DivTOD: Unleashing the Power of LLMs for Diversifying Task-Oriented Dialogue Representations

Anonymous ACL submission

Abstract

Pre-trained language models trained on general text have seen great success in various scenarios. However, the inherent linguistic differences between general text and task-oriented dialogues (TOD) make existing language models less useful in practice. Current task-oriented dialogue pre-training methods overlook the one-to-many property of conversations, where multiple responses can be appropriate given the same conversation context. In this paper, we propose a novel dialogue pre-training model called DivTOD, which collaborates with LLMs to learn diverse task-oriented dialogue representations. DivTOD guides LLMs in transferring diverse knowledge to smaller models while removing domain knowledge that contradicts task-oriented dialogues. Experiments show that our model outperforms strong TOD baselines on various downstream dialogue tasks and learns the intrinsic diversity of task-oriented dialogues. [1]

1 Introduction

Pre-trained language models (PLMs) (Devlin et al., 2019; Liu et al., 2019) based on massive general text corpora (Zhu et al., 2015) are commonly used in many NLP applications. These models are pre-trained in a self-supervised manner and then fine-tuned for supervised downstream tasks. The Pretrain and Finetune paradigm has significantly improved the performance of various downstream tasks. Despite their success, most current research efforts focus on general documents such as Wikipedia, which have a large linguistic gap with dialogues, particularly task-oriented dialogues. Directly using these PLMs is suboptimal and yields poor performance (Rashkin et al., 2019).

Compared to plain text, TOD aims to help users accomplish specific tasks with explicit goals (e.g. restaurant reservation), belief states, and database information. Thus, learning high-quality dialogue representations is crucial for understanding tasks in TOD. Previous methods pre-trained models using task-oriented dialogue datasets to improve dialogue understanding performance. SimCSE (Gao et al., 2021) uses a contrastive learning framework to learn sentence embeddings by generating positive pairs through Dropout (Srivastava et al., 2014) augmentation. TOD-BERT (Wu et al., 2020) considers the intrinsic properties of dialogue data by using dialogue history and corresponding responses as positive pairs for contrastive learning. DSE (Zhou et al., 2022) learns from dialogues by taking consecutive utterances of the same dialogue as positive pairs. Furthermore, FutureTOD (Zeng et al., 2023) proposes a new non-contrastive self-training framework to address the challenges faced by previous contrastive methods in selecting true positive and negative pairs.

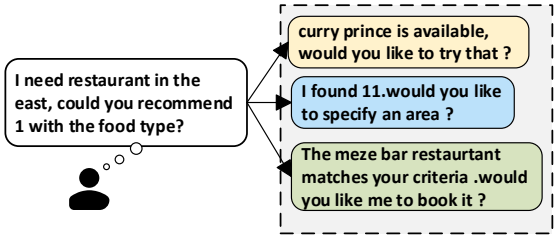


Figure 1: The same context may have multiple appropriate responses in a task-oriented dialogue, which we call one-to-many.

Despite previous TOD PLMs have made remarkable progress. Most work ignores the one-to-many property in the conversation where multiple responses can be appropriate under the same conversation context (shown in Figure 1). Our analysis shows that the lack of diversity in TOD datasets is the main reason for this. Specifically, (1) most TOD datasets only provide a single response for the same dialogue history, and (2) the style of system responses in TOD is often monotonous and dull. As a result, current TOD PLMs capture only the most common dialogue information and ignore

[1] We will release our source code, data, and pre-trained models after blind review to facilitate future research.

less frequent but still feasible user behaviors, which leads to duplicated and plain responses.

Large Language Models (LLMs) (Brown et al., 2020; Ouyang et al., 2022; Touvron et al., 2023) offer hope for addressing the problems mentioned above. LLMs have more parameters and are pre-trained and fine-tuned on a richer and wider corpus (Köpf et al., 2023; Chiang et al.; Ding et al., 2023). Consequently, LLMs possess a broader general background knowledge, which enables them to generate more diverse and feasible responses. However, it should be noted that LLMs have not been specifically fine-tuned for task-oriented dialogue systems (Hudeček and Dušek, 2023), resulting in a significant mismatch between their general knowledge and the domain knowledge required for task-oriented dialogue. Furthermore, LLMs typically have billions of parameters, making them too expensive to deploy at scale because of the overwhelming computational requirements, as well as the cost of fine-tuning and inference (Wei et al., 2022). To address these issues, a natural approach is to distill the rich background and domain-specific knowledge required for tasks from LLMs into smaller and more efficient models.

In this paper, we propose a new dialogue pre-training model, DivTOD, which enhances the ability of smaller models to model the intrinsic one-to-many diversity of human conversations by transferring rich general background knowledge and task-specific domain knowledge from LLMs. Our framework consists of three core steps: (1) Guiding LLMs to generate diverse system responses based on dialogue context in a "filling the blank" manner. (2) Using an LLM-based post-generation filter to align the generated responses with domain knowledge. (3) Allowing small models to imitate LLM's capabilities by observing diverse dialogues through self-training. We evaluated DivTOD on various task-oriented dialogue tasks, such as intent classification, dialogue state tracking, dialogue act prediction, and response selection. The results demonstrate that DivTOD consistently outperforms strong TOD baselines in all scenarios, indicating its generalization capability. Furthermore, we observed that DivTOD is able to capture a wider range of dialogue information and learn the intrinsic one-to-many diversity of TOD.

Our contributions are: (1) We propose a framework that distills task-specific domain knowledge and rich general background knowledge of LLMs into smaller models. We use this framework to pre-train DivTOD and model the intrinsic one-to-many diversity of human conversations. (2) Our DivTOD outperforms strong TOD baselines on diverse downstream dialogue tasks. It also learns the intrinsic diversity of task-oriented dialogues

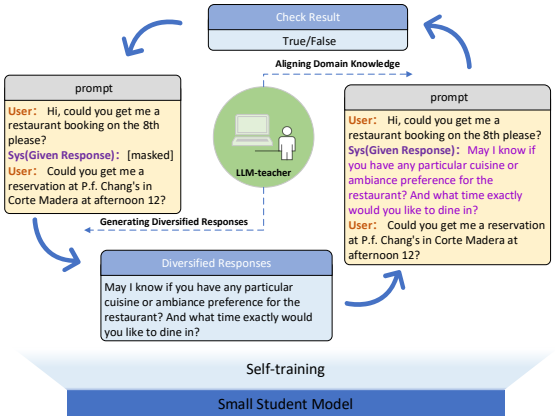


Figure 2: Overall architecture of DivTOD.

2 Model

2.1 Overall Architecture

Figure 2 shows the overall architecture of DivTOD. Our framework comprises a teacher model M_T based on LLM and a student model M_S based on a smaller model, initialized by Vicuna-7b[2] and BERT-base-uncased[3], respectively. First, we guide the M_T to generate diverse system responses based on the dialogue context, using a "filling the blank" approach. Then, we use the M_T as a filter to align the generated response with the domain knowledge of the task-oriented dialogue context. Finally, by continuously iterating the generate-filter steps, we enable the M_S to train on both the original dataset and the generated dataset using the self-training method proposed in Zeng et al. (2023).

