Target-constrained Bidirectional Planning for Generation of Target-oriented Proactive Dialogue

JIAN WANG, The Hong Kong Polytechnic University, Hong Kong, China
DONGDING LIN, The Hong Kong Polytechnic University, Hong Kong, China
WENJIE LI, The Hong Kong Polytechnic University, Hong Kong, China

Target-oriented proactive dialogue systems aim at leading conversations from a dialogue context toward a pre-determined target, such as making recommendations on designated items or introducing new specific topics. To this end, it is critical for such dialogue systems to plan reasonable actions to drive the conversation proactively, and meanwhile, to plan appropriate topics to move the conversation forward to the target topic smoothly. In this work, we mainly focus on effective dialogue planning for target-oriented dialogue generation. Inspired by decision-making theories in cognitive science, we propose a novel target-constrained bidirectional planning (TRIP) approach, which plans an appropriate dialogue path by looking ahead and looking back. By formulating the planning as a generation task, our TRIP bidirectionally generates a dialogue path consisting of a sequence of <action, topic> pairs using two Transformer decoders. They are expected to supervise each other and converge on consistent actions and topics by minimizing the decision gap and contrastive generation of targets. Moreover, we propose a target-constrained decoding algorithm with a bidirectional agreement to better control the planning process. Subsequently, we adopt the planned dialogue paths to guide dialogue generation in a pipeline manner, where we explore two variants: prompt-based generation and plan-controlled generation. Extensive experiments are conducted on two challenging dialogue datasets, which are re-purposed for exploring target-oriented dialogue. Our automatic and human evaluations demonstrate that the proposed methods significantly outperform various baseline models.

CCS Concepts: • **Computing methodologies** → **Discourse, dialogue and pragmatics**; **Natural language generation;**

Additional Key Words and Phrases: Target-oriented dialogue, dialogue generation, bidirectional planning

ACM Reference Format:
Jian Wang, Dongding Lin, and Wenjie Li. 2024. Target-constrained Bidirectional Planning for Generation of Target-oriented Proactive Dialogue. *ACM Trans. Inf. Syst.* 42, 5, Article 124 (April 2024), 27 pages.
https://doi.org/10.1145/3652598

1 INTRODUCTION

Human-machine dialogue systems have made significant progress in chatting with users for entertainment, e.g., open-domain dialogues [57, 58, 70], and assisting users in accomplishing specific

---

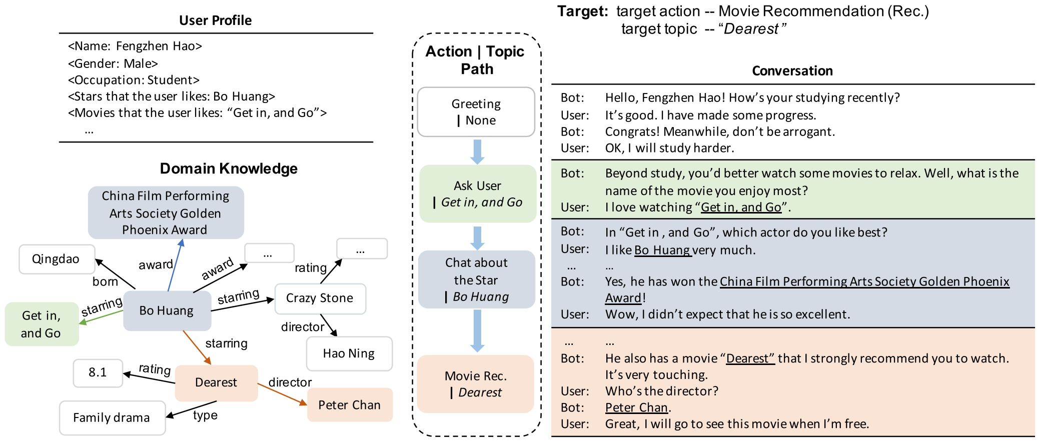
This work was supported by the Research Grants Council of Hong Kong (15207122, 15207920, 15207821, 15204018, 15213323) and National Natural Science Foundation of China (62076212). It was also supported in part by PolyU internal grants (ZVQ0, ZVVX).

Authors' address: J. Wang, D. Lin, and W. Li, The Hong Kong Polytechnic University, 11 Yuk Choi Road, Hung Hom, Kowloon, Hong Kong SAR, China; e-mails: jian-dylan.wang@connect.polyu.hk, dongding88.lin@connect.polyu.hk, cswjli@comp.polyu.edu.hk.

This work is licensed under a Creative Commons Attribution International 4.0 License.

© 2024 Copyright held by the owner/author(s).

ACM 1046-8188/2024/04-ART124
https://doi.org/10.1145/3652598



User Profile

<Name: Fengzhen Hao>
<Gender: Male>
<Occupation: Student>
<Stars that the user likes: Bo Huang>
<Movies that the user likes: "Get in, and Go">
...

Target: target action -- Movie Recommendation (Rec.)
target topic -- "*Dearest*"

Conversation

Bot: Hello, Fengzhen Hao! How's your studying recently?
User: It's good. I have made some progress.
Bot: Congrats! Meanwhile, don't be arrogant.
User: OK, I will study harder.
Bot: Beyond study, you'd better watch some movies to relax. Well, what is the name of the movie you enjoy most?
User: I love watching "Get in, and Go".
Bot: In "Get in , and Go", which actor do you like best?
User: I like Bo Huang very much.
... ...
Bot: Yes, he has won the China Film Performing Arts Society Golden Phoenix Award!
User: Wow, I didn't expect that he is so excellent.
... ...
Bot: He also has a movie "Dearest" that I strongly recommend you to watch. It's very touching.
User: Who's the director?
Bot: Peter Chan.
User: Great, I will go to see this movie when I'm free.

Fig. 1. An illustrative example from the re-purposed DuRecDial [29] dataset. Given a pre-determined target and a dialogue context, our objective is to generate utterances that proactively and smoothly lead the conversation to achieve the target.

tasks, e.g., task-oriented dialogues [32, 56, 61]. Despite passively responding to users, dialogue systems can also take a more proactive role [2, 62] to introduce new interesting topics to users. Such a target-oriented proactive dialogue system looks more intelligent, sociable, and capable of directing the users toward topic areas that the system knows how to talk about [8]. However, previous studies [8, 41, 47, 62] mainly focus on the scenario of open-domain dialogues. They define the target as a commonsense topic and explore bridging an initial dialogue context and the given topic. Such a scenario is difficult to be generalized to real-world applications.

In this work, we take a further step toward a more challenging target-oriented dialogue scenario, where the target is defined as an <action, topic> pair, such as providing recommendations for a specific topic that possibly attracts users. It requires the system to take more engaging actions to achieve the target, such as social chitchat, user exploration, topic elicitation, recommendation, and so on. As an example shown in Figure 1, suppose there is an explicit target, i.e., to recommend a specific movie named "*Dearest*", the system (i.e., Bot) is required to lead the conversation (e.g., "*greeting*" ⟶ "*ask user*" ⟶ "*chat about the star*" ⟶ "*movie recommendation*") so as to recommend the target movie when appropriate. It needs to consider the pre-determined target, dialogue history, and grounding domain knowledge (and user profile, if any). Particularly, the grounding domain knowledge associated with domain-specific topics and relevant attributes, is crucial to enable multiple topic transitions (e.g., warm-up chitchat ⟶ "*Get in, and Go*" ⟶ "*Bo Huang*" ⟶ "*Dearest*"). It is non-trivial to solve target-oriented dialogue generation for two reasons: (1) The system needs to keep the conversation engaging and proactively drive the conversation; (2) The system is desired to move the conversation forward to the target topic coherently and arouse the user's interest in the target topic to be recommended.

To address the above challenges, we observe that effective dialogue planning [54, 55] is essential for target-oriented dialogue generation. In order to achieve its target, the system needs to plan reasonable actions and appropriate topics to smoothly move the conversation forward to the target topic before generating each system utterance.[1] According to decision-making theories [14, 50] in

---

[1] We use the term "system utterance" in this article rather than "response" used in a lot of related work since the system needs to proactively lead the conversation in most cases.

cognitive science, humans tend to look ahead (forward) and look back (backward) when making decisions to achieve a long-term goal. Such bidirectional thinking alleviates short-sighted cognition and drives people to think about the complete decision path more. Similarly, in target-oriented proactive dialogue, the target has been designated in advance and should be bounded at the end of the dialogue path to be planned, backward path planning is effective in leveraging target-side information but insensitive to the coherence of the dialogue context. In contrast, forward path planning is more effective in generating a starting path point that is coherent with the dialogue context, while lacking the target-driven ability to enable the target to be bounded at the end of the dialogue path to be planned. With this in mind, we propose a **T**a**R**get-constrained b**I**directional **P**lanning (**TRIP**) method. The key point is to plan dialogue paths from both look-ahead (i.e., present-to-target) and look-back (i.e., target-to-present) directions. Generally, it is more appropriate when the look-ahead decision path is consistent with the look-back decision path. By formulating the planning as a generation task, our TRIP bidirectionally generates dialogue paths consisting of a sequence of <action, topic> pairs (see Figure 1) based on an encoder-decoder architecture. Concretely, we first take widely-used pre-trained language models, e.g., BERT [4], to encode complex input texts efficiently. Then, we employ two individual Transformer [53] decoders for dialogue path generation, with one to generate a dialogue path in the target-to-present direction and the other to generate one in the present-to-target direction. By minimizing the decision gap between the two directions, the two decoders are expected to provide supervision to each other and converge on a consistent dialogue path. In addition, we propose a contrastive generation mechanism (see Section 4.2) to enhance TRIP with the ability to better distinguish between the given target and non-targets. It enables TRIP to be more robust in generating the necessary target in the planned dialogue path accordingly. During inference, we propose a target-constrained decoding algorithm (see Section 4.3) with a bidirectional agreement, which reduces the gap between inference and training and facilitates the model to generate an appropriate dialogue path as the ultimate output.

Since each planned dialogue path outlines how to achieve the pre-determined target step by step, it is expected to help a dialogue system distill necessary knowledge and steer the system to generate more proper utterances with control. We adopt the planned dialogue path to guide dialogue generation in a pipeline manner, where we explore two variants: prompt-based generation (see Section 5.1) and plan-controlled generation (see Section 5.2). As part of this work, we re-purpose two existing recommendation-oriented dialogue datasets, namely DuRecDial [29] and DuRecDial 2.0 [28], for target-oriented dialogue generation through automatic target construction. Extensive experiments are conducted to show the effect of planning and the performance of different dialogue generation methods. Both automatic and human evaluations demonstrate that our proposed methods significantly outperform various baseline models.

Overall, our contributions are summarized as follows:

- We introduce the target-oriented dialogue generation task and discuss its relation and difference (Sections 2 and 3) compared to existing works.
- We propose a novel **target-constrained bidirectional planning** (**TRIP**) approach (Section 4) for target-oriented proactive dialogue systems. Our TRIP plans a dialogue path consisting of a sequence of <action, topic> pairs that outline how to achieve the designated target proactively and smoothly.
- We investigate both the prompt-based and plan-controlled methods (Section 5) to leverage planned dialogue paths to guide dialogue generation effectively.
- Experimental results show that our method achieves state-of-the-art performance in both automatic and human evaluations. Our extensive analysis provides some new insights into how planning affects target-oriented dialogue generation.

2 RELATED WORK

Our work is mainly related to target-oriented dialogue, recommendation-oriented dialogue, and content planning for natural language generation. We briefly review related work and clarify key differences compared with our work as follows.

2.1 Target-oriented Dialogue

Target-oriented dialogue systems work on the task of generating responses guided by the given target. According to the variety of the target, previous works have mainly focused on using a keyword [39, 47, 72], a topic [41, 62], and a concept or a sentence [8] as the guided target. For example, [47] introduced some coarse-grained keywords to control the intended content of the responses in open-domain dialogues, while [72] leveraged external commonsense knowledge graphs for keyword transitions. As a follow-up study, steering a dialogue toward a given keyword, or dialogue strategy learning, has also been explored in past work, including graph grounded policies [64, 65] and conversational lines [6]. For topic-guided dialogues, [62] investigated using an entity over a factual knowledge graph as the target topic, which requires the system to achieve a smooth transition from an initial topic to the given target topic. A new dataset called OTTers [41] was collected to explore one-turn topic transitions for open-domain response generation. More recently, [8] proposed to identify a bridging path of commonsense knowledge concepts between the dialogue context and the target sentence using data augmentation. Our work is more related to prior settings [25] on target topics and target sentences. However, existing works mainly focus on the scenario of open-domain target-guided dialogue, where they mainly consider guiding chitchat conversations to the target with transitions on commonsense topics. In comparison, we work on a more challenging setup that aims at achieving the target action for a designated target topic. It requires the system to take more engaging dialogue actions, such as social chitchat, user exploration, topic elicitation, and recommendation, to attract users so as to complete the target. We also clarify that existing studies on goal-oriented dialogue [7, 44] focus on the user-side goal or task, while our work explores the system-side target (or a specific goal).

