Relation Extraction Method Based on Pre-trained Large Language Model BERT

WU QIWEI, ZHANG YUCI, QIANG ZIYAN
University Of Macau
MC56730,MC56685,MC56461@um.edu.mo

Abstract

Relation extraction is a core task focused on identifying semantic relationships between entities within unstructured text. Traditional pipeline approaches treat named entity recognition and relation extraction as separate processes, whereas recent end-to-end methods have achieved superior performance by leveraging joint modeling techniques. This study proposes a relation extraction method based on the pre-trained large language model BERT and conducts experiments on the relation classification task in the Chinese medical domain.We employed Chinese-RoBERTa-wwm-ext as the pre-trained model and designed multiple text encoding strategies, such as entity tagging and question-answer encoding. Through model fine-tuning, we enhanced the accuracy of relation extraction. Experimental results demonstrate that the model based on Chinese-RoBERTa-wwm-ext exhibits excellent performance in relation classification tasks. Compared with other pre-trained models (BioBERT and MedBERT), our method achieves a Micro-F1 score improvement of 44.18% and 1.21%, respectively. This outcome highlights the significance of Chinese domain-specific optimization and entity tagging for this task.Furthermore, we effectively alleviated the "hard class" problem through a class enhancement strategy, which further improved the models performance on categories that are difficult to distinguish. This research provides novel methods and insights for relation extraction in Chinese medical texts, and holds substantial practical application value.

Keywords: Chinese Medical Texts; BERT; Relation Extraction; Pre-trained Language Models; Relation Classification

1 Introduction

Extracting entities and their semantic relations from unstructured text is a basic information extraction problem. This problem can be decomposed into 2 subtasks: Named Entity Recognition [1, 2]and Relation Extraction [3, 4].Early works adopted pipeline methods: one model is trained to extract entities[5], and another to classify relationships between them[6, 7]. Recently, end-to-end evaluation methods have become popular, where systems jointly model these two tasks[8, 9, 10, 11, 12]. It has long been believed that end-to-end joint models can better capture interactions between entities and relationships and help mitigate error propagation.

Recently, with the emergence of pre-trained large language models, some researchers have proposed a simpler approach. This method learns two encoders based on deep pre-trained language models[13, 14, 15]: one for named entity recognition (entity model) and the other for relation extraction (relation model). These encoders are trained independently, and the relation model relies only on the entity model to provide input features. The entity model is built on word-level representations, whereas the relation model is built on context representations of specific entity pairs. This is an efficient pipeline method but introduces connections between the two tasksusing the output of the entity model to assist the relation model in relation extraction, embodying a joint learning idea.

In the medical field, relation extraction faces special challenges: medical texts contain a large number of professional terms and nested entities[16]. In Chinese texts, there are also problems such as flexible vocabulary structure and ambiguous entity boundaries[17]. Moreover, labeled data is scarce due to privacy and cost issues. The existing pre-trained models still have room for improvement in adapting to the Chinese medical scenario, which also provides the core motivation for this study to adopt the Chinese-RoBERTa-wwm-ext model and design a dedicated text encoding method.

2 Related Work

Traditionally, extracting relationships between entities in text has been studied as two independent tasks: named entity recognition and relation extraction. The main workflow of pipeline-based (Figure 1) relation extraction involves performing relation extraction on sentences with pre-annotated target entity pairs and outputting triples with entity relationships as prediction results. A series of pipeline-based relation extraction models have been proposed, among which network structures based on RNN, CNN, and their improved variants have received significant attention in academia due to their high accuracy.

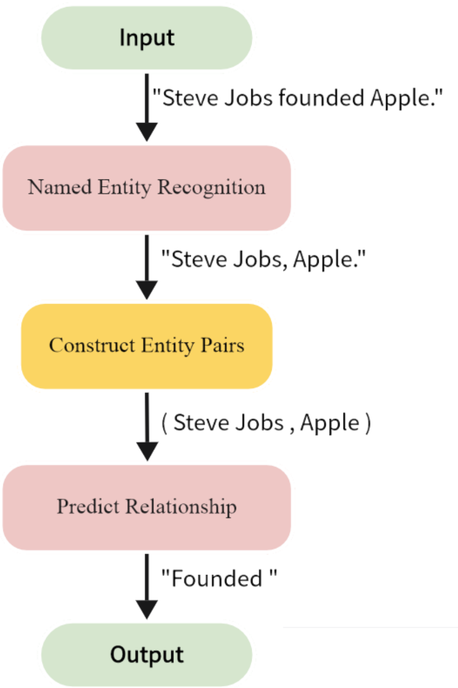

Figure 1: Pipeline-Based Methods

2.1 Pipeline-Based Methods

2.1.1 RNN-Based Entity Relation Extraction

The RNN-based relation extraction method was first proposed by Socher et al.[18] in 2012. This method assigns a vector and a matrix to each node in the parse tree: the vector captures the inherent semantics of components, while the matrix describes how the node affects the meaning of adjacent words or phrases. This matrix-vector RNN model (Figure 2) can learn the semantics of operators in propositional logic and natural language, effectively solving the problem that traditional word vector space models cannot capture the compositional semantics of long and short phrases, thereby advancing natural language understanding to a deeper level.

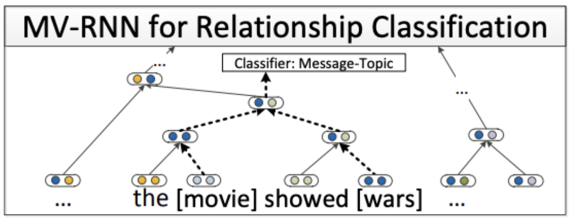

Figure 2: MV-RNN For Relation Classification.

