Learning to Deliberate: Multi-pass Decoding for Document-Grounded Conversations

1st Junyan Qiu
University of Chinese Academy of Sciences
Institute of Automation, Chinese Academy of Sciences
Beijing, China
qiujunyan2018@ia.ac.cn

2nd Yiping Yang
Institute of Automation, Chinese Academy of Sciences
Beijing, China
yiping.yang@ia.ac.cn

***Abstract*—Document-grounded conversations are designed to generate and engage in conversations based on specific documents or texts provided as context. The ability to incorporate documents into these conversations enables a deeper understanding of the subject matter, fostering more informed and meaningful discussions. However, prior approaches were predominantly rooted in auto-regressive models, overlooking the need for a comprehensive global perspective and the refinement of responses. In this paper, we introduce an innovative Multi-Pass Decoding (MPD) architecture, which iteratively updates background knowledge and enhances responses in document-grounded conversations. During each iteration, it starts by adaptively combining semantics derived from the context, documents, and previous responses. To address the issue of inadequate response quality, we have also developed two modules dedicated to identifying and refining inappropriate words or phrases in responses generated during the previous iteration. Furthermore, MPD is model-agnostic, enabling seamless integration with conventional sequence-to-sequence frameworks. Our empirical experiments on three document-grounded conversation datasets demonstrate that our methods facilitate the production of more contextually accurate and coherent responses.**

***Index Terms*—dialogue system, document-grounded conversations, deliberation network, sequence-to-sequence framework**

I. Introduction

Document-grounded conversations refer to a conversational AI paradigm where the dialogue system leverages external documents, such as articles, websites, or reference materials, as a primary source of information and context during interactions [1, 2]. In recent years, we have witnessed rapid advancements in dialogue systems. Many of these models are trained using the sequence-to-sequence framework in an end-to-end fashion, utilizing extensive datasets of human-to-human dialogues, and have achieved remarkable success.

However, there remains a significant journey ahead in realizing the ultimate objective of dialogue systems, which is the capability to engage in conversations that mimic human-like fluency. A pivotal aspect of attaining this goal hinges on the seamless integration of pertinent background knowledge in alignment with the context and the response to be generated [3, 1].

Existing techniques for incorporating background knowledge can be categorized into two primary approaches: knowledge selection and reasoning [4]. The former involves the process of identifying and selecting relevant information from external documents or sources that pertain to the ongoing conversation [3, 2]. Conversely, the latter endeavors to construct an interpretable path of reasoning through the evidence within the document [5]. However, these approaches are limited by their reliance on one-pass decoding, which can result in the loss of global information due to the auto-regression paradigm. In simpler terms, during the generation of each word, the model can only consider the words it has generated thus far and not those that will follow in the future [6]. Furthermore, they often fall short in fully harnessing background knowledge without further refinement [7].

Inspired by the cognitive processes inherent in human communication, some researchers have proposed deliberation networks [6] equipped with a refining mechanism, which reevaluates and potentially enhances the initial responses before delivering the final reply to the user. For example, [7] devised a two-pass decoder to enhance context coherence and ensure the accurate integration of knowledge. Nevertheless, these approaches encounter challenges in the identification of inappropriate words, potentially resulting in the retention of erroneous terms while replacing the correct ones during the deliberation process.

In this paper, we introduce a novel multi-pass decoding (MPD) architecture to dynamically incorporate more comprehensive background knowledge and enhance response refinement. On the one hand, MPD adaptively updates knowledge based on the context and previously generated responses to capture global information and maintain knowledge relevance. On the other hand, MPD refines the responses by identifying and addressing issues related to language fluency, context coherence, and factual correctness. This iterative process ensures that the generated responses not only remain contextually relevant but also exhibit improved linguistic quality and accuracy. MDP is a model-agnostic architecture, making it readily adaptable and integrable with a wide array of sequence-to-sequence frameworks.

Furthermore, we have observed that different responses may necessitate varying degrees of complexity to refine, with more challenging responses often requiring additional iterations for improvement. For instance, responding to a question like "How do you like *Avengers*?" is more intricate than addressing

a simpler query such as "Do you like *Avengers*?". Thus, we incorporate curriculum learning into the training stage, which gradually increases the complexity of the training examples presented to the model. This structured learning process helps the model develop skills to generate coherent responses across various user queries.

To conclude, our contributions are in three-fold: (1) We propose a novel multi-pass decoding (MPD) architecture to generate more coherent and informative responses. To our knowledge, this is the first attempt that introduces iterative knowledge updates, as well as the identification and correction of erroneous responses within the domain of document-grounded conversations. (2) MPD offers universal integration with existing sequence-to-sequence frameworks, ensuring seamless cooperation and compatibility, thereby enhancing the versatility of the system. (3) Extensive experimental results demonstrate that the proposed multi-pass decoding architecture significantly enhances response coherence and informativeness.

II. Related work

Our work is closely related to the field of deliberation networks, which draws inspiration from common human behaviors in everyday text creation and comprehension. In the process of writing or reading, humans often engage in iterative thinking, editing, and refinement to ensure text accuracy, fluency, and informativeness. Initially, this technology was employed to boost the performance of non-autoregressive machine translation models. Autoregressive models typically generate translation results word by word, resulting in slower generation. Non-autoregressive models can generate entire sequences at once, albeit at the cost of translation accuracy. [8] proposed to progressively improve the generated translation results through multiple iterations. Each iteration enhances the generated output, ultimately improving translation quality while maintaining speed. Subsequently, [7] integrated the deliberation network into dialogue generation, introducing an incremental transformer with two-pass decoders to enhance context coherence and ensure knowledge accuracy. However, these previous works faced challenges in accurately identifying erroneous words. Consequently, [9] made further advancements by introducing a locator to identify incorrect words and integrating a reviser into the deliberation process. This innovation ensures the correction of erroneous words while preserving the correctness of others.

