

收到 2019 年 11 月 4 日, 接受 2019 年 11 月 26 日, 发布日期 2019 年 11 月 29 日, 当前版本的日期 2019 年 12 月 13 日。

数字对象标识符 10.1109 / ACCESS.2019.2956831

从文本中提取事件的概述

伟乡 和王邦
华中科技大学电子信息与通信学院, 武汉 430074

通讯作者: 王望 (wangbang@hust.edu.cn)

这项工作得到了国家自然科学基金的部分资助, 资助号为 61771209。

摘要每天都有无处不在的重要事件发生, 但在不同媒体来源以不同叙事方式进行报道。如何检测文章和帖子中是否已报告真实事件是事件提取的主要任务之一。其他任务包括提取事件参数并确定其作用, 以及聚类跟踪来自不同文本的类似事件。作为自然语言处理和理解中最重要的研究主题之一, 事件提取在不同领域中具有广泛的应用, 并且已经进行了数十年的深入研究。本文为从文本中提取事件提供了全面而最新的调查。我们不仅为事件提取总结任务定义, 数据源和性能评估, 还为解决方案方法提供分类。在每个解决方案组中, 我们都会对最具代表性的方法进行详细分析, 尤其是它们的起源, 基础, 优点和缺点。最后, 我们还提出了有关未来研究方向的设想。

索引条款事件提取, 事件提取任务, 事件语料库, 自然语言处理。

I. 介绍

事件是在某个时间和某个地方发生的, 涉及一个或多个参与者的事物的特定发生, 通常可以将其描述为状态变化[1]。事件提取的目的是检测文本中的事件实例, 如果存在, 则标识事件类型及其所有参与者和属性。尽管可以使用不同的参数定义不同的事件类型, 但是事件提取的简单总结是从非结构化自然语言中获取事件的结构化表示, 从而帮助回答“5W1H”问题, 包括“谁, 何时, 何地”大量新闻来源(例如新闻文章, 社交媒体帖子等)中发生的现实事件的“什么, 为什么, 为什么”和“如何”。

作为自然语言处理(NLP)中信息检索的重要任务, 事件提取在不同领域中具有许多应用。例如, 结构化事件可以直接用于扩展知识库, 在此基础上可以进行进一步的逻辑推理和推断[2], [3]。事件检测和监视长期以来一直是政府公共事务管理的重点, 因为及时了解流行的社会事件的爆发和演变有助于当局迅速做出反应[4]–[8]。在业务和金融领域, 事件提取还可以帮助公司快速发现其产品的市场响应以及推断信号以进行风险分析和评估。

交易建议[9]–[11]。在生物医学领域, 事件提取可用于识别生物分子(例如基因和蛋白质)状态的变化或两个或多个生物分子之间的相互作用, 这在科学文献中以自然语言进行了描述, 以了解生理和发病机理机制[12]。简而言之, 许多领域都可以从事件提取技术和系统的进步中受益。

尽管事件提取具有广阔的前景, 但事件提取仍然是一项颇具挑战性的任务, 因为事件具有不同的结构和组件。而自然语言通常具有语义歧义和话语风格。此外, 事件提取还与其他 NLP 任务密切相关, 例如命名实体识别(NER), 词性(POS)标记, 语法分析等, 它们可以促进事件提取或对其性能产生反作用, 取决于这些任务的执行方式以及如何利用其输出。为了促进事件提取的发展和应用, 已经进行了许多公共评估程序来提供任务定义, 带注释的语料库以及公开竞赛, 以促进诸如事件提取之类的信息提取研究, 这也吸引了许多人才来贡献新颖的算法, 技术和系统。接下来, 我们简要介绍这些著名的程序。

A. 公众评价计划

消息理解会议(MUC)[13], [14]被普遍认为是首次公开评估

协理这份手稿的审查并批准发表的副编辑是 Nikhil Padhi。

信息提取程序由国防高级研究计划局 (DARPA) 组织和赞助。¹ 从 1987 年到 1997 年举行了七次。MUC 旨在从非结构化文本中提取信息并将其填充到预定义架构中的值结构。一些常见的时段包括实体, 属性, 实体关系, 事件等。1997 年, DARPA, 卡内基·梅隆大学, Dragon Systems 和 马萨诸塞州阿默斯特大学共同创立了另一个公共评估程序, 称为主题检测和跟踪 (TDT)。² 以促进在一系列广播新闻文章中发现和关注新事件 [15], [16]。后来, 美国国家标准技术研究院 (NIST)² 建立了针对 TDT 计划的完整评估系统。

自动内容提取 (ACE) 是迄今为止最具影响力的公共评估计划, 它是由 NIST 于 1999 年提出的, 后来又于 2009 年并入了新的公共评估计划文本分析会议 (TAC)。从 2000 年开始直到 2004 年, ACE 一直致力于实体和关系的检测和跟踪, 并且自 2005 年以来, 事件提取任务已附加到 ACE 中 [17]。继 ACE 之后, DARPA 的“文本深度探索和过滤” (DEFT) 程序提出了用于文本注释和信息提取的实体, 关系, 事件 (ERE) 标准。Light ERE 被定义为 ACE 注释的简化版本, 以便快速生成一致标记的数据。随后, Light ERE 已扩展到更复杂的 Rich ERE 规范 [18]。

此外, 事件抽取已成为知识库人口 (KBP) 公共评估计划的主要任务, 该计划自 2014 年至今已举行了四次。现在, KBP 已与 TAC 集成, 其目的是从大型文本语料库中提取信息, 以完成知识库的不足元素 [19]。此外, 还有一些其他事件公共评估程序, 用于特定领域中的事件提取, 例如生物医学领域中的 BioNLP [20], 用于提取事件的时间信息的 TimeBANK [21]。

B. 本调查摘要

本文提供了从文本中提取事件的最新调查。我们注意到, 有一些与此任务相关的调查文章, 但是每个文章都针对特定的应用领域而特别关注。Hogenboom 等。[22], [23] 回顾了用于各种决策支持系统中事件提取的文本挖掘技术。Vanegas 等。[12] 主要综述了生物分子事件的提取; 而张等。[24] 主要关注开放域事件提取; 一些人还专注于从社交媒体, 尤其是从 Twitter [25], [26] 提取事件。

¹<https://www.darpa.mil>

²<https://www.nist.gov>

与上述文章相比, 我们尝试提供更全面的调查和系统的技术分类法, 以从文本中提取事件, 不仅提供其任务定义, 数据源和性能评估, 而且从其角度对主要方法进行整个发展历史。此外, 我们介绍并分析每种技术类别中最具代表性的方法, 尤其是它们的起源, 基础, 优点和缺点。我们还讨论了事件提取的未来研究的有希望的方向。

本文的其余部分安排如下: 第二部分介绍了封闭域和开放域事件提取中的主要任务定义。第三部分介绍了最常用的语料库。我们将主要技术方法分为 5 组, 包括用于封闭域事件提取的方法, 第四节中的事件模式匹配的早期方法, 第五节中的机器学习方法, 第六节中的深度学习模型, 第二节中的半监督学习方案。第七节, 对于开放域事件提取, 第八节中的无监督学习方法。在第 IX 节中, 我们还比较了在 ACE 语料库上实验的那些算法的提取性能。最后, 第 X 节以有关未来研究方向的一些讨论来结束调查。

II. 事件提取任务

事件提取旨在检测文本中报告的事件的存在, 如果存在, 则从文本中发现与事件相关的信息, 例如有关事件的“5W1H” (即, 谁, 何时, 何地, 什么, 为什么如何)。有时, 特定的事件结构是预定义的, 不仅包括事件类型, 还包括事件参数的角色。事件提取不仅需要检测事件, 还需要提取相应的字符/单词/短语来填充给定的事件结构, 以便输出结构化的事件形式。这通常称为封闭域事件提取, 因为不同的域可能需要不同的事件结构。另一方面, 开放域事件提取的任务未采用预定义的事件结构, 主要任务是检测文本中事件的存在。在许多情况下, 它还会提取有关事件的关键字并将相似的事件聚类。

A. 封闭域事件提取

封闭域事件提取使用预定义的事件模式来发现并从文本中提取特定类型的所需事件。事件模式包含几种事件类型及其对应的事件结构。我们使用 ACE [1] 术语来介绍事件结构, 如下所示:

- **事件提及:** 描述事件的短语或句子, 包括触发器和多个参数。
- **事件触发:** 最清楚地表示事件发生的主要单词, 通常是动词或名词。

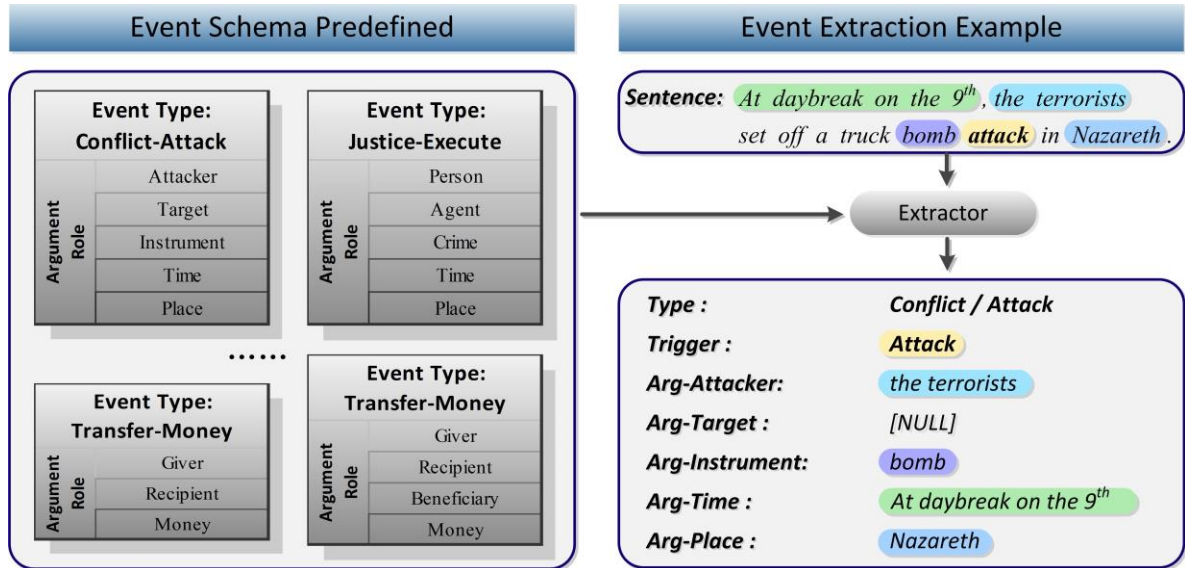


图 1. 闭域事件提取的示意图。左侧部分说明了 ACE 2005 中的一些预定义事件架构。右侧部分说明了用于触发检测，事件类型标识，参数检测和参数角色标识的四个子任务的提取结果。

- **事件自变量**：在事件中具有特定角色的实体提及，时间表达或值。
- **参数角色**：参数与其参与的事件之间的关系。

请注意，同一事件可能在不同的句子或文档中多次提及。如何区分同一事件的多个事件提及是另一个重要的 NLP 任务，称为事件共同引用。在本文中，我们不考虑事件提及共同引用解决方案。

Ahn [27] 首先提出将 ACE 事件提取任务分为四个子任务：触发检测，事件/触发类型识别，事件自变量检测和自变量角色识别。例如，考虑以下句子：

句子 1：在 9 日破晓时，恐怖分子拿撒勒发生卡车炸弹袭击。

句子 1 中存在类型为“冲突/攻击”的事件。事件提取器应通过检测句子中的触发词“攻击”并将其分类为事件，来发现此类事件并识别其类型。“冲突/攻击”的类型。接下来，应从文本中提取与此事件类型相关的所有自变量，并根据预定义的事件结构来标识它们各自的角色。

图 1 说明了结构化事件的封闭域提取。左侧部分说明了 ACE 2005 中的一些预定义事件架构。右侧部分说明了用于触发检测，事件类型标识，参数检测和参数角色标识的四个子任务的提取结果。

遵循 ACE [1] 的类似定义，现场定义和采用了许多其他事件类型和结构，例如 ERE 和 TAC-KBP [17] – [19] 所定义的事件类型和结构。除了组织，个别研究人员还为特定领域定义了事件类型和结构。

Petroni 等。[28] 定义了用于破坏事件的结构，包括 7 种事件类型，如“洪水”，“暴风雨”，“火灾”等，以及它们的“5W1H”属性，以便提取新闻报道和社交媒体上的重大事件。杨等。[29] 专注于提取金融领域的事件以帮助预测股票市场，投资决策支持等。他们定义了 9 种金融事件类型，例如“股权质押”，“股权冻结”等。以及它们具有不同作用的相应论点。Han 等。[30] 定义了一个业务事件分类法，该事件包含 8 个事件类型和 16 个子类型及其相应的参数，例如时态，时间，结果，实体等。

B. 开放域事件提取

在没有预定义事件模式的情况下，开放域事件提取旨在检测文本中的事件，并且在大多数情况下，还通过提取的事件关键字对相似事件进行聚类。事件关键字是指主要描述事件的那些词/短语，有时关键字又被进一步分为触发器和参数。

TDT 公众评估计划旨在自动发现以前未报告的事件，或者跟踪新闻文章中先前发现的事件的进展[15]。除了事件之外，TDT 还将故事定义为描述特定事件的新闻文章的一部分，将主题定义为与现实世界中的主题紧密相关的文章中的一组事件。基于这样的定义，它定义了以下任务：

- **故事细分**：从新闻文章中检测故事的边界。
- **第一个故事检测**：检测在新闻流中讨论新话题的故事。
- **主题检测**：根据他们讨论的主题对故事进行分组。

- **主题跟踪**: 检测讨论先前已知主题的故事。
- **故事链接检测**: 确定两个故事是否讨论相同的主题。

前两个任务主要集中在事件检测上。其余三个任务用于事件聚类。虽然这五个任务之间的关系是显而易见的, 但每个任务都需要一个不同的评估过程, 并鼓励采用不同的方法来解决特定问题。

