

融合上下文信息的篇章级事件时序关系抽取方法

王俊^{1,3} 史存会^{1,3} 张瑾² 俞晓明¹ 刘悦¹ 程学旗^{2,3}

¹(中国科学院计算技术研究所数据智能系统研究中心 北京 100190)

²(中国科学院网络数据科学与技术重点实验室(中国科学院计算技术研究所) 北京 100190)

³(中国科学院大学 北京 100049)

(wyswangjun@163.com)

Document-Level Event Temporal Relation Extraction with Context Information

Wang Jun^{1,3}, Shi Cunhui^{1,3}, Zhang Jin², Yu Xiaoming¹, Liu Yue¹, and Cheng Xueqi^{2,3}

¹(Data Intelligence System Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

²(CAS Key Laboratory of Network Data Science and Technology (Institute of Computing Technology, Chinese Academy of Sciences), Beijing 100190)

³(University of Chinese Academy of Sciences, Beijing 100049)

Abstract Event temporal relation extraction is an important natural language understanding task, which can be widely used in downstream tasks such as construction of knowledge graph, question answering system and narrative generation. Existing event temporal relation extraction methods often treat the task as a sentence-level event pair classification problem, and solve it by some classification model. However, based on limited local sentence information, the accuracy of the extraction of temporal relations among events is low and the global consistency of the temporal relations cannot be guaranteed. For this problem, this paper proposes a document-level event temporal relation extraction with context information, which uses the neural network model based on Bi-LSTM (bidirectional long short-term memory) to learn the temporal relation expressions of event pairs, and then uses the self-attention mechanism to combine the information of other event pairs in the context, to obtain a better event temporal relation expression for temporal relation classification. At last, that event temporal relation expression with context information will improve the global event temporal relation extraction by enhancing temporal relation classification of all event pairs in the document. Experiments on TB-Dense (timebank dense) dataset and MATRES (multi-axis temporal relations for start-points) dataset show that this method can achieve better results than the latest sentence-level methods.

Key words event temporal relation extraction; temporal relation classification; event relation identification; self-attention; bidirectional long short-term memory (Bi-LSTM)

摘要 事件时序关系抽取是一项重要的自然语言理解任务,可以广泛应用于诸如知识图谱构建、问答系统等任务。已有事件时序关系抽取方法往往将该任务视为句子级事件对的分类问题,而基于有限的局部句子信息导致其抽取的事件时序关系的精度较低,且无法保证整体时序关系的全局一致性。针对此问题,

收稿日期:2020-08-14;修回日期:2020-12-23

基金项目:国家自然科学基金面上项目(91746301,61772498);国家重点研发计划项目(29198220,2017YFC0820404)

This work was supported by the General Program of the National Natural Science Foundation of China (91746301, 61772498) and the National Key Research and Development Program of China (29198220, 2017YFC0820404).

通信作者:史存会(shicunhui@ict.ac.cn)

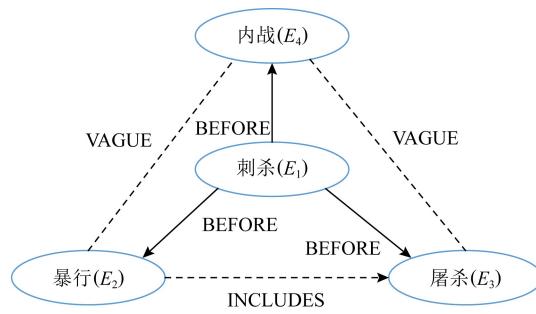
提出一种融合上下文信息的篇章级事件时序关系抽取方法,使用基于双向长短期记忆(bidirectional long short-term memory, Bi-LSTM)的神经网络模型学习文章中事件对的时序关系表示,再利用自注意力机制融入上下文中其他事件对信息,从而得到更丰富的事件对时序关系表示用于时序关系分类。通过 TB-Dense(timebank dense)和 MATRES(multi-axis temporal relations for start-points)数据集的实验表明:此方法能够取得比当前主流的句子级方法更佳的抽取效果。

关键词 事件时序关系抽取;时序关系分类;事件关系识别;自注意力;双向长短期记忆

中图法分类号 TP391

文章如新闻通常描述一系列事件的发生,这些事件看似离散地被叙述着,其实存在着一定的联系,其中最重要的一种事件联系为时序关系。时序关系表示事件发生的先后顺序,其串联了文章中事件的发展演化。如果能准确地抽取文章中的事件时序关系,将有助于理解文章信息,梳理事件脉络。因此,事件时序关系抽取成为了一项重要的自然语言理解任务,受到越来越多的关注。

事件时序关系抽取的目标为抽取文本中包含的事件时序关系,如图 1 中的例子,其包含 4 个事件:刺杀(E_1)、暴行(E_2)、屠杀(E_3)和内战(E_4),其中可抽取出时序图中的 6 对事件时序关系。图 1 中时序关系 BEFORE 表示事件在另一个事件之前发生;时序关系 INCLUDES 表示一个事件包含另一个事件;时序关系 VAGUE 表示两个事件之间不存在特定的时序关系。



S₁:刺杀行动(E_1)引发了胡图人安全部队的杀人暴行(E_2),他们主要屠杀(E_3)图西人,但也屠杀那些主张和解的胡图人。
S₂:它也重新点燃了内战(E_4)。

Fig. 1 An sample of event temporal relation extraction

图 1 事件时序关系抽取样例

目前,与实体关系抽取^[1]类似,已有的事件时序关系抽取方法往往将事件时序关系抽取任务视为句子级事件对的分类问题,以事件对和事件对所在的句子信息作为输入,使用基于规则、基于传统机器学习或基于深度学习的方法识别事件对的时序关系类别。然而,这种句子级的事件时序关系抽取方法使用的事件对所在的句子信息十分有限,往往不足以支