In 2013, Hashimoto et al.[19] proposed a recursive neural network model based on syntactic trees. Unlike the model by Socher et al.[18], Hashimotoś model does not use computationally complex word dependency matrices. Instead, it improves semantic representation by introducing additional features such as part-of-speech (POS) tags, phrase categories, and syntactic heads. Additionally, the model introduces an average parameter mechanism to assign higher weights to key phrases in the target task. Experiments show that adding features and introducing average parameters significantly improve model performance.

In further research, Cai et al.[20] proposed a deep learning relation extraction model based on the shortest dependency path (SDP. in 2016, namely the Bidirectional Recurrent Convolutional Neural Network (BRCNN)[20](Figure 3).

This model combines a Convolutional Neural Network (CNN) with a bidirectional re-

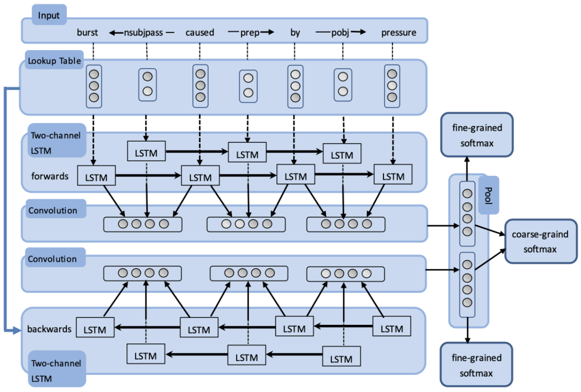

Figure 3: The overall architecture of BRCNN.

current neural network based on Long Short-Term Memory (LSTM) to fully extract information from the dependency relationship in the SDP. Among them, the bidirectional LSTM encodes the global patterns of the shortest dependency path, while the convolutional layer captures local features of adjacent words in dependency links. This fusion improves the models ability to classify the direction of relationships between pairs of entities, further advancing the field of relationship extraction.

2.1.2 CNN-Based Entity Relation Extraction

Zeng et al.[21] first proposed using CNN for relation extraction in 2014 and presented a method based on Convolutional Deep Neural Network[21]. By extracting lexical and sentence-level features, this model takes all word tokens as input and eliminates complex preprocessing steps, effectively avoiding feature error propagation caused by traditional preprocessing systems and significantly improving model performance robustness.

Building on this, Xu et al.[22] further proposed a CNN-based relation extraction model based on dependency parse trees in 2015[22]. The model maps input text to dependency parse trees and designs an innovative negative sampling strategy: first, it uses dependency paths to learn the directionality of relationships; second, it uses negative sampling to learn the positional distribution of subjects and objects. Specifically, the model takes the shortest dependency path from the object to the subject as negative sample input, thereby alleviating the impact of irrelevant information introduced by dependency parse trees on model performance when entity pairs are far apart. This strategy significantly improves the accuracy of relation extraction tasks.

In further research, Wang et al.[23] proposed a novel attention-based CNN architecture in 2016, aiming to capture multi-level attention information related to specified entities and aggregate attention features related to specific relationships through pooling operations (Figure 4).

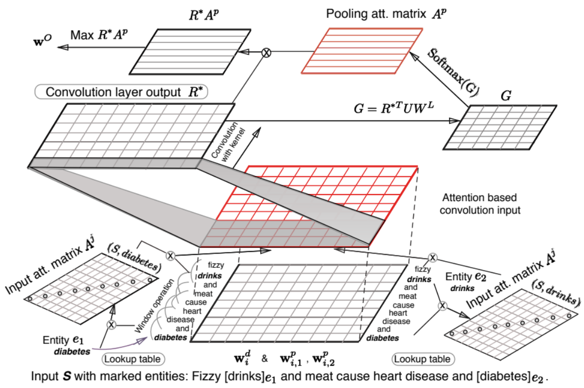

Figure 4: Schematic overview of our Multi-Level Attention Convolutional Neural Networks.

This method enables the model to automatically identify sentence segments related to target relationship categories in heterogeneous sentences. The attention-based pooling method introduced by the model can extract N-gram phrases most significant for relation classification. Additionally, the hybrid attention mechanism can efficiently extract triple fields related to relation classification during aggregation. Experimental results show that this method not only improves model performance but also significantly enhances the ability to capture fine-grained semantic cues.

2.2 Joint Learning-Based Methods

In recent years, due to issues such as error accumulation and propagation, neglect of dependencies between subtasks, and generation of redundant entities in pipeline-based methods, joint models (Figure 5) have gradually gained attention. We classify existing joint models into two categories: structured prediction and multi-task learning.

