



计算机科学

COMPUTER SCIENCE

低资源场景事件抽取研究综述

刘涛, 蒋国权, 刘姗姗, 刘浏, 环志刚

引用本文

刘涛, 蒋国权, 刘姗姗, 刘浏, 环志刚. [低资源场景事件抽取研究综述](#)[J]. 计算机科学, 2024, 51(2): 217-237.

LIU Tao, JIANG Guoquan, LIU Shanshan, LIU Liu, HUAN Zhigang. [Survey of Event Extraction in Low-resource Scenarios](#) [J]. Computer Science, 2024, 51(2): 217-237.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于核心句的端到端事件共指消解](#)

End-to-End Event Coreference Resolution Based on Core Sentence

计算机科学, 2023, 50(11): 185-191. <https://doi.org/10.11896/jsjcx.221000078>

[基于复合语义特征的事件图谱构建技术研究进展](#)

Overview About Composite Semantic-based Event Graph Construction

计算机科学, 2023, 50(9): 242-259. <https://doi.org/10.11896/jsjcx.230400046>

[基于多粒度实体异构图的篇章级事件抽取方法](#)

Document-level Event Extraction Based on Multi-granularity Entity Heterogeneous Graph

计算机科学, 2023, 50(5): 255-261. <https://doi.org/10.11896/jsjcx.220300154>

[门控机制融合多种特征的中文事件共指消解](#)

Employing Gated Mechanism to Incorporate Multi-features into Chinese Event Coreference Resolution

计算机科学, 2023, 50(3): 291-297. <https://doi.org/10.11896/jsjcx.220700146>

[医学知识图谱研究与应用综述](#)

Survey of Medical Knowledge Graph Research and Application

计算机科学, 2023, 50(3): 83-93. <https://doi.org/10.11896/jsjcx.220700241>

低资源场景事件抽取研究综述

刘涛^{1,2} 蒋国权² 刘姗姗² 刘浏² 环志刚²

1 南京信息工程大学计算机学院软件学院网络安全空间学院 南京 210044

2 国防科技大学第六十三研究所 南京 210007

(1987359802@qq.com)

摘要 事件抽取作为信息抽取的任务之一,旨在从非结构化文本中抽取结构化事件信息。当前基于机器学习和深度学习的自动化信息抽取方法过度依赖于标注数据,而大多数领域的标准数据集规模很小且分布不均匀,因此低资源场景成为了限制自动化信息抽取性能的瓶颈。虽然,近年来众多学者针对低资源场景进行了深入研究,并取得了许多显著的成果,但目前针对该场景下的事件抽取研究综述比较缺乏。文中对现有的学术成果进行了较为全面的总结分析,首先介绍了相关任务的定义,并将低资源场景事件抽取任务分为3类;其次围绕此分类重点阐述了6种相关技术方法,包括基于迁移学习、基于提示学习、基于无监督学习、基于弱监督学习、基于数据与辅助知识增强、基于元学习的方法,并指出了当前方法的不足和未来改进的方向;然后介绍了相关数据集及评价指标,并对典型技术方法的实验结果进行了总结分析;最后从全局角度总结分析了当前低资源场景事件抽取工作面临的挑战及未来研究的趋势。

关键词:事件抽取;低资源场景;数据处理;场景适应

中图分类号 TP391

Survey of Event Extraction in Low-resource Scenarios

LIU Tao^{1,2}, JIANG Guoquan², LIU Shanshan², LIU Liu² and HUAN Zhigang²

1 School of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China

2 The Sixty-third Research Institute, National University of Defense Technology, Nanjing 210007, China

Abstract As one of the tasks of information extraction, event extraction aims to extract structured event information from unstructured text. The current automated information extraction methods based on machine learning and deep learning rely on labeled data excessively, but standard datasets in most areas are small and unevenly distributed. So the low-resource scenarios become an important bottleneck that limits the performance of automated information extraction. Although in recent years, many scholars have conducted in-depth research on low resource scenarios and produced many remarkable results, there is still a lack of research on event extraction in this scenario at present. This paper makes a comprehensive summary and analysis of existing academic achievements. Firstly, it introduces the definition of related task, and the task of event extraction in low resource scenarios is divided into three categories. Then six kinds of related techniques and methods are discussed around this classification, including transfer learning based, prompt learning based, unsupervised learning based, weakly supervised learning based, data and auxiliary knowledge enhancement based, and meta learning based approaches. Subsequently, the shortcomings of current methods and strategies for future improvement are pointed out. Then the related datasets and evaluation metrics are introduced and the experimental results of typical techniques are summarized and analyzed. Finally, the challenges and future research trends about event extraction in low resource scenarios are summarized and analyzed from a global perspective.

Keywords Event extraction, Low-resource scenarios, Data processing, Scenarios adaptation

事件抽取是信息抽取的一个重要子任务,源于20世纪80年代末期开始的消息理解会议(Message Understanding Conference, MUC)^[1],由美国国防高级研究计划委员会(Defense Advanced Research Projects Agency, DARPA)赞助。

目前最具影响力的事件抽取评测是ACE(Automatic Content Extraction)评测,由美国国家标准技术研究所(National Institute of Standards and Technology, NIST)于1999年开始举办,2005年起,事件抽取任务被纳入ACE中,2009年被纳入

到稿日期:2022-12-24 返修日期:2023-04-13

基金项目:第四批军事科技领域青年人才托举工程项目(2021-JCJQ-QT-050);中国博士后科学基金资助项目(2021MD703983);江苏省高等学校自然科学研究面上项目(20KJB413003)

This work was supported by the Fourth Batch of Young Talent Lifting Project of Military Science and Technology(2021-JCJQ-QT-050), China Postdoctoral Science Foundation(2021MD703983) and Natural Science Foundation of the Jiangsu Higher Education Institutions of China(20KJB413003).

通信作者:蒋国权(jianggq2001@163.com)

文本分析会议 (Text Analysis Conference, TAC) 公共评估项目^[2]。之后 DARPA 的深度探索和文本过滤 (Deep Exploration and Filtering of Text, DEFT) 计划提出文本注释和信息抽取的实体、关系、事件 (Entities, Relations, Events, ERE) 标准^[3]。事件抽取也是 KBP (Knowledge Base Population) 公共评估项目的重要任务, 之后 KBP 与 TAC 被整合在一起^[4]。此外还有用于特定领域事件抽取的其他公共评估项目, 如在生物医学领域的 BioNLP^[5]、用于抽取事件时间信息的 TimeBANK^[6] 等。

早期的事件抽取方法采用模式匹配的方式, 需要领域专家设计模板且可移植性差^[7-8]。而传统的机器学习方法如支持向量机 (Support Vector Machine, SVM)^[9]、最大熵 (Maximum Entropy, ME)^[10] 等模型将事件抽取视为一个分类任务, 会面临特征工程的问题 (复杂特征的构建需要语言学 and 领域专家知识及特征选择的难题)。之后研究者们开始利用基于卷积神经网络 (CNN)、循环神经网络 (RNN)、图神经网络 (GNN)、Transformer 编码器等深度学习模型进行事件抽取, 特别是 ELMO (Embedding from Language Model)^[11]、GPT (General Pre-training)^[12]、BERT (Bidirectional Encoder Representations from Transformers)^[13] 等预训练语言模型可以获得 token 更通用的表示形式, 以更好地进行模型参数的初始化, 提升对目标任务的泛化能力。

上述方法中后两者特别是深度学习技术严重依赖于大量的注释数据, 然而在实际场景下可获得的标记数据极其有限, 低资源场景十分普遍, 往往会导致过拟合问题。低资源场景产生的原因有很多, 如在生物医学、金融、新闻、法律等特定领域, 由于保密性、安全性、数据质量要求高以及标注数据耗时耗力等因素, 可获得的注释数据是极其有限的。另外, 由于标注数据成本高等原因, 使得小语种语言和应用广泛的英文也面临标注数据缺乏的问题。同时, 现有的事件抽取数据集规模较小且分布不均匀等现象特别明显。这些问题很大程度上限制了事件抽取方法的应用, 导致现有的技术方法面临严峻的挑战。

基于上述挑战, 研究低资源场景事件抽取任务是十分有意义的。研究者设计了多种策略来解决真实应用场景下数据稀疏和冷启动等问题, 利用迁移学习、提示学习、无监督学习、弱监督学习、数据与辅助知识增强、元学习等技术, 有效地缓解了数据稀缺以及数据集规模较小且分布不均匀的问题。

尽管常规的事件抽取综述已经很丰富^[14-15], 但目前关于低资源场景下的事件抽取研究综述还比较缺乏。本文全面分析总结了目前关于低资源场景事件抽取的相关工作, 重点围绕当前较流行的技术方法展开。第 1 章给出了低资源场景事件抽取的任务定义及其分类; 第 2 章对近几年的技术方法进行了重点阐述, 同时简要介绍了各方法已有的工作, 并针对相关技术方法存在的问题及未来可改进和关注的重点进行了阐述; 第 3 章介绍了相关数据集及评价指标, 并总结分析了典型技术方法在 ACE2005 数据集^[2] 上的实验结果; 第 4 章从低资源场景事件抽取技术发展的全局角度, 对当前技术方法面临的挑战以及未来发展趋势进行了全面总结与展望; 最后总结全文。

1 低资源场景事件抽取任务定义及分类

1.1 任务定义

1.1.1 事件抽取任务定义

事件抽取任务的目标是从大量非结构化文本中抽取出

目标事件, 事件抽取包含事件检测和论元抽取两个子任务。其中事件检测包括触发词识别和触发词分类, 论元抽取包括论元识别和论元角色分类两个子任务。

触发词识别: 触发词是事件抽取的核心单元, 它清晰地表达一个事件的出现。触发词的识别任务是发现文本中的触发词。

触发词分类: 触发词的分类是根据存在的触发词, 确定相关句子是否是一个事件, 如果该句子是一个事件, 则需要进一步确定该句子属于哪一个或哪几个事件类型。

论元识别: 从文本中识别出包含一个事件类型的所有论元, 论元的识别通常依赖于触发词识别与分类的结果。

论元角色分类: 基于包含在事件抽取模式中的论元, 根据识别的论元, 将每一个论元分类为相应的类别。

1.1.2 低资源场景定义

低资源场景指可获取的资源非常稀少, 低资源任务^[16]具有平行数据稀缺、语义表达方式单一、数据噪声多、语料对齐质量差等特点。低资源任务不仅需要选取合适的神经网络结构, 还需要借助更多的训练技巧和策略来缓解数据稀疏等问题。

低资源场景可以分为 k-shot 场景与长尾场景^[17], k-shot 场景又分为 few-shot 与 zero-shot 场景。few-shot 场景指候选类只有小部分标记样本, 这些小的数量可以遵循 n-way-k-shot 设定固定, 或是相对于总量而言未固定的部分较小。形式上, 给定可见类 S_y 及其样本 X_a 和未见类 U_y 及其样本 X_b , 在训练 tr 和测试 te 的过程中, 样本满足 $X_{tr} \in X_a$, $X_{te} \in X_b$, 标签满足 $S_{tr} \in S_y$, $U_{te} \in U_y$, 并且未见类的 few-shot 标记样本 X_{te_w} 与测试样本 X_{te} 是不相交的, 即 $X_{te_w} \cap X_{te} = \emptyset$ 。而 zero-shot 场景意味着待预测的测试样本从未出现在训练样本中, 形式上类似于 few-shot 场景 $S_{tr} \in S_y$, $U_{te} \in U_y$, 在标准的零样本场景下, 只有未见类 U_y 在预测中被考虑, 在泛化的零样本场景下, 可见类 S_y 与未见类 U_y 都被考虑在预测中, 即 $U_{te} \in S_y \cup U_y$ 。除了 k-shot 场景外, 低资源场景还包括长尾场景, 即只有一小分类有足够丰富的标记样本, 多数类只有极少的标记样本数据。形式上, 给定候选类 $C = \{C_h, C_t\}$ (头类 C_h 和尾类 C_t), 分别使用 $|N_h|$ 和 $|N_t|$ 表示 C_h 和 C_t 标记样本的数量, 满足 $|N_h| \geq |N_t|$ 而 $|C_h| \leq |C_t|$ 。

1.1.3 低资源场景事件抽取任务定义

本文研究的是低资源场景事件抽取任务, 低资源场景事件抽取任务中事件类型只存在少量标注样本或者是无标注样本。低资源场景事件抽取的本质是: 依赖少量有标签的样本去学习泛化能力强的模型, 期望其在未知数据上表现良好。从模型角度来看, 旨在设计鲁棒的策略来提升模型的泛化能力, 从而充分利用已有的小样本数据进行事件抽取任务, 减少对样本的依赖。从数据角度来看, 利用内生 (基于规则的方法) 或外生 (如多语言、多模态数据) 的辅助资源生成更加丰富、准确的语义表示, 或是从模型和数据结合的角度利用现有的稀疏数据搜索更合适的学习策略。

结合事件抽取任务与低资源场景的定义, 以句子级事件抽取为准, 本文对低资源场景事件抽取给出如下定义: 在目标事件类型只有少量或零样本的情况下, 即给定低资源场景 L 和句子 $S(a_1, a_2, \dots, a_m)$ 以及执行事件抽取任务 E 。此时触发词 Tri 的识别任务可以表示为:

$$a_{k,\lambda}^* = \arg \max_{k \geq 1, \lambda \leq m} p(a_{k,\lambda} = Tri | S, L) \quad (1)$$

其中, $a_{k,\lambda}$ 表示句子 S 的子串, $p(\cdot)$ 表示 $a_{k,\lambda}$ 是触发词的概率,

$a_{k,\lambda}^*$ 表示识别得到的最优子串(即触发词)。

对于触发词的分类任务,给定一组预定义的事件类型集合 $E_e(e_1, e_2, \dots, e_n)$,判断触发词 Tri 属于事件类型集合 E_e 的最大概率,此时触发词的分类任务可以表示为:

$$e^* = \arg \max_{1 \leq i \leq n} p(Tri \in e_i | Tri, S, L) \quad (2)$$

其中, $p(\cdot)$ 表示已识别的触发词属于相应事件类型的概率, e^* 表示触发词所对应的最大概率的事件类型(即触发词的事件类型)。

对于每一个事件类型 e_i 的论元集合 $Re_i(re_1, re_2, \dots, re_n)$,此时论元识别任务可以表示为:

$$re^* = \arg \max_{1 \leq j \leq h} p(a_{k,\lambda} = re_j | Tri, e_i, S, L) \quad (3)$$

其中, $1 \leq i \leq n, k \geq 1, \lambda \leq m, p(\cdot)$ 表示句子 S 的子串 $a_{k,\lambda}$ 是相关事件所对应论元的概率, re^* 表示识别得到的最优子串(即事件对应的论元)。

对于每个事件类型 e_i 已经识别的论元集合 $Re_i(re_1, re_2, \dots, re_n)$,以及给定的论元角色集合 $Qe_i(qe_1, qe_2, \dots, qe_w)$,此时每一个事件类型 e_i 的论元角色分类问题可以表示为:

$$qe^* = \arg \max_{1 \leq j \leq w} p(re_j \in qe_j | re_j, Tri, e_i, S, L) \quad (4)$$

其中, $1 \leq j \leq h, 1 \leq i \leq n, p(\cdot)$ 表示相关事件的论元属于其对应论元角色集合的概率, qe^* 表示论元所对应的最大概率的论元角色(即论元所属的论元角色)。

