micromachines

MDPI

Article

Human–Machine Multi-Turn Language Dialogue Interaction Based on Deep Learning

Xianxin Ke, Ping Hu *, Chenghao Yang and Renbao Zhang

School of Mechanical and Electrical Engineering and Automation, Shanghai University, Shanghai 200444, China; xxke@staff.shu.edu.cn (X.K.); jingzhanhui2017@163.com (C.Y.); Zz3476502133@163.com (R.Z.)
* Correspondence: followapple@shu.edu.cn

Abstract: During multi-turn dialogue, with the increase in dialogue turns, the difficulty of intention recognition and the generation of the following sentence reply become more and more difficult. This paper mainly optimizes the context information extraction ability of the Seq2Seq Encoder in multi-turn dialogue modeling. We fuse the historical dialogue information and the current input statement information in the encoder to capture the context dialogue information better. Therefore, we propose a BERT-based fusion encoder ProBERT-To-GUR (PBTG) and an enhanced ELMO model 3-ELMO-Attention-GRU (3EAG). The two models mainly enhance the contextual information extraction capability of multi-turn dialogue. To verify the effectiveness of the two proposed models, we demonstrate the effectiveness of our model by combining data based on the LCCC-large multi-turn dialogue dataset and the Naturalconv multi-turn dataset. The experimental comparison results show that, in the multi-turn dialogue experiments of the open domain and fixed topic, the two Seq2Seq coding models proposed are significantly improved compared with the current state-of-the-art models. For specified topic multi-turn dialogue, the 3EAG model has the average BLEU value reaches the optimal 32.4, which achieves the best language generation effect, and the BLEU value in the actual dialogue verification experiment also surpasses 31.8. for open-domain multi-turn dialogue. The average BLEU value of the PBTG model reaches 31.8, the optimal 31.8 achieves the best language generation effect, and the BLEU value in the actual dialogue verification experiment surpasses 31.2. So, the 3EAG model is more suitable for fixed-topic multi-turn dialogues for the two tasks. The PBTG model is more muscular in open-domain multi-turn dialogue tasks; therefore, our model is significant for promoting multi-turn dialogue research.

Keywords: human–machine interaction; Seq2Seq; NLP; deep learning; context semantic coding

Citation: Ke, X.; Hu, P.; Yang, C.; Zhang, R. Human–Machine Multi-Turn Language Dialogue Interaction Based on Deep Learning. *Micromachines* **2022**, *13*, 355. https://doi.org/10.3390/mi13030355

Academic Editors: Zhangguo Yu and Marco Ceccarelli

Received: 27 January 2022
Accepted: 22 February 2022
Published: 23 February 2022

1. Introduction

Language communication is an integral part of people's daily life. With the development of artificial intelligence technology and natural language processing, the research on the human–machine dialogue has been transformed from single question–answer dialogue to multi-turn dialogue, which is more challenging. The applied dialogue model is concerned; it divides into two types in broadly human–machine dialogue [1]. The first is a task-based dialogue, and the second is an open-domain dialogue. Task-oriented dialogue is mainly task driven, and the machine needs to understand, ask, clarify to deal with users' needs. Task-based dialogue topics are relatively fixed and generally have poor generalization ability, but they have more advantages than non-task-based dialogues when dealing with questions–answers tasks. For non-task-based dialogues, they break through the topic limitation [2]. They can provide better responses between multiple topics and even in open domains, making human–machine dialogue resemble the natural communication between people. Still, the resulting methodological research is also more challenging. The research methods of non-task-based dialogue models divide into retrieval-based methods and neural generation-based methods. The retrieval-based method mainly completes the

reply matching of each dialogue turn, a discriminative model. The neural generation-based models mainly include Sequence-to-Sequence (Seq2Seq) Models [3–5], Dialogue Context [6], Response Diversity [7,8], Topic and Personality [9,10], Outside Knowledge Base. Dialogue context, Response Diversity, Topic or Personality, and other methods adopt multi-classification of contextual dialogues and then select and integrate the best alternative answers to return. This method is, in principle, a classification language model. At the same time, Outside Knowledge Base needs to build a massive amount of knowledge library to adapt to the diversity of the dialogue process. Sequence-to-Sequence models reduce manual preprocessing and post-processing. It tries to make the model from the original input to the final output as effective possible, giving the model more space for automatic adjustment according to the data. It has been widely used in various dialogue generation research [11,12] due to its advantages, such as increasing the model's overall fit [10].

Although many scholars have made some improvements and optimizations, the Transformer model [12] proposed by Google researchers has promoted the development of the Natural Language Generation. It offered a vast improvement; subsequently, the BERT model proposed by Devlin et al. [13] has become a natural language encoder in Seq2Seq. Later, with the proposal of the GPT model [14], the attention mechanism is used in the field of natural language generation. It has achieved excellent performance and significantly promoted the progress of the generation task. To combine the advantages of BERT and GPT, the MASS model proposed by Song [15] of Microsoft Research Asia shields some of the sentences and then regenerates this fragment in the decoder. They can easily change the structure of the model by adjusting the hyperparameters, but their specific achievements in multi-turn dialogue are still unknown. Still, the innovation of the model has expanded. The optimization of these related models is undoubtedly the transformation of the encoder and decoder. In any case, they still cannot rival the position of Seq2Seq in the field of Natural Language Generation.

