Guideline Compliance in Task-Oriented Dialogue: The Chained Prior Approach

Anonymous ACL submission

Abstract

Task-oriented dialogue (TOD) systems are widely used across various domains, including customer service, appointment scheduling, and technical support. In real-world scenarios, such systems must adhere to given operational guidelines. However, existing solutions based on large language models often cannot achieve strict guideline compliance, even when fine-tuned with domain knowledge. To address this issue, we introduce a novel TOD system named *GuidedTOD*, which explicitly considers domain-specific guidelines by integrating a policy module. This module employs a Markov Chain, termed *Chained Prior*, to efficiently encode and dynamically update guideline knowledge. During inference, the Chained Prior reranks outputs from the domain-expert language model using beam search, ensuring guideline adherence. Experimental results show that *GuidedTOD* significantly improves guideline compliance, achieving approximately 20% better action prediction accuracy than state-of-the-art solutions. [1]

1 Introduction

Task-Oriented Dialogue (TOD) systems are designed to facilitate specific tasks, such as scheduling appointments, booking flights, or providing customer support. Traditionally, these tasks were managed by human agents who relied on detailed operational guidelines provided by company policies to resolve issues efficiently and accurately. The demand for faster and more scalable solutions has driven significant interest in developing automated TOD systems within both industry and academia.

The emergence of generalist large language models (LLMs) like GPT-4 (Achiam et al., 2023), Gemini (Team et al., 2023), and Llama (Touvron et al., 2023) has revolutionized natural language processing. These models benefit from extensive pretraining on diverse tasks, which enhances their ability to learn and adapt to new contexts (Wei et al., 2021; Brown et al., 2020). Consequently, there has been a surge in research integrating LLMs into TOD systems to improve response fluency and handle complex scenarios that traditional methods struggle with (King and Flanigan, 2024; Cao, 2023; Zeng et al., 2024; Kawamoto et al., 2023). Despite their potential, studies show that even fine-tuned LLMs often fall short in adhering to guidelines (Hudeček and Dušek, 2023; Lee et al., 2022).

Various approaches have been proposed to address this limitation. For instance, direct stimulus prompting, as described by Li et al. (2024), involves training smaller-scale policy models to provide domain-specific guidance to LLMs using curated service data. Similarly, Ramakrishnan et al. (2023) suggests using compact language models(LMs) to predict multiple next actions to improve accuracy. While these methods demonstrate better performance, there is ample room for improvement in guideline compliance, as guidelines are not explicitly integrated into TOD systems.

In this paper, we introduce a novel TOD system named *GuidedTOD*, aiming to bridge the compliance gap by explicitly incorporating operational guidelines into the TOD framework. Similar to Li et al. (2024), our system features a policy module that provides domain-specific insights to generalist LLMs. The unique feature of our policy module is that we equip it with a *Chained Prior* mechanism, formulated as a Markov Chain. This mechanism consists of states derived from the actions specified in the guidelines and transition probabilities calibrated using curated service data. During inference, by re-ranking the predicted next actions during the beam search of the expert language model, our system ensures guideline adherence. To the best of our knowledge, *GuidedTOD* is the first system to explicitly leverage operational guidelines for improved compliance, achieving significant enhancements even in the absence of explicit guidelines.

[1] Code will be made accessible upon acceptance.

We evaluate our method on two benchmarks to measure action prediction accuracy and dialogue consistency. Our results indicate that *GuidedTOD* outperforms existing methods, performing 50% better than GPT models using in-context learning and achieving approximately 20% better action prediction accuracy compared to state-of-the-art solutions. Notably, the Chained Prior mechanism provides further improvements when initial dialogue data is sparse, demonstrating the advantage of *GuidedTOD* in reducing the need for extensive human labor.

2 Background

2.1 Datasets and Conventional TOD Systems

Before 2021, most task-oriented dialogue (TOD) datasets, such as MultiWOZ (Zang et al., 2020), were collected using the Wizard-of-Oz (WoZ) technique (Mrkšić et al., 2016a), where a human operator simulates an AI system's responses. While effective in capturing authentic dialogues, these datasets generally lack specific operational guidelines. In contrast, the ABCD dataset introduced by Chen et al. (2021) includes comprehensive guidelines covering various service scenarios, setting a precedent for TOD systems to explicitly utilize guidelines to enhance compliance with business logic.

Traditionally, TOD systems have employed specialized models for different components of the dialogue process: natural language understanding (NLU) (Bates, 1995; Storks et al., 2019), dialogue state tracking (DST) (Mrkšić et al., 2016b; Rastogi et al., 2017; Ren et al., 2018), and natural language generation (NLG) (Gatt and Krahmer, 2018; Ji et al., 2023). While these compartmentalized models are effective, they often face integration challenges and lack flexibility.

2.2 LLM-Powered TOD Systems

LLMs provide a unified, adaptable framework that enhances flexibility and scalability, making LLM-powered TOD systems increasingly prevalent as they overcome traditional system limitations.

In particular, in-context learning has significantly advanced the application of LLMs. Research by Brown et al. (2020) and Wei et al. (2021) demonstrates that LLMs can generalize to new tasks with minimal examples due to extensive pre-training. Techniques such as "Prefix-Tuning" by Li and Liang (2021) enhance task-specific performance by optimizing continuous prompts, while Reynolds and McDonell (2021) have developed a systematic taxonomy of prompting techniques and established best practices.

Despite these innovations, achieving strict guideline compliance when integrating LLMs into TOD systems continues to be challenging. Research by Hudeček and Dušek (2023) and Lee et al. (2022) indicates that even fine-tuned LLMs have difficulty adhering to stringent guidelines. Although Bang et al. (2023) and Chen et al. (2023) have fine-tuned LLMs with domain-specific data, their performance is still eclipsed by the methods of Li et al. (2024).

To better integrate domain-specific knowledge, Li et al. (2024) develops a compact language model that analyzes dialogue history and predicts the next action. These predicted actions, along with the dialogue history, are then used to guide a frozen generalist LLM in generating responses. Similarly, Ramakrishnan et al. (2023) trains an LM to predict multiple subsequent actions, selecting the one with the highest probability to improve accuracy.

2.3 The Need for Explicit Guideline Integration

Although datasets like ABCD introduce the task of guideline-driven task-oriented dialogue, subsequent studies (Hattami et al., 2022; Ramakrishnan et al., 2023; Li et al., 2024) have often overlooked these guidelines, opting instead to train models exclusively on curated service data.

ComplianceOPT (Min et al., 2023) implements online reinforcement learning with a reward model designed to evaluate whether responses adhere to the guidelines. While this approach has led to some improvements in guideline compliance, it relies implicitly on guidelines during the training of the reward model and is notably inefficient due to the demands of online learning. This highlights the need for innovative approaches that explicitly integrate operational guidelines into TOD systems to enhance compliance and efficiency.

3 Preliminaries & Problem Formulation

3.1 Preliminaries: Markov Chains

Markov chains model sequences of events as transitions between states with probabilities defined solely by the current state, a property known as the Markov property.

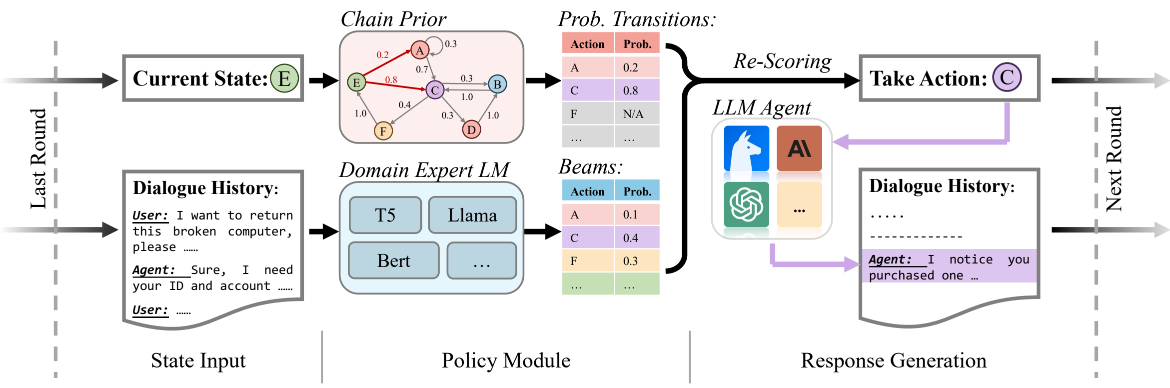


Figure 1: Framework that delineates the processes of the Chained Prior Guided Task-oriented Dialogue (*GuidedTOD*) System. Given the utterance context, the domain-expert LM predicts several possible actions using beam search techniques. The Chained Prior proposes the subsequent possible states with certain transition probabilities from the current state, and acts as a scorer to re-rank the beam actions. The predicted actions are then used to steer the LLM to generate the response that complies with the next utterance.

- **States:** Markov chains consist of a finite or countable number of states representing various conditions or system positions.
- **Transition Probabilities:** These define the likelihood of transitioning from one state to another and are crucial for predicting future states based on the current state.