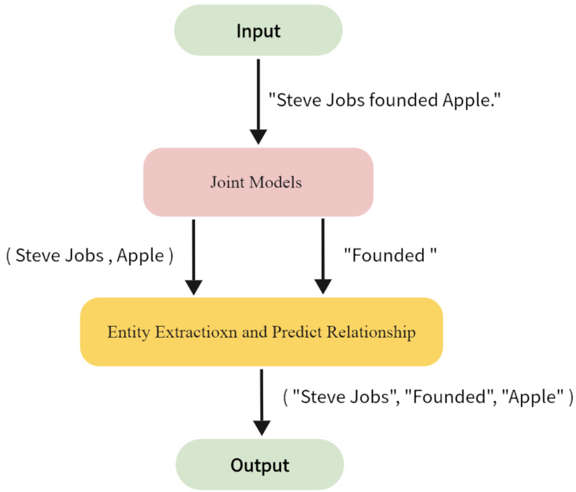

Figure 5: Joint Learning-Based Methods

2.2.1 Structured Prediction

Structured prediction methods integrate these two tasks into a unified framework, although they can be formulated in various ways. Li et al.[8] proposed an action-based system for joint entity and relation extraction in 2014, which can establish links with previous entities while identifying new entities. In 2017, Zhang et al.[10] adopted the table-filling method proposed by Miwa et al.[24] in 2014, integrating the two tasks into a unified framework and processing them simultaneously in a specific manner to achieve joint extraction. In 2017, Katiyar et al.[25] used sequence labeling for joint extraction. In 2019, Sun et al.[26] proposed a graph-based method to jointly predict entity and relation types, better capturing complex dependencies between entities and relationships by constructing graph structures. In 2019, Li et al.[27] transformed the relation extraction task into a multi-turn question answering problem, attempting joint extraction from a new perspective. These structured prediction methods all require solving global optimization problems during inference and use beam search or reinforcement learning for joint decoding to improve the accuracy of joint entity and relation extraction.

2.2.2 Multi-Task Learning

Essentially, multi-task learning constructs two separate models for entities and relationships and optimizes them together through parameter sharing. In 2016, Miwa et al.[9] proposed building separate models for entity recognition and relation extraction: a sequence labeling model for entity prediction and a tree-based LSTM model for relation extraction. The two models share an LSTM layer to obtain contextual word representations, and the performance of both models is optimized through this parameter sharing. The method proposed by Bekoulis et al.[28] in 2018 is similar: it also constructs two separate models but models relation classification as a multi-label head selection problem, and also improves overall performance through parameter sharing. These multi-task learning methods still retain certain pipeline characteristics during decoding, i.e., extracting entities first and then applying the relation model to the extracted entities.

In recent years, multi-task learning has further integrated with knowledge in the medical field. Some studies have introduced medical knowledge graphs to inject structured knowledge such as entity relationships and diagnostic logic into the model[29]. Through parameter sharing or feature fusion, multiple tasks including entity recognition, relationship classification, and diagnostic reasoning are simultaneously optimized. This integration not only alleviates the data sparsity problem of a single task but also enhances the professionalism and accuracy of the model's output. Especially in tasks that require strong domain knowledge support, such as medical relationship extraction, it demonstrates superior performance compared to traditional multi-task learning[30].

3 Task Analysis

For this course project, we selected the task of Chinese medical relation classification, which aims to identify and classify different types of medical relationships from Chinese medical texts. These relationships typically involve connections between entities such as drugs, symptoms, diseases, and treatment methods. For example, the task may require identifying and classifying relationships such as "drug-treats disease", "symptom-diagnoses disease", or "disease-causes complication"making it a multi-classification problem. By training a model to recognize and classify these relationships, we can assist in building medical knowledge graphs, supporting applications such as intelligent diagnosis, medical research, and health information management, which helps improve the efficiency and accuracy of medical data utilization.The data examples are as follows:

Example	
ID	2437
sentence	Food poisoning @ Fever may be caused by extragastrointestinal infection or overlapping infection.
(h)	Food Poisoning
(t)	stress ulcer
(r)	Etiology

Table 1: Data Example. Each piece of data includes: Sentence, Head Entity (h), Tail Entity (t), and Relationship (r).

There are 10 relationship categories: Clinical Manifestation, Drug Treatment, Synonym, Etiology, Complication, Pathological Subtype, Laboratory Test, Adjuvant Treatment, Related (Causes), and Imaging Test.

This dataset is divided into three jsonl files: `train.jsonl`, `val.jsonl`, and `test.jsonl` which contain 7,000, 2,000, and 1,000 pieces of data respectively.

As can be seen from the dataset, our task focuses on the relation extraction component: the data provides accurate head and tail entities for each sentence, and our goal is to determine the relationship between this entity pair in the sentence.

Therefore, we decided to adopt a pipeline-based method, mainly referring to the relation extraction module of the SOTA model[31]. We use a fine-tuned pre-trained large language model for our relation classification task and experiment with different text embedding methods to analyze their effects.

4 Method Description

4.1 Pre-trained Large Language Model BERT

BERT[13] is a pre-trained language representation model based on the Transformer architecture (Figure 6), designed to understand language through deep bidirectional learning. BERT aims to address the limitations of lexical representation in traditional modelsthese models are usually context-independent and cannot effectively capture the diverse meanings of words in different sentences. BERT trains the model through two pre-training tasks: Masked Language Model (MLM) and Next Sentence Prediction (NSP), enabling it to learn lexical representations in different contexts.

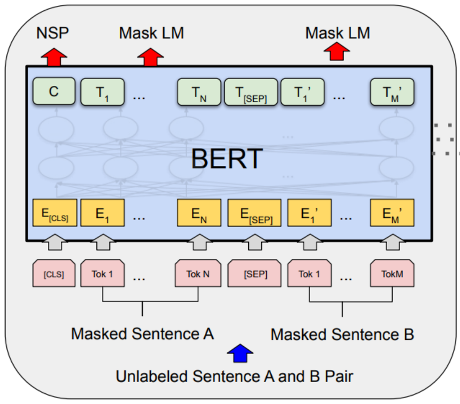

Figure 6: BERT pre-training using MLM and NSP

BERT training consists of two phases: pre-training and fine-tuning (Figure 7). In the pre-training phase, BERT uses a large amount of unannotated data and trains on both MLM and NSP pre-training tasks. The MLM task randomly masks some words in the input sentence and requires the model to predict the original content of these masked words, while the NSP task predicts whether two sentences are consecutive. These two methods can capture both word-level and sentence-level contextual representations. In the fine-tuning phase, based on the pre-trained model, the model is further trained using task-specific annotated data to better adapt to specific tasks.

