

融合 RACNN 和 BiLSTM 的金融领域事件隐式因果关系抽取

金方焱, 王秀丽

引用本文

金方焱, 王秀丽. 融合 RACNN 和 BiLSTM 的金融领域事件隐式因果关系抽取[J]. 计算机科学, 2022, 49(7): 179-186.

JIN Fang-yan, WANG Xiu-li. Implicit Causality Extraction of Financial Events Integrating RACNN and BiLSTM [J]. Computer Science, 2022, 49(7): 179-186.

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)

一种基于 Bottleneck Transformer 的轻量级微表情识别架构

Lightweight Micro-expression Recognition Architecture Based on Bottleneck Transformer
计算机科学, 2022, 49(6A): 370-377. <https://doi.org/10.11896/jsjcx.210500023>

基于 BERT-GRU-ATT 模型的中文实体关系分类

Chinese Entity Relations Classification Based on BERT-GRU-ATT
计算机科学, 2022, 49(6): 319-325. <https://doi.org/10.11896/jsjcx.210600123>

基于交互注意力图卷积网络的方面情感分类

Interactive Attention Graph Convolutional Networks for Aspect-based Sentiment Classification
计算机科学, 2022, 49(3): 294-300. <https://doi.org/10.11896/jsjcx.210100180>

基于注意力机制和 BiLSTM-CRF 的消极情绪意见目标抽取

Negative-emotion Opinion Target Extraction Based on Attention and BiLSTM-CRF
计算机科学, 2022, 49(2): 223-230. <https://doi.org/10.11896/jsjcx.210100046>

融入自注意力机制的深度学习情感分析方法

Self-attention-based BGRU and CNN for Sentiment Analysis
计算机科学, 2022, 49(1): 252-258. <https://doi.org/10.11896/jsjcx.210600063>

融合 RACNN 和 BiLSTM 的金融领域事件隐式因果关系抽取

金方焱¹ 王秀丽^{1,2}

1 中央财经大学信息学院 北京 102206

2 国家金融安全教育部工程研究中心 北京 102206

(2020212350@email.cufe.edu.cn)

摘要 金融领域的文本信息量大、价值高,尤其是其中的隐式因果关系事件包含着巨大的潜在利用价值。对金融领域文本进行隐式因果关系分析,挖掘隐式因果关系事件中隐含的重要信息,了解金融领域事件更深层的演化逻辑,进而构建金融领域知识库,对金融风险控制、风险预警等具有重要意义。为了提高金融领域中隐式因果关系事件识别的准确度,从特征挖掘的角度入手,提出了一种基于自注意力机制的融合循环注意力卷积神经网络(Recurrent Attention Convolution Neural Network, RACNN)和双向长短期记忆网络(Bidirectional Long Short-Term Memory, BiLSTM)的隐式因果关系抽取方法。该方法结合了基于迭代反馈机制能提取更重要文本局部特征的 RACNN,能更好地提取文本全局特征的 BiLSTM 以及能更深入地挖掘融合特征语义信息的自注意力机制,在 SemEval-2010 Task 8 数据集和金融领域数据集上进行了实验,结果表明,评估指标 F1 值分别达到了 72.98% 和 75.74%,均显著优于其他对比模型。

关键词: 金融领域;隐式因果关系抽取;循环注意力卷积神经网络;双向长短期记忆网络;迭代反馈机制;自注意力机制

中图分类号 TP391

Implicit Causality Extraction of Financial Events Integrating RACNN and BiLSTM

JIN Fang-yan¹ and WANG Xiu-li^{1,2}

1 College of Information, Central University of Finance and Economics, Beijing 102206, China

2 Engineering Research Center of State Financial Security, Ministry of Education, Beijing 102206, China

Abstract The financial field has a large amount of information and high value, especially the implicit causal events which contains huge potential useful value. Carrying out causal analysis on financial domain text to mine the important information hidden in the implicit causal events, understanding the deeper evolutionary logic of the financial field events, to build a financial field knowledge base, which plays an important role in financial risk control and risk early warning. In order to improve the accuracy of identifying the implicit causal events in the financial field, from the perspective of feature mining, based on self-attention mechanism, an implicit causality extraction method integrating recurrent attention convolution neural network (RACNN) and bidirectional long short-term memory (BiLSTM) is proposed. This method combines RACNN that can extract more important local features of text based on an iterative feedback mechanism, BiLSTM that can better extract global features of text, and a self-attention mechanism that can more deeply dig the semantic information of fused features. Experimental results on SemEval-2010 Task 8 and financial field datasets show that the evaluation index F1 value can reach 72.98% and 75.74% respectively, which is significantly better than other comparison models.

Keywords Financial field, Implicit causality extraction, RACNN, BiLSTM, Iterative feedback mechanism, Self-attention mechanism

1 引言

社会活动以事件为驱动,事件是社会活动的载体。金融领域事件中包含大量有价值的信息,如事件的原因、逻辑关系以及造成的损失等。对金融领域事件进行分析,演化事件文本之间的关系,可以实现对文本更深层次的理解。事理图谱是表示事件之间顺承、因果关系的有向图,其中每个节点表示

事件,每条边表示事件之间的因果关系。事件因果关系可分为显式因果关系和隐式因果关系,在金融领域都具有很高的利用价值。

显式因果关系指事件之间包含明显因果提示词的因果关系。例如,在文本“1997—1998 年的国际金融危机使得(e1)亚洲和世界其他一些地区的需求严重不足(e1),导致(e2)世界范围的商品价格下滑(e2),有利于美国的公司降低成本”

中,“导致”是因果提示词。隐式因果关系指事件之间不包含明显因果提示词的潜在因果关系。例如,在文本“〈e1〉1996年上半年当出口疲软〈/e1〉时,〈e2〉出口部门要求人民币贬值〈/e2〉的呼声尤其强烈”中,存在因果关系,但是不存在因果提示词。

