Enhancing Dialogue Generation via Multi-Level Contrastive Learning*

Xin Li¹, Piji Li², Yan Wang², Xiaojiang Liu² and Wai Lam¹
¹The Chinese University of Hong Kong
²Tencent AI Lab
{lixin,wlam}@se.cuhk.edu.hk
{pijili,brandenwang,kieranliu}@tencent.com

Abstract

Most of the existing works for dialogue generation are data-driven models trained directly on corpora crawled from websites. They mainly focus on improving the model architecture to produce better responses but pay little attention to considering the quality of the training data contrastively. In this paper, we propose a multi-level contrastive learning paradigm to model the fine-grained quality of the responses with respect to the query. A Rank-aware Calibration (RC) network is designed to construct the multi-level contrastive optimization objectives. Instead of emphasizing the sentence-level quality, which may erroneously encourage/suppress the generation of uninformative/informative words, we design an exquisite token-level strategy for estimating the instance loss more accurately. On the other hand, we build a Knowledge Inference (KI) component to capture the keyword knowledge from the reference during training and exploit such information to encourage the generation of informative words. We evaluate the proposed model on a carefully annotated dialogue dataset and the results suggest that our model can generate more relevant and diverse responses compared to the baseline models.

1 Introduction

Response generation for dialogue systems has stimulated great interests for researchers recently [Ritter et al., 2011; Shang et al., 2015; Vinyals and Le, 2015]. The core idea of dialogue generation is to formulate the task as a sequence translation problem and translate the query to a response. One common neural model is the sequence-to-sequence (S2S) encoder-decoder framework [Cho et al., 2014; Sutskever et al., 2014; Bahdanau et al., 2015]. Many approaches have been proposed to improve the basic S2S model for better human-computer conversation performance [Li et al., 2016; Li et al., 2017; Xing et al., 2017; Shao et al., 2017; Xu et al., 2017; Zhou et al., 2017; Pei and Li, 2018; Wang et al., 2018;]

*This work was done when Xin Li was an intern as Tencent AI Lab.

| Query | Response | Quality |
|-------|----------|---------|
| 又开始下雨了 It started to rain again | 1. 下雨虽然带来很多不便，可太阳也好无聊啊，建议别 | Relevant and interesting |
| | Although rains are inconvenient, sunny day is no better than rainy day since I am afraid of the sun. | |
| | 2. 我讨厌下雨天！I hate rainy days! | Relevant but simple |
| | 3. 你現在在衡阳还是广州？Are you in Qingdao or Guangzhou now? | Acceptable but universal |
| | 4. 嗯，又下雨了 Yes, it rains again | Quiet boring |
| | 5. 这所大学怎么样是胡园 | Irrelevant |
| | The university is near martyrs cemetery. | |

Table 1: Example of a sample from training-set: query with responses in different quality.

Gao et al., 2019a; Zhang et al., 2020; He and Glass, 2020; Tian et al., 2020.

Despite their popularity, these approaches assume that each training sample, namely, query-response pair, contributes equally to the model and ignore the consideration of different response quality contrastively. Table 1 depicts some example responses for a particular query in a dialogue dataset. Both of the first and the second response are relevant to the query but the first one is obviously better when considering informativeness and interestingness. The third response is acceptable for the conversation but quite universal, meaning that, it can also be used to answer other queries. Thus, its quality is not as good as that of the first two responses. The fourth response is poor since it directly copies part of the query. Although the fourth response is not acceptable, it is still better when compared with the fifth response, which is completely irrelevant to the query.