III. Methodology

As shown in Figure 1, we will introduce the knowledge enhanced seq2seq framework, LTD (Learning To Deliberate) module and model training in this section.

A. Knowledge Enhanced Seq2Seq Framework

Given the dialogue context $\mathcal{C}$ and background documents $\mathcal{D}$, the knowledge-enhanced Seq2Seq framework decomposes the distribution over potential output sentence $\mathcal{R} = \{\tilde{y}_1, \tilde{y}_2, \cdots, \tilde{y}_{\ell_r}\}$ into a chain of conditional probabilities with a left-to-right causal structure:

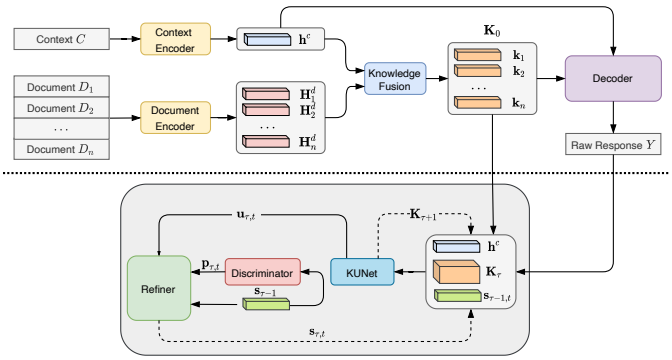

Fig. 1. Overview architecture of MPD. It consists of a knowledge enhanced seq2seq framework (top) and a LTD module (bottom).

$$p(\mathcal{R}|\mathcal{C},\mathcal{D}) = \prod_{i=1}^{\ell_r} p(\tilde{y}_i|\mathcal{C},\mathcal{D},\tilde{y}_1,\cdots,\tilde{y}_{i-1}) \tag{1}$$

Specifically, we firstly employ two encoders to convert the dialogue context and background documents into hidden vectors, $\mathbf{h}_c$ and $\mathbf{H}_d^i$, where $\mathbf{h}_c$ is a fixed-size vector representing the whole dialogue context, $\mathbf{H}_d^i = \{\mathbf{h}_1^i, \mathbf{h}_2^i, \cdots, \mathbf{h}_{\ell_i}^i\}$ is a word vector matrix corresponding to the i-th document. The two encoders can be implemented by GRU, Transformer [10] or pretrained language models (e.g., T5 encoder [11]). Then we use $\mathbf{h}_c$ to guide attention towards $\mathbf{H}_d^i$ to generate contextual representation $\mathbf{k}_i$ for each document:

$$\mathbf{k}_i = \sum_{j=1}^{\ell_i} \alpha_j^i \cdot \mathbf{h}_j^i \tag{2}$$

$$\alpha_j^i = \frac{\exp(\mathbf{h}_c \cdot \mathbf{h}_j^i)}{\sum_{u=1}^{\ell^i} \exp(\mathbf{h}_c \cdot \mathbf{h}_u^i)} \tag{3}$$

In the decodig process, we utilize the context vector $\mathbf{h}_c$ as the decoder's start token to avoid cold start problem at the initial time step when there is no previous output word [12]:

$$\mathbf{s}_t = \text{GRU}(y_{t-1}, \mathbf{c}_{t-1}) \tag{4}$$

$$\mathbf{c}_{t-1} = \sum_{i=1}^{k} \alpha_i \cdot \mathbf{k}_i \tag{5}$$

$$\alpha_i = \frac{\exp(\mathbf{s}_{t-1} \cdot \mathbf{k}_i)}{\sum_{j=1}^{n} \exp(\mathbf{s}_{t-1} \cdot \mathbf{k}_j)} \tag{6}$$

$$\tilde{y}_t = \arg\max(softmax(\mathbf{W}_V^T \mathbf{s}_t + \mathbf{b}_V)) \tag{7}$$

where $\mathbf{W}_V$ and $\mathbf{b}_V$ are trainable parameters used to map hidden states into probabilities over the vocabulary.

B. LTD Module

The LTD module is an iterative decoder that can be seamlessly integrated into a standard Seq2Seq framework. Its primary objective is to improve the quality of the raw

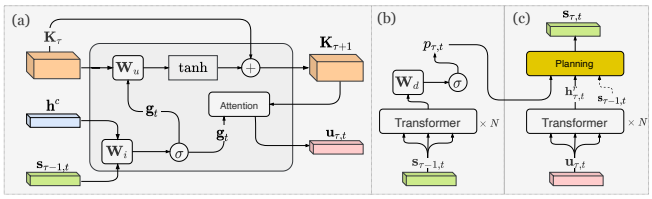

Fig. 2. The illustration of MDP, which is composed of (a) KUNet, (b) Discriminator and (c) Refiner.

responses generated by the encoder-decoder generative model within a maximum of K iterations. In the τ-th iteration, LTD takes the state vector $\mathbf{s}_{\tau-1,t}$ generated in the previous iteration and updates it to $\mathbf{s}_{\tau,t}$. Each word y_t in the response Y corresponds to a state vector $\mathbf{s}_{\tau,t}$, which is utilized to predict the response $\tilde{y}_{\tau,t}$ in the current iteration, as illustrated in Equation 7. The initial state vector $\mathbf{s}_{0,t}$ for the first iteration is initialized with the word embedding of the original response $\tilde{y}_t$.

