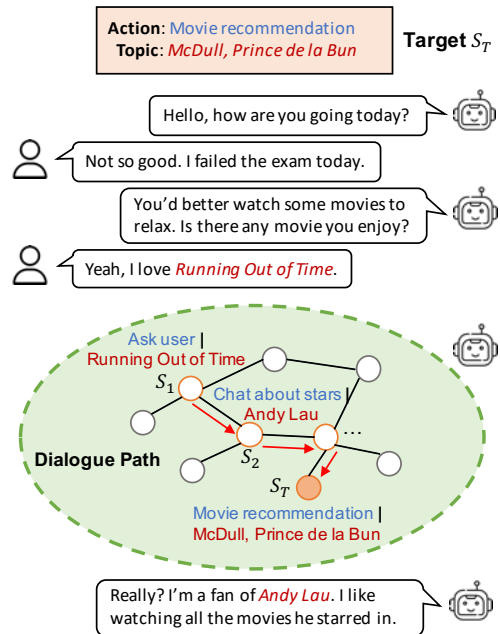


Figure 1: An example from the repurposed DuRecDial 2.0 (Liu et al., 2021b) dataset. Given a pre-determined target and current dialogue context, we expect to plan a dialogue path to direct the conversation.

For goal-directed dialogue systems, the objective is to proactively direct conversations towards a designated target. Previous work has primarily pre-determined the targets as specific keywords (Tang et al., 2019), topics (Wu et al., 2019b; Sevegnani et al., 2021), and dialogue action-topic pairs (Zhang et al., 2021; Wang et al., 2022a). To achieve this task, effective dialogue planning is essential, which requires taking reasonable actions and smoothly directing dialogue topics to the designated one. More importantly, the whole process is expected to be coherent and natural. Prior studies attempted to tackle this challenge through next-turn transition prediction (Tang et al., 2019), sub-goal generation (Zhang et al., 2021; Kishinami et al., 2022), and knowledge path reasoning (Gupta et al., 2022) to control dialogue generation. However, there are still open issues worth exploring. **First**, previous studies

adopted a greedy strategy with a single-turn topic prediction mechanism, which lacks global planning for the dialogue process (Yang et al., 2022). Consequently, these methods are often short-sighted, resulting in sub-coherent topic threads. **Second**, recognizing a user’s engagement and willingness to follow the system is crucial for achieving coherent transitions. However, current studies often overlook the importance of modeling such user feedback. Therefore, it is necessary to explore globally planned dialogue strategies while incorporating user feedback to improve the coherence of goal-directed dialogue systems.

In this work, our objective is to globally plan dialogue paths that connect the current context to the target at each turn. As illustrated in Figure 1, this dialogue path should strike a balance between coherence with the ongoing dialogue context and smooth transitions towards the target. Assuming that path trajectories without a target can be represented as Brownian motion (Revuz and Yor, 2013) in latent space, we expect the embeddings of neighboring trajectory points to be similar to each other, while those of distant trajectory points to be dissimilar. Drawing inspiration from Wang et al. (2022b), we view goal-directed dialogue behavior as a Brownian bridge (Revuz and Yor, 2013) stochastic process conditioned on fixed start and end points. As such, we can derive latent trajectories that follow coherent temporal dynamics.

Based on the above intuition, we propose a **coherent dialogue planning** approach via **Brownian bridge (COLOR)** stochastic process. It involves mapping dialogue path points, such as topics or action-topic pairs, into a latent space of Brownian bridge conditioned on the current context and designated target. To ensure goal-directed behavior and incorporate user feedback, we also map the latest user utterance into real-time user feedback representation using the same latent space. We leverage this feedback to perturb the density and uncertainty of the Brownian bridge, simulating its impact on the dialogue planning process. Our training process uses a contrastive objective, which helps retain global coherence. We then fine-tune pre-trained language models (PLMs) using the derived latent trajectories to plan dialogue paths explicitly. These paths provide step-by-step explanations for reaching the target and serve as natural language prompts for generating system utterances.

In summary, our main contributions are: (1) We

propose a novel approach called COLOR, which effectively models global coherence and incorporates user feedback in goal-directed dialogue planning. Our method utilizes the Brownian bridge stochastic process, and to the best of our knowledge, this is the first work to apply this method to the goal-directed proactive dialogue task. (2) We repurpose existing dialogue datasets by automatically constructing system goals and splitting them into in- and out-of-domain test sets. It facilitates research in the field and allows for more accurate evaluation of models. (3) Extensive experiments demonstrate that our proposed approach outperforms other methods, both in automatic and human evaluations.

2 Preliminaries

Problem Formulation We consider a corpus of goal-directed dialogues $\mathcal{D} = \{(\mathcal{K}_i, \mathcal{P}_i, \mathcal{C}_i)\}_{i=1}^N$, where N is the total number of dialogues. The domain knowledge facts relevant to the i -th dialogue are represented as $\mathcal{K}_i = \{k_{i,j}\}_{j=1}^{N_K}$, each in the form of a triple. The dialogue content for the i -th dialogue is $\mathcal{C}_i = \{c_{i,t}\}_{t=1}^{N_T}$, with a total of N_T turns. The whole dialogue path for the i -th dialogue is denoted as $\mathcal{P}_i = \{p_{i,l}\}_{l=1}^L$, where each path point is a topic or an action-topic pair. Here, dialogue topics are mainly constructed based on the domain knowledge \mathcal{K}_i . In some scenarios, there also exists a user profile \mathcal{U}_i , which can be user attributes or certain personal preferences.

Given a target \mathcal{T} consisting of an action-topic pair or a topic only, a dialogue context \mathcal{C} , and a set of relevant domain knowledge \mathcal{K} (and a user profile \mathcal{U} , if any), our objective is to generate coherent utterances to reach the target \mathcal{T} when appropriate. The problem can be decomposed into two sub-tasks: (1) **dialogue planning**, which involves planning suitable actions and topics to lead the dialogue proactively with coherent transitions to the target, and (2) **dialogue generation**, which involves generating an appropriate utterance to achieve the planned action and topic at each turn.

Brownian Bridge The standard Wiener process or Brownian motion $W(t)$ has a normal distribution with mean 0 and variance t , i.e., $W(t) \sim \mathcal{N}(0, t)$. A Brownian bridge (Revuz and Yor, 2013) is a continuous-time stochastic process pinned at fixed start and end points, where its distribution $B(t)$ is given by:

$$B(t) = W(t) - \frac{t}{T}W(T) \quad (1)$$

where $t \in [0, T]$, T denotes the end time. Furthermore, the transition distribution of a Brownian bridge process from an initial point z_0 at $t = 0$ to an end point z_T at $t = T$ is:

$$p(z_t|z_0, z_T) \sim \mathcal{N}\left(\left(1 - \frac{t}{T}\right)z_0 + \frac{t}{T}z_T, \frac{t(T-t)}{T}\right) \quad (2)$$

It implies that a trajectory point z_t follows a noisy linear interpolation between z_0 and z_T , with z_t closer to z_0 at the start and closer to z_T at the end. The uncertainty is higher in the middle of the time interval and lower near the start and end points. The time-controlled nature of the Brownian bridge process has led to its application in various fields, such as trajectory simulation (Sousa et al., 2015) and language modeling (Wang et al., 2022b).

3 Method

We propose a coherent dialogue planning approach via **Brownian bridge (COLOR)** stochastic process to steer goal-directed dialogue generation. The intuition behind COLOR is to learn a mapping (see §3.1) in the Brownian bridge latent space that captures coherent temporal dynamics for planning dialogue paths. Each dialogue path consists of a sequence of topics or action-topic pairs, starting from the current context and leading to the target. We generate these paths explicitly (see §3.2) based on representations derived from the latent space, and use them to guide the generation of dialogue utterances (see §3.3).

3.1 Stage 1: Brownian Bridge Mapping

A Brownian bridge latent space involves a non-linear mapping that transforms observations into a low-dimensional latent space, using the Brownian bridge stochastic process. Our objective is to utilize this mapping to train an encoder \mathcal{F} , to convert raw dialogue paths into latent representations that retain global coherence, with the overview depicted in Figure 2. In the following sections, we will introduce two crucial aspects of our approach: user feedback modeling and contrastive training.