The communication mode of multi-turn dialogue is widespread in daily communication. Table 1 shows an example of multi-turn dialogue in Tencent AI Lab [16]; the research of multi-turn dialogue has experienced template matching, task-driven dialogue, recommendation model dialogue management, Knowledge Graph, Seq2Seq, and other research processes. Template matching [17,18] completes the dialogue communication by building a vast multi-turn dialogue database to retrieve information. The task-driven dialogue [19–21] is mainly in the task-oriented dialogue field, such as booking an airline ticket and a hotel, and its context fills with fixed sentences. For the recommendation model dialogue management [22], it uses information, such as the extraction of input features, and historical conversations uses a search or retrieval model to extract the optimal answer. For knowledge graph-based dialogue, which manages the conversation by constructing a knowledge graph, Xu [23] disassembled the multi-turn open-domain dialogue into two sub-tasks: planning the dialogue target sequence and the in-depth dialogue for a given dialogue target for the first time. The knowledge graph-based dialogue introduces displayed and interpretable dialogue states and actions for dialogue policy learning, which facilitates the design of Reward factors related to dialogue goals and uses dialogue goals and fine-grained topics to guide response generation. Finally, the multi-turn generation of Seq2Seq [24,25] mainly takes the current sentence input plus historical context information and sends it to the encoder, then decodes it through the decoder to generate a reply. Wu [26] et al. developed sentence information, which is encoded using tokens in the encoder encoding process and then combined with the token information in the decoding process to generate dialogue.

Table 1. Examples of multi-turn dialogue.

Turns	Dialogue Text
Turn-1	hi~你好啊 (Hello)
Turn-2	嗯，你好，有什么事吗? (Hello, do you have any questions?)
Turn-3	看你一个人也挺无聊的，来聊会天吧。(You look bored, let us have a chat.)
Turn-4	好啊，聊点什么呢？(OK, what shall we talk about?)
Turn-5	你看网球吗，我可是很喜欢网球的。(Do you like tennis? I like it very much.)
Turn-6	网球啊，一般吧，我就知道个李娜。(Tennis is OK. I only know Li Na.)

2. Related Work

Google's Oriol Vinyals [27] first proposed a neural network dialogue model, which is the source of Seq2Seq. Then, Li Hang et al. [28] first applied the Seq2Seq translation model with attention to dialogue tasks based on Weibo comment data. Baidu and Université de Montréal [29,30] successively adopted the Seq2Seq framework to generate the Nth sentence using the first N-1 sentences, which divided the dialogue model into two layers. The entire dialogue in the first layer combines all penalties and the second layer. Each dialogue is a combination of all words. Still, the author believes that the fundamental reason for the low quality of dialogues generated by language models, such as RNNLM, is that the model does not deal with the random features and noise hidden in the dialogue. Hence, the following sentence causes the dialogue. The effect is not ideal, so the context layer RNN and the hidden state layer are embedded in the middle of Seq2Seq to improve the overall dialogue randomness. Université de Montréal, Georgia Institute of Technology, Facebook, and Microsoft Research [31] jointly trained a data-driven open-domain dialogue model. They believed that the current user query sentence and historical dialogue information should be considered when generating the current dialogue. This dialogue generation model has been generalized in open-domain multi-turn dialogue research. To solve the generation of meaningless sentences in Seq2Seq dialogue generation, Li Ji et al. [32] proposed to train Seq2Seq with maximum mutual information, which effectively solved the problem of generating irrelevant replies. However, it uses a traditional network model, which is more sensitive to the sequence length. This paper mainly combines the current popular pre-training mechanism to improve the semantic fusion effect in the multi-turn dialogue generation process.

3. Coding Model for Multi-Turn Dialogue

Since our work involves the problem of word vector representation, we tried Word2Vec [33] word vector training method and GolVe [34] word vector training method for word vector representation. These two types of model methods for word vector representation adopt fixed expressions. They are trained based on single corpus sentences. The word vector information of each word only captures all the information of the sentence, and the context information is relatively lacking. In addition, the ELMO [35] pre-training word vector method captures the word vector information of the dialogue model relatively well and can grasp the sentence semantics of the context better, so we used ELMO to represent the word vector.

The early structure of the traditional Seq2Seq dialogue model mentioned in this paper combines bidirectional GRU [36] and unidirectional GRU models. After the Transformer model was proposed, the dialogue model's coding part adopts the attention mechanism to capture the context information. Compared with the previous encoder, the capture ability has dramatically improved. After the BERT masked language model proposal, it adopts the context prediction method. The method has continuously enhanced the ability to generate word vectors and capture information, such as syntax and semantics.

This research focused on the context management model and word vector encoding learning optimization performance in generative dialogue. The improved encoding model based on BERT applied multi-turn dialogue. For the first time, we propose to embed the

multi-turn contextual positional encoding into the BERT model, which helps to improve the generator's decoding performance.

