applied sciences

MDPI

Article

An Adaptive Contextual Relation Model for Improving Response Generation

Meiqi Wang †, Shiyu Tian †, Caixia Yuan and Xiaojie Wang *

School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100083, China; wmq@bupt.edu.cn (M.W.); tiansy@bupt.edu.cn (S.T.); yuancx@bupt.edu.cn (C.Y.)

* Correspondence: xjwang@bupt.edu.cn

† These authors contributed equally to this work.

Abstract: Context modeling has always been the groundwork for the dialogue response generation task, yet it presents challenges due to the loose context relations among open-domain dialogue sentences. Introducing simulated dialogue futures has been proposed as a solution to mitigate the problem of low history–response relevance. However, these approaches simply assume that the history and future of a dialogue have the same effect on response generation. In reality, the coherence between dialogue sentences varies, and thus, history and the future are not uniformly helpful in response prediction. Consequently, determining and leveraging the relevance between history–response and response–future to aid in response prediction emerges as a pivotal concern. This paper addresses this concern by initially establishing three context relations of response and its context (history and future), reflecting the relevance between the response and preceding and following sentences. Subsequently, we annotate response contextual relation labels on a large-scale dataset, DailyDialog (DD). Leveraging these relation labels, we propose a response generation model that adaptively integrates contributions from preceding and succeeding sentences guided by explicit relation labels. This approach mitigates the impact in cases of lower relevance and amplifies contributions in cases of higher relevance, thus improving the capability of context modeling. Experimental results on public dataset DD demonstrate that our response generation model significantly enhances coherence by 3.02% in long sequences (4-gram) and augments bi-gram diversity by 17.67%, surpassing the performance of previous models.

Keywords: dialogue generation; context modeling; context relation

check for updates

Citation: Wang, M.; Tian, S.; Yuan, C.; Wang, X. An Adaptive Contextual Relation Model for Improving Response Generation. *Appl. Sci.* **2024**, *14*, 3941. https://doi.org/10.3390/app14093941

Academic Editor: Jyrki Nummenmaa

Received: 22 March 2024
Revised: 19 April 2024
Accepted: 25 April 2024
Published: 6 May 2024

1. Introduction

Open-domain dialogue has been a long-standing issue in natural language processing, and the end-to-end response generation methods have attracted increasing attention recently [1–5]. The relationship between dialogue sentences in open-domain dialogues is loosely coupled; a dialogue history can lead to multiple different responses (i.e., one-to-many phenomenon), and the dialogue history often contains irrelevant noise. Therefore, the correlation between history and response is not always closely related, so predicting response based on history alone is insufficient.

The research on enhancing context relations is mainly divided into two categories: (1) reducing history noise or increasing future information, such as improving the importance of key semantic information by filtering dialogue history [6–8], or introducing the dialogue future to provide more context information [9–11]; and (2) exploring the internal text features of the context and using these features to assist context modeling, such as using keywords [12], topic words [13,14], and hierarchical information [15] to achieve better contextual modeling. Among them, text features are closely related to the dialogue context relations, so analyzing and utilizing the context is the focus of these works. However, especially in work that introduces the dialogue future, they simply assume that the history and future of dialogue have the same impact on response generation, but we find that they have different effects in different dialogue contexts. For example, in some inquiry scenarios, it is often in the form of question–answer, and the next question (future) is often not very relevant to the previous answer. Therefore, we propose that when using history and the future to simultaneously predict responses, the relevance between history–response and response–future should be given, which we call the response's context relation. Otherwise, introducing less relevant information into the response prediction process will affect the prediction results.

Therefore, to address the aforementioned problems and better utilize the dialogue context, we first research the dialogue context relations and then propose a response generation method to utilize these relations. Our work encompasses two key aspects: (1) Context relations analysis and annotation: We employ three consecutive sentences as the basic unit. Based on the binary correlation between history–response and response–future, we obtain three context relations of response. For example, the above-mentioned history–response relevance but response–future irrelevance in the question–answer scenarios is one of our three context relations. We then annotate the context relation labels on a popular dialogue dataset DailyDialog [16]. (2) Context relations utilization: Building upon the annotated context relations, we propose a dialogue generation model that adaptively integrates the information of dialogue contexts, enhancing contextual modeling and enriching response prediction. Specifically, our model first generates two responses based on dialogue history and simulated future and learns to adaptively fuse these two generated responses through annotated relation labels to produce the final response. The primary contributions of this paper encompass the following:

- Propose three dialogue context relations based on three consecutive dialogue sentences (history, response, and future) and annotate them on a large amount of data.
- Propose a response generation model that adaptively integrates the information from dialogue contexts guided by the dialogue context relation labels. To the best of our knowledge, we are the first to introduce dialogue context relations to assist context modeling for dialogue generation.
- The experimental results demonstrate that our model outperforms previous strong baseline results on the DD dataset, achieving better context modeling ability by introducing the dialogue context relations.

Our research focuses on open-domain dialogue, with the main aim of providing more appropriate replies in response generation tasks. Among them, the key problem we address is that the relations between dialogue sentences in chitchat are loose, and it is difficult to perform high-quality context modeling. Therefore, we propose explicit contextual sentence relations to illustrate the binary relevance between the response and its context, which provides a reference for the degree of context utilization in response prediction tasks.

The remainder of this paper is organized as follows. In Section 2, we introduce some related studies of open-domain dialogue generation models from two types. In Section 3, we describe the context relations analysis from the aspect of task statement, data description, and annotation. Section 4 demonstrates our proposed model to predict response based on the context relations in detail. Section 5 states the experimental setup of baselines and evaluation metrics. The experimental results and analyses are given in Section 6. Finally, we summarize the models and suggest potential ways to improve the performance of the models in Section 7.

2. Related Work

Dialogue response generation has always been hot research in the field of natural language processing [17,18]. According to the degree of utilization of the dialogue context, the existing response generation models can be divided into two categories: models that generate responses based only on history (His2Res) [19–21], and models that generate responses relied on both history and future information (His&Fut2Res) [3,10,11,22]. Among them, the His2Res methods focus on mining the text features (such as keywords, topic, and hierarchical information) in the dialogue history to assist in response generation. The His&Fut2Res methods are dedicated to providing more complete dialogue information by introducing the dialogue future (the dialogue sentence after the response). In Table 1, we compare the two types of models and show the focus of each model.

Table 1. Comparison of dialogue response generation models.

