applied sciences

MDPI

Article

Research on Medical Text Parsing Method Based on BiGRU-BiLSTM Multi-Task Learning

Yunli Fan, Ruiqing Kang *, Wenjie Huang and Lingyan Li

School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China; m202210524@xs.ustb.edu.cn (Y.F.); m202320797@xs.ustb.edu.cn (W.H.); m202310527@xs.ustb.edu.cn (L.L.)
* Correspondence: krq@ustb.edu.cn

Abstract: With the development of technology, the popularity of online medical treatment is becoming more and more widespread. However, the accuracy and credibility of online medical treatment are affected by model design and semantic understanding. In particular, there are still some problems in the accurate understanding of complex structured texts, which affects the accuracy of judging users' intentions and needs. Therefore, this paper proposes a new method for medical text parsing, which realizes core tasks such as named entity recognition, intention recognition, and slot filling through a multi-task learning framework; uses BERT to obtain contextual semantic information; and combines BiGRU and BiLSTM neural networks, and uses CRF to realize sequence annotation and DPCNN to realize classification prediction. Thus, the task of entity recognition and intention recognition can be accomplished. On this basis, this paper builds a multi-task learning model based on BiGRU-BiLSTM, and uses CBLUE and CMID databases to verify the method. The verification results show that the accuracy of named entity recognition and intention recognition reaches 86% and 89%, respectively, which improves the performance of various tasks. The ability of the model to process complex text is enhanced. This method can improve the text generalization ability and improve the accuracy of online medical intelligent dialogue when it is used to analyze medical texts.

Keywords: multi-task learning; natural language processing; BERT; BiLSTM; BiGRU; DPCNN

check for updates

Citation: Fan, Y.; Kang, R.; Huang, W.; Li, L. Research on Medical Text Parsing Method Based on BiGRU-BiLSTM Multi-Task Learning. *Appl. Sci.* **2024**, *14*, 10028. https://doi.org/10.3390/app142110028

Academic Editor: Malik Yousef

Received: 18 September 2024
Revised: 24 October 2024
Accepted: 1 November 2024
Published: 3 November 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

With the development of technology and the progress of society, people pay more and more attention to health. Determining how to obtain medical information conveniently and accurately has become a hot topic in the medical field. With the popularization of intelligent medical technology and the Internet, the means of obtaining medical information are no longer limited to the traditional offline consultation. The popularity of online medical care benefits from the rapid development of information communication technology. Nowadays, many medical websites provide a lot of medical information, but the traditional online search methods need users to select them. In contrast, smart medicine enables users to get the information they need more quickly through man–machine dialogue. With the wide application of natural language processing in the medical field, smart medicine has made further development. It can not only learn and analyze a large amount of medical data, but also simulate the diagnostic process of doctors to a certain extent, provide users with evidence-based medical advice, and provide decision support for doctors, further improving the accuracy and efficiency of diagnosis and treatment. However, due to the ambiguity of language, the diversity of structure, and the dependence on context, the understanding of medical questions has become a technical difficulty in the research of intelligent medical systems. The existing medical text parsing methods still face challenges in understanding and identifying tasks in a complex context, especially in capturing long-term and short-term relationship dependencies and accurately identifying entities.

At present, solutions to this problem can be divided into two categories: traditional machine learning methods and large language model (LLMs). Traditional machine learning

methods include logical regression, support vector machine (SVM), and naive Bayesian (naive Bayes), as well as long-term and short-term memory network (LSTM), gated loop unit (GRU), and convolution neural network (CNN) based on deep learning. Although these models have achieved some success in classification, emotion analysis, and other tasks, they do not perform well in dealing with medical texts with complex grammatical structures and semantic dependencies, and it is difficult to deal with named entity recognition (NER), intention recognition, and other tasks that require context understanding.

With the development of deep learning technology, large language models (LLMs) such as the GPT series, BERT, T5, and PaLM have become a new solution [1]. They show powerful language understanding and generation ability through large-scale pre-training and a deep self-attention mechanism, and are widely used in named entity recognition, intention recognition, and other tasks. Singhal et al. [2] proposed a medical question answering system based on LLMs, which achieved high accuracy and reliability on multiple medical question answering datasets through prompt strategies and integration methods; however, there were still problems of performance degradation and insufficient interpretation ability in specific situations. Although mainstream large language models such as GPT and LLaMA perform well in a variety of NLP tasks, their training and fine-tuning processes are highly dependent on computational sources, especially for high-complexity tasks such as Chinese medical text processing, so large models require massive quantities of training data and powerful computational resources. It consumes a lot of manpower, material, and financial resources, which are difficult for ordinary research personnel to bear. In addition, in the Chinese medical field, the sensitivity of patient data requires us to pay extra attention to data privacy and security issues, and LLMs often need to upload data to the cloud for training, which may pose a potential privacy breach risk. Considering the lack of transparency in the process of large model reasoning, its application in medical decision-making is also limited due to the lack of explanation.

Based on these considerations, in order to improve the performance of the system with limited training data and computing resources, this paper proposes a medical text parsing method based on a multi-task learning framework. This method combines BERT, BiGRU, BiLSTM, and other models to enhance the understanding of medical text through the shared layer structure. At the same time, using the decoding advantage of a conditional random field (CRF) in named entity recognition, the deep pyramid convolutional neural network (DPCNN) is used to process long text information in the intention recognition task, and the context modeling problem in named entity recognition and the intention recognition task are solved respectively. Experiments show that for named entity recognition tasks, our method achieves an accuracy of 0.8607, which is 3.35% higher than that of the existing methods, and the F1 value reaches 83.64%, showing a good balance between accuracy and recall. In the intention recognition task, the accuracy of the model reached 0.8925; thus, the other methods increased by 2.91%, and the F1 value reached 88.78%. These results show that the proposed multi-task learning architecture can not only effectively alleviate the resource consumption of large language models, but also provide more accurate and efficient parsing capabilities according to the domain characteristics of medical texts. Under the condition of limited resources, this method can improve the performance of the model on multiple tasks, and has good interpretability and adaptability. In summary, our main contributions can be summarized as follows:

- The combination of BiGRU and BiLSTM, and the use of a multi-task learning framework and joint decoding strategy to comprehensively improve the capture and analysis of short-term and long-term dependency features of medical texts. The experimental results show that the proposed method is better than the benchmark method, improves the generalization ability of the model, and has the potential for practical application in personal health management systems.
- We propose a multi-task learning model of BiGRU-BiLSTM, which is used to parse medical texts and effectively improve the performance of named entity recognition, intention recognition, and slot filling in medical treatment.

- The experimental results show that the proposed method is better than the benchmark method, improves the generalization ability of the model, and has the potential for practical application in personal health management system.

The structure of this paper is as follows. Section 2 reviews the research status of semantic parsing and multi-task learning. Section 3 describes in detail the implementation process of the multi-task learning model proposed in this paper. Section 4 describes the lab settings, datasets, and evaluation criteria, and analyzes and discusses the results. Section 5 gives the conclusion.

2. Related Work

2.1. Semantic Analysis

In the field of medical application, named entity recognition (NER), intention recognition, and slot filling constitute the core framework of medical text semantic analysis. Medical named entity recognition is mainly used to extract specific entities related to health care from the text data input by users, such as diseases, symptoms, and drugs. The accuracy of named entity recognition directly affects the quality of medical text parsing, which determines whether the system can effectively extract valuable information from a large number of texts. Intention recognition refers to understanding the core purpose or intention of user input, which is generally regarded as a classification task to help determine the main needs of users. Intention recognition provides direction in medical text parsing to ensure that the system can understand the overall semantic structure of the text, thus accurately parsing useful information. As a sub-field of text classification, machine learning methods such as naive Bayes and Adaboost are often used in intention recognition tasks. Slot filling refers to filling a predefined slot according to the identified entity and context during a conversation, which is often regarded as a sequence tagging task. The accuracy and completeness of slot filling determines whether the user's needs can be fully understood and provides support for subsequent diagnosis or recommendations.

