山西大学学报(自然科学版)48(1):20－28,2025
Journal of Shanxi University(Nat. Sci. Ed.)

DOI:10.13451/j.sxu.ns.2024133

基于中心事件和句子感知的事件因果识别模型

尚舒鹏[1],普瑞丽[1],李旸[2],王素格[1,3],李德玉[1,3*]
(1.山西大学 计算机与信息技术学院,山西 太原 030006;
2.山西财经大学 金融学院,山西 太原 030006;
3.山西大学 计算智能与中文信息处理教育部重点实验室,山西 太原 030006)

摘 要:在金融领域的事件关系识别任务中,篇章级事件因果关系的识别是一项具有挑战性的任务。而事件因果关系关联于中心事件及其所在的句子,一方面,多个事件中的中心事件会影响事件间的因果关系,另一方面,事件所在的句子也会具有强化事件因果关系的作用,为此,本文提出了基于中心事件和句子感知的事件因果识别模型。该模型使用中心感知的高阶事件推理网络(Centrality-aware High-order Event Reasoning Network,CHEER),建立了中心事件的表示,并基于图注意力网络构建了事件关系图,建立了事件及其事件对的表示,以自适应模拟事件间的因果传递,从而识别中心事件及其因果相关事件。通过引入情感中心嵌入和关联词中心嵌入,句子的情感和结构对事件因果关系的表达得以增强,从而提高了金融文本中因果关系识别的准确性。本文在标注了中心事件的公开数据集EventStoryLine与自建数据集上进行实验,结果显示在EventStoryLine数据集上模型的句间准确率提高了1.1%,在自建数据集中整体提高1.8%,验证了中心事件对事件因果关系识别的准确性有较好提升。

关键词:事件因果关系;事件关系图;篇章级文本;金融领域;图注意力网络
中图分类号:TP391 **文献标志码:**A **文章编号:**0253-2395(2025)01-0020-09

Event Causal Identification Model Based on Central Event and Sentence Perception

SHANG Shupeng[1], PU Ruili[1], LI Yang[2], WANG Suge[1,3], LI Deyu[1,3*]
(1. *School of Computer and Information Technology, Shanxi University, Taiyuan* 030006*, China;*
2. *School of Finance, Shanxi University of Finance and Economics, Taiyuan* 030006*, China;*
3. *Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan* 030006*, China*)

Abstract: In the financial sector, recognizing causal relationships between events at the discourse level is a challenging task. Event causality is linked to central events and their corresponding sentences. On the one hand, central events among multiple events influence the causal relationships among them; on the other hand, the sentences in which events occur can also reinforce these causal links. To address this issue, this paper proposes a central event and sentence-aware model for event causality recognition. The model uses the Centrality-aware High-order Event Reasoning Network (CHEER) to establish the representation of the central event, and constructs the event relationship graph based on the graph attention network. This graph captures the representations of events and event pairs, adaptively modeling causal transmission among events, thereby identifying central events and their causally related counterparts. By introducing sentiment-centered and cue word-centered embeddings, the model enhances the representation of sen-

收稿日期:2024-06-08;**接受日期:**2024-10-24
基金项目:国家自然科学基金(62106130;62072294;62376143);山西省青年科学研究项目(20210302124084);山西省高等学校科技创新项目(2021L284)
作者简介:尚舒鹏(1998－),男,山西大同人,硕士研究生,研究方向为自然语言处理。E-mail: shangsp0220@foxmail.com
* **通信作者:**李德玉(LI Deyu),E-mail:lidy@sxu.edu.cn
引文格式:尚舒鹏,普瑞丽,李旸,等.基于中心事件和句子感知的事件因果识别模型[J].山西大学学报(自然科学版),2025,48(1):20-28. DOI:10.13451/j.sxu.ns.2024133.

tences' sentiment and structure, which improves the accuracy of causal relationship recognition in financial texts. Experiments conducted on the publicly available EventStoryLine dataset, annotated with central events, and a self-constructed dataset showed that the proposed model improved inter-sentence accuracy by 1.1% in the EventStoryLine dataset and achieved an overall improvement of 1.8% in the self-constructed dataset. These results confirm that central events significantly enhance the accuracy of event causality recognition.

Key words: event causality; event relationship graph; document-level text; financial domain; graph attention network

0 引言

事件因果关系识别(Event Causality Identification,ECI)是一个基本的自然语言处理(Natural Language Processing, NLP)任务,旨在识别文本中事件之间的因果关系。而因果关系的发现不仅有助于深入理解事件本身,也有助于影响和指导事件的发展。在金融领域中,金融文本蕴含着大量的信息,包括股票涨跌、市场波动、政治事件等,这些信息涵盖了各种事件和事件之间的关系。通过对这些文本中事件的因果关系识别,可为金融决策、风险管理、市场预测等方面的应用提供技术支持。对于篇章内事件之间的因果关系,有可能存在于同一个句子(Sentence Event Causality Identification, SECI)中,也可能跨越整个文档(Discourse Event Causality Identification, DECI)中的多个句子。DECI任务比SECI任务更复杂和更难实现。