Type	Model	Feature/Future	Stage
His2Res	DAWnet	Feature (keywords)	train
His2Res	KnowHRL	Feature (topic)	train
His2Res	DKGT	Feature (topic)	train
His2Res	TA-Seq2Seq	Feature (topic)	train
His2Res	HiSA-GDS	Feature (hierarchy)	train
His2Res	HSAN	Feature (hierarchy)	train
His2Res	IEHSA	Feature (hierarchy)	train
His2Res	HHKS	Feature (hierarchy)	train
His&Fut2Res	Posterior-GAN	Dialogue Future	train
His&Fut2Res	RegDG	Dialogue Future	train
His&Fut2Res	Prophetchat	Dialogue Future	train, inference
His&Fut2Res	HDLD	Dialogue Future	train
His&Fut2Res	H-F Prompt	Dialogue Future	inference

2.1. His2Res Methods

Incorporating keywords proves to be a straightforward yet efficient approach for enhancing response generation. Ref. [12] presents the DAWnet model, which first explores the deep and wide keywords for the current dialogue and then utilizes these keywords to deepen and widen the chatting topics. The utilization of dialogue topics is also a common practice in dialogue generation methods. For example, KnowHRL [13] pre-plans a set of conversation topics as chat targets to promote knowledge matching and response generation. DKGT [23] proposes a dynamic knowledge graph-based topic conversation model, which utilizes a static graph attention mechanism to combine knowledge triplets in each dialogue sentence to predict the next chat topic. TA-Seq2Seq [14] focuses on transforming the conversation topic to assist in response prediction. Combining multiple levels of dialogue context can achieve better context modeling and also yields notable effectiveness in response generation tasks, such as HiSA-GDS, HSAN, IEHSA, HDID and HHKS [15,24–27]. For example, HiSA-GDS utilizes the word-level and sentence-level history successively to interact with responses. These methods gain a more appropriate response through the interaction of responses and multiple levels of history.

His2Res are currently a widely used response prediction method type, which can enhance the correlation between conversational sentences by adding different information. However, the relationship between history and response is often loosely coupled in open-domain dialogues, and one history can always correspond to different responses; thus, the introduced features always have little effect.

2.2. His&Fut2Res Methods

Modeling the full History–Response–Future context by the neural network is a direct and effective method, but there is only a small amount of follow-up work. For example, Posterior-GAN [10] and RegDG [11] help the model learn the relations among History–Response–Future through adversarial training and imitative learning, respectively, in the training phase. Prophetchat [3] designs a beam-search-like roll-out strategy for dialogue future simulation and provides the future sentence in both the training and inference phases. Ref. [26] believes that people can infer follow-up sentences (or tokens) based on the previous text, and vice versa. Therefore, they propose Hierarchical Duality Learning for Dialogue (HDLD), aiming to maximize the mutual information between past and future sentences. H-F Prompt [28] proposes a lightweight dialogue generation framework named few-shot history–future prompt that utilizes useful histories and simulated futures to generate more informative responses without the need for fine-tuning or adding extra parameters. However, we think these methods lack the consideration of the specific contextual relations.

His&Fut2Res is a recently proposed response prediction method that can effectively alleviate the low relevance of conversational sentences in open fields by introducing conversational futures. These methods of introducing future sentences either combine explicit conversational futures only in the training phase or add similar or simulated futures in the inference phase. However, the premise of these methods is that the dialogue history and dialogue future contribute equally to response prediction, but in fact, the correlations between history–response and response–future have various applicable situations, and the utilization of both would be insufficient if treated equally.

3. Context Relations Analysis

This section provides the statement of the problem we addressed, the description of the dataset, and the related data preparation steps.

3.1. Problem Statement

As mentioned in the introduction, in this paper, we aim to tackle a dialogue generation problem. Given the dialogue history of the two speakers, the goal is to predict the most appropriate reply based on a pre-trained model. A simple and widely used method is to predict responses based solely on history, but since the connection between responses and history is often loose in the open domain, it is often necessary to introduce more information to obtain a suitable reply, such as conversational future information. However, in the case of low-conversational sentences, more information is often less helpful, and in the case of high relevance, more information is more effective. Therefore, we must first clarify the correlation between dialogue sentences. In order to obtain the relationship of conversational sentences, we first consider the basic unit composition of sentence correlation and the types of correlation corresponding to the correlation combination.

In our context relations analysis, considering that (1) selecting multi-sentence contexts introduces analytical and computational complexity, and (2) predicting multiple future sentences by the dialogue model increases the inaccuracy and easily introduces noise, we choose three dialogue sentences to form a concise unit that is relatively enough to express a basic context without making the analysis too complicated.

In addition, although the correlation between two sentences is divided into strong and weak, the specific numerical value is difficult to obtain because (1) it is difficult to grade the correlation, (2) different people have different definitions of the strength of sentence relevance, and (3) it is difficult to train annotators to have a unified understanding of complex annotations. Moreover, since our goal is to find relevant and irrelevant aids for computational models, we believe that simple labels that ensure high accuracy are most appropriate. Therefore, we use binary relevance to analyze two dialogue utterances, this not only reduces the difficulty of annotating but also makes the modeling of the utterance relations more straightforward.

Based on the above analysis, we define the contextual relationship in the three sentences as follows.

For the T turns dialogue sequence $u_1, u_2, u_3, \ldots, u_T$, note any three consecutive turns u_{t-1}, u_t, u_{t+1} as H, R, F, denoting dialogue history, response, and future, respectively. We analyze the binary relation of H-R and the binary relation of R-F. Based on the relation between H-R and R-F, we can draw three response-centered context relations:

(1) relation1, H and R are related, R and F are unrelated;
(2) relation2, H and R are unrelated, R and F are related;
(3) relation3, H and R are related, R and F are related.

Note that there are very few cases where H-R is irrelevant and R-F is irrelevant, so we do not consider it.

3.2. Data Description

We mainly performed our research study on the DailyDialog [16] released for the open-domain dialogue task in 2017. DailyDialog [16] comes from websites related to English learners and is collected by crawling a large number of dialogue practice content from English learning websites. These dialogue data focus on several major topics and contain real-life daily dialogues, covering ten topics, including tourism, politics, finance, etc. The dialogue topics are more concentrated and more helpful for training dialogue models. In addition, the conversations on these websites are written by English learners and are more grammatically rigorous than the datasets constructed from Weibo, such as Twitter and Chinese Weibo. There are 13,118 dialogues in DailyDialog, with an average of 7.9 turns per dialogue. The training set contains 11,118 dialogues, the dev set contains 1000 dialogues and the test set contains 1000 dialogues. As a public open-domain dialogue dataset, this dataset is widely used as evaluation data for various dialogue models.

Secondly, since DD involves chatting without knowledge, the meaning of some conversational sentences is easily unclear. Therefore, to improve the accuracy of annotation, we also extract some data from the Wizard of Wikipedia (WoW) [21] to supplement knowledge-based conversations. WoW is a document-based open-domain conversation dataset proposed by FaceBook in 2019. WoW uses Wikipedia as a knowledge base and covers a wide range of topics (1365 in total). Each data sample has a selected topic, a conversation history, a basic factual knowledge sentence, and a corresponding conversation reply. Each round of dialogue includes two roles: mentor and apprentice. Both parties conduct in-depth exchanges on a certain topic. Among them, the apprentice does not obtain the document information in advance, and the tutor needs to pass the core content to the apprentice through dialogue. The entire dataset has more than 20,000 conversations and 5.4 million documents.