除了 TDT 任务外, 还进行了许多其他研究来检测和聚类新闻文章中的开放域事件[6], [31]–[34]。例如, 欧洲委员会联合研究中心进行了调查, 从在线新闻中提取带有被杀, 受伤, 绑架等关键词的暴力事件, 以进行全球危机监控[6], [31]。Yu 和 Wu [33]将有关同一事件的新闻文章聚合到以主题为中心的集合中。而刘等。[34]根据有关政治, 经济, 社会, 体育, 娱乐等方面的日常重大事件, 将新闻报道聚类。

一些工作集中于句子级事件检测和聚类[35]–[37]。例如, Naughton 等。[35]在新闻文章中将引用同一事件的句子进行了分组, 其中非事件句子在启动聚类过程之前被删除。他们使用了一系列从不同来源收集来的新闻故事, 描述了与伊拉克战争有关的事件。此外, 在对事件相关句子进行分组时, 他们设计了聚类标签, 例如恐怖袭击, 轰炸, 射击, 空袭等。除了事件检测和聚类, Wang 等人。[36]还建议为每个事件提取关键字, 例如事件的类型, 位置, 时间和人员。

除了新闻专栏文章外, 许多在线社交媒体 (例如 Twitter 和 Facebook 等) 还提供了有关各种事件的大量及时信息。最近, 从社交媒体中检测和提取事件也已成为一项重要任务[28], [38]–[43]。值得注意的是, 由于社交网络中的帖子属于非官方文本, 具有很多缩写, 拼写错误和语法错误, 因此如何从此类在线帖子中提取事件比从新闻文章中提取事件面临更多的挑战。

尽管检测和聚类事件是开放域事件提取中的主要任务, 但一些研究人员还建议通过为每个事件聚类分配一个事件类型标签以及一个或多个事件, 来进一步从聚类的事件相关语句和文档中构造事件模式。事件属性标签[44]–[52]。请注意, 除了与封闭域事件提取中的结构清晰的预定义标签不同, 可以更好地将此类簇标签解释为一种根据每个簇的关键字进行的语义综合。

III. 事件提取语料库

本节主要介绍事件提取任务的语料库资源。通常, 公共评估程序会为事件提取的任务评估提供多个语料库。

语料库根据任务定义由公共评估程序手动注释, 也用于机器学习方法中的模型训练和验证, 样本注释由具有领域知识的专业人员或专家完成, 可以考虑注释后的样本与地面真相标签一样。但是, 由于注释过程成本高昂, 因此许多公共语料库规模小且覆盖率低。

A. ace 事件语料库

ACE 程序[1]为各种提取任务 (包括实体, 时间, 值, 关系和事件提取) 提供了带注释的数据和评估工具。ACE 中的实体分为 7 种类型 (人员, 组织, 位置, 地缘政治实体, 设施, 车辆和武器), 每种类型都有许多子类型。此外, 时间是根据 TIMEX2 标准[53], [54]进行注释的, 该标准是用于自然语言文本中事件和时间表达的丰富规范语言。每个文本样本均由两个独立的注释者进行双重注释, 而高级注释者将裁定它们之间的版本差异。

ACE 语料库中的事件具有复杂的结构和参数, 涉及实体, 时间和值。ACE 2005 事件语料库定义了 8 个事件类型和 33 个子类型, 每个事件子类型对应一组自变量角色。所有事件子类型共有 36 个自变量角色。在大多数基于 ACE 语料库的研究中, 事件的 33 个子类型通常被单独处理, 而无需进一步检索其层次结构。表 1 提供了这些事件类型及其对应的子类型。图 2 提供了一个带注释的事件示例的示例。ACE2005 语料库总共包含 599 个带注释的文档和大约 6000 个带标签的事件, 包括来自新闻媒体, 广播新闻等不同媒体来源的英语, 阿拉伯语和中文事件。表 2 提供了它们的源统计信息。

表 1. ACE 2005 中的事件类型和子类型。

SN	Event Type	SN	Event subtype
1	Life	1-5	Be-Born, Marry, Divorce, Injure, Die
2	Movement	6	Transport
3	Contact	7-8	Meet, Phone-write
4	Conflict	9-10	Attack, Demonstrate
5	Business	11-14	Merge-org, Declare-bankruptcy, Start-org, End-org
6	Transaction	15-16	Transfer-money, Transfer-ownership
7	Persosnnel	17-20	Elect, Start-position, End-position, Nominate
8	Justice	21-33	Arrest-jail, Execute, Pardon, Release-parole, Fine, Convict, Charge-indict, Trial-hearing, Acquite, Sentence, Sue, Extradite, appeal

B. tac-kbp 语料库

TAC-KBP 中的事件块检测任务着重于检测事件的显式提及及其类型和

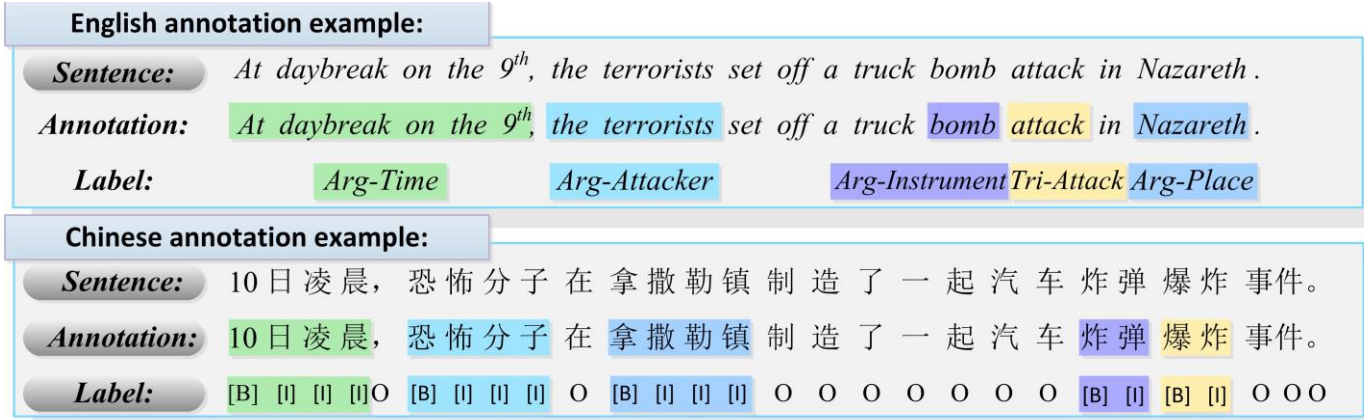


图 2. 事件注释的示例。在中文中，事件以字符 BIO 标签（开始/中间/其他）注释。

表 2. ACE 2005 中的数据源和数据统计。

Data sources	Language		
	English	Chinese	Arabic
Newswire (NW)	20%	40%	40%
Broadcast News (BN)	20%	40%	40%
Broadcast Conversation (BC)	15%	0%	0%
Weblog (WL)	15%	20%	20%
Usenet Newsgroups (UN)	15%	0%	0%
Conversational Telephone Speech (CTS)	15%	0%	0%

Rich ERE 中定义的子类型。语言数据协会 (LDC) 提供的 TAC-KPB 2015 产品包括 158 份文档是先前的培训集，而 202 份其他文档是测试集，用于接受新闻专栏文章和讨论论坛的正式评估[55]。TAC-KBP (Rich ERE) 中的事件类型和子类型是根据 ACE 语料库定义的，包括 9 个事件类型和 38 个子类型。此外，必须将事件提及分配到三个 REALIS 值之一：ACTUAL（实际发生），GENERIC（没有特定时间或地点）和 OTHERS（非通用事件，例如失败事件，未来事件和条件语句）等）[19]。TAC-KBP 2015 语料库仅用于英语评估，但是，TAC-KBP 2016 中为所有任务添加了中文和西班牙语。

C. tdt 语料库

在开放域事件提取中，LDC 还提供了一系列语料库来支持 TDT 从 TDT-1 到 TDT-5 的研究，包括中英文文本（普通话）和语音[15]，[56]。每个 TDT 语料库都包含数百万个新闻故事，这些新闻故事带有从新闻专线和广播文章等多个来源收集的数百个主题（事件）。此外，LDC 将所有音频材料转换为文本中介。如果故事讨论目标主题，则为这些故事主题标签分配值为 YES；如果该讨论包括

不到故事的 10%。否则，（默认）标签值为 NO。

D. 其他领域专用语料库

除上述众所周知的语料库外，还建立并发布了一些特定领域的事件语料库。BioNLP 共享任务 (BioNLP-ST) 的定义是从生物医学领域的科学文档中细粒度生物分子事件提取，该文件汇编了各种手动注释的生物语料库，包括 GENIA 事件语料库，BioInfer 语料库，基因调控事件语料库，GeneReg 语料库和 PPI 语料库[12]，[20]。TERQAS（问答系统的时间和事件识别）研讨会建立了一个名为 TimeBANK 的语料库，该语料库被注释了来自各种媒体来源的事件，时间和时间关系，用于突发新闻事件提取[21]。孟等人。[57]还注释了中文报道的突发新闻事件，称为 CEC（中文事件语料库）。其他特定领域的语料库包括用于军事情报领域的事件提取的 MUC 系列语料库，恐怖袭击，芯片技术和金融[13]，[14]，以及 Ding 等人的用于事件提取的语料库。音乐[58]。

IV. 基于模式匹配的事件提取早期的事件提取方法是一种模式匹配技术，该技术首先构造一些特殊的事件模板，然后执行模板匹配以从文本中提取带有单个参数的事件。如图 3 所示，事件模板可以从原始文本或带注释的文本构建，但是两者都需要专业知识。在在线提取阶段，如果事件及其自变量与预定义的模板匹配，则它们将被提取。

A. 手工图案构造

我们首先回顾一些典型的基于模式的事件提取系统，该系统使用具有专业知识的专家来手动构建不同领域的事件模式

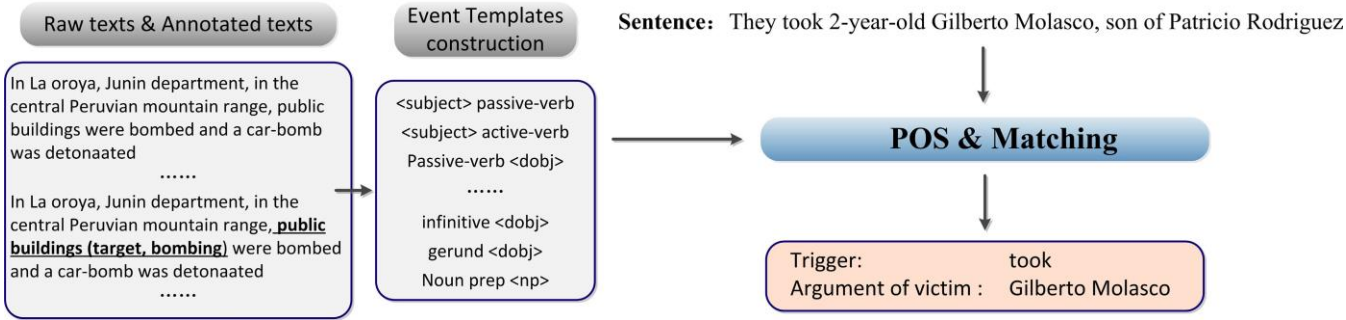


图 3. 事件模式构建和基于模式匹配的事件提取的图示。

应用程序。第一个基于模式的提取系统可以追溯到 Riloff 等人 在 1993 年开发的 AutoSlog。[59]，这是提取恐怖事件的特定领域。AutoSlog 利用一小套语言模式和一个手动注释的语料库来获取事件模式。如表所示。

3, 总共 13 lin-

定义了语言模式，例如“<subject>被动-动词”，意思是被动形式的短语动词，后面是作为主语的语法元素。请注意，语言模式与 AutoSlog 中的事件模式是不同的。语言模式用于根据手动注释的语料库自动建立事件模式；事件模式用于事件提取。此外，AutoSlog 旨在提取具有单个事件参数的事件，因此对它们的主体进行了注释。

每个事件都有一个参数。例如，在下面的句子中，带下划线的短语“公共建筑”是带有“炸弹”类型事件的注释为“目标”角色的参数。

句子 2: 在秘鲁中部山脉朱宁省拉奥罗亚省，炸毁了公共建筑（目标，轰炸），并引爆了汽车炸弹。

具体来说，AutoSlog 采用了句法分析器 CIRCUS [60] 识别手动注释的语料库中每个句子的词性 (POS)，例如主语，谓语，宾语，介词等。随后，它生成触发词词典和概念节点，可对其进行查看作为事件模式，涉及事件触发器，事件类型，事件自变量和自变量角色。在句子 2 中，“公共建筑”一词被 CIRCUS 识别为主题。由于其后跟随有被动动词“bombed”，因此它与预定义的语言模式“<subject>被动动词”相匹配。结果，它将“bombed”添加到触发词字典中，并生成带有轰炸类型的事件模式“<target>被轰炸”。

在模式匹配（事件提取）过程中，AutoSlog 首先使用触发词词典定位候选事件句子，然后通过 CIRCUS 对候选句子执行 POS。接下来，它将触发字周围的语法特征（即，使用 CIRCUS 的 POS 输出）与事件模式相关联，以提取自变量及其事件的作用。对于预定义的事件模式‘<victim>’具有绑架类型，其触发词为“took”

表 3. AutoSlog 中的语言模式和示例。方括号中的项目显示了找到字符串的语法成分（例如 <dobj>是直接对象，<np>是介词后的名词短语。在示例中，带括号的项目是可能与过滤器关联的广告位名称（例如，受试者是受害者）。带下划线的单词是触发器。