目前有关DECI的研究,通常采用构建跨句推理的事件关系图[1]。其中,事件关系图以每个事件对作为节点,事件对之间的关系作为边,在关系图上利用图神经网络(Graph Neural Network,GNN)实现高阶因果传递性,例如事件对(A,B)导致事件对(B,C)⇒事件对(A,C)。在事件关系的研究中,发现并非所有事件都同等重要,通常有一个或两个"中心"事件在整个篇章中占主导地位[2],其余事件为相关事件,这些事件可能作为原因解释,或作为后果解释。

中心感知的高阶事件推理网络(Centrality-aware High-order Event Reasoning Network,CHEER)[3]在识别中心事件时,根据事件本身的特点构建了度中心编码与位置中心编码,然而,CHEER框架在进行中心事件识别任务中忽视了事件所在句子的上下文对事件的影响,例如,句子中的情感以及句子中的关联词。在金融领域、金融机构或者金融媒体经常使用带有强烈情感的话语来煽动广大投资者进行交易。受文献[4]对情绪原因识别相关研究的结论启发,由于97%以上的情感关键词只有一个原因,并且95%以上的原因与情感关键词相近,因此,本文借助该结论,从情感关键词的上下文中提取候选词对事件进行中心嵌入。此外,句子中的关联词可分为显式因果和隐式因果,可以借用关联词前后两个事件建立因果联系。

本文的主要任务是在金融文本中识别中心事件与其具有因果关系的其他事件。除了考虑事件本身在识别中心事件中的作用外,还探讨了另一个维度:句子粒度上下文在中心事件识别中的作用。本文从事件粒度与句子粒度提出了两种处理方法:(1)对于事件粒度,沿用了CHEER框架中事件中心嵌入和度中心嵌入识别中心事件。通过采用基于图注意力网络(Graph Attention Network,GAT)的事件关系图,将事件对视为节点,事件对关系作为边,以自适应方式对不同事件间的重要性进行建模,从而有效实现了事件间因果传递的高阶推理。(2)在句子粒度方面,首先对句子进行情感中心嵌入和关联词中心嵌入。句子常含多个事件,而事件亦可跨越多句,通过均方误差损失实现事件与事件对的动态表示,并引入加权因子以平衡两个维度的训练。

本文的贡献可归纳为以下两个方面:

(1)对事件因果关系识别,利用中心事件在事件关系中的特殊地位,一方面对事件本身进行度中心编码和位置中心编码,另一方面对句子粒度进行情感编码和关联词编码,用于完善对中心事件的多维度表示。

(2)利用均方误差损失对事件与事件对进行动态表示,并在该损失函数中引入加权因子以平衡两个维度的训练,从而识别中心事件及其因果相关的其他事件。

1 相关工作

近年来,ECI引起了人们的广泛关注。就事件承载文本粒度,主要有句子级事件因果关系识别(SECI)和篇章级事件因果关系识别(DECI)。

1.1 句子级事件因果关系识别

句子级事件因果识别关注于单个句子内事件的因果关系。这种级别的识别任务比较直接,因为需要分析的文本较短并且因果关系通常在句子中比较明确。

早期基于特征的方法探索因果表达的不同形式,如词汇和句法模式[5]、因果线索标记[6-7]、时间模式[8]、统计信息[9]和弱监督数据[10]。最近,一些方法已经利用预训练语言模型(Pre-trained Language Model,PLM)用于ECI任务[11],为了处理隐含的因果关系[12],从外部因果中学习特定于上下文的因果模式。

1.2 篇章级事件因果关系识别

篇章级事件因果识别则更加复杂,它涉及跨越多个句子甚至多个段落进行因果关系识别。在这个级别上,因果关系可能不是直接在文本中表示出来,而是通过一系列事件和陈述隐含的展示。DECI需要更复杂的推理和分析能力。

随着句子级自然语言理解取得成功,许多任务被扩展到整个文档,如关系抽取[13]和事件论元抽取[14]。DECI对跨句推理缺乏明确的因果表示提出了新的挑战。Gao等[2]提出了一种基于特征的方法,使用整数线性规划来建模全局因果结构。方向敏感图卷积网络(Direction-sensitive Graph Convolutional Network,DSGCN)使用图推理机制来捕捉事件之间的交互[15]。多文档结构的图卷积网络(Rich Document-level Structures Graph Convolutional Networks,RichGCN)构建了一个事件图,在边的构建过程中可能会引入噪音,并且忽略了事件对之间的相互依赖[16]。文档级事件关联图传递(Event Relational Graph Transformer for Document-level,ERGO)构建了一个关系图,并建模了事件对之间的相互作用[1]。CHEER在事件推理过程中,提出了一种共指关系即中心事件识别[3]。在以上模型的基础上,本研究考虑了两个维度对中心事件的表示,完善了对中心事件的多维度理解。

