

# 脚本事件预测：方法、评测与挑战\*

刘玉婷<sup>1,2</sup>, 刘茗<sup>2</sup>, 王保卫<sup>1</sup>, 丁鲲<sup>2†</sup>, 刘珊珊<sup>2</sup>, 刘浏<sup>2,3</sup>

(1. 南京信息工程大学 计算机学院 软件学院 网络空间安全学院, 南京 210044; 2. 国防科技大学第六十三研究所, 南京 210007; 3. 宿迁学院 信息工程学院, 江苏 宿迁 223800)

**摘要:** 相较于问答等经典的自然语言处理任务,脚本事件预测并未受到广泛关注。脚本事件预测旨在通过给定的上下文事件对未来事件进行准确预测,进而为后续可能发生的事件作出相应部署。为此,其研究具有重要意义,需要进行归纳与总结。针对脚本事件预测任务,首先给出脚本事件预测的基本概念;然后,梳理分析了事件表征中论元构成和事件表示学习两个子任务,其中事件表示学习从嵌入编码和语义增强两个方面叙述;接着,又以脚本建模方法为主线,从事件对、事件链、事件图以及结合型建模四个角度对其进行总结;并且,对现有模型进行测试,分析各类模型效果;最后,总结当前研究存在的问题,并展望未来可能的研究方向。

**关键词:** 自然语言处理; 脚本事件预测; 事件表征; 脚本建模

**中图分类号:** TP181 **文献标志码:** A **文章编号:** 1001-3695(2023)05-003-09

**doi:** 10.19734/j.issn.1001-3695.2022.09.0494

## Script event prediction: methods, evaluation and challenges

Liu Yuting<sup>1,2</sup>, Liu Ming<sup>2</sup>, Wang Baowei<sup>1</sup>, Ding Kun<sup>2†</sup>, Liu Shanshan<sup>2</sup>, Liu Liu<sup>2,3</sup>

(1. School of Computer Science, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. The Sixty-Third Research Institute of National University of Defense Technology, Nanjing 210007, China; 3. School of Information Engineering, Suqian University, Suqian Jiangsu 223800, China)

**Abstract:** Compared to classic natural language processing tasks like Q&A, script event prediction has garnered less attention. In order to deploy subsequent events appropriately, script event prediction intends to accurately forecast future events with a given contextual event. This study is quite important and requires summarization. Based on the script event prediction task, this paper firstly proposed fundamental concept of script event prediction. Then it described argument composition and event representation learning in event characterization. Also, event representation learning was shown from embedded coding and semantic enhancement; Next, this paper summarized script modeling approaches from event pair, event chain, event graph, and integrated modeling in detail. In addition, it examined the impacts of different models, and horizontally evaluated the effects of the existing models. Finally, this paper provided a summary of the issues with the present study, along with an outlook on promising future research areas.

**Key words:** natural language processing; script event prediction; event characterization; script modeling

理解文本中描述的事件对许多人工智能任务来说至关重要,脚本事件预测正是其中一个具有挑战性的任务。脚本的概念由 Schank 等人<sup>[1]</sup>提出的,用来理解自然语言文本。一个典型的案例是描述一个人去电影院的场景,“顾客进入电影院”“顾客买电影票”“顾客进入影厅”和“顾客观看电影”等一系列活动相继发生,这种结构化的表示方式称之为脚本。脚本事件预测是对这种结构化事件的演化推理。2008年,Chambers 等人<sup>[2]</sup>提出了脚本事件预测的概念,将其定义为通过给定场景中已经发生的一系列上下文事件,预测接下来最有可能发生的事件;另外,他们还提出了一种新的结构化事件表示方法,称为叙事事件链。叙事事件链是以一个共同的主角为中心的部分有序的事件集,其中参与者和事件的结构化序列被称为脚本。他们的工作为脚本事件预测的发展奠定了重要基础。

脚本事件预测的处理通常是从非结构化文本中提取出与事件相关的信息后,将信息以结构化的形式展现,再利用文本来推断出事件之间的关系,预测未来可能发生的事件。

## 1 事件表征

事件是现实世界中信息更为结构化的表示形式,从事件中获取有效的事件表征有助于推理出正确的后续事件。为获取有效的事件表征,首先需确定事件要素构成形式,再对事件表示进行学习。下面将针对这两个部分分别进行介绍。

### 1.1 要素构成

事件通常由动词和名词短语等要素组成,动词描述主角的动作,而名词短语则描述相关的实体。事件的组成要素包括动作或事件的参与者、事件发生的时间或地点等,可以通过自然语言处理(natural language process, NLP)工具从事件的文本中提取。Chambers 等人<sup>[2]</sup>针对事件表示的方法进行了创新。他们认为,虽然一个叙事事件链有多个参与者,但事件链只有一个核心角色即主角。因此,他们提出用(predicate; dependency)来表示一个事件,这种表示方式被称做 Predicate-GR。其中

**收稿日期:** 2022-09-26; **修回日期:** 2022-11-22 **基金项目:** 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);国家自然科学基金面上项目(61972207);国家自然科学基金—通用联合基金重点项目(U1836208)

**作者简介:** 刘玉婷(1998-),女,江苏泰州人,硕士研究生,主要研究方向为知识图谱、事件推理;刘茗(1995-),女,山东威海人,助理研究员,硕士,主要研究方向为知识图谱、智能算法;王保卫(1982-),男,山东聊城人,教授,博导,博士,主要研究方向为物联网数据安全、可信数据交易、信息隐藏、数字水印;丁鲲(1978-),男(通信作者),江西景德镇人,研究员,硕士,主要研究方向为知识图谱、数据挖掘(dingkun18@nudt.edu.cn);刘珊珊(1994-),女,甘肃兰州人,助理研究员,硕士,主要研究方向为知识图谱、机器阅读理解;刘浏(1988-),男,江苏宿迁人,讲师,博士,主要研究方向为人工智能、知识图谱、信息安全。

predicate 指事件描述中的谓语动词, dependency 指动词和主角之间的语法依赖关系, 如“主语”“宾语”或“介词”等。

文献[2]所提出的事件表示方法虽然为后续的很多研究工作奠定了重要的基础, 然而这种表示也有其不足之处。由于某些事件文本缺乏一致的主题, 所以不同角色的事件就会被错误地混合到单个角色中。对于“X 去吃饭”和“Y 去度假”, 它们的表示方法都是 <go, obj>, 所以“go”表示的是“去吃饭”还是“去度假”就无从得知, 且无法区分事件的主角是 X 还是 Y。所以, Balasubramanian 等人[3]提出了 Rel-grams, 其事件表示模式为 (Arg1; relation; Arg2), 这种关系三元组提供了更为具体的事件表示方法。其中 Arg1 表示事件的主语, Arg2 表示事件的宾语, relation 表示事件的核心动词。例如, “He cited a new study that was released by UCLA in 2008 (他引用了加州大学洛杉矶分校 2008 年发布的一项新研究)”, 事件被表示为“(He, cited, a new study) (a new study, was released by, UCLA)”。虽然这种事件表示更为具体, 但增加了事件表示的稀疏性。为了减少稀疏性以及提高事件表示的泛化性能, 他们对事件表示使用规范化的关系字符串, 上述事件则表示为“( <person>, cite, study) (study, be release by, UCLA)”。