SN	Linguistic Pattern	Example
1	<subject> passive-verb	<victim> was <u>murdered</u>
2	<subject> active-verb	<perpetrator> was <u>bombed</u>
3	<subject> verb infinitive	<perpetrator> attempted to <u>kill</u>
4	<subject> auxiliary noun	<victim> was <u>victim</u>
5	passive-verb <dobj>	<u>killed</u> <victim>
6	active-verb <dobj>	<u>bombed</u> <target>
8	verb infinitive <dobj>	<u>to kill</u> <victim>
9	gerund <dobj>	<u>killing</u> <victim>
10	noun auxiliary <dobj>	<u>fatality</u> was <victim>
11	noun prep <np>	<u>bomb</u> against <target>
12	active-verb <np>	<u>killed</u> with <instrument>
13	passive-verb <np>	was <u>aimed at</u> <target>

会找到一个候选句子“他们花了 2 岁帕特里西奥·罗德里格斯 (Patricio Rodriguez) 的儿子吉尔伯托·莫拉斯科 (Gilberto Molasco)。使用 CIRCUS 进行 POS 的结果进一步确定“Gilberto Molasco”是 dobj (直接对象)。最后，结合 POS 结果和预定义的事件模式，AutoSlog 可以提取触发事件为“took”和受害者为“Gilberto Molasco”的绑架事件。

受 AutoSlog 系统的启发，针对不同应用领域开发了许多基于模式的事件提取系统，包括生物医学事件提取 [61]–[64]，财务事件提取 [65]，[66] 等。例如，Cohen 等。[61] 通过利用 OpenDMAP 语义解析器 [67] 通过生物医学本体分析提取生物医学事件，该分析器为生物医学概念及其属性提供了各种高质量的模板。卡西利亚斯等。[62] 应用了由 KYOTO 项目³ 开发的 Kybots (知识产生型机器人) 提取生物医学事件。Kybots 系统

³<http://www.kyoto-project.eu>

遵循基于规则的方法来手动构建事件模式。在金融领域, Borsje 等人。[65]提出了一种基于词汇语义模式的半自动财务事件提取方法。Arendarenko 和 Kakkonen [66]开发了一种基于本体的事件提取系统, 称为 BEECON, 用于从在线新闻中提取业务事件。由于构建事件模式通常需要专业知识, 因此这些基于模式的提取系统通常是特定的应用程序领域设计的。对于涉及多个域的应用, 曹等人。[68]提出通过将 ACE 语料库与其他专家定义的模式 (例如 TABARI corpus,⁴) 相结合来构建事件模式, 以产生更多的事件模式。

除了为英语事件提取设计的许多系统之外, 还针对其他语言设计了一些基于模式的提取系统[69], [70]。Tran 等。[69]通过将词汇语义规则与机器学习相结合, 从越南新闻中提取事件, 开发了一种称为 VnLoc 的实时提取框架。Saroj 等。[70]设计了一种基于规则的事件提取系统, 该系统从新闻专线和社交媒体文本中提取印度语言。Valenzuela-Escárcega 等。[71]观察到没有标准的语言来表达模式, 这可能会阻碍基于模式的事件提取任务的发展。他们提出了一个独立于域, 基于规则的框架, 并设计了用于事件提取的快速开发环境, 以降低新手的成本。

尽管基于模式的事件提取可以实现较高的提取精度, 但是模式构建存在可伸缩性问题, 因为模式通常取决于应用程序域。为了增加可扩展性, Kim 等人。[72]设计了 PALKA (并行自动语言知识获取) 系统, 以自动从带注释的语料库中获取事件模式。他们为事件模式定义了一种专门的表示形式, 称为 FP 结构 (框架短语和模式结构)。在 PALKA 中, 事件模式以 FP 结构的形式构造, 并通过语义约束的泛化来进一步调整。此外, FP 结构将单个事件自变量模式扩展为具有多个自变量的事件模式, 即一个事件可能包含多个自变量。例如, 句子 2 中的“爆炸”事件除了“目标”参数外, 还可能具有多个潜在的事件参数, 例如“代理”, “患者”, “工具”和“效果”。此外, PALKA 将一个句子中的多个子句转换为用于多事件提取的多句形式。此外, Aone 和 Ramos-Santacruz [73]开发了一个大型事件提取系统, 称为 REES, 该系统可以通过建立事件和关系本体模式以生成新的事件模式来提取多达 100 种事件类型, 仍需通过其系统的图形界面进行手动验证。

B. 自动图案制作

对于由具有专业知识的专家手动构建的事件模式, 由于它们具有高质量的良好定义, 因此基于模式匹配的事件提取通常可以针对特定领域的应用实现高精度。但是, 手动构造事件模式非常耗时且费力, 不仅导致生成大型模式数据库的可伸缩性问题, 还导致在其他领域中应用事件模式时的适应性问题。一些研究人员建议仅使用一些预先分类的训练语料库或种子模式, 即可应用弱监督方法或自举方法来自动获取更多模式。

Riloff 等。[74]开发了 AutoSlog-TS 事件提取系统来实现自动模式构建, 这是他们以前的 AutoSlog 系统的扩展。特别是, 它基于 AutoSlog 系统中定义的 13 种语言模式, 应用语法分析器 CIRCUS 从未标记的语料库中获取新的事件模式。以下示例说明了其新模式的构建过程。

句子 3: 世界贸易中心被恐怖分子炸毁。通过句法分析, 我们可以得到主题“世界贸易中心”, 动词短语“被炸”和介词短语“被恐怖分子”。以语言模式组合“<subject>被动-”和“被动动词准备<np>”, 我们可以获得潜在的事件模式: “<x>被轰炸”和“被<y>轰炸”>’。根据与域相关和与域无关的文档中的统计信息, 是否包括新事件模式取决于其评分。他们在 MUC-4 恐怖主义数据集⁵上进行的实验验证了 AutoSlog-TS 字典与手工制作的字典相比具有可比性。尽管仍然需要一些手动干预, 但 AutoSlog-TS 可以显著减少创建大型训练语料库的工作量。

已经提出了许多其他方法, 通过设计机器学习算法以基于一些种子模式来学习新模式来促进自动模式构建[6], [8], [31], [75]–[84]。例如, Yangarber 等人提出的 ExDisco。[75]提供了少数种子模式而不是语言模式来获得潜在事件模式。不仅种子模式, 而且种子术语或种子事件实例也已被用来构建潜在事件模式[78], [81]–[84]。由 Piskoriski 等人开发的 NEXUS 系统。[8], [31]通过基于熵最大化的机器学习算法从一小套带注释的语料库中学习了候选事件模式, 但仍手动检查并修改了候选模式以将其包括在模式数据库中。曹等。[79]提出了一种通过主动学习从外部语料库中导入频繁模式的模式技术。Li 等。[80]提出了一种从多个视图中提取中文事件的最小监督模型, 包括模式相似性视图 (PSV), 语义关系视图 (SRV) 和形态结构视图 (MSV)。此外, 每个

⁴<http://eventdata.parusanalytics.com/software.dir/tabari.html>

⁵https://www.nlpir.nist.gov/related_projects/muc/index.html

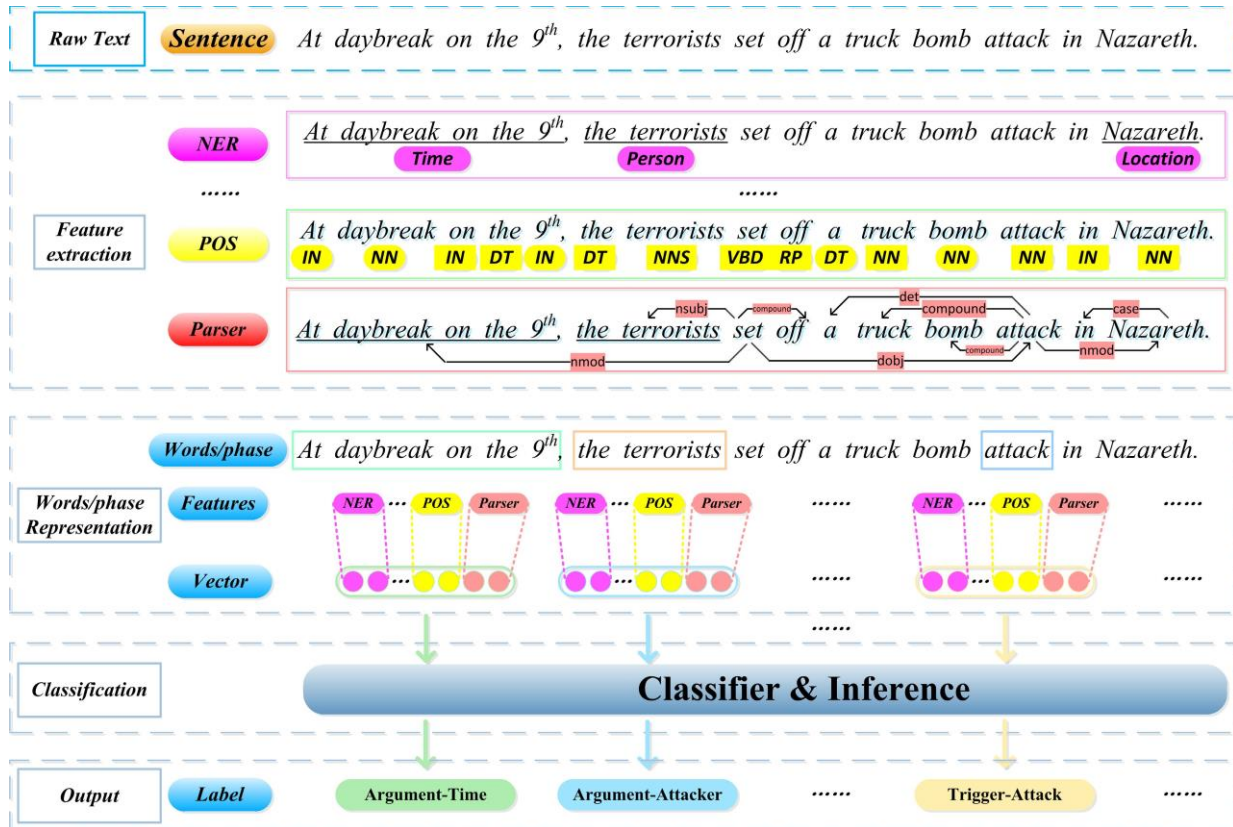


图 4. 基于机器学习的事件提取插图。单词/阶段特征是从执行特征工程获得的，然后输入到分类器以进行事件提取以输出触发器和参数。

视图还可以用于提取事件模式。PSV 根据其 与现有结构的相似性对每个候选模式进行排名，然后接受那些排名最高的新闻模式。相反，SRV 从相关文档中捕获相关事件提示，而 MSV 被合并以推断新模式。

基于模式的事件提取由于使用领域专家建立的高质量事件模式而具有很高的提取精度，因此已在许多工业应用中得到了应用。但是，构建大规模事件模式的成本很高。因此，近年来见证了各种基于机器学习的事件提取技术的快速发展。

V. 基于机器学习的事件提取本节介绍了使用传统的机器学习算法（例如支持向量机（SVM），最大熵（ME）等）进行事件提取的方法。下一节将对那些使用神经网络技术（或所谓的深度学习）的人进行回顾。请注意，它们都是一种有监督的学习技术，并且需要带有地面真相标签的训练数据，通常由具有专业知识的专家来完成。第三节介绍了如何注释事件及其论点。

机器学习方法的基本思想几乎是相同的，即从训练数据中学习分类器，并将分类器应用于从新文本中提取事件。

图 4 展示了基于机器学习的事件提取的总体结构。此外，基于机器学习的事件提取通常可以分为两个阶段和四个子任务：

- 第一阶段包括两个子任务：（1）触发检测，即检测事件是否存在，如果存在，则在文本中触发相应的事件；（2）触发/事件类型标识，即将触发/事件分类为给定事件类型之一；
- 第二阶段包括两个子任务：（3）参数检测，即，检测哪个实体，时间和值是参数；（4）自变量角色识别，即根据识别出的事件类型对自变量的角色进行分类。

可以以流水线方式执行这两个阶段和四个子任务，其中分别训练多个独立的分类器，并且一个分类器的输出也可以用作其后续分类器的输入的一部分。四个子任务也可以联合执行，其中一个分类器执行多任务分类并直接输出多个子任务的结果。

对于管道执行或联合执行，学习分类人员首先需要执行一些要素工程工作，即从文本中提取要素作为分类模型的输入。因此，在本节中，我们首先介绍用于分类器训练的一些常用功能；然后我们回顾

并比较文献中的管道和联合分类方法。

A. 学习模型的文本功能

常用的文本特征通常可以分为三种：词汇，句法和语义特征。它们中的大多数可以通过某些开源 NLP 工具获得。

一些常用的词汇功能包括：（1）完整词，小写词和接近词；（2）词形化词，将单个词的不同形式返回其根形式。例如，计算机一词是计算机的一种变形形式；

（3）POS 标签，用于将语料库中的单词标记为对应于特定词性，例如名词，动词，形容词，副词等。

句法特征是从依存关系解析中获得的，这是要找出句子的词法依存关系结构。简而言之，它在句子中表示不同关系类型的单词之间创建边缘，并将它们组织为树状结构。一些常用的语法特征包括：（1）依赖路径的标签；

（2）依赖词及其词法特征；（3）依赖树中候选词的深度。

一些常用的语义特征包括：（1）语言词典中的同义词及其词法特征；

（2）事件和实体类型特征，通常用于参数识别中。

每个特征可以表示为基于袋字模型的二进制向量。特征工程是关于如何选择最重要的特征以及如何将它们集成到高维向量中以表示每个单词的。最后，文本特征及其来自训练数据集的相应标签将用于训练事件提取分类器。

B. 管道分类模型

管道分类通常训练一组独立的分类器，每个分类器用于一个子任务。然而，一个分类器的输出也可以用作其后续分类器的输入的一部分。为了训练分类器，可以使用各种功能，包括句子中的局部特征和文档中的全局特征。根据局部和全局特征如何应用于分类器训练，我们将流水线分类方法分为句子级事件提取和文档级事件提取。

