Sustainable Energy Technologies and Assessments 55 (2023) 102896

Contents lists available at ScienceDirect

Sustainable Energy Technologies and Assessments

journal homepage: www.elsevier.com/locate/seta

A role distinguishing Bert model for medical dialogue system in sustainable smart city

Suixue Wang [a,1], Shuling Wang [b], Zhuo Liu [c], Qingchen Zhang [d,*]

[a] School of Information and Communication Engineering, Hainan University, Renmin Road, Haikou, 570228, Hainan, China
[b] Haikou Affiliated Hospital of Central South University Xiangya School of Medicine, Renmin Road, Haikou, 570208, Hainan, China
[c] The First Affiliated Hospital of Dalian Medical University, Zhongshan Road, Dalian, 116000, Liaoning, China
[d] School of Computer Science and Technology, Hainan University, Renmin Road, Haikou, 570228, Hainan, China

ARTICLE INFO

Keywords:
Smart sustainable city
Computing intelligence
Medical dialogue system
Bert

ABSTRACT

Smart medicine is a vital component for building sustainable smart city. In smart medicine, intelligent dialogue system is playing an important role in providing personalized and efficient healthcare services to improve the quality of life for human beings. The Bert model has become a popular way to construct intelligent medical dialogue system. However, the current Bert model is difficult to achieve desirable results for this task since its input does not reflect the difference in roles. To overcome this drawback, we present a role distinguishing Bert model for intelligent medical dialogue system to help construct the sustainable smart city. Particularly, we segment and label the utterances depending on different dialogue roles, and then construct the corresponding segment embedding as the input of our model. Furthermore, we substitute the NSP task with the SOP task to better learn the coherence between sentences. Finally, we verify the proposed model by comparing it with Ernie on some online E-commerce datasets for intent recognition, semantic matching, and session dialogue classification. The results demonstrate that our proposed model improves the average 1% accuracy for different tasks in dialogue system, proving the potential of the proposed model for establishing intelligent medical dialogue system in smart city.

1. Introduction

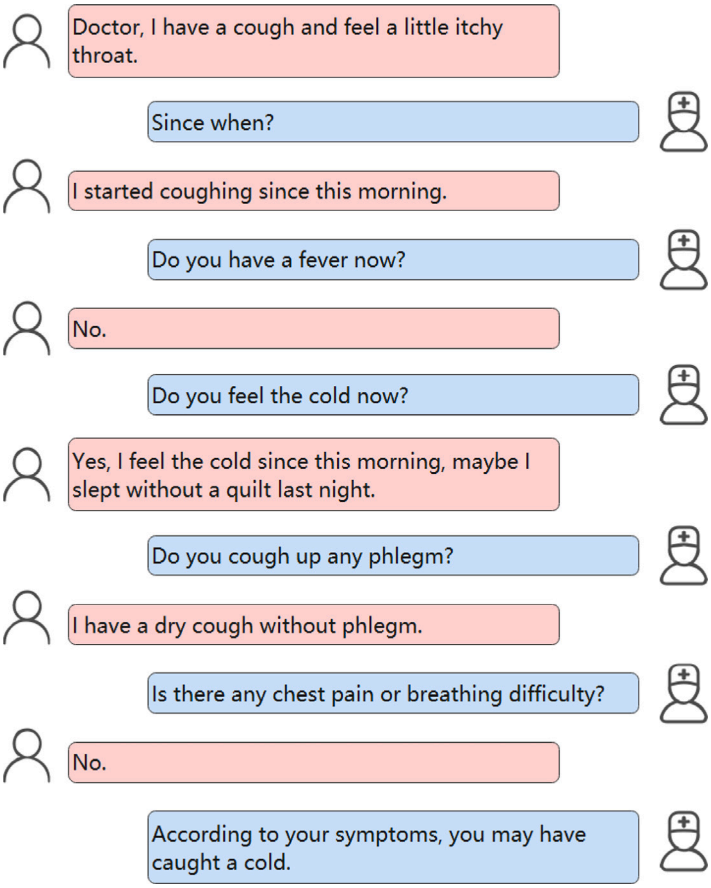
Smart sustainable city aims to improve the quality of life for human beings with various information technologies for improving the efficiency of resource utilization and optimizing urban management and services [1,2]. In detail, commonly used information technologies include computing intelligence especially deep learning, big data analytics, internet of things, and cloud computing, while important components of a smart sustainable city include smart medicine, intelligent transportation, smart factory, smart E-commerce and so on. For example, many sensors are deployed to construct intelligent transportation systems to address traffic congestion while the use of industrial robots efficiently enhances factory productivity and saves energy. As another example, smart E-commerce can recommend highly-quality services and goods to users with the help of personalized recommendation algorithms in big data, and furthermore, it provides the function of smart customer service such as intelligent dialogue system to save energy and improve work efficiency for sustainable urban development. Another vital component of smart sustainable city is smart medicine which offers precision diagnosis and treatment services to human beings by combining various artificial intelligence techniques especially deep learning with medical theory. Smart medicine consists of computer-aided diagnosis, medical dialogue system, and clinical decision support, etc. In particular, medical dialogue system can help to recognize the patient's intent for hospital guide, computer-aided diagnosis and doctor recommendation quickly, and more importantly, it can improve medical efficiency and save the medical resource for sustainable medicine development.