In order to improve the accuracy of medical text understanding, researchers introduced machine learning and pre-trained models to improve the accuracy and precision of semantic parsing of medical problems. Traditional machine learning models have shown certain effects in capturing semantic information. Shanavas et al. [3] proposed an SVM-based concept map medical text classification method, which uses a domain-specific similarity matrix to construct text concept maps and uses graph cores for classification. High-performance results are achieved in the classification of English biomedical literature and clinical reports. Lenivtceva et al. [4] explored the applicability of four shallow machine learning models (polynomial naive Bayes, logistic regression, support vector machine, and k-nearest neighbor) and two integrated classifier chains (ECCCLR and ECCSVM) in the classification of medical unstructured texts, which performed well in the multi-label classification task of Russian medical texts. This approach can improve the accuracy of classification and the recall rate, and it is planned to apply the NER model to extract entity information in medical free text to further improve model performance.

In the study of Chinese medical texts, Zhang et al. [5] proposed a hybrid model based on CNN and GRU to capture the semantic information and relationship between questions and answers, which improved the performance of the model, but there was still a problem of focusing on word matching while ignoring the relationship between words. The emergence of pre-trained language models (such as BERT) has greatly improved the performance of natural language processing tasks. Shi et al. [6] built a medical question answering system based on a knowledge graph, adopted the BERT model for word segmentation and vocabulary construction, and adopted the naive Bayes machine learning method for intention recognition. Wang et al. [7] proposed a multi-scale context-sensing interaction method based on BERT with multi-granularity embedment, which can effectively improve the accuracy of the TCM question answering matching task. Huang et al. [8] proposed the semantic parsing of ALBERT questions, which further enhanced the accuracy of named entity recognition, and the constructed intelligent question answering system can better

communicate with users. Guo et al. [9] proposed a joint model based on BERT, which solves the label dislocation problem caused by subword segmentation by introducing subword attention adapters and intent attention adapters, and achieves significant improvement in intention classification and slot filling tasks. Although BERT-based models perform well in many tasks, their ability to generalize to low-resource languages and domains is still limited. In addition, many models do not perform well when dealing with long text or complex context dependencies, resulting in excessive computational overhead.

In order to further improve the effect, researchers combined various models. Deng et al. [10] proposed an improved GL-GIN model, which realized bidirectional information flow by introducing slots into the unidirectional attention layer and BERT coding layer of intent, and significantly improved the performance of the multi-intent recognition model. In addition, the researchers found that the combination of BERT, BiGRU/BiLSTM, and CRF to identify the intention information and entity information in the question can improve the accuracy and task performance of the model. Huang et al. [11] proposed a question-and-answer model based on BERT-BiGRU-CRF, and the final F1 value of the model was 76.39%. Wang et al. [12] proposed a method based on BERT and the BERT-BiLSTM-CRF model to identify intention information and entity information in questions, which makes the returned answers more accurate. Hu et al. [13] proposed a Chinese medical entity recognition model based on BERT-BiGRU-CRF, which is improved compared with other models by combining the word vector representation of BERT, context feature extraction of BiGRU, and tag optimization of CRF.

Although the current methods have improved the accuracy of semantic understanding of medical texts to a certain extent, there is still room for improvement in the processing of long and complex sentences in complex natural languages. In addition, relying on a single model (such as LSTM or GRU) often faces the problem of insufficient understanding of context information and a limited ability to extract features. In this paper, a multi-task learning method based on BiGRU-BiLSTM is proposed to enhance the comprehensive understanding and parsing ability of medical texts.

Table 1 summarizes the current research on semantic parsing, which provides a reference for the follow-up research on entity recognition and intention recognition combined with the multi-task learning model.

Table 1. Comparison of current research status of semantic parsing.

Reference	Methods and Techniques	Research Findings	Prospect
Reference [3]	Medical text classification method based on concept map based on SVM, using domain specific matrix and graph kernel classification	High-performance results were achieved in English biomedical literature and clinical report classification	Extend the method to other languages and different medical fields to improve the generalization ability
Reference [4]	Four shallow machine learning models (naive Bayesian, logical regression, SVM, k-NN) and integrated classifier chain (ECCCLR, ECCSVM)	It performs well in the multi-tag classification of Russian medical texts, which improves the accuracy and recognition rate of classification	It is planned to use the NER model to extract entity information to further improve the performance of the model
Reference [5]	A hybrid model based on CNN and GRU to capture the semantic relationship between questions and answers	The performance of the model is improved, but there are some problems such as paying attention to word matching and interlexical association	Solve the problem of insufficient connection between words, and further improve the semantic comprehension ability of long text

Table 1. *Cont.*

Reference	Methods and Techniques	Research Findings	Prospect
Reference [6]	The medical question answering system based on knowledge graph uses BERT for word segmentation and vocabulary construction, and uses naive Bayes for intention recognition	The medical question answering system is constructed, and the meaning recognition ability of the system is improved. Solve the problem of insufficient connection between words, and further improve the semantic comprehension ability of long text	Continue to optimize the structure and efficiency of the model, explore sparse attention mechanism, dynamic context modeling and adaptive computing, in order to improve the performance of the model in dealing with long text and complex context dependencies, and reduce computational overhead.
Reference [7]	Multi-granularity embedding and multi-size context-aware interaction based on BERT	It improves the accuracy of traditional Chinese medicine question and answer matching task	
Reference [8]	Semantic parsing method of question based on ALBERT	It improves the accuracy of named entity recognition and improves the interactive experience between intelligent question answering system and users	
Reference [9]	Joint model based on BERT, using subwords and intention attention	The problem of subword segmentation label dislocation has been solved, and the task of intention classification and slot filling has been significantly improved	
Reference [10]	The improved GL-GIN model combines a BERT coding layer and single focus layer to realize the flow of individual information	It significantly improves the performance of the multi-intention recognition model	Furthermore, the two-way interaction between multi-intention and slot is modeled, and the optimization method of multi-intention recognition is explored.
Reference [11]	The F1 value of the question-and-answer model based on BERT-BiGRU-CRF reaches 76.39%	Furthermore, the two-way interaction between multi-intention and slot is modeled, and the optimization method of multi-intention recognition is explored	Further optimize the model structure to meet the challenge of complex sentence processing
Reference [12]	Based on BERT and BERT-BiLSTM-CRF model	It improves the accuracy of the question answering system and makes the returned answers more accurate.	
Reference [13]	Chinese medical entity recognition model based on BERT-BiGRU-CRF, optimized with BERT, BiGRU, and CRF	Combined with the characteristics of BERT, BiGRU, and CRF, the accuracy of Chinese medical entity recognition is improved	

2.2. Multi-Task Learning

Multi-task learning (MTL) [14] is a machine learning paradigm in which multiple related tasks are learned at the same time, which aims to improve the learning effect of each task by sharing information. Compared with single-task learning (STL) [14], MTL can make use of the shared information between different tasks to improve the generalization ability of the model and improve the learning effect of multiple tasks by sharing representation. By learning these tasks at the same time, the model can share presentation layers or parameters among tasks, thus improving the generalization ability of the model when the amount of data is limited.

Structurally speaking, multi-task learning is generally divided into two categories: hard parameter sharing and soft parameter sharing [15], as shown in Figure 1. Hard parameter sharing usually means that in the first layers of the model, all tasks share the same network parameters, while each task has its own specific parameters in the later layers. In soft parameter sharing, each task has an independent model, but in some way (such as adding constraints, regularization), the parameters of different tasks affect each other.

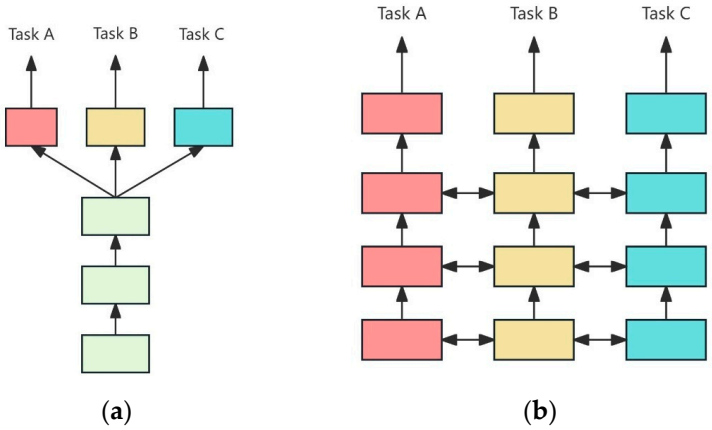

Figure 1. Multi-task learning structure: (**a**) hard parameter sharing structure; (**b**) soft parameter sharing structure.