1.2 低资源场景事件抽取任务分类

本节从低资源场景下不同任务类型的角度出发,将低资源场景事件抽取任务分为跨语言任务、跨领域任务以及新事件类型任务。

1.2.1 跨语言事件抽取任务

现实场景下存在大量语言,但只有很少的标注数据。世界上有7000多种语言,常用的语言也就几十种,剩下绝大多

数都是低资源语言,由于对相关语言数据进行标注耗时耗力,因此低资源的情况在小语种语言中是很常见的。研究者开始通过采用一系列的策略,利用高资源语言的数据(如英文)改善低资源语言数据缺乏的情况,从而提升对低资源语言的事件抽取效果。

1.2.2 跨领域事件抽取任务

由于特定领域(如医疗、金融、法律等)对数据质量的要求很高,而高质量的标注数据十分稀缺、昂贵。例如金融领域中,受限于标注成本以及数据保密等因素,很难获得大量的高质量标注数据。在生物医学领域,由于法律限制对患者病历记录的使用和共享,导致获得药物和疾病等相关事件的标注数据也极具挑战,因此特定领域通常会面临少样本低资源的问题。类似于跨语言的事件抽取任务,研究者利用通用领域、知识库、多模态等获得大规模数据来改善特定领域面临的资源缺乏问题,从而提升对特定领域事件抽取的效果。

1.2.3 新事件类型抽取任务

由于人工标注数据是耗时耗力的,因此现有的标准事件抽取数据集的规模普遍都很小。如ACE2005数据集只有8个类型和33个子类型,其对事件类型的覆盖规模很小,在实际场景下会面临大规模的事件类型只有少量标注数据,甚至是无标注数据的问题。而目前大多数事件抽取方法通过测量测试事件提及和注释事件提及之间编码特征的相似性,将事件抽取视为分类问题,因此这些新事件对模型来说是从未见过的,本文将这些在训练过程中模型从未见过的事件情形视为新事件类型,需要进一步设计策略来应对这些新事件。

2 低资源场景事件抽取技术方法

表1列出了各方法的思想、主要适用任务以及相关工作。

表1 低资源场景事件抽取技术方法总览

Table 1 Overview of event extraction techniques in low resource scenarios

方法类型		方法思想		主要适用任务			
				跨语言	跨领域	新事件类型	
迁移学习方法 (2.1节)	深度学习模型迁移方法	CNN	设计鲁棒的神经网络模型,捕获更多有用特征或利用资源丰富的数据、知识增强目标数据缺乏的情况	迁移学习技术旨在设计鲁棒的模型,将高资源、高质量的数据迁移到目标任务上,从而减少对目标注释数据的依赖、削减目标标注数据缺乏的情况	✓	✓	✓
		RNN					
		GNN					
	Transformer						
	预训练语言模型方法	采用预训练微调范式,首先在高资源数据上预训练,之后在低资源目标数据上微调	✓	✓	✓		
	机器阅读理解方法	将事件抽取任务建模为问答对,从给定的文本中查询正确的答案范围	—	—	✓		
	多任务学习方法	同时将多个子任务放在一起学习,使得子任务之间信息共享,相互补充,以实现相互促进	—	✓	—		
	提示(prompt)学习方法 (2.2节)	给预训练语言模型一些线索和提示,从而更好地利用预训练语言模型已有的知识让它理解下游任务	—	—	✓		
	无监督学习方法 (2.3节)	不训练基于标记语料库的事件抽取模型,主要集中在开放域事件抽取上,通过聚类策略归纳事件模式	—	—	✓		
弱监督学习方法 (2.4节)	半监督学习方法	旨在利用结构化知识库或者少量人工标注数据,自动生成大规模、高质量的标注数据,利用新生成的标注数据和原始的标注数据进行模型训练	—	✓	✓		
	远程监督学习方法						
数据与辅助知识增强方法 (2.5节)	多语言增强方法	主要利用高资源、高质量的数据和知识缓解目标数据稀疏的情况	✓	—	✓		
	多模态增强方法						
	辅助知识增强方法						
	元学习的方法 (2.6节)	从数据和任务的双重抽样中学习历史先验知识,使得模型可以快速地适应未来的任务(新事件类型)	—	—	✓		

注:其中“✓”表示常用,“—”表示不常用。

将低资源场景事件抽取方法分为6类,包括迁移学习、提示学习、无监督学习、弱监督学习、数据与辅助知识增强、

元学习,每种方法根据处理特点的不同又细分为若干种不同的子方法,本文对重要指标的分析如不具体说明,默认是在

ACE2005 数据集上实验得到的结果,具体的阐述见各自模块。

2.1 基于迁移学习的方法

传统机器学习、深度学习严重依赖大量的注释数据,但实际场景下可获得的注释数据是极其有限的,大大限制了其可用性。另外,训练数据和测试数据分布差异很大时无法满足独立同分布条件,需要重新搜集训练数据来训练新模型,这种方式实际代价很大,因此迁移学习作为一个新的学习方法应运而生。迁移学习旨在设计鲁棒的模型,挖掘更多有用信息,从而减少对目标数据的依赖或利用高质量、高资源的数据缓解目标数据缺乏的情况,其中较典型的方法如预训练、微调范式。最近也有研究者提出机器阅读理解(Machine Reading Comprehension, MRC)方法将事件抽取任务建模为问答对从文本中查询答案范围。本节针对事件抽取任务,将技术方法分为基于深度学习的迁移方法、基于预训练语言模型的方法、基于 MRC 的方法以及基于多任务学习方法。

2.1.1 基于深度学习模型的迁移方法

基于深度学习模型的迁移学习方法利用 CNN, RNN, GNN 以及 Transformer 等网络结构,来充分捕获大量有用的特征和信息,从而减少对注释数据的依赖,或利用资源丰富的数据、已学到的知识等缓解目标数据缺乏的情况。

1) 基于 CNN 的模型

卷积神经网络(CNN)可以很好地捕获句子的语法和语义特征,有效缓解训练数据缺乏的情况。CNN 的权值共享策略可以减少网络的训练参数,降低模型复杂度,适应性更强,还能减少对注释数据的依赖,从而扩展模型的泛化能力,特别是面对没有任何注释数据的新事件类型时,该方法很有效。

Nguyen 等^[18]提出首先利用辅助类型的训练数据学习一个特征抽取器,并对 CNN 的权重随机初始化,之后利用目标类型小部分标记数据和随机初始化的权重训练 CNN 模型,来预测目标事件类型和旧事件类型。Lai 等^[19]提出了一种从关键词(Learning from Keyword Formulation, LFK)检测新事件类型的方法,其优势在于只要涉及描述新事件的关键词被提供,就可以识别新事件类型,并提出一个注意力机制来增强 CNN 模型。

除了利用小部分新事件类型的标注数据外,还有研究者提出直接对没有任何注释的新事件类型进行识别。Huang 等^[20]利用 CNN 将事件提及和事件类型的结构表示映射到一个共享语义空间上,通过最小化两者之间的距离,将可见事件类型的知识迁移到未见类型上,在新事件上触发词识别的 $F1$ 值达到 55.6%,与使用 500 个事件提及注释的监督学习方法取得的效果相当。

CNN 可以有效地提升事件抽取的效果,但 CNN 模型最典型的缺点是不能很好地捕获远距离单词之间的依赖关系,并且事件检测和论元抽取两个子任务的管道执行方式也会面临错误传播的问题。

2) 基于 RNN 的模型

RNN 拥有很强的记忆能力,可以捕获上下文中每个单词的信息,建立起序列的长期依赖关系,更好地对单词表示建模。基于 RNN 的迁移学习技术利用 RNN 的优势捕获更多

有用的序列信息和特征,来改善事件抽取面临的标记数据缺乏的情况。

Feng 等^[21]利用双向长短期记忆(Bi-directional Long-short Term Memory, Bi-LSTM)网络和 CNN 网络分别捕获特定的序列信息和词块信息,使用它们训练一个多语言事件检测器,在英文触发词识别的 $F1$ 值及触发词分类的 P 值分别达到 75.9% 和 84.6%。在中文触发词识别的 $F1$ 值可达 68.2%。

Ramrakhiani 等^[22]提出首先使用通用领域事件标签数据预训练 BiLSTM-CRF(Conditional Random Field)序列标记器,提前学习到基于动词和名词事件的通用特征,之后迁移到一个小规模的工业领域事件标记数据集上对网络进行后训练,可以得到一个鲁棒的工业事件标记器,应用于后续的任务。

Hou 等^[23]提出使用迁移学习和结构相关性学习技术从两个不同的事件抽取领域生成一个通用的数据集,之后使用一个基于 LSTM 网络的分类器模型从生成的数据集中学习特征模式,可以很好地进行事件和关系的抽取任务。

基于 CNN 和 RNN 的迁移学习技术可以有效地捕获更丰富的上下文信息,但两者都很难处理图形数据结构,并且也不能完全模拟单词之间的依赖关系。

3) 基于 GNN 的模型

图神经网络通过应用多个神经元在一个图结构上操作,在非欧几里得空间实现几何深度学习,可以有效地解决上述 CNN 和 RNN 面临的难以完全建模单词依赖关系等问题。基于图神经网络模型进行事件抽取的方法主要利用句子的共性特征,如符号特征(词性(Part of Speech Tagging, POS)标记、单词依赖关系)、语义信息、类信息等,借助于相应的图操作来整合上述信息;另外还有方法将输入句子转换为依赖图、语言共性的图结构等来执行迁移任务。本节主要从基于图卷积神经网络、基于图注意力网络以及基于图表示学习的方法进行介绍。

(1) 基于图卷积神经网络(Graph Convolutional Neural Networks, GCN)的方法。GCN 通过节点之间边的连接,来预先设定相等的边权聚合邻居节点的信息,从而更新当前节点的特征表示,将卷积核与 GNN 结合可以有效地整合相关信息。

Lu 等^[24]提出使用 GCN 编码器将所有的实体提及、事件触发词和上下文表示编码到一个复杂和结构化的多语言通用空间上,通过利用语言共性特征,包括符号信息(POS 标记和依赖路径)和分布信息(类型表示和上下文表示),从源语言注释中训练一个关系和事件抽取器,并将其应用到目标语言上,用于解决低资源语言下复杂语义结构识别的问题。

受先前工作使用 GCN 进行跨语言事件论元抽取(Cross-lingual Event Argument Extraction, CEAE)的启发,Nguyen 等^[25]为 CEAE 模型引入了基于语义相似性和单词对普遍依赖关系两个语言独立的信息源,利用这两个信息源可以生成共享的句子结构,从而弥合语言之间的差异。

(2) 基于图注意力机制的方法。上述方法将中心结点与邻居结点的边权视为相等,但对于中心结点而言,不同的邻居应该具有不同的重要度,将注意力机制引入 GNN,以训练的

方式学习相邻节点间的边权,可以有效地解决该问题,同时增加了模型的可解释性。

类似 GCN 编码结构信息的方法,利用图注意力机制也可以有效地整合结构信息。Ahmad 等^[26]引入了一个图注意力 Transformer 编码器 GATE(Graph Attention Transformer Encoder),用于明确地融合结构信息,GATE 允许信息在多头注意结构中根据语法距离在不同头之间传播,以学习不同事件提及类型和目标标签之间的相关性,在由英文迁移到中文时论元角色标记的 F1 值达到 63.2%,相比之前的方法效果提升很显著。Xu 等^[27]则首先利用 TextRank 算法提取出包含新闻主旨的事件句,之后将依存句法树扩充成图得到节点之间的邻接矩阵,利用图注意力网络对节点特征进行训练,接着进行事件的联合抽取并取得了显著成效。

(3)基于图表示学习的方法。该方法主要是将文本句子转换为图结构的形式,之后使用相应的编码器利用图表示的信息。

Lu 等^[28]探索了基于普遍依赖解析和全连接图两种方法,如图 1 所示。该方法的优势在于其将任何一种语言中的句子转化为语言共性的图结构,然后使用跨语言的词嵌入来表示这些图结构中的结点,使得任何语言的句子都可以被表示在一个共享语义空间上,从而提升了跨语言事件抽取的效果。

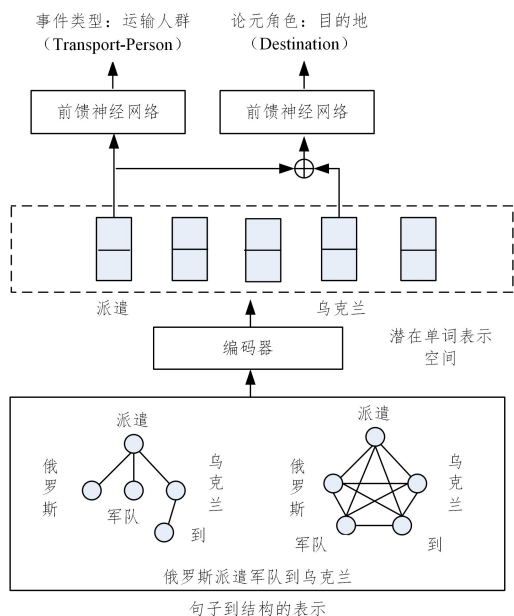


图 1 跨语言事件结构迁移整体框架

Fig. 1 Overall framework of cross-lingual event structure transfer

针对 Lu 等^[29]通过神经匹配方法将词义消歧(Word Sense Disambiguation, WSD)任务应用到小样本任务时面临性能大幅下降的问题,Lai 等^[30]提出让 WSD 模型继承最优的 WSD 结构得到最有效的表示,将 WSD 模型中的知识迁移到 ED 模型中执行训练,利用依赖图的训练信号正则化表示学习来提升新事件检测的效果。

Zhuang^[31]提出首先利用组件图模块神经网络(Compositional Graph Modular Network, CGMN)中的一个模块网络,将原子概念嵌入到潜在空间上,之后使用图注意力网络(Graph Attention Network, GAT)和 GCN 分别为依赖解析句

和构造的事件概念模式图获得组件语义表示,并将基于图的事件表示与事件匹配层的权重参数联系起来,在推理期间,一个新事件类型的表示可以通过 GNN 推理得到。

基于 GNN 的方法会导致依存句法解析树或解析器解析出来的图结构产生不可逆的误差,另外,随着 GCN 网络层数的增加会产生过平滑和过拟合的问题。因此,未来研究可以探索使得模型性能最佳的网络层数;利用数据增强技术增强输入数据的多样性和随机性,防止过拟合和过平滑问题;利用初始残差和恒等映射技术以及基于节点间的距离来解决过平滑的问题。

4)基于 Transformer 的模型

不同于传统的 CNN, RNN 和 GNN, Transformer 的网络结构主要由多头注意力机制和前馈神经网络组成。最具代表性的是基于 Transformer 编码器构建得到的 BERT 模型,其双层双向的结构可以充分地建模单词的上下文信息,以提升事件抽取的效果。

随着 Transformer 模型的提出,研究者开始探索利用 Transformer 语言模型进行事件抽取任务。Fincke 等^[32]提出了一个语言未知的方法来启动(“priming”)语言模型,并根据模型在运行时被询问的问题来增强 Transformer 堆栈语言模型的输入,很好地弥补了稀疏数据和噪声数据的缺点。

先前跨语言关系和事件抽取任务在单源语言上训练模型会导致单语言偏见的问题,可使用目标语言未标记的数据欺骗一个语言识别器来协助跨语言表示的对齐。但该方法没有以类别信息为条件,一个类的目标语言样本可能会被不正确地与一个不同类的源语言样本对齐。为了解决上述问题,Nguyen 等^[33]提出利用关系和事件抽取(Relation and Event Extraction, REE)任务的类信息进行表示学习。利用 BERT 模型为每一个类学习两个版本的表示向量,相关类的表示向量被对齐来为跨语言的表示获得类意识的对齐,同时为语言共性的单词种类(如 POS 标记和依赖关系)对齐表示向量,该方法由英文迁移到中文时在事件检测和论元抽取的 F1 值分别达到 72.1%和 65.5%。