In detail, a medical dialogue system firstly analyzes the semantics to recognize the patent's intent through the patient's questions or description. Accordingly, the system guides the patient to accurately describe their symptoms through a question-and-answer format. After several rounds of dialogue, the system informs the patient of a possible diagnosis or advises the patient to take the next steps, for example, making an appointment with a suitable doctor or taking the appropriate medical examination. Fig. 1 shows an example of a medical dialogue.

Apparently, the key to construct a medical dialogue system is semantic matching and intent recognition which can be realized by

[*] Corresponding author.
E-mail address: zhangqingchen@hainanu.edu.cn (Q. Zhang).
[1] Suixue Wang, Shuling Wang, and Zhuo Liu are the co-first authors.

https://doi.org/10.1016/j.seta.2022.102896
Received 17 July 2022; Received in revised form 5 October 2022; Accepted 27 October 2022
Available online 30 November 2022



Doctor, I have a cough and feel a little itchy throat.

Since when?

I started coughing since this morning.

Do you have a fever now?

No.

Do you feel the cold now?

Yes, I feel the cold since this morning, maybe I slept without a quilt last night.

Do you cough up any phlegm?

I have a dry cough without phlegm.

Is there any chest pain or breathing difficulty?

No.

According to your symptoms, you may have caught a cold.

Fig. 1. Example of medical dialogue.

some computing intelligence technologies such as deep learning. As a representative deep learning model, the Bert model has gained a wide range of applications and great success in the field of natural language processing [3]. It is worth mentioning that the Bert model has been broadly used to develop various intelligent dialogue systems, especially in smart e-commerce and smart medicine.

However, the Bert model and its variants are hard to achieve satisfactory results in these tasks, since they do not distinguish the different roles of the dialogue in their input. Essentially, the word distributions between user utterances and system responses are quite different. To overcome the drawback of the original Bert model, we propose a role distinguishing Bert model to construct a high-quality medical dialogue system for building a sustainable smart city. Especially, we separate the utterances in multi-turn conversations by assigning two different segmentation tags for two kinds of dialogue participants so that we refine the corresponding segment embeddings to differentiate tokens from sentences that are spoken by different roles. In addition, the original Bert model adopts Next Sentences Prediction (NSP) for a binary classification task during pre-training stage, typically predicting whether the sentence pairs within a document appear continuously or not, which is not suitable for building a multi-round medical dialogue system. Inspired by the variants of the original Bert model, namely ALBERT and Ernie, which use sentence-order prediction (SOP) task as a self-supervised loss to mainly focus on inter-sentence coherence [4, 5], we employ the SOP task instead of the NSP task in our model. We conduct an experiment on some online E-commerce datasets to evaluate the performance of the presented role distinguishing Bert model for three tasks closely associated with the medical dialogue system construction, namely intent recognition, semantic matching, and session dialogue classification. Results display that the presented model achieves 0.8%, 0.8%, and 1.3% higher accuracy than Ernie [5] on the three downstream tasks, respectively. Such results substantially prove the potential of the proposed model for establishing intelligent medical dialogue system in sustainable smart city.

