

# 融合依存句法和实体信息的临床时间关系抽取

黄汉琴<sup>1,2,3</sup>, 顾进广<sup>1,2,3</sup>, 符海东<sup>1,2,3</sup>

- (1. 武汉科技大学 计算机科学与技术学院, 湖北 武汉 430065;  
2. 国家新闻出版署富媒体数字出版内容组织与知识服务重点实验室, 北京 100038;  
3. 武汉科技大学 大数据科学与工程研究院, 湖北 武汉 430065)

**摘要:**在临床文本中,时间关系对于研究患者的病情和治疗方案至关重要。而目前的时间关系抽取基于简单时间比较,仅判断4种时间关系。考虑中文临床文本中还存在大量的复杂时间和关系,现有时间关系抽取任务不能全部表达临床事件的时间关系,参考CTO时间本体将抽取任务扩展为复杂时间关系抽取。同时针对中文临床文本语义的复杂性,提出了融合依存句法和实体信息的模型学习中文句子的整体信息和实体信息。该模型针对句内时间关系和句间时间关系设计依存特征矩阵引导BERT的编码器聚合全局信息和局部信息,然后导出句子表征向量,在此基础上使用内积和哈达玛积提取丰富的实体信息,最终将句子信息和实体信息导入分类器判断时间关系。与基线模型和其他深度学习模型相比,证明了该模型的有效性。

**关键词:**时间关系抽取;自注意力机制;依存句法;局部信息;实体信息

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2024)01-0128-08

doi:10.3969/j.issn.1673-629X.2024.01.019

## Extraction of Clinical Temporal Relation Fusing Dependency Syntax and Entity Information

HUANG Han-qin<sup>1,2,3</sup>, GU Jin-guang<sup>1,2,3</sup>, FU Hai-dong<sup>1,2,3</sup>

- (1. School of Computer Science, Wuhan University of Science and Technology, Wuhan 430065, China;  
2. Key Laboratory of Rich-media Knowledge Organization and Service of Digital Publishing Content National Press and Publication Administration of the People's Republic of China, Beijing 100038, China;  
3. Institute of Big Data Science and Engineering, Wuhan University of Science and Technology, Wuhan 430065, China)

**Abstract:**In clinical texts, temporal relation is crucial to the study of patient's conditions and treatment options. The current temporal relation extraction is based on the simple temporal comparison, and only four temporal relations are judged. Considering that there are still a large number of complex times and relations in Chinese clinical texts, the existing temporal relation extraction task cannot fully express the temporal relation of clinical events, referring to the CTO temporal ontology, the extraction task is expanded to complex time relationship extraction. At the same time, aiming at the semantic complexity of Chinese clinical texts, a model integrating dependency syntax and entity information is proposed to learn the overall information and entity information of Chinese sentences. The model scrambles to design dependency feature matrices for intra-sentence temporal relation and inter-sentence temporal relation to guide BERT's encoder to aggregate global and local information, derive sentence representation vectors on which rich entity information is extracted using the inner product and Hadamard product. Finally, the sentence information and entity information is imported into the classifier to determine temporal relation. Compared with baseline model and other deep learning model, the effectiveness of the proposed model is demonstrated.

**Key words:**temporal relation extraction; self-attention mechanism; dependent syntax; local information; entity information

收稿日期:2023-02-28

修回日期:2023-06-29

**基金项目:**国家自然科学基金委员会,联合基金项目(U1836118);国家新闻出版署富媒体数字出版内容组织与知识服务重点实验室开放基金项目(ZD2022-10/05);科技创新2030“新一代人工智能”重大项目(2020AAA0108500)

**作者简介:**黄汉琴(1999-),女,硕士研究生,研究方向为自然语言处理;顾进广(1974-),男,教授,研究方向为语义网与知识图谱;通信作者:符海东(1971-),男,教授,研究方向为机器学习、自然语言处理。

## 0 引言

在临床研究方面,从电子健康记录抽取时间关系用于多种应用,如临床问题解答、决策支持系统、临床时间线构建等。这些应用需要清晰的时间顺序做支撑。时间顺序在临床文本中有着重要的作用,很多的临床事件都需要在时间顺序下才彰显意义。

时间关系(TLINK)表示时间表达式和临床事件之间在时间线上的关系。临床事件是与临床相关的任何事件,包括患者的临床概念和医疗事件。时间表达式是指在临床文本中提供有关临床事件的发生时间、时间段、发生频率。现有的TLINK分为4种时间关系,考虑中文临床文本中存在大量的复杂时间和关系,现有的时间关系不能全部表达这些事件之间的关系,该文将时间关系抽取任务扩展。

TLINK抽取任务常用的方法有基于规则,机器,深度学习的方法。基于规则的方法依靠大量的人工抽取句子的语义特征制定规则,需要耗费大量时间且移植性差。近年来,基于深度学习的神经网络模型取得显著的成果,此类模型能自主挖掘出句子的语言特征。

文中数据集采用中文临床抑郁症病例文本,此类文本语义复杂,存在不均衡性,理解文本的语义需要学习全局上下文特征也要关注重要内容的局部特征。Lin等人使用动态或固定窗口来执行局部区域关注<sup>[1-2]</sup>;Bugliarello等人利用语法限制注意力获得更好的关注<sup>[3-4]</sup>。

两种方法对重要的局部区域进行更多关注得到局部特征再和全局信息融合。在此基础上,借助BERT模型的自注意力机制能从一个句子的众多信息中找到局部重要信息加权处理,该文引入依存句法分析临床文本的局部重要区域,抽取局部重要特征信息和整体句子信息共同学习。

