# Group-wise Contrastive Learning for Neural Dialogue Generation

**Hengyi Cai**[†,§,*]**, Hongshen Chen**[‡]
**Yonghao Song**[†]**, Zhuoye Ding**[‡]**, Yongjun Bao**[‡]**, Weipeng Yan**[‡]**,** and **Xiaofang Zhao**[†]
[†]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[§]University of Chinese Academy of Sciences, Beijing, China
[‡]JD.com, China
caihengyi@ict.ac.cn, ac@chenhongshen.com

## Abstract

Neural dialogue response generation has gained much popularity in recent years. Maximum Likelihood Estimation (MLE) objective is widely adopted in existing dialogue model learning. However, models trained with MLE objective function are plagued by the low-diversity issue when it comes to the open-domain conversational setting. Inspired by the observation that humans not only learn from the positive signals but also benefit from correcting behaviors of undesirable actions, in this work, we introduce contrastive learning into dialogue generation, where the model explicitly perceives the difference between the well-chosen positive and negative utterances. Specifically, we employ a pretrained baseline model as a reference. During contrastive learning, the target dialogue model is trained to give higher conditional probabilities for the positive samples, and lower conditional probabilities for those negative samples, compared to the reference model. To manage the multi-mapping relations prevailed in human conversation, we augment contrastive dialogue learning with group-wise dual sampling. Extensive experimental results show that the proposed group-wise contrastive learning framework is suited for training a wide range of neural dialogue generation models with very favorable performance over the baseline training approaches.

## 1 Introduction

Open-domain human-machine dialogue systems, especially the generation-based genre, have attracted extensive attention recently. Typically, following the neural encoder-decoder paradigm, contemporary dialogue generation models (Shang et al., 2015; Serban et al., 2016; Xing et al., 2017; Yan, 2018; Huang et al., 2020), more often than not,

are trained with Maximum Likelihood Estimation (MLE) principle to mimic human context-response pairs in the training corpus. While notable gains have been achieved under this learning framework, prior art (Li et al., 2016a, 2017; Zhang et al., 2018a) suggests that naive MLE objective used for training neural dialogue generation models is not that effective enough and tends to result in issues like dull response generation. By optimizing the likelihood of training dialogues, neural models are inclined to assign high probabilities to "safe" responses, due to the fact that vacuous responses like "I don't know" are of relatively high frequencies in conversational datasets (Li et al., 2016a).

One promising training framework for neural dialogue generation is adversarial learning (Goodfellow et al., 2014a; Li et al., 2017), where a discriminator provides rewards for the generator by contrastively distinguishing dialogues as human-generated or machine-generated. However, the learning ability of GANs in text is drastically limited due to training instability and model collapse (Nie et al., 2019; Caccia et al., 2020). *First*, the discriminator is usually unlikely to be fooled very easily, and the generator can hardly learn from those ineffective rewards. *Second*, the generator is sometimes encouraged to mimic the high-frequency generic responses in the training corpus, because in some cases, the discriminator fails to distinguish a good response from a bad one: it can easily recognize contentful but less-grammatical responses as machine-generated, yet treat those human-generated dull responses as the oracle.

In this paper, we introduce contrastive learning (Hadsell et al., 2006; Gutmann and Hyvärinen, 2012) into dialogue generation, where the model explicitly perceives the difference between the well-chosen positive and negative utterances. From the perspective of contrastive learning, the discriminator in adversarial learning considers human-
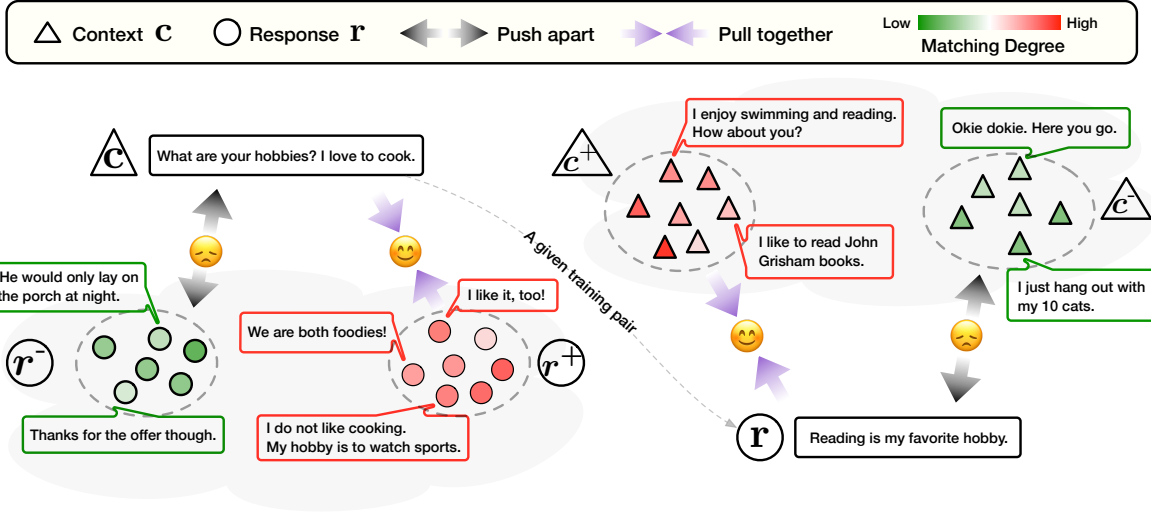
---

Figure 1: An illustration case of group-wise contrastive learning. For a given training instance, the proposed framework explicitly considers the multi-mapping relations in human conversations, by encouraging the dialogue generation model to pull the matched sample pairs together and push the mismatched pairs apart in the latent space.

generated responses as positive utterances and synthetic ones as negative samples. Instead, this work deems highly-matched context-response pairs as positive samples and mismatched training pairs as negative samples. In particular, we utilize a pretrained baseline model as a reference. During contrastive learning, for context $c$ and its response $r$, the target dialogue model is trained to give higher conditional probabilities $p(r|c)$ for the positive samples, and lower conditional probabilities for those negative samples, compared to the reference model. This training paradigm encourages the model to pull the positive data points together and push apart the negative samples, as exemplified in Figure 1. As a result, our proposed training scheme explicitly takes the semantic associations and differences among training examples into account for dialogue modeling. Besides, by contrastively characterizing the distinctions relative to a strong reference, our method implicitly enhances the distinctiveness of the generated responses as well, and ensures that the overall performance of the target model is not inferior to the reference.

