

基于深度学习的事件因果关系抽取综述

王朱君^{1,2}, 王石^{2*}, 李雪晴^{1,2}, 朱俊武¹

(1. 扬州大学 信息工程学院, 江苏 扬州 225000; 2. 中国科学院 计算技术研究所, 北京 100190)

(* 通信作者电子邮箱 wangshi@ict.ac.cn)

摘要: 因果关系抽取是自然语言处理(NLP)中的一种关系抽取任务,它通过构造事件图来挖掘文本中具有因果关系的事件对,已经在金融、安全、生物等领域的应用中发挥重要作用。首先,介绍了事件抽取和因果关系等概念,并介绍了因果关系抽取主流方法的演变和常用数据集;然后,列举了当前主流的因果关系抽取模型,并且在分别对基于流水线的模型和联合抽取模型进行详细分析的基础上,对比了各种方法和模型的优缺点;此外,对各模型的实验性能及相关实验数据进行了归纳分析;最后,给出了当前的因果关系抽取的研究难点和未来的重点研究方向。

关键词: 因果关系; 自然语言处理; 关系抽取; 事件对; 深度学习

中图分类号: TP391 **文献标志码:** A

Review of event causality extraction based on deep learning

WANG Zhujun^{1,2}, WANG Shi^{2*}, LI Xueqing^{1,2}, ZHU Junwu¹

(1. College of Information Engineering, Yangzhou University, Yangzhou Jiangsu 225000, China;

2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Causality extraction is a kind of relation extraction task in Natural Language Processing (NLP), which mines event pairs with causality from text by constructing event graph, and play important role in applications of finance, security, biology and other fields. Firstly, the concepts such as event extraction and causality were introduced, and the evolution of mainstream methods and the common datasets of causality extraction were described. Then, the current mainstream causality extraction models were listed. Based on the detailed analysis of pipeline based models and joint extraction models, the advantages and disadvantages of various methods and models were compared. Furthermore, the experimental performance and related experimental data of the models were summarized and analyzed. Finally, the research difficulties and future key research directions of causality extraction were given.

Key words: causality; Natural Language Processing (NLP); relation extraction; event pair; deep learning

0 引言

人们的社会活动以事件为驱动,事件是社会活动的载体。在自然语言处理中,事件作为一种信息表示形式显得十分重要,受到越来越多的重视。事件中包含了大量的内部组成结构(如参与者、时间、地点等)和外部关联(如因果、共指、时序等语义关系)。对包含大量事件的文本进行因果关系的抽取可以实现对文本的更加深层次的理解。在航空安全^[1]、医学^[2]等众多领域中,因果关系抽取也发挥着重要的作用。图灵奖获得者 Bengio 表示将因果关系集成到人工智能(Artificial Intelligence, AI)中是一件大事,了解因果关系可以使得现有的人工智能更加智能高效。Pearl 等^[3]讲到如果没有因果推理的能力,人工智能将从根本上受到限制。

因果关系一般可以组织为事理图谱。事理图谱是表示事件之间的顺承、因果关系的有向图^[4]。图中每个节点表示事件,而边用来表示事件之间的因果关系。事理图谱找出了事件的演化逻辑,从而形成大型的常识事理知识库用来直接刻

画人类行为活动。很多的学者在因果关系抽取上得到了理想的结果^[5-7]。

深度学习是机器学习领域如今最热门的研究方向之一,它能够模仿人类大脑的活动,使得模式识别很难解决的问题得到了合理的解决,使得自然语言处理(Natural Language Processing, NLP)相关技术取得了很大进步。近年来基于深度神经网络的因果关系抽取方法比较多,主要分为两大类:一类是基于流水线方式;另一类是基于联合抽取的方式。前者将抽取任务看作是实体识别和关系分类两个子任务,后者则是利用联合模型将因果关系三元组直接抽取出来。两者各有利弊,但是目前还缺乏对整个领域开展深入的梳理和总结。本文对现有的基于深度学习的因果关系抽取方法进行了总结:首先,分别对因果关系抽取的基本概念和方法演变进行了简要介绍,并对常用数据集进行了介绍;接着,对现有的基于深度学习的两种主要框架分别进行了详细介绍;最后,提出了未来的研究方向。

收稿日期:2020-07-23;修回日期:2020-11-09;录用日期:2020-11-17。

基金项目:国家自然科学基金资助项目(61872313);国家重点研发计划重点专项(2017YFB1002300, 2018YFC1700302)。

作者简介:王朱君(1996—),男,江苏东台人,硕士研究生,主要研究方向:自然语言处理;王石(1981—),男,山东博兴人,副研究员,博士,主要研究方向:语义分析、知识图谱;李雪晴(1995—),女,江苏泰州人,博士研究生,主要研究方向:自然语言处理;朱俊武(1972—),男,江苏江都人,教授,博士,CCF 高级会员,主要研究方向:知识工程、本体论。

1 因果关系抽取简介

因果关系抽取是关系抽取中的一类任务。根据原因事件和结果事件是否在文本中同时出现,将因果关系分为了隐式因果(文本中只出现了原因事件或结果事件)和显式因果(文本中原因事件和结果事件同时出现)。隐式因果关系抽取任务中缺失的原因事件或结果事件需要根据文本中的知识进行推理判断。对于显式因果关系,又可以根据原因事件和结果事件的对应关系,分为一对一(一件原因事件对应一件结果事件)、一对多(一件原因事件对应多件结果事件)、多对一(多件原因事件对应一件结果事件)、多对多(多件原因事件对应多件结果事件)^[8]。

1.1 因果关系抽取的基本概念

因果关系表示客观事件间存在的一种普遍联系。事件的因果关系主要由原因事件和结果事件两个部分构成。因果关系抽取任务从描述事件信息的文本中抽取原因事件和结果事件,并以结构化的形式将其呈现出来。一般可以将因果关系抽取分为两个子任务:候选事件对的抽取和因果关系分类。候选事件对抽取需要对文本中的事件触发词和相关论元进行定位抽取;关系分类需要对候选事件对进行过滤,找出属于因果关系的事件对,并确定原因事件和结果事件。

例如,文本“突然地降温使得小明感冒,昨晚去医院了”,事件检测发现句子中有3个事件分别为“降温”“感冒”和“去医院”,那么形成因果对(降温,感冒)、(感冒,去医院)和(降温,去医院);接着判断事件对间是否存在因果关系,因此排除(降温,去医院),抽取出了(降温,感冒)和(感冒,去医院)的因果关系。

