Mitigating Contradictions in Dialogue Based on Contrastive Learning

Weizhao Li[1,2], Junsheng Kong[1,2], Ben Liao[3], Yi Cai[1,2,*]
[1]School of Software Engineering, South China University of Technology
[2]Key Laboratory of Big Data and Intelligent Robot (SCUT), MOE of China
[3]Tencent Quantum Lab
{se_weizhao.li,sescut_kongjunsheng}@mail.scut.edu.cn
bliao@tencent.com
ycai@scut.edu.cn

Abstract

Chatbot models have achieved remarkable progress in recent years but tend to yield contradictory responses. In this paper, we exploit the advantage of contrastive learning technique to mitigate this issue. To endow the model with the ability of discriminating contradictory patterns, we minimize the similarity between the target response and contradiction related negative example. The negative example is generated with learnable latent noise, which receives contradiction related feedback from the pretrained critic. Experimental results show that our method helps to avoid contradictions in response generation while preserving response fluency, outperforming existing methods on both automatic and human evaluation.

1 Introduction

In recent years, with the advent of large training corpora and pretrain technology, chatbot models have evolved considerably in open domain (Bao et al., 2020; Roller et al., 2021). Current chatbots have achieved surprising results in generating fluent, engaging, informative responses, but still occasionally generate responses that are contradictory with history when interacting with human (Li et al., 2021b). Such contradiction issues are often jarring and severely disrupt communication. Therefore, it is essential to reduce contradiction for chat-bots in multi-turns dialogues.

Previous work (Li et al., 2016; Song et al., 2020) proposes to use the paradigm of RL to mitigate the gap between the training and contradiction avoiding objective. However, the RL-based methods are easy to degrade in deep neural network (Parisotto et al., 2020), leading to the decoder generates responses that deviate from human language (Lewis et al., 2017; Kottur et al., 2017). Other method (Li et al., 2020) aims to address dialogue logical contradictions via unlikelihood training (Welleck et al., 2019). While they reduce the probability of the labeled contradicting responses, it is less generalizable to different conversation scenarios with the limited coverage of labeled contradicting data.

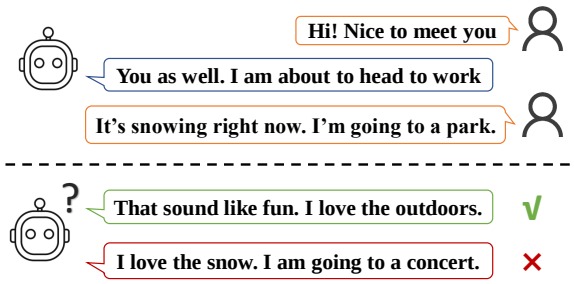

Figure 1: The similarity between correct and contradictory response is 0.9315 in blenderbot embedding space.

We argue that one of the reasons behind contradiction is that model lacks the ability to identify contradictory behavior clearly. As shown in Fig.1, the large pretrained chatbot blenderbot (Roller et al., 2021) still has high similarity between the correct and contradictory responses in embedding space. Chatbots are likely to cause contradictions when probed with unusual conversations during inference (Roller et al., 2021), while they are commonly trained to mimic human context-response pairs under the teacher-forcing algorithm (Williams and Zipser, 1989). Without being exposed to incorrect and contradictory context-response pairs, chatbots fail to learn the ability that discriminating contradictory response patterns directly, which hurts its robustness to avoid contradiction.

To tackle this challenging issue, we propose a novel method to Mitigate Contradiction via Contrastive Learning, namely MCCL. Our method explicitly perceives the difference between the self-contradiction negative example and semantic-aligned positive example. Instead of utilizing well-labeled contradicting examples (Li et al., 2020), we generate a self-contradiction negative example with a learnable latent noise. To capture contradic-

* Corresponding Author.

