Proceedings of the 3rd International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/5/20230575

Clinical reasoning auxiliary model based on deep learning

Yixin Yang

School of information studies, Syracuse University, NY, 13210, USA

yyang72@syr.edu

Abstract. Medical data is increasing rapidly. The rapid fusion of medical record text data, medical laboratory data and image data in many regional hospitals have brought light to the screening, diagnosis and treatment of diseases. Medical text data records the patient's detailed condition and treatment process, with abundant information. As a branch of artificial intelligence, deep learning is gradually penetrating the medical field. The application of deep learning in the medical field shows great application prospects. Therefore, mining the knowledge contained in medical texts can provide better medical services for the majority of patients, which has good theoretical and practical significance. In this paper, the content of medical texts was analyzed, and the test questions and standard answers of the USMLE® Step 2 Clinical Skills Exam were used as data sets to explore the core content of the medical chief texts. After comparing and analyzing different methods, such as BERT and DeBERTa, an intelligent follow-up system based on DeBERTa was finally proposed, hoping to help clinicians read medical data more effectively and weaken the problem of poor diagnosis results caused by personal academic ability differences.

Keywords：Clinical data, Natural language processing, BERT, DeBERTa.

1. Introduction

1.1. Background

Medical data is growing significantly in volume. The speedy integration of text-based medical records, test findings, and imaging data has improved disease screening, diagnosis, and treatment in many community hospitals. The status of the patient and the plan of therapy are extensively described in medical text data. People start to analyze the data using machine learning in order to uncover the important information buried behind these medical data.

A total of 215 journal papers published from 2000 to December 2021 may be found by searching the phrase "Machine learning in the medical area field" in Springer, Elsevier, IEEE, and other renowned publications[1]. And the number of releases is on the rise. Machine learning methods are used in computer-aided diagnosis (CAD) applications. This algorithm learns from many diagnostic samples collected in Medical Notes and combines expert diagnoses to support the prediction and diagnosis of future diseases by medical experts. Machine learning is useful for enhancing the dependability, effectiveness, and accuracy of diagnostic systems for certain illnesses.[2][3][4].

Deep learning-based models have achieved remarkable success in various medical tasks supporting disease detection and diagnosis. Existing deep learning technologies can be divided into four

DOI: 10.54254/2755-2721/5/20230575

categories: supervised learning, unsupervised learning, semi-supervised learning, and performance enhancement strategies. Common applications of deep learning include classification, segmentation, detection and registration. Deep learning has various uses in the medical industry, with categorization being the most popular one.

It can be found that people have done a lot of research on text classification in the medical field. In a review paper, only three of 72 studies of medical texts used deep learning[5]. This indicates that research in the future can use and study DL algorithms (e.g., CNN, RNN, LSTM, BERT) to classify clinical data. Different degrees of abstraction can be learned from data using deep learning computational algorithms. The creators of these features weren't actual people. As opposed to being manually constructed features, which might not be able to duplicate the vectors from the training set, the features are instead automatically learned from the training data by the generic learning process. The information at hand suggests that there is still significant opportunity for NLP growth in the medical area and that clinical NLP techniques trail behind those employed in the NLP community.

1.2. Research Contents

The research goal of this article is to understand Patient Notes, specifically, to identify specific features in patient notes. This research will focus on how to interpret the patient's symptoms and identify the clinical concepts in the case. Specifically, mapping clinical concepts from the exam scoring rubric to explicit clinical concepts of medicine written by medical students.

Since the precision of the detection and diagnosis of cancer or many other diseases depends on the experience of the particular doctor, reading and understanding medical material differs widely between readers. In order to assist physicians in reading medical data more efficiently and making diagnostic judgments in a more precise objective, and professional manner, computer-aided detection and diagnosis tools are used to help solve this clinical difficulty[6].

Data mining and free-text analysis have entered a new era thanks to natural language processing (NLP) tools. And the most cutting-edge approaches people use for medical text categorization are all related to the Transformer[7][8].