Some initial attempts [Shang et al., 2018; Liu et al., 2018] have been conducted to consider the quality of the training data. Following the idea of instance weighting [Jiang and Zhai, 2007], [Shang et al., 2018] pre-train a calibration network to calculate the response quality score for each training sample (i.e., query-response pair) and update the model with the weighted combination of the sample loss. Similarly, [Liu et al., 2018] estimate the instance score based on the corpus-level n-gram co-occurrence and the length of the response. Both of them are simple to implement but they still have some limitations: (1) The calibration network is only trained on relevant responses and irrelevant responses from other queries and therefore cannot capture the fine-grained
response quality, as exemplified in Table 1; (2) The instance weighting strategy treats all tokens in the response as equal importance to the query by assigning them with the same quality score, which may erroneously encourage the generation of some uninformative words in the relevant responses (e.g., “since” and “afraid” in the first response in Table 1). To tackle the issues mentioned above, we introduce the Contrastive Learning paradigm [Hadsell et al., 2006; He et al., 2019; Chen et al., 2020; Iler et al., 2020] to model the multi-level fine-grained quality of the responses with respect to the query. Specifically, we develop a Rank-aware Calibration (RC) network aiming for modeling the fine-grained quality and characterizing the response properties (e.g., relevance and informativeness) that will affect the conversation experience with a multi-scale response quality score. The rank-aware calibrator adopts the strategies of pointwise regression and pairwise ranking for gauging the quality of the query-response pair. Besides, to address the second limitation aforementioned, we design a more exquisite strategy to consider the different importance of tokens instead of simply scaling the training sample loss with the response-level quality score. Concretely, we propose to conditionally sample a response via Monte-Carlo Rollout [Yu et al., 2017; Lin et al., 2017] for each gold standard response token and deem the quality scores of the sampled responses as the importance of the tokens in the sample loss estimation.

It is also observed that some meaningful words such as “university” and “martyrs cemetery” in the fifth response in Table 1 are very likely to receive low quality scores due to the irrelevance to the query. Thus, we propose Knowledge Inference (KI) component to explicitly encourage the generation of the informative tokens in the gold standard responses. This component firstly associates the query and the decoder hidden representation with the memories of the informative tokens and then incorporates the summarized memories into each decoding step.

In summary, our contributions are as follows:

- To enhance the performance of dialogue generation, we propose a multi-level contrastive learning paradigm to model the fine-grained quality of the responses with respect to the query.
- We propose a Rank-aware Calibration (RC) network to construct the multi-level contrastive objectives. We further design a strategy to calibrate the model training with token-level quality information.
- We propose to reconsider the generation of some informative words erroneously punished by the calibrator via a tailor-made Knowledge Inference (KI) component.
- We build a dataset with fine-grained response annotations and conduct extensive evaluations. The experimental results validate the effectiveness of the proposed framework.

Code and the labelled dataset will be public to facilitate the research.

2 Model

2.1 Overview

Given a query-response pair \((X, Y)\), where the query word sequence \(X = \{x_1, x_2, \ldots, x_n\}\) and the response word sequence \(Y = \{y_1, y_2, \ldots, y_m\}\). The aim of the dialogue generation task is to maximize the conditional probability \(p(Y|X)\) and generate the relevant and meaningful responses with respect to the query.

As shown in Figure 1, the model proposed in this paper consist of three major components: the backbone S2S model, a rank-aware calibration network (RC) for response quality estimation, and a knowledge inference component (KI) for encouraging the generation of informative words. For RC-based response quality estimation, we apply Monte-Carlo Rollout with the policy network \(G^\beta\) to perform response sampling conditioned on the current token \(y_t\) and feed the sampled response \(\hat{Y}_t \in \mathbb{R}^m\) to the RC, where three contrastive quality estimation components \(D^{ns}, D^{pt}\) and \(D^{pair}\), namely, negative sampling component, pointwise ranking component, and pairwise ranking component are included. The produced quality scores of the \(\hat{Y}_t\) are \(q^{ns}, q^{pt}, q^{pair}\) respectively. For knowledge inference (KI), a knowledge memory \(K \in \mathbb{R}^{n_k \times \text{dim}_k}\) and a knowledge attention layer are designed. \(n_k\) refers to the number of the knowledge memories for the informative tokens. The knowledge summary \(c^k_t\) will also serve as the input for the \(t\)-th decoding step.

2.2 Response Generation

The backbone generation model is an attention-based sequence-to-sequence framework. Specifically, in encoding, we employ a bi-directional LSTM [Hochreiter and Schmidhuber, 1997] to map the query input \(x_i\) to the distributed hidden representation \(h_t \in \mathbb{R}^{2\text{dim}_h}\) \((i \in [1, n])\) as follows:

\[
h_t = [\text{LSTM}(x_i, \hat{h}_{i-1}); \text{LSTM}(x_i, \hat{h}_{i+1})]
\]

where the operation of an LSTM unit on \(x_i\) is denoted as \(\text{LSTM}(x_i, \hat{h}_{i-1})\). In decoding, the computational formula of the decoding hidden representation \(s_t\) at the \(t\)-th time step is below:

\[
s_t = f(y_{t-1}, s_{t-1}, c_t)
\]