最近,Wang 等^[34]提出将事件类型和论元角色作为自然语言查询与输入文本连接,使用 BERT 模型和注意力机制来利用查询中丰富的语义信息,该查询抽取范式可以很好地捕获事件类型、论元角色以及输入文本之间的语义关系,同时可以利用来自各种本体可用事件的注释。

基于 Transformer 的模型虽然可以明显提升事件抽取的性能,但其在真实的低资源场景下并不一定适用,原因在于其具有严重的数据依赖和算力资源依赖等问题,而在实际场景下由于资金限制等因素,优质的硬件设备不易获得。另外,Transformer 结构位置信息编码是人为设计的索引,在对词向量做线性变换时位置编码在语义空间中并不具有可变换性,将这种位置编码与词向量相加是不合理的,因此不能很好地表征位置信息。

5)小结

本节对基于深度学习模型的迁移方法进行了简要总结,如表 2 所列,主要从方法类型、方法优缺点以及改进措施等方面进行了分析。

表 2 基于深度学习模型的迁移方法总结

Table 2 Summary of transfer learning methods based on deep learning model

类型	优点	缺点	改进措施
基于 CNN 的模型	可以很好地捕获句子的语法和语义特征, 权重共享策略可以减少网络的训练参数, 降低网络模型的复杂度, 适应性更强	不能很好地捕获远距离单词之间的依赖关系, 管道的执行方式会面临错误传播问题, 很难处理图形数据结构	基于 RNN, GNN 等模型可以有效地解决其面临的问题
基于 RNN 的模型	能够处理变长输入序列, 并建立序列的长期依赖关系	与 CNN 一样, 很难处理图形数据结构, 且不能完全模拟单词之间的依赖关系	基于 GNN 的模型可以解决其面临的问题
基于 GNN 的模型	在非欧几里得空间实现几何深度学习, 可以有效地解决 CNN 和 RNN 面临的难以完全建模单词依赖关系等问题	依存句法解析树或解析器解析出来的图结构产生不可逆的误差, GCN 随着网络层数的增加会产生过平滑和过拟合的问题	探索削减解析时引入误差的策略, 用数据增强技术增强输入数据的多样性和随机性, 防止过拟合和过平滑问题, 或利用初始残差和恒等映射技术以及基于节点间的距离来解决过平滑问题
基于 Transformer 的模型	可以充分地建模单词的上下文信息来提升事件抽取的效果	过于依赖硬件和数据, 在实际场景应用时面临挑战, 位置信息编码做线性变换时不能很好地表征位置信息	基于提示学习的方法可以削减其依赖算力和数据的问题, 需进一步探索策略解决 Transformer 结构本身失去的位置信息

2.1.2 基于预训练语言模型的方法

随着深度学习技术的发展, 许多研究引入 ELMO^[11], GPT^[12] 等预训练模型, 利用微调范式来进行事件抽取任务, BERT 模型集成了两者的优点, 在多项自然语言处理任务中获得了较高的性能。微调与从头训练的本质区别在于模型参数的初始化, 微调无需重新设计网络结构, 只需稍微修改部分层即可。整合或微调预训练语言模型的嵌入已成为深度神经网络的新范式, 可以大大改善目标任务(语言、领域)资源缺乏的情况。

最近, Huang 等^[35] 提出了跨语言生成事件论元抽取模型(Cross-lingual Generative event Argument Extractor, X-GEAR)。X-GEAR 可以有效地编码事件结构并捕获论元之间的依赖关系, 其优势在于设计的语言未知模板, 以兼容任何形式的语言来表示事件论元结构, 通过微调多语言预训练生成语言模型, 从输入的文本片段中抽取论元来填充模板, 最后经过解码得到事件论元和角色的预测, 在由英文迁移到中文时论元分类的 $F1$ 值可达到 54.0%。

Huang 等^[36] 提出以边为条件的图注意力网络 GEANet (Graph Edge-conditioned Attention Net Works), 利用知识图谱中包含概念和语义推理路径的优势, 对不同的事件结构进行推理, 之后通过整合预训练语言模型 SciBERT^[37] 和知识图谱的表示, 来提高生物医学领域小样本事件抽取的效果。

另外, Sheng 等^[38] 提出了一个联合学习框架 SaltyFishes, 该方法可以共享不同事件类型之间的参数, 利用预训练微调框架进行触发词抽取和论元抽取任务, 其引入的条件层正则化机制可以有效地解决元素重叠问题。

尽管预训练微调范式可以减少不同目标类别分布的差异性, 但差异太大时并不一定能取得积极的效果, 在训练阶段使用 MASK 标志会造成预训练和微调之间的不匹配。另外, 预训练微调范式使用的 BERT 模型也继承了 2.1.1 小节中基于 Transformer 模型的缺点, 这里不再具体阐述。

因此, 研究重点可以集中在: 1) 弥合预训练和微调之间的差距, 促使其在下游任务上表现更好; 2) 减少微调大量参数的计算成本。2.2 节将要介绍的提示调优策略可以有效地解决上述问题。

2.1.3 基于机器阅读理解的方法

基于机器阅读理解模型的事件抽取方法, 通过将事件

抽取任务转化为查询问题 Q 形式, 根据给定的文本 C 回答问题 Q , 要求模型输出问题 Q 的答案 A : 即 $A = F(C, Q)$ 。将问答模型用于事件抽取任务, 可以有效地解决角色重合、论元重叠等问题; 同时可以在不改变模型结构的情况下处理各种事件类型; 另外, 机器阅读模型参数量较少, 在只有少量训练样本时收敛较快, 能够缓解过拟合问题。

1) 基于监督学习的 MRC 方法

针对先前事件论元抽取工作依赖于将实体识别作为一个预处理步骤导致错误传播的问题, Du 等^[39] 将事件抽取视为一个问答 (Question Answering, QA) 任务, 以端到端的方式在 Annot Guideline Q 设定下为未见论元角色抽取的 P 和 $F1$ 值分别达到 78.52% 和 67.79%。

针对先前基于分类的方法难以建模标准标签之间的语义以及捕获对事件有用的信息, 且不能泛化到新事件类型上的问题, Li 等^[40] 提出将事件抽取任务视为多轮问答问题 (Event Extraction as Multi-turn Question Answering, MQAEE), 可以捕获同一个事件类型不同论元角色论元之间的依赖关系, 充分利用触发词、事件类型、论元之间的相关联信息, 在论元抽取任务上的 $F1$ 值达到了 53.4%, 比先前最优的 DYGLIE++ 模型提升了 2%, 同时该方法也能很好地识别新事件类型。

此外, Feng 等^[41] 提出将事件检测和论元抽取分别视为一个文本蕴涵 (Textual Entailment, TA) 预测问题和问答 (QA) 问题, 可以有效地提取跟任务有关的丰富知识以及来自预训练语言模型中的标签语义。该方法首次探索了在不依赖触发词信息的前提下在 QA-Guide 设定下进行零样本事件论元抽取, $F1$ 值达到了 26.78%。

2) 基于无监督学习的 MRC 方法

值得注意的是, 上述 MRC 方法都涉及监督学习技术来对事件进行抽取。此外, 研究者也探索了直接利用无监督的 MRC 方式进行事件抽取。

Liu 等^[42] 提出基于无监督问题生成的 MRC 事件抽取方法。与先前使用模板的方法相比, 该方法将事件模式迁移为一组与主题和上下文有关的自然语言问题, 通过在 MRC 中引入 BERT 模型和大规模的数据集, 该学习模式能帮助强化事件抽取的推理过程并缓解数据缺乏的问题。

针对 Liu 等^[42] 提出的方法没有处理论元、触发词之间

长依赖关系以及 Du 等^[39]提出的方法不能很好地解决语法复杂性等问题,Mehta 等^[43]提出通过 MRC 模型本身指导无监督句子简化方法来缓解语法复杂性问题,并且改善在零样本事件抽取中基于 MRC 模型句子的长依赖问题。

之后,Lyu 等^[44]首次尝试将事件检测和论元抽取子任务分别视为文本蕴涵(TE)和问答(QA)查询问题,利用预训练的 TE/QA 模型直接进行迁移,两个子任务在零样本场景下 F1 值分别为 41.7%和 16.8%。实验结果表明,该方法与监督学习方法相比仍有很大差距,当前的 TE/QA 技术难以迁移到一个不同的领域。

总的来说,MRC 方法在事件抽取任务上取得了不错的效果,但也存在一定的缺陷,如针对目标事件类型中的每一个论元角色都需要构造相关查询,MRC 模型的效果依赖于查询的设计,想要取得好的效果,需要设计复杂规则来生成查询。上述方法中也提到研究者采用无监督方式直接进行迁移时其效果并不是很理想。

未来的研究可以集中在整合结构知识(如语义角色标记

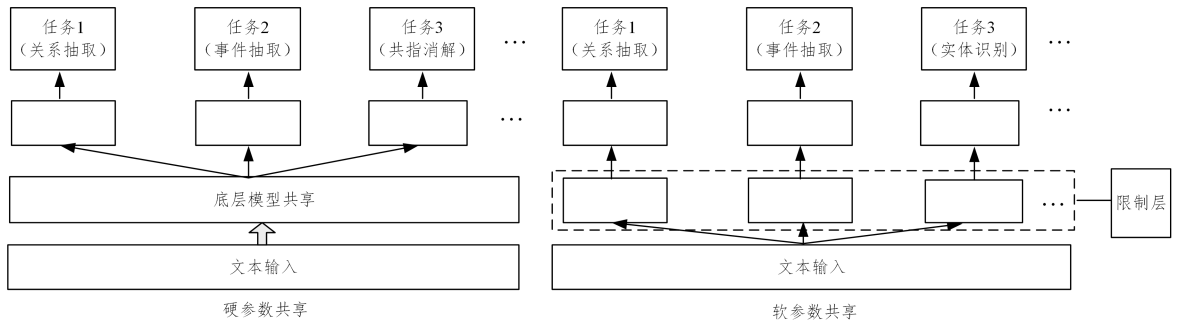


图2 基于多任务学习的事件抽取整体框架

Fig. 2 Overall framework of event extraction based on multi-task learning

针对先前管道的执行方式面临错误传播等问题,Zhang 等^[45]首次设计基于翻译的神经模型来联合预测事件抽取中嵌套的实体、事件触发词及与其有关的语义角色,利用增量方式依照从左到右的阅读顺序来捕获实体和事件提及之间的结构依赖关系。之后 Paolini 等^[46]设计了一个增强自然语言的翻译框架 TANL (Translation Between Augmented Natural Languages),将结构化的预测任务视为翻译任务,通过将输入的文本翻译为增强的自然语言,从而将输出文本编码成结构化信息。该框架不仅可以迁移标签语义知识来提升子任务的预测性能,还可以以统一的方式同时处理多个结构预测任务。

基于混合神经网络的深度学习模型依据其强大的特征学习和共享能力在多任务事件抽取中同样扮演着重要角色。Wu 等^[47]提出了一种混合神经网络模型,将采用双向 LSTM 识别获得的实体上下文信息传递到结合了自注意力的门控卷积神经网络中,以便同时对实体和事件进行抽取,挖掘两者之间的依赖关系。Zhang 等^[48]提出了一个两阶段的 GCN 结构,将实体和触发词的范围特征整合到自注意力机制中,生成以实体-触发词为中心的论元分类模型,该方法在触发词识别 $P, R, F1$ 这 3 个指标上的均值达到了 78.6%,在触发词分类和论元角色分类上的 F1 值分别达到了 75.4%和 55.5%。此外,Wadden 等^[49]提出了一个多任务知识抽取框架 DYGIE++,

和抽象语义表示)到 MRC 模型中,从而更加灵活地生成与上下文有关的查询,揭示并利用事件和论元之间的因果关系上,同时还可以关注如何有效地确保不同查询之间信息的共享,如何解决采用无监督方式直接进行迁移时对性能影响最严重的触发词识别效果差等问题。

2.1.4 基于多任务学习的方法

先前基于管道方式抽取事件的方法,以多阶段的形式执行子任务,并不能捕获子任务之间的依赖关系。多任务学习通过将多个相关任务放在一起学习,分为基于硬约束的多任务学习方法和基于软约束的多任务学习方法,如图 2 所示。前者使得多个任务共享底层模型,每个任务在顶端有自己独特的层来学习高层次的特征。软约束的多任务学习方法是隐藏层参数软共享,不同的任务使用不同的网络,但可以采用正则化约束不同任务的网络参数。与硬约束方法不同的是,软约束方法底层的参数不一定完全一致,而是鼓励参数相似化。多任务学习可以减少参数量,并通过信息共享来相互补充,以实现相互促进,从而提升事件抽取的效果。

通过设计枚举、精炼、评分文本范围来捕获局部(句内)和全局(跨句子)的上下文信息。DYGIE++ 的优势在于利用 BERT 模型提取句子、句间的实体关系特征,利用动态范围图更新上下文编码,确保模型能够整合跨句子的依赖关系。

中文事件抽取方面,He 等^[50]构建了基于 CRF 多任务学习的中文事件抽取联合模型。首先,采用分类训练策略解决联合模型中事件元素的多标签问题。其次,由于处于同一事件大类下的事件子类中的事件元素存在高度的相互关联性,因此提出了采用多任务学习方法对各事件子类进行互增强的联合学习,在中文触发词分类的 P 和 $F1$ 值分别达到了 87.4%和 68.9%,表明该方法可以有效解决中文触发词分类准确率差的问题。

多任务学习中非常重要的一点是如何选择子任务让它们相互促进。如果子任务之间没有关联或者相互影响,特别是对于硬参数共享的多任务学习,很可能导致目标任务效果下降,产生负迁移现象。另外,多任务学习还可能会出现跷跷板现象,对于一些任务相关性比较复杂的场景,在提升一部分任务性能的同时,可能会牺牲其他任务的效果。因此进行多任务迁移时应尽量保证任务直接相关联,避免负迁移现象,而对于出现的跷跷板现象可以重新设计多任务之间的共享机制和单任务的特定网络结构来解决。

2.1.1.5 小结

利用迁移学习技术进行事件抽取时,主要关注的重点是迁移什么、如何迁移、何时迁移。负迁移是迁移学习面临的主要挑战,另外还包括迁移界限等问题。关于迁移学习事件抽取的方法总结如表3所列。负迁移指在进行知识迁移时,目标性能没有提升反而下降的情况。出现负迁移的原因有

很多,如源任务(语言、领域)与目标任务(语言、领域)关联不大、迁移方法使用不当等,未来的研究可以集中在避免负迁移或降低负迁移带来的影响上。另外,量化影响迁移可行性和质量的迁移量也尤为重要,通过评估迁移学习与测量任务相关性的特定界限来估计迁移量,从而最大化迁移效果也是未来关注的重点。

表3 基于迁移学习事件抽取方法总结

Table 3 Summary of event extraction methods based on transfer learning

方法类型	不同点	相同点		
	方法思想	关注重点	方法缺点	改进措施
基于深度学习模型的迁移方法	利用 CNN, RNN, GNN 以及 Transformer 等神经网络结构执行迁移任务	在利用迁移技术指导模型迁移时,应重点关注迁移目标、如何迁移、何时迁移、迁移界限等	源任务(领域、语言)与目标(领域、语言)关联不大时,容易产生负迁移现象	尽可能选取与目标任务(领域、语言)关联性大的源任务(领域、语言)进行迁移
基于预训练语言模型的方法	利用 BERT 等预训练语言模型,通过预训练微调范式执行迁移任务			
基于机器阅读理解的方法	将事件抽取任务建模为问答对执行迁移任务			
基于多任务学习的方法	将多个相关任务放在一起学习,以实现相互促进			