Markov chains are widely used across various disciplines, such as physics and finance, due to their capability to model stochastic systems dynamically. Similarly, by systematically encoding and updating domain-specific guidelines, the application of Markov chains holds great potential to significantly enhance the accuracy and reliability of TOD systems, but it remains unexplored.

3.2 Problem Formulation

Given a dialogue context $X_t = [x_1, x_2, ..., x_t]$ and all the possible legal actions $A = [a_1, a_2, ..., a_N]$, our objective is to predict the subsequent action a_{t+1}. The context X_t contains the past system utterance s_i and user utterance u_i. An action $a = (a, v)$ includes an action name (for example, '*pull-up-account*'), and an optional list of action values v (for example, '[*crystal, minh*]').

To predict future actions, we assume access to pre-defined guidelines and a historical dataset of dialogues paired with ground truth actions. Given the currently available R dialogues, we extract all the actions from a dialogue and arrange them as a sequence using the workflow discovery proposed by Hattami et al. (2022). In this case, we may obtain R groups of actions: [[*pull-up-account*, *verify-identity*, ...], [*enter-details*, *log-out-in*, ...], ...], and formalize them as the data set of dialogues with action flows (X, F). A Chained Prior is a weighted, directed graph generated from both the guidelines and training set, with vertices $V = \{a_1, ..., a_N\}$ and edges $E = \{e_{a_i \to a_j}\}$. A guideline contains action sequences as workflows to achieve specific goals in different scenarios, as illustrated in Table 1. The structure and connectivity of Chained Prior are formed by merging the sequences of pre-defined actions on the guideline. All the possible actions are formed as the states in the Chained Prior. The weights of the edges reflect the actual frequency of transiting from one action to the other in the training set. A domain-expert LM using the beam search technique predicts the next action with $top - K$ candidates $C = [c_1, c_2, ..., c_K]$ given the context of X_t, where ideally the candidate set C should be a subset of A.

Given the context up to step t, X_t, a Chained Prior guided next action prediction is to select the action with the highest weighted sum of the probabilities from the Chained Prior and the policy model, that is,

$$P(C_i \mid X_t; \theta_{expert}, \theta_{cp}) \propto P(C_i \mid X_t) P(e_{a_t \to C_i}). \quad (1)$$

4 Methodology

As depicted in Figure 1, the new policy module in our *GuidedTOD* system consists of two components: the domain-expert language model and the Chained Prior.

The Chained Prior is developed based on task guidelines and statistical transition probabilities derived from the training set. Concurrently, the domain-expert LM is specifically trained solely on the training data to predict the likely next action. During inference, the policy module determines the subsequent action by jointly considering the outputs from both the Chained Prior and the domain-expert LM. This integrated approach ensures that decision-making is both contextually informed and statistically grounded.

4.1 Construction of Policy Module

4.1.1 Chained Prior

As a Markov Chain model, we construct the Chained Prior as Algorithm 1.

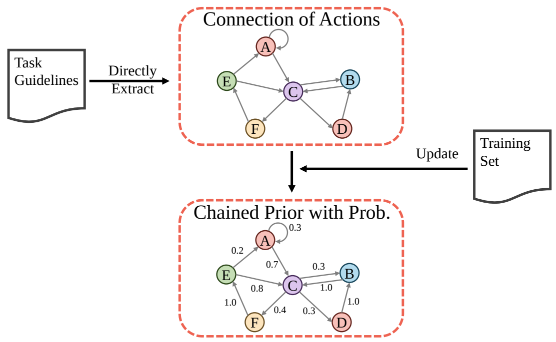


Figure 2: Initialize the Chained Prior using guidelines and update transition probabilities with the training set.

Step 1: Initialization with Guidelines. As shown in Figure 2, we first establish a graph of actions according to the guidelines. Initially, the Markov Chain consists of two states: "INIT" and "END". We incorporate pre-defined action flows by treating each action within the flow as a potential state. The "INIT" state is directly linked to the first action of each flow through a directed edge, integrating this action into the Markov Chain's state set. This process is repeated for every pair of adjacent actions, establishing a connection from each action to the next. The final action in each flow is connected to the "END" state. This construction results in a Markov Chain that includes actions as state nodes and connections as edges, although it initially lacks transition probabilities.

Step 2: Computing Transition Probabilities with Dialogue Data.

To establish a valid Markov Chain model, we still need the probabilities of each transition. Initially, we first count the transition counts between each pair of possible states from the action flows of each dialogue. These counts, denoted as $C(s, s')$, serve as the foundation for updating the transition probabilities of our initialized Chained Prior.

Subsequently, we convert these raw transition counts into probabilities. This step involves normalizing the counts so that the sum of the transition probabilities from any given state s to all other possible states s' is equal to one. This normalization is crucial to ensure that our model functions as an effective Markov Chain, accurately reflecting the stochastic nature of dialogue transitions based on empirical data.

Step 3: Building Sub-graph for Each Scenario To further extract useful knowledge from the guidelines and improve the quality of Chained Prior, we revisit the structure of the guidelines. The guideline that we can use usually divides all the dialogues into several scenarios, such as the "Product Defect" and "Shipping Issue" shown in Table 1. Using such prior knowledge, we can extract the pre-defined legal flows for each scenario, and repeat the above two steps to obtain refined Chained Priors for each scenario. These Chained Priors are sub-graphs of

---

Algorithm 1 Construct Chained Prior from Dialogue Data with Guidelines

Input: Training set of dialogues with action flows (X, F), Guidelines of pre-defined legal action flows F^*, All possible actions $A = \{a_1, a_2, ..., a_N\}$
Output: A Chained Prior M with probability P

// Initialization:
Initialize an empty Markov Chain M with state $S = \{\text{INIT}, \text{END}\}$
for each action flow f in F^* **do**
  Add all action in $A_f \subseteq A$ of the action flow f to S as new states
  Connect $a_i \rightarrow a_j \in A_f, i < j$ with directed edge in M without probabilities
end for
// Count transitions:
Initialize transition counts $C(s, s')$ for all $s, s' \in S$
for each dialogue action flow f in Training set (X, F) **do**
  Set current state $\bar{s} \leftarrow \text{INIT}$
  for each action a_f in dialogue action flow f **do**
    $C(\bar{s}, a_f) += 1$
    $\bar{s} \leftarrow a_f$
  end for
  $C(\bar{s}, \text{END})$
end for
// Compute transition probabilities:
for each state $s \in S$ **do**
  Total transitions from s: $T(s) = \sum_{s' \in S} C(s, s')$
  for each state $s' \in S$ where $C(s, s') > 0$ **do**
    Set transition probability $P(s \rightarrow s') = \frac{C(s,s')}{T(s)}$
  end for
end for

Scenario	Goal	Flow
Product Defect	Initiate Refund	1. Pull up Account
		2. Validate Purchase
		3. Record Reason
		4. Enter Details
		5. Offer Refund
	Return Due to Stain	1. Pull up Account
		2. Validate Purchase
		3. Membership Privileges
		4. End Conversation
		5. Enter Details
		6. Update Order
Shipping Issue	Manage Shipping	1. Pull up Account
		2. Shipping Status
		3. Validate Purchase
		4. Update Order

Table 1: Legal dialogue action flows in the guidelines of ABCD dataset.

the previous global one built on all the scenarios.

After defining and initializing the Chained Prior as outlined above, the domain-expert language model can leverage this structured probabilistic model as a scorer to re-rank its predictions. This enhances the performance in generating and predicting legal dialogue actions, thus aligning more closely with human expectations and real-world scenarios.

4.1.2 Domain Expert Language Model

A domain-expert language model (LM) is trained specifically for the task-oriented dialogue scenario using the training set $\mathcal{D} = \{(x, a)\}$ of pairs of dialogue annotated with the action flows. Given the current dialogue context, the domain-expert LM predicts the next action a and its slot values v. Similar to DSP (Li et al., 2024), We perform supervised fine-tuning on a language model to predict next actions. Each pair of data for training contains the utterance context $x \in X$ and the current action $a \in A$ with slot values $v = \{v_1, v_2, ..., v_m\}$. We then fine-tune the language model by maximizing the log-likelihood:

$$\mathcal{L}(\theta) = \sum_{(x,a)\in\mathcal{D}} \log P(a|x; \theta_{expert}) \qquad (2)$$

where θ_{expert} represents the parameters of the domain-expert LM, $P(a, \{v_1, v_2, \ldots, v_m\}|x; \theta)$ is the probability of predicting the action a and its corresponding slot values v given the dialogue context x, parameterized by θ.

4.2 Chained Prior Guided TOD System

During inference, we introduce a mechanism that harmonizes the strengths of both models. As shown in Figure 1, we jointly consider the output of Chained Prior and the domain-expert LM for the next step. Specifically, the Chained Prior acts as a scorer to re-rank the prediction by the domain-expert LM. The input (left of Figure 1) of each turn of inference contains a state indicator and the conversation context. The state indicator is the last action executed, while the conversation context includes all user-system interactions.