1) 句子级事件提取

在句子级事件提取中，首先将一个句子代币化为离散的标记，然后基于特征工程的结果用特征向量表示每个标记。在英语中，记号通常是一个单词；但是在某些语言中，令牌可以是一个字符或多个字符，如单词或短语。从语料库的带注释文本中训练分类器，然后将其用于确定标记是否是触发词（或事件自变量）及其事件类型（或自变量角色）。

David Ahn [27]提出了一个典型的管道处理框架，该框架由两个连续的分类器组成：第一个称为 TiMBL，它应用最近邻居学习算法来检测触发器；第二个分类器是 TiMBL。第二个分类器称为 MegaM，它采用最大熵学习器来识别参数。为了训练 TiMBL 分类器，使用了词汇功能，WordNet 功能，上下文功能，依赖功能和相关的实体功能；触发词，事件类型，实体提及，实体类型以及触发词与实体之间的依赖路径的特征用于训练 MegaM 分类器。已经提出了许多不同的管道分类器，并且它们通过各种类型的特征来训练。例如，Chieu 和 Ng [85]添加了 unigram, bigram 等作为特征，并采用了最大熵分类器。在生物医学领域，更多的特定领域特征和专业领域知识被用于分类训练，例如频率特征，令牌特征，路径特征等[86]–[93]。

一些研究人员建议将模式匹配集成到机器学习框架中[94]–[96]。如上一节所述，事件模式可以通过某些特定触发器及其关联参数之间的显式关系提供更精确的事件结构，尽管这种关系是手动设计的，没有太多可扩展性。Grishman 等。[94]，[97]提议首先执行模式匹配，以便预先分配一些潜在的事件类型⁶，然后应用分类器来识别剩余的事件提及。两种方法相互补充以增强事件提取。此外，在机器学习分类器检测到触发器后也可以执行模式匹配[95]。由于事件自变量与触发器类型密切相关，因此应用一些完善的触发器-自变量关系可以帮助改善自变量识别。此外，事件模式还可以与令牌特征合并，作为机器学习分类的输入[96]。

如图 2 所示，在某些语言（例如英语）中，句子标记化是一项琐碎的任务，因为单词由分隔符明确分隔。但是，在其他一些语言中，例如中文[98]–[100]和日语[101]，句子由连续字符组成，而无需使用分隔符来分隔单词。因此，通常需要单词分割，首先将一个句子分成许多离散的标记（单词/阶段）。由于可以独立于事件提取任务来实现单词分割，因此细分错误（如果有的话）可以传播到下游任务并降低其性能[99]，[100]。此外，由于自然语言的歧义性，单词分割后，一个单词可能由两个或多个触发器组成。触发器分为两个或多个单词[99]，[100]。

为了解决这种标记化问题，Zhao 等人。[98]提出，对于句子分割后的每个单词，首先要获得潜在的触发条件，每个触发条件都有其相应的事件

⁶<https://nlp.cs.nyu.edu/meyers/GLARF.html>

从同义词字典中输入。Chen 和 Ji [99] 提议同时使用单词级和字符级的 trigger 标记策略。此外，他们提出使用字符级特征，包括当前字符，其前一个和下一个字符等，来训练用于触发检测的最大熵马尔可夫模型分类器。Li 等。[100] 采用了构成语义的内部触发和触发提及之间的话语一致性来增强中文触发识别。具体来说，如果动词的组成部分（一个或多个汉字）被标记为中文触发器，则它是触发器。为了保证语篇的一致性，将动词的单个字符（如果属于带标签的触发器）与其上一个和下一个字符合并，以形成候选触发器。

2) 文档级事件提取

在前述方法中，仅将句子和句子级上下文中的单词或短语的本地信息应用于训练分类器。但是，如果我们放事件提取任务，即使是仅从一个句子中提取事件，在更大的背景下（例如包含多个句子的文档或包含多个文档的集合），也可以利用许多全局信息来增强提取准确性。对于文档级事件提取，两个关键设计问题包括：可以使用哪种全局信息；以及如何应用它们来辅助事件提取。对于第一个设计问题，可以从跨文档，跨语句，跨事件和跨实体推断中挖掘诸如词义，实体类型和论据提及之类的全局信息。第二，全局信息既可以用作本地分类器的补充模块，也可以用作本地分类器中的全局功能。

可以通过评估局部分类器输出的置信度来利用全局信息来构建附加的推理模型，以增强局部分类器 [102]–[105]。Ji 和 Grishman [102] 观察到事件自变量将在句子和文档之间保持某种一致性，例如不同句子和相关文档中的词义一致性，以及相同或相关事件的不同提及之间自变量和角色的一致性。为了利用这种观察，他们提议为一组局部相关文档建立两个全局推理规则，即每个簇一个触发意义和每个簇一个论点角色，以帮助提高句子级分类器的提取置信度。廖和格里什曼 [103] 后来扩展了全局信息推断的这种应用，并进一步提出了应用文档级跨事件信息的建议。刘等。[105] 探索了两种类型的全局信息，即事件-事件关联和主题-事件关联，以建立概率软逻辑 (PSL) 模型，以进一步处理来自本地分类器的初始判断以生成最终的提取结果。

全局信息还可以与语句级功能一起用作全局功能，以训练局部事件提取分类器 [106]–[108]。Hong 等。[106] 提出了一个跨实体推理模型来提取实体类型

一致性作为关系特征。他们认为，一致类型的实体通常以相同的角色参与类似的事件。为此，他们定义了 9 种新的全局功能，包括实体子类型，领域中的实体子类型共现，参数的实体子类型等，以训练一组句子级 SVM 分类器。在 [107]，[108] 中也采用了类似的方法。廖和格里什曼 [107] 提出首先通过潜在的狄利克雷分配 (LDA) 模型 [109] 从文档集合中计算每个文档的主题分布，并将这些主题特征编码为局部提取分类器。在 [108] 中，定义了三种类型的全局特征，包括词汇桥特征，话语桥特征和角色填充物分布特征，它们与局部特征一起用于训练提取分类器。

C. 联合分类模型

前述流水线分类模型可能会遇到错误传播问题，其中上游分类器中的错误很容易传播到那些下游分类器，并可能降低其性能。另一方面，下游分类器不会影响其先前的分类器的决策，并且不能很好地利用不同子任务的相互依赖性。为了解决此类流水线问题，已提出联合分类作为一种有前途的替代方案，可以从两个或多个子任务之间的紧密交互中获得可能的好处，因为一个子任务的有用信息既可以传递给下一任务，也可以传递给先前的任务。

在每个事件提取阶段，可以针对多个子任务训练联合分类模型 [110]–[112]。例如，Li 等。[110] 提出了一种用于触发检测和事件类型识别的联合模型，其中基于整数逻辑编程 (ILP) 的推理框架用于集成两种类型的分类器。特别是，他们提出了两种触发检测模型：一种基于条件随机场 (CRF)，另一种基于最大熵 (ME)。基于 ILP 的框架可以帮助找到约束变量的最佳值，以最小化加权目标函数。因此，可以在 ILP 推理框架内联合训练多个子任务，但每个子任务都有不同的约束和权重。Li 等人也采用了类似的方法。[111] 训练话语级论证确定和角色识别的联合模型。Chen 和 Ng [112] 训练了两个基于 SVM 的联合分类器，每个分类器可完成一个阶段的子任务，但具有更多的语言功能。

还可以训练联合分类模型，以根据预定义的事件结构同时提取事件的触发器和相应的自变量 [113]，[114]。例如，Li 等。[113] 将事件提取公式化为结构化学习问题，并提出了一种将局部和全局特征集成到结构化感知器模型中的联合提取算法 [115]，以同时预测事件触发器和参数。特别地，整个句子的结果可以被认为是一

图, 其中触发器或参数表示为节点, 参数角色表示为从触发器到其参数的键入边。他们应用波束搜索对图形执行不精确的解码, 以捕获触发器和参数之间的依赖关系。Judea 和 Strube [114]观察到事件提取在结构上与帧语义解析相同, 后者是从文本中提取语义谓词-自变量结构。因此, 他们优化并重新训练了 SEMAFOR [116]⁷ 一种框架语义解析系统, 用于结构事件提取。

结构化的预测方法也已广泛用于生物医学领域的联合提取模型[117]–[127]。例如, Riedel 等。[117], [118]将事件表示为句子的相关结构化标记, 并将基于马尔可夫逻辑的联合概率模型应用于生物医学事件提取。Venugopal 等。[119]将马尔可夫逻辑网络 (MLN) 和 SVM 结合起来作为联合提取模型, 他们利用 SVM 分类器来处理高维特征, 并通过 MLN 对关系进行建模。Vlachos 等。[122]采用了基于搜索的结构化预测框架来提供较高的建模灵活性。McClosky 等。[124]在重新排序的依赖分析器中利用事件参数的树结构来捕获全局事件结构属性。

除了事件提取之外, 联合分类模型还可以与其他 NLP 任务一起训练, 例如命名实体识别, 事件共参考解析, 事件关系提取等。一些研究人员建议训练单个模型来联合执行这些任务。任务[128] – [132]。例如, Li 等。[128]提出了一个框架, 以在单个模型中联合执行这些任务以及事件提取。特别是, 他们使用信息网络将实体、关系和事件表示为信息网络表示, 该信息网络表示基于结构化预测的单个模型来提取所有实体、关系和事件。前述的联合提取方法仅在句子级别操作, 但可能会丢失文档级别的有价值的信息。在 Li 的模型[128]的基础上, Judea 和 Strube [129]提出了一个全局推理模型, 将全局和文档上下文合并到句内基本系统中, 以共同提取实体的提及, 事件和关系。此外, 荒木和三村[131]发现事件及其共同引用为这两项任务提供了有用的语义和话语信息。他们提出了一个文档级联合模型来捕获事件触发器和事件共同引用之间的交互, 从而同时提高事件触发器识别和事件共同引用解析的性能。

VI. 基于深度学习的事件提取特征工程是基于机器学习的事件提取的主要难题。如前一节所述, 尽管词汇, 句法,

语义特征等可以作为分类器的输入, 它们的构造需要语言知识和领域专业知识, 这可能会限制训练后的分类模型的适用性和适应性。此外, 这些功能通常都具有“一站式”表示, 这不仅遭受数据稀疏性问题的困扰, 而且还简化了用于分类模型训练的特征选择。

最近, 对于各种分类任务, 已经深入研究了深度学习技术, 该技术使用多层连接的人工神经元来构建人工神经网络。在人工神经网络中, 最低层可以以非常简单的表示形式获取原始数据作为输入。每一层都可以学习将其较低层的输入转换为更抽象和复合的表示形式, 然后将其输入到其自己的较高层中, 直到将其输出特征用于分类的最高层为止。与传统的机器学习技术相比, 深度学习可以大大减少特征工程的难度。

深度学习已成功应用于各种 NLP 任务, 例如命名实体识别 (NER) [133], 搜索查询检索和问题解答[134], [135], 句子分类[136], [137], 名称标记和语义角色标签[138], 关系提取[139], [140]。对于事件提取, 最近还提出了许多深度学习方案[141]–[146]。一般过程是构建一个神经网络, 该神经网络将词嵌入作为输入并输出每个词的分类结果, 即, 对一个词是否是事件触发器 (或事件自变量) 进行分类, 如果是, 则对其事件类型 (或自变量角色)。

如何设计有效的神经网络架构是基于深度学习的事件提取的主要难题。我们将在本节的其余部分中回顾一些典型的神经网络。在此之前, 我们简要介绍单词嵌入技术, 因为大多数用于事件提取的神经网络共享一种使用单词嵌入作为原始数据输入的通用方法。单词嵌入技术用于将词汇表中的单词或短语转换为低维和实值向量。在实践中, 从大规模语料库训练单词的向量表示, 并且提出了各种单词嵌入模型, 例如连续词袋模型 (CBOW) 和连续跳过语法模型 (SKIP-GRAM) [147], [148]。

A. 卷积神经网络

卷积神经网络 (CNN) 是最常用的神经网络结构之一, 它由多层完全连接的神经元组成。也就是说, 下层的每个神经元都连接到上层的所有神经元。由于 CNN 能够基于连续和广义的词嵌入来学习文本隐藏特征, 因此已被证明可以有效地捕获句子的句法和语义[136]。

Nguyen 和 Grishman [141]可能是设计用于事件检测的 CNN 的第一批研究人员, 即。识别

⁷<http://www.cs.cmu.edu/~方舟/> SEMAFOR

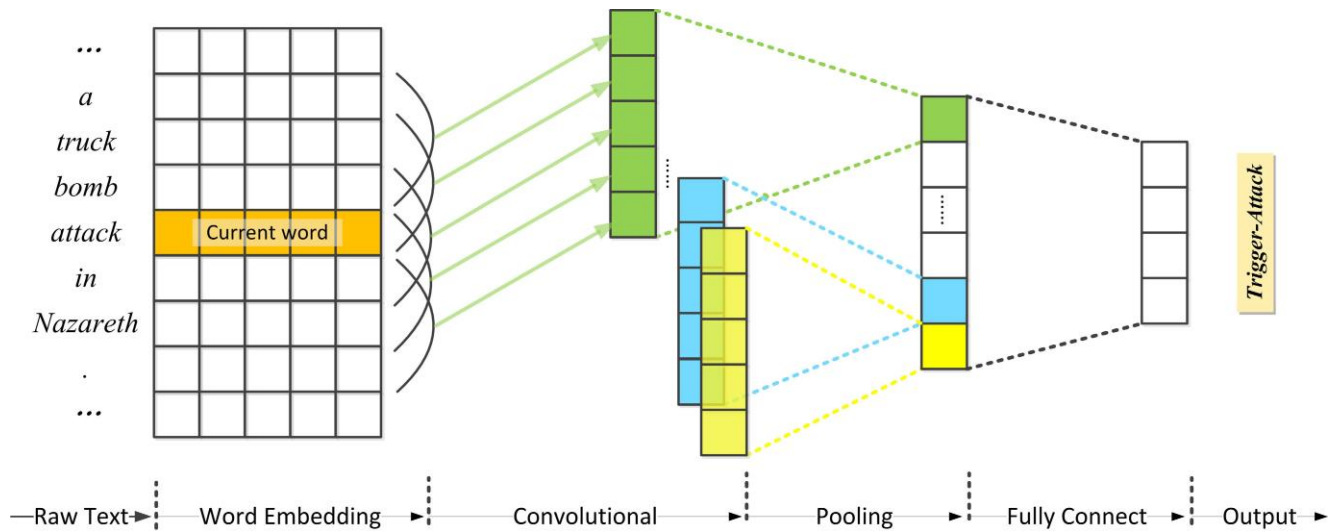


图 5. 使用卷积神经网络进行事件提取的图示。

触发器及其在句子中的事件类型。如图 5 所示，首先将每个单词转换为实值向量表示，它是单词嵌入，其位置嵌入和实体类型嵌入的串联，作为网络输入。CNN 由输入到输出包括卷积层，最大池化层和 softmax 层，并输出每个单词的分类结果。许多典型技术可用于训练 CNN，包括反向传播梯度，辍学规律化，随机梯度下降，随机小批处理，AdaDelta 学习速率自适应和权重优化。