2.2 Diversifying Task-Oriented Dialogue Representations

Notation We use the collected datasets by TOD-BERT (Wu et al., 2020) as our pre-training corpus. The corpus is the combination of 9 publicly available task-oriented datasets, including 100,707 dialogues and 1,388,152 utterances over 60 domains. For each dialogue, we first transform it into a token sequence $D = \{U_1, S_1, \ldots, U_n, S_n\}$. U_i and S_i denote the user utterance and system utterance with a prefix of two special role tokens [USR] or [SYS], respectively. n is the turn number of the dialogue.

[2]https://huggingface.co/lmsys/vicuna-7b-delta-v1.1
[3]https://huggingface.co/bert-base-uncased

Generating Diversified Responses We use a "filling in the blank" approach to guide M_T in generating diverse responses based on the dialogue context. For a given dialogue D, we randomly mask a system response S_i and use the remaining part as the input $D^{'}$. We design a few-shot prompt P consisting of a triplet (I^P, D^P, S^P) to instruct M_T in generating diverse responses $S_i^{'}$ based on $D^{'}$ [4]. The core component of the P is I^P, which describes the task to the model. I^P also constrains the behavior of the model, preventing it from generating irrelevant responses. D^P is the input example, and S^P is the corresponding generated response. For each input $D^{'}$, we append it to P and use it as input to prompt M_T to generate diverse responses. M_T can mimic the demonstrations D^P, S^P in P to generate new diverse responses. The complete prompt example is shown in Figure 5 in Appendix.

Aligning Domain Knowledge Although we can obtain more diverse responses using M_T, these responses may contradict the characteristics of task-oriented dialogue systems. For example, the generated responses may provide excessive information that the user does not need or answer questions that the user asks in the future. To ensure that the generated responses align with domain knowledge in TOD, we designed a filter based on M_T.We replace the masked parts in $D^{'}$ with the generated response $S_i^{'}$ to form a new input $D^{''}$. We have designed a few-shot prompt E consisting of a triplet (I^E, D^E, R^E) to prompt M_T to judge the contextual consistency of $D^{''}$ and whether it conflicts with the characteristics of TOD. The core part of the prompt is I^E, which describes the task to M_T. The prompt also provides logical knowledge related to task-oriented dialogue. D^E and R^E are the demonstrations provided to M_T. D^E represents the example input, and R^E represents the corresponding judgment result (either True or False). We append $D^{''}$ to E and determine whether to keep $S_i^{'}$ based on the filtering result. Figure 6 in the Appendix shows a complete example of this prompt.

Self Training We iterate through the generate-filter steps (summarized in Algorithm 1 in Appendix) described above and combine the newly generated dialogues with the original ones.

We train M_S using the self-training objective proposed by FutureTOD (Zeng et al., 2023) on the assembled dialogues. We initialize the new student model and teacher model using M_S. For each dialogue, we randomly split it into context and future sequences. The student model encodes the context and obtains the original dialogue representation, while the teacher model encodes both the context and future to obtain the target. The architectures of the student and teacher models are the same, but the weights of the teacher model are periodically updated by the student. The training goal is to align the original content representation with the full representation containing future knowledge. The generate-filter steps produce diverse responses, resulting in multiple reasonable full representations that can align with the same content representation.

Through the above framework of generation, filtering, and self-training, we transfer both general background knowledge and task-specific domain knowledge from M_T to M_S.

3 Experiment

3.1 Pre-training Corpus

We use the nine different task-oriented datasets collected by Wu et al. (2020) and show the full details in Appendix A.

3.2 Baselines

DivTOD is evaluated on various downstream tasks and compared to several well-established baselines, including both encoder-only and generative architectures. For details about the baselines, please refer to the Appendix B.

3.3 Implementation Details

LLM generating Details We use Vicuna as LLM to generate diverse system responses and to align domain knowledge. For experimental details and hyperparameter settings for this stage, please refer to the Appendix C.1.

Pre-training Details After diverse system response generation, all the dialogue will be merged into the original dataset as the new dataset to pre-train. The details of the hyperparameters for the pre-training can be found in the Appendix C.2.

Finetuning Details After completing pre-training on dialogue, we perform supervised fine-tuning on downstream dialogue tasks. However, it is important to note that we only use generated diverse dialogue during the pre-training phase. In the fine-tuning phase, we use datasets and settings that are identical to the previous baseline, including golden labels such as dialogue acts. The details

[4] We try different methods to instruct M_T, including zero-shot prompts. However, these methods are not very effective. For example, the pass rate of the zero-shot method is low in our post-filter.

Model	ACC(acc)	Acc(in)	Acc(out)	Recall(out)
BERT	84.9%	95.8%	88.1%	35.6%
DialoGPT	83.9%	95.5%	87.6%	32.1%
BERT-mlm	85.9%	96.1%	89.5%	46.3%
SimCSE	82.3%	94.7%	86.6%	26.6%
TOD-BERT	86.6%	**96.2%**	89.9%	43.6%
DSE	84.3%	95.8%	87.7%	32.5%
FutureTOD	87.2%	96.0%	90.0%	47.6%
DivTOD	**87.4%***	95.8%	**90.5%***	**49.5%***

Table 1: Intent recognition results on the OOS dataset. Acc(all), Acc(in), Acc(out) denotes the overall accuracy, in-domain intent accuracy, and out-of-domain intent accuracy. The numbers with * are significant using t-test with $p < 0.01$.

Model	Joint Acc	Slot Acc
BERT	45.6%	96.6%
BERT-mlm	47.7%	96.8%
SimCSE	48.0%	96.8%
TOD-BERT	48.0%	96.9%
DSE	49.9%	97.0%
FutureTOD	50.4%	97.1%
DivTOD	**50.9%***	**97.2%***

Table 2: Dialogue state tracking results on MWOZ 2.1. We report joint goal accuracy (Joint Acc) and slot accuracy (Slot Acc). The numbers with * are significant using t-test with $p < 0.01$.

of the hyperparameters for the pre-training can be found in the Appendix C.3.

3.4 Main Results

We evaluate all the pre-trained LMs on four core task-oriented dialogue tasks: intent recognition, dialogue state tracking, dialogue act prediction, and response selection. It is important to emphasize that our focus is on learning diverse dialogue representations. Therefore, we are more concerned with tasks related to dialogue understanding rather than tasks related to response generation. To ensure fairness in our evaluation, we adopt the same architecture for all baselines following TOD-BERT and only add simple components to the pre-trained model, such as a single-layer classification head. For each downstream task, we conduct experiments using the entire dataset. In addition, we also explored few-shot setting experiments in section 4.4. This allowed us to see how well these pre-trained language models generalize to multiple tasks and scenarios.

Intent Recognition is a multi-class classification task that takes a dialogue utterance as input and predicts an intent label. We use the [CLS] embeddings from the model as the dialogue representation. The model is trained with cross-entropy loss. We report classification accuracy and recall.