To summarize, our contributions are as follows:

- We propose a role distinguishing Bert model for intelligent medical dialogue system that can be used to construct a sustainable smart city.
- Considering there are two different dialogue roles in the medical dialogue system, we segment and label the utterances by two kinds of segmentation tags, and then construct the corresponding embedding as the input of our model.
- We conduct extensive experiments on several downstream tasks compared with the original Bert model and Ernie, and the results confirm the effectiveness of our proposed model.

The rest of this paper is organized as follows. In Section 2, we conduct a survey on the related work about the medical dialogue system, especially the work based on deep learning. Section 3 details the role distinguished Bert model with its training algorithm. Section 4 shows and analyzes the experimental results. Finally, Section 5 concludes the work and points out future work.

2. Related work

A good medical automatic dialogue system can significantly improve the efficiency of medical service, save medical resources and reduce the energy consumption, and eventually help to build a sustainable smart city. A large number of studies have been done to construct medical dialogue systems in recent years. In this study, we design a scheme based on the Bert model, a representative deep learning model for natural language processing, for a medical dialogue system, so we

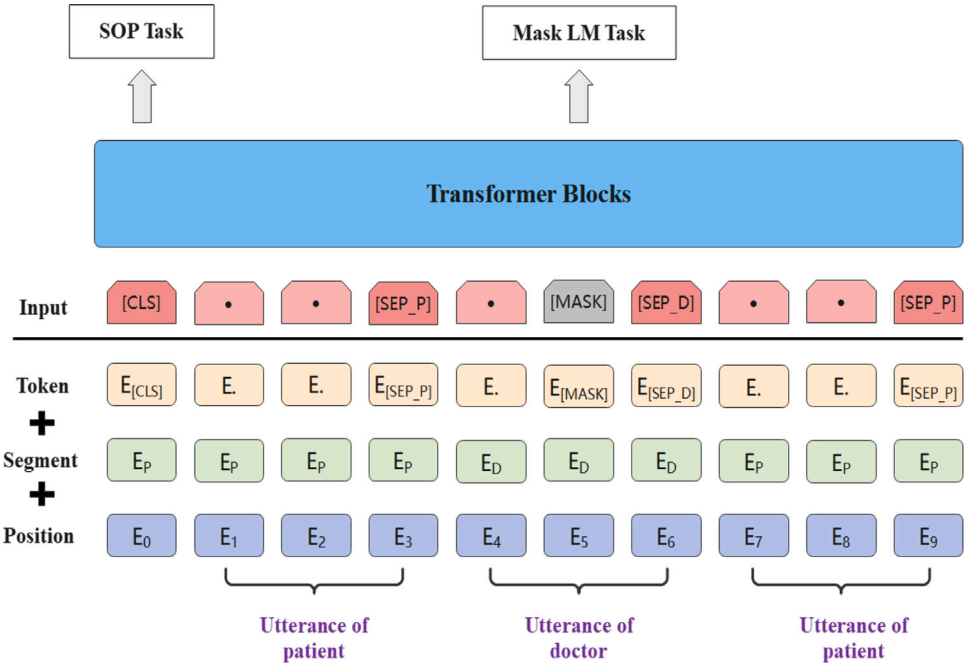


Fig. 2. Network overview of the Role Distinguishing Bert Model.

focus on surveying the representative related work for the medical dialogue system based on deep learning.