持事件时序关系的识别,限制了事件时序关系识别的精度。同时,句子级的事件时序关系抽取方法孤立地识别事件对的时序关系,未考虑文章中其他事件对的时序关系,其抽取的事件时序关系可能与其他事件时序关系冲突(即不满足时序关系的自反性、传递性等),使得全文的事件时序关系一致性难以得到保证。

为此,本文提出一种融合上下文信息的篇章级事件时序关系抽取方法。该方法使用基于双向长短期记忆(bidirectional long short-term memory, Bi-LSTM)的神经网络模型学习文章中每个事件对的时序关系表示,再使用自注意力机制学习待识别事件对与其文章上下文中其他事件对的联系,利用联系结合上下文中其他事件对的信息,得到更优的事件对的时序关系表示,从而增强事件时序关系抽取模型的效果。为了验证模型的性能,本文在 TB-Dense(timebank dense)数据集^[2]和 MATRES(multi-axis temporal relations for start-points)数据集^[3]上进行实验,实验结果证明本文方法的有效性。

1 相关工作

按照事件时序关系抽取研究的发展,其方法主要可以分为 3 类:基于规则的事件时序关系抽取方法、基于传统机器学习的事件时序关系抽取方法和基于深度学习的事件时序关系抽取方法。

1.1 基于规则的事件时序关系抽取方法

事件时序关系抽取研究发展的早期,由于缺乏相关的成熟语料库,主流的事件时序关系抽取方法通常是基于人工制定的规则展开的。Passonneau^[4]提出使用文本中的时态与体态信息来构造规则,用以推断事件的时序关系。Hitzeman 和 Moens 等人^[5]提出的话语时序结构分析方法,考虑了时态、体态、时间状语和修辞结构的影响。

基于规则的方法实现简单,但其效果极大地依赖于规则的数量和质量,其实用性不高。

1.2 基于传统机器学习的事件时序关系抽取方法

随着国际标准化组织制定了 TimeML^[6] (time markup language) 标注体系,以及时序关系抽取领域的一些重要标准数据集(如 TimeBank^[7])的构建。传统的统计机器学习方法开始被广泛应用于事件时序关系识别研究。Mani 与 Schiffman 等人^[8]提出基于决策树分类模型的新闻事件时序关系识别方法,该方法使用时序关系连接词、时态、语态和时间状语等特征。之后,Mani 等人^[9]又提出基于最大熵分类算法的时序关系分类方法,使用数据集给定的 TimeML 特征(如事件类别、时态、体态等)。Chambers 等人^[10]在 Mani 等人提出方法的特征之上,添加了词性、句法树结构等词法和句法特征,以及 WordNet 中获得的形态学特征,较之前的方法有了 3% 的提升。Ning 等人^[11]使用手工制作的特征作为输入,在训练过程中对全局的事件时序关系结构进行建模,利用结构化学习优化时间图的全局一致性。

基于传统机器学习的事件时序关系抽取的方法研究重点主要集中在特征工程,方法的有效性极大地依赖于特征工程的设计。

1.3 基于深度学习的事件时序关系抽取方法

随着深度学习技术的发展与兴起,神经网络模型被引入到事件时序关系抽取任务中来。Meng 等人^[12]提出一种简单的基于 LSTM(long short-term memory) 网络结构的事件时序关系抽取模型。该模型以事件之间的最短的句法依赖关系路径作为输入,识别相同句子或相邻句子中的事件时序关系类别,在 TB-Dense 数据集上取得不错效果。类似地,Cheng 等人^[13]采取最短句法依赖关系路径作为输入,构造了一种基于双向长期记忆的神经网络模型,同样取得不错的成绩。之后,Han 等人^[14]提出了一种结合结构化学习的神经网络模型。该模型包括递归神经网络来学习成对事件时序关系的评分函数,以及结构化支持向量机(structured support vector machine, SSVM) 进行联合预测。其中神经网络自动学习事件所在情境的时序关系表示,来为结构化模型提供鲁棒的特征,而 SSVM 则将时序关系的传递性等领域知识作为约束条件,来做出更好的全局一致决策,通过联合训练提升整体模型性能。

基于深度学习的事件时序关系抽取方法,相较于基于传统机器学习的方法,能自动学习并表示特征,无需繁杂的特征工程且模型性能更高而泛化能力更强,如今已经成为了事件时序关系抽取领域的研究趋势。

2 结合上下文的篇章级事件时序关系抽取

现有的事件时序关系抽取方法,无论是基于规则还是基于传统机器学习或深度学习,其本质上都是一个句子级的事件对的时序关系分类器。其实现方式为

$$r_{ij} = f(sent_{ij}), \quad (1)$$

其中, r_{ij} 表示任意事件对 $\langle e_i, e_j \rangle$ 的时序关系, f 为时序关系分类器, $sent_{ij}$ 为事件对所在句子。可以看出时序关系 r_{ij} 仅基于事件对所在的句子信息,而这些有限的局部信息导致识别的精度较低且无法保证全文整体的时序关系的一致性。

针对句子级事件时序关系抽取方法的局限性,本文提出结合上下文信息的篇章级事件时序关系抽取模型。可形式化为

$$r_{ij} = f(sent_{ij}, context_{ij}), \quad (2)$$

其中, $context_{ij}$ 为事件对 $\langle e_i, e_j \rangle$ 的上下文,即事件对所在文章信息。

2.1 上下文信息增强的事件时序关系抽取模型

本文提出篇章级的事件时序关系抽取模型,称为上下文信息增强的事件时序关系抽取模型(context information enhanced event temporal relation extraction model, CE-TRE)。模型的整体框架如图 2 所示,主要包含 3 个部分:1)事件对的时序关系编码模块;2)事件对的上下文增强模块;3)输出模块。其中,我们分别将每个事件对的句子级时序关系表示 S_k 通过全连接层得到 Query: Q_k , Key: K_k , Value: V_k 。CE-TRE 模型以文章作为基本输入单元。按照事件对分句后,事件对的时序关系编码模块使用基于 Bi-LSTM 的神经网络模型学习事件对的时序关系表示;事件对上下文增强模块利用自注意力机制结合上下文中其他事件对的信息,得到更完备的事件对时序关系表示;输出模块根据事件对时序关系表示预测事件对的时序关系类别。