Table 1 shows the results of intent recognition on the OOS dataset (Larson et al., 2019), which includes 151 intent classes across ten domains, including 150 in-domain intents and out-of-domain (OOD) intents. We find DivTOD outperforms all the baselines on 3 of 4 metrics, especially with significant improvements in overall accuracy and OOD metrics. All the results show the generalization ability of DivTOD both on in-domain and out-of-domain metrics.

Dialogue State Tracking is a multi-class classification task, which involves identifying the slot values for each (domain, slot) pair at each dialogue turn, based on a pre-defined ontology. The model takes dialogue history as input and is trained with cross-entropy loss summed over all the pairs. We use a widely-used TOD dataset MWOZ 2.1 (Budzianowski et al., 2018) across seven different domains. We report the Joint acc and Slot acc. The Joint acc considers true if and only if the predicted values exactly match its ground truth values at each dialogue turn. The slot acc individually compares each (domain, slot, value) triplet to its ground truth label.

Table 2 shows the results of dialogue state tracking on MWOZ 2.1. Our DivTOD achieves state-of-the-art results on all the metrics. We find SimCSE performs poorly because it ignores the intrinsic properties of dialogue data and can not model overall dialogue. Our method achieves a greater improvement on joint accuracy than on slot accuracy, indicating the strength of understanding the overall dialogue context. For example, DivTOD outperforms TOD-BERT by 0.3% on Slot Acc but 2.9% on Joint Acc in the full data setting, which indicates the superiority of dialogue modeling.

Dialogue Act Prediction is a multi-label classification task that takes dialogue history as input and predicts multiple dialogue acts corresponding to system response. The model is trained with binary cross-entropy loss over all possible actions. During inference, the threshold for triggering the dialogue act is set to 0.5. We use two datasets MWOZ (Budzianowski et al., 2018) and DSTC2 (Henderson et al., 2014). Following (Wu et al., 2020), we use the same data preprocessing to uniform the original dialogue acts to a general format. We report the micro-F1 and macro-F1.

Table 3 displays the results of dialogue act prediction on MWOZ and DSTC2 datasets. Our DivTOD method outperforms all other baselines in three out of four metrics. Specifically, our method surpasses FutureTOD on the DSTC2

Model	MWOZ		DSTC2	
	micro-F1	macro-F1	micro-F1	macro-F1
BERT	91.4%	79.7%	92.3%	40.1%
DialoGPT	91.2%	79.7%	93.8%	42.1%
BERT-mlm	91.7%	79.9%	90.9%	39.9%
SimCSE	91.6%	80.3%	91.5%	39.6%
TOD-BERT	91.7%	80.6%	93.8%	41.3%
DSE	91.7%	81.3%	92.6%	40.2%
FutureTOD	**92.0%**	81.9%	94.6%	44.6%
DivTOD	91.7%	**82.6%***	**95.8%***	**46.5%***

Table 3: Dialogue act prediction results on MWOZ and DSTC2. The numbers with * are significant using t-test with $p < 0.01$.

Model	MWOZ		DSTC2	
	1-to-100	3-to-100	1-to-100	3-to-100
BERT	47.5%	75.5%	46.6%	62.1%
DialoGPT	35.7%	64.1%	39.8%	57.1%
BERT-mlm	48.1%	74.3%	50.0%	65.1%
SimCSE	64.2%	85.4%	55.6%	70.5%
TOD-BERT	65.8%	87.0%	56.8%	70.6%
DSE	63.3%	85.3%	58.3%	72.0%
FuturueTOD	68.5%	87.9%	58.4%	72.6%
DivTOD	**71.3%***	**90.4%***	**59.5%***	**74.0%***

Table 4: Response selection evaluation results on MWOZ and DSTC. We report 1-to-100 and 3-to-100 accuracy, which represents the ratio of the ground-truth response being ranked at the top-1 or top-3 given 100 candidates. The numbers with * are significant using t-test with $p < 0.01$.

dataset, demonstrating significant improvement. It also exhibits improvement on MWOZ, with the macro-F1 increasing from 81.9% to 82.6%. However, we notice that different methods exhibit unclear distinctions in terms of micro-F1. We attribute this to the imbalanced distribution of dialogue action labels in MWOZ. In such cases, macro-F1 provides a more reasonable evaluation metric as it assigns equal weight to each label, regardless of the number of samples. In addition to the higher response quality, we also observe that DivTOD captures a wider range of dialogue policies and learns the intrinsic one-to-many diversity of TOD, as discussed in Section 4.6.

Response Selection is a ranking task that aims to retrieve the most relative system response from a candidate pool based on dialogue history. We also use MWOZ and DSTC2 as our evaluation datasets. We use a dual-encoder strategy, which calculates similarity scores between dialogue history and candidate responses. We train this model with random system responses from the corpus as negative samples. We report k-to-100 accuracy. This metric represents the ratio of the ground-truth response being ranked in the top-k positions when compared to 99 randomly sampled responses, as determined by the scores computed by the dual-encoder.

Table 4 displays the results of response selection on MWOZ and DSTC2. Our DivTOD method achieves state-of-the-art results on all metrics. Despite TOD-BERT being pre-trained with a response contrastive objective, our method still significantly outperforms it on both MWOZ and DSTC2 in full data settings. This indicates that Our Method has better generalization capabilities. Compared to FutureTOD, our method brings significant improvements in response selection, indicating that it can enhance the diversity of TOD representation and thus improve performance.

In summary, our method shows notable improvements in dialogue act prediction and response selection tasks. This indicates that considering the one-to-many nature of dialogues is essential for these tasks. Furthermore, our method also achieves enhancement in other important task-oriented dialogue tasks, such as intent classification and dialog state tracking. This further highlights the generalization of our method across various tasks.

4 Qualitative Analysis

4.1 Ablation Study of Domain Knowledge Alignment

Table 6 presents the ablation study results of the domain knowledge alignment, on the two downstream tasks, dialogue act prediction on DSTC2 and response selection on MWOZ [5]. DivTOD performs the best under various conditions when training using dialogue with aligned domain knowledge. However, the performance of DivTOD w/o Align is unsatisfactory. For example, in the dialogue act prediction task, DivTOD w/o Align is similar to the baseline and lower than DivTOD's performance. This suggests that aligning with domain knowledge may help maintain consistency in TOD dialogues, thereby contributing to the likelihood that the diverse dialogues generated by LLM have a beneficial influence on the pre-training process.

To visually represent the quality of the generated dialogues by different methods, we randomly selected dialogue samples, as shown in the Figure 3. From the dialogue examples, it can be seen that the DivTOD's dialogues are different from the original text and they are all consistent with the dialogue context. However, DivTOD w/o Alignment's Dialogue produces two problems. First, LLM may not answer according to the prompt instructions,

[5] Considering the cost of qualitative analysis, we select two classic task-oriented tasks. Furthermore, we use different datasets for each task to ensure generalizability.

Source	Unique 1-gram	Unique 2-gram	Unique 3-gram
Raw Dialogue	380	1512	2391
PPTOD	221	736	1189
Vicuna-7b	357	1793	3497

Table 5: The number of unique n-grams contained in the generated responses. "Raw Dialogue," "PPTOD," and "Vicuna-7b" refer to responses from the original dialogues, PPTOD, and the LLM we utilized.

