INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Conditional Response Augmentation for Dialogue using Knowledge Distillation

Myeongho Jeong[*], Seungtaek Choi[*], Hojae Han, Kyungho Kim, Seung-won Hwang

Department of Computer Science, Yonsei University, Seoul, Korea
`{wag9611, hist0613, stovecat, ggdg12, seungwonh}@yonsei.ac.kr`

Abstract

This paper studies dialogue response selection task. As state-of-the-arts are neural models requiring a large training set, data augmentation is essential to overcome the sparsity of observational annotation, where one observed response is annotated as gold. In this paper, we propose counterfactual augmentation, of considering whether unobserved utterances would "counterfactually" replace the labelled response, for the given context, and augment only if that is the case. We empirically show that our pipeline improves BERT-based models in two different response selection tasks without incurring annotation overheads.

1. Introduction

This paper studies the problem of response selection of the most appropriate answer given the dialogue history (or, context). A key challenge in this problem is, given the history, there can be multiple valid answers, which we denote as **one-to-many** property. However, training resources, based on "observational" annotations, annotate one of such valid answers.

As one way to address one-to-many problem, data augmentation strategies have been studied. To illustrate with a baseline augmentation **ctx-ctx**: One can find the semantic equivalent set of responses $R_{ctx-ctx}$, by finding an answer to another dialogue context with the same semantics, and treat all such answers as gold. However, this augmentation was reportedly problematic for task-[1] and persona-specific dialogues [2] where context cannot solely determine valid responses, because task goal and persona is **confounder** in determining responses, in addition to dialogue history.

One direction is to explicitly annotate such confounders, such as persona and dialogue state, so that augmentation can be done conditionally to the specified confounder, selecting a subset $R_c \subseteq R_{ctx-ctx}$. However, confounder annotation is expensive, such that in real-life benchmarks, counfounders often remain unannotated [3, 4], or annotated in a limited scale for evaluation only [5]. In either case, it is difficult to acquire a large scale annotation for training.

Alternatively, we consider a naive confounder estimation, using response similarity, giving the integrity of the augmented responses. However, response similarity is reportedly a poor estimator [4], augmenting a gold response "*Cheap please.*", with "*Could you find me a cheap restaurant?*". In isolation, they are similar, but when considering the context of asking "*Do you prefer a cheap or expensive restaurant?*", the latter does not qualify. Alternatively, we propose to leverage teacher model, which captures contexts too, for finding counterfactually plausible reponse, for both the given dialogue history and targeted response.

Our proposed method is evaluated on the two public benchmark datasets for next response selection task: `Advising` and `DailyDialog`. We observe the proposed method significantly boosts the performance of existing response selection approaches.

2. Preliminary

We first define response selection task, and describe a widely used baseline, Bi-encoder [6] architecture. Lastly, we state the challenge of selecting multiple responses.

2.1. Task: Response Selection

Next utterance selection task is the task that select proper utterances from candidates for given conversation context. Given a dataset $\mathcal{D} = \{(c_i, R_i)\}_{i=1}^{N}$, where c_i represents a conversation context, and R_i is a set of response candidates. Let $R_i = \{(r_{i,k}, y_{i,k}\}_{k=1}^{T}$, where T is the number of response candidates, determined in task setting. Each $r_{i,k}$ is a response candidate and $y_{i,k} \in \{0, 1\}$ denotes a label with $y_{i,k} = 1$ indicating $r_{i,k}$ is a proper response for c_i and $y_{i,k} = 0$ othrewise.

The goal of response selection task is to learn a matching model $s(\cdot, \cdot)$ from $\mathcal{D}$. For any context-response pair (c, r), the matching model gives a score $s(c, r)$ that reflects the matching degree between c and r, and thus allows one to rank a set of response candidates R_i according to the scores for response selection.

2.2. Architecture: Bi-Encoder

To design augmentation, we first introduce a widely adopted Bi-encoder [6] architecture for context-response matching $s(c, r)$. In a Bi-encoder, both the input context and the candidate response are encoded into vectors with BERT [7]:

$$\bar{c}_i = \text{BERT}_c(c_i) \tag{1}$$

$$\bar{r}_{i,k} = \text{BERT}_r(r_{i,k}) \tag{2}$$

where BERT_c and BERT_r are two transformers that have been pre-trained as described in [6]. It is noteworthy that the context and the response are encoded separately, allowing the precomputation of the embeddings of all contexts (and responses)[1].