3.3. Data Annotation

Since our basic unit is three sentences, we first divide the DD and WoW data into samples composed of three sentences $[H, R, F]$. Considering that manual annotation is too expensive, we adopt a method that combines manual and automatic annotation. We first obtain a part of high-quality data labels through manual annotation, and then train an annotation model through the data and labels, and the model completes the annotation of the remaining data. Through this annotation method, we can quickly obtain annotation results on a large number of datasets. However, this annotation method also has limitations. Since the annotation model is trained based on manual labels, its highest accuracy is often lower than manual annotation. Therefore, we recommend adopting an ensemble way of multiple annotation models for automatic annotation tasks that require high accuracy. For example, in our task, we obtain two classification models from the perspectives of semantic classification and feature classification and then combine their results. This method can greatly improve automatic annotation.

3.3.1. Manual Annotation

We select 11 annotators with a foundation in natural language processing. According to the analysis in the problem statement section, there are 3 relationships, so the corresponding relationship labels are 1/2/3. The relation labels and corresponding relevance between sentences are shown in Table 2.

Table 2. Explanation of the three relation labels.

	H-R	**R-F**	**Relation Label**
relation1	relevant	irrelevant	1
relation2	irrelevant	relevant	2
relation3	relevant	relevant	3

The rules for determining relevance include two levels: one is the existence of keyword correspondence and co-reference at the word level, and the other is a question-and-answer form or topic coherence at the sentence level. If the requirements of these two levels are not met, the two sentences are judged to be irrelevant with high probability.

In addition, to annotate the context relations more accurately, we also annotate the three binary dialogue attributes in two consecutive dialogue sentences: keywords (two sentences have similar keywords), topic shift (topic changes in two sentences), and specific information (the two sentences with specific meaning).

Finally, based on the above annotation rules, we crowdsource the labels of relations and attributes on the DD and WoW datasets, obtaining 6300 labels; among them, 3300 are from the DD and 3000 are from the WoW. To verify the consistency of the manual annotation, we also randomly select 2300 items from them and give them to different people for secondary annotation. As a result, we find a 62.87% agreement rate for relation labels and a 78.43% agreement rate for attribute labels.

3.3.2. Automatic Annotation

Based on the above-mentioned high-quality manually labeled labels, we train a classification model that can classify the relationship types of three consecutive sentences. This model is used to label the relationship labels of the remaining unlabeled samples in DD. Since we have also annotated three dialogue attributes, we can not only classify relationships based on semantics from the perspective of dialogue sentences but also classify relations from the perspective of attributes. Thus, our classification model is an ensemble model, containing a semantics-based XLNet [29] and a feature-based decision tree.

XLNet is a BERT-like model that can classify input text. It adopts a general autoregressive pre-training method, and the training process is divided into two stages: the first stage is the language model pre-training stage, and the second stage is the task data fine-tuning stage. In our experiment, the XLNet classification model inputs three consecutive dialogue sentences and outputs the relation label corresponding to these sentences.

The decision tree is a common machine learning method that uses features for classification. Each internal node of the decision tree represents an attribute, each branch represents a judgment condition, and each leaf node represents a category. For each leaf node, majority voting is used for classification, that is, the category with the largest proportion in the node is selected as the predicted category of the node. In our experiment, the decision tree inputs an attribute string (like $[1,1,0,1,0,0]$) and outputs the corresponding relation label.

The accuracy rates of XLNet and the decision tree are 67.38% and 66.43% respectively, which shows that our relations are learnable and computable for the computational models, both from the perspective of dialogue sentences and dialogue attributes. The ensemble model's classification accuracy is 71.83%, and we label the model to annotate the unlabeled data. Among them, XLNet simulates the way humans distinguish different relationships and classifies relation categories based on three sentences. However, during the manual annotation process, we find that it is difficult for humans to distinguish different relationship categories, so the classification accuracy of XLNet is in line with expectations. The decision tree model takes dialogue attributes as input, which is equivalent to extracting the key features of sentences in the form of natural language. Compared with the original dialogue sentences, the key information extraction by humans makes the relation classification problem simpler, so the decision tree achieves better results. For the ensemble model that finally combines the two models, it integrates the advantages of the two. It not only integrates judgment from a semantic perspective but also provides classification results from a feature perspective, so it achieves the highest accuracy.

For the two models, the decision tree is a machine learning model. The model scale is very small and only takes 10 to 30 min to predict relations. The pre-trained model XLNet needs further training. Our manually labeled data are 6300, which is divided into a training set and a dataset at a ratio of 8:2. On this data, we use XLNet-base, which requires about 7G of memory and takes about 2 h of training time to achieve the fitting of the model.

3.3.3. Annotation Result

After manual and automatic annotation, the proportions of the three types of relation1/2/3 in DD data are 27.22%/18.15%/54.63%, respectively. Therefore, all of our proposed dialogue relations exist in real dialogue scenarios and can be distinguished from a human perspective. Among them, the proportion of relation1 is only 27.22%, which indicates that the existing methods of generating responses based only on the dialogue history cannot model most dialogue scenarios well, as they underestimate the importance of context in dialogues and do not take full advantage of it. The proportion of relation3 is 54.63%, which indicates that the current model using both contextual information can handle close to half of the dialogue scenarios, but there is still a lack of learning for relation2, which also has a not low proportion (18.15%). Neither of these two generation methods can fully cover various dialogue relations in real dialogues, which shows that the existing response generation methods do not fully consider the correspondence between responses and contexts, and there is still a big gap between the dialogue modeling and the real situation.

4. Dialogue Generation Model

Based on the above analysis of dialogue relations, we can find that most of the existing dialogue generation models predict response with history as input, which is consistent with the context relation in relation1. However, it is difficult to effectively simulate the context relations in other relations (approximately 72.78%). There is also a small amount of work that introduces dialogue future; however, these methods implicitly model the relations between contexts without specific explicit contextual relations as guidance, and often cannot model all the dialogue relations proposed in this paper.

Based on the above analysis, we believe that there are reasons to put forward the following hypotheses.

We hypothesize that without explicit context guidance, it is difficult for the model to learn the conversation history and future utilization of simulated conversations during training. If history and future information is fed directly into the model without emphasizing which part is more important for response prediction, the model will be confused between the two types of information and will not know which content to lean toward when predicting the response. If the model is affected by noise brought by low-correlation data, it will affect the model effect. Therefore, when predicting responses, it is very important to guide history–response and response–future correlations. Based on the above assumptions and analysis, we propose an explicit context relations-guided dialogue generation model, using different context relevance features in different relations as a bridge to guide the model to learn the contribution of each relation's context to the response, thereby enhancing the effectiveness of response generation.

By introducing an explicit contextual relationship guidance mechanism, our model can better understand and exploit various relations in conversations. This helps improve the effectiveness and quality of conversation generation, making the responses generated more coherent and relevant.