以上工作主要针对的是单个主角事件。对于多主角事件, 例如“Mary emailed Jim and he responded to her immediately (Mary给 Jim 发了邮件, Jim 立即回复了她)”, 如果根据所涉及的实体进行分组就会产生两条事件链。第一条以 Mary 为实体, 其事件链表示为 (email, subject) (respond, object); 而对于 Jim, 事件链则表示为 (email, object) (respond, subject)。可以注意到单个动词可能会产生多个事件对, 并且离散的事件对之间彼此没有联系。因此, Pichotta 等人[4]提出带有多个论元的事件表示方法, 这种方法能够直接捕获多个实体之间的交互。他们将事件表示为  $v(e_s, e_o, e_p)$ , 其中  $v$  表示事件的核心动词,  $e_s$  表示事件的主语,  $e_o$  表示事件的宾语,  $e_p$  表示事件的介词宾语, 除  $v$  之外的任何参数都可以是 null(用-表示), null 表示所指参数与动词没有语义依赖关系。上述两条事件链则表示为 email(Mary, Jim, -) 和 respond(Jim, Mary, -)。由此可见, 这种事件表示方法可以获取更多的文本事件结构, 更能表达一个事件的具体含义, 非常适合作为脚本的表示。

接下来的研究也在不断改进多论元事件表示。文献[4]提出来的多论元事件表示方法虽然可以提高预测的准确性, 但是这种更丰富的事件表示方法仍有一些不足。事件表示是基于共指事件链的, 链中的实体在内部映射到变量。但是对于长度为 1 的实体而言, 所有单个实体映射到另一个公共类别后, 关于这些实体的信息都丢失了。以 order(people, food, -) 和 order(people, car, -) 为例, 如果所给文本中没有出现 food 和 car, 那么这两个事件无法进行区分。基于此问题, Ahrendt 等人[5]提出了基于参与者的模型。模型将文本表示为由参与者标记的事件链(PLE), PLE 由一个动词和它的参与者标签组成, 它的一般形式是  $\text{verb}(p_{\text{subj}}, p_{\text{dobj}}, p_{\text{iobj}})$ , 其中  $p_{\text{subj}}$ 、 $p_{\text{dobj}}$  和  $p_{\text{iobj}}$  不是直接表示事件的主语、直接宾语和间接宾语, 而是分别代表着主语、直接宾语和间接宾语的参与者标签。例如“The waitress brought us some water(服务员给了本文一些水)”, 其中直接宾语 water 的参与者标签就对应着 customer, 相应的 PLE 表示形式为 bring(waiter, drink, customer)。实验表明, 明确地标记事件参与者可以提高事件预测性能。为进一步地提高预测性能, Pichotta 等人[6]将事件表示为五元组  $v(e_s, e_o, e_p, p)$ , 与以往的事件表示方法的不同之处体现在多了一个介词论元  $p$ 。

事件中还有其他要素, 如情绪、有生性、事件时间和位置信息等, 这些要素都可以潜在地增强事件表示, 进而有效地表示事件。Lee 等人[7]提出特征事件嵌入学习(featured event embedding learning, FEEL)。FEEL 将事件表示为  $(\text{tok}(e), \text{sub}$

$(e), \text{obj}(e), \text{prep}(e), f_1(e), f_2(e))$  的形式, 它包含四个基本组件和两个细粒度属性, 其中  $\text{tok}(e)$  是将谓词及其依赖关系与事件联系起来的标记, 另外三个基本组件  $\text{sub}(e)$ 、 $\text{obj}(e)$  和  $\text{prep}(e)$  含义与之前工作类似, 分别代表事件的主语、宾语和介词宾语。两个细粒度属性  $f_1(e)$ 、 $f_2(e)$  分别代表句子级情感属性和事件主角的生命性信息。 $f_1(e)$  包含三种情感属性标签, 分别为消极、中立和积极,  $f_2(e)$  同样包含三种生命性信息类型, 分别为有生命、无生命和未知。以“Jenny went to a restaurant and ordered a lasagna plate(Jenny 去餐馆点了一份千层面)”为例, FEEL 可以捕获到以 Jenny 为事件主角的两个事件表示形式, 分别为  $((\text{go}, \text{subj}), \text{Jenny}, -, \text{restaurant}, \text{中立}, \text{有生命性})$  和  $((\text{order}, \text{subj}), \text{Jenny}, \text{plate}, -, \text{中立}, \text{有生命性})$ 。除了事件主角, 一个事件可能包含多个参与者, 每个参与者对事件的发生都有一定的影响。因此, 为了获得更全面的事件语义表示, Bai 等人[8]将事件出现的原句子集成到事件表示中, 即将事件表示成  $(v, a_0, a_1, a_2, t)$ , 其中  $v$ 、 $a_0$ 、 $a_1$  和  $a_2$  分别代表动词、主语、宾语和间接宾语, 而  $t$  则代表事件出现的原句。

最初的事件表示主要针对单个事件主角来构建叙事事件链, 忽略事件脚本中的很多重要信息。为了解决这一问题, 研究了更为丰富的事件表示方法, 考虑事件中更多的要素, 以此获得更全面的语义表示。各类事件要素构成方法如表 1 所示。

表 1 要素构成分类

Tab. 1 Classification of argument composition

编号	要素构成	举例
1	$(\text{predicate}; \text{dependency})^{[1]}$	(go, obj)
2	$(\text{Arg1}; \text{relation}; \text{Arg2})^{[3]}$	( <person>, cite, study)
3	$v(e_s, e_o, e_p)^{[4]}$	email(Mary, Jim, -)
4	$\text{verb}(p_{\text{subj}}, p_{\text{dobj}}, p_{\text{iobj}})^{[5]}$	bring(waiter, drink, customer)
5	$v(e_s, e_o, e_p, p)^{[6]}$	((pass, route, creek, north, in)
6	$(\text{tok}(e), \text{sub}(e), \text{obj}(e), \text{prep}(e), f_1(e), f_2(e))^{[7]}$	((go, subj), Jenny, NULL, restaurant, 中立, 有生命性)
7	$(v, a_0, a_1, a_2, t)^{[8]}$	(email, Mary, Jim, NULL, Mary emailed Jim)