Method	MWOZ		DSTC2	
	1-to-100	3-to-100	micro-F1	macro-F1
FutureTOD	68.5%	87.9%	94.6%	44.6%
DivTOD w/o Align	70.0%	90.1%	94.8%	44.2%
DivTOD	**71.3%**	**90.4%**	**95.8%**	**46.5%**

Table 6: Ablation Study of Domain Knowledge Alignment. DivTOD w/o Align denotes the DivTOD without domain knowledge alignment.

but may produce irrelevant answers such as "here's the rewritten response:". Second, LLM may produce answers that do not match the context, such as answering questions that users will only raise or provide information in the future.

4.2 Advantages of LLMs in Generating Diversified Responses

To demonstrate the advantage of LLMs over other models trained solely on TOD data in generating diversified responses, we randomly sample 500 TOD dialogue samples and generate responses using both PPTOD (Su et al., 2021) and LLM [6]. We compare the number of unique n-grams contained in the generated responses. Tabel 5 demonstrates that the responses generated by LLM contain more unique n-grams than those generated by PPTOD, even surpassing the number of unique n-grams present in the original dialogue. We analyze that PPTOD, being pre-trained on the TOD dataset, overfits the limitations of that dataset, resulting in a decrease in the diversity of responses it generates. This further supports the evidence that LLM is capable of generating more diverse responses.

4.3 Quantity of Diverse Dialogues

In our default experimental setting, we instructed LLMs to generate about 50k diverse dialogues for dialogue pre-training. Figure 4 shows the effect of varying the number of diverse dialogues during pre-training, on the two downstream tasks of dialogue act prediction on DSTC2 and response selection on MWOZ. We find the performance of DivTOD on both two tasks gradually improves as the number of diverse dialogues increases. This indicates that diverse dialogues generated by large language models can continuously improve the model's generalization.

4.4 Few Shot Learning

Table 7 displays the results of dialogue act prediction on DSTC2 and response selection on MWOZ in few shot setting. [7] Our DivTOD achieves state-of-the-art results on all the metrics. Although our DivTOD method and FutureTOD were both pre-trained using non-contrastive self-training frameworks, our DivTOD method significantly outperforms FutureTOD on both datasets in both 1% and 10% data settings. This demonstrates that DivTOD has a superior generalization capability across different scenarios.

We used the same pre-training corpus as the previous baseline, which includes MWOZ and CamRest676 (also part of MWOZ). However, concerns may arise about the reliability of few-shot performance due to this. To address this, we excluded MWOZ and CamRest676 from the pre-training corpus and evaluated the performance of FutureTOD and DivTOD on a 1% data setting. On the MWOZ dataset, the performance of FutureTOD† and DivTOD† decreased. This suggests that including MWOZ in the pre-training phase does enhance few-shot performance. However, our DivTOD† still achieves good performance and surpasses FutureTOD†. Surprisingly, on the DSTC2 dataset, DivTOD† and FutureTOD† also exhibit a slight decrease in performance. This highlights the high quality of MWOZ as a TOD dataset and provides some justification for its inclusion in the pre-training corpus in the previous baseline.

4.5 Zero Shot Learning

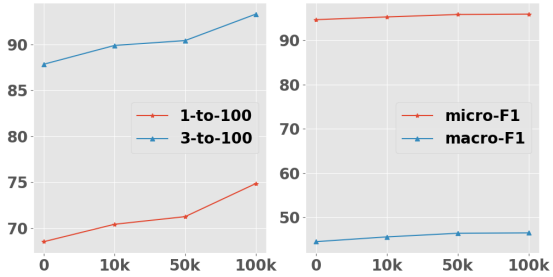
To validate the unsupervised embedding capability of our model, we performed zero-shot response selection on the MWOZ and DSTC2 datasets. The results are displayed in Table 8. BERT, FutureTOD, and DivTOD use an encoder architecture, while LLaMA and Vicuna use a decoder architecture. Therefore, for encoder models, we use the hidden state of the [CLS] layer as the embedding for inference retrieval, while for decoder models, we use the hidden state corresponding to the last input character as the embedding (consistent with the settings of DialoGPT). Our DivTOD outperforms BERT, FutureTOD, and LLaMA on all metrics,

[6] We do not include PPTOD in the main results because they use supervised labels. We focus on unsupervised dialogue pretraining and do not compare with it for fairness.

[7] TOD-BERT uses the response contrastive loss as the pre-training objective on full MWOZ training data so we don't report its results on few-shot setting.

Original Dialogue	DivTOD's Dialogue	DivTOD w/o Alignment's Dialogue
usr:Do you have any restaurants that serve caribbean food? sys:There are no Caribbean restaurants. Would you like something else? usr:What about a restaurant that serves Chinese food? What is their phone number? sys:There are many Chinese restaurants to choose from. Do you have a preferred area of town? usr:no, it doesn't matter sys:Let me provide you with a couple options, Charlie Chan- phone 01223 361763 or Golden House - phone 01842 753771 usr:Thank you, goodbye sys:Thank you for calling. Goodbye.	usr:Do you have any restaurants that serve caribbean food? sys:Yes, there are several Caribbean restaurants in the area. Would you like me to provide you with a list of options? usr:What about a restaurant that serves Chinese food? What is their phone number? sys:There are many Chinese restaurants to choose from. Do you have a preferred area of town? usr:no, it doesn't matter sys:Let me provide you with a couple options, Charlie Chan- phone 01223 361763 or Golden House - phone 01842 753771 usr:Thank you, goodbye sys:You're welcome. Let me know if you need any further assistance.	usr:Do you have any restaurants that serve caribbean food? sys:Sure, here's the rewritten response: usr:What about a restaurant that serves Chinese food? What is their phone number? sys:Sure, here are a few options for Chinese restaurants in the area. Charlie Chan's phone number is 01223 361763, and Golden House's phone number is 01842 753771. Let me know if you have any other questions or if you'd like me to book a table for you. usr:no, it doesn't matter sys:Let me provide you with a couple options, Charlie Chan- phone 01223 361763 or Golden House - phone 01842 753771 usr:Thank you, goodbye sys:Thank you for calling. Goodbye.

Figure 3: Different Dialogue Cases. Original Dialogues refers to the dialogues from the original TOD dataset. DivTOD's Dialogue refers to the dialogues generated using the complete generating and aligning steps. DivTOD w/o Alignment's Dialogue refers to the dialogues generated after removing domain knowledge alignment.



(a) MWOZ (b) DSTC2

Figure 4: The ablation experiment on the impact of the number of diverse dialogues generated by large language models on TOD.