2.2 时序关系编码模块

时序关系编码模块用于初步编码每个事件对的句子级时序关系表示。CE-TRE 模型以文章作为输入单元,对于每个文章输入,先将其按一个个事件对把文章分为一个个句子(存在一个句子包含多个事件对的情况),得到所有事件对的所在句子集合。时序关系编码模块就以每个事件对的所在句子作为输入,编码事件对的句子级时序关系表示。具体如图 3 所示。

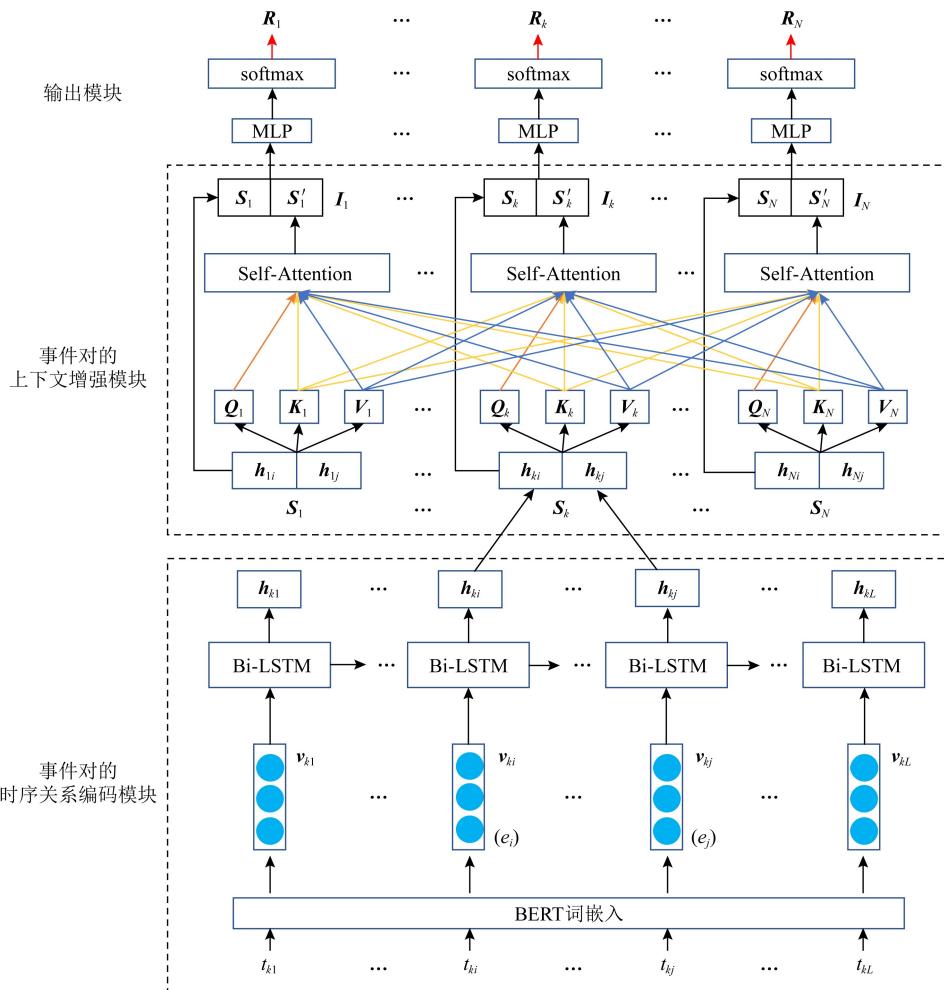


Fig. 2 Context information enhanced event temporal relation extraction model

图 2 上下文信息增强的事件时序关系抽取模型

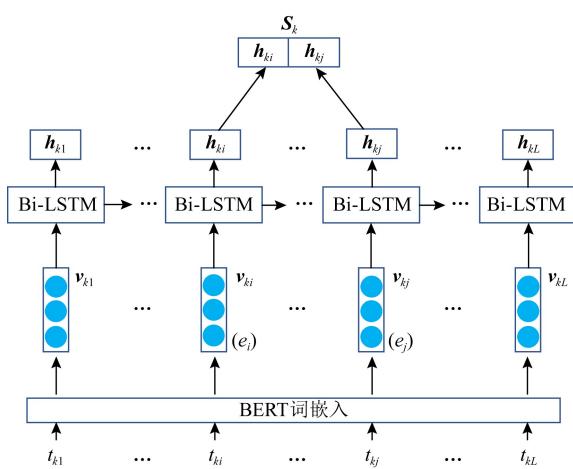


Fig. 3 Temporal relation coding module

图 3 时序关系编码模块

给定事件对所在的句子(可以是单独的1个句子或连续的2个句子),表示为 $t_k = (t_{k1}, \dots, t_{ki}, \dots, t_{kj}, \dots, t_{kL})$,其中 k 表示该句子是文章按事件对分

句后的第 k 个句子, t_{ki} 和 t_{kj} 为事件对应的词.