Findings of the Association for Computational Linguistics: ACL 2022, pages 2781 - 2788

tion actions, we employ the policy gradient method for rewarding the latent noise based on the feedback from a pre-trained critic. Furthermore, we construct an additional positive example by adding a small perturbation. The positive example has aligned semantic with the original context, which devotes to the training stability and robustness.

Overall, our contributions are summarized as follows: 1) To mitigate contradictions in dialogue, we propose a novel method named MCCL, which contrasts target response with negative pairs, to make chatbot models discriminate and refrain from contradictory response patterns. 2) Experiment results show that our method performs better than baselines in automatic metrics and manual evaluation, especially in contradiction score.

2 Related work

2.1 Consistent Conversation

It has been a long-standing goal of artificial intelligence to build an intelligent conversational system that passes the Turing test (Turing, 1950). Researchers improve chatbots intelligence according to dialogue consistency-related information like style (Wang et al., 2017), topic (Dziri et al., 2019) or persona fact (Zhang et al., 2018). Despite showing improvements in guided response generation based on consistency modeling, the issue of contradiction still remains challenging (Nie et al., 2021).

2.2 Contrastive Learning

The concept of contrastive learning has been widely used adopted in many tasks. SimCLR (Chen et al., 2020) shows that contrastive learning can boost the performance of self-supervised and semi-supervised learning in computer vision tasks. In recent years, contrastive learning has been been widely investigated for many NLP tasks, including language modeling (Gao et al., 2021; Li et al., 2021a), text summarization (Liu and Liu, 2021) and machine translation (Pan et al., 2021).

3 Approach

3.1 Encoder-decoder Architecture

Similar to conventional chatbots model (Roller et al., 2021; Bao et al., 2020), our response generation model employs the encoder-decoder architecture. Given the context history C and target response $Y = (y_1, \ldots, y_T)$, the encoder first transforms C into a sequence of hidden representations M. After that, the decoder predicts Y at word level. The decoding process at each time step t can be formalized as follows:

$$\begin{aligned} h_t &= Decoder(M, y_{t-1}) \\ P(y_t|y_{<t}, C) &= softmax(W_d h_t + b_d) \end{aligned} \tag{1}$$

where h_t is the hidden representation of y_t (the t-th word in the response). We maximize the conditional log likelihood for a given N observation $(C^{(i)}, Y^{(i)})_{i=1}^N$ as follows:

$$\mathcal{L}_{MLE} = -\sum_{i=1}^{N}\sum_{t=1}^{T} logP(y_t^{(i)}|y_{<t}^{(i)}, C^{(i)}) \tag{2}$$

3.2 Contrastive Learning Framework

In order to tackle the contradiction problem, we exploit contrastive learning framework to expose various incorrect dialogue pairs. Following (Chen et al., 2020), we can train the model to learn the response representation by contrasting the positive pairs with the negative pairs. A straightforward approach is to treat randomly selected responses from different conversations as semantic negative examples (Sinha et al., 2020). Then we have the base contrastive learning objective as follows:

$$\mathcal{L}_c = -\sum_{i=1}^{N} \log \frac{f(M^{(i)}, H^{(i)})}{\sum_{m \in S} f(m, H^{(i)})} \tag{3}$$

where $S = \{M^{(j)}\}_{j=1}^N$ is a set of context hidden representations randomly sampled from the same batch, $H = [h_1, ..., h_T]$ is the the concatenation of the hidden representations of the target tokens. The function $f(\cdot, \cdot)$ calculates the correlation between context and response as follows:

$$\begin{aligned} c &= Pool(\phi_x(M)) \\ z &= Pool(\phi_y(H)) \\ f(M, H) &= exp(sim(c, z)/\tau) \end{aligned} \tag{4}$$

where ϕ_x and ϕ_y are two fully connected layers with RELU activation and $Pool$ is the average pooling function, sim is the inner product between two vectors, τ is the temperature hyperparameter. Such contrastive learning objective guides chatbot model to learn a more accurate representation of the target response sequence, by identifying which features make the output positive or negative.