To increase the search space in producing the next action prediction, we employ the beam search to sample multiple outputs from the domain-expert LM. In each dialogue round, the LM first analyzes the previous dialogue history to predict multiple potential next actions and their associated probabilities. Subsequently, we retrieve the transition probabilities from the current state to each predicted action from the Chained Prior. Re-ranking is performed by calculating the weighted sum of these probabilities, as shown in Eq. 3. In Section 5, we demonstrate that this approach significantly enhances guideline compliance, as evidenced by markedly improved precision in action prediction.

$$\begin{aligned} \operatorname*{argmax}_{C_i} \, [\alpha \log(P(C_i \mid X_t)) + \\ (1-\alpha)\log(P(e_{a_t \to C_i}))]. \end{aligned} \qquad (3)$$

Finally, we utilize the action with greatest summed log probability, as the definitive action suggestion to direct the agent in producing subsequent responses (right of Figure 1).

5 Experiments

5.1 Experimental Settings

We use ABCD (Chen et al., 2021) and MultiWoz (Zang et al., 2020) as our two primary datasets for evaluation. Both datasets have a set of dialogues along with corresponding workflows created by Workflow Discovery (Hattami et al., 2022), making them ideal for testing the capability of our method in completing the TOD tasks and ensuring compliance with guidelines.

ABCD provides a large-scale benchmark for action-based conversational data. It comprises over 10K dialogues involving two human participants, covering a wide variety of interactions with 55 different user intents within the customer service field. It contains 10 different dialogue scenarios and 30

Datasets	Methods	Action Sequence level						Dialogue Level		
		Action CE	Value CE	Joint CE	Action EM	Value EM	Joint EM	BLEU	ROUGE-L	Bert Score
ABCD	ICL(GPT-3.5)	0.126	0.182	0.069	0.036	0.093	0.019	0.169	0.291	0.196
	ICL(GPT-4-Turbo)	0.184	0.262	0.114	0.062	0.131	0.023	0.234	0.345	0.239
	ICL(GPT-4)	0.198	0.279	0.135	0.077	0.135	0.048	0.260	0.388	0.275
	DSP	0.511	0.601	0.482	0.365	0.472	0.333	0.349	0.532	0.364
	Multi-step	0.501	0.599	0.469	0.354	0.473	0.322	0.337	0.523	0.349
	Ours	**0.737**	**0.728**	**0.692**	**0.682**	**0.663**	**0.619**	**0.432**	**0.624**	**0.430**
MultiWoz	ICL(GPT-3.5)	0.408	0.163	0.122	0.354	0.096	0.067	0.185	0.345	0.374
	ICL(GPT-4-Turbo)	0.474	0.080	0.063	0.465	0.043	0.028	0.204	0.381	0.420
	ICL(GPT-4)	0.619	0.302	0.25	0.254	0.126	0.102	0.194	0.364	0.405
	DSP	0.613	0.518	0.507	0.652	0.509	0.497	0.212	0.389	0.443
	Multi-step	0.567	0.470	0.449	0.591	0.449	0.425	0.225	0.409	0.461
	Ours	**0.665**	**0.565**	**0.548**	**0.730**	**0.571**	**0.551**	**0.238**	**0.426**	**0.479**

Table 2: The main results. We compare our method with the baselines of large language models using in-context learning (ICL), DSP, and Multi-Step within the TOD system setting. Our method shows significant improvement in both action prediction and conversation response generation.

unique actions. What distinguishes this dataset as particularly useful for our purposes is that it includes conversations where the agent adheres to specific guidelines, ensuring that an established workflow directs the exchanges.

MultiWoz offers a diverse set of dialogues within various domains, comprising over 10,000 dialogues. Previous studies have developed workflows for MultiWoz, which serve as the benchmark workflows for our training. However, we observed that MultiWoz offers a narrower range of workflow actions, featuring only 12 unique actions.

We choose two sets of metrics from the perspective of action prediction by the policy module and the response generation by the LLMs. Specifically, following the same setting in the work of Ramakrishnan et al. (2023), we use cascading evaluation (CE) and exact match (EM) to evaluate our method of predicting the next actions on the above two benchmarks. CE performs less strictly than EM since it gives partial credit to the sub-sequence of the predicted action sequences. In addition, we also employ BLEU, ROUGE, and BertScore to calculate the consistency of the generated responses by LLMs under the guidance of the policy module with the ground-truth utterances. Since the current implementations of both EM and CE haven't been fixed and are not very clear, we revisit the implementation and provide to make it more reasonable. The details of the metrics are in Appendix A.2.

The following models are used in our study: T5-small (Raffel et al., 2020), GPT-3.5-turbo, GPT-4, and GPT-4-turbo. We fine-tune the T5-small model as the domain-expert language model, and use the GPT models to serve as the LLM agents for response generation in a TOD system. All experiments are conducted using A800 GPUs and an Intel(R) Xeon(R) Platinum 8358 CPU. For more details on training and model settings, please refer to Appendix A.1.

5.2 Main Results

Table 2 demonstrates that our method significantly surpasses all baseline models in a range of metrics. By integrating a domain-expert language model (LM) with a Chained Prior within our new policy module, we achieve outstanding performance in Cascading Evaluation (CE) and Exact Match (EM) at the action sequence level, for action, value, and joint assessments. Notably, our method exceeds the best-performing baselines by approximately 20%.

As the results indicate, other policy-enhanced methods, such as DSP and Multi-step, also outperform traditional large language models utilizing in-context learning (ICL), as shown in Table 2. In contrast, our method shows a remarkable performance boost of over 50% compared to these ICL methods, showing a substantial 20% advantage compared to the baselines that incorporate a domain-expert LM in their policy modules. This underscores the effectiveness of our Chained Prior in enhancing action prediction accuracy.

At the dialogue level, our approach consistently delivers state-of-the-art performance across all metrics, indicating that the Chained Prior contributes to more precise and reliable responses that align well with human utterances.

It is important, however, to acknowledge variations in performance across different datasets. For example, in the MultiWoz dataset, where scenario definitions are less explicit, we implement a more generalized version of the Chained Prior, resulting in a more modest improvement of about 10%. This contrast is stark compared to the more than 20% improvement observed in the ABCD dataset. Such variance emphasizes the influence of dataset-specific characteristics on the efficacy of our approach and suggests that customizing the Chained Prior to specific dataset guidelines could further enhance performance.

5.3 Case Study: Chained Prior as a Plug-and-Play Module

The Chained Prior is designed to model guidelines and support the domain-expert language model (LM) by explicitly incorporating pre-defined knowledge of legal action flows during inference. Crucially, the Chained Prior operates effectively by requiring only the current state, derived from the historical context, to calculate the probabilities for the next transitions. This efficiency allows the Chained Prior to function as a plug-and-play module, readily integrable with most task-oriented dialogue (TOD) systems that utilize a policy module to predict subsequent actions.

We demonstrate this feature by directly integrating our Chained Prior with two domain-expert LM oriented baseline methods, i.e. DSP and Multi-step.

Methods	Strategy	Cascading Evaluation			Exact Match		
		Action	Value	Joint	Action	Value	Joint
DSP	w/o CP	0.511	0.600	0.481	0.364	0.472	0.332
	with CP	**0.604**	**0.662**	**0.560**	**0.462**	**0.552**	**0.411**
Multi-Step	w/o CP	0.500	0.599	0.469	0.353	0.473	0.321
	with CP	**0.604**	**0.651**	**0.568**	**0.480**	**0.546**	**0.438**

Table 3: Chained Prior works as a plugin to enhance other methods.

The experimental results shown in Table 3 support our insights. The results demonstrate that when using the Chained Prior, both baseline methods (DSP and Multi-Step) exhibit an approximately 10% improvement in action prediction accuracy as measured by both the CE and EM metrics, more details under different data scales and datasets can be found in Appendix A.3.2.

5.4 Ablation Studies

We conduct a series of ablation studies to show the effectiveness of our Chained Prior and the new policy module. Specifically, we first evaluate the impact of the Chained Prior in the policy module. Then, we explore the most proper hyper-parameters for balancing the Chained Prior and domain-expert LM. Finally, we quantify the benefit of creating the refined Chained Prior compared to the general global one, which does not consider the scenarios in the guidelines.

5.4.1 The Impact of Chained Prior

To show the significant boost of our Chained Prior, we conduct the comparison on the benchmarks with our method with or without the Chained Prior.