Wei et al. [6] develops a task-oriented dialogue system for automatic medical diagnosis whose three important modules are natural language understanding (NLU), dialogue management (DM), and natural language generation (NLG), respectively. This system first chats with a patient to collect his/her symptoms beyond the patient's self-reports, and then suggests a diagnosis result automatically. In this system, intent recognition and semantic matching are two key sub-modules in NLU, which are utilized to understand the user's intent from utterances. Khaldoon et al. [7] proposed an automated COVID-19 dialogue system based on deep learning, which aims to provide appropriate medical advice according to the patients' symptoms during the dialogue between patients and doctors. The model extracts two types of important vectors namely symptom vectors and doctor utterance vectors, and the two vectors are encoded by the same deep learning method. In [8], Yang et al. builds a deep learning-based medical recommendation system for modeling the diagnostic multi-round question answering (QA) records. This model represents the utterances from patients or doctors by using the same word embedding layer, therefore it is lack of capability to distinguish the different roles. Currently, nearly all top-performing biomedical question-answering systems use domain-specific pre-training language models such as BioBERT, PubMedBERT, or clinical BERT [9–11]. Nevertheless, these models are pre-trained on biomedical corpora such as PubMed abstracts, PubMed Central full-text articles and clinical narratives, the medical dialogue corpora were not used to pre-trained, so the models are not suitable for a multi-round dialogue system.

More recently, the Bert model has shown its powerful capability of natural language processing and obtained a broad application in this area since it has been devised. Not surprisingly, it also has been developed various dialogue systems in some fields such as online E-commerce, intelligent transportation, and smart medicine for building sustainable smart cities. However, it ignores the importance of different roles of utterances in its input during the training and it utilizes the NSP task that is not suitable for building a multi-round dialogue system. The above two drawbacks limit its ability for intent recognition and semantic matching. So we improve the original Bert model for a good medical dialogue system, which is described in the following section.

3. Role distinguishing Bert model

Given an initial description from a patient, such as the sentence "Doctor, I often feel a loss in appetite and nausea"., the goal of a medical dialogue system is to guide patients to accurately describe their symptoms by a multi-round questioning and answering, and accordingly offer the patient with a possible diagnosis or further examination suggestion. To achieve this goal, we design a role distinguishing Bert model that consists of three layers, namely input layer, hidden layer, and output layer, the network overview of the role distinguishing Bert model is shown in Fig. 2.

The input layer is composed of position embedding, segment embedding, and token embedding. The position embedding is responsible for capturing the order information of the sequence. Especially, we adopt the position embedding of the original Bert model [3] in our study, with the following equations :

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{\text{model}}}) \tag{1}$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}}) \tag{2}$$

where pos denotes the position of each token in the sequence and d_{model} is the dimension of a vector for encoding a token.

In the original Bert model, the segment embedding is used to separate two adjacent sentences. Different from the original Bert model, in our study, the segment embedding is devised to confirm the role of each sentence in a dialogue. In particular, a medical dialogue system has two roles of patients and doctors, so we have two kinds of segmentation embeddings such as E_p and E_d. Similar to the original Bert model, there is also a learned embedding for every segmentation token, which is used to denote whether it belongs to the utterance of patients or the utterance of doctors. In the original Bert model, only a sort of special segmentation tag such as $[SEP]$ is used in the token embedding. However, in our study, two special segmentation tags namely $[SEP\_P]$ and $[SEP\_D]$ are utilized to indicate the role of each sentence. Especially, $[SEP\_P]$ indicates the utterance in front of it is from a patient, and the utterance in front of $[SEP\_D]$ is from a doctor. In the token embedding, we also use learned token embedding to convert the input tokens to vectors of dimension d_{model}.

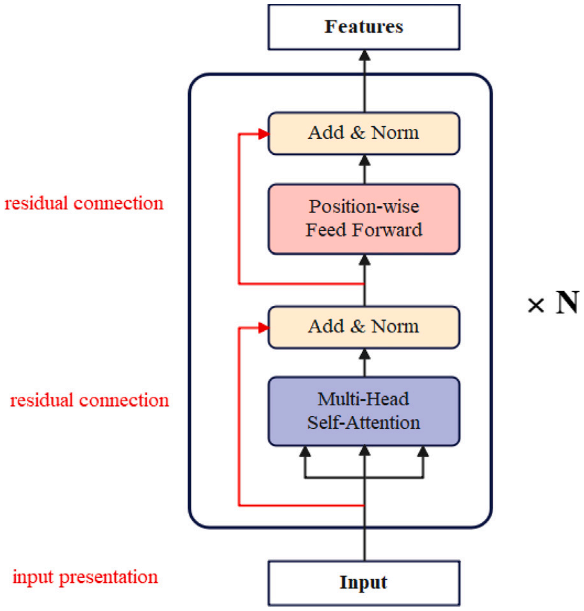


Fig. 3. Transformer encoder blocks.