	Model	DSTC2		MWOZ	
		micro-F1	macro-F1	1-to-100	3-to-100
1 % Data	BERT	77.1%	25.8%	7.8%	20.5%
	BERT-mlm	79.6%	26.4%	13.0%	34.6%
	SimCSE	78.9%	27.3%	17.2%	32.6%
	TOD-BERT	82.9%	28.0%	-	-
	DSE	72.4%	21.4%	7.9%	21.2%
	FutureTOD†	77.2%	26.2%	21.7%	40.6%
	FutureTOD	83.7%	31.0%	35.8%	53.5%
	DivTOD†	79.0%	26.9%	24.6%	45.2%
	DivTOD	**85.7%**	**36.5%**	**36.9%**	**59.4%**
10 % Data	BERT	88.2%	34.8%	20.9%	45.4%
	BERT-mlm	91.8%	39.4%	22.3%	48.7%
	SimCSE	92.3%	40.5%	37.2%	60.6%
	TOD-BERT	90.6%	38.8%	-	-
	DSE	91.1%	39.0%	24.8%	49.4%
	FutureTOD	93.6%	40.9%	50.0%	72.8%
	DivTOD	**95.1%**	**45.6%**	**52.0%**	**76.5%**

Table 7: Dialogue act prediction on DSTC2 and response selection on MWOZ for few-shot settings. DivTOD† and FutureTOD† are the models obtained by removing MWOZ and CamRest676 from the pre-training corpus. For DivTOD, we also excluded diversed dialogues generated from these two datasets.

and is comparable to Vicuna. This indicates that the model has already gained strong context representation ability from the diverse dialogue data pre-training provided by Vicuna. However, the time cost and parameter size are much smaller than LLM like Vicuna, with a 14-fold and 70-fold reduction respectively

4.6 Representation Diversity

To understand whether our DivTOD can capture more diverse dialogue information and learn the intrinsic one-to-many diversity of TOD, we perform a qualitative analysis on the MWOZ test set. For each dialogue history, we select 2000 randomly sampled responses. We then compute the cosine distance between the representations of the dialogue history and response using a pre-trained response selection model in Table 4. We select the top 10 responses according to the cosine distance and compute Diversity and Coherence as the automatic metrics. Diversity denotes the number of unique types of dialogue acts in the top 10 responses. Coherence denotes average relevance scores between history and top-10 responses using a fine-tuned dual encoder in the response selection task.[8] We combine these two metrics to get the combined scores to measure the overall automatic response diversity and quality. The left part of Table 9 shows the automatic results of different pre-trained models. Our model has advantages in all metrics, indicating that our model can capture rich dialogue policies without sacrificing response relevance. We also find TOD-BERT achieves comparable performance on coherence but performs worst on diversity, even worse than BERT. It proves that the noise introduced by the selection of positive and negative samples in contrastive learning may hurt the one-to-many diversity of dialogue representations.

Following Zhang et al. (2020a), we conduct a human evaluation to assess the appropriateness of individual responses and the diversity among selected responses. The appropriateness(App) is scored on a Likert scale of 1-3 for each response, while the diversity is scored on a Likert scale of 1-5 for all top 10 responses. We sample one hundred dialogue histories and corresponding top 10 responses retrieved by different pre-trained models. These samples are then scored by three judges given the dialogue history. The right part of Table 9 shows the results of the human evaluation. We can find that the results of the human evaluation have the same trend as the automatic evaluation. Both the automatic evaluation and the human evaluation prove that our DivTOD model can learn the intrinsic one-to-many diversity of task-oriented dialogues.

[8]We use TOD-BERT model in Table 4, but we observe similar results using other response selection models.

Model	Dataset	1-to-100	3-to-100	5-to-100	10-to-100	Inference efficiency	Parameter size
BERT	MWOZ	1.8%	6.0%	9.9%	20.0%	89.9	110M
FutureTOD		2.1%	6.3%	10.4%	20.7%	89.9	110M
LLaMA-7b		2.2%	6.3%	10.4%	20.7%	5.3	7B
Vicuna-7b		**2.6%**	**7.3%**	**11.8%**	**22.7%**	5.3	7B
DivTOD (50k)		2.5%	6.3%	11.0%	21.2%	89.9	110M
BERT	DSTC2	1.0%	3.3%	6.3%	16.5%	89.9	110M
FutureTOD		1.7%	5.1%	9.0%	17.9%	89.9	110M
LLaMA-7b		2.1%	6.0%	10.1%	19.6%	5.3	7B
Vicuna-7b		2.0%	6.2%	10.5%	20.4%	5.3	7B
DivTOD (50k)		**2.2%**	**6.6%**	**11.1%**	**21.1%**	89.9	110M

Table 8: Response Selection on DSTC2 and MWOZ for zero-shot setting. We report 1-to-100, 3-to-100, 5-to-100, and 10-to-100 accuracy, which represents the ratio of the ground-truth response being ranked at the top-1,top-3, top-5 and top-10 given 100 candidates. Inference efficiency denotes the number of samples a model can infer per second when deployed on an Nvidia Tesla A100 GPU.

Model	Automatic			Human	
	Diversity	Conherence	Combineed	Diversity	App
BERT	5.50	0.668	12.18	1.97	1.23
BERT-mlm	5.43	0.689	12.32	2.07	1.67
SimCSE	5.48	0.675	12.23	1.93	1.47
TOD-BERT	5.05	0.709	12.14	2.20	1.87
DSE	6.17	0.680	12.97	2.53	1.33
FutureTOD	6.33	0.706	13.39	2.67	1.91
DivTOD	**7.92**	**0.730**	**15.22**	**2.88**	**1.94**

Table 9: The automatic results and human evaluation results of response diversity on the MWOZ test set. The combined score is the overall automatic result which is calculated as follows: Combined score = Diversity + 10*Coherence.

5 Related Work

Dialogue Pre-trained Language Models Zhang et al. (2020b) use pre-trained GPT-2 model (Radford et al., 2019) on Reddit data for open-domain dialogue response generation. PLATO (Bao et al., 2019) pre-trains a dialog generation model with discrete latent variables using Twitter and Reddit data, which implicitly models dialog policy and solves the one-to-many mapping problem in open-domain dialog generation. However, since these models focus on chitchat dialogue, we do not compare them with our DivTOD. Wu et al. (2020); Zhou et al. (2022) use contrastive learning to learn TOD dialogue representations. Henderson et al. (2020); Liu et al. (2021) use similar ideas for dialogue retrieval and response selection. Zeng et al. (2023) proposes a non-contrastive framework that distills future knowledge into the representation of the previous dialogue. Apart from these unsupervised methods, Zhou et al. (2022); He et al. (2022) use labeled dialogue data for supervised or semi-supervised pre-training. Since we focus on unsupervised TOD pre-training in this paper, we do not compare these models and leave it to future work.

Enhancing small models with LLMs Large Language Models (LLMs) (Han et al., 2021; Bommasani et al., 2021), such as ChatGPT and GPT-4 (OpenAI, 2023), have demonstrated excellent generalization abilities in many language-related tasks. Recently, there have been many efforts to distill powerful LLMs for data augmentation, hoping to obtain equally powerful larger models through this approach without modifying the training objectives or model structures. For example, SelfInstruct (Wang et al., 2022) and Alpaca (Touvron et al., 2023) generate 52k high-quality instruction-response pairs by distilling Text-Davinci-003, based on 175 seed tasks. In another line of work, LLMs are used to improve the ability of small models for specific tasks. Ho et al. (2022) and Hsieh et al. (2023) use LLMs to generate rationales that enhance the model's reasoning ability. Liang et al. (2023) uses LLMs as a math tutor to improve the model's math ability. In impossible distillation (Jung et al., 2023), LLMs help models generate high-quality and controllable summarizations and paraphrases. In contrast to previous work, we transfer rich background knowledge from LLMs to smaller models while filtering out domain knowledge that is irrelevant to the task-oriented dialogue system.

