

doi: 10.7690/bgzdh.2025.06.005

## 基于注意力机制的中医实体关系抽取模型

李旻哲<sup>1</sup>, 刘 华<sup>2</sup>, 殷继彬<sup>1</sup>

(1. 昆明理工大学信息与自动化学院, 昆明 650500; 2. 昆明理工大学发展与规划处, 昆明 650500)

**摘要:** 针对中医实体关系复杂和多样导致实体关系抽取不佳的问题, 提出一种基于注意力机制与多模型融合的关系抽取模型(r-BERT-BiLSTM-attention-textCNN, RBBAT)。该模型由关系抽取预训练模型(r-BERT)、双向长短期记忆神经网络(BiLSTM)、注意力层(Attention)和TextCNN 4部分组成; 实验选取近年来各个医案平台上公开的消化科相关医案, 针对症状-病名、症状-证候、舌象-证候、脉象-证候、证候-治法 5个实体关系进行关系抽取。实验结果表明: 该模型与常用的关系抽取模型相比较, 在症状-病名、症状-证候、舌象-证候、证候-治法 4种实体关系上的抽取能力达到最优。

**关键词:** 关系抽取; 中医案例; 预训练模型; 注意力机制; TextCNN

**中图分类号:** TP183; TP391.4 **文献标志码:** A

## Model of TCM Entity Relation Extraction Based on Attention Mechanism

Li Minzhe<sup>1</sup>, Liu Hua<sup>2</sup>, Yin Jibin<sup>1</sup>

(1. College of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 2. Division of Development and Planning, Kunming University of Science and Technology, Kunming 650500, China)

**Abstract:** In order to solve the problem of poor entity relation extraction caused by the complexity and diversity of entity relations in traditional Chinese medicine (TCM), a relation extraction model (r-BERT-BiLSTM-attention-textCNN, RBBAT) based on attention mechanism and multi-model fusion is proposed. The model is composed of relation extraction pre-training model (r-BERT), bidirectional long/short-term memory neural network (BiLSTM), Attention layer and TextCNN; In the experiment, the relevant medical records of the department of gastroenterology published on various medical record platforms in recent years were selected, and five entity relationships were extracted, including symptom-disease name, symptom-syndrome, tongue-syndrome, pulse-syndrome, and syndrome-treatment. The experimental results show that compared with several commonly used relation extraction models, the proposed fusion model has the best extraction ability in the four entity relations of symptom-disease name, symptom-syndrome, tongue picture-syndrome, and syndrome-treatment method.

**Keywords:** relation extraction; TCM case; pre-training model; attention mechanism; TextCNN

### 0 引言

中医信息抽取技术是从中医医案资料中去抽取有用的、重要的相关信息, 分析不同医案中存在的用药规律, 从而对中医的临床、科研、教学和新药研发提供思路, 让中医知识更好地服务于医学相关领域的应用, 有益于推动中医的继承和发展, 使中医药与现代技术接轨<sup>[1-3]</sup>。关系抽取是信息抽取中十分重要的一个环节, 本质上是从海量文本数据中获取有关联的结构化文本并以三元组的形式出现, 即<实体, 关系, 实体>。早期使用的关系抽取有根据规则的关系抽取方法以及传统机器学习的关系抽取方法。

通过语言学规则和模板的编写, 可以有效地挖

掘出相关的实体之间的关系, 从而构建出一个完整的知识库<sup>[4-6]</sup>。这些规则和模板是由专家们精心设计的, 可以为特定领域的实体之间提供有效的模板支持, 从而更好地满足特定领域的需求。由于需要专家们进行大量的人工操作, 而且移植性不佳, 无法有效地应用于不同领域。

传统机器学习的关系抽取技术在3个主要研究领域得到了广泛应用: 监督学习<sup>[7-8]</sup>、半监督学习<sup>[9]</sup>、无监督学习<sup>[10-11]</sup>和远程监督<sup>[12]</sup>。由于数据集的标注质量、数量以及方法等因素的存在, 这种技术的训练效果往往会受到不可忽视的影响。随着技术的进步, 传统机器学习的关系抽取方法存在特征提取误差, 这种误差可能会向后传播, 从而影响算法的性能, 而深度学习<sup>[13]</sup>可以解决这种问题。