As depicted in Figure 2, the LTD comprises three attention-based modules: (a) a Knowledge Update Network (KUNet) responsible for updating background knowledge based on the context and the response generated in the previous iteration, (b) a Discriminator tasked with identifying inappropriate words or phrases in the previously generated response, and (c) a Refiner that corrects the erroneous words pinpointed by the Discriminator.

a) KUNet: When revising an academic paper, the author needs to consider the semantic information throughout the entire text and update relevant knowledge based on the modifications made. This ensures that the edited content remains consistent with the overall focus of the paper. To this end, we introduce the KUNet for updating background knowledge. Formally, at the τ-th iteration, KUNet firstly obtains the global information, denoted as $\mathbf{g}_{\tau,t}$, by merging the context vector $\mathbf{h}^c$ and the previous decoding output state vector $\mathbf{s}_{\tau-1,t}$. $\mathbf{g}_{\tau,t}$ can be regarded as a synthesized representation of the combined contextual and response information, which serves as a foundation for subsequent knowledge refinement and adaptation within the model. Subsequently, $\mathbf{g}_{\tau,t}$ is utilzed to update the background knowledge information $\mathbf{K}_{\tau+1} = \{\mathbf{k}_{\tau+1,2}, \mathbf{k}_{\tau+1,2}, \cdots, \mathbf{k}_{\tau+1,n}\}$:

$$\mathbf{g}_{\tau,t} = \tanh(\mathbf{W}_i^T([\mathbf{h}^c, \mathbf{s}_{\tau-1,t}] + \mathbf{b}_i)) \tag{8}$$

$$\mathbf{k}_{\tau+1,i} = (1 + \alpha_i)\mathbf{k}_{\tau,i} \tag{9}$$

$$\alpha_i = \tanh(\mathbf{W}_u^T([\mathbf{g}, \mathbf{k}_{\tau,i}]) + \mathbf{b}_u) \tag{10}$$

where $\mathbf{W}_i$,$\mathbf{b}_i$, $\mathbf{W}_u$,$\mathbf{b}_u$ are trainable parameters, $[\cdot,\cdot]$ denotes the concatenation operation. Finally, we utilize $\mathbf{g}_t$ to calculate attention over the updated knowledge information $\mathbf{K}_{\tau+1}$, resulting in the intermediate word-level response representation $\mathbf{u}_{\tau,t}$:

$$\mathbf{u}_{\tau,t} = \sum_{i=1}^{n} \beta_i \mathbf{k}_{\tau+1,i} \tag{11}$$

$$\beta_i = \frac{\exp(\mathbf{g} \cdot \mathbf{k}_{\tau+1,i})}{\sum_{j=1}^{n} \exp(\mathbf{g}_t \cdot \mathbf{k}_{\tau+1,j})} \tag{12}$$

The KUNet effectively filters out irrelevant or redundant knowledge, resulting in a more concise and refined knowledge base. Consequently, it ensures that the background knowledge used to reply remains aligned with the ongoing conversation. This, in turn, facilitates the generation of responses that are not only more relevant but also enriched with information.

b) Discriminator: The purpose of the discriminator is to identify incorrect words within the entire sentence. Specifically, it categorizes each word into two types: *revise* or *retain*. The revise category indicates that the word is inappropriate, possibly not in alignment with the context, or may contain errors, necessitating further examination and potential correction. On the other hand, the retain category designates that the word is deemed correct, suitable, or harmonious with the surrounding context, obviating the need for any modification.

Given the state vectors $\mathbf{s}_{\tau-1,1\sim\ell_y}$ generated in the previous iteration as input, we employ N layers of multi-head attention [10] blocks to encode them into hidden layer vectors $\mathbf{H}^d = [\mathbf{h}^d_{\tau,1}, \mathbf{h}^d_{\tau,2}, \cdots, \mathbf{h}^d_{\tau,\ell_y}]$ for capturing their contextual semantic information. Subsequently, a classifier is meticulously crafted to determine the category of each word, i.e., whether it should be revised or retained:

$$p_{\tau,t} = \text{sigmoid}(\mathbf{W}_d^T \mathbf{h}^d_{\tau,t} + \mathbf{b}_{\tau,t}) \tag{13}$$

where $\mathbf{W}_d$ and $\mathbf{b}_d$ are trainable parameters, $p_{\tau,t} \in [0, 1]$ indicates the likelihood of the word being classified as *revise*.