6 Conclusion

We propose a new dialogue pre-training called DivTOD to diversify task-oriented dialogue representations by modeling the intrinsic one-to-many diversity of human conversations. DivTOD guides LLMs to transfer diverse background knowledge to smaller models while filtering domain knowledge that conflicts with task-oriented dialogues. Our experiments on various task-oriented dialogue tasks show that DivTOD outperforms FutureTOD, TOD-BERT, DSE, and other strong baselines. We plan to release all pre-trained models and code to facilitate future research. In the future, we hope to explore larger pre-trained models and more task-oriented dialogue corpora and extend similar ideas to generative dialogue models.

Limitations

While DivTOD achieves significant improvements over existing baselines, there are still directions to explore for future work. (1) We have designed a simple and effective method for LLMs to help dialogue pre-train models capture the intrinsic one-to-many diversity of human conversations. However, we have not considered solving this problem through the structure of the dialogue pretraining model. In the future, we will explore designing more efficient architectures for dialogue pretraining models and more efficient methods of knowledge transfer. (2) DivTOD only focuses on dialogue understanding tasks, such as dialogue act prediction and response selection. In the future, we will expand the idea of LLM collaborating with small models to generative dialogue pre-trained models. (3) We attempt various instructions to constrain the responses of M_T, including zero-shot prompts. However, these methods have not been very effective. For instance, the pass rate of the zero-shot method is relatively low in our post-filter. So we did not report these results. In the future, we plan to explore more advanced prompt techniques, such as the CoT method, to enhance our approach.

Ethics Statement

We use Large Language Models (LLMs) to generate diverse responses. Despite our efforts to align domain knowledge, LLMs are inevitably prone to generating biased content. We anticipate that future research will focus on reducing the anti-social biases inherent in LLMs.

References

Layla El Asri, Hannes Schulz, Shikhar Sharma, Jeremie Zumer, Justin Harris, Emery Fine, Rahul Mehrotra, and Kaheer Suleman. 2017. Frames: a corpus for adding memory to goal-oriented dialogue systems. *ArXiv*, abs/1704.00057.

Siqi Bao, H. He, Fan Wang, and Hua Wu. 2019. Plato: Pre-trained dialogue generation model with discrete latent variable. In *Annual Meeting of the Association for Computational Linguistics*.

Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, T. J. Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeff Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *ArXiv*, abs/2005.14165.

Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gasic. 2018. Multiwoz - a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. In *EMNLP*.

Bill Byrne, Karthik Krishnamoorthi, Chinnadhurai Sankar, Arvind Neelakantan, Daniel Duckworth, Semih Yavuz, Ben Goodrich, Amit Dubey, Andy Cedilnik, and Kyu-Young Kim. 2019. Taskmaster-1: Toward a realistic and diverse dialog dataset. In *EMNLP*.

WL Chiang, Z Li, Z Lin, Y Sheng, Z Wu, H Zhang, L Zheng, S Zhuang, Y Zhuang, JE Gonzalez, et al. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality, mar. 2023.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Ning Ding, Yulin Chen, Bokai Xu, Yujia Qin, Zhi Zheng, Shengding Hu, Zhiyuan Liu, Maosong Sun, and Bowen Zhou. 2023. Enhancing chat language models by scaling high-quality instructional conversations. *arXiv preprint arXiv:2305.14233*.

Mihail Eric, Lakshmi. Krishnan, François Charette, and Christopher D. Manning. 2017. Key-value retrieval networks for task-oriented dialogue. *ArXiv*, abs/1705.05414.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Xu Han, Zhengyan Zhang, Ning Ding, Yuxian Gu, Xiao Liu, Yuqi Huo, Jiezhong Qiu, Yuan Yao, Ao Zhang, Liang Zhang, et al. 2021. Pre-trained models: Past, present and future. *AI Open*, 2:225–250.

Wanwei He, Yinpei Dai, Binyuan Hui, Min Yang, Zhen Cao, Jianbo Dong, Fei Huang, Luo Si, and Yongbin Li. 2022. Space-2: Tree-structured semi-supervised contrastive pre-training for task-oriented dialog understanding. In *COLING*.

Matthew Henderson, Iñigo Casanueva, Nikola Mrkvsi'c, Pei hao Su, Tsung-Hsien, and Ivan Vulic. 2020. Convert: Efficient and accurate conversational representations from transformers. *ArXiv*, abs/1911.03688.

Matthew Henderson, Blaise Thomson, and J. Williams. 2014. The second dialog state tracking challenge. In *SIGDIAL Conference*.

Namgyu Ho, Laura Schmid, and Se-Young Yun. 2022. Large language models are reasoning teachers. *arXiv preprint arXiv:2212.10071*.

Cheng-Yu Hsieh, Chun-Liang Li, Chih-Kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alexander Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. 2023. Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes. *arXiv preprint arXiv:2305.02301*.

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

Jaehun Jung, Peter West, Liwei Jiang, Faeze Brahman, Ximing Lu, Jillian Fisher, Taylor Sorensen, and Yejin Choi. 2023. Impossible distillation: from low-quality model to high-quality dataset & model for summarization and paraphrasing. *arXiv preprint arXiv:2305.16635*.

Andreas Köpf, Yannic Kilcher, Dimitri von Rütte, Sotiris Anagnostidis, Zhi-Rui Tam, Keith Stevens, Abdullah Barhoum, Nguyen Minh Duc, Oliver Stanley, Richárd Nagyfi, et al. 2023. Openassistant conversations–democratizing large language model alignment. *arXiv preprint arXiv:2304.07327*.

Stefan Larson, Anish Mahendran, Joseph J. Peper, Christopher Clarke, Andrew Lee, Parker Hill, Jonathan K. Kummerfeld, Kevin Leach, Michael A. Laurenzano, Lingjia Tang, and Jason Mars. 2019. An evaluation dataset for intent classification and out-of-scope prediction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1311–1316, Hong Kong, China. Association for Computational Linguistics.

Sungjin Lee, Hannes Schulz, Adam Atkinson, Jianfeng Gao, Kaheer Suleman, Layla El Asri, Mahmoud Adada, Minlie Huang, Shikhar Sharma, Wendy Tay, and Xiujun Li. 2019. Multi-domain task-completion dialog challenge.

Xiujun Li, Sarah Panda, Jingjing Liu, and Jianfeng Gao. 2018. Microsoft dialogue challenge: Building end-to-end task-completion dialogue systems. *ArXiv*, abs/1807.11125.

Zhenwen Liang, Wenhao Yu, Tanmay Rajpurohit, Peter Clark, Xiangliang Zhang, and Ashwin Kaylan. 2023. Let gpt be a math tutor: Teaching math word problem solvers with customized exercise generation. *arXiv preprint arXiv:2305.14386*.

