

# 知识与句法融合的因果关系抽取网络

汪诗蕊, 解博涵, 丁玲, 陈建廷, 向阳  
同济大学电子与信息工程学院, 上海 200000

## 摘要

因果关系抽取作为关系抽取的一个重要任务, 近年来得到了广泛关注。现有的因果关系抽取方法大多将句法结构和背景知识割裂开进行研究, 早期的因果关系抽取方法偏重于从句法结构层面进行分析, 随着深度学习技术的发展, 预训练模型结合背景知识的方法成为主流。然而上述两种方法均未完全融合句内信息和外部知识, 带来了不同程度的信息缺失。为了解决这一问题, 提出了结合句法结构和背景知识的因果关系抽取模型。该模型将句子解析为同时包含句法和知识的知识句法图结构, 使用图卷积网络进行信息融合。模型同时考虑了句法和知识两部分信息, 从而进一步丰富了实体嵌入, 达到了良好的因果关系抽取效果。本模型在EventStoryLine数据集上取得了良好效果, F1值达到0.445, 与现有方法相比提高了2.3%。

## 关键词

因果关系抽取; 预训练模型; 图卷积网络; 自然语言处理

中图分类号: TP391.1

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2024008

## *Event causality identification network based on knowledge and syntactic structure*

WANG Shirui, XIE Bohan, DING Ling, CHEN Jianting, XIANG Yang

School of Electronic and Information Engineering, Tongji University, Shanghai 200000, China

## *Abstract*

Event causality identification is an important task of relationship extraction, which has received much attention recent years. Most of the existing methods separate syntactic structure from the background knowledge information. The early causality extraction methods focus on the analysis of syntactic structure level. With the development of deep learning, the methods that use the pre-training model combined with background knowledge has become the mainstream. However, neither of the above two kinds of methods fully integrates the sentence information and external knowledge, resulting in different degrees of information loss. To address this problem, we proposed a novel model of event causality identification combining syntactic structure and background knowledge. Our model parses sentences into knowledge syntactic graph structures that contain both syntax and knowledge, and uses the graph convolution network for information fusion. It considers both syntax and knowledge information, which further enriches the event representation and performs effectively. In experiments on the widely-used dataset EventStoryLine, the F1 score of our model achieves 0.445, a 2.3% improvement over existing methods.

## *Key words*

event causality identification, syntactic structure, graph convolution network, natural language processing

## 0 引言

人类社会的知识信息大多以非结构化的形式存储在文本中,如新闻报道、科技文献等。关系抽取的主要任务就是从半结构化或非结构化数据中抽取两个或多个实体的语义关系,其中一个重要的子任务就是因果关系抽取(event causality identification, ECI)。该任务旨在判断自然语言文本中两个实体之间是否存在因果关系。常见的自然语言处理应用,如问答系统<sup>[1-2]</sup>、场景预测<sup>[3]</sup>、机器阅读理解<sup>[4]</sup>等,其中均涉及因果关系抽取,该方法改进有助于提升下游任务效果,具有重要的现实意义。

因果关系抽取任务通常被建模为分类问题,即给定自然语言文本中的两个或多个实体,通过模型学习推理出实体间存在因果关系的概率。ECI任务已经提出了多种方法。最初的基于模式匹配的抽取方法依赖于专家制定的规则模式,判断实体是否符合规则来实现因果关系抽取。随后的基于特征分类的抽取方法需要人工提取特征输入分类算法,常见的特征种类有词典特征、文档特征等。上述两类早期方法存在的问题是需要大量的专业知识和人工劳动,无论是规则制定,还是特征选择,均需要专业人士的大量手工工作,不适宜处理海量数据的情况,并且难以保持稳定的良好效果。

随着深度学习的广泛应用,基于深度学习的抽取方法成为主流。该类方法实现了端到端地进行因果关系抽取,适用于大量数据的场景。其主要思想是获取更完善的实体嵌入,从而达到更好的分类效果。特别是知识图谱出现以后,预训练模型结合背景知识添加进一步丰富了实体嵌入,取得了良好的分类效果。但这类方法大多抛弃了早期方法对句法结构的关注,将句

