

事件图谱的构建、推理与应用

胡志磊^{1,2,3}, 靳小龙^{1,2,3}, 陈剑赞⁴, 黄冠利⁵

1. 中国科学院网络数据科学与技术重点实验室, 北京 100190; 2. 中国科学院计算技术研究所, 北京 100190;
3. 中国科学院大学计算机科学与技术学院, 北京 100049; 4. 北京市信息技术研究所, 北京 100091;
5. 北京电子科技职业学院, 北京 100176

摘要

近些年,知识图谱的构建技术得到了极大的发展,构建好的知识图谱已经被应用到众多领域。在此基础上,研究者将目光从知识图谱转向事件图谱。事件图谱以事件为核心,准确地描述了事件信息以及事件之间的关联关系。基于此,总结了事件图谱在构建、推理与应用方面的关键技术,主要包括事件抽取、事件信息补全、事件关系推断以及事件预测技术。给出了事件图谱的具体应用场景,并且针对事件图谱研究中存在的挑战,对未来的研究趋势进行了展望。

关键词

知识图谱;事件图谱;事件知识;事件抽取;事件推理

中图分类号:TP391

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2021027

Construction, reasoning and applications of event graphs

HU Zhilei^{1,2,3}, JIN Xiaolong^{1,2,3}, CHEN Jianyun⁴, HUANG Guanli⁵

1. Key Laboratory of Network Data Science & Technology, Chinese Academy of Sciences, Beijing 100190, China
2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
3. School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China
4. Beijing Institute of Information Technology, Beijing 100091, China
5. Beijing Polytechnic, Beijing 100176, China

Abstract

In recent years, the construction technology of knowledge graphs have been greatly developed, and the constructed knowledge graphs have been applied to many fields. On this basis, the researchers turned their attention from the knowledge graph to the event graph. The event graph takes the event as the core and accurately describes the event information and the relationship between the events. The key technologies of event graphs construction, reasoning and applications were summarized, including event extraction, event information completion, event relationship inference and event prediction. Finally, the specific application scenarios of the event graphs were given, and the future research trends were prospected in view of the challenges existing in the event graph research.

Key words

knowledge graph, event graph, event knowledge, event extraction, event reasoning

1 引言

随着信息技术的飞速发展,目前整个社会已经迈入了大数据时代。大数据时代下每时每刻都在产生庞大的数据。在庞大数据的背后,蕴含着众多有价值的信息。但是由于数据种类繁多、数据量庞大,难以高效、准确地获取有用的信息。为了更加高效地获取数据背后的信息,提高获取信息的效率,研究人员开始研究并使用自动化的工具从原始数据中抽取有价值的信息。这种自动化的技术被称为信息抽取技术^[1],可以极大地提高工作效率,节省时间。与此同时,由于信息抽取技术逐渐走向成熟,一种新的数据组织形式逐渐形成,实现了知识互联,适应了用户的认知需求,其被称为知识图谱。

知识图谱的概念由Google公司在2012年正式提出^[2],目的是提高搜索引擎的性能,提供更加友好的搜索结果。随后知识图谱在学术界受到了极大的关注,其构建技术也在飞速发展。目前,知识图谱已经被广泛地应用到知识问答、智能搜索、个性化推荐、软件复用^[3]、政府治理^[4]等多个领域。随着技术的不断发展,现有研究内容已经从知识图谱的实体识别^[5]、关系抽取^[6]技术扩展到了事件图谱的构建与推理技术。事件图谱刻画了现实世界中发生的事件,对事件信息进行了准确描述。事件图谱中蕴含众多事件知识,事件知识的特点是拥有众多维度,例如时间维度、逻辑维度、关系维度等。

本文对现有的关于事件知识的研究做了总结,从事件图谱的构建、推理与应用3个方面阐述了相关技术的研究现状。最后,本文展望了事件图谱的发展方向。

2 知识图谱与事件图谱

知识图谱本质上是一种语义网络^[7],包含实体以及实体之间的关系。实体是知识图谱中最基本的概念,一个实体代表了现实世界中的一个事物或者一个概念。关系代表的是实体和实体之间的关联。另外,知识图谱中的属性代表的是实体具有的某个特征,其描述了实体的相关信息。知识图谱刻画了现实世界中的事物以及事物之间的各种关系,其主要表示形式是三元组,包括属性三元组和关系三元组。知识图谱存储了结构化数据,适合对数据进行搜索和推理,已经在知识问答、智能搜索等方面发挥了重要的作用。

事件是文本中包含的一种信息,其定义为在某个特定的时间以及特定的地点,由多个相关角色参与的一件事情或者一组事情^[8]。随着知识图谱技术的发展,越来越多的研究者开始关注一种特殊的基于事件的知识图谱,即事件图谱。在此基础上,本文将事件图谱定义为:一种以事件为中心,用来描述事件信息以及事件之间各种关系的图谱。事件图谱和知识图谱主要的不同点是事件图谱主要的研究对象是事件,描述了与事件相关的知识、事件的演变过程以及事件间的关联关系。而知识图谱主要的研究对象是实体,主要描述的是实体属性以及实体之间的关系。在事件图谱中,图的节点表示事件,图上的连边表示事件之间的时序、因果、顺承、包含等关系。事件图谱示例如图1所示,图1展示了收购事件、股价上涨事件和股价下跌事件的事件论元以及事件之间的关联关系。从图1中可以看到,收购事件的收购方是公司A,被收购方是公司B,收购金额

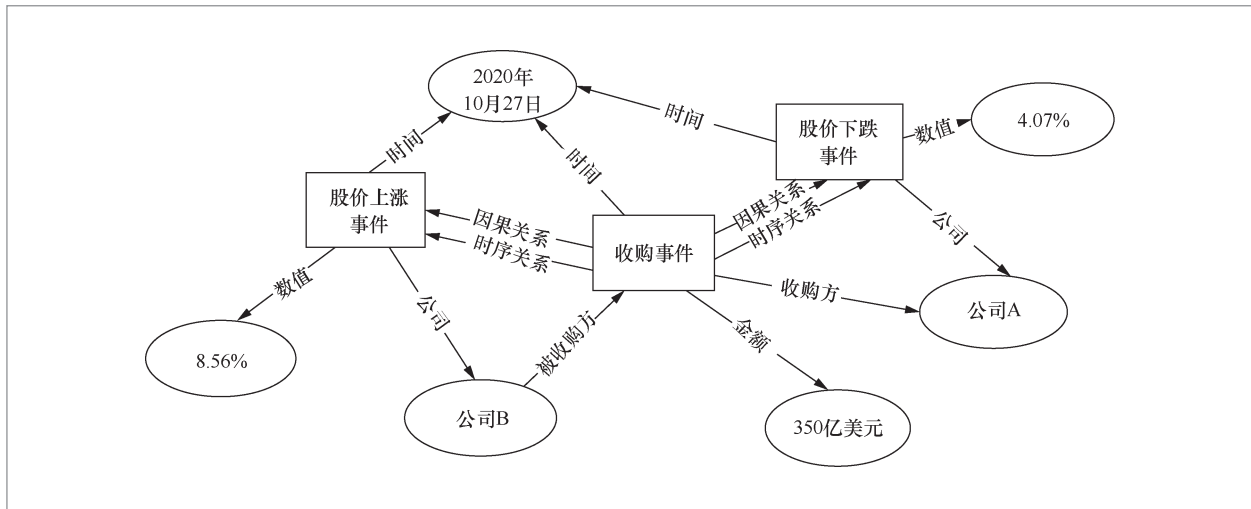


图1 事件图谱示例

是350亿美元，收购时间是2020年10月27日。另外，由于收购事件导致了股价上涨事件和股价下跌事件的发生，因此收购事件分别与股价上涨事件、股价下跌事件之间具有因果关系以及隐含的时序关系。