同时考虑时间关系抽取任务学习时间实体和临床事件实体的信息,现有的构造方法大多数使用三段式池化拼接方法来展开<sup>[5]</sup>。借鉴此方法,该文额外对BERT层的输入嵌入实体类型,并在输出表征对实体信息进行特征交互来获得更全面的实体信息。该文的贡献如下:

(1)针对中文临床文本提出复杂时间关系抽取任务,扩展现有抽取任务的时间关系类型。

(2)为追求全局信息和局部信息的兼具问题,针对临床文本设计依存特征抽取算法,提出将BERT的自注意力层和依存句法分析结合的方法,该模型能在现有的BERT架构上快速嵌入实现。

(3)为获得更全面的实体信息,从BERT表征抽取额外的两个实体信息,结合内积和哈达玛积实现特性交互,使用四段式池化拼接得到关系向量。

## 1 相关工作

当前时间关系抽取的相关工作与两个共享临床时间关系抽取任务有关: Informatics for Integrating Biology and Bedside(i2b2) challenge<sup>[6]</sup>和2015年开始的TempEval challenge<sup>[7]</sup>。自深度学习流行后,应用BI-LSTM, CNN, BERT等深度学习模型抽取时间关系成为主流,以下将介绍这些方法的研究现状。

基于BI-LSTM方法: Juline等人<sup>[8]</sup>使用BI-LSTM模型融合特征工程抽取包含关系,证明增加经典的特征收集更多信息是有益的方法,他们在<sup>[9]</sup>收集句子信息额外对每一个实体类型构建特定分类器。Chen等人<sup>[10]</sup>在自我训练框架内构建语义异构嵌入的循环神经网络,利用基本特征丰富句子信息。这些方法聚集在进一步收集局部信息,丰富模型语义。

基于CNN的方法: Dmitriy等人<sup>[11]</sup>将XML标签编码关系参数的位置引入CNN模型去提取时间关系,用这种标记成为神经网络的输入,后续的CNN方法输入大多借鉴于此。Chen等人<sup>[12]</sup>描述了一种单个伪标记表示时间表达式的方式修改了token的标签用于CNN的输入。此类神经网络方法能自动进行特征提取,但是池化层会丢失大量的价值信息,对长距离句子来说,输出的表征难以概括全局信息。

即使以BI-LSTM和CNN为主的方法在时间关系抽取中表现较好,但是和bert为主的预训练语言模型存在一定差距。基于BERT方法, Lin等人<sup>[13]</sup>在BERT模型上引入全局嵌入来帮助进行长距离关系推理,并通过多任务学习来提高模型性能和泛化性,从而增强模型抽取的效果。Chen等人<sup>[2]</sup>在BERT模型上开发基于连续标记的固定的句子边界不可知窗口的处理机制,展示了BERT出色的长距离推理能力。近2年的临床关系抽取综述<sup>[14-15]</sup>都总结了微调BERT及其变体等语言模型在从临床文本中提取关系方面表现最好。为提升下游任务的学习能力,通过具体任务设计简单堆叠模块实现微调。基于此研究思路,该文使用BERT现有架构,通过对自注意力机制层与依存句法结合的方法获取句子的全局信息和局部信息。同时,使用关系抽取中常用的实体特征交互方法获得更多的实体信息。

## 2 方法与模型

### 2.1 时间关系抽取任务定义

TLINK时间关系是指电子病历中时间点与临床事件或临床事件与临床事件之间的时序关系。临床文本中存在大量的复杂时间维度,现有的时间关系仅表达这些事件之间的简单时间关系,复杂时间本体定义可以挖掘临床文本中更多的语义信息。该文依据

CTO 临床时间本体<sup>[16]</sup>将现有的时间关系抽取中常见的 4 种时间关系类型扩展为 10 种时间关系抽取。

病历中的时间点是指带有时间信息的短语,参考现有病历常见的时间表达,根据 CTO 临床时间本体定义了各种类型的时间语义,将所有时间总分为 3 种类型:时间点,时间段,时间集合。病历中的临床事件是指病历中有关病情的相关记录和医疗记录,参考现有病历常见的临床事件,该文将临床事件分为疾病,症状,检查,药物,手术、医疗事件,治疗。病历的时间与临床事件分类实例表如表 1 所示。

参考 CTO 临床时间本体,设定每一对时间关系的两个实体为 A 和 B。临床事件实体依照时间线记录事件发生的起始时间,结束时间,频率等信息划分为在某一时刻完成这一件事,在某一段时间完成这一件事,在时间集合完成这一件事。使得时间实体和临床事件实体都能在时间维度上进行比较。时间关系为之前,之后,相等,遇见,共同开始,共同结束,相交,包含,全包含,部分包含。所有时间关系的定义如图 1 所示。根据每张图可以获取每一个时间关系的满足条件,例如时间点实体不可能和别的实体产生遇见,相交,全包含,部分包含这些时间关系。

表 1 时间与临床事件分类实例表

实体类型	分类	含义	实例
时间	instant	时间点	2010 年 3 月,昨天
	interval	时间段	2012 至 2016 年,3 个月
	collection	时间集合	每天,在 1975 年,1977 年
临床事件	disease	疾病	焦虑抑郁状态
	symptom	症状	经常哭泣,想自杀
	examination	检查	脑部 CT
	medication	药物	帕罗西汀 20ml
	operation	手术	外科手术
	medical event	医疗事件	在我院会诊中心诊断
	cure	治疗	MECT6 次