Datasets	Strategy	Cascading Evaluation			Exact Match		
		Action	Value	Joint	Action	Value	Joint
ABCD	w/o CP	0.517	0.608	0.487	0.375	0.487	0.345
	with CP	**0.737**	**0.728**	**0.692**	**0.682**	**0.663**	**0.619**
MultiWoz	w/o CP	0.613	0.518	0.507	0.652	0.509	0.497
	with CP	**0.665**	**0.565**	**0.548**	**0.730**	**0.571**	**0.551**

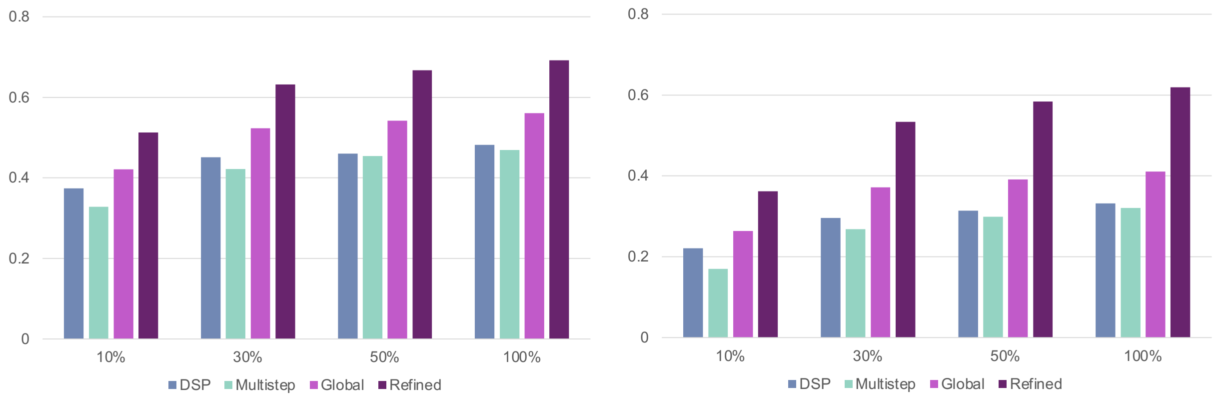
Table 4: Chained Prior enhances the action prediction accuracy.

Table 4 presents the impact of the Chained Prior (CP) on the TOD system for action prediction. We observe a dramatic decrease of up to 22% in the action prediction accuracy when excluding the Chained Prior. This observation demonstrates the essential role of the Chained Prior in the Guided-TOD system.

5.4.2 Balance between Chained Prior and Domain-Expert LM

It is crucial to balance the decision of two parts in our policy module which can lead to the optimal prediction result. Referring to Eq. 3, we have defined a hyper-parameter α, which indicates the weight of the domain-expert LM on the final output. We use such a notation in further experiments.

Tabel 5 shows the results of accuracy on the action prediction. For each dataset, we set four different weights, $\alpha \in \{0.6, 0.8, 0.9, 0.98\}$, for combining the Chained Prior and the domain-expert LM. For example, $\alpha = 0.6$ means that we add 0.6 of the log probability of the action beam of



(a) CE (Joint) (b) EM (Joint)

Figure 3: Refined Chained Prior improves the action (joint) prediction accuracy.

Dataset	α	Cascading Evaluation			Exact Match		
		Action	Value	Joint	Action	Value	Joint
ABCD	0.6	0.667	0.670	0.624	0.584	0.586	0.531
	0.8	0.718	0.714	0.674	0.657	0.643	0.597
	0.9	**0.737**	**0.728**	**0.692**	**0.682**	**0.663**	**0.619**
	0.98	0.699	0.699	0.659	0.616	0.609	0.561
MultiWoz	0.6	0.617	0.524	0.509	0.681	0.531	0.516
	0.8	0.643	0.544	0.529	0.709	0.553	0.537
	0.9	**0.665**	**0.565**	0.548	**0.730**	**0.571**	**0.551**
	0.98	0.664	0.562	**0.549**	0.724	0.566	0.550

Table 5: Evaluated using different weight parameters between the Chained Prior and the domain-expert LM.

the domain-expert LM to 0.4 of the log transition probabilities from the current state. The empirical results show that we get the best performance with $\alpha = 0.9$. We report all the results in other parts with this setting.

Scenarios		
Product Defect	Order Issue	Account Access
Troubleshoot Site	Manage Account	Purchase Dispute
Shipping Issue	Subscription Inquiry	Single-Item Query
Storewide Query		

Table 6: Scenarios in Guidelines.

5.4.3 Refined Chained Prior Considering Scenarios in Guideline

We further refine the Chained Prior based on the fine-grained scenarios defined in the guideline and obtain 10 distinct Chained Prior graphs, as shown in Table 6. Our experiment shows that the fine-grained Chained Prior can further boost the performance of GuidedTOD system.

Figure 3 shows the comparison results before and after refining the Chained Prior. We observe a 10% increase in the CE metric after using the refined Chained Prior. This demonstrates that considering the scenarios in the guideline benefits action prediction. Moreover, our *GuidedTOD* that trained with only 10% of the training data achieves comparable performance with the baselines. It outperforms the baselines by over 15% with only 30% of the training data, while the baselines are trained on the whole training set. Refer to Appendix A.3.3 for more results. These results demonstrate the efficiency of our method in reducing the need for human labor and resources in the real world, where data annotation is costly.

6 Conclusion

In this paper, we introduced *GuidedTOD*, a novel Chained Prior Guided TOD system designed to address the challenges of guideline adherence in automated customer support and similar applications. By integrating a policy module with a Markov Chain mechanism called *Chained Prior*, our system dynamically encodes and updates domain-specific guidelines, enhancing the guideline compliance of generalist LLMs. Experimental results demonstrate that *GuidedTOD* significantly surpasses existing solutions in action prediction accuracy, performing 50% better than GPT models using in-context learning and achieving approximately 20% better action prediction accuracy compared to state-of-the-art solutions. These findings highlight the efficacy of incorporating structured guideline knowledge directly into the model's decision-making process.

7 Limitations

In this section, we discuss the limitations of our methods and how they inform our future work. Our discussion focus on two main aspects: the accuracy of action prediction and the use of predicted actions to steer the LLMs.

Dataset	Subsequence Length	Action CE	Value CE	Joint CE
ABCD	1	0.873	0.866	0.836
	2	0.767	0.765	0.719
	3	0.570	0.553	0.520
	4	0.259	0.230	0.213
MultiWoz	1	0.904	0.808	0.798
	2	0.689	0.574	0.555
	3	0.403	0.312	0.291
	4	0.120	0.103	0.084

Table 7: Evaluation results on different sub-sequences.

When calculating the cascading evaluation (CE) metric, we define a hyper-parameter for the sub-sequence length, ranging from 1 to k. Different sub-sequence lengths indicate the rate of contiguous accurate action predictions. Table 7 displays the outcomes of our method as we modify the sub-sequence length for calculating the CE metric. The results indicate that as the sub-sequence length increases, there is a significant decrease in CE accuracy. This demonstrates that while our method has significantly outperformed existing solutions, the policy module still struggles to predict longer sequences of correct actions. Addressing this issue is a key area for future work. Moreover, the MultiWOZ dataset, which has fewer possible actions and dialogue scenarios, shows a more rapid decrease in accuracy compared to the ABCD dataset as the sub-sequence length increases. This might be attributed to the MultiWOZ dataset containing more unrelated random jumps between different scenarios, making it challenging to predict the next action from one scenario to another. This motivates us to explore a more powerful Chained Prior that could bridge different domains while maintaining high performance under complex action flows.

Despite the significant improvement in action prediction using the newly introduced Chained Prior, we believe that the current integration of the policy module and the LLMs may not fully leverage the potential of LLMs in generating conversation responses. Therefore, based on the current *GuidedTOD* system, we are working on developing a new module that enables LLMs to respond not only according to the actions but also in a manner that is preferred by humans.

Future Work In a more practical setting, we intend to deploy our *GuidedTOD* system within a simulated environment to enable authentic user-system interactions. Traditional testing and evaluation methods in existing benchmarks often rely on predetermined ground truth actions or user utterances to trigger subsequent system responses, an approach that does not accurately reflect real-world scenarios.

To address this, we propose using Large Language Models (LLMs) as agents to simulate the roles of both users and systems. Our goal is to generate genuine conversations, thereby improving both the Chained Prior and the domain-expert LM within a continual learning paradigm, particularly in situations with limited initial data. To facilitate this, we will introduce a novel metric focused on evaluating conversations according to how well they fulfill task guidelines and align with human preferences. By focusing on dialogues that rigorously follow specified action flows, we plan to enhance the Chained Prior by updating its transition probabilities and to refine the domain-expert LM through self-supervised fine-tuning or reinforcement learning. This strategy aims to significantly advance the *GuidedTOD* system automatically while reducing the reliance on annotating costs.

References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Namo Bang, Jeehyun Lee, and Myoung-Wan Koo. 2023. Task-optimized adapters for an end-to-end task-oriented dialogue system. *arXiv preprint arXiv:2305.02468*.

Madeleine Bates. 1995. Models of natural language understanding. *Proceedings of the National Academy of Sciences*, 92(22):9977–9982.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Lang Cao. 2023. Diaggpt: An llm-based chatbot with automatic topic management for task-oriented dialogue. *arXiv preprint arXiv:2308.08043*.

Derek Chen, Howard Chen, Yi Yang, Alex Lin, and Zhou Yu. 2021. Action-based conversations dataset: A corpus for building more in-depth task-oriented dialogue systems. *arXiv preprint arXiv:2104.00783*.

Zhi Chen, Yuncong Liu, Lu Chen, Su Zhu, Mengyue Wu, and Kai Yu. 2023. Opal: ontology-aware pretrained language model for end-to-end task-oriented dialogue. *Transactions of the Association for Computational Linguistics*, 11:68–84.