目前,金融领域事件的隐式因果关系抽取在 NLP 领域是一项巨大的挑战,它有助于发现事件之间的隐含逻辑,进一步构建金融领域的事理图谱,完善大规模金融领域的事理知识库,对金融风险预警、风险控制至关重要。充分考虑到事件文本局部特征和全局特征相结合的优势以及自注意力机制提取文本语义的全面性,本文提出了一种基于自注意力机制^[1]的融合循环注意力卷积神经网络^[2]和双向长短时记忆网络的隐式因果关系抽取方法,结合数据集位置标示(Position Indicator, PI),将该方法应用于金融领域事件隐式因果关系的抽取,旨在帮助完善金融领域的事理图谱。

本文的贡献是使用基于自注意力机制的融合 RACNN 和 BiLSTM 的隐式因果关系抽取方法,通过 RACNN 中的迭代反馈机制逐步挖掘文本更深层的局部特征,与 BiLSTM 捕获的全局特征相结合,能最大程度地保证文本特征的完整性,再利用自注意力机制提取融合特征中的重要因子,在不依赖外部条件的情况下能够更全面地捕获文本语义中的重要信息。最后,分别在 SemEval-2010 Task 8 公开数据集^[3]和金融领域数据集上进行了测试,在 SemEval-2010 Task 8 数据集中 F1 值达到了 0.729 762,在金融领域数据集中 F1 值达到了 0.757 4328。实验结果表明,本文模型优于实验中的所有对比模型。

2 相关研究

近年来,事件因果关系抽取,从基于模式匹配的方法到基于机器学习的方法,都取得了较大进展。

基于模式匹配的方法针对包含因果提示词的事件文本,实现了对事件显式因果关系的抽取,具有准确率较高、可解释性强等优点,但其模型结构复杂、领域性较强、普适性较差;基于机器学习的方法针对所有类型的事件文本,不依赖于已有的模式,通过机器学习算法自动学习特征,实现对事件因果关系的抽取,但其不对事件文本的显式因果关系和隐式因果关系进行区分,直接转换为事件分类或事件标注问题进行识别,忽略了显式因果关系特征和隐式因果关系特征之间的交叉影响。

2.1 基于模式匹配的方法

在基于模式匹配的因果关系抽取研究中,Sakaji 等^[4]提出了一种从财务报表发现因果关系的方法;Izumi 等^[5]提出了使用文本挖掘技术进行经济因果链搜索,从文本中提取出因果关系以构建因果关系数据库。该类方法依赖于因果关系提示词进行显式因果关系抽取,领域针对性较强,且需要完备的领域知识。

2.2 基于机器学习的方法

基于机器学习的因果关系抽取可转换为事件序列标注或事件分类两个方向。

2.2.1 事件序列标注方向

Cao 等^[6]提出了一种基于神经网络的药物实体和关系联合抽取方法,使用了一种新的序列标注方法,在关系识别实验上表现更好;Xu 等^[7]提出了一种基于图注意力网络的因果关系抽取方法,在序列标注关系识别任务中引入了图注意力网络,取得了较好的效果;Li 等^[8]提出了一种以 BiLSTM-CRF 模型为骨干的因果关系抽取器,引入了多头自注意力机制来学习因果词之间的依赖关系;Zhong 等^[9]提出了一种基于双层模型的维吾尔语突发事件因果关系抽取方法,证明了事件序列的两次模式识别标注对维吾尔语突发事件显式因果关系抽取的有效性。该类方法依赖于识别和标注因果关系触发词,进而了解事件文本的内部结构特征,适用于显式因果关系抽取。

2.2.2 事件分类方向

Zhou 等^[10]提出了一种加入注意力机制的双向长短时记忆网络模型(Attention-Based Bidirectional Long Short-Term Memory, Att_BiLSTM)来识别实体关系,将带有 PI 的文本集作为模型训练的输入,不需要依赖外部知识库或自然语言处理工具来提取高质量特征就能捕获句子的语义信息;Tian 等^[11]提出了一种基于双向 LSTM 的维吾尔语事件因果关系抽取方法,提取出事件中包含事件触发词的 10 项反映事件内部结构特征的信息,提高了事件因果关系的识别性能;Ning 等^[12]提出了一种基于多通道自注意力机制的电子病历实体关系抽取方法,该方法能够强化模型对句级语义特征的捕捉,提升模型对电子病历的学习能力;Wang 等^[13]提出了一种融合上下文信息的篇章级事件时序关系抽取方法,该方法将事件句与上下文信息相结合,保证了关系识别结果整体时序关系的一致性;Tourille 等^[14]提出了一种识别医学事件和时间表示之间包含关系的神经网络架构,以 BiLSTM 为框架融合多事件特征,提高了关系识别准确率;Feng 等^[15]提出了一种混合神经网络(Hybrid Neural Network, HNN),同时使用了 BiLSTM 和 CNN 来训练每个词的特征以进行事件检测。该方法将事件因果关系抽取转换为事件分类问题,可适用于全部类型事件的因果关系抽取,但对因果关系的显性和隐性的区分度不高。在关于隐式因果关系抽取的研究中,Guo 等^[16]提出了一种基于动态块最大池化的 BiLSTM-CNN 结构的隐式篇章关系识别方法,将通过 BiLSTM 获取的文本语义信息输入到 CNN 框架中以提取特征,同时兼顾了全局与上下文信息,提高了隐式篇章关系识别的有效性,但可能导致部分全局语义信息在卷积过程中丢失,从而影响最终的关系抽取性能。

上述研究表明,基于模式匹配的方法局限于已有模式,且需要大量的领域知识,只适用于事件显式因果关系抽取;基于机器学习的事件序列标注方法将因果关系抽取转换为事件序列标注,需要了解文本的内部结构特征,事件关系的判断受事件触发词的影响,也只适用于事件显式因果关系抽取;基于机器学习的事件分类方法将因果关系抽取转换为事件分类,大部分方法^[10-15]在没有考虑事件显式因果关系和隐式因果关系不同表现形式的前提下进行建模,可能导致模型训练仅考虑到