收稿日期: 2024-08-10; 修回日期: 2024-09-25

第一作者: 李旻哲(1997—), 男, 云南人, 硕士。

### 1 相关工作

针对实体关系抽取时遇到较远距离的命名实体出现在多个关系三元组内导致关系抽取效果不佳的问题，将中医实体关系抽取转化为序列标记任务，以注意力机制为切入点，构建基于注意力机制的关系抽取模型。RBBAT 是一种基于注意力机制的多模型融合实体关系抽取模型，它由 3 部分组成，如图 1 所示。1) 使用预训练的语言模型 r-BERT 来提取和强化中医语料的特征信息。2) 使用 BiLSTM-Attention 来训练输入的信息，以获取更多的特征信息；通过 BiLSTM，可以获取句子的上下文信息，并将它们拼接在一起，然后通过注意力层处理来计算出每个词的权重；最后将这些权重矩阵相乘，得到句子的向量表示，从而获取每个词的特征信息。3) 以 TextCNN 为特征筛选器，将得到的文本特征信息作为输入向量，使用卷积池化对重要程度高的语义特征进行选择并预测实体对关系类型。

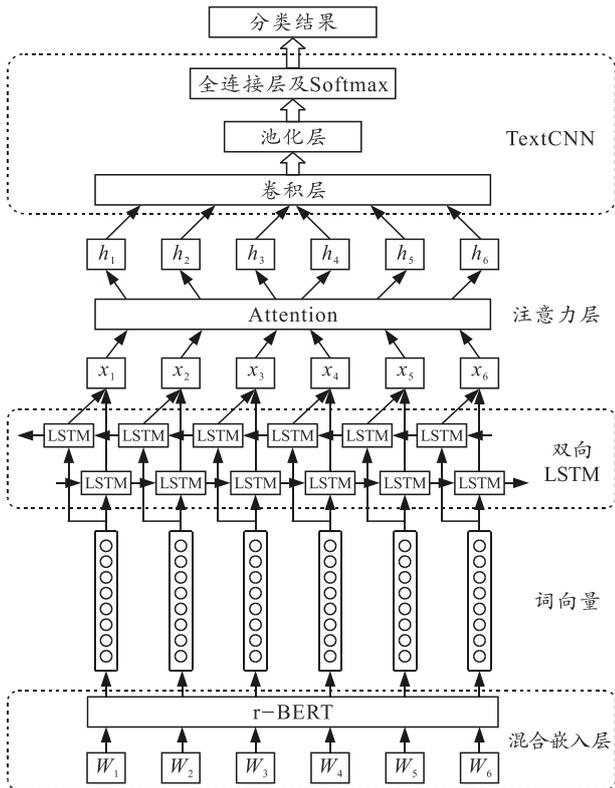


图 1 RBBAT 模型架构

#### 1.1 r-BERT 混合嵌入层

混合嵌入层是将词嵌入向量、类型向量及位置向量混合获取后进行拼接。r-BERT 模型是 BERT 用于关系抽取的一个衍生模型<sup>[14-15]</sup>，不仅使用了 BERT 句首的特殊符号[CLS]，而且在每个实体的两侧也插入特殊符号作为实体标识，BERT 处理后的