典型的 CNN 结构通常使用最大池层，该层在整个句子的表示上具有最大操作，但是，一个句子可能包含多个事件共享自变量，但角色不同。Chen 等。[149]提出了一个动态多池卷积神经网络（DMCNN），通过一个动态多池层来提取一个词法层次和句子层次特征来评估句子的每个部分。在 DMCNN 中，根据预测的触发将每个特征图分为三部分，并且保留每个部分的最大值以保留更多有价值的信息，而不是使用单个最大池值。此外，他们的 CNN 模型还使用跳跃语法单词模型来捕获单词有意义的语义规律。

在[141]和[149]中，卷积运算将早期的 k-gram 的向量映射到特征空间，在这里，这种 k-gram 的向量作为 k 个连续词表示的串联而获得。为了也利用远程和非连续的依赖关系，Nguyen 和 Grishman [150]提出对所有可能的非连续 k-gram 进行感知操作中的卷积运算，其中最大池函数用于计算卷积分数，用于区分事件检测中最重要的非连续 k-gram。

还提出了对典型 CNN 模型的其他一些改进[151]-[154]。例如，

Burel 等。[152]设计了一种语义增强的深度学习模型，称为 Dual-CNN，该模型在典型的 CNN 中添加了语义层以捕获上下文信息。Li 等。[153]提出了一个并行多池卷积神经网络（PMCNN），它可以捕获句子的组成语义特征以用于生物医学事件提取。PMCNN 还利用基于依存的词义和句法表示形式进行嵌入，并采用整流线性单元作为非线性函数。Kodelja 等。[154]按照自举方法建立了全局上下文的表示，并将表示形式集成到 CNN 模型中以进行事件提取。

B. 递归神经网络

许多基于 CNN 的事件提取方案都因流水线执行两个子任务（即事件检测优先和论据识别第二）而遭受错误传播问题。此外，CNN 结构通常将单词嵌入的级联作为输入，并对连续的单词执行卷积运算以捕获当前单词与其相邻单词的上下文关系。因此，它们无法很好地捕捉到远处单词之间可能存在的某些相互依存关系，从而无法将一个句子作为一个整体来共同提取触发器和论点。如前一节所述，联合提取可以利用事件触发器和事件参数之间的关系来往复执行这两个子任务。

在语言建模中，一个句子通常被视为一个单词序列，即从一个句子的开始到结尾一个接一个的单词。由一系列连接的神经元组成的递归神经网络（RNN）结构可以有效地利用此类顺序输入。如图 6 所示，一个简单的 RNN 由一系列连接的长期和短期记忆（LSTM）神经元组成，其中 LSTM 神经元的输出也用作其顺序连接的 LSTM 神经元的输入。因此，RNN 结构

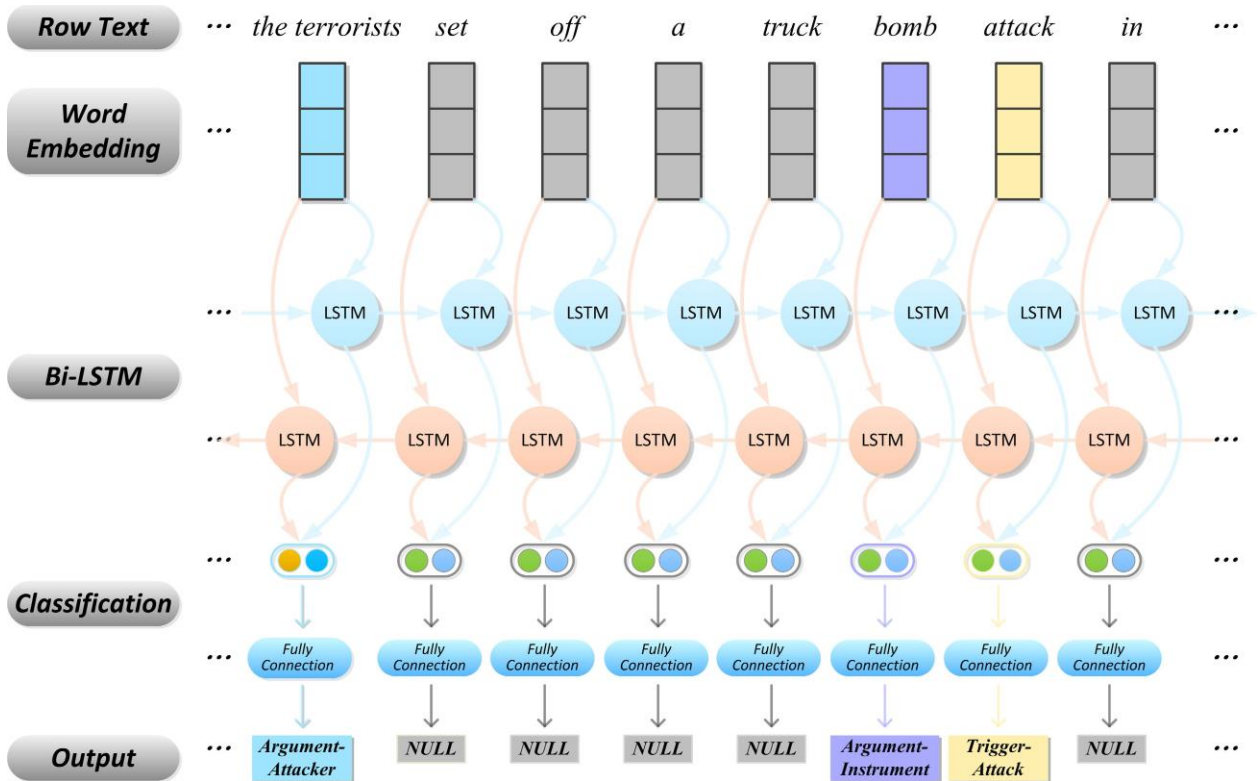


图 6. 使用递归神经网络进行事件提取的图示。

可以利用直接或间接连接的任意两个词之间的潜在依赖关系，从而使其在许多 NLP 任务中得以广泛应用[155]，包括命名实体识别[156]，词性标记[157]，关系提取[158]，句子解析[159]，序列标签[160]等。此外，RNN 结构还输出一个单词序列，每个单词都可以预测为触发或自变量，从而实现了联合触发检测和检测两个任务。参数识别可以共同执行。

对于事件提取，已经提出了一些基于 RNN 的模型，通过根据单词在句子中的向前或向后顺序输入单词来利用单词的相互依赖性[161]–[164]。例如，Nguyen 等。[161]设计了一种用于联合事件提取的双向 RNN 架构，如图 6 所示。它们的模型由两个单独的 RNN 在正反两个方向上运行，每个 RNN 都由一系列门控循环单元（GRU）组成[165]。联合提取包括两个阶段：编码阶段和预测阶段。在编码阶段，与 CNN 模型不同，它不采用位置特征，而是将其替换为二进制矢量来表示联合特征，以共同预测事件触发器和参数。在预测阶段，模型将触发器和参数之间的依赖关系分为三类：（1）触发器子类型之间的依赖关系；（2）自变量角色之间的依赖关系；（3）触发器子类型和参数角色之间的依赖关系。

单词之间的句法依赖性也可以用来增强基本的 RNN 结构。例如，Sha 等。[166]通过将两个 RNN 神经元的语法相关连接添加到双向 RNN 中，设计了 dbRNN（依赖关系桥 RNN）。如图 7 所示，通过将依赖关系桥包括到双向 RNN 中，可以利用句子 1 的语法依赖关系。

除了使用依赖关系桥外，还可以直接利用句子的句法依赖树来构建树结构的 RNN [167]。根据典型的 Bi-LSTM（双向 LSTM），Zhang 等人。[168]通过转换句法相关性分析器的原始相关性树以进行中文事件检测，进一步构建了一个以目标词为中心的 Tree-LSTM。Li 等。[169]提出用外部实体本体知识进一步扩展 Tree-LSTM，以用于生物医学事件提取。

RNN 结构不仅可以应用于句子级别，还可以应用于文档级别的事件提取。段等。设计了一个 DLRNN 模型（文档级 RNN）[170]，该方法通过使用用于文档表示的分布式矢量来提取交叉句子甚至跨文档线索，即通过无监督学习的 PV-DM 模型捕获文档的主题分布[171]。文档中的所有单词都使用相同的文档向量，单词嵌入和文档向量的串联被用作 Bi-LSTM 模型的输入。

RNN 结构还可用于训练联合分类模型，不仅用于事件提取（包括触发检测和参数识别），而且还用于实体

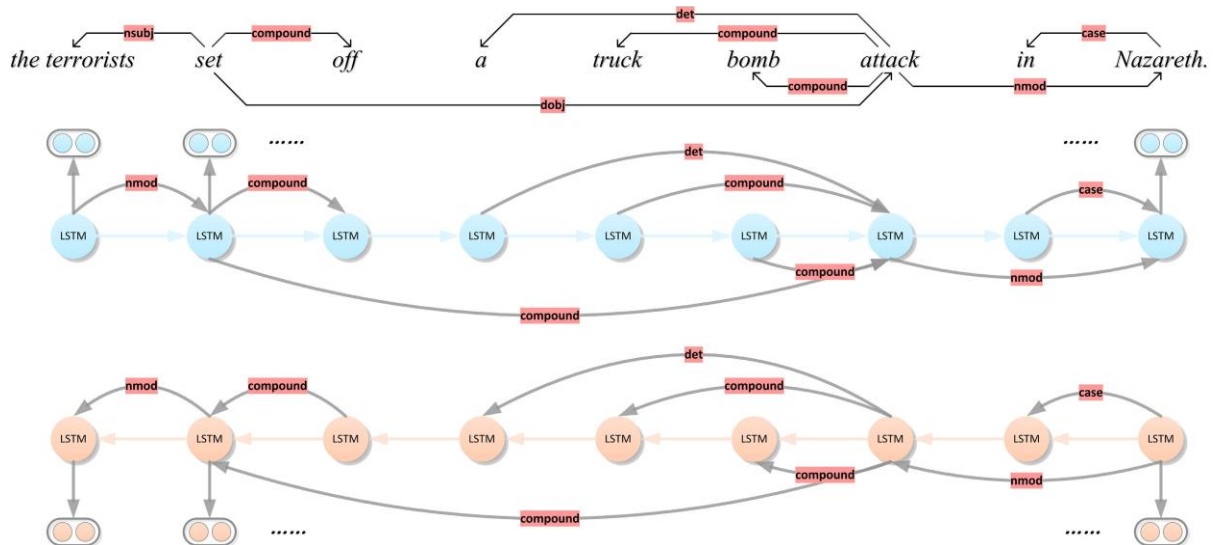


图 7. LSTM 上的依赖关系桥。除了最后一个 LSTM 单元，每个单元还从以前的语法相关单元接收信息。

提及检测[172]。通过训练单个 Bi-GRU (双向 GRU) 网络模型，以流水线方式按照实体提及检测器，触发器分类器和自变量角色分类器的顺序执行这三个子任务，其中所有三个网络共享了网络隐藏表示子任务以利用子任务之间的一些常识以及子任务之间的潜在依存关系或交互。

前述的 RNN 结构已采用 GRU 或 LSTM 作为基本组成单元，它使用门控策略来控制神经网络中的信息处理。但是，门运算不能并行进行，并且很耗时。张等。[173] 提出使用一种简单的递归单元 (SRU) 作为基本的组成神经元，因为它能够降低门运算的复杂度而不会产生依赖于先前单元的乘法运算[174]。他们建立了两个双向 SRU 模型 (Bi-SRU)：一个用于学习单词级表示形式，另一个用于字符级表示形式。

C. 图神经网络

近年来，许多在图上运行的神经网络，或称为图神经网络 (GNN) 作为通用参考，最近已被提倡广泛的应用领域[175]–[177]。简而言之，GNN 应用在图结构上运行的多个神经元来实现非欧几里得空间中的所谓几何深度学习。传统的神经元操作 (例如已在 RNN 和 CNN 中广泛使用的递归内核和卷积内核) 也可以应用于图结构中，以学习嵌入在图中的各种功能。

一些研究人员试图将 GNN 模型用于事件提取[178]–[180]。这样的核心问题

方法是首先为文本中的单词构建图形。Rao 等。[178] 采用了一种语义分析技术，称为抽象意义表示 (AMR) [181]，该技术可以规范文本中的许多词汇和句法变体，并输出有向无环图以捕获“谁对谁做了什么”的概念。在文本中。此外，饶等。认为事件结构是 AMR 图的子图，并将事件提取任务转换为子图识别问题。他们训练了图 LSTM 模型，以识别用于生物医学事件提取的事件子图。