显式因果关系特征而遗漏了隐式因果关系特征。针对文本隐式关系抽取的研究^[16]可能存在特征丢失的问题,另外,其仅针对具有一定格式结构的公共语料库,而对于金融领域这种涉及具体领域知识的隐式因果关系抽取的研究还较少。

因此,本文提出了一种融合 RACNN 和 BiLSTM 的金融领域事件隐式因果关系抽取方法。与 Zhou 等^[10]、Feng 等^[15]以往研究中文本特征的提取完整度、针对的文本类型不同,本文针对隐式因果关系文本,并且采用自注意力机制,能够更好地捕获文本语义信息;与 Guo 等^[16]以往研究中文本特征的提取顺序、采用最大池化进行特征选择不同,本文利用基于迭代反馈机制的 RACNN 进行局部特征提取,以提高文本局部特征的质量,与 BiLSTM 提取的全局特征进行并行融合,充分考虑了文本特征中局部特征和全局特征的完整性以及对关系抽取结果的直接影响,并且采用通道注意力模型(Channel Attention Module, CAM)^[17]来解决特征选择问题。该模型结合了最大池化和平均池化,既考虑了局部最大特征的重要性,又考虑了局部特征的总体分布情况。

3 数据预处理

3.1 数据来源

本文模型将用于金融领域事件隐式因果关系识别,实验数据来源于从中国经济 50 人论坛^[1]中爬取的宏观经济、产业经济以及金融与外贸 3 个方面的文章,共 2124 篇。

3.2 文本预处理

首先,对爬取的文章进行分句、去停用词处理,生成事件句。然后,通过识别事件句中的因果提示词,选择没有因果提示词的事件句用于文本因果事件抽取;利用依存句法分析从每个事件句中抽取事件组成事件对,根据事件对类型将其标注为 3 类,即前因后果、前果后因和其他,分别用数字 1, 2 和 0 表示。最后,将数据集以 8:1:1 的比例划分为训练集、验证集和测试集,用于事件隐式因果关系抽取实验。

4 隐式因果关系抽取模型搭建

如图 1 所示,本文提出的基于自注意力机制的融合了 RACNN 和 BiLSTM 的模型包括 5 个部分。

(1)输入层:将输入的事件对文本中的词映射成指定维度的词向量。

(2)RACNN 层:使用 3 种不同的卷积核+通道注意力模型,结合迭代反馈机制迭代更新每次卷积层的输入,每层卷积输出输入层中词向量的低阶特征。

(3)BiLSTM 层:使用两层 LSTM 输出输入层中词向量的高阶特征。

(4)自注意力层:根据输入的特征向量矩阵生成一个权重矩阵,与该特征向量相乘得到特征更明显的特征向量矩阵。

(5)输出层:将自注意力层中得到的特征向量矩阵最终用于关系分类。

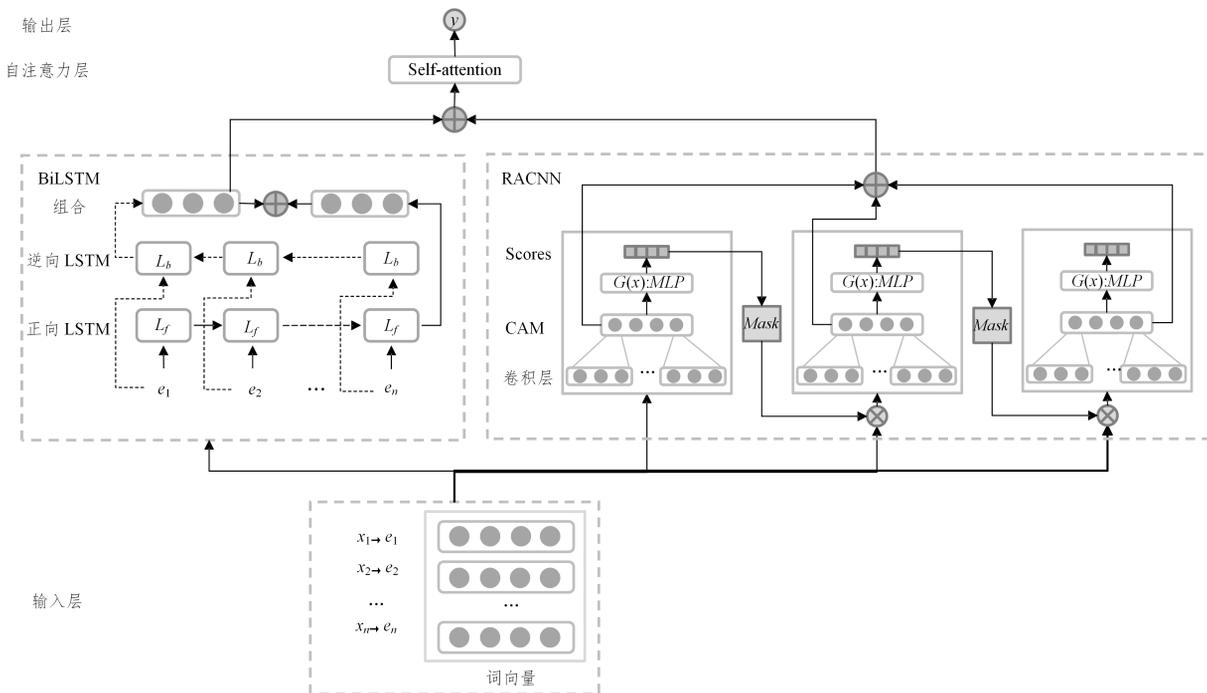


图 1 基于自注意力机制的融合 RACNN 和 BiLSTM 的模型

Fig. 1 Self-attention mechanism-based model with RACNN and BiLSTM

4.1 词向量

给定一个包含 n 个词的文本序列 $S = \{x_1, x_2, \dots, x_n\}$, 通过训练将 S 向量化, 映射为 $E_m = \{e_1, e_2, \dots, e_n\}$, 其中 $e_i = [\omega_1, \omega_2, \dots, \omega_m]$, $\omega \in [-1, 1]$, m 为指定的词向量维度。