In our study, we concatenate all utterances in a session of dialogue as an input sequence and use the corresponding segmentation tags [SEP_P] or [SEP_D] to separate different types of roles. For a given token in the input sequence, its input representation is constructed by summing the corresponding token, segment, and position embeddings.

In this work. we use a stack of N identical blocks, which is the same as the Transformer encoder that is implemented originally in [12] to construct the hidden layer. As displayed in Fig. 3, each block has two modules, including a multi-head self-attention module and a position-wise fully connected feed-forward network (FFN). Around each of the two modules, a residual connection is employed and followed by layer normalization.

In detail, Multi-head self-attention module passes through an attention mechanism multiple times concurrently. The outputs of all independent attentions are concatenated and followed by linear transformation into the dimension of expectation. In practice, multiple attention heads allow for capturing dependencies of various ranges (e.g. shorter-term dependencies vs longer-term dependencies) within a sequence. For a single self-attention head, the goal is to capture the dependencies between the tokens in the sequence and utilize that information to learn the internal structure of the sequence. For example, if we take the sequence “the cat didn’t cross the road because it was too tired”, we would expect the module to learn that “it” pays the most strong attention to its associated noun phrase “the cat”. Additionally to the multi-head self-attention module, a fully connected FNN which consists of two fully connected dense layers with a ReLU activation in between is added to the block, and the same dense layer is used for each position respectively and identically. The FFN is formalized as:

$$FFN(x) = \max\left(0,\ xW_1 + b_1\right) W_2 + b_2 \tag{3}$$

where W_1, b_1, W_2, b_2 are learnable parameters and x is input. The dimensions of input and output are identical, which is d_{model}, and the dimension of inner-layer is d_{ff}.

The residual connection and layer normalization are represented as:

$$x' = LayerNorm(x + FFN(x)) \tag{4}$$

There are two unsupervised tasks such as mask language model (MLM) and next sentence prediction (NSP) employed in the output layer of the original Bert model [3]. In this work, we use the same MLM task implemented in the original Bert model and substitute the sentence order prediction (SOP) task [4] for the NSP task.

MLM task randomly chooses and masks a certain percentage of the input tokens, and then follows by predicting those masked tokens. In detail, the vectors of the last hidden layer which corresponds to the mask tokens are entered into a softmax over the vocabulary, and the outputs of the softmax corresponding to the mask tokens are utilized to calculate the cross entropy loss L_{mask} with original tokens.

Papers like ALBERT, XLNET, and RoBERTa found that the NSP was ineffective when the pre-training model is applied to downstream tasks [4,13,14]. After removing the NSP task, the performance is improved across several tasks. In [4], ALBERT proposes a new task of SOP to substitute for the NSP task, and the SOP task can raise performance on downstream multi-sentences encoding tasks. Identically, the SOP task uses a binary classification loss L_{sop} which is the same as the NSP task to mainly focus on inter-sentence coherence.

The training loss is the sum of the mean MLM loss and the mean SOP loss, which can be represented as:

$$L = L_{mask} + L_{sop} \tag{5}$$

Finally, learning the parameters aims to minimize the objective function, so we apply the backpropagation algorithm to train the parameters.

4. Experiments

In the stage of pre-training, our proposed role distinguishing Bert model is pre-trained on a large-scale corpus for 25 days on six RTX 3090 GPUs with 24G memory, and we adopt a strategy of pre-training from scratch. In detail, the corpus is collected from a Chinese E-commence customer service, which contains over 50 million dialogue sessions and more than 600 million utterances from both users and customers. The average turn of a dialogue session is 10.6. The utterance of user and customer have different average lengths, which are 19 and 36.