图构建的另一种方法是基于对句法依存树的某种转换[179]，[180]。注意，句法依存关系可以被视为从首词到其依赖词的有向边，但也可以将边沿标签作为依存关系类型。首先，为每个单词创建一个新的[182]中提出的自环边，并将其包含在依存关系树中，即从单词开始到该单词结束的边。所有此类自环边均具有相同的新边类型。然后对于边类型为 T 的每个依赖边 (w_i, w_j) 单词 w_i 到单词 w_j ，类型 T 的反边 (w_j, w_i) 是创建并包含到依存关系树中。图的构造如图 8 所示。基于所构造的图，进行多跳图卷积运算，不仅可以从局部相邻词中获取上下文特征，还可以从与远程相关的词中捕获上下文特征，以生成新的表示向量每个字。

D. 混合神经网络模型

当用于捕获文本中的各种特征，关系和依存关系以进行事件提取时，上述三种基本类型的神经网络体系结构各有其优缺点。许多研究人员提出了混合神经网络模型，该模型结合了不同的神经网络来享受每个卓越的功能。常用方法

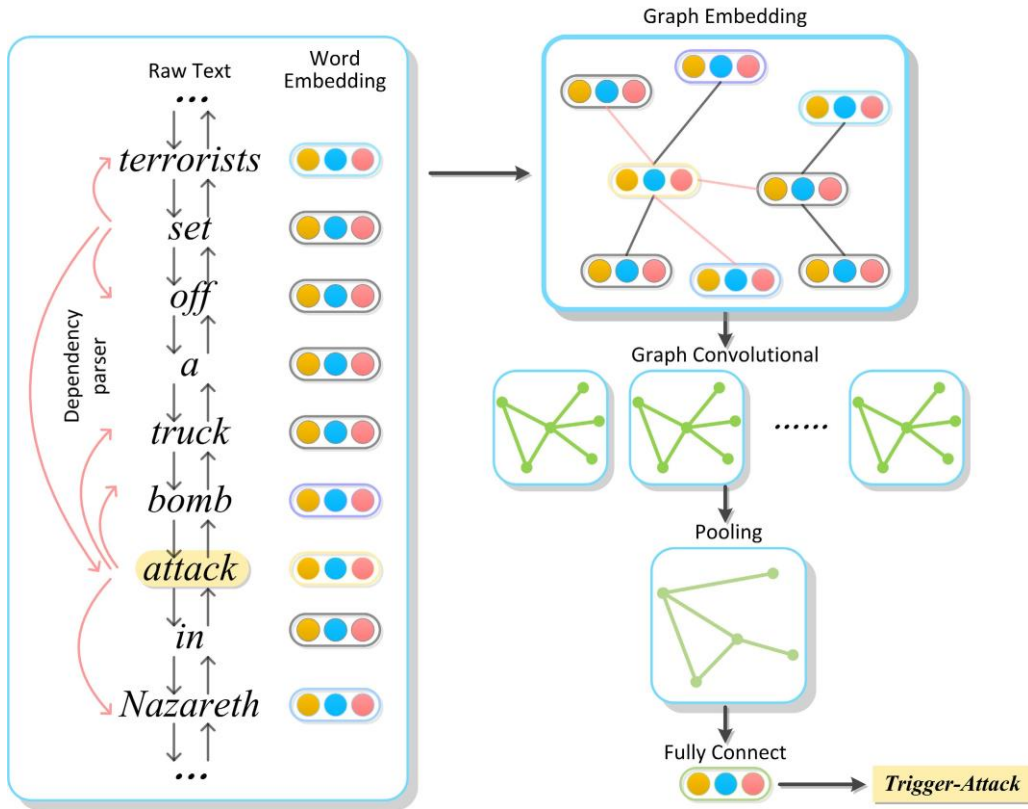


图 8. 使用图卷积神经网络进行事件提取的图示。

建立这种混合模型的方法是使用不同的神经网络来学习不同类型的单词表示。例如，[179]和[180]在执行图卷积运算之前，都首先应用了 Bi-LSTM 模型来获取初始单词表示。

有人提出将 CNN 和 RNN 结合起来作为混合模型[183]–[189]。例如，Zeng 等。[183]提出首先使用 CNN 来学习每个单词的局部词表表示，然后将其与另一个 Bi-LSTM 的输出连接起来以获得最终的词表进行分类。连接 CNN 输出和 Bi-LSTM 隐藏层的类似方法也已应用于中文，印度和其他语言的事件提取[184]–[187]。Nurdin 和 Manlivedevi [188]还设计了一个由 CNN 和 Bi-LSTM 组成的混合模型，用于从印尼新闻中提取事件，其参数如下：谁，做什么，何时，何地，为什么以及如何做。刘等。[189]还使用了 CNN 来获取每个单词的本地上下文表示；然而，他们建议使用 Bi-LSTM 来获取文档表示形式，作为前向和后向层级联隐藏状态的加权总和。然后，将单词本地表示和文档表示进一步串联在一起，以进行事件触发检测。

混合模型的另一个重要类型是生成对抗网络 (GAN) [190]，通常由

的两个神经网络相互竞争，一个被称为生成器 Gnn，另一个被称为鉴别器 Dnn。Gnn 产生候选结果；但是 Dnn 根据训练数据的真实性对它们进行了评估。Gnn 的训练过程引入了一些由训练数据集制造的嘈杂数据；虽然 Dnn 努力将与实际数据分布的差异最小化。生成器和鉴别器通常使用相同的神经网络结构，但是许多 GAN 模型提出了不同的训练策略。

一些研究人员建议将 GAN 框架应用于事件提取[191]–[193]。他们将 Bi-LSTM 之类的 RNN 结构用于生成器和判别器网络。在培训过程中，Hong 等人。[191]提出用两通道的自我调节学习策略来调节学习过程。在自我调节过程中，对发生器进行训练以产生最虚假的特征。而带有记忆抑制器的鉴别器经过训练可以消除假货。刘等。[192]提出了一种对抗模仿策略，将知识传播模块纳入特征编码过程。在他们的混合模型中，使用了分别为 Bi-GRU 网络的教师编码器和学生编码器。教师编码器通过金色注解进行训练，而学生编码器通过对抗式模仿学习，通过使其输出与教师编码器的输出之间的距离最小化来进行训练。张等。[193]使用强化学习 (RL)

在训练过程中更新 Q 表的策略, 其中 Q 表记录根据系统状态和操作计算的奖励值。

E. 注意机制

注意机制最早出现在计算机视觉领域, 其目的是模拟人脑的视觉注意机制。最近, 注意力机制已在许多 NLP 任务中广泛使用[194]–[196]。简而言之, 注意力是一种区分机制, 可根据其对给定任务的重要性, 引导神经模型对输入的每个成分进行不平等的对待。尽管通过为不同的神经元状态或输出分配不同的权重来实现, 但是权重实际上是从模型训练过程中自学的。

已经提出了许多单词级注意机制来学习句子中每个单词的重要性, 例如区分不同的论点单词, 单词类型, 单词关系[197]–[202]。它们的主要区别在于应更多关注哪些元素以及如何训练注意力向量。例如, 刘等。[197]认为, 触发因素的论点应该比其他词受到更多的关注。为此, 他们首先构造了金注意力向量, 以便仅对每个带注释的触发器编码带注释的自变量词及其上下文词。此外, 他们为每个单词设计了两个上下文关注向量: 一个是基于上下文单词; 另一个是基于上下文单词。另一个基于在转换后的实体类型空间中编码的上下文实体类型。然后, 将两个注意力向量与事件检测器连接在一起并进行训练, 以最小化事件检测和注意力差异的加权损失。Wu 等。[198]应用论点信息来训练 Bi-LSTM 网络的注意力。

在单词级注意机制中, 来自句法解析器的实体关系也可以用于训练注意力[199], [200]。基本思想是句法相关性可以在两个可能不连续但相距很远的词之间提供联系。依存关系类型可以帮助区分单词之间的句法重要性。对于中文事件提取, 因为没有像英语, Wu 等人这样的显式单词分割。[202]也提出了一种字符级注意机制来区分汉字中每个字符的重要性。

一些研究者还建议将单词和句子的注意力整合在一起, 以增强多句文档中事件的提取[203]–[205]。赵等。[203]认为, 在许多多句文档中, 一个文档中的句子通常与文档主题相关, 尽管它们可能包含不同类型的事件。他们提出了具有分层和监督注意机制的 DEEB-RNN 模型, 该模型将单词级别的注意力放在事件触发上, 将句子级别的注意力放在包含事件的句子上。为此, 他们构建了两个黄金注意力向量: 一个是基于句子触发的词级注意力; 另一个是基于单词触发的词级注意力。如果包含, 则另一个用于每个句子的句子级别关注

触发词。除了一个文档中句子之间的相互依赖性之外, 单个事件中可能还嵌入了多个事件。Chen 等。[204]认为同一句子中提到的事件在语义上往往是连贯的。为了捕获句子内相关性和句子间相关性, 他们提出了一种具有门控多级注意机制的 HBTNGMA 模型, 用于提取和融合句子内和句子间上下文信息以增强事件检测。

除了利用单词和句子之外, 一些研究人员还建议整合额外的知识以引起注意, 例如使用多语言知识[206]或先验触发语料库[207]。大多数事件提取模型都针对特定的语言进行了培训和应用, 这可能会遇到单个单词在不同上下文中具有不同含义的典型歧义问题。刘等。[206]在 ACE 2005 中检查了带注释的事件, 并观察到 57% 的触发词不明确。他们认为, 基于对多语言一致性和多语言互补的观察, 使用多语言方法可以帮助解决歧义问题。因此, 他们提出了一个封闭的多语言关注框架, 该框架既包含单语言上下文关注又包含封闭的跨语言关注。对于后者, 他们首先应用了机器翻译器来获取另一种语言的翻译文本。Li 等。[207]设计了一个先验知识集成网络, 将其收集的关键字编码为先验知识表示。然后将此类知识表示与自我注意网络结构集成在一起。

VII. 基于半监督学习的事件提取

基于机器学习和深度学习技术的上述算法是一种有监督的学习方法, 需要一种标记语料库进行模型训练。由于深度学习方法通常在神经网络中涉及大量参数, 因此标记语料库通常越大, 可以训练的模型越好。但是, 由于其耗时且费力的注释过程, 获得带标签的语料库是一项成本高昂的任务, 在大多数情况下, 这还需要领域专业知识和专业知识。因此, 许多标记语料库的规模较小且覆盖率较低。例如, 在 ACE 2005 语料库中, 仅定义了 33 种感兴趣的事件类型, 对于各种应用程序的覆盖率非常低; 此外, 在标记的事件中, 约 60% 的事件类型的实例少于 100 个, 而 3 个事件类型的实例甚至少于 10 个。

如何提高从少量标记的金数据中提取的准确性已经成为一个严峻的挑战。一个直接的解决方案是首先自动生成更多的训练数据, 然后使用包含原始黄金数据和新生成的数据的混合数据进行模型训练。在本节中, 我们将回顾文献中的此类解决方案, 尽管它们可能使用了不同的名称, 例如半监督, 弱监督, 远程监督等。

尽管使用混合数据会影响模型训练过程,但是这些学习分类算法的基础与前两节中介绍的相似。在本节中,我们主要集中于如何将少量标记数据扩展到较大的语料库,以及如何从混合数据中训练提取模型。

A. 联合数据扩展和模型训练

标记的金数据集可能很小,但是可以迭代地用于模型训练。通过数据替换进行的这种迭代模型训练是一种称为自举[208]的已知技术的一种变体。基本思想是首先训练带有少量标记数据的分类器,以对新的未标记数据进行分类。除分类标签外,分类器还输出新数据的分类置信度。然后,可以将具有非常高置信度的新数据包括到训练数据中,以进行下一轮模型训练。

使用分类结果进行数据扩展的主要挑战在于如何评估新数据的分类置信度[209]–[212]。由于事件具有复杂的结构,例如不同的事件类型包含具有不同角色的不同实体参数,因此计算事件提取的置信度通常具有较低的准确性。廖和格里什曼[209], [210]提出只使用新样本中的一部分提取结果进行数据扩展,尤其是最自信的触发因素。基于他们的提取模型[94],基于他们的触发器分类器和自变量分类器之间的概率乘积选择最有信心的“<角色,触发器>”对。此外,他们采用了一种称为INDRI [213]的信息检索系统来收集相关文档的簇,并应用了[102]中提出的跨文档推理算法来以高置信度包含新的训练数据。

Wang 等。[211]提出使用基于触发器的潜在实例来发现未标记的数据以进行扩展。如果单词在 gold 数据集中用作触发条件,则在未标记的数据集中提到该单词的所有实例也可能表示相同的潜在实例。基于这样的假设,他们提出了一种对抗训练流程,以过滤出嘈杂的实例,同时提取信息实例以包括新数据。尽管包括来自分类器输出的新数据具有成本效益,但实际上不能保证新包含的数据带有完全正确的注释。在这方面,手动检查非常必要。Liao 和 Grishman [212]提出了一种主动学习策略,称为伪协同测试,该策略采用最少的人工来仅注释那些高置信度的新数据。

前述算法集中于扩展训练数据,但是具有与标记为金色的事件类型相同的事件类型。最近,一些研究人员提出了一种转移学习算法,用于扩展训练数据的事件类型不同于黄金数据集中的参考数据[214]–[218]。例如,Nguyen 等。[216]提出了一种两阶段算法来训练 CNN 模型,该模型可以有效地将知识从旧事件类型转移到

目标类型(新类型)。具体来说,第一步是根据具有随机初始化的权重矩阵的旧事件类型的标记数据训练 CNN 模型。第二阶段是根据一小部分目标类型的标记数据训练 CNN 模型,并使用第一阶段初始化的权重矩阵。最后,经过两阶段训练的 CNN 模型可用于目标类型和旧类型事件检测。