子按顺序输入预训练模型。事实上,研究表明句法结构特别是句中依存关系对关系抽取类任务有极大的帮助。而放弃关注句法结构会遗失部分实体特征表示,从而降低模型的分类效果。

综上所述,为了同时考虑句法结构和背景知识信息,本文提出了一种融合依存句法和外部知识的因果关系抽取模型。该模型通过生成句子的知识句法融合图(以下简称知识句法图)组织句内句外信息,使用图卷积神经网络进行信息融合丰富实体嵌入,从而实现更精确的因果关系抽取。为构建知识句法图,首先使用依存分析解析句内结构,生成包含词语间依赖关系的依存句法树;之后引入外部知识,添加实体的背景描述信息,形成最终的包含句法、知识两方面的知识句法图结构;最后使用图卷积神经网络进行信息融合,同时采用句子依存结构和背景知识两方面的特征,得到最终的实体嵌入向量用于分类。

本文的主要贡献如下:

- 将句法结构和背景知识同时引入因果关系抽取任务,进一步丰富实体嵌入,构建出一个融合句法结构、背景知识的因果关系抽取模型;

- 提出将句子解析为知识句法图的形式,合理组织句内句外信息,并通过图卷积神经网络进行信息融合;

- 本文在EventStoryLine数据集上进行实验,与现有基线模型相比,F1值提高了2.3%,同时进行了消融实验分析,验证依存分析和知识添加对因果关系抽取任务的有效性。

## 1 相关工作

因果关系抽取是自然语言处理(natural language processing, NLP)中的一

种关系抽取任务,旨在挖掘文本中具有因果关系的实体对,在金融、安全、生物等领域的应用中发挥了重要作用。因果关系抽取方法主要可分为3类:基于模式匹配的方法、基于机器学习的方法和基于深度学习的方法。

### 1.1 基于模式匹配的方法

早期基于模式匹配的方法通常利用语义结构、词汇符号特征和自构约束进行因果关系的提取。Khoa等人<sup>[5]</sup>利用语言线索和模式匹配从华尔街日报中提取因果知识;Girju等人<sup>[6]</sup>使用句法模式提取因果关系,之后使用语义结构约束将候选对分类为因果对或非因果对。但这些完全依赖于规则模式匹配的方法通常有耗费大量人力、跨域适应性差等问题。

### 1.2 基于机器学习的方法

基于机器学习技术的方法主要以流水线的方式解决因果关系的抽取任务,将因果关系抽取分为两个子任务:候选因果对提取和关系分类。首先根据模板或一些线索词提取可能具有因果关系的候选实体对,然后根据语义特征和语法特征对候选因果对进行分类筛选出因果对。Girju等人<sup>[7]</sup>使用基于因果关系的约束触发词来提取英语文本中的因果关系;Luo等人<sup>[8]</sup>从大规模网络文本语料库中提取因果关系术语,然后使用因果线索来度量网络文本语料之间的因果强度。但是该类方法同样面临着泛化能力差等问题,同时难以识别出文本中隐含模糊的因果关系。

### 1.3 基于深度学习的方法

神经网络强大的表征学习能力,

有助于解决因果关系抽取中的隐含性和模糊性问题,近年来采用深度学习技术提取因果关系已渐渐成为主流。De Silva等人<sup>[9]</sup>使用卷积神经网络(convolution neural network, CNN)对文本中的因果关系进行分类;Kruengkrai等人<sup>[10]</sup>使用CNN从嘈杂的文本中提取背景知识来分类常识性的因果关系;Li等人<sup>[11]</sup>提出了一种面向知识的CNN,它结合了词汇知识库中的先验知识进行因果关系分类;Dasgupta等人<sup>[12]</sup>还通过基于长短期记忆(long short-term memory, LSTM)网络的深层模型,从语言的角度确定了文本中因果关系的语言表达。自2017年Transformer被提出以来,预训练模型在自然语言处理领域的应用越来越广泛,2019年刘健等人<sup>[13]</sup>提出了一种知识增强的因果关系抽取模型,将知识图谱信息引入关系抽取任务,丰富了实体嵌入,同时设置掩码提升模型的泛化效果。但在预训练模型广泛使用的同时,该类模型也抛弃了一部分句法结构信息。以Bert预训练模型为例,该模型编码时将句子顺序输入,在包含上下文语义的同时,消解了词语间的句法关系。然而许多研究证明,句法结构信息特别是依存句法分析,对关系抽取类任务具有明显作用<sup>[14-17]</sup>,因为句法结构提供了潜在长距离有用单词之间的连接,从而可以帮助模型更好地提取实体对之间的关系。此外,在背景知识的添加方面仍存在一些问题。2019年刘健等人<sup>[13]</sup>将知识直接嵌入句中,在一定程度上对句子结构造成破坏。