2.2 Recommendation-oriented Dialogue

As a special type of task-oriented dialogue system, a recommendation-oriented dialogue system is desired to make recommendations through natural conversations with users. It was the emergence of various recommendation-oriented dialogue datasets that helps push forward the research in this area, such as GoRecDial [15], TG-ReDial [74], INSPIRED [9], and DuRecDial [29]. As follow-up studies, CR-Walker [30] was proposed to perform tree-structured reasoning over knowledge graphs, which can then be mapped into hierarchical dialogue acts to guide both item and response generations. MGCG [29] and KERS [68] explored the transition policy from a non-recommendation dialogue to a recommendation-oriented one. There is another similar research area called **conversational recommender systems** (**CRS**) [22, 26, 46]. Compared with recommendation-oriented dialogue systems, the main task of CRS lies in discovering user preferences [63, 66], asking clarifying questions about item attributes [19, 23], and searching for optimal candidate items [24, 69, 73]. In addition, [3] unified item recommendation and response generation into the same **sequence-to-sequence** (**Seq2Seq**) paradigm using prompt-based learning. Nonetheless, most existing systems passively respond to a user, where they provide recommendations according to the user's expressed interests or requirements. Our work aims at endowing a dialogue system with a more proactive role that can attract the user's interests and naturally lead user-engaged dialogues to achieve a pre-determined target.

2.3 Content Planning for Language Generation

There is a line of work [12, 33, 38, 45] that separates natural language generation into *content planning* and *surface realization*. Content planning mainly focuses on selecting the key contents (e.g., key phrases and entities) and arranging their orderings [33, 38], followed by a neural generation stage that focuses only on realization. Different strategies have been explored for content planning. For example, [43] proposed a hierarchical variational model for planning-based data-to-text generation, where a global latent variable models the diversity of planning and a sequence of local latent variables controls sentence realization. [13] presented a planning framework with iterative refinement to leverage large pre-trained language models for argument generation and article writing. For long-form text generation tasks, several studies [10, 11] conducted dynamic content planning while generating the output based on mixed language models to bridge the gap between content planning and sentence realization. Compared to these prior studies, our work is more related to planning for dialogue generation [54, 67]. We aim at addressing a more challenging dialogue generation task, where we propose a novel target-constrained bidirectional planning method to guide pre-trained language models to generate dialogue utterances more effectively.

3 PRELIMINARIES

In this section, we aim at providing preliminaries about the problem formulation and introduce essential sub-tasks accordingly. Then, we briefly introduce our proposed method with respect to addressing the problem effectively.

Suppose we have a target-oriented dialogue corpus $\mathcal{D} = \{(\mathcal{K}_i, \mathcal{P}_i, \mathcal{H}_i)\}_{i=1}^{N}$, where N denotes the number of dialogue samples. $\mathcal{K}_i = \{k_{i,j}\}_{j=1}^{N_K}$ denotes a set of domain knowledge facts relevant to i-th dialogue with each element $k_{i,j}$ in form of a ⟨*subject, relation, object*⟩ triple. $\mathcal{H}_i = \{(X_{i,t}, Y_{i,t})\}_{t=1}^{T}$ denotes dialogue content with a total number of T turns. $\mathcal{P}_i = \{(a_{i,l}, z_{i,l})\}_{l=1}^{L}$ denotes an annotated dialogue path for ith dialogue, each path span specifies an action-topic pair (a dialogue action $a_{i,l}$ and a dialogue topic $z_{i,l}$). L is the number of unique action-topic pairs. Here, the dialogue topics are mainly constructed upon the domain knowledge $\mathcal{K}_i$, and each action/topic may affect multiple turns of dialogue. In some scenarios, there also exists a user profile $\mathcal{U}_i$ grounded on the ith dialogue, which can be personal attributes or certain preferences.

Given a target $\mathcal{G}' = (a_{T'}, z_{T'})$ consisting of a target action $a_{T'}$ and a target topic $z_{T'}$, a dialogue history $\mathcal{H}'$, 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 engage the user in the conversation so as to achieve the target $\mathcal{G}'$ when appropriate. Due to the complexity of the problem, it can be decomposed into three sub-tasks: (1) **action planning**, i.e., plan actions to determine where the conversation should go to lead the conversation proactively; (2) **topic planning**, i.e., plan appropriate topics to move forward to the target topic smoothly; (3) **dialogue generation**, i.e., generate an appropriate utterance to achieve the planned action and topic at each turn.

To address the above tasks, we propose a target-constrained bidirectional planning method to guide dialogue generation in a pipeline manner. The target-constrained bidirectional planning aims at simultaneously solving sub-tasks of action planning and topic planning, which plans a reasonable dialogue path consisting of a sequence of dialogue actions and topics with proper orderings. At each turn, the planned path drives the system to distill necessary knowledge from the grounding domain knowledge and meanwhile guides the system to generate a proper utterance. We describe the details of the target-constrained bidirectional planning in Section 4 and plan-guided dialogue generation in Section 5, respectively.

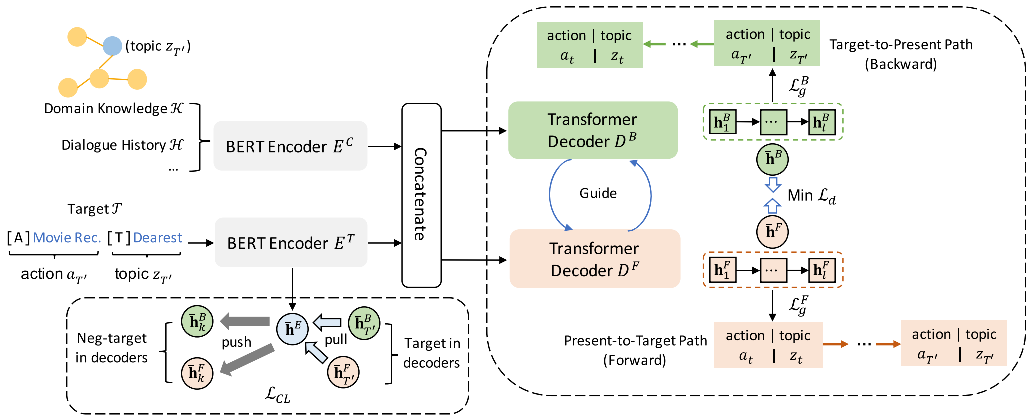

(topic $z_{\mathcal{T}'}$)
Domain Knowledge $\mathcal{K}$
Dialogue History $\mathcal{H}$
...
BERT Encoder E^C
Target $\mathcal{T}$
[A] Movie Rec. [T] Dearest
action $a_{\mathcal{T}'}$
topic $z_{\mathcal{T}'}$
BERT Encoder E^T
Concatenate
Neg-target in decoders
$\bar{\mathbf{h}}_k^B$
$\bar{\mathbf{h}}_k^F$
push
$\bar{\mathbf{h}}^E$
pull
$\bar{\mathbf{h}}_{\mathcal{T}'}^B$
$\bar{\mathbf{h}}_{\mathcal{T}'}^F$
Target in decoders
$\mathcal{L}_{CL}$
action | topic a_t | z_t
...
action | topic $a_{\mathcal{T}'}$ | $z_{\mathcal{T}'}$
Target-to-Present Path (Backward)
$\mathcal{L}_g^B$
Transformer Decoder D^B
$\mathbf{h}_1^B$ ... $\mathbf{h}_l^B$
$\bar{\mathbf{h}}^B$
Guide
Min $\mathcal{L}_d$
$\bar{\mathbf{h}}^F$
Transformer Decoder D^F
$\mathbf{h}_1^F$ ... $\mathbf{h}_l^F$
$\mathcal{L}_g^F$
Present-to-Target Path (Forward)
action | topic a_t | z_t
...
action | topic $a_{\mathcal{T}'}$ | $z_{\mathcal{T}'}$

Fig. 2. Overview of our target-constrained bidirectional planning (TRIP).

4 TARGET-CONSTRAINED BIDIRECTIONAL PLANNING

In this section, we propose a **Ta**R**get-constrained bIdirectional Planning (TRIP)** model to facilitate the system lead the conversation to achieve the pre-determined target. The overview of our TRIP is shown in Figure 2. Our TRIP is built with an encoder-decoder architecture, where we adopt two encoders to represent complex input texts and two individual decoders (i.e., a backward decoder and a forward decoder) to complete bidirectional planning.

4.1 Input Encoding

To efficiently represent various types of input, we take the widely-used pre-trained language model BERT [4] as our basic encoder. As shown in Figure 2, we concatenate the domain knowledge $\mathcal{K}$ and dialogue history $\mathcal{H}$ (and the user profile, if any) as the context. We separate them with a special token `[SEP]`, which is consistent with the processing in BERT. Then, the context sequence is encoded using a BERT encoder, denoted as E^C. For the given target consisting of a target action $a_{\mathcal{T}'}$ and a target topic $z_{\mathcal{T}'}$, we refer to the concatenated text of $a_{\mathcal{T}'}$ and $z_{\mathcal{T}'}$ as the target $\mathcal{T}$. We adopt two new special tokens `[A]` and `[T]` to differentiate $a_{\mathcal{T}'}$ and $z_{\mathcal{T}'}$, e.g., "`[A]` *Movie Recommendation* `[T]` *Dearest*". Then, the target $\mathcal{T}$ is encoded using another BERT encoder E^T. Briefly, the encoding of input is formulated as follows:

$$\mathbf{H}^C = \text{BERT Encoder } E^C([\mathcal{K};\mathcal{H}]), \tag{1}$$

$$\mathbf{H}^T = \text{BERT Encoder } E^T(\mathcal{T}), \tag{2}$$

where $\mathbf{H}^C = (\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_L) \in \mathbb{R}^{d \times L}$, $\mathbf{H}^T = (\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_{L'}) \in \mathbb{R}^{d \times L'}$, L and L' denote context length and target length respectively, d is the hidden size. Here, both $\mathbf{H}^C$ and $\mathbf{H}^T$ are token-level hidden representations. To maintain full input information for the subsequent planning, we concatenate $\mathbf{H}^C$ and $\mathbf{H}^T$ as the final input representation, denoted as $\mathbf{M} = [\mathbf{H}^C; \mathbf{H}^T]$.

4.2 Backward-forward Path Generation

Our TRIP aims at planning a reasonable dialogue path consisting of a set of dialogue actions and topics with proper orderings, and this planning process performs in a generation-based manner. We let TRIP generate a forward (present-to-target) path and a backward (target-to-present) path, respectively. It should be noted that the target action $a_{\mathcal{T}'}$ and target topic $z_{\mathcal{T}'}$ are bounded at the end of the path to be planned. For example, at tth turn, a forward dialogue path is "$a_t|z_t \longrightarrow a_{t+1}|z_{t+1} \longrightarrow$

$\cdots \rightarrow a_{T'}|z_{T'}$" while its backward dialogue path is "$a_{T'}|z_{T'} \rightarrow a_{T'-1}|z_{T'-1} \rightarrow \cdots \rightarrow a_t|z_t$". Planning a dialogue path from two opposite directions provides supervision to each other during training, and is expected to derive more reasonable dialogue action-topic pairs that compose the ultimate dialogue path, imitating humankind's bidirectional thinking.

In detail, our TRIP generates the two paths based on the Transformer [53] decoder architecture. We adopt two individual Transformer decoders D^B and D^F to generate the backward and forward paths, respectively. Both two decoders take the encoded hidden representation $\mathbf{M}$ as input and then output a dialogue path token by token, i.e., "`[A]`$a_1 a_2 \cdots$`[T]`$t_1 t_2 \cdots$`[EOS]`", in an autoregressive manner. Here, a_i denotes an action token, t_i denotes a topic token, `[A]` and `[T]` are two special tokens shared with the encoder E^T to differentiate an action and a topic, `[EOS]` denotes the end of the path sequence. For the backward decoder D^B, suppose the output dialogue path $\mathbf{y}$ is represented in token level, i.e., $\mathbf{y} = (y_1, y_2, \cdots, y_T)$ with a sequence length of T, and it is conditioned on the input text sequence (denoted as $\mathbf{x}$), the conditional distribution is approximated as follows:

$$p_\theta(y_t|\mathbf{y}_{<t}, \mathbf{x}) = \mathrm{softmax}(\mathbf{W}\mathbf{h}_t^B + \mathbf{b}), \tag{3}$$

$$\mathbf{h}_t^B = D^B(y_{t-1}, \mathbf{M}), \tag{4}$$

where $\mathbf{W} \in \mathbb{R}^{d\times d}$, $\mathbf{b} \in \mathbb{R}^d$ denote trainable parameters. We train the backward decoder D^B by minimizing the negative log likelihood for given N observations $\{(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})\}_{i=1}^N$ as follows:

$$\mathcal{L}_g^B(\theta) = -\sum_{i=1}^{N} p(\mathbf{y}^{(i)}) \log p_\theta(\hat{\mathbf{y}}^{(i)}|\mathbf{x}^{(i)}), \tag{5}$$

where $p(\mathbf{y}^{(i)})$ is the distribution of the ground-truth path sequence, while $p_\theta(\hat{\mathbf{y}}^{(i)})$ is the distribution of the approximated output path sequence, θ denotes all trainable parameters. Similarly, we train the forward decoder D^F following the above equations, with the loss function denoted as $\mathcal{L}_g^F$.

