

神经事件抽取综述

樊润泽

2021E8013282043

中国科学院计算技术研究所

摘要

随着网络文本数据量爆炸式的增长，信息抽取任务越来越受到研究者的重视，逐渐成为了自然语言处理领域的研究热点，而事件抽取是信息抽取中一个很重要的部分，有非常多的应用需求和场景。深度学习的出现带动了人工智能领域的又一个浪潮，并与事件抽取任务相结合促进其取得了很大的进步。本文将从 2015 年第一个使用深度学习解决事件抽取任务的工作开始总结目前为止所出现的较有影响力的工作，其中包括基于卷积神经网络 (CNN)、循环神经网络 (RNN)、注意力机制、图神经网络 (GNN)、预训练-微调范式、生成式范式与提示学习 (Prompt) 的事件抽取模型。随着深度学习的发展，出现了不同的事件抽取范式，包括分类任务、序列标注任务、问答任务、生成式任务等。最后，本文将简单讨论一下事件抽取现在面临的挑战以及未来发展趋势。

1 引言

随着互联网的蓬勃发展，网络数据量爆炸式增长，尤其是文本数据，例如新闻、文献、博客等。在大量数据下，用户很难快速检索到符合要求的数据和信息，若有结构性关键信息给予辅助，信息检索对于用户来说会方便很多。过去，往往是由人工的方式标注关键信息，但随着数据量的增大，人工标注的方式将大大增加成本，耗费人力资源，因此信息抽取 (Information Extraction, IE) 的研究将变得尤为重要。近些年里，因为信息抽取技术可以帮助人们自动从海量文本数据中抽取出潜在信息 [1]，所以逐渐成为了研究热点。事件抽取任务 (Event Extraction, EE) 是信息抽取研究中的一类，旨从非结构化的自然语言文本中自动抽取出用户感兴趣的事件信息并以结构化的形式表示 [2]。自动内容抽取 (Automatic Content Extraction, ACE) 国际评测会议将事件定义为：发生在某个特定时间点或时间段，某个特定地域范围内，由一个或多个角色参与的一个或者多个动作组成的事情或者状态的改变 [1]。现在事件抽取已经成为了一个非常重要也非常有挑战性的任务。

事件抽取任务有广泛的应用，在自然语言处理下游任务与其他学科领域都起着至关重要的作用。前者包括信息检索、推荐系统 [3]、智能问答 [4]、知识图谱构建 [5] 等；后者包括

安全领域 [6]、智能交通领域 [7]、金融领域 [8]、社交媒体领域 [9]、生物医学领域 [10]、法律领域 [11] 等。

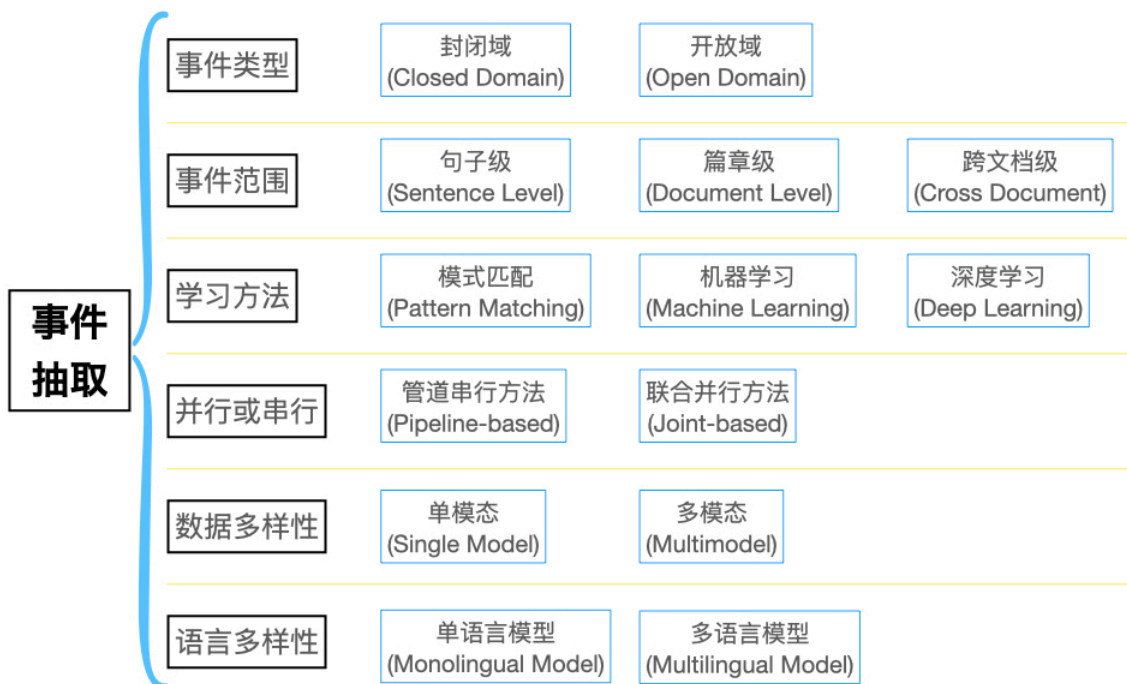


图 1: 事件抽取任务分类

根据不同的准则，事件抽取可以被分为多个不同的种类，如图1所示。根据事件类型和结构是否提前给定可以分为封闭域 (Closed Domain) 和开放域 (Open Domain) 的事件抽取：封闭域是指按照提前设计好的事件类型模版进行学习和预测；而开放域是指不提前给定事件类型，旨在检测出新的事件类型。根据一个事件在文档中的跨越范围可以分为句子级 (Sentence Level)、篇章级 (Document Level)、跨文档级 (Cross Document)：其中句子级是指事件触发词及其论元都在同一个句子中出现；篇章级指的是同一个事件的论元可能出现在了多个句子中；跨文档级指的是同一个事件的论元可能出现在多个文档中。根据学习方法可以分为模式匹配 (Pattern Matching)、统计机器学习 (Machine Learning)、深度学习 (Deep Learning)：其中模式匹配方法是提前手动设计好基于不同领域的事件模版以及大量的抽取规则，通常基于句法树或正则表达式进行抽取；基于机器学习的是指提前手动设计大量特征模版，然后在数据集上根据这些模版提取出大量特征，再选择统计机器学习方法进行训练和预测，比如支持向量机、最大熵等模型；基于深度学习的是近些年的主流方法，其优点是可以使用低维稠密向量表示单词、短语、句子，使用神经网络自动学习出隐式特

征进行预测,包括卷积神经网络 (CNN)、循环神经网络 (RNN)、注意力机制、预训练语言模型等。根据事件触发词与事件论元预测是否串行或并行可以分为管道串行方法 (Pipeline-based) 和联合并行方法 (Joint-based): 管道串行方法是指先预测事件触发词及其类型,再将触发词的类型信息用于事件论元的预测中,其优点是论元预测时可以利用先验知识,但同时也是个缺点,若触发词预测错误会导致错误传递;而联合并行方法是指同时预测事件触发词与事件论元,没有先验知识也不会导致错误传递。根据数据的多样性可以分为单模态模型 (Single Model) 与多模态模型 (Multimodel): 其中单模态模型是指只使用文本数据;而多模态模型是指在文本数据的基础上采用图像数据、声音数据等进行辅助。根据语言的多样性可以分为单语言模型 (Monolingual Model) 和多语言模型 (Multilingual Model): 其中单语言模型是指训练集和测试集都是同一种语言;多语言模型是指训练集是语料库丰富的语言,如英语和汉语,而测试集是缺乏数据的小语种。

传统机器学习的发展遇到最大的挑战是特征工程的问题,而深度学习的出现很好的解决了这个问题,原因有两点:第一是词嵌入技术 (Word Embedding) 可以用低维稠密的词向量表示每一个单词;第二是深度的神经网络模型可以很好的捕捉文本中的非线性特征。基于深度学习的事件抽取模型根据所使用神经网络的不同可以分为:卷积神经网络 (CNN)、循环神经网络 (RNN)、图神经网络 (GNN)、注意力机制 (Attention)、预训练语言模型 (Pre-Training Model)、提示学习 (Prompt)。在这之前也有很多全面的事件抽取综述 [12] [13] [14] [15],但他们不是针对传统机器学习就是针对特定领域,其中有的综述对于深度学习也只总结了 2019 年之前的神经网络模型,因此本文将主要聚焦于深度学习时代下的事件抽取任务,针对不同的神经网络模型在事件抽取任务上的应用一一进行介绍,特别是 2018 年以来的预训练语言模型和 2021 年以来的生成式范式与提示学习。