The context vector \(c_t\) is computed as the weighted combination of the encoder hidden memories \(H = [h_1, \ldots, h_n]\) along with the attention signal \(a_t \in \mathbb{R}^n\). Then \(s_t\) is fed to a softmax layer to calculate the probability distribution of the candidate words:

\[
p(Y|X) = \prod_{t=1}^{m} p(y_t|X, y_{1:t-1})
\]

2.3 Rank-Aware Quality Estimation

As shown in Figure 1, the Rank-aware Calibrator (RC) model is composed of three components: \(D^{ns}, D^{pt}\) and \(D^{pair}\). Each of the component contains a scoring network for calculating the quality score, either performing semantic matching or learning a joint semantic representation for a query and a response. The matching-based scoring network \(g_M\) maps a query and a response to vector representations and employs a bi-linear layer to measure the semantic relatedness between these two representations. The joint-representation-based scoring network \(g_1\) learns the joint semantic representation for the query and the response and utilizes a non-linear transformation layer to obtain the quality score. To exploit the fine-grained response quality, we propose three ranking
strategies, namely, **Negative Sampling**, **Pointwise Ranking** and **Pairwise Ranking**, to pre-train the components \( \text{D}^\text{ns}, \text{D}^\text{pt} \) and \( \text{D}^\text{pair} \) respectively.

**Negative Sampling**  Negative sampling is the strategy for optimizing the model \( \text{D}^\text{ns} \). Specifically, for each query response pair \((X, Y)\), we sample a response \( Y^- \) from other queries or a training dataset to construct a negative sample \((X, Y^-)\). Then, we augment the original training dataset with the sampled negative data and train the scoring network by maximizing the following objective:

\[
L^\text{ns} = b \cdot g^\text{ns}_i(X, Y^+) + (1-b) \cdot (1-g^\text{ns}_i(X, Y^-))
\]

where \( g^\text{ns}_i \) denotes the score obtained based on the joint representation of \( X \) and \( Y^+ \). The binary variable \( b \) is set as 1 if \( Y^+ \) is from the original training dataset. Otherwise, \( b \) is 0. Optimizing this classification-based objective can rank the relevant responses higher than the irrelevant ones.

**Pointwise Ranking**  As mentioned in the Introduction section, only differentiating relevant responses from the irrelevant responses is not enough. Since the multi-level quality annotations are available in the training dataset, one direct way for modeling the fine-grained response quality is to minimize the gap between the predicted quality scores and the gold standard multi-level scores. Thus, we propose a regression-based criterion for optimizing the model \( \text{D}^\text{pt} \). The objective function is given by:

\[
L^\text{pt} = (q^g - g^\text{pt}_s(X, Y))^2
\]

where \( q^g \) is the gold-standard quality score for the query-response pair \((X, Y)\) and \( g^\text{pt}_s \) is the matching-based scoring function. With this objective, the \( \text{D}^\text{pt} \) is enforced to pay attention to different levels of quality scores rather than just 0-1 relevance. In other words, \( \text{D}^\text{pt} \) can exploit how good the response \( Y \) is apart from its relevance/irrelevance to the query \( X \).

**Pairwise Ranking**  The pointwise ranking strategy adopted in \( \text{D}^\text{pt} \) is to directly model the absolute quality scores. Compared to the pointwise ranking, the pairwise ranking strategy implicitly exploits the multi-level quality scores by learning to assign preference for the pre-built paired responses \( Y^{++} \) and \( Y^+ \) from the same query \( X \). Here, \( Y^{++} \) and \( Y^+ \) denote the responses belonging to the adjacent quality levels and \( Y^{++} \) is the better one. To optimize the scoring component \( g^\text{pair}_M \) in \( \text{D}^\text{pair} \), the pairwise ranking strategy rewards the better response in the ranking pair and penalizes the other one so that the margin of their quality estimations is as close as possible to the pre-specified value \( \Delta \). The loss function is as follows:

\[
L^\text{pair} = \max \left( 0, \Delta - g^\text{pair}_M(X, Y^{++}) + g^\text{pair}_M(X, Y^+) \right)
\]

where the score function \( g^\text{pair}_M \) calculates the quality scores of the response via matching the semantic representations of the response and the query. Then, we traverse all of the possible response pairs under the query \( X \) to make the model \( \text{D}^\text{pair} \) aware of different quality levels.