3.3 Self-contradiction Negative Example

However, there is no explicit contradiction relationship between the randomly selected non-aligned context and target response. To expose the chatbot model with a contradiction-related negative example, we learn a latent noise ζ based on the input context. Inspired by (Zhao et al., 2019), we decouple the latent noise learning process from response generation. The latent noise ζ is taken as the form of continuous isotropic Gaussian distribution (Serban et al., 2017). We first determine the distribution of latent noise as follows:

$$\begin{aligned} \mu, log(\sigma^2) &= \pi(M) \\ P(\zeta|M) &= N(\mu, \sigma^2) \end{aligned} \tag{5}$$

where π is a feed forward network that projects M into μ and σ. The contradiction negative context representation $\hat{M}$ is formulated as follows:

$$\hat{M} = M + \epsilon\zeta \tag{6}$$

where ϵ is the balanced factor. After that, we sample a negative response $\hat{Y}$ from the decoder successively using the pseudo-Gibbs Markov chain (Ng et al., 2020). To capture the high-level contradiction action for the multi-turns context, we use the policy gradient theorem (Williams, 1992) to train the latent noise generation network, whose gradient can be estimated as follows:

$$\nabla_{\theta_{la}} J(\theta_{la}) = \mathbb{E}[R \cdot logP(\zeta|M, \theta_{la})] \tag{7}$$

where θ_{la} is the parameters in latent noise generation network, R is contradiction probability between C and $\hat{Y}$ measured by the external critic. We apply a pretrained MNLI [1] (Williams et al., 2018) model as critic in practice. With the help of the perturbed negative representation $\hat{M}$, we can augment the contrastive learning loss as follows:

$$\mathcal{L}_{cn} = -\sum_{i=1}^{N} \log \frac{f(M^{(i)}, H^{(i)})}{\sum_{m \in \{S \cup \hat{M}^{(i)}\}} f(m, H^{(i)})} \tag{8}$$

3.4 Semantic-aligned Positive Example

Moreover, we construct an additional positive example to improve the training robustness with a small, approximately worst-case perturbation. Following (Goodfellow et al., 2015), we obtain a perturbation with the linear approximation and generate our positive example $\tilde{M}$ as follows:

$$\begin{aligned} g &= \nabla_M logP(Y|C) \\ \tilde{M} &= M - \eta \frac{g}{||g||^2} \end{aligned} \tag{9}$$

where η is the balanced hyperparameter. We can argument the contrastive learning loss as follows:

$$\mathcal{L}_{cp} = -\sum_{i=1}^{N} \log \frac{f(\tilde{M}^{(i)}, H^{(i)})}{\sum_{m \in \{S \cup \tilde{M}^{(i)} \cup \hat{M}^{(i)}\}} f(m, H^{(i)})} \tag{10}$$

To ensure the positive examples can have aligned semantic, we also minimize the KL divergence between perturbed conditional distribution and the original conditional distribution as follows:

$$\mathcal{L}_{KL} = \sum_{i=1}^{N} KL[P(Y^{(i)}|M)||P(Y^{(i)}|\tilde{M})] \tag{11}$$

3.5 Training Objective

The overall training objective for the response generation model can be formulated as follows:

$$\mathcal{L}_{tot} = \mathcal{L}_{MLE} + \alpha\{\mathcal{L}_{cn} + \mathcal{L}_{cp}\} + \beta\mathcal{L}_{KL} \tag{12}$$

where α and β are balanced hyperparameters. We alternate the optimization of response generation model and the policy update of latent noise generation network (Lewis et al., 2017).

4 Experiment

4.1 Datasets

BST. (Smith et al., 2020) It is a crowdsourced dataset that blends three dialogue skills (engaging personality, empathy, and knowledge). Each conversation is collected with a guided and unguided human speaker. It contains 76k utterances, each with about 16 tokens on average. We use this dataset to finetune the response generation models.