该文将时间抽取任务定义为:(1)复杂句内时间关系抽取:指的是一个句子里面任意的时间点与临床事件的时间关系或临床事件与临床事件的时间关系都能在 10 种时间关系中找到答案;(2)复杂句间时间关系抽取:一个句子的医疗事件与相邻句子的医疗事件的时间关系能在 10 种时间关系中找到答案。后续实验与分析描述根据这两个任务分别标注数据集和训练模型进行实验。

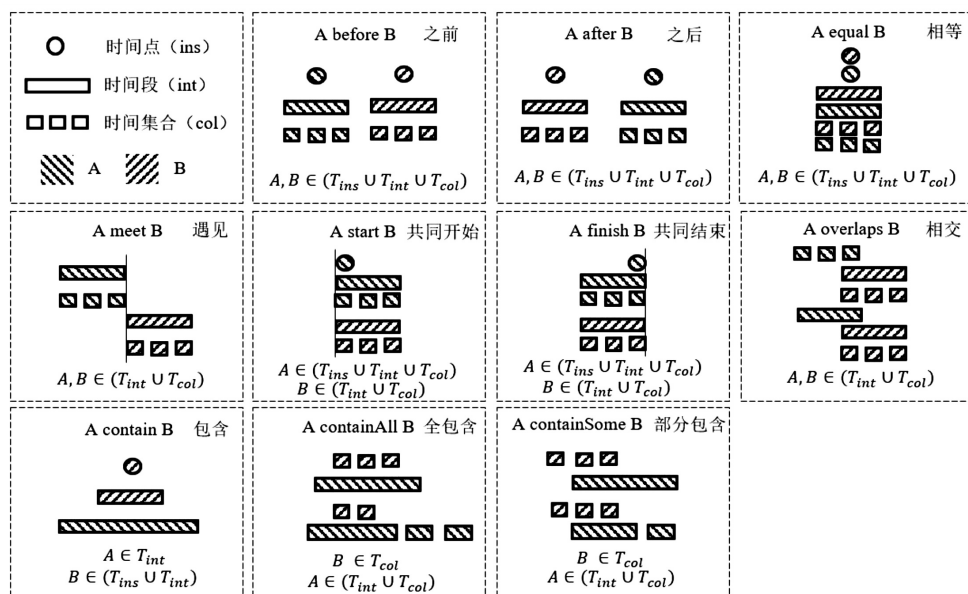


图 1 时间关系定义

## 2.2 DS-EI-BERT 模型

该文提出使用 BERT 预训练模型结合依存句法和实体信息对中文抑郁病临床文本进行时间关系抽取的模型(Enriching bert with dependency syntax and entity information, DS-EI-BERT)。DS-EI-BERT 模型结构如图 2 所示。将输入句子嵌入实体类型,转化成句子输入序列,具体描述在 2.2.1 节;然后将句子传入依存权重矩阵处理器导出每个句子的依存矩阵,具体描述在 2.2.2 节;此依存矩阵会引导 BERT 的 12 层

Transformer 编码器层来训练每一层的全局信息和局部信息,具体描述在 2.2.3 节;然后实体-句子信息层提取出实体信息做特征交互,最后将句子信息和实体信息拼接得到最终输出,使用 softmax 函数得到最终的时间关系,具体描述在 2.2.4 节。

### 2.2.1 句子输入序列

该文的输入采用预训练模型 BERT 的输入方式。BERT 模型的输入可以是单个句子,也可以包含一个句子对(句子 A 和句子 B)。而这两种方式刚好可以

对应句内关系和句间关系的输入。BERT 预训练参数给出的 vocab 中有一些特殊作用的标识符, [CLS] 标识符放在第一个句子的首位, 作为整个句子的语义表示。[SEP] 标识符用于分开两个句子, 每一个句子结束后会紧跟一个 [SEP] 标识符。此外 BERT 输入可以在字符字典设定有意义的标识符, 将时间类型和事件类型设置成特殊的标识符。时间分类中, instant 的标

识符为 [INS], interval 的标识符为 [INT], collection 的标识符为 [COL]。临床事件分类中, disease 的标识符为 [DIS], symptom 的标识符为 [SYM], examination 的标识符为 [EXAM], medication 的标识符为 [MED], operation 的标识符为 [OPE], medical event 的标识符为 [MEEV], cure 的标识符为 [CURE]。