词向量常被用于文本分类等下游任务。

模型架构如图 2 所示，由 3 部分组成。

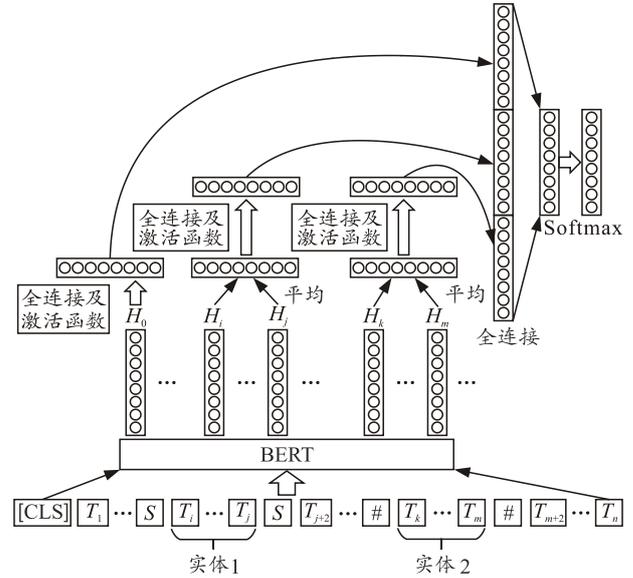


图 2 r-BERT 模型架构

- 1) BERT，用于提取文本的向量表示。
- 2) 全连接层，对从 BERT 获得的词向量进行处理。对于实体向量，由于实体可能包含不止一个词，所以将实体包含的词向量进行加和平均来得到实体的向量表示。以模型图为例，实体在未处理文本中的表示是 $[T_i, T_j]$ ，经 BERT 后的词向量为 $[H_i, H_j]$ ，将其加和平均后，再过一个 dropout, tanh 激活以及全连接层得到实体表示，如式(1)和(2)所示。

$$H'_1 = W_1 \left[ \tanh \left( \frac{\sum_{i=i}^j H_i}{(j-i+1)} \right) \right] + b_1; \quad (1)$$

$$H'_2 = W_2 \left[ \tanh \left( \frac{\sum_{i=k}^m H_i}{(m-k+1)} \right) \right] + b_2. \quad (2)$$

对于模型句首特殊标记[CLS]，模型经过类似的处理，得到了[CLS]最终的向量表示：

$$H'_0 = W_0 (\tanh(H_0)) + b_0. \quad (3)$$

- 3) 分类层，BERT 最后将[CLS]、实体 1、实体 2 的向量表示拼接，得到一个大小为 (batch\_size, 3\*hidden\_state) 的矩阵，将这个矩阵过一个全连接层后接一个 Softmax，训练时使用交叉熵。损失公式为：

$$h'' = W_3 [\text{concat}(H'_0, H'_1, H'_2)] + b_3; \quad (4)$$

$$p = \text{Softmax}(h''). \quad (5)$$

文中模型在使用 r-BERT 作为模型的混合嵌入层时，采用 1)和 2)，将获取的[CLS]、实体 1、实体 2 的向量拼接得到矩阵作为词向量输入到特征提取层 BiLSTM-Attention 中。

### 1.2 特征提取层 BiLSTM-Attention

特征嵌入层接收混合嵌入层中经过拼接后的词向量，利用 BiLSTM 进行分析，以求更好地理解上下文的联系；再通过注意力机制来获取所有来自 BiLSTM 隐藏层中具有代表性的特征。注意力矩阵  $A$  中的权重  $a_{t,j}$  表示当前第  $t$  个隐藏状态  $h_t$  与第  $j$  个隐藏状态之间的相似度， $a_{t,j}$  计算方法为：

$$a_{t,j} = \exp(\text{score}(x_t, x_j)) / \sum_{k=1}^L \exp(\text{score}(x_t, x_k)); \quad (6)$$

$$\text{score}(x_t, x_j) = W_a \tanh(W_b x_t + W_c x_j + b_a). \quad (7)$$

所有 BiLSTM 输出  $x_t$  的全局向量  $g_t$ ，计算公式为：

$$g_t = \sum_{j=1}^L a_{t,j} x_j. \quad (8)$$

通过将全局向量  $g_t$  与 BiLSTM 的输出  $x_j$  相互映射输入到  $\tanh$  函数，从而获得 Attention 层的输出  $h_t$ ，计算公式为：

$$h_t = \tanh(W_g [g_t, x_t]). \quad (9)$$

### 1.3 决策层 TextCNN

将 TextCNN 作为特征筛选器使用，TextCNN 结构如图 3 所示。TextCNN 主要结构包括输入层、卷积层、池化层和输出层<sup>[16]</sup>。

