Multi-turn Response Selection using Business Sequential Relations in Traffic Field

Yi Zhu[1,2], Zhengzheng Tu[1], Shilei Huang[2], Wanli Lv[1], Lin Yao[2*]
[1]*Anhui Provincial Key Laboratory of Multimodal Cognitive Computation,*
School of Computer Science and Technology, Anhui University, Hefei 230601, China
[2*]*IMSL Shenzhen Key Lab, PKU-HKUST Shenzhen Hong Kong Institute*
Email: [1,2]yi.zhu96@foxmail.com, [1]zhengzhengahu@163.com, [2]shilei.huang@isml.org.cn,
[1]wanly_lv@163.com, [2]yaolin@imsl.org.cn*

***Abstract*—BERT-based models play an essential role and achieve significant results in many tasks of natural language processing (NLP), including dialog tasks. However, there is a limitation of BERT to handle long dialogs. The paper studies how pre-trained model handles long conversations as the input in the multi-turn response selection task. Given the lack of data with long conversations and current dialog tasks are based on engineering applications, it is necessary to collect large dataset with long text sequences. Meanwhile, collecting dialog dataset is a time-consuming work, practical industrial settings usually need more accurate. In this paper, we cite the Chinese Multi-Intention Dialogue (The CMID-Transportation) dataset of transportation customer service, and change it to adapt to response selection task in the dialog systems. To this end, we propose a business-level strategy and use truncation methods to address this problem on the corpus. The experimental results on this corpus show that our proposed approach is fit for the BERT-based model and brings better performance.**

***Keywords*-dialog; response selection; business-level;**

I. INTRODUCTION

In recent years, building intelligent dialogue system is an academic and industrial long-term goal in the field of natural language processing. We are interested in retrieval-based response select ([1]; [2]; [3]; [4]), which is one of the tasks of dialog systems. The state-of-the-art deep learning models have achieved considerable improvement for the retrieval-based response selection tasks. Different matching models for existing response selection works ([1]; [2]; [5]; [6]) are designed for a better utterances-response matching degree. Recently, more and more attentions are paid to solving various problems of multi-turn response selection tasks([7]; [8]) based on the practical BERT model[9].

Due to the memory and computational demands of self-attention based architectures grow quadratically along with sequence length increases, the BERT-based models can't handle the sequences with long text. Meanwhile, the real-world dialog datasets may face the challenge that the input sequences exceed the maximum length of wordpiece token sequences of BERT. In addition, we find that in practical engineering applications, it is very difficult to obtain valuable information by analyzing these long and redundant sentences. Therefore, when BERT-based methods use long dialogs as the input, the entire dialog model might suffer from severe influence.

Table I
AN EXAMPLE FROM THE CORPUS.

	text
Q1	个人办卡(Personal card)
R1	新办...查看解答哦 (Apply for a card for the first time...see the explanation)
Q2	公司车辆(Company vehicle)
R2	您好,公司客车办卡... (Hello, the company bus cards...)
Q3	加急发货(Speed up the delivery)
R3	您好...耐心等待收货 (Hello...patiently wait for delivery)

In order to facilitate related research, a basic and important work is to build a corpus with long conversations. For this purpose, we make full use of CMID-Transportation dataset[10]. The CMID-Transportation dataset is mainly collected from the user's conversation records in the course of traffic customer service. We have further modified the dataset to meet the demands of the dialog task. The corpus is extracted from dialogues of question-answer pairs based on business knowledge. Unlike the previous public datasets (such as E-commerce Corpus[3]), there are a certain number of answers on the corpus. As shown in Table I, we find that the relevance between businesses usually is sequential, where the business is a set of question and answer in human-machine communication.

We have studied the effectiveness of pre-trained language models on this corpus. We propose a business-level strategy and use a truncation method to deal with the input with long sequence in the conversations. In order to adapt to the BERT model, we find that user's question in each business can well express the current intention in the entire conversation. Therefore, we replace the business with user's question, and each business becomes a refined chunk in the context. Then we take refined chunks of each business, last question and response as the input sequence of the pre-trained BERT model. Experimental results on the dialog corpus show that

978-1-6654-9489-2/21/$31.00 ©2021 IEEE

DOI 10.1109/CIS54983.2021.00128

our method is effective for long conversations and the BERT-based model achieves the best result.

