Regularizing Dialogue Generation by Imitating Implicit Scenarios

Shaoxiong Feng[1] Xuancheng Ren[2] Hongshen Chen[3] Bin Sun[1] Kan Li[1] Xu Sun[2]
[1]Beijing Institute of Technology
[2]MOE Key Laboratory of Computational Linguistics, School of EECS, Peking University
[3]JD.com
{shaoxiongfeng, binsun, likan}@bit.edu.cn {renxc, xusun}@pku.edu.cn
ac@chenhongshen.com

Abstract

Human dialogues are scenario-based and appropriate responses generally relate to the latent context knowledge entailed by the specific scenario. To enable responses that are more meaningful and context-specific, we propose to improve generative dialogue systems from the scenario perspective, where both dialogue history and *future conversation* are taken into account to implicitly reconstruct the scenario knowledge. More importantly, the conversation scenarios are further internalized using imitation learning framework, where the conventional dialogue model that has no access to future conversations is effectively regularized by transferring the scenario knowledge contained in hierarchical supervising signals from the scenario-based dialogue model, so that the future conversation is not required in actual inference. Extensive evaluations show that our approach significantly outperforms state-of-the-art baselines on diversity and relevance, and expresses scenario-specific knowledge.

1 Introduction

Neural dialogue generation has drawn increasing attention due to its vast commercial values and practical demands. Typically, given the dialogue history, neural dialogue models, such as plain Seq2Seq model (Sutskever et al., 2014) and Transformer (Vaswani et al., 2017), learn to predict responses via maximum likelihood estimation (Vinyals and Le, 2015; Shang et al., 2015).

Different from other sequence generation tasks, such as machine translation and paraphrase generation, the dialogue generation task can be regarded as a loose-coupling task, which has much freedom in the semantic and the linguistic aspects of the generated responses. However, it is often hard for the existing models to handle such freedom, compared to the fact that humans have no problem in giving specific yet varied responses even for open-ended dialogue history (Shen et al., 2018; Csaky et al., 2019). One important reason is that we can extend the given dialogue with many possible scenarios of enriched, imaginative background information from our experience and world knowledge, to which existing systems have no access.

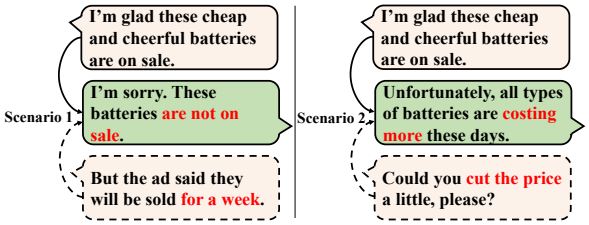

Figure 1: Examples of one dialogue history and its different responses followed by related future conversations. Different responses imply various conversation scenarios, which can be inferred by different relevant future conversations.

It is beneficial for the dialogue systems to build upon such scenarios to facilitate dialogue generation. However, manually annotating the scenario contexts is intractable in terms of both difficulty and quantity. In turn, we find that such scenarios are naturally contained in existing multi-turn dialogue corpora, where the entire dialogue of both dialogue history and future conversation with respect to the current utterance implicitly represents a specific dialogue scenario. An example is given in Figure 1. For Scenario 1, "for a week" in the future conversation suggests the response is related to time. For Scenario 2, "cut the price" indicates that the response contains price information.

Therefore, we reconstruct the dialogue task that only relies on dialogue history into a scenario-based response generation task. In order to enrich the conversation scenario, we employ future conversations together with dialogue histories to learn implicit conversation scenarios, which provide more

Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pages 6592–6604,
November 16–20, 2020.

semantic constraints to guide the response generation. We further propose a novel model to handle this new type of training data consisting of {*implicit scenario, response*} pairs.

It should be noted that the scenario-based dialogue model relies on future conversations that are inaccessible in inference. Rather than simply searching the training corpora for possible scenarios, we propose an imitation learning framework to drive the conventional dialogue model to absorb the corresponding scenario knowledge from the scenario-based dialogue model. Specifically, the scenario-based dialogue model serves as a teacher, and the conventional dialogue model that relies solely on dialogue history serves as a student that mimics the outputs of the teacher. Under the regularization of scenario knowledge, the student is effectively guided towards a wider local minimum that represents better generalization performance (Chaudhari et al., 2017; Keskar et al., 2017). To facilitate knowledge transfer, the student mimics the teacher on every layer instead of just the top layer, which alleviates the delayed supervised signal problem using hierarchical semantic information in the teacher (Li et al., 2019a). Besides containing the information of future conversations, the distilled knowledge (Hinton et al., 2015) is also a less noisy and more "deterministic" supervised signal in comparison to real-world responses (Lee et al., 2018; Guo et al., 2019), which provides the student with smoother sequence trajectories that are easier to fit.

We highlight our contributions as follows:

- We introduce future conversations together with dialogue histories to learn implicit conversation scenarios, which provide more semantic constraints to drive the responses to be meaningful and relevant to the real-world scenario-specific knowledge.
- We propose an imitation learning framework that bridges the gap between training and inference in the accessibility of future conversations. We also demonstrate why imitation learning works and further how to enhance the imitation learning.
- Our model achieves better results than state-of-the-art baselines on four datasets. Extensive analysis demonstrates the effectiveness and the scalability of the implicit conversation scenarios and the proposed imitation learning framework.

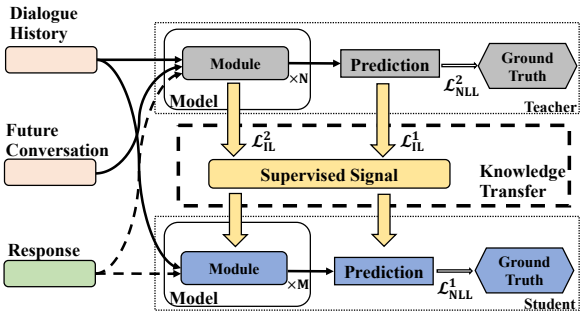

Figure 2: Illustration of the imitation learning framework transferring scenario knowledge from the teacher to the student. Top (Teacher): the scenario-based dialogue model. Bottom (Student): the conventional dialogue model.

2 Proposed Approach

In this section, we first introduce the scenario-based dialogue model, then describe the imitation learning framework shown in Figure 2, and finally present the training objective.

2.1 Scenario-Based Dialogue Model

The conventional dialogue model takes a sequence of dialogue history $X = \{x_1, \ldots, x_T\}$ as input, and generates a response $Y = \{y_1, \ldots, y_{T'}\}$ word by word, where T and T' represent the length of source side and target side respectively. The maximum likelihood estimation is usually used to train the model, which can also be expressed as minimizing the negative log-likelihood:

$$\mathcal{L}^1_{\mathrm{NLL}}(\theta_1) = -\sum_{i=1}^{T'}\sum_{k=1}^{|\mathcal{V}|} \mathbb{I}\{y_i = k\} \cdot \log p(y_i = k | y_{<i}, X_h; \theta_1), \quad (1)$$

where $|\mathcal{V}|$ is the size of vocabulary, θ_1 is a set of parameters, and X_h represents the input sequence that is from dialogue history.