Reducing Gap between Backward-Forward Paths. Although the backward and forward paths are different, agreement on the dialogue actions and topics derived from the two paths is necessary since the two paths are planned for the same dialogue. By minimizing the decision gap between the backward path and the forward path, the two decoders (i.e., D^B and D^F) are expected to provide supervision to each other and converge on consistent dialogue actions and topics. In detail, we adopt the composition of a linear transformation with the ReLU [34] activation function and an average pooling to obtain the fixed-sized representation of a path, given by

$$\bar{\mathbf{h}}^B = f(\mathbf{V}^B),\ \bar{\mathbf{h}}^F = f(\mathbf{V}^F), \tag{6}$$

$$f(\mathbf{V}) = \mathrm{AvgPool}([\mathbf{v}_1 \cdots \mathbf{v}_T]), \text{where } \mathbf{v}_t = \mathrm{ReLU}(\mathbf{W}_1\mathbf{h}_t + \mathbf{b}_1), \tag{7}$$

where $\mathbf{W}_1 \in \mathbb{R}^{d\times d}$, $\mathbf{b}_1 \in \mathbb{R}^d$ denote trainable parameters, $\mathbf{h}_t$ stands for decoder hidden state. $\bar{\mathbf{h}}^B \in \mathbb{R}^d$ and $\bar{\mathbf{h}}^F \in \mathbb{R}^d$ are fixed-sized representation of the backward path and the forward path, respectively. Then, we reduce the gap between the two paths by minimizing L_2 distance between $\bar{\mathbf{h}}^B$ and $\bar{\mathbf{h}}^F$ as follows:

$$\mathcal{L}_d = ||\bar{\mathbf{h}}^B - \bar{\mathbf{h}}^F||_2, \tag{8}$$

where the distance $\mathcal{L}_d$ is added to the training loss as a regularization term.

Contrastive Generation of Targets. Since our dialogue path generation model is trained with teacher forcing and never exposed to incorrectly generated actions or topics during training, it is insufficient to distinguish between the given target action/topic and other actions or actions. Hence, the model may struggle to constrain the given target generated in the path. To remedy such a situation, we propose a contrastive generation framework (see Figure 2) to expose the model to various incorrect output targets for a given input target $\mathcal{T}$. Following the contrastive learning framework [18] for conditional text generation, we train the model to learn the representations of the ground-truth dialogue path by contrasting the positives with the negatives. The critical difference is that, we construct the perturbed negative examples by replacing the target topic in the ground-truth path with multiple randomly sampled topics $\{z_k\}_{k=1}^{K}$ $(z_k \neq z_{T'})$ from the domain knowledge $\mathcal{K}$, such that the training paths are difficult for the model to distinguish correctly. By identifying which features make the output path negative, these perturbed negative examples are expected to leverage encoders and decoders to learn an adequate representation of the target. It tries to enable our model to generate the necessary target in the path accordingly.

In detail, for the text span consisting of the target action $a_{T'}$ and topic $z_{T'}$ separated with the special tokens `[A]` and `[T]` in the two decoders, we project their hidden representations into the latent space following Equation (6) and Equation (7), obtaining fixed-sized target representations $\bar{\mathbf{h}}_{T'}^{B}$ and $\bar{\mathbf{h}}_{T'}^{F}$ respectively. Similarly, for the constructed negative examples, we also project those negative targets into the latent space following Equation (6) and Equation (7), obtaining corresponding neg-target representations. Since the pre-determined target $\mathcal{T}$ is encoded by the encoder E^T, we pull the ground-truth target representations in the decoders to the encoded target representation while pushing the neg-target representations in the decoders far away from the encoded target representation (see Figure 2). Then, we maximize the similarity between the pair of the encoder-decoder targets, while minimizing the similarity between the negative pairs as follows:

$$\mathcal{L}_{CL}^{B} = -\log \frac{\exp(\mathrm{sim}(\bar{\mathbf{h}}^{E}, \bar{\mathbf{h}}_{T'}^{B})/\tau)}{\sum_{\bar{\mathbf{h}}_{k}^{B} \in S^{B}} \exp(\mathrm{sim}(\bar{\mathbf{h}}^{E}, \bar{\mathbf{h}}_{k}^{B})/\tau)}, \tag{9}$$

$$\mathcal{L}_{CL}^{F} = -\log \frac{\exp(\mathrm{sim}(\bar{\mathbf{h}}^{E}, \bar{\mathbf{h}}_{T'}^{F})/\tau)}{\sum_{\bar{\mathbf{h}}_{k}^{F} \in S^{F}} \exp(\mathrm{sim}(\bar{\mathbf{h}}^{E}, \bar{\mathbf{h}}_{k}^{F})/\tau)}, \tag{10}$$

where $\bar{\mathbf{h}}^{E}$ denotes the averaged representation of the target $\mathcal{T}$ in the encoder E^T after transformation following Equation (6) and Equation (7), $\bar{\mathbf{h}}_{T'}^{B}$ and $\bar{\mathbf{h}}_{T'}^{F}$ are the ground-truth target representations in the two decoders, respectively. S^B and S^F stand for a set of neg-target representations in the two decoders, respectively. $\mathrm{sim}(\cdot,\cdot)$ is a cosine similarity function, τ is a temperature coefficient. Furthermore, we use the averaged result between $\mathcal{L}_{CL}^{B}$ and $\mathcal{L}_{CL}^{F}$ as the contrastive generation loss:

$$\mathcal{L}_{CL} = \frac{1}{2}\left(\mathcal{L}_{CL}^{B} + \mathcal{L}_{CL}^{F}\right). \tag{11}$$

Training. During training, we train our TRIP model by minimizing all the losses introduced above. We use two hyperparameters β and γ to control the importance of gap reducing and contrastive generation, given by

$$\mathcal{L} = \mathcal{L}_{g}^{B} + \mathcal{L}_{g}^{F} + \beta\mathcal{L}_{d} + \gamma\mathcal{L}_{CL}. \tag{12}$$

4.3 Target-constrained Decoding

After training is done, our TRIP model can be directly used to generate a dialogue path consisting of a set of dialogue actions and topics during inference. Alternatively, we can either use the forward decoder D^F to generate a path from the present to the target (denoted as “forward generation”), or use the backward decoder D^B to generate a path from the target to the present (denoted as “backward generation”). In order to take advantage of the bidirectional decoders, we propose a simple yet effective target-constrained decoding algorithm with a bidirectional agreement based on the widely-used beam search decoding algorithm.

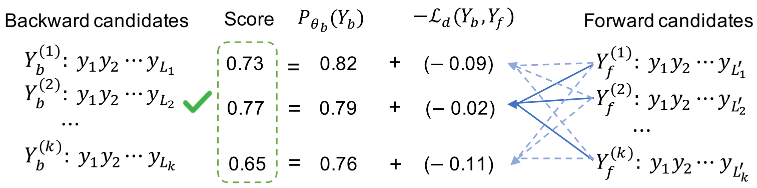

Backward candidates
Score
$P_{\theta_b}(Y_b)$
$-\mathcal{L}_d(Y_b, Y_f)$
Forward candidates
$Y_b^{(1)}$: $y_1 y_2 \cdots y_{L_1}$
$Y_b^{(2)}$: $y_1 y_2 \cdots y_{L_2}$ ✓
…
$Y_b^{(k)}$: $y_1 y_2 \cdots y_{L_k}$
0.73 = 0.82 + (– 0.09)
0.77 = 0.79 + (– 0.02)
0.65 = 0.76 + (– 0.11)
$Y_f^{(1)}$: $y_1 y_2 \cdots y_{L'_1}$
$Y_f^{(2)}$: $y_1 y_2 \cdots y_{L'_2}$
…
$Y_f^{(k)}$: $y_1 y_2 \cdots y_{L'_k}$

Fig. 3. Illustration of our target-constrained beam search decoding with bidirectional agreement.

First, each dialogue path is desired to be generated with lexical constraints, i.e., the target action and the target topic should be generated at the end of the path for "forward generation" while generated at the beginning of the path for "backward generation". To this end, we adopt two additional strategies to fulfill the lexical constraints. For the forward decoder D^F, we employ the **dynamic beam allocation** (**DBA**) [37] algorithm with a beam size of k to perform lexically constrained decoding, where the required constraint is defined as the given target action and topic. For the backward decoder D^B, we directly take the target tokens (i.e., a text span consisting of the target action and topic separated with the special tokens `[A]` and `[T]`) as the beginning input of the decoder, and then employ vanilla beam search decoding with the same beam size of k.

Second, after the two decoders finish the search process, we obtain k backward candidates (i.e., path sequences) and k forward candidates. As shown in Figure 3, to select the best one path sequence as the decoding output, we rank the backward candidates by the following scoring function:

$$Y_b = \mathop{\arg\max}_{Y_b^{(i)} \in S} P_{\theta_b}\left(Y_b^{(i)}\right) + \lambda \cdot \left(-\frac{1}{k}\sum_{j=1}^{k} \mathcal{L}_d\left(\bar{\mathbf{h}}\left(Y_b^{(i)}\right), \bar{\mathbf{h}}\left(Y_f^{(j)}\right)\right)\right), \tag{13}$$

where S denotes a set of backward candidates, $P_{\theta_b}(Y_b^{(i)})$ denotes the likelihood of the candidate $Y_b^{(i)}$. $\mathcal{L}_d(\cdot,\cdot)$ denotes L_2 distance between a backward candidate and a forward candidate, which is obtained by passing each pair of backward-forward candidates into the model and computed following Equation (8). Intuitively, the above scoring function ranks the backward candidates by likelihood and gives a partial reward to candidates that satisfy higher agreement (i.e., shorter distance) with the forward candidates, which reduces the gap between inference and training and facilitates the model to select a better one. Here, λ is a hyperparameter controlling the weight of the reward term. Note that we can also select the best one path sequence from the forward candidates using a similar scoring function, which performs a little inferior in most cases in our preliminary experiments. Therefore, by default, we select the best dialogue path sequence from the backward candidates as the ultimate planning output using Equation (13).

5 PLAN-GUIDED DIALOGUE GENERATION

As mentioned in the preliminaries, we adopt the planned dialogue path (denoted as "plan path" $\mathcal{P}$ for short) to guide dialogue generation in a pipeline manner. We expect these plan paths can help a dialogue system distill necessary knowledge and steer the system to generate more proper utterances with control. To achieve plan-guided dialogue generation, we devise two variants and describe them below.

5.1 Prompt-based Generation

Motivated by previous works that employ prompt-based learning for dialogue systems [31, 71], we regard each plan path $\mathcal{P}$ as the natural language prompt and then adopt a pre-trained generative **language model** (**LM**), e.g., GPT-2 [40], for dialogue generation. The overview of our

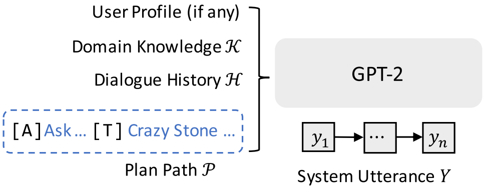

User Profile (if any)
Domain Knowledge $\mathcal{K}$
Dialogue History $\mathcal{H}$
[A] Ask ... [T] Crazy Stone ...
Plan Path $\mathcal{P}$
GPT-2
y_1
...
y_n
System Utterance Y

Fig. 4. Overview of our prompt-based dialogue generation.

prompt-based dialogue generation is shown in Figure 4. Formally, the plan path $\mathcal{P}$ is concatenated into the given dialogue history $\mathcal{H}$ and domain knowledge $\mathcal{K}$ (and the user profile, if any), formulating the input context X as follows:

$$X = [\mathcal{K}; \mathcal{H}; \mathcal{P}], \tag{14}$$

where ";" denotes concatenation. Here, the plan path $\mathcal{P}$ provides essential information that outlines how to achieve the target step by step. With the power of pre-trained LMs, the plan path $\mathcal{P}$ aims at distilling necessary knowledge from both input context and LMs. In particular, the input context X is fed into the pre-trained GPT-2 [40] model to generate the system utterance $Y = \{y_t\}_{t=1}^{n}$, where y_t is given by

$$y_t = \text{GPT-2}(y_{<t}, X). \tag{15}$$

We fine-tune GPT-2 for a few epochs using ground-truth plan paths in the dataset during training, while we adopt the plan paths generated by our TRIP model during inference.