输入层负责处理数据，使输入数据符合 TextCNN 模型的输入格式。卷积层则负责提取特征，卷积层中卷积核的宽度是特征的数量，并且只能以上到下滑动的方式进行卷积计算。池化层使用上述卷积核滑动窗口的宽度来模块化网络，对数据进行降维并取出最高值或均值。全连接层利用卷积神经网络和池化神经网络的特征信息进行整合，输出网络的最终结果。

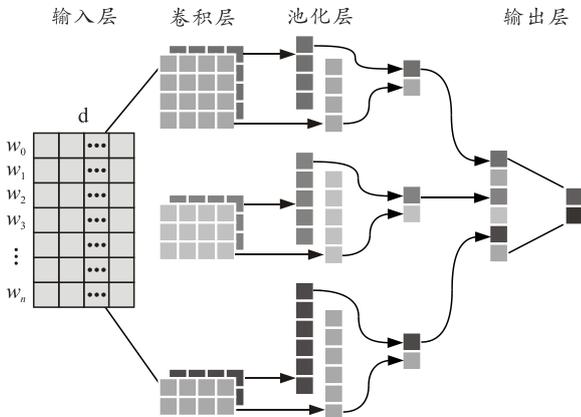


图 3 TextCNN 各个模块

将上一部分传输过来的  $h_t$  进行拼接，组成一个

矩阵  $X$ ，将输入向量  $X_n$  组合成  $[n, k, 1]$  3 维输入矩阵，其中  $n$  表示  $X_n$  的长度，1 表示通道数。

$$X_{1:n} = X_1 \oplus X_2 \oplus \dots \oplus X_n. \quad (10)$$

通过卷积层可以得到一系列 3 维矩阵，每 1 维矩阵的大小为  $[n-h+1, n-k+1, m]$ ，式中  $h$  为卷积核的高度； $k$  为卷积核的宽度；以及每 1 维矩阵的特征值； $m$  为卷积核的个数。

由于存在  $m$  个卷积核，因此卷积层的输出  $C_{out}$  可表示为一个形如  $[n-h+1, n-k+1, m]$  的 3 维矩阵。

$$C_i = f(W \cdot X_{i:i+h-1} + b); \quad (11)$$

$$C_j = [c_1, c_2, \dots, c_{n-h+1}]; \quad (12)$$

$$C_{out} = [C_1, C_2, \dots, C_m]. \quad (13)$$

式中： $f$  为一个  $\text{relu}$  函数； $W$  为卷积核； $X_{i:i+h-1}$  为输入矩阵第  $i$  至  $i+h-1$  行的向量； $b$  为权重常数； $C_i$  为卷积核第  $i$  次卷积所得的值； $C_j$  为第  $j$  个卷积核自上而下滑动，经过  $n-h+1$  次卷积后，最终形成了大小为  $[n-h+1, n-k+1]$  的矩阵。

通过池化处理，可以得到大小为  $[1, m]$  的 2 维输出矩阵，这些输出矩阵如式(14)和(15)所示，其中  $C_j$  表示第  $j$  个卷积核从上到下滑动，经过  $n-h+1$  次卷积，最终形成大小为  $[n-h+1, n-k+1]$  的 2 维矩阵。

$$\hat{C}_k = \max(C_j); \quad (14)$$

$$\hat{C}_{out} = [\hat{C}_1, \hat{C}_2, \dots, \hat{C}_m]. \quad (15)$$

$$y = (\hat{C}_{out} * r) \cdot w + b. \quad (16)$$

式中： $r$  为一个与  $\hat{C}_{out}$  大小相同的权重系数矩阵，矩阵的每个位置都相乘，而不改变矩阵的大小； $w$  为一个大小为  $[m, 1]$  的权重系数矩阵； $b$  为一个权重常数； $y$  为预测值。