1.2 因果关系抽取的方法演变

与传统关系分类^[9]不同的是,因果关系抽取的方法主要分为三种:基于模式匹配的方法、基于模式匹配与机器学习组合的方法和基于深度学习的方法。

1.2.1 基于模式匹配的方法

基于模式的方法利用语义特征、词汇符号特征和自构约束通过模式匹配进行因果关系的提取。Khoo等^[10]利用语言线索和模式匹配从华尔街日报中提取因果知识;Khoo等^[11]在医学领域使用图形模式从医疗数据库中提取因果知识;Girju等^[12]使用句法模式提取因果关系,然后使用语义约束将候选对分类为因果对或非因果对;Ittoo等^[13]提出一种基于词性、句法分析和因果模板的因果对提取方法。完全依赖于模式匹配规则的方法通常跨域适应性差,在解决特定领域的问题时可能需要广泛的领域知识。

1.2.2 基于模式和机器学习相结合的方法

基于模式和机器学习技术相结合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事件对,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^[14]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^[15]从大规模网络文本语料库中提取因果关系的术语,然后基于点态相互信息的统计度量使用因果线索来度量网络文本语料之间的因果强度。

1.2.3 基于深度学习的方法

由于深度神经网络具有强大的表征学习能力,可以强有

力地捕捉隐含的和模棱两可的因果关系,因此近年来采用深度学习技术提取因果关系已成为一种流行选择。De Silva等^[16]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^[17]选择使用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^[18]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^[19]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。

1.3 常用数据集

基于深度学习的因果关系抽取所使用的数据集有ACE 2005数据集^[20]、SemEval 2010 task 8数据集^[21]等。

ACE由美国国家标准和技术学会(National Institute of Standards and Technology, NIST)、中央情报局(Central Intelligence Agency, CIA)和国家安全局(National Security Agency, NSA)主管,该会议的主要关注点并非是会议本身而是信息与关系抽取的评测。随着会议影响力的不断提升,语料数据得到了丰富的扩展。该会议提供已经标注好的多种类型实体、关系和事件的语料数据集,主要用于事件和关系的抽取任务中。SemEval 2010 task 8数据集^[21]共包含8 000条训练数据和2 717条测试数据。共有如下10种关系:实体-起源(Entity-Origin)、实体-目的地(Entity-Destination)、原因-结果(Cause-Effect)、成员-集合(Member-Collection)、信息-主题(Message-Topic)、产品-生产者(Product-Producer)、部分-整体(Component-Whole)、仪器-代理商(Instrument-Agency)、内容-容器(Content-Container)和其他(Other),其中训练集和测试集中包含的因果关系文本分别占比12.54%和12.07%。

2 基于深度学习的因果关系抽取模型

因果关系抽取是自然语言处理(NLP)的重要任务,当前主要有两类框架来完成此任务:一种是流水线方法,一种是联合抽取方法。流水线方法中,因果关系抽取任务被划分为两个步骤:确定句子中的事件触发词及相关论元,根据触发词构造候选事件对并进行因果关系判别。这两个步骤分别定义为两个相关的任务:事件检测(Event Detection, ED)和关系分类(Relation Classification, RC)。如果只需要对已标注事件对进行分类,则因果关系抽取可以简化为因果关系分类,联合抽取模型则是将两个子任务统一构建成一个模型,在建模时进一步利用两个任务间的潜在关联,缓解错误传播问题。

2.1 基于流水线的因果关系抽取

事件检测任务是自然语言处理中信息获取的一个常见任务,从非结构化的信息中抽取事件触发词和相关角色的触发词论元;关系分类任务则是依据标注好的语料进行事件对间的相关类型判别。下面是对当前主流的事件检测任务和关系分类任务模型的详细介绍。

2.1.1 基于神经网络的事件检测模型

事件检测的主要工作是识别语料中的事件触发词和相关论元。CNN通过捕捉句子中重要特征从而获得句子表示,传统的CNN模型在池化操作后获得的向量表示会错过有价值线索。为了解决这一问题,Chen等^[22]提出了动态多池卷积神经网络(Dynamic Multi-pooling Convolutional Neural Network, DMCNN)来提取句子级特征,如图1所示。DMCNN使用一个动态的多池层来获取句子各部分的最大值,句子表示的各

分依据事件的触发词和论元进行分割。与传统CNN模型相比,DMCNN无需借助NLP工具的帮助,能够捕获更多有价值的特征。但DMCNN模型中的语料需要预先标记好触发词和相关论元,可以将其简单地理解成对触发词和论元之间的角色进行分类。

大量研究希望通过使用循环神经网络进行事件触发词的抽取,但起初只关注单一的事件触发词提及。Ghaeini等^[23]使用双向循环神经网络(Bi-directional Recurrent Neural Network, Bi-RNN)来检测可以是单词或短语的事件,这是对多字事件

抽取的首次尝试,也是第一次尝试使用双向神经网络进行事件检测。为了在上下文中对单词语义进行编码,摆脱特定于语言的知识已有的自然语言处理工具,Feng等^[24]提出了一种混合神经网络(Hybrid Neural Network, HNN),如图2所示。该模型同时使用了LSTM和CNN来训练每个字的特征,并将其拼接作为字的表示,同时传入到Softmax层进行事件触发词的预测。实验结果最终证明了双向长短期记忆(Bidirectional Long Short-Term Memory, BiLSTM)网络在远距离捕获前后上下文时,对触发器提取非常有效。