2.2 基于提示学习的方法

简单来说,提示(prompt)学习就是给预训练语言模型一些线索和提示,从而利用预训练语言模型已有的知识让它更好地理解下游任务。prompt-tuning 启发于文本分类任务,并试图将所有的分类任务转换为与掩码语言模型(Masked Language Model, MLM)一致的完形填空的形式。基于提示学习的事件抽取方法启发于基于预训练语言模型的生成式方法^[51-52],之前的工作主要利用人工设计的模版或者自动化搜索的模版来进行预训练,但人工设计的未知性比较影响模型性能且自动化搜索成本较高,之后研究者开始探索基于连续模版的事件抽取方法。

2.2.1 基于离散提示模版的方法

Lu 等^[52]提出的基于 TEXT2EVENT 的模型以及 Paolini 等^[46]提出的基于 TANL 的模型等生成式方法面临的问题如下:1)其在解码阶段整合标签语义,并没有利用编码阶段的信息,如触发词和事件类型的信号;2)当一个简单句子中触发词和论元语法或语义相关性不大时,此类方法会面临挑战;3)TANL 生成与任务无关的 Token 比例较大以及 TEXT2EVENT 结构太复杂难以扩展等。Si 等^[53]首次将基于 prompt 学习策略引入事件抽取中,提出利用 prompts 生成解耦论元(Generates Disentangled Arguments with Prompts, GDAP)的方法,在对文本编码时通过各种标签语义有效地注入知识。GDAP 将触发词和论元抽取解开的机制极大地增强了计算并行性,同时解决了重叠问题。

Hsu 等^[54]提出了一个基于高效数据生成的事件抽取框架 DEGREE(Data-efficient Generation-based Event Extraction),将一个文本片段和手动设计的提示模板作为输入,并遵循模板设计规则首先学习将段落总结为自然句子形式,之后使用一个确定算法来从生成句子中抽取事件触发词和论元。

顺序导致的子优化问题,以及 Hsu 等^[54]、Li 等^[51]提出的方法中一个论元角色有多个提及导致的错误传播问题,Lin 等^[55]提出了一种基于提示的知识引发方法(Prompt-based Knowledge Eliciting Approach, PoKE),采用单个论元提示策略和联合论元提示策略增强对事件论元的抽取效果,在 PoKE (large)模型下其 P 和 $F1$ 值分别达到 71.91% 和 69.19%。文章还提出了一个问题意识的序列标记方法,用于获得一个论元角色的所有论元范围,当文本中出现多个论元提及时该方法很有效。

2.2.2 基于连续提示模版的方法

Li 等^[51]、Hsu 等^[54]的方法也面临两个挑战,首先,为每一个事件类型手工设计的离散 prompts 若没有调整则会导致子优化问题。其次,当抽取一个特定类型事件时,这些方法忽视了与其他事件类型之间的联系。为了解决上述问题,Liu 等^[56]提出了基于动态前缀的模板生成事件抽取方法(Generative Template-based Event Extraction Method with Dynamic Prefix, GTEE-DYNPREF),与先前设计的离散 prompts 不同,GTEE-DYNPREF 通过整合所有特定类型前缀的上下文信息来学习一个特定上下文的前缀,形成了连续的 prompts,每一个特定类型的前缀提供了单个可调事件类型的信息,可以动态地连接所有可能相关事件类型的信息,指导预训练语言模型对新事件类型的抽取。

类似于当前 Liu 等^[57]提出的 p-tuning v2 工作, Ma 等^[58]提出了一个文档级事件论元抽取模型 PAIE(Prompting Argument Interaction for Event Argument Extraction),通过利用基于 prompt 的两个范围选择器为每一个角色从输入文本中选择开始/结束的 token,并通过多角色 prompts 来捕获论元之间的相互关系,利用双方匹配损失函数以最优范围分配为目标执行联合优化。另外,PAIE 代替了传统启发式阈值调优方法,可以同时抽取一个角色的多个论元。

针对 Li 等^[40]提出将事件抽取任务视为多轮问答时问题

基于提示学习的事件抽取方法总结如表4所列,手工

设计的提示模板依赖于领域/语言的见解且耗时耗力,而连续提示优化模板也需要标记数据优化参数且容易产生复杂的标签空间,自动搜索的提示模板缺乏可解释性且具有脆弱性,当训练数据不平衡时也会面临困境。未来在经济成本和环境因素限定的前提下,可以探索利用大量无标记数据以及注入更加

复杂的知识(如知识图谱中的符号规则)来构造更强大的提示模板,另外已有研究者在其他相关任务上探索了基于自动搜索的零样本多任务提示预训练方法^[59]、基于预训练的软提示嵌入初始化方法^[60]以及基于迁移学习的软提示调优方法^[61],未来可以考虑将这些方法应用到事件抽取任务上。

表 4 基于提示学习事件抽取方法的总结

Table 4 Summary of event extraction methods based on prompt learning

方法类型	方法思想	方法缺点	改进措施
基于离散提示模板的方法	采用人工构造或自动搜索的离散模板指导预训练语言模型	手工构造模板耗时耗力且质量不高,自动搜索的离散模板缺乏可解释性;训练数据不平衡时会面临困境	设计连续型提示模板,通过梯度优化获得最优的模板,未来可探索基于可解释的提示模板方法;并设计策略解决提示模板依赖训练数据的问题
基于连续提示模板的方法	设计连续可微的提示模板来获得最优的模板,从而指导预训练语言模型	需要标记数据优化参数且容易产生复杂的标签空间,自动搜索的模板同样缺乏可解释性	未来可探索不依赖数据且不易产生复杂标签空间以及具备可解释性的模板

2.3 基于无监督学习的方法

无监督学习不训练基于标记语料库的事件抽取模型,主要集中在开放域事件抽取的研究上,同时也包括一些在常规数据集上经典的无监督学习事件抽取方法。采用无监督学习技术对事件抽取的研究主要集中在事件模式归纳上,其方法体系大概可以分为四大类:基于表示学习及聚类的方法、基于概率图模型的方法、基于模式自动归纳与映射的方法以及基于对抗与对比训练的方法。

2.3.1 基于表示学习及聚类的方法

最近的研究表明,可以从原始的文本中自动地归纳事件模式。研究者开始探索概念生成方法^[62-64]来联合建模谓词和论元分配任务,以及专门的聚类算法^[65-67]来归纳事件模式。

Sha 等^[65]提出了基于实体表示的事件模式归纳方法,利用实体在高维空间上表示的距离来衡量两个实体是否属于同一个模板/论元角色。文献^[64]将实体之间的距离视为论元角色和模板级别的距离,使用实体词向量余弦距离和 PMI (Point Wise Mutual Information) 测量实体在模板级别的距离,利用词向量余弦距离、上位词相似度以及所依赖谓词的相似度测量实体在论元角色级别的距离,使用图像分割中的标准更精确地对实体进行分类,并利用模板和论元角色约束规则指导聚类过程。

Huang 等^[66]提出了一个自由事件抽取和模式归纳框架,其结合了抽象语义和分布式语义表示来检测和表示事件结构,采用同一类型框架同时抽取事件类型和论元角色并发现事件模式,将分布相似性作为聚类的距离度量,试图将事件触发词和事件论元聚类。

Yuan 等^[67]提出利用无参贝叶斯模型得到新闻文章中实体槽表达式信息来抽取事件,同时使用实体共现信息学习论元角色模式的嵌入表示,之后对论元角色表示聚类,将每一个聚类簇作为事件模式的论元角色,最后选择相应论元角色构造事件模式。

2.3.2 基于概率图模型的方法

上述方法大都依赖于手工创建的指标特征,没有充分

利用文本丰富性来进行事件模式归纳。基于概率图模型的事件模式归纳方法一般利用主题模型或神经隐变量模型建模事件模式的隐含结构,从而更好地利用文本冗余特征。

Liu 等^[68]设计了一种新颖的无监督图模型,由神经网络产生的潜在变量来获得更好的表示能力,以便从新闻集群中抽取无约束事件类型。该模型具有潜在的事件类型矢量以及实体的文本冗余特征,可以从新闻聚类中抽取多个事件,在开放领域中事件可以有不同数量的论元角色槽(slots),另外该方法也考虑到了在一个新闻类中,来自于不同报道的同一个实体的提及信息。

Zhou 等^[69]提出了一个无监督贝叶斯潜变量模型 LECM (Latent Event & Category Model),用于从推特中抽取并聚类事件实例,首先利用基于词典的方法过滤掉不相关的推特事件,之后基于与事件有关的推特抽取事件结构表示,在没有使用任何标签数据的情况下,利用无监督贝叶斯模型对其分类,最后给分类的事件分配事件类型标签。但是该方法面临两个问题:1)在真实的应用场景下无法预先知道要抽取的事件数量;2)不能识别在多个提及中出现的同一个命名实体,在使用这些提及会导致错误分配事件。之后 Zhou 等^[70]提出了一个无参贝叶斯狄利克雷混合模型,利用词嵌入来进行事件抽取,此时事件的数量可以被自动地推断且通过整合词嵌入来处理多个提及指向同一个命名实体的问题,其在 FSD 数据集上的 R 和 $F1$ 值分别达到了 90% 和 88.06%。

2.3.3 基于模式自动归纳与映射的方法

上述基于概率图模型的方法归纳的事件模式难以从新文本中抽取事件论元,开放域文本大都依赖于上下文信息,词汇共现信息的缺乏导致很难有效地训练主题模型。另外,基于表示学习的聚类方法仅以表示触发词为目的,无法给出文本主题信息。针对上述问题,Hao^[71]提出了一种基于模式自动归纳的开放域事件抽取方法,该方法包括一个融合主题模型和预训练模型的事件类型归纳方法,利用神经变分推理主题模型得到候选事件触发词所在事件提及的隐含事件类型分布表示,利用预训练语言模型对候选事件触发词的上下文进行表示,基于两种模型的融合表示进行事件类型归纳。同时,

文献[71]还提出利用基于图表示学习的事件论元角色归纳方法和论元角色表达式匹配的方法进行事件论元的抽取。

而基于映射的无监督技术不依赖任何标注数据,可以直接进行事件类型的识别^[20,72]。Zhang 等^[72]提出利用预训练语言模型得到标签语义,并计算表示相似性将识别的事件映射到目标类型上,同时使用事件定义作为全局约束来正则化预测的结果,在零样本触发词识别和分类上的 F1 值分别为 58.3% 和 53.6%。

2.3.4 基于对抗与对比训练的方法

基于对抗训练的方法主要利用生成对抗网络(Generative Adversarial Nets, GAN),通过对抗训练的方式来使得生成网络产生的样本服从真实数据的分布。如 Wang 等^[73]提出了一个基于生成对抗网络的事件抽取模型 AEM(Adversarial-neural Event Model)。AEM 使用狄利克雷先验建模事件并使用生成网络捕获潜在事件的模式,利用识别器识别由潜在事件重构的文档和原始文档。另外,由识别器学习到的特征可以对抽取的事件进行可视化。

而基于对比训练方法所遵循的原则是为“邻居”学习相似特征,为“非邻居”学习不同特征。如 Wang 等^[74]提出了一个对比预训练框架,来充分利用大规模无监督数据的丰富事件

知识和语义结构。具体地,使用一个文本编码器,利用自监督对比学习,通过考虑事件和单词之间的相似性来学习事件的语义表示。采用一个图编码器,通过在一个解析的事件相关的语义结构上进行图对比预训练,来学习事件的结构表示。上述得到的两个预训练表示可以在没有任何已知事件模式和注释实例的情况下直接抽取事件并发现新的事件图式,其在事件检测和论元抽取任务上的 F1 值分别达到了 53.7% 和 35.1%,比上述 Huang 等^[66]提出的 LiberalEE 方法分别提升了 3.9% 和 4.5%。

总的来说,无监督学习技术可以有效地克服事件抽取任务缺乏事件类型、事件模板与标注数据的困难,但抽取的事件通常也会面临精确度和召回率低等问题。本节最后对基于无监督学习事件抽取方法进行了总结对比,如表 5 所列。未来研究可以集中在以下几点:1)选取最优的评估度量指标来提升聚类结果的准确性;2)设计更鲁棒的模型、规则或引入关键因素(情感信息、因果关系等)来充分建模文本的信息;3)减少大规模非结构化无标注数据中的脏标记和噪声样本,以及处理文本过程中相关技术引入噪声(如 AMR(Abstract Meaning Representation)解析器引入噪声)对事件抽取结果有着较大的影响。

表 5 基于无监督学习事件抽取方法对比

Table 5 Comparison of event extraction methods based on unsupervised learning

方法类型	方法思想	方法缺点	最终目标
基于表示学习及聚类的方法	检测并表示关键词或命名实体,采用聚类策略归纳事件模式	大都依赖于手工创建的指标特征,仅以表示为目的,无法给出文本主题信息,没有充分利用文本丰富性	归纳事件模式
基于概率图模型的方法	利用主题模型或神经隐变量模型建模事件模式的隐含结构,更好地利用文本冗余特征	难以从新文本中抽取事件论元,开放域文本大都依赖于上下文信息,词汇共现信息缺乏,很难有效训练主题模型	归纳事件模式
基于模式自动归纳与映射的方法	融合主题模型和预训练模型的事件类型归纳方法,可自动从大量文本语料中发现事件模式和类型;或采用映射方法识别事件	抽取出的事件不能直接归类到已有的知识库体系中	归纳事件模式/映射
基于对抗与对比训练的方法	对抗训练:利用对抗训练策略捕获潜在事件模式 对比训练:采用对比预训练框架,充分利用大规模无监督数据的丰富事件知识和语义结构	不可避免地会引入噪声	归纳事件模式

2.4 基于弱监督学习的方法

弱监督学习事件抽取方法旨在利用结构化知识库或少量人工标注数据,自动生成大规模、高质量的标注数据。早期的弱监督方法更倾向于模板迭代的方式,只需要人工对语料进行一定的预分类或利用部分种子模板,让机器根据已有的预分类语料或种子模板进行模式学习,利用新生成的标注数据和原始的标注数据进行模型的训练^[75-77]。本节重点讨论了典型的半监督、远程监督等弱监督学习事件抽取方法。

2.4.1 基于半监督学习的方法

半监督学习是一种典型的弱监督学习方法,其中最具有代表性的方法是模板迭代的自训练方法,但基于模板迭代扩展训练数据的方法通常集中在与标注数据事件类型一致的情形上,当面临新事件类型时,其性能会大大下降。因此,通过设计合适的策略来应对新事件类型也尤为重要,此外还有研究者提出采用聚类策略进行数据扩展的方法。

1) 基于模板迭代的自训练方法

先前弱监督方法依赖于复杂的预定义规则和知识库中的已有实例来自动注释数据,通常会面临低覆盖、主题偏见、数据噪声等问题。针对上述问题,Wang 等^[78]认为如果在标注数据集中一个单词是一个触发词,那么在未标记的数据中包含这个单词的句子都可能表达这样的事件。利用上述思想,通过对抗训练让模型具备更好的降噪能力之后,就能更好地识别那些不是噪声、能够提供有用信息的新数据进行模型的训练,在事件检测上的 P 和 F1 值分别为 77.9%,75.1%,相比在原始数据集上训练的基线模型,F1 值提升了 0.5%。

从模型角度来看,现有的论元抽取方法面临共享参数不充分以及没有考虑角色的语义问题。从数据的角度来看,现有的方法集中在严重依赖额外资源的数据生成和数据增强方法上,但是这些资源通常是不完整的并且构建耗时耗力。Zhou 等^[79]设计了一个半监督的框架 DualQA(Dual Question

Answering),该框架通过将事件论元识别和论元角色识别两个任务视为问答问题,使得可以尽可能地共享更多的参数并利用角色的语义信息,同时也解决了传统半监督方法错误传播的问题。

Wang等^[80]提出了一个包含语法和语义特征向量的集合,并通过事件特征耦合泛化(Event Feature Coupling Generalization, EFCG)算法,利用来自标记数据的两类原始特征的相关性和大量未标记数据相关性生成新的特征,从而解决了生物医学事件抽取面临的数据稀缺问题。