Che Liu, Rui Wang, Jinghua Liu, Jian Sun, Fei Huang, and Luo Si. 2021. Dialoguecse: Dialogue-based contrastive learning of sentence embeddings. *ArXiv*, abs/2109.12599.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *ArXiv*, abs/1907.11692.

Nikola Mrksic, Diarmuid Ó Séaghdha, Tsung-Hsien Wen, Blaise Thomson, and Steve J. Young. 2017. Neural belief tracker: Data-driven dialogue state tracking. In *ACL*.

OpenAI. 2023. Gpt-4 technical report.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke E. Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Francis Christiano, Jan Leike, and Ryan J. Lowe. 2022. Training language models to follow instructions with human feedback. *ArXiv*, abs/2203.02155.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *ACL*.

Abhinav Rastogi, Xiaoxue Zang, Srinivas Sunkara, Raghav Gupta, and Pranav Khaitan. 2020. Towards scalable multi-domain conversational agents: The schema-guided dialogue dataset. In *AAAI*.

Lina Maria Rojas-Barahona, Milica Gavsic, Nikola Mrksic, Pei hao Su, Stefan Ultes, Tsung-Hsien Wen, Steve J. Young, and David Vandyke. 2017. A network-based end-to-end trainable task-oriented dialogue system. In *EACL*.

Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15:1929–1958.

Yixuan Su, Lei Shu, Elman Mansimov, Arshit Gupta, Deng Cai, Yi-An Lai, and Yi Zhang. 2021. Multi-task pre-training for plug-and-play task-oriented dialogue system. *arXiv preprint arXiv:2109.14739*.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aur'elien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models. *ArXiv*, abs/2302.13971.

Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2022. Self-instruct: Aligning language model with self generated instructions. *arXiv preprint arXiv:2212.10560*.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou. 2022. Chain of thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903*.

Sean Welleck, Jason Weston, Arthur D. Szlam, and Kyunghyun Cho. 2019. Dialogue natural language inference. In *ACL*.

Adina Williams, Nikita Nangia, and Samuel R. Bowman. 2018. A broad-coverage challenge corpus for sentence understanding through inference. In *NAACL*.

Chien-Sheng Wu, Steven C. H. Hoi, Richard Socher, and Caiming Xiong. 2020. Tod-bert: Pre-trained natural language understanding for task-oriented dialogue. In *EMNLP*.

Weihao Zeng, Keqing He, Yejie Wang, Chen Zeng, Jingang Wang, Yunsen Xian, and Weiran Xu. 2023. FutureTOD: Teaching future knowledge to pre-trained language model for task-oriented dialogue. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6532–6546, Toronto, Canada. Association for Computational Linguistics.

Yichi Zhang, Zhijian Ou, and Zhou Yu. 2020a. Task-oriented dialog systems that consider multiple appropriate responses under the same context. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9604–9611.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and William B. Dolan. 2020b. Dialogpt : Large-scale generative pre-training for conversational response generation. In *ACL*.

Zhihan Zhou, Dejiao Zhang, Wei Xiao, Nicholas Dingwall, Xiaofei Ma, Andrew O. Arnold, and Bing Xiang. 2022. Learning dialogue representations from consecutive utterances. In *NAACL*.

Yukun Zhu, Ryan Kiros, Richard S. Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 19–27.

Name	Dialogue	Utterance	Avg. Turn	Domain
MetaLWOZ	37,884	432,036	11.4	47
Schema	22,825	463,284	20.3	17
Taskmaster	13,215	303,066	22.9	6
MWOZ	10,420	71,410	6.9	7
MSR-E2E	10,087	74,686	7.4	3
SMD	3,031	15,928	5.3	3
Frames	1,369	19,986	14.6	3
WOZ	1,200	5,012	4.2	1
CamRest676	676	2,744	4.1	1

Table 10: Data statistics for our pre-training task-oriented dialogue datasets.

A Pre-training Data Statistics

We use the nine different task-oriented datasets collected by (Wu et al., 2020): MetaLWOZ (Lee et al., 2019), Schema (Rastogi et al., 2020), Taskmaster (Byrne et al., 2019), MWOZ (Budzianowski et al., 2018), MSR-E2E (Li et al., 2018), SMD (Eric et al., 2017), Frames (Asri et al., 2017), WOZ (Mrksic et al., 2017), CamRest676 (Rojas-Barahona et al., 2017). We show the full statistics in Table 10.

B Baselines

DivTOD is evaluated on a variety of downstream tasks and compared to several well-established baselines. One such baseline is BERT (Devlin et al., 2019), which is the original BERT-base-uncased model that was pre-trained on a large text corpus. Another baseline is BERT-mlm, which is a version of BERT that underwent continual pre-training using MLM on our pre-training dialogue corpus. DialoGPT (Zhang et al., 2020b) is also included as a baseline, it is a decoder-only dialogue generation model that utilizes a language modeling target. SimCSE (Gao et al., 2021) constructs positive pairs using Dropout and undergoes further pre-training on the same TOD corpus. TOD-BERT (Wu et al., 2020) employs a contrastive response selection objective, treating a response utterance and its dialogue context as a positive pair. DSE (Zhou et al., 2022) takes consecutive utterances from the same dialogue as positive pairs.[9] FutureTOD(Zeng et al., 2023) uses a non-contrastive self-training framework with a self-distillation mechanism. It should be noted that some dialogue pre-training methods adopt an encoder-decoder architecture, but they usually use supervised settings, i.e. using labeled NLI datasets (Williams et al., 2018; Welleck et al., 2019) or dialogue act labels (He et al., 2022). How-

[9] In the interest of fairness, we use the unsupervised version of DSE, as done by Zeng et al. (2023).

ever, our focus is on unsupervised dialogue pre-training, and for fairness, we do not compare with them.

To validate the unsupervised embedding capability of our model, we also compared it with the 7B model LLaMA (Touvron et al., 2023) and Vicuna (Chiang et al.) in a zero-shot response selection task. LLaMA is a powerful open-source large-scale model trained on a large corpus, while Vicuna is fine-tuned based on LLaMA using 70K high-quality conversation data.

C Implementation Details

C.1 LLM generating Details

We use Vicuna as LLM to generate diverse system responses and to align domain knowledge. For generation settings, the maximum generation length is 1024, the temperature is 0.7, and in order to ensure dialogue diversity, we choose to perform sampling. For verification settings, we obtain the logits corresponding to True and False in the first word of Vicuna's decoding as the basis. In addition, if the model does not understand the task[10], it will also be considered as not passing verification. If the response does not pass verification, Vicuna will generate a response again and verify it. The original dialogue will be retained if the generated response fails verification 5 times. Half responses in the dialogue will be rewritten. After diverse system response generation, all the dialogue will be merged into the original dataset as the new dataset to pre-train.