Intuitively, we can utilize the RC model to estimate the quality of each training sample and leverage the quality scores to adjust the weight of the training sample loss. However, as mentioned above, such response-level strategy ignores the consideration of the importance of different tokens in the same response. In this paper, we give a solution that performing token-level calibration with the fine-grained quality estimation.

### 2.4 Token-Level Calibration

The aim of the vanilla sequence-to-sequence (S2S) model is to minimize the following negative log-likelihood (NLL) objective function:

\[
L^\text{NLL} = \frac{1}{|\mathcal{D}|} \sum_{(X, Y) \in \mathcal{D}} -\log(p(Y|X))
\]

where \( \mathcal{D} \) denotes the training dataset or a training batch and \( p(Y|X) \) is the conditional generation probability given by Eq 3.

Recall that response-level quality estimation treats all of the tokens in the response as equal importance and cannot
accurately reflect the importance of the token in the response, which may erroneously encourage the generation of some uninformative words. In order to integrate the quality information more accurately, we propose to calibrate the S2S training with token-level quality information. Specifically, we propose to measure the quality of the sampled response conditioned on each token and deem the quality score estimated by rank-aware calibrator as the importance of the corresponding token in the sample loss estimation.

Given the training sample \(X, Y\) and the gold standard response token \(y_t\) at the \(t\)-th decoding step, we firstly apply Monte-Carlo search with a rollout policy \(G^\beta\) (see the “Rollout” component in Figure 1), which is set identical to the current generator (i.e., S2S model), to conduct the sampling conditioned on \(y_t\) (that is, perform generation with the starting token as \(y_t\)). Let \(\hat{Y}_t\) be the sampled response, then, we employ the pre-trained rank-aware calibrator to calculate the fine-grained quality score for \(\hat{Y}_t\). Specifically, \(\hat{Y}_t\) is fed to the components \(D^r, D^d\) and \(D^{pairex}\) separately and the final quality score is the mean value of the outputs from the scoring models:

\[
q_t = \frac{1}{3}[D^r(X, \hat{Y}_t) + D^d(X, \hat{Y}_t) + D^{pairex}(X, \hat{Y}_t)]
\]

To reduce the variance of the sampling, we perform the Monte-Carlo search \(N\) times, yielding a set of simulated responses \(\{\hat{Y}_t^1, \ldots, \hat{Y}_t^N\}\) for the response token \(y_t\). The final token-level quality estimations for \(y_t\) is:

\[
q_t = \frac{1}{N} \sum_{i=1}^{N} q_t^i
\]

where \(q_t^i\) is the fine-grained quality score of the sampled response \(\hat{Y}_t^i\). Here, \(q_t\) characterizes the contribution of the \(t\)-th generation step to computing the training sample loss. Compared to [Shang et al., 2018; Liu et al., 2018], which treat each generation step equally, performing the weighted sum of the losses at each generation step can alleviate the incorrect bonus on some uninformative words. The calculations of the sample loss \(L_{RC}(X, Y)\) and the total loss \(L_{RC}\) are given below:

\[
L_{RC}(X, Y) = - \sum_{t=1}^{m} q_t \ast \log(p(y_t|X, y_{1:t-1}))
\]

\[
L_{RC} = \frac{1}{|D|} \sum_{(X, Y) \in D} L_{RC}(X, Y)
\]

2.5 Knowledge Inference

Recall the issue that some meaningful words in the irrelevant responses may be erroneously penalized due to the topical irrelevance. For example, in the fifth response in Table 1, the tokens “university” and “martyrs” will be assigned low quality scores and thus the generation of these words will be suppressed after the training. To reduce such kind of side effect, we propose Knowledge Inference (KI) component. The key idea of this component is to guide the generation with the useful knowledge. In this paper, we regard keywords (i.e., informative words) as the knowledge. Firstly, we utilize TextRank model [Mihalcea and Tarau, 2004] to extract the keywords from the responses in the training dataset. Then, we construct a knowledge memory \(K \in \mathbb{R}^{n_k \times \text{dim}_k}\), where \(n_k\) memory vectors correspond to \(n_k\) informative words obtained from the training corpus. After building the knowledge memory, we leverage the context-aware decoding hidden representation \(s_t\) to pay attention to the memory vectors \(k_j (j \in \{1, \ldots, n_k\})\) and summarize the knowledge via weighted combination:

\[
p(k_j|s_t) = \text{tanh}(s_t^T W_K k_j)
\]

where \(W_K\) is a parameter matrix in the KI component and \(\text{tanh}\) denotes the hyperbolic tangent function. \(c_t^k\) is the summarized knowledge at the \(t\)-th generation timestep. To incorporate the knowledge into the generation, we feed the concatenated \(c_t^k\) and decoder hidden representation \(s_t\) to the ultimate softmax layer for producing the token \(y_t\).