## 2 实验

### 2.1 实验数据

实验数据来源于各大知名中医案例平台，均为名老中医治疗疾病的临床病案共 800 例中医确诊案例，实体及实体关系数量如表 1 所示。

表 1 实体及实体关系的数量

命名实体	实体数量	实体关系	关系数量
病名	2 852	症状-病名	11 627
证候	4 903	症状-证候	9 159
症状	13 653	舌象-证候	4 814
治法	5 321	脉象-证候	2 284
舌象	2 614	证候-治法	6 877
脉象	1 336		

由于训练集规模小，可能会造成模型泛化能力

不足的情况；因此，通过引入噪声，增加 2 个相互抵消的干扰因素使得参数得到正则化，从而提高模型的泛化能力。通过对抗训练(adversarial training)可大大提高神经网络的鲁棒性。该过程中，将一些微小的干扰因素混入到模型中，以便让模型能够更好地适应外部环境的变化，从而提高模型的鲁棒性。通过对抗训练，不仅能提高系统的鲁棒性，而且能增强泛化能力。对抗训练如式(17)所示：

$$\min_{\theta} E_{(x,y) \sim D} [\max_{\Delta x \in \Omega} L(x + \Delta x, y, \theta)] \quad (17)$$

式中： $D$  为训练集； $x$  为输入； $y$  为标签； $\theta$  为模型参数； $L(x, y, \theta)$  为单个样本的 loss； $\Delta x$  为对抗扰动； $\Omega$  为扰动空间。实验采用快速梯度法(fast gradient method, FGM)。在模型中，向 Embedding 层在梯度方向添加扰动，可使 Word Embedding 的质量更好，不仅提高了置信度，而且还避免了模型过拟合，使模型性能得到提升。

### 2.2 模型参数

RBBAT 参数设置如表 2 所示。BERT 中共有 12 层 Transform，使用 hugging face4 中的 bert-base-Chinese 相关模块进行初始化，BERT 模型的参数不变。

表 2 RBBAT 模型参数

名称	值
句子的最大长度	128
实体关系分类数	5
激活函数	relu
Learning rate	1e-5
Batch_size	32
Dropout	0.2
Optimizer	Adam
Max_epoch	20
bert 层输出维度	768
TextCNN 的卷积输出	2
TextCNN 的卷积核大小	[2, 3, 4]

### 2.3 评价指标

通过对比  $P$ 、 $R$ 、 $F_1$  3 个参数，可以更好地评估该模型的可靠性和可行性。各评价指标计算公式为：

$$P = \frac{m}{n} \times 100\% \quad (18)$$

$$R = \frac{m}{N} \times 100\% \quad (19)$$

$$F_1 = \frac{2PR}{P+R} \times 100\% \quad (20)$$

式中： $m$  为正确抽取实体关系的数量； $N$  为数据集中实体关系的总数量； $n$  为抽取出的实体关系数量。

### 2.4 实验对比设置

为检验基于注意力机制的融合模型 RBBAT 的有效性，选取以下多种不同的神经网络模型进行对比。所有模型都是基于 NER 后的结果进行关系抽取。

TextCNN：使用 Word2vec 作为词向量输入，TextCNN 进行关系抽取。

BERT-Softmax：使用 BERT 作为词向量输入，Softmax 进行关系抽取。

BERT-TextCNN：使用 BERT 去获取词向量，TextCNN 进行关系抽取。

r-BERT：使用 r-BERT 模型解决关系抽取。

图 4—8 和表 3—7 为 5 个模型分别在[症状-病名，症状-证候，舌象-证候，脉象-证候，证候-治法] 5 种关系分类上的精确率( $P$ )、召回率( $R$ )、 $F_1$  值的对比情况。