C.2 Pre-training Details

In DivTOD, we utilize a batch size of 48 and set the maximum input length to 512. The models are initialized using BERT-base-uncased and optimized using the Adam optimizer and a linear learning rate scheduler with an initial learning rate of 5e-5. A dropout ratio of 0.2 is employed and the mask ratio is set to 15%. The predictor MLP head consists of two linear layers and a ReLU activation layer with an input dimension of 768 and a middle hidden dimension of 512. Upon completion of pre-training, all parameters of the Bert encoder are saved and the MLP head module is dropped for fine-tuning downstream tasks. Pre-training takes three days using an early-stopped strategy based on perplexity scores of a held-out development conducted on eight NVIDIA Tesla A100 GPUs. We use pre-trained models including BERT-MLM and TOD-BERT released by (Wu et al., 2020), DSE model released by (Zhou et al., 2022), and FutureTOD model released by (Zeng et al., 2023). We re-implement SimCSE(Gao et al., 2021) using Dropout to construct positive pairs and augment every single utterance obtained through Dropout on our pre-training corpora. In terms of computational efficiency during pre-training, our DivTOD model is comparable to other baselines.

C.3 Finetuning Details

For BERT-mlm and TOD-BERT, we use the results reported by TOD-BERT (Wu et al., 2020) directly. We adopt the same hyperparameters for all downstream tasks except the batch size and learning rate. We finetune all downstream tasks with the original dataset for 50 epochs with an early-stopped strategy evaluated on the validation set every 100 steps with patience set to 10. We respectively set the batch size to 8, 25, 16, and 100 for intent recognition, dialogue state tracking, dialogue act prediction, and response selection and keep the learning rate to 5e-5 for all the tasks.

D Prompt Examples

We provided prompts for generating diversified responses and aligning domain knowledge in Figure 5 and Figure 6, respectively.

[10] For example, answering "I am a large language model" or "Okay, here is the written response you need", etc.

Generation Instruction (I^P)

"A chat between a curious user and an artificial intelligence assistant. "
"The assistant gives helpful, detailed, and polite answers to the user's questions. "
"I will give you a dialogue between a user and a system. USR means user, SYS means system. "
"However, there is a responce made by system in the conversation covered in ink, expressed as' masked' in the dialogue. "
"As a talent writer, your task is to create a response to replace the [masked] within a dialogue to make it more appropriate and logically coherent within the its corresponding context. "
"Be sure the system want to reply users query just before the SYS(Given Response) in the dialogue, so you should act as the system, which means you can complete all task system can do like booking a table. "
"The system may also want to ask questions to the user, if user just apply relative information to the SYS(Given Response) in the dialogue. "
"Do not responce to the user's query that after the SYS(Given Response), just rewrite the system response. "
"Do not ask information that user already provided before the SYS(Given Response). "
"Do not provide information that user did not ask for before the SYS(Given Response). "
"Do not write fine-grained information like address,time etc. that user provide after [masked]. "
"Do not provide too much information. "
"Do not write any prefix that is not a part of the system response, like 'Sure, here's the rewritten response'"
"Do not write any user responce. The response should be one line, only focus on SYS(Given Response) line. "
"I will give you the query you should responce and what user reply to your responce. "
"Before your work, I will give you a example and a good assistant answer to help you understand the task.",

Demonstrations (D^P,S^P)

#example#
#Given Dialogue#
USR: Hi, could you get me a restaurant booking on the 8th please?
SYS(Given Response): [masked]
USR: Could you get me a reservation at P.f. Chang's in Corte Madera at afternoon 12?
SYS: Please confirm your reservation at P.f. Chang's in Corte Madera at 12 pm for 2 on March 8th.
#Response to#: Hi, could you get me a restaurant booking on the 8th please?
#Response#:"Thanks for considering booking with us! May I know if you have any particular cuisine or ambiance preference for the restaurant? And what time exactly would you like to dine in?"

Input(D')

#your task#
#Given Dialogue#
USR:I am looking for a restaurant that is moderately priced and serves Cantonese food.
SYS:There are no restaurants that serve Cantonese food in the moderate price range.
USR:How about chinese type of food?
SYS:the Golden wok serves chinese food and is in the moderate price range. Would you like their location?
USR:Their phone number please.
SYS(Given Response): [masked]
USR:What is the area?
SYS:it is located in the north part of town
USR:Thank you for your help. Good bye.
SYS:Goodbye.
#Response to#: Their phone number please.
#Response#:

Figure 5: The complete prompt example for generating diversified responses.

Algorithm 1 Generating-Aligning Steps

Initialization: Generation Prompt with Example P, Eval Prompt with Example E, Try Turns T, Model M_T
Input: a Dialogue D, Dialogue Turns Number n
for N in $[1, \lfloor n/2 \rfloor]$ **do**
  $try_number = 0$, $filtering_result$=False
  while $try_number < T$ **do**
    D' = Replace S_i in D into [masked]
    $S'_i = M_T(P, D')$
    D'' = Replace S_i in D into S'_i
    $filtering_result = M_T(E, D'')$
    try_number += 1
    if $filtering_result$ is True **then**
      $D = D''$
      break
    end if
  end while
end for
Output: D

Eval Instruction (I^E)

"A chat between a curious user and an artificial intelligence assistant. "
"The assistant gives helpful, detailed, and polite answers to the user's questions. "
"I want you to act as a System Response Checker. "
"You need to check if a given system response(Given Response) within a dialogue(Given Dialogue) is logically coherent with its corresponding context."
"you should assure that the Given Response should not appear information that system did not know logically, like the information provided by user after the responce(Given Response). "
"If the responce(Given Response) is a question, the check the logic between the question and user's answer. "
"IF the responce(Given Response) is not a part of the system response in a dialogue but a responce to rewrite commamd like 'Sure, here's the rewritten response:', you should answer 'False' and should not answer 'True'. "
"Your answer should be either 'True' or 'False'. Do not output any other words. I will give you 2 examples to help you understand the task."

Demonstrations (D^E,S^E)

#example#
#Given Dialogue#
USR: I would like to book a hotel in Washington D.C.
SYS(Given Response): Great, I'm glad that you are considering our hotel options. May I know what type of room or amenities are you looking for? And when would you like to check in?
USR: Yes, that room is good.
SYS: Would you like to make a reservation?
#Check Result#: False

#example#
#Given Dialogue#
USR: Hi, could you get me a restaurant booking on the 8th please?
SYS(Given Response): Any preference on the restaurant, location and time?
USR: Could you get me a reservation at P.f. Chang's in Corte Madera at afternoon 12?
SYS: Please confirm your reservation at P.f. Chang's in Corte Madera at 12 pm for 2 on March 8th.
#Check Result#: True

Input(D'')

#your task#
#Given Dialogue#
USR:I am looking for a restaurant that is moderately priced and serves Cantonese food.
SYS:There are no restaurants that serve Cantonese food in the moderate price range.
USR:How about chinese type of food?
SYS:the Golden wok serves chinese food and is in the moderate price range. Would you like their location?
USR:Their phone number please.
SYS(Given Response): Golden Wok's phone number is [masked].
USR:What is the area?
SYS:it is located in the north part of town
USR:Thank you for your help. Good bye.
SYS:Goodbye.
#Check Result#:

Figure 6: The complete prompt example for aligning domain knowledge.