The score of a response $r_{i,k}$ is given by the dot-product $\hat{s}(c_i, r_{i,k}) = \bar{c}_i \cdot \bar{r}_{i,k}$. Fine-tuning goal is trained to minimize a cross-entropy loss $\mathcal{L}$ in which the logits are $\hat{s}(c_i, r_{i,1}), ..., \hat{s}(c_i, r_{i,T})$, where $r_{i,1}$ is the only correct response:

$$\mathcal{L} = \sum_{\mathcal{D}} y_{i,k} \log \hat{s}(c_i, r_{i,k}) \tag{3}$$

To follow the convention of [6], during training, we consider all other gold responses of other contexts in the same batch as negative responses.

[*] The authors contribute equally to this paper.

[1] An alternative architecture, cross-encoder, is reportedly more effective, but it is not the case in our empirical study. In addition to that, cross-encoder cannot precompute the embeddings, which limits the efficiency in performing response selection. We thus do not consider cross-encoder as a solution in this work.

http://dx.doi.org/10.21437/Interspeech.2020-1968

2.3. Challenge: Multiple Valid Responses

As overviewed in Section 1, though there can be more than one valid responses (**one-to-many**), training resources are usually "observational", annotating one gold response per context (only one $y_{i,k} = 1$ for a context c_i). Using these resources as is for training neural models is reported to make the models follow a skewed dialog policy, ignoring other (unseen) feasible user behaviors [1]. Our hypothesis is there is **unobserved** multiple valid responses, such that $y_{i,1}, y_{i,2}, ... = 1$ and $y_{i,P+1}, y_{i,P+2}, ... = 0$ with the number of valid responses P. To deal with one-to-many property, some existing work aims to collect P multiple valid annotations explicitly, using meta-annotation such as dialogue states [1] or human paraphrasing [8]. Contrary to that, our goal is acquire "counterfactual" P observations from a single factual observation.

3. Our Method

Recall that the observed annotation is $\mathcal{D} = \{(c_i, R_i)\}_{i=1}^{N}$ where for each context c_i, and R_i consists of one gold annotation, denoted as $r_{i,1}$, and $T - 1$ negatively sampled examples. Our goal is to expand $\mathcal{D}$, a $N \times T$ matrix, into counterfactual observations of $N \times N$ matrix, where each context may have up to P positive labels.

1. Train teacher model $s^{(T)}$ on labeled dataset $\mathcal{D}$
2. Expand $\mathcal{D}$ into couterfactual pairs $\mathcal{D}'$
3. Filter $\mathcal{D}'$ to obtain $f(\mathcal{D}')$ with $s^{(T)}$
4. Generate soft-labels $\hat{s}^{(T)}(f(\mathcal{D}'))$ with $s^{(T)}$
5. Train student model $s^{(S)}$ on the mix of $\hat{s}^{(T)}(f(\mathcal{D}'))$ and $\mathcal{D}$.
6. Trained student model can be a teacher for another iteration[2].

Note unlike knowledge distillation aiming at a smaller or faster student, noisy student training [9] is considered knowledge expansion, giving the student model the same or higher capacity and tougher training environments with noises. We thus use a student network of the same size as the teacher, and following filtering and counterfactual estimation methods.

3.1. Data Filtering

Our goal is to get diverse responses without increasing labeling cost. Toward the goal, we leverage a fine-tuned bi-encoder architecture following intuition: if a context c_j is similar with the given context c_i, the response $r_{j,1}$ may be appropriate for the context c_i. In other words, human annotations for an unobserved gold response on the given context is close to their annotation on another similar context.

Formally, we set our first step as to find the response set that have semantic equivalent context by its context similarity. In order to conduct such objective, we first encode all the contexts $\bar{c}_i$ in $\mathcal{D}$. Then we build a matrix of context similarity $M^{\text{ctx}} \in \mathbb{R}^{N \times N}$ by comparing two contexts c_i and c_j. The (i, j)-th entry is calculated as $M^{\text{ctx}}_{i,j} = \text{sim}(\bar{c}_i, \bar{c}_j)$, where j denotes the index of another context in the dataset. In this work, we use cosine similarity as similarity measure, *i.e.*, $\text{sim}(a, b) = \cos(a, b) = \frac{a \cdot b}{\|a\| \cdot \|b\|}$.