4.1. Overview

Notations. As in Section 3.1, the dialogue history, response, and future are still represented as H, R, and F, while the corresponding hidden vectors obtained in the model are h, r, and f. We use $Rel = (w_{hr}, w_{rf})$ to represent the annotated relevance labels, where w_{hr} is the relevance label of H-R, w_{rf} is the relevance label of R-F. In addition, the simulated response and future by the model are labeled by $*$, and the variables without $*$ are ground truth sentences. And we note the forward generator which predicts the response by history as G_F, and the generator which predicts the future by history as G_{F2}. The G_B is the generator predicting the response by using future information.

Model overview. This paper proposes a response generation model with adaptive context relations, which is composed of three components. The first component, Bidirectional Response Generation, predicts two hidden representations of the response with history and simulated future respectively, which represent the contribution of two sentences to the response. The second component, Bidirectional Relevance Learning, obtains the relevance strength of $[H, R]$ and $[R, F]$ with the annotated relevance labels as constraints, which represent the contextual relation of different dialogue relations. The third component Adaptive Response Generation adaptively fuses two representations of response by introducing the response–history and response–future relevance strength to obtain the final response. The goal of the model is to predict an appropriate response R^* under the conditions of given dialogue history H, simulated future F^*, and context relevance labels (Rel). Thus, the objective is:

$$\arg\max_{\theta} p(R^*|H, F^*, Rel, \theta) \tag{1}$$

where θ is the parameters of our model. The model framework is shown in Figure 1. BART-base [30] is used as the backbone of the encoder–decoder generators G_F, G_{F2} and G_B.

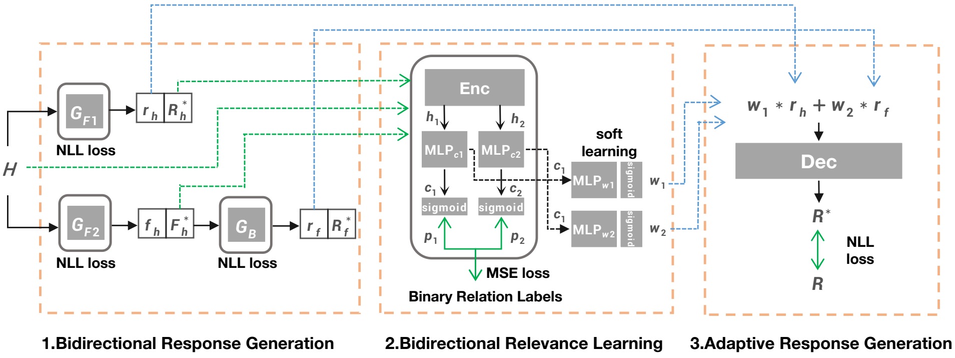


Figure 1. The overall framework of the dialogue generation model.

4.2. Bidirectional Response Generation

In this section, we consider generating the response from history2response and future2response directions. The forward response r_h is predicted by the history, and the backward response r_f is predicted by the simulated future.

Forward response generation. We use history H as input of G_F to predict the forward response:

$$r_h, {R_h}^* = G_F(H) \tag{2}$$

where r_h is the hidden vector of response and ${R_h}^*$ is the generated response sentence.

Backward response generation. Since the dialogue future is not available, we have to obtain a simulated future first and then predict the backward response using it.

We use a cross-step approach to simulate the future directly from the dialogue history:

$$f_h, {F_h}^* = G_{F2}(H) \tag{3}$$

where f_h is the hidden vector of future, and ${F_h}^*$ is the simulated future sentence.

Then, the backward response is obtained by generator G_B based on the hidden vector f_h of the future:

$$r_f, {R_f}^* = G_B(f_h) \tag{4}$$

where r_f is the hidden vector of response, and ${R_f}^*$ is the generated response sentence. All three generators G_F, G_{F2}, and G_B are constrained by the standard Negative Log Likelihood (NLL) loss with the gold response or gold future in the training phase.

4.3. Bidirectional Relevance Learning

This component learns the relevance of history–response and response–future through the representations H, ${R_h}^*$, and ${F_h}^*$. In order to simulate the strength of different contextual correlations in different relations, we use the annotated $0/1$ relevance w_{hr} and w_{rf} as the gold labels to constrain the model to learn the bidirectional relevance hidden representations c_1 and c_2 and then obtain the history–response relevance score w_1 and response–future relevance score w_2 from c_1 and c_2 through soft learning.

Forward relevance learning. The quality of ${R_h}^*$ generated from the gold history is better than ${R_f}^*$ generated from the simulated future in Section 4.2, so we use ${R_h}^*$ to represent the response in relevance learning. We first concatenate the history and response as the input of the BART-encoder (Enc in Figure 1) to obtain the hidden representation h_1:

$$h_1 = Enc([H, {R_h}^*]) \tag{5}$$

Then, h_1 is fed into multiple MLPs to obtain the 1-dimensional relevance hidden vector c_1 of history and response. The specific relevance value p_1 is obtained with sigmoid activation:

$$c_1 = MLP_{c_1}(h_1) \tag{6}$$

$$p_1 = sigmoid(c_1) \tag{7}$$

We constrain p_1 by the Mean Square Error (MSE) loss with the binary $0/1$ annotated relevance label w_{hr}, where N is the number of training examples:

$$Loss_{MSE} = \frac{\sum_1^N (w_{hr} - p_1)^2}{N} \tag{8}$$

To reduce the impact of noise in the annotation and error in the model learning, instead of using p_1 directly, the model learns the history–response relevance w_1 from the hidden vector c_1 with the soft learning method. We feed c_1 into MLPs and sigmoid to learn a soft relevance value w_1 of the history and response, then pass it to the next component of the model:

$$w_1 = sigmoid(MLP_{w_1}(c_1)) \tag{9}$$

Note that to guarantee the effect of relevance labels, the hidden vector c_1 is only constrained by MSE loss. The NLL loss in Section 4.4 is not back-propagated to c_1.

Backward relevance learning. Similar to the calculation process of w_1, we use $[{R_h}^*, {F_h}^*]$ as input to learn the relevance value w_2 of the response and future by the constraint of label w_{rf}.

4.4. Adaptive Response Generation

We obtain the bidirectional hidden vectors r_h and r_f in Section 4.2, and the bidirectional relevance w_1 and w_2 of the response–history and response–future in Section 4.3. In this section, we use them together to generate the final response adaptively.

We use w_1 and w_2 as weights to weighted r_h and r_f to obtain the final response hidden vector r_{hf}:

$$r_{hf} = w_1 * r_h + w_2 * r_f \tag{10}$$

We use a simple decoder Dec in Figure 1 as the final generator which consists of linear layers to predict the final response. We feed r_{hf} into Dec to map r_{hf} to the probability of the vocabulary and then obtain the generated response sentence R^*:

$$R^* = Dec(r_{hf}) \tag{11}$$

This simple response generator Dec is also constrained by NLL loss with the gold response in the training phase.