c) Refiner: The refiner is designed to make modifications to the words identified as *revise* in the deliberate process. Similarly, N layers of multi-head attention [10] blocks are employed to encode $\mathbf{u}_{\tau,1\sim\ell_y}$ into hidden layer vectors $\mathbf{H}^r = [\mathbf{h}^r_{\tau,1}, \mathbf{h}^r_{\tau,2}, \cdots, \mathbf{h}^r_{\tau,\ell_y}]$. Each vector can be regarded as the word-level response representation that integrates updated contextual and background knowledge information. [7] predicted the current round's response solely based on the output of the previous round's decoder (i.e., $\mathbf{H}^r$). Nevertheless, it may face challenges in discerning between accurate and erroneous words from the preceding decoder output, potentially resulting in alterations to accurate words while preserving the erroneous ones. To address that issue, we introduce a planning approach to enhance the precision of word-level correction in the Refiner. Formally, it performs an adaptive integration of $\mathbf{s}_{\tau-1,t}$ and $\mathbf{h}_{\tau,t}$ based on the result of the discriminator:

$$\mathbf{h}^o_t = p_{\tau,t} \cdot \mathbf{h}^r_s + (1 - p_{\tau,t}) \cdot \mathbf{s}_{\tau-1,t} \tag{14}$$

$$\mathbf{s}_{\tau,t} = \tanh(\mathbf{W}_o^T \mathbf{h}^o_t + \mathbf{b}_o^T) \tag{15}$$

where $\mathbf{W}_o$ and $\mathbf{b}_o$ are trainable parameters. The planning comprehensively incorporates information from both the current and previous iterations, blending them based on the discriminative outputs. It not only addresses glaring errors

but also implements subtle adjustments, even when it hasn't explicitly identified an issue with a specific word.

C. Model Training

We employ the cross entropy $\mathcal{H}$ loss between the model's predicted output $\tilde{y}_t$ and the true response label y_t to train the seq2seq framework, i.e., $\mathcal{L}_{seq2seq} = \mathcal{H}(\tilde{y}_t, y_t)$. To facilitate training the multi-pass decoding network, we retain the output representation of the discrimintor $p_{\tau,t}$ and that of the refiner $\mathbf{s}_{\tau,t}$ for each iteration. Then we construct two types of supervised signals to instruct the training of these two modules, specifically, for the discriminator, we calculate the mean squared error (MSE) between the predicted probability $p_{\tau,t}$ and true label $y^d_{\tau,t}$ as the training objective:

$$\mathcal{L}_d = \frac{1}{k \cdot \ell_y} \sum_{\tau=1}^{k} \sum_{t=1}^{\ell_y} (p_{\tau,t} - y^d_{\tau,t})^2 \tag{16}$$

where $y^d_{\tau,t}$ is obtained by comparing the previously generated token $\tilde{y}_{\tau,t}$ and ground truth $y_{\tau,t}$:

$$y^d_{\tau,t} = \begin{cases} 1 & \tilde{y}_\tau = y_t \\ 0 & \tilde{y}_\tau \neq y_t \end{cases} \tag{17}$$

Similar to the seq2seq framework, the training objective of the refiner is defined as the cross-entropy between the predicted response and the ground truth:

$$\mathcal{L}_r = \frac{1}{k \cdot \ell_y} \sum_{\tau=1}^{k} \sum_{t=1}^{\ell_y} (-y_t \log(\mathbf{p}_{\tau,t})) \tag{18}$$

$$\mathbf{p}_{\tau,t} = \text{softmax}(\mathbf{W}_V^T \mathbf{s}_{\tau,t} + \mathbf{b}_V) \tag{19}$$

where k in equation (16) and (19) is the number of iterations. During the τ-th iteration, if the discriminator determines that all words in the previously generated response are correct (i.e., $\forall t \in [1, \ell_y]$, $p_{\tau,t} < 0.5$), the iteration process terminates, and we take the output for that iteration as the model's final output, with $k = \tau$. If the termination condition is not met even after reaching the maximum number of iterations K, the iteration process is forcefully terminated, setting $k = K$. The training loss is the sum of these three components:

$$\mathcal{L} = \mathcal{L}_{seq2seq} + \mathcal{L}_d + \mathcal{L}_r \tag{20}$$

Furthermore, we employ curriculum learing [13] to train the model to improve performance and convergence rate [14]. Formally, $\mathcal{L}_v = v_i \cdot \mathcal{L}$, where $v_i = \mathbb{I}(k \leq \lambda)$. If k is less than the threshold λ, the sample is included as a training sample; otherwise, it is excluded. In this paper, the value of λ is linearly increased as training progresses. Considering that in the early stages of training, the model's capacity is relatively limited, and the number of iterations is typically higher, setting a smaller threshold may result in fewer samples participating in training. Therefore, we introduce a warm-up training phase in which the threshold is set to the maximum number of iterations K, ensuring that all samples are included in training during this phase. In summary, for a given current training step t and the total number of training steps T, the definition of λ is as follows:

$$\lambda = \begin{cases} K & t \leq t_0 \\ K \cdot (\lambda_0 + \dfrac{1 - \lambda_0}{T} \cdot t) & t > t_0 \end{cases} \tag{21}$$

where λ_0 defines the initial value of λ. After conducting multiple tests, we set λ_0 to 0.3.

IV. Experiment

A. Datasets

We conduct experiments on three publicly avaliable document-grounded conversation datasets, Wizard-of-Wikipedia (WoW) [15], CMU_DoG [16] and KdConv [17]. WoW is built around conversations that are grounded in Wikipedia articles, which covers a wide range of topics, reflecting the diversity of Wikipedia articles. CMU_DoG involves two interlocutors, with one participant selecting a topic from 30 movie-related Wikipedia documents and steering the conversation around that document. KdConv focuses on knowledge-driven conversation modeling in the context of Chinese dialogue, where both dialogue participants have access to the knowledge graph during the conversation.