Our models were divided into two forms: (1) multi-turn word-sentence vector encoding model, contextual syntactic and semantic capture for historical multi-turn dialogues; (2) PBTG network model, for historical conversation semantic multi-sentence information encoding the device model, jointly encodes the historical sentence information and current input sentence.

3.1. Context Semantic Encoding Model

3.1.1. PBTG Context Semantic Coding Model

We used historical turn-based sentence coding combined with the current input sentence to encode multi-turn contextual word coding. Based on the word encoding of BERT [13], the historical turn sentence information and the current input information encoding were added. It started from the first turn of dialogue: the first sentence started with ["CSL"] and ended with ["SEP"] between each turn, where the encoding result was the vector sum of Token encoding, Segment encoding, Sentences encoding, and Position encoding. Furthermore, the model's architecture is shown in Figure 1.

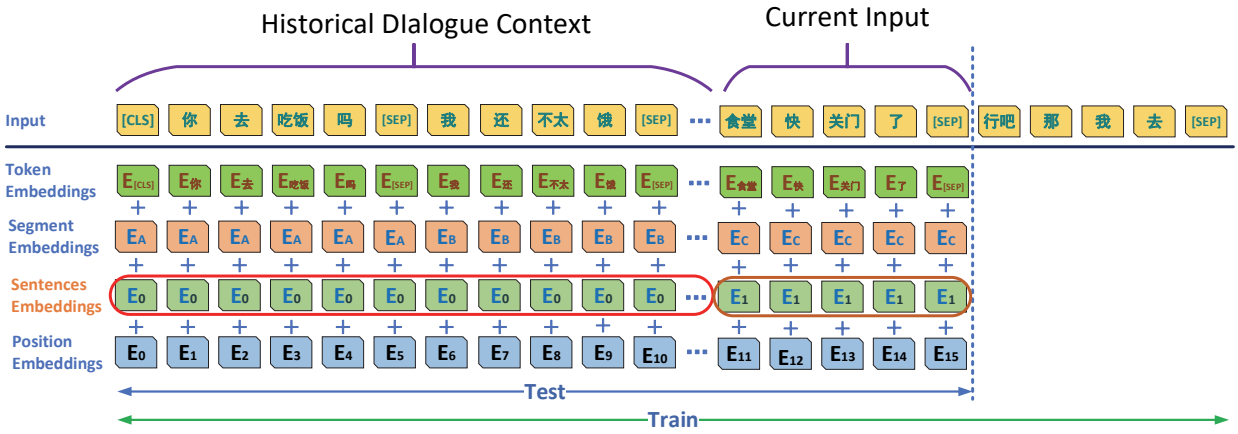

Figure 1. Context Encoding Structure of the PBTG Model.

Among the parameters, "Historical Dialogue" is all the historical information of the dialogue from the first round; "Current Input" is the input dialogue sentence of the current turn, followed by the decoder input information of teacher forcing; the encoding result is presented in Equation (1).

$$W_E = f_{Token}(input) \oplus f_{Seg}(input) \oplus f_{Sent}(input) \oplus f_{Posi}(input) \quad (1)$$

3.1.2. Context Semantic Encoding Structure of 3EAG Model

The traditional ELMO [35] model pre-training adopts a bidirectional two-layer LSTM [37] model to capture contextual information. The contextual information mainly includes the context in a single sentence and lacks information capture between sentences. Therefore, this study used the traditional ELMO model to capture contextual information. The number of layers was increased to capture context information, and a 3-layer GRU [36] bidirectional network as used to capture word information, segment information, and context information. The model's structure is shown in Figure 2.

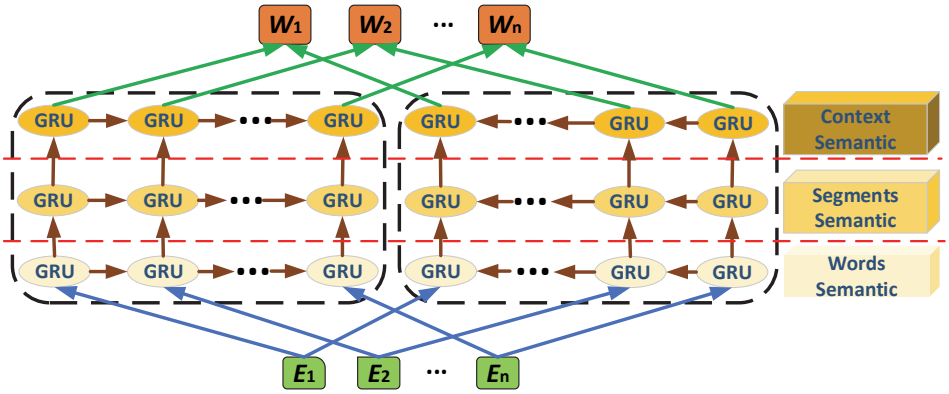

Figure 2. Context Semantic Encoding Structure of the 3EAG Model.