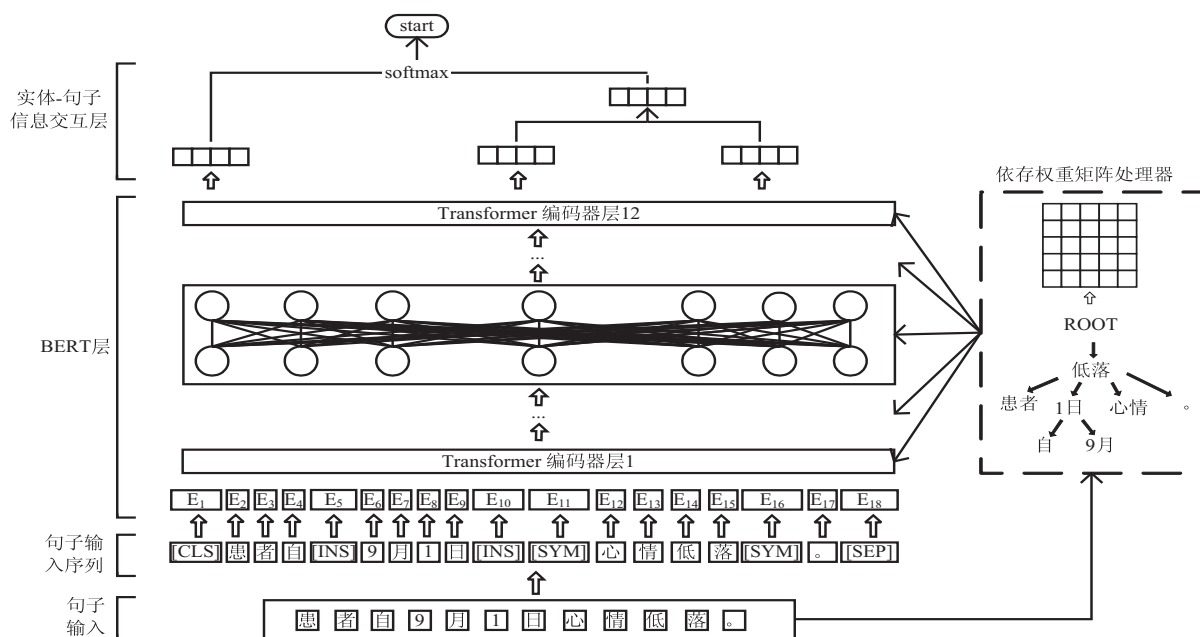


图2 DS-EI-BERT 模型结构

举例说明句内的输入模式: “[CLS] 患者自 [INS] [INS] 9 月 1 日 [SYM] 心情低落 [SYM]。 [SEP]”。句间的输入模式: “[CLS] 8 月服 [MED] 奥氮平 [MED]。 [SEP] 一月后 [MEEV] 自行停药 [MEEV]。 [SEP]”通过这种方式能让模型更好学习两个实体的关系。

### 2.2.2 依存权重矩阵处理器

为了更好地挖掘句子包含的整体价值, 时间关系抽取需要更关注重要的词语, 根据现有研究表明, BERT 预训练模型中的注意力在训练时会关注不重要的虚词, 对重要的局部区域进行更多关注, 保证句子中重要的语义被充分学习利用。该文利用依存句法分析获取句子重要的依存特征, 然后转化成依存权重矩阵, 此矩阵传入 BERT 每一层的自注意力机制中限制注意力来帮助模型获得更好的关注。

对输入的句子进行依存分析, 能将一个句子分析成一棵依存句法树, 这棵树的最顶层节点是整个句子的核心词, 其他词汇直接或者间接与它产生联系, 整个树描述出词汇与词汇之间的相互依存作用。通过依存句法分析得到的依存句法树如图3所示。

参考孙健等人在中文时间关系抽取研究中对依存特征的使用<sup>[17]</sup>, 针对不同的关系类别制定句子上重要

的依存路径, 如表2所示。这些路径上的词汇与关系抽取的两个实体存在直接或者间接的支配与从属的关系, 属于语句中语义丰富的词汇。

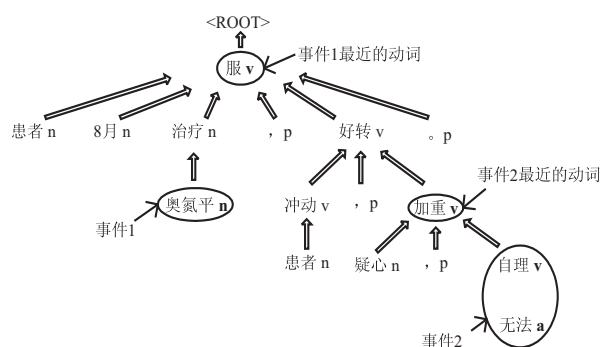


图3 依存句法树

假设一个句子 S, 句子中的两个目标实体分别为 A、B, 根据不同的依存特征找到所用路径上面的词汇, 把这些词汇的下标做成一个集合 Key, 这个 Key 代表的就是这个句子依存特征词的下标集合。图3展示一个句子的依存句法树: A 表示“奥氮平”, B 表示“无法自理”, 根据上表句内事件-事件时间关系的依存特征, 从“奥氮平”追溯依存关系得到路径 (“奥氮平”→“治疗”→“服”→“好转”→“加重”→“自理”→“无法”), 计算所有路径最后得到 Key 集合 [“奥氮平”,



“治疗”,“服”,“无法”,“自理”,“加重”,“好转”]。以上假设符合句内关系抽取的情况,句间关系抽取涉及两个句子,则分别得到两个句子的集合 Key 合并再进行如下计算。

表 2 句内/句间依存特征路径

关系类别	依存特征
句内	(1) 时间到事件的路径
	(2) 时间与事件最近的动词之间的路径
	(3) 时间最近的动词到事件的路径
	(4) 时间最近的动词到事件最近的路径
事件-事件	(1) 事件 1 到事件 2 的路径
	(2) 事件 1(2) 到事件 2(1) 最近的动词的路径
	(3) 事件 1(2) 到事件 1(2) 最近的动词的路径
	(4) 事件 1 最近动词与到事件 2 最近动词的路径
句间	(1) 两个句子的最短依存路径
	(2) 两个事件与句子 root 节点的路径
	(3) 两个事件最近动词和句子 root 节点的路径