Albert Gatt and Emiel Krahmer. 2018. Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 61:65–170.

Amine El Hattami, Stefania Raimondo, Issam Laradji, David Vázquez, Pau Rodriguez, and Chris Pal. 2022. Workflow discovery from dialogues in the low data regime. *arXiv preprint arXiv:2205.11690*.

Vojtěch Hudeček and Ondřej Dušek. 2023. Are llms all you need for task-oriented dialogue? *arXiv preprint arXiv:2304.06556*.

Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38.

Tatsuki Kawamoto, Takuma Suzuki, Ko Miyama, Takumi Meguro, and Tomohiro Takagi. 2023. Application of frozen large-scale models to multimodal task-oriented dialogue. *arXiv preprint arXiv:2310.00845*.

Brendan King and Jeffrey Flanigan. 2024. The power of the noisy channel: Unsupervised end-to-end task-oriented dialogue with llms. *arXiv preprint arXiv:2404.15219*.

Sang-Woo Lee, Sungdong Kim, Donghyeon Ko, Donghoon Ham, Youngki Hong, Shin Ah Oh, Hyunhoon Jung, Wangkyo Jung, Kyunghyun Cho, Donghyun Kwak, et al. 2022. Can current task-oriented dialogue models automate real-world scenarios in the wild? *arXiv preprint arXiv:2212.10504*.

Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation. *arXiv preprint arXiv:2101.00190*.

Zekun Li, Baolin Peng, Pengcheng He, Michel Galley, Jianfeng Gao, and Xifeng Yan. 2024. Guiding large language models via directional stimulus prompting. *Advances in Neural Information Processing Systems*, 36.

Do June Min, Paloma Sodhi, and Ramya Ramakrishnan. 2023. Workflow-guided response generation for task-oriented dialogue. *arXiv preprint arXiv:2311.08300*.

Nikola Mrkšić, Diarmuid O Séaghdha, Tsung-Hsien Wen, Blaise Thomson, and Steve Young. 2016a. Neural belief tracker: Data-driven dialogue state tracking. *arXiv preprint arXiv:1606.03777*.

Nikola Mrkšić, Diarmuid O Séaghdha, Tsung-Hsien Wen, Blaise Thomson, and Steve Young. 2016b. Neural belief tracker: Data-driven dialogue state tracking. *arXiv preprint arXiv:1606.03777*.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

Ramya Ramakrishnan, Ethan Elenberg, Hashan Narangodage, and Ryan McDonald. 2023. Multi-step dialogue workflow action prediction. *arXiv preprint arXiv:2311.09593*.

Abhinav Rastogi, Dilek Hakkani-Tür, and Larry Heck. 2017. Scalable multi-domain dialogue state tracking. In *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 561–568.

Liliang Ren, Kaige Xie, Lu Chen, and Kai Yu. 2018. Towards universal dialogue state tracking. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2780–2786.

Laria Reynolds and Kyle McDonell. 2021. Prompt programming for large language models: Beyond the few-shot paradigm. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–7.

Shane Storks, Qiaozi Gao, and Joyce Y Chai. 2019. Commonsense reasoning for natural language understanding: A survey of benchmarks, resources, and approaches. *arXiv preprint arXiv:1904.01172*, pages 1–60.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Yonghui Wu, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. 2021. Finetuned language models are zero-shot learners. *arXiv preprint arXiv:2109.01652*.

Xiaoxue Zang, Abhinav Rastogi, Srinivas Sunkara, Raghav Gupta, Jianguo Zhang, and Jindong Chen. 2020. Multiwoz 2.2: A dialogue dataset with additional annotation corrections and state tracking baselines. *arXiv preprint arXiv:2007.12720*.

Weihao Zeng, Dayuan Fu, Keqing He, Yejie Wang, Yukai Xu, and Weiran Xu. 2024. Divtod: Unleashing the power of llms for diversifying task-oriented dialogue representations. *arXiv preprint arXiv:2404.00557*.

A Appendix

A.1 Training Details

Model	Strategy	Hyper-parameters
T5-small	Supervised fine-tuning	60M Batch size: 32 Epochs: 100 Dataset size: 8000 dialogues Training data: 10%, 30%, 50%, 100% Beam num: 4
GPT-3.5-turbo GPT-4 GPT-4-turbo	In-context learning for action prediction	Temperature: 0.3 Temperature: 0.7 Max tokens: 256
GPT-3.5-turbo	In-context response generation	Temperature: 0.7 Max tokens: 256

Table 8: Model settings.

A.2 Metrics

these are detailed descriptions of metrics that are used to evaluate our methods.

Algorithm 2 Exact Match for Action Prediction in Dialogue Systems

Input: Predicted actions with slot values in all dialogues: $A_{\text{label}}, A_{\text{pred}}, V_{\text{label}}, V_{\text{pred}}$
Output: Exact match metrics: $EM_{\text{action}}, EM_{\text{value}}, EM_{\text{joint}}$

// Preparation:
Group **A**ctions by dialogue ids ID: $GA_{\text{label}}, GA_{\text{pred}}$
Group **V**alues by dialogue ids ID: $GV_{\text{label}}, GV_{\text{pred}}$
Successful dialogue action counter $C_{action} \leftarrow 0$
Successful dialogue value counter $C_{value} \leftarrow 0$
Successful dialogue joint counter $C_{joint} \leftarrow 0$
Get all the possible dialogue IDs: ID

// Count successful dialogues:
for each dialogue id in ID **do**
  if $GA^{id}_{pred} = GA^{id}_{label}$ **then**
    $C_{action} = C_{action} + 1$
  end if
  if $GV^{id}_{pred} = GV^{id}_{label}$ **then**
    $C_{value} = C_{value} + 1$
  end if
  if $GA^{id}_{pred} = GA^{id}_{label}$ & $GV^{id}_{pred} = GV^{id}_{label}$ **then**
    $C_{joint} = C_{joint} + 1$
  end if
end for

// Calculate metrics:
Total number of dialogues: $T(ID)$
$EM_{action} = \frac{C_{action}}{T(ID)}$
$EM_{value} = \frac{C_{value}}{T(ID)}$
$EM_{joint} = \frac{C_{joint}}{T(ID)}$

A.2.1 Exact Match (EM)

Exact Match (EM) is the process of performing a precise comparison between the true actions and the predicted actions. Different versions of EM evaluate the action name (action), slot values (value), and both simultaneously (jointly).

A.2.2 Cascading Evaluation (CE)

Algorithm 3 Cascading Evaluation for Action Prediction in Dialogue Systems

Input: Predicted actions with slot values in all dialogues: A_{label}, A_{pred}, V_{label}, V_{pred}, Sub-sequence length: $L = \{1, 2, 3, \cdots, k\}$
Output: Cascading Evaluation metrics: CE_{action}, CE_{value}, CE_{joint}

// Preparation:
Group **A**ctions by dialogue ids ID: GA_{label}, GA_{pred}
Group **V**alues by dialogue ids ID: GV_{label}, GV_{pred}
Get all the possible dialogue IDs: ID
for each sub-sequence length l in L **do**
    // Count successful sub-sequences:
    Set successful sub-sequence counter: C^l_{action}, C^l_{value}
    for each dialogue id in ID **do**
        Obtain the **G**rouped **A**ctions in sub-**S**equence of length l: $GAS^{l,id}_{\text{label}}$, $GAS^{l,id}_{\text{pred}}$
        Obtain the **G**rouped **V**alues in sub-**S**equence of length l: $GVS^{l,id}_{\text{label}}$, $GVS^{l,id}_{\text{pred}}$
    end for
    Obtain all the possible sub-sequences of actions and values of length l: GAS^l_{label}, GAS^l_{pred}, GVS^l_{label}, GVS^l_{pred}
    // Calculate EM metrics:
    Regard the set of sub-sequences as new "dialogues"
    Calculate EM of both actions and values in the sub-sequence length of l using Algorithm 2
    $(EM^l_{action}, EM^l_{value}, EM^l_{joint}) \leftarrow$ Algorithm 2
end for
// Calculate CE metrics:
counter $\leftarrow 0$
for each sub-sequence length l in L **do**
    $CE_{action} += EM^l_{action}$
    $CE_{value} += EM^l_{value}$
    $CE_{joint} += EM^l_{joint}$
    counter = counter + 1
end for
$CE_{action}, CE_{action}, CE_{action} /=$ counter

Cascading Evaluation (CE) is more lenient than EM as it assigns partial credit to correct sub-sequences. It awards an exact match for 3 predicted and 3 true steps, likewise for 2 steps, and finally for 1 step, then calculates the average scores. When we calculate the CE metric, we can first separate all the predicted actions into dialogues. Then, based on the pre-defined sub-sequence length, we can do the overlapped action sequence separations for each dialogue. Subsequently, we can calculate the EM metric on these sequence separations. Finally, the CE metrics can be obtained by averaging the EM metrics calculated on each kind of separation.

