

基于 BERT 和图注意力网络的 医疗文本因果关系抽取算法

刘位龙¹,王玘²,赵超³,王宁²,张旭¹,苏萍²,宋书典²,张娜²,迟蔚蔚²

(1.山东财经大学管理科学与工程学院,山东 济南 250014;2.山东健康医疗大数据管理中心,山东 济南 250002;
3.潍坊市卫生健康委员会,山东 潍坊 261071)

摘要: **目的** 提出一种能够有效抽取因果关系的算法,以提高医疗领域文本处理的准确性。**方法** 提出基于 Transformer 的双向编码器(bidirectional encoder representations from Transformers, BERT)和因果图注意力网络(causal graph attention networks, CGAT)的 BERT-CGAT 算法。首先构建因果关系图,利用医疗文本对 BERT 模型进行微调,以获得优化的实体嵌入表示;随后通过知识融合通道整合文本编码信息与因果结构,输入至图注意力网络;采用多头注意力机制并行处理不同子空间信息,增强复杂语义关系捕捉能力;最后通过双通道解码层实现实体及因果关系的同步抽取。**结果** 在自建的糖尿病因果实体数据集上的实验表明,模型在准确率(99.74%)与召回率(81.04%)上较传统 BiLSTM-CRF 基线提升 0.65% 和 16.73%,F1 分数达 80.83%。**结论** BERT-CGAT 算法通过结合 BERT 的语义特征提取能力和图神经网络的关系建模优势,有效提升了医疗文本因果关系抽取的准确性,验证了该方法的有效性。

关键词: 医疗文本;BERT 模型;图注意力网络;因果关系抽取

中图分类号: TP391.1 **文献标志码:** A

Causality extraction algorithm of medical text based on BERT and graph attention network

LIU Weilong¹, WANG Ding², ZHAO Chao³, WANG Ning², ZHANG Xu¹, SU Ping²,
SONG Shudian², ZHANG Na², CHI Weiwei²

(1. School of Management Science and Engineering, Shandong University of Finance and Economics,
Jinan 250014, Shandong, China; 2. National Administration of Health Data, Jinan 250002, Shandong, China;
3. Weifang Municipal Health Commission, Weifang 261071, Shandong, China)

Abstract: **Objective** To propose an algorithm capable of effectively extracting causal relationships to improve the accuracy of medical text processing. **Methods** The study proposed a bidirectional encoder representations from Transformers (BERT)-causal graph attention networks (CGAT) algorithm based on BERT and graph attention network. First, a causal relationship graph was constructed, and the BERT model was fine-tuned on medical texts to obtain optimized entity embeddings. Subsequently, a knowledge fusion channel integrated textual encoding information with causal structures, which were then fed into the graph attention network. A multi-head attention mechanism was employed to process information from different subspaces in parallel, enhancing the ability to capture complex semantic relationships. Finally, a dual-channel decoding layer was adopted to simultaneously extract entities and their causal relationships. **Results** Experiments on the self-built diabetes causal entity dataset showed that the model employing the BERT-CGAT algorithm had an improvement of 0.65% and 16.73% in precision rate (99.74%) and recall rate (81.04%) compared with the traditional BiLSTM-CRF baseline, and the F1 value were 80.83%. **Conclusion** The BERT-CGAT algorithm

effectively enhances the accuracy of causal relationship extraction from medical texts by combining BERT's semantic feature extraction capability with the relational modeling advantages of graph neural networks, thereby validating the efficacy of the proposed method.

Key words: Medical text; BERT model; Graph attention network; Causality extraction

实体因果关系抽取主要是从医疗文本中识别和提取医疗实体之间的因果关系,例如“肥胖是糖尿病肾脏病变的独立危险因素”,表明肥胖是糖尿病肾脏病变可能的病因,即“肥胖”与“糖尿病肾脏病变”之间存在因果关系。这种关系反映实体之间的因果逻辑,在疾病预测、医疗问答机器人等应用中具有重要价值。但是,由于医疗文本中专业术语多、因果实体长距离依赖、数据集缺乏等问题,设计一种稳定有效的因果关系挖掘框架用于医疗文本值得研究。

目前因果关系抽取研究主要聚焦于基于规则、基于统计学、基于深度学习3种方法。基于规则的方法依赖于文本特征的人工筛选,如词汇表、句法、匹配模式等。这种方法尽管进行了大量人工建模,但仍然不能列举出所有可能会出现的关系表达式,导致在进行因果关系抽取时,由于触发词不同、主客体长距离依赖、词汇缩写等问题,出现规则的不匹配。基于统计学的方法在进行因果关系抽取之前需要精心设计特征,并对语料库中的数据进行大量的特征工程操作,以求从中学到因果关系。