User Feedback Modeling Suppose we obtain the user feedback representation z_u and have an engagement indicator $\delta_u \in (0, 1)$, which reflects the user’s level of engagement and likelihood of following the system, we newly define the transition distribution of the Brownian bridge process between a start point z_{s_0} at $t = 0$ and end point

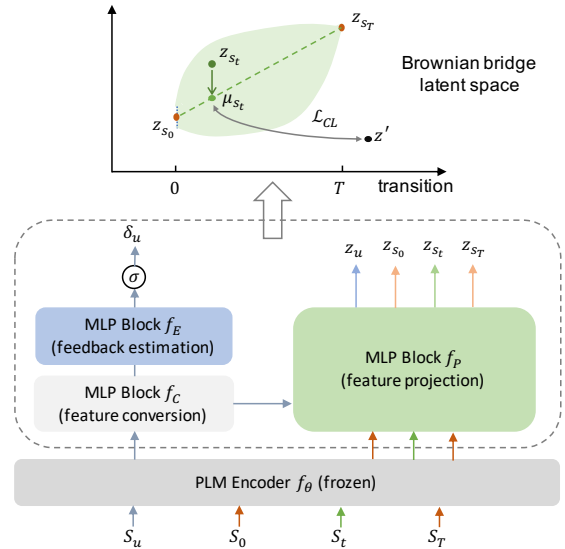


Figure 2: Overview of Stage 1: Mapping observations to Brownian bridge latents. S_u is the latest user utterance, S_0 is the concatenated text of domain knowledge and dialogue context, S_T is the designated target, S_t denotes a sampled path point in the dialogue path with $0 < t < T$. We differentiate data flow by colored arrows.

z_{s_T} at $t = T$ as:

$$p(z_{s_t}) \sim \mathcal{N}\left(\underbrace{\left(1 - \frac{t}{T}\right)(z_{s_0} + z_u)}_{\mu_{s_t}} + \underbrace{\frac{t}{T}z_{s_T}}_{\sigma^2}, \frac{t(T-t)}{T} + \varphi(\delta_u)\right) \quad (3)$$

where $0 < t < T$, $\varphi(\cdot)$ is a decaying function. Here, z_u is used to perturb the density (the mean μ_{s_t}) of the Brownian bridge process, and δ_u is used to perturb its uncertainty (the variance σ^2), with perturbation strength decaying over time. This decay means that the impact of the current user feedback on future planning is reduced. Alternatively, $\varphi(\cdot)$ can be implemented with the linear decaying, i.e., $\varphi(\delta_u) = \delta_u(1 - t/T)$, or the exponential decaying, i.e., $\varphi(\delta_u) = \delta_u e^{-t/(\lambda T)}$, where $\lambda \in (0, 1)$ is a scaling factor.

Contrastive Training For a tuple of observations (S_u, S_0, S_t, S_T) , our objective is to ensure that their latent representations $(z_u, z_{s_0}, z_{s_t}, z_{s_T})$ follow the Brownian bridge transition distribution described in Eq. (3). Here, S_u is the latest user utterance (and the concatenation of the user profile, if applicable), which may embody real-time user feedback information. S_0 consists of the concatenated domain knowledge and dialogue context, revealing the start of the dialogue path. S_T is the designated target, representing the end of the dialogue path. A *path point*, by default, refers to a topic or action-topic

pair specific to the dataset. S_t denotes a sampled path point in the dialogue path, *s.t.*, $0 < t < T$. Here, T denotes the number of transitions required to reach the target.

As shown in Figure 2, we build our encoder \mathcal{F} on top of a frozen PLM encoder, which is followed by specific trainable multilayer perceptron (MLP) blocks. All the necessary latents are given by:

$$z_{s_0} = f_P\left(\text{AvgPool}(f_\theta(S_0))\right), \quad (4)$$

$$z_{s_t} = f_P\left(\text{AvgPool}(f_\theta(S_t))\right), \quad (5)$$

$$z_{s_T} = f_P\left(\text{AvgPool}(f_\theta(S_T))\right), \quad (6)$$

$$z_u = f_P\left(f_C\left(\text{AvgPool}(f_\theta(S_u))\right)\right), \quad (7)$$

$$\delta_u = \sigma\left(f_E\left(f_C\left(\text{AvgPool}(f_\theta(S_u))\right)\right)\right) \quad (8)$$

where f_θ denotes a frozen PLM encoder such as BERT (Devlin et al., 2019) or BART (Lewis et al., 2020) encoder, $\text{AvgPool}(\cdot)$ denotes the average pooling operation. f_P , f_C , and f_E are MLP blocks that produce output with a latent dimension of d . σ is the Sigmoid activation function. The intuition behind the training is to ensure that the representation z_{s_t} of a positive path point S_t sampled from the same dialogue is close to the expected embedding μ_{s_t} (the mean in Eq. (3)). In contrast, the representation z'_t of a negative random path point S'_t from a different dialogue is far from μ_{s_t} (see Figure 2) because it does not align with the Brownian bridge pinned by z_{s_0} and z_{s_T} . We consider a contrastive objective proposed in Wang et al. (2022b) for training. Formally, given input batches $\mathcal{B} = \{(S_u, S_0, S_t, S_T)\}$ consisting of randomly sampled positive path points S_t where $0 < t < T$, we optimize our encoder \mathcal{F} as follows:

$$\mathcal{L}_{CL} = -\log \frac{\exp(d(S_t^+; \mathcal{F}))}{\sum_{S_t^- \in \mathcal{B}} \exp(d(S_t^-; \mathcal{F}))}, \quad (9)$$

$$d(S_t; \mathcal{F}) = -\frac{1}{2\sigma^2} \left\| z_{s_t} - \mu_{s_t} \right\|_2^2 \quad (10)$$

where S_t^+ denotes a positive tuple (S_u, S_0, S_t, S_T) , S_t^- denotes a negative tuple (S_u, S_0, S'_t, S_T) , σ^2 is the variance in Eq. (3), μ_{s_t} is the mean in Eq. (3).

3.2 Stage 2: Planning Dialogue Paths

The Brownian bridge latent space makes it easy to derive a coherent latent trajectory with temporal

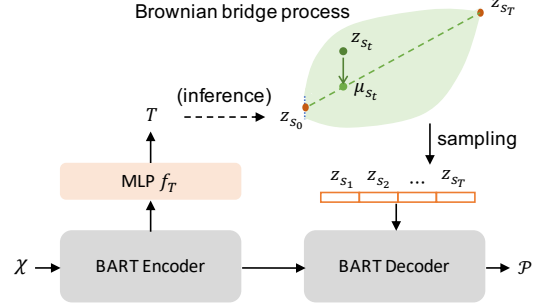


Figure 3: Overview of Stage 2: Planning the dialogue path \mathcal{P} , where \mathcal{X} is the required input, T denotes the number of transitions required to reach the target.

dynamics. We feed the start point S_0 , designated target S_T , and observed S_u , into the trained encoder \mathcal{F} respectively, then sample a latent trajectory $z = (z_{s_1}, z_{s_2}, \dots, z_{s_T})$ that follows Eq. (3), where $z_{s_t} \in \mathbb{R}^d$, $t = 1, 2, \dots, T$. Here, z acts like the transition-level latent representation that connects the ongoing dialogue context to the target, i.e., the dialogue path \mathcal{P} to be planned.

To generate the path \mathcal{P} , we define the required input as $\mathcal{X} = [\mathcal{C}; \mathcal{K}; \mathcal{T}]$, which is the concatenated text of the dialogue context \mathcal{C} , domain knowledge \mathcal{K} , and target \mathcal{T} . As shown in Figure 3, we feed \mathcal{X} into a pre-trained BART (Lewis et al., 2020) model for fine-tuning, with the encoded hidden states being $h = (h_1, h_2, \dots, h_m)$. We discuss the generation of \mathcal{P} by conditioning on h and z below.