Learning the alignment between the memory \(K\) and the decoder hidden representation \(s_t\) from scratch is difficult because the model has no prior knowledge of the informative words. Thus, we regard the extracted informative words from the current ground-truth response \(Y\) as additional supervision signal to help the training of \(K\). Given the informative words \(Y^K\) extracted from \(Y\), the objective of the KI component is defined as follows:

\[
L_{KI} = - \frac{1}{|D|} \sum_{(X, Y) \in D} \sum_{t=1}^{n} \sum_{j=1}^{n_k} b_j^K \ast \delta(p(k_j|s_t))
\]

where the binary variable \(b_j^K\) is the indicator of the existence of the \(j\)-th informative word and \(\delta\) denotes the sigmoid activation function. \(p(k_j|s_t)\) is the alignment score between \(k_j\) and \(s_t\). By introducing the objective \(L_{KI}\), the model can learn better knowledge memory \(K\) and its alignment with \(s_t\) based on more feedback from both the S2S component and the KI component. Consequently, there will be a higher potential generating informative words.

2.6 Joint Training

The parameters of the pre-trained rank-aware calibrator model are kept fixed all the time. Although the loss with rank-aware calibration, i.e., \(L_{RC}\), reduce the possibility of generating unimportant words, it may weaken the language model constraint on the generated responses. Thus, we combine the NLL loss \(L_{NLL}\) with \(L_{RC}\) during the training. The loss \(L_{KI}\) is also included to encourage the generation of the informative words. The final training objective \(J(\theta)\) of the proposed framework is as follows:

\[
J(\theta) = L_{NLL} + L_{RC} + L_{KI}
\]

3 Experimental Setup

3.1 Dataset

To investigate the effectiveness of the proposed framework, we build a dialogue dataset with multi-level fine-grained
quality labels. As shown in Table 1, we consider 5-level response quality: (1) L5: relevant and interesting; (2) L4: relevant but simple; (3) L3: acceptable but universal; (4) L2: quiet boring; (5) L1: irrelevant. To process such quality labels, we convert L1 to L5 to the normalized quality scores, namely, 0.0, 0.25, 0.5, 0.75 and 1.0 respectively.

We crawl 534,381 query-response pairs from Douban Group, a famous forum in China\(^1\), and recruit human annotators to annotate the multi-level quality labels for each sample. We split the annotated dataset into three parts, where 522,881 query-response pairs are for training, 10000 query-response pairs are for validation and the remaining 1500 query-response pairs are for testing. There is no overlap among the queries in training, validation and testing set.

### 3.2 Settings

We compare our model with the following baselines and comparison models:

- **S2S**: It is the standard attention-based sequence-to-sequence model [Bahdanau et al., 2015].
- **S2S-RW** [Liu et al., 2018]: **S2S** with Re-Weighting. It is an extended S2S model with response quality measurement based on the corpus-level n-gram co-occurrence statistics and the length of the response.
- **S2S-CN** [Shang et al., 2018]: **S2S** with Calibration Networks. It is an enhanced S2S model where a pre-trained calibration networks is introduced to adjust the weight of the sample loss.
- **S2S-GT**: **S2S** with Ground Truth. It is a variant of the models [Liu et al., 2018] and [Shang et al., 2018], where we replace the instance quality measurement with the ground truth quality scores.
- **S2S-DF**: **S2S** with Data Filtering. It is a S2S model trained on the high quality data where we adopt the data filtering strategy proposed in [Xu et al., 2018] to obtain the topically related query-response pairs.\(^2\)
- **S2S-DF\(^+\)**: A variant of **S2S-DF**. We perform data filtering according to the gold standard multi-level quality score.