在事件图谱的构建、推理与应用的过程中，需要用到多种智能化信息处理技术^[9-10]，核心技术主要包括事件抽取技术、信息补全技术、关系推断技术以及事件预测技术。面对开放网络上的文本数据，首先要做的任务是事件抽取。事件抽取技术可以从非结构化的文本数据中提取出与事件有关的信息，并将信息以结构化的形式呈现出来。而信息补全技术是利用事件图谱中已有的知识，推理补全事件图谱中缺失的事件论元知识。关系推断技术则是利用文本中的信息来推断出事件之间的共指、时序以及因果等关系。最后，事件预测技术被用来预测未来可能发生的事件，分析事件的演变趋势。构建好的事件图谱具有广阔的应用场景，主要包括热点事件检测、事件脉络分析以及未来事件预测等。

3 事件抽取

事件是文本中包含的一种特殊信息，事件抽取就是从非结构化的文本数据中抽取与事件有关的各种角色，将信息用结构化数据表示。按照确定事件类别的方法，事件抽取可以被分为限定域事件抽取和开放域事件抽取。

3.1 限定域事件抽取

限定域事件抽取是指在进行抽取任务之前，已经确定好了相应的目标事件类型和相应的结构。另外，限定域事件抽取任务还会给出一些标注数据。因为事件标注较为复杂，需要耗费一定的人力物力，所以数据集规模一般较小。在事件抽取领域中较为常用的标准数据集是ACE 2005语料库。关于限定域事件抽取的研究较多，目前研究主要采用的是深度学习方法，几种代表性的神经网络方法如下。

(1) 基于注意力机制的方法

在进行事件抽取时,许多研究者使用了注意力机制,以此来提升神经网络模型的效果。注意力机制是一种仿生技术,借鉴了人类的选择性注意行为。选择性注意行为是指人类在进行观察时,视觉会快速扫描全局图像,从而确定要重点关注的内容,抑制或忽略其他无关的信息。研究者受此启发,提出了深度学习中的注意力机制,核心目标就是在众多信息中选择对于当前任务来说最关键的信息。

ACE 2005语料库给出了每个事件的事件触发词和事件论元。但是之前的研究者并没有充分利用数据集中的标注信息,更多地依赖句子的语义信息,忽略了被标注的论元信息。因此,Liu S等人^[11]提出了一种新的方法,利用论元信息来加强对触发词的识别和分类。该方法将句子中的单词信息、上下文的单词信息、上下文的实体信息结合起来,组成单词的触发词候选项。他们还采用了一种有监督的注意力机制更加深入地提取句子中的有效信息,句子中的事件触发词会比其他上下文单词获得更多的注意。最后该方法使用了一个多分类的神经网络模型完成事件抽取任务,充分利用了语料库中被标注的论元信息。

上述方法将句子中的多个事件视为独立事件,只是利用单个句子内部的信息来检测事件。但是句子中表达的事件是相互关联的,单纯地利用句子内部的信息不能很好地区分某些事件。因此Chen Y等人^[12]提出了一种分层的基于门控注意力机制的偏差标记网络,目的是融合句子和文档的信息,从而进行多事件识别和抽取。该方法采用了一种新的思路,将事件抽取看作一种序列标注问题。模型中首次添加了一个层次化的基于循环神经网络(recurrent neural network, RNN)的标签层来捕捉所有事件的依赖关系,同时

设计了一种偏差目标函数来增强触发标签对模型的影响。除此之外,为了充分利用事件候选项的上下文信息,该方法采用了一种基于门控的多层次注意力机制,可以自动提取句子和文档中的信息,并进行动态的集成。该方法充分利用了事件之间的关联关系,融合了文档信息来增强事件识别的结果。可以看到,基于注意力机制的方法实现了对信息的有效提取,使得事件识别更加准确。

(2) 基于预训练模型的方法

一般来说,为了更好地训练神经网络模型,需要为模型提供大规模的标注数据。但是构建大规模的标注数据耗时耗力,难以满足要求。相比之下,大规模未标注的语料却很容易构建。为了利用大量的未标注数据,研究人员提出了预训练模型。预训练模型可以从大规模的语料中提取隐含的语义信息,学习到更好的通用语义表示向量,从而提高下游任务的表现。

Yang S等人^[13]针对现有远程监督事件抽取方法中存在的问题,将目光转向预训练的语言模型,希望利用从大规模语料库中学习到的知识表示向量来提高模型的性能。其设计了一种基于预训练语言模型的事件抽取(pre-trained language model based event extractor, PLMEE)模型。该模型的结构如图2所示^[13]。他们将事件抽取看作由两个子任务组成,两个子任务分别是触发词抽取和论元抽取,并提出了以预训练语言模型为基础的触发词抽取器和论元抽取器。基于预训练模型的方法会使事件的语义表示更加精确。但是现有的方法将事件抽取看作两个子任务,构建的是流水线模型,存在明显的错误传递问题。