5. Experimental Setup

5.1. Baselines

The strong baseline models compared in this paper can be divided into two groups:

(1) Introducing a dialogue future during response generation.

NEXUS [9] introduces an auxiliary continuous code space and maps such a code space to a learnable prior distribution for estimating the future in the inference phase. This method introduces the dialogue future in the training stage, uses the model to implicitly learn the characteristics of the dialogue future, and uses the ability of the model to indirectly assist in response prediction during the inference stage.

RegDG [11] is a response generation model which utilizes gold future in the training phase. They use the teacher–student model to realize the utilization of future information, in which the teacher model predicts responses based on historical and future information, and the student model only has history as the input. During the training process, the student model is constrained to learn the teacher model's use of future information.

ProphetChat [3] is a DialoGPT-based model that utilizes the simulated future information in the response inference phase. Compared with the above two baseline models that introduce explicit dialogue futures in the training stage, this method can also introduce explicit dialogue futures in the inference stage by predicting the future in advance, improving the utilization efficiency of future information.

HDLD [26] uses hierarchical duality learning for dialogue to simulate human cognitive ability, estimating future information implicitly. Compared with ProphetChat, this method also models the dialogue prediction from the direction of the future to history, making more thorough use of history and future sentences.

(2) Classic models that only consider the dialogue history.

HRED [6] is a hierarchical neural network architecture model. HRED uses an Encoder RNN mainly to encode the input sentence. The middle layer context RNN is used to encode dialogue-level information such as the status and intention of the entire dialogue. The hidden layer of context RNN Vectors can remember previous dialogue information and become context vectors.

Transformer [31] is an encoder–decoder model based on the attention mechanism. Its encoder is used to encode the input dialogue history into context vectors, and the decoder decodes the context vectors to predict the sequence for the response. Each layer of the encoder and decoder consists of a multi-head self-attention layer and a fully connected feed-forward network layer. Compared with traditional RNN methods, the transformer has better position awareness and representation ability, and can capture long-distance dependencies.

D2GPO [32] adds a data-dependent Gaussian prior objective function to the maximum likelihood estimation objective function to enhance training. Compared with models based on maximum likelihood estimation, this method alleviates the suppression of maximum likelihood estimation in terms of diversity by introducing an extra Kullback–Leibler and makes effective use of a more detailed prior in the data.

AdaLabel [33] uses another decoder of bidirectional attention to dynamically estimate a token distribution at each time step. In order to alleviate the poor generation diversity problem, this method proposes an adaptive Label Smoothing (AdaLabel) approach that can adaptively estimate a target label distribution at each time step for different contexts, improving the diversity of response predictions.

iVAE$_{MI}$ [34] is a VAE-based model with regularization of maximizing mutual information. This method proposes sample-based representations of variational distributions for natural language, leading to implicit latent features, which can provide flexible representation power compared with Gaussian-based posteriors and can alleviate the "posterior collapse" issue.

PLATO [35] is a pre-trained dialogue generation model. PLATO $_{w/oLatent}$ with discrete latent variables is designed to tackle the inherent one-to-many mapping problem in response generation.

Dual [36] is a model that exploits abstract meaning representation to help dialogue modeling. Compared with the textual input, abstract meaning representation explicitly provides core semantic knowledge and reduces data sparsity. It shows the superiority of our model. Dual shows better performance in dialogue understanding and response generation tasks.

5.2. Evaluation Metrics

Automatic Evaluation. We choose BLEU-1/2/3/4 [37] to measure the consistency between the generated response and the reference. We use evaluation codes from model Dual [36] to calculate our model results. Dist-1/2 [38] are used to assess the uni-gram/bi-gram diversity of generated response.

6. Experimental Results

6.1. Overall Performance

Table 3 shows the main experimental results of our model. For other models, we use the results directly from their papers. It shows that our method consistently achieves the best results on all metrics, especially in long sequence consistency (BLEU-4) and diversity (Dist), surpassing the previous models.

Table 3. Main results of our model with all evaluation metrics, the bold font in the table represents the best result value.

Models	BLEU-1	BLEU-2	BLEU-3	BLEU-4	Dist-1	Dist-2
HRED	17.19	7.76	4.35	2.53	1.67	6.41
NEXUS	17.33	7.40	4.10	2.34	2.46	8.13
Transformer	19.69	10.15	6.43	4.37	2.18	13.94
RegDG	21.14	11.04	7.39	5.48	2.12	9.74
D2GPO	22.17	11.98	7.86	5.55	2.08	14.87
ProphetChat	23.21	14.17	10.62	8.25	4.05	22.67
AdaLabel	24.16	14.80	10.85	8.63	3.95	22.20
HDLD	27.52	18.26	14.30	12.02	4.37	23.23
iVAE$_{MI}$	30.90	24.90	-	-	2.90	25.00
PLATO	39.70	31.10	-	-	5.30	29.10
PLATO$_{w/oLatent}$	40.50	32.20	-	-	4.60	24.60
Dual	40.80	35.00	32.70	31.50	6.60	33.00
Ours	**41.05**	**35.61**	**33.51**	**32.45**	**7.01**	**38.83**

6.2. Ablation Study

Our model contains three components: component1 contains two response generators (G_F and G_B) and a future generator G_{F2}; component2 implements the learning of sentence relevance by annotated labels; component3 uses a simple generator Dec to predict the final response. In this part, we analyze the performance of the 3 response generators and the effect of our annotated labels.

The performance of response generators. G_F, G_B, and Dec are 3 response generators fed with history, future, and [history, future, relevance label] as inputs, respectively. We first analyze their performance and then discuss their functionality from the perspective of different dialogue relations.

Table 4 shows the results of the 3 generators. We first compare Ours$_{G_F}$ with BART because they have the same model relation and input, the only difference being that Ours$_{G_F}$ is also constrained by the final NLL loss from the Dec through back-propagating. Ours$_{G_F}$ shows significantly better results on BLEU-2∼4 (+0.16, +0.3, +0.34), which indicates that the relevance-aware final loss facilitates the context modeling between history and response. Then, we compare Ours$_{G_B}$ with Ours$_{G_F}$. The difference between them is that Ours$_{G_B}$ uses the simulated future generated by Ours$_{G_{F2}}$ as input. However, the performance of Ours$_{G_B}$ is not decreased much, which not only shows that our method of simulating the future by history in Ours$_{G_{F2}}$ is feasible but also demonstrates that the extremely difficult backward generation can still perform well via our model design. Ours$_{Dec}$ achieves the best results in all metrics, which proves that it can reasonably combine the strengths of both Ours$_{G_F}$ and Ours$_{G_B}$ with the help of relevance labels to achieve better results.

Table 4. Ablation results of different components of our model, the bold font in the table represents the best result value.