| Model | BLEU-1 | BLEU-2 | BLEU-3 | Average | Extrema | Greedy | Dist-1 | Dist-2 |
|-------|--------|--------|--------|---------|---------|--------|--------|--------|
| S2S   | 17.3   | 3.9    | 1.3    | 0.481   | 0.286   | 0.403  | 0.059  | 0.232  |
| S2S-RW| 17.1   | 3.4    | 1.1    | 0.519   | 0.294   | 0.409  | 0.052  | 0.213  |
| S2S-CN| 18.1   | 4.0    | 1.0    | 0.525   | 0.302   | 0.401  | 0.058  | 0.224  |
| S2S-GT| 18.9   | 4.3    | 1.5    | 0.521   | 0.317   | 0.411  | 0.062  | 0.245  |
| S2S-DF| 16.9   | 3.0    | 0.8    | 0.491   | 0.295   | 0.398  | 0.060  | 0.238  |
| S2S-DF\(^+\)| 20.2   | 4.5    | 1.7    | 0.501   | 0.308   | 0.417  | 0.054  | 0.213  |
| OURS  | 20.9   | 4.9    | 1.8    | 0.559   | 0.342   | 0.471  | 0.077  | 0.288  |

Table 2: Results on automatic metrics.

| Model | +2  | +1  | 0  | Avg. |
|-------|-----|-----|----|------|
| S2S   | 7.50| 61.00| 31.50| 76.00|
| S2S w/MMI | 10.00| 59.50| 30.50| 79.50|
| S2S-GT w/MMI | 10.50| 60.00| 29.50| 81.00|
| OURS w/MMI | 13.50| 58.00| 28.50| 85.00|

Table 3: Human evaluation results. (%)

Specifically, we only preserve query-response pairs of L4 or L5 quality for training the model.\(^3\)

We divide the comparisons with the baseline models into two groups according to using the Maximum Mutual Information (MMI) decoding [Li et al., 2016] or not. The first group of comparison is under the setting **w/o MMI**, where the decoding is identical to the normal beam search. The second group is under the setting **w/ MMI**, where an inverse S2S (response-to-query) model is introduced to rerank the \(N\)-best (\(N\) is set as 50) hypothesis generated from the standard S2S (query-to-response) model. We employ a two-layer bi-directional LSTM as the encoder and a two-layer unidirectional LSTM as the decoder. The dimension of both the encoder and the decoder hidden representations is 500. The word embeddings are randomly initialized and the size of each word vector is 300. For the model training, we employ Adam [Kingma and Ba, 2014] as optimizer, with the initial learning rate being 0.0001 and the decay rate being 0.9. The dimension of the keyword memory vector \(d_{\text{emb}}\) is 30. In the rank-aware calibrator, we employ multi-channel CNN [Kim, 2014] to extract sentence-level features. The sizes of convolutional filter are 1, 2, 3 and the number of filters is 100 for each size.

### 3.3 Evaluation Metrics

We evaluate our model and the comparison models using the following evaluation metrics:

**Word Overlap**: we employ BLEU-1, BLEU-2 and BLEU-3 to measure the word overlap between the generated and the gold standard responses.

**Embedding Similarity**: Actually, BLEU metric does not

\(^1\)https://www.douban.com/group/explore

\(^2\)https://github.com/XinnuoXu/CVAE_Dial

\(^3\)By doing this, the size of the training dataset reduces from 522,881 to 334,471.
### 4 Results and Discussions

#### 4.1 Main Results

Table 2 depicts the experimental results on the automatic metrics. Under the settings of both w/o MMI (the first group in Table 2) and w/ MMI (the second group in Table 2), our model consistently outperforms the baseline and the comparison models on all of the metrics, demonstrating the effectiveness of the proposed components. Specifically, our model is better than all of the non MMI-based baselines on Dist-1 and Dist-2, suggesting that the proposed model can generate more diverse and informative responses. The probable reason is that we introduce a knowledge inference component to encourage the generation of some informative words erroneously punished by the models adopting instance-weighting strategies (i.e., S2S-RW, S2S-CN and S2S-GT). We also observe that OURS achieves better performance on BLEU scores and Embedding Similarity compared with S2S-GT. This finding shows that token-level training loss estimation together with rank-aware calibration is a better strategy than just scaling the sample loss with an instance weight. Another interesting finding is that training model with only the high-quality data, as done in S2S-DF and S2S-DF”, is still worse than the instance weighting strategy adopted in S2S-GT, indicating that soft data filtering via quality estimation is a more suitable way for integrating quality information into the model training.

