



# 计算机科学

COMPUTER SCIENCE

## **PosNet:基于位置的因果关系抽取网络**

朱广丽, 许鑫, 张顺香, 吴厚月, 黄菊

### 引用本文

朱广丽, 许鑫, 张顺香, 吴厚月, 黄菊 [PosNet:基于位置的因果关系抽取网络](#) [J]. 计算机科学, 2022, 49(12): 305-311.

ZHU Guang-li, XU Xin, ZHANG Shun-xiang, WU Hou-yue, HUANG Ju. [PosNet:Position-based Causal Relation Extraction Network](#) [J]. Computer Science, 2022, 49(12): 305-311.

---

### 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

#### Similar articles recommended (Please use Firefox or IE to view the article)

#### [基于高秩特征和位置注意力的RGBT目标跟踪](#)

RGBT Object Tracking Based on High Rank Feature and Position Attention

计算机科学, 2022, 49(12): 236-243. <https://doi.org/10.11896/jsjcx.220600037>

#### [基于全局增强图神经网络的序列推荐](#)

Sequence Recommendation Based on Global Enhanced Graph Neural Network

计算机科学, 2022, 49(9): 55-63. <https://doi.org/10.11896/jsjcx.210700085>

#### [融合RACNN和BiLSTM的金融领域事件隐式因果关系抽取](#)

Implicit Causality Extraction of Financial Events Integrating RACNN and BiLSTM

计算机科学, 2022, 49(7): 179-186. <https://doi.org/10.11896/jsjcx.210500190>

#### [基于区域激活策略的Tiny YOLOv3目标检测算法](#)

Tiny YOLOv3 Target Detection Algorithm Based on Region Activation Strategy

计算机科学, 2021, 48(6A): 118-121. <https://doi.org/10.11896/jsjcx.200700122>

#### [多级字典存储的未知文本协议候选关键词链式合并方法](#)

Chain Merging Method for Unknown Text Protocol Candidate Keyword Stored in Multi-level Dictionary

计算机科学, 2020, 47(12): 332-335. <https://doi.org/10.11896/jsjcx.190900116>

# PosNet: 基于位置的因果关系抽取网络

朱广丽 许鑫 张顺香 吴厚月 黄菊

安徽理工大学计算机科学与工程学院 安徽淮南 232001

**摘要** 因果关系抽取是一种从文本中抽取因果实体对的自然语言处理技术,被广泛应用于金融、医疗等领域。传统的因果关系抽取技术需要人工选取文本特征进行因果匹配或使用神经网络多次提取特征,导致模型结构较为复杂,抽取效率不高。针对这一问题,提出一种基于位置的因果关系抽取网络(Position-based Causal Extraction Network, PosNet),以期提高因果关系的抽取效率。首先,预处理文本,构建多粒度文本特征作为网络的输入;然后,将文本特征传入位置预测网络,使用经典的浅层卷积神经网络预测因果实体的开始位置和结束位置;最后,通过组装算法按起始位置组装因果实体,抽取全部因果实体对。实验结果证明 PosNet 可以提升因果关系抽取的效率。

**关键词:** 因果关系抽取;位置信息;文本特征表示

**中图分类号** TP391

## PosNet: Position-based Causal Relation Extraction Network

ZHU Guang-li, XU Xin, ZHANG Shun-xiang, WU Hou-yue and HUANG Ju

School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001, China

**Abstract** Causal relation extraction is a natural language processing technology to extract causal entity pairs from text, which is widely used in financial, medical and other fields. Traditional causal relationship extraction technology needs to manually select text features for causal matching or use neural networks to extract features many times, resulting in complicated model structure and low extraction efficiency. To solve this problem, this paper proposes a position-based causal relation extraction network(PosNet) to improve the efficiency of causal relation extraction. Firstly, it preprocesses the text and constructs multi-granularity text features as the input of the network. Then passing the text features into the position prediction network, and predicting the start and end positions of causal entities by the classical shallow convolution neural network. Finally, the causal entities are assembled according to the start and end positions by the assembling algorithm, so that all causal entity pairs are extracted. Experimental results show that PosNet can improve the efficiency of causal relation extraction.

**Keywords** Causal relation extraction, Position information, Text feature representation

## 1 引言

近年来,全球大数据储量每年都保持 40%左右的增长。海量文本数据中包含了大量因果关系,例如“特朗普政府对待疫情的消极政策导致民众支持率下降”这段文本包含了(消极政策,支持率下降)因果关系对。从海量文本中挖掘事件之间的因果关系,可用于现代企业的智能问答、风险预警、系统评估、事件推理等方面,加速产业的智能化发展。然而,面对数据的爆发式增长,传统的抽取方式不再满足大数据场景下的应用需求,提高因果关系抽取效率逐渐成为各领域的战略追求。

在目前的因果抽取任务中,文本之间的因果关系可分为显式和隐式。其中,显式因果特征较为直观,可以通过模式

匹配获取;而针对隐式因果特征不易获取这一问题,大多数模型通过增加网络层数或宽度来解决,导致模型的结构较为复杂且效率不高。复杂的网络结构会增加模型整体的参数量,同时参数初始化的合理性也可能导致训练时网络拟合速度较慢,进而影响抽取效率。

为了构建一个简单高效的因果抽取网络,需要考虑以下几点:1)如何将文本转为语义信息丰富、上下文关联的特征表示? 2)如何通过神经网络预测因果实体的关键位置信息,以降低模型结构的复杂度? 3)如何通过预测得到的位置信息来抽取完整的因果实体?