得到这些重要的依存特征词汇集合 Key 后,将 Key 中的词汇和 BERT 输入序列对应,这些序列下标从 1 到  $s$ ,如果 BERT 下标对应的字在 Key 的词汇中,则将下标记录在 keyPoint 集合,最后构建一个  $s \times s$  维的二维数组  $D_s$ ,这个矩阵的行下标为  $x$ ,列下标为  $y$ ,每一行和每一列都包含一个字对其他字是否存在依存关系需要关注的含义,计算如式 1 所示:

$$D_s[x, y] = \begin{cases} 1 & x \in \text{keyPoint} \& y \in \text{keyPoint} \text{ or } x = y \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

### 2.2.3 Transformer 编码器层

不同于田园等人<sup>[18]</sup>将几种模型拼接一起得到语句更全面的信息,该文利用 BERT-BASE 模型<sup>[4]</sup>层层堆叠的 12 层 Transformer 编码器层来不断地训练全局有用信息。通过上一节得到的依存矩阵不断强化局部注意力与 BERT 本身的全局注意力结合,每一层编码器的具体结构如图 4 所示。

在每一层编码器层中,会计算所有 token 的 3 个向量(查询向量  $Q$ ,值向量  $V$ ,关键字向量  $K$ )。使用  $Q$  和  $K$  做点积,能得到每一个 token 需要对其他 token 关注的权重系数  $\text{Score}^g$ ,这也是每一层自注意力层的全局 attention 分数,  $\text{Score}^g$  计算公式如式 2 所示。

$$\text{Score}^g = Q \cdot K^T \quad (2)$$

该文还需要获取每一层具有依存句法信息的局部 attention 分数  $\text{Score}^l$ ,使用依存句法矩阵对  $\text{Score}^g$  做点积获取包含句法信息的注意力权重,可以得到任意一个 token 对其余所有存在依存关系的 token 的权重和矩阵,  $\text{Score}^l$  计算公式如式 3 所示。

$$\text{Score}^l = D_s \cdot \text{Score}^g \quad (3)$$

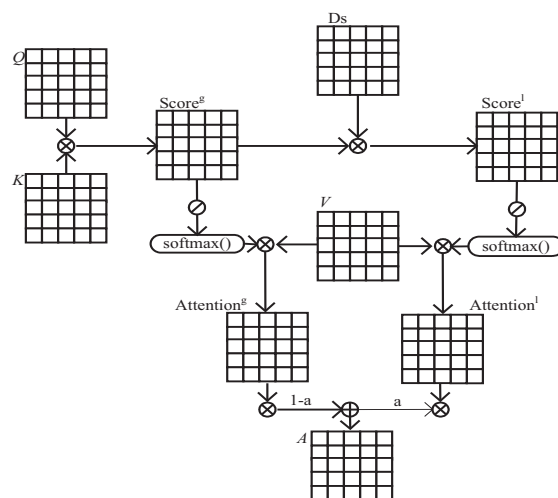


图 4 Transformer 编码器层结构

对这两个 attention 分数做归一化处理,然后对  $V$  加权求和分别得到这一层编码器注意力表征  $\text{Attention}^g$  和依存句法局部注意力表征  $\text{Attention}^l$ 。计算公式如式 4 和式 5 所示。这里的  $d$  是指嵌入维度。

$$\text{Attention}^g = \text{softmax}\left(\frac{\text{Score}^g}{\sqrt{d}}\right) \cdot V \quad (4)$$

$$\text{Attention}^l = \text{softmax}\left(\frac{\text{Score}^l}{\sqrt{d}}\right) \cdot V \quad (5)$$

而最后的输出表征是在两个表征代表的全局信息和局部信息找到一个平衡。针对每一个 token,设置阈值  $\alpha_k, k \in [0, s]$  用于组合全局和局部注意力。这个值是可变的,更大的阈值意味着更多的注意力集中在句法信息上,对于与依存句法矩阵没有关联的 token,这个  $\alpha_k$  为 0;对于有关联的 token,这个  $\alpha_k$  为 0.5。计算公式如式 6 所示。

$$\alpha_k = \begin{cases} 0.5 & k \in \text{keyPoint} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

每一个 token 得到的注意力输出如下,然后汇总可得此层的注意力表征输出  $A_k$ ,最后传给下一层继续此循环,最后得到整个 BERT 层的输出表征  $H$ 。与原始架构相比,文中方法只是在每一层编码器中加入依存矩阵  $D_s$  和  $\alpha$  进行运算,使得此方法较容易在现有的 BERT 代码中嵌入实现,比起很多论文复杂的模型叠加来训练更全面的表征无疑是精简的。计算公式如式 7 所示。

$$A_k = (1 - \alpha_k) \cdot \text{Attention}^g_k + \alpha_k \cdot \text{Attention}^l_k \quad (7)$$

### 2.2.4 实体-句子信息交互层

时间关系抽取已获得丰富的句子信息,接下来需要获取句子中时间实体和临床事件实体的信息,常用方法<sup>[19]</sup>将句子信息和两个实体信息池化后拼接进行关系分类。然而这样的拼接并不能得到实体的交互信息,为得到这些实体的交互信息,现有方法会使用全连

接,外积,内积,哈达玛积。该文在 FiBiNET 的 Bilinear-Interaction 基础上<sup>[20]</sup>,提出利用现有的依存矩阵引导权重结合内积和哈达玛积来获得实体交互信息,内积和哈达玛积交互会考虑不同特征对于预测目标的重要性程度,给不同的特征根据重要性程度进行加权。