Since this research area was proposed by Caruana et al. in the early 1990s, a wealth of theoretical and practical results has been achieved. At present, because it is often difficult to obtain high-quality tagged data in the medical field, we can use the available data more effectively by sharing knowledge through multi-task learning. Wu et al. [16] proposed a multi-task learning model, MTL-BERT, based on a task weight adaptive framework. BERT is used to extract features and improve the generalization of the model, which verifies the feasibility of multi-task learning in the field of natural language. In the field of named entity recognition, many scholars improve the accuracy of entity recognition through multi-task learning. Zhang et al. [17] proposed a named entity recognition model based on radical features and lexical enhancement, which combined multi-task learning and semi-supervised learning methods to effectively improve the effect of named entity recognition in Chinese electronic medical records. Gao et al. [18] proposed a nested named entity recognition method based on multi-task learning and a double affine mechanism to solve the problem of low recall rate caused by insufficient boundary supervision. Zeng et al. [19] proposed a multi-dimensional objective semantic learning model based on BERT, which combines a complex semantic enhancement mechanism and adaptive local attention mechanism to significantly improve the effect of fine-grained emotion analysis. Peng et al. [20] proposed a multi-task learning model based on multiple decoders, which achieved 2.0% and 1.3% higher results than BERT and its variants for text similarity, relation extraction, named entity recognition, and text reasoning tasks in biomedical and clinical fields. Liao et al. [21] proposed a Chinese named entity recognition model integrating label information, which significantly improves feature representation by combining BERT embedded representation, a Transformer decoder, and multi-task learning.

Researchers have applied multi-task learning to the study of natural language classification, which has performed well in many classification tasks. Li et al. [22] proposed a speech emotion recognition model based on attention frame pooling and multi-task learning, which significantly improved the extraction ability of emotion information. Zhu et al. [23] proposed a multi-dimensional objective semantic learning model based on BERT, which combines a complex semantic enhancement mechanism and adaptive local attention mechanism to significantly improve the effect of fine-grained emotion analysis. Myint et al. [24] proposed an emotion and emotion classification method based on multi-task learning and an attention mechanism, and combined the BERTweet model and COMETATOMIC 2020 to identify the intention of crisis tweets, which significantly improved the classification and prediction effect in crisis scenarios. Song et al. [25] proposed a classification model based on BERT-DPCNN-MMOE. By introducing a deep pyramid convolutional network and multi-gate control unit mechanism, the classification effect was significantly improved in multi-task and transfer learning experiments. However, most of the existing studies are focused on specific languages or domains. Determining how to improve the generalization

ability of multi-task learning models in cross-language or open domain tasks will be an important research direction in the future.

Table 2 summarizes the current research on the application of multi-task learning in the field of NLP, which provides a reference for the follow-up research of entity recognition and intention recognition combined with a multi-task learning model.

Table 2. Comparison of current research on multi-task learning applied to NLP: (**a**) named entity recognition; (**b**) natural language classification.

(a)			
Reference	**Methods and Techniques**	**Research Findings**	**Prospect**
Reference [17]	NER model based on radical features and vocabulary enhancement, combining multi-task learning and semi-supervised learning	It improves the effect of named entity recognition in Chinese electronic medical records	Further explore how to popularize this method in NER tasks in other areas or languages
Reference [18]	Nested named entity recognition method based on multi-task learning and double affine mechanism	It solves the problem of low recall rate caused by insufficient border supervision	
Reference [19]	Joint entity and relationship extraction Model CopyMTL based on multi-task learning	The problems of inaccurate entity recognition and difficult prediction of multi-labeled entities in CopyRE model are solved	
Reference [20]	Multi-task learning model based on Multi-Decoder.	It is 2.0% and 1.3% better than BERT and its variants in the tasks of text similarity, relationship extraction, named entity recognition, and text reasoning in biomedical and clinical fields	
Reference [21]	A Chinese named entity recognition model combining tag information, BERT embedded representation, Transformer decoder, and multi-task learning	It significantly improves the effect of feature representation in the Chinese named entity recognition task	
(b)			
Reference	**Methods and Techniques**	**Research Findings**	**Prospect**
Reference [22]	Speech emotion recognition model based on attention frame pooling and multi-task learning	It improves the ability of extracting emotional information	Explore how to apply the multi-task learning model to more complex affective analysis and cross-language, open-domain classification tasks.
Reference [23]	A multi-dimensional objective semantic learning model based on BERT, which combines a complex semantic enhancement mechanism and adaptive local attention mechanism	It significantly improves the effect of fine-grained emotional analysis	
Reference [24]	Emotion and emotion classification based on multi-task learning and attention mechanism, combined with BERTweet model and COMET-ATOMIC 2020	The purpose of classification and prediction effect of tweets in crisis scenarios are improved	
Reference [25]	Classification model based on BERT-DPCNN-MMOE, combined with deep pyramid convolution network and multi-gate control unit mechanism	The classification effect is significantly improved in multi-tasking and transfer learning experiments	

3. Materials and Methods

Because entity recognition, intention recognition, and slot filling are three highly related tasks, by sharing model parameters, MTL can improve the performance of each task. In order to improve the accuracy of entity recognition and intention recognition, and improve the generalization ability of the model, the method constructed in this paper includes a BERT pre-training model, BiGRU-BiLSTM layer, and task-specific decoding layer, and the framework diagram is shown in Figure 2. The model adopts a hard parameter

sharing model. During the training, the model alternates between intention classification and named entity recognition tasks.

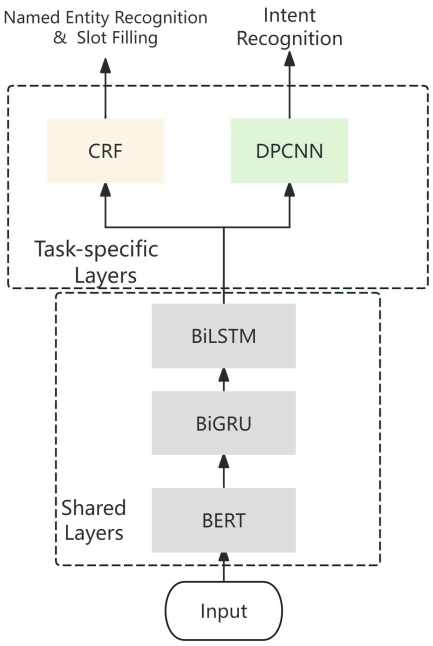

Figure 2. Text parsing method implementation framework.

The input text is encoded by a word vector through the BERT embedding layer in turn, and then input to the bidirectional gated loop unit (BiGRU) and bidirectional long-term and short-term memory network (BiLSTM) layers to learn the long-term and short-term dependencies of each time step in the sequence through its loop structure. Finally, for the named entity recognition task and slot filling task, CRF uses the forward–backward algorithm to constrain the label distribution output of the model to ensure that the predicted label sequence conforms to the global dependency; for the intention recognition task, DPCNN further processes the extracted text features through the convolution network, and finally makes a classification prediction.

3.1. Language Sequence Processing

Bidirectional Encoder Representations from Transformers (BERT) [26] is a pre-trained language model proposed by Google based on the Transformer architecture. The bidirectional encoder learns contextual information, which is widely used for natural language processing tasks. The structure of BERT is shown in Figure 3.

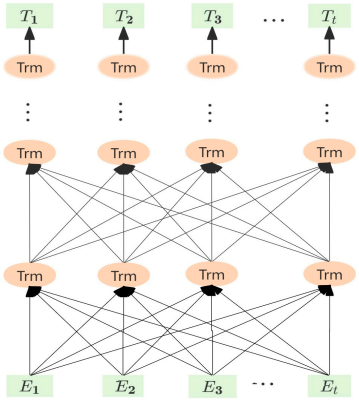

Figure 3. BERT structure diagram.

BERT is optimized by two pre-training tasks, the Masked Language Model (MLM) and Next Sentence Prediction (NSP). MLM randomly selects 15% of the words in the input sequence, replaces them with a special [MASK] tag, and then asks the model to predict these masked words. NSP provides sentences in pairs in the input sequence and requires the model to predict whether the second sentence is the subsequent sentence of the first sentence, which can help the model to understand the relationship and logic at the sentence level.