(3) 基于图神经网络的方法

过去几年深度学习技术兴起,以神经网络为代表的技术被用来提取欧氏空间中

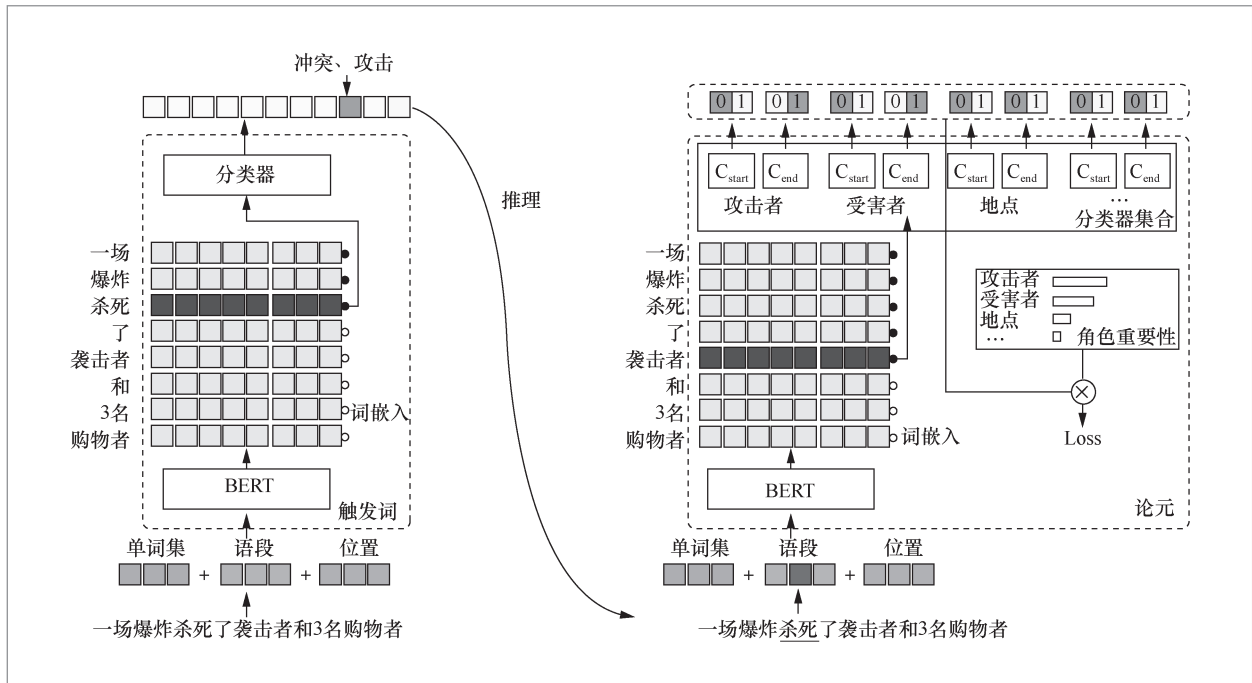


图 2 PLMEE 模型结构^[13]

的数据特征^[14]。但是现实世界中还有众多场景使用的是非欧氏空间数据，其中具有代表性的就是图数据。图数据被广泛应用于多种场景，如电子商务的推荐系统、知识图谱的在线推理等。但是图数据结构复杂性较高，之前的神经网络方法无法直接被使用。因此，研究人员借鉴了卷积神经网络、循环神经网络以及深度自动编码器的思想，设计了一种专门用来处理图结构数据的神经网络，即图神经网络。图卷积神经网络 (graph convolutional network, GCN) 是指在图数据中应用卷积操作，其核心思想是学习到一种函数对节点进行表示。通过函数变换，一个节点自身的特征可以结合其临近节点的特征，从而生成节点新的表示。

在事件抽取、关系抽取等任务中，一般利用词嵌入等方法将输入序列转换为连续的向量，并没有使用句子的结构信息。为了在神经网络中引入句法结构特征，

Nguyen T等人^[15]提出了一个基于句法依存树的GCN模型用于事件抽取。在GCN中，每个节点的卷积向量是由相邻节点的表示向量计算出来的，可以作为该节点的唯一特征进行分类。另外，模型中通过对当前单词的卷积向量以及句子中提到的实体进行池化操作，克服实体指称无法捕捉的问题。池化操作聚合了卷积向量，从而为事件类型预测生成了单个向量表示。该方法在事件抽取中引入了GCN模型，将句法依存树上的信息进行聚合，首次利用了句子中的结构信息。

除此之外，Liu X等人^[16]设计了一个新的联合多事件抽取 (jointly multiple events extraction, JMEE) 框架。该框架利用基于注意力机制的图卷积神经网络进行建模，并通过引入句法依存树中的句法捷径弧来增强信息流，以此来提升在一个句子中抽取多个事件的效果。使用句法捷径弧可以减少将信息流从一个节点转换到

目标节点的转换次数。与基于序列的模型相比,该方法会使在同一个句子中从一个事件触发词跳到另一个事件触发词的跳数明显减少。GCN会利用输入的句法捷径弧,聚合目标节点一阶邻居的信息,为每个节点学习到其句法上下文的表示。之后,模型通过自注意力机制进行信息聚合,保留了多个事件之间的信息,用于抽取事件触发词和论元。基于图神经网络的方法有效地利用了句法依存树中包含的信息。但是基于图神经网络的模型的计算量比较大,且只适用于对构建好的静态图进行处理。

(4) 其他神经网络方法

除了上述方法,还有众多方法被应用到事件抽取任务中,也取得了良好的效果。例如,Liu J等人^[17]设计了一种新的基于对抗模仿的知识蒸馏方法,目的是从句子中获取知识来进行事件抽取。该方法首先构建了一个教师模块,充分利用标注数据来学习知识表示,之后建立相应的学生模块用于测试。在训练过程中,鉴别器通过检测教师模块和学生模块的输出来区分两者。同时,学生模块会尽可能地模仿教师模块,生成与教师模块相似的向量来迷惑鉴别器。该方法有效地完成了知识蒸馏,得到的新模型参数量少,且性能接近复杂模型。

Hong Y等人^[18]利用具有自我调节机制的生成式对抗网络来完成事件抽取任务,提高事件抽取的性能。一般来说,生成模型产生的虚假特征往往来自语义上的伪相关上下文,在训练过程中神经网络可能会错误地、不自觉地保留记忆,从而产生虚假的特征。因此该模型采用了一种双通道自调节的学习策略来调节学习过程,还添加了一对生成判别模型。在自学习过程中,生成模型被用来生成虚假特征,而判别模型被用来消除错误。该方法减轻了虚假特征对结果的影响,提升了事件抽取的效果。

3.2 开放域事件抽取

开放域事件抽取和限定域事件抽取的不同是事件类型不需要预先指定。因此,开放域事件抽取一般没有人工标注数据,主要使用无监督的方法。这种方法主要基于统计的思想,当事件触发词和事件论元相似时,其表达的事件类型也相似。在此过程中,聚类方法和概率模型的使用较为广泛。

开放域事件抽取的目标是抽取所有类型的事件,不对事件类型进行限定,有更高的事件覆盖率。Araki J等人^[19]提出了一种基于远程监督的自动生成训练数据的方法,不依赖人工标注的数据。该方法利用WordNet^[20]进行远程监督,以此来生成高质量的事件分类数据。之后,生成的数据被用来对事件判别器进行训练。该方法可以生成大量的训练数据,但是训练数据的准确程度无法保证,会导致判别器的结果较差。