整个 BERT 层最后的输出词向量为  $H$  :

$$H = [h_0, h_1, \dots, h_i, \dots, h_j, \dots, h_l, \dots, h_n, \dots, h_s] \quad (8)$$

由于每一条输入句子前加上了 [CLS] 标识符表征整个句子,所以取第一个向量  $h_0$  代表整个句子信息。对于句子中第一个实体  $e_1$ , 获取  $e_1$  在句子的下标为  $i$  到  $j$ , 那么  $h_i$  到  $h_j$  是实体  $e_1$  的最终隐藏向量。对于第二个实体  $e_2$ , 获取  $e_2$  在句子的下标为  $l$  到  $n$ , 那么  $h_l$  到  $h_n$  是实体  $e_2$  的最终隐藏向量, 取这些向量的平均向量  $h_{e1}, h_{e2}$  来获得目标实体的向量表示。然后通过激活函数和全连接层分别得到最终的句子向量和实体向量表示。  $W_0 \in Rd \times d$ ,  $W_1 \in Rd \times d$ ,  $W_2 \in Rd \times d$ , 其中  $d$  是 BERT 的隐藏向量大小。然后将实体 1 的最终表示与依存代表的权重矩阵进行内积, 再与实体 2 进行哈达玛积就能获得两个实体重要维度上的信息。计算公式如式 9 ~ 式 14 所示。

$$h_{e1} = \frac{1}{j-i+1} \sum_{i=i}^j H \quad (9)$$

$$h_{e2} = \frac{1}{n-l+1} \sum_{l=l}^n H \quad (10)$$

$$H_0 = W_0 [\tanh(h_0)] + b_0 \quad (11)$$

$$H_{e1} = W_1 [\tanh(h_{e1})] + b_1 \quad (12)$$

$$H_{e2} = W_2 [\tanh(h_{e2})] + b_2 \quad (13)$$

$$H_{bili} = H_{e1} \cdot DS \odot H_{e2} \quad (14)$$

然后将  $H_0, H_{e1}, H_{e2}, H_{bili}$  连接起来获得包含重要信息的句子信息和丰富实体信息的向量  $H'$ , 这里设置  $W_3 \in RL \times 3d$  ( $L$  为时间关系总数 10), 以上  $b_0, b_1, b_2, b_3$  为偏置。添加一个全连接层和 softmax 层, 该文使用交叉熵损失函数。计算公式如式 15 和式 16 所示。

$$H' = W_4 [\text{concat}(H_0, H_{e1}, H_{e2}, H_{bili})] + b_4 \quad (15)$$

$$p = \text{softmax}(H') \quad (16)$$

### 3 实验与分析

#### 3.1 数据集

鉴于目前没有中文时间关系数据集, 该文在私有数据集和公有数据集上进行中文临床文本时间关系研究。私有数据集源于合作医院抑郁症临床数据 (Chinese Medical Record of Depression, CMRD), 应医院要求不予公开。主要内容包括患者的个人信息、入院日期、出院日期、现病史、个人史等信息。抑郁症文本包含病人所有的病史, 拥有较长且完整的时间线描

述病人的病情和住院情况, 适用于时间关系抽取任务。公开数据集源于 2020CCKS 的骨科临床文本, 本身有标注好的实体信息。

抽取每一个句子的时间实体和事件实体或者事件实体和其他事件实体作为关系候选对, 然后标注它们的句内和句间时间关系。最后共标注了 100 份电子病历数据。标注出的句内和句间数据集如表 3 所示, 总样本以 8 : 2 分为训练集和测试集。

表 3 句内和句间数据集数量

时间关系	句内数据集		句间数据集	
	CMRD	CCKS	CMRD	CCKS
before	652	197	1 827	242
after	171	42	1 113	187
equal	1 447	469	296	72
meet	124	12	241	54
start	231	26	63	13
finsh	96	39	101	24
overlaps	1 375	254	295	63
contain	518	107	55	7
containAll	895	134	83	15
containSome	68	4	39	3

#### 3.2 实验设置

所有实验运行在搭载了 Tesla P-100 PCIE 显卡和 Intel(R) Xeon(R) CPU E5-2600 v4 @ 2.00 GHz 型号 CPU 的机器上, 使用 PyTorch1.4.0 作为深度学习框架, 依赖 Python 3.7 环境下完成实验。

文中方法实现基于 huggingface 官网的 bert-base-chinese 模型, 依存句法分析使用 spacy 的 zh\_core\_web\_sm-3.3.0 版本。

句内时间关系抽取实验和句间时间关系抽取实验是分开实现的。实验基线模型为原始 BERT<sup>[4]</sup> 和丰富实体信息的 R-BERT<sup>[5]</sup>。为评测 DS-EI-BERT 模型在时间关系抽取任务的有效性, 将与 BERT 和 R-BERT 在相同数据集上进行比较。为验证文中模型的竞争力, 在文中数据集上对比了现有的时间关系抽取最好的模型 BERT-TS<sup>[2]</sup>。除此之外, 对比了相关工作中其他深度学习模型。Juline 等人<sup>[8]</sup>提出的 BI-LSTM 方法同样是在已知实体情况下识别实体之间存在的关系, 使用 TensorFlow 实现模型。Dmitriy 等人提出的 CNN 方法<sup>[11]</sup>使用时间表达式的类型代替时间实体作为 CNN 的输入, 文中亦在 BERT 的输入层融入实体类型信息。