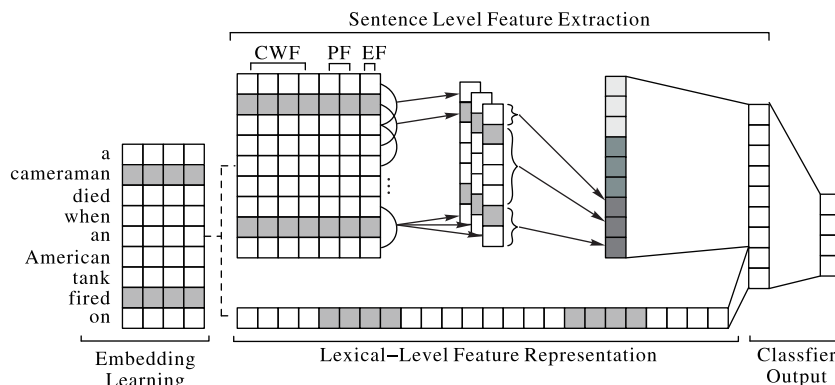


图1 基于卷积神经网络的事件抽取框架

Fig. 1 Event extraction framework based on convolutional neural network

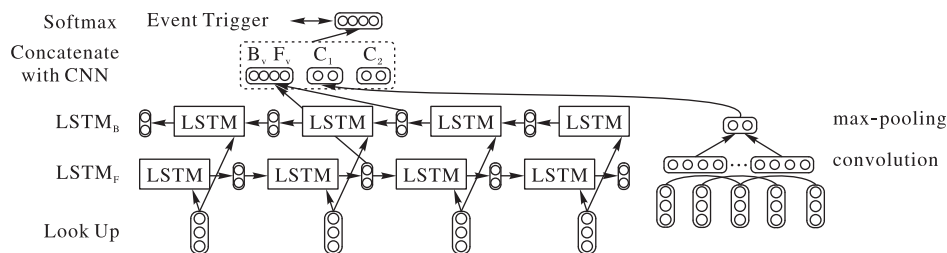


图2 HNN模型

Fig. 2 HNN model

然而研究者考虑到句子的局部上下文不足以解决特定事件类型识别中的歧义问题,Duan等^[25]提出了一种新的文档级递归神经网络(Document Level Recurrent Neural Network, DLRNN)模型,该模型可以在不设计复杂推理规则的情况下自动提取跨句线索,提高句子级事件检测的效率。Zhao等^[26]引入文档特征来丰富词的信息,加入到原本的事件检测任务中,通过设计注意力机制,从单词和句子两个层次学习文档的分布式表示。为了提高事件抽取的效率,Yang等^[27]首次指出事件抽取模型中的角色存在重叠问题,提出的一种基于预训练语言模型的事件抽取器将事件抽取分为两个阶段,包括触发器提取和论元提取,实验性能超过了大多数抽取方法。

2.1.2 基于神经网络的关系分类模型

在流水线方式的因果关系抽取中,关系分类任务使用事件检测阶段标注出的语料,对已标记事件的语料进行因果关系的判别。当前对因果关系抽取的研究相对较少,因果关系分类是特殊的关系分类,关系分类的主要任务是抽取语料中实体对间存在的关系,与因果关系抽取事件间的关系的任务比较类似。

卷积网络在文本分类或自然语言处理中的应用已有文献进行了探索^[28]。研究表明,卷积网络可以直接应用于语料的分布式表示,而不需要对文本的句法或语义结构有任何了解。

Liu等^[29]将关系抽取任务分为了实体抽取和关系分类两个部分串行处理,提出了一种结合词汇特征的卷积网络,采用同义词词典对输入词进行编码,将语义知识集成到了神经网络中,克服了词嵌入忽略词汇之间的语义关系的缺陷,但是该卷积模型的结构比较简单,且没有设置池化层,受噪声影响比较明显。Zeng等^[30]首次提出使用位置特征(Position Feature, PF)来编码到目标词的相对距离,用卷积神经网络提取词汇和句子级的特征,将所有的单词标注作为输入(无预处理操作)。通过词嵌入将单词标记转换为向量,根据给定的词提取出词汇级别的特征,同时使用卷积神经网络学习句子级特征。连接两个层次的特征输入到Softmax分类器预测两个标记实体的关系。实验结果证明了位置特征在关系抽取中的有效性。

往往输入文本序列的含义与单词的顺序是有联系的,因此使用循环神经网络来进行因果关系的抽取也是可行的。基于循环神经网络的关系抽取使用不同的循环神经网络模型获取文本的信息,进而对每个时序的隐藏状态进行组合,从而获得句子级的特征。在因果关系的抽取任务中,对于每一个时序的输入,模型通常只在序列的最后得到关系的标记特征。Zhang等^[31]首次利用循环神经网络来进行关系的抽取,提出了双向循环神经网络模型,如图3所示,模型包含3个组成部分:1)单词嵌入层,将句子中的每个单词映射成低维的词向

量;2)双向递归层,对单词序列进行建模并生成词级特征(表示);3)一种最大池化层,通过在每个维度的所有单词级特征中选择最大值,将每个时间步(每个单词)的单词级特征合并成句子级特征向量。最后利用句子级特征向量进行关系分类。双向循环神经网络模型中的每一时刻的输出向量不仅依赖于序列中当前时刻之前的输入,也依赖于后续时间的输入。文中使用更加简单的位置指示器(Position Indicator, PI)取代了位置特征。在不使用任何词汇特征的情况下,结合位置指示器的双向循环神经网络模型与使用外部信息 WordNet 的 Zeng 提出的模型具有相似的性能效果,并证明了词嵌入的维度可以很明显影响 F 值。Hou 等^[32]也使用双向循环神经网络模型进行关系抽取,将输入窗口设置到了两个实体的前后固定长度的文本序列,并根据距离远近赋予单词不同的权重。

循环神经网络模型随着时间会出现权重指数级爆炸或消失的问题,难以捕捉长期时间关联,而 LSTM 可以很好解决这个问题。Xu 等^[33]首次将 LSTM 模型用于关系抽取任务,证明了 LSTM 结合最短依存路径能够有效提取句子级别的关系。Zhang 等^[34]在此之上提出了使用双向循环神经网络模型抽取句子双向的隐藏状态输出。Zhou 等^[35]在 BiLSTM 的基础上增加了注意力机制^[36],提出了 Att-BiLSTM 模型。

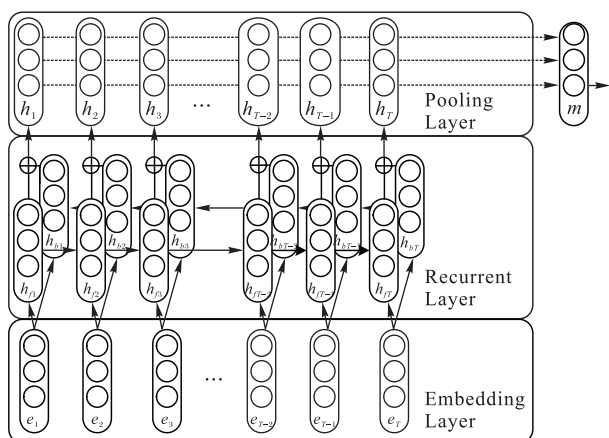


图3 基于循环神经网络的关系抽取框架

Fig. 3 Relation extraction framework based on recurrent neural network

研究证明利用深度神经网络从句子中学习句法语义特征是有用的,但是不可避免地存在信息冗余。Xiao 等^[37]使用多级循环神经网络,并将注意力机制引入其中,选择性地关注句子中的有用信息。根据标记的实体标记的位置将语料分成了 5 个部分,分别传输到 BiLSTM-Attention 层学习特征表示,之后传输进双向循环神经网络(Bi-RNN)层进行 3 个表示的语义组合,最终 Softmax 层进行分类操作。类似的,Chen 等^[38]提出了一个多通道框架,根据两个实体将句子分为多个通道,每个通道都用层次神经网络进行处理。这些通道在循环传播期间不相互作用,这使得神经网络能够学习不同的表示。为了更好地抽取语料特征,研究人员将循环神经网络、卷积神经网络、图神经网络(Graph Neural Network, GNN)等与机器学习的相关方法进行组合,建模语言模型进行因果关系抽取。Huang 等^[39]首次将 BiLSTM-CRF 模型应用于 NLP 基准序列标记数据集,其中条件随机场(Conditional Random Field, CRF)^[40]使用句子级的标记信息,带来了较高的精度。条件随机场如图 4 所示,是一种标注算法,输入一段序列后输出目标序列,在 NLP 的标注任务中,输入序列为一段文本,输出相应