For the context semantic encoding of 3EAG, its forward model was:

$$p(w_1, w_2, ..., w_N) = \prod_{k=1}^{N} p(w_k | w_1, w_2, ..., w_{k-1}). \tag{2}$$

The backward model was:

$$p(w_1, w_2, ..., w_N) = \prod_{k=1}^{N} p(w_k | w_{k+1}, w_{k+2}, ..., w_N). \tag{3}$$

The context semantic encoding output $W_{ELMo,k}$ was:

$$W_{ELMo,k} = W_{words,k} \oplus W_{segs,k} \oplus W_{context,k} \tag{4}$$

$$\begin{cases} W_{words,k} = \left\{ x_k^{LM}, \overrightarrow{h}_{k,j}^{LM}, \overleftarrow{h}_{k,j}^{LM} \mid j = 1 \right\} \\ W_{Segs,k} = \left\{ x_k^{LM}, \overrightarrow{h}_{k,j}^{LM}, \overleftarrow{h}_{k,j}^{LM} \mid j = 2 \right\} \\ W_{context,k} = \left\{ x_k^{LM}, \overrightarrow{h}_{k,j}^{LM}, \overleftarrow{h}_{k,j}^{LM} \mid j = 3 \right\} \end{cases} \tag{5}$$

where k is the position of each word in the coder, with a value range in $\{0, 1, 2, .., N\}$.

The encoding model adopted a three-layer Markov chain model, and its semantic representation depended on the dialogue information of the previous turn. Due to the hierarchical dialogue structure, the contextual correlation is not high in multi-turn dialogues with uncertain topics [38,39]. We tried to conduct experiments in open-domain and fixed-topic multi-turn dialogues. The follow-up experimental results also prove that it is more suitable for fixed-topic multi-turn dialogues. This paper performed a comparative investigation in the fixed-topic multi-turn dialogue experiment and the open-domain dialogue experiment. The detailed results are shown in the experimental Results Analysis Section.

3.2. Encoder Network Model

3.2.1. PBTG Network Model

In this study, the contextual sentence encoding result was used as the input of the BERT model, and the Encoder part of the Transformer [6] model as the model framework. The model structure is shown in Figure 3. This paper tested the coding effect of the number of model layers in a variety situations.

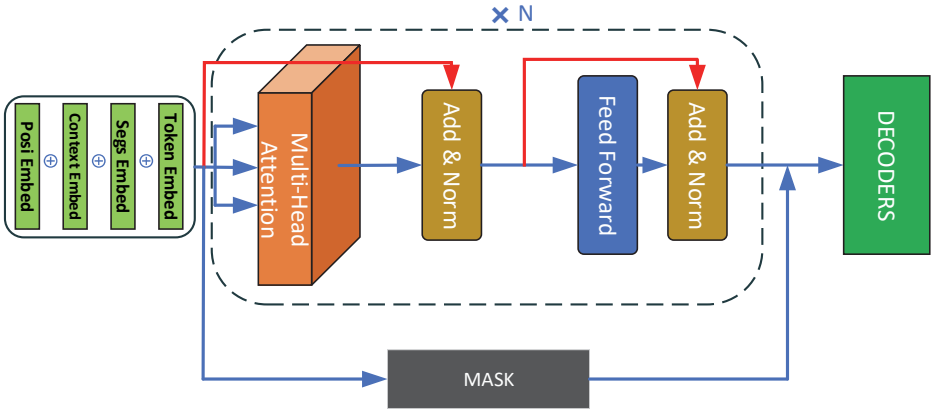

Figure 3. PBTG model structure.

3.2.2. EAG Network Model

We also used the ELMO variant model as the research effect of the encoder. First, the 3-layer forward and backward bidirectional GRU [36] model captured the context information. The Attention mechanism and input jointly encoded the historical information and the current input as DECODERS. The model structure is shown in Figure 4.

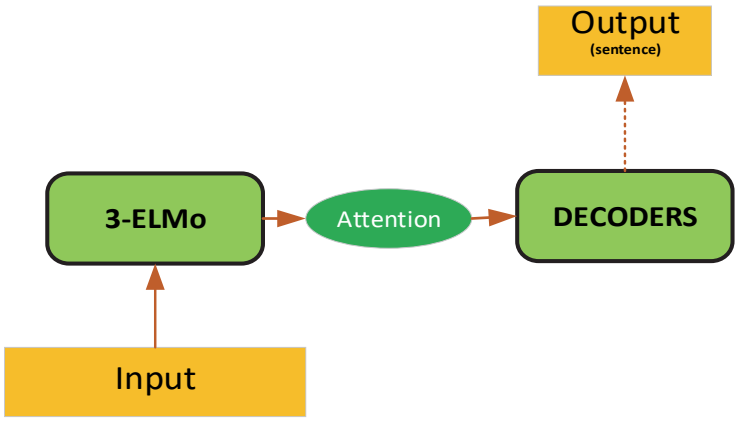

Figure 4. 3EAG model structure.

4. Experiments

This section describes the experiments conducted on multi-turn dialogue data and shows the promising results.

4.1. The Data Set

Our experimental data adopts the open-source LCCC-large [40] Chinese multi-turn dialogue dataset of Tsinghua University and the open-source Naturalconv [16] multi-turn dialogue dataset of Tencent AI Lab. We mainly cleaned the two data comprehensively, and the total number of Naturalconv databases is 19,919. When we split multiple turns, we expanded the data into 2\3\4\5\6 by splitting and combining each turn with a data volume of 50,000. Our final data contained 2~6 turns of dialogue data, and the processed data had a data volume of 170,000 per turn; the total population was 850,000 (counting the dialogue example in Table 1 as a data volume of 6 turns). The number of sentences after processing is shown in Table 2. The overall sentence length distribution is shown in Figure 5.