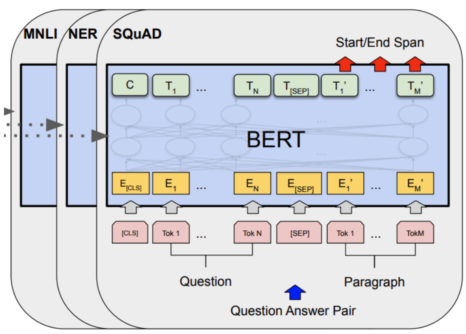

Figure 7: BERT fine-tuning on a variety of tasks

The main innovation of BERT lies in its bidirectional training strategy, which allows the model to consider both left and right contextual information simultaneously, thereby better understanding the contextual meaning of words. Additionally, BERT adopts the Transformer architecture, which enables it to effectively handle long-distance dependency issues and train on large-scale datasets.

4.2 Multiple Downstream Tasks of BERT

BERTs downstream tasks can be divided into four categories: text classification, named entity recognition, question answering systems, and text generation. BERT adapts to specific downstream tasks by fine-tuning the pretrained model.

4.2.1 Text Classification

In text classification tasks(Figure 8), the output of BERTs [CLS] token is used as the basis for classification, as the [CLS] token contains contextual information of the entire sentence. To adapt to text classification tasks, BERTs input data requires specific preprocessing, such as segmenting text into tokens, adding special tokens (e.g., [CLS] and [SEP]), and fine-tuning to adapt to the classification task.

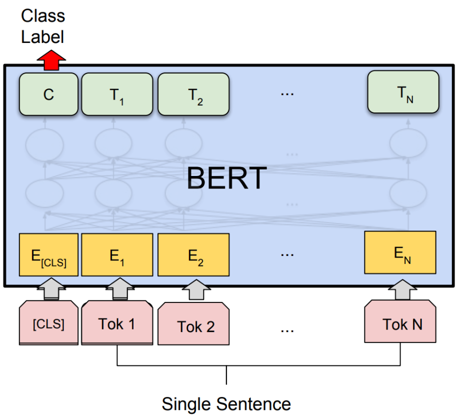

Figure 8: Text Classification

4.2.2 Named Entity Recognition

Named Entity Recognition (NER) is the task of identifying specific types of entities (e.g., person names, place names, organization names) in text (Figure 9). The application of BERT in NER involves inputting text into the model and outputting predicted labels for various entity categories in sequence. When processing NER tasks, the output of each token in BERT is used to predict whether the token belongs to a specific entity category and which category it belongs to.

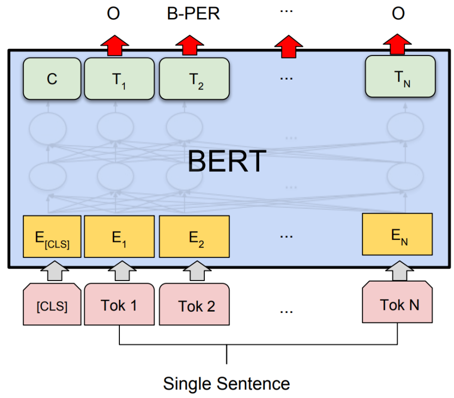

Figure 9: Named Entity Recognition

4.2.3 Question Answering Systems

The application of BERT in question answering systems mainly lies in its ability to capture the contextual relationship between questions and answers (Figure 10).

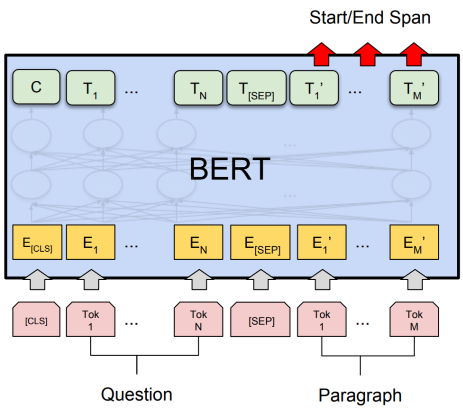

Figure 10: Question Answering Systems

Through fine-tuning, BERT can learn to predict the start and end positions of answers based on questions, thereby processing question-answer pairs.

4.2.4 Text Generation

BERT can also be used for text summarizationspecifically, the BERTSUM model (Figure11), which is built on BERT, specializes in extracting the most important information from long texts for extractive summarization.

This typically involves using the decoder part of BERT or a BERT-based sequence-to-sequence model, and training the model to

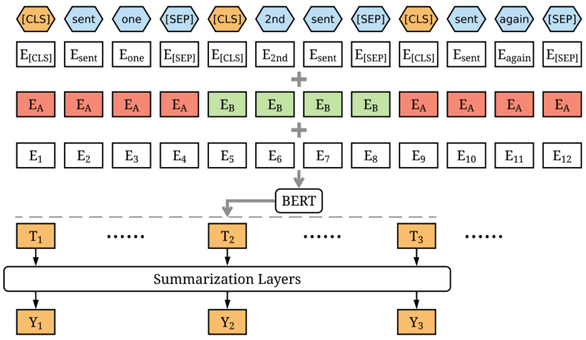

Figure 11: Text Generation

learn how to generate summaries from input text.

Our task is relation classification, so we can leverage BERTs excellent performance in text classification. However, there are certain differences and connections between relation classification and text classification: relation classification focuses on identifying relationships between entities in text, while text classification classifies the sentiment or topic of the entire text. In relation classification, BERTs output includes not only the output of the [CLS] token but also the output of tokens corresponding to entities. Therefore, we can design entity tags such as [E1] and [\E1] for BERT to learn, and the output of BERT for these tags is used for relation classification.