除此之外,目前贝叶斯图形模型(Bayesian graphical model)被用于对开放域事件进行结构化的表示,但是这种方法假定所有的单词都是由单个事件产生的,具有一定的局限性。因此Wang R等人^[21]提出了一种基于生成式对抗网络的事件抽取模型,称之为对抗神经事件模型(adversarial-neural event model, AEM)。该模型使用狄利克雷分布对事件进行建模,并利用生成器网络捕捉潜在事件。该模型可以用来处理不同长度的文档,适用范围比较广泛。另外,Naik A等人^[22]将对抗域自适应(adversarial domain adaptation, ADA)框架应用到开放域的事件触发词识别任务中。该方法是一个无监督的过程,不需要目标域的标记数据,实验结果显示了该框架的有效性。

事件抽取技术是用于构建事件图谱的基础技术,影响着后续任务的精度,是极其重要的。针对这一任务,研究者提出了众多新方法。研究者期望通过不断探索,设计出更多适合此任务的神经网络,不断提高事件抽取的准确率。虽然目前已有的方法取得了不错的结果,但是该领域的评估标准还不是很明确,所使用的标准数据集规模也比较小。未来需要进一步明确评估标准,提出规模更大的标准数据集用于研究。

4 事件信息补全

事件抽取技术主要用于获取事件知识,但是事件抽取的结果往往是不完整的,存在部分论元缺失、论元抽取不准确等情况。为了对事件抽取的结果进行补全,就需要用到信息补全技术。信息补全技术依托于相应的事件图谱,根据图谱中的相关知识对缺失的知识进行补全。目前,信息补全和链接预测也是知识图谱领域的研究热点,但是研究者们主要关注二元关系事实。二元关系事实通常表示为三元组,即(头部实体,关系,尾部实体)。而在事件图谱当中,存在众多的多元关系事实。多元关系事实一般通过引入虚拟实体分解为多个三元组,例如Freebase^[23]中的复合值类型(compound value type, CVT)实体。目前研究者已经开展了对多元关系推理的研究,提出了一些方法。而这些多元关系推理方法可以应用在事件信息补全任务中,多元关系事实的预测对于事件图谱的补全具有重要意义。

在将实体转换为多个三元组实体的过程当中,Wen J等人^[24]认为在转换过程中会造成结构信息的丢失,可能导致链接预测不准确。因此他们提出了一种基于翻

译的方法m-TransH来对这些实例进行建模。在该方法中,关系(二元关系或多元关系)是由对应于该类型关系的一系列角色到其值的映射来定义的,每个特定的映射都是此关系的一个实例。但是m-TransH并没有考虑在同一多元关系事实中各成分之间的相关性。在此基础上,Zhang R C等人^[25]提出了一种新颖的相关性关联嵌入(relatedness affiliated embedding, RAE)模型,还提出了一种可伸缩实例重构(scalable instance reconstruction, SIR)算法。RAE通过对各成分之间的相关性进行互补建模,进一步改进了m-TransH方法。这里的相关性是指两个值在共同实例中共同参与的可能性。

而Guan S P等人^[26]并没有对多元关系事实进行分解,而是将每个多元关系事实表示为一组角色-值对。他们提出了一种对多元关系数据进行链接预测的方法NaLP,该方法对同一多元关系事实中所有角色-值对的相关性进行了建模。在该方法中,他们认为一个角色与其值是紧密相连的,因此应该绑定在一起。同样地,如果在同一个组中的所有角色-值对都紧密相关,就假定这组角色-值对很有可能构成有效的关系事实。基于上述假设,NaLP方法中包含两个关键组件,分别是角色值对嵌入和相关性评估。对于给定的关系事实,需要首先学习到角色-值对的嵌入表示,再在相关性评估组件中计算角色-值对的成对相关性的。之后还需要估计关系事实的整体相关性,获得最终的评估得分,并用于计算损失函数。该方法有效地建模了多元关系事实中角色和值之间的相关性,充分利用了多元关系事实中的内部信息。另外,由于公开可用的多元关系数据集有限,他们基于Wikidata中的原始数据构建了一个实用的多元关系数据集WikiPeople。

Guan S P等人^[27]还进一步提出了神经

网络模型NeuInfer, 用于对多元关系事实进行知识推断。NeuInfer不仅可以用于推断整个事实的未知元素, 还可以用于新型任务, 进行灵活知识推断。该模型使用的事实是由主要三元组以及任意数量的辅助描述组成的。他们假设一个有效事件的主三元组是有效的, 而不管其有没有辅助描述。另外, 每个辅助描述都与主三元组相关, 可以作为主三元组的某个特征。该模型首先对主三元组的有效性进行评估, 得到相应的有效性得分。之后再对主三元组与描述信息的兼容性进行评估, 得到兼容性得分。最后, 模型将有效性得分和兼容性得分的加权和作为最终的分值。该方法主要利用了事实中的主三元组以及辅助描述, 可以根据部分知识进行知识推理。

在事件信息补充的过程中, 不仅需要二元关系事实进行补充, 还需要对多元关系事实进行补充。目前多数研究者认为将多元关系事实分解为多个三元组会导致多元关系事实内部的关系信息丢失, 可能会加剧信息补充精度不够高的问题。因此, 上述方法都将多元关系事实看作一个整体, 保留了原有数据的信息。目前用于事件信息补充的方法还比较少, 补充精度还不够高, 需要进行深入研究。

5 事件关系推断

对于一篇给定的文本, 文本中存在多个事件。与此同时, 事件之间有可能是相互关联的, 可能存在多种关系。事件关系推断技术则利用文本中的信息来推断事件之间的关系, 主要包括共指关系、时序关系以及因果关系等。早期基于规则的方法实现简单, 但依赖于人工制定的规则, 实用性不高。随着深度学习技术的发展, 众多深

度学习方法被应用到关系推断任务中, 本文简要介绍了相关的方法。

5.1 事件共指关系

事件共指关系指的是给定描述事件的文本, 如果两个事件指向同一个事件实例, 则这两个事件之间存在共指关系。当文本中的多个事件指向一个事件实例时, 则多个事件组成了一个共指事件链。在共指事件链中的任意两个事件都具有共指关系。事件共指关系可以分为文档内事件共指和跨文档事件共指。进行事件共指关系的推断有助于计算机更好地理解事件发展的脉络, 对于综合全局信息、推测事件演变、预测未来事件具有重要的意义。

Zeng Y T等人^[28]提出了一种新的基于事件转述和论元感知语义嵌入的EPASE模型。该模型会在特定事件的上下文中识别深层次的转述关系, 并且可以涵盖更多情况的事件转述。另外, 由于自变量角色的嵌入被编码为事件嵌入, 无须依赖固定数量和类型的自变量, 因此该模型具有良好的可伸缩性。该方法首次将转述关系引入了事件共指任务, 通过句子中的完整语义信息来识别句子之间的转述关系, 充分利用了上下文信息。该方法有效地将自注意力机制和特殊事件的标识结合起来, 只对选取的特定事件进行关注, 排除了文本中噪声信息的干扰。