First, sampling the latent trajectory z requires the value T , i.e., the number of transitions to reach the target. We obtain this value by adding an MLP layer f_T to the BART encoder as a predictor, which outputs the probability of T :

$$p(T) = \text{softmax}(W_1 f_T(\bar{h}) + b_1) \quad (11)$$

where \bar{h} is the average pooled representation of h , W_1 and b_1 are trainable parameters. We optimize the predictor using cross-entropy loss \mathcal{L}_c .

Second, our BART decoder conditions on h and the derived latent trajectory z , then generates the dialogue path \mathcal{P} with encoder-decoder attentions. The output distribution is approximated as follows:

$$p_\theta(\hat{y}_t) = \text{softmax}(W_2 h_t^o + b_2), \quad (12)$$

$$h_t^o = \text{Decoder}(y_{t-1}, H), \quad (13)$$

$$H = [h; W^T z] \quad (14)$$

where W_2 , b_2 are trainable parameters, W denotes a linear transformation that maps the dimension of z to be identical to h , and $[\cdot; \cdot]$ denotes concatenation.

The decoder is trained by minimizing the negative log-likelihood below:

$$\mathcal{L}_g = - \sum_{i=1}^N p(y^{(i)}) \log p_{\theta}(\hat{y}^{(i)}) \quad (15)$$

where $p(y^{(i)})$ is the distribution of the ground-truth dialogue path, while $p_{\theta}(\hat{y}^{(i)})$ is the distribution of the approximated dialogue path.

In addition, for the decoder’s all hidden states $h^o = (h_1^o, h_2^o, \dots, h_n^o)$ (see Eq. (13)) and the transformed latent trajectory $z^o = W^T z$ (see Eq. (14)), they inevitably both represent the dialogue path \mathcal{P} though at different levels. We minimize the Kullback–Leibler (KL) divergence between h^o and z^o :

$$\mathcal{L}_{KL} = \sum_{i=1}^N D_{KL}(\bar{h}^{o(i)} || \bar{z}^{o(i)}) \quad (16)$$

where \bar{h}^o and \bar{z}^o denote the average pooled representation of h^o and z^o , respectively.

For training, our model is optimized as follows:

$$\mathcal{L} = \alpha \mathcal{L}_c + \beta \mathcal{L}_g + \gamma \mathcal{L}_{KL} \quad (17)$$

where α , β , and γ are hyperparameters. During inference, we obtain the value T inferred by the predictor f_T , then sample a latent trajectory $z = (z_{s_1}, \dots, z_{s_T})$ given $t = 1, \dots, T$. The decoder then generates a dialogue path token by token. Additionally, no transition is needed to reach the target if $T = 0$. In such cases, we directly generate the dialogue path by copying the given target \mathcal{T} .

3.3 Stage 3: Generating Dialogue Utterances

Motivated by prior work on prompt-based learning for dialogue generation (Zheng and Huang, 2021; Madotto et al., 2021), we regard each dialogue path \mathcal{P} as a natural language prompt to guide a generative PLM for dialogue generation. Here, \mathcal{P} serves as a global prompt that outlines the dialogue actions and topics needed to reach the target step by step. With the power of the PLM, \mathcal{P} helps to distill the necessary knowledge from both the input text and the PLM. To formulate the newly input \mathcal{X}' , we append \mathcal{P} to the given dialogue context \mathcal{C} and domain knowledge \mathcal{K} , and concatenate them as:

$$\mathcal{X}' = [\mathcal{K}; \mathcal{C}; \mathcal{P}] \quad (18)$$

where $[\cdot]$ denotes concatenation. We then feed \mathcal{X}' into a pre-trained GPT-2 (Radford et al., 2019) or DialoGPT (Zhang et al., 2020) for supervised fine-tuning. We adopt the planned dialogue paths generated by our COLOR during inference.

Dataset		#Dial.	#Utter.	Dial. Turn	
				#Max.	#Avg.
DuRecDial 2.0	Train	4,256	68,781	13	8.1
	Valid	608	9,677	14	8.0
	Test-ID	770	12,299	13	8.0
	Test-OOD	446	7,962	12	8.9
TGConv	Train	15,197	70,205	9	3.8
	Valid	2,681	12,167	9	3.7
	Test	1,000	5,132	9	3.9

Table 1: Overview of the datasets.

4 Experiments and Results

4.1 Experimental Setup

Datasets The task of goal-directed proactive dialogue is under-explored, making it challenging to find feasible benchmarks for evaluation. After careful consideration, we have identified the DuRecDial 2.0 (Liu et al., 2021b) and TGConv (Yang et al., 2022) datasets as appropriate for our experiments. DuRecDial 2.0 (Liu et al., 2021b) is a crowdsourced dataset of human-to-human dialogues in recommendation-oriented scenarios. We repurpose the dataset by defining the targets as action-topic pairs. We obtain two types of splits for the test set: *in-domain* (ID) and *out-of-domain* (OOD), similar to Sevegnani et al. (2021). The OOD split ensures that none of the target topics in the test set are present in the training set, whereas the ID split allows them to appear. The TGConv (Yang et al., 2022) dataset contains high-quality open-domain dialogues on a variety of common-sense topics. Each dialogue is designed to direct the conversation towards a specific keyword or topic through coherent keyword transitions, which are categorized as either easy-to-reach or hard-to-reach based on their difficulty level. Table 1 summarizes the statistics of both datasets. More details are available in Appendix A.

Baseline Methods For dialogue generation, our baselines include: **GPT-2** (Radford et al., 2019), **DialoGPT** (Zhang et al., 2020), and **BART** (Lewis et al., 2020). On the repurposed DuRecDial 2.0 dataset, we also compared our method with three competitive methods: **MCGG_G** (Liu et al., 2020), **KERS** (Zhang et al., 2021), and **TCP-Dial** (Wang et al., 2022a). We chose these methods because they are highly relevant to our problem setting, and TCP-Dial is currently the state-of-the-art model in our knowledge. Given that our method is generalizable to the existing TGConv dataset, we evaluated its effectiveness against four competitive mod-

Split	Model	PPL (\downarrow)	F1 (%)	BLEU-1 / 2	DIST-1 / 2	Know. F1 (%)	Succ. (%)
ID	MGCG_G (Liu et al., 2020)	25.32	35.13	0.316 / 0.211	0.016 / 0.053	39.53	29.49
	KERS (Zhang et al., 2021)	20.15	31.27	0.288 / 0.196	0.017 / 0.061	41.18	33.75
	GPT-2 (Radford et al., 2019)	5.33	36.86	0.314 / 0.222	0.024 / 0.081	43.62	41.80
	DialoGPT (Zhang et al., 2020)	5.26	38.12	0.324 / 0.252	0.023 / 0.076	44.71	46.46
	BART (Lewis et al., 2020)	6.46	36.11	0.279 / 0.181	0.030 / 0.096	43.33	58.40
	TCP-Dial (Wang et al., 2022a)	5.88	34.46	0.293 / 0.201	0.027 / 0.091	45.75	60.49
	Ours (COLOR w/ GPT-2)	5.17	40.43*	0.337* / 0.243*	0.026 / 0.084	50.81*	69.14*
	Ours (COLOR w/ DialoGPT)	5.22	43.14*	0.371* / 0.277*	0.024 / 0.073	57.89*	73.20*
	MGCG_G (Liu et al., 2020)	28.21	30.84	0.276 / 0.167	0.015 / 0.046	20.53	8.46
	KERS (Zhang et al., 2021)	24.35	27.91	0.259 / 0.160	0.016 / 0.058	26.88	14.15
OOD	GPT-2 (Radford et al., 2019)	5.86	33.06	0.276 / 0.193	0.023 / 0.077	28.79	32.79
	DialoGPT (Zhang et al., 2020)	5.37	34.27	0.283 / 0.176	0.021 / 0.068	31.75	32.47
	BART (Lewis et al., 2020)	8.09	32.38	0.244 / 0.149	0.026 / 0.081	30.02	43.08
	TCP-Dial (Wang et al., 2022a)	8.24	29.24	0.255 / 0.165	0.027 / 0.089	21.36	18.40
	Ours (COLOR w/ GPT-2)	5.63	34.44*	0.285* / 0.198*	0.025 / 0.082	34.04*	57.41*
	Ours (COLOR w/ DialoGPT)	5.30	37.97*	0.320* / 0.227*	0.024 / 0.072	41.35*	52.36*

Table 2: Automatic local and global evaluation results of dialogue generation on the DuRecDial 2.0 dataset with different test splits. Significant improvements over backbone models are marked with * (t-test, $p < 0.05$).

els specific to that dataset: **MultiGen** (Ji et al., 2020), **DKRN** (Qin et al., 2020), **CKC** (Zhong et al., 2021), and **TopKG** (Yang et al., 2022). More details about the above methods are shown in Appendix B.1. For dialogue planning, we compared our COLOR with the planning models proposed in the above methods using a planning-enhanced paradigm. We also included **BERT** (Devlin et al., 2019) and **BART** (Lewis et al., 2020) as our baselines. More details about them are described in Appendix B.2.