Table 2. Data sources and processing results.

Data Sources	Turns	Data Quantity
Naturalconv [10]	2; 3; 4; 5; 6	50,000; 50,000; 50,000; 50,000; 50,000
LCCC-large [26]	2; 3; 4; 5; 6	120,000; 120,000; 120,000; 120,000; 120,000
Ours	**2; 3; 4; 5; 6**	**170,000; 170,000; 170,000; 170,000; 170,000**

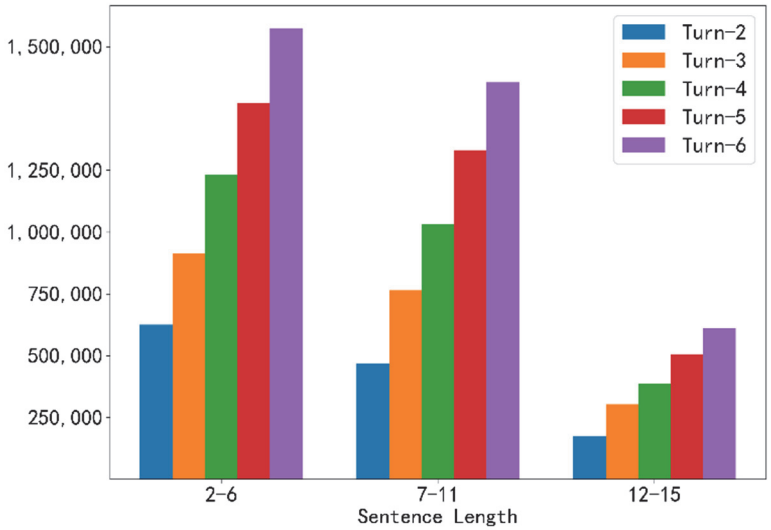

Figure 5. Sentence length distribution of the data sets.

For the fixed topic dialogue experiment, we screened out 50,000 dialogues on sports topics, health topics, and science and technology topics for experiments. Two thousand pieces of dialogue data were used as verification dialogues. The statistics of our processed sentences are shown in Table 3.

Table 3. Statistics of conversation data on fixed topics.

Topics	Turn	Data Quantity
Sport	2; 3; 4; 5; 6	10,000; 10,000; 10,000; 10,000; 10,000
Health	2; 3; 4; 5; 6	10,000; 10,000; 10,000; 10,000; 10,000
Tech	2; 3; 4; 5; 6	10,000; 10,000; 10,000; 10,000; 10,000
Verification	random	2000

4.2. Experimental Parameters and Results Analysis

4.2.1. Experimental Parameters

For the experiments with the 3EAG model, we adopted the context-encoded representation $E \in R^{M \times N}$ as the model input representation. The single input of sentence length was increased to 15, and the word vector dimension was set to 512. We used the forward and backward three-layer GRU combined with self-attention as an encoder, while the decoder used a three-layer unidirectional GRU.

For the experiments with the PBTG model, we adopted the context-encoded representation $E \in R^{M \times N}$ as the model input representation. The single input sentence length was increased to 15, and we set the word vector dimension to 512. We set the number of ENCODER layers to 8–12 layers, the number of attention heads to 8, the masking rate of MASK to 0.17, and the DECODER layer to use four layers of unidirectional GRU as the generator.

Based on the above parameter settings, we testes the language effects of the traditional Seq2Seq (LSTM to LSTM) model, Transformer model, and our 3EAG model and PBTG model in open-domain dialogue generation and fixed-topic dialogue generation.

In this experiment, BLEU [41] was used to judge the quality of the model generation effect. BLEU is one of the commonly used evaluation indicators of the Seq2Seq model. With an improvement in the effect, we used the BLEU value to evaluate the similarity between the response generated by the model and the target sentence. We used BLEU-2, BLEU-3, and BLEU-4 for the specific evaluation in this experiment.

4.2.2. Results Analysis

As shown from Table 4, the average BLEU value of our 3EAG model is 2.3 higher than the traditional Seq2Seq and 0.4 higher than the average BLEU value of the Transformer model. The average BLEU value of our improved PBTG model is 3.3, 1.4, and 1.0 higher, respectively, compared with the conventional Seq2Seq, Transformer, 3EAG the models. In the actual dialogue verification experiments, our average BLEU value also achieves a score of 31.2, which outperforms the previous three models. It can be seen from these results that our model has a stronger ability to identify topics in the contextual information capture ability in multi-turn dialogue generation.

Table 4. BLUE's evaluation results of open-domain dialogue.

Model	BLEU-2	BLEU-3	BLEU-4	Average BLEU
Seq2Seq	39.2	29.1	17.3	28.5
Transformer	40.3	31.4	19.5	30.4
3EAG(Our)	**40.7**	**31.6**	**20.2**	**30.8**
PBTG(Our)	**41.3**	**32.7**	**21.5**	**31.8**
PBTG (Verification)	**40.9**	**32.1**	**20.7**	**31.2**

For the PBTG model, we tested the encoding performance between encoder layers 8–12, as shown in Table 5.