### 1.2 事件表示学习

为保留事件丰富的语义信息, 需将结构化的信息表示为计算机可以理解的形式, 即事件表示学习。通常是采用编码的思想, 使事件表示中尽可能地保留事件元素的信息。另外为了获取更准确的事件表示, 还会考虑在事件表示中融合外部知识, 即语义增强。

#### 1.2.1 嵌入编码

事件表示通过事件结构对事件元素的向量进行语义组合, 计算事件向量表示。按照组合方式的不同, 可以分为基于词向量与基于神经网络的语义组合方式。

##### 1) 基于词向量的语义组合

语义组合的方式最简单莫过于自然组合。以“人民日报”为例, 将“人民”与“日报”进行自然组合后得到了“人民日报”。但由于“日报”可能也是一份报纸的名称, 所以它们组合后可能会得到另一份不同的报纸, 所以使用自然组合的处理方式不能用来表示多个单词组成的短语。

研究发现, 使用向量来表示整个短语会有效避免这种情况, 即将词转换成一个分布式表示。基于此, Mikolov 等人[9,10]提出了一种高效训练词向量的模型—word2vec, 来学习短语向量表示, 根据词汇的“共现”信息将词汇编码成一个向量。在一定程度上, 他们的工作是对用矩阵向量表示短语的补充, 推进了推理工作的发展。除了提出 word2vec 来学习短语向量表示, Mikolov 等人[9]还发现了向量的“加法”属性, 即简单的向量加法可以产生有意义的结果, 如  $\text{vec}(\text{“中国”}) + \text{vec}(\text{“首都”})$  的结果接近  $\text{vec}(\text{“北京”})$ 。通过将事件元素的词向量进行相加或拼接后映射到事件向量空间, 可以充分利用词向量信

息,但在建模事件元素间的交互上较为薄弱。具体来说,通过与其相邻词的共现频次统计得到的向量难以捕获较长短语的语义,因此限制了其对句意的理解。

## 2) 基于神经网络的语义组合

虽然文献[9]提出的向量的“加法”属性可以获得有意义的结果,但利用“加法”属性很难针对事件表面形式的细微差异进行建模。例如,“She throw basketball(她扔篮球)”与“she throw bomb(她扔炸弹)”会得到相近的向量表示,但是这两个事件在语义上并不相近。为了解决这一问题,Weber等人<sup>[11]</sup>提出了基于张量(tensor)的神经网络组合模型,核心思想是:事件嵌入由主语元素和宾语元素的乘法之和构成,其中的权重取决于谓语,通过使用权重来获取事件中谓宾的关键信息。与以隐式或相加的方式学习语义组合相比,该模型可以通过乘法组合事件的语义。这种方式下,即使事件论元只有细微的表面差异,也能够在事件表示中体现出语义上的较大差别。

随着神经网络的发展,它被广泛应用于语义特征提取任务上,从文本数据中提取特征。Socher等人<sup>[12]</sup>提出一种用于组合语义的递归神经网络模型(matrix-vector recursive neural network, MVRNN),将每个单词和短语都用一个向量和一个矩阵表示,通过解析树的语法结构,自下而上递归组合单词,从而学习任意语法类型和长度的短语或句子的组合向量表示。2014年,Modi等人<sup>[13]</sup>提出使用一个简单的组合模型来表示事件谓词及其论元的语义,然后通过组合神经网络将事件单个论元的向量表示组合得到事件的向量表示。类似地,Modi等人<sup>[14]</sup>又提出在计算事件表示的过程中,参数和模型的排序组件是从文本中联合估计的,可以用事件排序任务来学习嵌入和评估模型。由于事件不是一个牢不可破的单位,而是由一些包含谓词及其参数的可分离的组件组成,所以所有的组件可以表示为嵌入向量,并从预测原型事件排序中学习得到,组件的嵌入是在相同的向量空间中,再通过组合神经网络获得整个事件的向量表示。文献[15]则是联合使用 word2vec 和组合神经网络来学习事件嵌入,将事件元素的词向量拼接后,输入多层全连接神经网络,对事件元素的词向量进行组合。文献[16]认为,事件是由一系列子事件组成,而且事件通常在两个层次上表现出顺序结构:a)描述特定子事件的词按语义顺序排列;b)属于同一事件的多个子事件按顺序进行。由于这种顺序结构的存在,给准确表示一个子事件带来了难度,所以,Hu等人<sup>[16]</sup>提出上下文感知的层次长短期记忆网络(context-aware hierarchical long short-term memory, CH-LSTM),用于对未来的子事件进行预测。CH-LSTM模型是一个两层的LSTM架构,第一层用来编码子事件,将每个子事件映射到一个向量空间中;第二层则是对观察到的子事件序列进行编码,其中还结合了上下文的主题特征。

### 1.2.2 语义增强

基于神经网络的方法核心在于将事件嵌入到向量空间进行编码。嵌入的向量保留了事件中的语义信息,克服了事件元素建模较为薄弱的问题。虽然以嵌入编码的方式可以表达客观事实,但人类的主观情感可能会对客观事件的发生产生影响,且不同事件背后的意图也有所不同。为更好地建模事件语义,在嵌入编码工作的基础上进一步增加语义增强的工作,为事件表示学习带来了新的突破。

不考虑事件语义的情况下,难以区分事件之间微妙的差别。一方面,如果两个事件中单词重叠较少,就会被映射为距离较远的两个向量,例如,“X threw bomb”(X扔炸弹)和“Y attacked school”(Y袭击学校);另一方面,如果两个事件单词重叠较多,即使它们关联很小,也容易被映射为距离较近的两个向量,认为具有很高的相似度,例如,“X broke record”(X打破记录)和“Y broke glass”(Y打破杯子)。但从事件发生的意图