Implementation Details Our proposed COLOR model is implemented by PyTorch. In both Stage 1 and Stage 2, we adopt the BART-base model (768 dimensions, 6 encoder/decoder layers, and 12 attention heads) released in Huggingface’s Transformers (Wolf et al., 2020) library. The latent dimension d is set to 16. The MLP blocks f_P , f_C , and f_E are all stacked to 3 layers. The decaying function $\varphi(\cdot)$ employs the linear decaying. The hyperparameters α , β and γ are set to 0.1, 1.0 and 1.0, respectively. For training in Stage 2, we construct the dialogue path \mathcal{P} in the format of $[A]a_1[T]t_1 \cdots [A]a_T[T]t_T$ on the DuRecDial 2.0, and of $[T]t_1 \cdots [T]t_T$ on the TGConv. Here, $[A]$ is a special token to separate an action a_i , $[T]$ is a special token to separate a topic t_i . During inference, we generate a dialogue path token by token. Further details on training and inference are provided in Appendix C.

4.2 Evaluation of Dialogue Generation

Evaluation Metrics To evaluate the performance of next-turn system utterance generation, we adopt

Model	Easy Target		Hard Target	
	Succ. (%)	Coh.	Succ. (%)	Coh.
GPT-2 [†] (G)	22.3	0.23	17.3	0.21
DialoGPT (D)	32.3	0.30	23.8	0.25
MultiGen [†]	26.7	0.21	19.6	0.24
DKRN [†]	38.6	0.33	21.7	0.31
CKC [†]	41.9	0.35	24.8	0.33
TopKG [†]	48.9	0.31	27.3	0.33
Ours (COLOR w/ G)	54.2	0.34	28.8	0.33
Ours (COLOR w/ D)	66.3	0.36	30.1	0.35

Table 3: Automatic global evaluation results of dialogue generation on the TGConv dataset. G and D are short for GPT-2 and DialoGPT, respectively. Models marked with [†] are reported from Yang et al. (2022).

commonly used local evaluation metrics, including perplexity (**PPL**), distinct (**DIST**) (Li et al., 2016), **BLEU- n** (Papineni et al., 2002), word-level **F1** and knowledge F1 (**Know. F1**) (Liu et al., 2020). To evaluate models’ goal-directed performance, we use the goal success rate (**Succ.**) as the global evaluation metric. In the repurposed DuRecDial 2.0 dataset, Succ. measures the proportion of correct target topic generation within the target turn and the two adjacent turns in the test set, as per Wang et al. (2022a). For the TGConv dataset, we perform self-play simulations, following Yang et al. (2022), to simulate multi-turn conversations and compute the success rate of generating the target keyword within 8 turns. Additionally, we adopt coherence (**Coh.**) (Yang et al., 2022) as another global evaluation metric, which measures the average contextual semantic similarity between the last utterance in the context and generated utterance.

Results and Discussion Table 2 shows evaluation results on the DuRecDial 2.0 dataset. We observe that MGCG_G and KERS achieve comparable performance to PLM-based models on the in-domain (ID) split. One main reason is that they use the predicted dialogue action and topic to guide the model in utterance generation. However, they perform poorly in terms of goal success rate due to a lack of dialogue-level planning. We note that BART and TCP-Dial obtain much better DIST-1/2 scores than others because they seldom generate repeated words, making the generated utterances more diverse. In comparison, our models achieve remarkable improvements over most evaluation metrics. For example, our COLOR with DialoGPT achieves much better knowledge F1 scores, indicating that our method is more likely to generate utterances with correct knowledge. Regarding the goal success rate, our models obtain a large margin of improvement on both ID and OOD splits. It shows that using prompts with appropriate dialogue paths effectively steers PLMs to generate proper utterances for goal-directed dialogue.

As shown in Table 3, we notice that directing a dialogue to reach the target seems challenging in the context of open-domain chitchat for all models. However, with the guidance of our dialogue planning approach, COLOR, our models are able to produce more coherent utterances and reach the target at a significantly higher success rate.

4.3 Evaluation of Dialogue Planning

Evaluation Metrics To evaluate the performance of dialogue planning, we first adopt **F1** to measure the micro-averaged precision and recall of the predicted action or topic. For generation-based models, we extract the action or topic at the evaluated turn from the generated dialogue path for a fair comparison. Due to the nature of dialogue, multiple temporary planning strategies can be reasonable before reaching the target. Following Zhou et al. (2020), we also expand gold labels by considering the system’s actions or topics within the previous and subsequent turns. As such, we then compute bigram action F1 (**Bi-act. F1**) and bigram topic F1 (**Bi-top. F1**) for evaluation.

Results and Discussion Table 4 reports the evaluation results on the DuRecDial 2.0 dataset. We find that predicting or generating dialogue topics is more challenging than dialogue actions. Further analysis reveals that the dialogue actions follow

Split	Model	Action		Topic	
		F1	Bi-act. F1	F1	Bi-top. F1
ID	MGCG	90.26	92.47	74.93	79.24
	KERS	90.33	91.54	77.85	80.35
	BERT	91.68	92.37	80.64	82.59
	TCP	92.25	93.82	85.77	87.25
	BART	95.40	96.31	90.96	92.21
	Ours (COLOR)	96.86	97.68	93.30	94.26
OOD	MGCG	82.30	87.25	36.03	42.00
	KERS	84.21	86.39	34.20	37.85
	BERT	92.23	94.19	46.55	52.12
	TCP	89.93	92.09	44.49	50.71
	BART	92.63	93.18	58.57	62.37
	Ours (COLOR)	93.43	93.82	79.09	83.46

Table 4: Results of dialogue planning on the DuRecDial 2.0 with different test splits.

Model	F1	Bi-top. F1
BERT (Devlin et al., 2019)	45.90	49.17
BART (Lewis et al., 2020)	43.20	47.69
TopKG-Plan (Yang et al., 2022)	46.06	48.04
Ours (COLOR)	47.17	52.85

Table 5: Results of dialogue planning on the TGConv.

a similar transition pattern in the dialogue paths, making it easier for all models to predict actions with an F1 score of over 80%. On the other hand, the variation in dialogue paths is primarily related to topics, which requires complex reasoning of domain knowledge, dialogue context, and target for accurate prediction. When evaluating on the OOD split, all baselines show lower F1 and Bi-top. F1 scores for topics. However, our proposed COLOR achieves substantial improvements. We observe similar trends in Table 5 when evaluating on the TGConv dataset. Overall, our COLOR outperforms the baselines by generating more reasonable actions and appropriate topics, making it a promising approach for planning dialogue paths.

Analysis of Model Variants We analyze the following variants of our model: (1) **COLOR_{d=?}**, which varies the value of the latent dimension d in $\{8, 32, 128\}$ (The d in our COLOR is set to 16 as described in §4.1); (2) **w/o** Brownian bridge (**BB**), which removes the operation of conditioning on the derived Brownian bridge latent representation z ; (3) **w/o** user feedback modeling (**UFM**), which removes z_u and $\varphi(\delta_u)$ in our Brownian bridge process as defined in Eq. (3); (4) **w/o** \mathcal{L}_{KL} , which means the model is trained without the loss \mathcal{L}_{KL} .