Meanwhile, the above estimation yields small, yet non-zero scores for dissimilar pairs of contexts, generating inappropriate supervisions. To select more likely candidates, we filter the inappropriate candidates by redefining M^{ctx} with an introduction of threshold ϵ, where an entry with value smaller than ϵ becomes 0:

$$M^{\text{ctx}}_{ij} = \begin{cases} \text{sim}(\bar{c}_i, \bar{c}_j), & \text{if } \text{sim}(\bar{c}_i, \bar{c}_j) > \epsilon, \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

where we empirically set the threshold ϵ to 0.6. For the pair of having high similarity $M^{\text{ctx}}_{ij} > \epsilon$, we append each corresponding response $r_{i,1}$ and $r_{j,1}$ to the set of responses R_j and R_i respectively. We empirically tune the maximum number of augmented responses K as a hyper-parameter. At this step, we replace the sampled negative candidates with the augmented responses. To be fair, we do not increase the total number of candidates in one batch.

3.2. Counterfactual Estimation

We now discuss how to label $\mathcal{D}'$ with the teacher model $s^{(T)}$, which we argue to provide counterfactual estimation of how much the response is confounding to its context. Formally, we replace the objective function, designing a model to predict the soft-labels with minimal error, represented by objective function $\mathcal{L}$.

$$\mathcal{L} = \sum_{\mathcal{D}'} \bar{y}_{i,k} \log \hat{s}^{(S)}(c_i, r_{i,k}), \tag{5}$$

where $\bar{y}_{i,k}$ is the soft-labels from the teacher model $s^{(\text{T})}$.

In this work, we explore two counterfactual estimation methods of generating soft-labels for the filtered responses: 1) response similarity $\text{sim}(r_{i,1}, r_{j,1})$, and 2) context-response relevance $s(c_i, r_{j,1})$.

1) Response Similarity (*rsp-rsp*): Response similarity has a potential as confounder estimation– if the pair is semantically similar, the gold response $r_{i,1}$ and each augmented response $r_{i,2}, ..., r_{i,K+1}$ is more likely to share the confounder. Based on the assumption that the higher similarity the augmented response gets, the more the augmented response confounds to the gold response, we re-label the augmented responses with its response similarity. Formally, we add the augmented response $r_{j,1}$ with new label $\bar{y}_{i,k} = \text{sim}(\bar{r}_{i,1}, \bar{r}_{j,1})$, appending it to the response set R_i. We denote this dataset as *rsp-rsp* in later section.

2) Context-Response Relevance (*ctx-rsp*): However, the response similarity is not a good estimator, as it has risk of discarding an outlier, yet valid response. In contrast, we use better signals of confounder estimation [11, 12] to give soft-labels, *i.e.*, context-response relevance $\hat{s}^{(\text{T})}(c_i, r_{j,1})$. Compared to *rsp-rsp*, we argue that the predicted relevance *ctx-rsp* captures missing parts of the dialogue contexts too, which we call *latent confounder*. The written dialogue is just a partial observation of the true dialogue contexts, missing the rich annotations about agent's persona [2], background knowledges [13], and conversation skills. Giving the relevance can be considered as injecting latent confounders, accompanied with the *causal chains* between missing parts and the target responses.

Formally, the teacher model predicts $\hat{s}^{(T)}(c_i, r_{j,1})$ regardless of the gold response $r_{i,1}$. Then we generate the label of $r_{j,1}$ for c_i with $\bar{y}_{i,k} = \hat{s}^{(T)}(c_i, r_{j,1})$. We denote this dataset as *ctx-rsp*.

4. Experiments

We empirically validate the effectiveness of conditional response augmentation, using two widely adopted benchmark

[2] Iterations and mix ratio are determined empirically.