From the human evaluation results in Table 3, our model generates the most satisfactory responses (responses labeled as +2) and the least invalid responses (responses labeled as 0). Specifically, OURS w/ MMI increases 3.5% +2 responses and reduces 2.0% 0 responses compared with S2S w/ MMI. Without introducing the additional knowledge and diversity modeling, the pure S2S model tend to produce more generic responses than the MMI-based models, i.e., S2S w/ MMI and OURS w/ MMI. We also find that taking the instance quality into consideration can lead to less generation of the 0 responses, for example, OURS w/ MMI / S2S-GT w/ MMI versus S2S w/ MMI. The above observations are somewhat consistent with those in the Table 2.

#### 4.2 Ablation Study

To further investigate the effectiveness of the proposed component, we compare our model with the ablated models. The results of the ablation study are listed in Table 4. OURS significantly outperforms the ablated models on the relevance metrics (i.e., Embedding Similarity) but the performance gap on the diversity metrics is small. It is reasonable because the aim of the proposed components is not to model the diversity. Removing either the KI component or the RC component will result in performance degradation, especially for the case removing the RC component, our model degenerates to the S2S model together with the knowledge inference (KI) component. Comparing with OURS, the relevance scores of the OURS w/o RC drastically decrease (-6%, -6%, -7% on Average, Extrema and Greedy respectively), indicating that our token-level rank-aware calibration is helpful for improving the relevance of the generated responses. Without RC & KI components, our model will become equivalent to the standard S2S model and the metric scores will drop for further step.

#### 4.3 Case Visualization and Analysis

In Figure 2, we visualize the token-level quality scores calculated from Eq 8. Each cell corresponds to a Chinese word and it uses depth of color to represent the importance of this word. We can observe that the importance score for the token “Sichuan”, a province in China, is significantly larger than those for other tokens when the query is talking about “hometown” (the upper query). Similarly, the token “cloudy” and “rainy” are highlighted by the rank-aware calibrator when the topic of the query is related to the weather. These examples demonstrate that our token-level strategy armed with Monte-Carlo Rollout can accurately characterize the quality of token with respect to the query, leading to better calibration for the training.

Table 5 shows some example outputs. The informative words appearing in the training dataset are wrapped in the
What I use is browser.

Table 5: Example output, color printing is preferred. The informative words are wrapped in brackets and coupled with underline.
[Hadsell et al., 2006] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In Proceedings of CVPR, volume 2, pages 1735–1742, 2006.

[He and Glass, 2020] Tianxue He and James Glass. Negative training for neural dialogue response generation. In Proceedings of ACL, pages 2044–2058, 2020.

[He et al., 2019] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. arXiv preprint arXiv:1911.05722, 2019.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and J¨urgen Schmidhuber. Long short-term memory. Neural computation, 9(8):1735–1780, 1997.

[Iter et al., 2020] Dan Iter, Kelvin Guu, Larry Lansing, and Dan Jurafsky. Pretraining with contrastive sentence objectives improves discourse performance of language models. arXiv preprint arXiv:2005.10389, 2020.

[Jiang and Zhai, 2007] Jing Jiang and ChengXiang Zhai. Instance weighting for domain adaptation in NLP. In Proceedings of ACL, pages 264–271, 2007.

[Kim, 2014] Yoon Kim. Convolutional neural networks for sentence classification. In Proceedings of EMNLP, pages 1746–1751, 2014.

[Kingma and Ba, 2014] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Proceedings of ICLR, 2014.

[Li et al., 2016] Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. A diversity-promoting objective function for neural conversation models. In Proceedings of NAACL, pages 110–119, 2016.

[Li et al., 2017] Jiwei Li, Will Monroe, Tianlin Shi, Sébastien Jean, Alan Ritter, and Dan Jurafsky. Adversarial learning for neural dialogue generation. In Proceedings of EMNLP, pages 2157–2169, 2017.

[Lin et al., 2017] Kevin Lin, Diaqi Li, Xiaodong He, Zhengyou Zhang, and Ming-Ting Sun. Adversarial ranking for language generation. In Proceedings of NeurIPS, pages 3155–3165, 2017.

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

[Liu et al., 2018] Yuhui Liu, Wei Bi, Jun Gao, Xiaojiang Liu, Jian Yao, and Shuming Shi. Towards less generic responses in neural conversation models: A statistical re-weighting method. In Proceedings of EMNLP, pages 2769–2774, 2018.