4.3 BERT Variants for Different Domains

Different optimized versions of BERT[13] are pre-trained on domain-specific data to improve performance in these domains.

BioBERT[32] is an English BERT model pre-trained on biomedical literature and clinical notes (Figure 12).

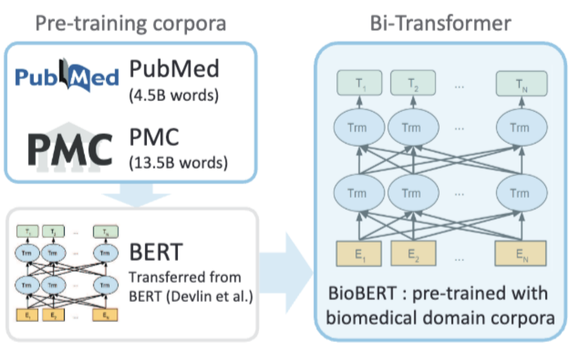

Figure 12: BioBERT

Its improvement lies in its ability to better capture professional terms and context in the biomedical field. BioBERT performs excellently in medical domain named entity recognition and relation extraction tasks. The development of BioBERT is of great significance to the field of biomedical text processing: it not only retains BioBERTs strong ability to process biomedical text but also significantly reduces the model size, enabling it to be used in resource-constrained environments.

Chinese-RoBERTa-wwm-ext[33] is a Chinese BERT pre-trained model released by the Harbin Institute of Technology and iFLYTEK Joint Laboratory (Figure 13).

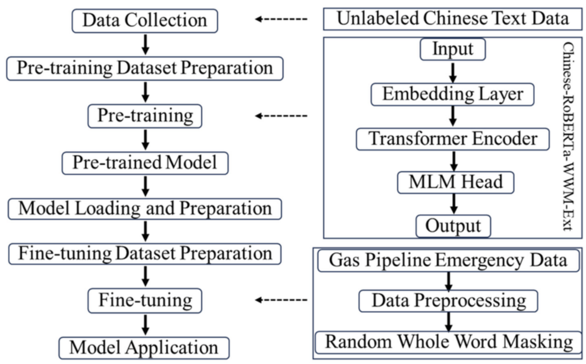

Figure 13: Chinese-RoBERTa-wwm-ext

Its improvement lies in the adoption of whole-word masking technology. Compared with character-level masking, it can more effectively learn word-level semantics and improve the effect of Chinese natural language processing. This model is particularly suitable for processing Chinese text because it can better understand and handle the linguistic characteristics of Chinese, and has achieved further performance improvements on multiple benchmark tests.

MedBERT is a pre-trained Transformer model for biomedical named entity recognition [34], which is initialized based on Bio ClinicalBERT and pre-trained using biomedical-related texts from N2C2, BioNLP, CRAFT challenges, and Wikipedia scraping (Figure 14). Through comparative experiments with BERT, DistilBERT, BioBERT, and Bio ClinicalBERT on 10 biomedical datasets from BioNLP and CRAFT tasks, it was found that MedBERT achieved the best performance on 9 datasets, and the model weights and fine-tuning code are publicly available, providing a reference baseline for biomedical NER tasks.

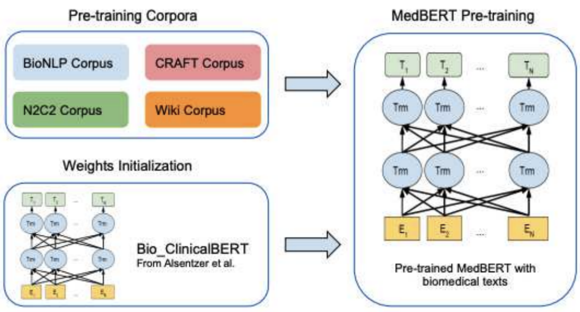

Figure 14: Overview of the pre-training of MedBERT

4.4 Our Approach

We mainly use Chinese-RoBERTa-wwm-ext [33] as the pre-trained large language model for three reasons: first, it is a language model trained on Chinese corpora, making it particularly suitable for Chinese natural language tasks; second, it adopts whole-word masking technology, which performs masking training on complete words, improving the ability to understand word-level informationespecially performing better in domain-specific named entity recognition and relation extraction tasks.

Meanwhile, we experiment with 5 different text encoding modes, including adjusting the input order of text and entities and adding custom entity tags (e.g., [E1], [\E1]), to analyze changes in model performance under fine-tuning with different encoding modes. The implementation of different text encoding methods is shown in Table 2.

Method	Pattern
Basic1	*[CLS] {h} [SEP] sentence [SEP] {t}*
Basic2	*[CLS] {h} [SEP] {t} [SEP] sentence*
QA	*[CLS] What is the relationship between {h} and {t}? [SEP] sentence*
Entity1[1]	*[CLS] s1 [E1] {h} [/E1] s2 [E2] {t} [/E2] s3 [SEP]*
Entity2[2]	*[CLS] s1 [Entity1] {h} [/Entity1] s2 [Entity2] {t} [/Entity2] s3 [SEP]*

Table 2: Encoding schemes. s1/2/3 denote sentence segments around entity spans.

1.Entity1 is Entity_marked1
2.Entity2 is Entity_marked2

During experiments, we noticed that the models performance in distinguishing between "Complication" and "Related (Causes)" categories was significantly lower than that in other categories. Therefore, we conducted enhanced training on these two poorly performing categories, including designing a weighted loss and a data augmentation method. Finally, we obtained a model with more balanced performance across different categories.