DECODE. (Nie et al., 2021) This dataset offers a new domain for NLI. It contains human-written dialogues, which are labeled as "contradiction" or "non-contradiction". This dataset has 27,184/4,026/4,216 pairs for train/validation/test. To explore the contradiction situation, we only select the context in contradiction pairs from the validation/test sets, namely *DECODE-C*.

[1] https://huggingface.co/roberta-large-mnli

4.2 Implement Details

We use the Blender (Roller et al., 2021) as our backbone chatbot model. We choose the 400M-distill version [2], whose hidden dimension is 1,280. We employ Adam to optimize the model parameters, with the learning rate of 1e-5. For contrastive learning, the temperature τ is set as 0.1, the perturbation factor ϵ is set to 0.4 and η is set to 3. For the hyperparameters in the overall objective, We set α as 0.5 and β as 1. During the inference stage, we use beam search of width 10 to generate the target responses. All the methods are trained in 10 epochs with an NVIDIA Tesla V100.

4.3 Baselines

We compare our method against state-of-the-art baselines: **Blender** (Roller et al., 2021): a pretrained model that maximizes log likelihood. **PersonaCat** (Zhang et al., 2018): a method that prepends all possible persona texts to the input message. **R3F** (Aghajanyan et al., 2021): a method that minimizes the negative log likelihood and symmetric KL-divergence. **CLASP** (Lee et al., 2021): a method that minimizes the similarity between the output sequence and adversarial negative sample, which is generated by adding a small perturbation. **LaRL** (Zhao et al., 2019): a flexible latent variable RL-based method that uses the positive consistent score as reward.

4.4 Evaluation Metrics

The evaluation of logical consistent conversation is mainly about two aspects: contradiction performance and text generation metrics. For contradiction performance, we calculate the contradiction score (C.S) following (Nie et al., 2021). We re-implement the structured utterance-based approach, which finetunes the pretrained RoBERTa (Liu et al., 2019) on DECODE training set, to detect contradictions automatically. Our re-implementation achieves accuracy of 92.33% on test set, which is aligned with the reported accuracy 93.19%. The C.S is calculated as follows:

$$C.S = \frac{\sum_{i=1}^{D} P_i}{D} \tag{13}$$

where D denotes the size of test set, P_i is the label of the i_{th} test case (0: non-contradiction, 1: contradiction). To evaluate the fluency and relevance of responses, we adopt PPL (Adiwardana et al., 2020), BLEU-1/2 (Papineni et al., 2002) and Embedding Greedy metrics (E.grd) (Liu et al., 2016).

Table 1: Automatic evaluation results for compared methods in BST dataset. "B" indicates the BLEU metrics. Bold scores are the best overall.

	C.S(%) ↓	B1 ↑	B2 ↑	PPL ↓	E.grd ↑
Blender	13.81	16.13	5.93	10.96	69.04
Persona	12.69	16.27	6.03	10.99	69.00
CLASP	13.13	16.23	5.88	9.97	69.53
R3F	12.23	16.08	5.88	10.58	69.01
LaRL	11.72	16.37	6.12	10.13	69.37
MCCL	**10.88**	16.42	6.09	9.59	**69.93**
naive	11.70	16.30	6.01	9.86	69.41
+ pos	11.64	**16.51**	**6.21**	**9.45**	69.63
+ neg	11.31	16.29	6.08	9.62	69.70

4.5 Results and Analysis

Table.1 shows that our method outperforms all baselines on BST dataset. We also compare with ablation study about contrastive learning objective. **naive** only maximizes the naive objective from Eq 3; **+ pos/neg** utilizes additional positive or negative examples solely. As we can see, the performance of the **naive** model is not outstanding. When we integrate the **pos** module and the **neg** module, the performance achieves the best.