2 事件因果关系识别模型

基于事件粒度和句子粒度两个维度对事件因果关系进行识别,整个模型如图1所示。该模型包括五个部分:(1)篇章预处理;(2)基于图注意力网络的事件关系图的事件因果关系识别;(3)句子粒度在事件因果关系识别中的作用;(4)事件表示融合;(5)实验与训练。

2.1 篇章预处理

对于篇章的处理,主要利用预训练语言模型作为基本编码器获得上下文的嵌入。给一篇文章 $D=[x_t]_{t=1}^{L}$,其中 x_t 表示第 t 个句子,L 表示长度。D 的开头和结尾添加特殊标记"[CLS]"和"[SEP]",为了标记篇章中的事件,在所有事件的开始和结束时使用"<t>"和"</t>"标记事件位置。使用预训练模型(Bidirectional Encoder Representations from Transformers,BERT)[17]获取每个句子进行嵌入表示,得到整个篇章的嵌入表示 H,如公式(1)所示:

$$H=[h_1,h_2,\cdots,h_L]=\text{Encoder}([x_1,x_2,\cdots,x_L]), \tag{1}$$

其中 h_i 表示句子 x_i 的嵌入。

考虑到BERT处理文本的长度不能超过512个字符,因此,在处理篇章级文档,利用动态窗口对整个文档进行编码。具体来说,根据特定的步长将文本 D 分成几个跨度,并将它们分别输入BERT。找到不同跨度的标记"[CLS]"或"<t>"的所有嵌入,对于在不同跨度中的相同事件,计算<t>标记所有嵌入的平均值,以获得事件 e_i 的最终事件表示 h_{e_i}。

2.2 基于GAT的事件关系图的事件因果关系识别

2.2.1 中心嵌入

考虑到中心事件在事件推理过程中扮演较为重要的角色。因此,本文沿用CHEER[3]的中心感知机制。获得由文档编码器输出的上下文事件嵌入 h_{e_i} 后,执行以下操作:

位置中心编码:中心事件通常位于文章或段落的开头或结尾,以突出其核心要旨。为每个位置随机初始化一个向量,依据事件在文档中的具体位置,为每个事件分配一个位置嵌入向量 c_{pos}。