An advancement in NLP uses the bidirectional encoder Representation from the Converter, a representation model based on self-attention (BERT)[9]. The BERT model concentrates on general human language understanding and discriminates between various word use scenarios by pre-training a sizable text-only corpus [10]. BERT has completed a variety of NLP missions with state-of-the-art outcomes[11]. And involved in cancer, genetics and many other areas of medicine. Among them, the BERT model is often used. Bert can achieve excellent performance on biomedical and clinical sentence similarity and short document classification tasks[12].

2. Method

2.1. BERT

BERT is a Transformer-based deep bidirectional language representation model. It uses the Transformer encoder to construct a multilayer bidirectional network consisting of multiple layers of Transformer encoders piled one on top of the other, each consisting of a multi-headed self-attentive sublayer and a feedforward neural network sublayer. The BERT model differs from models that obtain only the semantic information of words, but focuses on obtaining comprehensive semantic information of the whole text. The model randomly masks words in a sentence and then trains the model to predict the removed words in order to understand the relationship between two sentences. Figure 1 shows this BERT model.

DOI: 10.54254/2755-2721/5/20230575

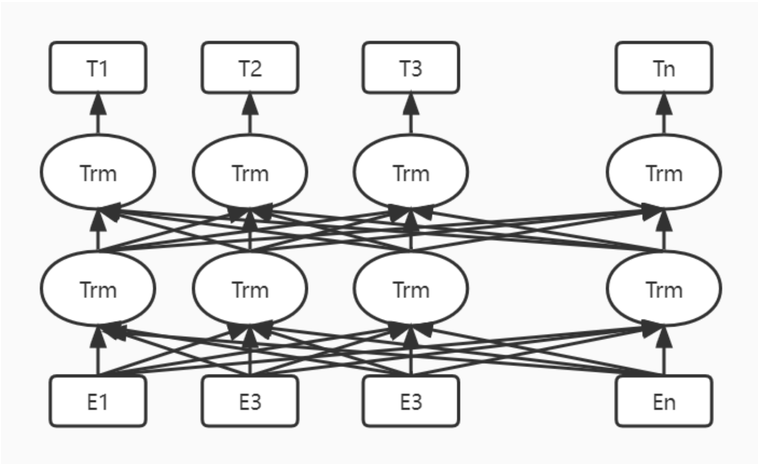

Figure 1. BERT Model.

2.2. DeBERTa

In the pre-training process of BERT, BERT discovers the intricate and delicate links between every word in the pre-training corpus[13]. But many of these connections might not be necessary for the classification process and might even distract from important keywords. This leads to the poor performance of the BERT model.

To improve performance, an improved technology based on the BERT, model DeBERTa will be used. DeBERTa has enhanced BERT with disentangled attention. This improved algorithm has two new mechanisms. The first mechanism is the disentangled attention mechanism. The weight of attention between words is determined using a disentangled matrix to calculate each word's content and relative location. Two vectors that encode the content and location of each word are used to represent it.

The second DeBERTa technique makes use of the context word's content and location data for the Masked Language Model. Given a sample“The white cat is hiding behind the white dog.” with the words “dog” and “cat” masked for prediction. Using only the local context is insufficient for the model to distinguish store and mall in this sentence, because both follow the word “white”. To address this limitation, the model needs to take into account absolute positions. For example, the subject of the sentence is “cat” not “dog”. Figure 2 shows this mechanism.