A.3 Experimental Results

A.3.1 Different scales of training data

Datasets	Methods	Action CE	Value CE	Joint CE	Action EM	Value EM	Joint EM
ABCD	DSP	0.414	0.493	0.374	0.261	0.331	0.221
	Multi-step	0.373	0.373	0.327	0.207	0.288	0.170
	Ours	**0.630**	**0.567**	**0.513**	**0.530**	**0.420**	**0.362**
MultiWoz	DSP	0.548	0.374	0.356	0.561	0.328	0.309
	Multi-step	0.536	0.353	0.334	0.540	0.298	0.277
	Ours	**0.599**	**0.391**	**0.375**	**0.630**	**0.340**	**0.325**

Table 9: Comparison with baselines with 10% of the training data.

Datasets	Methods	Action CE	Value CE	Joint CE	Action EM	Value EM	Joint EM
ABCD	DSP	0.490	0.570	0.450	0.340	0.429	0.296
	Multi-step	0.460	0.547	0.422	0.306	0.406	0.268
	Ours	**0.709**	**0.674**	**0.632**	**0.647**	**0.582**	**0.534**
MultiWoz	DSP	0.579	0.452	0.432	0.612	0.423	0.398
	Multi-step	0.568	0.439	0.420	0.591	0.38	0.402
	Ours	**0.633**	**0.486**	**0.464**	**0.681**	**0.455**	**0.427**

Table 10: Comparison with baselines with 30% of the training data.

Datasets	Methods	Action CE	Value CE	Joint CE	Action EM	Value EM	Joint EM
ABCD	DSP	0.498	0.580	0.459	0.355	0.453	0.313
	Multi-step	0.491	0.578	0.454	0.340	0.441	0.299
	Ours	**0.719**	**0.706**	**0.667**	**0.661**	**0.605**	**0.584**
MultiWoz	DSP	0.587	0.481	0.465	0.622	0.462	0.444
	Multi-step	0.578	0.471	0.457	0.604	0.451	0.435
	Ours	**0.649**	**0.526**	**0.509**	**0.709**	**0.514**	**0.493**

Table 11: Comparison with baselines with 50% of the training data.

A.3.2 Chained Prior as Plug-and-Play Module

A.3.3 Refined Chained Prior

The experimental results show that the refined Chained Prior can further improve action prediction performance on both CE and EM metrics compared with the global Chained Prior built without considering scenarios.

Methods	Strategy	Cascading Evaluation			Exact Match		
		Action	Value	Joint	Action	Value	Joint
DSP	w/o CP	0.414	0.493	0.374	0.261	0.221	0.331
	with CP	**0.509**	**0.536**	**0.421**	**0.347**	**0.383**	**0.263**
Multi-Step	w/o CP	0.373	0.373	0.327	0.207	0.288	0.170
	with CP	**0.413**	**0.465**	**0.366**	**0.246**	**0.291**	**0.199**

Table 12: Chained Prior works as a plugin to enhance other methods (10% of the data, ABCD).

Methods	Strategy	Cascading Evaluation			Exact Match		
		Action	Value	Joint	Action	Value	Joint
DSP	w/o CP	0.490	0.570	0.450	0.340	0.429	0.296
	with CP	**0.577**	**0.616**	**0.522**	**0.430**	**0.494**	**0.371**
Multi-Step	w/o CP	0.460	0.547	0.422	0.306	0.406	0.268
	with CP	**0.540**	**0.588**	**0.501**	**0.404**	**0.464**	**0.363**

Table 13: Chained Prior works as a plugin to enhance other methods (30% of the data, ABCD).

Methods	Strategy	Cascading Evaluation			Exact Match		
		Action	Value	Joint	Action	Value	Joint
DSP	w/o CP	0.498	0.580	0.459	0.355	0.453	0.313
	with CP	**0.593**	**0.642**	**0.542**	**0.447**	**0.527**	**0.391**
Multi-Step	w/o CP	0.491	0.578	0.454	0.340	0.441	0.299
	with CP	**0.584**	**0.633**	**0.546**	**0.454**	**0.518**	**0.412**

Table 14: Chained Prior works as a plugin to enhance other methods (50% of the data, ABCD).

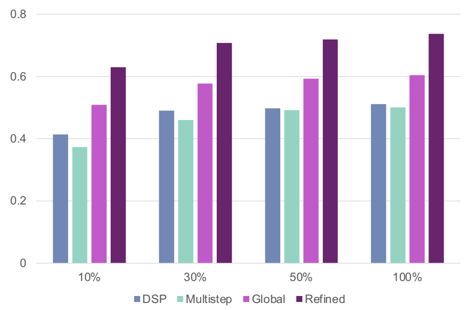


Figure 4: Refined Chained Prior improves the action (CE, Action) prediction accuracy.

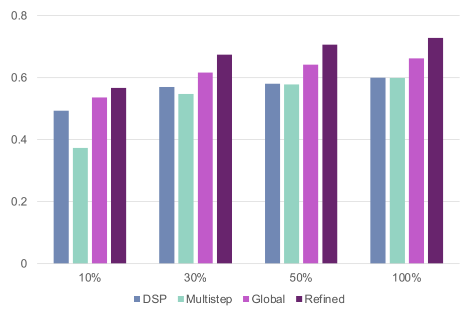


Figure 5: Refined Chained Prior improves the action (CE, Value) prediction accuracy.

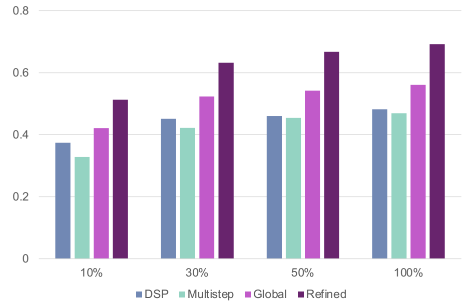


Figure 6: Refined Chained Prior improves the action (CE, Joint) prediction accuracy.

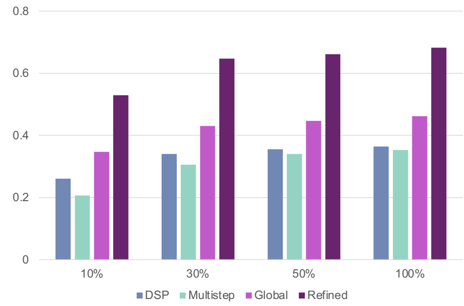


Figure 7: Refined Chained Prior improves the action (EM, Action) prediction accuracy.

Figure 8: Refined Chained Prior improves the action (EM, Value) prediction accuracy.

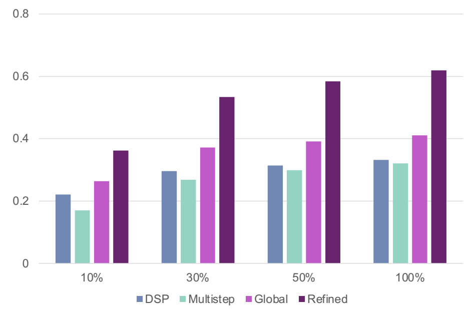


Figure 9: Refined Chained Prior improves the action (EM, Joint) prediction accuracy.

A.4 Prompts

Prompts for Action Prediction (ABCD)

The following are conversations between a user and an assistant. Indicated by the dialog acts, the assistant can help the user with checking in or providing information of temperature, time, price, location, and so on. You should predict the next action the assistant should take based on the context of the conversation. The action should be taken from the list of dialog acts provided below. Also, you need to fill in the slot value along with the action, if any, if no slot value is required, you should make the slot value be none. The format is action_name [none].

Available Dialog acts:
- pull-up-account: account has been pulled up for <name>.
- enter-details: details of <username> have been entered.
- verify-identity: identity verification in progress ...
- make-password: a password has been generated.
- search-timing: system action: search timing, I need to ask a certain question about timing.
- search-policy: system action: search policy, what kind of policy does the customer want to know?
- validate-purchase: purchase validation in progress ...
- search-faq: Answers can be found in the faq pages, searching the faq pages ...
- membership: membership level of <level> has been noted.
- search-boots: system action: search boots, click the boots toggle switch
- try-again: agent is looking for solutions ...
- ask-the-oracle: querying the system for an answer ...
- update-order: order has been updated with <change>.
- promo-code: a promo code has been created.
- update-account: account has been updated with <change>.
- search-membership: system action: search membership, I need to know the membership level of the customer.
- make-purchase: a purchase of <item> was made.
- offer-refund: a refund has been made for the amount of $<amount>.
- notify-team: the website team has been notified.
- record-reason: a reason of <reason> has been recorded.
- search-jeans: system action: search jeans, click the jeans toggle switch
- shipping-status: shipping status of <status> has been noted.
- search-shirt: system action: search shirt, click the shirt toggle switch
- instructions: agent is looking for solutions ..., I will give you some instructions.
- search-jacket: system action: search jacket, click the jecket toggle switch
- log-out-in: agent is looking for solutions ..., instruct the customer to log out of their account and log back in.
- select-faq: faq answer related to <faq> was selected.
- subscription-status: querying the system for subscription status ...
- send-link: a link will be sent.
- search-pricing: system action: search pricing, price of something.