Contrastively learning from one pair of positive and negative samples is quite straightforward, however, multi-mapping relations prevail in human-human conversations, where there exist multiple appropriate responses for a given context, and a response sometimes fits well to several contexts, known as one-to-many and many-to-one relations. Such complex multi-mapping relations are overlooked in previous learning framework, which hampers effective dialogue response learning. Furthermore, if a potential highly-matched utterance pair is treated as the negative sample or an outlier is used as the positive sample, the model may be confused. Therefore, in order to consider the multi-mapping phenomenon in human conversations and remedy the potential problematic false learning samples, and enhance the training stability, we augment contrastive learning with group-wise dual sampling, where groups of positive and negative instances are sampled regarding both the context and the response, respectively. To further depict subtle differences between instances in the group, we adapt the instance importance with the matching scores, and optimize the weighted loss.

We show an illustration case to understand our learning framework in Figure 1. Given a training context-response pair $(c, r)$, for context "What are your hobbies? I love to cook", multiple highly-matched responses are organized as the positive samples $r^+$, and the mismatched utterances are deemed as the negatives $r^-$. On the dual direction, regarding the response "Reading is my favorite hobby", multiple sampled context utterances are similarly divided into $c^+$ and $c^-$. Compared with the reference baseline, the target dialogue model is trained to give higher generation probabilities for positive instances, $(c, r^+)$ and $(c^+, r)$, and lower probabilities for negatives $(c, r^-)$ and $(c^-, r)$. By this mean, the target model is actually induced to pull the positive sample pairs together and push the mismatched pairs apart, and thus learns from the

distinctions between the positives and negatives.

The proposed group-wise contrastive learning framework is suited for training various neural dialogue generation models. We conduct extensive studies on three large-scale conversation datasets using four popular dialogue models to assess the proposed approach. The experimental results confirm the effectiveness of our learning framework with very favorable performance over the baseline training approaches[1].

## 2 Contrastive Dialogue Learning

### 2.1 Dialogue Learning by Comparison

Given training data $\mathbb{D}$ containing context-response pairs $\{(c, r)_i\}_{i=1}^{N}$, a dialogue model parameterized by $\theta$ aims to map from the input context $c$ to the output response $r$. To achieve this, conventional dialogue learning approaches search the parameter $\theta$ by maximizing the conditional probability $p_\theta(r|c)$ over the training samples. MLE maximizes the log-likelihood of training pairs while adversarial based approaches rely on the discriminator to distinguish between good responses and bad ones. To combat the aforementioned drawbacks of traditional training approaches in dialogue learning, we advocate the use of contrastive learning to explicitly perceive the difference between the positive and negative samples. Inspired by Gutmann and Hyvärinen (2012); Dai and Lin (2017), we utilize a pretrained baseline model $p_n(\cdot; \phi)$, to provide the target dialogue model $p_m(\cdot; \theta)$ a strong reference when contrasting the positive samples and the negatives. Humans not only learn from the positive signals but also benefit from correcting behaviors of undesirable actions. Intuitively, the target dialogue model is expected to give higher conditional probabilities $p(r|c)$ for the positive samples, and lower conditional probabilities for those negative samples, compared to the reference model. Towards this end, we define the difference between $p_m(r|c, \theta)$ and $p_n(r|c, \phi)$ as:

$$\mathcal{D}((c, r); \theta, \phi) = \log \frac{p_m(r|c, \theta)}{p_n(r|c, \phi)}. \qquad (1)$$

We wish that $\mathcal{D}((c, r); \theta, \phi) > 0$ for any positive pair and vice versa for any negative pair. Concretely speaking, we minimize the following loss

function:

$$\mathcal{L}''(\theta; \mathbb{D}, \phi) =$$
$$- \frac{1}{N} \sum_{(c,r) \in \mathbb{D}} \log \sigma(\mathcal{D}((c, r)^+; \theta, \phi))$$
$$- \frac{1}{N} \sum_{(c,r) \in \mathbb{D}} \log [1 - \sigma(\mathcal{D}((c, r)^-; \theta, \phi))] \qquad , \qquad (2)$$

where $\sigma(\cdot)$ is the sigmoid activation function, the given training pair $(c, r)$ can be used as the positive sample $(c, r)^+$ and the negative sample $(c, r)^-$ can be obtained through negative sampling using the given instance $(c, r)$.

Optimizing the dialogue model with the above objective function is reminiscent of nonlinear logistic regression in Noise-Contrastive Estimation (NCE) (Gutmann and Hyvärinen, 2012). The underlying motivation of our formulation and NCE are essentially different. The reference model in our work is utilized to constrain the behaviors of the target model, rather than serve as a noise distribution to provide noise data. Another difference is that, instead of using the log-ratio between $p_m(\cdot; \theta)$ and $p_n(\cdot; \phi)$ to compute posterior classification probabilities as in NCE, we introduce the function $\mathcal{D}((c, r); \theta, \phi)$ to characterize the distinctions of intrinsic dialogue properties relative to the reference, and encourage the generation of positive samples as well as penalize the negative ones through minimizing the loss in Eq.(2). Besides, by contrastively characterizing the distinctions relative to a strong reference, our method implicitly enhances the distinctiveness of the generated response as well, and ensures that the overall performance of the target model is not inferior to the reference.