5.2 Plan-controlled Generation

Considering that the plan path $\mathcal{P}$ outlines how to achieve the target step by step with a sequence of dialogue actions and topics, we expect to better leverage such critical information to control the attribute (e.g., switching or target topics) of dialogue generation. Inspired by plug-and-play language models for controllable language generation [1], we propose a plan-controlled dialogue generation method (see Figure 5). Built upon the pre-trained LM $p(y)$, e.g., GPT-2, we employ a simple plan model $p(a|y)$ to act as the attribute controller, which guides the generation of the LM $p(y)$ through gradients. Considering that the generation of system utterances follows the conditional form of $p(y|a) \propto p(y) \cdot p(a|y)$, we shift the hidden states of generation in the direction of the sum of two gradients: one toward higher log-likelihood of the unmodified LM $p(y)$ and one toward higher log-likelihood of the attribute a under the conditional plan model $p(a|y)$. Combining the two factors together makes it controllable to guide dialogue generation in a given direction (i.e., the plan path $\mathcal{P}$) with specified strength.

Concretely, as shown in Figure 5, we take the concatenated context X following Equation (14) as input and employ the pre-trained GPT-2 (denoted as LM $p(y)$) for dialogue generation. Let us define the cached hidden representations $\mathbf{H}_t$ of the LM $p(y)$ as all key-value pairs from the past, i.e., $\mathbf{H}_t = [(\mathbf{K}_t^{(1)}, \mathbf{V}_t^{(1)}), \cdots, (\mathbf{K}_t^{(l)}, \mathbf{V}_t^{(l)})]$, where $(\mathbf{K}_t^{(i)}, \mathbf{V}_t^{(i)})$ corresponds to the key-value pairs from the ith layer generated at all time-steps from 0 to t. Efficient computations of the LM $p(y)$ to generate the next token y_{t+1} using the cached $\mathbf{H}_t$ are summarized as

$$\mathbf{o}_{t+1}, \mathbf{H}_{t+1} = \text{LM}(y_t, \mathbf{H}_t), \tag{16}$$

$$y_{t+1} \sim p(y_{t+1}) = \text{softmax}(\mathbf{W}\mathbf{o}_{t+1}), \tag{17}$$

where $\mathbf{W}$ is a linear transformation that maps the hidden vector $\mathbf{o}_{t+1}$ to a vector of vocabulary size. On top of that, we build a simple plan model (denoted as $p(a|y)$) using a Transformer [53] decoder. The plan model $p(a|y)$ aims at re-generating the given plan path $\mathcal{P}$ conditioning on hidden

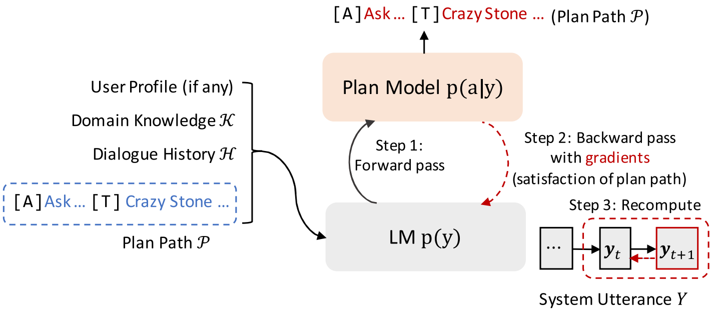


Fig. 5. Overview of our plan-controlled dialogue generation.

vectors $\{\mathbf{o}_0, \mathbf{o}_1, \cdots, \mathbf{o}_t\}$ of the LM $p(y)$ across all time-steps from 0 to t. Here, the plan model $p(a|y)$ performs as a generative discriminator that gives the LM $p(y)$ a higher reward for having the desired generation direction, i.e., the plan path $\mathcal{P}$. During training, we jointly train the plan model $p(a|y)$ and fine-tune the LM $p(y)$ by maximizing log-likelihood.

During inference, we use the plan model $p(a|y)$ to control the output of the LM $p(y)$ at every generation step t, following [1]. As shown in Figure 5, a forward pass is performed first through the LM $p(y)$ to compute the unmodified likelihood. In step 2, a backward pass updates the cached key-value pairs $\mathbf{H}_t$ with gradients from the plan model $p(a|y)$. Let $\Delta\mathbf{H}_t$ be the update to $\mathbf{H}_t$, such that the generation with $(\mathbf{H}_t + \Delta\mathbf{H}_t)$ shifts the distribution of the generated utterance so that it is more likely to satisfy the plan path $\mathcal{P}$. $\Delta\mathbf{H}_t$ is initialized at zero and updated as follows:

$$\Delta\mathbf{H}_t \leftarrow \Delta\mathbf{H}_t + \alpha \frac{\nabla_{\Delta\mathbf{H}_t} \log p(a|\mathbf{H}_t + \Delta\mathbf{H}_t)}{\|\nabla_{\Delta\mathbf{H}_t} \log p(a|\mathbf{H}_t + \Delta\mathbf{H}_t)\|}, \tag{18}$$

where α is the step size. This updating step can be repeated multiple times while in practice we update once for computational efficiency. Subsequently, we use the updated key-value pairs to recompute the perturbed hidden vector $\widetilde{\mathbf{o}}_{t+1}$, given by

$$\widetilde{\mathbf{o}}_{t+1}, \mathbf{H}_{t+1} = \text{LM}(y_t, \widetilde{\mathbf{H}}_t), \text{where } \widetilde{\mathbf{H}}_t = \mathbf{H}_t + \Delta\mathbf{H}_t. \tag{19}$$

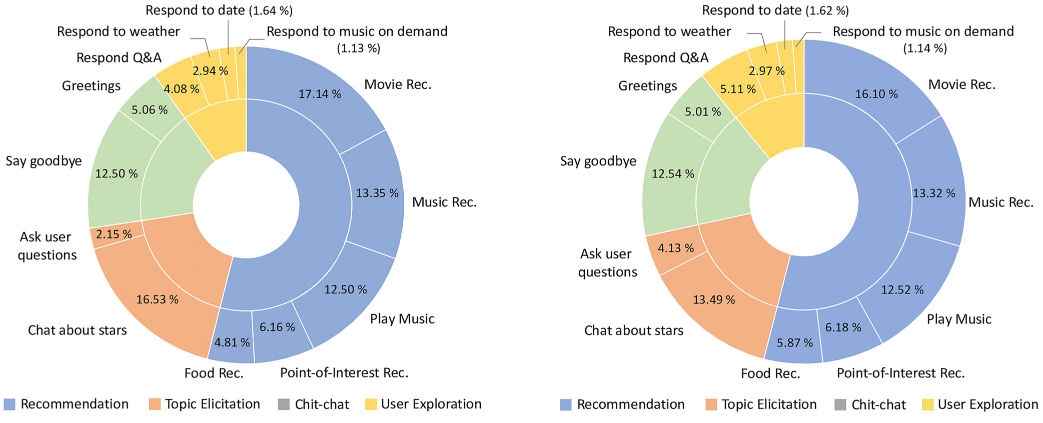
The perturbed $\widetilde{\mathbf{o}}_{t+1}$ is then used to generate the next token y_{t+1} following Equation (17).

6 EXPERIMENTAL SETUP

6.1 Datasets and Processing

Datasets. The task of target-oriented dialogue generation is still relatively under-explored. Although many publicly available dialogue datasets exist, we find the DuRecDial [29] and DuRecDial 2.0 [28] are the most suitable datasets for this task to the best of our knowledge. The system often leads the dialogue proactively instead of passively responding to users in the DuRecDial and DuRecDial 2.0 datasets, with rich interactive actions such as chitchat, question answering, recommendation, and so on. We first briefly introduce the two datasets and then introduce how we re-purpose the datasets for target-oriented dialogue generation.

The original DuRecDial and DuRecDial 2.0 datasets were collected from crowdsourced human-to-human dialogues. One person was defined as the seeker (the user's role) and the other as the recommender (the system's role) in a dialogue. The recommender was required to proactively lead the dialogue and make recommendations by introducing new topics. Each seeker was equipped with a user profile containing user attributes (e.g., name, age range) and his/her past preference information. In order to perform smooth conversations with the seeker, the recommender has a



(a) (b)

Fig. 6. Statistics of the system's dialogue actions on the re-purposed (a) DuRecDial and (b) DuRecDial 2.0 datasets.

domain knowledge graph consisting of domain-specific topics (e.g., movies, music, and food) with related attributes. More importantly, a dialogue path composed of dialogue actions and topics was annotated for the recommender (or the system) from the beginning to the end of the dialogue. The original DuRecDial dataset contains about 10 k multi-turn Chinese dialogues and 156 k utterances, while the DuRecDial 2.0 dataset has 8.2 k dialogues aligned across English and Chinese languages. In this work, we adopt the DuRecDial dataset in Chinese and the DuRecDial 2.0 dataset in English for experiments.

Data Processing. Since no explicit targets are annotated in the original DuRecDial and DuRecDial 2.0 datasets, we re-purpose the two datasets through automatic target construction for target-oriented dialogue generation, following [54]. 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) as the target action. We filter out those dialogues without introducing any new recommendation topic. In addition, we discard all user reviews in the original domain knowledge triples because user reviews do not belong to domain knowledge. We further enrich existing grounding domain knowledge triples corresponding to each dialogue with more knowledge triples sampled from the triples within two hops of the target topics in the dataset. Hence, it is more challenging for knowledge selection and topic planning. Note that each target topic is guaranteed to ground on the domain knowledge triples corresponding to the dialogue. Statistics of all the system's dialogue actions on the re-purposed DuRecDial and DuRecDial 2.0 datasets are shown in Figure 6. The total numbers of topics are 640 (including a `NULL` topic) and 628 (including a `NULL` topic) in the DuRecDial and DuRecDial 2.0, respectively.

Following the splitting criterion in [29, 54], we split the re-purposed DuRecDial dataset into the train/dev/test sets with 4,440/633/1,266 dialogues, respectively. Similarly, we obtain the train/dev/test sets of the DuRecDial 2.0 dataset with 4,256/608/1,216 dialogues, respectively. To investigate the performance of different methods for target-oriented dialogue generation, we further use the processed datasets with two types of splits for the test set: (1) **In-Domain (ID)** split and (2) **Out-Of-Domain (OOD)** split, similar to [8, 41]. The OOD split ensures that none of the target topics in the test set are present in the train set. In the ID split, the target topics in the test set are

Table 1. Statistics of the Re-Purposed DuRecDial and DuRecDial 2.0 Datasets

Dataset		#Dial.	#Utter.	Plan Path	Dial. Turn		Knowledge Triples	
				#Avg.	#Max.	#Avg.	#Max.	#Avg.
DuRecDial	Train	4,440	72,466	4.4	13	8.2	47	35.1
	Dev	633	10,467	4.5	12	8.3	46	35.2
	Test-ID	780	12,633	4.4	13	8.1	45	34.8
	Test-OOD	486	7,500	4.3	14	7.7	43	34.1
DuRecDial 2.0	Train	4,256	68,781	4.4	13	8.1	69	55.0
	Dev	608	9,677	4.3	14	8.0	68	55.4
	Test-ID	770	12,299	4.3	13	8.0	69	55.0
	Test-OOD	446	7,962	4.8	12	8.9	68	56.0

Here, "Dial." denotes "dialogue", "Utter." denotes "utterance", "ID" and "OOD" are short for "In-Domain" and "Out-Of-Domain", respectively.

allowed to appear in the train set. In total, statistics of the two re-purposed datasets are reported in Table 1. We can observe an average of 4.3 ∼ 4.8 action-topic transitions (i.e., the average length of the plan path) from the beginning toward the target.

6.2 Baseline Methods

Dialogue Generation. To validate the effectiveness of our proposed two variants for target-oriented dialogue generation, we first compare them with the following dialogue generation methods based on pre-trained language models:

- *DialoGPT* [70]: It is an autoregressive generation model pre-trained using large-scale dialogue corpora for conversational response generation. We adopt the pre-trained model[2] for fine-tuning the dataset in English. For fine-tuning the dataset in Chinese, we adopt the Chinese version [59] pre-trained model.[3]
- *GPT-2* [40]: It is a pre-trained autoregressive generation model for language generation. We use the publicly available GPT-2 base[4] model and Chinese GPT-2 base[5] model for fine-tuning the English and Chinese datasets, respectively.
- *BART* [20]: It is an encoder-decoder pre-trained language model with denoising for natural language generation. We use the publicly available BART-base[6] model and Chinese BART-base[7] model for fine-tuning the English and Chinese datasets, respectively.

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

We also compare our plan-guided dialogue generation methods with several competitive models that are closely related to target-oriented dialogue generation, where they follow the predict-then-generate paradigm or planning-enhanced generation paradigm:

- *MGCG_G* [29]: It employs the predicted next dialogue action and next topic to guide system utterance generation. Following our problem setting, we re-run the officially released code[8] on the two re-purposed datasets.