深度学习技术在自然语言处理(natural language processing, NLP)领域取得了显著进展。卷积神经网络(convolutional neural networks, CNN)^[1]利用n-gram特征在关系抽取任务中得到广泛应用;通用语言模型基于Transformer的双向编码器(bidirectional encoder representations from Transformers, BERT)^[2]被用于神经网络,在许多NLP任务中取得SOTA成绩。除此之外,图神经网络(graph neural networks, GNN)^[3]由于出色的关系表达能力也被广泛地应用于实体关系的任务抽取。然而,无论是CNN、BERT亦或是GNN,在医疗文本因果关系抽取方面都存在不足:①BERT需要大规模数据进行预训练,而对于医疗领域缺乏可利用的数据支持;②CNN无法利用上下文信息,难以解决文本中实体的长距离依赖;③GNN建模需要结构化数据,而医疗文本是非结构化;④由于医疗文本中复杂的医理解释,导致医疗文本中的实体普遍具有长距离依赖及大量重复的实体,仅靠单个模型很难全面抽取医疗文本中的实体因果关系。

为解决上述问题,本文结合BERT强大的特征

表示能力和GNN强大的关系表示能力,提出一种实体因果关系抽取模型BERT-因果图注意力网络(causal graph attention networks, CGAT)。BERT-CGAT融合了从文本中获得的句子语义、实体语义和从因果关系图中获得的实体因果关系,利用文本描述和图结构数据为因果关系抽取提供了丰富的特征表示。利用数据的全局语义信息建立实体因果关系图,并利用预训练模型BERT生成输入文本的嵌入表示。将文本的嵌入表示和对应的因果关系图进行信息融合作为图注意力网络(graph attention networks, GAT)的输入。通过双通道解码层抽取因果关系,获得因果关系三元组。与传统方法相比,本研究模型充分利用了句子的编码表示及因果关系图,对长距离实体及其因果关系更加敏感;GAT中包含的注意力机制增强了对医疗文本中的重复实体的筛选能力。

1 资料与方法

1.1 资料

本研究以糖尿病为例构建了医疗文本因果实体数据集,使用来源于阿里天池的DiaKG数据集^[4],利用ChatGPT^[5]作为虚拟专家从中筛选包含因果关系的文本,然后使用BIO标注法对筛选出的句子进行标注和整理,包括对句子长度的处理。标注文本包括疾病、病因、发病部位、发病机制、临床表现、检查方法、检查指标、用药名称、治疗方法等实体类型以及疾病与病因、疾病与检查指标、疾病与治疗方法、疾病与临床表现等九种因果关系类型。从中筛选出具有因果关系的数据共18 226条,以8:2的比例划分训练集和测试集。其中共标注26 452个医学实体、16 457对因果关系,并详细标注实体的位置及组成要素。

1.2 方法

1.2.1 基于BERT和GAT医疗文本因果关系抽取算法框架

实体因果关系抽取是从文本中识别并抽取实体之间的因果联系,有助于构建更丰富的知识库和理解文本的深层含义。实体通常是指文本中由一个词或多个词组成的有意义的序列,代表现实世界中的一个特定对象、概念或类别。在实体因果关系抽取

中,用 $S=(w_1, w_2, w_3, \dots, w_n)$ 表示由多个词元 w_i 组成的语句。 $l=(B-C, I-C, B-E, I-E, O)$ 表示通过 BIO 标注法获得的标签集合,其中, B-C 代表 w_i 是原因的开始、I-C 代表 w_i 是原因的延续、B-E 代表 w_i 是

结果的开始、I-E 代表 w_i 是结果的延续、O 代表 w_i 是非目标词。实体因果关系抽取模型需要输出 w_i 属于 l 中每个标签的概率,旨在从文本中提取包含因果关系的实体三元组,见图 1。