In order to better identify sentences in the training process, the input of BERT is divided into two parts: text sequence and position embedding. The text sequence processed by BERT is separated by a special separator [SEP] between two sentences, and each word input is embedded in each tag by Token Embeddings (meaning of each word), Segment Embeddings (embedding of the sentence to which the distinction belongs) and Position Embeddings (indicating the position of words in the sequence). Finally, the Embedding representation of BERT is shown in Figure 4:

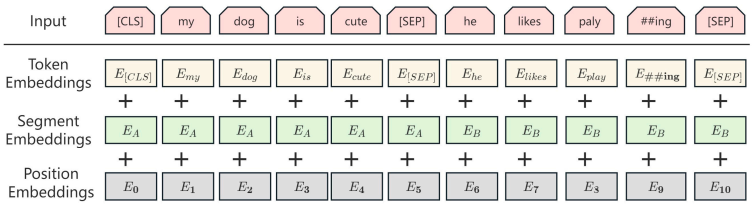

Figure 4. Embedding schematic diagram of BERT.

BERT is based on the Encoder part of Transformer, and each Encoder consists of an attention mechanism and feed-forward neural network. The final output of BERT is that each word goes through multiple layers of Transformer to generate a context-sensitive representation.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

$$MultiHead(Q, K, V) = Concat(head_1, head_2, \cdots, head_h)W^O \tag{2}$$

$$FFN(x) = ReLU(xW_1 + b_1)W_2 + b_2 \tag{3}$$

where Q (query), K (key), and V (value) are vectors obtained from the input by different linear transformations; d_k is the dimension of the key vector; W^O is a trainable parameter matrix; W_1, W_2, b_1, b_2 are trainable parameters.

3.2. Semantic Feature Extraction

3.2.1. Bidirectional Long Short-Term Memory

Long short-term memory (LSTM) [27] is a variant of recurrent neural networks (RNNs) specifically designed to solve long-term dependency problems. By introducing gating mechanisms to control the input, output, and forgetting of information, it can effectively capture and retain in long-term memory the dependencies in the sequence.

The LSTM unit consists of three gates (forget gate, input gate, and output gate) and a memory unit, which work together to process and remember long-term dependencies in the sequence. Their structure is shown in Figure 5.

The forget gate determines how many memories of the past to retain, and decides whether to forget the past information based on the current input x_t and the hidden state of the previous time step h_{t-1}. When f_t approaches 0, it indicates that the model will forget the memory information of the previous time step. As f_t approaches 1, the model retains more historical information. The input gate determines what information from the current input x_t needs to be added to the memory unit, and updates the current memory with the candidate memory $\widetilde{C}_t$. The input gate determines how to update the memory unit by

i_t, and $\widetilde{C}_t$ generates a new candidate memory that is added to the memory state under the input gate's control. The closer i_t is to 1, the more information currently entered is retained and added to the memory unit. The output gate determines the current hidden state h_t and the final output. When the output gate o_t approaches 1, it indicates that the current memory state C_t will greatly affect the output of the current time step. The specific calculation formula is:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{4}$$

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{7}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{8}$$

$$h_t = o_t * \tanh(C_t) \tag{9}$$

where σ is the sigmoid function, W and b are the weights and biases of the forgetting gate, input gate and input gate, tanh is the hyperbolic tangent function, and $*$ represents element-by-element multiplication.

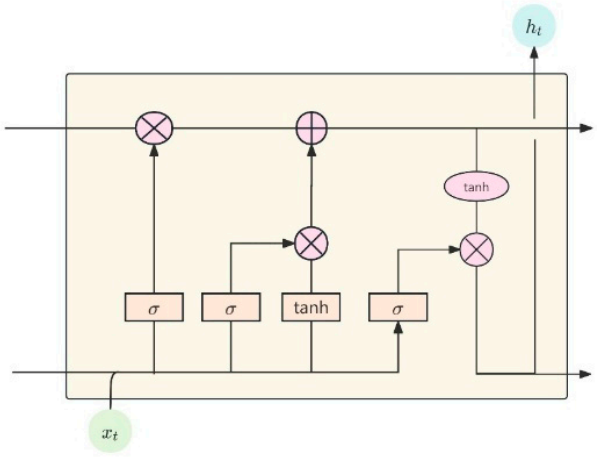

Figure 5. LSTM structure diagram.

Bidirectional long short-term memory (BiLSTM) is a variant of a recurrent neural network that combines the advantages of standard LSTM and bidirectional propagation, and is particularly suitable for sequential data modeling, such as text processing tasks in natural language processing. BiLSTM improves on traditional one-way LSTM by considering both past and future contextual information at each time step. The output of BiLSTM is a concatenation of forward and backward hidden states:

$$\text{Forward LSTM: } \overrightarrow{h_t} = \text{LSTM}\left(\overrightarrow{h_{t-1}}, x_t\right) \tag{10}$$

$$\text{Backward LSTM: } \overleftarrow{h_t} = \text{LSTM}\left(\overleftarrow{h_{t+1}}, x_t\right) \tag{11}$$

$$h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right] \tag{12}$$

where $[\,;\,]$ indicates the connection operation.

3.2.2. Bidirectional Gated Recurrent Unit

A gated recurrent unit (GRU) [28] is another RNN variant commonly used to process sequential data, similar to long short-term memory, but with a simpler structure, as shown

in Figure 6. GRU integrates the forgetting gate and the input gate into an update gate, and controls when to update the hidden state, thus realizing long-term dependent learning.

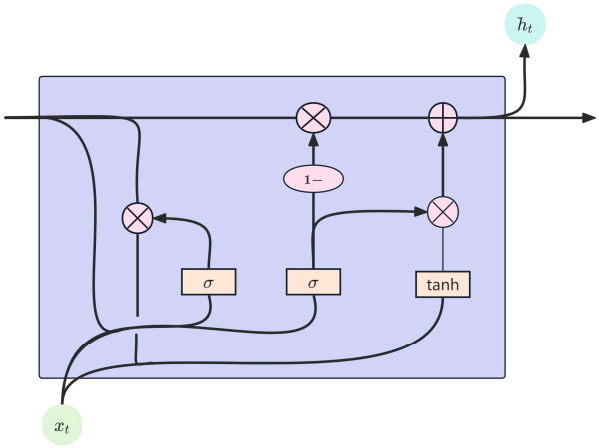

Figure 6. GRU structure diagram.

The GRU unit consists of the update gate and the reset gate, which together determine the hidden status update mode of the current time step.

The reset gate r_t controls the influence of the hidden state h_{t-1} of the previous time step on the current candidate hidden state $\tilde{h}_t$. When r_t approaches 0, it ignores the previous hidden state and focuses on the current input x_t; conversely, when r_t approaches 1, it means more information about the previous time step needs to be considered to generate the current candidate hidden state, which means that there is a strong relationship between the current input and the historical state.

The update gate z_t determines whether the hidden state needs to be updated in the current time step. When z_t is close to 1, the model tends to retain the hidden state h_{t-1} of the previous time step and ignore the current input x_t. On the contrary, when z_t is close to 0, the new hidden state h_t is closer to the candidate hidden state $\tilde{h}_t$. The specific calculation formula is:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \tag{13}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \tag{14}$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t] + b) \tag{15}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{16}$$

where σ is the sigmoid function, W_r and b_r are the weight and bias of the reset gate, and W_z and b_z are the weight and bias of the update gate. tanh is the hyperbolic tangent function, and W and b are the weights and biases associated with the candidate hidden state.

The combination of these two gates helps the GRU efficiently deal with the problem of disappearing gradients, especially to capture important long-range dependencies in longer sequences. By flexibly regulating the flow of information, GRUs can effectively retain long-term dependencies without losing short-term information.

Bidirectional gated recurrent unit (BiGRU) is a neural network model that combines GRU and bidirectional propagation. By considering both forward and backward information of the sequence data at each time step, it is able to capture the context in the sequence more comprehensively. BiGRU consists of two GRU layers: a forward GRU and a backward GRU. The output of each time step is a concatenation of the output of these two GRU layers, which is similar in structure to BiLSTM.

3.3. Exclusive Task Layer Design

3.3.1. Conditional Random Field

Conditional random field (CRF) is a probabilistic graph model for sequence annotation tasks. It can model the output sequence under the condition of a given input sequence and fully consider the context information to improve the accuracy of annotation. Its structure is shown in Figure 7.