4.2 RACNN 变体

RACNN 是由 Fu 等^[2]提出的一种卷积神经网络结构,用于图像识别。本文提出了一种 RACNN 变体,如图 2 所示,将其用于文本局部特征提取。相比用于文本局部特征提取的

¹⁾ <http://www.50forum.org.cn/home/>

传统 CNN, 该变体采用了迭代反馈机制($G(x) \rightarrow Mask$), 连续对文本词向量进行掩膜(Mask)处理, 模拟人注意力特性, 能够更好地捕获文本中关键的局部特征。

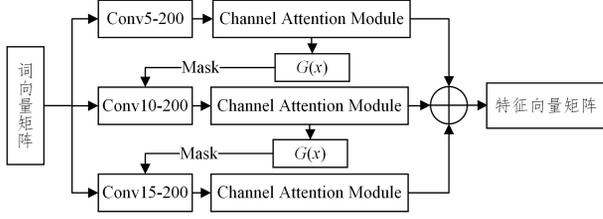


图2 RACNN变体

Fig. 2 Variant of RACNN

RACNN 单次卷积层如图3所示, 首先将包含 n 个词的输入文本中的词 x_i 通过训练映射为词向量 e_i , 组合形成词向量矩阵 $E_m = \{e_1, e_2, \dots, e_n\}$; 其次, 用 $r \times m$ 的过滤器来对 E_m 进行卷积, 过滤器的数量为 u , 提取出 E_m 局部特征。

$$c_{v(i)} = \text{relu}(W_v \cdot e_{(i,i+r-1)} + b_v) \quad (1)$$

其中, $v \in (1, u)$ 表示过滤器的序号; W_v 表示第 v 个 $r \times m$ 的

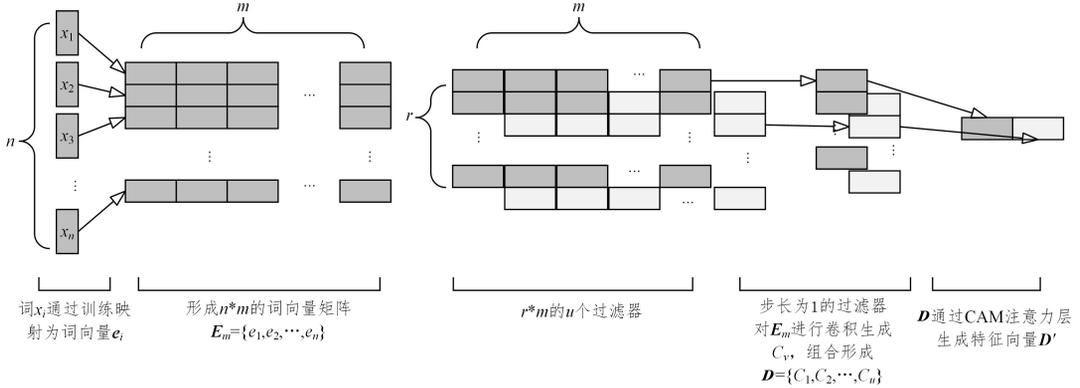


图3 RACNN单次卷积层

Fig. 3 Single convolution layer of RACNN

RACNN 中迭代反馈函数 $G(x)$ 可以通过全连接层来实现, 先计算文本中词 x_i 对应的分数 $score_i$, 然后生成一个矩阵 $Mask$, 用于识别分数低的词。

$$Scores = \{score_1, score_2, \dots, score_n\} \quad (10)$$

$$Mask = Matrix(IODS(Scores)) \quad (11)$$

其中, $IODS$ 表示获取 $Scores$ 中分数较低值索引的函数; $Matrix$ 表示将索引列表映射为矩阵。

将 $Mask$ 反馈到下一次卷积, 卷积层的输入更新为 E_m' 。

$$E_m' = Mask \cdot E_m \quad (12)$$

根据上述步骤进行迭代, 直到 3 次卷积操作完成, 将 3 次卷积操作输出的特征向量 D' 组合, 形成最终的局部特征向量。

$$E = [D_1', D_2', D_3'] \quad (13)$$

其中, D_i' 表示第 i 次卷积操作输出的特征向量。

4.3 BiLSTM

LSTM 由 Hochreiter 等^[18]首次提出, 用于解决梯度爆炸的问题。但 LSTM 只关心文本中当前词的历史信息而忽略了未来信息。因此, 本节采用 BiLSTM 替代 LSTM, 不仅解决了梯度问题, 而且考虑了当前词的前后文语义信息。BiLSTM 是一个正向 LSTM 和一个反向 LSTM 的组合, 可以

获取到整个文本序列的信息, 在隐式因果关系抽取任务中可以充分挖掘文本全局特征, 进而能更准确地识别事件的关系类型。将文本词向量矩阵 E_m 输入到 BiLSTM 中, 分别输出正向结果 f_w 和反向结果 b_w , 连接 f_w 和 b_w 作为 BiLSTM 的输出。

$$output_i = [f_w \oplus b_w] \quad (14)$$

$$H = [output_1, output_2, \dots, output_n] \quad (15)$$

其中, $output_i$ 表示词向量 e_i 的 BiLSTM 输出, H 表示 E_m 的 BiLSTM 输出。

$$C_v = \{c_{v(1)}, c_{v(2)}, \dots, c_{v(n-r+1)}\} \quad (2)$$

$$D = \{C_1, C_2, \dots, C_u\} \quad (3)$$

如图3所示, 将 D 输入到 RACNN 的注意力层, 采用 CAM 提取新的局部特征 D' 来代替 D 。

$$D_{Max} = Max\ pool(D, axis=1) \quad (4)$$

$$D_{Avg} = Avg\ pool(D, axis=1) \quad (5)$$

$$D'_{Max} = w_2 \cdot (w_1 \cdot D_{Max} + b_1) + b_2 \quad (6)$$

$$D'_{Avg} = w_2 \cdot (w_1 \cdot D_{Avg} + b_1) + b_2 \quad (7)$$

$$D' = \text{sigmoid}(D'_{Max} + D'_{Avg}) \quad (8)$$

其中, $axis=1$ 表示按维度 1 对 D 池化; w_1, b_1, w_2, b_2 为共享 MLP 的参数, 最终 D' 表示为一个长度为 u 的局部特征向量。

$$D' = \{d_1, d_2, \dots, d_u\} \quad (9)$$

D 通过 CAM 注意力层生成特征向量 D'