的标记序列。由于选择使用双向 LSTM 组件,该模型可以同时使用过去和将来的输入特性。此外,由于 CRF 层的存在,该模型可以使用句子级的标签信息。

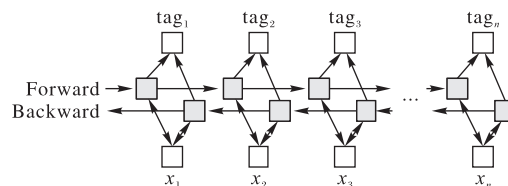


图4 BiLSTM-CRF 模型

Fig. 4 BiLSTM-CRF model

图神经网络(GNN)大多只能处理预定义图上的多跳关系推理,而无法直接应用于自然语言处理。Zhang 等^[41]通过编码依赖树将 GNN 应用于关系提取,通过使用依存树对长距离关系进行捕捉。针对关系抽取使用了一种图卷积网络,它能有效地并行存储任意依赖结构的信息,保留两个可能存在关系的实体之间的最短路径中的词作为输入树。Zhu 等^[42]提出了一种新的带生成参数的图神经网络模型,称为 GPGNNs (Graph Neural Network with Generated Parameters),该模型首先根据语料中的实体构造全连接图,然后采用编码模块、传播模块、分类模块对关系进行推理。

流水线方法首先分别抽取事件触发词和相关论元,再判断触发词间因果关系,模型灵活性高,事件抽取模型和关系分类模型可以使用独立的数据集,并不需要同时标注实体和关系的数据集。但是流水线方式中事件抽取的错误会直接影响下一步的关系分类的性能,另外对抽取的事件进行两两配对,然后再进行关系分类带来冗余信息,造成错误率提高、计算复杂度提高。此外,流水线方式独立进行两项任务,忽略了两个任务的内在的关联。

2.2 基于联合抽取方式的因果关系抽取模型

联合抽取模型简单讲就是用一个模型将两个子任务统一建模,利用两个任务之间的潜在联系来缓解错误传播的问题。现有联合抽取模型总体上有两大类:一类是以共享参数,即共享输入特征和内部隐藏状态,从而实现联合抽取,此类方法仍然使用独立的解码算法,因此两个任务之间的交互不强;另一类为了加强两个子任务的交互,采用标注策略。

2.2.1 基于参数共享的联合抽取模型

Zheng 等^[43]采用联合学习框架进行命名实体识别和关系分类,提出混合神经网络模型来提取实体和它们之间的关系,而不需要任何手工制作的特征。Zheng 等的模型如图 5 所示,包含 1 个用于实体提取的双向编码器-解码器 LSTM 模块和 1 个用于关系分类的 CNN 模块。在 BiLSTM 中获得的实体上下文信息进一步传递给 CNN 模块,以改进关系分类。与传统的流水线方法相比,该模型不仅考虑了命名实体识别(Named Entity Recognition, NER)模块和关系分类模块的相关性,而且考虑了实体标签之间的长距离关系,不需要复杂的特征工程。

训练语料中的关系往往是复杂的,不同的关系三元组在句子中会存在重叠的情况。Zeng 根据三元组的重叠情况将句子分为三类,包括正常(Normal)、实体对重叠(EntityPairOverlap)和单个实体重叠(SingleEntityOverlap)。此前的方法主要集中在 Normal 这一类别上,无法精确地提取关系三元组。Zeng 等^[44]提出了一种基于复制机制的端到端学习模型联合提取关系事实。在解码过程中采用了两种不同的策

略:使用一个联合解码器或应用多个分离解码器。该结构使用双向RNN对语料进行编码,然后使用解码器预测关系并从源句中复制实体从而实现三元组的抽取。解码的工作流程分为3步:1)预测关系;2)复制一个实体;3)复制第二个实体。

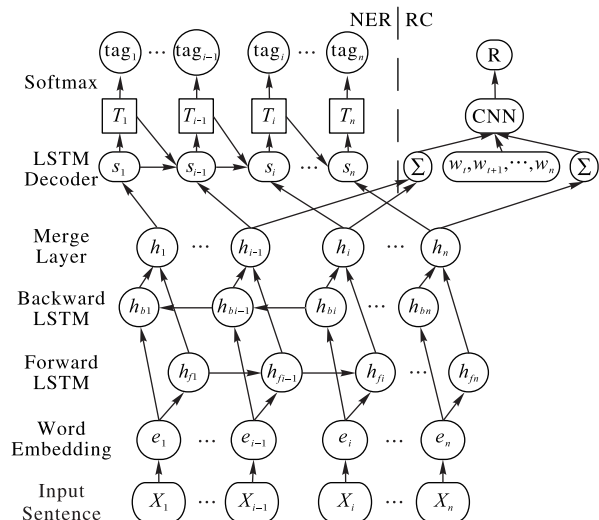


图5 基于参数共享的联合抽取模型

Fig. 5 Joint extraction model based on parameter sharing

解码器重复上述过程,可以生成多个三元组,直到所有有效的三元组生成,解码器将生成NA三元组,这意味着“停止”,从而解决了重叠实体和重叠关系的问题。复制机制结合神经网络成功解决了关系抽取中的实体和关系重叠问题,但是模型仍然存在缺陷,实体抽取只针对单个单词,若是实体是由多个单词组成的,则无法进行关系抽取。