Conversation:
Context: hello, how may i help you? i want to know the state of my refund. let me help you with that. i have an existing refund of $100 + i want to refund another $<amount>. did you want to add an extra item to your current refund? yes. could i have your full name or account id? albert sanders. account id 123445.
Assistant: pull-up-account [albert sanders]

Conversation:
Context: Context: hello, how may i help you? i want to know the state of my refund. let me help you with that. i have an existing refund of $100 + i want to refund another $<amount>. did you want to add an extra item to your current refund? yes. could i have your full name or account id? albert sanders. account id 123445. thanks. could i have your username, email address and order id to validate your order? <username>. <email>. and the order id? <order_id>. thank you. what is the item that you want to return? jeans. <name>.
Assistant: record-reason [guess jeans]

Conversation:
Context: hi. i want to manage my shipping details as my situation has changed. welcome to acmebrands! how may i help you today? i see. what is your name please? i want to change my shipping address. rodriguez domingo. and what is the shipping status please? order received. thanks.
Assistant: shipping-status [order received]

Conversation:
Context: i would like to know more about a product. hello. how may i help you today? sure. i would like to know if the buttons are brown or black. i see. so you are looking to purchase buttons? is there a drop down menu to select the color buttons you want to buy? no im looking to buy a shirt and asking if the button on the shirt is brown or black. product: shirt brand: michael_kors amount: $<amount>. oh the buttons on a shirt? should have mentioned that at the beginning. let me take a look for you. that shirt has dark brown buttons on them.
Assistant: select-faq [shirt_other_3]

Conversation:
Context: hi! how may i help you? hello. i recently signed up for a subscription but it looks like you guys charged me twice for it. i see, let's fix that. may i have your full name, account and order ids? sure, it's albert sanders and my account id is <account_id> do you have an order id? yes its <order_id>
Assistant: verify-identity [albert sanders, <account_id>, <account_id>]

Conversation:
Context: hello, thank you for contacting us today. how can i help you? how do you cancel a subscription? i'm sorry to hear that you might want to cancel your subscription. did something happen that made you want to do this? no, not at all. i was just thinking of ordering some things and i don't want to if the cancelation process is too hard. alright let me see what i can find for you.
Assistant: search-policy [none]

Conversation:
CURRENT DIALOG

Prompts for Action Prediction (MultiWoz)

The following are conversations between a user and an assistant. Indicated by the dialog acts, the assistant can help the user with checking in or providing information of temperature, time, price, location, and so on. You should predict the next action the assistant should take based on the context of the conversation. The action should be taken from the list of dialog acts provided below. Also, you need to fill in the slot value along with the action, if any, if no slot value is required, you should make the slot value be none. The format is action name [none].

Available Dialog acts:
- search for hotel: customers are looking for hotels with specific requirements
- book hotel: customers are going to booking hotels
- search for trains: customers are looking for trains with specific requirements
- book train ticket: customers are going to booking train tickets
- search for attractions: customers are looking for attractions with specific requirements
- search for restaurants: customers are looking for restaurants with specific requirements
- book table at restaurant: customers are going to booking tables at restaurants
- search for hospital: customers are looking for hospitals with specific requirements
- book taxi: customers are going to booking taxis
- search for taxi: customers are looking for taxis with specific requirements
- search for bus: customers are looking for buses with specific requirements
- search for police station: customers are looking for police stations

Conversation:
Context: i need a list of cheap place -s to stay that include free parking . alexander bed and breakfast is in the cheap price range in the centre of town . okay , does that place include free wifi and it is 4 stars ? yes , the alexander has free wifi and is a 4 star hotel . how many nights will you be staying ? i will be staying 5 nights starting from saturday .
Assistant: search for hotel [with parking, cheap, with internet, alexander bed and breakfast, 4 stars]

Conversation:
Context: can you let me know if a place called the gonville hotel is still around ? yes and it is as popular as ever . it is 3 stars and quite expensive . would you like me to book a room for you ? what area of town is it in ? it s in town centre . ok , thanks . also , are there any indian restaurant -s in the centre ?
Assistant: search for restaurants [centre, indian]

Conversation:
Context: i am looking for a train leaving on saturday from kings lynn . do you have any time preferences ? i need to leave after 13:00 . what will your destination be ? i am wanting to go to cambridge . the tr1499 leaves at 17:11 . can i book some tickets for you ? yes book for 4 people you are reference number is biazmbuc . is there anything else i can help you with ? i am looking for an expensive restaurant in the centre city . i have several restaurant -s in the centre in the expensive range . what type of cuisine would your prefer ? no preference . can you recommend 1 & book a table for 4 people at 17:45 on saturday . & may i have the reference # please ? i would recommend british cuisine . would that be okay with you ? that would be fine . is it available saturday at 17:45 ? the restaurant fitzbillies is available , and i have made you a reservation . your reference number is 4wgdgosa . thank you for your help , have a nice day .
Assistant: book table at restaurant [saturday, 4 people, 17:45]

Conversation:
Context: hi , i am looking to visit the colleges in town , and need to know the names of some of them . there are 18 matches . do you have an area in mind ? are there any in the centre ? there are 13 in the centre , including christ s college and emmanuel college . the area actually does not really matter , i just need the address of 1 of your recommendations , please ? sure ! the address for christ s college is saint andrew s street , postcode cb23bu . thank you ! i am also looking for a hotel with 4 stars . the university arms hotel is 4 starts and in the centre where the college is , would you like me to book you or provide more information ? do they have internet and could i have the phone number please ? yes they have internet and phone 01223351241 i need to book a taxi also . it should leave the hotel by 11:30
Assistant: search for taxi [university arms hotel, christ s college, 11:30]

Conversation:
Context: i am looking for a train that will depart from birmingham new street and go to cambridge . can you help me ? there are 133 trains from birmingham new street and to cambridge . is there a particular day and and or time you are interested in ? yes , i want to leave on monday and i need to arrive by 11:15 train tr5747 departs at 7:40 and would arrive in cambridge at 10:23 . would that work for you ? yes , that would be perfect , thanks . here is your reference number: qqvpn13m . total fee will be 75.09 gbp payable at the station . is there anything else i can help you with today ? can you help me find a place to stay ? i am looking a 0 star hotel with a cheap price . i found 2 zero star guest houses that are cheap . 1 is in the north and 1 in the centre . which area would you prefer ? are you able to find me a hotel in the east with free parking ? there are 3 cheap guest houses in east with free parking . they are all 4 stars . would you like me to book 1 of those ? no , i would like that zero star cheap guest house in the north . sure , i can help you with that . when would you like to stay ? monday . please give me their phone number , postcode and address
Assistant: book train ticket [none]

Conversation:
Context: hi , i am looking for the nearest police station . hi , the nearest police station is at parkside , cambridge postcode: cb11jg . is there anything else i can help with ? can you please give me the phone number as well ?
Assistant: search for police station [none]

Conversation:
CURRENT DIALOG

Prompts for Response Generation (ABCD)

The following are conversations between a user and an assistant. Indicated by the dialog acts, the assistant can help the user with checking in or providing information of temperature, time, price, location, and so on. The response should be coherent, engaging, diverse, informative, and overall good and should be in line with the next action. The response should be concise and to the point and not exceed 30 words. If there is a slot value, such as <item>, <username>, it should be filled in with the correct value.

Available Dialog acts:
- pull-up-account: account has been pulled up for <name>.
- enter-details: details of <username> have been entered.
- verify-identity: identity verification in progress ...
- make-password: a password has been generated.
- search-timing: system action: search timing, I need to ask a certain question about timing.
- search-policy: system action: search policy, what kind of policy does the customer want to know?
- validate-purchase: purchase validation in progress ...
- search-faq: Answers can be found in the faq pages, searching the faq pages ...
- membership: membership level of <level> has been noted.
- search-boots: system action: search boots, click the boots toggle switch
- try-again: agent is looking for solutions ...
- ask-the-oracle: querying the system for an answer ...
- update-order: order has been updated with <change>.
- promo-code: a promo code has been created.
- update-account: account has been updated with <change>.
- search-membership: system action: search membership, I need to know the membership level of the customer.
- make-purchase: a purchase of <item> was made.
- offer-refund: a refund has been made for the amount of $<amount>.
- notify-team: the website team has been notified.
- record-reason: a reason of <reason> has been recorded.
- search-jeans: system action: search jeans, click the jeans toggle switch
- shipping-status: shipping status of <status> has been noted.
- search-shirt: system action: search shirt, click the shirt toggle switch
- instructions: agent is looking for solutions ..., I will give you some instructions.
- search-jacket: system action: search jacket, click the jecket toggle switch
- log-out-in: agent is looking for solutions ..., instruct the customer to log out of their account and log back in.
- select-faq: faq answer related to <faq> was selected.
- subscription-status: querying the system for subscription status ...
- send-link: a link will be sent.
- search-pricing: system action: search pricing, price of something.