在开始介绍具体模型之前,本文第2节将首先介绍事件抽取的任务定义、常用的数据集总结以及模型评价指标。第3节到第8节分别介绍基于卷积神经网络、循环神经网络、注意力机制、图神经网络、预训练-微调范式和生成式范式与提示学习的事件抽取模型,第9节将汇总一下神经事件抽取的经典工作在 ACE2005 数据集上的性能,第10节将简单总结一下深度学习下事件抽取任务面对的主要挑战和发展方向,最后第11节将对本文做一个总结。

2 事件抽取

2.1 基本概念

事件抽取任务 (Event Extraction, EE) 是信息抽取研究中的一类, 旨从非结构化的自然语言文本中自动抽取出用户感兴趣的事件信息并以结构化的形式表示 [2]。事件是指发生在某个特定时间点或时间段, 某个特定地域范围内, 由一个或多个角色参与的一个或者多个动作组成的事情或者状态的改变 [1]。简单来说就是从一段文本中提取出事件的触发词 (Trigger) 及其论元 (Argument), 如图2所示, 其中“attack”是事件触发词, 其所触发的事件类型 (Event Type) 是“Attack”, 这个事件对应的论元: “At daybreak on the 9th”的论元角色 (Argument Role) 是“Time”, “the terrorists”的论元角色是“Attacker”, “bomb”的论元角色是“Instrument”, “Nazareth”的论元角色是“Place”。根据这个例子可以看到其中有一些专业名词, 自动内容抽取 (ACE) 国际评测会议规定了这些专业名词, 如下所示:

- 实体 (Entity): 一个或一组按照语义分类的物体, 包括人物 (People)、组织 (Organizations)、地点 (Places)、时间 (Times) 等。
- 事件提及 (Event Mention): 一个描述事件的短语或句子, 其包含事件触发词及其对应的论元。
- 事件类型 (Event Type): 表示一个事件的类别, 通常用事件触发词的类型表示事件类型。
- 事件触发词 (Event Trigger): 事件触发词是事件抽取的核心, 通常是一个动词或者名词, 触发词识别通常也是管道串行方法的核心步骤。
- 事件论元 (Event Argument): 事件论元是一个事件的参数, 通常指这个事件发生的时间、地点以及参与者等。
- 论元角色 (Argument Role): 指论元在这个事件中充当的角色, 这也反映了论元与论元、论元与触发词的关系。

2.2 任务定义

事件抽取是在研究如何从非结构化数据中自动抽取出结构化的事件信息, 按照事件的不同组成部分可以将这个任务分为四个子任务:

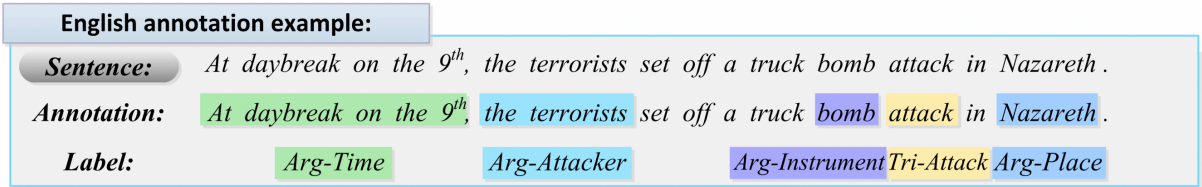


图 2: 事件抽取示例 [15]

- 触发词识别 (Trigger Identification): 从文本中找到触发词, 也就意味着有事件的出现。
- 触发词分类 (Trigger Classification): 对触发词的类别进行判断, 也就是判断事件类型, 这也是个多分类问题。
- 论元识别 (Argument Identification): 从文本中找到属于对应事件的所有事件论元。
- 论元分类 (Argument Classification): 对每一个找到的论元进行分类, 判断其在事件中充当的角色类型, 是一个多分类问题。

根据这四个子任务的执行顺序可以分为管道串行方法 (Pipeline-based) 和联合并行方法 (Joint-based)。管道串行方法是将每个子任务独立对待, 表示为一个多步分类的过程, 其包含事件触发词分类器和事件论元分类器, 先进行触发词的识别与分类, 然后对论元进行识别和分类, 这种方法的优点是触发词的分类结果可以帮助论元的识别与分类, 但同时这也是个缺点, 若触发词分类错误会导致触发词也分类错误。联合并行方法是为了克服管道串行方法中错误传播的缺点而提出的, 其构造一个联合学习模型同时去识别触发词和论元, 但它的缺点就是在论元分类中无法使用触发词的信息, 不过经典模型 JRNN [16] 的实验证明了其优于之前的管道并行方法。

2.3 数据集

在深度学习时代下, 有标注的高质量数据集是一个领域快速发展中必不可少的东西, 在事件抽取领域的发展过程中同样也产生了很多具有不同特点的数据集, 这些数据集通常是由相关领域的专业人士标注出的, 这一小节将一一介绍这些数据集。本文要介绍的数据集汇总如表1所示。

ACE2005: ACE2005 数据集是最常用的句子级事件抽取数据集, 覆盖了英语、汉语、阿拉伯语三个语种, 其标注了实体类别、关系类别以及事件类别, 包含 599 个英文文档, 633

表 1: 事件抽取常用数据集汇总

数据集	文档数	句子数	事件类型数量	语言	事件范围	相关论文
ACE2005	599	18117	33	英语/汉语/阿拉伯语	句子级	[1]
TAC KBP 2015	360	12976	38	英语	句子级	[17]
TAC KBP 2016	505	9042	18	英语/汉语/西班牙语	句子级	[18]
TAC KBP 2017	500	8951	18	英语/汉语/西班牙语	句子级	[19]
MAVEN	4480	118732	168	英语	句子级	[20]
FewFC	-	8982	10	汉语	句子级	[21]
DuEE	-	17000	65	汉语	句子级	-
Google	11909	-	30	英语	篇章级	[22]
Twitter	1000	-	20	英语	篇章级	[22]
ChFinAnn(Doc2EDAG)	32040	-	5	汉语	篇章级	[23]

个中文文档，18117 个句子。事件部分包含 8 个事件大类，33 个子类，35 种论元角色，这 33 个事件类别也是最常用的类别设置之一。

TAC KBP: TAC KBP 数据集也是句子级事件抽取常用数据集之一，其旨在开发和评估从非结构化数据中提取知识的技术，事件知识是其中的一部分。TAC KBP 2015 只有英文语料，包含 360 个文档，12976 个句子，共有 38 个事件类型。TAC KBP 2016 与 TAC KBP 2017 增加了汉语和西班牙语，其包含 500 个文档，9042 个句子，事件类型更改为了 18 种。

MAVEN: MAVEN 数据集是由清华提出的事件检测数据集，其包含 4480 个维基百科文档，118732 个事件实例以及 168 个事件类型。MAVEN 中事件类型所覆盖的领域远超其他数据集，提供了一个更接近真实场景的数据集。

FewFC: FewFC 数据集是一个面向中文金融领域的事件抽取数据集，其包含 8982 个金融相关的句子，10 种金融事件类型。

DuEE: DuEE 数据集来自于由百度发布的一个中文事件抽取竞赛，其包含 17000 个句子，65 种事件类型，事件类型来源于百度的热搜板块，数据来源于百家号新闻，比起传统新闻表达更加自由，使事件抽取任务更有挑战性。

Google: Google 数据集是篇章级事件抽取常用的数据集之一，是 GDELT 事件数据库的一个子集，包含 30 个事件类型，11909 个新闻篇章。

Twitter: Twitter 数据集是篇章级事件抽取常用的数据集之一，都是在 Twitter 上发表的篇章，包含 20 个事件类型，1000 个 Twitter 篇章。

ChFinAnn(Doc2EDAG): ChFinAnn 数据集是一个中文篇章级事件抽取数据集，其

包含 32040 个中文篇章，有 5 个事件类型。

2.4 评价指标

事件抽取任务中常用的评价指标是准确率 (Precision, P), 召回率 (Recall, R) 和 F1 值, 但在四个子任务下的使用各有一些差异, 以下将一一介绍:

定义 $I()$ 为指示函数, 括号内为 *True* 则返回 1, 否则返回 0; X 表示数据集中的原短语, \hat{X} 表示预测出来的短语; X_s 表示 X 这个短语的开始索引, X_e 表示 X 这个短语的结束索引; N_X 表示 X 这个短语的个数; X_t 表示 X 这个短语的类别。