In the experiment of fine-tuning, we evaluate the presented role distinguishing Bert model by comparing it with the original Bert model and Ernie [5], Ernie is a popular variant of the Bert model that is suitable for Chinese NLP tasks. Especially, we compare their performance for intent recognition and semantic matching which are key components to build a good medical dialogue system. Furthermore, we evaluate the performance of the presented model for another task of session dialogue classification. These tree downstream tasks are conducted on several E-commerce datasets collected from some large-scale online e-commerce platforms. All the experiments are conducted on a RTX 3090 GPU successively.

For the stages of pre-training and fine-tuning, Adam optimizer is used to optimize the loss with learning rate 5e-5 and a linear decay scheduler is applied to adjust the learning rate automatically. Our model has the same configuration as the original Bert model: 12 transformer blocks, a hidden size of 768, and 12 self-attention heads. In the pre-training stage, the maximum sequence length is fixed to 512 and the batch size is set to 50. During the stage of fine-tuning, the setting of the maximum sequence length and the batch size depends on the type of downstream task. Specifically, for the session dialogue classification task, the maximum sequence length and the batch size are the same as the pre-training stage. For the other downstream tasks, the maximum sequence length is fixed to 50 and the batch size is set to 200.

First, we conduct the experiments on the dataset for three tasks of intent recognition, semantic matching and session dialogue classification for five times. The results on average accuracy are reported in Table 1.

From the results, the designed role distinguishing Bert model yields 0.8%, 0.8%, and 1.3% average higher accuracy than Ernie for intent recognition, semantic matching, and session dialogue classification, respectively. Furthermore, the proposed model are 1.5%, 1.4%, and

Table 1
Average accuracy on three tasks.

Task Type	Original Bert	Ernie	Our model
Intent recognition	91.2%	91.9%	**92.7%**
Semantic matching	86.5%	87.1%	**87.9%**
Session dialogue classification	71.2%	72.1%	**73.4%**

Table 2
Average accuracy on word-of-mouth evaluation classification.

Model	30 per category	3000 per category
Ernie	81.6%	91.5%
Our model	**86.2%**	92.1%

Table 3
Average accuracy on work order labeling.

Model	1% training set	5% training set	10% training set	100% training set
Ernie	61.6%	78.2%	81.3%	83.4%
Our model	**75.3%**	83.9%	85.4%	85.6%

2.2% better than the original Bert model, respectively. Such results fully demonstrate that the proposed model is more effective for building automatic dialogue system than Ernie and the original Bert model. More importantly, our model produces significantly outperforms Ernie for the third task of session dialogue classification which sufficiently proves the importance of role distinguishing introduced into the Bert model for building an intelligent dialogue system.

Generally speaking, the number of training samples in the area of medicine is limited. To evaluate the generalization performance of the designed role distinguishing Bert model, we perform our proposed model and Ernie for word-of-mouth evaluation classification and work order labeling which are two tasks closely related to dialogue generation on datasets with a small number of labeled training samples. Tables 2 and 3 report the results.

Not surprisingly from the results, as the number of labeled training samples increases, the average accuracy produced by the designed role distinguishing Bert model and Ernie gradually improves. When the labeled training samples are few, our designed model produces significantly higher accuracy than Ernie for the two tasks. For example, our model obtains 4.6% higher accuracy than Ernie for word-of-mouth evaluation classification when the training set contains only 30 labeled samples per category. The more representative example is shown in Table 3. Our model achieves 13.7% higher accuracy than Ernie for work order labeling when only 1% of the labeled training samples are used. Such results fully argue that our designed model has good generalization and is suitable for cold start tasks, which is vital to build a medical dialogue system.