针对上述问题,本文在以上研究的基础上,提出了构建知识句法图进行因果关系抽取的方法。首先提取出句子的依存关系,将背景知识作为图的一部分进行添加,之后使用图卷积神经网络进行信息融合。在保留句子结构的同时,考虑了背景知识对实体嵌入的影响。

## 2 模型架构

本文提出的方法模型的总体框架如图1所示。本模型可分为以下3个步骤：首先构建知识句法图，将句子解析为依存树结构，并在知识图谱中检索背景知识进行添加，在保留句子原本句法关系的前提下进行知识增强，由Bert编码得到图中节点的特征向量；之后进行特征聚合，使用图卷积网络提取节点特征和边特征，之后聚合特征，多层堆叠卷积得到包含实体间关系信息的嵌入向量；最后进行分类预测，将聚合之后的嵌入向量输入全连接层，得到实体间存在因果关系的概率。

### 2.1 知识句法图构建

本模型的输入是语句 $s$ 和需要判断是否具有因果关系的实体对 $\langle e_1, e_2 \rangle$ 。模型首先生成语句 $s$ 关于实体 $e_1$ 、 $e_2$ 的知识句法图。

依存句法分析有助于模型从语句结构层面理解语义，依存分析通过找到句子中每一个词语所依赖的部分来描述句子结构。依存句法树是知识句法图的骨架，使用依存关系分析将句子解析为树结构，得到知识句法图的基础框架，具体操作可见式(1)、式(2)、式(3)。