### 2.2 Contrastive Dual Sampling

Nevertheless, in the presence of multi-mapping relations in human dialogues, effectively sampling the positive and negative pairs in conversation is not that straightforward and even runs the risk of introducing false learning samples. To manage the complex multi-mapping phenomenon in human conversations and enhance the training stability, we augment the contrastive learning with group-wise dual sampling, where groups of positive and negative instances are sampled regarding both the context and the response, respectively. To put it concretely, for each training instance $(c, r)$, we find a group of positive examples $\{(c, r^+)_i\}_{i=1}^{k}$ with highest matching degree and a group of negative examples $\{(c, r^-)_i\}_{i=1}^{k}$ with lowest matching degree, using an off-the-shelf pretrained matching
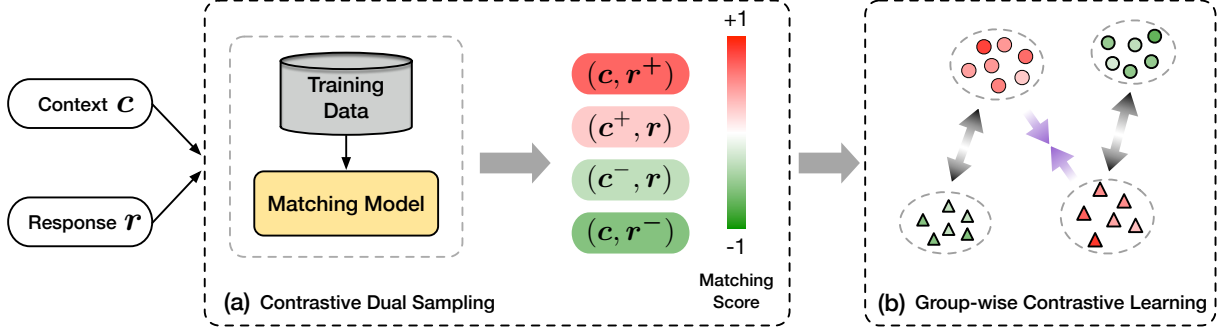
Figure 2: A demonstration of the proposed group-wise contrastive dialogue learning pipeline. For each training pair, it first samples a group of highly-matched examples and another group of most mismatched utterances regarding both the context and response to build the contrastive samples, using an off-the-shelf conversation matching model (§2.2). The target dialogue model is then trained with group-wise contrastive learning (§2.3).

model to compute the matching scores between the given context and candidate responses. Similarly, $\{(c^+, r)_i\}_{i=1}^k$ and $\{(c^-, r)_i\}_{i=1}^k$ are also retrieved from the training set to serve as the context-side contrastive examples, as shown in Figure 2(a). In this work, we adopt MSN (Yuan et al., 2019), a context-response matching network based on multi-hop selection, as the off-the-shelf matching model. Note that other sophisticated matching models can also be applied, e.g., deep attention matching network (Zhou et al., 2018).

## 2.3 Group-wise Contrastive Learning

For each training instance $(c, r)$, as describe in §2.2, we sample $k$ different positive and negative pairs regarding both the dialogue context and its response to manage multi-mapping relations in conversation and stabilize the model training. The resultant well-chosen samples are composed of positive samples, $\{(c, r^+)_i\}_{i=1}^k$ and $\{(c^+, r)_i\}_{i=1}^k$, and the negatives, $\{(c, r^-)_i\}_{i=1}^k$ and $\{(c^-, r)_i\}_{i=1}^k$. Then, the loss function is updated as:

$$
\begin{aligned}
\mathcal{L}'(\boldsymbol{\theta}; \mathbb{D}, \boldsymbol{\phi}) = \\
-\frac{1}{N} \sum_{(c,r) \in \mathbb{D}} \frac{1}{2k+1} \sum_{i=1}^{2k+1} \log \sigma(\mathcal{D}((c, r)_i^+; \boldsymbol{\theta}, \boldsymbol{\phi})) \\
-\frac{1}{N} \sum_{(c,r) \in \mathbb{D}} \frac{1}{2k} \sum_{i=1}^{2k} \log[1 - \sigma(\mathcal{D}((c, r)_i^-; \boldsymbol{\theta}, \boldsymbol{\phi}))]
\end{aligned} \quad (3)
$$

Given varied matching degrees of the collected context-response pairs in open-domain dialogue, indiscriminately training on such data impedes the model to perceive intra-group differences of these samples. We thus utilize the matching score $s$ attached with each sample to adapt its instance effect on the group-wise contrastive dialogue learning. Specifically, for a given training example $(c, r)$,

the matching score $s^+$ of its positive pair lies in $(0, 1]$ and the negative score $s^-$ lies in $[-1, 0]$. To induce the model learning from sample pairs with varied matching degrees discriminately, the loss function is finally defined to be:

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}; \mathbb{D}, \boldsymbol{\phi}) = \\
-\frac{1}{N} \sum_{(c,r) \in \mathbb{D}} \frac{1}{2k+1} \sum_{i=1}^{2k+1} \log[s_i^+ \cdot \sigma(\mathcal{D}((c, r)_i^+; \boldsymbol{\theta}, \boldsymbol{\phi}))] \\
-\frac{1}{N} \sum_{(c,r) \in \mathbb{D}} \frac{1}{2k} \sum_{i=1}^{2k} \log[1 + s_i^- \cdot \sigma(\mathcal{D}((c, r)_i^-; \boldsymbol{\theta}, \boldsymbol{\phi}))]
\end{aligned} \quad (4)
$$

The loss function $\mathcal{L}(\boldsymbol{\theta})$ reaches its lower bound when the positive and negative pairs can be perfectly distinguished, i.e., $p_m(\boldsymbol{r}|\boldsymbol{c}, \boldsymbol{\theta}) \gg p_n(\boldsymbol{r}|\boldsymbol{c}, \boldsymbol{\phi})$ for the positive samples and $p_m(\boldsymbol{r}|\boldsymbol{c}, \boldsymbol{\theta}) \ll p_n(\boldsymbol{r}|\boldsymbol{c}, \boldsymbol{\phi})$ for the negatives, which indicates that the target dialogue model is able to clearly contrast a group of positive candidates from the negative ones and generate highly-distinctive responses for the given contexts.