		Advising					DailyDialog				
Train Data	Teacher	MAP	R@1	R@10	R@50	EM	MAP	R@1	R@10	R@50	EM
Oracle											
ESIM [10]	-	0.3862	0.0973	0.5462	0.9089	0.1280	-	-	-	-	-
Bi-encoder	-	0.4570	0.1290	0.6157	0.9412	0.1720	-	-	-	-	-
Scarce											
Bi-encoder	-	0.3836	0.1308	0.5183	0.8659	0.1000	0.7838	0.1868	0.8575	0.9793	0.1932
+ Augmented (*ctx-ctx*)	-	0.4344	**0.1327**	0.6038	**0.9291**	0.1320	0.7809	0.1862	0.8541	0.9805	0.1835
	rsp-rsp	0.4311	0.1227	0.6036	0.9230	0.1220	0.7806	0.1860	0.8543	0.9886	0.1803
	ctx-rsp	**0.4485**	0.1264	**0.6149**	0.9265	**0.1380**	**0.8024**	**0.1884**	**0.8702**	**0.9890**	**0.2182**

Table 1: *First two rows trained on **Oracle** annotations for valid responses (upper bound), and the rest is for **Scarce** scenario.*

datasets from DSTC– **Advising** and **DailyDialogue**. respectively, the latter is more confounded. Section 4.1 will describe these sets in further details. Based on these sets, we address the following two key questions:

- **RQ1:** Does response augmentation improve model accuracy? Is our proposed augmentation more effective than existing methods? (Section 4.2)
- **RQ2:** Does teacher model contribute to confounder estimation? (Section 4.3)

4.1. Datasets

This section describes two datasets, where one is widely used next response selection task from DSTC 7 and the other one is originally built for response generation task. Here, we also describe how we repurposed it for response selection.

- **Advising**: **Advising** dataset [8] is introduced at DSTC 7, as the 3rd subtask. This dataset is constructed by paraphrasing each utterance in conversation. The training split is constructed by the same strategy introduced in [14], which considers each utterance as a potential response to all the previous utterances (dialogue context), resulting in multiple training instances from one dialogue session. With this dataset, we perform **Oracle** and **Scarce** evaluation: To validate our framework in **Scarce** annotation scenario, we sample only one gold response for each instance, by selecting one out of 1 ∼ 10 (avg: 3.6) gold responses. Meanwhile, we report the results of using all annotations for training, marked as **Oracle**, which can be used as an upper bound accuracy.
- **DailyDialog**: **DailyDialog** dataset with multi-reference test set [5] is constructed to evaluate semantic diversity of *generated* responses. Though this dataset is for generation task, we repurpose it by constructing new training and testing scheme suitable for response selection task, sampling negative candidates from other dialogue contexts. As there is no available training annotations of multi-reference, we only report the evaluation results of one-to-one approaches. The final evaluation requires to select 5 gold responses out of given 100 candidates.

For evaluation, we employ generally used metrics: mean average precision (MAP), recall at position k in 100 candidates ($R@k$), and exact match (EM).

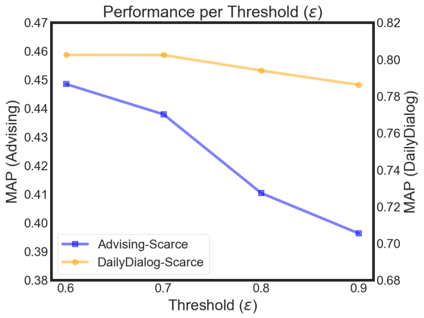

Figure 1: *Empirical study of threshold ϵ. Lower threshold (0.6) yields better performance.*

4.1.1. Implementation Details

We strictly follow the original settings of public bi-encoder implementation[3], specifically using `bi_model_huge_reddit` pretrained weights. However, BERT architecture requires large GPU memories, we modify the batch size and the number of response candidates to fit in our experimental environments. First, we modify batch size 512 to 32, processing 32 dialogue contexts in a batch. However, to prevent performance drop from reduced number of candidates, we additionally sample negative candidates from other contexts having up to 224 candidates for one context. Following original implementations, we use AdaMax [15] optimizer with 5e-05 learning rate for training on Advising dataset and Adam [15] optimizer with 5e-05 learning rate on DailyDialog dataset.

4.2. RQ1: Response Selection Performance

We first evaluate how our conditional augmentation contributes to the ranking performance. Table 1 shows the ranking performance for two multi-reference evaluation datasets. Compared to **Oracle**, using all human annotations for multiple valid annotations for training, our work samples only one gold response and still performs comparably, with our proposed augmentation: For confounder estimation, context-response relevance outperforms mostly, despite using all augmented responses is favored with respect to recall in some scenario. Empirically, we find that a noisier augmentation (with lower threshold ϵ) is favored (Figure 1), as the augmented responses are labeled with estimated confounder, giving the integrity to the soft-labels.