Table 5. BLEU performance of different layers for PBTG's encoders.

#Layers	BLEU-2	BLEU-3	BLEU-4	Average BLEU
8	40.2	29.4	19.1	29.5
9	40.7	30.8	20.2	30.5
10	40.5	31.2	19.3	30.3
11	41.3	30.4	20.6	30.7
12	**41.3**	**32.7**	**21.5**	**31.8**

Table 5 and Figure 6 show that, in open-domain dialogue generation, for our PBTG model, the number of layers of the encoder network is between 8–12 layers, and its BLEU value keeps increasing as the number of layers increases. Additionally, the changes of BLEU-2, BLEU-3, and BLEU-4 all show an upward trend; their growth rates are also stable in the region 0~0.08. It can be seen that the more layers the encoder has, the better the generation effect of the model. From the experimental results, we can see the effectiveness of our two encoding models for the multi-turn dialogue generation task.

For multi-turn dialogue experiments with fixed topics, we used the same experimental parameters to conduct experiments, and the obtained BLEU performance is shown in Table 6.

From the multi-turn dialogue experiments on fixed topics, we can conclude that the PBTG model and 3EAG model proposed in this paper achieved better results. Their average BLEU values are higher than the previous two models; PBTG model is better than Seq2Seq and Transformer, respectively. The model outperformed the other two models by 0.9 and 0.5, respectively, while the average BLEU improvement of 3EAG was 3.1 and 1.7, respectively. Finally, we conducted an actual dialogue verification experiment on the 3EAG model. According to the experimental results, the BLEU-4 value of the actual dialogue reached 21.0, and the average BLEU value also achieved a good score of 31.8.

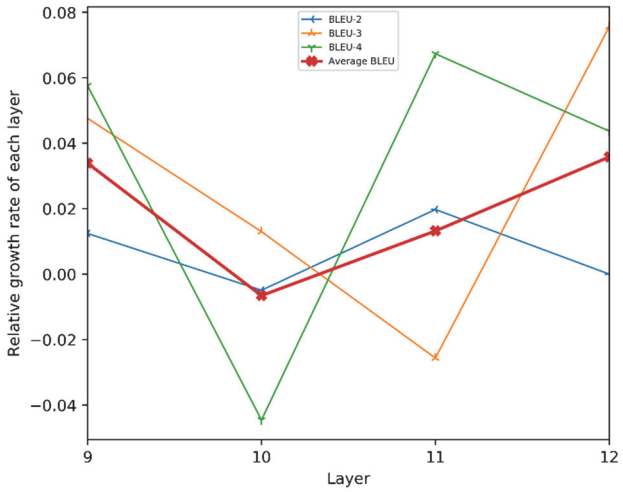

Figure 6. The growth rate of BLEU value varies with the number of network layers.

Table 6. BLUE's evaluation results of fixed topics.

Model	BLEU-2	BLEU-3	BLEU-4	Average BLEU
Seq2Seq	40.4	29.5	18.2	29.3
Transformer	41.2	31.7	19.3	30.7
PBTG(Our)	41.4	31.9	20.3	31.2
3EAG(Our)	**42.6**	**33.2**	**21.6**	**32.4**
3EAG (Verification)	**42.1**	**32.5**	**21.0**	**31.8**

Through the above open-domain and fixed-topic multi-turn dialogue experiments, we can conclude that, for the two models proposed in this paper, PBTG and 3EAG, the BLEU-2, BLEU-3, and BLEU-4 values of PBTG were the best experimental results (in the open domain dialogue experiment), and the average BLUE value was 1.0 higher than that of 3EAG. Therefore, the PBTG model is more suitable for open-domain dialogue tasks and has the best effect. In the fixed-topic dialogue experiment, the BLEU-2, BLEU-3, and BLEU-4 values of 3EAG obtained the best experimental results, and the average BLUE value was 0.8 higher than that of 3EAG. The average BLEU value of the actual dialogue experiment also achieved better results. The value of 31.8 is excellent, so the 3EAG model is more suitable for fixed-topic dialogue tasks and has the best effect.

5. Conclusions

Aiming to achieve a multi-turn dialogue generation system model, we proposed a semantic encoding model as the inner encoder of the dialogue generation model. It improves the ability of contextual semantic extraction and can integrate historical dialogue information and current input information in multi-turn dialogue, which enhances the dialogue context. We combined the LCCC-large multi-turn dialogue dataset and Naturalconv multi-turn data for our research purpose. We, then, adopted the combined split method to construct our open-domain and fixed-topic multi-turn datasets. Additionally, to extract contextual semantics in multi-turn dialogues, we proposed two contextual semantic fusion models, 3EAG and PBTG. To further validate the performance of our model, we evaluated it on our multi-turn dialogue dataset. The experimental results show that our proposed 3EAG model achieves the optimal language effect for fixed-topic dialogues. The PBTG model achieves the best dialogue effect in the open-domain dialogue generation experiment and verifies the effectiveness of dialogue context information extraction and related model design. Therefore, the model we proposed has a good significance for promoting multi-turn dialogue research. We will carry out in-depth research in the direction of

multi-turn emotional dialogue and topics recognition of multi-turn dialogue in the future to test the capabilities of our model.