2) 基于数据聚类的自训练方法

先前的半监督学习方法大都采用模板迭代的方式,最近有学者提出基于数据聚类的自训练策略。Ferguson等^[81]提出对已有的一些文章进行聚类,然后对聚类结果打标签并划分到各个已有的类上。该方法同时考虑了时间和实体信息,之后训练一个简单分类器对这些聚类结果进行分类,对未标注的数据进行标注,并根据已有的触发词集合从句子中选取事件触发词,得到完整的事件抽取标注数据,与已标注的数据一起联合训练,该方法在触发词识别上的F1值相比先前最优的基线模型提升了1.3%。

3) 基于新事件类型泛化的方法

先前扩展训练数据的方法主要集中在与标准标记数据事件类型一致性上。最近,有研究集中在扩展与标准注释数据事件类型不一致的新事件类型的训练数据上^[82-84]。对于新的未见类型,通常的方法是付出大量的努力注释新类型的数据,并且重新训练模型。但是这类方法阻碍了事件检测系统在新事件类型上的应用,因此急需提出可以在新事件类型上扩展的模型来解决事件检测问题。

Huang等^[84]提出了一个半监督向量量化变分自动编码器(Semi-supervised Vector Quantized Variational Autoencoder, SS-VQ-VAE)框架,自动地为可见类型和未见类型学习一个离散的潜在变量表示,并且使用可见类型的事件注释来优化它们。为了防止类型的预测对可见类型过拟合,文献^[84]引入一个变分自动编码器,以每一个事件提及的潜在类型分布为条件,对该事件提及进行重构,最终可以为每一个候选的触发词预测其类型。

总的来说,选取高质量初始种子是半监督学习方法的关键,另外,如何降低迭代过程中的噪声问题也是研究者面临的难题,具体的阐述见本节的小结部分。

2.4.2 基于远程监督学习的方法

由于手工标记数据耗时耗力,事件类型的覆盖很小,因此,远程监督(Distant Supervision, DS)技术被提出,其最早被Mintz等^[85]用在通用领域关系抽取中。基于远程监督的事件抽取方法类似于关系抽取中采用的远程监督策略,旨在利用结构化知识自动标注数据来扩充用于事件抽取的标记数据规模。因为有知识库、外部词典等结构化数据可以获得,所以对命名实体识别和关系抽取任务来说该方法很流行。但当远程监督技术应用到事件抽取中时会面临以下挑战:1)已存在的知识库没有给出触发词,因此不能直接用事件类型和论元来进行回标,在用DS前,必须先检测触发词;2)一个事件中的论元可能出现在多个句子中,如果用所有论元来进行句子的回标,那么能抽出的训练数据就非常少。

对于句子级事件抽取,为了解决上述两个关键问题,Chen等^[86]提出通过联合使用领域知识Freebase^[87]和语言学知识FrameNet^[88]优先考虑选择关键论元进行回标。首先使用Freebase选择出关键论元标记事件并找出触发词,然后使用FrameNet来过滤噪声触发词并扩展更多的触发词,最后使用软远程监督(Soft Distant Supervision, SDS)技术来自动地标记训练数据进行事件抽取。

对于文档级事件抽取,为了解决标记数据缺乏以及一个事件分布在文档中的多个句子中的问题,Yang等^[89]提出了一个文档级中文金融事件检测系统,首先对公告中文本数据的触发词构造一个预定义的字典,以便将金融事件知识库中的结构知识映射到公告中的文本数据上自动识别事件提及。然后将文档中事件提及标记为正样本并将其作为句子级别的数据,其他句子标记为负样本作为文档数据。最后将以上两种级别的数据整合起来用于事件抽取模型的训练。

对于开放域事件抽取,其人工标注的成本极大,通常面临没有人工标注的数据供监督学习方法学习的问题。Araki等^[90]利用WordNet^[91]和Wikipedia^[92]来生成新的训练数据,提出了一种基于DS自动生成训练数据的方法,其利用WordNet进行监督,不依赖于其他人工标注的数据集,从而生成高质量的事件二分类训练数据,并从简单的英文Wikipedia中选取文章进行事件标注,实验结果表明采用DS方法后,模型的迁移性能更加稳定。

以上方法存在明显的问题,即所有远程监督方法都不可避免地会引入数据噪声,影响模型的表现,因此,降噪是DS场景抽取的一个最核心的能力。此外,在筛选和检测中,无论是复杂的规则或类似TF-IDF的算分方法,都会导致低频但有信息量的实例被筛选掉,使得结果越来越偏向最高频的数据,而这种数据往往无益于训练模型。基于上述挑战,Wang等^[78]提出了一种对抗训练的方法,用于改善采用DS方式时事件抽取的效果,该方法能在无监督情况下学习降低数据噪声、抵抗数据噪声的能力。

远程监督事件抽取方法主要面临的难题是噪声数据与抽取特征误差传播的问题,利用知识库可能会错误地标记数据,从而产生噪声数据,而自然语言处理工具抽取的特征也存在误差,会引起特征的误差传播与错误积累。如何解决上述问题是未来研究的重点。

2.4.3 小结

总的来说,目前弱监督事件抽取方法的性能与监督学习方法仍然有一定的差距。本节对基于半监督与远程监督的弱监督学习方法进行总结对比,结果如表6所列。类似于无监督学习方法,弱监督学习方法不可避免地会受到噪声数据的影响,如果利用的无标签样本的质量较差或是完全不同分布的数据,反而很容易降低模型的泛化性能。如何降噪是弱监督学习研究的核心问题,目前已有研究者利用对抗训练等方法进行降噪处理,未来可以探索更好的降噪策略来提升弱监督学习事件抽取的效果。一个值得探索的方向是引入强化学学习设定优化目标,利用强化学学习建模伪平行语料的质量评测指标,设计相应的训练过程如质量评估(Quality-estimation, QE),在此基础上训练得到质量更好的伪平行语料。

表6 基于半监督与远程监督学习事件抽取方法对比

Table 6 Comparison of event extraction methods based on semi-supervised and distant supervision learning

方法类型	相同点	不同点
基于半监督学习的方法	1. 两者都属于弱监督学习的范畴,通过利用少部分高质量标注数据,生成大量新的标注数据,从而扩展训练数据的规模用于模型的训练	半监督学习方法主要利用一小部分种子模式或预分类的语料,采用模板迭代、数据聚类等自训练方式对大规模无标注数据进行标注,来扩展训练数据的规模,或设计鲁棒的模型来识别事件
基于远程监督学习的方法	2. 两者都不可避免地会引入噪声数据 3. 两者的核心问题都是降噪	不同于半监督学习扩展训练数据的方式,远程监督学习方法主要利用知识库、外部词典等结构化知识自动标注数据,从而扩展训练数据的规模

2.5 基于数据与辅助知识增强的方法

基于数据与辅助知识增强的方法主要利用高资源、高质量的数据和知识来缓解目标数据稀疏的情况,具体的分析见各自模块。

2.5.1 基于多语言增强的方法

先前事件抽取的方法大都集中在单语言线索上,忽视了其他语言提供的大量有用信息。多语言增强的本质是利用现有大量的高质量、高资源的标注语料(如英文)来获得更丰富和鲁棒的样本表示,从而改善低资源语言数据缺乏、质量低下等问题。早期研究者主要利用模型引导、机器翻译、嵌入映射等方法进行跨语言事件抽取,之后研究者开始利用深度学习模型进行跨语言事件抽取。

1) 基于机器翻译与映射的方法

早期跨语言事件抽取的方法主要利用模式引导(Bootstrapping)的思想,Chen等^[93]提出了一个基于跨语言信息映射的协同引导框架,基于中英文单语事件抽取系统和并行语料,构建跨语言同义谓词集合的方法,用于辅助进行中文事件抽取。Zhu等^[94]提出了一个基于机器翻译的双语言分类方法,使用双语言特征生成统一的文本表示,通过合并两种语言中的训练样本来扩展标记数据的尺寸,并从单语言和双语言两个角度来解决触发词定位困难的问题,该方法可以有效地解决中文事件抽取数据稀缺的问题。Hsi等^[95-96]提出嵌入映射的方法,结合语言依赖和语言特有的特征,利用多语言数据训练模型,在中文触发词抽取上 Macro-Average P 和 $F1$ 值分别为 44.3% 和 25.8%,相比使用中文单语言训练分别提升了 2.2% 和 2.5%。

2) 基于深度学习模型的方法

上述跨语言事件抽取方法严重依赖于用于映射的机器模型,并且需要对齐映射单词以及大量的并行资源,大大限制了其适用性。随着深度学习技术的发展,研究者开始探索利用深度学习模型进行跨语言事件抽取。

Liu等^[97]提出了一种上下文依赖的翻译方法,用于构造不同语言之间的词汇映射。该方法替换了传统的单词到单词的翻译形式,采用一个上下文有关的选择注意力机制,对不同的翻译候选对进行排名,来选取最接近的翻译对。另外,文献^[97]设计了一个共享语法顺序的事件检测器,用于处理翻译中单词顺序的差异化问题,从而可以进行多语言的协同训练,在中文事件检测上的 $F1$ 值可达 45.4%,比之前的 LexMap^[96]方法提升了 5.8%。

近几年,随着 BERT 模型问世,Hambardzumyan等^[98]通过使用一个注意力机制整合句子 Token 的 mBERT 嵌入来

获得该句子的一个表示,之后在句子表示上方使用一个线性层尝试将 mBERT 的嵌入映射到一个共享空间上,使得在不同语言上相似的句子在该共享空间上具有相同的位置,从而将不同语言句子的表示对齐,可以有效地提升跨语言事件抽取的效果。

总的来说,多语言增强方法可以有效地提升对低资源语言事件抽取的效果。构建多语言统一模型能大大降低在新语言上冷启动的成本,相比从零学习,该方法可以减小对标注数据的依赖,跨语言迁移的效果最终还是依赖语言之间内在的联系,未来可以集中在设计更加强大通用的多语言模型上,减少对语言关系的依赖并结合高资源语言数据进行跨语言学习,从而解决目标语言面临的数据缺乏难题。

2.5.2 基于多模态增强的方法

利用多模态数据(文本、图像、语音、视频等)进行增强的方法,其思想很直观,即通过引入其他模态的数据来缓解当前数据可用信息不足的情况。低资源场景下的多模态学习指利用多个模态的互补信息学习更好的特征表示,丰富文本的可用特征,消除模态之间的冗余信息,通过融合多模态信息来提升模型对小样本数据的感知能力,利用多模态数据来改善事件抽取的效果。

其中一种最直接的利用多模态进行数据增强的方式就是利用文本本身的多模态数据。Zhu等^[99]提出了一个小样本事件检测模型 CADA-fVAE-GAN,通过从事件提及、类型结构、类型名字 3 个模态中抽取语义特征,利用对抗训练增强的变分自动编码器(Variational Autoencoder, VAE)将特定模态对齐并映射到一个共享的低维潜在空间上,并通过训练一个 CNN 分类器来评估潜在表示的质量。

图片的视觉特征可以有效地改善文本特征信息缺乏的情况。Zhang等^[100]提出将外部可视化存储库的视觉特征与传统的文本相对应,首先以弱监督的方式从大规模的文本-图片中对挖掘视觉模式,之后提出一个多模态事件抽取算法,利用文本特征和视觉模式联合训练事件抽取器,从而优化事件抽取的效果。Tong等^[101]提出了一个双重循环多模态模型(Dual Recurrent Multi Modal Model, DRMM),利用交替的双重注意力机制在每一步选择有用的文本信息并过滤掉不相关的噪声用于特征的抽取,同时能有效地捕获图片与句子之间更深层的相互关系,用于模态特征的整合,从而提升事件抽取的效果。Li等^[102]提出了一种弱对齐结构嵌入方法(Weakly Aligned Structured Embedding, WASE),将文本和视觉数据语义信息的结构表示编码到通用的嵌入空间上,采用弱监督

训练策略,利用已有的单模态注释语料库,在没有跨模态注释的情况下执行联合推理。

但通过一个单独的快照图片特征难以确定一些关键的事件论元角色,而视频包含关于事件的更多信息,有助于消除在事件中不同论元扮演的角色歧义问题。Chen 等^[103]提出联合文本和视频,从中抽取多媒体事件和论元,首先执行多模态的共指消解,即利用自监督训练策略来识别与同一个事件有关的句子和视频片段,之后引入一个多模态 Transformer 结构,利用特定模式的解码器联合进行文本和视频事件抽取。Wu 等^[104]提出了一个零样本学习的通用框架,利用多个多模态特征包括视觉和音频特征,将一个视频映射到一个中间语义属性空间上,之后将从大规模文本语料库中学到的统计数据映射到高维的概念空间上,在此空间上计算属性和文本查询概念之间的相似性,将不同属性集合计算得到的分数连接获得最终的融合分数,该方法可以有效地对齐文本查询与不同模态之间不相交的词汇,从而提升零样本事件检测的效果。

目前基于多模态增强的事件抽取技术还不太成熟,仍然存在许多需要改进的地方,具体阐述见第 4 章。

2.5.3 基于辅助知识增强的方法

考虑到 ACE 语料库定义了非常有限和粗粒度的事件模式(8 个类型和 33 个子类型),当实际面临大规模无标注的新事件类型时,将会错过大量有用的事件信息。由于 Framet-Net^[88]中的框架与 ACE 语料库中的事件模式在结构上是高度一致的,研究者开始利用 FrameNet 等框架语义知识来增强当前标记数据缺乏的样本,另外还有研究者探索基于开放域以及基于本体和规则的知识增强方法。

1) 基于框架语义与开放域知识增强的方法

Liu 等^[105]首先利用 ACE 语料训练的分类器去判定 FrameNet 中句子的事件类别,再利用全局推断将 FrameNet 的语义框架和 ACE 中的事件类别进行映射,进而利用 FrameNet 中人工标注的事件样例扩展训练数据以提升事件检测性能。后来,Li 等^[106]提出了基于 FrameNet 重定义事件模式的方法。首先从 FrameNet 中抽取表达事件信息的框架,并利用框架到框架的关系建立一个比 ACE 语料库粒度更细、覆盖更广的分层事件模式,基于新定义的事件模式,利用 FrameNet 中事件模式的层次结构和框架到框架的关系,基于马尔可夫逻辑网(Markov Logic Network,MLN)联合抽取事件。

还有研究者利用开放域丰富的知识来提供额外的信息,

从而缓解标注数据缺乏的情况。Tong 等^[107]提出利用开放域触发词知识来处理事件抽取中的长尾问题,通过给未见的或稀疏的标记触发词提供额外的语义支持,在未见和稀疏标记场景下触发词识别的 P 值分别达到 79.0% 和 80.8%,这表明该方法可以有效提升触发词识别的准确率并减少对注释中频繁出现触发词的内在偏见。

2) 基于本体和规则增强的方法

针对大多数事件检测方法严重依赖训练实例,忽略了事件类型的相关性,会面临数据缺乏且无法处理新事件类型的问题,Deng 等^[108]提出了基于本体嵌入的事件检测框架 OntoED(Event Detection Framework with Ontology Embedding)。基于事件本体,OntoED 可以利用数据丰富的事件类型知识,并将其传播到数据稀缺的事件类型上,使得本体可以很好地应对新事件类型。