基于上述考虑,为了提高因果关系的抽取效率,本文提出了基于位置的因果关系抽取网络 PosNet。首先构造含有

到稿日期:2021-11-26 返修日期:2022-03-11

基金项目:国家自然科学基金面上项目(62076006);安徽高校协同创新项目(GXXT-2021-008);安徽省重点研发计划国际科技合作专项(202004b11020029)

This work was supported by the National Natural Science Foundation of China(62076006), University Synergy Innovation Program of Anhui Province(GXXT-2021-008) and Anhui Provincial Key R & D Program(202004b11020029).

通信作者:朱广丽(glzhu@aust.edu.cn)

位置信息的文本特征,然后使用经典的浅层卷积神经网络预测因果实体的起始位置,最后通过组装算法抽取因果实体。模型的框架如图 1 所示,网络由 3 个模块组成。

(1)嵌入模块(Embedding Module)。嵌入模块将网络的原始输入文本转换为多粒度的文本特征表示。文本特征由 3 个部分计算得到:1)预训练模型生成的 Embedding(E);2)基于相对位置信息的 Relative Position Embedding(RPE);3)基于字符信息的 Character Embedding(CE)。

(2)位置预测模块(Position Prediction Module)。将嵌入模块构造的文本特征传入卷积神经网络,预测因果实体的起始位置。该模块设有两个并行的 CNN 通道 Start Net 和 End Net,分别用于预测因果实体的开始位置和结束位置。

(3)组装模块(Assembler Module)。使用组装器算法将 Start Net 和 End Net 预测的因果实体的起始位置组合起来,抽取全部的因果实体对。

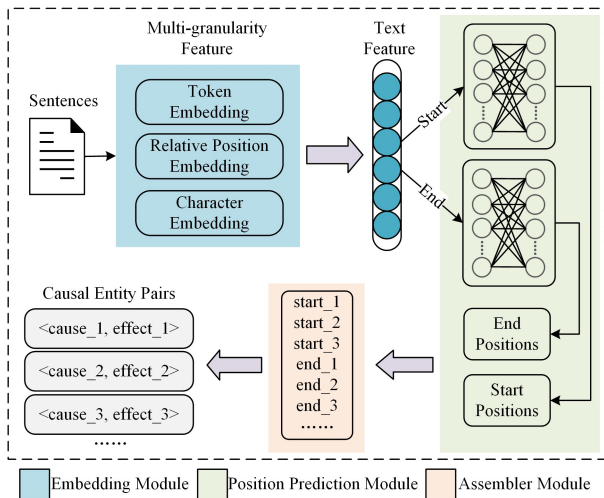


图 1 PosNet 模型

Fig. 1 PosNet model

本研究工作的主要贡献包括:

(1)提出了一种基于预测因果实体起始位置进行组装的因果抽取方法。

(2)定义了 CTS 系数,计算文本之间的内容相似度和语义相似度,用于度量因果抽取任务中预测因果与真实因果之间的匹配程度。

(3)定义了惩罚窗口,在预测因果实体位置时,用于减小正负样本不平衡导致的损失误差。

## 2 相关工作

因果关系抽取方法可分为模式匹配和机器学习两类<sup>[1]</sup>。基于模式匹配的方法需要人工总结因果关系关键标识词,构建因果关系模板,通过因果关系模板与原句匹配发现因果关系。基于机器学习的方法需要构建统计模型或神经网络自动学习文本的因果特征,无须预先学习语义知识即可实现因果关系的抽取。

### 2.1 基于模式匹配的方法

在早期的因果关系抽取工作中,Ittoo 等<sup>[2]</sup>使用因果动词构建因果关系匹配模板,针对显式因果和隐式因果进行标识,

在稀疏文本中取得了较好的抽取效果。Luo 等<sup>[3]</sup>从大量文本语料中学习因果关系术语并构建因果网络作为因果关系的度量标准,用于识别短语之间、短语和句子之间的因果关系。Zhao 等<sup>[4]</sup>使用贝叶斯模型,结合上下文特征、句法特征、位置特征和因果连接词特征,进行因果关系的抽取。Seol 等<sup>[5]</sup>结合事件因果关系模式和临床语义单元,基于条件随机场从叙事文本中提取事件。Lee 等<sup>[6]</sup>提出了一个因果疾病网络,通过对因果关系强度和频率进行分析,从生物医学文献中抽取因果关系。Lee 等<sup>[7]</sup>在多个句子中对相邻的句子进行编码,提出一种基于实体链接的从多个句子中提取  $N$  元关系的方法,在 PubMed 等数据集上验证了该方法的有效性。Fratini 等<sup>[8]</sup>基于句法标准,提出了一个自动因果抽取系统,该方法可以从工程领域中按需求提取关系,但其仅在理论上得到证明,缺乏实际运用。Heindorf 等<sup>[9]</sup>使用因果模板构建的一个大型的因果关系图网络 CauseNet,可用于计算论证、智能问答等任务。