1. 触发词识别 (Trigger Identification): 触发词预测正确表示触发词字符串、触发词开始的索引和触发词结束的索引相等。

$$P_{TI} = \frac{\sum I(\hat{T} = T \wedge \hat{T}_s = T_s \wedge \hat{T}_e = T_e)}{N_{\hat{T}}} \quad (1)$$

$$R_{TI} = \frac{\sum I(\hat{T} = T \wedge \hat{T}_s = T_s \wedge \hat{T}_e = T_e)}{N_T} \quad (2)$$

$$F1_{TI} = \frac{2 \times P_{TI} \times R_{TI}}{P_{TI} + R_{TI}} \quad (3)$$

其中 \hat{T} 表示模型预测的触发词, T 表示数据集中标注的正确触发词。

2. 触发词分类 (Trigger Classification): 触发词分类正确表示在触发词识别正确的基础上, 其事件类型也预测正确, 即触发词字符串、触发词开始的索引、触发词结束的索引和触发词类型相等。

$$P_{TC} = \frac{\sum I(\hat{T} = T \wedge \hat{T}_s = T_s \wedge \hat{T}_e = T_e \wedge \hat{T}_t = T_t)}{N_{\hat{T}}} \quad (4)$$

$$R_{TC} = \frac{\sum I(\hat{T} = T \wedge \hat{T}_s = T_s \wedge \hat{T}_e = T_e \wedge \hat{T}_t = T_t)}{N_T} \quad (5)$$

$$F1_{TC} = \frac{2 \times P_{TC} \times R_{TC}}{P_{TC} + R_{TC}} \quad (6)$$

3. 论元识别 (Argument Identification): 论元识别正确是指论元字符串、论元开始的索引、论元结束的索引和论元所属事件的类别均相等。

$$P_{AI} = \frac{\sum I(\hat{A} = A \wedge \hat{A}_s = A_s \wedge \hat{A}_e = A_e \wedge \hat{T}_t = T_t)}{N_{\hat{A}}} \quad (7)$$

$$R_{AI} = \frac{\sum I(\hat{A} = A \wedge \hat{A}_s = A_s \wedge \hat{A}_e = A_e \wedge \hat{T}_t = T_t)}{N_A} \quad (8)$$

$$F1_{AI} = \frac{2 \times P_{AI} \times R_{AI}}{P_{AI} + R_{AI}} \quad (9)$$

其中 A 表示标注数据中正确的论元短语, \hat{A} 表示模型预测的论元短语, T_t 是指标注数据中正确的此论元 A 所属事件的类别, \hat{T}_t 是模型预测的此论元 A 所属事件的类别。

4. 论元分类 (Argument Classification): 论元分类正确是指在论元识别正确的基础上, 论元在事件中充当的论元角色也分类正确。

$$P_{AC} = \frac{\sum I(\hat{A} = A \wedge \hat{A}_s = A_s \wedge \hat{A}_e = A_e \wedge \hat{T}_t = T_t \wedge \hat{A}_t = A_t)}{N_{\hat{A}}} \quad (10)$$

$$R_{AC} = \frac{\sum I(\hat{A} = A \wedge \hat{A}_s = A_s \wedge \hat{A}_e = A_e \wedge \hat{T}_t = T_t \wedge \hat{A}_t = A_t)}{N_A} \quad (11)$$

$$F1_{AC} = \frac{2 \times P_{AC} \times R_{AC}}{P_{AC} + R_{AC}} \quad (12)$$

其中 A_t 表示标注数据中正确的论元类型, \hat{A}_t 表示模型预测的论元类型。

3 基于卷积神经网络的事件抽取模型

在这之前, 事件抽取任务一直被用传统的判别式模型来解决, 而传统模型面临的问题是需要大量的特征工程技术、错误传播问题很严重以及数据稀疏性。当卷积神经网络在计算机视觉领域大放异彩 [24] 后, 人们意识到神经网络可以有效的自动学习数据的表示, 于是 Chen et al. [25] 提出了使用 CNN 来做事件抽取任务, 模型称为 DMCNN, 如图3所示, 他首先使用 CNN 和动态多池化层自动提取文本中词汇和句子信息, 然后把事件抽取任务

的触发词分类和论元分类看成单词的多分类问题，采用管道串行的方式用全连接层来预测，在 ACE2005 数据集上取得了当时 SOTA 的结果。

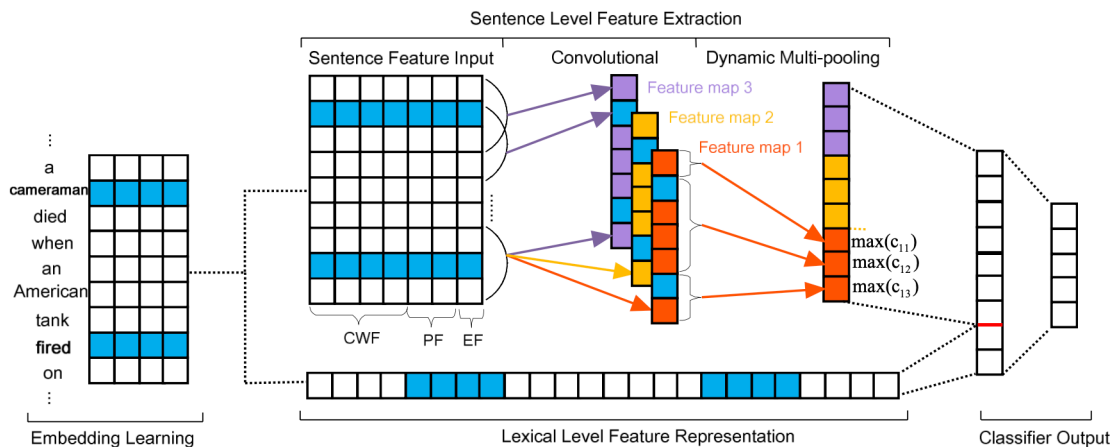


图 3: DMCNN 模型图 [25]

在同时期，Nguyen et al. [26] 也提出了采用 CNN 做事件检测的模型，但他依然使用了很多手动提取的特征，尽管如此也超过了传统方法的结果。之后，Nguyen et al. 认为之前的 CNN 模型只考虑了连续的 K-gram，而不连续的 K-gram 可能也存在事件抽取相关信息，因此他提出了一种非连续 K-gram 的 CNN 模型 [27] 做事件检测。Zhang et al. [28] 意识到管道串行方法依然存在错误传播问题，而传统的联合并行方法需要依靠大量特征工程，此外基本的 CNN 模型只能捕捉句子级别的特征，因此他提出了一种基于跳跃窗口的 CNN 模型，采用联合并行的方法解决了事件抽取任务，可以较好的捕捉全局信息以及减缓错误传播问题。

4 基于循环神经网络的事件抽取模型

比起 CNN 来说，RNN 及其类似模型 LSTM 和 GRU 更适合对序列数据进行处理，而事件抽取任务中的触发词与论元分类都可以看作是序列标注任务。Nguyen et al. [16] 提出了名为 JRNN 的模型，模型如图4所示，由两部分组成：编码和预测。编码层将词向量、实体类别编码、依存树关系编码拼接起来送入双向 RNN 中，得到了整个上下文信息。预测层使用 RNN 输出的表示上下文信息的向量采用联合的方式预测事件触发词和论元，在 ACE2005

数据集上取得了当时 SOTA 的结果。

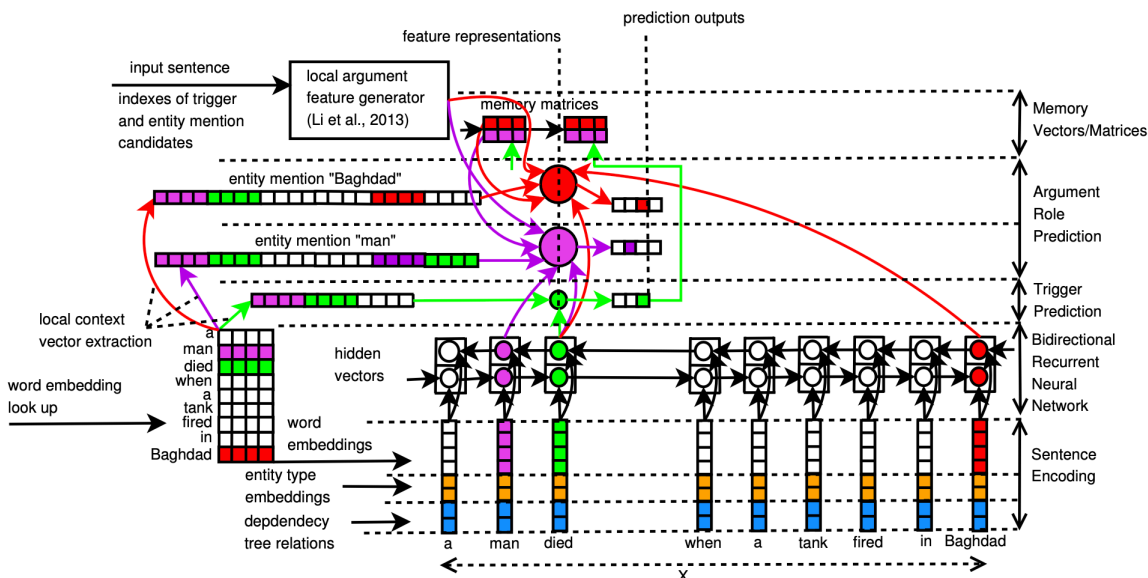


图 4: JRNN 模型图 [16]