---

[2] https://huggingface.co/microsoft/DialoGPT-small

[3] https://github.com/thu-coai/CDial-GPT

[4] https://huggingface.co/gpt2

[5] https://huggingface.co/uer/gpt2-chinese-cluecorpussmall

[6] https://huggingface.co/facebook/bart-base

[7] https://huggingface.co/fnlp/bart-base-chinese

[8] https://github.com/PaddlePaddle/Research/tree/master/NLP/ACL2020-DuRecDial

- *KERS* [68]: It has a knowledge-enhanced mechanism for recommendation dialogue generation built upon Transformer [53] architectures. Similarly, we re-run the officially released code[9] on the two re-purposed datasets.
- *TCP-Dial* [54]: It proposes a target-driven conversation planning method to explicitly extract necessary knowledge and then guides dialogue generation built upon various backbone models. We adopt the GPT-2 [40] as the backbone model for comparisons in this work, and re-run the officially released code[10] on the two re-purposed datasets.

Dialogue Planning. To explore the performance of planning for target-oriented dialogue systems, we compare our TRIP model with the following dialogue planning methods:

- *MGCG* [29]: It employs a convolutional neural network [16] to conduct multi-task predictions for the next dialogue action and the next topic. However, it assumes that ground-truth historical dialogue actions and topics are known for a system. In this work, we only provide the target (i.e., a target action paired with a target topic), while the system itself should plan all interim dialogue actions and topics to achieve the target. For a fair comparison, we take the same input as our problem definition to conduct multi-task predictions.
- *KERS* [68]: It aims at generating the next dialogue action and the next topic based on a Transformer [53] network. Similarly, we take the same input as our problem definition for KERS.
- *BERT* [4]: Based on the intuition of multi-task predictions, we fine-tune the widely-used pre-trained language model BERT [4] by adding two fully-connected layers to jointly predict the system's next dialogue action and topic. We use the publicly available BERT-base-uncased[11] model and the Chinese BERT-base[12] model for fine-tuning the English and Chinese datasets, respectively.
- *TCP* [54]: It is a target-driven planning framework that aims at planning a path consisting of dialogue actions and topics in a generation-based manner. To the best of our knowledge, TCP is the most related work to ours on dialogue planning for the target-oriented dialogue generation task.

6.3 Evaluation Metrics

Automatic Evaluation. Following many previous studies [29, 54] in dialogue generation, we adopt widely-used metrics for automatic evaluation as follows:

- ***Perplexity*** (***PPL***) and ***distinct*** (***DIST***) [21]: The *perplexity* and *distinct* measure the fluency and the diversity of generated system utterances, respectively.
- *F1*: The *F1* score estimates the precision and recall of each generated utterance at the word level (the character level if evaluating Chinese datasets).
- *BLEU* [36]: The *BLEU* score calculates *n*-gram overlaps between generated utterances and gold utterances.
- ***Knowledge F1*** (***Know. F1***) [29]: It evaluates the performance of generating correct knowledge (e.g., topics, attributes) from the domain knowledge triples. However, there is no labeled knowledge annotated in gold system utterances in the datasets. We first conduct strict string matching to search for the entities from the domain knowledge that also occur in each gold system utterance as the knowledge label. Since some knowledge entries (*object* in the triple

---

[9] https://github.com/z562/KERS

[10] https://github.com/iwangjian/Plan4RecDial

[11] https://huggingface.co/bert-base-uncased

[12] https://huggingface.co/bert-base-chinese

⟨*subject, relation, object*⟩) are in form of long texts (e.g., topic-associated attributes) and they can be paraphrased during conversations, we then compute word-based recall scores between knowledge entries and gold system utterances. We take the knowledge entries whose recall scores are greater than a threshold of 0.55 as the pseudo label. For evaluating knowledge F1, we take the same threshold (i.e., 0.55) to examine whether a knowledge entry is hit in the generated utterances.

- ***Goal success rate*** (***Goal Succ.***): It is essential to validate a model of how well it achieves the pre-determined target, where the target topic can be used for automatic evaluation. Similar to [54], we choose the dialogues at the target turn in the test dataset to compute the ratio of generating the target topic correctly for each model as the *goal success rate*.

To evaluate dialogue planning, we adopt the following metrics:

- *F1*: It estimates the micro-averaged precision and recall of the predicted action or topic. For generation-based models, we take the generated action or topic at the evaluating turn for a fair comparison. We report dialogue action F1 and topic F1 scores in the experimental results, respectively.
- ***Bigram F1*** (***Bi. F1***): Due to the nature of dialogues, multiple temporary planning strategies can be reasonable before completing the target. Following [75], we also expand gold labels by taking the system's actions and topics within the previous turn and the following turn into account, formulating the *bigram F1*.

Human Evaluation. Similar to [29], we conduct human evaluation from both turn-level and dialogue-level aspects. For turn-level evaluation, we randomly select 50 samples from the test-ID dataset and 50 samples from the test-OOD dataset and ask each model to produce system utterances. Three well-educated annotators are required to mark scores for different models from the aspects of both (i) *appropriateness* and (ii) *informativeness*. The *appropriateness* measures if a generated system utterance can complete the current plan and respond to the dialogue context appropriately, and the *informativeness* measures if a model can make full use of the grounding domain knowledge to generate an informative utterance. For fairness, all model names are masked to annotators during the evaluation process.

For dialogue-level evaluation, we let each model interact with human annotators, which indicates that a model's generated utterance in the current turn will be further used as a part of the dialogue history in the next turn. To ensure that the evaluation covers a wide range of targets, we randomly sample 5 different target actions from the test sets, with each action consisting of 10 different target topics. In total, 50 different dialogue targets are evaluated. To examine whether a model can lead the conversation to achieve the pre-determined target proactively and smoothly, we do not expose the target to human annotators during human-model interactions. Besides, human annotators are asked to be consistent with each given user profile, if any. All human-model dialogues are limited to no more than 15 turns. At the end of each dialogue, we will expose the pre-determined target to human annotators and ask each annotator to mark scores for different models from (i) *proactivity*, which measures if a model can proactively lead new actions and topics in the conversation, (ii) *coherence*, which measures the overall fluency and naturalness of the whole dialogue generation, and (iii) *goal success*, which estimates how well the pre-determined target is achieved.

For all the above metrics, human evaluation scores are settled from {0, 1, 2}, where a higher score denotes better performance. The agreement among the annotators is measured by Fleiss's kappa [5]. The averaged score of different human annotators is reported as the ultimate human evaluation result for each model.

6.4 Implementation Details

Our TRIP model and plan-guided generation methods are implemented by PyTorch. During planning, we adopt the BERT-base model (12 layers, 768 dimensions, 12 heads, and 110 M parameters) and the Chinese BERT-base model released in Huggingface's Transformers [60] library as input encoders for the DuRecDial 2.0 dataset and the DuRecDial dataset, respectively. Both the backward and forward decoders are stacked to 6 layers with 8 attention heads. The hidden size is set to be consistent with BERT encoders, i.e., 768. The embeddings of the two decoders are randomly initialized, with the vocabulary size consistent with the BERT encoders. For the contrastive generation of targets, we randomly sample 3 negatives. The temperature coefficient τ is set to 0.1. The two hyperparameters β and γ are set to 0.1 and 1.0, respectively. We set the batch size to 6 due to memory constraints and train our TRIP model with a maximum of 10 epochs. We adopt the Adam [17] optimizer with an initial learning rate of $2e$-5 and warm up over the first 3,000 training steps with linear decay. We select the best model based on the performance of the validation set. For the target-constrained decoding, the beam size is set to 3, with a maximum decoding length of 80. The hyperparameter λ that controls the weight of the agreement reward is set to 1.0.

For our plan-guided dialogue generation, we fine-tune the GPT-2 base model and Chinese GPT-2 base model released in Huggingface's Transformers [60] library on the DuRecDial 2.0 dataset and the DuRecDial dataset, respectively. The length of the concatenated input text is limited to 512. In addition, the plan model $p(a|y)$ in the plan-controlled generation employs a lightweight Transformer decoder with 3 layers and 8 attention heads. The embeddings of $p(a|y)$ are copied from the embeddings of the LM $p(y)$ (i.e., GPT-2). The step size α is set to 0.01. Both variants employ greedy search decoding during generation, with a maximum decoding length of 100. All the experiments are conducted on a single NVIDIA GeForce 3,090 GPU machine. Our code and data are available at https://github.com/iwangjian/TRIP.

7 EXPERIMENTAL RESULTS

Our experiments and detailed analysis aim at answering the following research questions:

- RQ1: How is the performance of the proposed planning for generation on the end task of target-oriented dialogue generation compared to existing methods?
- RQ2: How is the performance of the proposed TRIP model on each sub-task, including action planning and topic planning, compared to existing methods?
- RQ3: How does each proposed component or strategy contribute to the overall performance?
- RQ4: What are the merits and limitations of the pipelined approach in this work?

7.1 Evaluation Results of Dialogue Generation (RQ1)

Our automatic evaluation results of dialogue generation on the DuRecDial and DuRecDial 2.0 datasets are reported in Table 2 and Table 3, respectively. The best result in terms of the corresponding metric is highlighted in boldface. As shown in Table 2, MGCG_G and KERS are capable of obtaining better results than DialoGPT on the **in-domain** (**ID**) test set in terms of F1, BLEU, and DIST. Considering that the two models are trained without using pre-trained language models, their competitive performance mainly benefits from the prediction of the next dialogue action and topic, which guides the model to generate more informative and reasonable utterances. However, MGCG_G, KERS, and DialoGPT obtain poor goal success rates, which drop sharply on the **out-of-domain** (**OOD**) test set in particular. It shows that they still struggle to lead dialogues to reach the target when necessary. In comparison, GPT-2 and BART perform much better than other baseline models over various metrics when evaluated on both ID and OOD test sets. We note that in terms of DIST-1/2 scores, BART is significantly better than other baselines because BART seldom

Table 2. Evaluation Results of Dialogue Generation on the DuRecDial Dataset

	Model	**PPL (↓)**	**F1 (%)**	**BLEU-1 / 2**	**DIST-1 / 2**	**Know. F1 (%)**	**Goal Succ. (%)**
Test-ID	MGCG_G	17.81	40.07	0.352 / 0.273	0.012 / 0.058	41.14	37.22
	KERS	12.39	38.24	0.356 / 0.277	0.011 / 0.047	45.04	44.16
	DialoGPT	5.64	34.25	0.314 / 0.237	0.009 / 0.045	33.90	38.55
	GPT-2	4.42	39.57	0.370 / 0.297	0.012 / 0.062	45.84	59.97
	BART	5.01	38.68	0.341 / 0.268	**0.013** / **0.075**	43.84	68.31
	TCP-Dial	4.41	39.10	0.379 / 0.303	0.011 / 0.058	49.90	69.88
	Ours (prompt)	**4.38**	42.34*	0.387* / **0.312***	0.011 / 0.059	53.12*	77.40*
	Ours (controlled)	4.40	**43.11***	**0.388*** / **0.312***	0.012 / 0.062	**53.69***	**77.55***
Test-OOD	MGCG_G	18.57	36.72	0.339 / 0.257	0.012 / 0.045	32.24	10.24
	KERS	14.06	36.28	0.340 / 0.262	0.010 / 0.042	38.38	14.07
	DialoGPT	5.57	35.52	0.325 / 0.252	0.010 / 0.041	35.16	37.16
	GPT-2	4.48	40.90	0.388 / 0.317	0.013 / 0.055	47.42	58.60
	BART	5.04	40.04	0.360 / 0.288	**0.015** / **0.067**	46.62	65.45
	TCP-Dial	4.46	34.65	0.358 / 0.279	0.012 / 0.055	32.29	16.72
	Ours (prompt)	4.46	41.90*	0.396* / **0.322***	0.012 / 0.054	48.32*	78.30*
	Ours (controlled)	**4.45**	**42.40***	**0.397*** / **0.322***	0.013 / 0.055	**48.93***	**79.80***

Significant improvements over baseline models are marked with * (t-test, $p < 0.05$).