4.5 Class Enhancement Strategy

Two categories (Complication and Related (Causes)) show lower discriminability. We apply weighted cross-entropy and weak sample augmentation (duplicating or lightly perturbing minority/difficult instances). The weighted loss is:

$$\mathcal{L} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C}\omega_c\, y_c^{(n)} \log p_c^{(n)} \tag{1}$$

with class weights ω_c emphasizing hard categories.

5 System Implementation

This course project provided sentences and the head and tail entities of the relations to be extracted in the dataset, and limited the relation categories to a fixed 10 types. Therefore, we could use the pre-trained BERT for relation (text) classification. Since it is a Chinese task, I chose the `hfl/chinese-roberta-wwm-ext` model, and compared it with the `BioBERT` model and the `MedBERT` model.

During the training process, by using the `BertForSequenceClassification class` from the `transformers` library, it is very easy to perform fine-tuning and inference in the text classification task. We used this method to fine-tune on the training data and tried five different encoding methods for the training statements.

5.1 Environment configuration

The required environment configuration for this project is as follows:

- torch==2.2.2
- transformers==4.40.2
- scikit-learn==1.4.2
- numpy==1.24.4
- Other common packages

5.2 Project file structure

Package	Main File
.venv	bin / lib / share / include
checkpoint	.getkeep
output	loss / log
dataset	train.jsonl / val.jsonl / test.jsonl
run	data.py / train.py / test.py / log.py

Table 3: File Structure.This project development and implementation document structure table.

5.3 Pipeline Summary

This section summarizes the concrete implementation based on `data.py`, `train.py`, and `test.py`.

	Details
Tokenizer	Load pretrained; Add special tokens for entity-marked modes
Samplers	BalancedBatchSampler (uniform across classes); PiorityBatchSampler (focus on hard classes 4,8)
Loss	Weighted cross-entropy (w[4]=w[8]=2.0)
Two Stages	First balanced epochs Then priority-focused epochs
Metrics	Accuracy (dev), full classification report, Macro/Micro-F1
Persistence	Per-mode checkpoints `./checkpoint/model{v}`

Table 4: System components of the training/evaluation pipeline.

5.4 Implementation Highlights

We employ a two-stage curriculum with balanced sampling followed by hard-class priority sampling, weighted cross-entropy for classes (Complication, Related), and per-encoding checkpoints. Full training details remain in Section *train.py*.

5.5 Design Rationale

Balanced sampling prevents early bias; priority sampling concentrates gradient updates on hard categories (Complication, Related). Weighted loss amplifies minority/error-prone signal; marker tokens provide explicit entity boundaries improving attention focus.

6 Results Analysis

6.1 Pretrained Model Comparison

Chinese-RoBERTa-wwm-ext outperforms BioBERT and MedBERT (Table 5), highlighting domain and language alignment importance. BioBERT's English biomedical pretraining transfers poorly to Chinese.

Table 5: Scores of Chinese-RoBERTa-wwm-ext with Different Text Encoding Methods.

Model	BioBERT[1]	Chinese-RoBERTa[2]	MedBERT[3]
Macro-P	46.00%	**90.93**%	89.79%
Macro-R	36.30%	**90.20**%	88.70%
Macro-F1	30.80%	**90.24**%	88.80%
Micro-F1	46.28%	**90.46**%	89.25%

1.BioBERT : Entity_marked1
2.Chinese-RoBERTa : Entity_marked2
3.MedBERT : QA

It can be seen that BioBERT performs significantly lower than the other two models in all indicators. Although BioBERT is a model pre-trained specifically for the biomedical field, it is mainly pre-trained on English corpora, while this task deals with Chinese medical texts, so there is an obvious lack of corpus adaptability. Chinese-RoBERTa-wwm-ext achieved the highest scores in all indicators, and the entity_marked2 encoding mode further optimized the entity tagging method, improving the model's adaptability to the relation classification task through annotation forms more in line with the Chinese context. MedBERT is a model pre-trained specifically for Chinese medical texts, and has excellent performance in medical question matching and named entity recognition. Its ability to understand professional terms and perform relation extraction is comprehensive. The QA encoding mode helps the model capture the relationship between entities by explicitly constructing a "question-answer" form, and this encoding method generally has a better effect for models that are not specifically optimized for text classification tasks.

6.2 Encoding Strategy Impact

Entity marker strategies and QA prompting yield gains over basic concatenation; Entity_marked2 performs best (Table 6). Ordering (Basic2 vs. Basic1) modestly affects performance.

It can be seen that different encoding methods have a certain impact on model performance. The advantage of the entity tagging method lies in explicitly marking the entity range, reducing the model's uncertainty in

Table 6: Chinese-RoBERTa performance under encoding variants.

	Macro-P	Macro-R	Macro-F1	Micro-F1
Basic1	88.94%	88.60%	86.71%	89.01%
Basic2	89.33%	88.00%	88.09%	88.46%
QA	89.23%	86.80%	87.06%	89.25%
Entity1[1]	88.34%	87.30%	87.60%	87.81%
Entity2[2]	**90.72**%	**90.20**%	**90.24**%	**90.46**%

1.Entity1 is Entity_marked1
2.Entity2 is Entity_marked2

entity positioning. At the same time, Entity_marked2 further enhances the model's ability to capture the relationship between entities by using the unique linguistic structure of Chinese. The QA encoding method converts relation classification into a question-answer form through explicit task prompts, which significantly improves the model's understanding of the task. Basic2 outperforms Basic1, indicating that adjusting the order of entities has a certain impact on the model's ability to capture relationship information. Placing the tail entity in advance may be more conducive to the model's understanding of the relationship between entities.