5.2 事件时序关系

事件时序关系是指两个事件发生的先后关系。时间是事件的一个天然属性, 事件发生时间的不同揭示了事件之间存在的先后关系。通过对时序关系的分析可以获取事件从开始到结束的演化过程, 有助于

对未来事件进行预测。事件时序关系抽取是自然语言处理领域中的重要研究方向,受到了越来越多研究者的关注。

Cheng F等人^[29]提出了一个基于双向长短期记忆的神经网络模型进行时序关系抽取。该方法使用了句子的依存路径,将词向量、词性向量和依存关系向量拼接,并将拼接后的向量作为模型输入的向量。在此基础上,为了解决跨句子实体之间不存在依存路径的问题,假设两个相邻的句子共享一个根节点。通过在模型中加入多种特征,该方法的实验结果取得了有效的提升。Han R J等人^[30]为了解决在事件抽取和事件时序关系识别两阶段任务中存在的误差传递问题,提出了一种联合学习方法,同时进行事件抽取以及时序关系识别。他们在两个子任务中共享了事件表达,利用结构化约束和整数线性规划来优化问题,提升了事件表达效果,缓解了误差传递的问题。另外,他们还提出了深度结构化学习框架^[31]的方法,利用递归神经网络学习事件的时序关系表示,同时采用结构化支持向量机(structured support vector machine, SSVM)进行预测。该方法在多个数据集上取得了优秀的结果,作者还通过消融实验进行了广泛的误差分析。

在此基础上Han R J等人^[32]还指出,之前的方法中只利用了例如硬约束的结构化知识,同时因为训练数据有限,进行时序关系分类时会引起偏差。因此他们提出了一个新的框架,通过概率领域知识构建的分布约束来增强深度神经网络的性能。新的方法还将拉格朗日松弛方法应用到时序关系抽取任务中,取得了最优性能。Wang H Y等人^[33]设计了一种新的联合约束学习框架,利用时间和子事件关系之间的逻辑约束对模型进行正则化。同时,他们设计了新的事件复合结构,用来描述事件提及之间的关系结构。

5.3 事件因果关系

事件因果关系是指两个事件之间的因果联系,是一种复杂的语义关系。通过对事件因果关系进行分析,事件发生的前因后果会更加清晰。清晰的因果关系有助于进一步认识事件发展的过程,使得事件推理与事件预测更加准确,可以避免风险、提高收益,为决策者提供有力的支持。事件因果关系识别在事件预测中的重要性,吸引了众多学者对其进行研究。

因为在事件关系推断任务中缺少标准数据集,所以Caselli T等人^[34]提出了一个用于事件时序和因果关系检测的数据集ESC(event storyline corpus)。ESC数据集的构建为之后的研究提供了便利。针对文档级事件因果关系的识别任务,Gao L等人^[35]建模了粗粒度和细粒度的文档级因果结构,在ESC数据集上取得了不错的效果。他们设计了丰富的特征进行事件因果关系识别,包括词法特征、因果潜在特征、句法特征,还利用事件共指链接将效果较差的句间预测问题转换为句内预测问题。最后,他们使用整数线性规划的方法来进行文档级全局推理,用来抽取文档中所有事件对的因果关系。该方法专门用于对事件因果关系进行识别,但是其并没有对因果关系的方向进行区分。

Liu J等人^[36]还利用外部知识来提升事件的表示,设计了一种事件提及屏蔽机制来挖掘历史数据中存在的因果关系。该方法将ConceptNet^[37]的知识引入了原有文本数据中,扩展了已有事件,涵盖了更多的信息。另外,为了处理数据中新出现的事件,该方法利用事件提及屏蔽机制来发现因果关系模式,增加对上下文信息的关注。最后,模型对两个组件做了权衡,用来充分发挥两个组件的优点。实验结果显示,该

方法效果显著,在跨主题应用中也展现出很强的鲁棒性。

事件关系推断用于对事件之间的关系进行判断,在此基础上可以对事件图谱进行完善,更好地刻画事件之间的关系。之前研究者主要利用规则的方法进行事件关系推断,现在已经引入了深度学习的方法,取得了不错的效果。但是总体来说,目前对事件关系推断技术的研究还不够充分,尤其是事件因果关系这一方面。在事件因果关系的研究领域中缺少明确的定义以及足够的标准数据集,这就需要研究人员在此领域中投入更多的精力,广泛开展研究。

6 事件预测

事件预测主要指的是根据历史事件来预测未来发生的事件。对未来事件进行准确预测具有十分重要的意义,既可以减少突发事件带来的损失,也可以针对未来事件做出相应的应对部署,为整个社会带来巨大的经济效益。但与此同时,对未来事件进行预测是十分困难的,需要面临很多的问题。这是因为对于许多事件来说,很难知道其发生的机制以及其背后的原因。随着大数据时代的来临,相关数据的规模不断扩大,为事件预测的研究提供了基础。因此,众多研究者开始探索如何进行事件预测,提出了众多有意义的方法。后文主要介绍脚本事件预测技术和基于图卷积网络的事件预测技术。

脚本事件预测技术根据给出故事的上下文来推断出故事的结尾。在这一任务当中,事件是用脚本结构描述的,即利用脚本描述事件的参与者、事件类型、事件触发词等多个元素,之后利用现有脚本对未来的事件进行预测。脚本事件预测是由Chambers N等人^[38]在2008年提出的,需要

在候选事件列表中选择最有可能发生的事件,利用填空式的评估标准来评估模型,这种思路被称为统计脚本学习。Chambers N等人^[38]设计了点互信息(pointwise mutual information, PMI)来计算事件对之间的关系。

目前脚本事件预测任务广泛使用的评测方法是多选项完形填空(multiple choice narrative cloze, MCNC),该方法是由Granroth-Wilding M等人^[39]提出的。在MCNC任务中,完整的事件链中存在某处缺失,给定多个候选的后续事件,模型需要从多个候选项中选择最符合逻辑的一项内容。Granroth-Wilding M等人^[39]采用深度学习的方法,设计了一种神经网络来对事件链进行建模。他们设计了一个EventComp模型,使用孪生神经网络(siamese network)代替原有的PMI。模型结构包含3层,分别是输入层、论元组合层和事件组合层。最终,训练后的模型与统计学习方法相比取得了极大的进步,证明深度学习方法在这一领域是有效的。

上述方法对事件对之间的关系进行了建模,并取得了不错的效果,但是文本中事件之间丰富的连接信息并没有被充分地利用。正因如此,研究者开始利用图的方法对事件之间的连接信息进行建模。在事件构成的图中,图的节点表示单个事件,节点之间的边表示事件之间的关系。Li B等人^[40]提出了一种众包的构建事件图的方法。另外,Glavaš G等人^[41]提出了一种从文本中构建事件图的有效方法。在此基础上,Li Z Y等人^[42]提出了一种新的方法来充分利用事件之间稠密的连接信息。该方法首先构建了叙事事理图谱,之后基于事理图谱进行网络表示学习,最后再利用得到的表示向量预测后续事件,模型结构如图3所示^[42]。具体来说,在构建叙事事理图谱的过程中,首先需要抽取脚本事件链