Table 3. Evaluation Results of Dialogue Generation on the DuRecDial 2.0 Dataset

	Model	**PPL (↓)**	**F1 (%)**	**BLEU-1 / 2**	**DIST-1 / 2**	**Know. F1 (%)**	**Goal Succ. (%)**
Test-ID	MGCG_G	25.32	35.13	0.316 / 0.211	0.016 / 0.053	39.53	20.51
	KERS	20.15	31.27	0.288 / 0.196	0.017 / 0.061	41.18	28.75
	DialoGPT	5.26	35.12	0.304 / 0.212	0.023 / 0.076	42.71	30.09
	GPT-2	5.33	36.86	0.314 / 0.222	0.024 / 0.081	43.62	31.64
	BART	6.46	36.11	0.279 / 0.181	**0.030** / **0.096**	43.33	33.05
	TCP-Dial	5.88	34.46	0.293 / 0.201	0.027 / 0.091	45.75	29.49
	Ours (prompt)	**5.17**	37.40*	0.326* / 0.233*	0.026 / 0.083	47.03*	36.13*
	Ours (controlled)	5.23	**37.48***	**0.331*** / **0.238***	0.025 / 0.080	**47.44***	**38.67***
Test-OOD	MGCG_G	28.21	30.84	0.276 / 0.167	0.015 / 0.046	20.53	5.65
	KERS	24.35	27.91	0.259 / 0.160	0.016 / 0.058	26.88	11.06
	DialoGPT	5.37	31.27	0.283 / 0.176	0.021 / 0.068	30.75	26.57
	GPT-2	5.86	31.26	0.266 / 0.193	0.023 / 0.077	28.79	26.30
	BART	8.09	32.38	0.244 / 0.149	0.026 / 0.081	30.02	28.10
	TCP-Dial	8.24	29.24	0.255 / 0.165	**0.027** / **0.089**	21.36	6.97
	Ours (prompt)	5.63	33.05*	0.292* / 0.198*	0.025 / 0.079	31.81*	31.17*
	Ours (controlled)	**5.59**	**33.30***	**0.297*** / **0.202***	0.024 / 0.078	**32.82***	**33.44***

Significant improvements over baseline models are marked with * (t-test, $p < 0.05$).

generates repeated words, making the generated utterances more diverse in many cases. However, GPT-2 performs better in most cases in generating n-gram overlapped utterances (see BLEU-1/2) with correct knowledge (see Know. F1). We employ GPT-2 as our backbone model due to its strong generation ability and ease of incorporation in our plan-controlled generation. For the TCP-Dial, the goal success rate deteriorates remarkably on the OOD test dataset (16.72%) compared to the ID test dataset (69.88%). It is because TCP-Dial explicitly extracts topic-centric knowledge triples according to the planned topic, which may discard necessary domain knowledge when the target topic is not correctly planned especially on the OOD test dataset, making it difficult to generate a proper utterance containing the target topic.

Compared to baseline methods, our proposed plan-guided generation methods achieve significant improvements over most evaluation metrics. For example, our prompt-based generation method achieves much better knowledge F1 scores, i.e., 53.12% and 48.32% on the ID and OOD

Table 4. Experimental Results of Dialogue Planning on the DuRecDial Dataset

Model	Test-ID				Test-OOD			
	Action		Topic		Action		Topic	
	F1	Bi. F1	F1	Bi. F1	F1	Bi. F1	F1	Bi. F1
MGCG	87.30	93.16	71.21	76.99	88.29	91.75	44.50	49.88
KERS	92.13	94.08	80.34	82.60	90.34	91.88	41.25	44.36
BERT	94.82	95.99	83.37	84.43	92.73	93.65	48.70	50.97
TCP	91.76	94.17	85.71	87.26	92.41	94.86	46.63	47.46
Ours (TRIP)	**95.84***	**97.28***	**90.23***	**91.39***	**95.61***	**96.87***	**69.76***	**70.48***

Significant improvements over baseline models are marked with * (t-test, $p < 0.05$).

test sets (see Table 2). It shows that our model is more likely to generate correct knowledge (e.g., topics, attributes) from the domain knowledge triples. In terms of the goal success rate according to Table 2, our prompt-based generation method obtains a much higher score of 77.40%, which significantly outperforms existing baseline models. It indicates that we successfully stimulate the potential of the existing pre-trained language model (i.e., GPT-2) to generate more proper utterances for target-oriented dialogue generation by enriching appropriate dialogue paths as prompts. More importantly, our model is still able to maintain a high goal success rate when evaluated on the OOD test set. In contrast to GPT-2, our model mainly benefits from our dialogue planning, which verifies the effectiveness of the proposed planning for generation on the end task of target-oriented dialogue generation. Moreover, our plan-controlled generation method further improves the performance of the prompt-based generation method, demonstrating that each planned dialogue path can further steer the model by controlling the generation process of each utterance.

We observe similar trends in Table 3 regarding automatic evaluation results on the DuRecDial 2.0 dataset. Both our prompt-based generation and plan-controlled generation methods outperform existing baseline models over most evaluation metrics. We note that all baseline models and our methods perform inferior to that on the DuRecDial dataset in terms of the goal success rate. It is because, in the DuRecDial 2.0 dataset, the domain knowledge triples grounding on each dialogue are noisier than that in the DuRecDial dataset, making it non-trivial for these models to distinguish the target topic and to generate the target topic in the utterance accordingly when necessary. Nonetheless, our methods still achieve better goal success rates, especially when evaluated on the OOD test set. Overall, experimental results reported in Table 2 and Table 3 demonstrate that compared to existing methods, our proposed two variants are effective in generating more appropriate utterances on the end task of target-oriented dialogue generation.

7.2 Evaluation Results of Dialogue Planning (RQ2)

To validate the performance of dialogue action planning and topic planning, we compare our proposed TRIP model with existing dialogue planning models. The automatic evaluation results on the DuRecDial and DuRecDial 2.0 datasets are reported in Table 4 and 5, respectively. As shown in Table 4, it is more difficult for all models to predict or generate dialogue topics correctly than dialogue actions because the total size of the topics is much larger than that of the actions in the dataset. For example, MGCG and KERS achieve comparable F1 and Bi. F1 scores on action planning while they perform much inferior on topic planning compared to other baseline models (i.e., BERT and TCP) that employ pre-trained language models. More obviously, we find that all models obtain much lower F1 and Bi. F1 scores in terms of topic planning when evaluated on the OOD test set. Since the target topics in the OOD test set are not allowed to appear in the train set, all models are challenging to capture the semantics of the target topics and predict or generate the target topics

Table 5. Experimental Results of Dialogue Planning on the DuRecDial 2.0 Dataset

Model	**Test-ID**				**Test-OOD**			
	Action		**Topic**		**Action**		**Topic**	
	F1	**Bi. F1**	**F1**	**Bi. F1**	**F1**	**Bi. F1**	**F1**	**Bi. F1**
MGCG	90.26	92.47	74.93	79.24	82.30	87.25	36.03	42.00
KERS	90.33	91.54	77.85	80.35	84.21	86.39	34.20	37.85
BERT	91.68	92.37	79.21	81.22	92.23	94.19	46.55	52.12
TCP	92.25	93.82	85.77	87.25	89.93	92.09	44.49	50.71
Ours (TRIP)	**94.49***	**95.89***	**91.83***	**93.51***	**93.27***	**95.18***	**70.65***	**74.47***

Significant improvements over baseline models are marked with * (t-test, $p < 0.05$).

Table 6. Ablation Study Results of our TRIP Model on the DuRecDial Dataset

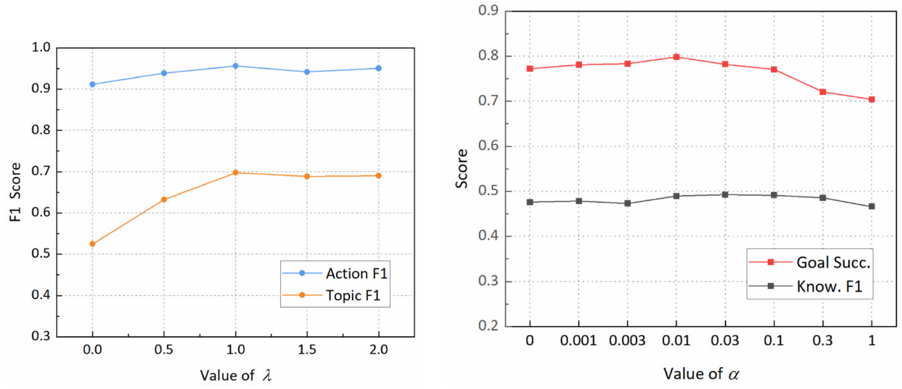
Model	**Test-ID**				**Test-OOD**			
	Action		**Topic**		**Action**		**Topic**	
	F1	**Bi. F1**	**F1**	**Bi. F1**	**F1**	**Bi. F1**	**F1**	**Bi. F1**
TRIP (full)	95.84	97.28	90.23	91.39	95.61	96.87	69.76	70.48
w/o D^F	93.71	95.55	86.26	87.40	92.09	93.17	46.52	50.22
w/o D^B	92.89	94.68	85.89	87.02	91.16	92.87	45.09	48.67
w/o $\mathcal{L}_d$	95.33	96.81	88.15	90.06	94.03	95.11	68.12	69.80
w/o $\mathcal{L}_{CL}$	95.45	96.90	88.76	90.13	93.15	94.09	67.30	68.22
w/o LC	91.31	93.08	84.20	85.66	91.02	93.23	51.34	53.06
w/o BA	92.06	94.35	85.46	86.89	91.15	93.20	52.49	54.13

correctly. In contrast, our TRIP model achieves substantial improvements in both dialogue action planning and topic planning. Particularly, TRIP improves the topic F1 score from 70%–80% to over 90% on the ID test set. It still maintains a much higher topic F1 score of 69.76% on the challenging OOD test set. Similar trends are also observed in Table 5 when all these methods are evaluated on the DuRecDial 2.0 dataset. We can conclude that our TRIP is able to plan a dialogue path consisting of more accurate dialogue actions and more reasonable topics. It is our effective dialogue planning that makes it possible to steer the system to lead the conversation toward the target proactively and smoothly.

7.3 Ablation Study of TRIP (RQ3)

To explore why our TRIP achieves superior performance in dialogue planning, we conducted an ablation study to verify the effectiveness of the modules and mechanisms proposed in TRIP. We focus on the following settings for ablation experiments: **(1)** without the forward decoder (w/o D^F), which denotes that only the backward decoder is employed to generate the dialogue path from the target turn to the present turn, followed by vanilla beam search decoding (the proposed target-constrained decoding algorithm is invalid in such a case); **(2)** without the backward decoder (w/o D^B), which denotes that only the forward decoder is employed to generate the dialogue path from the present turn to the target turn, followed by vanilla beam search decoding similarly; **(3)** without reducing the gap between backward-forward paths (w/o $\mathcal{L}_d$); **(4)** without the contrastive generation of targets (w/o $\mathcal{L}_{CL}$); **(5)** without the lexical constraints in the target-constrained decoding (w/o LC); **(6)** without the bidirectional agreement in the target-constrained decoding (w/o BA).

From the ablation study results shown in Table 6, we observe that each module or mechanism contributes to dialogue planning. The performance of TRIP sharply dropped when removing either



(a) Impact of the hyperparameter λ

(b) Impact of the step size α

Fig. 7. Quantitative results by varying the value of different parameters.

the backward decoder D^B or the forward decoder D^F. In particular, the topic F1 score decreased from 69.76% to 46.52% (w/o D^F) and 45.09% (w/o D^B) on the OOD test set. Such ablation results prove that our basic idea of employing two decoders for bidirectional planning is viable and effective. We also observe that the absence of D^B performs worse than that of D^F. It is because D^B directly takes the target as the beginning input of the decoder and generates the dialogue path in a target-to-present direction, which is of benefit to leverage the target-side information to guide planning more effectively. For the ablation results without $\mathcal{L}_d$ and $\mathcal{L}_{CL}$, both reducing the gap between backward-forward paths and contrastive generation of targets can benefit the model in planning as we expect. In terms of the target-constrained decoding, we find that the ultimate performance deteriorated rapidly when removing the lexical constraints (w/o LC) or bidirectional agreement (w/o BA), especially the topic F1 score decreased from 69.76% to 51.34% (w/o LC) and 52.49% (w/o BA) on the OOD test set. It indicates that our target-constrained decoding performs a vital role in dialogue planning since it controls the model's attention to the target-side information during inference even when handling out-of-domain target topics.

7.4 Analysis of Parameters (RQ3)

We quantitatively analyzed some critical parameters of our methods, including (1) the hyperparameter λ that controls the weight of the bidirectional agreement reward in the planning stage and, (2) the step size α that controls the updating step in the plan-controlled dialogue generation.

Impact of the hyperparameter λ. To investigate the impact of the hyperparameter λ in the planning stage, we conducted target-constrained decoding by varying λ in {0, 0.5, 1.0, 1.5, 2.0}. Experimental results are shown in Figure 7(a). We observe that our model achieves the best action F1 and topic F1 scores when $\lambda = 1.0$ and a smaller value of λ results in lower action F1 and topic F1 scores. Particularly, the model performs much inferior without any reward of bidirectional agreement, i.e., $\lambda = 0$, indicating that our target-constrained decoding with a bidirectional agreement is crucial in generating a more reasonable dialogue path.

Impact of the step size α. To investigate the impact of the step size α in the plan-controlled dialogue generation, we varied α by selecting its value in {0, 0.001, 0.003, 0.01, 0.1, 0.3, 1.0}. Experimental results are shown in Figure 7(b). We observe that the step size α mainly affects the goal success rate while it has a slighter impact regarding the knowledge F1 score. If no updating step is

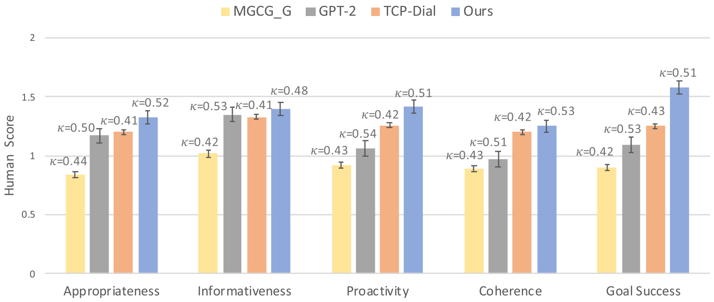


Fig. 8. Human evaluation results of different models. κ denotes Fleiss's kappa.