首先本文使用预训练好的 BERT^[15] 模型计算每个词的词嵌入 v_{ki} , 得到句子的向量表示 $sent_k$. 使用 BERT 计算词嵌入的目的是为了得到上下文有关的词向量表示,使得后续模块能更好地学习事件对的时序关系表示. 因为预训练的 BERT 只用于计算词向量,不参与整体模型的训练,因此不会影响整体模型的训练难度. 形式化过程为

$$sent_k = (v_{k1}, v_{k2}, \dots, v_{kL}) = Embedding_{BERT}(t_k). \quad (3)$$

接着,将事件对所在句子 $sent_k$ 通过 Bi-LSTM 模型,可以得到句子长度的隐藏状态序列 $\mathbf{H}_k = (h_{k1}, \dots, h_{ki}, \dots, h_{kj}, \dots, h_{kL})$. 将 2 个事件的位置(i 和 j)相对应的隐藏状态 h_{ki} 和 h_{kj} 串联起来,得到事件对的句子级时序关系向量表示 $\mathbf{S}_k = (h_{ki}; h_{kj})$. 上述过程可形式化为

$$(h_{k1}, h_{k2}, \dots, h_{kL}), h_{kL} = Bi_LSTM(sent_k, h_{k0}), \quad (4)$$

其中, \mathbf{h}_{ki} 是由正向编码向量 $\overrightarrow{\mathbf{h}}_{ki}$ 和反向编码向量 $\overleftarrow{\mathbf{h}}_{ki}$ 拼接而成, 即 $\mathbf{h}_{ki} = (\overrightarrow{\mathbf{h}}_{ki}; \overleftarrow{\mathbf{h}}_{ki})$, 而 $\mathbf{S}_k = (\overrightarrow{\mathbf{h}}_{ki}; \overleftarrow{\mathbf{h}}_{ki}; \overrightarrow{\mathbf{h}}_{kj}; \overleftarrow{\mathbf{h}}_{kj})$. 通过拼接 Bi-LSTM 的前后双向的隐层输出, 可以得到综合句子前后文信息的事件对的时序关系表示.

最后, 1 个句子可能包含多个事件对, 因此本模块针对 1 个句子输入, 可能输出多个事件对向量表示, 一起汇总成文章的事件对表示集合.

2.3 事件对上下文增强模块

本文将时序关系编码模块得到的每个事件对的句子级的时序关系表示, 输入到事件对上下文增强模块, 得到上下文信息增强的事件对时序关系表示. 具体如图 4 所示:

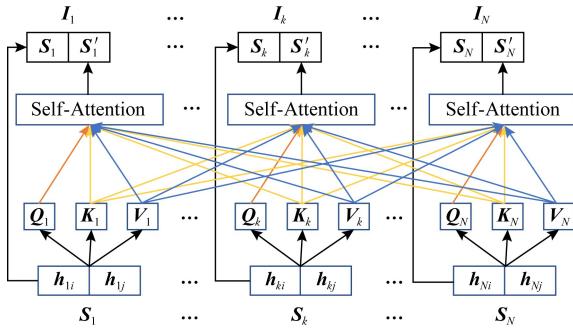


Fig. 4 Context enhancement module for event pairs

图 4 事件对上下文增强模块

为了利用上下文中其他事件对的时序关系信息, 本文利用 Self-Attention 机制^[16]学习事件对与上下文中其他事件对的联系权重, 使用权重对上下文中的事件对时序关系表示进行加权求和, 得到一

个综合上下文信息的事件对时序关系表示 \mathbf{S}'_k . 而最终的上下文增强后的事件对时序关系表示还需要通过残差连接, 把 Self-Attention 的输入 \mathbf{S}_k 与输出 \mathbf{S}'_k 拼接, 得到 $\mathbf{I}_k = (\mathbf{S}_k; \mathbf{S}'_k)$. 不同于常规残差连接的将模块的输入与输出直接相加, 这里使用拼接式的残差网络. 这种残差连接的目的是让模型训练时既能学习较好的事件对的句子级时序关系表示, 又能学习事件对之间的联系, 从而更好地增强事件对的时序关系表示, 使得模型能够优化得更好.

具体的 Self-Attention 计算流程如图 5 所示, 使用注意力机制加权融合上下文中每个事件对的时序关系表示 \mathbf{V}_i , 得到该事件对的上下文增强的时序关系表示向量 \mathbf{S}'_i , 该过程形式化表示为:

- 1) 对于事件对 \mathbf{S}_i , 首先计算其与其他事件对 \mathbf{S}_j 之间的权重 w'_{ij} , 然后将权重归一化为 w_{ij} . 其中, d_k 表示 $\mathbf{Q}_k, \mathbf{K}_k, \mathbf{V}_k$ 的数据维度, 点积除以 d_k , 用于缩放点积 $\mathbf{Q}_i \mathbf{K}_j^T$, 防止点积过大时模型梯度更新不稳定或难以有效地优化模型.

$$\mathbf{w}'_{ij} = \frac{\mathbf{Q}_i \mathbf{K}_j^T}{\sqrt{d_k}}, \quad \mathbf{w}_{ij} = \frac{e^{w'_{ij}}}{\sum_t e^{w'_{it}}}. \quad (5)$$

- 2) 使用权重 w_{ij} 对与所有事件对的 \mathbf{V}_j 进行加权求和, 得到事件对 \mathbf{S}_i 的上下文增强的时序关系表示向量 \mathbf{S}'_i .