为了解决多词实体的抽取问题,以及针对关系抽取的重叠问题,Pang等^[45]提出了一种基于序列到序列学习的深层神经网络模型,即混合双指针(Hybrid Dual Pointer,HDP)网络。如图6所示,HDP通过生成混合双指针序列从给定的句子中提取多对三元组,能够在重叠问题下联合提取多对三元组,打破了多词实体无法被提取的限制。将多对关系三元组的联合抽取问题转化为序列生成问题,将实体和关系串行关联,实现了对关系三元组的联合解码。HDP模型使用序列到序列的学习机制,包括编码语义表示向量的编码器和生成多对三元组的解码器。特别地,对于每个三元组,采用混合双指针网络来联合提取一对实体及其关系。在混合双指针网络中,有一个双指针预测器和一个关系预测器。双指针预测器用于提取实体提及,包括单词实体和多词实体。关系预测器用于预测两个实体之间的语义关系,其中关系来自预定义的关系集。受实体提及是从句子的某一特定位置开始并具有连续跨度的启发,模型利用双指针提取实体来预测实体的位置,减少了信息冗余。模型分别使用BiLSTM和Transformer两种编码器进行对比实验,最终的结果证实Transformer效果更佳。

实体信息是学习实体对之间关系的重要信息,当实体包含多个单词时,单个词不能覆盖实体的全部语义信息,可能导致关系类型提取错误。此外端到端模型大多是基于一个带注意的递归神经网络(RNN)。在关系抽取中,RNN必须在多个时间步长上累积才能连接两个实体。当两个实体之间的距离太远时,RNN很难学习它们之间的相关性,这可能会影响关系提取。Bai等^[46]提出的具有双指针模块的端到端模型联合

提取整个实体和关系。双指针模块与多个解码器组合以预测输入语句中实体的起始和结束位置。在嵌入过程中,通过预训练获得输入句子中每个单词的嵌入。为了充分利用输入句子中的位置信息,嵌入语句时加入了“位置编码”^[47]。为了有效地学习远距离实体之间的相关性,编码器选择采用多层卷积和自注意力机制的组合,而不是双向RNN。卷积捕获文本的局部结构,而自注意力学习每对单词之间的全局交互。在解码器方面,采用了Zeng等^[44]提出的多译码器策略。每个解码器都能产生一个三元组。在解码过程中,引入了双指针模块,通过预测输入语句的起始位置和结束位置,从输入语句中复制出一个完整的实体。

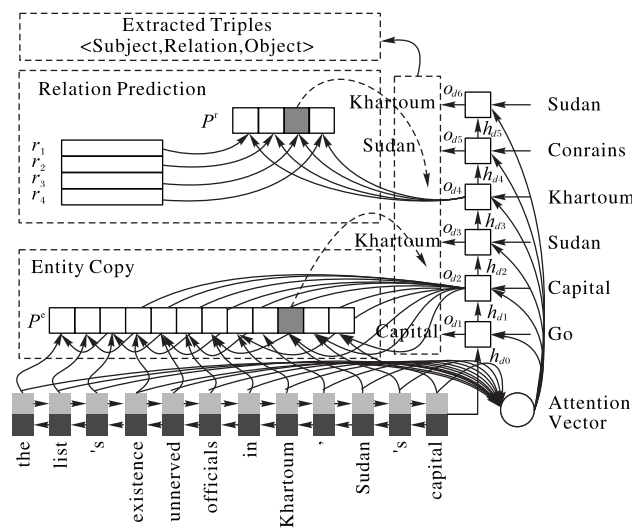


图6 HDP模型

Fig. 6 HDP model

2.2.2 基于标注方案的联合抽取模型

最初的参数共享的联合抽取方法还是存在两个子任务,只是这两个子任务之间通过参数共享有了交互,依然会产生没有关系的实体对这种冗余信息。为了减少冗余信息,提高抽取效率,减少运算量,Zheng等^[48]提出了一种新的标签方案,可以将联合抽取任务转化为标签问题,如图7所示。Zheng方案研究了基于LSTM的不同的端到端模型来直接提取实体及其关系,以联合提取实体和关系,而不需要分别识别实体和关系。此外还修了解码方法,增加了一个有偏的损失,使得该模型更适合针对关系抽取所提出的标签。基于神经网络的端到端的模型被广泛应用于序列标注任务中,Zheng针对关系抽取研究了一个端到端的模型,包含了双向LSTM层、基于LSTM的编码层和基于LSTM的解码层。

Zheng等第一次将关系抽取任务转变成标记任务,但是该方案无法解决关系的重叠。Dai等^[49]提出了一种新颖的统一联合抽取模型,如图8所示。该模型根据查询词位置 P 标记实体和关系标签,即设定一个位置 P ,然后在 P 处检测实体,并识别与 P 处实体有关系的其他位置的实体。为了实现这种模式,文中还提出了一种标记方案,为 n 个单词的句子生成 n 个标记序列。然后引入位置注意机制为每个查询位置生成不同的句子表示,以对这 n 个标签序列进行建模。该模型可以同时提取实体及其类型以及所有重叠关系。对于查询词 P ,建立一个长度为 n 的标签序列来表示与 P 处的实体相对应的

所有可能的重叠关系。如果 P 在实体的开始处,则在 P 处标记该实体类型;对于其他的位置,如果它们与 P 处的实体具有关系,则在其余的单词上标记关系类型,此标记方案可以对所