$$T(V_T, E_T) = \text{dsp}(s) \quad (1)$$

$$V_e = V_T = [\text{node}_1, \dots, \text{node}_n] \quad (2)$$

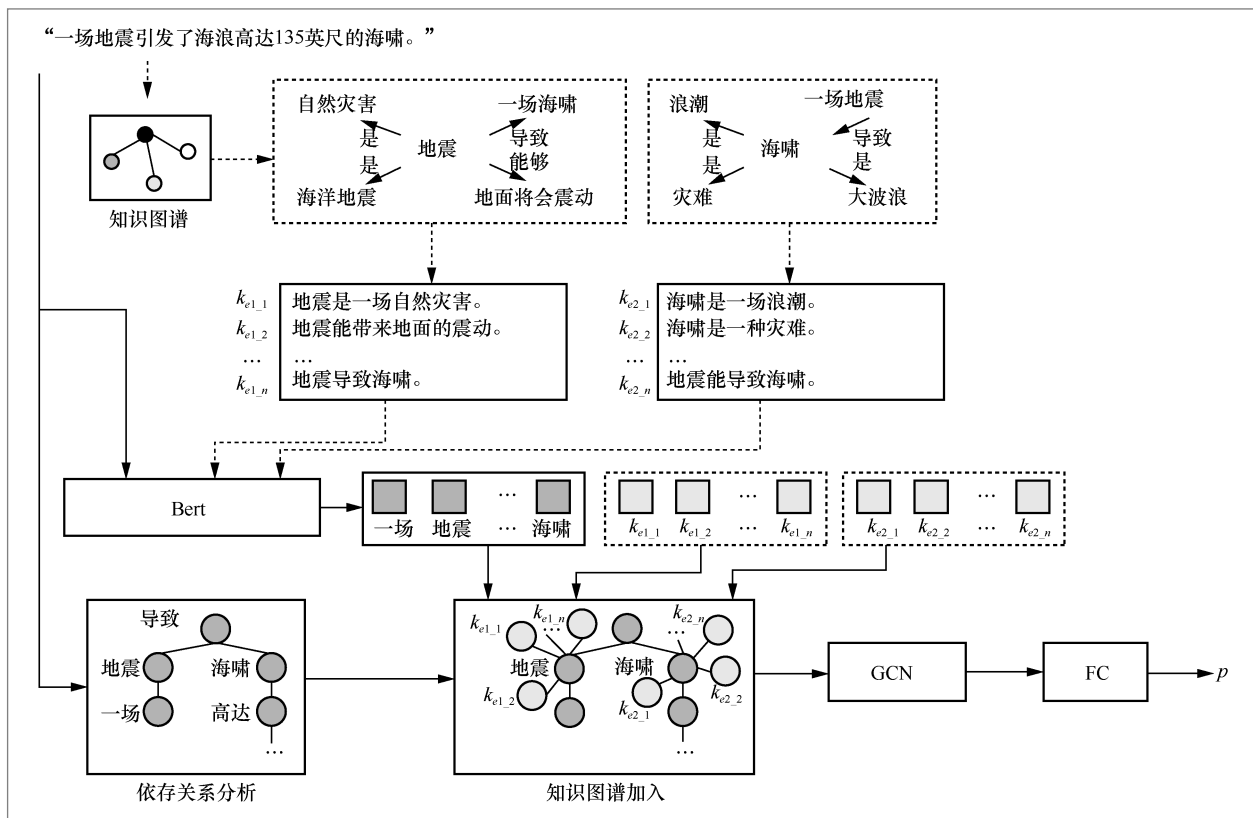


图1 模型总体架构

$$E_e = E_T = [\text{edge}_1, \dots, \text{edge}_{n-1}] \quad (3)$$

其中,  $T(V_T, E_T)$  为语句  $s$  经过依存关系分析 (dependency syntactic parsing, DSP) 得到的依存树,  $n$  为节点个数。其中  $V_T$ 、 $E_T$  分别为依存树的节点和边。由于依存关系树中仅有一个特殊节点没有入边, 其余节点有且仅有一条入边, 故句法边的边数为  $n-1$ 。 $V_e$  为知识句法图中的实体节点列表,  $E_e$  为知识句法图的句法边列表。同时可得句法结构图的邻接矩阵  $A'_{n \times n}$ , 定义为:

$$A'_{i,j} = \begin{cases} 1, & \text{依存树中节点 } \text{node}_{i+1} \text{ 和} \\ & \text{node}_{j+1} \text{ 之间存在边} \\ 0, & \text{其他} \end{cases} \quad (4)$$

$i \in [0, n-1], j \in [0, n-1]$

以“一场地震引发了海浪高达135英尺的海啸”为例, 模型构建的句法如图2所示。

2018年, Devlin等提出了BERT模型。该模型是一类大规模预训练语言模型, 基于多层Transformer的双向编码, 在对一个词进行编码的同时考虑到前文和后文的双向信息, 解决了Word2Vec等词向量表示方法无法解决“一词多义”的问题, 语义获取更加充分。因此本文采用预先训练的BERT模型作为词向量嵌入层来编码词语

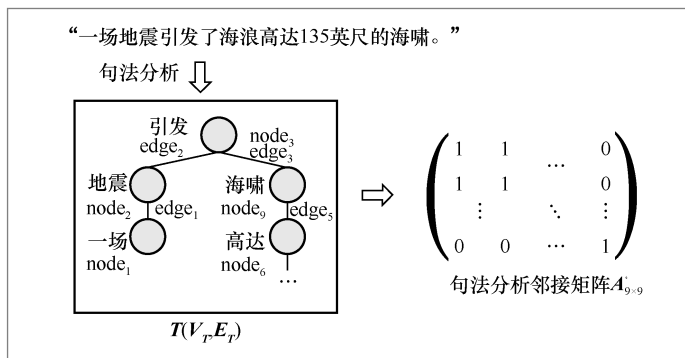


图2 句法分析示例

节点, 获得节点的特征向量。

$$h_0^e = W_t^e + W_s^e + W_p^e \quad (5)$$

$$h_{i+1}^e = \text{Transformer}(h_i^e), i \in [0, N-1] \quad (6)$$

$$F_e = W_e \cdot h_N^e \quad (7)$$

其中,  $h_i^e$  为第  $i$  层 Transformer 的输出结果,  $N$  为 Transformer 的层数。 $h_0^e$  为BERT预训练模型的初始输入, 由  $W_t^v, W_s^v, W_p^v$  相加得到, 三者分别表示语句经过Token Embedding、Segment Embedding、Position Embedding得到的编码矩阵。 $W_e$  为BERT模型中token的编码和依存分析得到的实体节点之间的映射矩阵。得到BERT编码的最终输出后, 需要将其中token的编码和实体节点进行映射, 才能得到节点编码矩阵  $F_e$ 。