However, dialogue task allows responses to continue the dialogue topic from many aspects or even introduce a new topic depending on various conversation scenarios or semantic constraints, which dramatically increases the difficulty of prediction without any specific scenario information besides the hints from the given dialogue history. Moreover, labeling the scarce scenario information is labor-consuming and impractical. Instead, we resort to easy-to-access but underutilized future conversations that exist in all multi-turn dialogue corpora. By combining the dialogue history and its corresponding future conversation, we introduce the implicit conversation scenario into existing dialogue

models to provide more semantic constraints and reduce the difficulty of prediction.

Concretely, we enforce the model to use implicit conversation scenarios to generate responses from two aspects. Different from previous dialogue models only based on the dialogue history X_h to predict the response Y, the future conversation X_f is also considered as part of input so that the model can look ahead and predict more purposefully. Intuitively, our training pair $\{(X_h, X_f), Y\}$ induces the model to imitate humans to produce the scenario-specific response.

We also redesign the sequence generation architecture to handle the proposed training pair. The attention module in each layer calculates the weight of the contextualized token representations from the encoder based on the information that has been generated in the decoder, and then returns the context c_h. In order to consider the future conversation X_f, we apply another encoder to produce the contextualized token representations of X_f, which will be further extracted as the context c_f by the attention module. The new encoder shares the parameters with the original encoder. Meanwhile, the output of the attention module is the concatenation of the past context c_h and the future context c_f. Finally, the training criterion is formulated as the following negative log-likelihood:

$$\mathcal{L}^2_{\mathrm{NLL}}(\theta_2) = -\sum_{i=1}^{T'}\sum_{k=1}^{|\mathcal{V}|} \mathbb{I}\{y_i = k\} \cdot \log p(y_i = k | y_{<i}, X_h, X_f; \theta_2), \tag{2}$$

where θ_2 is a set of parameters to minimize the NLL loss for the scenario-based dialogue model.

2.2 Imitation Learning

In inference, future conversations are inaccessible, which means implicit conversation scenarios cannot be constructed. Thus, the performance improvement from the scenario-based dialogue model cannot facilitate the generation of high-quality responses in practice. In order to bridge this gap between training and inference, we propose an imitation learning framework, in which we regard the scenario-based dialogue model as a teacher and the conventional dialogue model as a student. Through step-by-step imitation, including fine-grained prediction imitation and intermediate representation imitation, scenario knowledge distilled from the teacher regularizes the student to reach a robust local minimum and obtain significant generalization performance in inference.

2.2.1 Fine-Grained Prediction Imitation

Compared with the ground-truth labels, the soft predictions (i.e., the probability distribution from the output layer) contain more fine-grained and valuable information, such as the similarity of labels and potential future conversations. Moreover, the soft predictions provide less noisy and more "deterministic" targets that are easy to mimic. To transfer knowledge from the teacher, instead of taking the one-hot representation of Y as the target, we minimize the cross-entropy of the predicted probability distribution between the teacher and the student:

$$\mathcal{L}^1_{\mathrm{IL}}(\theta_1, \theta_2) = -\sum_{i=1}^{T'}\sum_{k=1}^{|\mathcal{V}|} p(y_i = k | y_{<i}, X_h, X_f; \theta_2) \cdot \log p(y_i = k | y_{<i}, X_h; \theta_1) \tag{3}$$

2.2.2 Intermediate Representation Imitation

Only transferring knowledge from the output layer has a limited effect on the student to use implicit conversation scenarios. When the student network is very deep, the supervised signals from the output layer hardly conduct an effective update and regularization on the parameters of intermediate layers, which will make the imitation learning framework quickly reach saturation (Romero et al., 2015; Sun et al., 2019). This problem prevents the student from scaling to deeper models to further improve the model performance.

To tackle this problem, we extend the range of imitation learning from the soft predictions in the output layer to the output h of intermediate layers to guide the imitation process. Specifically, we penalize the discrepancy of hidden states in intermediate layers between the teacher and the student:

$$\mathcal{L}^2_{\mathrm{IL}}(\theta_1, \theta_2) = \sum_{i=1}^{T'}\sum_{l=1}^{|\mathcal{O}|} f\big(h^t_{il}(X_h, X_f; \theta_2), h^s_{il}(X_h; \theta_1)\big), \tag{4}$$

where $|\mathcal{O}|$ is the number of intermediate layers, h^t_{il} and h^s_{il} are the outputs of intermediate layers in the teacher and the student respectively, and $f(\cdot)$ is the measurement function.

$$f(\cdot) = \begin{cases} \phi(h^s_{il}, h^t_{il}), & \text{if } \phi(h^s_{il}, h^t_{il}) \geq \alpha; \\ 0, & \text{else .} \end{cases} \tag{5}$$

where $\phi(\cdot)$ is the mean-squared-error (MSE) loss. Because we observe that directly applying the MSE loss as an additional loss hurts the stability of the imitation learning process, we set a scalar threshold α to loose this constraint.

2.3 Training

Combining the NLL loss in Equation (1) with the IL losses in Equation (3) and Equation (4), the final objective function of the student is formulated as:

$$\mathcal{L} = \mathcal{L}^1_{NLL} + \lambda_1(\mathcal{L}^1_{\text{IL}} + \mathcal{L}^2_{\text{IL}}), \tag{6}$$

where λ_1 is a hyper-parameter that balances the importance of the NLL loss and the IL losses.

Because the scenario knowledge is only transferred from the teacher by hierarchical supervised signals, our imitation framework has the following three advantages: (1) Compared with the fine-tuning style of knowledge transfer (Dai and Le, 2015; Howard and Ruder, 2018), the proposed imitation framework does not affect the teacher, i.e., the knowledge learned from the teacher will not be forgotten. (2) The proposed method is model agnostic. Thus, the imitation object can be extended from one teacher to multiple teachers, such as incorporating a language model besides the scenario-based dialogue model. (3) The imitation process does not change the current objective function, which means the previous work of modifying objective function can serve as a complementary to improve the model performance further.

3 Experiment

3.1 Datasets

DailyDialog It is provided by Li et al. (2017b), which contains various dialogue topics about daily life. We randomly select 27K, 2.5K, and 1.5K pairs for training, validation, and testing.

PersonaChat It is gathered by assigning two Amazon Turkers with their personas to chat with each other (Zhang et al., 2018a). We only use the conversation section and split it to 67K, 8.5K, and 8K pairs for training, validation, and testing.

OpenSubtitles It is collected from movie subtitles and consists of more than 60M scripted lines (Lison and Tiedemann, 2016). We randomly extract 1500K, 50K, and 25K pairs for training, validation, and testing.

For all datasets, every seven consecutive dialogue turns form a training example, in which the first three turns, the middle turn, and the last three turns are taken as *dialogue history*, *response*, and *future conversation*, respectively.

We also conducted the experiment on a multi-domain goal-oriented dataset called **MultiWOZ**, which is simplified by us as a general dialogue generation task. The detailed description of MultiWOZ and data pre-processing is provided in Appendix A.

3.2 Baselines

We re-implemented two classes of six baselines for comparison. The detailed settings of baselines are provided in Appendix B.