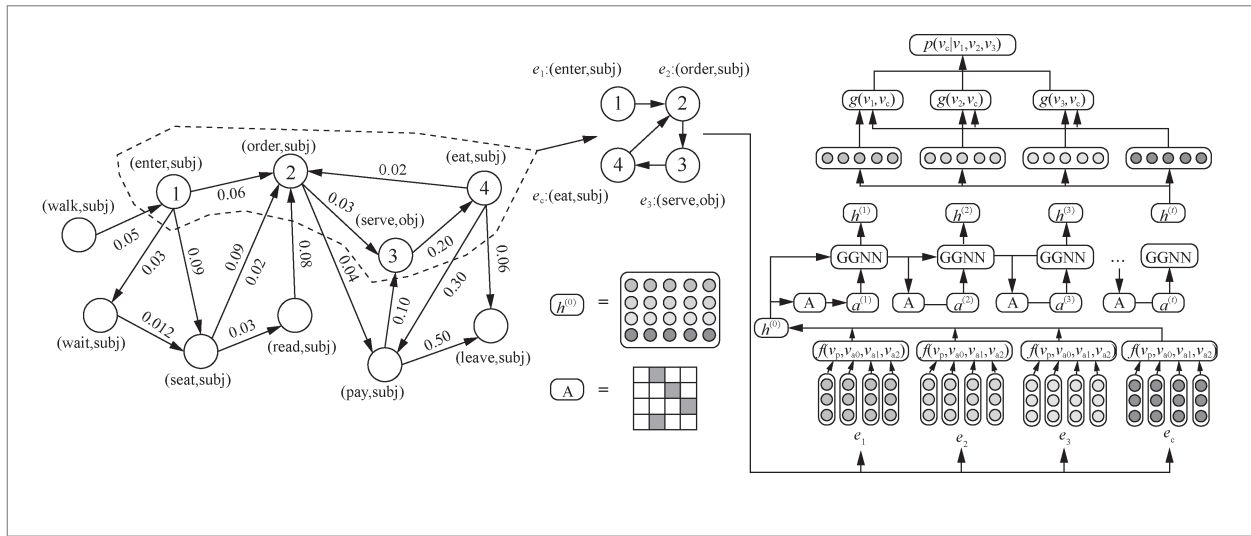


图3 SGNN 模型结构^[42]

条，之后利用抽取出来的链条构建图谱。在构建好事理图谱之后，模型每次会从中抽取一个子图，学习并更新所选子图上的事件表示。其中每个子图都包括故事上下文、所有候选事件的节点以及这些节点之间的有向边。该方法通过构建好的事理图谱来学习事件表示，有效地利用了事件之间的关系。但是构建事理图谱的过程较为复杂，构建好的图谱规模较大，每次只能选取一个子图来更新子图上的事件表示。

基于图卷积网络的事件预测技术主要指的是在时序知识图谱上利用图卷积网络进行推理的技术。时序知识图谱主要由四元组构成，是在原有知识图谱的三元组上添加了时间维度得到的。通过引入时间维度，时序知识图谱可以更好地描述事件信息。因此对时序知识图谱的未来状态进行推理就是对未来的事件进行预测。

Han Z等人^[43]提出了一种全新的图霍克斯神经网络 (graph Hawkes neural network, GHNN)。该神经网络首次利用霍克斯过程^[44]进行建模，主要用于捕捉过去事件对未来事件的影响。通过捕捉时序知识图谱上潜在的动态关系，该神经网络

可以用于对大规模时序多关系型数据进行分析，更加准确地对未来某一时刻发生的事件进行预测。与此同时，Jin W等人^[45]提出了一种新的循环事件网络 (recurrent event network, RE-NET) 来预测将要发生的事件。该网络主要分为两个部分，一部分是对历史事件的信息进行循环编码，另一部分是对时序相邻事件的信息进行聚合。RE-NET充分利用了历史事件的信息、时序相邻事件的信息以及同时发生事件的信息。通过聚合多种信息，该方法给出了所有事件的联合概率分布，用来预测未来事件发生的概率。基于图卷积网络的事件预测方法可以利用历史数据中蕴含的信息，捕捉到历史事件对未来事件的影响。但是目前这种方法的准确率还比较低，需要开展进一步的研究。

对事件预测的研究是十分重要的，其具有广阔的应用场景，事件预测是对事件图谱的高级应用。目前在脚本事件预测任务中已经有了相对清晰的评测方法，研究者也提出了许多方法对事件链以及事件图进行建模。除此之外，研究者还提出了基于图卷积网络的事件预测方法，采用新的

方法对将来的事件进行预测。总体来说，研究者对事件预测的关注度较高，期待取得更多的研究进展。

7 事件图谱的应用场景

目前，知识图谱已经被应用到各行各业，成为人工智能技术的重要组成部分。但是知识图谱具有一定的局限性，只能描述与实体相关的知识。而现实世界中存在着大量的事件，事件是日常生活中不可或缺的一部分，充分利用与事件相关的知识会更加真实、具体地刻画现实世界。因此，事件知识构成的事件图谱具有广阔的应用前景^[9]。将事件图谱应用到人工智能技术中可以使产品的智能化水平更高，更加方便人们的生产和生活。具体来说，事件图谱主要有热点事件检测、事件脉络分析以及未来事件预测等应用场景。

7.1 热点事件检测

事件图谱可以用于对热点事件进行检测。热点事件发生突然，会在短时间内形成，而且时时刻刻在发生变化，传统的方法难以应对。而采用与事件图谱相关的分析技术可以对网络上的舆论内容进行分析，实时地捕捉热点事件。另外，还可以对行业热点、地域热点事件进行检测，从多个方面展现热点事件。该技术可以被应用到媒体机构中，将检测到的热点事件实时地推送给用户，让用户在第一时间获得更多的热点资讯。

7.2 事件脉络分析

事件图谱的另一重要应用场景是事件脉络分析。针对特定的事件，事件图谱可

以利用事件之间的关系，对事件的前因后果进行关联，形成事件发展的脉络，并展示给用户。具体来说，可以将事件图谱与搜索引擎结合，当用户对感兴趣的事件进行搜索时，就可以得到事件发生的来龙去脉，清晰地反映事件发展的脉络，极大地提高知识检索的效率，给用户提供更多的便利。