背景知识作为句法结构的补充, 为查询时间对添加背景知识信息, 其包括知识节点和知识边两部分。

对于给定的查询实体对  $\langle e_1, e_2 \rangle$ , 在知识图谱CONCEPTNET中进行检索得到该实体背景知识, 用三元组 (head, relation, tail) 表示。由于查询到的三元组可能非常多, 本文对关系进行筛选, 仅保留18个特定关系的三元组: CapableOf、IsA、HasProperty、Causes、MannerOf、Causes-Desire、Used-For、HasSubevent、HasPrerequisite、Not-Desires、PartOf、HasA、Entails、Receive-sAction、UsedFor、CreatedBy、MadeOf和Desires。

本文提取三元组中tail节点作为知识节点, 将其和实体节点列表拼接得到知识句法图的节点列表。具体如式(8)所示, 其中  $V$  为知识句法图的节点列表,  $m(r)$  为实体  $e_1$  ( $e_2$ ) 经过筛选后保留的三元组数量。

$$V = [\text{node}_1, \dots, \text{node}_n, \text{node}_{n+1}, \dots, \text{node}_{n+m}, \text{node}_{n+m+1}, \dots, \text{node}_{n+m+r}] \quad (8)$$

知识边连接实体节点和依据该实体进行检索得到的知识节点,即上述三元组(head, relation, tail)中的relation部分。进行知识添加后,知识句法图的邻接矩阵 $A_{(n+m+r) \times (n+m+r)}$ 定义如式(9)所示,其中 $i \in [0, n+m+r-1], j \in [0, n+m+r-1]$ 。

$$A_{i,j} = \begin{cases} 1, & i \leq n, j \leq n, A_{i,j} = 1 \text{ 或节点 } i+1 \text{ 和} \\ & \text{节点 } j+1 \text{ 间存在知识边} \\ 0, & \text{其他} \end{cases} \quad (9)$$

本文仍然使用BERT预训练模型对知识节点进行编码。

$$h_0^k = W_t^k + W_s^k + W_p^k \quad (10)$$

$$h_{j+1}^k = \text{Transformer}(h_j^k) \quad (11)$$

$i \in [0, m+r-1], j \in [0, N-1]$

$$F_k = \text{concat}([W_{k_0} \cdot h_{k_0}^k, \dots, W_{k_{m+r-1}} \cdot h_{k_{m+r-1}}^k]) \quad (12)$$

其中, $h_j^k$ 为三元组 $k_i$ 第 $j$ 层Transformer的输出结果, $N$ 为Transformer的层数。 $h_0^k$ 为该三元组BERT预训练模型的初始输入, $W_t^k$ 、 $W_s^k$ 、 $W_p^k$ 三者分别表示三元组 $k_i$ 经过Token Embedding、Segment Embedding、Position Embedding得到的编码矩阵。 $W_{k_i}$ 为依据三元组 $k_i$ 的编码矩阵得到tail节点向量的映射矩阵。

此时已经完成了知识句法图构建的全过程,将 $F_e$ 和 $F_k$ 进行拼接,得到最终的知识句法图节点特征向量矩阵 $F$ 。

$$F = \text{concat}(F_e, F_k) \quad (13)$$

## 2.2 图卷积获取嵌入

对待非结构化的图数据,使用图卷积方法可以更好地实现节点特征聚合,其

通过聚合邻居节点和连接的边的信息可以得到中心节点的更准确的表达。本节介绍对前文产生的知识依存图进行图卷积运算,从而获得信息融合后的实体嵌入表示。

本文中图卷积操作的分层传播式为:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}} \bar{A} \tilde{D}^{-\frac{1}{2}} H^l W^l\right), l \in [0, L-1] \quad (14)$$