虽然已经有很多工作在事件抽取上取得了较好的成绩，但他们都是针对一种特定语言，特别是英语，并且非常依赖之前的 NLP 处理工具，因此 Feng et al. [29] 提出了一种与语言无关的神经网络。他使用双向 LSTM 去捕捉句子信息，再使用 CNN 去得到短语信息，结合这两者去预测事件触发词。这个模型没有使用任何手动提取的特征，在 ACE2005 英语数据集上得到了 SOTA 的结果，此外在中文和西班牙语上得到了较好的结果。

之前的基于 RNN 的模型都只考虑了句子或单词级的信息，或者将提取好的依存关系添加到输入中，而没有动态考虑句法信息，因此 Sha et al. [30] 提出了一种在 RNN 基础上使用语法依赖桥进行增强的模型，同时增加了一种张量层来捕捉候选论元的信息，与 JRNN 对比有明显提升。

事件抽取任务一般都需要结合实体信息，而大部分的工作在训练过程都使用的人工标注的实体信息，在使用时需要借用其他的命名实体识别工具进行预处理，而现存的命名实体识别工具都不可能做到百分百准确，若发生错误会导致错误传播，以至于事件抽取任务也得到错误的结果。因此，Nguyen et al. [31] 提出了一种联合模型，同时预测实体、事件触发词和事件论元。他把这三个任务结合起来看成多任务学习，使用双向 GRU，并且共享隐层表示信息，在 ACE2005 数据集上取得了 SOTA 的结果。除此之外，Wei et al. [10] 将神经事件抽取应用于了临床文本数据中提取其中的药物不良事件，他将事件抽取任务看成序

列标注任务，采用双向 LSTM 和 CRF 模型，在 MIMIC-III 数据集上得到了较好的结果。

5 基于注意力机制的事件抽取模型

基于 CNN 和 RNN 的模型都只能对有限长度的句子进行建模，捕捉局部信息，并且大部分工作都是对单个事件进行处理，没有考虑事件之间的关系信息。为了解决这一问题，研究人员发现注意力机制很适合，它不仅可以对较长序列进行建模，没有长距离依赖，并且可以像 CNN 一样进行并行处理。除此之外，注意力机制可以对句子中每一部分赋予相应的权重，这可以使得模型更富有解释性，使人们可以从中发现事件之间、论元之间的关系，从而进一步分析与改进。

Liu et al. [32] 使用自注意力机制捕获全局信息，提出了 JMEE 模型，如图5所示，他使用自己提出的捷径弧去加强信息流，使用基于注意力机制的图卷积神经网络对图信息进行建模，来联合提取多个触发词与论元。此外，他还针对数据不平衡问题，对损失函数进行了优化，减少了偏差。最终，在 ACE 2005 数据集上得到了 SOTA 的结果。

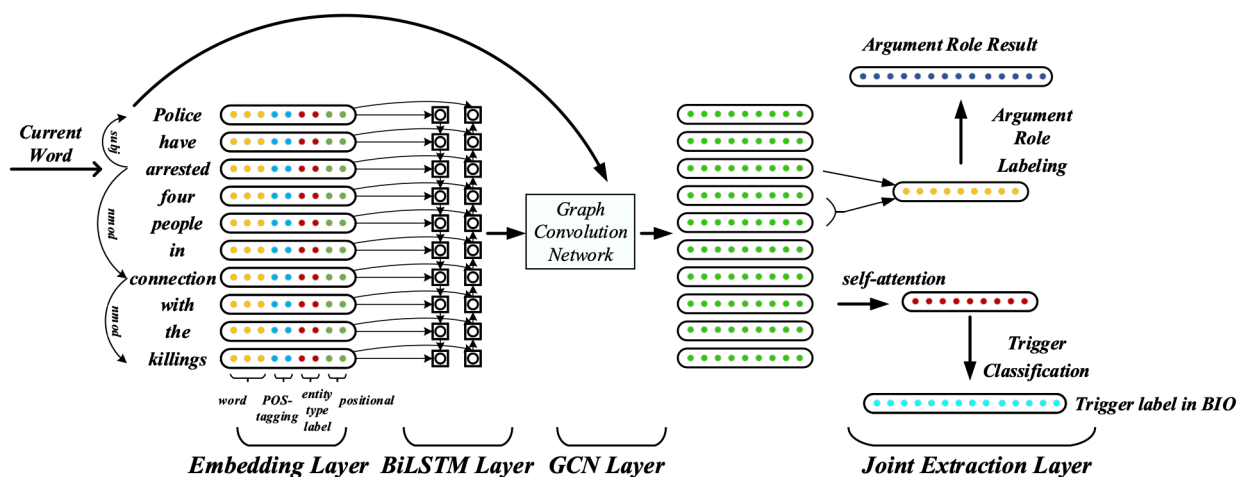


图 5: JMEE 模型图 [32]

随着互联网的发展，事件相关的任务在各个领域的需求都逐渐增加，比如金融、医疗、法律等，此外这些真实数据上有很多事件的论元是跨越多个句子的，即使句子级别的事件抽取模型做的很好也无法解决这个问题。因此，针对这个挑战，Zheng et al. [23] 提出了一种端到端的篇章级事件抽取模型 Doc2EDAG，模型如图6所示，与之前工作不同的地方是，他使用纯注意力机制 Transform [33] 模型代替以前的基于 RNN 的模型来获取篇章级信息，

同时他贡献了一个已标注的基于中文金融篇章的大规模篇章集事件抽取数据集，实验结果远高于之前的篇章级事件抽取模型，为这一块贡献了一个很好的基准系统。

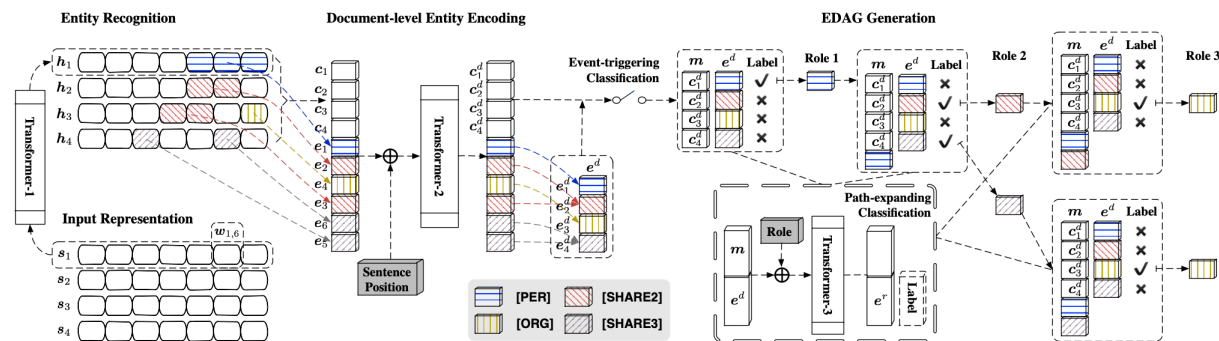


图 6: Doc2EDAG 模型图 [23]

Lu et al. [34] 发现之前的工作总是将事件抽取任务分为多步任务，每个任务之间会存在间隙，因此他从生成的角度提出了一个很新颖的方法，使用生成范式建立了直接从文本到结构性数据的端到端的模型，实验结果展示得到了和之前工作差不多的结果。