获取到整个文本序列的信息, 在隐式因果关系抽取任务中可以充分挖掘文本全局特征, 进而能更准确地识别事件的关系类型。将文本词向量矩阵 E_m 输入到 BiLSTM 中, 分别输出正向结果 f_w 和反向结果 b_w , 连接 f_w 和 b_w 作为 BiLSTM 的输出。

$$output_i = [f_w \oplus b_w] \quad (14)$$

$$H = [output_1, output_2, \dots, output_n] \quad (15)$$

其中, $output_i$ 表示词向量 e_i 的 BiLSTM 输出, H 表示 E_m 的 BiLSTM 输出。

4.4 自注意力层

注意力机制旨在模拟人的注意力特性, 在特征提取等多个场景中得到了较好的应用。谷歌机器翻译团队于 2017 年首次提出了自注意力机制^[1], 相比注意力机制, 自注意力机制以全连接的形式计算文本中每个词与其他词之间的注意力值, 解决了长期依赖问题, 使得词与词之间存在相关性, 因此每个词具有不同的重要性。

$$HE = \text{contact}(H, E) \quad (16)$$

$$A = \text{softmax}(\text{transpose}(\tanh(HE \cdot w_3)) \cdot w_4) \quad (17)$$

$$M = A \cdot HE \quad (18)$$

其中, $contact$ 为矩阵拼接函数; $transpose$ 为矩阵转置函数; w_3 和 w_4 为参数矩阵。

4.5 分类

本文使用 softmax 分类器来预测文本 S 中事件对的关系 \hat{y} , 将 M 输入得:

$$M' = \text{relu}(M \cdot w_5 + b_5) \quad (19)$$

$$\hat{p}(y|S) = \text{softmax}(w^{(S)} \cdot M' + b^{(S)}) \quad (20)$$

$$\hat{y} = \arg \max \hat{p}(y|S) \quad (21)$$

其中, w_5 和 b_5 为全连接层的参数和偏置项。

5 实验与分析

5.1 实验环境

本文的实验环境如表 1 所列。

表 1 实验环境

Table 1 Experiment environments

名称	配置	名称	配置
操作系统	Windows10	内存	16 GB
CPU	Intel Core i7-6700	开发环境	Tensorflow 1.4.1
GPU	GeForce GTX 2080Ti		

5.2 实验数据

本文先后采用两个不同的数据集进行对比实验: 1) SemEval-2010 Task 8 数据集^[3]; 2) 金融领域数据集。

(1) SemEval-2010 Task 8 数据集。该数据集为公开数据集, 包括 9 种关系, 其中训练集有 8000 条, 测试集有 2717 条。

(2) 金融领域数据集。该数据集为金融领域事件数据集, 每条数据包含两个事件, 对应 3 种关系, 即前因后果、前果后因和其他, 其中训练集有 1437 条, 测试集有 358 条。

本文利用上述两种数据集进行隐式因果关系抽取实验。为达到更好的对比效果, 在公开数据集实验中使用由 Pennington 等^[19]训练的词向量进行实验; 在金融领域数据集实验中使用由 Li 等^[20]训练的词向量进行实验。

5.3 实验参数

实验参数的选取较大程度地影响了模型的性能。表 5 列出了本文模型中 RACNN 部分的参数设置, 表 6 列出了本文模型中 BiLSTM 部分的参数设置。其中, 对部分重要参数 (RACNN 滑动窗口、RACNN Masksize 和 BiLSTM 隐藏层) 进行实验对比展示, 结果分别如表 2—表 4 所列。

表 2 RACNN 滑动窗口

Table 2 RACNN sliding window

滑动窗口	F1/%
3,4,5	71.49
5,6,7	73.88
5,10,15	75.08
10,15,20	74.24
15,20,25	74.77

表 3 RACNN Masksize

Table 3 RACNN Masksize

Masksize	F1/%
5	70.91
10	72.12
15	77.71
20	74.62
25	73.57

表 4 BiLSTM 隐藏层大小

Table 4 BiLSTM hidden layer size

隐藏层大小	F1/%
100	70.49
200	75.58
300	73.75

表 5 RACNN 参数

Table 5 RACNN parameters

参数	值	参数	值
词向量维度	300	L2	1×10^{-5}
滑动窗口	5,10,15	Epoch	50
滑动窗口数量	200	Masksize	15
激活函数	RELU		

表 6 BiLSTM 参数

Table 6 BiLSTM parameters

参数	值	参数	值
词向量维度	300	隐藏层大小	200
层数	2	损失函数	交叉熵
优化函数	Adadelta	Epoch	50
学习率	1	L2	1×10^{-5}
dropout_ratio	0.3		

5.3.1 滑动窗口

调整 RACNN 模型中的滑动窗口参数以进行实验对比, 结果如表 2 所列。当滑动窗口选择为 5, 10, 15 时效果最好, 原因可能是: 当滑动窗口过小时, 卷积过程侧重于对细节特征的捕捉, 而忽略了对局部特征的整体把握; 而当滑动窗口过大时, 卷积过程则侧重于局部特征的整体性而忽略了细节特征。

5.3.2 Masksize

调整 RACNN 模型中的 Masksize 参数以进行实验对比, 结果如表 3 所列。当 Masksize 选择为 15 时效果最好, 原因可能是: Masksize 过小时, 上一个卷积的反馈对下一个卷积的特征聚焦效果不够明显; 而 Masksize 过大时, 上一个卷积的反馈会导致大部分特征缺失, 进而影响最终性能。