[3] `https://github.com/facebookresearch/ParlAI/tree/master/projects/polyencoder`

Dialog Context:		
Agent A: Good coming, sir. What can I do for you?		
Agent B: I'm Mr.Bob, Room 309. I'm checking out today. Can I have my bill now?		
Agent A: Certainly. Please wait a moment. Here is your bill.		
Agent B: What's the 30 yuan for?		
Gold Response:		
This is the charge for your laundry service on Nov.		
Augmented Responses:	***rsp-rsp***	***ctx-rsp***
That's for the breakfast you ordered from the service.	**0.626**	**0.999**
That's for the wine.	0.492	**0.996**
It's for the drinks.	0.501	**0.995**
For three bottles of Tsingtao beer.	0.516	**0.951**
Let me see... it's $ 50.	0.511	0.011

Table 2: *Illustration of teacher effectiveness on `DailyDialog`. **ctx-rsp** scores high on counterfactual answers.*

4.3. RQ2: Estimation Analysis

We measure effectiveness of pseudo labels.

- Would our estimation correlate with another dataset with annotated confounder?
- Would user find our counterfactual augmentation plausiable?

4.3.1. *Persona Estimation*

In this section, we treat persona statement as an explicitly stated confounder, and evaluate how well we can estimate such confounder when annotated information is deleted (and used as a ground-truth evaluation).

The most widely adopted dataset for such evaluation is PersonaChat dataset [2], specifically with the help of Dialog NLI [16], which annotates entailment relationship between persona and each utterance in dialogue history. We reconstruct the PersonaChat dataset as unseen persona selection task, where unseen persona means it is not used in all of the utterances of dialogue, though it is provided in persona list of the dialogue agent. Specifically, given a dialogue history, if the model indeed capture the hidden knowledge in the contexts, the model should select the unseen persona rather than randomly sampled one negative persona. To evaluate only how much Bi-encoder could estimate such hidden context, we use the Bi-encoder model trained on DailyDialog dataset.

As a result, the model ranks unseen persona higher than irrelevant persona with 61% chance. As claimed in [13], this suggests that pre-trained language model has an ability of catching extra information, even not explicitly represented in the dialogue utterances.

4.3.2. *User study*

We performed a qualitative analysis to gain more insight into the augmentation process. To examine the counterfactual response augmentation, we randomly select 50 dialogues from DailyDialog dataset with its augmented response and manually analyze them according to its validity.

First, *rsp-rsp* corrects 43.2% of the invalid labels[4] by giving low scores for the invalid responses. On the other hand, *ctx-rsp* improves the correcting chance up to 69.5% of the same responses. This result is consistent with Table 1 where *ctx-rsp* contributes more to the ranking accuracy than *rsp-rsp*.

Table 2 is an example of counterfactual response augmentation on `DailyDialog`. From the dialogue context, one can imagine the hidden contexts, such as "The hotel where Agent A works charge the usage of mini bar", such that selecting the following responses is also natural: "That's for the wine.". However, in this example, *rsp-rsp* seems to mainly concern about the "charge" for the service, giving high values to responses with specific payment like "$ 50". On the other hand, *ctx-rsp* labels low values for such unnatural responses, which manifests the effects of knowledge expansion for generating counterfactual labels.

5. Related Work

This paper studies one-to-many problem in training dialogue systems. An extreme example is "I don't know" being valid to all questions, yet rarely useful.

The goal is thus, to create diverse valid answers, but specific to the given dialogue context, task, and persona. Existing approaches mostly aim at modeling responses with confounder, specified as meta information, such as response specificity [17, 18]. For task-oriented system, dialogue state is annotated [19] to generate more diverse dialog responses, conditional to both dialogue history and task completion policy. [1, 4] maps dialogue system into valid system actions, and create conversation conditional to this action. The difference is that [1] aims to generate diverse action-specific responses, while [4] targets to improve generation performance using only paraphrase. Ours shares the same goal of generating diverse yet action-specific responses, but we do not require additional annotations. [20] tackles one-to-one mapping between query-response in single turn dialogue task by combining the common features between different valid responses.