LSTM-Based One class is based on LSTM, including **Seq2Seq+Att**, which contains a vanilla Seq2Seq model (Sutskever et al., 2014) with attention mechanism (Bahdanau et al., 2015), **VHRED+BOW** (Serban et al., 2017), which introduces a continuous latent variable attached to the response information into HRED (Serban et al., 2016) and applies BOW loss (Zhao et al., 2017) as a complementary with KL annealing, and **NEXUS** (Shen et al., 2018), which further uses the future conversation to incorporate more scenario information into the latent variable.

Transformer-Based The other class is based on **Transformer** (Vaswani et al., 2017), including itself, **ReCoSa** (Zhang et al., 2019a), and **CHMAM** (Tao et al., 2018), which consists of Multi-Head Attention Mechanism (MHAM) and an attention weight regularizer. Both ReCoSa and CHMAM aim to extract more relevant and diverse scenario information from dialogue history.

3.3 Experiment Settings

Based on the performance including the loss and metrics on the validation dataset, we trained baselines and our models with the following hyper-parameters. According to the scale of the dataset, the vocabulary sizes for OpenSubtitles, DailyDialog, PersonaChat, and MultiWOZ are set to 50k, 20k, 20k, and 18k, respectively. We use separate word embeddings for the encoder and the decoder, and the word embedding dimension is 256. All the parameters are initialized randomly from a normal distribution $\mathcal{N}(0, 0.0001)$. All models are trained using Adam (Kingma and Ba, 2015) with a learning rate of 0.001 and gradient clipping at 2.0. The batch size is 128. The hyper-parameters in our proposed appraoch are set as $\alpha = 0.01$ and $\lambda_1 = 2.0$. Our models, i.e., **RegDG**, the imitating student conventional model, and **Transformer-IF**, the imitated teacher scenario-based model, are based on Transformer.

DailyDialog	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	D^u_{kl} ↓	D^b_{kl} ↓	PPL ↓	BLEU ↑	GRE ↑	AVE ↑	EXT ↑	COH ↑
Seq2Seq+Att	0.42	1.66	3.83	13.865	29.096	116.72	16.7	0.4729	0.5308	0.3131	0.6096
VHRED+BOW	0.95	3.34	6.93	12.933	28.366	111.91	17.9	0.4801	0.5214	0.3008	0.5981
NEXUS	0.92	3.45	7.48	13.204	28.615	114.85	18.6	**0.4827**	0.5415	0.3105	0.6254
Transformer	0.65	1.69	2.81	19.222	34.002	90.36	15.9	0.4605	0.5174	0.2961	0.6010
ReCoSa	0.66	2.18	3.97	16.471	32.429	92.76	18.3	0.4528	0.5205	0.2965	0.5465
CHMAM	0.82	2.43	4.40	15.288	30.835	97.42	17.7	0.4486	0.5112	0.2901	0.5920
RegDG	**1.15**	**4.45**	**9.22**	**10.983**	**26.846**	**80.61**	**19.1**	0.4820	**0.5477**	**0.3178**	**0.6324**
PersonaChat	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	D^u_{kl} ↓	D^b_{kl} ↓	PPL ↓	BLEU ↑	GRE ↑	AVE ↑	EXT ↑	COH ↑
Seq2Seq+Att	0.11	0.41	0.94	14.488	27.562	136.83	19.2	0.4898	0.5099	0.2599	0.5998
VHRED+BOW	0.25	0.76	1.55	12.772	**26.616**	137.08	20.6	0.4952	0.4827	0.2622	0.5743
NEXUS	0.26	0.89	2.02	11.325	27.134	145.32	**21.3**	0.4940	0.4923	0.2654	0.6015
Transformer	0.28	0.64	1.01	25.076	36.985	90.52	16.0	0.4614	0.4976	0.2449	0.5824
ReCoSa	0.25	0.76	1.26	13.039	28.505	196.09	20.9	**0.4953**	0.4519	0.2321	0.4906
CHMAM	0.42	1.27	2.22	11.558	27.009	159.59	20.5	0.4909	0.4453	0.2569	0.5358
RegDG	**1.11**	**4.39**	**9.43**	**10.352**	26.641	**83.36**	21.1	**0.4953**	**0.5257**	**0.2665**	**0.6163**
OpenSubtitles	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	D^u_{kl} ↓	D^b_{kl} ↓	PPL ↓	BLEU ↑	GRE ↑	AVE ↑	EXT ↑	COH ↑
Seq2Seq+Att	0.09	0.47	1.28	9.125	22.437	88.35	22.1	0.5194	0.5161	0.3156	0.6216
VHRED+BOW	0.10	0.50	1.42	8.990	21.903	113.82	22.5	0.5265	0.5148	0.3072	0.6149
NEXUS	0.09	0.53	1.60	9.106	21.986	97.31	22.8	0.5291	0.5212	0.3096	0.6237
Transformer	0.07	0.33	0.75	11.229	26.480	88.56	21.0	0.5295	0.4952	0.3033	0.5911
ReCoSa	0.07	0.38	0.91	10.188	25.144	**84.26**	22.5	**0.5349**	0.5029	0.3126	0.4746
CHMAM	0.09	0.41	0.98	10.129	24.885	89.33	22.3	0.4792	0.4512	0.2542	0.5612
RegDG	**0.12**	**0.61**	**1.61**	**8.278**	**21.709**	85.68	**22.9**	0.5300	**0.5282**	**0.3175**	**0.6345**

Table 1: The automatic evaluation results at the lowest point of the validation loss. The proposed approach achieves substantial improvements across all the dialogue datasets. "↑" means higher is better. "↓" means lower is better.

3.4 Evaluation Metrics

We conducted both automatic and human evaluation to compare the performance of the models.

Automatic Evaluation The evaluation of open-domain dialogue generation has no well-defined automatic metrics. Thus, we employ two kinds of automatic metrics to evaluate all models. The reference-based metrics, perplexity (**PPL**), **BLEU** (%) (Papineni et al., 2002), and the embedding metrics (including embedding average (**AVE**), embedding greedy (**GRE**), embedding extrema (**EXT**)) (Liu et al., 2016), and coherence (**COH**) (Xu et al., 2018b), are widely adopted to reflect the *grammaticality and semantic relevance* of the responses (Serban et al., 2017; Csaky et al., 2019). The count-based metrics, distinct (**Dist-**{**1,2,3**} (%)) (Li et al., 2016) and KL divergence (Csaky et al., 2019), are used to evaluate the *lexical diversity* and the *distribution distance* of the responses (Xu et al., 2018a; Zhang et al., 2018b). We report the unigram and bigram version of KL divergence, i.e., $\boldsymbol{D^u_{kl}}$ and $\boldsymbol{D^b_{kl}}$. Please refer to Appendix C for the detailed settings of automatic metrics.

Human Evaluation We conducted human evaluation to assess the quality of response. We randomly selected 200 test examples from each dataset and asked three annotators to judge which generated response in each pair (RegDG and baseline) is better (i.e., win, lose or tie) in terms of **Diversity** (how much the generated response contains meaningful information), **Relevance** (how likely the generated response is coherent to both dialogue history and future conversation), and **Fluency** (how likely the generated response is from human).