The main contributions of our paper are summarized as follows:

- Considering the lack of long dialogs data, we choose and optimize some data with long conversations from CMID-Transportation dataset to meet the needs of multi-turn response selection task.
- We propose a business-level strategy and use a truncation method to process long dialogs in the BERT-based model. The experiments show that the BERT-based model gets better effects for this corpus.

II. Proposed Method

Recently, the pre-trained models, such as BERT[9] and RoBERTa[11], have attracted more and more attention in the academic and the industrial fields. Many BERT-based dialog response selection models hava proposed and achieved better performances than traditional NLP models for dialog response selection and other diverse NLP tasks. In this study, we make use of the BERT-based model for dialog tasks on the corpus.

A. Training Response Selection Models

Given the corpus D, we take a conversational example from the selected data as $(c,\ r,\ y)$, where $c = \{b_1, b_2, ..., b_n, q\}$ represents a context in the conversation. $\{b_k\}_{k=1}^{n}$ is a single business with question and q is the last question from the customer, r is defined as a response candidate and $y \in \{1, 0\}$ denotes a binary label. $y = 1$ means that r is a positive response for c, otherwise $y = 0$. Following previous research [8], we take each utterances-response pair of target language as a token sequence input X for the pre-trained model, as shown as follows:

$$X = [[CLS] b_1\ b_2\ ...\ b_n\ q\ [SEP]\ r\ [SEP]]$$

The input tokens also contain a special $[EOT]$ token at the end of each business. Multi-turn response selection requires learning a match model from any dialog-response pair $(c,\ r)$ and predicting an proper response to the context.

B. Truncation method

We make the analysis for the corpus, and find users' questions are short and obvious. The machine chooses the response from a certain amount of answers, the longest in those answers has 190 words. The texts with excessive length have many useless symbols or noisy information, so it is important to select the key information from the long text sequences. Therefore, the truncation method is used on this corpus, which can be divided into two kinds:

- head-only: Select the first 50 words in the utterance
- tail-only: Select the last 50 words in the utterance

If the important information in a long text sequences is at the beginning, head-truncation is better than tail-truncation. Similarly, when the key information is always at the end of long text sequences, tail-truncation works well for long text sequences.

C. Business-level Strategy

As shown as Table I, the entire dialog is a process of multiple question-answer pairs between the user and the machine. The user asks questions based on their own needs, and the machine chooses the appropriate answer according to the user's question. We define single question-answer pair as a complete business. From these data on the corpus, we can find that business intents contain a business sequential relationship in a whole conversation (such as: card application method -> activation method -> activation success status, etc). Moreover, we also find that the user's questions can often well express the intention of the business, most of these are concise and short. Therefore, we replace the user's questions with the entire business, which can reduce the influence from long content in the entire conversation process.

III. Results

A. The Dialog Corpus

In this paper, we optimize some new data in CMID-Transportation dataset, and build a foundation for the research and application in the aspect of multi-turn response selection.

1) Data Collection: We select some data from the training data of CMID-Transportation. According to the customers' request and experience, the machine's replys in the dataset are replaced with new answers and selected in a fixed number of answers. We divide these data into 13 main categories and there are many small businesses under each major categories. To ensure a higher quality, there are more than 0.2 million conversations in these data, and a total of more than 200 categories of intents have been relabeled. More importantly, each intent can be matched to the corresponding business.

2) Corpus Creation: In these data, we select 0.1 million traffic conversations as the training data, 5K traffic conversations as the validation data, and 1K traffic conversations as the testing data. We count the number of businesses in each dialogue (see Fig. 1). In order to ensure that the business distribution on the corpus is homogeneous and the conversations have no overlap, we have done a good calculation and classification before labeling. For each dialogue, the final reply is taken as the true response, and the wrong responses are decided by randomly selecting the answer from the associated categories. The positive and negative ratios we set on the corpus are the same as the public dataset(such as E-commerce Corpus[3]), training data and validation data are set to 1:1, and testing data are set to 1:9. In addition, chinese word segmentation for a text sequences or a sentence