5.3.3 隐藏层大小

调整 BiLSTM 模型中的隐藏层大小以进行实验对比, 结果如表 4 所列。当隐藏层大小选择为 200 时效果最好, 原因可能是: 隐藏层的大小与模型的复杂度有关, 过小可能导致模型欠拟合, 而过大可能导致模型过拟合, 两种情况都会影响模型的效果。

5.3.4 其他参数

对于 RACNN 部分, 还分别比较了滑动窗口数量 30, 50, 100, 200 和激活函数 relu, tanh, 最终发现设置为表 5 所列的参数能使模型表现出较好的性能; 对于 BiLSTM 部分, 还分别比较了 dropout_ratio 0.3 和 dropout_ratio 0.7, 优化函数 Adadelta 和 Adam, 最终发现设置为表 6 所列的参数能使模型表现出较好的性能。另外, 模型迭代次数 Epoch 默认为 50, 默认层数参考 Zhou 等^[10]实验中的设置, 还比较了模型每次输入的批量 (batch_size) 10 和 20, 实验对比发现 batch_size 设置为 10 时能够更好地收敛。

5.4 对比方法

为了更好地对比本文模型的性能, 本文设置了横向和纵向两种对比方式。横向对比指本文模型与其次级模型进行

比较,可以验证模型中模块优化的有效性;纵向比较指本文模型与当前领域最前沿的研究成果进行比较,从特征提取的角度进一步验证了模型的核心竞争力,并且对不适合进行对比实验的研究成果进行解释说明。

5.4.1 横向对比

(1)Att_BiLSTM。Zhou等^[10]提出了一种加入了注意力机制的双向长短时记忆网络模型,用于抽取实体关系,其中引出了文本PI,可以将其作为机器学习模型用于隐式因果关系抽取。

(2)Self-Att_BiLSTM。Vaswani等^[1]根据传统的注意力机制,首次提出了自注意力机制的概念;Cao等^[21]首次将自注意力机制用于中文实体识别任务,该自注意力机制在任务中表现良好,可作为机器学习模型用于隐式因果关系抽取。

(3)Self-Att_BiLSTM+RACNN(无迭代反馈)+PI。本文提出了该对比模型,用于测试本文模型(Self-Att_BiLSTM+RACNN(迭代反馈)+PI)中迭代反馈机制对隐式因果关系抽取结果的影响。

5.4.2 纵向对比

(1)BMGF-RoBERTa。Liu等^[22]提出了一种融合上下文表示、双边多角度匹配以及全局信息融合的模型,全面地对文本中的词句进行重要性分析,进而提高模型的效果。该模型的输入中CLS和EOS部分是BERT和RoBERTa特有的标记向量,并且实验数据基于RoBERTa预训练模型进行向量映射,这与本文模型向量映射环节不匹配,不能直接比较特征挖掘对分类的影响,因此该模型不适合用于对比实验。

(2)KANN。Guo等^[23]通过模拟人利用相互关联的长、短记忆机制来理解文本,设计了一种知识增强神经网络模型。该模型结合文本内部和文本外部两方面的信息进行关系识别任务,涉及外部知识库的应用,这与本文模型聚焦于文本内部信息挖掘不同,不适合用于对比实验。

(3)BERT+BiLSTM+Tree。Jiang等^[24]提出了一种结合BERT和BiLSTM的模型,该模型通过BERT处理文本向量映射任务,且在分类中运用了树型结构,其优势在于能更好地处理多类别分类任务(19类),但其对于少类别分类的任务优势不够明显,不适合用于对比实验。

(4)DC-BCNN。Guo等^[16]提出了一种基于动态块最大池化的BiLSTM-CNN模型,将通过BiLSTM获取的文本语义信息输入到CNN框架中以提取特征,最终用于关系分类任务。该模型从特征提取的角度入手,验证了模型的有效性,可用于对比实验。

(5)Syntax+self-attention。Fan等^[25]提出了一种融合双仿射深度神经网络和BiLSTM的关系分类模型,该模型运用双仿射神经网络抽取句法信息,再通过BiLSTM对句法信息和原文本进行特征提取,最终用于关系分类任务。该模型考虑了文本的依存句法信息在语义理解过程中的作用,能提升分类任务的效果,可用于对比实验。

5.5 结果分析

本文模型分别在SemEval-2010 Task 8数据集和金融领域数据集上进行了实验,实验结果如表7和表8所列。本文采用官方评估指标对实验进行测评,包括精确率(Precision,

Pre)、召回率(Recall,Rec)和F1值(F1-score,F1),其中:

$$F1 = 2 * Pre * Rec / (Pre + Rec) \quad (22)$$

F1采用的是Macro-F1,表示对除“0”标签以外的所有类别的F1值计算平均值。此外,考虑到选择一种方法训练的最优模型在未知数据上表现较差,本文在实验中对每种方法都训练5次,选取评估指标的平均值作为最终的指标,这样既避免了单一模型的随机性,又提高了模型的鲁棒性。

表7 各方法在公共数据集上的实验结果

Table 7 Experimental results of each method on public datasets (单位:%)

方法	Pre	Rec	F1
Att_BiLSTM	53.46	67.97	59.37
Att_BiLSTM+PI	63.70	70.82	66.73
Self-Att_BiLSTM+PI	67.08	75.99	70.57
Self-Att_BiLSTM+RACNN (无迭代反馈)+PI	68.21	76.61	71.77
DC-BCNN	64.41	79.00	70.92
Syntax+self-attention	69.28	73.12	71.14
本文模型	71.58	75.12	72.98

表8 各方法在金融领域数据集上的实验结果

Table 8 Experimental results of each method on financial datasets (单位:%)

方法	Pre	Rec	F1
Att_BiLSTM	70.39	68.69	69.48
Att_BiLSTM+PI	74.96	70.99	72.69
Self-Att_BiLSTM+PI	75.95	71.94	73.58
Self-Att_BiLSTM+RACNN (无迭代反馈)+PI	77.10	73.37	74.96
DC-BCNN	71.68	68.78	69.47
Syntax+self-attention	69.72	74.90	71.55
本文模型	75.90	75.89	75.74