$$\mathbf{S}'_i = \sum_{j=1}^N w_{ij} \mathbf{V}_j. \quad (6)$$

该过程可简写为

$$\text{Self_Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}. \quad (7)$$

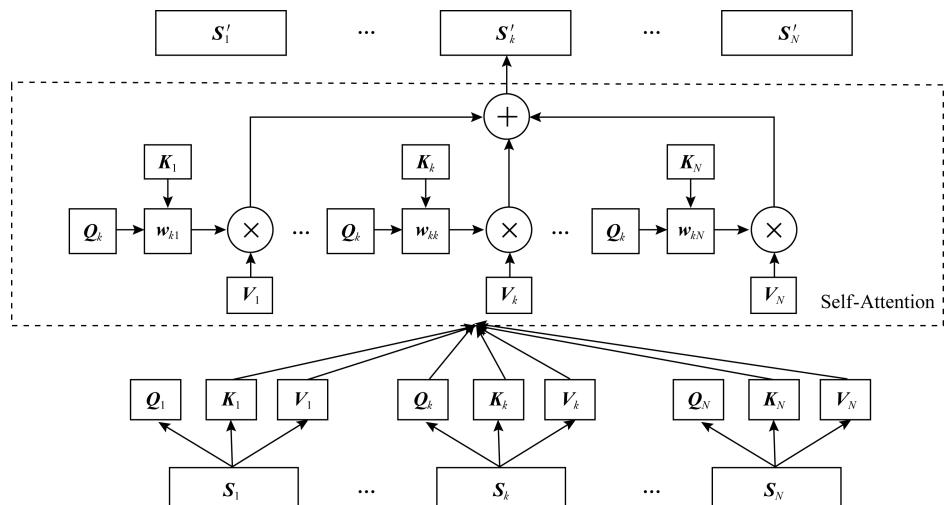


Fig. 5 Self-Attention diagram

图 5 Self-Attention 示意图

2.4 输出模块

本模块用于输出事件对的时序关系的分类预测值。如图 6 所示,本模块将上下文增强后的事件对时序关系表示,通过一层全连接层和用于分类的 softmax 层,得到该事件对的 T 维的时序关系概率向量 \mathbf{R} ,其中第 j 个元素表示将其判别为第 j 个时序关系类型的概率值。我们采用真实时序关系类型和预测概率的交叉熵误差作为损失函数:

$$J(\mathbf{y}, \mathbf{r}) = \sum_{i=1}^N \sum_{j=1}^T I(y_i=j) \ln r_i^{(j)}, \quad (8)$$

其中, N 为事件对总数, I 为指示函数。

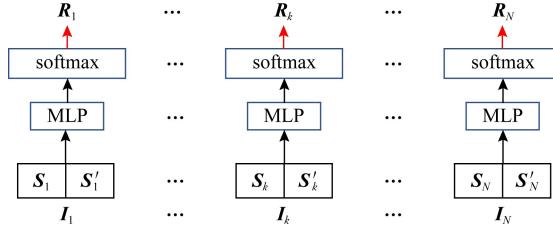


Fig. 6 Output module

图 6 输出模块

3 实验

本节主要介绍实验部分的相关细节。首先介绍本文所使用的公开数据集与实验的评价指标,接着介绍实验对比的基线方法,随后介绍实验的参数设置,最后对实验结果进行分析与探讨。

3.1 数据集与评价指标

本文在公开的 TB-Dense 和 MATRES 数据集上对模型的效果和性能进行评估,这 2 个数据集的规模信息如表 1 所示:

Table 1 Statistics of TB-Dense and MATRES Datasets

表 1 数据集 TB-Dense 和 MATRES 的规模统计

统计维度	数据拆分	TB-Dense	MATRES
文章数	训练集	22	183
	开发集	5	
	测试集	9	20
事件对数	训练集	4 032	6 332
	开发集	629	
	测试集	1 427	827

表 1 详细说明为:

1) TB-Dense 数据集是 Cassidy 等人^[2] 基于

TimeBank 数据集,通过标注相同或相邻句子中的所有事件对的时序关系,构建的稠密标注语料,解决了 TimeBank 的标注稀疏性,近年来已被广泛应用于事件时序关系抽取研究。TB-Dense 包含 6 类时序关系:VAGUE, BEFORE, AFTER, SIMULTANEOUS, INCLUDES, IS_INCLUDED。

2) MATRES 是 Ning 等人^[3] 在 2018 年基于 TempRels3 数据集(TempEval 第 3 次评测任务^[17])构建的数据集,包括 TimeBank, AQUAINT(advanced question-answering for intelligence), Platinum 构建的新数据集。该数据集通过使用多轴注释方案并采用事件起点比较时序来改进注释者之间的一致性,进一步提高了数据质量,成为近几年来值得关注的一个新数据集。MATRES 只包含 4 类时序关系:VAGUE, BEFORE, AFTER, SIMULTANEOUS。

为了与已有相关研究进行对比,本实验采用 2 套微平均 $F1$ 值作为评价指标,具体为:

1) 针对数据集 TB-Dense, 使用全部 6 个时序关系类别的 Micro- $F1$ 。

2) 针对数据集 MATRES, 使用除了“VAGUE”外的 3 个时序关系类别的 Micro- $F1$ 。

3.2 基准方法

1) CAEVO. Chambers 等人^[18] 于 2014 年提出的有序筛网式流水线模型。其中每个筛子可以是基于规则的分类器,也可以是机器学习模型。

2) CATENA. Mirza 等人^[19] 于 2016 年提出了多重筛网式的事件关系抽取系统,可利用时序关系抽取模型和因果关系抽取模型之间的交互作用,增强时序和因果关系的提取和分类。

3) 文献[13]方法.Cheng 等人^[13] 于 2017 年提出了一种基于双向长短期记忆(Bi-LSTM)的神经网络模型,该模型采取最短句法依赖关系路径作为输入,能达到与基于人工特征相当的性能。

4) 文献[20]方法.Vashishtha 等人^[20] 于 2019 年提出了一个用于建模细粒度时序关系和事件持续时间的新语义框架,该框架将成对的事件映射到实值尺度,以构建文档级事件时间轴。基于此框架,训练模型联合预测细粒度的时序关系和事件持续时间。

5) 文献[21]方法.Meng 等人^[21] 于 2018 年提出了上下文感知的事件时序关系抽取的神经网络模型,模型以事件最短依存路径作为输入,按叙述顺序储存处理过的时序关系,用于上下文感知。

6) 文献[22]方法.Han 等人^[22] 于 2019 年提出了一种具有结构化预测的事件和事件时序关系联合

抽取模型.该模型让事件抽取模块和事件时序关系抽取模块共享相同的上下文嵌入和神经表示学习模块,从而改善了事件表示.利用结构化的推理共同分配事件标签和时序关系标签,避免了常规管道模型中的错误传递.