3.5 Experimental Results

Automatic Evaluation The results obtained at the lowest point of the validation loss are shown in Table 1. Our proposed model significantly outperforms all baselines on all datasets. The LSTM-based baselines obtain better performance than Transformer-based baselines in terms of diversity, distribution distance, and relevance, while they lose in grammaticality. It suggests that the LSTM-based model still has a certain advantage in the loose coupling dialogue task. Compared with CHMAM, ReCoSa achieves higher scores on BLEU and embedding metrics but weaker results on Dist-{1,2,3} and KL divergence, which means that only extracting scenario information from dialogue history cannot provide sufficient semantic constraints to improve model performance across all metrics. Although NEXUS and VHRED+BOW enrich the latent variable and bring more diversity and relevance, they show a distinct decline in PPL. It verifies that our method not only effectively uses the implicit conversation scenario to boost the performance but also indeed transfers this advantage to the inference phase. The improvements of our model on all datasets are significant with $p \leq 0.01$ (t-test). The results of MultiWOZ, reported in Appendix D, show similar improvements.

Datasets	vs. Models	Diversity			κ	Relevance			κ	Fluency			κ
		Win (%)	Lose (%)	Tie (%)		Win (%)	Lose (%)	Tie (%)		Win (%)	Lose (%)	Tie (%)	
DailyDialog	VHRED+BOW	53.25	19.00	27.75	0.484	52.00	22.25	25.75	0.442	39.25	19.50	41.25	0.532
	NEXUS	50.50	18.00	31.50	0.500	47.75	23.75	28.50	0.556	38.00	18.75	43.25	0.523
	ReCoSa	45.75	19.50	34.75	0.568	47.75	25.75	26.50	0.360	29.50	22.75	47.75	0.371
	CHMAM	38.25	23.25	38.50	0.572	43.75	26.00	30.25	0.374	28.50	20.25	51.25	0.451
PersonaChat	VHRED+BOW	46.00	30.50	23.50	0.433	43.25	39.75	17.00	0.440	34.50	26.50	39.00	0.442
	NEXUS	41.25	23.00	35.75	0.677	40.75	34.25	25.00	0.557	30.00	23.50	46.50	0.652
	ReCoSa	44.25	26.75	29.00	0.491	36.50	35.50	28.00	0.510	35.50	18.75	45.75	0.476
	CHMAM	40.25	31.75	28.00	0.374	38.50	35.75	25.75	0.449	33.50	26.75	39.75	0.418
Opensubtitles	VHRED+BOW	47.50	25.75	26.75	0.464	41.25	28.75	30.00	0.430	47.50	19.75	32.75	0.496
	NEXUS	42.50	33.25	24.25	0.529	35.50	34.50	30.00	0.468	45.25	23.50	31.25	0.455
	ReCoSa	31.00	21.25	47.75	0.445	40.00	13.00	47.00	0.372	30.50	11.25	58.25	0.324
	CHMAM	29.25	28.25	42.50	0.448	35.75	27.75	36.50	0.469	28.00	18.00	54.00	0.457

Table 2: The human evaluation results. Our model has higher percentages of Win than the baselines.

Human Evaluation The results are shown in Table 2. We only report the results of VHRED+BOW, NEXUS, ReCoSa, and CHMAM, which are more related to our work. From the results, we can observe that our model performs better than baselines in all datasets. In particular, our model obtains the most significant win-lose difference on diversity, which demonstrates that the implicit conversation scenario induces the response containing more tokens that are meaningful. We calculate the Fleiss's kappa (Fleiss, 1971) to measure the inter-annotator agreement, and the results are mainly distributed in $[0.4, 0.6]$ (i.e., moderate agreement range) with the significance $p \leq 0.01$.

3.6 Experimental Analysis

In this section, we further quantitatively analyze the effectiveness of future conversations and explore why imitation learning works and how to enhance it. For limited space, we select a set of complementary metrics, Dist-$\{1,2,3\}$, PPL, and BLEU, to report the results. The rest of the results is in Appendix F.

3.6.1 Case Study

Table 3 presents some generated responses. The responses generated by baselines are usually dull and meaningless, while the responses generated by our model show diverse and coherent semantic information that indicates distinct relations with those topics in future conversation. The improvements of our model demonstrate the effectiveness of implicit conversation scenarios and our imitation learning framework. Due to limited space, we provide more examples in Appendix E.

3.6.2 Ablation Study

We evaluate the performance of our method without fine-grained prediction imitation (**FPI**) or intermediate representation imitation (**IRI**). The ablation study results, reported in Table 4, show that both types of imitation are beneficial for knowledge transfer. Without IRI, the model converges to a weaker performance than RegDG.

Dialogue history: Well, I am. How much will that cost? // The pass is free. // I don't have to pay for anything?
Seq2Seq+Att: I don't like it. **VHRED+BOW:** Yes, I'm sorry. **NEXUS:** Yes, we have to pay the cash, please. **Transformer:** I don't know what I want to do. **ReCoSa:** I'm going to check out this magazine. **CHMAM:** I want to get a deposit. **RegDG:** You need to pay for the monthly sticker.
Future Conversation: How much is the monthly sticker? // It's $ 24 for each month. // I'll take the student bus pass.
Dialogue history: We're considering of ordering 200 computers, but I'm wondering about the price you'll possibly offer. // Our price will be not less than $ 5000. // Your price is higher than I expected. Could you give us a little discount?
Seq2Seq+Att: Yes, I'm afraid I'm going to get it. **VHRED+BOW:** Yes, we have a credit card. **NEXUS:** Yes, I need to order the price, but I need to pay the goods in the price. **Transformer:** I see. I'll take it. **ReCoSa:** Well, we have to pay a discount for our products. **CHMAM:** Yes, we'll take a 20%. **RegDG:** I'm afraid I can't. We don't have any reduction of quality.
Future Conversation: But the price is always negotiable and you should consider our quantity of order. // Well, what would you suggest? // Could you make it $ 4500.

Table 3: Examples of the generated responses. The responses generated by our model imply the implicit conversation scenario and contain meaningful information.

3.6.3 Effect of Future Conversation

Effect of the Informativeness of Future Conversation We first investigate the situation under which future conversations benefit the generation of high-quality responses. Intuitively, if the true responses or the future conversations are general safe responses, the future conversions contribute little useful information to current dialogue, thereby playing a limited role in the current response prediction. Thus, we classify examples into two sets, i.e., **Uninformative** and **Other**. Specifically, Un-

Models	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	PPL ↓	BLEU ↑
RegDG	1.1	4.4	9.2	80.61	19.1
- FPI	0.7	3.3	7.1	94.44	18.0
- IRI	0.9	4.1	8.7	87.93	18.4

Table 4: Results of the ablation study.

Sets	Exact Match	Word Overlap	Sent. Cluster
Uninformative	×1.035	×1.047	×1.059
Other	×**1.072**∗∗	×**1.078**∗∗	×**1.080**∗

Table 5: The average of improvements across all metrics on the Uninformative set and the Other set. "∗" and "∗∗" indicate $p \leq 0.05$ and $p \leq 0.01$, respectively.

1-1-1	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	PPL ↓	BLEU ↑
Transformer	0.1	0.4	0.6	121.93	16.4
Transformer-IF	**0.2**	**0.6**	**1.2**	**120.43**	**17.7**
Improvement	+0.1	+0.2	+0.6	+1.50	+1.3
3-1-3	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	PPL ↓	BLEU ↑
Transformer	0.6	1.6	2.8	90.36	15.9
Transformer-IF	**0.8**	**1.9**	**3.6**	**87.24**	**20.2**
Improvement	+0.2	+0.3	+0.8	+3.12	+4.3

Table 6: Results on DailyDialog (1-1-1) and (3-1-3).