6.3 Class Enhancement Effects

Weighted loss reduces inter-class F1 range (variance) while modestly improving global metrics; combining weighted loss with weak sample augmentation yields further gains (Table 7). As can be seen from Table 7 , the weighted loss adjusts the category weights to focus more on the difficult-to-classify minority categories, thereby significantly improving the model's precision and recall rates. The increase in Macro-F1 indicates that the weighted loss has a remarkable effect on the balanced optimization of the overall classification performance; the significant increase in Micro-F1 demonstrates that the model's accuracy in the overall classification task has been greatly improved. The reduction in the range of F1 values is due to the increase in the weight of the minority categories, reflecting the targeted role of the weighted loss in improving the recognition effect of minority categories.

	Regular Loss	Weighted Loss	Variation
Macro-P	88.02%	89.62%	+1.60%
Macro-R	88.1%	88.60%	+1.70%
Macro-F1	87.01%	88.71%	+1.70%
Micro-F1	86.90%	89.01%	+2.11%
F1 Range	26.29%	18.01%	-8.28%

Table 7: Chinese-RoBERTa-wwm-ext employs weighted Loss to enhance the scores.

	Weighted Loss	Weighted+Aug	Variation
Macro-P	89.62%	90.93%	+1.31%
Macro-R	88.6%	90.20%	+1.60%
Macro-F1	88.71%	90.24%	+1.53%
Micro-F1	89.01%	90.46%	+1.45%
F1 Range	18.01%	9.45%	-8.56%

Table 8: Chinese-RoBERTa-wwm-ext improves the score through sample augmentation.

As can be seen from Table 8, the weighted loss adjusts the category weights to focus more on the difficult-to-classify minority categories, thereby significantly improving the model's precision and recall rates. The increase in Macro-F1 indicates that the weighted loss has a remarkable effect on balancing the overall classification performance; the significant increase in Micro-F1 demonstrates that the model's accuracy in the overall classification task has been greatly improved. The wider range of F1 values reflects the expansion of the model's ability to handle minority categories, demonstrating the targeted role of the weighted loss in enhancing the recognition effect of minority classes.

7 Conclusion

This study proposes a relation extraction method based on a large pre-trained BERT model, specifically for the relation classification task in the Chinese medical field. Medical texts have problems such as a large number of professional terms, ambiguous boundaries of Chinese entities, and insufficient labeled data. The existing pre-trained models have insufficient adaptability. Therefore, we use Chinese-RoBERTa-wwm-ext as the base model and fine-tune it by combining entity tagging, question answering encoding, etc., which effectively improves the accuracy of relation extraction.

During the model evaluation, we found

that the 10 categories in the test set were evenly distributed, which resulted in the micro-Precision, micro-Recall, and micro-F1 values calculated under the original logic being exactly the same. After verification, it was because in the balanced data, the high accuracy of the model made the numerator and denominator of the index calculation form a fixed ratio, unable to distinguish the performance of different dimensions. To address this, we optimized the evaluation logic: first calculate the Precision and Recall for each category, and then use category weights for correction. This not only resolves the issue of numerical coincidence but also retains the original return structure, without affecting subsequent usage. Additionally, for the issue of class imbalance, we adopted the strategy of "weighted loss + weak sample enhancement", by adjusting the category weights and expanding the dataset, not only improved the recognition performance of difficult-to-classify categories, but also narrowed the differences in F1 values among various categories, achieving dual optimization of overall performance and class balance.

Experiments have shown that optimization in the Chinese domain, entity tagging, category enhancement, and improvement of evaluation logic are all effective. The proposed method outperforms models such as BioBERT and MedBERT. This study provides a more accurate and more generalized solution for extracting relationships in Chinese medical texts, which can support applications such as medical knowledge graph construction and intelligent diagnosis. The optimized evaluation logic also provides a reasonable reference for model evaluation in balanced data scenarios.

Acknowledgments

We acknowledge the valuable contributions of the medical professionals who provided domain expertise and validation support throughout this research. We also thank the open-source community for developing and maintaining the tools and frameworks that made this work possible.At the same time, we would like to express our sincere gratitude to the CISC7021 teaching team for the guidance and computing resources they provided for this research.

Use of Generative AI

We acknowledge the use of Kimi and DeepL to help us mainly for language polishing, grammar checking, code debugging, and optimization.

References

[1] Erik F. Tjong Kim Sang and Fien De Meulder. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147, 2003.

[2] Lev Ratinov and Dan Roth. Design challenges and misconceptions in named entity recognition. In Suzanne Stevenson and Xavier Carreras, editors, *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 147–155, Boulder, Colorado, June 2009. Association for Computational Linguistics.

[3] Dmitry Zelenko, Chinatsu Aone, and Anthony Richardella. Kernel methods for relation extraction. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pages 71–78. Association for Computational Linguistics, July 2002.

[4] Razvan Bunescu and Raymond Mooney. A shortest path dependency kernel for relation extraction. In Raymond Mooney, Chris Brew, Lee-Feng Chien, and Katrin Kirchhoff, editors, *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 724–731, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics.

[5] R. Florian, H. Hassan, A. Ittycheriah, H. Jing, N. Kambhatla, X. Luo, N. Nicolov, and S. Roukos. A statistical model for multilingual entity detection and tracking. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL 2004*, pages 1–8, Boston, Massachusetts, USA, May 2 - May 7 2004. Association for Computational Linguistics.

[6] GuoDong Zhou, Jian Su, Jie Zhang, and Min Zhang. Exploring various knowledge in relation extraction. In Kevin Knight, Hwee Tou Ng, and Kemal Oflazer, editors, *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 427–434, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.

[7] Yee Seng Chan and Dan Roth. Exploiting syntactico-semantic structures for relation extraction. In Dekang Lin, Yuji Matsumoto, and Rada Mihalcea, editors, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 551–560, Portland, Oregon, USA, June 2011. Association for Computational Linguistics.