来看,“扔炸弹”和“袭击学校”都属于暴力事件,从而应具有很高的相似度;从事件参与者的情感来看,“打破记录”可能会很高兴,但“打破杯子”可能心情是很沮丧的,因此应被认为关联很小。为了避免这种情况,研究者在事件中考虑事件意图以及参与者情感等信息来对事件的语义进行增强。文献[7]提出的FEEL模型正是归属于语义增强的工作之一。模型所采用的六元组表示方法考虑到了事件背后的意图等细微信息,其表示方法可以更好地将文本信息用做常识性推理。类似地,2019年,Ding等人<sup>[17]</sup>为了更好地建模事件表示,提出学习事件表示时融入人的情感及意图等外部常识知识。在学习事件表示过程中使用张量神经网络(neural tensor network, NTN)<sup>[18]</sup>作为事件表示模型,模型使用双线性变换显式地建模谓语与主语、谓语与宾语及三者间的交互关系。研究表明,具有顺承关系的事件间情感的一致性可以帮助预测后续的事件,所以融入人的情感等外部常识性知识在脚本事件预测等任务上能取得更好的效果。

除了事件背后的意图等信息,语义增强还会考虑到事件之间的逻辑关系,如事件之间的因果关系。因果关系不仅是一种知识,也是推理和理解未知事件的基础。Zhao等人<sup>[19]</sup>尝试使用事件因果关系来预测事件,他们建立了一个抽象的新闻事件因果网络,从这个因果网络中可以得到一般的因果模式,另外还将因果网络嵌入到连续向量空间,简化了事件匹配过程,使其较易于其他应用。经观察发现,外部知识在理解和预测事件时起着重要的作用,为充分利用额外的事件知识,运用的方法可以分为使用预训练和使用图神经网络融合外部知识两种类型。针对将常识整合到语言模型中的预训练方法,文献[20]提出了K-ADAPTER,其保持了预训练模型的原始参数,支持持续的知识注入。关于整合外部知识的方法,文献[21]提出了“只限尾”“事件模板”和“关系嵌入”三种方法,将外部知识集成到模型中,结果表明,与不使用外部知识相比,三种方法的效果均有超0.44%的提升。虽然预训练的方法取得了巨大成功,但其缺乏可解释性,即难以明确说明使用到了训练语料库中的哪些知识。而图神经网络能提供良好的可解释性,因此文献[22]提出从事件图中学习知识,以引入一个潜在变量的方式对中间事件的信息建模,从而实现融合外部知识。

### 1.2.3 小结

本节针对事件表示学习方法中存在的问题与挑战进行了总结,如表2所示。

表2 事件表示学习方法

Tab. 2 Event representation learning method

事件表示学习	方法类别	问题与挑战	代表成果
嵌入编码	基于词向量	建模事件元素间的交互较为薄弱	文献[9,10]
	基于神经网络的语义特征提取	事件语义保留不充分	文献[11~16]
语义增强	使用意图、参与者状态等信息	将外部信息元素融合	文献[7,17~22]

无论是用嵌入编码的形式表示事件还是将外部信息融入事件表示进行语义增强,目的都在于将结构化的事件信息表示为机器可理解的形式。基于词向量的方法将词转换成稠密向量,优点在于包含更多的信息,但面临着元素间交互较为薄弱的问题;基于神经网络的方法使用神经网络模型从文本中提取特征进行编码,有效解决了上述问题,但由于事件中包含丰富的语义信息,所以缺乏对事件语义的充分保留;为保留事件的语义信息,在建模事件语义时考虑事件外部信息,即语义增强。

## 2 脚本建模

确定事件的表征后,对脚本进行建模。根据建模方法的不同,将其分为基于统计学习的方法和基于深度学习的方法。



### 2.1 基于统计学习

早期的脚本建模方法主要是统计学习方法,这类模型通过统计训练集中事件发生的概率来进行后续的预测。Chambers 等人<sup>[2]</sup>提出使用 PMI(pairwise mutual information)学习叙事关系,度量事件和叙事链之间的关系。PMI 统计训练集任意两个事件同时发生的频率作为这两个事件同时发生的概率,进行后续事件的预测。叙事链是一组叙述事件,由元组  $(L, O)$  构成,其中  $L$  由一系列的事件-关系对组成, $O$  是事件的部分时序关系。为了生成叙事事件链,首先将依存句法分析器与共指消解相结合收集事件脚本统计数据并预测脚本事件,对于其训练语料库中的每个文档,使用共指消解识别所有实体,并使用依存句法分析器识别具有实体作为主语或宾语的所有动词。另外将事件定义为动词加上依赖类型(主语或宾语),并为每个实体收集共同论元所参与的事件链。然后,计算在语料库中发生的事件链中所有事件对的 PMI,再通过最大化来找到下一个最有可能发生的事件。给定一个包含  $n$  个事件的事件链  $E = (e_1, e_2, \dots, e_{n-1}, e_n)$  以及一个候选事件  $c$ ,事件  $e_i$  采用 Predicate-GR 的表示方法。PMI 计算过程如式(1)~(3)所示,其中, $n$  是事件链中的事件数, $e_i$  代表着第  $i$  个事件, $m$  是训练语料库中的候选事件  $c$  的数量, $C(e_i, c)$  则是事件对  $(e_i, c)$  的共现次数。

$$PMI(E, c) = \sum_{i=1}^n \log \frac{P(e_i, c)}{P(e_i)P(c)} \quad (1)$$

$$P(e_i, c) = \frac{C(e_i, c)}{\sum_{e_i} \sum_{c_j} C(e_i, c_j)} \quad (2)$$

$$\max_{j:0 < j < m} \sum_{i=0}^n \text{pmi}(e_i, c_j) \quad (3)$$

由 PMI 可知,通过给定部分事件链,可以预测属于该脚本的其他事件。进一步地,Jans 等人<sup>[23]</sup>基于 C&J08 的工作提出了 Bigram。Bigram 采用二元条件概率来表示两个事件的关联强度,并按照观察事件链的顺序建模事件链。其建模思想为:以个别先前已有事件为条件计算后来的每个事件元组的最大似然概率。这种方法按照观察事件链的顺序建模事件链。

计算事件间的条件概率如式(4)所示(以  $e_1$  和  $e_2$  为例)。

$$P(e_1 | e_2) = \frac{C(e_1 | e_2)}{C(e_1)} \quad (4)$$

候选事件  $c$  的得分则是由上下文事件  $e_i$  和候选事件之间的平均概率得到的,如式(5)所示。

$$s(c) = \frac{1}{n} \sum_{i=0}^{n-1} P(c | e_i) \quad (5)$$

总结起来,PMI 和 Bigram 这类基于统计学习方法的模型是通过统计训练集中的事件发生的概率来进行后续预测,但对训练集中没有出现过的事件则无法进行预测。

### 2.2 基于深度学习

由于统计学习方法是简单地从概率论的角度给出每个事件发生的概率,而忽略了事件之间内在的联系。随着研究的进一步深入,基于统计学习的方法逐步被基于深度学习的方法所代替。神经网络具有强大的表征能力,同时数据中的信息可以在模型中进行有效编码。本节中根据任务发展阶段,将基于深度学习的脚本事件预测方法划分为基于事件对、事件链、事件图和结合型建模四个部分进行回顾。