## 2.4 Discussion

Neural sequence-to-sequence models trained with the MLE objective function are plagued by the low-diversity issue when it comes to the open-domain conversational setting, in which bland and generic utterances usually dominate the data distribution. Since the objective of MLE is to maximize only the probabilities of ground-truth context-response pairs, it fails to capture the multi-mapping nature of human dialogues, not to mention the semantic differences among various candidates for a given example. While the proposed group-wise contrastive learning framework explicitly explores multiple variants of a given dialogue example by leveraging an off-the-shelf matching model, and implicitly

|  | PersonaChat | Douban | OpenSubtitles |
|---|---|---|---|
| #context-response pairs | 140,248 | 218,039 | 353,046 |
| Avg. #turns per context | 2.69 | 3.94 | 3.79 |
| Avg. #words per utterance | 11.96 | 15.28 | 6.85 |
| Training Pairs | 113,558 | 198,039 | 293,129 |
| Validation Pairs | 13,602 | 10,000 | 29,960 |
| Test Pairs | 13,088 | 10,000 | 29,957 |
| #Tokens | 18,029 | 40,000 | 40,000 |

Table 1: Data statistics for PersonaChat, Douban and OpenSubtitles datasets.

guarantees the ground-truth generation probabilities through the contrastive constraints in Eq.(4).

Adversarial learning approaches and our proposed framework both involve an auxiliary model during the training process. However, GANs are learned via a competition between the target generator and the counteracting discriminator, which needs careful tuning to prevent model collapse in text modeling (Caccia et al., 2020), whereas in our framework, the auxiliary reference system models conversation data in the same direction with the target dialogue model, and is stable during the learning procedure.

## 3 Experiments

### 3.1 Experiment Settings

**Datasets** We perform experiments on three conversation datasets: PersonaChat (Zhang et al., 2018b), Douban Corpus (Wu et al., 2017) and OpenSubtitles (Lison and Tiedemann, 2016). **PersonaChat**, an English-language dataset, contains multi-turn dialogues between pairs of speakers, collected via Amazon Mechanical Turk. **Douban** consists of daily conversations from a popular social networking service—Douban group[2] in China. **OpenSubtitles** contains human-human conversations converted from movie transcripts in English. Data statistics are listed in Table 1.

**Experimental Models** We apply the proposed group-wise contrastive learning framework to several state-of-the-art models, including (i) **SEQ2SEQ**: a LSTM-based sequence-to-sequence model with attention mechanisms (Bahdanau et al., 2015), (ii) **HRED**: a hierarchical recurrent neural dialogue generation model (Serban et al., 2016), (iii) **TRANSFORMER**: an encoder-decoder architecture relying solely on attention mechanisms (Vaswani et al., 2017), (iv) **HRAN**: a hierarchical recurrent attention network for multi-turn

response generation (Xing et al., 2018). Each model is trained using two protocols: the vanilla MLE training procedure and the proposed group-wise contrastive learning procedure, keeping other configurations the same.

**Baselines** We compare our group-wise contrastive learning framework against the following dialogue learning approaches: (i) **ADVERSARIAL**: an adversarial training approach for response generation (Li et al., 2017), (ii) **MMI**: a training objective which maximums the mutual information between the dialogue context and its response (Li et al., 2016a; Zhang et al., 2018c), (iii) **DEEPRL**: a reinforcement learning framework for neural response generation with heuristic reward functions to boost response qualities (Li et al., 2016b), (iv) **CVAE**: a conditional variational auto-encoder learning framework to maximize the data likelihood, augmented with the KL-annealing technique (Bowman et al., 2016) and a BOW loss (Zhao et al., 2017), and (v) **DIALOGWAE**: a conditional Wasserstein auto-encoder framework, modeling the distribution of dialogues by training a GAN within the latent variable space (Gu et al., 2019).

**Automatic Evaluation Metrics** We adopt several standard metrics widely used in existing works to measure the performance of dialogue generation models, including BLEU (Papineni et al., 2002), embedding-based metrics (Average, Extrema, Greedy and Coherence)[3] (Liu et al., 2016; Xu et al., 2018; Sedoc et al., 2019), entropy-based metrics[4] (Ent-{1,2}) (Serban et al., 2017) and distinct metrics (Dist-{1,2,3}) (Li et al., 2016a).

### 3.2 Implementation and Reproducibility

We implement our model in ParlAI (Miller et al., 2017) and train them on Nvidia P40 GPUs. All the models use pretrained word embeddings produced by fastText (Bojanowski et al., 2017), and the dimensionality of word vectors is 300. For experimental models, 2-layer LSTM-based encoder and decoder with hidden size 256 are used in SEQ2SEQ. We use the base Transformer configuration described in Vaswani et al. (2017), i.e., 6 blocks with 8 attention heads and 512 hidden units. 2-layer GRU-based RNNs are employed to build the word-level encoder, utterance-level encoder and

---

[2]https://www.douban.com/group

[3]https://chateval.org/

[4]We compute the entropy value for the empirical distribution of n-grams based on the maximum likelihood estimation on the training data.