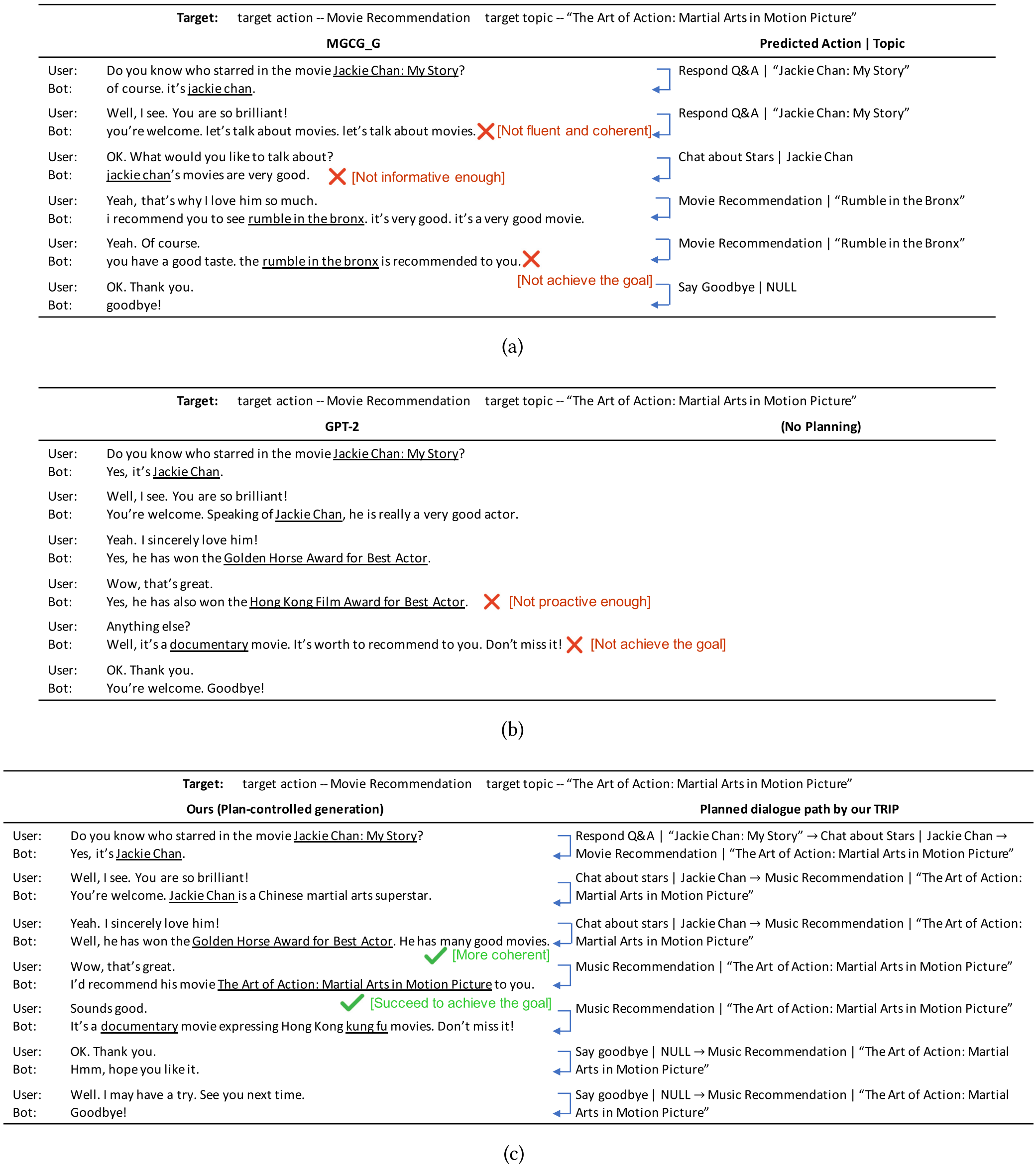
performed during plan-controlled dialogue generation, i.e., $\alpha = 0$, the dialogue generation model (i.e., LM $p(y)$) has no control of the output distribution, especially for those utterances that the target topics should explicitly appear. By default, we choose 0.01 as the most proper step size since neither a larger value nor a smaller one will bring any performance gain.

7.5 Human Evaluation Results (RQ4)

We selected several representative models for human evaluation, including MGCG_G, GPT-2, TCP-Dial, and ours. The evaluation results are shown in Figure 8. The Fleiss's kappa scores are mainly distributed in [0.4, 0.6], denoting moderate inter-annotator agreement. For turn-level evaluation, we observe that GPT-2, TCP-Dial, and ours obtain comparable scores in informativeness since they utilize powerful pre-trained language models and thus can generate informative utterances. In terms of appropriateness, our method obtains the highest human score on average, demonstrating the ability to generate more appropriate system utterances in response to dialogue context. On the other hand, dialogue-level evaluation (i.e., proactivity, coherence, and goal success) is more challenging for all models because errors might be propagated as the dialogue goes on. We find that our method obtains better results on average compared to all baseline models. Notably, our method achieves the highest proactivity and goal success scores, indicating that our method is more likely to drive the dialogue to reach the target successfully.

7.6 Case Study (RQ4)

To illustrate the quality of different methods for target-oriented dialogue generation, we conducted some case studies. We selected the same target with the same initial dialogue context and investigated the generated utterances by three different models from dialogue-level human evaluation, including MGCG_G, GPT-2, and ours (plan-controlled generation). Here, we show some generated cases in Figure 9. As shown in Figure 9(a), we observe that MGCG_G is incapable of generating fluent and coherent utterances. Although MGCG_G conducts planning first to predict the next dialogue action and topic, it fails to predict a correct topic when necessary, causing the model fails to achieve the target (i.e., recommend the movie "*The Art of Action: Martial Arts in Motion Picture*") at the end of the dialogue. For the case of GPT-2 shown in Figure 9(b), we find that GPT-2 is able to generate more fluent and informative utterances in general. However, it fails to achieve the target since it has no dialogue planning, making it not proactive enough to lead the dialogue towardthe pre-determined target. In such cases, GPT-2 is not effective to generate the target topic as the dialogue goes on. In contrast, the case shown in Figure 9(c) demonstrates that our TRIP model can plan a dialogue path with reasonable actions and appropriate topics that outlines how to achieve the target step by step. With the guidance of the planned path, our plan-controlled generation method can know when and what to talk about to move the dialogue forward proactively. More



Target: target action -- Movie Recommendation target topic -- "The Art of Action: Martial Arts in Motion Picture"

	MGCG_G	**Predicted Action | Topic**
User:	Do you know who starred in the movie Jackie Chan: My Story?	Respond Q&A | "Jackie Chan: My Story"
Bot:	of course. it's jackie chan.	
User:	Well, I see. You are so brilliant!	Respond Q&A | "Jackie Chan: My Story"
Bot:	you're welcome. let's talk about movies. let's talk about movies. ✗ [Not fluent and coherent]	
User:	OK. What would you like to talk about?	Chat about Stars | Jackie Chan
Bot:	jackie chan's movies are very good. ✗ [Not informative enough]	
User:	Yeah, that's why I love him so much.	Movie Recommendation | "Rumble in the Bronx"
Bot:	i recommend you to see rumble in the bronx. it's very good. it's a very good movie.	
User:	Yeah. Of course.	Movie Recommendation | "Rumble in the Bronx"
Bot:	you have a good taste. the rumble in the bronx is recommended to you. ✗	
User:	OK. Thank you. [Not achieve the goal]	Say Goodbye | NULL
Bot:	goodbye!	

(a)

Target: target action -- Movie Recommendation target topic -- "The Art of Action: Martial Arts in Motion Picture"

	GPT-2	**(No Planning)**
User:	Do you know who starred in the movie Jackie Chan: My Story?	
Bot:	Yes, it's Jackie Chan.	
User:	Well, I see. You are so brilliant!	
Bot:	You're welcome. Speaking of Jackie Chan, he is really a very good actor.	
User:	Yeah. I sincerely love him!	
Bot:	Yes, he has won the Golden Horse Award for Best Actor.	
User:	Wow, that's great.	
Bot:	Yes, he has also won the Hong Kong Film Award for Best Actor. ✗ [Not proactive enough]	
User:	Anything else?	
Bot:	Well, it's a documentary movie. It's worth to recommend to you. Don't miss it! ✗ [Not achieve the goal]	
User:	OK. Thank you.	
Bot:	You're welcome. Goodbye!	

(b)

Target: target action -- Movie Recommendation target topic -- "The Art of Action: Martial Arts in Motion Picture"

	Ours (Plan-controlled generation)	**Planned dialogue path by our TRIP**
User:	Do you know who starred in the movie Jackie Chan: My Story?	Respond Q&A | "Jackie Chan: My Story" → Chat about Stars | Jackie Chan → Movie Recommendation | "The Art of Action: Martial Arts in Motion Picture"
Bot:	Yes, it's Jackie Chan.	
User:	Well, I see. You are so brilliant!	Chat about stars | Jackie Chan → Music Recommendation | "The Art of Action: Martial Arts in Motion Picture"
Bot:	You're welcome. Jackie Chan is a Chinese martial arts superstar.	
User:	Yeah. I sincerely love him!	Chat about stars | Jackie Chan → Music Recommendation | "The Art of Action: Martial Arts in Motion Picture"
Bot:	Well, he has won the Golden Horse Award for Best Actor. He has many good movies. ✓ [More coherent]	
User:	Wow, that's great.	Music Recommendation | "The Art of Action: Martial Arts in Motion Picture"
Bot:	I'd recommend his movie The Art of Action: Martial Arts in Motion Picture to you.	
User:	Sounds good. ✓ [Succeed to achieve the goal]	Music Recommendation | "The Art of Action: Martial Arts in Motion Picture"
Bot:	It's a documentary movie expressing Hong Kong kung fu movies. Don't miss it!	
User:	OK. Thank you.	Say goodbye | NULL → Music Recommendation | "The Art of Action: Martial Arts in Motion Picture"
Bot:	Hmm, hope you like it.	
User:	Well. I may have a try. See you next time.	Say goodbye | NULL → Music Recommendation | "The Art of Action: Martial Arts in Motion Picture"
Bot:	Goodbye!	

(c)

Fig. 9. Illustrative cases from the dialogue-level human evaluation. The bot's utterances are generated by (a) MGCG_G, (b) GPT-2, and (c) Ours (plan-controlled generation), respectively. The topics and topic-related attributes that also appear in the domain knowledge are marked with underlines.

importantly, our method succeeds in achieving the target since our TRIP plans a correct topic (i.e., the target topic "*The Art of Action: Martial Arts in Motion Picture*") when appropriate.

7.7 Additional Discussions (RQ4)

According to the human evaluation results and case study, our proposed methods effectively plan reasonable dialogue paths to guide dialogue generation. The advantages of such a pipelined framework are: (1) It provides our model with better explanations because each planned dialogue path tells the dialogue generation model how to achieve the target step by step with specific actions and essential topics. (2) It is controllable for the end task of target-oriented dialogue generation. Our methods divide the complicated end task into two stages, making it more flexible to improve the overall performance stage by stage. Therefore, our methods are more practical and can be extended to real-world applications. After analyzing those cases with low human evaluation scores, we also identify some limitations and discuss the potential solutions: (1) Our pipelined framework has error propagation, which might be a typical issue of most existing pipelined methods. We find that the performance of dialogue generation is prone to drop once our TRIP model fails to plan a dialogue path appropriately. We intend to alleviate this issue by introducing some techniques in the cascaded generation, such as noisy channel models [27, 42]. (2) Our plan-guided dialogue generation method is still not robust enough. Although we have achieved significant good planning results with a large margin compared to baseline models on both datasets, we observe that the performance gain in terms of the goal success rate is much less prominent on one dataset than on another. One possible direction is to study how to improve dialogue generation with adaptive control when it is the turn with the target action and the target topic.

8 CONCLUSION AND FUTURE WORK

In this work, we explore the task of target-oriented proactive dialogue and focus on effective dialogue planning for dialogue generation. We propose a novel TRIP approach to plan dialogue paths from both backward and forward directions. Our TRIP formulates planning as a generation task and bidirectionally generates dialogue paths consisting of reasonable actions and appropriate topics. To better control path generation, we devise a novel target-constrained decoding algorithm to achieve bidirectional agreement. We adopt the planned dialogue paths to guide dialogue generation in a pipeline manner, with two explored variants: prompt-based generation and plan-controlled generation. Experimental results on two re-purposed datasets show that the proposed methods achieve state-of-the-art performance on all sub-tasks. Extensive analysis and discussions demonstrate the advantages of our methods.