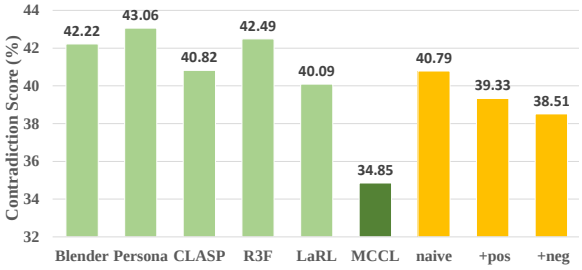

Figure 2: C.S of compared methods on DECODE-C.

Furthermore, MCCL is mainly designed for solving the contradiction problem in dialogue. To verify the effectiveness of the self-contradiction negative component in our method, we take experiment on DECODE-C dataset which is hard for chatbots to generate consistent responses. We only shows the contradiction score since DECODE-C lacks the consistent ground truth while PPL, BLEU and E.grd are reference-based metrics. The results are shown in Fig.2. From this result, we can get some observations. First, our method has a significant advantage (>10%) on contradiction score compared

[2] https://huggingface.co/facebook/blenderbot-400M-distill

Table 2: Manual Evaluation Comparison results.

	Ours Win(%)	Tie(%)	Ours Lose(%)
Blender	25.3	65.3	9.3
CLASP	32.7	54.0	13.3
R3F	25.3	57.3	17.3

with all baselines. Secondly, we find that Persona model has a high contradiction score. This indicates that only adding personal profile information is not enough to resolve dialogue contradiction problem. Lastly, the ablated methods suffer from the ablations on contradiction score which proves that every component is essential for our method.

We further randomly select 50 conversation examples and ask 3 annotators to compare the contradiction performance. As shown in Table 2, our method performs better than other baselines, which is consistent with automatic evaluation results. The kappa score (Fleiss, 1971) is 0.478, showing moderate agreement between the annotators.

5 Conclusion

In this paper, we propose a new method named MCCL to mitigate the contradiction problem in open domain chatbots. Our method minimizes the similarity between the target response and self-contradiction negative example, and maximizes the similarity with semantic-aligned positive example. Experiment results show that our contrastive loss helps to avoid contradiction and obtain better response generation metrics on two different datasets. In the future, we will investigate how to improve the interpretability of negative examples.

Acknowledgement

This work was supported by National Natural Science Foundation of China(62076100), and Fundamental Research Funds for the Central Universities, SCUT(D2210010,D2200150,and D2201300), the Science and Technology Planning Project of Guangdong Province(2020B0101100002), 2020 The industrial technology Basic public service platform project aims at the construction of public service platform in the field of artificial intelligence.

References

Daniel Adiwardana, Minh-Thang Luong, David R So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, et al. 2020. Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977*.

Armen Aghajanyan, Akshat Shrivastava, Anchit Gupta, Naman Goyal, Luke Zettlemoyer, and Sonal Gupta. 2021. Better fine-tuning by reducing representational collapse. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*.

Siqi Bao, Huang He, Fan Wang, Hua Wu, and Haifeng Wang. 2020. Plato: Pre-trained dialogue generation model with discrete latent variable. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 85–96.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR.

Nouha Dziri, Ehsan Kamalloo, Kory Mathewson, and Osmar R Zaiane. 2019. Augmenting neural response generation with context-aware topical attention. In *Proceedings of the First Workshop on NLP for Conversational AI*, pages 18–31.

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

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. Simcse: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821*.

Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. 2015. Explaining and harnessing adversarial examples. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Satwik Kottur, José Moura, Stefan Lee, and Dhruv Batra. 2017. Natural language does not emerge 'naturally'in multi-agent dialog. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2962–2967.

Seanie Lee, Dong Bok Lee, and Sung Ju Hwang. 2021. Contrastive learning with adversarial perturbations for conditional text generation. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*.

Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or no deal? end-to-end learning of negotiation dialogues. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2443–2453.