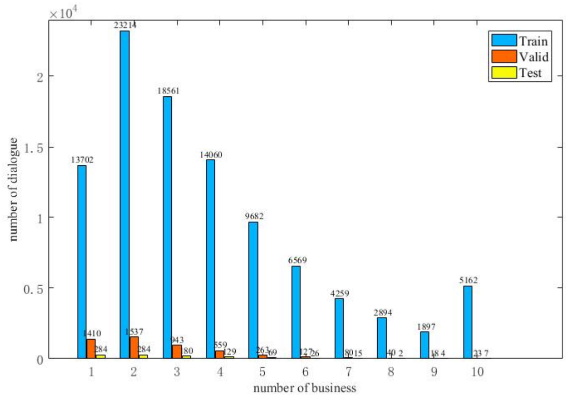

Figure 1. Distribution of business in our selected data. The business is a question-answer pair.

is the primary step by using the Jieba[1] Toolkit. Table II gives the statistics of the corpus.

B. Setting

For the evaluation, we employ recall at position k in n candiates $(R_n@k)$ as evaluation metrices, which are are used in previous researches([1]; [2]; [5]; [6]; [8]). Based on researches for the corpus and dialogs, we set important parameters in the compared methods. The maximum length of context is set to 10 for computational efficiency. We also set the maximum utterance length to 50 and capture the first 50 words in the long sentence.

Our pre-trained model uses the whole-word masking (WWM) strategy for Chinese BERT[2]. We use the first 50 words in the each utterance. In the fine-tuning course, the batch size is set to 20, training model uses Adam algorithm as the model optimizer and the initial learning rate is set to 0.00001. Considering that there are more words in the context and response, we truncate them in experiments.

We introduce the compared methods, including: Sequential Matching Network (SMN)[1] is proposed to utilize each utterance in the context to match the response. Deep Attention Matching Network(DAM)[5] uses attention mechanism to obtain matching vectors between the context and response in the conversations, which is inspired by the Transformer architecture [12]. Interaction-Over-Interaction Network(IOI)[2] obtains matching information within the conversational context-response pair via stacking multiple interaction blocks. The Utterance-To-Utterance Interactive Matching Network(U2U-IMN)[13] gets the matching information between contexts and responses to take them as the sequences of utterances. Multi-hop Selector Network(MSN)[6] matches the filtered context by a multi-hop selector with the given response candidates.

[1]https://github.com/fxsjy/jieba
[2]https://github.com/ymcui/Chinese-BERT-wwm

Table II
STATISTICS OF THE CORPUS.

	Train	Valid	Test
# context-responses pairs	200K	10K	10K
# candidates per context	2	2	10
Max # turns per context	19	19	19
Avg # turns per context	6.63	4.16	4.27
Avg # words per question	7.94	4.77	4.74
Avg # words per context	50.39	43.93	44.01
# the number of answers	228	212	185
Max # words of answer	192	192	192
Avg # tokens of answers	127.37	134.41	129.12

Table III
COMPARISON OF DIFFERENT MODELS ON THE CORPUS.

	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$
SMN-200	0.412	0.561	0.881
SMN	0.499	0.708	0.919
DAM	0.538	0.761	0.921
IOI	0.614	0.760	0.921
MSN	0.612	0.762	0.934
U2U-IMN	0.636	0.779	0.948
Business-model	0.685	0.826	0.953

C. Experimental Results

Table III lists the relevant experimental results of our proposed Business-Strategy approach and the compared models on the corpus. Only using first 50 words in each utterance is more effective than using 200 words of the utterance used by the SMN-method, since it is not easy to capture useful words and utterances in the long sentence. Our Business-Strategy model also selects the first 50 words to fine-tune it, outperforms the previous non-pretrained models and achieves improvement of 4.9% in $R_{10}@1$ on this corpus. These results show that our solution based on business-policy and truncation can effectively solve the data problem of long utterance on the large corpus.

D. Further Analysis

1) Truncation methods: We select 50 words for truncation methods and investigate the effect of head-truncation and tail-truncation in the SMN method. The experimental results on this corpus are summarized in Table IV. Depending on the given answer, we can find that head-truncation achieves the better performances than tail-truncation, as shown in Table IV. The results demonstrate that the head-truncation can capture more critical information, and some important words at the beginning of sentences could be selected for multi-turn response selection task.