B. Baselines

We compare our methods with the following baseline models: **GRU:** an encoder-decoder architecture using GRU as the backbone with global attention [18]. **Transformer (Trans):** the standard paradigm for modeling long sequences based on multi-head attention [10]. **T5:** state-of-the-art pretrained language model that frames almost all NLP tasks as a text-to-text problem. **HRED:** a hierarchical encoder-decoder model that encodes dialogue context in both word level and utterance level [19]. **GPT2** [20]: a decoder-only Transformer architecture capable of generating human-like text in response to a given prompt. **DialoGPT:** an extension of the GPT architecture, specifically designed for engaging in natural language conversations [21]. Both GPT-2 and DialoGPT concatenate the document and context, separated by a special token, as their input.

C. Training Details

In this paper, our approach is integrated with various seq2seq-based models, including GRU, Transformer [10], and T5 [11], to showcase the model-agnostic superiority of MDP. Specifically, The GRU model is configured with a dimension of d_m = 512 and ℓ = 3 layers. For the transformer, the number of hidden nodes d_m is set to 768 and number of layers ℓ is set to 12. We employ the base version of T5 in this paper with d_m = 768, ℓ = 12, and the total number of parameters is 220M. The dropout rate is set to 0.1 for all these models. The maximum length for both dialogue history and external knowledge is constrained to 512 tokens, with any exceeding portions being truncated. The maximum length of response is set to 128.

TABLE I

AUTOMATIC EVALUATION RESULTS OF ALL COMPARED MODELS. ALL METRICS ARE IN %. BEST RESULTS ARE MARKED IN BOLD, AND SECOND BEST ARE UNDERLINED. YELLOW BACKGROUND NUMBERS INDICATE THAT THE BASELINE MODELS ARE SIGNIFICANTLY IMPROVED WHEN COMBINED WITH OUR METHOD.

Model	WoW					CMU_DoG					KdConv				
	PPL	BLEU-2/4		Dist-1/2		PPL	BLEU-2/4		Dist-1/2		PPL	BLEU-2/4		Dist-1/2	
GRU	19.61	6.54	2.98	6.77	14.57	22.45	7.81	2.41	13.13	28.14	**18.11**	**28.71**	7.77	4.11	7.81
Trans	17.81	6.72	3.56	6.89	17.15	18.37	9.87	2.56	11.41	35.51	19.01	27.91	7.98	4.41	8.13
T5	16.66	7.91	3.79	6.77	16.11	15.76	16.77	2.95	14.71	38.43	21.78	26.09	6.91	4.79	8.95
HRED	20.57	6.34	2.58	5.98	12.91	19.91	8.91	2.31	12.41	25.66	18.39	27.01	7.35	3.91	7.51
GPT2	18.64	7.45	3.11	6.98	18.93	17.34	15.48	2.45	14.55	28.99	21.24	25.91	6.41	4.45	8.41
DialoGPT	**15.14**	8.17	3.67	7.43	**20.91**	16.01	15.41	3.21	15.67	39.12	20.47	26.17	6.91	5.49	8.89
GRU+MPD	17.18	6.77	3.05	6.99	17.77	20.11	7.97	2.45	14.17	33.77	18.45	27.91	7.74	4.78	8.31
Trans+MPD	16.98	7.01	4.01	7.14	18.19	17.17	11.01	2.77	13.68	35.71	19.07	30.12	**9.08**	4.79	8.37
T5+MPD	15.41	**8.35**	**4.41**	**7.47**	19.97	**15.61**	**17.91**	**3.37**	**15.94**	**43.17**	20.31	28.19	7.25	**5.71**	**9.01**

We utilize Adam optimizer [22] to train the models. The learning rate for GRU and Transformer models anneals linearly in the range of $[1e-3, 1e-4]$ and $[1e-4, 1e-5]$ respectively. The learning rate for T5 is set to $1e-5$. Batch size is set to 64, 16 and 8 for GRU, Transformer and T5 respectively. All models are trained for 10 epochs, and training will be halted if the loss fails to decrease for 10 consecutive steps. In the MDP model, the number of layers for both the discriminator and corrector Transformers is configured as N = 1, and the maximum number of iterations is set at 5.

D. Evaluation Metrics

a) Automatic metrics: : We adopt widely used metrics to automatically evaluate the response generation performance, including perplexity (PPL), BLEU and Distinct (Dist) [23]. PPL quantifies how well a model can predict a sequence of words in a given text or language. Lower perplexity values indicate that the model is better at predicting the text, suggesting a better understanding of the language and context. BLEU calculates the precision of n-grams in the generated response compared to the golden response. In addition, we employ Distinct-n (Dist, n=1, 2), which is the ratio of unique n-grams among the generated responses, to evaluate the response diversity.

b) Human evaluation:: We randomly select 100 conversations from the test set for human evaluation in WoW and KdConv dataset. Five professional annotators are invited to assess the generated responses from three key perspectives: (1) Fluency (Flu.): Assessing the naturalness and grammatical correctness of the response; (2) Coherence (Coh.): Evaluating the response's alignment with the context and its ability to guide subsequent utterances; (3) Informativeness (Inf.): Gauging the extent to which the response provides valuable information. Each annotator is asked to rate the response on a scale of 1 to 5, with 5 indicating the highest quality and 1 representing the lowest. The final results are presented as the average score given by all annotators.

E. Experiment Results

Table I reports the automatic evaluation results. From the table, we have the following observations: (1) T5+MPD consistently outperforms other baselines or ranks as the second-best performer on WoW and CMU_DoG in terms of all automatic metrics. (2) When combined with MPD, all seq2seq generative models show improved performance across nearly all datasets. This underscores the effectiveness of our proposed method in improving the coherence and informativeness of generated responses. (3) Comparing the performance of three seq2seq generative models, typically T5 outperforms Transformer and GRU in WoW and CMU_DoG datasets. However, in KdConv, there is a significant drop in PPL and BLEU scores for T5. This may be due to T5's pretraining on a large English corpus, which is less effective on Chinese datasets. A similar observation can be made for GPT2 and DialoGPT.