#### 3.3 评价指标

实验采用精确率 (Precision)、召回率 (Recall) 和 F1 值 (F-score) 3 个指标来评价中文电子病历中时间关系抽取的性能。计算公式如式 17 ~ 式 19 所示。

$$P = \frac{TP}{TP + FP} \times 100\% \quad (17)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (18)$$

$$F1 = \frac{2RP}{R + P} \times 100\% \quad (19)$$

### 3.4 结果与分析

使用不同的参数训练模型后,以下实验中 BERT 模型的超参数设置:最大序列长度为 128,学习率为  $1e-5$ ,batch 为 16,dropout 为 0.1。CNN 模型的超参数设置:最大序列长度为 128,学习率为  $1e-4$ ,batch 为 50,dropout 为 0.25。BI-LSTM 模型的超参数设置:最大序列长度为 128,学习率为  $1e-3$ ,batch 为 64,dropout 为 0.4。

#### 3.4.1 句内时间关系抽取

实验结果如表 4 所示。从表 4 可以看出,BI-LSTM 和 CNN 取得的  $P$ 、 $R$  和  $F1$  值都比预训练 BERT 及衍生方法的低。CNN 能捕捉局部语言信息,但是对于中文这种具有长距离的句子,不能很好地捕捉上下文语义,实验效果并不好, $P$  和  $R$  相比之下存在较大差距;而 BI-LSTM 通过递归不断捕捉句子上下文语义, $P$  值相较 CNN 提高 4.4 百分点,但是和 BERT 的自注意力机制捕捉信息能力存在差距,此外 BI-LSTM 的训练时间过长,和其他模型比较起来需要更多的计算资源。这里也能看出对于时间关系抽取这一项任务,模型学习句子整体信息极为重要。

表 4 句内与句间时间关系对比实验 %

模型	句 内			句 间		
	$P$	$R$	$F1$	$P$	$R$	$F1$
BERT <sup>[4]</sup>	78.68	76.95	77.80	71.07	69.54	70.29
R-BERT <sup>[5]</sup>	81.81	78.38	80.05	73.28	72.92	73.09
BERT-TS <sup>[2]</sup>	83.57	80.10	81.79	80.14	76.58	78.31
CNN <sup>[11]</sup>	69.87	67.18	68.49	62.99	59.90	61.40
BI-LSTM <sup>[8]</sup>	74.27	71.25	72.72	64.95	61.48	63.16
DS-EI-BERT	87.98	85.91	86.93	82.84	78.78	80.75

预训练语言 BERT 模型系列表现优异,BERT 原始架构也要比 CNN,BI-LSTM 表现优异,其  $F1$  值达到 77.8%。R-BERT 对比 BERT 模型学习句子中的实体信息,进一步提升时间关系抽取的效果,在  $F1$  提升 2.25 百分点。这两个基线模型实验证明实体信息能帮助时间关系抽取,笔者认为抽取实体信息能改善时间关系实体跨度大的问题。

文中模型 DS-EI-BERT 在表 4 中在  $P$ 、 $R$  和  $F1$  上优于其他模型。原因在于自注意力机制层使用依存句法进行全局信息和局部信息提取,可以帮助 BERT 更好地学习句子语义。而使用两个实体进行特征交互

进一步丰富了实体信息,让模型进一步提升使得  $F1$  值达到 86.93%。

最后和现有时间关系抽取最好的模型 BERT-TS 对比, $P$ 、 $R$  和  $F1$  值都有不错的提升,分别提升 4.41 百分点,5.81 百分点和 5.14 百分点,验证此方法在现有的时间关系抽取模型中都有不错的竞争力。

#### 3.4.2 句间时间关系抽取

实验结果如表 4 所示,句间实验的整体效果较句内实验在  $P$ 、 $R$ 、 $F1$  值都有退步,分别降低 5.14 百分点,7.13 百分点,6.18 百分点。原因在于句间的语句比较句内的语句,无论从输入句子长度还是整体句子复杂度都有提升,对模型的语义理解有更高的要求。BERT 系列实验效果比其他深度学习模型还是有不错的提升,但是与句内时间关系相比, $P$ 、 $R$  和  $F1$  值都有退步。

文中模型 DS-EI-BERT 在表 4 中在  $P$ 、 $R$  和  $F1$  上依旧优于其他模型,分别达到 82.84%、78.78%、80.75%。但是 BERT-TS 也有非常好的效果, $F1$  值达到 78.81%。可以看出,时间关系抽取不仅从模型提升上能够提高抽取效果,从关系实体候选策略入手也能提升不错的效果。

#### 3.4.3 消融实验

最后为了验证依存句法和实体信息的有效性,进行了消融实验,一共设计 4 组实验,设计只保留依存句法矩阵引导自注意力机制的 DS-BERT 和只保留实体信息层的 EI-BERT 进行消融实验,结果如表 5 所示。

表 5 句内/句间时间关系消融实验 %

任务	模型	$P$	$R$	$F1$
句内时 间关系	BERT <sup>[4]</sup>	78.68	76.95	77.80
	DS-BERT	84.48	81.32	82.86
	EI-BERT	82.69	79.75	81.19
	DS-EI-BERT	87.98	85.91	86.93
句间时 间关系	BERT <sup>[4]</sup>	71.07	69.54	70.29
	DS-BERT	77.69	76.94	77.31
	EI-BERT	75.73	71.52	73.56
	DS-EI-BERT	82.84	78.78	80.75