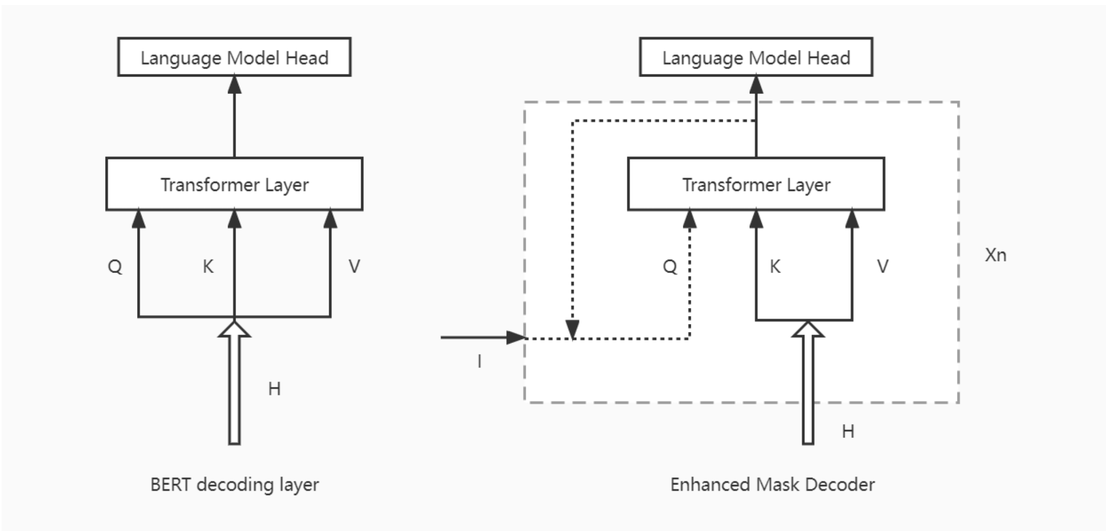

Figure 2. Comparison of the decoding layer.

DOI: 10.54254/2755-2721/5/20230575

3. Experiment and Results

3.1. Dataset

The learning of the deep neural network model depends on the data set, and the model's creation will be significantly influenced by superior data collection. The dataset for this experiment was derived from the USMLE® Step 2 Clinical Skills exam. USMLE is a Medical Licensing Examination. This professional test is required for all doctors who want to work in the US. The exam consists of three sections, the American Medical Licensing Examination Stage 1, the American Medical Licensing Examination Stage 2 and the American Medical Licensing Examination Stage 3. The exam is administered by FSMB, which is the Federation of State Medical Boards and NBME®, which is the National Board of Medical Examiners®. The exam measures the trainee's ability to identify relevant clinical facts when meeting with standardized patients.

The dataset contains a total of 42,146 pieces of data. Essential terms include: Clinical Case (The situation's signs, complaints, and worries, that's what people say when they go to the doctor.), Patient Note (crucial material for the patient to read during the appointment), annotation string(The real string in the patient notes of the corresponding feature), feature (labels, 143 total)

3.2. Evaluation

In this paper, Micro-F1 is used as an assessment metric to assess how well the models perform in terms of categorization. The evaluation index involves parameters TP, FN, FP and TN. TP: It is expected that the positive category will really be positive. FP: It is anticipated that the negative category will actually be positive. FN: It is expected that the positive category will really be negative. TN: It is anticipated that the negative category will actually be negative. TP and TN indicate that the prediction is correct, while FP and FN indicate that the prediction fails.

Precision is defined as the fraction of accurately predicted positive samples in the actual anticipated positive samples, or how many of the projected positive samples are accurate. Therefore, formula (1) is the calculation method of Precision rate.

$$Pecision = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + \sum_{i=1}^{n} FP_i} \tag{1}$$

The percentage of accurately anticipated positive samples in positive samples is known as the recall ratio, that is, how many positive samples in the samples are correctly predicted. Therefore, Equation (2) is the calculation formula for the recall ratio

$$Recall = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + \sum_{i=1}^{n} FP_i} \tag{2}$$

Micro-F1 is the weighted harmonic average of Precision and Recall. The meaning of F1 is that when F1 is higher, the experimental method is more ideal. Therefore, formula (3) is the formula of the F1 calculation method.

$$\text{Micro-}F1 = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{3}$$

3.3. Result

In this paper, two groups of experiments, BERT and DeBERTa, were conducted. In the two groups of experiments, the training set, test set and validation set remained unchanged, and the common parameter values of the experiments remained unchanged. DeBERTa model is based on the improved model, BERT by this group of contrast tests shows that the disentangled attention mechanism and the enhanced mask decoder help to the semantic understanding. It can be seen from the results that the DeBERTa model achieves a higher micro-F1 value than the BERT model. The size of F1 reflects the stability of the model. More stability is indicated by a higher number. Therefore, it can be concluded that the DeBERTa model performs much better than BERT. In Table 1, the experimental findings are displayed.