| | Models | BLEU-1 / 2 / 3 / 4 | Dist-1 | Dist-2 | Dist-3 | Avg. | Ext. | Gre. | Coh. | Ent-1 / 2 |
|---|---|---|---|---|---|---|---|---|---|---|
| (a) | SEQ2SEQ | 12.040 / 3.9950 / 0.8815 / 0.2312 | 0.4309 | 2.045 | 4.303 | 36.33 | 28.66 | 63.64 | 41.66 | 6.891 / 10.81 |
| | SEQ2SEQ (►) | **13.660 / 4.9160 / 1.5970 / 0.6122** | **0.8492** | **5.093** | **12.000** | **39.76** | **31.74** | **64.76** | **49.39** | **7.016 / 10.90** |
| | HRED | 12.410 / 3.8360 / 0.8455 / 0.2364 | 0.4744 | 2.546 | 6.127 | 36.52 | 28.37 | 64.12 | 39.08 | 6.792 / 10.66 |
| | HRED (►) | **13.180 / 4.3220 / 1.0360 / 0.3274** | **0.7130** | **4.468** | **11.220** | **38.54** | **29.65** | **64.26** | **44.07** | **6.931 / 10.84** |
| | TRANSFORMER | 11.460 / 3.2080 / 0.5389 / 0.1476 | 0.4813 | 2.544 | 6.146 | 35.72 | 27.38 | 63.61 | 38.02 | 6.804 / 10.55 |
| | TRANSFORMER (►) | **12.660 / 3.8920 / 0.8406 / 0.2577** | **0.7859** | **4.562** | **10.950** | **37.42** | **28.95** | **64.02** | **41.96** | **6.918 / 10.80** |
| | HRAN | 12.190 / 3.8290 / 0.7752 / 0.2171 | 0.5074 | 2.883 | 7.104 | 36.53 | 28.08 | 63.58 | 40.22 | **6.964** / 10.83 |
| | HRAN (►) | **13.430 / 4.5030 / 1.0630 / 0.3513** | 0.7713 | 4.974 | 12.380 | 39.04 | 30.08 | 64.48 | 46.63 | 6.942 / **10.87** |
| (b) | SEQ2SEQ | 5.585 / 0.7887 / 0.1008 / 0.0296 | 1.1610 | 6.105 | 13.100 | 46.75 | 36.80 | **53.52** | 52.13 | 7.225 / 11.13 |
| | SEQ2SEQ (►) | **5.821 / 0.7910 / 0.1053 / 0.0377** | **1.3010** | **7.935** | **18.070** | **46.96** | **36.99** | 53.41 | **53.40** | **7.464 / 11.66** |
| | HRED | **5.899** / 0.7925 / 0.0786 / 0.0206 | 0.8334 | 5.147 | 14.160 | 48.12 | 36.50 | **54.20** | 49.99 | 7.107 / 10.90 |
| | HRED (►) | 5.778 / **0.7968 / 0.0996 / 0.0387** | **1.2910** | **7.461** | **19.450** | **48.23** | **36.51** | 53.34 | **50.31** | **7.436 / 11.10** |
| | TRANSFORMER | 5.229 / 0.6443 / 0.0764 / 0.0240 | 1.1140 | 5.658 | 13.830 | 45.45 | 35.45 | 53.04 | 48.04 | 7.084 / 11.15 |
| | TRANSFORMER (►) | **5.386 / 0.6460 / 0.0889 / 0.0274** | **1.3280** | **6.723** | **15.800** | **45.96** | **36.11** | **53.33** | **48.92** | **7.238 / 11.16** |
| | HRAN | 5.366 / 0.7229 / **0.0860** / 0.0182 | 1.0960 | 6.679 | 17.250 | 47.44 | 36.35 | **53.93** | 50.25 | 7.202 / **11.15** |
| | HRAN (►) | **5.541 / 0.7424** / 0.0723 / **0.0194** | **1.6630** | **10.030** | **24.240** | **48.01** | **36.99** | 53.46 | **51.81** | **7.394** / 10.94 |
| (c) | SEQ2SEQ | 5.666 / 1.0870 / **0.2471** / 0.0416 | 0.2880 | 2.110 | 5.566 | 54.22 | 46.11 | 63.96 | 56.82 | 6.685 / 10.54 |
| | SEQ2SEQ (►) | **5.696 / 1.1290** / 0.2199 / **0.0476** | **0.4495** | **3.681** | **10.860** | **54.32** | **47.13** | **64.54** | **58.60** | **6.792 / 10.80** |
| | HRED | 5.489 / 0.9953 / 0.2206 / 0.0711 | 0.3020 | 2.179 | 6.355 | **54.61** | 54.36 | 67.91 | 56.45 | 6.699 / 10.74 |
| | HRED (►) | **5.670 / 1.0930 / 0.2461 / 0.0828** | **0.4490** | **3.099** | **8.949** | 54.19 | 54.36 | **68.16** | **57.26** | **6.722 / 10.80** |
| | TRANSFORMER | 4.619 / **0.8294** / 0.1500 / 0.0307 | 0.3470 | 2.038 | 5.028 | 52.29 | 44.21 | 63.16 | 53.40 | 6.677 / 10.40 |
| | TRANSFORMER (►) | **4.712** / 0.8197 / **0.1744 / 0.0314** | **0.3897** | **2.437** | **6.188** | **52.34** | **45.12** | **63.52** | **54.11** | **6.722 / 10.50** |
| | HRAN | 5.090 / 0.8424 / 0.1665 / 0.0405 | 0.3205 | 2.604 | 8.188 | **54.74** | 54.52 | 68.16 | 56.58 | 6.556 / 10.53 |
| | HRAN (►) | **5.423 / 0.9192 / 0.1913 / 0.0529** | **0.5034** | **3.935** | **11.920** | 54.40 | **54.54** | **68.30** | **57.48** | **6.699 / 10.89** |

Table 2: Automatic evaluation results (%) on the test set of three datasets: (a) PersonaChat, (b) Douban and (c) OpenSubtitles. "►" denotes that the model is trained using our proposed framework. The metrics Average, Extrema, Greedy and Coherence are abbreviated as Avg., Ext., Gre. and Coh., respectively. The best results in each group are highlighted with **bold**.

response decoder for both the HRED and HRAN. The GRU hidden size is set to 256. For both models using different training procedures, we pretrain them by MLE and the result checkpoint is adopted as the reference model used in our framework. We employ BM25 (Robertson and Zaragoza, 2009) to construct the index used during the contrastive dual sampling procedure. The group size $k$ is set to 3. Regarding comparison models, we adopt the default configurations used in the original papers. We optimize models by Adam (Kingma and Ba, 2015) with an initial learning rate of 0.001 and the batch size of 128. All systems are trained until the validation loss fails to decrease for 5 checkpoints. We compute the loss on the validation set at every 0.5 epochs and save the parameters for the top model. Evaluation scores on the test set from the saved model are finally reported.