其中, $\bar{A} = A + I$ , $A$ 为前文生成的知识依存图的邻接矩阵。由于聚合后的节点特征和该节点自身的特征向量联系密切,添加自环机制,在邻接矩阵的基础上考虑了节点与自身的联系。 $L$ 为卷积层数, $H^l$ 为第 $l$ 层的输出矩阵, $\tilde{D}$ 为 $A$ 的归一化矩阵, $W^l$ 为模型可学习的参数。 $H^0$ 为节点的特征矩阵,依前文定义有 $H^0 = F$ 。

使用图卷积网络更新后的节点表示不仅包含了BERT中的上下文信息,还融合了背景知识和句法结构信息。使用实体对 $\langle e_1, e_2 \rangle$ 的实体映射矩阵 $W_{e_1, e_2}$ ,从更新后的节点特征向量中提取出实体 $e_1, e_2$ 的融合特征向量,用 $F_{KD}(e_1)$ 和 $F_{KD}(e_2)$ 表示。

$$[F_{KD}(e_1), F_{KD}(e_2)] = W_{e_1, e_2} \cdot H^L \quad (15)$$

## 2.3 分类预测

对于上述实体对 $\langle e_1, e_2 \rangle$ 通过图卷积网络得到的特征表示 $F_{KD}(e_1)$ 、 $F_{KD}(e_2)$ ,使用全连接层进行特征分类,得到 $e_1, e_2$ 之间存在因果关系的概率 $p$ 。

$$p = \text{fc}(F_{KD}(e_1), F_{KD}(e_2)) \quad (16)$$

其中,fc表示全连接函数。

由于数据存在严重的不平衡性,为了提升模型对少数正样本的关注,使用加权损失函数进行训练评估。损失函数定义

如下:

$$L = -[k*y*\log p + (1-y)\log(1-p)] \quad (17)$$

其中,  $y$ 代表语句 $s$ 中实体对 $\langle e_1, e_2 \rangle$ 的真实标签值,  $k$ 是加权参数, 反映正样本相较于负样本的重要性。

### 3 实验

#### 3.1 实验设置

本文使用Caselli和Vossen构建的EventStoryline数据集<sup>[18]</sup>进行模型测试。该数据集是现有的实体因果关系抽取方面数据量最大的数据集。其中包含22个主题的258个文档, 共5 334个实体, 7 805对实体对中有1 770对是因果相关的。本文按照主题对数据进行训练测试划分, 为了充分评估模型效果, 使用五折交叉验证进行评估。

评价指标方面, 本文采用Precision (P)、Recall (R)和F1-score (F1)作为评价指标, 与以往的方法相同以确保可比性, 其中F1为主要的评价指标。评价指标计算式具体如式(18)、式(19)和式(20)所示:

$$P = \frac{TP}{TP + FP} \quad (18)$$

$$R = \frac{TP}{TP + FN} \quad (19)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (20)$$

本模型使用斯坦福大学开发的Stanford Parser对句子进行依存分析, 生成句子的依存树。使用CONCEPTNET5.0作为背景知识图谱。预训练模型使用BERT-BASE, 其包括12层Transformer结构。超参数选择方面, 选取batch size=64,

learning\_rate=0.0001, BERT微调learning\_rate=0.00005, 以线性速率0.01进行衰减, 损失函数权重 $k=2$ 。训练过程使用Adam优化器进行优化。

#### 3.2 实验结果

本文提出的方法在EventStoryline数据集上进行测试, 同时与一些现有方法进行了对比。其中, DP方法指2017年Cheng等人<sup>[19]</sup>在序列建模的过程中引入依赖路径的方法, 使用双向长短期记忆网络(BiLSTM)进行预测; LSTM方法指2017年Choubey等人<sup>[20]</sup>提出的一种新的序列建模方法, 该模型使用上下文序列、词性特征和词语依赖关系作为LSTM的输入, 进而预测包含因果关系的概率; ILR方法指2019年Gao等人<sup>[21]</sup>提出的使用整数线性规划对因果关系建模的方法, 该方法在文档级因果关系抽取任务上取得了良好效果; MMG方法指2020年Liu等人<sup>[13]</sup>提出的使用知识增强的因果关系抽取方法, 同时加入掩码遮蔽机制提升模型泛化能力。此外, 本文还实现了仅使用BERT模型进行编码的因果关系抽取模型用于结果对比。实验结果见表1。