DOI: 10.54254/2755-2721/5/20230575

Table 1. Experimental results (BERT and DeBERTa).

Model	Micro-$F1$
BERT	0.774
DeBERTa	0.883

4. Conclusion

This paper mining features for patient notes, mining annotated strings containing important information such as "3-4 months of", "Nausea", "No sick contacts" and so on in long patient notes, which can play a decisive role in the diagnosis of illness. These strings are then divided into the 143 features that have been defined. Thus, unified, standardized and clear labels were given to a paragraph of patient notes. And these are the things that doctors need to know when they are making a diagnosis. In this paper, BERT and the DeBERTa models were used to carry out experiments and comparisons, and F1 values of 0.77 and 0.88 were obtained respectively. The DeBERTa model shows a more stable effect. Thus, this paper argues that the use of the DeBERTa model on this medical problem has improved the reliability, performance and accuracy of patient notes content recognition. Thus, medical professionals can be assisted in completely examining the concealed information in patient situations and revealing the data relevant to clinical skills evaluation.

References

[1] Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M. A., Shambour, M. K. Y., Alsalibi, A. I., & Gandomi, A. H. (2022). Machine learning in medical applications: A review of state-of-the-art methods. Computers in Biology and Medicine, 145, 105458.

[2] M. Li, Z.-H. Zhou. Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. IEEE Trans. Syst. Man Cybern. Syst. Hum., 37 (6) (2007), pp. 1088-1098.

[3] A. Choudhury, D. Gupta. A survey on medical diagnosis of diabetes using machine learning techniques Recent Developments in Machine Learning and Data Analytics, Springer (2019), pp. 67-78

[4] Z. Obermeyer, E.J. Emanuel. Predicting the future big data, machine learning, and clinical medicine. N. Engl. J. Med., 375 (13) (2016), p. 1216

[5] Mujtaba,Z; Shuib,L; Idris,N, 2019. Clinical text classification research trends: Systematic literature review and open issues. Expert Systems with Applications.Volume 116, 2019, pp. 494-520, ISSN 0957-4174

[6] Chen, X., Wang, X., Zhang, K., Fung, K.-M., Thai, T. C., Moore, K., Mannel, R. S., Liu, H., Zheng, B., & Qiu, Y. (2022). Recent advances and clinical applications of deep learning in medical image analysis. Medical Image Analysis, 79, 102444.

[7] Pons E, Braun LM, Hunink MG, et al. Natural language processing in radiology: a systematic review. Radiology. 2016;279(2):329–43.

[8] Chen TL, Emerling M, Chaudhari GR, et al. Domain specific word embeddings for natural language processing in radiology. J Biomed Inform. 2021;113: 103665.

[9] Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding.

[10] Han X, Zhang Z, Ding N, et al. Pre-trained models: past, present and future. AI Open. 2021;2:225 – 50.

[11] Qiu X, Sun T, Xu Y, et al. Pre-trained models for natural language processing: a survey. Science China Technol Sci. 2020;63(10):1872–97.

[12] Roussinov,D; Conkie,A; Patterson,A; Sainsbury,C.(2021) Predicting Clinical Events Based on Raw Text: From Bag-of-Words to Attention-Based Transformers. Frontiers in Digital Health.

DOI: 10.54254/2755-2721/5/20230575

[13] He, P., Liu, X., Gao, J., & Chen, W. (2021). Deberta: Decoding-enhanced bert with disentangled attention.