We report evaluation results on the OOD split of the DuRecDial 2.0 dataset, as shown in Table 6.

Model	Action		Topic	
	F1	Bi-act. F1	F1	Bi-top. F1
COLOR _{d=8}	93.21	93.73	79.21	83.30
COLOR _{d=32}	91.24	92.82	78.03	83.34
COLOR _{d=128}	93.57	94.30	78.67	82.89
COLOR	93.43	93.82	79.09	83.46
w/o BB	93.66	93.93	62.45	64.27
w/o UFM	92.42	92.84	77.21	80.57
w/o \mathcal{L}_{KL}	92.95	93.01	77.34	80.97

Table 6: Dialogue planning performance of our COLOR with different variants.

We observe that a larger value of d brings fewer performance gains. Hence, the d in our COLOR is set to 16 after making a trade-off between effectiveness and efficiency. We note that each module or mechanism of COLOR contributes to dialogue planning. In particular, the performance of COLOR sharply drops without the Brownian bridge (BB). It is because the derived Brownian bridge latent trajectory serves as a transition-level latent representation of the dialogue path to be planned. More importantly, it follows coherent temporal dynamics and thus benefits planning the dialogue path.

4.4 Human Evaluation

We recruit three well-educated graduate students as annotators for human evaluation. We ask the annotators to score different models based on turn-level and dialogue-level metrics, following Liu et al. (2020). The turn-level evaluation measures appropriateness (**Appr.**) and informativeness (**Info.**). The dialogue-level evaluation measures proactivity (**Proact.**), coherence (**Coh.**), and goal success (**Succ.**). More details on the metrics and evaluation procedure are described in Appendix D.

Table 7 shows human evaluation results on the DuRecDial 2.0 dataset. The Fleiss’s kappa (Fleiss, 1971) scores are mainly distributed between [0.41, 0.60], indicating moderate inter-annotator agreement. We observe that DialoGPT, TCP-Dial, and ours obtain comparable scores in informativeness since they all utilize powerful PLMs. However, our method is able to generate more appropriate utterances in response to dialogue context. For dialogue-level evaluation, our method obtains better results on average compared to all baseline models. Notably, our method achieves the highest coherence score and goal success rate, indicating that our method is more likely to direct the dialogue to reach the target coherently and successfully.

Model	Appr.	Info.	Proact.	Coh.	Succ.
MGCG_G	0.84	1.02	0.92	0.92	0.90
DialoGPT	1.17	1.35	1.06	1.17	1.19
TCP-Dial	1.20	1.24	1.26	1.20	1.02
Ours	1.33	1.40	1.42	1.35	1.38
kappa	0.48	0.52	0.46	0.56	0.53

Table 7: Human evaluation results. The Fleiss’s kappa measures the agreement among the annotators.

4.5 Case Study

To better analyze goal-directed dialogue generation, we show some cherry-picked cases in Appendix E due to space limitation. We observe that some baseline models can generate fluent and informative utterances. However, they still fail to direct the dialogue to reach the target and are ineffective to maintain coherence. In comparison, our COLOR model can plan a dialogue path with reasonable actions and appropriate topics that outlines how to reach the target step by step. With the guidance of the planned dialogue path, our system better knows when and what to talk about to proactively move the dialogue forward. More importantly, our method succeeds in achieving the goal (see Appendix E).

5 Related Work

Goal-directed Dialogue Generation In the goal-directed or target-oriented setting, existing studies mainly predetermine the targets as specific keywords (Tang et al., 2019; Qin et al., 2020; Zhong et al., 2021), topics (Wu et al., 2019b; Sevegnani et al., 2021; Lei et al., 2022), and dialogue action-topic pairs (Zhang et al., 2021; Wang et al., 2022a). The key to the task is dialogue planning, which leads the dialogue towards the target smoothly and coherently. Prior work pays attention to next-turn transition strategy (Tang et al., 2019), hierarchical policy (Xu et al., 2020a,b), and sub-goal generation (Zhang et al., 2021; Kishinami et al., 2022). For this knowledge-rich task, recent work (Gupta et al., 2022; Yang et al., 2022; Wang et al., 2022a) further concerns planning a dialogue path based on grounded knowledge to guide every turn of response generation.

Planning for Language Generation There is a line of work (Puduppully et al., 2019; Hua and Wang, 2019; Moryossef et al., 2019; Su et al., 2021) that separates text generation into content planning and surface realization. Content planning mainly concerns selecting key content (e.g., key entities)

and arranging their orders. Several planning frameworks (Hua et al., 2021; Hu et al., 2022; Li et al., 2022) have been studied to control complex language generation tasks. Our work is more related to planning for dialogue generation (Kishinami et al., 2022; Yang et al., 2022; Cohen et al., 2022). Our proposed COLOR is a novel dialogue-level planning method that steers dialogue generation.

6 Conclusion

In this work, we explore the task of goal-directed proactive dialogue and focus on planning dialogue paths that direct conversations towards the designated target. We propose a novel approach called COLOR, which models coherent temporal dynamics for dialogue paths in the defined latent space, and considers the impact of user feedback on the dialogue planning process. We employ the planned dialogue paths as prompts to steer dialogue generation. Experiments show that our proposed method outperforms other methods significantly.

Limitations

Though our proposed method exhibits superior performance, we also recognize its limitations and discuss potential solutions. Our proposed method for goal-directed dialogue generation suffers from error propagation since the three stages perform in a pipeline manner. After analyzing those generated utterances with low human evaluation scores, we find that the performance of dialogue generation is prone to drop when our COLOR model fails to plan an appropriate dialogue path. We intend to alleviate this issue by introducing some techniques in the cascaded generation, such as noisy channel models (Shannon, 1948; Liu et al., 2021a). In addition, other issues, such as how to make existing goal-directed dialogue systems more engaging and personalized, are worth further exploring.

Ethical Considerations

Goal-directed dialogue systems can be used for creating non-obtrusive recommendations for specific products and services, introducing interesting new topics and educating users about those topics, and so forth. Developing such systems requires careful consideration since it has a broad impact on applications. The intention of our work is not to force the system to reach the designated target nor force users to accept recommendations. Instead, we aim to build better assistive technologies to improve the

proactiveness of dialogue systems. Furthermore, our experimental datasets are publicly available. They have been filtered for sensitive and private information during dataset construction.

We hope to raise awareness of the potential for misuse of such systems with toxic intentions. For example, such systems may be used to pose as humans and actively manipulate users’ perceptions on specific issues or political inclinations. To mitigate these risks, we emphasize the importance of improving transparency through regulations. It is essential to inform users that they are conversing with a bot instead of a human, and regulations on target designation are crucial when deploying these systems in specific domains. It is necessary to ensure that setting a target does not violate factual accuracy, user privacy rules, or human laws.

Acknowledgments

This work was supported by the Research Grants Council of Hong Kong (15207122, 15207920, 15207821, 15204018) and National Natural Science Foundation of China (62076212). It was also supported in part by PolyU internal grants (ZVQ0, ZVVX).