度中心编码:中心事件在整个文章中被多

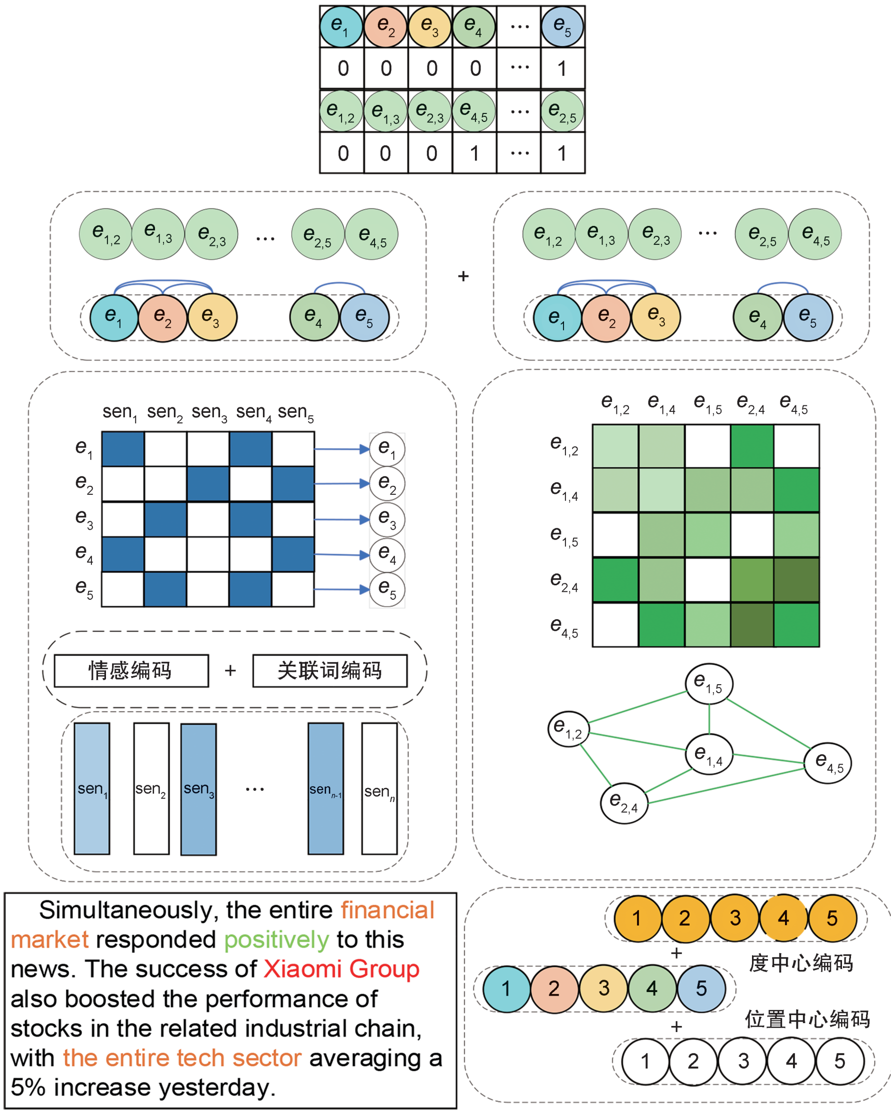

图 1 事件因果关系识别模型图

Fig. 1 Diagram of the event causality recognition model

次提及。依据每个事件在事件关系图中的节点度,为每个事件分配一个度中心嵌入向量 c_{deg}。

中心感知事件嵌入表示:鉴于中心性编码适用于每个事件,故直接将其融合进事件的上下文嵌入中。形式上,对于事件 e_i 及其对应的嵌入 h_{e_i},通过以下方式获得中心感知事件嵌入:

$$c_{e_i} = h_{e_i} \| c_{\text{pos}(e_i)} \| c_{\text{deg}(e_i)}, \tag{2}$$

其中 c_{pos}, c_{deg} 分别通过事件 e_i 的位置中心和度中心编码获得。

2.2.2 事件关系图构建

给定文档 D 的所有事件,构造事件关系图 $G=\{V, E\}$,其中 V 是节点集,E 是边集。对于 V 中的每个节点,引用 D 中不同的事件对,而不是单个事件。目的是了解事件之间的关系,即因果传递性这一高阶推理。因果传递性主要是指,事件A导致事件B,事件B导致事件C,就可以得到事件A导致事件C。对于边 E,将任何两个节点之间的所有边添加到 E 中,不需要事件之间的任何先验关系,因此,G 被初始化为完全图。

2.2.3 基于GAT的事件关系图

GAT[18]结合了图卷积与注意力机制的网络结构,在待处理图的输入上计算注意力。在每个节点更新其隐藏层输出时,对相邻节点进行注意力计算,以便为每个相邻节点分配不同的权重,进而关注更重要的节点并忽略作用较小的节点。

为了预测事件 e_i 和 e_j 之间是否存在因果关系,将文档中事件和事件对进行嵌入。首先使

用事件对节点嵌入初始化节点特征向量。形式上，对于事件对(e_1,e_2)和对应的上下文嵌入(h_{e_1},h_{e_2})，对其进行中心嵌入得到(c_{e_1},c_{e_2})，事件对节点嵌入表示见公式(3)。

$$\nu_{e_{1,2}}^{(0)}=\left[\boldsymbol{c}_{e_1}\|\boldsymbol{c}_{e_2}\right], \tag{3}$$

其中‖表示拼接，0表示神经层的初始状态。

事件对节点的嵌入揭示了两个事件之间的隐式关系信息，从而可以集成到事件对的表示学习中。另外，结构化推理通过全面考虑所有事件对，不仅进一步提升了节点特征的区分能力，也增强了因果关系的传递性。

利用GAT网络，对事件对节点进行嵌入表示，其中自注意机制用于度量事件对v_i与事件对v_j关联关系的程度，其度量见公式(4)。

$$\mathrm{att}_{ij}=\frac{(v_i\boldsymbol{W}_q)(v_j\boldsymbol{W}_k)^{\mathrm{T}}}{\sqrt{d_k}}, \tag{4}$$

其中$\boldsymbol{W}_q$，$\boldsymbol{W}_k$是权重关系矩阵，$\sqrt{d_k}$是缩放因子。对于因果不相关节点将分配较低的注意力权重。

使用softmax函数，对事件对v_i与事件对v_j所有的邻居节点关联关系的程度进行归一化，见公式(5)。

$$a_{ij}=\mathrm{softmax}_j(\mathrm{att}_{ij})=\frac{\exp(\mathrm{att}_{ij})}{\sum z\in N_i\exp(\mathrm{att}_{iz})}, \tag{5}$$

其中N_i是节点i的所有一阶邻居。

为了从邻域信息中聚集事件对的关系知识，计算嵌入的加权线性组合：

$$v_i^{(l)}=\sum_{j\in N_i}a_{ij}(v_j\boldsymbol{W}_v), \tag{6}$$

其中$\boldsymbol{W}_v$是权重参数矩阵。

进一步，使用多头注意力机制对事件v_i的节点进行如下表示：

$$v_i^{(l)}=(\|_{C=1}^{C}\sum_{j\in N_i}a_{ij}(v_j\boldsymbol{W}_v))\boldsymbol{W}_o, \tag{7}$$

其中‖表示拼接，C表示头数，$\boldsymbol{W}_o$表示权重参数矩阵。

通过同时计算所有事件对节点的嵌入$v^{(l)}$，通过多层堆叠实现高阶连接，从而捕获复杂的相互作用。通过事件状态的表征，可以建立单个事件与事件对之间的关系，用于中心事件的识别。