有实体和重叠关系进行注释。模型使用词嵌入和基于字符的嵌入的组合,将其输入Bi-LSTM网络中,结合位置注意力后传到CRF层进行标注任务。

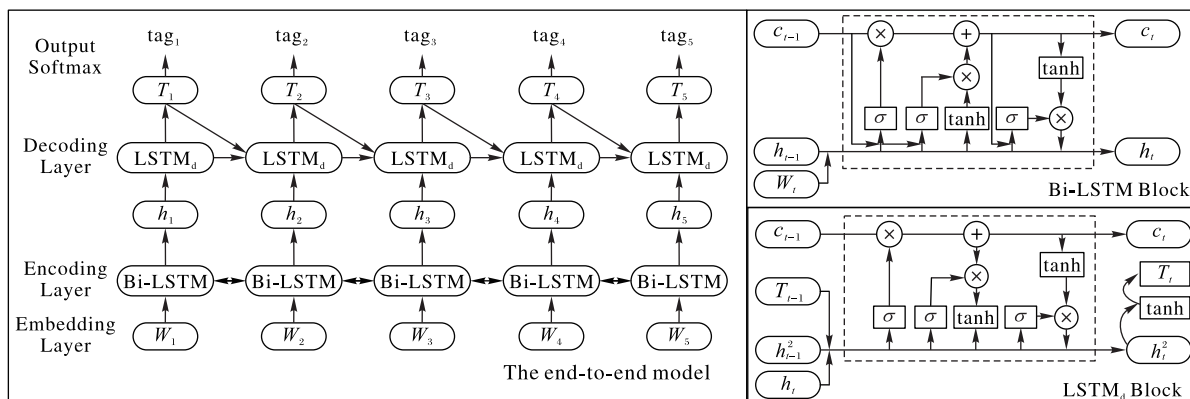


图7 具有偏置目标函数的端到端模型

Fig. 7 End-to-end model with biased objective function

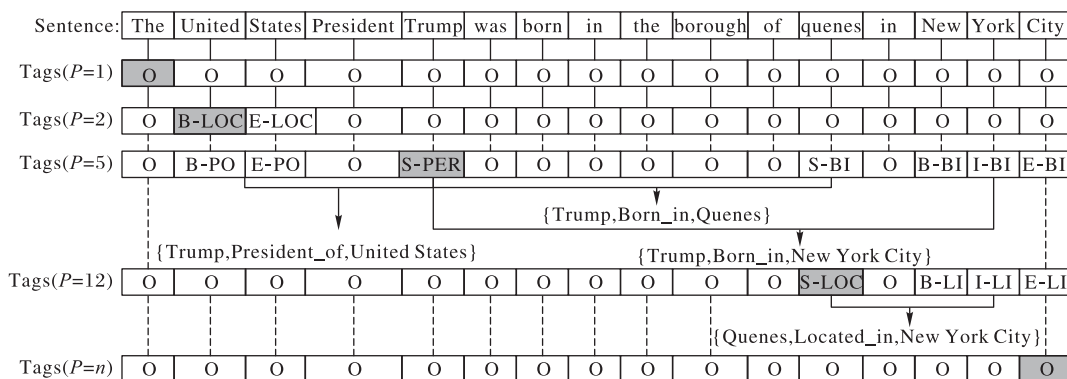


图8 P标签方案

Fig. 8 P tagging scheme

现有语料库数据不足,很难构造出一个没有先验知识的高级深度学习模型。因果关系之间的实体距离有时也相距甚远,因果三元组中的长程依赖关系给深度学习模型带来了困难和模糊。为了准确地提取出因果三元组,Li等^[50]提出了一种以BiLSTM-CRF模型为骨干的因果关系抽取器 SCIFI (Self-Attentive BiLSTM-CRF with FlaIr Embeddings),如图9所示。模型使用一种新的因果关系标注方案直接抽取因果关系,使用“BIO”标注和“C,E,Emb”标签来表示词的位置信息和因果事件中的语义角色,而不需要提取候选因果对并分别识别它们之间的关系。为了解决数据不足的问题,将上下文字符串嵌入应用到了到因果关系抽取任务中。为了捕获特定任务的字符特征,采用单层CNN结构与最大池化操作来学习字符级表示。通过双向的LSTM进行特征抽取。之后经过多头自注意力层(Multi-Head Self-Attention, MHSA)学习给定句子中因果关系的依存关系,最终传入到CRF层进行实体的标注进行因果关系的实体对的抽取。

基于联合抽取的信息抽取模型将抽取任务简化成了分类和标注两种模型。分类任务利用底层知识的共享,将抽取的特征同时用来识别事件和候选事件对之间的因果关系,而基于标注的方案将在标记的设计中同时将事件类型和相关关系进行了联合标注,减少了模型中的冗余信息,对实验结果的提升起到了很大的帮助。

2.3 小结

在基于深度学习的关系抽取任务的早期,序列结构作为输入是卷积神经网络和循环神经网络是主要的做法。近年来,随着研究的深入,图神经网络、指针网络也逐渐被运用到了关系抽取的任务中。在NLP任务中,卷积神经网络考虑了所有可能的短语组合,由于训练过程中参数较少,效果优异,成为在关系抽取领域模型构建中获取句子特征表示时的常用模型。循环神经网络则是根据当前词的上下文来获得有效特征。通过两种模型的组合在一定程度上提升抽取效果,注意力机制的加入也变成了关系抽取领域的常见做法。为了得到更好的句子的向量表示,神经网络模型的构建仍然要考虑反映自然语言处理方面的先验知识,如 n -gram特征、依存关系特征、单词的上下文特征等各种特征。

基于流水线的关系抽取方法,将关系抽取分为了实体的抽取任务和关系分类任务。相较于流水线方式,联合方法抽取使用一个模型,将两个子模型统一建模。流水线方式筹建两个模型,灵活性高,且对于实体抽取与关系分类模型,可以分别使用独立的数据集。流水线方式存在3个缺陷:1)误差积累,实体抽取的误差会影响到关系分类的准确度;2)信息冗余,在实体抽取过后,需要对实体两两配对并进行关系分类,对于不存在关系的候选实体会造成大量冗余信息,增加了计算复杂度,造成计算资源的浪费;3)交互缺失,流水线方式的关系抽取忽略了两个子任务之间的内在关联性。联合关系抽取增强了实体模型和关系模型之间的交互,主要分为两大类:

1)共享参数的联合抽取;2)基于标注策略的联合抽取。传统的关系抽取任务使用召回率(Recall, R)、精确度(Precision, P)和F值来对模型的好坏进行评判。与传统关系抽取模型评判标准一致,因果关系抽取任务同样选择这三个指标对模型

进行评判,但往往高的召回率和高的精确度是互相矛盾的,因此F值被用来作为评判标准更加可靠。针对本文介绍的基于流水线方式和联合抽取方式构造的模型,表1与表2分别展示了模型的性能以及各实验使用的数据集。