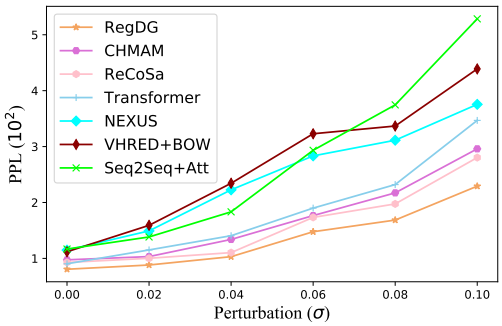

Figure 3: Analysis on model generalization.

Models	Similarity
Seq2Seq+Att	0.553
VHRED+BOW	0.689
NEXUS	0.700
Transformer	0.594
ReCoSa	0.633
CHMAM	0.656
RegDG	**0.732**

Table 7: Effect of regularization reflected by cosine similarity between the generated and real-world word distributions.

informative includes the examples in which the {*dialogue history, response*} or the {*response, future conversation*} is a many-to-one pair. Generally, the second sequence in many-to-one pairs is dull and meaningless (Csaky et al., 2019). To determine the many-to-one pairs, we need to judge whether sentences are of the same meaning and we adopt three measures, that is, whether if the strings match (**Exact Match**), the words overlap more than 80% (**Word Overlap**), or the sentences are in the same embedding cluster (**Sent. Cluster**). For the detailed settings of the above strategies, please refer to Appendix F.

Table 5 shows the results of Transformer-IF on DailyDialog. We can see that the average of all metric improvements of Transformer-IF on the Uninformative set is lower than the Other set, which verifies the assumption that the informativeness of the future conversation supplementing the conversation scenario is crucial to the proposed approach.

Effect of the Capacity of Future Conversation In order to demonstrate the impact of the information content of the implicit conversation scenario on model performance, we conducted the training and testing of both Transformer and Transformer-IF on DailyDailog (1-1-1) and DailyDailog (3-1-3), respectively. "3-1-3" represents that both dialogue history and future conversation consist of three turns, and response only contains one turn. "1-1-1" represents that all sequences in the training examples consist of one turn.

The results are shown in Table 6. Compared with the results on DailyDialog (1-1-1), both models on DailyDialog (3-1-3) achieve overall improvements. The absolute improvements in multi-turn conversation are higher than those in single-turn conversation, which means that Transformer-IF performs better when the implicit conversation scenario contains rich semantic information. Because the automatic metrics may still improve after the lowest point of validation loss (Csaky et al., 2019), the results of both models after 50 epochs of training are reported in Appendix F. It can be observed that Transformer-IF still substantially outperforms Transformer across all metrics under this setting.

3.6.4 Effect of Imitation Learning

Why does the imitation learning work? According to observations in previous work (Chaudhari et al., 2017; Keskar et al., 2017), the model generalization is related to the width of the local minimum achieved by the model. Wider local minima suggests that the model can effectively resist perturbations and obtain better performance on unseen datasets. Therefore, we inject perturbations into the student to judge whether it is guided to a wider local minimum based on the regularization of knowledge transfer. Specifically, we add Gaussian noise with varying magnitude to the parameters of the trained model and observe the perplexity drop on the test set. The results in Figure 3 show that the perplexity of all baselines rapidly increases while the perplexity of our student model grows slowly,

Models	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	PPL ↓	BLEU ↑
RegDG	1.1	4.4	9.2	**80.61**	**19.1**
+ Word-Emb	1.6	6.5	12.6	101.73	18.4
+ Encoder	**1.8**	**7.8**	**17.7**	118.00	17.9

Table 8: Results using the hard transfer. With more hard-transferred modules, the diversity gradually improves, while the relevance gradually weakens.

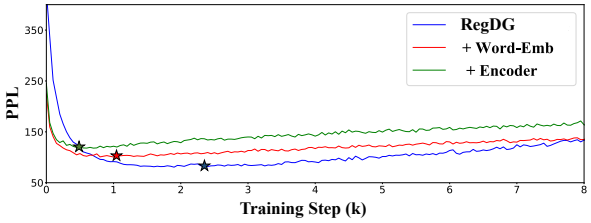

Figure 4: Convergence analysis of the hard transfer. The convergence is faster with more hard-transferred modules.

indicating that the student model reaches a wider local minimum to gain better generalization.

We also analyze the word distributions of the generated responses to intuitively reflect the effect of regularization from imitation learning. Concretely, we use a vector to represent all generated responses, and each element in the vector represents the frequency of a word. Only 2350 most frequent words are considered as Feng et al. (2020). Then, we calculate the distance between the word distributions from each model and the real-world data. From Table 7, it can be seen that our model significantly outperforms plain Transformer and other baselines, which indicates that knowledge transfer effectively regularizes the model so that the model avoids sticking in a relatively centralized word distribution.

Can imitation learning be accelerated? Before the student mimics the teacher, the teacher is usually well pre-trained. According to our observation, this is a redundant workflow that almost doubles the training time. It is worse if we should train a larger model on a huge dataset. In order to accelerate the training process, instead of transferring knowledge via supervised signals to train a student from scratch, we initialize the specified module of the student directly using the parameters of the teacher, and the transferred parameters are kept from updates during the training process. We call this the **hard transfer**. We first apply the hard transfer operation on word embedding (Word-Emb), and further extend it to the encoder. The results of DailyDialog in Table 8 indicate that the performance has been further improved on the diversity with a slight drop on the relevance. Figure 4 shows the variation curve of PPL on the validation set with the training step. The full results are provided in Appendix F. With more hard-transferred modules, the model reaches the lowest point of validation loss faster. It demonstrates that the hard transfer distinctly accelerates the convergence.

Models	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	PPL ↓	BLEU ↑
RegDG	1.1	4.4	9.2	80.61	19.1
+ LM	**1.3**	**5.0**	**9.6**	**79.31**	**19.4**

Table 9: Results of multiple teachers on DailyDialog. With the help of a pre-trained LM, the performance is improved consistently.

Do multiple teachers work? To take advantage of more diverse and richer prior knowledge, we consider extending the teacher from one to many. We pre-train a transformer-based language model as another teacher. The results are shown in Table 9 with full results in Appendix F. It is clear that with the help of the language model, the student further improves on all metrics, except for a weak decline in relevance, because the language model conducts unconditional sequence generation and does not consider the mapping between the dialogue history and the response. We defer the exploration of balancing multiple teachers in future work.

4 Related Work

Diversified Dialogue Generation Recently, various researches have focused on neural dialogue models to generate diverse, informative, and relevant responses. One line of research attempts to extract relevant contexts from redundant dialogue history accurately (Xing et al., 2018; Tao et al., 2018; Zhang et al., 2019a). Another line of research tries to explicitly incorporate a latent variable to inject the variability of response in the decoding process (Serban et al., 2017; Zhao et al., 2017). Shen et al. (2018); Gu et al. (2019); Gao et al. (2019) further enriched the latent variable approach. Also, some works redesigned the objective function or automatically learned it by adversarial learning (Li et al., 2016, 2017a; Xu et al., 2018a; Feng et al., 2020), which improves diversity but brings a fragile training process. Finally, some researchers have adapted external knowledge, such as topic information (Xing et al., 2017), persona (Zhang et al., 2018a), knowledge base (Ghazvininejad et al., 2018). Unlike the above models to pre-

dict responses given a dialogue history, our method combines the future conversation with the dialogue history as the implicit conversation scenario, which contains comprehensive background information to guide the response generation.