Author Contributions: Conceptualization, P.H. and X.K.; methodology, P.H.; model, P.H. and C.Y.; validation, R.Z., P.H. and C.Y.; investigation, X.K.; resources, X.K.; data curation, P.H.; writing—original draft preparation, P.H.; writing—review and editing, X.K. and P.H.; visualization, P.H.; supervision, X.K.; project administration, X.K.; funding acquisition, X.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Science and Technology Innovation Action Plan Foundation of Shanghai (Grant No. 19DZ2330500).

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

References

1. Chen, H.; Liu, X.; Yin, D.; Tang, J. A survey on dialogue systems: Recent advances and new frontiers. *ACM SIGKDD Explor. Newsl.* **2017**, *19*, 25–35. [CrossRef]
2. Ritter, A.; Cherry, C.; Dolan, W.B. Data-Driven Response Generation in Social Media. In Proceedings of the Empirical Methods in Natural Language Processing, EMNLP 2011, Edinburgh, UK, 27–31 July 2011; Association for Computational Linguistics (ACL): Edinburgh, UK, 2011; pp. 583–593.
3. Xiang, Z.; Yan, J.; Demir, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour. Res.* **2020**, *56*, e2019WR025326. [CrossRef]
4. Liu, X.; Wang, L.; Wong, D.F.; Ding, L.; Chao, L.S.; Tu, Z. Understanding and improving encoder layer fusion in sequence-to-sequence learning. *arXiv* **2020**, arXiv:14768.
5. Lee, S.; Lim, D.-E.; Kang, Y.; Kim, H.J. Clustered Multi-Task Sequence-to-Sequence Learning for Autonomous Vehicle Repositioning. *IEEE Access* **2021**, *9*, 14504–14515. [CrossRef]
6. Sun, X.; Ding, B. Neural Network with Hierarchical Attention Mechanism for Contextual Topic Dialogue Generation. *IEEE Access* **2022**, *10*, 4628–4639. [CrossRef]
7. Qu, C.; Yang, L.; Qiu, M.; Croft, W.B.; Zhang, Y.; Lyyer, M. BERT with History Answer Embedding for Conversational Question Answering. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, 21–25 July 2019; Association for Computing Machinery: Paris, France, 2019; pp. 1133–1136.
8. Li, J.; Galley, M.; Brockett, C.; Spithourakis, G.P.; Gao, J.; Dolan, B. A persona-based neural conversation model. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, Berlin, Germany, 7–12 August 2016; Association for Computational Linguistics (ACL): Berlin, Germany, 2016; pp. 994–1003.
9. Zhou, X.; Li, L.; Dong, D.; Liu, Y.; Chen, Y.; Zhao, W.X.; Yu, D.; Wu, H. Multi-turn response selection for chatbots with deep attention matching network. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, VIC, Australia, 15–20 July 2018; Association for Computational Linguistics (ACL): Melbourne, VIC, Australia, 2018; pp. 1118–1127.
10. Ayana; Shen, S.-Q.; Lin, Y.-K.; Tu, C.-C.; Zhao, Y.; Liu, Z.-Y.; Sun, M.-S. Recent Advances on Neural Headline Generation. *J. Comput. Sci. Technol.* **2017**, *32*, 768–784. [CrossRef]
11. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. In Proceedings of the 28th Annual Conference on Neural Information Processing Systems 2014, NIPS 2014, Montreal, QC, Canada, 8–13 December 2014; Neural information processing systems foundation: Montreal, QC, Canada, 2014; pp. 3104–3112.
12. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the 31st Annual Conference on Neural Information Processing Systems, NIPS 2017, Long Beach, CA, USA, 4–9 December 2017; Neural Information Processing Systems Foundation: Long Beach, CA, USA, 2017; pp. 5999–6009.
13. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2019, Minneapolis, MN, USA, 2–7 June 2019; Association for Computational Linguistics (ACL): Minneapolis, MN, USA, 2019; pp. 4171–4186.
14. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. Available online: https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf (accessed on 26 January 2022).
15. Song, K.; Tan, X.; Qin, T.; Lu, J.; Liu, T.-Y. MASS: Masked Sequence to Sequence Pre-Training for Language Generation. In Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, CA, USA, 9–15 June 2019; International Machine Learning Society (IMLS): Long Beach, CA, USA, 2019; pp. 10384–10394.
16. Wang, X.; Li, C.; Zhao, J.; Yu, D. Naturalconv: A chinese dialogue dataset towards multi-turn topic-driven conversation. *arXiv* **2021**, arXiv:02548.
17. Weizenbaum, J. ELIZA—A computer program for the study of natural language communication between man and machine. *Commun. ACM* **1966**, *9*, 36–45. [CrossRef]