从表1可以看出, 相较于传统的未添加外部知识的因果关系抽取方法, 添加知识的因果关系抽取模型表现出了更好的效果。本文模型和DP、LSTM、ILR、BERT相比, 实验F1指标分别提升了19.0%、17.7%、9.3%、9.1%。外部知识的引入添加了实体的背景信息, 丰富了实体特征表示, 有助于更加精准预测文本中实体间的因果概率。

在均添加了外部知识的前提下, 本文模型和MMG相比, F1指标提升了2.3%。MMG模型简单地将外部知识信息插入句中, 破坏了文本结构, 造成句法信息的损

失。而本文模型通过依存句法分析将句子解析为图结构，再以添加节点和边的方式补充背景信息，在保留句法结构信息的基础上使用外部知识丰富实体表征，取得了良好效果。

此外，通过对比发现，使用BERT预训练模型进行编码的方法效果均好于前期使用word2vec编码的方法，主要原因是BERT具有强大的上下文表征能力，在对一个词进行编码的同时考虑到了前文和后文的双向信息，语义获取更加充分。

### 3.3 消融实验

为了进一步证明模型中各个模块的有效性，本文在EventStoryline数据集上进行了消融实验，对模型的各个模块进行测试分析。-dsp表示不添加依存树结构，仅使用背景知识构建知识图；-kl表示不添加背景知识信息。消融实验的结果见表2，去掉各模块之后，模型的性能结果均呈现不同程度的下降，证明了本模型中各模块的有效性。

## 4 结束语

现有的因果关系抽取方法将依存句法和背景知识割裂研究，本文提出了构建知识句法图进行因果关系抽取的方法，将两者结合丰富实体嵌入，从而获得更好的模型效果。本模型首先解析句子的依存树结构，并将背景知识作为图的一部分进行添加，使用图卷积神经网络进行信息融合。在保留句子结构的同时，考虑了背景知识对实体节点的影响。在EventStoryline数据集上的实验中，F1值取得了良好结果。下一步的工作是测试该模型在不同领域文本下的泛化能力，并进行改进。

表1 EventStoryline 实验结果

方法	<i>P</i>	<i>R</i>	F1
DP	0.340	0.415	0.374
LSTM	0.327	0.449	0.378
ILR	0.370	0.452	0.407
BERT	<b>0.559</b>	0.321	0.408
MMG	0.441	0.428	0.435
Ours	0.309	<b>0.793</b>	<b>0.445</b>

表2 消融实验结果

方法	<i>P</i>	<i>R</i>	F1
本文方法	<b>0.309</b>	0.793	<b>0.445</b>
-dsp	0.262	<b>0.917</b>	0.407
-kl	0.294	0.808	0.431

### 参考文献:

- [1] GIRJU R. Automatic detection of causal relations for Question Answering[C]// Proceedings of the ACL 2003 workshop on Multilingual Summarization and Question Answering. Morristown: Association for Computational Linguistics, 2003: 76–83.
- [2] OH J H, TORISAWA K, HASHIMOTO C, et al. A semi-supervised learning approach to why-question answering[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2016, 30(1).
- [3] HASHIMOTO C, TORISAWA K, KLOETZER J, et al. Toward future scenario generation: extracting event causality exploiting semantic relation, context, and association features[C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2014: 987–997.