Human evaluation results are presented in Table II. The results clearly demonstrate that T5+MPD excels, particularly in terms of fluency and coherence, on the WoW dataset, surpassing the second-best results by 0.22 and 0.27, respectively. For the KdConv dataset, Trans+MPD generally delivers the best performance. Overall, MPD significantly enhances the performance of seq2seq models, aligning with our findings from automatic evaluation.

TABLE II

HUMAN EVALUATION RESULTS ON ALL WOW AND KDCONV DATASETS.

Model	WoW			KdConv		
	Flu.	Coh.	Inf.	Flu.	Coh.	Inf.
GRU	3.31	2.89	2.31	2.51	2.23	1.48
Trans	3.14	2.71	2.26	2.67	2.56	1.91
T5	3.61	3.04	2.91	2.63	2.31	1.71
GRU+MPD	3.41	2.71	2.81	2.54	2.49	1.76
Trans+MPD	3.65	2.49	2.51	2.71	**3.01**	**2.01**
T5+MPD	**3.87**	**3.31**	**2.98**	**2.72**	2.57	1.89

F. Ablation Study

To assess the effectiveness of each module within the proposed MPD, we conducted ablation experiments on the WoW test set. In particular, upon removal of the KUNet, the input to the refiner shifts from $\mathbf{u}_{\tau,t}$ to $\mathbf{s}_{\tau-1,t}$. Similarly, with the discriminator removed, the refiner exclusively relies on the output of the refiner to revise the response. Formally, Equation 15 becomes $\mathbf{s}_{\tau,t} = \tanh(\mathbf{W}_o^T \mathbf{s}_{\tau-1,t} + \mathbf{b}_o^T)$.

TABLE III
ABLATION STUDY OF MPD ON THE WOW DATASET. 'W/O' IS AN ABBREVIATION FOR 'WITHOUT,' INDICATING THE REMOVAL OF THE CORRESPONDING MODULE. 'KUN.,' 'DIS.,' AND 'REF.' STAND FOR 'KUNET,' 'DISCRIMINATOR,' AND 'REFINER,' RESPECTIVELY. THE NUMBER IN THE BOTTOM-RIGHT CORNER INDICATES THE PERFORMANCE DECLINE IN COMPARISON TO MPD+T5.

Model	WoW		
	PPL	BLEU-4	Dist-2
MPD+T5	15.41	4.41	19.97
w/o CL.	$15.57_{(+0.18)}$	$4.31_{(-0.10)}$	$18.97_{(-1.00)}$
w/o Kun.	$15.79_{(+0.38)}$	$4.39_{(-0.02)}$	$18.77_{(-1.20)}$
w/o Dis.	$16.41_{(+1.00)}$	$4.13_{(-0.28)}$	$18.41_{(-1.58)}$
w/o Dis. & Ref.	$16.68_{(+1.27)}$	$3.90_{(-1.51)}$	$16.84_{(-3.13)}$
T5	$16.66_{(+1.25)}$	$3.79_{(-0.62)}$	$16.11_{(-3.86)}$

The results are presented in Table III. We observe a slight decline in performance after removing curriculum learning, which demonstrates its capability to aid the model in its gradual adaptation to more complex conversations, thereby enhancing its overall performance. Furthermore, when the refiner and discriminator are removed, a more substantial decline in performance is observed. This highlights their role in integrating knowledge and evaluating the generated responses, consequently improving the informativeness and fluency of the model's output. Notably, in the last two rows of the table, it is noticeable that removing more than one module results in a performance decline that is either greater or comparable to the cumulative decline caused by the three individual reductions mentioned above. This observation underscores the interdependence and collective impact of these modules, emphasizing their essential roles in the model's performance.

G. Case Study

Figure 3 lists some responses generated by different models alongside the corresponding reference responses (Gold). It is evident that T5+MPD consistently produces high-quality responses, outperforming other baselines in terms of context consistency and knowledge relevance. Specifically: (1) Transformer and GRU often generate generic and less informative responses, such as "Yes, it is" (Transformer in case 1) and "I don't know" (GRU in case 2). (2) Additionally, these two models tend to produce grammatically erroneous (GRU: "sorry sorry" in case 2), factually incorrect responses (In case 2, the Transformer mistakenly believed that the Batman movie came out in 2008) and repetitive pieces. (3) Comparatively, T5 excels in generating contextually relevant responses. Its ability to understand and incorporate knowledge from a wide range of sources makes it particularly adept at providing insightful and informative replies. (4) When combined with MPD, T5 is capable of producing more informative and engaging responses, as can be inferred from the underlined text.

Table IV presents the results from various iterations produced by GRU+MPD. It is evident that in the first few iterations, the model struggles to generate fluent and coherent responses. Specifically, the model tends to produce ungrammatical responses with frequent word repetitions, which is consistent with the experimental results discussed in previous sections. In contrast, responses generated in the last two iterations are notably more fluent and informative. Notably, in iteration 5, a grammatical error present in iteration 4, where "he" is used instead of "her," has been successfully rectified.

TABLE IV
RESPONSES GENERATED BY GRU+MPD IN MULTIPLE ITERATIONS ON CMU_DOG TEST SET.