### 2.2 基于机器学习的方法

近年来,基于机器学习的抽取方法逐步占据主导地位。Kruengkrai 等<sup>[10]</sup>从网络文本中学习因果知识并提取因果关系候选句,然后使用多层卷积神经网络识别事件之间的因果关系。Zheng 等<sup>[11]</sup>提出了一个混合神经网络,使用双向 LSTM 编解码器提取事件实体,通过 CNN 分类器获取实体之间的关系,实现了事件之间的关系抽取。Zheng 等<sup>[12]</sup>提出了一种因果事件标注策略,将因果事件抽取问题转换为序列标注问题。Zeng 等<sup>[13]</sup>使用复制机制从文本中复制因果实体,将其组合成为因果实体关系对。Dasgupta 等<sup>[14]</sup>使用 BiLSTM 网络进行因果抽取,并且将抽取出的因果事件构建为因果事理图谱。Li 等<sup>[15]</sup>提出了一种卷积神经网络,以先验知识为驱动,获取因果关系的语言线索,借此进行因果关系的抽取。Li 等<sup>[16]</sup>提出了 SCITE 网络,采用序列标注方法,使用 BiLSTM+CRF 提取因果关系。Sahu 等<sup>[17]</sup>提出一种自适应图卷积网络,使用注意力机制来进行编码,实现了实体之间端到端的关系抽取。Zhao 等<sup>[18]</sup>将事件因果关系识别转化为图的边缘预测问题,利用上下文信息进行基于图的因果推理。Cao 等<sup>[19]</sup>提出一种嵌套关系提取方法,使用有向无环图抽取因果关系。Jiao 等<sup>[20]</sup>提取模拟电路设置中的因果实体,用于理解电路设计中的拓扑结构。Kim 等<sup>[21]</sup>从专利文件中提取技术之间的因果关系。Maisonave 等<sup>[22]</sup>使用基于 BERT 的循环神经网络,分析 4 种因果抽取方法的差异和共性。Nasar 等<sup>[23]</sup>指出,当前很多先进模型都是离线运行且抽取效率不高,需要对模型的架构进行有效调整,以使其适应当今社会文本数据的爆发式增长需求。

基于模式匹配的方法对文本的相似性有较高的要求,只能抽取显式因果,无法抽取隐式因果。基于机器学习的方法神经网络结构较为复杂,没有考虑到因果关系抽取这一任务的特殊性,对模型参数的优化方法不够合理。本文提出 PosNet,结合预训练文本特征和经典卷积神经网络抽取因果关系。

### 3 嵌入模块

嵌入模块负责获取文本的语义特征和位置信息,框架图如图 2 所示。

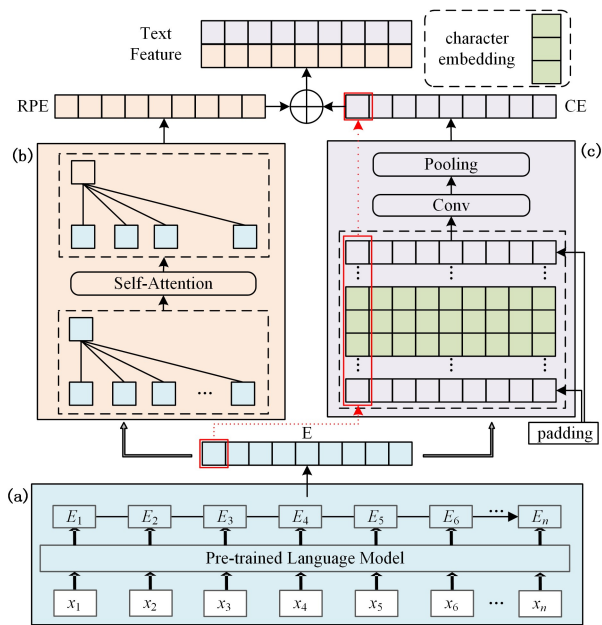


图2 嵌入模块

Fig. 2 Embedding module

图2(a)获取预训练词向量。为了使输入序列的词向量表示方法更加准确,使用预训练模型生成输入序列中每个词的 Embedding(E)。预训练语言模型由海量语料训练得到,得到的词向量含有更加丰富的语义信息。

图2(b)向词向量添加相对位置信息。为了解决词义混淆问题,将(a)中得到的 Embedding 输入 Self-Attention 层,计算得到 Relative Position Embedding(RPE)。这种词向量采用相对的位置计算方式,可以区分词在句中的前后位置关系,获取基于位置的语义信息。

图2(c)获取字符级特征。为了获取粒度更细的语义信息,将每个英文单词拆分为字符,再通过卷积层和池化层计算得到每个词的 Character Embedding(CE)。单词的字符级特征可以丰富文本特征,为因果抽取提供更多隐含信息。

### 3.1 Embedding(E)

首先将文本转换为词向量,主流的方法是通过个人训练或预训练语言模型获取。Word2Vec<sup>[24]</sup>模型通过构建词袋模型来统计词频信息,将每个词映射为一个词向量;BERT<sup>[25]</sup>预训练语言模型通过对大量语料执行预训练任务,将单词映射为词向量。

在“an agreed account”和“on account of”这两段文本中,两个“account”显然表达了不同的意思。然而,通过 Word2Vec 模型得到的两个“account”的词向量完全相同,忽略了词的顺序。BERT 模型得到的两个“account”的词向量会根据上下文语境不同而变化,解决了“一词多义”问题,但模型的参数量较大。