Orthogonally to augmentation, there have been modelling efforts for general one-to-many generation tasks. [21] explicitly separates diversification process from generation using plug-and-play module, employing mixture of experts. [22] also uses BERT-based semantic similarity to propose an evaluation that is more flexible than cross-entropy objective for sentence generation task.

6. Conclusion

This paper proposes counterfactual augmentation of acquiring training instances by labeling unobserved utterances, using knowledge expansion from noisy student training from pre-trained language model. Our empirical results validate that our proposed augmentation improves response selection accuracy in real-life benchmarks.

7. Acknowledgements

This work is supported by AI Graduate School Program (2020-0-01361) and IITP grant (No.2017-0-01779, XAI) supervised by IITP. Hwang is a corresponding author.

[4] out of 43.5% of augmented responses that are invalid after filtering

8. References

[1] Y. Zhang, Z. Ou, and Z. Yu, "Task-oriented dialog systems that consider multiple appropriate responses under the same context," *arXiv preprint arXiv:1911.10484*, 2019.

[2] S. Zhang, E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston, "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)*, 2018, pp. 2204–2213.

[3] A. Madotto, Z. Lin, C.-S. Wu, and P. Fung, "Personalizing dialogue agents via meta-learning," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 5454–5459.

[4] S. Gao, Y. Zhang, Z. Ou, and Z. Yu, "Paraphrase augmented task-oriented dialog generation," *arXiv preprint arXiv:2004.07462*, 2020.

[5] P. Gupta, S. Mehri, T. Zhao, A. Pavel, M. Eskenazi, and J. P. Bigham, "Investigating evaluation of open-domain dialogue systems with human generated multiple references," *arXiv preprint arXiv:1907.10568*, 2019.

[6] S. Humeau, K. Shuster, M.-A. Lachaux, and J. Weston, "Poly-encoders: Transformer architectures and pre-training strategies for fast and accurate multi-sentence scoring," *arXiv preprint arXiv:1905.01969*, 2019.

[7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186.

[8] K. Yoshino, C. Hori, J. Perez, L. F. D'Haro, L. Polymenakos, C. Gunasekara, W. S. Lasecki, J. K. Kummerfeld, M. Galley, C. Brockett *et al.*, "Dialog system technology challenge 7," *arXiv preprint arXiv:1901.03461*, 2019.

[9] D. S. Park, Y. Zhang, Y. Jia, W. Han, C.-C. Chiu, B. Li, Y. Wu, and Q. V. Le, "Improved noisy student training for automatic speech recognition," *arXiv preprint arXiv:2005.09629*, 2020.

[10] Q. Chen and W. Wang, "Sequential attention-based network for noetic end-to-end response selection," *arXiv preprint arXiv:1901.02609*, 2019.

[11] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[12] Q. Xie, E. Hovy, M.-T. Luong, and Q. V. Le, "Self-training with noisy student improves imagenet classification," *arXiv preprint arXiv:1911.04252*, 2019.

[13] T. H. Trinh and Q. V. Le, "Do language models have common sense?" 2018.

[14] R. Lowe, N. Pow, I. Serban, and J. Pineau, "The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems," *arXiv preprint arXiv:1506.08909*, 2015.

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[16] S. Welleck, J. Weston, A. Szlam, and K. Cho, "Dialogue natural language inference," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 3731–3741.

[17] G. Zhou, P. Luo, R. Cao, F. Lin, B. Chen, and Q. He, "Mechanism-aware neural machine for dialogue response generation," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[18] G. Zhou, P. Luo, Y. Xiao, F. Lin, B. Chen, and Q. He, "Elastic responding machine for dialog generation with dynamically mechanism selecting," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[19] J. Rajendran, J. Ganhotra, S. Singh, and L. Polymenakos, "Learning end-to-end goal-oriented dialog with multiple answers," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 3834–3843.

[20] L. Qiu, J. Li, W. Bi, D. Zhao, and R. Yan, "Are training samples correlated? learning to generate dialogue responses with multiple references," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 3826–3835.

[21] J. Cho, M. Seo, and H. Hajishirzi, "Mixture content selection for diverse sequence generation," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 3112–3122.

[22] G. Yasui, Y. Tsuruoka, and M. Nagata, "Using semantic similarity as reward for reinforcement learning in sentence generation," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, 2019, pp. 400–406.