#### 2.2.1 事件对

基于事件对的方法研究主要集中在对事件对的关联进行建模,其建模方法是候选事件  $c$  与整个脚本的关系拆解为和脚本中每个事件  $e_i$  之间的连贯性分数的聚合,公式如式(6)所示。

$$\text{coh}(c, e) = \text{Agg}_{e_i \in e} (\text{coh}(c, e_i)) \quad (6)$$

word2vec<sup>[9]</sup> 和 eventcomp 模型<sup>[15]</sup>正是其中的典型方法。word2vec 采用两个事件嵌入的余弦相似性作为事件对的关联

强度;eventcomp 模型则是由输入层、论元组合层和事件组合层组成,学习单词的嵌入向量表示和事件的向量表示,以此来预测两个事件是否会出现同一个事件链中。如图 1 所示,以两个事件为例。在模型中,谓词和论元进行非线性组合,再通过一个共享权重的孪生神经网络(siamese network)将两个事件的向量表示映射到新的空间,将其组合到一起,形成新的向量空间表示。最后利用损失函数计算出这两个输入事件在同一链中出现的关联强度。训练后的 eventcomp 模型与基于统计学习的方法相比,在 NYT 英文数据集上,Acc 指标提升了 9.05%,证明了将深度学习应用在脚本事件预测领域是有效的。

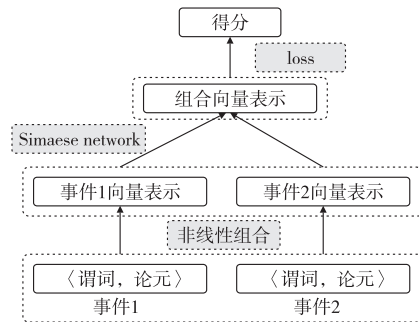


图 1 eventcomp 模型示例

Fig. 1 Example of eventcomp

基于事件对的方法重点关注事件和候选事件之间的相关性,因而可以用来应对更灵活的事件顺序。但它忽略了叙事事件链中各个事件之间的时序信息,而叙事事件链中各个事件的发生顺序相互影响,先发生的事情必然影响后续事件的预测。

#### 2.2.2 事件链

基于事件链的方法重点关注叙事事件链中各个事件之间的时序信息,将已知的事件组织成事件序列,并认为待预测事件是序列的下一个元素。而递归神经网络(RNN)则常被用来处理时序信息。原因在于 RNN 将上个时间节点的隐节点状态也作为神经网络单元的输入。但它也存在着一些问题,首先是“长距离依赖”问题,所谓依赖,即某个单词可能与其距离较远的某个单词具有强依赖关系。以图 2 为例,第一句话中,was 受 dog 影响所以是单数;而第二句话中,were 则是受 dogs 影响,并非它们之前的单词 food。当神经网络的节点经过许多阶段的计算后,之前较长的时间片的特征已经被覆盖,所以难以建立跨度较大的依赖性。由于句子之间跨度很大,RNN 则容易出现梯度消失问题从而捕捉不到句一和句二之间的依赖关系,最终造成语法错误。所以,为了解决 RNN 所存在的问题,提出了长短期记忆(LSTM)<sup>[24]</sup>。LSTM 引入了门控(gate)机制用于控制特征的流通和损失,从而解决了长距离依赖和梯度消失问题。

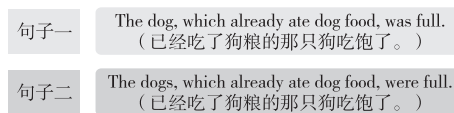


图 2 长距离依赖问题

Fig. 2 Long-distance dependencies

Pichotta 等人<sup>[25]</sup>率先提出将 LSTM 用于脚本事件预测任务,在任务中直接合并事件论元的名词信息,并在模型中引入 LSTM 和 GRU,实现编码远距离传播事件的同时,而不丢失历史信息。具体来说,他们将脚本学习任务定义为一个序列建模任务,使用训练模型来预测下一个输入,在每个时间步长中,都有一个事件组件被输入到 LSTM 模型中。在输入整个事件链后,模型将输出一个附加事件的预测。在推断已有事件和新事件这两项任务上进行评估,使用 LSTM 后的方法,表现均优于之前的方法。但随着事件文本数量的增多,序列模型进行推断的难度也会增大。进一步地,Pichotta 等人<sup>[26]</sup>提出事件链的原始文本可以被直接用来预测缺失事件。他们采用句子级 RNN

编码器—解码器<sup>[27]</sup>模型进行文本预测,在实验中将由原始文本训练的系统与由结构化事件表示的相同文本的 LSTM 模型进行了比较,研究结果表明两者结果相当,仅在 Acc 指标上有 0.1 的差别。

事件通常由一系列子事件组成,例如,地震事件就会包含险情预报、地震发生和救援工作等一系列的子事件,因此利用历史事件来自动预测未来的子事件预测是非常有必要的。Hu 等人<sup>[16]</sup>提出了一个端到端的模型 CH-LSTM (context-aware hierarchical long short-term memory),它无须手动标记特征,并且可以生成训练集中没有的新事件。CH-LSTM 正是使用了已发生的子事件进行预测,与文献<sup>[26]</sup>的工作类似,他们将事件链原始的子事件的文本作为输入,未来可能发生的子事件的文本描述作为输出。CH-LSTM 模型使用两级 LSTM 架构,在第一层,编码子事件后将子事件映射到另一个嵌入事件当中。而在第二层,则是对观察到的子事件顺序进行编码,另外还结合了上下文的主题特征来增强语义。由于 CH-LSTM 模型考虑了两个层次上的顺序结构,还合并了额外的上下文特征,在单词分类错误任务上和困惑度 (perplexity) 这两个指标上相比 LSTM 分别降低了 5.17% 和 73.26%,所以证明考虑事件的两级结构可以有效提升模型的预测能力。