3.3 实验参数设置

CE-TRE 中,Bi-LSTM 层的输出维度为 100,线性层的输入维度为 400 和 100;模型训练的 *batch_size* 大小为 1(因为篇章级的事件时序关系抽取以文章为基本单位,一篇文章包含若干的事件对).

此外由于篇章级时序关系抽取以文章为输入单位,而不同文章包含的事件对的数目是不同的.这种差异导致不同 *batch* 包含的事件对数目存在明显差异,如果使用一样的学习率进行训练,会使得模型更新不均衡,导致模型优化过程不稳定.因此,本文使用的解决策略为设置动态学习率,为包含事件对多的文章样例增大学习率,即学习率与事件对数目正相关:

$$lr = rel_num \times lr_base \times (lr_decay^{\lfloor \frac{epoch}{5} \rfloor}), \quad (9)$$

其中,*lr_base* 为基准学习率,设置为 0.0001;*lr_decay* 为学习率衰减,设置为 0.9,每 5 轮进行 1 次衰减;*rel_num* 为文本包含的事件对数目,学习率与其成正比.

3.4 实验结果与分析

为了验证本文提出的模型 CE-TRE 在不同数据集上的抽取效果,本文选取了近几年的主流模型(如 3.2 节所述)与 CE-TRE 在公开的 TB-Dense 和 MATRES 数据集上进行对比实验.

3.4.1 验证 CE-TRE 模型性能

实验 1. 在 TB-Dense 数据集上验证 CE-TRE 模型性能.

在数据集 TB-Dense 上实验的原因是该数据是事件时序关系抽取任务的传统数据集,有着充分的前人工作可以对比.实验结果如表 2 所示,与近几年的主流模型对比,本文提出的 CE-TRE 模型在 TB-Dense 取得了最高的 Micro-F1 评分,这说明 CE-TRE 模型很有竞争力.

实验 2. 在 MATRES 数据集上验证 CE-TRE 模型性能.

在数据集 MATRES 上实验的原因是该数据是个新的标注质量更高且数据规模更大的数据集,可以进一步验证本文提出模型的效果.实验结果如表 3 所示,在 MATRES 数据集上,本文提出的 CE-TRE 模型的效果优于 2019 年最优的文献[22]方法^[22].

Table 2 Comparative Experiment of CE-TRE and 6 Baselines on TB-Dense

表 2 CE-TRE 模型与 6 个主流模型在数据集 TB-Dense 上的对比实验

模型	Micro-F1
CAEVO	0.494
CATENA	0.519
文献[13]方法	0.529
文献[20]方法	0.566
文献[21]方法	0.57
文献[22]方法	0.645
CE-TRE	0.657

注:粗体值为最优值.

Table 3 Comparative Experiment of CE-TRE and the Best Baseline on MATRES

表 3 CE-TRE 模型与最优基线在数据集 MATRES 上的对比实验

模型	Micro-F1
文献[22]方法	0.755
CE-TRE	0.766

注:粗体值为最优值.

3.4.2 验证结合上下文信息对模型的增强效果

实验 3. 验证结合上下文的有效性.

为了验证结合上下文信息,是否对事件时序关系抽取模型有着增益效果,本实验对比有无上下文信息增强的模型的效果.实验结果如表 4 所示,在 2 个数据集上,结合上下文信息的 CE-TRE 模型均优于未结合上下文信息的 TRE(without CE)模型.此结果表明,结合上下文信息确实能够增强事件对时序关系抽取模型的效果.

Table 4 Comparative Experiment Between CE-TRE and TRE (without CE)

表 4 CE-TRE 模型与去除上下文增强后的 TRE 模型 (without CE) 的对比实验

模型	Micro-F1	
	TB-Dense	MATRES
TRE(without CE)	0.619	0.757
CE-TRE	0.657	0.766

注:粗体值为最优值.

此外,通过对比 2 个数据集上的实验结果可以发现,CE-TRE 模型相比于 TRE 模型,在数据集 TB-Dense 上的性能提升较为显著(采用 McNemar

检验,所得 $p < 0.005$),而在数据集 MATRES 上的性能差距较小.本文认为原因是:MATRES 数据集的标注类别只包含 4 类,相较于 TB-Dense 的 6 类,MATRES 中的事件时序关系的识别难度较低.同时,MATRES 数据集中的句子平均长度长于 TB-Dense 数据集,通常情况下,MATRES 的句子包含的时序关系信息也会更多.这 2 点使得未结合上下文信息的 TRE 模型在 MATRES 数据集上也能取得较高的得分,从而 CE-TRE 模型相比 TRE 模型提升较小.同时,TRE(without CE)模型在 2 个数据集上的性能还优于许多基线模型,本文认为:TRE 虽未结合上下文信息,但模型的训练方式和 CE-TRE 一样,以文章作为单位输入,使得模型的每次更新优化只针对 1 篇文本的数据,可以集中学习每篇文章所属文意的时序关系,从而学习出更优的模型.