### 3.3 Evaluation Results

**Performance on Experimental Models**  We instantiate the proposed framework on several state-of-the-art models for dialogue generation. Table 2 reports the automatic evaluation results of our learning framework and the conventional MLE training procedure. By training dialogue models using the proposed learning framework, we witness solid performance boosts on three conversation datasets in terms of almost all the evaluation metrics, compared to the vanilla training. Such improvements are also consistent across various experimental architectures, affirming the general applicability and superiority of the proposed framework.

**Comparison with Baseline Approaches**  We compare our proposed framework with existing learning approaches designed for dialogue generation task. Table 3 summarizes the evaluation results. We observe that our learning framework outperforms previous approaches regarding the majority of evaluation metrics. It is worth noting that the proposed framework brings a relatively large improvement regarding both the response diversity and conversation coherence, indicating that our approach helps the dialogue model to generate not only informative but also context-relevant responses, which confirms our hypothesis that the group-wise contrastive learning encourages distinctiveness.

**Human Evaluation**  We further conduct human evaluations to assess the proposed learning framework. We choose the PersonaChat as our evaluation

| Learning Approaches | BLEU-1 / 2 / 3 / 4 | Dist-1 | Dist-2 | Dist-3 | Avg. | Ext. | Gre. | Coh. | Ent-1 / 2 |
|---|---|---|---|---|---|---|---|---|---|
| ADVERSARIAL | 12.190 / 4.0060 / 0.8950 / 0.2644 | 0.6269 | 3.357 | 7.374 | 35.93 | 29.00 | 63.65 | 42.38 | 6.980 / 10.88 |
| MMI | **14.030** / 4.6460 / 1.3340 / 0.5022 | 0.4734 | 2.443 | 5.515 | 39.34 | 30.92 | **64.84** | 45.16 | 6.874 / 10.65 |
| DEEPRL | 12.660 / 4.0150 / 1.0140 / 0.3314 | 0.6838 | 3.838 | 8.581 | 37.23 | 29.68 | 64.30 | 44.13 | 6.885 / 10.85 |
| CVAE | 11.570 / 2.8100 / 0.6357 / 0.1714 | 0.2876 | 2.326 | 7.506 | 39.29 | 30.61 | 63.67 | 41.76 | 6.869 / 10.82 |
| DIALOGWAE | 11.430 / 2.9260 / 0.5676 / 0.1436 | **0.9936** | 5.080 | 9.928 | 38.68 | 28.70 | 63.39 | 41.06 | 7.009 / **11.09** |
| **Ours** | 13.660 / **4.9160 / 1.5970 / 0.6122** | 0.8492 | **5.093** | **12.000** | **39.76** | **31.74** | 64.76 | **49.39** | **7.016** / 10.90 |

Table 3: Performance (%) of our approach instantiated on naive SEQ2SEQ and baseline approaches on PersonaChat.

| Opponent | Win | Loss | Tie | Kappa |
|---|---|---|---|---|
| Ours *vs.* VANILLA MLE | 53% | 10% | 37% | 0.5750 |
| Ours *vs.* ADVERSARIAL | 47% | 15% | 38% | 0.5495 |
| Ours *vs.* MMI | 43% | 12% | 45% | 0.5863 |
| Ours *vs.* DEEPRL | 40% | 22% | 38% | 0.6036 |
| Ours *vs.* CVAE | 40% | 15% | 45% | 0.5510 |
| Ours *vs.* DIALOGWAE | 45% | 18% | 37% | 0.4216 |

Table 4: The results of human evaluation on the test set of PersonaChat.

corpus since its expressions are more close to the style of daily chat and are easier for the annotators to make judgments. Three crowd-sourced graduate students are employed to evaluate the quality of generated responses for 100 randomly sampled input contexts. During the evaluation, the annotators are requested to select a preferred response, or vote a tie, considering the following aspects of response quality: fluency, informativeness, coherence and engagingness. Table 4 summarizes the evaluation results and the Cohen's kappa scores (Cohen, 1960) to measure the intra-rater reliability. We observe that our learning framework brings more preferable replies compared with the competitors. This indicates that training the dialogue model with the proposed group-wise contrastive learning framework does improve the response quality.

### 3.4 Model Analysis

**Effect of the Group-wise Learning Strategy**
To manage the multi-mapping relations in human-human conversation and stabilize the model training with noisy data, the dialogue model is induced to contrast a group of positive samples from the negative ones, pulling the matched sample pairs together and pushing the mismatched pairs apart. We ablate the group-wise learning from the framework by using only one pair of positive and negative samples to verify its effectiveness. As presented in Table 5 (a), we can see that disabling group-wise learning hurts performance significantly on all evaluation metrics. Note that ablating either

the group-wise positive sampling (Table 5 (b)) or group-wise negative sampling (Table 5 (c)) also leads to a performance drop with respect to the evaluation metrics. It demonstrates that the group-wise learning strategy plays a key role in achieving strong performance.

**Effect of the Dual Sampling** In our framework, the contrastive samples can be organized regarding either the dialogue context or response, allowing the dialogue model to explore both the many-to-one and one-to-many relations in human conversation. We investigate different sampling strategies in Table 5 (d) and (e). We notice that when both the response-side and context-side samplings together incorporated into the learning framework, the model achieves its best performance, verifying the effectiveness of the contrastive dual sampling.

**Impact of Matching Scores** To discriminatively exploit the sampled context-response pairs with varied matching degrees, we utilize the matching score attached with each sample to adapt its instance effect on the model training. We conduct the ablation test of this learning strategy by simply discarding the impact of matching scores as in Eq.(3). As shown in Table 5 (f), training without considering matching degrees of samples leads to a consistent performance drop, which suggests that the system can benefit from perceiving fine-grained differences within the group during contrastive learning.