由表7可以看出,在SemEval-2010 Task 8数据集上的实验中,加入位置标示的方法(Att_BiLSTM+PI)比Att_BiLSTM方法的F1值高7.36%,加入自注意力机制的方法(Self-Att_BiLSTM+PI)比Att_BiLSTM+PI方法的F1值高约3.84%,Self-Att_BiLSTM+RACNN(无迭代反馈)+PI方法比Self-Att_BiLSTM+PI方法的F1值高约1.2%,在加入迭代反馈机制之后,本文模型相比Self-Att_BiLSTM+RACNN(无迭代反馈)+PI方法的F1值提高了约1.21%;与最新研究成果进行比较,本文模型分别比DC-BCNN方法和Syntax+self-attention方法的F1值提高了约2.06%和1.84%。

由表8可以看出,在金融领域数据集的实验中,Att_BiLSTM+PI方法比Att_BiLSTM方法的F1值提高了3.21%,Self-Att_BiLSTM+PI方法比Att_BiLSTM+PI方法的F1值提高了约0.89%,Self-Att_BiLSTM+RACNN(无迭代反馈)+PI方法比Self-Att_BiLSTM+PI方法的F1值提高了约1.38%,在加入迭代反馈机制之后,本文模型比Self-Att_BiLSTM+RACNN(无迭代反馈)+PI方法的F1值提高了约0.78%;在与最新研究成果的比较中,本文模型分别比DC-BCNN方法和Syntax+self-attention方法的F1值提高了约6.27%和4.19%。

在横向比较中,相比Self-Att_BiLSTM+PI方法,Self-Att_BiLSTM+RACNN(无迭代反馈)+PI方法融入了RACNN结构,能提取文本局部特征的特性,同时考虑了文本局部

特征和全局特征,即 RACNN 的基本结构是由 3 个卷积层组成,其滑动窗口逐渐增大,所捕获的文本局部特征的区域也逐渐扩大,进而提高了文本局部特征挖掘的全面性;相比 Self-Att_BiLSTM+RACNN(无迭代反馈)+PI 方法,本文模型在 RACNN 中加入迭代反馈机制后,其挖掘文本局部特征的优势被进一步放大,即 RACNN 的前两个卷积层都会对下一个卷积层产生一个正向反馈,该反馈指前一个卷积层生成的对输入文本中所有词的权重分数矩阵表示,能减少下一次卷积中非重要特征的提取,进一步提升了文本局部特征的质量。

在纵向比较中,与 DC-BCNN 方法相比,本文模型通过模型并行拼接方式来减少全局语义信息在卷积过程中的部分丢失,可提高最终的关系抽取性能;与 Syntax+self-attention 方法相比,本文模型利用 RACNN 对文本局部特征进行深度挖掘,相比 Syntax+self-attention 方法更加注重文本上下文全局语义,本文模型更加全面地考虑了局部特征对模型性能的影响。

实验结果表明,融合了 RACNN 和 BiLSTM 的方法比单 BiLSTM 方法在提取特征上表现更好,迭代反馈机制能够提高 RACNN 提取局部特征的性能,并且本文模型在与前沿研究成果的比较中表现良好,因此本文模型能够有效提升金融领域中隐式因果关系事件抽取的准确度。

5.6 损失值和准确率分析

本文横向比较了 4 种模型,即 Att_BiLSTM+PI,Self-Att_BiLSTM+PI,Self-Att_BiLSTM+RACNN(无迭代反馈)+PI 和本文模型训练集的损失值和验证集的准确率,数据曲线由可视化工具 TensorBoard 绘制,如图 4 和图 5 所示,其中 Step(步)表示训练步数,指每个迭代(Epoch)中一个批量大小(batch_size)数据的训练过程。

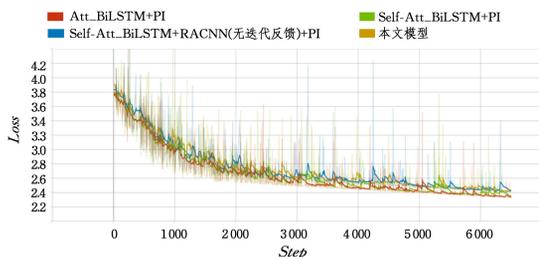


图 4 训练步数-损失值(训练集)

Fig. 4 Training step-loss(training set)

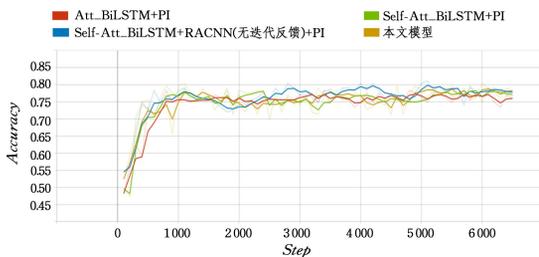


图 5 训练步数-准确率值(验证集)

Fig. 5 Training step-accuracy(dev set)

由图 4 可知,本文模型相比其他模型在训练过程中收敛速度较慢,模型损失值略大;相比 Self-Att_BiLSTM+RACNN(无迭代反馈)+PI 方法,本文模型在训练过程中的收敛

速度无明显差异,但损失值略低。由图 5 可知,除 Att_BiLSTM+PI 以外的模型在训练初始阶段验证集的准确度上升速度较快,所有模型在约 1000 步以后准确率都趋于平缓,随着迭代次数增多,在 5000 步左右时本文模型和 Self-Att_BiLSTM+RACNN(无迭代反馈)+PI 方法的准确率逐渐超过其他模型,在 5100 步左右时达到顶峰,最终在 6400 步左右时本文模型的准确率超过了其他所有模型。

在基于自注意力机制的 BiLSTM 模型的基础上,本文模型引入了特征融合的概念,在利用基于迭代反馈机制的 RACNN 和 BiLSTM 共同提取文本特征的同时,神经网络模型结构变得更加复杂,训练参数增多,因此在训练过程中本文模型的收敛速度略慢于其他模型,损失值增大,但本文模型抽取的融合特征表现力更强,故最终验证集的准确率相比其他模型更高。