- [4] BERANT J, SRIKUMAR V, CHEN P C, et al. Modeling biological processes for reading comprehension[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2014: 1499–1510.
- [5] KHOO C S G, KORNFILT J, ODDY R N, et al. Automatic extraction of cause–effect information from newspaper text without knowledge–based inferencing[J]. *Literary and Linguistic Computing*, 1998, 13(4): 177–186.
- [6] GIRJU R, MOLDOVAN D. Text mining for causal relations[C]// Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference. Palo Alto: AAAI Press, 2002: 360–364.
- [7] GIRJU R. Automatic detection of causal relations for Question Answering[C]// Proceedings of the ACL 2003 Workshop on Multilingual Summarization and Question Answering. Morristown: Association for Computational Linguistics, 2003: 76–83.
- [8] LUO Z, SHA Y, ZHU K Q, et al. Commonsense causal reasoning between short texts[C]// Proceedings of the 15th International Conference on Principles of Knowledge Representation and Reasoning. Palo Alto: AAAI Press, 2016: 421–430.
- [9] DE SILVA T N, XIAO Z, ZHAO R, et al. Causal relation identification using convolutional neural networks and knowledge based features[J]. *World Academy of Science, Engineering and Technology: International Journal of Mechanical and Mechatronics Engineering*, 2017, 11(6): 703–708.
- [10] KRUEGKRAI C, TORISAWA K, HASHIMOTO C, et al. Improving event causality recognition with multiple background knowledge sources using multi–column convolutional neural networks[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017, 31(1): 3466–3473.
- [11] 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.
- [12] 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. Stroudsburg, PA, USA: Association for Computational Linguistics, 2018: 306–316.
- [13] LIU J, CHEN Y B, ZHAO J. Knowledge enhanced event causality identification with mention masking generalizations[C]// Proceedings of the Twenty–Ninth International Joint Conference on Artificial Intelligence. New York: ACM, 2021: 3608–3614.
- [14] MIWA M, BANSAL M. End–to–end relation extraction using LSTMs on sequences and tree structures[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2016: 1105–1116.
- [15] ZHANG Y H, QI P, MANNING C D. Graph convolution over pruned dependency trees improves relation extraction[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2018: 2205–2215.
- [16] SUN K, ZHANG R C, MAO Y Y, et al. Relation extraction with convolutional network over learnable syntax–transport graph[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*,

- 2020, 34(5): 8928–8935.
- [17] CHEN G M, TIAN Y H, SONG Y, et al. Relation extraction with type-aware map memories of word dependencies[C]// Proceedings of Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Stroudsburg: Association for Computational Linguistics, 2021: 2501–2512.
- [18] CASELLI T, VOSSSEN P. The event StoryLine corpus: a new benchmark for causal and temporal relation extraction[C]// Proceedings of the Events and Stories in the News Workshop. Stroudsburg: Association for Computational Linguistics, 2017.
- [19] CHENG F, MIYAO Y. Classifying temporal relations by bidirectional LSTM over dependency paths[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2017: 1–6.
- [20] CHOUBEY P K, HUANG R H. A sequential model for classifying temporal relations between intra-sentence events[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2017: 1796–1802.
- [21] GAO L, CHOUBEY P K, HUANG R H. Modeling document-level causal structures for event causal relation identification[C]// Proceedings of the 2019 Conference of the North. Stroudsburg: Association for Computational Linguistics, 2019: 1808–1817.

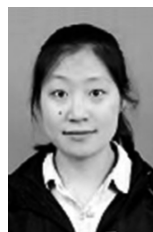
## 作者简介



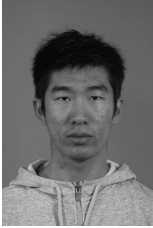
汪诗蕊(2001- ),女,同济大学电子与信息工程学院硕士生,主要研究方向为数据挖掘、知识图谱。



解博涵(2000- ),女,同济大学电子与信息工程学院硕士生,主要研究方向为知识图谱、自然语言处理。



丁玲(1995- ),女,同济大学电子与信息工程学院博士生,主要研究方向为自然语言处理、信息抽取。



陈建廷 (1995- ), 男, 博士, 同济大学电子与信息工程学院博士后, 主要研究方向为数据挖掘、大数据、深度学习。



向阳 (1962- ), 男, 博士, 同济大学, 教授, 主要研究方向为数据挖掘、自然语言处理、知识图谱。

收稿日期: 2023-02-16

通信作者: 向阳, shxiangyang@tongji.edu.cn; 解博涵, hey\_bohan@tongji.edu.cn

基金项目: 国家自然科学基金项目 (No.72071145)

Foundation Item: The National Natural Science Foundation of China (No.72071145)