2.3 基于句子的中心嵌入的事件因果关系识别

语句的情感和因果关联词在一定程度上影响了该句中所包含的事件。对于中心事件的识别，从另一个维度即句子粒度出发，对句子进行中心感知，对事件和事件对进行表示。

2.3.1 基于情感词和关联词的中心嵌入

情感中心嵌入：事件的发生会使所在句子具有情感，而句子的情感色彩对句子中包含的事件会产生影响。比如在金融领域，金融机构或者金融媒体经常会用带有强烈情感的话语煽动广大投资者进行交易，产生的交易又会引发整体板块的变化。

识别文本中的情感词和否定词，并将其输入模型，以生成预测标签(正面、负面或中性)，为每个句子分配嵌入向量c_{emo}。

金融情感词典[19]主要包括停用词、积极情感词和消极情感词。该词典总计包含9 228个词语，其中包括5 890个消极词语和3 338个积极词语，以及“那么”“怎么”等连接词和标点符号的停用词。词典包含的数据越多，情感预测的也就越准确。

关联词中心嵌入：在因果关系识别中，句子根据是否包含因果连接词被分为显式因果和隐式因果。1)显式因果包括显式连接词，如“cause”“result”等具有明显因果含义的动词；模糊连接词：无明显的因果含义；可以通过结果性和工具性的动词模式来实现因果含义的连接词，如“decrease”“trigger”等；使因果代理与产生情况不可分割的连接词，如“generate”等；非动词模式，如“from”等。2)隐式因果无因果连接词。

在关联词嵌入过程中，依据连接词词典[20]为句子中的因果连接词打标签，并为每个含关键词的句子分配嵌入向量c_{key}。

2.3.2 基于句子粒度对于事件影响

基于句子粒度的特征，情感和关联词被嵌入到文本的初始句子中。由于中心性编码被应用于每个句子，它们直接被添加到事件的上下文嵌入中。嵌入后的句子表示为：

$$c_{\mathrm{sen}}=h_{[\mathrm{CLS}]}\|c_{\mathrm{emo}}\|c_{\mathrm{key}}, \tag{8}$$

其中c_{emo}，c_{key}分别通过情感中心和关键词中心编码获得。

在训练过程中，利用以下方式对事件e_i进行表示：

$$c'_{e_i}=\frac{1}{N}\sum_{j=1}^{N}c_{\mathrm{sen}_j}\boldsymbol{W},\tag{9}$$

其中 $\boldsymbol{W}$ 是事件对于句子的标记矩阵,N 是句子的个数。

根据句子对事件 e_i 表示后,每个事件两两组成事件对 $e_{i,j}$。事件对节点嵌入表示见公式(10)。

$$\nu'_i=\left[c_{e_i}\|c_{e_j}\right]。\tag{10}$$

通过事件状态的表征,可以建立单个事件与事件对之间的关系,用于中心事件的识别。

2.4 事件表示的融合

结合事件关系图和句子本身对事件因果关系的分析,我们得到了事件与事件对之间的表示。由公式(7)和公式(10)将这两种事件对 v 的表示进行叠加,并对结果的叠加状态进行判别。

$$v_j=v_i+v'_i,\tag{11}$$

$$p_{e_{i,j}}=\mathrm{softmax}(v_j),\tag{12}$$

其中 $p_{e_{i,j}}$ 是事件 e_i 和 e_j 是否存在因果关系的预测概率。判别的最终目的是筛选出中心事件相关的事件对。

2.5 模型的损失函数

根据ERGO[1]采用焦点损失以减轻假阴性问题(即,在训练期间负样本的数量远远超过正样本的数量)。采用权重因子 $\beta\in[0,1]$ 用于平衡焦点损失变量。损失函数 $\mathcal{L}_1$ 可以写成:

$$\mathcal{L}_1=-\sum_{e_i,e_j\in\mathcal{D}}\beta_{e_{i,j}}\left(1-p_{e_{i,j}}\right)^{\tau}\log\left(p_{e_{i,j}}\right),\tag{13}$$

其中 τ 是聚焦超参数,β 是加权超参数,其值与正样本和负样本的比率有关。

我们对中心事件进行预测的损失函数如下:

$$\mathcal{L}_2=-\sum_{e_i\in D}\log(p_{e_i})。\tag{14}$$

鉴于单个句子可能包含多个事件,并且一个事件可能跨越多个句子,为了更有效地表示事件,在模型优化过程中引入了以下损失函数项:

$$\mathcal{L}_3=|\boldsymbol{E}\boldsymbol{S}^{\mathrm{T}}-\boldsymbol{W}|_F^2,\tag{15}$$

其中 $\boldsymbol{E}$ 是事件矩阵,其每一行由 c_{ei} 构成。$\boldsymbol{S}$ 是句子中心嵌入矩阵,$\boldsymbol{W}$ 是事件对于句子的标记矩阵。