除了利用少量带有新类型的标记数据外,Huang 等人。[218]提出了一种零样本迁移学习,用于提取具有新的看不见类型的事件,它仅需要手动构造新事件类型的定义(例如,事件模式中的事件类型名称和自变量角色名称)。特别地,他们首先通过训练基于金的 CNN 模型来构造事件提及结构(事件实例的触发器,参数及其关系结构)和事件类型结构(事件模型的类型,角色及其关系结构)的矢量表示。数据。接下来,他们使用优化的 CNN 模型来表示新数据的事件提及结构和事件类型结构,并为每个新事件提及找到最接近的事件类型。

B. 从知识库扩展数据

许多现有的知识库都存储了大量的结构化信息,例如 FrameNet,⁸Freebase,⁹Wikipedia¹⁰WordNet,¹¹可以利用这些信息来生成新的标记数据作为事件提取的训练数据。

FrameNet 定义了许多完整的语义框架,每个框架都由一个词法单元和一组名称元素组成。这样的框架与事件共享高度相似的结构。由于 FrameNet 中的许多帧实际上表示某些事件,因此一些研究人员已经探索了将帧映射到事件以进行数据扩展的方法[219]–[221]。例如,刘等。通过使用从 FrameNet 中给定示例语句检测到的事件,扩展了[220] ACE 训练数据。具体来说,他们首先学习了基于 ACE 标记数据的事件检测模型,然后将其用于产生示例语句的初始判断。然后,基于这样的假设,将一组软约束应用于全局推论:“相同的词汇单元,相同的框架和相关的框架倾向于表达相同的事件。”然后将初始判断和软约束形式化为一阶公式,并通过概率软逻辑(PSL)建模[222]进行全局推理。最后,他们从给定的例句中检测事件以进行数据扩展。

类似地,可以将 Freebase(语义知识库)中的复合值类型(CVT)视为事件模板,并将 CVT 实例视为事件实例。CVT 的类型,值和角色分别被视为事件类型,事件中的自变量和事件中所扮演的自变量的角色。Zeng 等[223]利用 Freebase 中 CVT 的结构信息来自动

⁸<https://framenet.icsi.berkeley.edu/fndrupal/>

⁹<https://developers.google.com/freebase/> ¹⁰<https://www.wikipedia.org>

¹¹<https://www.wordnet.org>

注释事件提及以进行数据扩展。他们首先确定了 CVT 的关键论点，它们在一个事件中起着重要作用。如果一个句子包含 CVT 的所有关键参数，则很可能表示 CVT 所呈现的事件。结果，他们记录了该句子中的单词或短语，以使 CVT 的属性与所涉及的自变量相匹配，并具有注释的作用。

而且，Chen 等。[221]提出通过同时利用 FrameNet 和 Freebase 来扩展训练数据。Araki 和 Mitamura [224]利用 WordNet 和 Wikipedia 生成了新的训练数据。除了使用通用知识库以外，一些研究还集中在使用相关知识库来进行领域特定事件提取 [29]，[178]，[225]，[226]。例如，饶等人。[178]利用包含蛋白质之间关系的生物途径交换 (BioPax) 知识数据库来扩展 PubMed¹² 中央文章的训练数据。在金融领域，Yang 等人。[29]利用金融事件知识数据库进行数据扩展，其中包含 9 种常见的金融事件类型。

C. 从多语言数据扩展数据的原因是，同一事件可能用不同的语言描述，并且来自一种语言的标记数据极有可能以另一种语言传达相似的信息，因此，人们提出了一些利用这种多语言的方法解决数据稀疏和低覆盖率问题的信息 [227] - [231]。例如，Zhu 等。[227]使用 Google 翻译消除了中文和英文之间的语言鸿沟，并产生了具有双语单词特征的统一文本表示。以这种方式，它可以合并来自两种语言的训练数据以进行模型训练。

相比之下，一些研究建议通过利用跨语言数据（即跨语言引导）来引导事件提取，而无需使用机器翻译或手动对齐的知识库 [228] - [230]。例如，Chen and Ji [228]提出了一种联合训练自举框架，其中包含两个单语言自我训练自举事件提取系统，一个用于英语，另一个用于中文。可以将已标记的事件从一种语言转换为另一种语言，以便生成所谓的计划触发器和自变量，以另一种语言进行事件提取。Hsi 等。[229]，[230]提出了通过使用多语言功能（例如通用 POS 标签，通用依赖项，双语词典，多语言单词嵌入等）来利用多语言训练来增强分类器的标准事件提取管道的方法。

VIII. 基于无监督学习的事件提取

与监督学习和半监督学习不同，无监督学习不会训练基于标记语料库的事件提取模型。相反，无监督的学习方法

主要专注于开放域事件提取任务，例如基于单词分布表示检测触发器和自变量以及对事件实例进行聚类，并根据它们的相似性进行提及。

A. 事件提及检测和跟踪

事件提及是可以从一个或多个句子描述事件的关键字的集合。这些任务包括检测文章中的事件提及和跟踪不同文章中的类似事件提及。请注意，不需要触发器类型和/或参数角色的分类，因为在此类任务中通常未预定义类型和角色。

TDT 程序 [15] 将主题定义为一组新闻和/或故事，这些新闻和/或故事与某些开创性的现实世界事件密切相关。然后，它会将 TDT 任务定义为确定给定的文章是否与同一主题的事件的聚类有关，并为 TDT 语料库提供简单的标签，例如 {Yes, No, Brief}，以表明每篇文章与该主题的相关性，其中，Brief 表示部分相关。

在许多 TDT 算法中，首先将句子转换为向量表示形式，并计算向量距离以测量与某个主题的相似度以进行事件检测 [232] - [236]。例如，Yang 等。[232]和 Nallapati 等。[233]通过文档的 TF-IDF 向量表示文档，该文档是使用术语袋表示的常规向量空间模型。具体而言，使用术语频率和反向文档频率 (IDF) 对文档中的术语（单词或短语）进行统计加权。它们将 k 个排名最高的术语保留在术语包中，作为文档的表示向量。

斯托克斯 (Stokes) 和凯西 (Carthy) 提出用词汇链表示文档 [237]，该词汇链探索文本的内聚结构以创建一系列语义相关的单词 [234]。例如，关于飞机的文档的词汇链可能由以下单词组成：飞机，飞机，飞行员，驾驶舱，空中小姐，机翼，发动机。他们使用 WordNet 识别了文档中的词汇链，这些词汇链用单个唯一标识符表示同义词。

按照 TDT 程序的任务定义，已经进行了其他一些研究，以检测各种网站中的新文章是否与某些已经确定的事件相关，而无需使用 TDT 语料库 [32] - [37]，[238]。例如，Naughton 等。[35]提出了使用凝聚的层级聚类算法 [239]，使用新闻文章和聚类句子中的词袋编码对句子进行矢量化处理。

除了使用单词和句子嵌入之外，一些人还建议利用新闻报道的一些其他信息（例如时间和位置）来增强事件提及检测。Ribeiro 等。[32]提出将时间，位置和内容维度集成到文本表示中，并应用全对相似性搜索算法和马尔可夫聚类算法对描述同一事件的新闻文章进行聚类。同样，Yu 和 Wu [33]提出了一种 Time2Vec 表示技术来构造文章

¹²<https://www.ncbi.nlm.nih.gov/> 发布

通过上下文向量和时间向量表示, 并采用双层聚类算法进行事件检测。

最近, 已经提出了几种用于事件检测和跟踪的新颖方法, 包括否定实例修剪支持向量机[238], 基于卷积神经网络的多实例学习模型[36]以及基于加权无向二部图的排名。框架[34]。

B. 事件提取和聚类

事件提及检测的任务主要集中于检测事件关键字, 以聚类表达同一事件的句子或文章。一些研究建议进一步将事件触发与将相似事件聚类并为相似事件构建事件模式的参数区分开来[44]–[51]。

一种直接的方法是将句子的动词视为事件触发。例如, Rusu 等。[47]使用句子中的动词作为事件触发, 并通过利用触发器和命名实体, 时间表达, 句子主题和句子对象之间的依赖路径来识别事件自变量。而且, 一些知识库可以被应用来增强触发和论点的区别。Chambers 和 Jurafsky [44]认为动词及其在 WordNet 中的同义词是触发因素。而句法对象的实体是自变量。黄等。[45]认为 OntoNotes¹³ 中的所有名词和动词概念以及 FrameNet 中的言语和词汇单元都是候选事件触发条件。此外, 他们将与候选触发器具有语义关系的所有概念视为候选参数。然后, 他们计算了每对候选日期触发器和参数之间的相似度, 以标识最终触发器和参数。

基于提取的触发器和自变量, 事件实例可以聚类到不同的事件组中, 每个事件组都有一个潜在但又独特的主题。例如, Chambers 和 Jurafsky [44]提出使用概率潜在 Dirichlet 分配主题模型来为它们的聚类计算事件相似度。Romadhony 等。[49]提出利用结构化知识库来聚类触发器和论点。Jans 等。[240]识别和聚类的事件链, 可以将其看作是部分事件结构, 由基于动词表述及其统计信息的动词及其从属角色组成。

此外, 对于每个事件组, 也可以使用插槽值结构建立事件模式, 其中插槽可以表示某个自变量角色, 而值可以表示事件实例的相应自变量。Yuan 等。[46]首先介绍了一个新的事件分析问题, 以填补开放域新闻文档中的空值事件结构。他们还通过利用实体共现信息来提取插槽模式, 提出了一种模式归纳框架。

Glavass^h和 S^hnajder [52]认为事件通常包含

一些常见的论点角色, 如代理, 目标, 时间和位置, 并建议构造一个事件图结构以识别同一事件, 但在不同的文档中已提到。

C. 从社交媒体提取事件

许多在线社交网络, 例如 Twitter, Facebook 等, 都提供了大量最新信息。Twitter 是一个代表。据报道, 每天大约发布 2 亿条推文[241], 其中 1%的推文涵盖了新闻网上报道的事件的 95% [242]。接下来, 我们以 Tweet 为例, 并回顾一些有关从社交媒体提取事件的工作。

与新闻专栏文章相比, Tweet 帖子具有自己的特征, 事件提取面临着新的挑战。推文主要由个人用户发布, 每个推文都有字符限制。因此, 推文通常带有缩写, 拼写错误和语法错误, 这会导致许多零散且嘈杂的文本, 而没有足够的上下文来进行事件提取[28], [38]–[41]。与从新闻专线提取事件相反, 实体, 日期, 位置和关键字是从推文中提取的主要组成部分。由于推文简短, 因此推文中的所有实体均被视为感兴趣事件的参与者。

Weng 和 Lee [42]提出首先分析单个单词的频率, 然后根据它们之间的相关性过滤掉琐碎的单词。之后, 通过基于模块的图形分区技术将剩余的单词聚类以形成事件。里特等。[38]利用命名实体标记器和 TempEx 工具[243]解析时间表达式。周等。[40], [41]建议通过词典匹配过滤掉嘈杂的推文。该词典包含事件关键字(或触发器), 这些事件关键字是从与推文相同时期发布的新闻专线文章中提取的。此外, Zhou 等。[40]建议首先从新闻专线中识别命名实体, 并建立字典以匹配推文中的命名实体。特别是, 他们使用 SUTime [244]解决了时间表达的歧义, 并提出了一种无监督的潜在事件模型来从推文中提取事件。Guille 等。[43]提出了一种基于提及异常的事件检测算法, 该算法通过针对给定的连续时间序列检测事件来计算单词频率中的出现异常。

IX. 事件提取性能比较如前面各节所述, 事件提取可能具有不同的任务定义, 并且可能已经在不同的语料库上进行了实验。因此, 不可能对所有算法都进行公平的比较。但是, 由于有了这些公共评估程序, 性能指标和开放数据集的标准化才使研究人员可以比较他们的算法。在本节中, 我们主要将在公共 ACE 2005 数据集上实验的算法与标准评估程序进行比较, 如下所示:

¹³<https://catalog.ldc.upenn.edu/LDC2013T19>

表 4. 在 ACE 2005 数据集上英语事件提取的性能比较。

System	Approach	Results-F1			
		Trigger	Event Type	Argument	Argument Role
David Ahn (2006) [27]	Machine Learning	62.6%	60.1%	57.3%	-
Ji & Grishman (2008) [102]	Machine Learning	-	67.3%	46.2%	42.6%
Liao & Grishman (2010) [103]	Machine Learning	-	68.8%	50.3%	44.6%
Das et al. (2010) [116]	Machine Learning	-	60.0%	-	41.0%
Hong et al. (2011) [106]	Machine Learning	-	68.3%	53.2%	48.4%
Liao et al. (2011) [107]	Machine Learning	-	61.7%	39.1%	35.5%
Li et al. (2013) [113]	Machine Learning	70.4%	67.5%	56.8%	52.7%
Li et al. (2014) [128]	Machine Learning	-	65.2%	-	46.8%
Cao et al. (2015) [79]	Pattern Matching	-	70.4%	-	-
Chen et al. (2015) [149]	Convolutional Neural Networks	73.5%	69.1%	59.1%	53.5%
Nguyen & Grishman (2015) [141]	Convolutional Neural Networks	-	69.0%	-	-
Judea & Strube (2016) [129]	Machine Learning	66.5%	63.7%	53.1%	41.8%
Sha et al. (2016) [96]	Machine Learning	-	68.9%	61.2%	53.8%
Liu et al. (2016) [105]	Machine Learning	-	69.4%	-	-
Yang & Mitchell (2016) [130]	Machine Learning	71.0%	68.7%	50.6%	48.4%
Zhang et al. (2016) [151]	Convolutional Neural Networks	74.8%	69.1%	58.6%	53.1%
Nguyen et al. (2016) [161]	Recurrent Neural Networks	71.9%	69.3%	62.8%	55.4%
Chen et al. (2016) [163]	Recurrent Neural Networks	72.2%	68.9%	60.0%	54.1%
Feng et al. (2016) [185]	Hybrid Neural Networks	75.9%	73.4%	-	-
Liu et al. (2017) [142]	Artificial Neural Networks	72.3%	69.6%	-	-
Liu et al. (2017) [197]	Artificial Neural Networks	-	71.7%	-	-
Duan et al. (2017) [170]	Recurrent Neural Networks	-	70.5%	-	-
Sha et al. (2018) [166]	Recurrent Neural Networks	-	71.9%	67.7%	58.7%
Wu et al. (2018) [198]	Recurrent Neural Networks	73.4%	71.6%	-	-
Zhang et al. (2018) [200]	Recurrent Neural Networks	76.1%	73.9%	-	-
Ding et al. (2018) [201]	Recurrent Neural Networks	74.9%	71.2%	64.8%	56.6%
Zhao et al. (2018) [203]	Recurrent Neural Networks	-	74.0%	-	-
Chen et al. (2018) [204]	Recurrent Neural Networks	-	73.3%	-	-
Li et al. (2018) [207]	Recurrent Neural Networks	-	75.6%	-	-
Liu et al. (2018) [206]	Recurrent Neural Networks	74.1%	72.4%	-	-
Liu et al. (2018) [179]	Graph Neural Networks	75.9%	73.7%	68.4%	60.3%
Nguyen & Grishman (2018) [180]	Graph Neural Networks	-	73.1%	-	-
Liu et al. (2018) [189]	Hybrid Neural Networks	65.4%	-	-	-
Hong et al. (2018) [191]	Hybrid Neural Networks	77.0%	73.0%	-	-
Nguyen & Nguyen (2019) [172]	Recurrent Neural Networks	72.5%	69.8%	59.9%	52.1%
Zhang et al. (2019) [173]	Recurrent Neural Networks	72.9%	71.6%	-	-
Zhang et al. (2019) [193]	Hybrid Neural Networks	74.6%	72.9%	67.9%	59.7%
Liu et al. (2019) [192]	Hybrid Neural Networks	-	74.8%	-	-