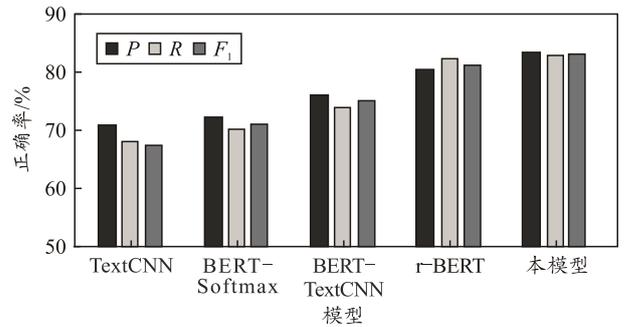


图 4 症状-病名实验结果

表 3 症状-病名关系分类 %

模型	P	R	F1
TextCNN	70.8	64.7	67.6
BERT-Softmax	72.4	70.3	71.3
BERT-TextCNN	76.3	74.1	75.2
r-BERT	80.8	82.4	81.6
本模型	83.6	82.9	83.2

如图 4 和表 3 所示，在基于注意力机制的 r-BERT 与 TextCNN 融合模型中，可以得出本文中模型  $P$ 、 $R$ 、 $F_1$  指标最高，对比其他模型具有一定的优势。

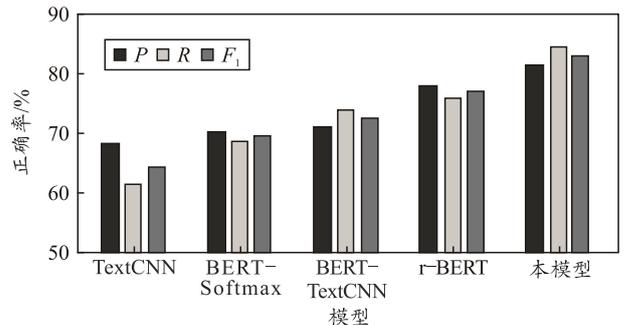


图 5 症状-证候实验结果

表 4 症状-证候关系分类 %

模型	P	R	F <sub>1</sub>
TextCNN	68.3	61.1	64.5
BERT-Softmax	70.4	68.7	69.5
BERT-TextCNN	71.5	74.2	72.8
r-BERT	78.4	76.2	77.3
本模型	81.6	84.6	83.1

如图 5 和表 4 所示，在基于注意力机制的 r-BERT 与 TextCNN 融合模型中，本文中模型的 P、R、F<sub>1</sub> 指标同样最高，对比其他模型具有优势。

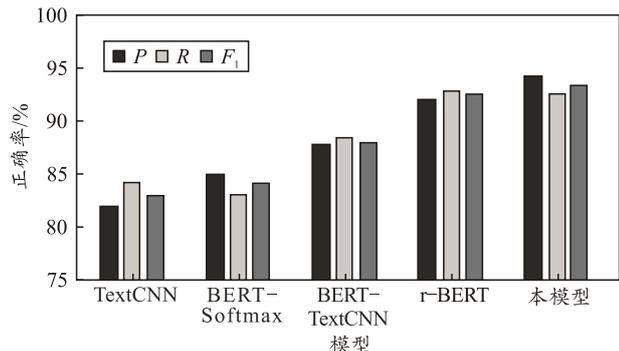


图 6 舌象-证候实验结果

表 5 舌象-证候关系分类 %

模型	P	R	F <sub>1</sub>
TextCNN	82.1	84.3	83.2
BERT-Softmax	85.0	83.3	84.1
BERT-TextCNN	87.8	88.5	88.1
r-BERT	92.1	93.0	92.5
本模型	94.3	92.6	93.4

如图 6 和表 5 所示，在基于注意力机制的 r-BERT 与 TextCNN 融合模型中，可以看出，本文中模型的指标衡量模型优秀程度 F<sub>1</sub> 最高，相比其他模型存在优势，但优势不大。