References

- Deborah Cohen, Moonkyung Ryu, Yinlam Chow, Orgad Keller, Ido Greenberg, Avinatan Hassidim, Michael Fink, Yossi Matias, Idan Szpektor, Craig Boutilier, et al. 2022. Dynamic planning in open-ended dialogue using reinforcement learning. *arXiv preprint arXiv:2208.02294*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- 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.
- Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.
- Prakhar Gupta, Harsh Jhamtani, and Jeffrey Bigham. 2022. [Target-guided dialogue response generation](#)

- using commonsense and data augmentation. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1301–1317, Seattle, United States. Association for Computational Linguistics.
- Zhe Hu, Hou Pong Chan, Jiachen Liu, Xinyan Xiao, Hua Wu, and Lifu Huang. 2022. **PLANET: Dynamic content planning in autoregressive transformers for long-form text generation**. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2288–2305, Dublin, Ireland. Association for Computational Linguistics.
- Xinyu Hua, Ashwin Sreevatsa, and Lu Wang. 2021. **DYPLOC: Dynamic planning of content using mixed language models for text generation**. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6408–6423, Online. Association for Computational Linguistics.
- Xinyu Hua and Lu Wang. 2019. **Sentence-level content planning and style specification for neural text generation**. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 591–602, Hong Kong, China. Association for Computational Linguistics.
- Haozhe Ji, Pei Ke, Shaohan Huang, Furu Wei, Xiaoyan Zhu, and Minlie Huang. 2020. **Language generation with multi-hop reasoning on commonsense knowledge graph**. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 725–736, Online. Association for Computational Linguistics.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Yosuke Kishinami, Reina Akama, Shiki Sato, Ryoko Tokuhisa, Jun Suzuki, and Kentaro Inui. 2022. **Target-guided open-domain conversation planning**. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 660–668, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.
- Wenqiang Lei, Yao Zhang, Feifan Song, Hongru Liang, Jiaxin Mao, Jiancheng Lv, Zhenglu Yang, and Tat-Seng Chua. 2022. Interacting with non-cooperative user: A new paradigm for proactive dialogue policy. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 212–222.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. **BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.
- Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016. **A diversity-promoting objective function for neural conversation models**. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 110–119, San Diego, California. Association for Computational Linguistics.
- Qintong Li, Piji Li, Wei Bi, Zhaochun Ren, Yuxuan Lai, and Lingpeng Kong. 2022. **Event transition planning for open-ended text generation**. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 3412–3426, Dublin, Ireland. Association for Computational Linguistics.
- Qi Liu, Lei Yu, Laura Rimell, and Phil Blunsom. 2021a. **Pretraining the noisy channel model for task-oriented dialogue**. *Transactions of the Association for Computational Linguistics*, 9:657–674.
- Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, and Wanxiang Che. 2021b. **DuRecDial 2.0: A bilingual parallel corpus for conversational recommendation**. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 4335–4347, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, Wanxiang Che, and Ting Liu. 2020. **Towards conversational recommendation over multi-type dialogs**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1036–1049, Online. Association for Computational Linguistics.
- Andrea Madotto, Zhaojiang Lin, Genta Indra Winata, and Pascale Fung. 2021. Few-shot bot: Prompt-based learning for dialogue systems. *arXiv preprint arXiv:2110.08118*.
- Andrea Madotto, Chien-Sheng Wu, and Pascale Fung. 2018. **Mem2Seq: Effectively incorporating knowledge bases into end-to-end task-oriented dialog systems**. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1468–1478, Melbourne, Australia. Association for Computational Linguistics.
- Amit Moryossef, Yoav Goldberg, and Ido Dagan. 2019. **Step-by-step: Separating planning from realization in neural data-to-text generation**. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2267–2277, Minneapolis, Minnesota. Association for Computational Linguistics.

- Kishore Papineni, Salim Roukos, Todd Ward, and Weijing 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, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.
- Ratish Puduppully, Li Dong, and Mirella Lapata. 2019. Data-to-text generation with content selection and planning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 6908–6915.
- 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*, 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.
- Daniel Revuz and Marc Yor. 2013. Continuous martingales and brownian motion. *Springer Science & Business Media*, 293.
- Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. **Recipes for building an open-domain chatbot**. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325, Online. Association for Computational Linguistics.
- Karin Sevegnani, David M. Howcroft, Ioannis Konstas, and Verena Rieser. 2021. **OTTers: One-turn topic transitions for open-domain dialogue**. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2492–2504, Online. Association for Computational Linguistics.
- Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.
- J Belezza Sousa, Manuel L Esquivel, and RM Gaspar. 2015. Brownian bridge and other path-dependent gaussian processes vectorial simulation. *Communications in Statistics-Simulation and Computation*, 44(10):2608–2621.
- 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*.
- Yixuan Su, David Vandyke, Sihui Wang, Yimai Fang, and Nigel Collier. 2021. **Plan-then-generate: Controlled data-to-text generation via planning**. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 895–909, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- 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, Florence, Italy. Association for Computational Linguistics.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
- Jian Wang, Dongding Lin, and Wenjie Li. 2022a. Follow me: Conversation planning for target-driven recommendation dialogue systems. *arXiv preprint arXiv:2208.03516*.
- Rose E Wang, Esin Durmus, Noah Goodman, and Tatsunori Hashimoto. 2022b. Language modeling via stochastic processes. In *International Conference on Learning Representations (ICLR)*.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP): System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.
- Chien-sheng Wu, Richard Socher, and Caiming Xiong. 2019a. Global-to-local memory pointer networks for task-oriented dialogue. In *7th International Conference on Learning Representations (ICLR)*.
- Wenquan Wu, Zhen Guo, Xiangyang Zhou, Hua Wu, Xiyuan Zhang, Rongzhong Lian, and Haifeng Wang. 2019b. **Proactive human-machine conversation with explicit conversation goal**. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3794–3804, Florence, Italy. Association for Computational Linguistics.
- 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, Online. Association for Computational Linguistics.
- 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*, 05, pages 9338–9345.

Zhitong Yang, Bo Wang, Jinfeng Zhou, Yue Tan, Dongming Zhao, Kun Huang, Ruifang He, and Yuexian Hou. 2022. **TopKG: Target-oriented dialog via global planning on knowledge graph**. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 745–755, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Jun Zhang, Yan Yang, Chencai Chen, Liang He, and Zhou Yu. 2021. **KERS: A knowledge-enhanced framework for recommendation dialog systems with multiple subgoals**. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1092–1101, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2020. **DIALOGPT: Large-scale generative pre-training for conversational response generation**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 270–278, Online. Association for Computational Linguistics.

Chujie Zheng and Minlie Huang. 2021. Exploring prompt-based few-shot learning for grounded dialog generation. *arXiv preprint arXiv:2109.06513*.

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

Yiheng Zhou, Yulia Tsvetkov, Alan W Black, and Zhou Yu. 2020. Augmenting non-collaborative dialog systems with explicit semantic and strategic dialog history. In *International Conference on Learning Representations (ICLR)*.

A Dataset Descriptions and Pre-processing

DuRecDial 2.0 The DuRecDial 2.0 (Liu et al., 2021b) dataset is collected from crowdsourced human-to-human dialogues. In each dialogue, one person is defined as the seeker (the user’s role) and the other as the recommender (the system’s role). The recommender needs to proactively lead the dialogue and make recommendations by introducing new topics. Each seeker is equipped with a user profile containing user attributes (e.g., age range) and his/her past preference information. In order to smoothly converse with the seeker, the recommender has a domain knowledge graph consisting of domain-specific topics (e.g., movies, music) with related attributes. More importantly, a dialogue path composed of dialogue actions and topics is annotated for the recommender from the beginning to the end of the dialogue. All dialogues are

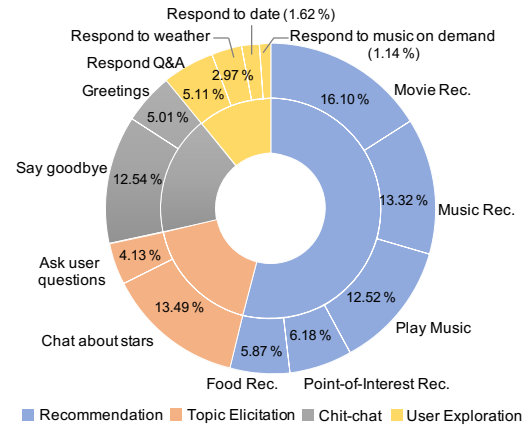


Figure 4: Statistics of the system’s dialogue actions on the repurposed DuRecDial 2.0 dataset.

aligned across the English and Chinese languages. We adopt the dataset in English for experiments.