[Mihalcea and Tarau, 2004] Rada Mihalcea and Paul Tarau. TextRank: Bringing order into text. In Proceedings of EMNLP, pages 404–411, 2004.

[Mou et al., 2016] Lili Mou, Yiping Song, Rui Yan, Ge Li, Lu Zhang, and Zhi Jin. Sequence to backward and forward sequences: A content-introducing approach to generative short-text conversation. In Proceedings of COLING, pages 3349–3358, 2016.

[Pei and Li, 2018] Xiaojin Pei and Chenchang Li. S2SPMN: A simple and effective framework for response generation with relevant information. In Proceedings of EMNLP, pages 745–750, 2018.

[Ritter et al., 2011] Alan Ritter, Colin Cherry, and William B. Dolan. Data-driven response generation in social media. In Proceedings of EMNLP, pages 583–593, 2011.

[Serban et al., 2016] Iulian V Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, and Joelle Pineau. Building end-to-end dialogue systems using generative hierarchical neural network models. In Proceedings of AAAI, pages 3776–3783, 2016.

[Shang et al., 2015] Lifeng Shang, Zhengdong Lu, and Hang Li. Neural responding machine for short-text conversation. In Proceedings of ACL, pages 1577–1586, 2015.

[Shang et al., 2018] Mingyue Shang, Zhenxin Fu, Nanyun Peng, Yansong Feng, Dongyan Zhao, and Rui Yan. Learning to converse with noisy data: generation with calibration. In Proceedings of IJCAI, pages 4338–4344, 2018.

[Shao et al., 2017] Yuanlong Shao, Stephan Gouws, Denny Britz, Anna Goldie, Brian Strope, and Ray Kurzweil. Generating high-quality and informative conversation responses with sequence-to-sequence models. In Proceedings of EMNLP, pages 2210–2219, 2017.

[Sutskever et al., 2014] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In Proceedings of NeurIPS, pages 3104–3112, 2014.

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

[Tian et al., 2020] Zhiliang Tian, Wei Bi, Dongkyu Lee, Lanqing Xue, Yiping Song, Xiaojiang Liu, and Nevin L. Zhang. Response-anticipated memory for on-demand knowledge integration in response generation. In Proceedings of ACL, pages 650–659, 2020.

[Vinyals and Le, 2015] Oriol Vinyals and Quoc Le. A neural conversational model. arXiv preprint arXiv:1506.05869, 2015.

[Wang et al., 2018] Wenjie Wang, Minlie Huang, Xin-Shun Xu, Fumin Shen, and Liqiang Nie. Chat more: Deepening and widening the chatting topic via a deep model. In Proceedings of SIGIR, pages 255–264, 2018.

[Wojciechowski and Zakrzewicz, 2002] Marek Wojciechowski and Maciej Zakrzewicz. Dataset filtering techniques in constraint-based frequent pattern mining. In Pattern detection and discovery, pages 77–91. Springer, 2002.

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

[Xu et al., 2017] Zhen Xu, Bingquan Liu, Baoxun Wang, Chengjie Sun, Xiaolong Wang, Zhaoran Wang, and Chao Qi. Neural response generation via GAN with an approximate embedding layer. In Proceedings of EMNLP, pages 617–626, 2017.

[Xu et al., 2018] Xinnuo Xu, Ondˇrej Duˇsek, Ioannis Konstas, and Verena Rieser. Better conversations by modeling, filtering, and optimizing for coherence and diversity. In Proceedings of IJCAI, pages 3981–3991, 2018.

[Yu et al., 2017] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. Seqgan: Sequence generative adversarial nets with policy gradient. In Proceedings of AAAI, pages 2852–2858, 2017.

[Zhang et al., 2020] Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. DIALOGPT: Large-scale generative pre-training for conversational response generation. In Proceedings of ACL: System Demonstrations, pages 270–278, 2020.
[Zhou et al., 2017] Ganbin Zhou, Ping Luo, Rongyu Cao, Fen Lin, Bo Chen, and Qing He. Mechanism-aware neural machine for dialogue response generation. In Proceedings of AAAI, pages 3400–3406, 2017.

[Zhou et al., 2018] Ganbin Zhou, Ping Luo, Yijun Xiao, Fen Lin, Bo Chen, and Qing He. Elastic responding machine for dialog generation with dynamically mechanism selecting. In Proceedings of AAAI, 2018.