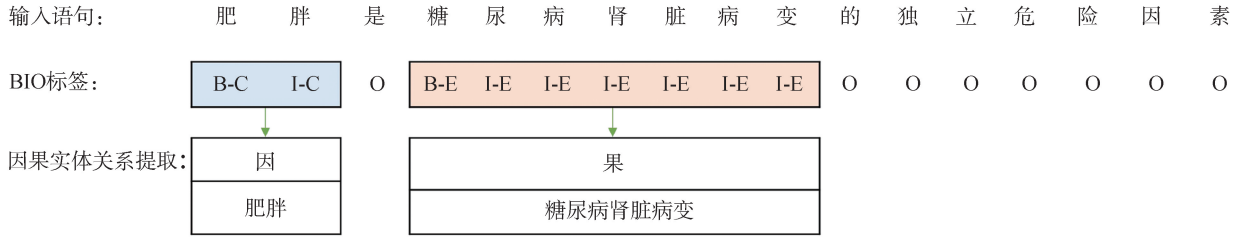


图 1 实体关系抽取示意图
Figure 1 Entity relationship extraction diagram

本研究提出一种基于 BERT 和 GAT 的医疗文本因果关系抽取算法 BERT-CGAT,在文本编码层采用了 BERT 模型,并使用微调后的 BERT 模型以

获得文本编码,最后结合 GAT 学习文本的因果结构以提高因果关系抽取的准确性。模型的具体结构,见图 2。

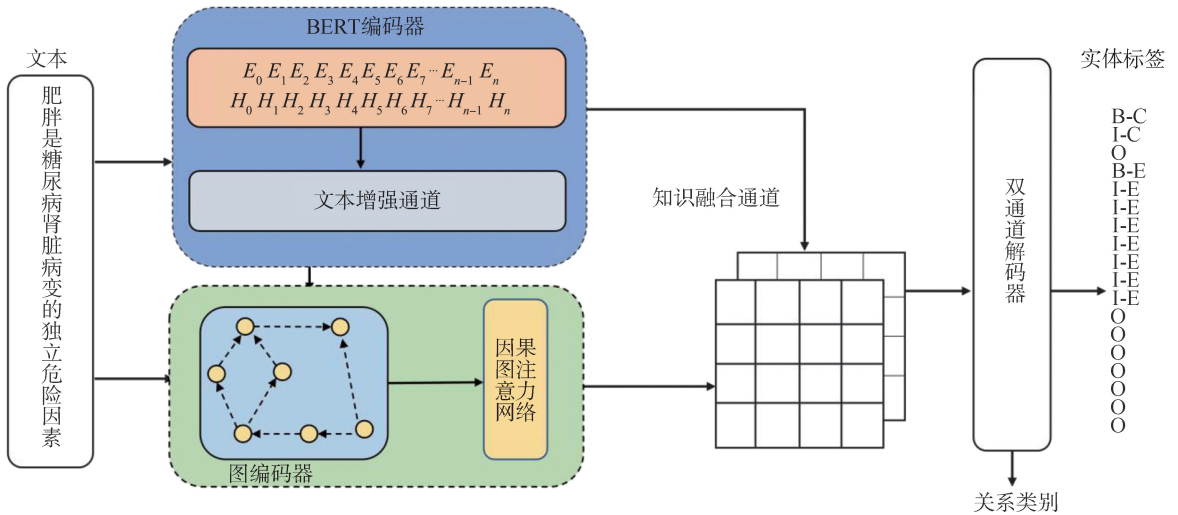


图 2 BERT-CGAT 网络结构图
Figure 2 BERT-CGAT network architecture diagram

①将文本输入 BERT 模型获得词嵌入向量,并将嵌入向量使用文本增强通道进行处理,文本增强通道的输出为节点特征向量;同时,将文本输入因果图编码器对数据集内的数据进行因果图建模,获得文本数据的边索引及边初始权重信息,结合文本增强通道的输出生成完整的因果关系图数据。②使用因果图注意力网络学习因果关系图中的因果结构。③使用知识融合通道将 BERT 模型的实体预测结果和因果图注意力网络的输出进行融合,并输入双通道解码层获得实体和关系标签,进而得到因果三元组。

1.2.2 文本编码层

本研究使用由多层双向 Transformer 堆叠而成的 BERT 模型作为文本编码器,每层利用上下文信息获得特征表示,通过获取输入文本的动态词向量捕获文本的深层语义信息,并且 BERT 模型使用大规模数据进行预训练,因此 BERT 模型可以捕获文

本更丰富的语义表达。本研究使用通用 BERT-Chinese 的预训练模型作为基础模型。由于医疗文本中包含许多医疗领域特有的实体、医疗文本数据集多标签及实体重叠严重,使用数据集中的文本对通用 BERT-Chinese 模型进行微调。假设输入的文本为 W ,其词元集合为 $[w_1, w_2, \dots, w_n]$,将词元集合输入 BERT 模型得到词嵌入表示为:

$$[h_1, h_2, \dots, h_n] = \text{BERT}([w_1, w_2, \dots, w_n]), \quad (1)$$

其中, h_i 为经过 BERT 模型的隐藏层嵌入表示, w_i 为句子中每个词的 token 表示, n 为文本长度。

使用微调后的 BERT 模型获得句子的词嵌入向量。然后,词嵌入向量会输入到文本增强通道,通过迭代扩张卷积网络 (iterated dilated convolutional neural network, IDCNN) 进行进一步的特征提取^[6]。IDCNN 通过在卷积操作中引入膨胀率来扩展感受野,使得模型能够在捕捉全局上下文信息

的同时,保持对局部特征的敏感性。IDCNN 通过多层膨胀卷积叠加,可以有效捕获文本中的细粒度和全局特征,这对于处理复杂的医疗文本中的因果关系尤为重要。最后,将 IDCNN 输出的增强特征作为后续模块输入,以构建更加精确的因果关系图。

1.2.3 因果图建模

本文通过三个步骤来完成对句子的因果图建模:因果句识别、实体检测和因果关系权重计算,建模过程见图 3。因果句识别旨在通过规则匹配,从标注的长文本中提取包含因果关系的句子,包括定