Albert<sup>[26]</sup>是在 BERT 基础上设计的精简模型,通过参数因式分解和跨层参数共享技术,在保证性能的同时,解决了 BERT 参数过多、训练缓慢的问题。因此,本文的词向量通过 Albert 模型获取,每个词向量的维度为 768。

### 3.2 Relative Position Embedding(RPE)

相对位置编码可以解决经典 Transformer 无法感知相同字符前后关系的问题。自注意力的输入序列为  $x = (x_1,$

$x_2, \dots, x_n), x_i \in \mathbb{R}^{d_x}$ , 输出序列为  $z = (z_1, z_2, \dots, z_n), z_i \in \mathbb{R}^{d_z}$ 。Shaw 等<sup>[27]</sup>在经典的 Self-Attention 计算公式中加入元素之间的相对位置信息,其中  $z_i$  由式(1)计算得到。 $\alpha_{ij}^V = \omega_{j-i}^V, \alpha_{ij}^K = \omega_{j-i}^K (\alpha_{ij}^V, \alpha_{ij}^K \in \mathbb{R}^{d_a})$  表示输入序列中元素  $x_i$  和  $x_j$  之间的相对位置; $W^Q, W^K, W^V \in \mathbb{R}^{d_x \times d_z}$  为参数矩阵。

$$\begin{cases} e_{ij} = \frac{x_i W^Q (x_j W^K + \alpha_{ij}^K)^T}{\sqrt{d_z}} \\ \alpha_{ij} = \frac{e^{e_{ij}}}{\sum_{k=1}^n e^{e_{ik}}} \\ z_i = \sum_{j=1}^n \alpha_{ij} (x_j W^V + \alpha_{ij}^V) \end{cases} \quad (1)$$

从输入文本的第一个词开始,依次计算这个词和左右若干个词(最多  $2k+1$  个词)的 Self-Attention Score,并将其作为权重赋予该词。一般情况下,具有因果关系的实体相距较近,因此注意力机制的计算范围被限制在  $k$  以内,以降低模型的计算复杂度。相对位置编码表示一定范围内单词之间的前后位置关系,可以在模型抽取因果关系时提供位置信息。

### 3.3 Character Embedding(CE)

单词可以拆分为多个字符的组合,字符级语义特征比单词级语义特征的粒度更细。事件之间的因果关系往往较为复杂,因此因果关系特征难以捕捉。为了丰富输入文本的特征表示,本文通过卷积神经网络生成单词的字符级特征 CE,为模型预测提供更多的语义线索。

首先将单词拆分为字符,通过预训练模型将每个字符映射为向量;然后使用卷积神经网络和池化层获得这个单词的字符级特征 CE。

CE 与 RPE 的长度相同,但维度不同。本文采用对位拼接的方式,依次拼接输入序列中每个词的 RPE 和 CE。拼接后每个词的词向量包含了上下文信息、相对位置信息和字符级语义信息,将其作为预处理后的文本特征传入位置预测网络,用于模型预测因果实体的起始位置。

## 4 位置预测网络

传统的序列标注方法需要预测因果实体所有字符的标签,并检查标签顺序的合法性,导致计算过程较为复杂。一段文本由开始位置和结束位置唯一确定,因此本文对因果实体的起始位置进行预测,可以显著降低抽取过程的复杂度。

### 4.1 预测因果实体位置

位置预测网络分为上下两个并行通道:Start Net 和 End Net。它们的网络结构相同,分别负责预测因果实体的开始位置和结束位置。

位置预测网络的结构如图3所示,上半部分为 Start Net,下半部分为 End Net。

位置预测共分为两步进行。

(1)将嵌入模块构造的文本特征传入 Start Net,首先通过卷积神经网络对文本特征进行降维,然后通过 SoftMax 层计算序列中每个位置的分类损失,进行二分类操作,判断输入序列中每个位置是否为因果实体的开始位置。图中正样本用深灰色表示,负样本用白色表示。End Net 的计算方法与 Start Net 完全相同且计算过程并行进行,区别在于 End Net

判断输入序列中每个位置是否为因果实体的结束位置。

(2)将 Step(1)的输出结果传入卷积神经网络,然后通过 SoftMax 层计算分类损失,进行二分类操作,判断 Step(1)预

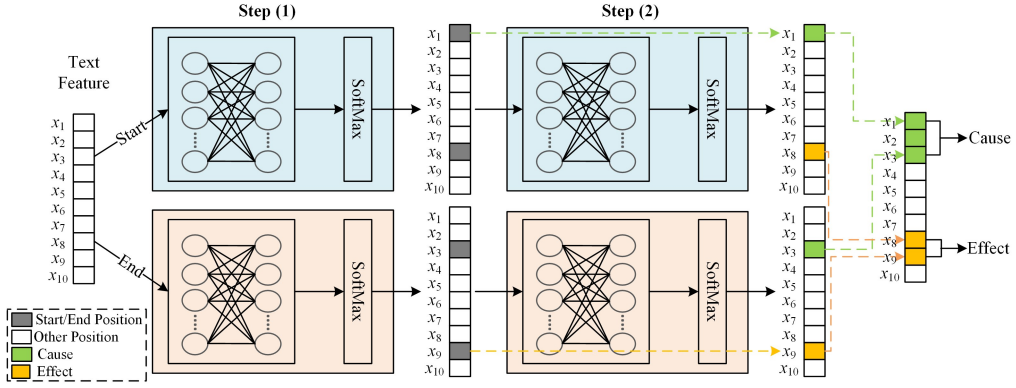


图3 位置预测网络(电子版为彩图)