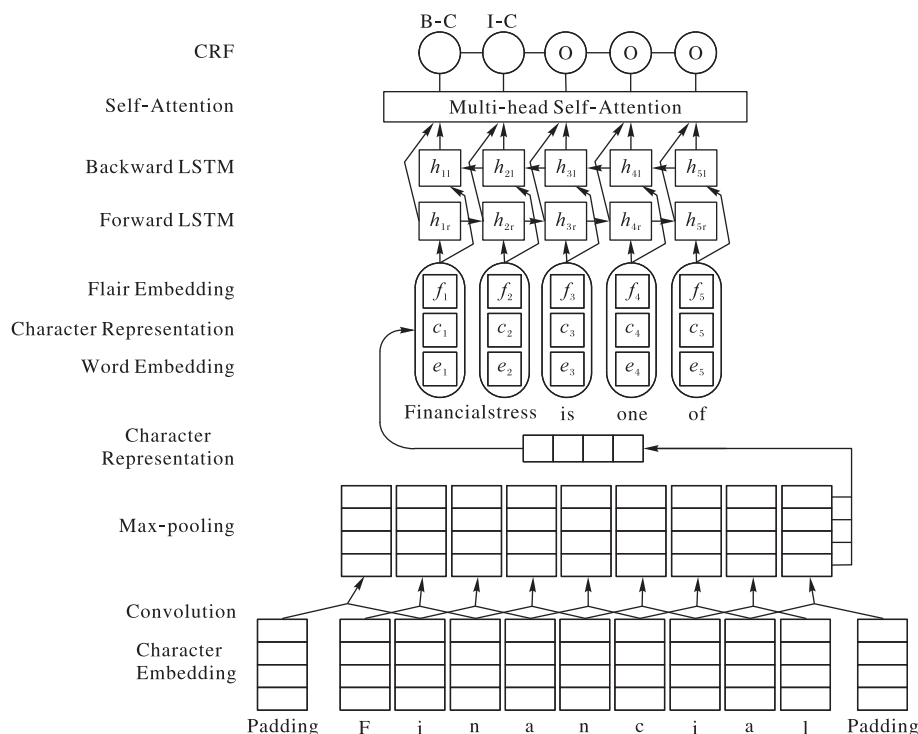


图9 SCIFI因果关系抽取器

Fig. 9 SCIFI causality extractor

表1 基于流水线方式的模型关系抽取性能

Tab. 1 Performance of relation extraction models based on pipeline

模型	F值/%	语料
CNN(2-3-4-5)	61.2	ACE2005
BiRNN	79.6	SemEval-2010 Task8
DMCNN	69.1	ACE2005
Att-BiLSTM	84.0	SemEval-2010 Task8
BiLSTM-BiRNN	84.3	SemEval-2010 Task8
Multi-channel	84.6	SemEval-2010 Task8

表2 基于联合抽取方式的模型关系抽取性能

Tab. 2 Performance of relation extraction models based on joint extraction

模型	F值/%	语料
BiLSTM-CNN	52.1	ACE2005
MultiDecoder	58.7	NYT
HDP	73.8	NYT
LSTM-LSTM-Bias	50.1	NYT
SCIFI	84.5	SemEval-2010 Task8

3 未来研究方向

近年来,无论是基于流水线的方式,还是基于联合抽取的方式,基于深度学习的因果关系抽取技术有很好的发展,在关系抽取的任务中获得了较好的性能,然而在实际应用中还有很多不足。

针对流水线方法中的关系分类任务,首要的问题是降低计算复杂度,解决候选实体对中的冗余信息问题。对于联合抽

取可以利用两个实体抽取与关系分类两个任务之间的潜在关系,但是在联合抽取任务中仍然需要加强实体模型与关系模型之间的交互能力,构建实体模型和关系模型的输出之间的约束,提高联合模型的性能。引入图神经网络是一个很好的选择,已经有不少的研究^[51-52]对此做了尝试。跨语言问题、领域自适应、不平衡数据集下的因果关系抽取等都是当前的研究方向。此外,跨句子、跨段落等更粗粒度的关系抽取也是未来的研究方向。

在使用标记方案进行因果关系的抽取中,由于标记只能对同一个实体进行一次标注,这使得一句话中出现多处因果关系,且某一关系中的原因事件是另一关系的结果事件这样的例子无法被准确识别,已有研究在设计模型的标记方案时对于同一句话中存在两个因果关系进行了解决,但是仍然无法解决超出两个的关系的例句,这是未来需要突破的一大要点。参考现有的事件抽取模型^[53]中的方法,对于句子进行层级标注是一个很好的方法。

参考文献(References)

- [1] 郭笑丹. 基于贝叶斯网络的民航突发事件因果关系分析方法研究[D]. 天津:中国民航大学, 2018:1-57. (GUO X D. Research on causal reasoning method of domain ontology based on Bayesian network [D]. Tianjin: Civil Aviation University of China, 2018: 1-57.)
- [2] 刘加新. 生物医学领域的实体因果关系抽取[D]. 苏州:苏州大学, 2018:1-86. (LIU J X. Causal relation extraction between biomedical entities [D]. Suzhou: Soochow University, 2018:1-86.)
- [3] PEARL J, MACKENZIE D. The book of why: the new science of

- cause and effect[J]. *Science*, 2018, 361(6405): 855-855.
- [4] 刘挺. 从知识图谱到事理图谱[R]. 哈尔滨: 哈尔滨工业大学社会计算与信息检索研究中心, 2017. (LIU T. From knowledge graph to event evolutionary graph[R]. Harbin: Research Center for Social Computing and Information Retrieval of Harbin Institute of Technology, 2017.)
 - [5] CHANG D S, CHOI K S. Incremental cue phrase learning and bootstrapping method for causality extraction using cue phrase and word pair probabilities [J]. *Information Processing and Management*, 2006, 42(3): 662-678.
 - [6] 付剑锋, 刘宗田, 刘炜, 等. 基于层叠条件随机场的事件因果关系抽取[J]. *模式识别与人工智能*, 2011, 24(4): 567-573. (FU J F, LIU Z T, LIU W, et al. Event causality extraction based on cascaded conditional random fields [J]. *Pattern Recognition and Artificial Intelligence*, 2011, 24(4): 567-573.)
 - [7] 冯冲, 康丽琪, 石戈, 等. 融合对抗学习的因果关系抽取[J]. *自动化学报*, 2018, 44(5): 811-818. (FENG C, KANG L Q, SHI G, et al. Causality extraction with GAN [J]. *Acta Automatica Sinica*, 2018, 44(5): 811-818.)
 - [8] 许晶航, 左万利, 梁世宁, 等. 基于图注意力网络的因果关系抽取[J]. *计算机研究与发展*, 2020, 57(1): 159-174. (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.)
 - [9] 庄传志, 靳小龙, 朱伟建, 等. 基于深度学习的关系抽取研究综述[J]. *中文信息学报*, 2019, 33(12): 1-18. (ZHANG C Z, JIN X L, ZHU J W, et al. Deep learning based relation extraction: a survey [J]. *Journal of Chinese Information Processing*, 2019, 33(12): 1-18.)
 - [10] 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.
 - [11] KHOO C S G, CHAN S, NIU Y. Extracting causal knowledge from a medical database using graphical patterns[C]// *Proceedings of 38th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics, 2000: 336-343.
 - [12] GIRJU R, MOLDOVAN D. Text mining for causal relations[C]// *Proceedings of the 15th International Florida Artificial Intelligence Research Society Conference*. Palo Alto, CA: AAAI Press, 2002: 360-364.
 - [13] ITTOO A, BOUMA G. Extracting explicit and implicit causal relations from sparse, domain-specific texts[C]// *Proceedings of 16th International Conference on Applications of Natural Language to Information Systems, LNCS 6716*. Berlin: Springer, 2011: 52-63.
 - [14] GIRJU R. Automatic detection of causal relations for question answering [C]// *Proceedings of the ACL 2003 Workshop on Multilingual Summarization and Question Answering*. Stroudsburg, PA: Association for Computational Linguistics, 2003: 76-83.
 - [15] 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, CA: AAAI Press, 2016: 421-430.
 - [16] 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.
 - [17] KRUEGKRAI 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 31st AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press, 2017: 3466-3473.
 - [18] 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.
 - [19] 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: Association for Computational Linguistics, 2018: 306-316.
 - [20] AGUILAR J, BELLER C, MCNAMEE P, et al. A comparison of the events and relations across ACE, ERE, TAC-KBP, and FrameNet annotation standards [C]// *Proceedings of the 2nd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*. Stroudsburg, PA: Association for Computational Linguistics, 2014: 45-53.
 - [21] 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 2009 Workshop on Semantic Evaluations: Recent Achievements and Future Directions*. Stroudsburg, PA: Association for Computational Linguistics, 2009: 94-99.
 - [22] CHEN Y, XU L, LIU K, et al. Event extraction via dynamic multi-pooling convolutional neural networks [C]// *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics/ the 7th International Joint Conference on Natural Language Processing*. Stroudsburg, PA: Association for Computational Linguistics, 2015: 167-176.
 - [23] GHAEINI R, FERN X, HUANG L, et al. Event nugget detection with forward-backward recurrent neural networks[C]// *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics, 2016: 369-373.
 - [24] FENG X, QIN B, LIU T. A language-independent neural network for event detection [J]. *SCIENCE CHINA Information Sciences*, 2018, 61(9): No. 092106.
 - [25] DUAN S, HE R, ZHAO W. Exploiting document level information to improve event detection via recurrent neural networks [C/OL]// *Proceedings of the 8th International Joint Conference on Natural Language Processing*. [2020-06-12]. <https://www.aclweb.org/anthology/I17-1036.pdf>.
 - [26] ZHAO Y, JIN X, WANG Y, et al. Document embedding enhanced event detection with hierarchical and supervised attention [C]// *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics, 2018: 414-419.
 - [27] YANG S, FENG D, QIAO L, et al. Exploring pre-trained language models for event extraction and generation [C]// *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics, 2019: 5284-5294.
 - [28] ZHANG X, ZHAO J, LECUN Y. Character-level convolutional networks for text classification [C]// *Proceedings of the 28th International Conference on Neural Information Processing Systems*. Cambridge: MIT Press, 2015: 649-657.