**Impact of Group Size** We explore the impact of using different group size $k$ in our group-wise contrastive learning framework in Figure 3. We observe that increasing the group size $k$ leads to continuous improvement on the Distinct metric while other reference-based metrics achieve the best results at a moderate group size. We conjecture that a larger group size allows the dialogue model to learn from more diverse expressions, meanwhile it also risks introducing more utterances that are

| Framework variants | BLEU-1 / 2 / 3 / 4 | Dist-1 | Dist-2 | Dist-3 | Avg. | Ext. | Gre. | Coh. | Ent-1 / 2 |
|---|---|---|---|---|---|---|---|---|---|
| (a) *w/o* group-wise sampling | 12.870 / 4.102 / 0.9564 / 0.2308 | 0.3965 | 2.070 | 4.633 | 36.52 | 29.09 | 64.21 | 42.40 | 6.836 / 10.62 |
| (b) *w/o* group-wise positive sampling | 13.120 / 4.800 / 1.4180 / 0.5967 | 0.4632 | 2.270 | 5.002 | 38.26 | 31.18 | 64.66 | 43.03 | 6.812 / 10.49 |
| (c) *w/o* group-wise negative sampling | 13.210 / 4.698 / 1.3970 / 0.5587 | 0.7175 | 3.532 | 7.473 | 38.23 | 30.96 | 64.62 | 46.27 | 6.911 / 10.68 |
| (d) *w/o* response-side sampling | 13.340 / 4.730 / 1.4820 / 0.5779 | 0.8487 | 4.964 | 11.340 | 39.31 | 31.51 | 64.66 | 48.35 | 6.938 / 10.75 |
| (e) *w/o* context-side sampling | 13.170 / 4.539 / 1.4160 / 0.5308 | 0.8455 | 4.892 | 11.210 | 39.57 | **31.81** | 64.56 | 47.19 | 6.904 / 10.66 |
| (f) *w/o* impact of matching scores | 13.560 / 4.359 / 1.1140 / 0.3823 | 0.6086 | 3.809 | 9.037 | 38.78 | 30.35 | 64.44 | 46.88 | 6.952 / **10.90** |
| **Full version** | **13.660 / 4.916 / 1.5970 / 0.6122** | **0.8492** | **5.093** | **12.000** | **39.76** | 31.74 | **64.76** | **49.39** | **7.016 / 10.90** |

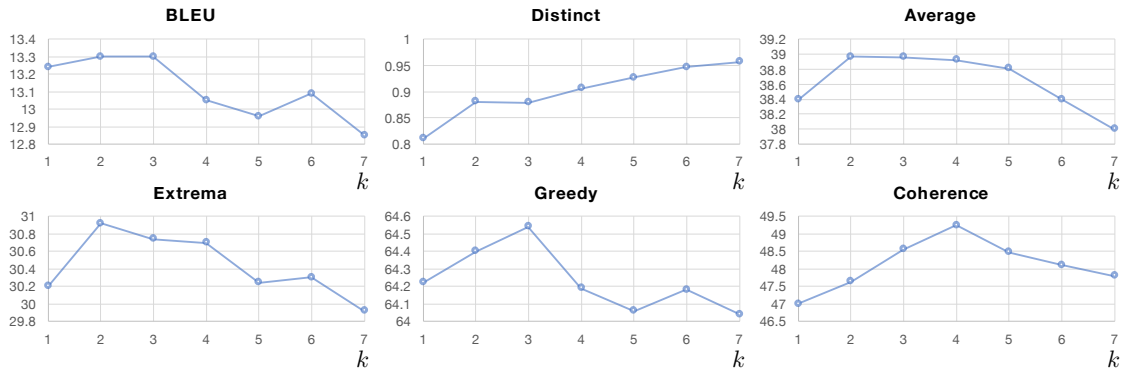Table 5: Ablation test (%) using SEQ2SEQ with different framework variants on PersonaChat.



Figure 3: Evaluation results (%) with different group size $k$ on the validation set of PersonaChat using the proposed framework instantiated on SEQ2SEQ. BLEU-1 and Dist-1 are denoted as "BLEU" and "Distinct", respectively.

inconsistent with the references.

## 4 Related Work

**Learning Methods for Dialogue Generation**
Typically, state-of-the-art neural dialogue generation models adopt *Maximum Likelihood Estimation (MLE)* as their learning approach, maximizing log-likelihood of model parameters on the training data. Though effective, well-known issues, including the notorious general dull response problem, are reported by prior art (Li et al., 2016a; Zhao et al., 2017) on dialogue models trained with MLE.

Alternative dialogue learning approaches are proposed to tackle the issues. Li et al. (2016a); Zhang et al. (2018c) introduce the Maximum Mutual Information as the objective function to promote response diversity. Techniques of reinforcement learning (RL) and adversarial learning have been introduced into dialogue generation by Li et al. (2016b, 2017) to better approximate the optimization goal of dialogue models. Conditional variational framework (Zhao et al., 2017; Shen et al., 2017; Park et al., 2018) has also shown a promise in dialogue generation. Gu et al. (2019) further introduce a conditional Wasserstein autoencoder that employs GAN (Goodfellow et al., 2014b) to model the multimodal latent structures. Cai et al. (2020) design a multi-curriculum learning frame-

work to facilitate the dialogue model training. Contrasted with existing learning methods for dialogue generation, the proposed framework in this work encourages the model to learn from the difference between well-chosen contrastive pairs, which explicitly models the multi-mapping relations in conversation and promotes the distinctiveness of the generated responses in the meantime.

**Contrastive Learning** The concept of learning by contrasting positive pairs against negative pairs (Hadsell et al., 2006; Gutmann and Hyvärinen, 2012) has been successfully adopted in many tasks. For example, contrastive learning in language modeling task (Mnih and Teh, 2012; Vaswani et al., 2013; Baltescu and Blunsom, 2015) aims to approximate the negative log-likelihood by training the model to correctly classify between generated noise samples and words observed in the training data. Contrastive visual representation learning (van den Oord et al., 2018; Chen et al., 2020) trains a generative model to score real data points higher than negative samples. Dai and Lin (2017) propose to use contrastive learning for image caption. Clark et al. (2020) use contrastive learning to train a discriminative model for language representation learning. Compared with existing work, samples used in this paper, instead of being sampled randomly, are carefully chosen to ex-

hibit particular properties of human dialogues. Another difference is that, we manage multi-mapping relations prevailed in human conversation using many positives and many negatives, which captures both the intra-group and inter-group variability.