Imitation Learning Imitation learning, acquiring skills from observing demonstrations, has proven to be promising in structured prediction, such as alleviating the exposure bias problem (Bengio et al., 2015; Zhang et al., 2019b), transferring knowledge to guide non-autoregressive translation model (Gu et al., 2018; Wei et al., 2019), and automatically learning the reward of the dialogue system (Li et al., 2019b). In our work, the conventional dialogue model as a student mimics the scenario-based dialogue model on both the output layer and intermediate layers.

5 Conclusion

In this work, we introduce the future conversation with the corresponding dialogue history to learn the implicit conversation scenario, which entails latent context knowledge and specifies how people interact in the real world. To incorporate such scenario knowledge without requiring future conversation in inference, we propose an imitation learning framework. The scenario-based teacher model first learns to generate responses with access to both the future conversation and the dialogue history and then a conventional student model is trained to imitate the teacher by hierarchical supervisory signals. As a result, the student is effectively regularized to reach a robust local minimum that represents better generalization performance. Evaluation on four datasets demonstrates the effectiveness and the scalability of our approach, compared to the state-of-the-art baselines. The proposed framework enables the generation of responses that pertain more closely to the scenario indicated by the given dialogue history. Moreover, detailed analyses illustrate how imitating implicit scenarios regularizes the student model. For future work, we will incorporate pre-trained models into our framework (e.g., BERT as a teacher and GPT as a student) to further unlock the performance improvement and explore how to balance diverse prior knowledge from multiple teachers.

Acknowledgment

This research is supported by Beijing Natural Science Foundation (No. L181010 and 4172054), National Key R&D Program of China (No. 2016YFB0801100), and National Basic Research Program of China (No. 2013CB329605). This work is partly supported by Beijing Academy of Artificial Intelligence (BAAI). Xu Sun and Kan Li are the corresponding authors.

References

Lei Jimmy Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. 2016. Layer normalization. *CoRR*, abs/1607.06450.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *ICLR*.

Samy Bengio, Oriol Vinyals, Navdeep Jaitly, and Noam Shazeer. 2015. Scheduled sampling for sequence prediction with recurrent neural networks. In *NIPS*, pages 1171–1179.

Pawel 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*, pages 5016–5026.

Pratik Chaudhari, Anna Choromanska, Stefano Soatto, Yann LeCun, Carlo Baldassi, Christian Borgs, Jennifer T. Chayes, Levent Sagun, and Riccardo Zecchina. 2017. Entropy-SGD: Biasing gradient descent into wide valleys. In *ICLR (Poster)*. OpenReview.net.

Richard Csaky, Patrik Purgai, and Gábor Recski. 2019. Improving neural conversational models with entropy-based data filtering. In *ACL (1)*, pages 5650–5669.

Andrew M. Dai and Quoc V. Le. 2015. Semi-supervised sequence learning. In *NIPS*, pages 3079–3087.

Shaoxiong Feng, Hongshen Chen, Kan Li, and Dawei Yin. 2020. Posterior-GAN: Towards informative and coherent response generation with posterior generative adversarial network. In *AAAI*, pages 7708–7715. AAAI Press.

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

Xiang Gao, Sungjin Lee, Yizhe Zhang, Chris Brockett, Michel Galley, Jianfeng Gao, and Bill Dolan. 2019. Jointly optimizing diversity and relevance in neural response generation. In *NAACL-HLT (1)*, pages 1229–1238.

Marjan Ghazvininejad, Chris Brockett, Ming-Wei Chang, Bill Dolan, Jianfeng Gao, Wen-tau Yih, and Michel Galley. 2018. A knowledge-grounded neural conversation model. In *AAAI*, pages 5110–5117.

Jiatao Gu, James Bradbury, Caiming Xiong, Victor O. K. Li, and Richard Socher. 2018. Non-autoregressive neural machine translation. In *ICLR (Poster)*.

Xiaodong Gu, Kyunghyun Cho, Jung-Woo Ha, and Sunghun Kim. 2019. DialogWAE: Multimodal response generation with conditional Wasserstein auto-encoder. In *ICLR (Poster)*.

Junliang Guo, Xu Tan, Di He, Tao Qin, Linli Xu, and Tie-Yan Liu. 2019. Non-autoregressive neural machine translation with enhanced decoder input. In *AAAI*, pages 3723–3730.

Geoffrey E. Hinton, Oriol Vinyals, and Jeffrey Dean. 2015. Distilling the knowledge in a neural network. *CoRR*, abs/1503.02531.

Jeremy Howard and Sebastian Ruder. 2018. Universal language model fine-tuning for text classification. In *ACL (1)*, pages 328–339.

Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. 2017. On large-batch training for deep learning: Generalization gap and sharp minima. In *ICLR*. OpenReview.net.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *ICLR (Poster)*.

Jason Lee, Elman Mansimov, and Kyunghyun Cho. 2018. Deterministic non-autoregressive neural sequence modeling by iterative refinement. In *EMNLP*, pages 1173–1182.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016. A diversity-promoting objective function for neural conversation models. In *HLT-NAACL*, pages 110–119.

Jiwei Li, Will Monroe, Tianlin Shi, Sébastien Jean, Alan Ritter, and Dan Jurafsky. 2017a. Adversarial learning for neural dialogue generation. In *EMNLP*, pages 2157–2169.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017b. DailyDialog: A manually labelled multi-turn dialogue dataset. In *IJCNLP(1)*, pages 986–995.

Zhuohan Li, Zi Lin, Di He, Fei Tian, Tao Qin, Liwei Wang, and Tie-Yan Liu. 2019a. Hint-based training for non-autoregressive machine translation. *CoRR*, abs/1909.06708.

Ziming Li, Julia Kiseleva, and Maarten de Rijke. 2019b. Dialogue generation: From imitation learning to inverse reinforcement learning. In *AAAI*, pages 6722–6729.

Pierre Lison and Jörg Tiedemann. 2016. Opensubtitles2016: Extracting large parallel corpora from movie and TV subtitles. In *LREC*. European Language Resources Association (ELRA).

Chia-Wei Liu, Ryan Lowe, Iulian Serban, Michael Noseworthy, Laurent Charlin, and Joelle Pineau. 2016. How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. In *EMNLP*, pages 2122–2132.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A method for automatic evaluation of machine translation. In *ACL*, pages 311–318.

Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio. 2015. FitNets: Hints for thin deep nets. In *ICLR (Poster)*.

Iulian Vlad Serban, Alessandro Sordoni, Yoshua Bengio, Aaron C. Courville, and Joelle Pineau. 2016. Building end-to-end dialogue systems using generative hierarchical neural network models. In *AAAI*, pages 3776–3784.