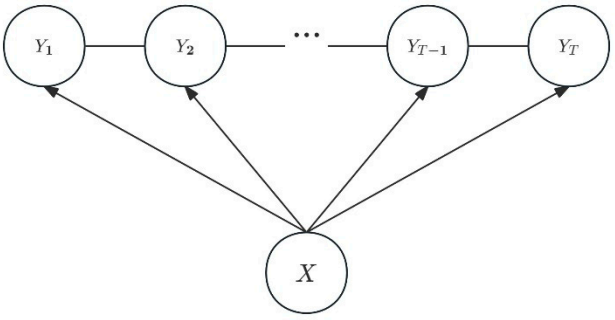

Figure 7. CRF structure diagram.

CRF directly models conditional probability $P(Y \mid X)$, that is, the probability of predicting label sequence Y given input sequence X. The model describes the relationship between the output labels and the input sequence through the feature function, as well as the dependency between the labels. By optimizing the weight parameters, CRF is able to maximize the log-likelihood function of the training data. The Viterbi Algorithm is often used to efficiently find the optimal label sequence during the push process.

Given the input sequence $X = (x_1, x_2, \ldots, x_T)$ and the corresponding output sequence $Y = (y_1, y, \ldots, y_T)$:

$$P(Y \mid X) = \frac{1}{Z(X)} exp\left(\sum_{t=1}^{T}\sum_{k}\lambda_k f_k(y_t, y_{t-1}, X, t)\right) \tag{17}$$

$$Z(X) = \sum_{Y} exp\left(\sum_{t=1}^{T}\sum_{k}\lambda_k f_k(y_t, y_{t-1}, X, t)\right) \tag{18}$$

$$P(Y \mid X) = \sum_{t=1}^{T}\sum_{k}\lambda_k f_k(y_t, y_{t-1}, X, t) - \log Z(X) \tag{19}$$

where $f_k(y_t, y_{t-1}, X, t)$ is the characteristic function, which describes the relationship between the output tag and the input sequence, as well as the dependency between the tags, λ_k is the weight of the characteristic function, and $Z(X)$ is a normalized factor, ensuring a normal probability distribution that sums to 1.

The characteristic function can be divided into two categories: the state characteristic function $f(y_t, X, t)$ is used to describe the label y_t and input X t relationship in place; and the transfer characteristic function $f(y_t, y_{t-1}, X, t)$ is used to describe the label y_t, the previous label y_{t-1}, and the relations among them.

3.3.2. Deep Pyramid Convolutional Neural Network

Deep pyramid convolutional neural network (DPCNN) [29] is a deep convolutional neural network designed for text classification tasks. DPCNN captures global semantic information in text through pyramid-like convolution and a pooling structure, and maintains gradient flow in deep networks by using residual connections.

DPCNN consists of the following five parts, and its structure is shown in Figure 8:

- Embedding Layer: Converts the input text into an embedding vector.

- Region Embedding Layer: By combining the embedding vector of adjacent words with one-dimensional convolution, the features of the office are extracted.
- Pyramid Block: Consists of multiple residual blocks. Each residual block contains two layers of convolution and one layer of pooling, which halves the size of the feature map.
- Global Max Pooling Layer: The feature map is maximized to generate a fixed-length feature vector.
- Fully Connected Layer: Classifies the pooled vectors.

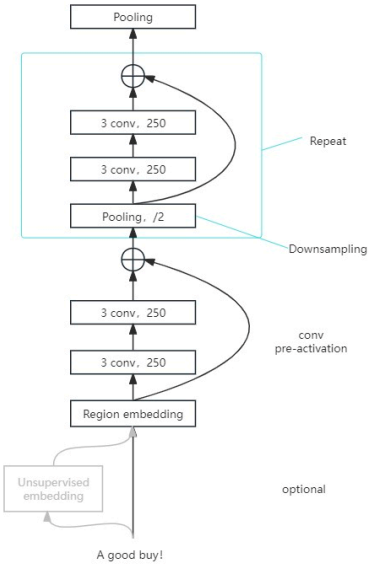

Figure 8. DPCNN structure diagram.

The input text is transformed into a low-dimensional dense vector through the embedding layer to obtain the word vector matrix. After that, DPCNN extracts text local features through multi-layer stacked convolutional networks, and uses residual connections to alleviate the gradient disappearance problem and enhance the feature representation capability of deep networks. The spatial dimension of the feature graph is gradually reduced layer by layer, and finally a fixed length vector is obtained by global average pooling. The vector passes through the full connection layer and SoftMax function to output the classification probability distribution.

For sequence $\{w_1, w_2, \cdots, w_n\}$ having a length of n, text is entered, each word w_i is mapped for the d word vector x_i, the word embedded $X = [x_1, x_2, \cdots, x_n]$ is obtained, and the classification results y as follows:

$$C = ReLU(W * X + b) \tag{20}$$

$$H_{l+1} = ReLU(H_l + Conv(H_l)) \tag{21}$$

$$P = MaxPooling(C) \tag{22}$$

$$G = \frac{1}{Z}\sum_{i=1}^{Z} P[i] \tag{23}$$

$$y = softmax(W_{out} \cdot G + b_{out}) \tag{24}$$

where $X \in \mathbb{R}^{n \times d}$ is the input word vector matrix, n is the length of the text, and d is the dimension of the word vector. $*$ is the said convolution operation, $W \in \mathbb{R}^{k \times d}$ is the convolution kernel weight, $b \in \mathbb{R}$ is bias, $C \in \mathbb{R}^{(n-k+1) \times d}$ is the characteristic of volume after the product figure, H_l is the input feature map for layer l, $Conv$ is the convolution of the two-stack operation, P is the pooling characteristics after the figure, Z is the size of the

feature graph, G is the output vector after global averaging pooling, and W_{out} and b_{out} are the weights and bias terms of the fully connected layer, respectively.

3.4. Multi-Task Learning Implementation

3.4.1. Model Construction

The text parsing model based on multi-task learning includes the input layer, BERT coding layer, BiGRU layer, BiLSTM layer, and task-specific layer. The model schematic diagram is shown in Figure 9.

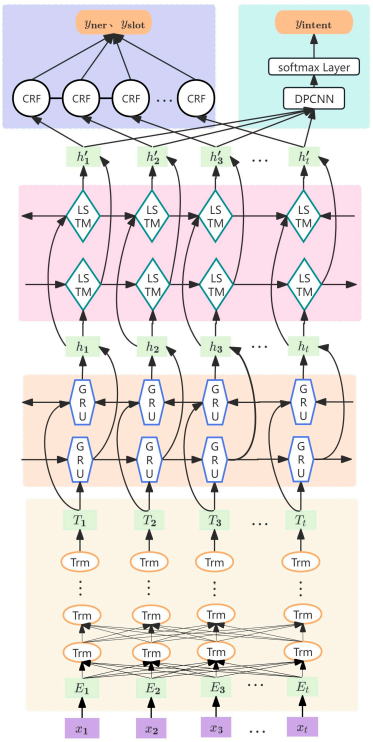

Figure 9. Multi-task learning model framework.

For the hypothesis of the input sequence for $X = (x_1, x_2, ..., x_t)$, first, extract the context representation of the input sentence through BERT:

$$H = BERT(x) = [h_1, h_2, ..., h_t]$$

The data processed by BERT are used as the input of the BiGRU layer, and based on the output of BiGRU, the BiLSTM structure is used to enhance the capture of sequence features, and the output of BiGRU is used as the input of BiLSTM.

Forward GRU output: $\overrightarrow{h_t} = GRU_f(\text{BERT}(x_t), \overrightarrow{h_{t-1}})$

Backward GRU output: $\overleftarrow{h_t} = GRU_b(\text{BERT}(x_t), \overleftarrow{h_{t-1}})$

Combine hidden states in both directions: $h_t = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right]$

Forward LSTM output: $\overrightarrow{h_t}' = \text{LSTM}\left(h_t, \overrightarrow{h_{t-1}}'\right)$

Backward LSTM output: $\overleftarrow{h_t}' = \text{LSTM}\left(h_t, \overleftarrow{h_{t+1}}'\right)$

Combine hidden states in both directions: $h_t' = \left[\overrightarrow{h_t}'; \overleftarrow{h_t}'\right]$

The output from the shared layer is passed to a dedicated layer for each task.