## 5 Conclusion

In this work, we propose a group-wise contrastive dialogue learning approach, that explicitly perceives the difference between the well-chosen positive and negative utterances, and manages the multi-mapping relations in human conversations simultaneously. Given a training instance, the proposed learning framework first organizes a group of positive samples and negative samples regarding context-response matching degrees, and then trains the target dialogue model to give higher conditional probabilities for positive pairs and lower probabilities for the negatives. Extensive experimental results show that the proposed learning framework brings a solid favorable performance boost amongst various strong baseline approaches.

## References

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

Paul Baltescu and Phil Blunsom. 2015. Pragmatic neural language modelling in machine translation. In *NAACL-HLT*.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew Dai, Rafal Jozefowicz, and Samy Bengio. 2016. Generating sentences from a continuous space. In *SIGNLL*.

Massimo Caccia, Lucas Caccia, William Fedus, Hugo Larochelle, Joelle Pineau, and Laurent Charlin. 2020. Language gans falling short. In *ICLR*.

Hengyi Cai, Hongshen Chen, Cheng Zhang, Yonghao Song, Xiaofang Zhao, Yangxi Li, Dongsheng Duan, and Dawei Yin. 2020. Learning from easy to complex: Adaptive multi-curricula learning for neural dialogue generation. In *AAAI*.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. A simple framework for contrastive learning of visual representations. *CoRR*, abs/2002.05709.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: pretraining text encoders as discriminators rather than generators. In *ICLR*.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Bo Dai and Dahua Lin. 2017. Contrastive learning for image captioning. In *NIPS*.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014a. Generative adversarial nets. In *NIPS*.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014b. Generative adversarial nets. In *NIPS*.

Xiaodong Gu, Kyunghyun Cho, JungWoo Ha, and Sunghun Kim. 2019. Dialogwae: Multimodal response generation with conditional wasserstein autoencoder. In *ICLR*.

Michael Gutmann and Aapo Hyvärinen. 2012. Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. *J. Mach. Learn. Res.*, 13:307–361.

Raia Hadsell, Sumit Chopra, and Yann LeCun. 2006. Dimensionality reduction by learning an invariant mapping. In *CVPR*.

Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in building intelligent open-domain dialog systems. *ACM Trans. Inf. Syst.*, 38(3).

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

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

Jiwei Li, Will Monroe, Alan Ritter, Dan Jurafsky, Michel Galley, and Jianfeng Gao. 2016b. Deep reinforcement learning for dialogue generation. In *EMNLP*.

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

Pierre Lison and Jörg Tiedemann. 2016. Opensubtitles2016: Extracting large parallel corpora from movie and TV subtitles. In *LREC*.

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*.

Alexander Miller, Will Feng, Dhruv Batra, Antoine Bordes, Adam Fisch, Jiasen Lu, Devi Parikh, and Jason Weston. 2017. ParlAI: A dialog research software platform. In *EMNLP*.

Andriy Mnih and Yee Whye Teh. 2012. A fast and simple algorithm for training neural probabilistic language models. In *ICML*.

Weili Nie, Nina Narodytska, and Ankit Patel. 2019. Relgan: Relational generative adversarial networks for text generation. In *ICLR*.

Aäron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *CoRR*, abs/1807.03748.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *ACL*.

Yookoon Park, Jaemin Cho, and Gunhee Kim. 2018. A hierarchical latent structure for variational conversation modeling. In *NAACL-HLT*.

Stephen E. Robertson and Hugo Zaragoza. 2009. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, 3(4):333–389.

Jo ao Sedoc, Daphne Ippolito, Arun Kirubarajan, Jai Thirani, Lyle Ungar, and Chris Callison-Burch. 2019. Chateval: A tool for chatbot evaluation. In *NAACL (Demonstrations)*.

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*.

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*.

Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural responding machine for short-text conversation. In *ACL*.

Xiaoyu Shen, Hui Su, Yanran Li, Wenjie Li, Shuzi Niu, Yang Zhao, Akiko Aizawa, and Guoping Long. 2017. A conditional variational framework for dialog generation. In *ACL*.

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*.

Ashish Vaswani, Yinggong Zhao, Victoria Fossum, and David Chiang. 2013. Decoding with large-scale neural language models improves translation. In *EMNLP*.

Yu Wu, Wei Wu, Chen Xing, Ming Zhou, and Zhoujun Li. 2017. Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots. In *ACL*.

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

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

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

Rui Yan. 2018. "chitty-chitty-chat bot": Deep learning for conversational ai. In *IJCAI*.

Chunyuan Yuan, Wei Zhou, Mingming Li, Shangwen Lv, Fuqing Zhu, Jizhong Han, and Songlin Hu. 2019. Multi-hop Selector Network for Multi-turn Response Selection in Retrieval-based Chatbots. In *EMNLP-IJCNLP*.

Hainan Zhang, Yanyan Lan, Jiafeng Guo, Jun Xu, and Xueqi Cheng. 2018a. Reinforcing coherence for sequence to sequence model in dialogue generation. In *IJCAI*.

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

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

Tiancheng Zhao, Ran Zhao, and Maxine Eskénazi. 2017. Learning discourse-level diversity for neural dialog models using conditional variational autoencoders. In *ACL*.

Xiangyang Zhou, Lu Li, Daxiang Dong, Yi Liu, Ying Chen, Wayne Xin Zhao, Dianhai Yu, and Hua Wu. 2018. Multi-Turn Response Selection for Chatbots with Deep Attention Matching Network. In *ACL*.