Jiwei Li, Will Monroe, Alan Ritter, Dan Jurafsky, Michel Galley, and Jianfeng Gao. 2016. Deep reinforcement learning for dialogue generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1192–1202.

Margaret Li, Stephen Roller, Ilia Kulikov, Sean Welleck, Y-Lan Boureau, Kyunghyun Cho, and Jason Weston. 2020. Don't say that! making inconsistent dialogue unlikely with unlikelihood training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4715–4728.

Wei Li, Can Gao, Guocheng Niu, Xinyan Xiao, Hao Liu, Jiachen Liu, Hua Wu, and Haifeng Wang. 2021a. UNIMO: towards unified-modal understanding and generation via cross-modal contrastive learning. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 2592–2607. Association for Computational Linguistics.

Zekang Li, Jinchao Zhang, Zhengcong Fei, Yang Feng, and Jie Zhou. 2021b. Addressing inquiries about history: An efficient and practical framework for evaluating open-domain chatbot consistency. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1057–1067, Online. Association for Computational Linguistics.

Chia-Wei Liu, Ryan Lowe, Iulian Vlad Serban, Mike 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 *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2122–2132.

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 preprint arXiv:1907.11692*.

Yixin Liu and Pengfei Liu. 2021. SimCLS: A simple framework for contrastive learning of abstractive summarization. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 1065–1072, Online. Association for Computational Linguistics.

Nathan Ng, Kyunghyun Cho, and Marzyeh Ghassemi. 2020. Ssmba: Self-supervised manifold based data augmentation for improving out-of-domain robustness. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1268–1283.

Yixin Nie, Mary Williamson, Mohit Bansal, Douwe Kiela, and Jason Weston. 2021. I like fish, especially dolphins: Addressing contradictions in dialogue modeling. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1699–1713, Online. Association for Computational Linguistics.

Xiao Pan, Mingxuan Wang, Liwei Wu, and Lei Li. 2021. Contrastive learning for many-to-many multilingual neural machine translation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 244–258, Online. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318.

Emilio Parisotto, Francis Song, Jack Rae, Razvan Pascanu, Caglar Gulcehre, Siddhant Jayakumar, Max Jaderberg, Raphael Lopez Kaufman, Aidan Clark, Seb Noury, et al. 2020. Stabilizing transformers for reinforcement learning. In *International Conference on Machine Learning*, pages 7487–7498. PMLR.

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. Recipes for building an open-domain chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325, Online. Association for Computational Linguistics.

Iulian Serban, Alessandro Sordoni, Ryan Lowe, Laurent Charlin, Joelle Pineau, Aaron Courville, and Yoshua Bengio. 2017. A hierarchical latent variable encoder-decoder model for generating dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Koustuv Sinha, Prasanna Parthasarathi, Jasmine Wang, Ryan Lowe, William L. Hamilton, and Joelle Pineau. 2020. Learning an unreferenced metric for online dialogue evaluation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2430–2441, Online. Association for Computational Linguistics.

Eric Michael Smith, Mary Williamson, Kurt Shuster, Jason Weston, and Y-Lan Boureau. 2020. Can you put it all together: Evaluating conversational agents' ability to blend skills. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2021–2030, Online. Association for Computational Linguistics.

Haoyu Song, Wei-Nan Zhang, Jingwen Hu, and Ting Liu. 2020. Generating persona consistent dialogues by exploiting natural language inference. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8878–8885.

AM Turing. 1950. Computing machinery and intelligence. *Mind*, 59(236):433–433.

Di Wang, Nebojsa Jojic, Chris Brockett, and Eric Nyberg. 2017. Steering output style and topic in neural

response generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2140–2150.

Sean Welleck, Ilia Kulikov, Stephen Roller, Emily Dinan, Kyunghyun Cho, and Jason Weston. 2019. Neural text generation with unlikelihood training. In *International Conference on Learning Representations*.