- **触发检测**：如果触发的偏移量（即触发字在文本中的位置）与参考触发相匹配，则可以正确检测到触发。
- **类型识别**：如果触发器的偏移量和事件类型都与参考触发器及其事件类型匹配，则可以正确识别事件类型。

表 5. 在 ACE 2005 数据集上进行中文事件提取的性能比较。

System	Approach	Results-F1			
		Trigger	Event Type	Argument	Argument role
Zhao et al. (2008) [98]	Machine Learning	-	61.2%	-	64.6%
Chen & Ji (2009) [99]	Machine Learning	62.7%	59.9%	46.5%	43.8%
Chen et al. (2012) [112]	Machine Learning	66.7%	63.2%	49.5%	44.6%
Li et al. (2012) [100]	Machine Learning	70.5%	66.9%	55.3%	50.8%
Li et al. (2012) [110]	Machine Learning	73.1%	70.2%	58.1%	53.9%
Li et al. (2013) [111]	Machine Learning	-	-	63.2%	57.4%
Li et al. (2013) [104]	Machine Learning	-	-	64.4%	58.7%
Li et al. (2016) [80]	Pattern Matching	-	58.4%	-	-
Ghaeini et al. (2016) [162]	Convolutional Neural Networks	69.0%	64.8%	-	-
Zeng et al. (2016) [184]	Hybrid Neural Networks	69.3%	64.5%	52.6%	46.9%
Feng et al. (2016) [185]	Hybrid Neural Networks	68.2%	63.0%	-	-
Zhang et al. (2018) [168]	Recurrent Neural Networks	70.9%	63.9%	-	-

- 自变量检测：如果自变量的偏移量与任何参考自变量提及相匹配（即正确识别事件中的参与者），则可以正确检测到自变量。
- 角色标识：如果参数角色的事件类型，偏移量和角色与任何参考参数提及相匹配，则将正确识别该参数角色。

表 4 和表 5 分别以标准 F 度量（F1）表示了通过英文和中文数据集的不同算法报告的事件提取结果，该标准是通过两个常用性能指标即 Precision 和 Recall 获得的。我们注意到并非所有算法都报告所有结果，因为其中一些算法仅设计用于实现某些子任务。还值得注意的是，事件提取还取决于某些上游任务的结果，例如命名实体识别，实体提及分类等。这些算法中的大多数已假定直接在实体，时间，值等中使用金色注释。ACE 2005 作为事件提取任务输入的一部分。但是，我们注意到 ACE 2005 是一个很小的数据集，即使有很少的错误注释也是如此。

从表中可以看出，总体而言，英语算法要比中文算法更好。这可能是由于对中文句子的事件提取也取决于分词结果。此外，还可以注意到，最新的高级神经模型可以实现更好的性能，这可以归因于其强大的能力，可以学习更深，更全面的上下文感知和/或句法感知的单词和句子表示形式。最后，我们可以观察到所有子任务都没有达到很高的 F1 值，这一方面表明了事件提取任务的困难，另一方面却促使人们开发了更高级的算法。

X. 结论与讨论

事件提取是自然语言处理中的一项重要任务，其目的是检测句子是否提及某个现实事件，如果是，则对事件类型进行分类并确定事件参数。由于其广泛的应用，几十年前就对事件提取进行了深入研究，并且由于诸如深度学习等许多新技术的迅速发展，最近它吸引了越来越多的研究兴趣。

在本文中，我们试图为从文本中提取事件提供全面而最新的评论。我们首先介绍了用于封闭域和开放域事件提取的公共评估程序及其任务定义和带注释的数据集。我们将解决方案方法分为五组，包括模式匹配算法，机器学习方法，深度学习模型，半监督学习技术和无监督学习方案。我们已经介绍并分析了每个组中最具代表性的方法，尤其是它们的起源，基础，优点和缺点。此外，我们介绍了评估方法并比较了在 ACE 2005 语料库上实验的典型算法。

模式匹配方法通常可以实现较高的提取精度，但是，模式构建要付出巨大的人力和专业知识。另一方面，包括机器学习和深度学习在内的监督学习方法基于大型带注释语料库的可用性，尽管它们似乎可以实现更好的性能。此外，这些封闭域提取算法通常针对特定域进行训练，并且可能无法直接应用于其他域。对于开放域事件提取，如何处理诸如社交帖子之类的嘈杂文本以及如何组织时空事件仍然是实践中的巨大挑战。

接下来, 我们讨论事件提取的未来可能的研究方向。

A. 知识增强型深度学习

尽管深度学习技术已证明是自动特征学习中用于事件提取的强大工具, 但这些神经模型通常具有太多的学习参数以及网络配置, 这不仅需要大量带注释的原始数据, 而且还需要仔细调整众多的网络配置。另一方面, 模式匹配可以很好地利用专家的知识来指定准确的事件模式, 尽管这需要付出更多人工的代价。尽管这两种方法似乎相互矛盾, 但一个有希望的方向可能是如何在专家的知识参与下增强深度学习模型。尽管可以设想在神经模型训练过程中使用模式的初步尝试, 但需要付出更多的努力才能享受高质量的人类知识。

B. 领域自适应转移学习

在封闭域事件提取中, 通常使用详细的事件类型和事件参数的角色来预定义各种事件模式。尽管模式为事件提供结构化的表示, 但明确定义模式涉及特定领域的知识。此外, 尽管许多带有不同标签的参数即使在不同类型的事件中也扮演相似的角色, 但特定于域的架构很难从一个域扩展到其他域。转移学习似乎是开发领域自适应事件提取系统的一种有前途的方法, 其中在一个领域中训练的提取模型可以很容易地应用于其他领域, 而只需进行少量调整即可。尽管如此, 将常见的用于多类型分类的训练模型首先转换为多任务学习模型可能是有见地的尝试。

C. 资源感知事件聚类

在开放域事件提取中, 首先从不同来源检测事件, 然后提取事件提及或关键字。然后, 通常根据事件关键字之间的相似性来执行事件聚类, 即, 实施一种以主题为中心的事件聚类。但是, 我们注意到, 不仅事件可以用其关键字进行简短描述, 而且还可以使用文本中的其他资源(例如发布者)进行描述。一方面, 我们同意以主题为中心的事件聚类仍然应该是开放域事件提取的基本操作; 另一方面, 我们注意到通过包含其他资源, 可以实现多焦点聚类, 这可能会进一步提示其他任务, 例如事件推理以回答此类事件发生的原因以及事件推理以回答下一类事件的发生。被期待。

参考资料

- [1] *Ace (事件的自动内容提取) 英语注释准则*, 语言数据协会, 美国宾夕法尼亚州费城, 2005 年。
- [2] M. Rospocher, M. van Erp, P. Vossen, A. Fokkens, I. Aldabe, G. Rigau, A. Soroa, T. Ploeger 和 T. Bogaard, “通过新闻构建以事件为中心的知识边缘图”, *J. Web 语义*, 第一卷, 37-38, 第 132-151 页, 2016 年 3 月。
- [3] Z. Li, X. Ding 和 T. Liu, “为脚本事件预测构建叙事事件进化图”, *Proc. 27 日联合会议 Artif. Intell.*, 2018 年, 第 4201-4207 页。
- [4] SJ Conlon, AS Abrahams 和 LL Simmons, “从在线报告中提取恐怖主义信息”, *J. Comput. Inf. 系统*, 卷, 55, 不, 3, 第 20-28 页, 2015 年。
- [5] M. Atkinson, J. Piskorski, H. Tanev, EV der Goot, R. Yangarber 和 V. Zavarella, “边界安全领域的自动事件提取”, *Proc. 诠释 Conf. 用户中心媒体*, 2009 年, 第 321-326 页。
- [6] H. Tanev, J. Piskorski 和 M. Atkinson, “用于全球危机监控的实时新闻事件提取”, *Proc. 诠释 Conf. 应用自然语言处理. Inf. Syst.*, 2008, 第 207-218 页。
- [7] M. Atkinson, M. Du, J. Piskorski, H. Tanev, R. Yangarber 和 V. Zavarella, 《计算语言学》中的“从在线新闻中提取与多语言安全性相关的事件的技术”。德国柏林: 施普林格, 2013 年, 第 163-186 页。
- [8] J. Piskorski, H. Tanev 和 PO Wennerberg, “从在线新闻中提取暴力事件以获取本体论人群”, *Proc. Nasdaq. 诠释 Conf. 总线. Inf. Syst.*, 2007, 第 287-300 页。
- [9] W. Nuij, V. Milea, F. Hogenboom, F. Frasincar 和 U. Kaymak, “将新闻纳入股票交易策略的自动框架”, *IEEE Trans. 知道数据工程*, 第 1 卷, 26 号 4, 第 823-835 页, 2014 年 4 月。
- [10] P. Capet, T. Delavallade, T. Nakamura, A. Sandor, C. Tarsitano 和 S. Voyatzi, *Proc. 中的“具有事件类型的自动提取的风险评估系统”*, 诠释 Conf. 智力 Inf. 流程, 2008 年, 第 220-229 页。
- [11] AS Abrahams, “开发和执行出现的电子商务应用程序”, 博士学位论文, 计算机系, 科学大学技术学院剑桥, 英国剑桥, 2002 年 9 月。
- [12] JA Vanegas, S. Matos, F. González 和 JL Oliveira, “从科学文献中提取生物分子事件的概述”, *Comput. 数学. 方法医学*, 第一卷, 2015, 八月 2015, 艺术. 没有. 571381。
- [13] BM Sundheim 和 NA Chinchor, *Proc. 中的“信息理解会议的调查”*, 车间嗡嗡声. 郎技术, 1993 年, 第 56-60 页。
- [14] L. Hirschman, “评估的演变: 消息理解会议的经验教训”, *Comput. 演讲和语言*, 第一卷 12 号 4, 第 281-305 页, 1998 年。
- [15] J. Allan, 主题检测和跟踪: 基于事件的信息组织, 第一卷, 12. 德国柏林: 施普林格, 2012 年。
- [16] H. Yu, Y. Zhang, L. Ting 和 L. Sheng, “主题检测和跟踪审查”, *J. Chin. Inf. 过程*, 6 号 21, 第 77-79 页, 2007 年。
- [17] J. Aguiar, C. Beller, P. McNamee, BV Durme, S. Strassel, Z. Song 和 J. Ellis, *Proc. 中的“ace, ere, tac-kbp 和 framenet 注释标准之间的事件和关系的比较”*, 第二届研讨会活动定义检测会议。代表, 2014 年, 第 45-53 页。
- [18] Z. Song, A. Bies, S. Strassel, T. Riese, J. Mott, J. Ellis, J. Wright, S. Kulick, N. Ryant 和 X. Ma, *Proc. 中的“从轻到重: 实体, 关系和事件的注释”*, 第三研讨会活动定义, 检测会议。代表, 2015 年, 第 89-98 页。
- [19] T. Mitamura, Z. Liu 和 E. Hovy, *Proc. 中的“TAC KBP 2015 事件块轨迹概述”*, 文字肛门. Conf., 2015 年, 第 1-11 页。
- [20] C. Nédellec, R. Bossy, J.-D. 金俊杰 Kim, T. Ohta, S. Pyysalo 和 P. Zweigenbaum, *Proc. 中的“BioNLP 共享任务概述 2013”*, BioNLP 共享任务研讨会, 2013 年, 第 1-7 页。
- [21] J. Pustejovsky, P. Hanks, R. Sauri, A. 见 R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro 和 M. Lazo, “语料库”, 语料库语言学。英国兰开斯特: 兰开斯特大学, 2003 年, 第 2 页。40
- [22] F. Hogenboom, F. Frasincar, U. Kaymak 和 FD Jong, 《过程中

- 从文本中提取事件的概述》，Proc. 车间检测代表。剥削事件语义网 (DeRiVE)，2011 年，第 48-57 页。
- [23] F. Hogenboom, F. Frasincar, U. Kaymak, FD Jong 和 E. Caron, “从决策支持系统的文本中提取事件方法的概述”，Decis. 支持系统, 卷 85, 第 12-22 页, 2016 年 5 月。
- [24] Z. Zhang, Y. Wu 和 Z. Wang, “开放域事件提取调查”，计算部。大学科学伊利诺伊州厄巴纳香槟, 美国伊利诺伊州香槟市, 技术。2018 年 1 月。