	BLEU-1	BLEU-2	BLEU-3	BLEU-4
BART	39.84	34.43	32.34	31.27
Ours$_{G_F}$	39.73	34.59	32.61	31.61
Ours$_{G_B}$	39.43	34.14	32.14	31.12
Ours$_{Dec}$	**40.12**	**34.98**	**33.04**	**32.05**

Further, we compare the 3 response generators on different dialogue relations in Table 5. Ours$_{G_F}$ obtains better results than Ours$_{G_B}$ on relation1, while Ours$_{G_B}$ has better performance on relation2. This shows that with the help of sentence relevance labels, Ours$_{G_F}$ and Ours$_{G_B}$ in our framework are more focused on learning the relations that they correspond to. And Ours$_{Dec}$ obtains the best performance in all the relations, which indicates that it can make good use of the annotated labels and further handle different dialogue relations flexibly.

Table 5. Evaluation results (BLEU-4) of the G_F, G_B, and Dec of our model on different dialogue relations, the bold font in the table represents the best result value.

	Relation1	Relation2	Relation3
Ours$_{G_F}$	31.00	26.34	33.13
Ours$_{G_B}$	30.41	26.53	32.51
Ours$_{Dec}$	**31.65**	**26.85**	**33.43**

The effect of sentence relevance labels. To validate the effect of the relevance labels, we design two alternative methods for obtaining w_1 and w_2 without the help of labels based on our model framework: (1) assigning fixed values 0.5 to w_1 and w_2, called Ours$_{fix}$; and (2) letting the model learn w_1 and w_2 by itself implicitly, called Ours$_{learn}$.

As the results show in Table 6, Ours$_{fix}$ does not perform well because the equal-weight fusion method is less appropriate for relation1 and relation2. Then, Ours$_{learn}$ does not achieve good results either, which proves that dialogue sentence relevance cannot be learned well by the generation model itself implicitly. Finally, Ours shows the best performance, proving that explicit supervised labels on the relevance learning can make the model obtain suitable relevance weights corresponding to different dialogue contexts.

Table 6. Evaluation results of the different relevance obtaining models, the bold font in the table represents the best result value.

	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Ours$_{fix}$	38.51	33.93	32.22	31.41
Ours$_{learn}$	39.26	34.02	31.99	30.93
Ours	**40.12**	**34.98**	**33.04**	**32.05**

6.3. More Analysis

We discuss the response generation model performance and dialogue relation from two perspectives: (1) the performance of generation models on different dialogue relations, and (2) the impact of different dialogue relations on generation models.

6.3.1. Model Performance on Different Dialogue Relations

We take the BART and our model as the representatives of the history-based and history&future-based response generation models and show their results on different dialogue relations in Table 7.

Table 7. Models performance (BLEU-4) on different dialogue relations, the bold font in the table represents the best result value.

	Relation1	**Relation2**	**Relation3**
BART	31.03	26.14	32.62
Ours	**31.65**	**26.85**	**33.43**

From the results, it can be concluded that for both generation models, the difficulty of dialogue relations is relation2 > relation1 > relation3. relation1 is related to history, relation2 is related to the future, and they both rely only on unilateral information. While real future information is not available in practice, relation2 is more difficult than relation1. relation3 is easy to learn because it is related to both the history and future, which contain much information and account for the largest proportion of data.

Compared to BART, our model has the most significant improvement with relation2 (+2.7%) due to the introduction of modeling future information and the assistance of context sentence relevance. The second is the improvement of relation3 (+2.5%), which is because our model can learn good relevance values of the history and future to fit the relation3 data. For relation1, which only uses the history, our model also improves 2% more than BART. This indicates that our model is not only better because of modeling future information but also because the use of dialogue sentence relevance labels guides the model training process of each component so that the full model can learn different abilities according to different dialogue relations.

6.3.2. Impact of Different Dialogue Relations on Models

To investigate the potential of context modeling, we test the upper limit of different context modeling methods with gold data corresponding to different dialogue relation combinations. For example, BART$_{(H,F)}$ corresponds to relation3, where the response is related to both history and future, so [H, F] is used as the input. And BART$_{(H,F,gold\,Rel)}$ corresponds to all the relations because sentence relevance labels *Rel* can distinguish which relation it is.

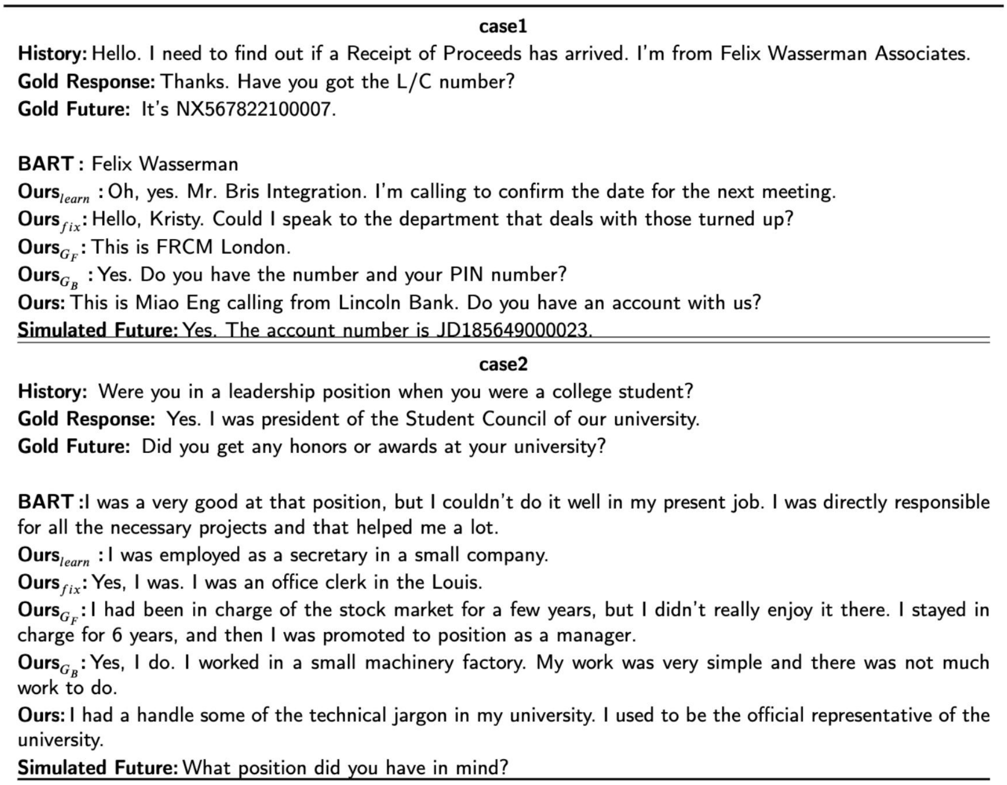
From Table 8, we conclude the following: (1) The relation2-based model BART$_{(F)}$ is better than model BART$_{(H)}$, which is based on relation1 because the explicit information in future can alleviate the one-to-many problem. (2) Model BART$_{(H,F)}$ based on relation3 is the best because it utilizes both the history and future. (3) BART$_{(H,F,gold\,Rel)}$, which is based on all the relations by the hard 0/1 dialogue sentence labels, is suboptimal because the hard 0/1 labels cannot reflect the specific relevant strength.So, it needs to be exploited in a specially designed way, e.g., our soft learning method. (4) There is a gap between Ours and the ceiling BART$_{(H,F)}$, the main reason being that it is difficult to simulate the future, where great research potential still lies.