$\boldsymbol{E}\boldsymbol{S}^{\mathrm{T}}$ 计算了每个事件与每个句子的内积,通过最小化 $\boldsymbol{E}\boldsymbol{S}^{\mathrm{T}}$ 与 $\boldsymbol{W}$ 的均方误差,使得能够在模型优化过程中被动态调整。最终的损失函数如下:

$$\mathcal{L}=\mathcal{L}_1+\mathcal{L}_2+\lambda\mathcal{L}_3,\tag{16}$$

其中 λ 是一个超参数,用来调控 $\mathcal{L}$。

3 实验

3.1 数据集

本文研究的数据主要为自建数据集和公开数据集 EventStoryLine[21]。

表1自建数据集:本文研究的金融事件数据主要通过经典主题聚类算法(Latent Dirichlet Allocation,LDA)进行聚类处理,数据来源包括《人民日报》金融板块、《每日财经》《每日政经新闻》《东方财富网》等2022—2023年的金融新闻。数据源的新闻文本经过词频与逆文档词频(Term Frequency-Inverse Document Frequency,TF-IDF)统计方法和Jieba分词工具处理后,低频词被统计出并用于筛选关键词和触发词。经过人工校对和清洗,包括中心事件的标注以及剔除与事件抽取无关的组织和数据等,进一步保证了数据质量,事件识别任务的精确性得到提升。

数据集 EventStoryLine:该数据包含22个主题,不包含金融主题,但部分主题中含有少量与金融领域有关的篇章。其中,数据集中包括258个文档,5 334个事件,7 805个句内事件对和62 774个句间事件对(其中1 770个和3 885个分别用因果关系注释)。

表1 自建数据集事件类型与数量

Table 1 The type and number of events in the dataset

事件类型	篇章数量	事件类型	篇章数量
商业信息泄露	1 000	履行担保责任	1 000
金融监管	1 000	资产负面	1 000
债务违约	1 000	非法集资	1 000
金融诈骗	1 000	实际控制人涉诉仲裁	1 000
持有人会议	1 000	股份增持	1 000

在对篇章中心事件进行标注时,考虑到以下几个原则[2]:(1)中心事件必须是篇章的焦点;(2)篇章中描述的所有事件都必须与之相关;(3)中心事件的共指也将被视为中心事件;(4)在正确完整表达主要内容的前提下,中心事件的数量应尽可能少,一般为1到2个。

对数据集进行人工标注中,难免会带有个人的主观性,所以我们有两个标注者,如果两个标注者的答案不一致,则重新进行标注,以

确保数据标注的准确性,标注者之间的平均一致率为85%。

3.2 实验设置

基于Pytorch实现本文的模型。使用未封装的BERT[17]作为文档编码器。基于BERT的文档编码器,将动态窗口大小设置为256,并将文档划分为几个步长为32的重叠窗口。基于Huggingface Transformer的Pytorch版本实现方法[22]。使用AdamW[23]优化模型,使用0.000 02的学习率。应用dropout[24]将模型参数的梯度裁剪为最大范数1.0。根据开发集的性能通过网格搜索执行早期停止并调整超参数:关系图Transformer模型的头$C\in\{1,\mathbf{2},4,8\}$,丢弃率$\in\{0.1,\mathbf{0.2},0.3\}$,聚焦参数$\gamma\in\{0,1,2,\mathbf{3}\}$,权重因子$\beta\in\{0.25,0.50,\mathbf{0.75}\}$,$\lambda\in\{\mathbf{0.2},0.4,0.6,0.8\}$。其中黑体参数为模型最后使用的超参数。

采用精确度(P)、召回率(R)及$F1$分数($F1$)作为评估指标。

3.3 基线方法

DECI基线(1)文本呈现顺序(Order of Presentation, OP)[21]是一个为事件对分配因果关系的虚拟模型。(2)逻辑回归(Logistic Regression, LR)和整数线性规划(Integer Linear Programming, ILP)[2]基于特征的方法,构建文档级结构并使用各种类型的资源。(3)BERT(文中实现)是一个利用动态窗口和事件标记技术的基线方法。(4)RichGCN[16]构建了文档级交互,使用GCN关联事件。(5)ERGO[1]构建了一个关系图,并对事件对之间的交互进行建模。(6)CHEER[3]未公开代码,在自建数据集上难以实验,故不做对比实验,只在公开数据集对比。

4 分析与讨论

4.1 实验结果分析

为了验证本文方法的有效性,与第4.3节的基线方法在自建的数据集上进行对比实验,实验结果见表2。

从表2中,我们可以观察到:(1)我们的模型在自建数据集中表现都优于其他基线,与数据集上的最佳模型(State of The Arts, SOTA)方法相比,我们的模型在$F1$分数提高了1.8%。证明了方法的有效性。(2)与基于特征的方法OP、LR+和LIP相比,使用PLM的模型的性能得到了很大的提高,这验证了BERT可以为ECI任务提取有用的文本特征。

表2 自建数据集对比方法实验结果

Table 2 Experimental results of the comparison methods of the personal data set

Method	P	R	$F1$
OP	22.7	41.7	29.1
LR+	23.4	42.3	40.7
LIP	31.5	44.5	44.6
BERT	32.7	45.7	46.9
RichGCN	35.9	48.2	47.6
ERGO	38.2	50.3	49.3
Our Model	41.6	50.7	51.1