句内和句间消融实验模型 EI-BERT 与 BERT 对比  $F1$  值分别提高 3.39 百分点和 3.27 百分点。进一步验证了提取出更多的实体信息能提升抽取效果。句内和句间消融实验模型 DS-BERT 与 BERT 对比  $F1$  值分别提高 5.06 百分点和 7.01 百分点,也显示依存句法和实体信息的有效性,且对于句间这类长距离的抽取任务,依存句法能帮助模型学习长句子的重要部分获得更好的结果。DS-EI-BERT 集中句子信息和实体信息,最大程度提高模型效果,使得  $P$ 、 $R$ 、 $F1$  值达到最高。

## 4 结束语

临床时间关系抽取作为临床信息抽取中重要的分支,如何从一个复杂冗余的临床病例文本抽取精确的时间关系有着重要的作用。该文发现中文数据集的复杂时间描述难以使用现有的时间关系任务展示时间线,于是扩充了复杂时间关系抽取任务。并且针对复杂时间关系句子语义的复杂性,从句子信息和实体信息提升模型实验效果。利用 BERT 模型的自注意力机制融合输入句子的依存句法特征,帮助句子充分利用依存句法信息,引导 BERT 关注更有用的信息;利用输出表征的实体信息进行特征交互来丰富实体信息。在句间时间关系抽取中只考虑相邻句子的时间关系抽取,在后续研究中,考虑使用滑动窗口来扩大关系候选对的面积,从而研究一段文本之中包含的时间关系。

### 参考文献:

- [1] 潘理虎,陈亭亭,闫慧敏,等. 基于滑动窗口注意力网络的关系分类模型[J]. 计算机技术与发展,2022,32(6):21-27.
- [2] LIN C, MILLER T, DLIGACH D, et al. A BERT-based universal model for both within- and cross-sentence clinical temporal relation extraction [C]//Proceedings of the 2nd clinical natural language processing workshop. Minneapolis: ACL,2019:65-71.
- [3] LI Z, ZHOU Q, LI C, et al. Improving BERT with syntax-aware local attention [C]//Findings of the association for computational linguistics. [s. l.]: ACL,2021:645-653.
- [4] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding [J]. arXiv:1810.04805v2,2018.
- [5] WU S, HE Y. Enriching pre-trained language model with entity information for relation classification[J]. arXiv:1905:08284,2019.
- [6] SUN W, RUMSHISKY A, UZUNER O. Evaluating temporal relations in clinical text: 2012 i2b2 challenge[J]. Journal of the American Medical Informatics Association,2013,20(5):806-813.
- [7] BETHARD S, DERCZYNSKI L, SAVOVA G, et al. Semeval-2015 task 6: clinical tempeval [C]//Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015). Denver: ACL,2015:806-814.
- [8] TOURILLE J, FERRET O, NEVEOL A, et al. Neural architecture for temporal relation extraction: a Bi-LSTM approach for detecting narrative containers [C]//Proceedings of the 55th annual meeting of the association for computational linguistics. Vancouver: ACL,2017:224-230.
- [9] TOURILLE J, FERRET O, TANNIER X, et al. LIMSI-COT at SemEval-2017 task 12: neural architecture for temporal information extraction from clinical narratives [C]//Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017). Vancouver: ACL,2017:597-602.
- [10] LIN C, MILLER T, DLIGACH D, et al. Self-training improves recurrent neural networks performance for temporal relation extraction [C]//Proceedings of the ninth international workshop on health text mining and information analysis. Brussels: ACL,2018:165-176.
- [11] DLIGACH D, MILLER T, LIN C, et al. Neural temporal relation extraction [C]//Proceedings of the 15th conference of the European chapter of the association for computational linguistics; volume 2. Valencia: ACL,2017:746-751.
- [12] LIN C, MILLER T, DLIGACH D, et al. Representations of time expressions for temporal relation extraction with convolutional neural networks [C]//BioNLP 2017. Vancouver: ACL,2017:322-327.
- [13] LIN C, MILLER T, DLIGACH D, et al. A BERT-based one-pass multi-task model for clinical temporal relation extraction [C]//Proceedings of the BioNLP 2020 workshop. [s. l.]: ACL,2020:70-75.
- [14] GUMIEL Y B, SILVA E OLIVEIRA L E, CLAVEAU V, et al. Temporal relation extraction in clinical texts: a systematic review [J]. ACM Computing Surveys,2021,54(7):1-36.
- [15] GALVAN-SOSA D, MATSUDA K, OKAZAKI N, et al. Empirical exploration of the challenges in temporal relation extraction from clinical text [J]. Journal of Natural Language Processing,2020,27(2):383-409.
- [16] HU D, WANG M, GAO F, et al. Knowledge representation and reasoning for complex time expression in clinical text [J]. Data Intelligence,2022,4(3):573-598.
- [17] 孙健,高大启,阮彤,等. 中文电子病历中的时间关系识别 [J]. 计算机应用,2018,38(3):626-632.
- [18] 田园,马文. 基于 Attention-BiLSTM 的电网设备故障文本分类 [J]. 计算机应用,2020,40(S2):24-29.
- [19] 厉晓妍,张德平. 基于位置编码与实体交互信息的关系抽取方法 [J]. 计算机系统应用,2022,31(6):238-244.
- [20] LIAN D, XIE X, ZHENG V W, et al. CEPR: a collaborative exploration and periodically returning model for location prediction [J]. ACM Transactions on Intelligent Systems and Technology,2015,6(1):8.1-8.27.