1. Implementing sequence tagging with CRF

A CRF layer is added on top of the BiGRU-BiLSTM layer for labeling sequences, enabling entity recognition and slot filling tasks. The CRF layer can capture the dependencies between labels in the sequence, further improving the accuracy of the annotation task.

The CRF layer is designed for the NER and slot filling tasks, and the output is as follows:

$$y_{ner} = CRF_{ner}(H_{seq}) \tag{25}$$

$$y_{slot} = CRF_{slot}(H_{seq}) \tag{26}$$

2. DPCNN implementation intention recognition

For the intention recognition task, DPCNN is introduced for classification to help capture key short words and lexical patterns in intention recognition, improve the accuracy of intention recognition, and output the result through the SoftMax activation function, whose output result is as follows:

$$y_{intent} = softmax(Wh_{DPCNN} + b) \tag{27}$$

where y_{intent} is the output probability distribution; W is the weight matrix of all connected layers; h_{DPCNN} is the feature representation obtained after convolution and pooling of input data by DPCNN; and b is the offset of the fully connected layer.

3.4.2. Loss Function

In order to balance the learning among multiple tasks, a loss function is introduced to measure the model performance in multi-task learning, promote shared learning among tasks, and prevent overfitting. The loss function of the multi-task learning framework is composed of the weighted summation of the loss function of each task.

Named entity recognition loss: The CRF loss function is used to calculate the loss of the entity recognition task.

Intention recognition loss: The cross-entropy loss function is used to calculate the loss of the intention recognition task.

Slot filling loss: The CRF loss function is used to calculate the loss of the slot filling task.

The total loss function is calculated as follows:

$$L_{total} = \alpha L_{ner} + \beta L_{intent} + \gamma L_{slot} \tag{28}$$

where, α, β, γ are weight hyperparameters, which are used to balance the loss of different tasks.

3.4.3. Model Training

For named entity recognition and intention classification tasks, each task has its own training dataset, a separate output layer (CRF layer or full connection layer) is created for each task, and the loss function of each task is defined. In the process of training, the parameters of the sharing layer and task-specific layer are optimized step by step by training different tasks alternately. In each batch of each round, the model randomly selects a task (NER, intention classification, slot filling) for training. For the selected task, the model extracts a batch of data from the training dataset of the task, then only carries out forward propagation and back propagation for the data of the task, calculates the gradient of the loss function, and optimizes the parameters of the model by back propagation. In each round of training, the model trains multiple batches of data, and the data of each batch are extracted from the training sets of different tasks. Although the optimization of each task is carried out independently, the shared parameters are cumulatively optimized by alternating training of different tasks, which ultimately improves the performance of the model on all tasks.

4. Experiments and Results Analysis

4.1. Dataset

The data of the experiment come from two parts: a public dataset and self-built data. The data for constructing the knowledge graph come from the dataset crawled by 39 health networks, including 15 items of information encompassing 7 types of entities, about 33,000 entities, and 170,000 entity relations. The entity relationship includes seven categories: diagnostic examination items, departments, diseases, drugs, food, disease sites, and disease symptoms, which are used to construct the knowledge graph of intelligent dialogue.The label distribution is shown in Table 3.

Table 3. Database label distribution.

Entity Type	Meaning	Number of Entities
Checks	Check item	3353
Departments	Department	82
Diseases	Illness	14,336
Drugs	Restoratives	4625
Foods	Food	4870
Body	The body part of the disease	82
Symptom	Disease symptom	5998
Total		33,346

Named entity recognition uses the CBLUE dataset [30], including 15,000 pieces of open training set data, 5000 items of verification set data, and 3000 items of test set data. The dataset of medical text naming entities is divided into nine categories, namely disease (dis), clinical manifestation (sym), drug (dru), medical equipment (equ), medical procedure (pro), body (bod), medical laboratory item (ite), microbiology (mic), and department (dep).The label distribution is shown in Table 4.

Table 4. Distribution of CBLUE datasets.

Entity Type	Label	Training Set	Verification Set
Body	bod	24,106	7361
Diseases	dis	19,371	6328
Clinical manifestation	sym	17,118	5297
Medical procedure	pro	9688	3319
Drugs	dru	4410	1566
Medical laboratory item	ite	4379	1339
Microorganisms	mic	2343	621
Medical equipment	equ	889	164
Department	dep	345	86

Intention recognition uses the CMID Chinese medical intention dataset [31]; this contains 12,254 medical questions, which can be divided into 4 categories and 36 subcategories. According to the relationship types of knowledge graph constructed in this paper, some of them are selected. These are definition, etiology, prevention, clinical manifestation (disease symptoms), related diseases, treatment methods, departments to which they belong, infectivity, cure rate, drug taboos, laboratory/physical examination programs, treatment time, and others, totaling 8105 items. This is divided into training set, verification set, and test set according to the ratio of 8:1:1.The label distribution is shown in Table 5.

Table 5. Distribution of CIMD datasets.

Category	Quantity	Maximum Length	Minimum Length
definition	440	131	5
etiology	970	157	5
prevention	138	111	6

Table 5. *Cont.*

Category	Quantity	Maximum Length	Minimum Length
clinical manifestation (disease symptoms)	1313	217	5
related diseases	242	139	8
treatment methods	2101	185	5
departments	26	60	6
infectivity	43	111	5
cure rate	144	135	6
drug taboos	235	102	7
laboratory/physical examination programs	80	107	7
treatment time	41	119	9
others	2332	221	5
total	8105	221	5

4.2. Evaluation Indicators

In order to evaluate the effectiveness of the model, the accuracy P, recall rate R, and F1 are used to evaluate the effectiveness of the model. The accuracy rate P is the proportion of the samples predicted by the model as a positive class, the recall rate R is the proportion that is successfully identified as a positive class by the model, and F1 is the harmonic average of accuracy and recall rate. The calculation formulas are as follows:

$$P = \frac{TP}{(TP + FP)} \tag{29}$$

$$R = \frac{TP}{(TP + FN)} \tag{30}$$

$$F1 = \frac{2PR}{(P + R)} \tag{31}$$

where TP is the part of the correct identification of the model; FP represents the part of the wrong recognition of the model; and FN represents the missing part of the model.

4.3. Experimental Environment and Parameter Settings

The experiments in this article were based on Python 3.7 and implemented using the Pytorch deep learning framework. The CPU was an Intel CORE i9. The GPU graphics card was a GeForce RTX 3050 and the running memory was 16 GB. The parameters of the multi-task model mainly included the parameters of BERT, BiGRU, and BILSTM. When other parameters were unchanged, the values of variable parameters were changed successively to obtain the optimal parameters of the model. The experiment adopted the Chinese pre-training model "BERT-Base" released by Harbin Engineering University, which contains 12 layers, a hidden layer with 768 dimensions, 12 head models, a total of 110 million parameters, and 128 nodes in the hidden layer. The number of BiGRU layers is 1, the number of BiLSTM layers is 1, the fall rate is set to 0.1 in the entity recognition task, and the fall rate is set to 0.5 in the DPCNN layer. The batch size is set to 32, the global learning rate is 2×10^{-5}, the Adam optimizer is used as the adaptive learning rate optimization algorithm, the convolution kernel of DPCNN is set to 3, and the number of training epochs is set to 20.

In order to verify the validity of the proposed question answering model based on multi-task learning, several models with better performance in intention recognition and named entity recognition are selected for comparison.

4.4. Experimental Results and Analysis

We trained our model using named entity recognition and intent classification training datasets, and tested the performance of named entity recognition and intent classification

tasks using F1 values, recall rates, and accuracy rates to evaluate model performance in this study.

The model proposed in this paper was validated in the test set. For the entity recognition task, the accuracy of the model reached 0.8607 in the test set, and for the intention recognition task, the accuracy of the model reached 0.8925.

In order to verify the superiority of the model, the constructed model was compared with BERT, BERTCRF, Bert-BiLSTM and Bert-BiLSTM-CRF models for the entity recognition task, and the results are shown in Table 6:

Table 6. Comparative experimental results of entity recognition.