6 基于图神经网络的事件抽取模型

长距离依赖一直以来是事件抽取任务中最大的挑战之一，比如多个事件在同一个句子、一个事件的论元分布在多个句子中以及篇章级事件抽取任务，它们都需要对较长的文本进行建模才行，之前有人采用依存树 [16]、句法依赖桥 [30] 尝试来解决这个问题，虽然得到了缓解，但还是不够。Yao et al. [35] 提出了使用 GCN 做文本分类任务，取得了不错的结果，因此 Nguyen et al. [36] 提出了一种基于依存树的 GCN 模型以及一种新颖的池化方法，可以很好的将句子中的依存关系融合进上下文信息中缓解长距离依赖问题，基于 ACE 2005 数据集在事件检测任务上取得了 SOTA 的结果。

针对同一个句子中存在多个事件的问题是很难解决的，因为多个事件之间的关系很难去建模。Liu et al. [32] 在 GCN 模型上增加了一个捷径弧的设计很好的解决了这个问题，可以同时提取多个触发词和论元。

精确的估计事件发生的时间线是一个很重要的事情，但很少有人做这个工作，因为自然语言的歧义性以及事件之间信息传播的困难，导致这个问题很难解决。Wen et al. [37] 形式化了这个问题，同时提出了一种图注意力神经网络在篇章级上传播事件之间的关系信息来预测事件发生的时间线，为这个问题做出了较好的探索。

在多语言事件抽取任务上, Ahmad et al. [38] 发现只使用图卷积网络是很难解决长距离依赖问题的, 因此他提出了图注意力 transformer 编码器 (GATE) 利用自注意力机制学习词与词之间的依赖关系, 同时使用 GCN 学习到与特定语言无关的上下文表示, 然后可以迁移到其他语言进行使用。作者在 ACE 2005 数据集上测试了该模型的多语言迁移性, 发现结果远远高于其他多语言事件抽取模型, 具有很好的鲁棒性。

7 基于预训练-微调范式的事件抽取模型

自从 2018 年 ELMo [39] 的提出, 预训练语言模型越来越受到人们的关注, 特别是在 BERT [40] 被提出后在 11 个自然语言理解任务上取得了 SOTA 的结果, 预训练语言模型成为了整个自然语言处理领域的主流, 各个任务纷纷在向预训练模型靠齐, 预训练-微调范式也成为了 NLP 的主流。

借助预训练语言模型的力量, 事件抽取任务也取得了很大的进步。Yang et al. [41] 第一个将 BERT 应用于事件抽取任务, 称为 PLMEE, 模型如图7所示, 首先使用 BERT 的动态词向量预测触发词类型, 然后将触发词信息拼接到原来的句子信息中继续使用 BERT 来预测论元, 但是在这一步作者为了解决论元角色重叠问题, 将论元角色分开预测, 很好的克服了这一问题。此外, 作者针对数据不充分的问题, 提出使用 MASK 的方法将原数据中的论元 MASK 掉, 再用 BERT 去预测填补这个 MASK, 得到一个新的数据, 这是一种采用 BERT 做数据增强的方法, 同时很好的利用了 BERT 预训练阶段采用的 MASK 语言模型, 最终在 ACE2005 数据集上取得了 SOTA 的结果。

Wadden et al. [42] 利用 BERT 同时对多个句子进行表示, 学到范围较大的上下文信息, 然后同时对命名实体识别、事件抽取、关系抽取三个任务做多任务学习, 中间还加入了图传播层来缓解长距离依赖问题, 该模型在三个任务、四个数据集上取得了 SOTA 的结果, 其中包括常用于事件抽取的 ACE 2005 数据集。

由于 NLP 任务的每个模型都只能解决单独的任务, 不能对多个任务进行统一, 因此 McCann et al. [43] 提出了一种全新的范式: 多任务问答网络, 可以同时多个 NLP 任务上迁移, 并取得很不错的结果。基于这个工作, Du et al. [44] 和 Li et al. [45] 提出了使用 QA 模型来解决事件抽取任务。因为之前的事件抽取模型都需要依靠实体信息, 会导致错误传播问题, 而这种 QA 模型不需要实体信息, 直接对文本建模, 提前设定好问题模版, 通过 BERT 做 QA 的能力来抽取出需要的元素: 触发词和论元, [44] 的模型图如图8所示。实验展示, 该模型的结果稍优于以前的方法。

QA 范式下的事件抽取模型需要给出问题模版, 因此 [46] 提出了一种基于阅读理解范

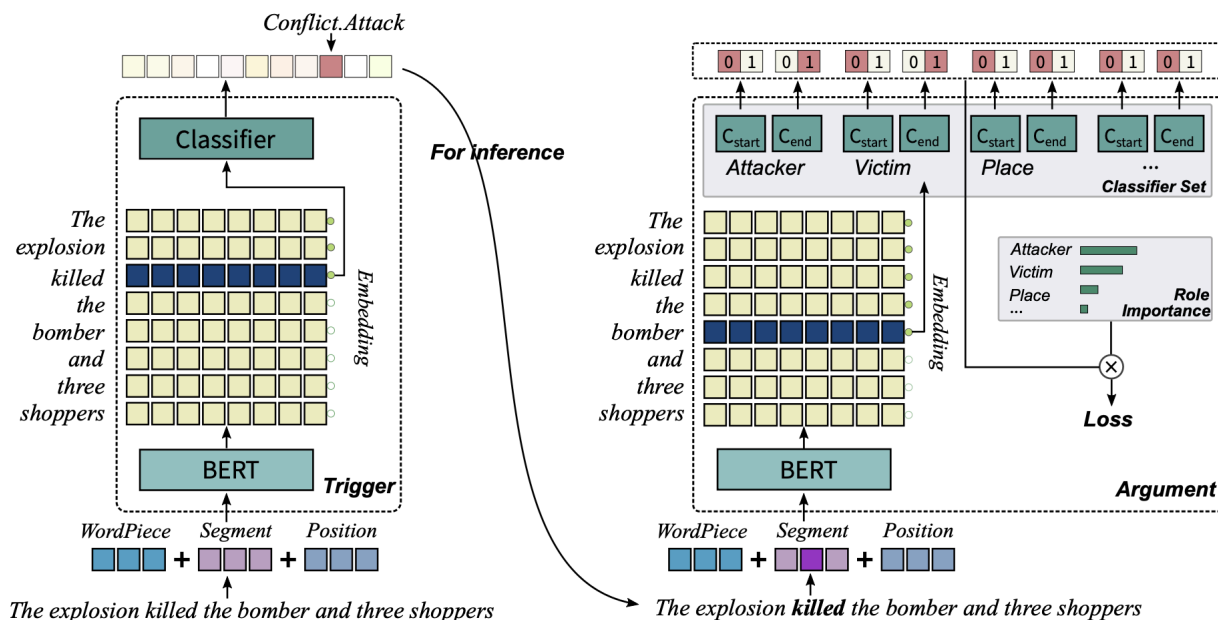


图 7: PLMEE 模型图 [41]

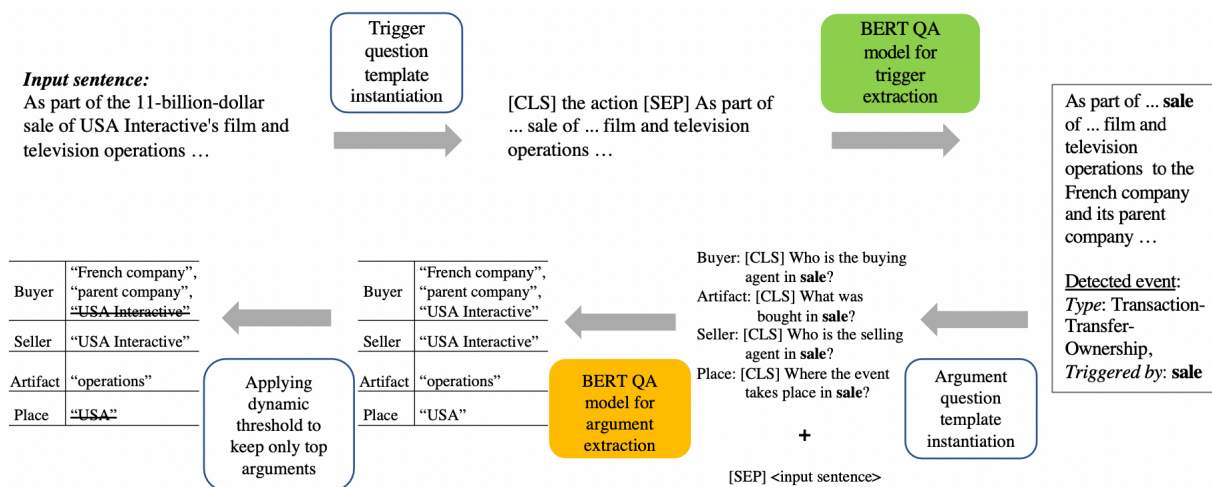


图 8: QA 范式下事件抽取模型图 [44]