Iulian Vlad Serban, Alessandro Sordoni, Ryan Lowe, Laurent Charlin, Joelle Pineau, Aaron C. Courville, and Yoshua Bengio. 2017. A hierarchical latent variable encoder-decoder model for generating dialogues. In *AAAI*, pages 3295–3301.

Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural responding machine for short-text conversation. In *ACL (1)*, pages 1577–1586.

Xiaoyu Shen, Hui Su, Wenjie Li, and Dietrich Klakow. 2018. Nexus network: Connecting the preceding and the following in dialogue generation. In *EMNLP*, pages 4316–4327.

Siqi Sun, Yu Cheng, Zhe Gan, and Jingjing Liu. 2019. Patient knowledge distillation for BERT model compression. *CoRR*, abs/1908.09355.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *NIPS*, pages 3104–3112.

Chongyang Tao, Shen Gao, Mingyue Shang, Wei Wu, Dongyan Zhao, and Rui Yan. 2018. Get the point of my utterance! Learning towards effective responses with multi-head attention mechanism. In *IJCAI*, pages 4418–4424.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS*, pages 5998–6008.

Oriol Vinyals and Quoc V. Le. 2015. A neural conversational model. *CoRR*, abs/1506.05869.

Bingzhen Wei, Mingxuan Wang, Hao Zhou, Junyang Lin, and Xu Sun. 2019. Imitation learning for non-autoregressive neural machine translation. In *ACL (1)*, pages 1304–1312.

Chen Xing, Wei Wu, Yu Wu, Jie Liu, Yalou Huang, Ming Zhou, and Wei-Ying Ma. 2017. Topic aware neural response generation. In *AAAI*, pages 3351–3357.

Chen Xing, Yu Wu, Wei Wu, Yalou Huang, and Ming Zhou. 2018. Hierarchical recurrent attention network for response generation. In *AAAI*, pages 5610–5617.

Jingjing Xu, Xuancheng Ren, Junyang Lin, and Xu Sun. 2018a. Diversity-promoting GAN: A cross-entropy based generative adversarial network for diversified text generation. In *EMNLP*, pages 3940–3949.

Xinnuo Xu, Ondrej Dusek, Ioannis Konstas, and Verena Rieser. 2018b. Better conversations by modeling, filtering, and optimizing for coherence and diversity. In *EMNLP*, pages 3981–3991.

Hainan Zhang, Yanyan Lan, Liang Pang, Jiafeng Guo, and Xueqi Cheng. 2019a. ReCoSa: Detecting the relevant contexts with self-attention for multi-turn dialogue generation. In *ACL (1)*, pages 3721–3730.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018a. Personalizing dialogue agents: I have a dog, do you have pets too? In *ACL (1)*, pages 2204–2213.

Wen Zhang, Yang Feng, Fandong Meng, Di You, and Qun Liu. 2019b. Bridging the gap between training and inference for neural machine translation. In *ACL (1)*, pages 4334–4343.

Yizhe Zhang, Michel Galley, Jianfeng Gao, Zhe Gan, Xiujun Li, Chris Brockett, and Bill Dolan. 2018b. Generating informative and diverse conversational responses via adversarial information maximization. In *NeurIPS*, pages 1815–1825.

Tiancheng Zhao, Ran Zhao, and Maxine Eskénazi. 2017. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In *ACL (1)*, pages 654–664.

A Datasets

MultiWOZ This dataset is a large-scale multi-turn conversation corpus that contains highly natural conversation across 7 goal-oriented scenarios written by human (Budzianowski et al., 2018). It is split to 58K, 15K, and 5K pairs for training, validation, and testing, respectively.

Two rules are applied to process the four datasets in the experiment:

1. Every seven consecutive dialogue turns form a training example, in which the first three turns, the middle turn, and the last three turns are taken as *dialogue history*, *response*, and *future conversation*, respectively.
2. The lengths of response, dialogue history, and future conversation are limited to $[5, 25]$, $[25, 80]$ and $[25, 80]$, respectively.

B Baselines

Seq2Seq+Att We use a plain Seq2Seq model (Sutskever et al., 2014) with attention mechanism (Bahdanau et al., 2015). The encoder consists of a 2-layer bidirectional LSTM with 256 hidden units. The decoder is based on a 4-layer unidirectional LSTM with 256 hidden units. This baseline is enhanced with multiple techniques from related work and should be considered as a strong baseline.

VHRED+BOW VHRED is proposed by Serban et al. (2017), which introduces conditional variational auto-encoder (CVAE) into the HRED model (Serban et al., 2016) with a continuous latent variable attached to the response. We also adopt the BOW loss (Zhao et al., 2017) as a complementary with KL annealing. The latent variable is 256.

NEXUS NEXUS (Shen et al., 2018) enriches the latent variable with both dialogue history and future conversation through mutual information maximization.

Transformer Transformer (Vaswani et al., 2017) is based solely on the attention mechanism. The number of blocks and heads is 2 and 4, respectively. The hidden size is set to 256. The dimension of the feed-forward layer is 1024.

ReCoSa ReCoSa is proposed by Zhang et al. (2019a), which consists of a word-level LSTM encoder, a self-attention based context-level encoder, and a self-attention based context-response decoder.

CHMAM CHMAM (Tao et al., 2018) applies Multi-Head Attention Mechanism (MHAM) to capture multiple semantic aspects from the dialogue history with a regularizer penalizing the redundancy of attention weight vectors across different aspects of the source sequence.

We adopt residual connection, Layer Normalization (Ba et al., 2016), and Dropout in the LSTM-based baselines, which significantly boost the performance of Seq2Seq+Att, VHRED+BOW, and NEXUS. RegDG and Transformer-IF use the same settings as Transformer. The parameters of Transformer-IF are kept fixed during imitation learning.

MultiWOZ	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	D^u_{kl} ↓	D^b_{kl} ↓	PPL ↓	BLEU ↑	GRE ↑	AVE ↑	EXT ↑	COH ↑
Seq2Seq+Att	0.18	0.73	1.62	4.269	14.332	20.82	23.2	0.5710	0.6557	0.4104	0.6520
VHRED+BOW	0.24	0.92	2.15	3.796	12.831	20.26	**24.4**	0.5687	0.6555	0.3999	0.6546
NEXUS	0.28	1.01	2.16	4.020	13.656	18.64	23.1	0.5556	0.6339	0.4029	0.6547
Transformer	0.21	0.72	1.39	4.998	16.258	16.48	24.0	0.5634	0.6535	0.4085	0.6597
ReCoSa	0.20	0.79	1.61	4.567	14.684	**15.19**	24.2	0.5731	0.6613	0.4044	0.5463
CHMAM	0.26	0.96	1.90	4.274	14.554	18.68	24.0	0.5612	0.6484	0.4151	0.6564
RegDG	**0.27**	**1.08**	**2.37**	**3.756**	**12.744**	16.23	24.2	**0.5738**	**0.6619**	**0.4234**	**0.6601**

Table 10: The automatic evaluation results at the lowest point of the validation loss. "↑" means higher is better. "↓" means lower is better.

C Automatic Evaluation Metrics

Embedding Metrics The word embedding used in embedding metrics is pre-trained on each dataset using fasttext.