Fig. 3 Position prediction network

## 4.2 组装因果实体

在模型获取 Start Net 和 End Net 的预测位置后,需要组装出正确的因果实体。首先将预测位置传入 Assembler 模块,检查错误并删除多余的位置;然后根据组装算法寻找距离最近的起始位置进行配对,组装出全部因果实体;最后将因果实体交叉组合为因果实体对,作为模型的抽取结果。算法 1 给出了组装模块的具体描述。

### 算法 1 组装模块因果实体组装算法

输入:因果起始位置;cause\_startPos, effect\_startPos; cause\_endPos, effect\_endPos

输出:因果关系对 causalityPairs

1. calculate extra position number Num<sub>c</sub>, Num<sub>e</sub>
2. fetch score of each start/end position Scores<sub>c</sub>, Scores<sub>e</sub>
3. get larger\_collection between cause\_startPos and cause\_endPos
4. generate to\_del\_list=[td<sub>1</sub>, td<sub>2</sub>, ...]
5. sort(to\_del\_list, Scores<sub>c</sub>, ascend)
6. for pos into\_del\_list:
7. if(del pos make residual pos matched){
8. delete pos from larger\_collection
9. break }
10. for cs in cause\_startPos:
11. for ce in cause\_endPos:
12. find the nearest paired position:{cs, ce}
13. all\_causes.add({cs, ce})
14. use the same way to getall\_effects
15. causalityPairs=Assemble()

在算法 1 中,首先获取基本参数。计算起始位置的数量差值 Num<sub>c</sub> 和 Num<sub>e</sub>,用于确定待删除的位置数量,获取位置预测网络输出的位置分类损失 Scores<sub>c</sub> 和 Scores<sub>e</sub> (第 1 行、第 2 行)。

然后,组装出所有的原因实体。确定待删除的是开始位置还是结束位置,根据 Num<sub>c</sub> 构造待删除位置组合列表,每个位置组合按照 Scores<sub>c</sub> 计算平均 SoftMax 得分,并将列表按总分升序排序(第 3-5 行)。采取低分淘汰原则,依次从待删除位置组合列表中取出一个位置组合,判断删除后是否可以用

测的起始位置属于原因实体还是结果实体。图中,原因实体使用绿色表示,结果实体使用橙色表示。这一步骤中,Start Net 和 End Net 仅区分原因和结果实体,操作完全相同。

剩余位置组装出合法的因果实体,如果可以,则删除这个位置组合(第 6-9 行)。采取就近匹配原则,每个开始位置依次寻找最近的结束位置,组装出原因实体(第 10-13 行)。结果实体使用相同的算法得到(第 14 行)。

最后,通过 Assemble 函数交叉组合出所有的因果关系对,并输出抽取结果(第 15 行)。

## 5 损失优化

**定义 1(CTS 系数)** 给定两段文本  $t_1$  和  $t_2$ , CTS(Causal Text Similarity)系数定义为  $t_1$  与  $t_2$  内容相似度和语义相似度的加权和。 $CTS_{(t_1, t_2)} \in [0, 1]$ 用于描述两段文本之间的因果关系匹配程度,计算式见式(2)-式(4)。

$$TIOU_{(t_1, t_2)} = \frac{t_1 \cap t_2}{t_1 \cup t_2} \quad (2)$$

$$TSS_{(t_1, t_2)} = \text{CosineSimilarity} \left( \frac{\sum_{i=1}^{n_{t_1}} x_{1i}}{n_{t_1}}, \frac{\sum_{j=1}^{n_{t_2}} x_{2j}}{n_{t_2}} \right) \quad (3)$$

$$CTS_{(t_1, t_2)} = \begin{cases} \omega_1 * TIOU_{(t_1, t_2)} + \omega_2 * TSS_{(t_1, t_2)}, & TIOU_{(t_1, t_2)} \geq 0.7 \\ 0, & TIOU_{(t_1, t_2)} < 0.7 \end{cases} \quad (4)$$

其中,  $t_1$  是预测文本,  $t_2$  是真实文本,  $TIOU_{(t_1, t_2)}$  表示  $t_1$  和  $t_2$  两段文本的内容相似度,  $TSS_{(t_1, t_2)}$  用余弦相似度表示  $t_1$  和  $t_2$  的语义相似度,  $n_{t_1}$  是  $t_1$  的长度,  $x_{1i}$  是  $t_1$  第  $i$  个单词,  $n_{t_2}$  是  $t_2$  的长度,  $x_{2j}$  是  $t_2$  第  $j$  个单词,  $\omega_1$  和  $\omega_2$  分别是 TIOU 和 TSS 的权重系数。

给定两段文本  $t_1$  = “financial crisis”,  $t_2$  = “financial crisis broke out”, 假设  $TSS_{(t_1, t_2)} = 0.7$ ,  $\omega_1 = 0.7$ ,  $\omega_2 = 0.3$ , 则可计算得出  $CTS_{(t_1, t_2)} = 0.56$ 。