7.3 未来事件预测

事件图谱还可以在未来事件预测方面发挥重要的作用。通过分析历史事件的发展过程，可以预测未来可能发生的事件。在金融领域，可以基于历史事件掌握行业动态，预测行业发展的趋势。这样就能够准确地把握市场发展动向，针对市场的变化做出相应的调整。在舆情预警方面，可以在现阶段发生事件的基础上预测未来可能会发生的事件，对事件的演化过程做出预测^[46]。一旦事件发生不良演变或者有不可控的趋势，监测系统可以及时发出预警，由相关部门第一时间介入处理。未来事件预测在现实生活中有重要的作用，可以做到预测预警，及时规避风险，创造出巨大的社会价值。

8 事件图谱的研究展望

目前，在研究者的共同努力下，对事件图谱的研究取得了不错的进展，越来越多的学者开始关注这一领域。对事件图谱的研究有助于人们进一步了解事件的发展历程以及事件背后的影响因素，可以更加容易地获取知识，提高生产效率。虽然事件图谱具有广阔的应用场景，但是现在对事件图谱相关技术的研究还不够深入，距离事件图谱的大规模应用还有很长的路要

走,还面临很多新的挑战。未来事件图谱可向以下几个方向发展。

(1) 提升事件抽取与事件推理的精度

目前事件抽取的准确率还不够高,而事件抽取是事件图谱构建的基础,只有从文本中准确地抽取事件知识,才能推动后续技术的发展,因此需要进一步提升事件抽取的准确率。之后是事件推理,虽然研究者对其进行了众多有意义的探索,但是目前提出的方法也存在准确率低、限制条件多等问题。正因如此,需要继续进行广泛的研究,不断提高事件推理的精度,进行更加准确的事件预测。

(2) 自动构建事件抽取与事件关系推断数据集

目前在事件抽取与事件关系推断的研究中,缺少数据集是一个严重的问题。数据集的匮乏制约了事件抽取与事件关系推断技术的发展。因此,未来应当关注数据集的构造方法,目标是实现自动化地构建高质量的数据集。自动构建高质量的事件抽取、事件关系推断数据集具有重要的意义,是未来技术发展的重要方向。

(3) 广泛开展事件时序、因果关系研究

在事件图谱中,事件之间的时序关系和因果关系占据重要的地位,但是目前对其进行的研究还不够多。只有获取到准确的时序关系以及因果关系,才能推断出事件的发展历程,实现对事件的溯源。另外,事件预测的研究也离不开事件时序关系以及因果关系的支撑。因此,广泛开展事件时序、因果关系研究是十分必要的,也是十分有意义的,需要研究人员的共同努力。

(4) 研究事件间关系可信性的界定方法

事件之间的关系是事件图谱的重要组成部分,反映了复杂事件背后的关联关系。在事件推理的过程中,必然要用到事

件之间的关系。而事件之间的关系是否可信、如何界定事件间关系的可信性是个十分重要的问题,对事件推理技术的发展有重要的影响。因此事件间关系的可信性的界定方法值得研究者关注,需要在深入研究的基础上提出较为合理的界定方法。

(5) 深入研究复杂事件的推理

目前事件推理研究的内容大多是简单事件,缺乏对复杂事件的研究。为了更好地利用事件图谱,对复杂事件的研究是不可或缺的。因为现实世界中的事件错综复杂,不再是单一的简单事件,所以未来需要开展对复杂事件的研究,挖掘复杂事件之间的关联。只有这样才能不断提高事件推理技术的水平,最终实现对事件的智能分析。

9 结束语

事件图谱中包含大量的事件知识,现代社会智能化水平的提升需要事件图谱的辅助。事件图谱中的事件涉及的维度较多,导致事件图谱的构建与推理具有一定的难度。基于事件图谱的重要性,研究人员已经开始对这一领域展开积极的探索,推动着该领域的快速发展。本文简单介绍了知识图谱和事件图谱,从事件图谱的构建、推理与应用方面对事件图谱的研究现状做了总结。之后,本文给出了事件图谱具体的应用场景,并展望了事件图谱的研究方向。

目前,事件图谱是自然语言处理领域中的前沿研究方向。为了更好地发挥事件图谱的价值,利用事件图谱服务社会,需要更多研究者进行深入研究。希望本文可以为事件图谱相关的研究提供一些帮助。

参考文献:

- [1] COWIE J, LEHNERT W. Information extraction[J]. *Communications of the ACM*, 1996, 39(1): 80-91.
- [2] AMIT S. Introducing the knowledge graph[R]. *America: Official Blog of Google*, 2012.
- [3] 邹艳珍, 王敏, 谢冰, 等. 基于大数据的软件项目知识图谱构造及问答方法[J]. *大数据*, 2021, 7(1): 22-36.
- ZOU Y Z, WANG M, XIE B, et al. Software knowledge graph construction and Q & A technology based on big data[J]. *Big Data Research*, 2021, 7(1): 22-36.
- [4] 陈成, 陈跃国, 刘宸, 等. 意图知识图谱的构建与应用[J]. *大数据*, 2020, 6(2): 57-68.
- CHEN C, CHEN Y G, LIU C, et al. Constructing and analyzing intention knowledge graphs[J]. *Big Data Research*, 2020, 6(2): 57-68.
- [5] 孙镇, 王惠临. 命名实体识别研究进展综述[J]. *现代图书情报技术*, 2010(6): 42-47.
- SUN Z, WANG H L. Overview on the advance of the research on named entity recognition[J]. *New Technology of Library and Information Service*, 2010(6): 42-47.
- [6] ETZIONI O, BANKO M, SODERLAND S, et al. Open information extraction from the Web[J]. *Communications of the ACM*, 2008, 51(12): 68-74.
- [7] BERNERS-LEE T, HENDLER J, LASSILA O. The semantic Web[J]. *Scientific American*, 2001, 284(5): 34-43.
- [8] 赵军, 刘康, 何世柱. 知识图谱[M]. 北京: 高等教育出版社, 2018.
- ZHAO J, LIU K, HE S Z. *Knowledge graph*[M]. Beijing: Higher Education Press, 2018.
- [9] 项威. 事件知识图谱构建技术与应用综述[J]. *计算机与现代化*, 2020(1): 10-16.
- XIANG W. Reviews on event knowledge graph construction techniques and application[J]. *Computer and Modernization*, 2020(1): 10-16.
- [10] ROSPOCHER M, VAN ERP M, VOSSSEN P, et al. Building event-centric knowledge graphs from news[J]. *Journal of Web Semantics*, 2016, 37: 132-151.
- [11] LIU S, CHEN Y, LIU K, et al. Exploiting argument information to improve event detection via supervised attention mechanisms[C]//*Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: ACL Press, 2017: 1789-1798.
- [12] CHEN Y, YANG H, LIU K, et al. Collective event detection via a hierarchical and bias tagging networks with gated multi-level attention mechanisms[C]//*Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: ACL Press, 2018: 1267-1276.
- [13] YANG S, FENG D, QIAO L, et al. Exploring pre-trained language models for event extraction and generation[C]//*Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: ACL Press, 2019: 5284-5294.
- [14] WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(1): 4-24.
- [15] NGUYEN T, GRISHMAN R. Graph convolutional networks with argument-aware pooling for event detection[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI Press, 2018.
- [16] LIU X, LUO Z C, HUANG H Y. Jointly multiple events extraction via attention-based graph information aggregation[C]//*Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: ACL Press, 2018: 1247-1256.
- [17] LIU J, CHEN Y B, LIU K. Exploiting the ground-truth: an adversarial imitation