[8] Qi Li and Heng Ji. Incremental joint extraction of entity mentions and relations. In Kristina Toutanova and Hua Wu, editors, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 402–412, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[9] Makoto Miwa and Mohit Bansal. End-to-end relation extraction using LSTMs on sequences and tree structures. In Katrin Erk and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1105–1116, Berlin, Germany, August 2016. Association for Computational Linguistics.

[10] Meishan Zhang, Yue Zhang, and Guohong Fu. End-to-end neural relation extraction with global optimization. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1730–1740, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[11] David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. Entity, relation, and event extraction with contextualized span representations. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5784–5789, Hong Kong, China, November 2019. Association for Computational Linguistics.

[12] Jue Wang and Wei Lu. Two are better than one: Joint entity and relation extraction with table-sequence encoders. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1706–1721, Online, November 2020. Association for Computational Linguistics.

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[14] Iz Beltagy, Kyle Lo, and Arman Cohan. SciBERT: A pretrained language model for scientific text. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3615–3620, Hong Kong, China, November 2019. Association for Computational Linguistics.

[15] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. ALBERT: A lite BERT for self-supervised learning of language representations. *CoRR*, abs/1909.11942, 2019.

[16] LIU Jun, RUAN Tong, and ZHANG Huanhuan. Prompt learning-based generative approach towards medical dialogue understanding. *Computer Science*, 51(5):258–266, 05 2024.

[17] REN Fanghui, GUO Xitong, PENG Xin, and YANG Jinfeng. A survey of spoken language understanding in medical field. *Chung Wen Hsin Hsi Hsueh Pao*, 38(01):24–35, 2024.

[18] Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. Semantic compositionality through recursive matrix-vector spaces. In Jun'ichi Tsujii, James Henderson, and Marius Paşca, editors, *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1201–1211, Jeju Island, Korea, July 2012. Association for Computational Linguistics.

[19] Kazuma Hashimoto, Makoto Miwa, Yoshimasa Tsuruoka, and Takashi Chikayama. Simple customization of recursive neural networks for semantic relation classification. In David Yarowsky, Timothy Baldwin, Anna Korhonen, Karen Livescu, and Steven Bethard, editors, *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1372–1376, Seattle, Washington, USA, October 2013. Association for Computational Linguistics.

[20] Rui Cai, Xiaodong Zhang, and Houfeng Wang. Bidirectional recurrent convolutional neural network for relation classification. In Katrin Erk and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 756–765, Berlin, Germany, August 2016. Association for Computational Linguistics.

[21] Daojian Zeng, Kang Liu, Siwei Lai, Guangyou Zhou, and Jun Zhao. Relation classification via convolutional deep neural network. In Junichi Tsujii and Jan Hajic, editors, *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 2335–2344, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics.

[22] Kun Xu, Yansong Feng, Songfang Huang, and Dongyan Zhao. Semantic relation classification via convolutional neural networks with simple negative sampling. In Lluís Màrquez, Chris Callison-Burch, and Jian Su, editors, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 536–540, Lisbon, Portugal, September 2015. Association for Computational Linguistics.

[23] Linlin Wang, Zhu Cao, Gerard de Melo, and Zhiyuan Liu. Relation classification via multi-level attention CNNs. In Katrin Erk and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1298–1307, Berlin, Germany, August 2016. Association for Computational Linguistics.

[24] Makoto Miwa and Yutaka Sasaki. Modeling joint entity and relation extraction with table representation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Process-*

ing (EMNLP), pages 1858–1869, Doha, Qatar, October 2014. Association for Computational Linguistics.

[25] Arzoo Katiyar and Claire Cardie. Going out on a limb: Joint extraction of entity mentions and relations without dependency trees. In Regina Barzilay and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 917–928, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[26] Changzhi Sun, Yeyun Gong, Yuanbin Wu, Ming Gong, Daxin Jiang, Man Lan, Shiliang Sun, and Nan Duan. Joint type inference on entities and relations via graph convolutional networks. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1361–1370, Florence, Italy, July 2019. Association for Computational Linguistics.

[27] Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. Entity-relation extraction as multi-turn question answering. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1340–1350, Florence, Italy, July 2019. Association for Computational Linguistics.

[28] Giannis Bekoulis, Johannes Deleu, Thomas Demeester, and Chris Develder. Adversarial training for multi-context joint entity and relation extraction. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2830–2836, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

[29] MU Tianyang, LI Bao'an, YOU Xindong, and Lü Xueqiang. An end-to-end dialogue system based on medical knowledge graph. *Journal of Beijing Information Science*, 36(06):14–18, 2021.

[30] Lu Xueqiang, Zhang Jian, MU Tianyang, LI Bao'an, and YOU Xindong. A dialogue system in the medical field embedded with knowledge semantics. *Computer Engineering and Design*, 44(12):3794–3799, 2023.

[31] Zexuan Zhong and Danqi Chen. A frustratingly easy approach for entity and relation extraction. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou, editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 50–61, Online, June 2021. Association for Computational Linguistics.

[32] Jinhyuk Lee, WonJin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36:1234 – 1240, 2019.

[33] Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, Ziqing Yang, Shijin Wang, and Guoping Hu. Pre-training with whole word masking for chinese bert. *ArXiv*, abs/1906.08101, 2019.

[34] Charangan Vasantharajan, Kyaw Zin Tun, Ho Thi-Nga, Sparsh Jain, Tong Rong, and Chng Eng Siong. Medbert: A pre-trained language model for biomedical named entity recognition. In *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1482–1488, 2022.