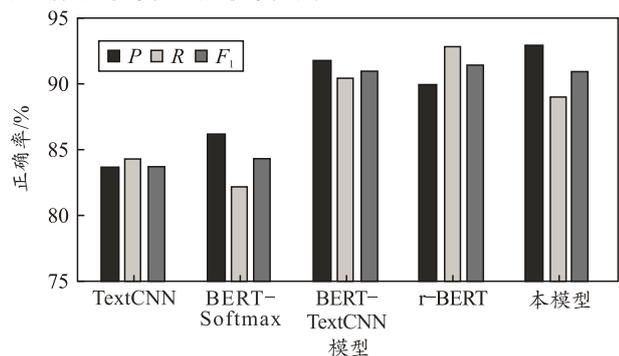


图 7 脉象-证候实验结果

表 6 脉象-证候关系分类 %

模型	P	R	F <sub>1</sub>
TextCNN	83.7	84.3	84.0
BERT-Softmax	86.3	82.3	84.3
BERT-TextCNN	91.8	90.5	91.1
r-BERT	90.1	93.0	91.5
本模型	93.0	89.0	91.0

如图 7 和表 6 所示，在基于注意力机制的

r-BERT 与 TextCNN 融合模型中，可以看出，BERT-Text 和 r-BERT 模型的 F<sub>1</sub> 指标均超过本文中模型，但 3 个模型的 F<sub>1</sub> 相差不多。

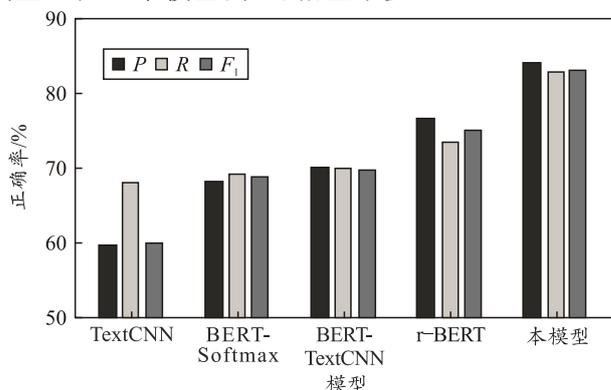


图 8 证候-治法实验结果

表 7 证候-治法关系分类 %

模型	P	R	F <sub>1</sub>
TextCNN	59.7	61.1	60.4
BERT-Softmax	68.2	69.2	68.7
BERT-TextCNN	70.1	69.9	70.0
r-BERT	76.7	73.5	75.1
本模型	83.9	82.6	83.2

如图 8 和表 7 所示，在基于注意力机制的 r-BERT 与 TextCNN 融合模型中，可以看出，笔者所提模型的指标最高，相比其他模型优势比较大。

综上，在不使用 BERT 系列模型进行数据处理时或仅使用 BERT 模型抽取信息后进行分类，该类模型相比其他模型 F<sub>1</sub> 值普遍较低；对舌象-证候及脉象-证候这 2 类实体关系分类时，由于其较为绝对和单一的关系分类导致上述 5 个模型都能取得较好的结果，所提模型在这种单一划分任务中优势并不明显；在分类复杂的情况下，如症状-病名、症状-证候、证候-治法，所提模型对比其他 4 个模型就拥有较大优势。

## 2.5 消融实验

通过消融实验，可以得到提出的方法对系统的影响。在这项实验中，删除了 BiLSTM 和注意力机制模块或者 TextCNN 模块，其实验结果如表 8 所示。

表 8 消融实验对比 %

研究内容	P	R	F <sub>1</sub>
r-BERT-TextCNN (去除 BiLSTM 和注意力机制模块)	78.4	77.8	78.1
BERT-BiLSTM-Att-Softmax (去除 TextCNN 模块)	82.3	82.1	82.2
BERT-BiLSTM-Att-TextCNN (提出的模型)	85.3	85.0	85.2

从表 8 可以看出：去除 BiLSTM 和注意力机制

导致模型的  $F_1$  降低了 7.1%，说明 BiLSTM-Attention 不仅能有效提取片段的深层特征提取，而且能加强片段之间的交互及依赖关系；通过去除 TextCNN 模块发现模型的  $F_1$  下降 3%，这表明 TextCNN 可以准确地处理带有深层特征的信息。消融实验结果证明，模型中添加辅助模块可以显著提高关系抽取模型的性能。