事件链中存在着丰富的事件片段关系,若能充分利用这些事件片段之间的关系,则能提升后续事件预测的准确性。而自注意力机制<sup>[28]</sup>则可以用来提取不同的事件片段并将事件链表示成事件片段的组合,学习更准确的事件表示。基于此,Lyu 等人<sup>[29]</sup>提出 SAM-Net,采用动态记忆网络来提取事件链的特征,将事件链表示为一组事件片段。由于不同的个体事件或事件片段可能与后续事件有不同的语义关系,SAM-Net 使用两种注意力机制为每个个体事件和事件片段分配不同的对应权重,而模型对后续事件的预测则是基于这两种注意力机制的结合来实现的。SAM-Net 解决了两个很重要的问题:

a) 如何准确地表示事件链。由于事件链是由一系列事件组成,所以事件可能比句子中的单词更稀疏。研究者在使用自注意力机制,从事件链中获取不同的事件片段,为了减少冗余,再采用 DenseNet<sup>[30]</sup>进行特征提取,同时得到事件链的向量表示。

b) 如何将事件链和候选事件集成在一起并表示它们之间的关系。SAM-Net 模型如图 3 所示,首先使用事件级注意力机制匹配候选事件与单个事件,获取事件级的上下文表征;然后,用链级注意力机制匹配候选事件和事件片段,获得链级上下文表征;最后,整合事件级和链级上下文表征,与事件链表征进行交互预测后续事件。

鉴于事件表示学习的方法主要关注事件或链级的粗粒度连接,而忽略了事件之间更细粒度的连接。Wang 等人<sup>[31]</sup>提出了一种新的事件表示学习范式 MCer,通过集成事件在多个粒度级别上的连接,包括论元级别、事件级别和链级别,来增强事件的表示学习。之前的大部分工作仅考虑单一参与者所对应的事件链。在事件层面,以往的研究是将事件视为一个带有参与者的动词,而忽略了其他有用的属性,如参与者的意图和状态等;在脚本层面,大多数研究也是只考虑与一个共同的主角对应的单一事件序列,这种事件表示方法缺乏必要的信息来获得更准确的预测。因此,Bai 等人<sup>[8]</sup>提出了一个基于 Transformer 的模型,利用不同参与者所对应的叙事事件链来建模各自的行为趋势,称为 MCPredictor,该模型集成了深度事件级和脚本级信息。事件级信息包含描述事件的必要元素,如动词及其参与者,而脚本级信息则描述了事件是如何连接和结构化的,例如通过时间顺序或共同参与者。在事件层面,MCPredictor 保留文本中描述事件的所有组成部分来获得更全面的事件语义表示,它包含一个事件编码组件和一个文本编码组件,通

过对两个组件的输出进行聚合,就可以得到更全面的事件语义表示;在脚本层面上,通过注意力机制聚集了多个叙事事件链,以刻画不同参与者各自的行为趋势。集成了事件级和脚本级的信息后,模型预测性能实现了超越 11.45% 的提升。

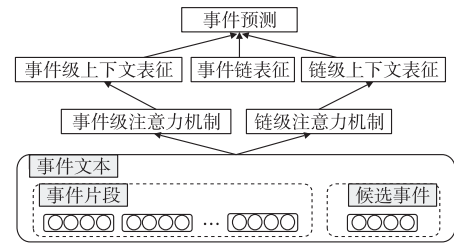


图 3 SAM-Net 模型示例

Fig. 3 Model example of SAM-Net

### 2.2.3 事件图

与基于事件对和事件链的模型相比,基于图的模型可以表达事件之间更密集、更广泛的联系,其中包含更丰富的脚本知识。

过去的研究大多是从语料库中自动地学习脚本知识,由于一些脚本事件属于常识性知识,并未在文本中详细阐述,这在一定程度上限制了脚本学习的发展。鉴于此,Regneri 等人<sup>[32]</sup>提出从众包数据中构建特定场景的时间脚本图,在构建图的过程中,关注哪些短语可以用来描述脚本中的同一事件,以及这些事件发生的时间顺序有何约束条件。在这之后利用事件序列描述 (event sequence descriptions, ESD) 来区分发生在故事脚本中不同时间点上的事件,然后使用多序列比对 (multiple sequence alignment, MSA) 算法来识别相应的事件描述,提取特定场景的释义和时间顺序信息,从而计算出该场景的时间脚本图。这个脚本图说明了哪些短语可以描述场景中的同一事件,以及这些事件可以以什么顺序发生。

现实生活中的两个事件通常存在着各种各样的关系,比如时序关系等,但这种关系往往不能从文本中推断出来。进一步地,如果脚本事件以图结构表示,就可以使用各种图算法来解决与脚本相关的任务,并利用事件之间的关系。为了更好地处理面向事件的文本和信息需求,将文档级事件表示为由句子级事件构建的结构,Glavaš 等人<sup>[33]</sup>提出用事件图—EvGraph 构造文本信息的新方法,其中事件图的节点表示单个事件提及,而边表示事件提及之间的语义关系 (例如时间关系)。不同的是,以往研究主要关注于单个事件提取任务,而文献<sup>[33]</sup>则是描述了一个完整的端到端系统,用于从文本中提取事件图,通过引入事件图的方法,弥补了目前句子级事件和文档级事件之间存在的差距。

事件的演变和发展存在基本规则,发现事件之间的演变模式对事件预测、策略决断和情景对话系统的发展有极大价值。随着技术的发展,越来越多的研究者开始关注一种基于事件的知识图谱,即事理图谱。相较于传统的知识图谱,事理图谱更加关注谓词性事件及其关系。知识图谱与事理图谱的不同点以及相同点如表 3 所示。

表 3 知识图谱与事理图谱对比

Tab. 3 Comparison between knowledge graphs and event logic graph

对比	不同点	相同点
知识图谱	名词性实体及其关系	有向图
事理图谱 <sup>[34]</sup>	谓词性事件及其关系	

基于此,Li 等人<sup>[34]</sup>提出了一种新的方法来学习事件之间更稠密的连接信息,他们首先从语料库中抽取叙事事件链来构建叙事事理图谱 (narrative event evolutionary graph, NEEG),之后基于图学习事件表示,图中的节点代表事件,边表示事件之间的关系,最后再利用得到的表示向量预测后续事件。由于事理图谱规模较大,借鉴分治的思想提出了缩放图神经网络



(scaled graph neural network, SGNN), 将当前所需子图作为样本以解决 NEEG 上的推理问题, 即每次从事理图谱中抽取一个子图来建模事件交互并学习事件表示, 如图 4 所示。该研究进一步表明了事件图结构可以显著提高事件预测性能, 并增加了模型的鲁棒性。进一步地, Xiong 等人<sup>[35]</sup>发现使用预训练语言模型与图神经网络相结合的框架进行脚本事件预测会产生更好的效果。