为了进一步说明本文方法有较好的泛化性,选取了公开数据集EventStoryLine进行验证。在实验中,针对句间(Inter)因果关系的识别结果和句内(Inter+Intra)因果关系的识别结果分别进行实验以说明对于篇章级文本的效果。实验结果见表3所示。

表3 模型在EventStoryLine中句间和句内的表现

Table 3 Inter-sentence and intra-sentence performance of the model in EventStoryLine

Model	Inter			Intra+Inter		
	P	R	$F1$	P	R	$F1$
OP	8.4	99.5	15.6	10.5	99.2	19.0
LR+	25.2	48.1	33.1	27.9	47.2	35.1
LIP	35.1	48.2	40.6	36.2	49.5	41.9
BERT	36.8	29.2	32.6	41.3	38.3	39.7
RichGCN	39.2	45.7	42.2	42.6	51.3	46.6
ERGO	43.2	48.8	45.8	46.3	50.1	48.1
CHEER	45.2	52.1	48.4	49.7	53.3	51.4
Our Model	46.3	53.1	49.5	48.1	52.4	50.7

从表3中,我们可以观察到:(1)在公开数据集EventStoryLine中,我们的模型在句间(Inter)因果关系的识别结果优于所有基线,在句内(Inter+Intra)因果关系的识别结果低于CHEER。原因是CHEER研究事件本身,对于句内事件关系更容易识别,情感嵌入和关联词嵌入是在句子粒度上,使得跨句推理时识别效果更好。(2)句子间(Inter)设置的总体$F1$分数远低于句子内(Inter+Intra)设置的$F1$分数,这表明篇章级ECI仍具有改进空间。(3)BERT基线在句内表现很好,但在句子间的性能不及句

内,这表明篇章级文本使用图结构有助于捕获因果关系预测的全局交互。(4)我们注意到OP在EventStoryLine上实现了最高的召回,这可能是由于通过模仿文本顺序简单的分配因果关系,这导致许多假阳性,从而导致低精度。

4.2 消融实验

为了分析我们的模型提出的每个主要组件的效果,在自建数据集上进行消融实验。具体实验结果见表4所示。

表4 消融实验

Table 4 Ablation experiments

Model	$F1$
Our Model	51.7
w/o 事件关系图GNN	50.5
w/o损失优化	50.1
w/o嵌入句子情感	49.9
w/o嵌入关联词	49.7

由表4可以看出:

(1)事件中心性效应(无事件中心性),将使用的方法替换成GNN网络,其性能下降证明了GAT网络可以捕捉事件之间的因果关系,对于该任务要强于GNN网络。(2)去除损失$\mathcal{L}_3$后性能下降,因为$\mathcal{L}_3$主要是对表示的优化,其本质是一个优化过程,其证明了优化的有效性。(3)去掉句子情感嵌入后性能下降,原因是情感影响中心事件与相关事件的关系。(4)去掉关联词嵌入后性能下降,原因是在篇章文本里,关联词常常直接连接句子前后两个因果事件。

5 结论

本文提出了一种新型因果识别模型,旨在面向金融篇章文本中探讨事件间的关系。本模型以中心事件为核心,通过分析事件本身及其所在句子来推断中心事件,并进一步匹配该中心事件对应的事件对。该模型通过精心设计的特性和多任务学习方法评估事件的中心性。大量实验数据显示,该模型在特定领域显著提升了事件因果识别(ECI)的性能。进一步分析验证了模型各主要组件的有效性。

参考文献:

[1] CHEN M Q, CAO Y X, DENG K Q, *et al*. Ergo: Event Relational Graph Transformer for Document-level Event Causality Identification[C]//Proceedings of the 29th International Conference on Computational Linguistics. Stroudsburg, PA, USA: International Committee on Computational Linguistics, 2022: 2118−2128. DOI: 10.48550/arXiv.2204.07434.

[2] 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, PA, USA: Association for Computational Linguistics, 2019: 1808−1817. DOI: 10.18653/v1/n19-1179.

[3] CHEN M Q, CAO Y X, ZHANG Y, *et al*. CHEER: Centrality-aware High-order Event Reasoning Network for Document-level Event Causality Identification[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA, USA: Association for Computational Linguistics, 2023: 10804−10816. DOI: 10.18653/v1/2023.acl-long.604.

[4] GUI L, WU D Y, XU R F, *et al*. Event-driven Emotion Cause Extraction with Corpus Construction[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA, USA: Association for Computational Linguistics, 2016: 1639−1649. DOI: 10.18653/v1/d16-1170.

[5] RIAZ M, GIRJU R. Toward a Better Understanding of Causality between Verbal Events: Extraction and Analysis of the Causal Power of Verb-verb Associations[C]//Proceedings of the SIGDIAL 2013 Conference. Stroudsburg, PA, USA: Association for Computational Linguistics, 2013: 21−30.