Adina Williams, Nikita Nangia, and Samuel Bowman. 2018. A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1112–1122.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256.

Ronald J Williams and David Zipser. 1989. A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1(2):270–280.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2204–2213.

Tiancheng Zhao, Kaige Xie, and Maxine Eskenazi. 2019. Rethinking action spaces for reinforcement learning in end-to-end dialog agents with latent variable models. 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 1208–1218.

A Appendix

A.1 Case Analysis

Case study is shown in Table 3 and Table 4. The gray text indicates the speaker role. The context history consists of the utterances between two different speakers, interleaving with each other. The chatbot models need to generate the next response for the speaker2. Table 3 shows the case 1. The context talks about the dangers of drinking alcohol. The speaker2 is a sober and claims that he don't drink at all. The baseline chatbots fail to avoid contradiction by talking about the last drinking time. Table 4 shows the case 2. From the context, we can know that the speaker2 is afraid of getting bodily injury in sports. The baseline chatbots generate contradictory responses by speaking "not worried about ...", while there are no contradictions in our response.

A.2 Type of Latent Noise

To generate additional self-contradiction negative example, we add a learnable latent noise to the context representation. The latent noise is optimized based on policy gradient method under the feedback from a MNLI (Williams et al., 2018) model. To verify the effectiveness of our method, we sample a random noise as latent noise from Gaussian distribution. This method can get a contradiction score of 39.82% which is far worse than our model (34.85%). This indicates that the contradiction reward is essential for the generation of self-contradiction negative example.

Table 3: Generated Results 1.

Context	[S2] I don't drink alcohol at all, I'm all about sobriety. [S1] Thats a good choice. Alcohol is just never the answer for anything. [S2] I agree with you! Sobriety is such a natural state of a human being given at birth. [S1] Yes, unfortunately many mothers drink alcohol during pregnancy which can cause the baby to have some health issue. [S2] Gosh that's so horrible! my father used to be an alcoholic so i'm pretty traumatized, that's why i all for sobriety and i want everybody i care about to be sober. [S1] Sorry to hear about your trauma. I once drank alcohol and hit my friend in a drunken rage. I have vowed to be sober since then. [S2] Good for you! Stay sober, my friend! because someone who abstains from alcohol but wants to start drinking again isn't truly sober. [S1] I suppose that is one way to see it, when was your last drink?
Blender	about a month ago, and i still want to be a part of it but i just don't want to go through with it.
Persona	about a month ago, but i want to stay sober for as long as possible. i don't want to relapse.
CLASP	The last time I had a beer was in December of this year. I don't drink as much now that I'm sober.
R3F	about a month ago, and i've been sober for almost a year now. i used to drink all the time, but now i just don't drink at all.
LaRL	about a month ago, but i'm trying to stay sober for as long as i can.
MCCL	i don't drink so i can't say when it was, but i know it's bad for me.

Table 4: Generated Results 2.

Context	[S1] I was so happy to watch my friend do a backflip [S2] I've always been too scared to do a backflip. I always worry I'm going to break my neck or something. Did it take him a long time to get it right?And did he get hurt? [S1] I think he just took gymnastics classes. He luckily did not get hurt! [S2] Oh, that makes sense. And I'm glad to hear that. I always worry about that sort of thing. You ever think about trying to do it netx? [S1] no, i never tried to so it netx. Are you worry about your neck or some of your body parts?
Blender	I'm not too worried about my body parts, but I do worry about my balance.
Persona	I'm not really worried about my body parts, but I do worry about my neck. It's something that I worry about a lot.
CLASP	I'm not worried about my body parts, I'm more concerned about my neck.
R3F	I'm not really worried about my body parts, but I do worry about my balance.
LaRL	I'm not worried about my body parts, but I do worry about things like my balance, balance, and coordination.
MCCL	I'm not sure what you mean by body parts. Do you mean the bones, organs, and skin?