Table 8. Ceiling and actual results for different context input. The input information of each BART-based model is in the subscript bracket. H, F, and *gold Rel* denote the gold history, gold future, and gold binary sentence relevance labels, respectively. All results are unreachable, in fact, except for BART$_{(H)}$ and Ours.

	BLEU-1	**BLEU-2**	**BLEU-3**	**BLEU-4**
BART$_{(H)}$	39.84	34.43	32.34	31.27
BART$_{(F)}$	40.54	34.80	32.59	31.45
BART$_{(H,F)}$	44.22	38.25	35.77	34.43
BART$_{(H,F,gold\ Rel)}$	42.96	36.72	34.14	32.73
Ours	40.12	34.98	33.04	32.05

6.4. Case Study

We present two cases in Figure 2. The responses generated by our model are more appropriate than others. Case1 corresponds to relation3, where the response is related to both the history and future, and case2 corresponds to relation1, where the response is only related to the history. Through the observation of the two cases, we can find that (1) BART and Ours$_{G_F}$ have difficulty generating responses that elicit the future, (2) Ours$_{G_B}$ with the simulated future as input has difficulty in predicting the response related to the history, (3) Ours$_{learn}$ and Ours$_{fix}$ have the complete context as input but no relevance label as constraint, which leads to the lacking of targeted learning for different context relations, so the responses are not appropriate, and (4) due to the help of relevance labels, our model utilizes the history and the simulated future reasonably. Our model focuses on modeling the complete context in case1, so it generates a response that not only responds to history but also elicits the future. For case2, our model targets how to respond to the history without eliciting the future, thus grasping the keyword "college" in history and giving a more appropriate response.



case1

History: Hello. I need to find out if a Receipt of Proceeds has arrived. I'm from Felix Wasserman Associates.
Gold Response: Thanks. Have you got the L/C number?
Gold Future: It's NX567822100007.

BART : Felix Wasserman
Ours$_{learn}$: Oh, yes. Mr. Bris Integration. I'm calling to confirm the date for the next meeting.
Ours$_{fix}$: Hello, Kristy. Could I speak to the department that deals with those turned up?
Ours$_{G_F}$: This is FRCM London.
Ours$_{G_B}$: Yes. Do you have the number and your PIN number?
Ours: This is Miao Eng calling from Lincoln Bank. Do you have an account with us?
Simulated Future: Yes. The account number is JD185649000023.

case2

History: Were you in a leadership position when you were a college student?
Gold Response: Yes. I was president of the Student Council of our university.
Gold Future: Did you get any honors or awards at your university?

BART : I was a very good at that position, but I couldn't do it well in my present job. I was directly responsible for all the necessary projects and that helped me a lot.
Ours$_{learn}$: I was employed as a secretary in a small company.
Ours$_{fix}$: Yes, I was. I was an office clerk in the Louis.
Ours$_{G_F}$: I had been in charge of the stock market for a few years, but I didn't really enjoy it there. I stayed in charge for 6 years, and then I was promoted to position as a manager.
Ours$_{G_B}$: Yes, I do. I worked in a small machinery factory. My work was very simple and there was not much work to do.
Ours: I had a handle some of the technical jargon in my university. I used to be the official representative of the university.
Simulated Future: What position did you have in mind?

Figure 2. Two cases of the generated responses from different models.

7. Conclusions

In this paper, we propose a new response-centered dialogue relations analysis for three consecutive dialogue sentences. We present three basic dialogue relations that are applicable to computational models and perform annotation and relation analysis on large-scale data. Based on the analysis, we propose a response generation model with adaptive contextual relations, which learns the two relevance scores of history–response and response–future with the help of annotated labels, and then adaptively fuses the responses predicted by the dialogue history and the simulated future via the scores to generate the final response. Experimental results show that our method achieves significant improvement in long sequence consistency and diversity, surpassing the previous baseline model and proving the effectiveness of our model. In future work, we have two exploration directions, where one is to explore more characteristics of dialogue context to help response generation, such as tone of voice and facial expressions, and the other is to provide more support to response prediction by introducing more history and future sentences, i.e., expanding the basic unit of three consecutive sentences centered on the response to five or seven sentences.

Author Contributions: M.W.: conceptualization, methodology, coding, validation, investigation, data curation, writing—original draft preparation; S.T.: coding, validation, investigation, data curation; C.Y. and X.W.: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: The work is partially supported by State Grid Corporation of China's Science and Technology Project "Construction of Electric Power Cognitive Large Model and key Techniques of Its Applications on Operation, Maintenance and Detection" (Project No: 5700-202313288A-1-1-ZN).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets utilized in this study are based on publicly available datasets and interested researchers can access them through the provided references.

Acknowledgments: The authors would like to thank Yaowu Zhang, Ruonan Wang, Hao Shen, Yuxi Qian, Wenjun Wu, Shuyu Qiao, Shaoting Guo, Enhao Tan, Fangkun Zhao, Xiaoshuang Ji, and Xiaobiao Gao for their contribution to labels collection and valuable comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Prabhumoye, S.; Hashimoto, K.; Zhou, Y.; Black, A.W.; Salakhutdinov, R. Focused Attention Improves Document-Grounded Generation. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 6–11 June 2021; pp. 4274–4287.
2. Meng, C.; Ren, P.; Chen, Z.; Ren, Z.; Xi, T.; de Rijke, M. Initiative-Aware Self-Supervised Learning for Knowledge-Grounded Conversations. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual, 11–15 July 2021; pp. 522–532.
3. Liu, C.; Tan, X.; Tao, C.; Fu, Z.; Zhao, D.; Liu, T.; Yan, R. ProphetChat: Enhancing Dialogue Generation with Simulation of Future Conversation. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Dublin, Ireland, 22–27 May 2022; pp. 962–973.
4. Chen, X.; Meng, F.; Li, P.; Chen, F.; Xu, S.; Xu, B.; Zhou, J. Bridging the Gap between Prior and Posterior Knowledge Selection for Knowledge-Grounded Dialogue Generation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020; pp. 3426–3437.
5. Yang, Z.; Liu, Y.; Ouyang, C.; Ren, L.; Wen, W. Counterfactual can be strong in medical question and answering. *Inf. Process. Manag.* **2023**, *60*, 103408. [CrossRef]
6. Serban, I.V.; Sordoni, A.; Bengio, Y.; Courville, A.C.; Pineau, J. Building End-to-End Dialogue Systems Using Generative Hierarchical Neural Network Models. *Proc. AAAI Conf. Artif. Intell.* **2016**, *30*, 3776–3784. [CrossRef]
7. Moghe, N.; Arora, S.; Banerjee, S.; Khapra, M.M. Towards Exploiting Background Knowledge for Building Conversation Systems. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 2322–2332.