Conversation:
Context: hello, how may i help you? i want to know the state of my refund. let me help you with that. i have an existing refund of $100 + i want to refund another $<amount>. did you want to add an extra item to your current refund? yes. could i have your full name or account id? albert sanders. account id 123445.
Assistant(pull-up-account [albert sanders]): account has been pulled up for albert sanders.

Conversation:
Context: Context: hello, how may i help you? i want to know the state of my refund. let me help you with that. i have an existing refund of $100 + i want to refund another $<amount>. did you want to add an extra item to your current refund? yes. could i have your full name or account id? albert sanders. account id 123445. thanks. could i have your username, email address and order id to validate your order? <username>. <email>. and the order id? <order_id>. thank you. what is the item that you want to return? jeans. <name>.
Assistant(record-reason [guess jeans]): a reason of guess jeans has been recorded.

Conversation:
Context: hi. i want to manage my shipping details as my situation has changed. welcome to acmebrands! how may i help you today? i see. what is your name please? i want to change my shipping address. rodriguez domingo. and what is the shipping status please? order received. thanks.
Assistant(shipping-status [order received]): shipping status of order received has been noted.

Conversation:
Context: i would like to know more about a product. hello. how may i help you today? sure. i would like to know if the buttons are brown or black. i see. so you are looking to purchase buttons? is there a drop down menu to select the color buttons you want to buy? no im looking to buy a shirt and asking if the button on the shirt is brown or black. product: shirt brand: michael_kors amount: $<amount>. oh the buttons on a shirt? should have mentioned that at the beginning. let me take a look for you. that shirt has dark brown buttons on them.
Assistant(select-faq [shirt_other_3]): faq answer related to shirt_other_3 was selected.

Conversation:
CURRENT DIALOG

Prompts for Response Generation (MultiWoz)

The following are conversations between a user and an assistant. Indicated by the dialog acts, the assistant can help the user with checking in or providing information of temperature, time, price, location, and so on. The response should be coherent, engaging, diverse, informative, and overall good and should be in line with the next action. The response should be concise and to the point and not exceed 30 words. If there is a slot, such as <item>, <username>, <location>, it should be filled in with the correct value.

Available Dialog acts:
- search for hotel: customers are looking for <price> hotels with <requirements>, <level>, in <location>, <date> <time>, the hotel should have <requirements>.
- book hotel: customers are going to booking hotels for <number> people, <number> nights starting from <date>.
- search for trains: customers are looking for trains from <location> to <location> on <date> <time>. - book train ticket: customers are going to booking train tickets for <number> people.
- search for attractions: customers are looking for <type> attractions in <location> with <requirements>.
- search for restaurants: customers are looking for <type> restaurants in <location>, <price> range, with <requirements>.
- book table at restaurant: customers are going to booking tables at restaurants for <number> people, on <date> at <time>.
- search for hospital: customers are looking for <type> hospitals in <location>.
- book taxi: customers are going to booking taxis
- search for taxi: customers are looking for a taxi at <time> from <location> to <location>.
- search for bus: customers are looking for a bus from <location> to <location> on <date> <time>.
- search for police station: customers are looking for police stations

Conversation:
Context: i need a list of cheap place -s to stay that include free parking . alexander bed and breakfast is in the cheap price range in the centre of town . okay , does that place include free wifi and it is 4 stars ? yes , the alexander has free wifi and is a 4 star hotel . how many nights will you be staying ? i will be staying 5 nights starting from saturday .
Assistant(search for hotel [with parking, cheap, with internet, alexander bed and breakfast, 4 stars]): customers are looking for cheap hotels with free parking and wifi, 4 stars, in the centre of town, for 5 nights starting from saturday

Conversation:
Context: i need a list of cheap place -s to stay that include free parking . alexander bed and breakfast is in the cheap price range in the centre of town . okay , does that place include free wifi and it is 4 stars ? yes , the alexander has free wifi and is a 4 star hotel . how many nights will you be staying ? i will be staying 5 nights starting from saturday . thank you so much for that information . how many people in your party so i can make the booking for you ? yes , can you book it for 6 people ?
Assistant(book hotel [saturday, 6 people, 5 stay]): customers are going to booking hotels for 6 people, 5 nights starting from saturday.

Conversation:
Context: i need a list of cheap place -s to stay that include free parking . alexander bed and breakfast is in the cheap price range in the centre of town . okay , does that place include free wifi and it is 4 stars ? yes , the alexander has free wifi and is a 4 star hotel . how many nights will you be staying ? i will be staying 5 nights starting from saturday . thank you so much for that information . how many people in your party so i can make the booking for you ? yes , can you book it for 6 people ? booking was successful . your reference number is qit1w09h . do you need anything else ? thank you . i also need a train to kings lynn that leaves on sunday . there are 19 trains leaving cambridge on sunday . when would you like to depart ? i want the train to leave after 16:30 .
Assistant(search for trains [sunday, kings lynn, 16:30]): customers are looking for trains to kings lynn on sunday, leaving after 16:30

Conversation:
Context: what trains arrive in cambridge by 10:30 ? where are you traveling from and on what day ? i am traveling from kings lynn on sunday . please give me your day and time of departure to help me to narrow down to a suitable result . i do not have a departure time . i just need to be in cambridge by 10:30 . i suggest the tr8092 that will arrive in cambridge at 08:58 . this will give you ample time . do you want me to book this ? please book for 5 people , i will also need the reference number .
Assistant(book train ticket [5 people]): customers are going to booking train tickets for 5 people.

Conversation:
Context: hello , i am looking for something to do in the west part of town . it sould involve multiple sports . unfortunately none of those place -s exist here . any other preferences ? hm , can you tell me about what entertainment venue -s might be on the west side of town instead ? there s a fun place called whale of a time at unit 8 , viking way , bar hill . thank you , can i please get an address and postal code . is there an entrance fee that will be charged .
Assistant(search for attractions [west, entertainment, whale of a time]): customers are looking for entertainment venues on the west side of town, whale of a time

Conversation:
Context: i am looking for a chinese restaurant please . or 1 that serves chinese food . what area of town would you like the restaurant to be in ? i would like to be in the centre of town . jinling noodle bar is in the centre area serving chinese in the moderate price range . would you like to reserve a table ? i want something expensive .
Assistant(search for restaurants [chinese, centre, expensive]): customers are looking for chinese restaurants in the centre of town, expensive

Conversation:
Context: can you let me know if a place called the gonville hotel is still around ? yes and it is as popular as ever . it is 3 stars and quite expensive . would you like me to book a room for you ? what area of town is it in ? it s in town centre . ok , thanks . also , are there any indian restaurant -s in the centre ? there are several with various price range -s . whatever you recommend . give me your best recommendation and go ahead and book me for a table for 6 people . i want to go on saturday at 15:30 . saffron brasserie has your reservation . it will be held for 15 minutes . the reference number is , pryp175n . can i get you the address or phone number ? no , that will be all . thank you !
Assistant(book table at restaurant [saturday, 6 people, 15:30]): customers are going to booking tables at restaurants for 6 people, on saturday at 15:30

Conversation:
Context: i need to find a hospital here in the area . the nearest hospital is located at hills rd , cambridge . can i get thhe phone number and postcode which department are you looking for so i can give you the correct phone number ? i need the paediatric clinic please . the phone number for the paediatric clinic is 01223348313 , is there anything else i can help with ? yes , may i please have the post code ?
Assistant(search for hospital [paediatric clinic]): customers are looking for the paediatric clinic.

Conversation:
Context: i am looking for a hotel in cambridge called the cambridge belfry that hotel is in the west and listed as cheap , but still has 4 stars . they provide free parking and internet . would you like to make a reservation ? yes , could you please book me a room on tuesday for 5 people and for 4 nights . booking was successful . reference number: mcitlhi8 . great ! i also need information on multiple sports in the centre . there are no multiple sports attractions in the centre . should we try another area ? how about a museum ? how about broughton house gallery ? it s free of charge . that sounds great . can you please give me the phone number ? sure . their phone number is 01223314960 . i would also like a taxi to commute and i would like it 17:45

please and the contact number and the car type , thank you okay , where do you want the taxi to pick you up and where will you be going ? i need the taxi to pick me up at the hotel by 17:45 .
Assistant(search for taxi [17:45, cambridge belfry, broughton house gallery]): customers are looking for a taxi at 17:45 from cambridge belfry to broughton house gallery.

Conversation:
Context: can you help me find a train that leaves cambridge after 9:45 pm ? thanks . i can help with that . what is the destination and what day would you like to travel ? i am going to bishops storford on wednesday . i actually need to leave after 21:45 though . i do not have any trains that match your request . that s disappointing . can you recommend a taxi or bus service ?
Assistant(search for bus [wednesday, cambridge, bishops stortford, 21:45]): customers are looking for a bus from cambridge to bishops stortford on wednesday, leaving after 21:45.

Conversation:
Context: i am looking for the parkside police station parkside police station is located in parkside , cambridge , within the postcode of cb11jg . may i help with something else ? yes , can you please provide their phone number and physical address ?
Assistant(search for police station [none]): customers are looking for police stations.

Conversation:
CURRENT DIALOG