Since there are no explicit annotated targets, we repurpose the original dataset automatically. For all those dialogues that are proactively led by the system, we treat the topic that the user has accepted at the end of each dialogue as the target topic, and view the system’s corresponding action (e.g., movie recommendation, point-of-interest recommendation, etc.) as the target action. Each target topic is guaranteed to be grounded in the domain knowledge triples corresponding to the dialogue. We filter out those dialogues without introducing any new recommendation topics. The total number of topics is 628 (including a NULL topic). Figure 4 shows the statistics of all the system’s actions. We observe an average of 4.3 ~ 4.8 action-topic transitions from the beginning to reaching the target.

Following the splitting criterion (Liu et al., 2021b), we obtain training/validation/test sets with 4,256/608/1,216 dialogues, respectively. To investigate the performance of different methods for goal-directed dialogue generation, we further use the dataset with two types of splits for the test set: *in-domain (ID)* split and *out-of-domain (OOD)* split, similar to Sevegnani et al. (2021); Gupta et al. (2022). The OOD split ensures that none of the target topics in the test set are present in the training set. In contrast, the target topics in the ID split are allowed to appear in the training set.

TGConv The TGConv (Yang et al., 2022) dataset is extracted based on the chit-chat corpus ConvAI2 (Dinan et al., 2020), and the external commonsense KG ConceptNet (Speer et al., 2017). In the TGConv dataset, all target-oriented sam-

ples are identified by the dialogue utterances containing a go-through keyword/concept sequence that aligns with the KG path over the ConceptNet. Suppose the designated global target keyword is w_n , a transition path of keywords or concepts $\mathcal{P} = \{w_1 \rightarrow \dots \rightarrow w_n\}$ is annotated for each dialogue. Here, each neighbor word pair (i.e., w_i and w_{i+1}) is direct or low-order connected in the ConceptNet. On average, the number of transitions from the start context to the target is approximately 5. Furthermore, the target keywords are distinguished into “*easy-to-reach*” and “*hard-to-reach*”. Specifically, the *easy-to-reach* targets refer to target keywords with high frequency in the corpus. In comparison, target words with low frequency (less than 800) in the corpus are classified as *hard-to-reach* targets because there are fewer cases to learn the transition to low-frequency target words. In this work, we follow the same data splitting as in Yang et al. (2022) for experiments.

B Baseline Methods

B.1 Dialogue Generation

To evaluate dialogue generation quality, we first consider the following PLMs-based methods:

- **GPT-2** (Radford et al., 2019): It is an autoregressive generation model for language generation. We use the GPT-2 base² model for fine-tuning.
- **DialoGPT** (Zhang et al., 2020): It is an autoregressive dialogue generation model pre-trained using large-scale dialogue corpora. We adopt the pre-trained small³ model for fine-tuning.
- **BART** (Lewis et al., 2020): It is a denoising encoder-decoder model for language generation. We use the BART-base⁴ model for fine-tuning.

Note that these models concatenate all parts of input texts described in the problem formulation as the model input and are fine-tuned to generate utterances directly.

On the DuRecDial 2.0 dataset, we additionally consider several competitive models that follow the planning-enhanced generation paradigm:

- **MGCG_G** (Liu et al., 2020): It employs the predicted next dialogue action and next topic to

²<https://huggingface.co/gpt2>

³<https://huggingface.co/microsoft/DialoGPT-small>

⁴<https://huggingface.co/facebook/bart-base>

guide utterance generation. We re-run the officially released code⁵ on the repurposed dataset.

- **KERS** (Zhang et al., 2021): It leverages a knowledge-enhanced mechanism to guide dialogue generation. We re-run the available code⁶ on the repurposed dataset.
- **TCP-Dial** (Wang et al., 2022a): It builds a target-driven conversation planning method to explicitly extract necessary knowledge and then guides dialogue generation. We re-run the available code⁷ on the repurposed dataset.

On the TGConv dataset, we consider the following competitive models:

- **MultiGen** (Ji et al., 2020): It is a language generation model with multi-hop reasoning on commonsense knowledge graphs.
- **DKRN** (Qin et al., 2020): It builds a dynamic knowledge routing network for topic transitions.
- **CKC** (Zhong et al., 2021): It is a keyword-guided neural conversational model that leverages ConceptNet for keyword transitions.
- **TopKG** (Yang et al., 2022): It employs global planning on ConceptNet to guide dialogue generation and is the state-of-the-art approach⁸ on the TGConv dataset.

B.2 Dialogue Planning

To compare the performance of dialogue planning for goal-directed dialogues, we consider the following dialogue planning methods:

- **MGCG** (Liu et al., 2020): It makes multi-task predictions for the next-turn’s dialogue action and topic. However, it assumes that ground-truth historical dialogue actions and topics are known for a system. For a fair comparison in this work, we adopt the same input as our problem definition to conduct multi-task predictions.
- **KERS** (Zhang et al., 2021): It generates the next-turn’s dialogue action and topic based on a Transformer (Vaswani et al., 2017) network.

⁵<https://github.com/PaddlePaddle/Research/tree/master/NLP/ACL2020-DuRecDial>

⁶<https://github.com/z562/KERS>

⁷<https://github.com/iwangjian/Plan4RecDial>

⁸<https://github.com/yyyyyyzt/topkgchat>

- **TCP** (Wang et al., 2022a): It is a target-driven planning framework that plans a dialogue path consisting of dialogue actions and topics in a generation-based manner.
- **TopKG-Plan** (Yang et al., 2022): It employs reinforcement learning to plan a commonsense keyword path based on ConceptNet.
- **BERT** (Devlin et al., 2019): Based on the intuition of multi-task predictions, we fine-tune the widely-used BERT model by adding two fully-connected layers to jointly predict the next-turn’s dialogue action and topic. We use the uncased BERT-base⁹ model for fine-tuning.
- **BART** (Lewis et al., 2020): Based on the intuition of generation, we use the BART-base model for fine-tuning, which is then used to generate a dialogue path similar to ours.

C Training and Inference Details

In Stage 1, we set the batch size for contrastive training to 64 and adopt the Adam (Kingma and Ba, 2014) optimizer with a learning rate of $2e-4$. We train our encoder \mathcal{F} for 10 epochs. For training in Stage 2, we adopt the Adam optimizer with an initial learning rate of $2e-5$ and warm up over the first 10% training steps. We train our COLOR for a maximum of 10 epochs with a batch size of 16. The best checkpoint is chosen based on its performance on the validation set. For inference, we employ greedy decoding to generate a dialogue path token by token, with a maximum decoding length of 80. In Stage 3, we employ GPT-2 base and DialoGPT-small (see Appendix B.1) as our backbone models. We follow the description in §3.3 and fine-tune backbone models for 10 epochs. For a fair comparison, we use greedy decoding with a maximum decoding length of 100 for all models. We conduct experiments on one NVIDIA 3090 GPU machine.

D Procedure of Human Evaluation

For turn-level evaluation, we randomly sampled 50 dialogues from the ID test split and 50 dialogues from the OOD test split from the DuRecDial 2.0 dataset. We then compared the generated utterances of the following models: MGCG_G, DialoGPT, TCP-Dial, and ours (COLOR w/ DialoGPT). For a fair comparison, the models were randomly renamed as “model-1”, “model-2”, and

so forth. The annotators were then asked to mark scores for the compared models from (1) appropriateness (**Appr.**), which measures whether the utterance responds to the dialogue context appropriately, and (2) informativeness (**Info.**), which measures whether the utterance is informative by making full use of the grounded knowledge.

For dialogue-level evaluation, we asked our annotators to act as users and converse with the models. Each model’s generated utterance in the current turn was used as part of the dialogue context in the next turn. Our annotators were asked to maintain consistency with the equipped user profile. To ensure diverse evaluation targets, we randomly selected 5 target actions from the DuRecDial 2.0 test set, each paired with 10 different target topics, resulting in a total of 50 targets evaluated. We did not expose the targets to them during human-model conversations and restricted all conversations to no more than 12 turns. We finally released the designated targets to the annotators and asked them to score the models based on three evaluation metrics: (1) proactivity (**Proact.**), which measures whether a model proactively leads the dialogue; (2) coherence (**Coh.**), which manually examines whether the whole dialogue is fluent, coherent, and smoothly transited; (3) goal success (**Succ.**), which estimates whether a model effectively reaches the target.