8. Zhang, H.; Lan, Y.; Pang, L.; Guo, J.; Cheng, X. ReCoSa: Detecting the Relevant Contexts with Self-Attention for Multi-turn Dialogue Generation. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 3721–3730.
9. Shen, X.; Su, H.; Li, W.; Klakow, D. Nexus Network: Connecting the Preceding and the Following in Dialogue Generation. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 4316–4327.
10. Feng, S.; Chen, H.; Li, K.; Yin, D. Posterior-GAN: Towards Informative and Coherent Response Generation with Posterior Generative Adversarial Network. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 7708–7715. [CrossRef]
11. Feng, S.; Ren, X.; Chen, H.; Sun, B.; Li, K.; Sun, X. Regularizing Dialogue Generation by Imitating Implicit Scenarios. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020; pp. 6592–6604.
12. Wang, W.; Huang, M.; Xu, X.; Shen, F.; Nie, L. Chat More: Deepening and Widening the Chatting Topic via A Deep Model. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018; pp. 255–264.
13. Xu, J.; Wang, H.; Niu, Z.; Wu, H.; Che, W. Knowledge Graph Grounded Goal Planning for Open-Domain Conversation Generation. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 9338–9345. [CrossRef]
14. Ling, Y.; Cai, F.; Hu, X.; Liu, J.; Chen, W.; Chen, H. Context-controlled topic-aware neural response generation for open-domain dialog systems. *Inf. Process. Manag.* **2021**, *58*, 102392. [CrossRef]
15. Wang, M.; Tian, S.; Bai, Z.; Yuan, C.; Wang, X. Hierarchical history based information selection for document grounded dialogue generation. *Appl. Intell.* **2023**, *53*, 17139–17153. [CrossRef]
16. Li, Y.; Su, H.; Shen, X.; Li, W.; Cao, Z.; Niu, S. DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset. In Proceedings of the Eighth International Joint Conference on Natural Language Processing, Taipei, Taiwan, 27 November–1 December 2017; pp. 986–995.
17. Cambria, E.; Malandri, L.; Mercorio, F.; Mezzanzanica, M.; Nobani, N. A survey on XAI and Natural Language Explanations. *Inf. Process. Manag.* **2023**, *60*, 103111. [CrossRef]
18. Sun, K.; Guo, C.; Zhang, H.; Li, Y. HVLM: Exploring human-like visual cognition and language-memory network for visual dialog. *Inf. Process. Manag.* **2022**, *59*, 103008. [CrossRef]
19. Li, L.; Xu, C.; Wu, W.; Zhao, Y.; Zhao, X.; Tao, C. Zero-Resource Knowledge-Grounded Dialogue Generation. In Proceedings of the 34th Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–12 December 2020.
20. Zhao, X.; Wu, W.; Tao, C.; Xu, C.; Zhao, D.; Yan, R. Low-Resource Knowledge-Grounded Dialogue Generation. In Proceedings of the Eighth International Conference on Learning Representations ICLR 2020, Virtual, 26 April–1 May 2020.
21. Dinan, E.; Roller, S.; Shuster, K.; Fan, A.; Auli, M.; Weston, J. Wizard of Wikipedia: Knowledge-Powered Conversational Agents. In Proceedings of the Seventh International Conference on Learning Representations ICLR, New Orleans, LA, USA, 6–9 May 2019.
22. Li, Z.; Kiseleva, J.; de Rijke, M. Improving Response Quality with Backward Reasoning in Open-domain Dialogue Systems. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual, 11–15 July 2021; pp. 1940–1944.
23. Wu, J.; Zhou, H. Augmenting Topic Aware Knowledge-Grounded Conversations with Dynamic Built Knowledge Graphs. In Proceedings of the Deep Learning Inside Out (DeeLIO): The 2nd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures, Online, 10 June 2021; pp. 31–39.
24. Kong, Y.; Zhang, L.; Ma, C.; Cao, C. HSAN: A hierarchical self-attention network for multi-turn dialogue generation. In Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 7433–7437.
25. Wang, J.; Sun, X.; Chen, Q.; Wang, M. Information-enhanced hierarchical self-attention network for multiturn dialog generation. *IEEE Trans. Comput. Soc. Syst.* **2023**, *10*, 2686–2697. [CrossRef]
26. Lv, A.; Li, J.; Xie, S.; Yan, R. Envisioning Future from the Past: Hierarchical Duality Learning for Multi-Turn Dialogue Generation. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, Toronto, ON, Canada, 9–14 July 2013; pp. 7382–7394.
27. Shen, L.; Zhan, H.; Shen, X.; Feng, Y. Learning to select context in a hierarchical and global perspective for open-domain dialogue generation. In Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 7438–7442.
28. Wang, Y.; Li, Y.; Wang, Y.; Mi, F.; Zhou, P.; Liu, J.; Jiang, X.; Liu, Q. History, Present and Future: Enhancing Dialogue Generation with Few-Shot History-Future Prompt. In Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
29. Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.G.; Salakhutdinov, R.; Le, Q.V. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019; pp. 5754–5764.
30. Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; Zettlemoyer, L. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 7871–7880.

31. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
32. Li, Z.; Wang, R.; Chen, K.; Utiyama, M.; Sumita, E.; Zhang, Z.; Zhao, H. Data-dependent Gaussian Prior Objective for Language Generation. In Proceedings of the Eighth International Conference on Learning Representations, Addis Ababa, Ethiopia, 26 April–1 May 2020.
33. Wang, Y.; Zheng, Y.; Jiang, Y.; Huang, M. Diversifying Dialog Generation via Adaptive Label Smoothing. *arXiv* **2021**, arXiv:2105.14556.
34. Fang, L.; Li, C.; Gao, J.; Dong, W.; Chen, C. Implicit Deep Latent Variable Models for Text Generation. *arXiv* **2019**, arXiv:1908.11527.
35. Bao, S.; He, H.; Wang, F.; Wu, H.; Wang, H. PLATO: Pre-trained Dialogue Generation Model with Discrete Latent Variable. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 85–96.
36. Bai, X.; Chen, Y.; Song, L.; Zhang, Y. Semantic Representation for Dialogue Modeling. *arXiv* **2021**, arXiv:2105.10188.
37. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W. Bleu: A Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; pp. 311–318.
38. Li, J.; Galley, M.; Brockett, C.; Gao, J.; Dolan, B. A Diversity-Promoting Objective Function for Neural Conversation Models. In Proceedings of the NAACL-HLT 2016, San Diego, CA, USA, 12–17 June 2016; pp. 110–119.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.