与 Jaccard 系数相比, CTS 系数同时考虑文本之间的内容相似度和语义相似度。在计算模型预测损失时,通过为内容匹配得分和语义匹配得分设定适当的权重,可以较为准确地评估抽取结果,更加适合因果抽取任务。

**定义 2(惩罚窗口)** 规定以因果实体的起始位置为中心的周围区域为惩罚窗口(Penalty Window),惩罚窗口的左右

范围大小是动态的,预测位置是否位于窗口内部由预测文本和真实文本的CTS系数所决定。在本文中,CTS阈值被设定为0.5。

如图4所示,绿色为真实文本,红色虚线框代表惩罚窗口。橙色和粉色是由两个惩罚窗口内的预测位置组成的两个预测文本,它们与真实文本的CTS系数均大于或等于0.5。显然,由定义1可知,从两个真实位置的惩罚窗口内任意取出的位置组成的预测文本,与真实文本均有一半以上的因果关系匹配度。

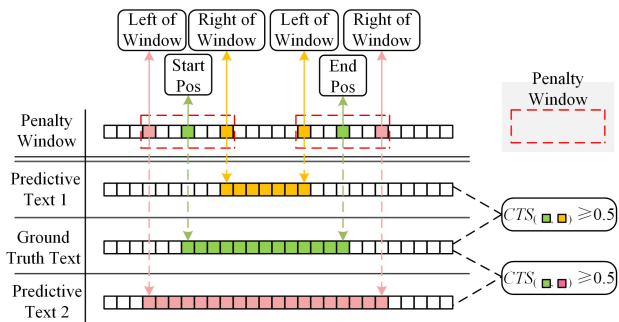


图4 惩罚窗口(电子版为彩图)

Fig. 4 Penalty window

对句子预处理后的文本特征执行二分类时,因果实体的起始位置是正样本,标签为1;其余位置都是负样本,标签为0。在位置预测网络的训练过程中,当预测起始位置 $y_{pi}$ 与真实起始位置 $y_i$ 接近时,预测标签为0。抽取的因果实体虽然不完全正确,但仍然包含了大部分正确的内容。本文在负样本被错误预测为正样本时,根据预测位置的不同采取不同的惩罚策略对预测损失进行调整。具体而言,预测位置越靠近真实位置,模型此次预测的损失越小。

模型损失 $Loss = Loss_{cls} + Loss_{pred}$ 。其中, $Loss_{cls}$ 是对原因和结果分类时产生的损失,使用交叉熵损失函数计算得到,如式(5)所示,其中 $N$ 为样本数量, $n$ 为第 $i$ 个样本中待分类位置个数, $y_{ij}$ 为第 $i$ 个样本的SoftMax得分, $y_{pij}$ 为原始数据中样本预测为正的的概率。 $Loss_{pred}$ 是对起始位置和其余位置分类时产生的损失,式(6)对其进行了优化。

$$Loss_{cls} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{n_i} [(1-y_{ij}) \log(1-y_{pij}) + y_{ij} \log y_{pij}] \quad (5)$$

$$Loss_{pred} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{n_i} \begin{cases} \log y_{pij}, & y_{ij} = 1 \\ \log(1-y_{pij}), & y_{ij} = 0 \text{ (not in penalty window)} \\ 1 - e^{-\text{offset}/3}, & y_{ij} = 0 \text{ (in penalty window)} \end{cases} \quad (6)$$

其中, $N$ 为样本数量, $n_i$ 为第 $i$ 个样本的长度, $y_{ij}$ 为第 $i$ 个样本在第 $j$ 个位置的真正标签分数, $y_{pij}$ 为第 $i$ 个样本在第 $j$ 个位置的预测标签分数, $offset$ 为假正例与正例的相对距离。

基于定义1和定义2,本文使用CTS系数作为阈值来确定惩罚窗口范围。当预测位置为假正例且位于惩罚窗口内部时,通过修改预测位置的预测损失,使假正例的惩罚项随着靠近真实位置而减小。

## 6 实验及结果分析

### 6.1 实验数据

本文使用 SemEval-2010 Task8<sup>1)</sup> 和 SemEval-2020 Task5<sup>2)</sup> 的数据作为基础数据集,FNP2021 Share Task2<sup>3)</sup> 中含有因果关系的文本作为扩充数据集,每条样本中含有一个或多个因果实体对,数据统计结果如表1所列。

表1 数据统计  
Table 1 Statistics of data

Dataset	Number of Example
SemEval-2010 Task 8	1 003
SemEval-2020 Task 5	3 021
FNP 2021	998
Total	5 022

### 6.2 评价指标

本文采用的实验评价指标包括:准确率(Precision)、召回率(Recall)、F1和抽取速度(Speed)。

$$Precision = \frac{CEC}{EC} \quad (7)$$

$$Recall = \frac{CEC}{AEC} \quad (8)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

其中, $CEC$ 代表抽取出的因果关系对中正确的数量, $EC$ 代表抽取出的因果关系对, $AEC$ 代表数据集中总共包含的因果关系对数。本文统计的抽取速度 $Speed$ 指在相同的机器配置下,从一个样本中抽取因果关系所消耗的时间。