	Precision	Recall	Fl
BERT	0.7999	0.7271	0.7617
BERT-CRF	0.7901	0.7752	0.7825
BERT-BiLSTM	0.7703	0.7395	0.7545
BERT-BiLSTM-CRF	0.8254	0.8196	0.8224
Ours	0.8607	0.8135	0.8364

As can be seen from Table 6, the method proposed in this paper has higher accuracy and F1 value, and maintains a high recall rate of 0.8035, while the accuracy reaches 0.8607, which shows its superiority. This shows that the method proposed in this paper can improve the performance of entity recognition. From the comparison, the main conclusions are as follows:

(1) The output of BERT is often the result of word-by-word classification. Adding a CRF layer can optimize the output and make the dependency between tag sequences more reasonable, so as to improve the overall consistency of sequence tagging. Through comparison, it can be found that CRF can effectively capture more relevant information, reduce the missed detection rate, and improve the overall F1 score.

(2) Compared with using BERT or BiLSTM alone, it has the advantage of capturing long-distance dependencies and context information, and is more able to identify complex sequence dependencies. However, compared with simple BERT, the performance improvement of the BERT-BiLSTM model is limited. With the combination of BiLSTM and CRF layers, the model can make use of both the contextual feature extraction ability of BiLSTM and the tag dependency modeling ability of CRF. BiLSTM handles long-distance dependencies and CRF optimizes the dependencies of output tags. The combination of the two makes the model significantly improve in entity recognition tasks. In particular, both precision and recall are improved, which shows that the model has achieved good results in reducing missed detection and false detection.

(3) By adding BiGRU and BiLSTM layers to the output of BERT, the model can make full use of two different cyclic neural network (RNN) structures to process sequence data. BiGRU is relatively simple and can quickly capture short-term dependencies, while BiLSTM is better at capturing long-term dependencies. This dual RNN architecture can effectively balance the capture ability of short-term and long-term dependencies and make the model more robust when dealing with serial data with complex dependencies. This characteristic is particularly evident in the identification of long-distance dependent entities. From the experimental results, this model achieves a good balance between accuracy and recall rate. Compared with other models, this model performs well in capturing the correct entities and covering as many entities as possible. This balance makes the model more practical in real-world application, as it can not only reduce false detection (improve accuracy), but also reduce missed detection (improve recall rate), so as to provide users with more reliable entity recognition results.

For the intention recognition task, we compared the proposed model with CNN, BERT, BiLSTM, TextCNN, and TextRNN, and the results are shown in Table 7.

Table 7. Intention recognition and comparison of experimental results.

	Precision	Recall	Fl
CNN	0.8315	0.7921	0.8113
BILSTM	0.8377	0.8086	0.8229
BERT	0.8582	0.8391	0.8485
TextCNN	0.8625	0.8511	0.8567
TextRNN	0.8634	0.8602	0.8618
Ours	0.8925	0.8832	0.8878

As can be seen from Table 7, the method proposed in this paper shows good results in accuracy rate, recall rate, and F1 value, and the recall rate was maintained at 88% while the accuracy rate reached 89%.

Through comparative experiments, the following can be concluded:

(1) From the introduction of BERT, the CNN and BILSTM models provide relatively low accuracy, and these results show their limitations in capturing contextual information and capturing sequential dependencies. The CNN model mainly extracts local features through convolution operations, which may lose long-distance dependent information. BILSTM is good enough to capture better sequence-dependent information, but its handling of long-distance dependencies may be inadequate. With the introduction of the BERT model, the effect is significantly improved, possibly because the BERT pre-trained model is better able to understand contextual information and capture the bidirectional dependencies of words in a multi-layer bidirectional structure to improve the understanding of the global context of the sentence.

(2) By combining BERT, BiGRU, BiLSTM, and DPCNN, the model can take advantage of the advantages of various frames. BERT provides powerful context understanding, BiGRU and BiLSTM enhance the capture of long- and short-term dependencies, and DPCNN further extracts deep textual features. This multi-layered, hybrid architecture makes the model perform well in intention recognition tasks. In particular, the introduction of DPCNN improves the feature extraction ability of the model, meaning the model can more accurately identify complex intentions.

4.5. Ablation Experiment

In order to test the effect of the proposed method on performance and explore the effect on contextual information and complex text, ablation experiments on the above models were carried out on CBLUE and CMID datasets. The experimental results are shown in Table 8.

Table 8. Ablation results.

Name	Precision	Recall	Fl
Module1	0.8368	0.8017	0.8189
Module2	0.8319	0.8203	0.8261
Ours-N	0.8607	0.8135	0.8364
Module3	0.8838	0.8673	0.8754
Module4	0.8628	0.8503	0.8565
Ours-I	0.8925	0.8832	0.8878

Among them, Modules 1 and 3 are the single task test results of entity recognition and intention recognition, respectively. For entity recognition and intention recognition tasks, F1 values decrease by 1.8% and 1.2%, respectively, compared with the model proposed in this paper, which verifies that multi-task learning can effectively capture the correlation between different tasks. The model's ability to understand specific tasks is improved. It can be seen from the ablation experiment that in the entity recognition task, the auxiliary task of intention recognition enables the model to better understand the semantic relationship in

the context, thus improving the accuracy of entity recognition. In the intention recognition task, the additional features provided by entity recognition enhance the model's understanding of the user's input intention, and further improve the accuracy of the intention recognition. Ablation results also show that multi-task learning can make better use of training data, especially in the case of relatively limited data. By training multiple tasks at the same time, the model can more fully learn the common features between different tasks, thereby improving the overall performance. However, it has to be admitted that although multi-task learning has the potential to improve the performance of the model, negative transfer is a potential problem that needs to be paid attention to. Future research should focus on how to effectively alleviate this phenomenon and find ways to optimize task combination and the sharing mechanism to ensure that the impact of negative migration on model performance is reduced while maximizing task synergy.

Modules 2 and 4 are the experimental results of entity recognition and intention recognition after the BiGRU layer is removed from the sharing layer, respectively. For entity recognition and intention recognition tasks, F1 values are reduced by 1% and 3.1%, respectively, but the recall rate of entity recognition tasks is increased by 0.7%, which verifies the ability of two-way information flow with BiGRU. The ability to capture forward and backward context dependencies simultaneously plays an important role in deep secondary feature extraction in the framework of multi-task learning. Through the ablation experiment, we can see that after removing the BiGRU layer, although the recall rate in the entity recognition task is slightly improved, the overall F1 value decreases; in particular, in the intention recognition task, the F1 value decreases significantly by 3.1%. The BiGRU layer can capture the long-term dependency information in the sequence data; in particular, in the intention recognition task, it helps the model to understand the global information of the context more accurately and improve the ability to identify complex intentions. Although the recall rate of entity recognition tasks is improved after removing the BiGRU layer, this can be due to the weakening of the overfitting of some entities by the model, but it also leads to a decrease in the accuracy of some entities, resulting in a decline in F1 values as a whole.

In addition, although complex deep learning models such as BERT, BiGRU, and BiLSTM perform well in processing medical data, their lack of interpretability may pose risks. Especially in the medical field, the transparency of model decision-making is very important to the trust of clinicians and the acceptability of results. Therefore, we need to find a balance between model performance and interpretability. In order to enhance the interpretability of the model, consideration is introduced into the attention mechanism to enable the model to show the input characteristics concerned in the decision-making process. In addition, interpretable techniques such as LIME (locally interpretable model unknowable) and SHAP (Shapley value) can be used to help interpret the prediction results of the model, thereby enhancing doctors' trust in the AI system.

5. Conclusions

In order to improve the accuracy of entity recognition and intention classification in medical question processing, this paper proposes a medical text parsing method based on BiGRU-BiLSTM multi-task learning, which consists of a sharing layer and specific task layer. The sharing layer uses the pre-trained BERT model to generate the input sequence, captures the longer-distance context information, generates a more fine-grained representation of the above and the following through the BiGRU-BiLSTM hybrid neural network, and constructs a multi-task learning sharing layer. For the intention recognition task, DPCNN is used on the output of the BiGRU-BiLSTM layer, which predicts the category of intent for each input text. For entity recognition tasks, in order to better understand the needs of users, CRF is used for sequence tagging, and the output of the BiLSTM layer is used as the input of CRF to obtain the tags of entities in the sequence, which improves the performance of intelligent man–machine dialogue. According to the method proposed in this paper, a hybrid neural network model based on multi-task learning is constructed

and experiments are carried out on CBLUE and CMID datasets. The experimental results show that the accuracy of named entity recognition and intention recognition is 86% and 89%, respectively, which effectively improves the accuracy of semantic understanding, the generalization ability of the model, and the understanding of long and difficult sentences. Through multi-task learning, the processing ability of fine-grained information is improved, the meaning of input text can be better understood, the context information in text can be captured more easily, and the processing of complex text and long sentences is improved to some extent.