[6] DO Q, YEE S C, ROTH D. Minimally Supervised Event Causality Identification[C]//Proceedings of the 49st Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011: 294−303.

[7] HIDEY C, MCKEOWN K. Identifying Causal Relations Using Parallel Wikipedia Articles[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA, USA: Association for Computational Linguistics, 2016: 1424−1433. DOI: 10.18653/v1/p16-1135.

[8] NING Q, FENG Z L, WU H, *et al*. Joint Reasoning for Temporal and Causal Relations[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA, USA: Association for Computational Linguistics, 2018: 2278−2288. DOI: 10.18653/v1/p18-1212.

[9] HU Z C, RAHIMTOROGHI E, WALKER M. Inference of Fine-grained Event Causality from Blogs and Films [C]//Proceedings of the Events and Stories in the News Workshop. Stroudsburg, PA, USA: Association for Computational Linguistics, 2017: 52-58. DOI: 10.18653/v1/w17-2708.

[10] ZUO X Y, CAO P F, CHEN Y B, *et al*. LearnDA: Learnable Knowledge-guided Data Augmentation for Event Causality Identification[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Stroudsburg, PA, USA: Association for Computational Linguistics, 2021: 3558-357. DOI: 10.18653/v1/2021.acl-long.276.

[11] 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. California: International Joint Conferences on Artificial Intelligence Organization, 2020: 3608-3614. DOI: 10.24963/ijcai.2020/499.

[12] ZUO X Y, CAO P F, CHEN Y B, *et al*. Improving Event Causality Identification via Self-supervised Representation Learning on External Causal Statement[C]//Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Stroudsburg, PA, USA: Association for Computational Linguistics, 2021: 2162-2172. DOI: 10.18653/v1/2021.findings-acl.190.

[13] YAO Y, YE D M, LI P, *et al*. DocRED: a Large-scale Document-level Relation Extraction Dataset[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019: 764-777. DOI: 10.18653/v1/p19-1074.

[14] MA Y B, WANG Z H, CAO Y X, *et al*. Prompt for Extraction? PAIE: Prompting Argument Interaction for Event Argument Extraction[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA, USA: Association for Computational Linguistics, 2022: 6759-6774. DOI: 10.18653/v1/2022.acl-long.466.

[15] ZHAO K, JI D H, HE F Z, *et al*. Document-level Event Causality Identification *via* Graph Inference Mechanism [J]. *Inf Sci*, 2021, **561**: 115-129. DOI: 10.1016/j.ins.2021.01.078.

[16] TRAN PHU M, NGUYEN T H. 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. Stroudsburg, PA, USA: Association for Computational Linguistics, 2021: 3480-3490. DOI: 10.18653/v1/2021.naacl-main.273.

[17] DEVLIN J, CHANG M W, LEE K, *et al*. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[EB/OL]. (2019-5-24) [2024-5-10]. https://doi.org/10.48550/arXiv.1810.04805.

[18] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, *et al*. Graph Attention Networks[EB/OL]. (2018-2-4)[2024-5-10]. https://doi.org/10.48550/arXiv.1710.10903.

[19] 姜富伟, 孟令超, 唐国豪. 媒体文本情绪与股票回报预测[J]. 经济学(季刊), 2021, **21**(4): 1323-1344. DOI: 10.13821/J.cnki.ceq.2021.04.10.
JIANG F W, MENG L C, TANG G H. Media Textual Sentiment and Chinese Stock Return Predictability[J]. *China Econ Q*, 2021, **21**(4): 1323-1344. DOI: 10.13821/J.cnki.ceq.2021.04.10.

[20] 杨纪星, 杨波, 朱剑林, 等. 金融领域事件因果关系发现及事理图谱构建与应用[J]. 中文信息学报, 2023, **37**(7): 131-142. DOI: 10.3969/j.issn.1003-0077.2023.07.016.
YANG J X, YANG B, ZHU J L, *et al*. Event Causality Extraction, Eventic Graph Construction and Application in Financial Domain[J]. *J Chin Inf Process*, 2023, **37**(7): 131-142. DOI: 10.3969/j.issn.1003-0077.2023.07.016.

[21] CASELLI T, VOSSEN 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, PA, USA: Association for Computational Linguistics, 2017: 77-86. DOI: 10.18653/v1/w17-2711.

[22] WOLF T, DEBUT L, SANH V, *et al*. Transformers: State-of-the-art Natural Language Processing[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Stroudsburg: Association for Computational Linguistics, 2020: 38-45. DOI: 10.18653/v1/2020.emnlp-demos.6.

[23] LOSHCHILOV I, HUTTER F. Decoupled Weight Decay Regularization[EB/OL]. (2019-01-04) [2024-05-10].https://doi.org/10.48550/arXiv.1711.05101.

[24] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, *et al*. Dropout: a Simple Way to Prevent Neural Networks from Overfitting[J]. *J Mach Learn Res*, 2014, **15**: 1929-1958.