先前数据驱动的方法大都忽视了事件之间内在的相关联知识,导致对不常见的事件类型和角色识别的性能相对较差。Deng 等^[109]提出了一个知识意识的超球面原型网络 K-HPN(Knowledge-aware Hyperspherical Prototype Network),通过在超球面上进行成对原型学习,利用事件之间因果关系的相关知识来将类中高资源(源)的知识迁移到低(目标)资源上,从而减少对数据的依赖。Li 等^[110]提出利用一个文档级的图注意力网络,通过连接事件类型的结点来建立类型信息的共享,构造一个通用的模式来统一所有事件类型的角色模式,从而实现角色之间知识的共享,并进一步提出一个参数继承机制,用于考虑事件类型的知识协助类型偏好角色的识别。该方法在事件类型分类、触发词识别上的 $F1$ 值分别高达 82.37% 和 86.71%,比 Du 等^[39]、Li 等^[40]提出的模型在事件类型分类上的 $F1$ 值分别提升了 9.98% 和 10.67%。

总的来说,基于辅助知识增强的方法主要利用内生或外生的知识更加准确地学习样本的语义表示,从而提升事件抽取的效果。引入外部知识带着先验知识进行推理是一个很好的任务增强方式,引入外部知识时,需要避免引入太多的噪声数据。另外,还可以设计更有效的策略、规则等挖掘更多事件之间的潜在知识(如时序关系)来丰富事件的表示。

2.5.4 小结

本节对基于数据与辅助知识增强的事件抽取方法进行了总结,如表 7 所列,主要从方法思想、所用数据/知识类型以及方法缺点等方面进行分析。

表 7 基于数据与辅助知识增强的事件抽取方法总结

Table 7 Summary of event extraction methods based on data and auxiliary knowledge enhancement

方法类型	方法思想	所用数据/知识类型	方法缺点
基于多语言增强的方法	利用高质量、高资源的标注语料(如英文)改善低资源语言数据缺乏、质量低下等问题	高质量、高资源语言数据	依赖语言之间的内在联系,依赖源语言的质量和资源
基于多模态增强的方法	利用多模态数据(文本、图像、语音、视频等)进行增强,通过引入其他模态的数据来缓解当前数据可用信息不足的情况	多模态数据	难以联合学习异源领域的的数据,难以处理在连接不同模态期间不同程度的噪声
基于辅助知识增强的方法	采用框架语义、开放域知识以及利用本体和规则等方式更加准确地学习样本的语义表示	框架语义、开放域、本体以及内生知识	引入外部知识时容易产生噪声数据

2.6 基于元学习的方法

基于元学习的事件抽取方法将不同的事件类型视为不同的任务(下文以“任务”进行叙述)。由于预训练模型在面对新任务时更新了参数,因此原先在旧任务上训练好的参数被新的信息覆盖,容易产生灾难性遗忘的问题。元学习^[111-112]从数据

和任务的双重抽样中学习历史先验知识,然后提取元知识应用于未来的任务。元学习集中在使用先验知识来提升对于未见任务的泛化能力,如果先验知识被用来教模型如何在一个特定任务上学习,那么元学习就可以被视为小样本学习的一个变体。基于元学习的事件抽取方法的整体框架如图3所示。

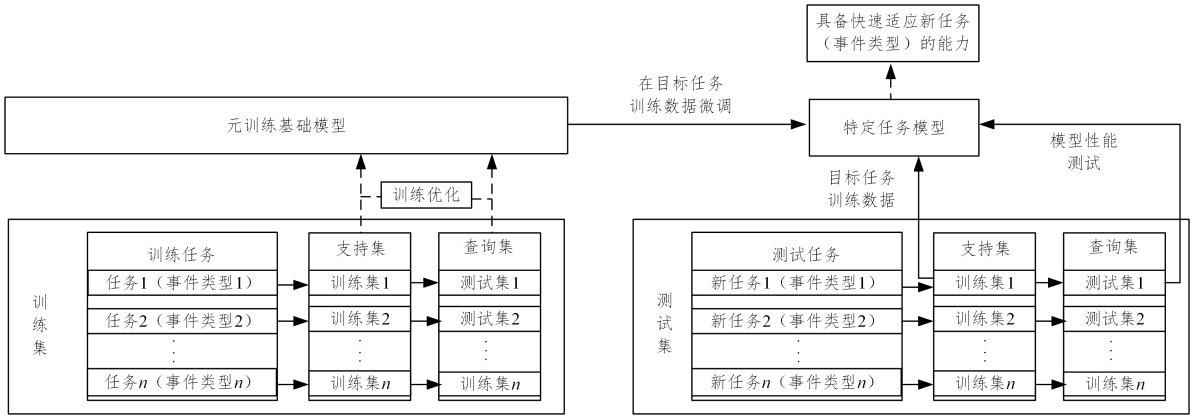


图3 基于元学习的事件抽取框架

Fig. 3 Overall framework of event extraction based on meta learning

先前的小样本学习方法面临的问题是模型仅依赖于查询实例和支持集合之间的训练信号,支持集中样本之间的匹配信息尚未得到充分利用。Lai等^[113]首次提出利用支持集中匹配样本的两个损失因素来给模型提供更多的训练信号,这些训练信号可以显著地提升小样本事件检测的性能。

另外,利用动态原型网络进行元学习事件抽取也取得了显著成效。Yang等^[114]提出使用一个动态存储网络来编码特定领域的信息,从而可以更好地学习同一个事件领域任务之间的结构相似性,增强模型未知的元学习算法(Model-agnostic Meta-learning, MAML)^[111]对新事件类型的适应性和事件论元的抽取效果。Deng等^[115]提出利用动态存储原型网络(Dynamic-memory-based Prototypical Network, DMB-PN)中的动态存储网络(Dynamic Memory Network, DMN)多次使用事件提及编码来为事件学习更好的原型,DMN在句子编码方面更健壮,可以充分利用事件实例并进行记忆存储和更新,在5-Way-10-Shot设定下事件分类的F1值可达73.99%。

上述方法支持集中的样本多样性通常被限制,从而导致事件检测的性能并不理想。为了解决上述问题,Shen等^[116]

提出一种自适应贝叶斯增强元学习框架(Adaptive Knowledge-enhanced Bayesian Meta-learning, AKE-BML),通过采用启发式规则对齐支持集合与FrameNet中的事件类型,并利用编码器将样本与知识库编码在一个语义空间上,同时提出了一个可学习的偏移量来修正对齐的知识表示,从而建立事件类型知识的先验分布,生成事件类型原型表示的后验分布。在预测阶段采用学习到的原型表示的后验分布对查询实例分配相应的事件类型,从而解决了采用知识增强方法时知识库覆盖不完整以及预定义知识库与不同应用不能准确对齐的问题。

类似于迁移学习,基于元学习的事件抽取方法对训练任务和测试任务的相关性也有要求,当训练任务和测试任务存在明显差距时,很少使用元学习来初始化参数权重,很容易导致模型的负迁移。本节最后对元学习与迁移学习事件抽取方法进行了对比,结果如表8所列。当前元学习的思想基础是要有充足的历史任务,若特定问题任务缺乏时,元学习技术可能失效。另外元学习高度依赖网络的结构,因此模型结构的好坏对元学习技术的影响很大,未来可以利用神经架构搜索技术得到最适合元任务的网络结构。

表8 基于元学习与迁移学习事件抽取方法的对比

Table 8 Comparison of event extraction methods based on meta learning and transfer learning

方法类型	相同点	不同点	
		训练目标	损失函数
元学习	1. 要求源任务与目标任务在分布上尽可能相似 2. 两者本质都是如何让模型用到更广泛的数据分布,从而提高模型泛化性能	元学习强调模型在测试任务上表现好	1. 损失函数明显针对的是任务,根据任务的表现更新参数 2. 元学习在训练阶段就为测试阶段的更新做准备,迁移学习则没有
迁移学习	3. 在训练阶段,两者都会训练一组参数,在测试阶段,会根据少量数据更新参数	迁移学习强调模型在预训练任务上表现好	迁移学习没有任务观念,只有数据观念

3 实验结果及分析

3.1 数据集介绍

目前对于低资源场景事件抽取任务的研究主要采用

ACE2005^[2], TAC-KBP^[4]等通用的标准数据集,另外还有一些特定的数据集如 OntoEvent^[108], Causal-EE^[109], Few-Event^[115], GENIA^[117], FewFC^[79], CEC^[118]等,具体如表9所列。

1)ACE2005。ACE2005 数据集由语言数据联盟(Language Data Alliance, LDC)发布,由实体、关系和事件等多种数据类型组成,包含英语、阿拉伯语和中文 3 种语言。其定义了 8 种大类事件类型和 33 种小类事件子类型,包含 599 篇文档,是事件抽取任务使用最广泛、最具影响力的基准数据集。

2)TAC KBP。TAC KBP 事件追踪(2015—2017)的目标是抽取适合输入到知识库中的事件信息。TAC KBP 2015 以英文的方式定义了 9 种事件类型和 38 种事件子类型,TAC KBP 2016 和 TAC KBP 2017 包括英语、汉语和西班牙语 3 种语言,拥有 8 种事件类型和 18 种事件子类型。

3)OntoEvent。该数据集创建于 2021 年,其典型特点是带有事件之间的关系信息,包含来自 4 115 个文档中的 13 个事件类型,分为 100 个子类型和 60 546 个事件实例。数据集的创建基于 MAVEN^[119] 和 FewEvent 两个数据集,并使用 TCR 模型^[120] 来抽取时间和因果关系,使用 JCL 模型^[121] 来抽取层次关系。

4)Causal-EE。该数据集创建于 2021 年,主要特点是采用事件之间的因果注释增强 ACE-2005, TAC-KBP-2017 和 FewEvent 数据集,同时使用事件类型的信息来扩展 MATRES^[120], Causal-TB^[122], EventCausality^[123] 事件因果数据集。数据集包含 80 个事件类型,3 700 个事件实例以及 112 条事件因果关系对。

5)FewEvent。该数据集采用远程监督技术和手工注释的方式创建,包含 19 种事件类型,分为 100 个事件子类型和 70 852 个实例。每个事件类型的注释平均大约有 700 个实例。数据集构建过程大概分为两个步骤:首先使用已存在的 ACE-2005 语料库和 TAC-KBP2017 事件追踪数据中扩展事件类型的数量,之后采用自动标记数据生成策略^[86],在限制领域(如音乐、电影、体育、教育等)的前提下,从 Freebase 和 Wikipedia 中扩展一些新事件类型。

6)GENIA。该数据集创建于 2003 年,属于生物医学领域的语义注释语料库,GENIA3.0 由 1 999 个文档摘要组成,目前 GENIA 语料库的事件注释只覆盖了 1 000 个文档摘要,在 9 372 个句子上标记了 36 114 个事件。

7)FewFC。该数据集集中的数据来自于真实的金融新闻数据,由金融领域专业人员对其进行标注,被用于 CCKS 2020 金融领域跨类迁移事件抽取评测任务。FewFC 是一个经典的金融领域小样本数据集,包含了 10 种金融领域的事件类型,如投资、股份股权转让、质押等事件,共 8 982 个句子以及 12 890 个事件。

8)CEC。该数据集是由上海大学语义智能实验室创建的中文突发事件数据集,对交通事故、火灾、地震等 5 类突发事件的新闻报道进行标注,共 332 篇。与 ACE 数据集相比,CEC 较小,但其在事件和事件要素的注释方面更加全面。

表 9 数据集介绍

Table 9 Introduction of datasets

数据集	年份	介绍	链接
ACE2005	2005	8 大类事件类型和 33 小类事件子类型,包含 599 篇文档	https://catalog.ldc.upenn.edu/LDC2006T06
TAC KBP 2017	2017	8 种事件类型和 18 种事件子类型	https://tac.nist.gov/2017/KBP/
OntoEvent	2021	13 个事件类型,100 个子类型和 60 546 个事件实例	https://github.com/231sm/Reasoning_In_EE
Causal-EE	2021	80 个事件类型,3 700 个事件实例以及 112 条事件因果关系对	https://github.com/231sm/Reasoning_In_KE
FewEvent	2020	19 个事件类型,分为 100 个事件子类型和 70 852 个实例,每个事件类型的注释平均大约有 700 个实例	https://github.com/231sm/Low_Resource_KBP
GENIA	2003	47 个事件类型,1 999 个文档,18 545 个句子以及 36 114 个事件	http://www-tsujii.is.s.u-tokyo.ac.jp/GENIA
FewFC	2020	10 个事件类型,8 982 个句子以及 12 890 个事件	https://github.com/TimeBurningFish/FewFC
CEC	2015	5 种事件类型,332 个文档,5 991 个事件	https://github.com/shijiebei2009/CEC-Corpus

3.2 评价指标

本文采用事件抽取常用的评估指标准确率 P 、召回率 R 和 $F1$ 值来评估结果,计算式如下:

$$P = \frac{TP}{TP + FP} \quad (5)$$

$$R = \frac{TP}{TP + FN} \quad (6)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (7)$$

其中, TP 表示被模型预测为正类的正样本个数, FP 表示被模型预测为正类的负样本个数, FN 表示被模型预测为

负类的正样本个数。准确率 P 表示模型抽取出的正确个数占抽取总数的比例,用来衡量抽取的准确程度;召回率 R 表示模型正确抽取的个数占所有正确总数的比例,用来衡量抽取的全面程度; $F1$ 值是准确率和召回率的加权平均值,用来总体评估系统的性能, $F1$ 值越大表示模型的性能越好。

3.3 实验结果分析与对比

本小节总结了典型低资源场景事件抽取技术方法在 ACE2005 数据集上触发词分类(事件检测)与论元角色分类的实验结果,如表 10 所列。

表 10 典型低资源场景事件抽取技术在 ACE2005 数据集上的实验结果对比

Table 10 Comparison of experimental results on ACE2005 dataset of typical event extraction techniques

方法类型		年份-方法	触发词分类(事件检测)			论元角色分类			
			P	R	F1	P	R	F1	
迁移学习方法 (2.1节)	深度学习模型 迁移方法	CNN	2017-Huang 等 ^[20]	75.5	36.3	49.1	16.1	15.6	15.8
		RNN	2016-Feng 等(en) ^[21]	84.6	64.9	73.4	—	—	—
		GNN	2019-Lu 等(en2zh) ^[24]	—	—	—	—	—	59.0*
		Transformer	2021-GATE(en2zh) ^[26]	—	—	—	—	—	63.2*
			2021-IE-PRIME(G. Truth) (en,en2ar) ^[32]	—	—	—	—	—	72.4 (50.3*)
			2021-CCCAR(en2zh) ^[33]	—	—	72.1*	—	—	65.5*
			2022-Wang 等 ^[34]	—	—	47.8	—	—	43.0
	预训练语言模型方法	2022-X-GEAR(mT5-large, G. Truth)(en,en2zh) ^[35]	—	—	—	—	—	71.2 (54.0*)	
	机器阅读理解方法	2020-BERT_QA ^[39]	71.12	73.70	72.39	56.77	50.24	53.31	
		2020-MQAAE(en) ^[40]	—	—	73.8	—	—	55.0	
		2020-Feng et al. (G. Truth) ^[41]	73.28	76.29	75.43	57.69 (72.20)	51.90 (65.79)	54.61 (68.79)	
		2020-RCEEER(G. Truth) ^[42]	75.6	74.2	74.9	63.0 (71.2)	64.2 (69.1)	63.6 (70.1)	
	多任务学习方法	2019-Transition ^[45]	74.4	73.2	73.8	55.7	51.1	53.3	
		2019-MTL-CRF(zh) ^[50]	87.4	56.8	68.9	66.6	44.1	53.1	
		2020-Zhang 等 ^[48]	76.2	74.6	75.4	58.4	52.6	55.5	
提示(prompt)学习方法(2.2节)		2022-DEGREE(G. Truth) (com10) ^[54]	—	—	73.3 (54.5)	—	—	73.5 (31.4)	
		2022-PAIE(G. Truth) ^[58]	—	—	—	—	—	72.7	
无监督学习方法(2.3节)		2021-CLEVE (supervised, G. Truth) ^[74]	62.0 (78.1)	47.3 (81.5)	53.7 (79.8)	41.6 (55.4)	30.3 (68.0)	35.1 (61.1)	
		2021-Zhang 等 ^[72]	54.1	53.1	53.6	4.6	10.0	6.3	
弱监督学习 方法 (2.4节)	半监督学习方法	2019-DMBERT+ Boot ^[78]	77.9	72.5	75.1	—	—	—	
		2020-SS-VQ-VAE ^[84]	75.7	77.8	76.7	—	—	—	
	远程监督学习方法	2021-DualQA(G. Truth) ^[79]	—	—	—	49.1	42.3	45.4	
		2017-DMCNN trained with ACE+ED ^[86]	75.7	66.0	70.5	62.8	50.1	55.7	
数据与辅助 知识增强 方法 (2.5节)	多语言增强方法	2019-CLTrans(en2zh) ^[97]	62.5*	35.7*	45.4*	—	—	—	
	多模态增强方法	2020-CADA-fVAE-GAN (few- shot) ^[99]	43.21	52.36	50.51	—	—	—	
	辅助知识增强方法	2020-DRMM ^[101]	77.9	74.8	76.3	—	—	—	
		2021-AEE ^[110]	83.34	81.96	82.37	67.82	55.66	61.42	
元学习方法(2.6节)		2020-Proto+ Att(LSTM+Linter + Lintra + 5 + 1-way-5- shot) ^[113]	—	—	78.44	—	—	—	
		2020-DMB-PN(5-Way-15- Shot) ^[115]	—	—	76.03	—	—	—	