式的模型，其实与 QA 类似，不同的是他使用无监督的方式自动生成问题，可以将事件规则迁移为问题，不仅在有监督方法上得到了 SOTA 的结果，还在低资源数据下与零样本情况下取得了很好的结果。

8 基于生成式范式与提示学习的事件抽取模型

提示学习 (Prompt) 首次出现在 2020 年 GPT-3 的文章中 [47], 其主要思想是修改输入, 使其适应预训练语言模型, 比如分类问题可以增加一个带有 MASK 标签的句子, 使其变成完形填空的问题, 这样就非常适合 BERT 做掩码预测。根据 Liu et al. [48], 如图9所示, 机器学习与神经网络范式下预训练语言模式和任务是没有直接关系的, 到了预训练微调范式下各个任务都在预训练语言模型的基础上做, 而提示学习是指让各个任务去适配预训练语言模型。


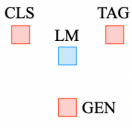
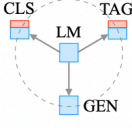
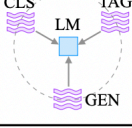
Paradigm	Engineering	Task Relation
a. Fully Supervised Learning (Non-Neural Network)	Features (e.g. word identity, part-of-speech, sentence length)	
b. Fully Supervised Learning (Neural Network)	Architecture (e.g. convolutional, recurrent, self-attentional)	
c. Pre-train, Fine-tune	Objective (e.g. masked language modeling, next sentence prediction)	
d. Pre-train, Prompt, Predict	Prompt (e.g. cloze, prefix)	

图 9: 四种 NLP 范式图 [48]

而抽取式任务通常采用对 Span 分类或序列标注的方式去解决, 若想加入 Prompt 作为辅助会有一些困难。而受到 T5 [49] 的启发, 可以将抽取式任务转换为生成式任务采用 Seq2Seq 的方法去做, 在生成式范式下再去采用 Prompt 就非常容易了。

第一个采用生成式范式做事件抽取的工作是 2021 年来自 AWS 的 Paolini et al. 的工作 TANL [50], 他们将结构预测任务 (命名实体识别, 实体关系抽取, 语义角色标注, 共指消解, 事件抽取, 对话状态追踪等) 均转换为类似于机器翻译的 Seq2Seq 形式, 并且采用了多任务学习, 其生成格式是对输入文本做了补充, 即在对应 Span 上进行标注, 如图10所示。图中的实体关系联合抽取任务在头实体上标注出实体信息, 在尾实体上标注实体信息、

关系以及对应头实体；语义角色标注任务会标注出主语、谓语等。在做事件抽取任务时也类似，将触发词抽取看做一种和实体抽取类似的标注任务，将论元抽取看成和触发词之间的关系抽取进行处理。

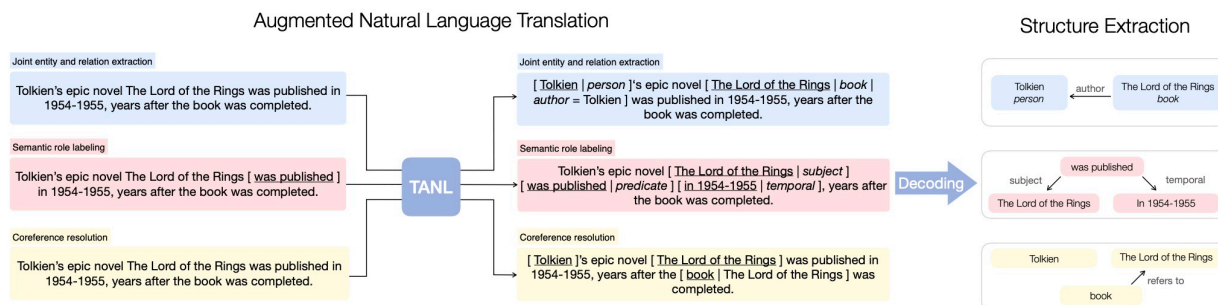


Figure 1: Our TANL model translates between input and output text in *augmented natural language*, and the output is then decoded into structured objects.

图 10: TANL 模型图 [50]

随后 Lu et al. 的工作 [34] 将事件的结构化信息转换成了一种线性的结构: (*EventType Trigger(Role Argument)...*), 再采用约束解码算法限制生成 Token 的范围, 取得了不错的结果。2022 年, Lu et al. 将结构化的工作进行了扩展, 提出了 UIE [51], 一统信息抽取任务。如图 11 所示, 他将命名实体识别、关系抽取和事件抽取采用两个原子操作进行表示, 并且均使用上述结构化格式进行表示, 同时加了 prompt 以区别不同的任务及进行对应类别的提示, 最终在几乎 13 个数据集上全部拿到了 SOTA 结果。

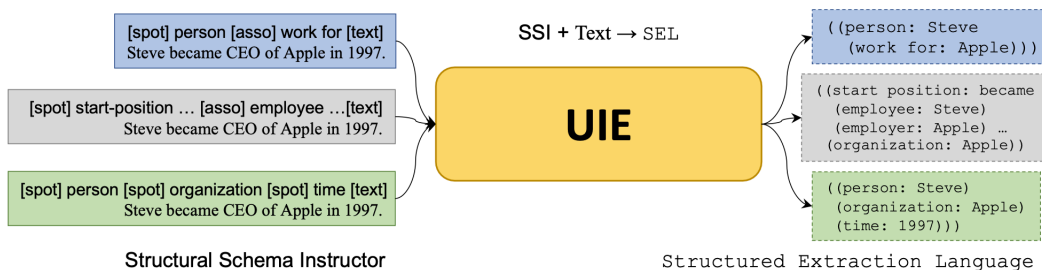


Figure 3: The overall framework of UIE.

图 11: UIE 模型图 [51]

除此之外, Huang et al. 在 2022 年提出了一种采用填补 Special Tokens 的方式来做生

成式事件抽取 [52], 由于 Special Tokens 是与语言无关的, 因此作者将其应用于跨语言的事件论元抽取上, 取得了非常好的效果, 特别是在 Zero-Shot 下。

上述生成式方法所生成出的文本格式均不是符合人类的自然语言的, 也就是说与预训练语言模型学到的自然语言知识之间存在一定的 Gap, 因此 Hsu et al. 提出了一种采用转述方式生成事件抽取信息的方式 [53]。模型如图12所示, 分为 pipeline-based 和 joint-based 两种方式。pipeline-based 是将事件检测和事件论元抽取分开做, 均是在输入文本后面拼接一个 Prompt, 事件检测的 Prompt 是由一句事件描述加上该事件的关键词以及模版组成的, 输出是将模版上的占位符换为对应的触发词。事件论元抽取的 Prompt 是由事件类型描述、触发词描述以及论元模版组成, 输出将模版中的占位符换为对应的论元。joint-based 是采用 End2End 方式同时抽取事件触发词和论元, 其 Prompt 是由事件类别描述、事件相关的关键词以及模版组成, 这里的模版是将 pipeline-based 中的模版拼接而成, 输出就是将模版中的占位符换成了对应的触发词与论元。虽然这篇文章没有在全监督设置下取得很好的结果, 但是在低资源情况下性能很好。

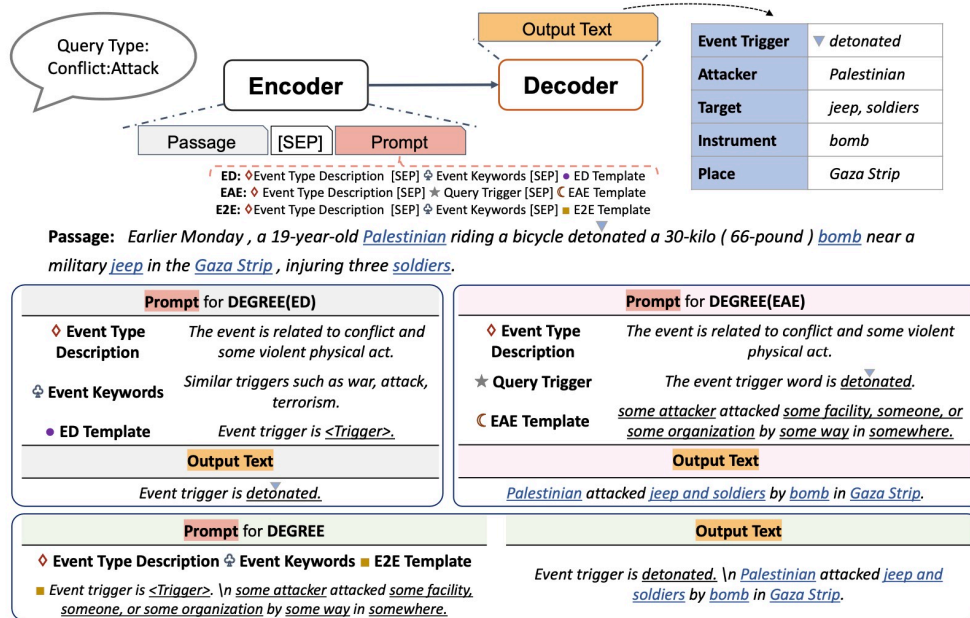


Figure 2: An illustration of our method. Given an event type, we feed a sequence containing a passage and a set of prompts to the model. The prompts vary depending on the queried type and the task (ED, EAE, or end-to-end). The model is trained to generate output texts following event-type-specific templates, which contain placeholders for filling triggers or arguments. The final event predictions can be made by comparing the template and the output text.