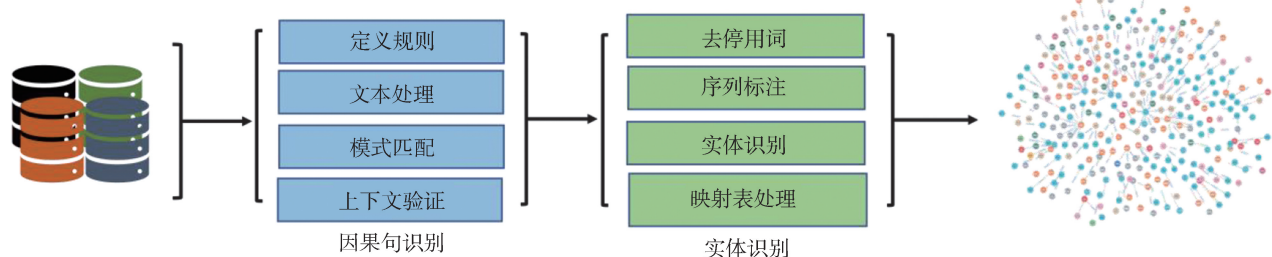


图3 因果图建模概述

Figure 3 Overview of causal diagram modeling

通过以下步骤提取因果句中的实体。首先,删除停用词和特定实体(例如人名、组织名和地名),然后对句子进行序列标注和命名实体识别,从句子中挑选出可以作为实体的部分。最后使用映射表处理实体中包含的同义词和术语缩写,得到可以用来作为网络节点的实体。

通过因果句的识别和实体检测,得到大量的因果句及句中带有因果关系的实体。由于因果关系具有明确的方向性,根据实体在句子中的位置,可以分为从因到果的顺序和从果到因的顺序。本研究基于贝叶斯网络设计了因果关系图,以贝叶斯网络中的先验概率作为因果关系图中的边权重。因果关系图可以用 $g=(w, e)$ 来表示, w_i 为网络节点的实体, e_{ij} 表示从节点 w_i 到节点 w_j 的边,计算实体之间的共

义规则、文本处理、模式匹配、上下文验证。① 根据文本中的因果表达方式定义了因果连接词、条件表达、语境依赖、统计频率等 10 类判定因果句的规则。② 对数据中的同义词和术语缩写进行扩展,将文本中的同义词与术语缩写统一为一个标准词并映射到一个标准词表。③ 使用语言技术平台(language technology platform, LTP)^[7] 识别判定规则及标准词表中的词性。通过判断因果指示词是否成为实体的一部分或者标准词的词性是否发生变化,消除处理过程中产生的歧义部分,尽量避免通过模式匹配得到错误因果句。

现概率作为边权重,共现概率按照以下方式构造:

$$D^{w_i \rightarrow w_j} = \frac{\text{co}(w_i \rightarrow w_j)}{\sqrt{\text{DF}(w_i) * \text{DF}(w_j)}}, \quad (2)$$

其中, $D^{w_i \rightarrow w_j}$ 中的节点 w_i 和节点 w_j 表示共现概率, $\text{co}(w_i \rightarrow w_j)$ 中的节点 w_i 和节点 w_j 表示共现次数, $\text{DF}(w_i)$ 表示节点 w_i 在数据中出现的次数。

1.2.4 因果图注意力层

为了获得更多的语义信息来捕获医疗文本中词与词、词与关系和关系之间的复杂相互作用,本研究在图注意力层中采用了多头注意力机制,通过多个注意力头并行处理信息,每个头学习不同的表示子空间,从而增强 BERT-CGAT 模型对医疗文本中复杂语义关系的表征能力,每个节点参数更新过程,见图 4。