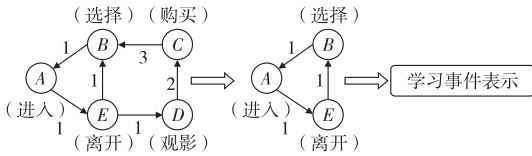


图 4 抽取子图  
Fig. 4 Extracting sub-graph

基于事件图的方法主要是使用图神经网络对图结构信息进行传递、聚合以及更新, 这种方法在建模上显示出了较明显的优势。相比基于事件对的方法来说, 性能有了超越 21.93% 的提升; 相比基于事件图的方法而言, 性能有了超越 1.62% 的提升。

### 2.2.4 结合型建模

#### 1) 结合事件对和事件链

顺序灵活的事件链可能存在过拟合问题, 而事件对可以将链中的时间顺序作为事件对建模的特征, 缓解 LSTM 过拟合问题, 同时 LSTM 具有编码无界序列且不丢失历史事件的优势。所以, Wang 等人<sup>[36]</sup>提出了 PairLSTM, 整合了基于链的时间顺序学习和基于事件对的一致性学习的优势。另外, 事件的多论元结构还存在稀疏性问题, 因此模型使用隐藏层来学习事件嵌入, 在计算事件对关系中使用 LSTM 隐藏状态作为现有事件的特征表示。文献<sup>[36]</sup>还考虑到不同的事件对候选事件的重要性不同, 使用了一个动态记忆网络(dynamic memory network)自动地为每一个事件计算事件权值, 以推断出最佳的候选事件。

#### 2) 结合事件链和事理图谱

单独使用基于事件链或者是事件图的方法来进行事件预测时, 面临着以下问题, 首先是无法充分考虑论元之间的相关性, 其次是不能同时利用事件链和事件图中的信息。为了解决以上问题, 克服事件表示不全面、信息融合不充分的问题, 孙盼等人<sup>[37]</sup>提出了一种结合事件链和事理图谱的脚本事件预测方法 (ECGNet)。具体来说, 为得到更准确的事件表示, 先将每个事件的各个元素组成一个短句, 使用 Transformer 获取元素之间的序列信息; 然后, 构建一个长程时序模块 (LRTO) 学习事

件链中的时序信息; 同时, 构建一个全局事件演化模块 (GEEP), 先根据训练集构建叙事事件链, 再通过可缩放的图神经网络 (scaled graph neural network, SGNN) 获取隐藏在事理图谱中的演化模式; 最后, 通过门控注意力机制动态融合时序信息和演化模式进行脚本事件预测。

现有的方法主要关注候选事件和上下文事件之间的语义相似性, 忽略了事件链尾部的事件, 而尾部事件也很容易影响后续事件的发展方向。因此, Huang 等人<sup>[38]</sup>提出了一个新的脚本事件预测模型。模型由 Token 编码层、事件表示层和预测层组成。其中事件预测层由事件得分和尾部事件得分两个部分组成。事件得分模块利用包含一个单一候选事件的整个事件链的信息来进行预测, 而尾部事件得分模块则是通过构建事理图谱, 利用邻接矩阵来计算尾部事件得分。

#### 3) 多模型结合

将不同建模方法结合起来可以充分利用它们的优势, 有效解决事件表示不充分等问题, 有助于后续事件的推理工作。文献<sup>[31]</sup>提出的 Mcer, 通过集成论元级别、事件级别和链级别的表示, 增强事件的表示学习。此外, 研究表明采用 Mcer 与其 LSTM 变体 (即 Mcer-LSTM) 相结合会产生更好的效果, 结合后的模型在 NYT 数据集上与单独使用 Mcer 相比, 准确率提升了 4.12%。预训练语言模型具有更强的提取特征的能力, 很多自然语言处理任务使用预训练 + 微调的机制产生了更好的效果, 脚本事件预测也不例外。文献<sup>[21]</sup>将预先训练过的 RoBERTa 模型转移到事件序列的模型中, 相较于 SGNN 模型, 在 MCNC 任务上的准确率上升 2.63%。Du 等人<sup>[39]</sup>考虑使用 BERT 模型自动构建事件图, 在 BERT 中加入了一个额外的结构化变量学习预测训练过程中的事件连接, 称之为 GraphBERT, 其可以预测不可见事件的连接强度, 从而避免事件图的稀疏性, 提高了事件预测精度。与 RoBERTa 相比, 准确率提升了 2.06%。

### 2.2.5 小结

如表 4 所示, 将脚本事件预测方法的发展历程划分为基于统计学习和深度学习的方法。早期的脚本建模方法主要是基于统计学习方法, 即通过统计训练集中的事件发生的概率来进行后续预测, 其中典型的方法为 PMI (pointwise mutual information) 和 Bigram 方法。但是基于统计学习的方法忽略了叙事事件链中各个事件之间的时序信息, 仅仅考虑各个事件和候选事件的相关性, 这在很大程度上限制了模型的预测能力。

表 4 脚本建模方法总结  
Tab. 4 Summary of script modeling methods

脚本建模	分类	模型	建模思路
基于统计学习		PMI <sup>[1]</sup>	PMI 统计训练集任意两个事件同时发生的频率作为这两个事件同时发生的概率, 以进行后续事件的预测
		Bigram <sup>[23]</sup>	采用二元条件概率来表示两个事件的关联强度, 并按照观察事件链的顺序建模事件链
	基于事件对	Eventcomp <sup>[15]</sup>	将候选事件与整个脚本的关系拆解为与脚本中每个事件之间的连贯性分数的聚合
	基于事件链	LSTM <sup>[25]</sup>	将脚本学习任务定义为一个序列建模任务, 使用训练模型来预测下一个输入
		CH-LSTM <sup>[16]</sup>	将事件链的原始子事件作为输入, 未来可能发生的子事件的描述作为输出
基于深度学习		SAM-Net <sup>[29]</sup>	采用动态记忆网络来提取事件链的特征, 将事件链表示为一组事件片段
	基于事件图	时间脚本图 <sup>[32]</sup>	从众包数据中构建了特定场景的时间脚本图
		EvGraph <sup>[33]</sup>	一个完整的端到端系统, 用于从文本中提取事件图
		NEEG <sup>[34]</sup>	构建叙事事理图谱后基于图学习事件表示, 最后再利用得到的表示向量预测后续事件
	结合型建模	PairLSTM <sup>[36]</sup>	整合了基于链的时间顺序学习和基于成对的一致性学习的优势
		ECGNet <sup>[37]</sup>	捕获事件链中的时序信息和事理图谱中的事件演化模式进行事件预测
		Mcer <sup>[31]</sup>	集成事件在多个粒度级别上的连接来增强事件的表示学习
	GraphBERT <sup>[39]</sup>	使用 BERT 模型自动构建事件图	