5. Conclusion

In this paper, we presented a role distinguishing Bert model for medical dialogue system to build sustainable smart city. Especially, we construct the segment embedding as the input of the presented model by labeling the utterances depending on different dialogue roles. In this way, the model can learn the difference of the utterances of different dialogue roles to improve the accuracy of intent recognition and semantic matching which are two important components to build a good medical dialogue system. Experimental results on online E-commerce datasets clearly argue that our designed model produced an average 1% higher accuracy than the original Bert model. Such results prove that our presented model has great potential to build a good medical dialogue system because an online E-commerce dialogue system is similar to a medical dialogue system. Furthermore, our presented model can potentially improve the efficiency of medical services and reduce the energy consumption which is important to build sustainable smart city. In future work, we will evaluate the presented model on real medical datasets to further validate its performance.

CRediT authorship contribution statement

Suixue Wang: Methodology, Validation, Writing – original draft, Data curation. **Shuling Wang:** Conceptualization, Investigation. **Zhuo Liu:** Conceptualization, Writing – review & editing. **Qingchen Zhang:** Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This paper is supported by the Key Research and development Project of Hainan Province (Grant No. ZDYF2021GXJS016) and National Natural Science Foundation of China (Grant No. 6216070295).

References

[1] Silva BN, Khan M, Han K. Towards sustainable smart cities: A review of trends architectures, components, and open challenges in smart cities. Sustain Cities Soc 2018;38:697–713.

[2] Chang J, Kadry SN, Krishnamoorthy S. Reivew and synthesis of big data analytics and computing for smart sustainable cities. IET Intell Transp Syst 2020;14(11):1363–70.

[3] Devlin Jacob, Chang Ming-Wei, Lee Kenton, Toutanova Kristina. 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, 1(Long and Short Papers). 2019, p. 4171–86.

[4] Lan Zhenzhong, Chen Mingda, Goodman Sebastian, Gimpel Kevin, Sharma Piyush, Soricut Radu. Albert: A lite bert for self-supervised learning of language representations. In: Proceedings of international conference on learning representations. 2019.

[5] Sun Yu, Wang Shuohuan, Li Yukun, Feng Shikun, Chen Xuyi, Zhang Han, et al. Ernie: Enhanced representation through knowledge integration. In: Proceedings of ACL. 2019, p. 1441–51.

[6] Wei Zhongyu, Liu Qianlong, Peng Baolin, Tou Huaixiao, Chen Ting, Huang Xuan-Jing, et al. Task-oriented dialogue system for automatic diagnosis. In: Proceedings of the 56th Annual meeting of the association for computational linguistics, vol. 2. 2018, p. 201–7.

[7] Alhussayni Khaldoon H, Alshamery Eman S. Automated COVID-19 dialogue system using a new deep learning network. Periodi Eng Nat Sci 2021;9:667–77.

[8] Zhan Yang, Wei Xu, Runyu Chen. A deep learning-based multi-turn conversation modeling for diagnostic Q & A document recommendation. Inf Process Manage 2021;58(3):102485.

[9] Lee Jinhyuk, Yoon Wonjin, Kim Sungdong, Kim Donghyeon, Kim Sunkyu, So Chan Ho, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020;36(4):1234–40.

[10] Gu Yu, Tinn Robert, Cheng Hao, Lucas Michael, Usuyama Naoto, Liu Xiaodong, et al. Domain-specific language model pretraining for biomedical natural language processing. ACM Trans Comput Healthcare 2022;3(1):1–23.

[11] Alsentzer Emily, Murphy John, Boag William, Weng Wei-Hung, Jindi Di, Naumann Tristan, et al. Publicly available clinical BERT embeddings. In: Proceedings of the 2nd Clinical natural language processing workshop. Minneapolis, Minnesota, USA: Association for Computational Linguistics; 2019, p. 72–8.

[12] Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, Gomez Aidan N, et al. Attention is all you need. Adv. Neural Inf. Process. Syst. 2017;30:5998–6008.

[13] Liu Yinhan, Ott Myle, Goyal Naman, Du Jingfei, Joshi Mandar, Chen Danqi, et al. Roberta: A robustly optimized bert pretraining approach. In: Proceedings of international conference on learning representations. 2020d.

[14] Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov RR, Le QV. Xlnet: Generalized autoregressive pretraining for language understanding. Adv Neural Inf Process Syst 2019;32:5753–63.