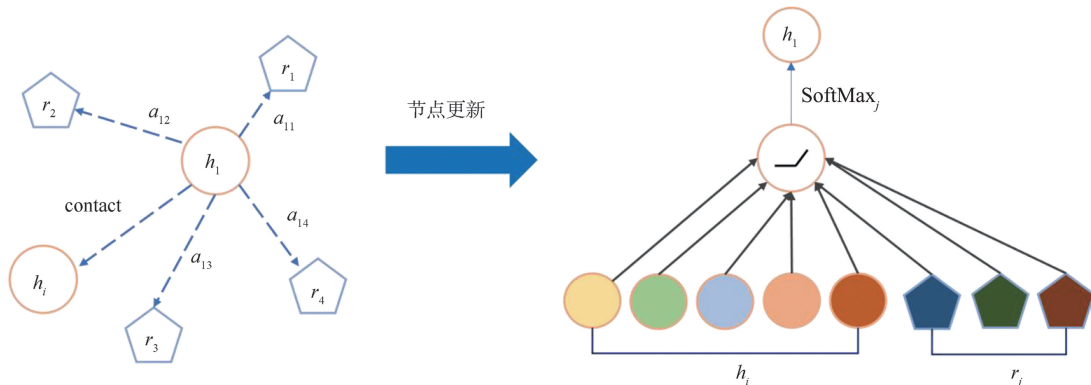


图4 节点参数更新示意图

Figure 4 Schematic diagram of node parameter update

图注意力层中,对于每个节点 i ,聚合其所有邻居节点 j 的特征,

$$\mathbf{z}_i = \sum_{j \in N(i)} \mathbf{W} \mathbf{h}_j, \quad (3)$$

其中, $N(i)$ 是节点 i 的邻居节点的集合。每个节点 i 通过 BERT 模型获得特征向量 \mathbf{h}_i , 同时每条边关联一个权重矩阵 \mathbf{W} 。

图 4 中 a_{ij} 表示节点与其邻居节点之间的注意力分数, 计算公式如下:

$$a_{ij} = \mathbf{a}^T [\mathbf{h}_i \parallel \mathbf{z}_i]. \quad (4)$$

通过 SoftMax 函数对注意力分数进行归一化, 确保每个节点的邻居贡献之和为 1, 使得注意力权重可以合理地分配给每个邻居。计算公式如下:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N(i)} \exp(e_{ik})}. \quad (5)$$

为了反应节点在图中的局部结构和内容信息, 本研究使用归一化的注意力权重对邻居特征进行加权求和, 更新每个节点的特征表示, 计算公式如下

$$\mathbf{h}'_i = \sum_{j \in N(i)} \alpha_{ij} \mathbf{W} \mathbf{h}_j. \quad (6)$$

为了避免训练过程中出现梯度消失的问题, 本研究使用了残差计算连接。具体计算公式如下:

$$\mathbf{h}_i^{\text{residual}} = \mathbf{h}_i + \mathbf{h}_i^{\text{dropped}}, \quad (7)$$

其中, $\mathbf{h}_i^{\text{residual}}$ 为残差计算后的节点特征, \mathbf{h}_i 为更新后的节点特征, $\mathbf{h}_i^{\text{dropped}}$ 为经过 dropped 层后的节点特征。

1.2.5 知识融合通道

为了用更广的感知域提取实体的特征信息, 文本编码层引入了文本增强通道, 在充分考虑上下文信息的情况下提取完整的实体, 因果图建模可以捕捉实体之间的因果关系结构。医疗文本存在大量的隐式因果关系, 利用这些因果关系图可以提高模型在实体因果关系提取中的准确率和鲁棒性。因果图注意力网络可以为实体抽取的学习过程提供指导信息, 并且两部分信息拥有不同的优先级。本研究设计了一个知识融合通道, 将文本的编码信息与因果图注意力网络的输出相融合。使用多头注意力机制^[8]将两部分链接在一起, 以表示实体因果关系的全局偏好。

首先, 对每个注意力头计算 BERT 和 CGAT 的输出之间的注意力分数, 计算公式如下:

$$\text{Attention}_{\text{BERT-CGAT}} = \text{SoftMax} \left(\frac{\mathbf{Q}_{\text{BERT}} \times \mathbf{K}_{\text{CGAT}}^T}{\sqrt{d_k}} \right), \quad (8)$$

其中, $\text{Attention}_{\text{BERT-CGAT}}$ 为每个注意力头的注意力分数, \mathbf{Q}_{BERT} 为 BERT 模型输出的查询矩阵, \mathbf{K}_{CGAT} 为图注意力网络输出的键矩阵, d_k 为键矩阵的维度。

然后, 使用注意力分数对值向量进行加权求和, 计算公式如下:

$$\mathbf{H}_i = \text{Attention}_{\text{BERT-CGAT}} \times \mathbf{V}_{\text{CGAT}}, \quad (9)$$

其中, \mathbf{H}_i 表示第 i 个注意力头的输出, \mathbf{V}_{CGAT} 为因果图注意力网络输出的值矩阵。

最后, 将所有注意力头的输出拼接在一起, 并通过线性变换得到最终的融合表示, 计算公式如下:

$$\mathbf{H} = \mathbf{W}_0 \times \text{Contact}(\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_n), \quad (10)$$

其中, \mathbf{H} 为融合后的注意力表示, \mathbf{W}_0 为线性变换矩阵。

1.2.6 双通道解码层

本研究设计了一个双通道解码层, 使用因果图注意力层的输出作为输入, 通过一个额外的实体识别层来预测每个词是否为实体的开始或结束。在实体识别的基础上, 使用另一个解码通道来预测实体之间的关系。实体识别通道将因果图注意力层输出的特征从节点映射到实体类别的得分。应用 SoftMax 获取每个节点属于每个实体类别的概率。计算公式如下:

$$S_e = \text{SoftMax}(\mathbf{W}_e \cdot \mathbf{h} + b_e), \quad (11)$$

其中, S_e 表示每个节点属于每个实体类别的概率, \mathbf{h} 是 BERT-CGAT 输出的节点特征, \mathbf{W}_e 和 b_e 是实体识别层的权重和偏执。