注:表中加数据数据表示最优结果,斜体表示零样本场景,*表示跨语言场景,G. Truth表示标准触发词给定场景,en表示英文,zh表示中文,en2zh表示英文→中文,en2ar表示英文→阿拉伯语,(en,en2zh)等表示在单语言和跨语言下的评估指标;仅有G. Truth限定时若存在两种指标,则()表示在G. Truth下的实验结果。

由表 10 可知,零样本事件抽取方法与监督学习方法相比仍然存在很大的差距。Wang 等^[84]提出的基于 Transformer 的迁移方法在触发词和论元角色分类上的 F1 值分别达到 47.8%和 43.0%,Huang 等^[20]提出的基于 CNN 的迁移方法在触发词分类任务上比其提升了 1.3%,但在论元角色分类任务上的效果却很差,主要原因在于该方法严重依赖于论元路径和论元概念的语义信息,而大多论元角色没有充分的信息来匹配这些论元概念。Wang 等^[74]提出的 CLEVE 模型和 Zhang 等^[72]提出的零样本分类模型等无监督学习方法在触发词分类上的效果相比迁移学习方法有所提升,其 F1 值分别达到了 53.7%和 53.6%,但两者在论元角色分类任务上的效果依然很差。Hsu 等^[54]提出的 DEGREE 模型(提示学习方法)稍微优于上述无监督学习方法的效果,其 F1 值为 54.5%。

目前的事件抽取方法在触发词分类上的 F1 值集中在

70%~80%之间。Feng 等^[21]提出的基于 RNN 的模型、Feng 等^[41]提出的基于 MRC 的模型以及 Zhang 等^[48]提出的 Soft tree + Fine-tuned 模型等迁移学习方法在触发词分类上的 F1 值分别为 73.4%、75.43%、75.4%。Wang 等^[74]提出的基于对比预训练的方法、Huang 等提出的 SS-VQ-VAE 模型(半监督学习方法)以及 Lai 等^[113]提出的基于 Proto+ Att 模型(LSTM+Linter+ Lintra + 5 + 1-way-5-shot 设定,元学习方法)在触发词分类上的 F1 值分别为 79.8%、76.7%、78.44%。Li 等^[110]提出的 AEE 模型(辅助知识增强方法)采用一个通用的论元角色模式将多个角色连接,以利用论元之间的共享信息,有效削减了错误信息传播和数据不平衡问题,同时还采用一个参数继承机制来增强类型偏好角色的识别,从而抽取不同事件类型的元素。该方法可以有效地学习事件类型和论元角色之间的共享信息,在 P,R,F1 指标上分别达到了 83.34%、81.96%、82.37%。在准确率方面,He 等^[50]

提出的 MTL-CRF 模型(多任务学习方法)在中文事件抽取触发词分类上的准确率高达 87.4%,在表中所有方法中最优,这得益于相关子类别事件相互增强缓解了数据规模小和数据不平衡的问题,从而提升了事件抽取的效果。

由表 10 可知,在论元角色分类方面,在标准触发词给定时,Hsu 等^[64]提出的 DEGREE 模型(提示学习方法)在论元角色分类效果上的 F1 值可达 73.5%,明显高于在无标准触发词给定时进行论元角色标记的效果,如 Du 等^[39]提出的 BERT_QA 模型(MRC 方法)在论元角色分类上的 F1 值为 53.31%。这表明论元抽取的效果严重依赖于事件检测的效果,此时错误传播的问题是阻碍论元角色标记性能下降的主要原因。另外,Nguyen 等^[33]提出的 CCCAR 模型在无标准触发词给定时,由英文迁移到中文的论元角色标记的 F1 值达到了 65.5%,在所列方法中最优。

4 未来研究趋势

第 3 章已分析对比了不同方法的优缺点,并针对性地为相关方法的不足提供了可解决的方案,同时也列出了未来研究可以关注的重点。本节从低资源场景事件抽取的全局角度,分析其未来主流的发展趋势。

1) 低资源场景事件抽取通用语料库的扩展

目前对低资源场景事件抽取的研究主要还是采用事件抽取任务标准的数据集,如 ACE2005^[2]等。但在实际研究时可能会用到不同事件之间的知识,即因果关系、时间关系等,此时采用通用领域的数据集就不能很好地表征上述知识,使得研究者需要重新构造特定的低资源场景事件抽取数据集,这十分耗时耗力。目前低资源场景下通用的标准数据集很缺乏,如何构造像 ACE2005 标准数据集那样能广泛使用的低资源场景事件抽取数据集是未来研究的热点与难点。可考虑首先利用外部知识增强策略对常用领域的事件进行扩展,并采用自动化工具标注出事件之间的多种常用关系(如时序、因果),接着对结果进行人工校对,后续可根据需求采用上述方式进一步扩展语料库,使得所构建的低资源事件抽取语料库具备通用性、可靠性和可扩展性。

2) 基于数据增强的事件抽取技术研究

目前,只有少量工作将数据增强技术用到事件抽取中,如 Luo 等^[124]提出了一种基于 BERT 的混合文本对抗训练方法(Bert Based Mix-text Adversarial Training, BMAD),采用回译的传统数据增强方法从原始数据构建真实无标注的数据,之后利用 Mix-Text 的数据增强方法生成虚拟训练数据和标签,通过训练不准确的数据和带噪学习来提高模型的泛化能力并尽量避免过拟合。

数据增强技术在事件抽取任务上的应用并不广泛,原因在于自然语言中的数据是离散的,不能像计算机视觉领域中对输入数据直接进行转换。利用类似计算机视觉中的裁剪法可能会改变语义,容易删除最重要的部分,变换文本的表现形式时展开变换的过程可能存在歧义问题。另外,使用 MixUp 方法(利用线性插值的方式产生更多的训练样本)建模文本中间语义时在句子或单词角度上会面临可解释性差等问题。因此,在引入数据增强时既要保证数据质量又要保证多样性,

未来的研究可以考虑将事件论元和触发词与知识图谱、事理图谱中的实体、动词融合,采用新形式的数据增强技术来提升模型的识别能力和泛化能力。

3) 基于多模态增强的事件抽取技术研究

目前已有研究者利用多模态数据(图像、语音、视频等)来缓解当前事件抽取任务面临的数据缺乏问题,基于多模态增强的事件抽取技术还处于发展阶段,主要集中在多模态嵌入以及生成丰富的语义信息上,其面临的主要挑战包括:如何更好地联系来自异源领域的的数据,如何有效地处理在连接不同模态期间不同程度的噪声,如何将两者放在一起学习。未来的研究可以聚焦在通过跨模态统一建模方法来增强模型的跨模态语义对齐与降噪能力,提升事件抽取的效果。

4) 联合神经网络与符号知识的事件抽取技术研究

目前大多事件抽取方法都是基于神经网络的,其缺乏透明度和健壮性,而事件抽取中的目标知识是符号的形式。未来的研究可以考虑将传统基于规则和模式匹配的方法与深度学习技术相结合,以有效地利用符号知识来削减深度学习模型对于注释数据的依赖,改善事件抽取面临的标注数据缺乏的问题。

结束语 随着事件抽取技术的发展,人们不再局限于利用已标注好的数据来达成既定的事件抽取任务目标,而是更加关注与实际应用场景的联系。传统机器学习和深度学习方法都需要标注好的训练数据,特别是深度学习模型需要大量的注释数据,在标注数据充足的情况下确实能显著地提升事件抽取的效果,但在现实场景下,相关领域和语言标注数据极度缺乏,标准数据集规模很小,低资源场景普遍存在。研究者寻求设计更鲁棒的方法使得模型具备跨领域、跨语言以及新事件类型的应对能力。本文全面总结归纳了近年来低资源场景事件抽取任务的相关工作,首先介绍了其任务定义和分类,之后重点阐述了当前流行的技术方法及先前工作,并给出了当前方法的不足和未来研究的重点;接着介绍了相关数据集及评价指标,并对典型技术方法的实验结果进行总结分析;然后从低资源场景事件抽取任务发展的全局角度分析了当前所面临的问题和未来的发展趋势,期望促进低资源场景事件抽取及其相关领域的协同发展,并为后续研究提供指导。

参考文献

- [1] GRISHMAN R, SUNDHEIM B M. Message Understanding Conference-6: A Brief History [C]// COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics. 1996:466-471.
- [2] 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 Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation. 2014:45-53.
- [3] SONG Z, BIES A, STRASSEL S, et al. From Light to Rich Ere: Annotation of Entities, Relations, and Events [C]// Proceedings of the the 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation. 2015:89-98.
- [4] MITAMURA T, LIU Z, HOVY E H, et al. Overview of Tackbp

- 2015 Event Nugget Track [C]//TAC. 2015:1-11.
- [5] NEDELLEC C, BOSSY R, KIM J D, et al. Overview of BioNLP Shared Task 2013[C]//Proceedings of the BioNLP Shared Task 2013 Workshop. 2013:1-7.
- [6] PUSTEJOVSKY J, HANKS P, SAURI R, et al. The Timebank Corpus[C]//Corpus Linguistics. 2003.
- [7] RILOFF E. Automatically Constructing a Dictionary for Information Extraction Tasks[C]//AAAI. 1993.
- [8] YANGRBER R, GRISHMAN R, TAPANAINEN P, et al. Automatic Acquisition of Domain Knowledge for Information Extraction[C]//COLING 2000 Volume 2: The 18th International Conference on Computational Linguistics. 2000:1-7.
- [9] CHEN C, NG V. Joint Modeling for Chinese Event Extraction with Rich Linguistic Features [C]// Proceedings of COLING 2012. 2012:529-544.
- [10] YU J D, LI X Y, FAN X Z, et al. Event Classification of Maximum Entropy Model[J]. Journal of University of Electronic Science and Technology of China, 2010, 39(4):612-616.
- [11] SARZYNSKA-WAWER J, WAWER A, PAWLAK A, et al. Detecting Formal Thought Disorder by Deep Contextualized Word Representations[J]. Psychiatry Research, 2021, 304:114135.
- [12] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving Language Understanding by Generative Pre-training [J/OL]. [2022-07-18]. [https://s3-us-west-2. amazonaws. com/ openai-assets/research covers/language_ unsupervised/language_ understanding_paper. pdf](https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language_ unsupervised/language_ understanding_paper.pdf).
- [13] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding [J]. arXiv:1810. 04805, 2018.
- [14] XIANG W, WANG B. A Survey of Event Extraction From Text [J]. IEEE Access, 2019, 7:173111-173137.
- [15] LI Q, LI J, SHENG J, et al. A Compact Survey on Event Extraction: Approaches and Applications [J]. arXiv:2107. 02126, 2021.
- [16] JI Y T. Research on Key Issues in Low-resource Neural Machine Translation [D]. Hohhot: Inner Mongolia University, 2020.
- [17] DENG S, ZHANG N, CHEN H, et al. Knowledge Extraction in Low-resource Scenarios: Survey and Perspective [J]. arXiv: 2202. 08063, 2022.
- [18] NGUYEN T H, FU L, CHO K, et al. A Two-stage Approach for Extending Event Detection to New Types Via Neural Networks[C]//Proceedings of the 1st Workshop on Representation Learning for NLP. 2016:158-165.
- [19] LAI V D, NGUYEN T H. Extending Event Detection to New Types with Learning From Keywords[J]. arXiv:1910. 11368, 2019.
- [20] HUANG L, JI H, CHO K, et al. Zero-shot Transfer Learning for Event Extraction[J]. arXiv:1707. 01066, 2017.
- [21] FENG X, QIN B, LIU T, et al. A Language-Independent Neural Network for Event Detection [J]. Science China Information Sciences, 2018, 61(9):1-12.
- [22] RAMRAKHIYANI N, HINGMIRE S, PATIL S, et al. Extracting Events From Industrial Incident Reports[C]//Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2021). 2021:58-67.
- [23] HOU W J, CEESAY B. Domain Transformation on Biological Event Extraction by Learning Methods[J]. Journal of Biomedical Informatics, 2019, 95:103236.
- [24] SUBBURATHINAM A, LU D, JI H, et al. Cross-lingual Structure Transfer for Relation and Event Extraction[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019:313-325.
- [25] VAN NGUYEN M, NGUYEN T H. Improving Cross-lingual Transfer for Event Argument Extraction with Language-universal Sentence Structures[C]// Proceedings of the Sixth Arabic Natural Language Processing Workshop. 2021:237-243.
- [26] AHMAD W U, PENG N, CHANG K W, et al. GATE: Graph Attention Transformer Encoder for Cross-lingual Relation and Event Extraction[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021:12462-12470.
- [27] Joint Extraction of Sudden Hot Events Based on Graph Attention Network[J]. Journal of Chinese Computer Systems, 2023, 44(5):902-909.
- [28] LU D, SUBBURATHINAM A, JI H, et al. Cross-lingual Structure Transfer for Zero-resource Event Extraction[C]// Proceedings of The 12th Language Resources and Evaluation Conference. 2020:1976-1981.
- [29] LU W Y, NGUYEN T H. Similar but not the Same: Word Sense Disambiguation Improves Event Detection Via Neural Representation Matching[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018:4822-4828.
- [30] LAI V D, NGUYEN M V, NGUYEN T H, et al. Graph Learning Regularization and Transfer Learning for Few-shot Event Detection[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021:2172-2176.
- [31] ZHUANG Y. Zero-shot Compositional Event Detection Via Graph Modular Network[D]. Georgia Institute of Technology, Atlanta, 2021.
- [32] FINCKE S, AGARWAL S, MILLER S, et al. Language Model Priming for Cross-lingual Event Extraction[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2022:10627-10635.
- [33] VAN NGUYEN M, NGUYEN T N, MIN B, et al. Cross Lingual Transfer Learning for Relation and Event Extraction Via Word Category and Class Alignments[C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021:5414-5426.
- [34] WANG S, YU M, CHANG S, et al. Query and Extract: Refining Event Extraction as Type-oriented Binary Decoding[J]. arXiv: 2110. 07476, 2021.
- [35] HUANG K H, HSU I, NATARAJAN P, et al. Multilingual Generative Language Models for Zero-shot Cross-lingual Event Argument Extraction[J]. arXiv:2203. 08308, 2022.