2) Utterance length: What we found from the data is that the machine's replies are very long and could bring more noisy messages. Therefore, we carry out sentence length

Table IV
COMPARISON OF TRUNCATION METHODS ON THIS CORPUS.

	R_{10}@1	R_{10}@2	R_{10}@5
tail-truncation	0.448	0.667	0.904
head-truncation	0.499	0.708	0.919

experiments of 200 words and the first 50 words in the SMN method, and the result shows that the first 50 words is more effective than 200 words. Considering that the pre-trained model has the fixed input, we also select first 50 words for context utterances and response in the BERT-based model. However, this is a pity that few key information in latter half of answer have not been chosen. In future work, we will study that if category information could be necessary for the development of dialogue systems on the corpus.

IV. CONCLUSIONS

In this paper, we consider of the limitation of BERT to handle longer input sequences and the lack of real-world dialog datasets with long converstaions. We optimize the CMID-Transportation dataset of transportation customer service, which can meet the needs of dialog tasks. Considering the particularity of the data, we propose an effective business-level strategy and use truncation methods for multi-turn response selection tasks, which can take full advantage of large-scale pre-trained BERT model. The experiment results prove the practicability of business-strategy and truncation method, the pre-trained model achieves best result on the corpus. Our optimized corpus is released at https://github.com/zhuyi96/dialog-data.

ACKNOWLEDGMENT

The work was supported by the Natural Science Foundation of Anhui Higher Education Institution of China (No.KJ2020A0033), the Anhui Natural Science Foundation Anhui energy Internet joint fund(No.2008085UD07), the Joint Funds of the National Natural Science Foundation of China (No.U20B2068), the National Natural Science Foundation of China (No.61976003), NSFC Key Project of International (Regional) Cooperation and Exchanges (No.61860206004), the Anhui Provincial Natural Science foundation (No.2108085MF211), Shenzhen Basic Research Program (No.JCYJ20180507182908274).

REFERENCES

[1] Y. Wu, W. Wu, C. Xing, M. Zhou, and Z. Li, "Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2017, pp. 496–505.

[2] C. Tao, W. Wu, C. Xu, W. Hu, D. Zhao, and R. Yan, "One time of interaction may not be enough: Go deep with an interaction-over-interaction network for response selection in dialogues," in *Proceedings of the 57th annual meeting of the association for computational linguistics*, 2019, pp. 1–11.

[3] Z. Zhang, J. Li, P. Zhu, and H. Zhao, "Modeling multi-turn conversation with deep utterance aggregation," in *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*, 2018, pp. 3740–3752.

[4] J.-C. Gu, Z.-H. Ling, and Q. Liu, "Interactive matching network for multi-turn response selection in retrieval-based chatbots," in *proceedings of the 28th ACM international conference on information and knowledge management*, 2019, pp. 2321–2324.

[5] X. Zhou, L. Li, D. Dong, Y. Liu, Y. Chen, W. X. Zhao, D. Yu, and H. Wu, "Multi-turn response selection for chatbots with deep attention matching network," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018, pp. 1118–1127.

[6] C. Yuan, W. Zhou, M. Li, S. Lv, F. Zhu, J. Han, and S. Hu, "Multi-hop selector network for multi-turn response selection in retrieval-based chatbots," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 111–120.

[7] J. Lu, X. Ren, Y. Ren, A. Liu, and Z. Xu, "Improving contextual language models for response retrieval in multi-turn conversation," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, pp. 1805–1808.

[8] T. Whang, D. Lee, D. Oh, C. Lee, K. Han, D.-h. Lee, and S. Lee, "Do response selection models really know what's next? utterance manipulation strategies for multi-turn response selection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

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

[10] J. Song, Q. Luo, and J. Nie, "Research and application of multi-round dialogue intent recognition method," in *2020 16th International Conference on Computational Intelligence and Security (CIS)*. IEEE, 2020, pp. 131–135.

[11] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[12] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[13] J.-C. Gu, Z.-H. Ling, and Q. Liu, "Utterance-to-utterance interactive matching network for multi-turn response selection in retrieval-based chatbots," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 369–379, 2019.