Our annotators were required to score the generated dialogues rating in $\{0, 1, 2\}$, where higher is better. The agreement among the annotators is measured by Fleiss’s kappa (Fleiss, 1971). We reported each model’s average score from different annotators as the ultimate human evaluation result. In addition, we transparently informed all annotators of our research intent. We paid reasonable wages and provided enough time for the annotators to complete the evaluation.

E Case Study

Table 8 and Table 9 show some cases on the DuRecDial 2.0 and TGConv datasets, respectively.

⁹<https://huggingface.co/bert-base-uncased>

Target: <u>Movie recommendation</u> (target action) <u>The Art of Action: Martial Arts in Motion</u> (target topic)		
Model	Example	Dialogue Planning
MGCG_G	User: Do you know who starred in the movie <u>Jackie Chan: My Story</u> ?	
	Bot: of course. it's jackie chan.	Respond Q&A Jackie Chan: My Story
	User: Well, I see. You are so brilliant!	
	Bot: you're welcome. <u>let's talk about movies. let's talk about movies.</u> (not fluent and coherent)	Respond Q&A Jackie Chan: My Story
	User: OK. What would you like to talk about?	
	Bot: jackie chan's movies are very good.	Chat about stars Jackie Chan
	User: Yeah, that's why I love him so much.	
DialoGPT	Bot: i recommend you to see <u>rumble in the bronx</u> . it's very good. it's a very good movie.	<u>Movie recommendation</u> <u>Rumble in the Bronx</u>
	User: Yeah. Of course.	
	Bot: you have a good taste. <u>the rumble in the bronx</u> is recommended to you.	<u>Movie recommendation</u> <u>Rumble in the Bronx</u>
	... (not reach the target)	
	User: Do you know who starred in the movie <u>Jackie Chan: My Story</u> ?	
	Bot: Yes, it's <u>Jackie Chan</u> .	
	User: Well, I see. You are so brilliant!	
Bot: You're welcome. Speaking of <u>Jackie Chan</u> , he is really a very good actor.		
User: Yeah. I sincerely love him!		
Bot: Yes, he has won the <u>Golden Horse Award for Best Actor</u> .	(no dialogue planning)	
User: Wow, that's great.		
Bot: He also won the <u>Hong Kong Film Award for Best Actor</u> .		
User: Anything else?		
Bot: Well, it's a <u>documentary movie</u> . It's worth to recommend to you. Don't miss it! (not coherent with context)		
...	(get trapped)	
Ours	User: Do you know who starred in the movie <u>Jackie Chan: My Story</u> ?	\mathcal{P}_1 : [A] Respond Q&A [T] Jackie Chan: My Story [A] Chat about stars [T] Jackie Chan [A] Movie recommendation [T] The Art of Action: Martial Arts in Motion Picture
	Bot: Yes, it's <u>Jackie Chan</u> .	\mathcal{P}_2 : [A] Chat about stars [T] Jackie Chan [A] Movie recommendation [T] The Art of Action: Martial Arts in Motion Picture
	User: Well, I see. You are so brilliant!	
	Bot: You're welcome. <u>Jackie Chan</u> is a Chinese martial arts superstar.	
	User: Yeah. I sincerely love him!	
	Bot: Well, he has <u>won the Golden Horse Award for Best Actor</u> . He has many good movies. (more coherent)	\mathcal{P}_3 : (same as \mathcal{P}_2)
	User: Wow, that's great.	
	Bot: I'd recommend his movie <u>The Art of Action: Martial Arts in Motion Picture</u> to you.	\mathcal{P}_4 : [A] <u>Movie recommendation</u> [T] <u>The Art of Action: Martial Arts in Motion Picture</u>
	User: Sounds good.	
	Bot: It's a <u>documentary movie</u> expressing Hong Kong kung fu movies. Don't miss it. (succeed to reach the target)	\mathcal{P}_5 : (same as \mathcal{P}_4)

Table 8: Generated cases from dialogue-level human evaluations on the DuRecDial 2.0 dataset. The bot's utterances are generated by different models, including (a) MGCG_G, (b) DialoGPT, and (c) Ours (COLOR w/ DialoGPT). The topics and topic-related attributes that also appear in the domain knowledge are marked with underlines.

Start Keywords: wear → shoes Easy Target: computer		
Context	A: it is freezing outside, i enjoy wearing skinny jeans. B: i am often gunning for new shoes lol.	
Model	Example	Dialogue Planning
TopKG	A: i love tee shirts with rock band logos. B: that sounds more cozy. what is your favorite band? A: i'm young but i love old stuff, beatles. B: that is nice. A: it's cool! i like working on the computer (succeed to reach the target)	shirt → band → stuff → computer
Ours	A: oh, me too. my favorite shirts are flannel ruffle sleeves. B: really that's sick love them, why do you like that? A: i enjoy playing video games and listening to music. B: is that game rolls royce's ghost tho am i right. A: yeah, it is the wallpaper for my computer. (succeed to reach the target more coherently)	\mathcal{P}_1 : [T] shirt [T] band [T] music [T] computer \mathcal{P}_2 : [T] like [T] music [T] computer \mathcal{P}_3 : [T] play [T] game [T] computer \mathcal{P}_4 : [T] game [T] computer \mathcal{P}_5 : [T] computer
Start Keywords: day → living Hard Target: barbershop		
Context	A: i enjoy lawn mowing on sunny days. B: i enjoy green living and just put solar panels on roof.	
Model	Example	Dialogue Planning
TopKG	A: i do live rpg events over the weekends. B: wow, are you living in game or like that? A: i gonna make it a big one day. B: amazing dream, huh? ... (get trapped)	event → game → day → dream → ...
Ours	A: cool. do you only live with your parents? B: yes i do. i also have two dogs myself. A: ha, i love dogs. but my italian wife doesn't. B: not easy. dogs must spend a lot of time outdoors. A: lol, but she love going to the pet barbershop for pictures. (succeed to reach the target)	\mathcal{P}_1 : [T] live [T] family [T] outdoor [T] barbershop \mathcal{P}_2 : [T] dog [T] room [T] outdoor [T] barbershop \mathcal{P}_3 : [T] wife [T] outdoor [T] barbershop \mathcal{P}_4 : [T] outdoor [T] barbershop \mathcal{P}_5 : [T] outdoor [T] barbershop

Table 9: Generated cases from self-play simulations on the TGConv dataset. The compared models include (a) TopKG and (b) Ours (COLOR w/ DialoGPT).

ACL 2023 Responsible NLP Checklist

A For every submission:

- A1. Did you describe the limitations of your work?
Section "Limitations".
- A2. Did you discuss any potential risks of your work?
Section "Ethical Considerations".
- A3. Do the abstract and introduction summarize the paper's main claims?
Section "Abstract" and Section 1 "Introduction".
- A4. Have you used AI writing assistants when working on this paper?
Left blank.

B Did you use or create scientific artifacts?

Section 4.1.

- B1. Did you cite the creators of artifacts you used?
Section 4.1; Appendices A, B.1-B.2, C.
- B2. Did you discuss the license or terms for use and / or distribution of any artifacts?
We will include the license and terms of use when releasing our code and data.
- B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
Appendix A.
- B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?
Appendix A.
- B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Appendix A.
- B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.
Section 4.1 and Appendix A.

C Did you run computational experiments?

Section 4.

- C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
Appendix C.

The Responsible NLP Checklist used at ACL 2023 is adopted from NAACL 2022, with the addition of a question on AI writing assistance.

- C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Appendix C.
 - C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
Section 4.2.
 - C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?
Appendix B.1-B.2.
- D** **Did you use human annotators (e.g., crowdworkers) or research with human participants?**
Section 4.4.
- D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
Section 4.4. and Appendix D.
 - D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
Section 4.4. and Appendix D.
 - D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?
Section 4.4. and Appendix D.
 - D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
Not applicable. Left blank.
 - D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
Not applicable. Left blank.