- [36] HUANG K H, YANG M, PENG N, et al. Biomedical Event Extraction with Hierarchical Knowledge Graphs[J]. arXiv:2009.09335,2020.
- [37] BELTAGY I, LO K, COHAN A, et al. SciBERT: A Pretrained Language Model for Scientific Text[J]. arXiv:1903.10676,2019.
- [38] SHENG J, LI Q, HEI Y, et al. A Joint Learning Framework for the CCKS-2020 Financial Event Extraction Task[J]. Data Intelligence,2021,3(3):444-459.
- [39] DU X, CARDIE C. Event Extraction by Answering (almost) Natural Questions[J]. arXiv:2004.13625,2020.
- [40] LI F, PENG W, CHEN Y, et al. Event Extraction as Multi-turn Question Answering[C]// Findings of the Association for Computational Linguistics; EMNLP. 2020:829-838.
- [41] FENG R, YUAN J, ZHANG C, et al. Probing and Fine-tuning Reading Comprehension Models for Few-shot Event Extraction[J]. arXiv:2010.11325,2020.
- [42] LIU J, CHEN Y, LIU K, et al. Event Extraction as Machine Reading Comprehension[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020:1641-1651.
- [43] MEHTA S, RANGWALA H, RAMAKRISHNAN N, et al. Improving Zero-shot Event Extraction Via Sentence Simplification[J]. arXiv:2204.02531,2022.
- [44] LYU Q, ZHANG H, SULEM E, et al. Zero-shot Event Extraction Via Transfer Learning: Challenges and Insights[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). 2021:322-332.
- [45] ZHANG J, QIN Y, ZHANG Y, et al. Extracting Entities and Events as a Single Task Using a Transition-based Neural Model[C]// IJCAI. 2019:5422-5428.
- [46] PAOLINI G, ATHIWARATKUN B, KRONE J, et al. Structured Prediction as Translation between Augmented Natural Languages[J]. arXiv:2101.05779,2021.
- [47] WU W T, LI P F, ZHU Q M. Joint Extraction of Entities and Events by a Hybrid Neural Network[J]. Journal of Chinese Information Processing,2019,33(8):77-83.
- [48] ZHANG J, HE Q, ZHANG Y. Syntax Grounded Graph Convolutional Network for Joint Entity and Event Extraction[J]. Neurocomputing,2021,422:118-128.
- [49] WADDEN D, WENNBURG U, LUAN Y, et al. Entity, Relation, and Event Extraction with Contextualized Span Representations[J]. arXiv:1909.03546,2019.
- [50] HE R F, DUAN S Y. Joint Chinese Event Extraction Based Multi-task Learning[J]. Journal of Software,2019,30(4):1015-1030.
- [51] LI S, JI H, HAN J. Document-level Event Argument Extraction by Conditional Generation[J]. arXiv:2104.05919,2021.
- [52] LU Y, LIN H, XU J, et al. Text2Event: Controllable Sequence-to-structure Generation for End-to-end Event Extraction[J]. arXiv:2106.09232,2021.
- [53] SI J, PENG X, LI C, et al. Generating Disentangled Arguments with Prompts: A Simple Event Extraction Framework that Works[C]// ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022:6342-6346.
- [54] HSU I H, HUANG K H, BOSCHÉE E, et al. DEGREE: A Data-efficient Generation-based Event Extraction Model[C]// Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. 2022:1890-1908.
- [55] LIN J, JIAN J, CHEN Q. Eliciting Knowledge From Language Models for Event Extraction[J]. arXiv:2109.05190,2021.
- [56] LIU X, HUANG H, SHI G, et al. Dynamic Prefix-tuning for Generative Template-based Event Extraction[J]. arXiv:2205.06166,2022.
- [57] LIU X, JI K, FU Y, et al. P-tuning v2: Prompt Tuning can be Comparable to Fine-tuning Universally Across Scales and Tasks[J]. arXiv:2110.07602,2021.
- [58] MA Y, WANG Z, CAO Y, et al. Prompt for Extraction? PAIE: Prompting Argument Interaction for Event Argument Extraction[J]. arXiv:2202.12109,2022.
- [59] XU H, CHEN Y, DU Y, et al. ZeroPrompt: Scaling Prompt-based Pretraining to 1,000 Tasks Improves Zero-shot Generalization[J]. arXiv:2201.06910,2022.
- [60] GU Y, HAN X, LIU Z, et al. Ppt: Pre-trained Prompt tuning for Few-shot Learning[J]. arXiv:2109.04332,2021.
- [61] VU T, LESTER B, CONSTANT N, et al. Spot: Better Frozen Model Adaptation Through Soft Prompt Transfer[J]. arXiv:2110.07904,2021.
- [62] CHAMBERS N. Event Schema Induction with a Probabilistic Entity-driven Model[C]// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013:1797-1807.
- [63] NGUYEN K H, TANNIER X, FERRET O, et al. Generative Event Schema Induction with Entity Disambiguation[C]// 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). 2015:188-197.
- [64] JIN X, LI M, JI H. Event Schema Induction with Double Graph Autoencoders[C]// 2022 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies (NAACL 2022). Association for Computational Linguistics(ACL),2022:2013-2025.
- [65] SHA L, LI S, CHANG B, et al. Joint Learning Templates and Slots for Event Schema Induction[J]. arXiv:1603.01333,2016.
- [66] HUANG L, CASSIDY T, FENG X, et al. Liberal Event Extraction and Event Schema Induction[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016:258-268.
- [67] YUAN Q, REN X, HE W, et al. Open-schema Event Profiling for Massive News Corpora[C]// Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018:587-596.
- [68] LIU X, HUANG H, ZHANG Y. Open Domain Event Extraction

- Using Neural Latent Variable Models[J]. arXiv:1906.06947, 2019.
- [69] ZHOU D, CHEN L, HE Y. An Unsupervised Framework of Exploring Events on Twitter: Filtering, Extraction and Categorization[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2015.
- [70] ZHOU D, ZHANG X, HE Y. Event Extraction from Twitter Using Non-parametric Bayesian Mixture Model with Word Embeddings[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1. Long Papers, 2017:808-817.
- [71] HAO D K. Research on Key Technologies of Schema Induction Based Open-domain Event Extraction [D]. Harbin Institute of Technology, 2021.
- [72] ZHANG H, WANG H, ROTH D. Zero-shot Label-aware Event Trigger and Argument Classification [C]//Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. 2021:1331-1340.
- [73] WANG R, ZHOU D, HE Y. Open Event Extraction from Online Text Using a Generative Adversarial Network[J]. arXiv:1908.09246, 2019.
- [74] WANG Z, WANG X, HAN X, et al. CLEVE: Contrastive Pre-training for Event Extraction[J]. arXiv:2105.14485, 2021.
- [75] HUANG R, RILOFF E. Bootstrapped Training of Event Extraction Classifiers[C]//Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. 2012:286-295.
- [76] LIAO S, GRISHMAN R. Can Document Selection Help Semi-supervised Learning? A Case Study on Event Extraction[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011:260-265.
- [77] LIAO S, GRISHMAN R. Using Prediction from Sentential Scope to Build a Pseudo Co-testing Learner for Event Extraction [C]//Proceedings of 5th International Joint Conference on Natural Language Processing. 2011:714-722.
- [78] WANG X, HAN X, LIU Z, et al. Adversarial Training for Weakly Supervised Event Detection [C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019:998-1008.
- [79] ZHOU Y, CHEN Y, ZHAO J, et al. What the Role is vs. What Plays the Role: Semi-supervised Event Argument Extraction Via Dual Question Answering [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021:14638-14646.
- [80] WANG J, XU Q, LIN H, et al. Semi-supervised Method for Biomedical Event Extraction [J]. Proteome Science, 2013, 11(1): 1-10.
- [81] FERGUSON J, LOCKARD C, WELD D S, et al. Semi-supervised Event Extraction with Paraphrase Clusters [J]. arXiv:1808.08622, 2018.
- [82] PENG H, SONG Y, ROTH D. Event Detection and Co-reference with Minimal Supervision[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016:392-402.
- [83] MIWA M, PYYSALO S, OHTA T, et al. Wide Coverage Biomedical Event Extraction Using Multiple Partially Overlapping Corpora[J]. BMC Bioinformatics, 2013, 14(1):1-12.
- [84] HUANG L, JI H. Semi-supervised New Event Type Induction and Event Detection[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020:718-724.
- [85] MINTZ M, BILLS S, SNOW R, et al. Distant Supervision for Relation Extraction without Labeled Data[C]//Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. 2009:1003-1011.
- [86] CHEN Y, LIU S, ZHANG X, et al. Automatically Labeled Data Generation for Large Scale Event Extraction[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017:409-419.
- [87] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a Collaboratively Created Graph Database for Structuring Human Knowledge[C]//Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. 2008:1247-1250.
- [88] BAKER C F, FILLMORE C J, LOWE J B, et al. The Berkeley Framenet Project[C]//COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics. 1998: 86-90.
- [89] YANG H, CHEN Y, LIU K, et al. Dcfec: A Document-level Chinese Financial Event Extraction System Based on Automatically Labeled Training Data[C]//Proceedings of ACL 2018, System Demonstrations. 2018:50-55.
- [90] ARAKI J, MITAMURA T. Open-domain Event Detection Using Distant Supervision[C]//Proceedings of the 27th International Conference on Computational Linguistics. 2018: 878-891.
- [91] MILLER G A, BECKWITH R, FELLBAUM C, et al. Introduction to Wordnet: an On-line Lexical Database[J]. International journal of lexicography, 1990, 3(4): 235-244.
- [92] MILNE D, WITTEN I H. Learning to Link with Wikipedia [C]//Proceedings of the 17th ACM Conference on Information and Knowledge Management. 2008:509-518.
- [93] CHEN Z, JI H. Can One Language Bootstrap the Other; a Case Study on Event Extraction [C] // Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing. 2009:66-74.
- [94] ZHU Z, LI S, ZHOU G, et al. Bilingual Event Extraction; a Case Study on Trigger Type Determination[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2014:842-847.
- [95] HSI A, CARBONELL J G, YANG Y. Modeling Event Extraction Via Multilingual Data Sources[C]//TAC. 2015.
- [96] HSI A, YANG Y, CARBONELL J G, et al. Leveraging Multilingual Training for Limited Resource Event Extraction[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics; Technical Papers. 2016:1201-1210.
- [97] LIU J, CHEN Y, LIU K, et al. Neural Cross-lingual Event De-

- tection with Minimal Parallel Resources[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing(EMNLP-IJCNLP). 2019;738-748.
- [98] HAMBARDZUMYAN K, KHACHATRIAN H, MAY J. The Role of Alignment of Multilingual Contextualized Embeddings in Zero-shot Cross-lingual Transfer for Event Extraction [J/OL]. <https://d-nb.info/1216910162/34#page=109>.
- [99] ZHU X, HOU M, ZENG X, et al. CADA-fVAE-GAN: Adversarial Training for Few-shot Event Detection [J]. *Computer Science & Information Technology (CS & IT)*, 2020, 10(14): 13-24.
- [100] ZHANG T, WHITEHEAD S, ZHANG H, et al. Improving Event Extraction Via Multimodal Integration[C]//Proceedings of the 25th ACM International Conference on Multimedia. 2017: 270-278.
- [101] TONG M, WANG S, CAO Y, et al. Image Enhanced Event Detection in News Articles[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020:9040-9047.
- [102] LI M, ZAREIAN A, ZENG Q, et al. Cross-media Structured Common Space for Multimedia Event Extraction [J]. arXiv: 2005. 02472, 2020.
- [103] CHEN B, LIN X, THOMAS C, et al. Joint Multimedia Event Extraction from Video and Article [J]. arXiv: 2109. 12776, 2021.
- [104] WU S, BONDUGULA S, LUISIER F, et al. Zero-shot Event Detection Using Multi-modal Fusion of Weakly Supervised Concepts[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014:2665-2672.
- [105] LIU S, CHEN Y, HE S, et al. Leveraging Framenet to Improve Automatic Event Detection[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016:2134-2143.
- [106] LI W, CHENG D, HE L, et al. Joint Event Extraction Based on Hierarchical Event Schemas from Framenet[J]. *IEEE Access*, 2019, 7: 25001-25015.
- [107] TONG M, XU B, WANG S, et al. Improving Event Detection Via Open-domain Trigger Knowledge [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020:5887-5897.
- [108] DENG S, ZHANG N, LI L, et al. Ontoed: Low-Resource Event Detection with Ontology Embedding [J]. arXiv: 2105. 10922, 2021.
- [109] DENG S, ZHANG N, CHEN H, et al. Low-resource Extraction with Knowledge-aware Pairwise Prototype Learning[J]. *Knowledge-Based Systems*, 2022, 235: 107584.
- [110] LI Q, GUO S, WU J, et al. Event Extraction by Associating Event Types and Argument Roles[J]. arXiv:2108. 10038, 2021.
- [111] FINN C, ABBEEL P, LEVINE S. Model-agnostic Meta-learning for Fast Adaptation of Deep Networks[C]//International Conference on Machine Learning. PMLR, 2017: 1126-1135.
- [112] SONG Y, WANG T, MONDAL S K, et al. A Comprehensive Survey of Few-shot Learning: Evolution, Applications, Challenges, and Opportunities[J]. arXiv:2205. 06743, 2022.
- [113] LAI V D, DERNONCOURT F, NGUYEN T H. Extensively Matching for Few-shot Learning Event Detection [J]. arXiv: 2006. 10093, 2020.
- [114] YANG H, CHEN Y, LIU K, et al. Meta Learning for Event Argument Extraction Via Domain-specific Information Enhanced [C]//China Conference on Knowledge Graph and Semantic Computing. Springer, Singapore, 2020: 160-172.
- [115] DENG S, ZHANG N, KANG J, et al. Meta-learning with Dynamic-memory-based Prototypical Network for Few-shot Event Detection[C]//Proceedings of the 13th International Conference on Web Search and Data Mining. 2020: 151-159.
- [116] SHEN S, WU T, QI G, et al. Adaptive Knowledge-enhanced Bayesian Meta-learning for Few-shot Event Detection [J]. arXiv: 2105. 09509, 2021.
- [117] KIM J D, OHTA T, TATEISI Y, et al. GENIA corpus—a semantically annotated corpus for bio-textmining[J]. *BIOINFORMATICS*, 2003, 19(1): i180-i182. .
- [118] MENG H J. Research on Event Extraction Technology in the Field of Unexpected Events[D]. Shanghai: Shanghai University, 2015.
- [119] WANG X, WANG Z, HAN X, et al. MAVEN: A Massive General Domain Event Detection Dataset [J]. arXiv: 2004. 13590, 2020.
- [120] NING Q, FENG Z, WU H, et al. Joint Reasoning for Temporal and Causal Relations[J]. arXiv: 1906. 04941, 2019.
- [121] WANG H, CHEN M, ZHANG H, et al. Joint Constrained Learning for Event-event Relation Extraction [J]. arXiv: 2010. 06727, 2020.
- [122] MIRZA P, TONELLI S. An Analysis of Causality Between Events and Its Relation to Temporal Information[C]//Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers. 2014: 2097-2106.
- [123] DO Q, CHAN Y S, ROTH D. Minimally Supervised Event Causality Identification[C]//Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. 2011: 294-303.
- [124] LUO P, DING L, YANG X, et al. Chinese Event Detection Based on Data Augmentation and Weakly Supervised Adversarial Training [J/OL]. <http://kns.cnki.net/kcms/detail/51.1307.tp.20211221.2112.004.html>.



LIU Tao, born in 1998, master candidate. His main research interests include natural language processing and knowledge graph.



JIANG Guoquan, born in 1978, master, associate professor, is a member of CCF (No. B2492M). His main research interests include data engineering and knowledge graph.