### 6.3 结果与分析

本文基于相同的硬件设施和环境配置,将 PosNet 与一些基准模型进行了对比实验。各模型所使用的词向量相同,均由 Albert 模型获取。实验统计结果如表2所列。

表2 实验结果统计

Table 2 Statistics of experimental results

Model	Precision	Recall	F1	Time/ms
Rules+Bayesian	0.6042	0.5878	0.5959	—
CausalNet	0.6211	0.5372	0.5761	—
LSTM	0.6019	0.6730	0.6355	282
LSTM+CRF	0.6407	0.7012	0.6696	310
BiLSTM	0.7744	0.7622	0.7682	367
PosNet(E)	<b>0.7513</b>	<b>0.7327</b>	<b>0.7419</b>	<b>146</b>
PosNet(E+CE)	<b>0.7927</b>	<b>0.7750</b>	<b>0.7838</b>	<b>175</b>
PosNet(E+CE+RPE)	<b>0.8111</b>	<b>0.8071</b>	<b>0.8208</b>	<b>187</b>

通过对比实验结果,可以得出以下结论。

(1) PosNet 超过了一般基准模型的准确率。Rules+Bayesian 和 CausalNet 是基于模式匹配的基准模型,在因果

<sup>1)</sup> <https://www.kaggle.com/drtoshi/semeval2010-task-8-dataset>

<sup>2)</sup> <https://zenodo.org/record/3932442#.YPeLcvkzaUk>

<sup>3)</sup> <http://wp.lancs.ac.uk/cf7ie/fincausal2021/>

抽取任务上达到了一定的准确率,但模型泛化性较差。LSTM, LSTM+CRF 和 BiLSTM 是基于机器学习的基准模型,其中 LSTM 经常被用于 NLP 任务,但单向 LSTM 受到方向限制,其准确率不高。LSTM+CRF 对文本标签的分类更加准确,相比 LSTM 提升了约 4% 的准确率。BiLSTM 不再受限于文本的处理方向,获得了相对完整的上下文语义特征,相比其他模型提升较为明显。

(2) PosNet 的抽取速度比基准模型提升了约一倍。PosNet 强调在保持一定准确性的基础上提高抽取效率,随着模型复杂性的提高,抽取效率必然会下降,因此本文没有和较为复杂的因果抽取网络进行对比实验。当准确率相近时,PosNet 的部署效率和运行效率都更高。

图 5 是训练过程中的模型损失,可以得出以下结论。

(1) 在使用惩罚窗口技术后,模型的参数优化路线更加高效。训练前期(前 20 轮),模型的准确率较低,交叉熵损失和惩罚窗口损失的计算结果相近。训练中期(20~50 轮),模型的准确率逐渐提升,使用交叉熵损失函数计算得到的损失值不利于模型的参数优化,而使用惩罚窗口计算的损失值更加符合因果抽取的任务特性。训练后期(50~80 轮),模型的准确率接近稳定,此时两种损失计算方式区别不大,均可得出较为精确的模型损失。

(2) 惩罚窗口越大,模型对假正例的容忍度越高。通过实验发现,对于本文的数据集而言,CTS 阈值取 0.5 时,模型的拟合曲线最佳。不同领域对假正例的容忍度可能不同,在应用到其他领域时,可以通过实验为惩罚窗口设置不同的 CTS 系数阈值,以达到最佳抽取效果。

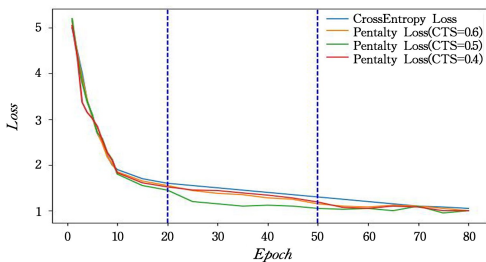


图 5 模型损失  
Fig. 5 Model loss

**结束语** 随着业务和用户规模的扩大,文本数据开始出现爆发式增长。因果关系抽取作为自然语言处理领域的重要任务,对抽取效率逐渐有了更高的要求。为了在保证抽取准确性的前提下提高因果抽取的效率,本文提出了基于位置的因果关系抽取网络 PosNet。网络构建了含有位置信息的文本特征,通过预测因果实体的起始位置实现因果关系的抽取。研究工作的主要贡献包括:

(1) 通过预测因果实体的关键位置,降低了因果抽取模型的复杂度。传统抽取任务需要对因果实体的每个词进行分析处理,导致模型结构较为复杂。本文通过向文本特征中添加位置信息,构造了含有位置信息的文本特征,提升了网络预测起始位置的准确性,根据起始位置抽取因果实体。

(2) 针对因果抽取任务,提出了更加高效的参数优化方案。针对预测结果近似正确这种情况,大多数模型会认为

模型预测结果错误。本文认为这种做法过于绝对,利用 CTS 系数和惩罚窗口技术,综合考虑文本之间的语义相似度和内容相似度,调整惩罚窗口内部的预测损失,实现模型参数的合理优化。

实验结果表明,PosNet 可以达到较高的抽取准确率,并提升抽取效率。在模型的训练过程中,本文提出的参数优化方案加快了模型的拟合速度。在未来,我们仍会进行一些与因果抽取相关的研究,例如嵌套因果关系抽取、篇章级因果关系抽取等。