关系识别通道使用实体的特征表示作为输入, 来预测实体队之间的关系类别, 最后应用 SoftMax 函数来获取实体对属于每个可能关系类别的概率。计算公式如下:

$$S_r = \text{SoftMax}(\mathbf{W}_r \cdot (\mathbf{h}_i \oplus \mathbf{h}_j) + b_r), \quad (12)$$

其中, S_r 表示实体对属于每个可能关系类别的概率, \mathbf{h}_i 和 \mathbf{h}_j 是一对实体的特征表示, \oplus 表示特征的点积, \mathbf{W}_r 和 b_r 是关系识别层的权重和偏置。

因果图注意力层的输出经过双通道解码后得到因果关系三元组。

1.3 对比实验

1.3.1 实验设置

对于收集数据集, 使用 BERT 对文本进行编码以生成嵌入表示, 并针对数据集构建了关于糖尿病及其并发症的中文因果关系图。该模型在训练过程中的超参数如下: ①模型训练的 epoch 数为 60 个; ②训练集的批量大小为 8; ③优化器 Adam 的学习率为 1×10^{-4} 。为了保证实验结果的可靠性, 在训练前将所有数据与不同的随机种子进行洗牌, 使用抽取实体和关系的准确率、召回率和 F1 分数作为判定模型性能指标。

1.3.2 对比模型

为验证本文 BERT-CGAT 模型抽取实体因果关系的性能, 本文选取 5 个模型做对比实验, 其中包括 3 个基准模型: 图卷积网络 (graph convolutional network, GCN)、BERT、BERT+长短期记忆网络 (long short-term memory networks, LSTM); 以及两个主

流模型 BiLSTM+CRF、BiLSTM+self-ATT。本文针对数据集的特征对上述模型数据输入部分做了部分修改,以保证模型正常训练。

2 结果

本文主要考虑因果关系的抽取,因此主要关注

表1 实验结果
Table 1 Experimental results

| 模型 | 实体 | | | 关系 | | |
|-----------------|---------|---------|---------|---------|---------|---------|
| | 准确率 | 召回率 | F1 分数 | 准确率 | 召回率 | F1 分数 |
| GCN | 0.824 2 | 0.733 3 | 0.728 6 | 0.730 9 | 0.789 5 | 0.742 6 |
| BERT | 0.982 3 | 0.624 2 | 0.687 5 | 0.731 9 | 0.727 9 | 0.724 9 |
| BERT+LSTM | 0.984 7 | 0.664 2 | 0.793 3 | 0.785 5 | 0.782 5 | 0.783 5 |
| BiLSTM+CRF | 0.990 9 | 0.643 1 | 0.612 4 | 0.581 6 | 0.575 6 | 0.573 4 |
| BiLSTM+self-ATT | 0.760 6 | 0.507 5 | 0.571 6 | 0.675 7 | 0.669 7 | 0.660 9 |
| BERT+CGAT | 0.997 4 | 0.810 4 | 0.808 3 | 0.803 6 | 0.793 6 | 0.796 8 |

3 讨论

基于规则的方法是因果关系挖掘中的传统方法之一,其通常依赖于人工制定的规则来识别文本中的因果关联。这些规则可以是简单的词汇匹配,也可以是更复杂的句法结构分析。Gopalan 等^[9]通过显性话语连接词实现了因果关系的自动提取,但是这种通过触发词来提取因果的方式会忽略文本中隐性的因果。Guo 等^[10]考虑文本中显性和隐性的因果关系,在计算中证实实体类别和因果之间的联系;Kabir 等^[11]针对医学文献中因果短语组成复杂的特点,以因果依赖模式集合为模板,提取因果短语的第一个词,继而获得完整的因果短语。Radinsky 等^[12]则是将语义自然语言建模技术应用用于建立因果模式模板,通过已有的因果模式来预测因果关系。但是,基于规则的方法普遍需要人工建立因果触发词词汇库和因果关系模式模板,且词汇库及因果模式的模板通用性差,模板匹配的准确性也有待提高。

基于统计学方法,可通过分析大量数据中的相关性来推断因果关系。这些方法包括回归分析、关联规则挖掘等。Spirtes 等^[13]提出一种基于约束的因果发现算法——PC 算法。这种算法基于因果网络,通过独立性检验和学习因果结构的骨架图来描述多个变量之间的因果关系。Peters 等^[14]为了从联合分布中学习因果关系图,提出连续加性噪声模型,其中每个变量由其父变量的加性函数和 1 个独立的噪声项组成。为了确定因果方向,Zhao 等^[15]提出分层加性噪声模型,通过变分方法将多个因子推广