- based knowledge distillation approach for event detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2019: 6754–6761.
- [18] HONG Y, ZHOU W X, ZHANG J L, et al. Self-regulation: employing a generative adversarial network to improve event detection[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL Press, 2018: 515–526.
- [19] ARAKI J, MITAMURA T. Open-domain event detection using distant supervision[C]//Proceedings of the 27th International Conference on Computational Linguistics. New York: ACM Press, 2018: 878–891.
- [20] 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.
- [21] WANG R, ZHOU D Y, HE Y L. Open event extraction from online text using a generative adversarial network[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg: ACL Press, 2019: 282–291.
- [22] NAIK A, ROSE C. Towards open domain event trigger identification using adversarial domain adaptation[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL Press, 2020: 7618–7624.
- [23] 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. New York: ACM Press, 2008: 1247–1250.
- [24] WEN J, LI J, MAO Y, et al. On the representation and embedding of knowledge bases beyond binary relations[C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2016: 1300–1307.
- [25] ZHANG R C, LI J P, MEI J J, et al. Scalable instance reconstruction in knowledge bases via relatedness affiliated embedding[C]//Proceedings of the 2018 World Wide Web Conference. New York: ACM Press, 2018: 1185–1194.
- [26] GUAN S P, JIN X L, WANG Y Z, et al. Link prediction on N-ary relational data[C]//Proceedings of the World Wide Web Conference. New York: ACM Press, 2019: 583–593.
- [27] GUAN S P, JIN X L, GUO J F, et al. NeuInfer: knowledge inference on N-ary facts[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL Press, 2020: 6141–6151.
- [28] ZENG Y T, JIN X L, GUAN S P, et al. Event coreference resolution with their paraphrases and argument-aware embeddings[C]//Proceedings of the 28th International Conference on Computational Linguistics. Stroudsburg: ACL Press, 2020: 3084–3094.
- [29] CHENG F, MIYAO Y. Classifying temporal relations by bidirectional LSTM over dependency paths[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL Press, 2017: 1–6.
- [30] HAN R J, NING Q, PENG N Y. Joint event and temporal relation extraction with shared representations and structured prediction[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg: ACL

- Press, 2019: 434–444.
- [31] HAN R J, HSU I H, YANG M, et al. Deep structured neural network for event temporal relation extraction[C]// Proceedings of the 23rd Conference on Computational Natural Language Learning. Stroudsburg: ACL Press, 2019.
- [32] HAN R J, ZHOU Y C, PENG N Y. Domain knowledge empowered structured neural net for end-to-end event temporal relation extraction[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL Press, 2020: 5717–5729.
- [33] WANG H Y, CHEN M H, ZHANG H M, et al. Joint constrained learning for event-event relation extraction[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL Press, 2020: 696–706.
- [34] CASELLI T, VOSSSEN P. The event StoryLine corpus: a new benchmark for causal and temporal relation extraction[C]// Proceedings of the Events and Stories in the News Workshop. Stroudsburg: ACL Press, 2017: 77–86.
- [35] GAO L, CHOUBEY P K, HUANG R H. Modeling document-level causal structures for event causal relation identification[C]// Proceedings of the 2019 Conference of the North. Stroudsburg: ACL Press, 2019: 1808–1817.
- [36] LIU J, CHEN Y B, ZHAO J. Knowledge enhanced event causality identification with mention masking generalizations[C]// Proceedings of the 29th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2020: 3608–3614.
- [37] LIU H, SINGH P. ConceptNet—a practical commonsense reasoning tool-kit[J]. BT Technology Journal, 2004, 22(4): 211–226.
- [38] CHAMBERS N, JURAFSKY D. Unsupervised learning of narrative event chains[C]// Proceedings of ACL–08: HLT. Stroudsburg: ACL Press, 2008: 789–797.
- [39] GRANROTH-WILDING M, CLARK S. What happens next? Event prediction using a compositional neural network model[C]// Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2016.
- [40] LI B, LEE-URBAN S, JOHNSTON G, et al. Story generation with crowdsourced plot graphs[C]// Proceedings of the 27th AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2013: 598–604.
- [41] GLAVAŠ G, ŠNAJDER J. Construction and evaluation of event graphs[J]. Natural Language Engineering, 2015, 21(4): 607–652.
- [42] LI Z Y, DING X, LIU T. Constructing narrative event evolutionary graph for script event prediction[C]// Proceedings of the 27th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2018: 4201–4207.
- [43] HAN Z, WANG Y Y, MA Y P, et al. The graph Hawkes network for reasoning on temporal knowledge graphs[C]// Proceedings of the Automated Knowledge Base Construction. [S.l.:s.n.], 2020.
- [44] HAWKES A G. Spectra of some self-exciting and mutually exciting point processes[J]. Biometrika, 1971, 58(1): 83–90.
- [45] JIN W, QU M, JIN X S, et al. Recurrent event network: autoregressive structure inference over temporal knowledge graphs[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL Press, 2020: 6669–6683.
- [46] 王瑞. 网络舆情事件知识图谱构建技术及应用研究[D]. 泉州: 华侨大学, 2020.
- WANG R. Research on construction technology and application of knowledge graph for internet public opinion event[D]. Quanzhou: Huaqiao University, 2020.

作者简介



胡志磊 (1996-), 男, 中国科学院计算技术研究所博士生, 主要研究方向为知识图谱、信息抽取、自然语言处理。



靳小龙 (1976-), 男, 博士, 中国科学院计算技术研究所研究员, 主要研究方向为知识工程、知识计算、知识图谱。



陈剑贇 (1977-), 女, 博士, 北京市信息技术研究所高级工程师, 主要研究方向为智能信息处理、系统工程。



黄冠利 (1975-), 女, 博士, 北京电子科技职业学院基础学院数学部副教授, 主要研究方向为计算数学、智能信息处理。

收稿日期: 2020-03-22

通信作者: 靳小龙, jinxiaolong@ict.ac.cn

基金项目: 国家重点研发计划资助项目 (No.2016QY02D0404); 国家自然科学基金资助项目 (No.U1911401, No.61772501, No.62002341, No.U1836206)

Foundation Items: The National Key Research and Development Program of China(No.2016QY02D0404), The National Natural Science Foundation of China(No.U1911401, No.61772501, No.62002341, No.U1836206)