3.4.3 实验结果样例分析

如表 5 中的样例,需判断事件 E_1 与事件 E_2 之间的时序关系.对于未结合上下文信息的事件对时序关系识别模型 TRE,其预测结果为“VAGUE”,预测错误.而结合上下文信息的事件对时序关系识别模型 CE-TRE 的预测结果为“BEFORE”,预测正确.通过分析样例可知,TRE 预测错误的原因是样例的句子信息不足以判别事件对的时序关系.而 CE-TRE 通过结合上文中的“1991 年,查尔斯·基廷在州法院被判协助诈骗数千名投资者,这些投资者购买了基廷在林肯储蓄贷款公司员工出售的高风险垃圾债券.”的信息可知,基廷于 1991 年被判协助诈骗,导致投资者购买了垃圾债券,之后债券变得一文不值.由此推断 1996 年基廷的判决被推翻发生在债券变得一文不值之后.

Table 5 A Sample from TB-Dense Dataset

表 5 来自 TB-Dense 数据集的样例

选项	内容
事件对	E_1 : bonds became worthless(债券变得一文不值) E_2 : convictions were thrown out(判决被推翻)
事件对所在句子	The bonds became worthless when the bankrupt thrift was seized by government regulators. Keating's convictions were thrown out in 1996 on a technicality. 译文:当破产的储蓄银行被政府监管机构接管时,这些债券变得一文不值.基廷的判决在 1996 年因为一个技术细节而被推翻.
真实标签	BEFORE
TRE	VAGUE
CE-TRE	BEFORE

由 TRE 和 CE-TRE 在样例上的预测结果进一步验证了结合上下文的有效性.

综合实验 1 和实验 2 的结果,本文提出的 CE-TRE 模型在 2 个数据集上均取得优于近年来的最新模型,说明了本文方法的有效性.其次通过对比有上下文增强的 CE-TRE 模型和无上下文增强的 TRE(without CE)模型的评分,以及两者在表 5 中样例的预测结果可以看出,结合上下文信息确实能够增强事件对时序关系抽取模型的效果.

4 总 结

事件时序关系抽取技术是一种从文本中获取事件的时序结构信息的重要手段,有着很高的研究价值和实用价值.现有的事件时序关系抽取方法往往都是句子级的抽取方法,存在时序关系识别精度低且无法保证全文一致性的问题.本文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,并通过 TB-Dense 数据集和 MATRES 数据集上的多组实验验证了本文方法的有效性.

目前事件时序关系抽取的数据集规模都较小,限制了神经网络模型的抽取性能.在未来的工作中,我们将进一步研究如何丰富相关数据集或如何引入外部资源来进一步提升事件时序关系抽取模型的性能.

参 考 文 献

- [1] Li Dongmei, Zhang Yang, Li Dongyuan, et al. Review of entity relation extraction methods [J]. Journal of Computer Research and Development, 2020, 57(7): 1424–1448 (in Chinese)
(李冬梅, 张扬, 李东远, 等. 实体关系抽取方法研究综述 [J]. 计算机研究与发展, 2020, 57(7): 1424–1448)
- [2] Cassidy T, McDowell B, Chambers N, et al. An annotation framework for dense event ordering [C] //Proc of the 52nd Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2014: 501–506
- [3] Ning Qiang, Wu Hao, Roth D. A multi-axis annotation scheme for event temporal relations [C] //Proc of the 56th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2018: 1318–1328
- [4] Passonneau R J. A computational model of the semantics of tense and aspect [J]. Computational Linguistics, 1988, 14(2): 44–60
- [5] Hitzeman J, Moens M, Grover C. Algorithms for analysing the temporal structure of discourse [C] //Proc of the 7th Conf on European Chapter of the ACL. Stroudsburg, PA: ACL, 1995: 253–260

- [6] Saurí R, Littman J, Knippen B, et al. TimeML annotation guidelines version 1.2.1 [J/OL]. 2006 [2020-11-03]. https://www.researchgate.net/publication/248737128_TimeML_annotation_Guidelines_Version_121
- [7] Pustejovsky J, Hanks P, Saurí R, et al. The timebank corpus [C/OL] //Proc of Corpus Linguistics. 2003: 647–656 [2020-11-03]. https://www.researchgate.net/publication/228559081_The_TimeBank_corpus
- [8] Mani I, Schiffman B, Zhang Jianping. Inferring temporal ordering of events in news [C] //Proc of HLT-NAACL. Stroudsburg, PA: ACL, 2003: 55–57
- [9] Mani I, Verhagen M, Wellner B, et al. Machine learning of temporal relations [C] //Proc of the 21st Int Conf on Computational Linguistics and the 44th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2006: 753–760
- [10] Chambers N, Wang Shan, Jurafsky D. Classifying temporal relations between events [C] //Proc of the 45th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2007: 173–176
- [11] Ning Qiang, Feng Zhili, Roth D. A structured learning approach to temporal relation extraction [C] //Proc of the 2017 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2017: 1027–1037
- [12] Meng Yuanliang, Rumshisky A, Romanov A. Temporal information extraction for question answering using syntactic dependencies in an LSTM-based architecture [C] //Proc of the 2017 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2017: 887–896
- [13] Cheng Fei, Miyao Y. Classifying temporal relations by bidirectional LSTM over dependency paths [C] //Proc of the 55th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2017: 1–6
- [14] Han Rujun, Hsu I H, Yang Mu, et al. Deep structured neural network for event temporal relation extraction [C] //Proc of the 23rd Conf on Computational Natural Language Learning (CoNLL). Stroudsburg, PA: ACL, 2019: 666–106
- [15] Devlin J, Chang Mingwei, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C] //Proc of the 2019 Conf of the North American Chapter of the ACL: Human Language Technologies. Stroudsburg, PA: ACL, 2019: 4171–4186
- [16] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C] //Proc of the Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2017: 5998–6008
- [17] UzZaman N, Llorens H, Derczynski L, et al. Semeval-2013 task 1: TempEval-3: Evaluating time expressions, events, and temporal relations [C/OL] //Proc of the 2nd Joint Conf on Lexical and Computational Semantics. Stroudsburg, PA: ACL, 2013 [2020-11-03]. https://www.researchgate.net/publication/228059143_TempEval-3_Evaluating_Events_Time_Expressions_and_Temporal_Relations
- [18] Chambers N, Cassidy T, McDowell B, et al. Dense event ordering with a multi-pass architecture [J]. Transactions of the Association for Computational Linguistics, 2014, 2(11): 273–284
- [19] Mirza P, Tonelli S. Catena: Causal and temporal relation extraction from natural language texts [C] //Proc of the 26th Int Conf on Computational Linguistics. Stroudsburg, PA: ACL, 2016: 64–75
- [20] Vashishtha S, Van Durme B, White A S. Fine-grained temporal relation extraction [C] //Proc of the 57th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2019: 2906–2919
- [21] Meng Yuanliang, Rumshisky A. Context-aware neural model for temporal information extraction [C] //Proc of the 56th Annual Meeting of the ACL. Stroudsburg, PA: ACL, 2018: 527–536
- [22] Han Rujun, Ning Qiang, Peng Nanyun. Joint event and temporal relation extraction with shared representations and structured prediction [C] //Proc of the 2019 Conf on Empirical Methods in Natural Language Processing and the 9th Int Joint Conf on Natural Language Processing (EMNLP-IJCNLP). Stroudsburg, PA: ACL, 2019: 434–444