However, the method proposed in this paper also has some shortcomings. Large-scale and high-quality tagged data are needed in the process of deep learning model training, but the data labeling in the medical field is relatively complex and requires a professional background. Therefore, the scale of the data used in this paper is not large enough; as a result, in this method there is still some deviation in understanding in text parsing.

In our future work, we will introduce deeper context understanding to enhance the model's capture of long-distance dependency relationships, and because data from some specific diseases or areas in the medical field may be scarce, we will study the optimization model in a low-resource environment, and explore ways to introduce a large language model to achieve higher accuracy in the case of balancing resources and efficiency.

6. Patents

This research result has been applied to a patent (patent application number: 202411054736.7) which is currently in the process of acceptance.

Author Contributions: Y.F., R.K., W.H. and L.L. conceived and designed the method; Y.F. implemented the entire model; W.H. and L.L. processed the dataset as required; Y.F., W.H. and L.L. debugged relevant parameters and drew data graphs; Y.F. wrote the paper; R.K. revised the paper; W.H. and L.L. helped write the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset supporting the conclusions of this article is available at https://github.com/CBLUEbenchmark/CBLUE and https://github.com/IMU-MachineLearningSXD/CMID, accessed on 18 September 2024.

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

References

1. He, K.; Mao, R.; Lin, Q.; Ruan, Y.; Lan, X.; Feng, M.; Cambria, E. A survey of large language models for healthcare: From data, technology, and applications to accountability and ethics. *arXiv* **2023**, arXiv:2310.05694.
2. Singhal, K.; Tu, T.; Gottweis, J.; Sayres, R.; Wulczyn, E.; Hou, L.; Clark, K.; Pfohl, S.; Cole-Lewis, H.; Neal, D. Towards expert-level medical question answering with large language models. *arXiv* **2023**, arXiv:2305.09617.
3. Shanavas, N.; Wang, H.; Lin, Z.; Hawe, G. Ontology-based enriched concept graphs for medical document classification. *Inf. Sci.* **2020**, *525*, 172–181. [CrossRef]
4. Lenivtceva, I.; Slasten, E.; Kashina, M.; Kopanitsa, G. Applicability of machine learning methods to multi-label medical text classification. In Proceedings of the Computational Science–ICCS 2020: 20th International Conference, Amsterdam, The Netherlands, 3–5 June 2020; Proceedings, Part IV. pp. 509–522.
5. Zhang, Y.; Lu, W.; Ou, W.; Zhang, G.; Zhang, X.; Cheng, J.; Zhang, W. Chinese medical question answer selection via hybrid models based on CNN and GRU. *Multimed. Tools Appl.* **2019**, *79*, 14751–14776. [CrossRef]
6. Shi, H.; Liu, X.; Shi, G.; Li, D.; Ding, S. Research on medical automatic Question answering model based on knowledge graph. In Proceedings of the 2023 35th Chinese Control and Decision Conference (CCDC), Yichang, China, 20–22 May 2023; pp. 1778–1782.
7. Meiling, W.; Xiaohai, H.; Yan, L.; Linbo, Q.; Zhao, Z.; Honggang, C. MAGE: Multi-scale Context-aware Interaction based on Multi-granularity Embedding for Chinese Medical Question Answer Matching. *Comput. Methods Programs Biomed.* **2023**, *228*, 107249.
8. Huang, X. Design and Implementation of Medical Question Answering System Based on ALBERT. Master's Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2022.
9. Guo, Y.; Xie, Z.; Chen, X.; Chen, H.; Wang, L.; Du, H.; Wei, S.; Zhao, Y.; Li, Q.; Wu, G. ESIE-BERT: Enriching sub-words information explicitly with BERT for intent classification and slot filling. *Neurocomputing* **2024**, *591*, 127725. [CrossRef]

10. Deng, F.; Chen, Y.; Chen, X.; Li, J. The joint model of multi-intention recognition and slot filling of GL-GIN is improved. *Appl. Comput. Syst.* **2023**, *32*, 75–83.
11. Huang, Y.; Kui, J.; Guan, C. Intelligent medical question answering system based on BERT-BiGRU model. *Softw. Eng.* **2024**, *27*, 11–14+25. [CrossRef]
12. Wang, Z.; Zheng, K. Chinese medical question answering system based on BERT. *Comput. Syst. Appl.* **2023**, *32*, 115–120. [CrossRef]
13. Hu, W.; Zhang, Y. Medical entity recognition method based on BERT-BiGRU-CRF. *Comput. Age* **2023**, 24–27.
14. Kollias, D. ABAW: Learning from Synthetic Data & Multi-task Learning Challenges. In Proceedings of the Computer Vision—ECCV 2022 Workshops, Tel Aviv, Israel, 23–27 October 2022; Springer: Cham, Switzerland, 2023; pp. 157–172.
15. Vandenhende, S.; Georgoulis, S.; Proesmans, M.; Dai, D.; Van Gool, L. Revisiting multi-task learning in the deep learning era. *arXiv* **2020**, arXiv:2004.13379.
16. Wu, Q.; Peng, D. MTL-BERT: A Multi-task Learning Model Utilizing Bert for Chinese Text. *Mini-Microcomput. Syst.* **2021**, *42*, 291–296.
17. Zhang, J. Research on Named Entity Recognition Method for Chinese Electronic Medical Records. Master's Thesis, Zhejiang University of Science and Technology, Hangzhou, China, 2022.
18. Gao, W.; Li, Y.; Guan, X.; Chen, S.; Zhao, S. Research on Named Named entity recognition Based on Multi-Task Learning and Biaffine Mechanism. *Comput. Intell. Neurosci.* **2022**, *2022*, 2687615. [CrossRef] [PubMed]
19. Zeng, D.; Zhang, H.; Liu, Q. CopyMTL: Copy Mechanism for Joint Extraction of Entities and Relations with Multi-Task Learning. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 9507–9514. [CrossRef]
20. Peng, Y.; Chen, Q.; Lu, Z. An empirical study of multi-task learning on BERT for biomedical text mining. *arXiv* **2020**, arXiv:2005.02799.
21. Liao, M.; Jia, Z.; Li, T. Chinese named entity recognition based on tag information fusion and multi-task learning. *Comput. Sci.* **2024**, *51*, 198–204.
22. Li, Y. Speech Emotion Recognition with Multitask Learning. Master's Thesis, China University of Mining and Technology, Xuzhou, China, 2023.
23. Zhu, Q.; Wang, X.; Liu, X.; Du, W.; Ding, X. Multi-task learning for aspect level semantic classification combining complex aspect target semantic enhancement and adaptive local focus. *Math. Biosci. Eng. MBE* **2023**, *20*, 18566–18591. [CrossRef]
24. Myint, P.Y.W.; Lo, S.L.; Zhang, Y. Unveiling the dynamics of crisis events: Sentiment and emotion analysis via multi-task learning with attention mechanism and subject-based intent prediction. *Inf. Process. Manag.* **2024**, *61*, 103695. [CrossRef]
25. Song, D.; Hu, M.; Ding, J.; Qu, Z.; Chang, Z.; Qian, L. Research on cross-type text classification technology based on multi-task learning. *Data Anal. Knowl. Discov.* **2024**, 1–19. [CrossRef]
26. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2018**, arXiv:1810.04805.
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
28. Cho, K. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.
29. Johnson, R.; Zhang, T. Deep pyramid convolutional neural networks for text categorization. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, BC, Canada, 30 July–4 August 2017; pp. 562–570.
30. Zhang, N.; Chen, M.; Bi, Z.; Liang, X.; Li, L.; Shang, X.; Yin, K.; Tan, C.; Xu, J.; Huang, F. Cblue: A chinese biomedical language understanding evaluation benchmark. *arXiv* **2021**, arXiv:2106.08087.
31. Chen, N.; Su, X.; Liu, T.; Hao, Q.; Wei, M. A benchmark dataset and case study for Chinese medical question intent classification. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 125. [CrossRef] [PubMed]

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