We observe that the emergence of **large language models (LLMs)** [35, 48, 49] has unprecedentedly boosted the research field of dialogue systems. LLMs will generally perform better for dialogue generation in terms of some aspects, such as fluency, informativeness, and human likeness. However, for the target-oriented proactive dialogue generation task, more critical dimensions should be considered, including proactivity, coherence, and target achievement success rate. Our work shows that dialogue planning plays a vital role in improving dialogue generation performance in these dimensions. Recent studies [51, 52] indicate that the planning capabilities of LLMs are still far from that of humans. In the future, we intend to incorporate our proposed bidirectional approach based on LLMs for dialogue planning and generation since our methods are model-agnostic to backbone models. We are also interested in empowering the planning capabilities of LLMs to solve other complex tasks.

REFERENCES

[1] Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. 2020. Plug and play language models: A simple approach to controlled text generation. In *Proceedings of the International Conference on Learning Representations.*

[2] Yang Deng, Wenqiang Lei, Wai Lam, and Tat-Seng Chua. 2023. A survey on proactive dialogue systems: Problems, methods, and prospects. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence, IJCAI-23.* International Joint Conferences on Artificial Intelligence Organization, 6583–6591.

[3] Yang Deng, Wenxuan Zhang, Weiwen Xu, Wenqiang Lei, Tat-Seng Chua, and Wai Lam. 2023. A unified multi-task learning framework for multi-goal conversational recommender systems. *ACM Transactions on Information Systems* 41, 3 (2023), 1–25.

[4] 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*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186.

[5] Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76, 5 (1971), 378.

[6] Sarik Ghazarian, Zixi Liu, Tuhin Chakrabarty, Xuezhe Ma, Aram Galstyan, and Nanyun Peng. 2021. DiSCoL: Toward engaging dialogue systems through conversational line guided response generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations*. Association for Computational Linguistics, Online, 26–34.

[7] Ana Valeria Gonzalez, Isabelle Augenstein, and Anders Søgaard. 2019. Retrieval-based goal-oriented dialogue generation. arXiv:1909.13717. Retrieved from https://arxiv.org/abs/1909.13717

[8] 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*. Association for Computational Linguistics, Seattle, United States, 1301–1317.

[9] Shirley Anugrah Hayati, Dongyeop Kang, Qingxiaoyang Zhu, Weiyan Shi, and Zhou Yu. 2020. INSPIRED: Toward sociable recommendation dialog systems. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online, 8142–8152.

[10] 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*. 2288–2305.

[11] 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*. 6408–6423.

[12] 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*. 591–602.

[13] Xinyu Hua and Lu Wang. 2020. PAIR: Planning and iterative refinement in pre-trained transformers for long text generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. 781–793.

[14] L. T. Hunt, N. D. Daw, P. Kaanders, M. A. MacIver, U. Mugan, E. Procyk, A. D. Redish, E. Russo, J. Scholl, K. Stachenfeld, C. R. E. Wilson, and N. Kolling. 2021. Formalizing planning and information search in naturalistic decision-making. *Nature Neuroscience* 24, 8 (2021), 1051–1064.

[15] 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. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, Hong Kong, China, 1951–1961.

[16] Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Doha, Qatar, 1746–1751.

[17] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *The International Conference for Learning Representations (ICLR)*.

[18] Seanie Lee, Dong Bok Lee, and Sung Ju Hwang. 2021. Contrastive learning with adversarial perturbations for conditional text generation. In *The International Conference for Learning Representations (ICLR)*.

[19] Wenqiang Lei, Xiangnan He, Yisong Miao, Qingyun Wu, Richang Hong, Min-Yen Kan, and Tat-Seng Chua. 2020. Estimation-action-reflection: Towards deep interaction between conversational and recommender systems. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 304–312.

[20] 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*. 7871–7880.

[21] 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*. Association for Computational Linguistics, San Diego, California, 110–119.

[22] Raymond Li, Samira Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. Towards deep conversational recommendations. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. 9748–9758.

[23] Yu Li, Shirley Anugrah Hayati, Weiyan Shi, and Zhou Yu. 2021. DEUX: An attribute-guided framework for sociable recommendation dialog systems. arXiv:2105.00825. Retrieved from https://arxiv.org/abs/2105.00825

[24] Zujie Liang, Huang Hu, Can Xu, Jian Miao, Yingying He, Yining Chen, Xiubo Geng, Fan Liang, and Daxin Jiang. 2021. Learning neural templates for recommender dialogue system. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 7821–7833.

[25] Dongding Lin, Jian Wang, and Wenjie Li. 2021. Target-guided knowledge-aware recommendation dialogue system: An empirical investigation. In *Proceedings of the 3rd Edition of Knowledge-aware and Conversational Recommender Systems & 5th Edition of Recommendation in Complex Environments Joint Workshop @ RecSys 2021*.

[26] Dongding Lin, Jian Wang, and Wenjie Li. 2023. COLA: Improving conversational recommender systems by collaborative augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 4462–4470.

[27] Qi Liu, Lei Yu, Laura Rimell, and Phil Blunsom. 2021. Pretraining the Noisy Channel Model for Task-Oriented Dialogue. *Transactions of the Association for Computational Linguistics* 9 (2021), 657–674.

[28] Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, and Wanxiang Che. 2021. DuRecDial 2.0: A bilingual parallel corpus for conversational recommendation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 4335–4347.

[29] 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*. Association for Computational Linguistics, Online, 1036–1049.

[30] 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*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 1839–1851.

[31] Andrea Madotto, Zhaojiang Lin, Genta Indra Winata, and Pascale Fung. 2021. Few-shot bot: Prompt-based learning for dialogue systems. arXiv:2110.08118. Retrieved from https://arxiv.org/abs/2110.08118

[32] 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*. 1468–1478.

[33] 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*. 2267–2277.

[34] Vinod Nair and Geoffrey E. Hinton. 2010. Rectified linear units improve restricted boltzmann machines. In *ICML*.

[35] OpenAI. 2022. Introducing ChatGPT. Retrieved November 30, 2022 from https://openai.com/blog/chatgpt

[36] 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*. Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, 311–318.

[37] Matt Post and David Vilar. 2018. Fast lexically constrained decoding with dynamic beam allocation for neural machine translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. 1314–1324.

[38] 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*. 6908–6915.

[39] 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*. 8657–8664.

[40] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI Blog* 1, 8 (2019), 9.

[41] 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*. Association for Computational Linguistics, Online, 2492–2504.

[42] Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27, 3 (1948), 379–423.

[43] Zhihong Shao, Minlie Huang, Jiangtao Wen, Wenfei Xu, and Xiaoyan Zhu. 2019. Long and diverse text generation with planning-based hierarchical variational model. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. 3257–3268.

[44] Charlie Snell, Sherry Yang, Justin Fu, Yi Su, and Sergey Levine. 2022. Context-aware language modeling for goal-oriented dialogue systems. In *Findings of the Association for Computational Linguistics: NAACL 2022*. 2351–2366.

[45] 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*. 895–909.

[46] Yueming Sun and Yi Zhang. 2018. Conversational recommender system. In *Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval*. 235–244.

[47] 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* . Association for Computational Linguistics, Florence, Italy, 5624–5634.

[48] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. arXiv:2302.13971. Retrieved from https://arxiv.org/abs/2302.13971

[49] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models. arXiv:2307.09288. Retrieved from https://arxiv.org/abs/2307.09288

[50] Evangelos Triantaphyllou. 2000. Multi-criteria decision making methods. In *Proceedings of the Multi-Criteria Decision Making Methods: A Comparative Study*. Springer, 5–21.

[51] Karthik Valmeekam, Alberto Olmo, Sarath Sreedharan, and Subbarao Kambhampati. 2022. Large language models still can't plan (a benchmark for llms on planning and reasoning about change). In *Proceedings of the NeurIPS 2022 Foundation Models for Decision Making Workshop*.

[52] Karthik Valmeekam, Sarath Sreedharan, Matthew Marquez, Alberto Olmo, and Subbarao Kambhampati. 2023. On the planning abilities of large language models - a critical investigation. In *Proceedings of the Advances in Neural Information Processing Systems*. 75993–76005.

[53] 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 *Proceedings of the Advances in Neural Information Processing Systems*. 5998–6008.

[54] Jian Wang, Dongding Lin, and Wenjie Li. 2022. Follow Me: Conversation planning for target-driven recommendation dialogue systems. arXiv:2208.03516. Retrieved from https://arxiv.org/abs/2208.03516

[55] Jian Wang, Dongding Lin, and Wenjie Li. 2023. A target-driven planning approach for goal-directed dialog systems. *IEEE Transactions on Neural Networks and Learning Systems* (2023). https://doi.org/10.1109/TNNLS.2023.3242071

[56] Jian Wang, Junhao Liu, Wei Bi, Xiaojiang Liu, Kejing He, Ruifeng Xu, and Min Yang. 2020. Dual dynamic memory network for end-to-end multi-turn task-oriented dialog systems. In *Proceedings of the 28th International Conference on Computational Linguistics*. Donia Scott, Nuria Bel, and Chengqing Zong (Eds.), International Committee on Computational Linguistics, Barcelona, Spain (Online), 4100–4110.

[57] Jian Wang, Junhao Liu, Wei Bi, Xiaojiang Liu, Kejing He, Ruifeng Xu, and Min Yang. 2020. Improving knowledge-aware dialogue generation via knowledge base question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 9169–9176.

[58] Wenjie Wang, Minlie Huang, Xin-Shun Xu, Fumin Shen, and Liqiang Nie. 2018. Chat More: Deepening and widening the chatting topic via a deep model. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 255–264.

[59] Yida Wang, Pei Ke, Yinhe Zheng, Kaili Huang, Yong Jiang, Xiaoyan Zhu, and Minlie Huang. 2020. A large-scale chinese short-text conversation dataset. In *Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing*. Springer, 91–103.

[60] 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: System Demonstrations*. Association for Computational Linguistics, Online, 38–45.

[61] Chien-sheng Wu, Richard Socher, and Caiming Xiong. 2019. Global-to-local memory pointer networks for task-oriented dialogue. In *Proceedings of the 7th International Conference on Learning Representations*.

[62] Wenquan Wu, Zhen Guo, Xiangyang Zhou, Hua Wu, Xiyuan Zhang, Rongzhong Lian, and Haifeng Wang. 2019. Proactive human-machine conversation with explicit conversation goal. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 3794–3804.

[63] Hu Xu, Seungwhan Moon, Honglei Liu, Bing Liu, Pararth Shah, Bing Liu, and Philip Yu. 2020. User memory reasoning for conversational recommendation. In *Proceedings of the 28th International Conference on Computational Linguistics* . International Committee on Computational Linguistics, Barcelona, Spain (Online), 5288–5308.

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

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

[66] Kerui Xu, Jingxuan Yang, Jun Xu, Sheng Gao, Jun Guo, and Ji-Rong Wen. 2021. Adapting user preference to online feedback in multi-round conversational recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 364–372.

[67] Denis Yarats and Mike Lewis. 2018. Hierarchical text generation and planning for strategic dialogue. In *Proceedings of the International Conference on Machine Learning*. 5591–5599.

[68] 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*. Association for Computational Linguistics, Punta Cana, Dominican Republic, 1092–1101.

[69] Tong Zhang, Yong Liu, Peixiang Zhong, Chen Zhang, Hao Wang, and Chunyan Miao. 2021. KECRS: Towards knowledge-enriched conversational recommendation system. arXiv:2105.08261. Retrieved from https://arxiv.org/abs/2105.08261

[70] 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*. Association for Computational Linguistics, Online, 270–278.

[71] Chujie Zheng and Minlie Huang. 2021. Exploring prompt-based few-shot learning for grounded dialog generation. arXiv:2109.06513. Retrieved from https://arxiv.org/abs/2109.06513

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

[73] Kun Zhou, Wayne Xin Zhao, Shuqing Bian, Yuanhang Zhou, Ji-Rong Wen, and Jingsong Yu. 2020. Improving conversational recommender systems via knowledge graph based semantic fusion. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1006–1014.

[74] Kun Zhou, Yuanhang Zhou, Wayne Xin Zhao, Xiaoke Wang, and Ji-Rong Wen. 2020. Towards topic-guided conversational recommender system. In *Proceedings of the 28th International Conference on Computational Linguistics* . International Committee on Computational Linguistics, Barcelona, Spain (Online), 4128–4139.

[75] Yiheng Zhou, Yulia Tsvetkov, Alan W. Black, and Zhou Yu. 2020. Augmenting non-collaborative dialog systems with explicit semantic and strategic dialog history. In *Proceedings of the International Conference on Learning Representations* .

Received 15 September 2022; revised 18 December 2023; accepted 6 March 2024