图 12: DEGREE 模型图 [53]

9 模型性能汇总

本节将神经事件抽取的经典工作在 ACE2005 数据集上的结果做一个汇总，包含每个工作采用的模型以及触发词分类 (TC) 和论元分类 (AC) 的结果，汇总如表2所示，加粗的为 SOTA 结果，结果均来自原论文。

表 2: 神经事件抽取经典模型在 ACE2005 数据集的性能汇总

Year - Method	Model	TC			AC		
		P	R	F1	P	R	F1
2015 - DMCNN [25]	CNN	75.6	63.6	69.1	62.2	46.9	53.5
2016 - JRNN [16]	RNN	66.0	73.0	69.3	54.2	56.7	55.4
2018 - JMEE [32]	LSTM&GCN	76.3	71.3	73.7	66.8	54.9	60.3
2019 - Joint3EE [31]	GRU	68.0	71.8	69.8	52.1	52.1	52.1
2019 - DYGIE++ [42]	BERT-large	-	-	69.7	-	-	48.8
2020 - OneIE [54]	BERT-large	-	-	74.7	-	-	56.8
2020 - EEQA [44]	2×BERT-base	71.1	73.7	72.4	56.8	50.2	53.3
2020 - MQAEE [45]	3×BERT-large	-	-	71.7	-	-	53.4
2021 - TANL [50]	T5-base	-	-	68.4	-	-	47.6
2021 - Multi-Task TANL [50]	T5-base	-	-	68.5	-	-	48.5
2021 - Text2Event [34]	T5-base	67.5	71.2	69.2	46.7	53.4	49.8
2021 - Text2Event [34]	T5-large	69.6	74.4	71.9	52.5	55.2	53.8
2021 - DEGREE-pipeline [53]	BART-large	-	-	72.2	-	-	56.0
2021 - DEGREE-joint [53]	BART-large	-	-	70.9	-	-	54.4
2022 - UIE [55]	UIE-large	-	-	73.4	-	-	54.8

10 未来展望

事件抽取任务是自然语言处理领域很重要一个部分，他建立了无结构的文本和事件信息之间的联系。随着深度学习的到来，严重耗费人力资源的特征模版设计部分已经逐渐消失，同时让事件抽取任务可以端到端的完成。在预训练语言模型的帮助下，可以利用大语料库中学习到的信息来解决事件抽取任务，并且不再局限为原本的分类任务和序列标注任

务，可以使用基于 QA 的范式和基于生成的模型来解决。随着 T5 和提示学习的出现，现在学术界事件抽取任务基本上转换为了生成式范式。即使现在事件抽取任务发展速度很快，SOTA 模型层出不穷，但仍然存在很多待解决的问题，以下将一一描述：

- 随着预训练语言模型的迅速发展，可以利用其庞大的语言知识量来提高事件抽取模型的性能，但预训练模型的训练任务都与事件抽取无关，尽管提示学习可以帮助任务靠近预训练模型，但目前的工作还不够，未来应该从提示学习或针对事件抽取的专用预训练语言模型入手，使得事件抽取可以更好的利用预训练模型。
- 现存的事件抽取数据集规模还是较小，而深度学习方法都需要大规模数据，因此标注一个大规模事件抽取语料库或设计相应模型自动生成一个是很重要的事情，可以更好的利用深度模型。
- 事件抽取任务分为句子级、篇章级和跨文档级，目前句子级的模型很多，但篇章级的较少，跨文档级的几乎没有，而真实场景应用中篇章级的事件抽取是非常常见的，也是一个很迫切需要解决的问题。这个任务的难点是长距离依赖问题，即使目前已经有人提出一些方法来解决了，但还远远没有达到使用的标准。
- 目前事件抽取任务大都集中在语料库丰富的英语文本中，针对汉语的模型也逐渐增多，但对于缺乏标注数据的小语种依然很难解决。因此，研究多语言模型或跨语言模型是很必要的。
- 事件抽取分为封闭域下和开放域下，目前大部分工作都是基于事件规则的封闭域任务，而事件规则无法覆盖所有的情况，很多专业领域的事件都是独立的，因此开放域的研究是很重要的。开放域事件抽取研究遇到的难点是：一是缺少大规模、高质量的标注数据；二是需要一个合理的成熟的评价指标来评估开放域事件抽取任务的结果；三是开放域任务需要去发展更适合它的模型、技术。
- 真实场景下的数据形式非常丰富，不仅只有文本数据，比如在电影中有视频、声音和字幕三种模态的数据，在这种场景下做事件抽取最好可以结合多模态进行分析，而目前多模态事件抽取的研究还相对较少，需要未来继续去进一步研究。
- 即使预训练语言模型已经在大规模语料中学习到了大量知识，但是在下游任务阶段，由于数据集大小的有限，所学习到的知识也是有限的，因此可以引入外部知识，想办法对下游任务进行辅助增强。

11 总结

本文主要介绍了将深度学习用于事件抽取任务的模型，神经网络强大的表示能力可以很好的解决传统机器学习繁琐的特征工程问题。本文先介绍了事件抽取任务的定义、相关数据集以及评价指标，然后分别介绍了卷积神经网络、循环神经网络、注意力机制、图神经网络、预训练-微调范式以及生成式范式与提示学习在事件抽取方向的应用以及发展过程。从这一发展过程其实可以看到事件抽取任务从一开始的分类任务或序列标注任务到预训练模型之后的 QA 任务、生成式任务，所使用的范式在逐渐改变，新出现的范式使得事件抽取任务不再依赖实体标签。自 BERT 以来的很多模型都可以用在各种任务上，自然语言处理领域的任务有逐渐统一的趋势。

参考文献

- [1] G. R. Doddington, A. Mitchell, M. A. Przybocki, L. A. Ramshaw, S. M. Strassel, and R. M. Weischedel, “The automatic content extraction (ace) program-tasks, data, and evaluation.,” in *Lrec*, vol. 2, pp. 837–840, Lisbon, 2004.
- [2] D. Ahn, “The stages of event extraction,” in *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pp. 1–8, 2006.
- [3] C.-Y. Liu, C. Zhou, J. Wu, H. Xie, Y. Hu, and L. Guo, “Cpmf: A collective pairwise matrix factorization model for upcoming event recommendation,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1532–1539, IEEE, 2017.
- [4] J. Boyd-Graber and B. Börschinger, “What question answering can learn from trivia nerds,” *arXiv preprint arXiv:1910.14464*, 2019.
- [5] A. Bosselut, R. Le Bras, and Y. Choi, “Dynamic neuro-symbolic knowledge graph construction for zero-shot commonsense question answering,” in *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, 2021.
- [6] H. Tanev, J. Piskorski, and M. Atkinson, “Real-time news event extraction for global crisis monitoring,” in *International Conference on Application of Natural Language to Information Systems*, pp. 207–218, Springer, 2008.
- [7] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, “Real-time event extraction for driving information from social sensors,” in *2012 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, pp. 221–226, IEEE, 2012.
- [8] J. Sheng, S. Guo, B. Yu, Q. Li, Y. Hei, L. Wang, T. Liu, and H. Xu, “Casee: A joint learning framework with cascade decoding for overlapping event extraction,” *arXiv preprint arXiv:2107.01583*, 2021.
- [9] H. Peng, J. Li, Y. Song, R. Yang, R. Ranjan, P. S. Yu, and L. He, “Streaming social event detection and evolution discovery in heterogeneous information networks,” *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 15, no. 5, pp. 1–33, 2021.