18. Colby, K.M.; Weber, S.; Hilf, F.D. Artificial paranoia. *Artif. Intell.* **1971**, *2*, 1–25. [CrossRef]
19. Lowe, R.; Pow, N.; Serban, I.; Pineau, J. The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. *arXiv* **2015**, arXiv:08909.
20. Yang, Z.; Choi, J.D. FriendsQA: Open-Domain Question Answering on TV Show Transcripts. In Proceedings of the 20th Annual SIGdial Meeting on Discourse and Dialogue, Stockholm, Sweden, 11–13 September 2019; pp. 188–197.
21. Sun, K.; Yu, D.; Chen, J.; Yu, D.; Choi, Y.; Cardie, C. Dream: A challenge data set and models for dialogue-based reading comprehension. *Trans. Assoc. Comput. Linguist.* **2019**, *7*, 217–231. [CrossRef]
22. Liu, Z.; Wang, H.; Niu, Z.-Y.; Wu, H.; Che, W.; Liu, T. Towards conversational recommendation over multi-type dialogs. *arXiv* **2020**, arXiv:03954.
23. Xu, J.; Wang, H.; Niu, Z.; Wu, H.; Che, W. Knowledge Graph Grounded Goal Planning for Open-Domain Conversation Generation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; pp. 9338–9345.
24. Ma, Z.; Du, B.; Shen, J.; Yang, R.; Wan, J. An encoding mechanism for seq2seq based multi-turn sentimental dialogue generation model. *Procedia Comput. Sci.* **2020**, *174*, 412–418. [CrossRef]
25. Han, Z.; Zhang, Z. Multi-turn Dialogue System Based on Improved Seq2Seq Model. In Proceedings of the 2020 International Conference on Communications, Information System and Computer Engineering (CISCE), Kuala Lumpur, Malaysia, 3–5 July 2020; IEEE: New York, NY, USA, 2020; pp. 245–249.
26. Wu, T.-W.; Su, R.; Juang, B.-H. A Context-Aware Hierarchical BERT Fusion Network for Multi-turn Dialog Act Detection. *arXiv* **2021**, arXiv:01267.
27. Vinyals, O.; Le, Q. A neural conversational model. *arXiv* **2015**, arXiv:05869.
28. Shang, L.; Lu, Z.; Li, H. Neural responding machine for short-text conversation. *arXiv* **2015**, arXiv:02364.
29. Serban, I.; Sordoni, A.; Bengio, Y.; Courville, A.; Pineau, J. Building End-to-End Dialogue Systems Using Generative Hierarchical Neural Network Models. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
30. Serban, I.V.; Sordoni, A.; Lowe, R.; Charlin, L.; Pineau, J.; Courville, A.; Bengio, Y. A Hierarchical Latent Variable Encoder-Decoder Model for Generating Dialogues. In Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI 2017, San Francisco, CA, USA, 4–10 February 2017; AAAI Press: San Francisco, CA, USA, 2017; pp. 3295–3301.
31. Sordoni, A.; Galley, M.; Auli, M.; Brockett, C.; Ji, Y.; Mitchell, M.; Nie, J.-Y.; Gao, J.; Dolan, B. A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2015, Denver, CO, USA, 31 May–5 June 2015; Association for Computational Linguistics (ACL): Denver, CO, USA, 2015; pp. 196–205.
32. Li, J.; Galley, M.; Brockett, C.; Gao, J.; Dolan, B. A Diversity-Promoting Objective Function for Neural Conversation Models. In Proceedings of the 15th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016, San Diego, CA, USA, 12–17 June 2016; Association for Computational Linguistics (ACL): San Diego, CA, USA, 2016; pp. 110–119.
33. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of the 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, AZ, USA, 2–4 May 2013; International Conference on Learning Representations, ICLR: Scottsdale, AZ, USA, 2013.
34. Pennington, J.; Socher, R.; Manning, C.D. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, Doha, Qatar, 25–29 October 2014; Association for Computational Linguistics (ACL): Doha, Qatar, 2014; pp. 1532–1543.
35. Peters, M.E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; Zettlemoyer, L. Deep Contextualized Word Representations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2018, New Orleans, LA, USA, 1–6 June 2018; Association for Computational Linguistics (ACL): New Orleans, LA, USA, 2018; pp. 2227–2237.
36. Cho, K.; Van Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, Doha, Qatar, 25–29 October 2014; Association for Computational Linguistics (ACL): Doha, Qatar, 2014; pp. 1724–1734.
37. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; Woo, W.-C. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In Proceedings of the 29th Annual Conference on Neural Information Processing Systems, NIPS 2015, Montreal, QC, Canada, 7–12 December 2015; Neural Information Processing Systems Foundation: Montreal, QC, Canada, 2015; pp. 802–810.
38. Grosz, B.J.; Sidner, C.L. Attention, intentions, and the structure of discourse. *Comput. Linguist.* **1986**, *12*, 175–204.
39. Roulet, E. A modular approach to discourse structures. *Pragmatics* **1997**, *7*, 125–146. [CrossRef]
40. Wang, Y.; Ke, P.; Zheng, Y.; Huang, K.; Jiang, Y.; Zhu, X.; Huang, M. A Large-Scale Chinese Short-Text Conversation Dataset. In Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Zhengzhou, China, 14–18 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 91–103.
41. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.-J. Bleu: A Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; pp. 311–318.