Wang Jun, born in 1994. Master. His main research interests include natural language processing, event relation extraction.

王俊,1994年生,硕士,主要研究方向为自然语言处理、事件关系抽取。



Shi Cunhui, born in 1987. PhD. Engineer. His main research interests include network science, information recommendation, and event extraction.

史存会,1987年生,博士,工程师,主要研究方向为网络科学、信息推荐、事件抽取。



Zhang Jin, born in 1978. PhD. Senior engineer. His main research interests include public opinion analysis, natural language processing, automatic summarization, and distributed processing. (jinzhang@ict.ac.cn)

张瑾,1978年生,博士,高级工程师,主要研究方向为舆情分析、自然语言处理、自动文摘、分布式处理。



Yu Xiaoming, born in 1977. PhD. Senior engineer. His main research interests include information retrieval and big data. (yuxiaoming@ict.ac.cn)

俞晓明,1977年生,博士,高级工程师,主要研究方向为信息检索、大数据。



Liu Yue, born in 1971. PhD. Associate professor. Her main research interests include information retrieval, data mining and social computing. (liuyue@ict.ac.cn)

刘 悅,1971年生.博士,副研究员.主要研究方向为信息检索、数据挖掘、社会计算.



Cheng Xueqi, born in 1971. PhD. Professor and PhD supervisor. Member of CCF. His main research interests include network information security, large-scale information retrieval and knowledge mining. (cxq@ict.ac.cn)

程学旗,1971年生.博士,研究员,博士生导师,CCF会员.主要研究方向为网络信息安全、大规模信息检索与知识挖掘.

2022 年《计算机研究与发展》专题(正刊)征文通知 ——数据挖掘前沿进展

数据挖掘旨在利用智能数据分析技术,从海量数据中提取或挖掘潜在的知识和规律,为决策任务提供有效支撑。在大数据时代背景下,数据挖掘技术已在工业、金融、医疗、教育、交通、媒体等领域取得广泛应用.然而,大数据的复杂多样性也为数据挖掘研究在理论、方法、应用等多个层面提出了新的挑战。

为及时反映国内同行在数据挖掘方面的前沿研究成果,《计算机研究与发展》将于2022年出版“数据挖掘前沿进展”专题.欢迎数据挖掘及相关研究领域的专家学者、科研人员踊跃投稿.此外,专题组稿与第九届中国数据挖掘会议(CCDM 2022)合作,所有专题录用稿件均需在CCDM 2022会议报告交流.

征文范围(但不限于)

1) 数据挖掘理论与方法

分类、聚类、排序、集成学习、强化学习、关联分析、链接分析、频繁模式挖掘、动态数据挖掘、交互式与联机挖掘、并行与分布式挖掘、大规模数据挖掘、数据预处理、数据挖掘语言等.

2) 特定类型数据挖掘与分析

关系数据挖掘、图模式挖掘、空间与时序数据挖掘、数据流与增量挖掘、多媒体数据挖掘、社交网络分析与挖掘、文本挖掘、隐私保护数据挖掘、生物信息数据挖掘、推荐系统、数据仓库等.

3) 数据挖掘技术应用

面向工业、金融、医疗、教育、交通、旅游、管理、电商、电信等领域的数据挖掘技术及应用.

征文要求

1) 论文应属于作者的科研成果,数据真实可靠,具有重要的学术价值与推广应用价值,且未在国内外公开发行的刊物或会议上发表,不存在一稿多投问题.作者在投稿时,需向编辑部提交版权转让协议.

2) 论文一律用word格式排版,论文格式体例参考近期出版的《计算机研究与发展》的要求(<http://crad.ict.ac.cn/>).

3) 论文须通过期刊网站(<http://crad.ict.ac.cn/>)投稿,投稿时提供作者的联系方式,留言中务必注明“数据挖掘2022专题”(否则按自由来稿处理).

4) 论文预录用后,至少有一位作者注册参加CCDM 2022会议并做口头报告.否则,论文视为退出专刊.

重要日期

论文截稿日期:2022年1月8日

录用通知日期:2022年2月28日

最终稿提交日期:2022年3月10日

论文出版日期:2022年8月

特邀编委

张长水 教 授 清华大学 zcs@mail.tsinghua.edu.cn

杨 博 教 授 吉林大学 ybo@jlu.edu.cn

钱 超 副教授 南京大学 qianc@nju.edu.cn

联系方式

编 辑 部:crad@ict.ac.cn, 010-62620696, 010-62600350

通信地址:北京2704信箱《计算机研究与发展》编辑部

邮 编:100190