结束语 本文提出了一种基于自注意力机制的融合 RACNN 和 BiLSTM 的金融领域事件隐式因果关系抽取方法,该方法利用 RACNN 中迭代的反馈机制逐步挖掘更深层的文本局部特征,与 BiLSTM 提取文本的全局特征相融合,并结合自注意力机制对融合特征进行语义挖掘,充分考虑了文本的全局语义信息。本文模型在 SemEval-2010 Task 8 公共数据集和金融领域数据集上做了对比实验,其 F1 值分别达到了 72.98% 和 75.74%,优于其他对比模型,结果表明本文模型有效提升了抽取金融领域中隐式因果关系事件的准确度,能对进一步构建金融领域事理图谱发挥重要作用。

本文也存在一些不足,从特征挖掘的角度入手,基于当前经典的 BiLSTM 提取文本全局特征,单方面研究了 RACNN 提取文本局部特征性能以及与 BiLSTM 全局特征融合的表现能力,但在全局特征挖掘上的改进考虑得不充分。图 5 表明,本文模型的整体稳定性有待提升,模型之间的融合方式可能存在一些缺陷。因此,未来将深入研究金融领域事件的全局语义分析,重点了解模型内部结构和模型之间的关联方式,进一步提升模型效果并着手构建金融领域的事理图谱。

参考文献

- [1] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2017: 5998-6008.
- [2] FU J L, ZHENG H L, MEI T. Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition[C]// Proceedings of the CVPR. Piscataway, NJ: IEEE, 2017: 4438-4446.
- [3] HENDRICKX I, KIM S N, KOZAREVA Z, et al. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals [C]// Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009). Stroudsburg, PA: ACL, 2009: 94-99.
- [4] SAKAJI H, MURONO R, SAKAI H, et al. Discovery of rare causal knowledge from financial statement summaries[C]// Proceedings of 2017 IEEE Symp Series on Computational Intelligence (SSCI). Piscataway, NJ: IEEE, 2017: 1-7.

- [5] IZUMI K, SAKAJI H. Economic causal-chain search using text mining technology[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. Amsterdam; Elsevier, 2019; 23-35.
- [6] CAO M Y, YANG Z H, LUO L, et al. Joint drug entities and relations extraction based on neural networks[J]. Journal of Computer Research and Development, 2019, 56(7): 1432-1440.
- [7] XU J H, ZUO W L, LIANG S N, et al. Causal relation extraction based on graph attention networks[J]. Journal of Computer Research and Development, 2020, 57(1): 159-174.
- [8] LI Z N, LI Q, ZOU X T, et al. Causality extraction based on self-attentive BiLSTM-CRF with transferred embeddings[J]. Neurocomputing, 2019, 423: 207-219.
- [9] ZHONG J, YU L, TIAN S W, et al. Causal relation extraction of uyghur emergency events based on cascaded model[J]. Acta Automatica Sinica, 2014, 40(4): 771-779.
- [10] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification [C]// Proceedings of the 54th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2016; 207-212.
- [11] TIAN S W, ZHOU X F, YU L, et al. Causal relation extraction of uyghur events based on bidirectional long short-term memory model[J]. Journal of Electronics and Information Technology, 2018, 40(1): 200-208.
- [12] NING S M, TENG F, LI T R. Multi-channel self-attention mechanism for relation extraction in clinical records[J]. Chinese Journal of Computers, 2020, 43(5): 916-929.
- [13] WANG J, SHI C H, ZHANG J, et al. Document-level event temporal relation extraction with context information[J]. Journal of Computer Research and Development, 2021, 58(11): 2475.
- [14] 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 ACL. Stroudsburg, PA: ACL, 2017; 224-230.
- [15] FENG X C, HUANG L F, TANG D Y, et al. A language-independent neural network for event detection[C]// Proceedings of the 54th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2016; 66-71.
- [16] GUO F Y, HE R F, DANG J W. Implicit discourse relation recognition via a BiLSTM-CNN architecture with dynamic chunk-based max pooling[J]. IEEE Access, 2019, 7: 169281-169292.
- [17] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]// Proceedings of the 15th European Conference on Computer Vision (ECCV). Berlin, Germany: Springer, 2018; 3-19.
- [18] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [19] PENNINGTON J, SOCHER R, MANNING C. Glove: Global vectors for word representation[C]// Proceedings of Conference on the 2014 Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2014; 1532-1543.
- [20] LI S, ZHAO Z, HU R F, et al. Analogical reasoning on Chinese morphological and semantic relations[C]// Proceedings of the 56th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2018; 138-143.
- [21] CAO P F, CHEN Y B, LIU K, et al. Adversarial transfer learning for Chinese named entity recognition with self-attention mechanism[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2018; 182-192.
- [22] LIU X, OU J, SONG Y, et al. On the Importance of Word and Sentence Representation Learning in Implicit Discourse Relation Classification[C]// Proceedings of the 29th International Joint Conference on Artificial Intelligence. Amsterdam; Elsevier, 2020; 3830-3836.
- [23] GUO F, HE R, DANG J, et al. Working memory-driven neural networks with a novel knowledge enhancement paradigm for implicit discourse relation recognition [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2020; 7822-7829.
- [24] JIANG D, HE J. Tree Framework With BERT Word Embedding for the Recognition of Chinese Implicit Discourse Relations[J]. IEEE Access, 2020, 8: 162004-162011.
- [25] FAN Z W, ZHANG M, LI Z H. BiLSTM-based Implicit Discourse Relation Classification Combining Self-attention Mechanism and Syntactic Information[J]. Computer Science, 2019, 46(5): 214-220.



JIN Fang-yan, born in 1998, postgraduate. His main research interests include financial technology and natural language processing.



WANG Xiu-li, born in 1977, Ph.D, professor, is a senior member of China Computer Federation. His main research interests include financial technology, artificial intelligence and security.

(责任编辑:何杨)