### 3 结束语

笔者提出一种基于注意力机制的实体关系抽取模型 RBBAT：首先，利用 r-BERT 模型获取带有实体标签信息的特征向量；然后，通过 BiLSTM 以及注意力层 Attention 在获取上下文的同时提取深层特征向量的信息；最后，将获取的信息输入到 TextCNN 中将关系抽取问题转化为多标签分类问题，得到更优秀的关系抽取效果。使用处理好的数据集进行实验，对比验证使用 r-BERT 词向量、BERT 词向量和 Word2vec 词向量的多种不同算法模型的性能表现，证明基于注意力机制的实体关系模型 RBBAT 的结果优化效果，经测试数据验证取得了 85.3% 的识别准确率，85.0% 的识别召回率及 85.2% 的  $F_1$  值。通过消融实验，展现模型构建后的优化效果，证明每个模块设置的意义。

### 参考文献：

- [1] 扶莉. 基于中医医案的信息抽取方法研究[J]. 信息技术与信息化, 2015(4): 139, 144.
- [2] 王宁. 基于电子病历的中医临床决策支持研究[D]. 合肥: 中国科学技术大学, 2021.
- [3] 顾铮, 顾平. 信息抽取技术在中医研究中的应用[J]. 医学信息, 2007, 20(1): 27-30.
- [4] 朱玲, 朱彦, 杨峰. 基于中医疾病相关语义关系的正则表达式及知识抽取研究[J]. 世界科学技术-中医药现代化, 2016, 18(8): 1241-1250.
- [5] 高凤帅, 杨化斌. 结合词语规则和 SVM 模型的军事命名实体关系抽取方法[J]. 信息通信, 2017(11): 162-163.
- [6] 庄成龙, 钱龙华, 周国栋. 基于树核函数的实体语义关系抽取方法研究[J]. 中文信息学报, 2009, 23(1): 3-8, 34.
- [7] KAMBHATLA N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations[C]. Proc of ACL on Interactive Poster and Demonstration Sessions. Stroudsburg, PA: Association for Computational Linguistics, 2004: 22-26.
- [8] ZELENKO D, AONE C, RICHARDELLA A. Kernel methods for relation extraction[J]. Journal of Machine Learning Research, 2003, 3(3): 1083-1106.
- [9] 肖丹. 基于半监督学习的中文电子病历实体关系抽取研究[D]. 绵阳: 西南科技大学, 2021.
- [10] 张金壬, 章韵, 王宇. 基于 sentence2vec 与半监督算法的中文问答提问模式抽取[J]. 计算机应用研究, 2019, 36(7): 2035-2039.
- [11] HASSAN H, HASSAN A, EMAM O. Unsupervised information extraction approach using graph mutual reinforcement[C]//Proc of Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2006: 501-508.
- [12] 马超. 基于 Web 信息使用改进的无监督关系抽取方法构建交通本体[J]. 计算机系统应用, 2015, 24(12): 273-276.
- [13] 赵国清, 何佳洲, 乔慧, 等. 基于远程监督关系抽取的开源情报处理[J]. 指挥控制与仿真, 2021, 43(1): 69-73.
- [14] 隗昊, 周爱, 张益嘉, 等. 深度学习生物医学实体关系抽取研究综述[J]. 计算机工程与应用, 2021, 57(21): 14-23.
- [15] LI Z G, CHEN H, QI R H, et al. DocR-BERT: Document-Level R-BERT for Chemical-Induced Disease Relation Extraction via Gaussian Probability Distribution[J]. Journal on Biomedical and Health Informatics (J-BHI), 2022, 26(3): 1341-1352.
- [16] 耿唯佳, 宋玉蓉, 周伟伟. 融合 TextCNN 与 TextRNN 模型的谣言识别方法[J]. 微电子学与计算机, 2022, 39(1): 31-38.