- [10] Q. Wei, Z. Ji, Z. Li, J. Du, J. Wang, J. Xu, Y. Xiang, F. Tiryaki, S. Wu, Y. Zhang, *et al.*, “A study of deep learning approaches for medication and adverse drug event extraction from clinical text,” *Journal of the American Medical Informatics Association*, vol. 27, no. 1, pp. 13–21, 2020.
- [11] E. Filtz, M. Navas-Loro, C. Santos, A. Polleres, and S. Kirrane, “Events matter: Extraction of events from court decisions,” *In Legal Knowledge and Information Systems, Legal Knowledge and Information Systems*, pages 33–42. IOS Press, 2020.
- [12] F. Hogenboom, F. Frasinca, U. Kaymak, F. de Jong, and E. Caron, “A survey of event extraction methods from text for decision support systems,” *Decision Support Systems*, vol. 85, pp. 12–22, 2016.
- [13] F. Hogenboom, F. Frasinca, U. Kaymak, and F. De Jong, “An overview of event extraction from text.,” in *DeRiVE@ ISWC*, pp. 48–57, Citeseer, 2011.
- [14] J. Deng, F. Qiao, H. Li, X. Zhang, and H. Wang, “An overview of event extraction from twitter,” in *2015 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*, pp. 251–256, IEEE, 2015.
- [15] W. Xiang and B. Wang, “A survey of event extraction from text,” *IEEE Access*, vol. 7, pp. 173111–173137, 2019.
- [16] T. H. Nguyen, K. Cho, and R. Grishman, “Joint event extraction via recurrent neural networks,” in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 300–309, 2016.
- [17] T. Mitamura, Z. Liu, and E. H. Hovy, “Overview of tac kbp 2015 event nugget track.,” in *TAC*, 2015.
- [18] H. Ji, J. Nothman, H. T. Dang, and S. I. Hub, “Overview of tac-kbp2016 tri-lingual edl and its impact on end-to-end cold-start kbp,” *Proceedings of TAC*, 2016.
- [19] T. Mitamura, Z. Liu, and E. H. Hovy, “Events detection, coreference and sequencing: What’s next? overview of the tac kbp 2017 event track.,” in *TAC*, 2017.

- [20] X. Wang, Z. Wang, X. Han, W. Jiang, R. Han, Z. Liu, J. Li, P. Li, Y. Lin, and J. Zhou, “MAVEN: A massive general domain event detection dataset,” in *Proceedings of EMNLP 2020*, 2020.
- [21] Y. Zhou, Y. Chen, J. Zhao, Y. Wu, J. Xu, and J. Li, “What the role is vs. what plays the role: Semi-supervised event argument extraction via dual question answering,” in *Proceedings of AAAI-21*, AAAI Press, 2021.
- [22] S. Petrovic, M. Osborne, R. McCreadie, C. Macdonald, I. Ounis, and L. Shrimpton, “Can twitter replace newswire for breaking news?,” in *Seventh international AAAI conference on weblogs and social media*, 2013.
- [23] S. Zheng, W. Cao, W. Xu, and J. Bian, “Doc2edag: An end-to-end document-level framework for chinese financial event extraction,” in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 337–346, EMNLP-IJCNLP, 2019.
- [24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
- [25] Y. Chen, L. Xu, K. Liu, D. Zeng, and J. Zhao, “Event extraction via dynamic multi-pooling convolutional neural networks,” in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 167–176, 2015.
- [26] T. H. Nguyen and R. Grishman, “Event detection and domain adaptation with convolutional neural networks,” in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 365–371, 2015.
- [27] T. H. Nguyen and R. Grishman, “Modeling skip-grams for event detection with convolutional neural networks,” in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 886–891, 2016.

- [28] Z. Zhang, W. Xu, and Q. Chen, “Joint event extraction based on skip-window convolutional neural networks,” in *Natural Language Understanding and Intelligent Applications*, pp. 324–334, Springer, 2016.
- [29] X. Feng, B. Qin, and T. Liu, “A language-independent neural network for event detection,” *Science China Information Sciences*, vol. 61, no. 9, pp. 1–12, 2018.
- [30] L. Sha, F. Qian, B. Chang, and Z. Sui, “Jointly extracting event triggers and arguments by dependency-bridge rnn and tensor-based argument interaction,” in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [31] T. M. Nguyen and T. H. Nguyen, “One for all: Neural joint modeling of entities and events,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 6851–6858, 2019.
- [32] X. Liu, Z. Luo, and H.-Y. Huang, “Jointly multiple events extraction via attention-based graph information aggregation,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 1247–1256, 2018.
- [33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in neural information processing systems*, pp. 5998–6008, 2017.
- [34] Y. Lu, H. Lin, J. Xu, X. Han, J. Tang, A. Li, L. Sun, M. Liao, and S. Chen, “Text2event: Controllable sequence-to-structure generation for end-to-end event extraction,” *arXiv preprint arXiv:2106.09232*, 2021.
- [35] L. Yao, C. Mao, and Y. Luo, “Graph convolutional networks for text classification,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 7370–7377, 2019.
- [36] T. H. Nguyen and R. Grishman, “Graph convolutional networks with argument-aware pooling for event detection,” in *Thirty-second AAAI conference on artificial intelligence*, 2018.
- [37] H. Wen, Y. Qu, H. Ji, Q. Ning, J. Han, A. Sil, H. Tong, and D. Roth, “Event time extraction and propagation via graph attention networks,” in *Proceedings of the 2021*

- Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 62–73, 2021.
- [38] W. U. Ahmad, N. Peng, and K.-W. Chang, “Gate: Graph attention transformer encoder for cross-lingual relation and event extraction,” in *The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)*, 2021.
- [39] M. Peters, M. Neumann, M. Iyyer, M. Gardner, and L. Zettlemoyer, “Deep contextualized word representations,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2018.
- [40] J. D. M.-W. C. Kenton and L. K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of NAACL-HLT*, pp. 4171–4186, 2019.
- [41] S. Yang, D. Feng, L. Qiao, Z. Kan, and D. Li, “Exploring pre-trained language models for event extraction and generation,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 5284–5294, 2019.
- [42] D. Wadden, U. Wennberg, Y. Luan, and H. Hajishirzi, “Entity, relation, and event extraction with contextualized span representations,” in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5784–5789, 2019.
- [43] B. McCann, N. S. Keskar, C. Xiong, and R. Socher, “The natural language decathlon: Multitask learning as question answering,” *arXiv preprint arXiv:1806.08730*, 2018.
- [44] X. Du and C. Cardie, “Event extraction by answering (almost) natural questions,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 671–683, 2020.
- [45] F. Li, W. Peng, Y. Chen, Q. Wang, L. Pan, Y. Lyu, and Y. Zhu, “Event extraction as multi-turn question answering,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pp. 829–838, 2020.

- [46] J. Liu, Y. Chen, K. Liu, W. Bi, and X. Liu, “Event extraction as machine reading comprehension,” in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1641–1651, 2020.
- [47] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.*, “Language models are few-shot learners,” *arXiv preprint arXiv:2005.14165*, 2020.
- [48] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, “Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing,” *arXiv preprint arXiv:2107.13586*, 2021.
- [49] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, “Exploring the limits of transfer learning with a unified text-to-text transformer,” *J. Mach. Learn. Res.*, vol. 21, pp. 140:1–140:67, 2020.
- [50] G. Paolini, B. Athiwaratkun, J. Krone, J. Ma, A. Achille, R. Anubhai, C. N. dos Santos, B. Xiang, and S. Soatto, “Structured prediction as translation between augmented natural languages,” in *9th International Conference on Learning Representations, ICLR 2021*, 2021.
- [51] Y. Lu, Q. Liu, D. Dai, X. Xiao, H. Lin, X. Han, L. Sun, and H. Wu, “Unified structure generation for universal information extraction,” in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 2022.
- [52] K.-H. Huang, I.-H. Hsu, P. Natarajan, K.-W. Chang, and N. Peng, “Multilingual generative language models for zero-shot cross-lingual event argument extraction,” in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2022.
- [53] I. Hsu, K.-H. Huang, E. Boschee, S. Miller, P. Natarajan, K.-W. Chang, N. Peng, *et al.*, “Degree: A data-efficient generative event extraction model,” in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2022.

-
- [54] Y. Lin, H. Ji, F. Huang, and L. Wu, “A joint neural model for information extraction with global features,” in *Proceedings of The 58th Annual Meeting of the Association for Computational Linguistics*, 2020.
- [55] Y. Lu, Q. Liu, D. Dai, X. Xiao, H. Lin, X. Han, L. Sun, and H. Wu, “Unified structure generation for universal information extraction,” in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, 2022.