Document	Cast: Gal Gadot as Diana Prince / Wonder Woman
Context	Hey there! What did you think of Wonder Woman?
Response	iter.=1: I think yes, yes, yes ... iter.=2: I think it it yes ... iter.=3: I think it is great, what about you yes yes ... iter.=4: I think it is great, I think Gal Gadot is great, I like <u>he</u> very much. iter.=5: I think it is great, Gal Gadot is beautiful, I like <u>her</u> very much.

H. Analysis of the Maximum Number of Iterations

Additionally, we are intrigued by the potential impact of different choices for K, which represents the maximum number of iterations, on the final results. In pursuit of this, we have conducted experiments by adjusting different values of K in the T5+MPD model on the WoW dataset. As illustrated in Figure 4, the performance generally improves as K increases from 1 to 5. This suggests that the quality of responses can be enhanced through multi-pass refinements. However, as K continues to increase, performance improvements become marginal, highlighting the need for a careful exploration of the optimal K value. Since larger values of K typically demand more time and computational resources. In this paper, we have set $K = 5$ to strike a balance between performance and efficiency.

I. Performance Analysis of Different Iterations

To evaluate whether and how multiple iterations can improve the quality of generated responses, we randomly selected 100 responses generated by GRU+MPD from the WoW test set, and assess the following metrics during multi-pass decoding: the percentage of consecutive repetitive words (rep.)[24], the length of the generated responses (len.) and the BERTScore[25] in relation to the number of iterations (bs.).

The results are presented in Figure 5. As can be seen from the figure, the word repetition decreases, and BERTScore increases as the decoding iteration continues. This suggests

Case 1

Context
u_1: Hello! how did you like the movie home alone?

Response (u_2):
GRU: I think think this is good good good ...
Trans: I like it very much very much ...
T5: I think this movie is funny.
T5+MPD: I think this is a funny movie, what about you?
Gold: It was entertaining, funny and moving. How did you like it?

Response (u_1):
GRU: I like like it to to ...
Trans: I think so, and you?
T5: I think it was entertaining.
T5+MPD: I think so, and I am a fan of Kevin McCallister
Gold: I though Culkin did a nice job as Kevin McCallister. It was entertaining.

Case 2

Context:
u1: Hello! Have you ever seen Batman Begins?
u2: I'm not sure, is that the one that came out in 2008?

Response (u_1):
GRU: Yes, it is not, do you like it it ...?
Trans: Yes, it is a super hero film based on the DC comics ...
T5: Christian Bale stars Batman, Michael Caine stars Alfred Pennyworth, Liam Neeson stars Henri Ducard ...
T5+MPD: No, it's a 2005 superhero film starring Christian Bale as Batman. I am a big fan.
Gold: No, it actually came out in 2005.
----------MULTIPLE TURNS----------
Context:
u_2: Oh, then I definitely didn't see it. Who's in it?

Response (u_1):
GRU: Sorry sorry, I don't know, do you like like like ...
Trans: Christopher Nolan
T5: Christian Bale stars Batman, Michael Caine stars Alfred Pennyworth, Liam Neeson stars Henri Ducard ...
T5+MPD: Christopher Nolan has directed this film, Christian Bale stars Batman, Michael Caine stars ...
Gold: It was one of the Christopher Nolan films. Christian Bale plays Batman and there's also Michael Caine

Fig. 3. Responses generated by different models on CMU_DOG test set.

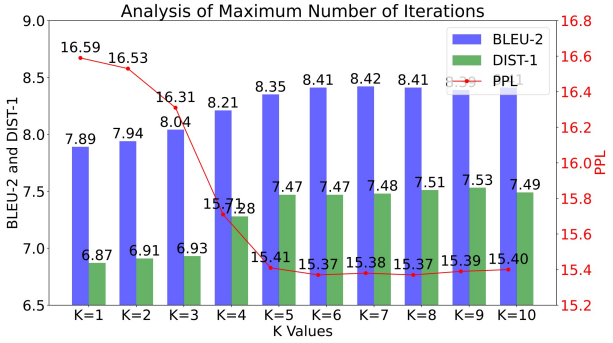

Fig. 4. Impact of the maximum number of iterations (K) on the Performance of T5+MPD.

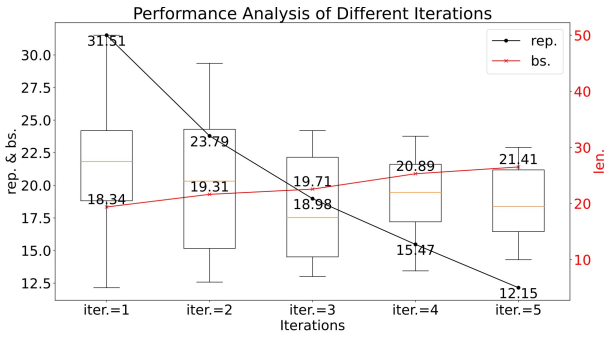

Fig. 5. Repetitions (rep.), BERTScore (bs.) and length (len.) variation with respect to iterations.

that the iterative decoding process effectively mitigates the issue of repetitive language patterns and enhances the overall semantic coherence of the generated responses. Conversely, a noticeable trend emerges in the distribution of response lengths as the iterations progress. The response lengths become progressively more tightly clustered. This can be attributed to the reduction of repetitive words and a more precise alignment between generated content and context.