- [29] LIU C, SUN W, CHAO W, et al. Convolution neural network for relation extraction [C]// Proceedings of the 9th International Conference on Advanced Data Mining and Applications, LNCS 8347. Berlin: Springer, 2013: 231-242.
- [30] ZENG D, LIU K, LAI S, et al. Relation classification via convolutional deep neural network [C]// Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers. Stroudsburg, PA: Association for Computational Linguistics, 2014: 2335-2344.
- [31] ZHANG D, WANG D. Relation classification via recurrent neural network [EB/OL]. [2020-08-05]. <https://arxiv.org/pdf/1508.01006.pdf>.
- [32] HOU X, ZHAO Z, LIU T, et al. Classifying relation via bidirectional recurrent neural network based on local information [C]// Proceedings of the 18th Asia-Pacific Web Conference, LNCS 9931. Cham: Springer, 2016: 420-430.
- [33] XU Y, MOU L, LI G, et al. Classifying relations via long short term memory networks along shortest dependency paths [C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2015: 1785-1794.
- [34] ZHANG S, ZHENG D, HU X, et al. Bidirectional long short-term memory networks for relation classification [C/OL]// Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation. [2020-08-05]. <https://www.aclweb.org/anthology/Y15-1009.pdf>.
- [35] 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 Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2016: 207-212.
- [36] MNIH V, HEES N, GRAVES A, et al. Recurrent models of visual attention [C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2204-2212.
- [37] XIAO M, LIU C. Semantic relation classification via hierarchical recurrent neural network with attention [C]// Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers. Stroudsburg, PA: Association for Computational Linguistics, 2016: 1254-1263.
- [38] CHEN Y, WANG K, YANG W, et al. A multi-channel deep neural network for relation extraction [J]. IEEE Access, 2020, 8: 13195-13203.
- [39] HUANG Z, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging [EB/OL]. [2020-08-09]. <https://arxiv.org/pdf/1508.01991v1.pdf>.
- [40] LAFFERTY J, MCCALLUM A, PEREIRA F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data [C]// Proceedings of the 18th International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc., 2001: 282-289.
- [41] ZHANG Y, 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, PA: Association for Computational Linguistics, 2018: 2205-2215.
- [42] ZHU H, LIN Y, LIU Z, et al. Graph neural networks with generated parameters for relation extraction [C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2019: 1331-1339.
- [43] 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.
- [44] 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. Stroudsburg, PA: Association for Computational Linguistics, 2018: 506-514.
- [45] PANG Y, LIU J, LIU L, et al. A deep neural network model for joint entity and relation extraction [J]. IEEE Access, 2019, 7: 179143-179150.
- [46] BAI C, PAN L, LUO S, et al. Joint extraction of entities and relations by a novel end-to-end model with a double-pointer module [J]. Neurocomputing, 2020, 377: 325-333.
- [47] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2017: 6000-6010.
- [48] 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. Stroudsburg, PA: Association for Computational Linguistics, 2017: 1227-1236.
- [49] DAI D, XIAO X, LYU Y, et al. Joint extraction of entities and overlapping relations using position-attentive sequence labeling [C]// Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019: 6300-6308.
- [50] 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.
- [51] SUN C, GONG Y, WU Y, et al. Joint type inference on entities and relations via graph convolutional networks [C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2019: 1361-1370.
- [52] FU T J, LI P H, MA W Y. GraphRel: modeling text as relational graphs for joint entity and relation extraction [C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2019: 1409-1418.
- [53] WEI Z, SU J, WANG Y, et al. A novel cascade binary tagging framework for relational triple extraction [C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics, 2020: 1476-1488.

This work is partially supported by the National Natural Science Foundation of China (61872313), the Key Project of National Key Research and Development Program of China (2017YFB1002300, 2018YFC1700302).

WANG Zhujun, born in 1996, M. S. candidate. His research interests include natural language processing.

WANG Shi, born in 1981, Ph. D., associate research fellow. His research interests include semantic analysis, knowledge mapping.

LI Xueqing, born in 1995, Ph. D. candidate. Her research interests include natural language processing.

ZHU Junwu, born in 1972, Ph. D., professor. His research interests include knowledge engineering, ontology.