Perplexity We calculate the perplexity of each individual sentence and then calculate the average perplexity, which is different from the other way that takes the whole test set (or each batch of data) as one long sentence and then calculates the perplexity (e.g., the official code of Zhao et al. (2017)). The score calculated by the former way is usually larger than the latter one. The difference will increase as the test set (or the size of each batch) gets larger.

KL Divergence It calculates the KL divergence $D_{kl} = {}^1/_{|U_r|} \sum_{w \in U_r} \log_2 {}^{p_r(w)}/_{p_m(w)}$ for each reference response between the reference and the generated response sets, where p_r and p_m are estimated based on the reference response set and the generated response set respectively.

D Automatic Evaluation on MultiWOZ

Table 10 shows the automatic evaluation results on MultiWOZ. Our model beats all baselines on 9 out of 11 metrics.

E Case Study

We provide more generated examples in Table 11.

F Experimental Analysis

Effect of the Informativeness of Future Conversation We use the cosine similarity as the similarity measure and set 0.8 and 0.98 as the thresholds for one-turn sequence and three-turn sequence, respectively. The sentence embedding used in single-pass algorithm is the sum of word embedding with a corresponding weight estimated on the training set. We also adopt the k-means algorithm, which achieves similar clustering proportion.

Effect of the Capacity of Future Conversation Table 12 demonstrates that our model works better when the implicit scenario contains rich information. Table 13 indicates that our model still outperforms Transformer after a fixed epochs (50) of training.

Effect of Imitation Learning Table 14 and Table 15 report the full version of the experimental results in Section 3.6.4.

Dialogue history: Sure. The fitting room is on your right. // It fits well. I like it a lot. What do you think? // You look pretty in red.
Seq2Seq+Att: I don't know. I don't want to do that. **VHRED+BOW:** Oh, I don't know. **NEXUS:** I don't know it. **Transformer:** No, I don't think it's too heavy. **ReCoSa:** I think I'll have a glass of wine. **CHMAM:** It looks like this one. **RegDG:** I'll take it. How much is it?
Future Conversation: $ 29.99. // Okay, I'll take it. Thank you very much for your help. // You're welcome.
Dialogue history: First off, I'm going to need a pound of ground beef. // A pound of ground beef is $ 2.48. // That's perfect.
Seq2Seq+Att: How much do you need to do for the next week? **VHRED+BOW:** How much do you want? **NEXUS:** How much do you like? **Transformer:** How much do you want to get? **ReCoSa:** I'm going to need to get a ticket. **CHMAM:** What kind of ground would you like? **RegDG:** What kind of chicken do you want?
Future Conversation: I'm also going to need three pounds of chicken breasts. // The chicken breasts cost $ 4.05 per pound. // What's the total price for the chicken?
Dialogue history: How much dose it cost for a bus pass? // It'll be $ 65 for a monthly pass. // Is there anything cheaper than that?
Seq2Seq+Att: Yes, I have to take a taxi. **VHRED+BOW:** Well, it's the same thing. **NEXUS:** That's right. How much do you want to pay the bus? **Transformer:** I'm not sure. **ReCoSa:** It's about two years ago. **CHMAM:** We'll get a student pass. **RegDG:** You can take a student pass.
Future Conversation: Well, I am. How much will that cost? // The pass is free. // I don't have to pay for anything?
Dialogue history: I don't want to have a specific title. I just want to enjoy what I am doing. // That sounds very reasonable. // It's the most important thing to me.
Seq2Seq+Att: I don't know that. **VHRED+BOW:** Well, I'll have a good idea. You can have a good idea. **NEXUS:** What do you have to do? **Transformer:** I don't know. But i don't know what to do. **ReCoSa:** I think I should get a part-time job. **CHMAM:** What kind of relationship do you have? **RegDG:** But you know that that's a good idea. I think you should have a good impression on the job.
Future Conversation: That obviously depends on how things go. Whether I'm suited to the firm and firm to me. // Tell me about some of your recent goals and what you do to achieve them. // I want to put my knowledge and experience to use in a challenging position. In order to achieve this goal, I just want work step by step.

Table 11: Examples of the generated responses.

1-1-1	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	D^u_{kl} ↓	D^b_{kl} ↓	PPL ↓	BLEU ↑	GRE ↑	AVE ↑	EXT ↑	COH ↑
Transformer	0.1	0.4	0.6	20.559	34.814	121.93	16.4	**0.4688**	0.4943	0.2889	**0.4943**
Transformer-IF	**0.2**	**0.6**	**1.2**	**17.664**	**32.575**	**120.43**	**17.7**	0.4636	**0.4978**	**0.3000**	0.4883
Improvement	+0.1	+0.2	+0.6	+2.895	+2.239	+1.50	+1.3	-0.0052	+0.0035	+0.0111	-0.0060
3-1-3	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	D^u_{kl} ↓	D^b_{kl} ↓	PPL ↓	BLEU ↑	GRE ↑	AVE ↑	EXT ↑	COH ↑
Transformer	0.6	1.6	2.8	19.222	34.002	90.36	15.9	0.4605	0.5174	0.2961	0.6010
Transformer-IF	**0.8**	**1.9**	**3.6**	**15.858**	**31.455**	**87.24**	**20.2**	**0.4658**	**0.5388**	**0.3063**	**0.6227**
Improvement	+0.2	+0.3	+0.8	+3.364	+2.547	+3.12	+4.3	+0.0053	+0.0214	+0.0102	+0.0217

Table 12: The results on DailyDialog (1-1-1) and DailyDialog (3-1-3).

3-1-3	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	D^u_{kl} ↓	D^b_{kl} ↓	PPL ↓	BLEU ↑	GRE ↑	AVE ↑	EXT ↑	COH ↑
Transformer	6.2	29.2	54.0	3.357	16.374	618.99	26.0	0.5225	0.5954	0.3716	0.6413
Transformer-IF	**6.3**	**31.7**	**60.7**	**2.896**	**14.831**	**597.19**	**31.3**	**0.5539**	**0.6320**	**0.4027**	**0.6574**

Table 13: The results on DailyDialog after 50 epochs of training.

Models	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	D^u_{kl} ↓	D^b_{kl} ↓	PPL ↓	BLEU ↑	GRE ↑	AVE ↑	EXT ↑	COH ↑
RegDG	1.1	4.4	9.2	10.983	26.846	**80.61**	**19.1**	**0.482**	**0.547**	**0.317**	**0.632**
+ Word-Emb	1.6	6.5	12.6	9.506	26.011	101.73	18.4	0.476	0.526	0.306	0.608
+ Encoder	**1.8**	**7.8**	**17.7**	**8.831**	**24.776**	118.00	17.9	0.475	0.541	0.315	0.623

Table 14: The results about the hard transfer operation on DailyDialog.

Models	Dist-1 ↑	Dist-2 ↑	Dist-3 ↑	D^u_{kl} ↓	D^b_{kl} ↓	PPL ↓	BLEU ↑	GRE ↑	AVE ↑	EXT ↑	COH ↑
RegDG	1.1	4.4	9.2	10.983	26.846	80.61	19.1	**0.482**	**0.547**	**0.317**	**0.632**
+ LM	**1.3**	**5.0**	**9.6**	**10.378**	**26.369**	**79.31**	**19.4**	0.475	0.535	0.311	0.616

Table 15: The results about multiple teachers on DailyDialog.