实体和关系的各项粗粒度指标,实验结果列于表 1。由表 1 可见,本研究提出的 BERT-CGAT 模型优于其他模型,在所构建的数据集上有较好的识别效果,实体抽取的准确率为 99.74%、召回率为 81.04%、F1 分数为 80.83%,关系抽取的准确率为 80.36%、召回率为 79.36%、F1 分数为 79.68%。

到 1 个中间变量中,并利用模型中的不对称性来识别因果关系方向。与基于规则的方法相比,基于统计学的方法可以处理更复杂的数据,并且能够发现数据中的隐含模式,但是需要对数据进行大量的特征工程操作以从中学到因果关系,这降低了方法泛化能力,同时基于统计的方法难以捕捉上下文信息和深层次的语义关系,从而影响抽取的准确性。

传统的方法(如基于规则或统计)在因果关系挖掘领域取得显著的成果,但仍然存在一定的局限性。近年来,基于深度学习的方法在因果关系挖掘领域取得突破性进展。一些模型基于深度学习的方法使用神经网络从复杂的数据中提取有用信息,例如 CNN^[16]、递归神经网络(recurrent neural network, RNN)^[17-18]和 LSTM^[19-20]。这些模型有强大的特征学习和模式识别能力,可以捕捉文本中的深层次特征和复杂结构。然而,CNN 网络的感知范围受到卷积核大小的限制,只能以步进方式获得局部特征。这种方式使得 CNN 无法很好地提取并处理文本的上下文语义特征,从而影响最终结果。为了提高 CNN 对上下文信息的感知,Strubell 等^[6]提出迭代扩张卷积神经网络(iterated dilated convolutional neural network, ID-CNNs),通过以循环方式重复使用相同参数的方法提升了输入宽度,有效提升了 CNN 的感知范围。

由于文本中的实体关系可以表示为图结构数据,为了使用图数据强大的关系表示能力,大量研究着眼于如何在图数据上使用深度学习方法。由于异构图数据的挑战,人们对如何将深度学习方法应用于图进行了深入的研究^[21]。既往研究将图上的神

神经网络视为递归神经网络的一种形式,然而其需要重复应用收缩图作为激活函数,直到节点的表示达到稳定状态。通过添加门控循环单元和改进反向传播优化策略,缓解了这一限制^[22]。Bruna等^[23]将卷积神经网络扩展到图,Defferrard等^[24]使用切比雪夫多项式获得图卷积以替换计算复杂的拉普拉斯特征分解,Kipf等^[25]通过局部一阶近似进一步简化了谱图卷积。近年来,人们对图神经网络在NLP上的应用做了大量的研究,如文本分类^[26]、事件参数抽取^[27]、用于关系抽取的文档级图^[28]、用于事件因果关系识别的文档级图^[29]等。

本文提出一种基于因果图注意力网络的实体因果关系抽取方法,利用因果关系图完成模型结构的构建,融合文本的词嵌入信息和因果图结构信息更新节点信息,在解码阶段采用双通道解码方法获取文本中的实体信息和实体因果关系三元组。通过在糖尿病因果实体数据集上进行训练和验证,证明本文提出的方法相较于其他方法具有一定的优势,适用于医疗文本的实体因果关系抽取。对比实验效果最好的是BERT+LSTM和BERT-GCN,这主要是因为BERT模型有大量的预训练语料,使构建的词向量表征能力更强,更好地学习到了文本中的因果关系,证明使用大规模的预训练模型对因果关系抽取任务有显著提升。实验结果表明,GCN模型相比LSTM在因果关系抽取任务上的性能更好,主要是因为图卷积神经网络中的卷积操作相比LSTM中的单向按语序学习到更多的上下文信息,有助于因果关系的抽取,使用Softmax对卷积结果进行分类筛选对模型的性能也有提升。

然而,本文的研究发现,基于因果图注意力网络的抽取方法对数据集的要求较高,需要大量精确标注的训练数据。除此之外,医疗文本中存在超长距离的依赖关系以及同义词与缩写数量大等问题,依赖人工筛选并建立映射。在未来的研究中,将探索小样本学习方法或者自监督学习方法,降低模型对有标注数据及人工筛选的依赖,使模型可以自动在少量的样本中进行快速有效的学习。

参考文献:

- [1] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences [C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Baltimore, Maryland: Association for Computational Linguistics, 2014: 655-665.
- [2] Devlin J, Chang MW, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]// 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). Minneapolis, Minnesota: Association for Computational Linguistics, 2019: 4171-4186.
- [3] Scarselli F, Gori M, Tsoi AC, et al. The graph neural network model [J]. IEEE Trans Neural Netw, 2008, 20 (1): 61-80.
- [4] Chang D, Chen M, Liu C, et al. DiaKG: an annotated diabetes dataset for medical knowledge graph construction [C]//Knowledge Graph and Semantic Computing: Knowledge Graph Empowers New Infrastructure Construction. Singapore: Springer, 2021: 308-314.
- [5] OpenAI. ChatGPT (v4) [EB/OL]. (2024-05) [2025-02-01]. <https://openai.com/chatgpt>
- [6] Strubell E, Verga P, Belanger D, et al. Fast and accurate entity recognition with iterated dilated convolutions [EB/OL]. (2017-07-22) [2025-02-01]. <https://arxiv.org/abs/1702.02098>
- [7] Che W, Li Z, Liu T. LTP: a chinese language technology platform [C]// Coling 2010: demonstrations. Beijing, China: Coling 2010 Organizing Committee, 2010: 13-16.
- [8] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [EB/OL]. (2017-12-06) [2025-02-01]. <https://arxiv.org/abs/1706.03762v5>
- [9] Gopalan S, Devi SL. Cause and effect extraction from biomedical corpus [J]. CyS, 2018, 21(4). doi: 10.13053/cys-21-4-2854
- [10] Guo Y, Wang ZH, Shao ZQ. Improving causality induction with category learning [J]. Sci World J, 2014, 2014: 650147. doi:10.1155/2014/650147
- [11] Kabir MA, Almulhim A, Luo X, et al. Informative causality extraction from medical literature via dependency-tree-based patterns [J]. J Healthc Inform Res, 2022, 6 (3): 295-316.
- [12] Radinsky K, Davidovich S, Markovitch S. Learning causality for news events prediction [C]//Annual Conference on World Wide Web. Lyon, France: CS Department Technion-Israel Institute of Technology Haifa, Israel. 2012: 909-918.
- [13] Spirtes P, Glymour C. An algorithm for fast recovery of sparse causal graphs [J]. Soc Sci Comput Rev, 1991, 9 (1): 62-72.
- [14] Peters J, Mooij JM, Janzing D, et al. Causal discovery with continuous additive noise models [J]. J Mach Learn Res, 2014, 15(58): 2009-2053.
- [15] Zhao BX, Wang SL, Chi LH, et al. HANM: hierarchi-

- cal additive noise model for many-to-one causality discovery [J]. *IEEE Trans Knowl Data Eng*, 2023, 35 (12): 12708-12720.
- [16] Gu JX, Wang ZH, Kuen J, et al. Recent advances in convolutional neural networks [J]. *Pattern Recognit*, 2018, 77: 354-377. doi:10.1016/j.patcog.2017.10.013
- [17] Socher R, Lin CC, Manning C, et al. Parsing natural scenes and natural language with recursive neural networks [EB/OL]. (2024-04-09) [2025-02-01]. https://nlp.stanford.edu/pubs/SocherLinNgManning_ICML2011.pdf
- [18] Luong MT, Socher R, Manning CD. Better word representations with recursive neural networks for morphology [EB/OL]. (2024-04-09) [2025-02-01]. <https://aclanthology.org/W13-3512.pdf>
- [19] He DC, Zhang HJ, Hao WN, et al. Distant supervised relation extraction via long short term memory networks with sentence embedding [J]. *Intell Data Anal*, 21 (5): 1213-1231.
- [20] Zheng SC, Xu JM, Zhou P, et al. A neural network framework for relation extraction: learning entity semantic and relation pattern [J]. *Knowl Based Syst*, 2016, 114: 12-23. doi:10.1016/j.knosys.2016.09.019
- [21] Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains [EB/OL]. (2024-08-06) [2025-02-01]. <https://ieeexplore.ieee.org/document/1555942>
- [22] Li Y, Tarlow D, Brockschmidt M, et al. Gated Graph Sequence Neural Networks [EB/OL]. (2024-08-06) [2025-02-01]. <https://www.semanticscholar.org/paper/492f57ee9ceb61fb5a47ad7aebfec1121887a175>
- [23] Bruna J, Zaremba W, Szlam A, et al. Spectral networks and locally connected networks on graphs [EB/OL]. (2014-05-21) [2025-02-01]. <http://www.xueshufan.com/publication/1662382123>
- [24] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering [C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2016: 3844-3852.
- [25] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks [EB/OL]. (2017-02-22) [2025-06-16]. <http://arxiv.org/abs/1609.02907>
- [26] Yao L, Mao C, Luo Y. Graph convolutional networks for text classification [C]//Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu, Hawaii, USA: AAAI Press, 2019: 7370-7377.
- [27] Veyseh APB, Nguyen TN, Nguyen TH. Graph transformer networks with syntactic and semantic structures for event argument extraction [EB/OL]// (2020-10-26) [2025-02-01]. <https://arxiv.org/abs/2010.13391>
- [28] Christopoulou F, Miwa M, Ananiadou S. Connecting the dots: document-level neural relation extraction with edge-oriented graphs [C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, 2019: 4925-4936.
- [29] Phu MT, Nguyen TH. Graph convolutional networks for event causality identification with rich document-level structures [C]. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Bangkok, Thailand: Association for Computational Linguistics, 2021: 3480-3490.

(编辑:相峰)