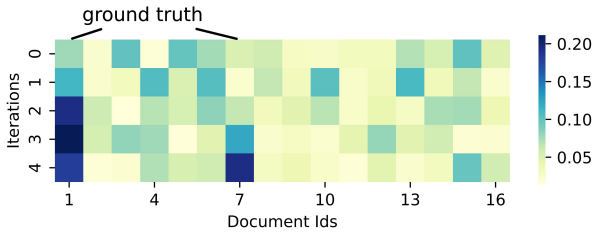

Fig. 6. Attention visualization for KUNet.

J. Visual Analysis

Figure 6 illustrates how the attention weights in KUNet change with respect to the related documents during the iterative processes. Specifically, these attention weights are determined by calculating $\gamma_i = \exp(1+\alpha_i)/\sum_j \exp(1+\alpha_j)$, where i and j represent document indices, and α is derived from Equation 10. A higher value signifies that the corresponding document is more relevant to the response. The ground truth is manually evaluated by determining which document is most related to the responses and context. As depicted in the figure, the model initially distributes attention equally among all the documents without differentiation. As the iterations progress, the model gradually shifts its focus to the correct document. This explains why our approach is capable of generating responses with higher knowledge relevance

V. Conclusion

In this paper, we introduced a novel Multi-Pass Decoding (MPD) architecture, iteratively conducting knowledge updates and the identification and refinement of erroneous responses. On the one hand, our method provides a versatile approach to iteratively enhance the quality of generated responses. By integrating this framework with various seq2seq models, we effectively address the challenges of generating coherent and contextually relevant text. Moreover, our approach incorporates a curriculum training strategy that further refines the model's performance during training. The results of extensive experiments indicate the substantial improvement over several seq2seq models.

REFERENCES

[1] B. Kim, D. Lee, D. Kim, H. Kim, S. Kim, O. Kwon, and H. Kim, "Generative model using knowledge graph for document-grounded conversations," *Applied Sciences*, vol. 12, no. 7, p. 3367, 2022.

[2] Y. Zhang, H. Fu, C. Fu, H. Yu, Y. Li, and C.-T. Nguyen, "Coarse-to-fine knowledge selection for document grounded dialogs," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[3] P. Ren, Z. Chen, C. Monz, J. Ma, and M. de Rijke, "Thinking globally, acting locally: Distantly supervised global-to-local knowledge selection for background based conversation," *arXiv preprint arXiv:1908.09528*, 2019.

[4] L. Ma, W.-N. Zhang, M. Li, and T. Liu, "A survey of document grounded dialogue systems (dgds)," *arXiv preprint arXiv:2004.13818*, 2020.

[5] Y. Zou, Z. Liu, X. Hu, and Q. Zhang, "Thinking clearly, talking fast: Concept-guided non-autoregressive generation for open-domain dialogue systems," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 2215–2226.

[6] Y. Xia, F. Tian, L. Wu, J. Lin, T. Qin, N. Yu, and T.-Y. Liu, "Deliberation networks: Sequence generation beyond one-pass decoding," *Advances in neural information processing systems*, vol. 30, 2017.

[7] Z. Li, C. Niu, F. Meng, Y. Feng, Q. Li, and J. Zhou, "Incremental transformer with deliberation decoder for document grounded conversations," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 12–21.

[8] J. Lee, E. Mansimov, and K. Cho, "Deterministic non-autoregressive neural sequence modeling by iterative refinement," *arXiv preprint arXiv:1802.06901*, 2018.

[9] X. Geng, X. Feng, and B. Qin, "Learning to rewrite for non-autoregressive neural machine translation," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 3297–3308.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[11] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, pp. 1–67, 2020.

[12] Q. Zhu, W. Zhang, and T. Liu, "Learning to start for sequence to sequence based response generation," in *Information Retrieval: 24th China Conference, CCIR 2018, Guilin, China, September 27–29, 2018, Proceedings 24*. Springer, 2018, pp. 274–285.

[13] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 41–48.

[14] X. Wang, Y. Chen, and W. Zhu, "A survey on curriculum learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[15] E. Dinan, S. Roller, K. Shuster, A. Fan, M. Auli, and J. Weston, "Wizard of wikipedia: Knowledge-powered conversational agents," in *International Conference on Learning Representations*, 2018.

[16] K. Zhou, S. Prabhumoye, and A. W. Black, "A dataset for document grounded conversations," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 708–713.

[17] H. Zhou, C. Zheng, K. Huang, M. Huang, and X. Zhu, "Kdconv: A chinese multi-domain dialogue dataset towards multi-turn knowledge-driven conversation," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 7098–7108.

[18] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 1412–1421.

[19] I. Serban, A. Sordoni, Y. Bengio, A. Courville, and J. Pineau, "Building end-to-end dialogue systems using generative hierarchical neural network models," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 30, no. 1, 2016.

[20] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.

[21] Y. Zhang, S. Sun, M. Galley, Y.-C. Chen, C. Brockett, X. Gao, J. Gao, J. Liu, and W. B. Dolan, "Dialogpt: Large-scale generative pre-training for conversational response generation," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2020, pp. 270–278.

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

[23] J. Li, M. Galley, C. Brockett, J. Gao, and W. B. Dolan, "A diversity-promoting objective function for neural conversation models," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 110–119.

[24] M. Ghazvininejad, O. Levy, Y. Liu, and L. Zettlemoyer, "Mask-predict: Parallel decoding of conditional masked language models," 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. 6112–6121.

[25] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "Bertscore: Evaluating text generation with bert," in *International Conference on Learning Representations*, 2019.