## 参考文献

- [1] WANG Z J, WANG S, LI X Q, et al. Review of Event Causality Extraction Based on Deep Learning[J]. Journal of Computer Applications, 2021, 41(5): 1247-1255.
- [2] ITTOO A, BOUMA G. Extracting explicit and implicit causal relations from sparse, domain-specific texts[C]// International Conference on Application of Natural Language to Information Systems. Berlin: Springer, 2011: 52-63.
- [3] LUO Z, SHA Y, ZHU K Q, et al. Commonsense causal reasoning between short texts[C]// Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning. Palo Alto, CA: AAAI Press, 2016: 421-430.
- [4] ZHAO S, LIU T, ZHAO S, et al. Event causality extraction based on connectives analysis [J]. Neurocomputing, 2016, 173(3): 1943-1950.
- [5] SEOL J W, YI W, CHOI J, et al. Causality patterns and machine learning for the extraction of problem-action relations in discharge summaries[J]. International Journal of Medical Informatics, 2017, 98: 1-12.
- [6] LEE D G, SHIN H. Disease causality extraction based on lexical semantics and document-clause frequency from biomedical literature[J]. BMC Medical Informatics and Decision Making, 2017, 17(1): 1-9.
- [7] LEE S, SEO S, OH B, et al. Cross-sentence N-ary Relation Extraction using Entity Link and Discourse Relation[C]// Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM, 2020: 705-714.
- [8] FRATTINI J, JUNKER M, UNTERKALMSTEINER M, et al. Automatic extraction of cause-effect-relations from requirements artifacts[C]// Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering. NJ: IEEE, 2020: 561-572.
- [9] HEINDORF S, SCHOLTEN Y, WACHSMUTH H, et al. Causenet: Towards a causality graph extracted from the web [C]// Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM, 2020: 3023-3030.
- [10] KRUENCKRAI C, TORISAWA K, HASHIMOTO C, et al. Improving event causality recognition with multiple background knowledge sources using multi-column convolutional neural networks[C]// Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2017: 3466-3473.

- [11] ZHENG S,HAO Y,LU D,et al. Joint entity and relation extraction based on a hybrid neural network[J]. *Neurocomputing*, 2017,257:59-66.
- [12] ZHENG S,WANG F,BAO H,et al. Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme[C]// *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics(Volume 1: Long Papers)*. PA: ACL, 2017: 1227-1236.
- [13] ZENG X,ZENG D,HE S,et al. Extracting relational facts by an end-to-end neural model with copy mechanism[C]// *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics(Volume 1: Long Papers)*. PA: ACL, 2018: 506-514.
- [14] DASGUPTA T, SAHA R, DEY L, et al. Automatic extraction of causal relations from text using linguistically informed deep neural networks[C]// *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. PA: ACL, 2018:306-316.
- [15] LI P,MAO K. Knowledge-oriented convolutional neural network for causal relation extraction from natural language texts [J]. *Expert Systems with Applications*, 2019,115:512-523.
- [16] LI Z,LI Q,ZOU X,et al. Causality extraction based on self-attentive bilstm-crf with transferred embeddings[J]. *Neurocomputing*, 2021,423:207-219.
- [17] SAHU S K, THOMAS D, CHIU B, et al. Relation extraction with self-determined graph convolutional network[C]// *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. New York: ACM, 2020: 2205-2208.
- [18] ZHAO K,JI D,HE F,et al. Document-level event causality identification via graph inference mechanism[J]. *Information Sciences*, 2021,561:115-129.
- [19] CAO Y,CHEN D,XU Z,et al. Nested relation extraction with iterative neural network[J]. *Frontiers of Computer Science*, 2021,15(3):1-14.
- [20] JIAO F,LI H,DOBOLI A. Modeling and extraction of causal information in analog circuits[J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2017, 37(10):1915-1928.
- [21] KIM H, JOUNG J, KIM K. Semi-automatic extraction of technological causality from patents[J]. *Computers & Industrial Engineering*, 2018, 115:532-542.
- [22] MAISONNAVE M, DELBIANCO F, TOHMÉ F, et al. Assessing Causality Structures learned from Digital Text Media[C]// *Proceedings of the ACM Symposium on Document Engineering 2020*. New York: ACM, 2020: 1-4.
- [23] NASAR Z, JAFFRY S W, MALIK M K. Named Entity Recognition and Relation Extraction: State-of-the-Art[J]. *ACM Computing Surveys(CSUR)*, 2021, 54(1): 1-39.
- [24] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. *arXiv: 1301.3781*, 2013.
- [25] KENTON J D M W C, TOUTANOVA L K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]// *Proceedings of NAACL-HLT*. PA: ACL, 2019: 4171-4186.
- [26] LAN Z, CHEN M, GOODMAN S, et al. Albert: A lite bert for self-supervised learning of language representations[J]. *arXiv: 1909.11942*, 2019.
- [27] SHAW P, USZKOREIT J, VASWANI A. Self-Attention with Relative Position Representations[C]// *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. PA: ACL, 2018: 464-468.



**ZHU Guang-li**, born in 1971, master, associate professor, master supervisor. Her main research interests include Web mining, semantic search, and calculation theory.

(责任编辑:柯颖)