随着深度学习进入研究者的视野, 上述的统计学习方法逐渐被基于深度学习的方法所替代。基于深度学习的方法可以自动学习出模式特征, 并将特征学习融入到建立模型的过程中, 事件预测能力取得了极大的进步。因此, 按照发展历程将后续脚本建模方法从基于事件对、事件链、事件图和结合型建模的角度进行阐述。基于事件对的方法核心为计算事件对

之间的关联强度, 但忽略了各个事件之间的时序信息; 基于事件链的方法则是对基于事件对方法的改进, 关注叙事事件链中各个事件的时序信息, 但事件间的连接信息未得到充分利用; 基于事件图的方法出发点正在于此, 利用图的方法对事件之间的连接信息进行建模, 但进行事件预测时同样存在着信息融合不充分的问题; 结合型建模的方法在一定程度上克服了上述问题,

取长补短,将多种建模方法结合起来,有效提高了事件预测效率。

### 3 实验分析

为了分析各种模型在脚本事件预测任务上的表现,本章对经典的方法进行了分析。实验主要关注如下方面:采用对比实验的手段,对基于统计学习和深度学习的方法分别进行了分析;通过在相同数据集下各种方法的性能,证明深度学习方法对事件预测的重要意义;通过对比深度学习方法下的不同建模方法,讨论各种模型的准确度的区别,证明多建模结合的方法显示出更好的效果。

#### 3.1 数据集

数据集中的事件链来自 Gigaword 语料库的纽约时报 (NYT) 部分,其中每个事件元组由 C&C 工具进行词性标注和依赖解析,使用 OpenNLP 进行短语结构解析和共指消解。按照 140331 : 10000 : 10000 的比例划分训练集、验证集和测试集。每个上下文事件链有 5 个候选事件,其中只有一个是正确的。

#### 3.2 评估方法及评价指标

为了对脚本事件预测任务进行评测,对不同模型进行比较,研究者们提出了不同的评测方法。Chambers 等人<sup>[2]</sup>首先提出用叙事完形填空 (narrative cloze, NC) 评测脚本事件预测任务。叙事完形填空给定文档中的一系列叙事事件 (其中 1 个事件已被删除) 和 5 个随机排序的候选事件 (其中 1 个为该事件上下文的正确的后续事件,其余 4 个事件是从数据集中随机抽样出的处于别的事件上下文中的事件)。这些事件中,使用随机抽样出的事件主体替换为当前事件上下文的主体。具体来说,以 McCann 为主角的文章为例:a) McCann threw two interceptions early (McCann 提前进行了两次拦截); b) Toledo pulled McCann aside and told him he'd start (Toledo 把 McCann 拉到一边,告诉他自己要开始了); c) McCann quickly completed his first two passes (McCann 很快就完成了他的前两次传球)。

以上叙事模型由五个事件表示:(threw, subject)(pulled, object)(told, object)(start, subject)(completed, subject)。这些动词/依赖事件构成了一个叙事总结模型,本文可以删除 (threw, subject), 并使用剩下的四个事件来对这个丢失的事件进行排序。通过删除一对这样的配对,以此来对模型进行评估。但是 NC 任务评测仍存在不足之处,首先对于任意给定事件,后续事件都有多个可信的选择;其次它需要搜索整个事件词汇表,非常大的词汇量会导致计算问题。

叙事完形填空的另外一个问题是有时会产生多个可信答案,这就需要手动评估系统输出的答案,导致成本过高。为了解决这一问题,Modi<sup>[40]</sup>提出了对抗性叙事完形填空 adversarial narrative cloze (ANC) 任务。ANC 任务显示两个事件序列,一个是正确的事件序列,另一个是除了有一个事件被一个随机事件取代外,其他事件均与正确序列相同的序列。其任务是猜测这两个事件中的哪一个是正确的事件序列。

与之前评估方法有所不同,为了评估模型对文本的理解能力,Mostafazadeh 等人<sup>[41]</sup>提出了故事填空任务 (story cloze test, SCT) 来评估模型的效果,即预测一个未完成的故事的结局。SCT 并不是预测一个事件,而是选择一个完整的句子来完成给定的故事。因此 NC 任务可以看做是故事结束任务中的一个子任务,并且事件链之外的信息对于 SCT 任务也很有用。

另外,2016 年,Granroth-Wilding 等人<sup>[15]</sup>还提出了一种被称为多选完形填空 (multiple choice narrative cloze, MCNC) 的评估标准来评价脚本事件预测模型的有效性 (如图 5 所示)。MCNC 根据特定场景下已经发生的一系列事件,从给定的候选事件集中选择接下来最有可能发生的后续事件,候选事件集中有且仅有一个事件是标准答案,并且候选事件集中的事件共享

主语。MCNC 对脚本事件预测任务进行了简化,大大缩小了候选事件集的范围,把候选事件集从所有的事件缩小至有限个事件。在 MCNC 中,系统能够利用事件上下文中更丰富的信息。

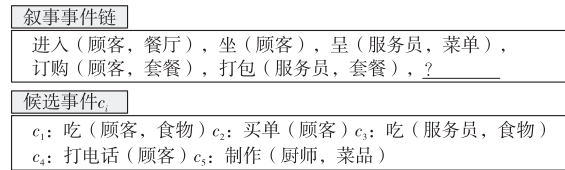


图 5 多选完形填空示例

Fig. 5 Example of MCNC

为了评估学习模型的质量, Lee 等人<sup>[42]</sup>基于 Granroth-Wilding 等人<sup>[15]</sup>的工作,提出了额外的两种评测指标:多选叙事序列 (multiple-choice narrative sequence, MCNS) 和多选叙事解释 (multiple-choice narrative explanation, MCNE)。与 MCNC 不同的是,MCNS 任务是除了第一个事件外,为每个事件生成多个选项,再将每个事件链建模为马尔可夫链,最后用推理算法识别得分最高的事件链。MCNE 任务则是同时提供了开始事件和结束事件,而预测任务则是推断两者之间发生了什么。总的来说,MCNS 和 MCNE 旨在评估模型推断更长事件序列的能力,从而更好地解释叙事结构。

#### 3.3 实验结果

选取 Gigaword 语料库中 NYT 的一部分作为数据集,MCNC 作为评估方法,准确度 (accuracy) 作为评估指标。为了对不同方法的性能差异进行说明,针对每一类方法选取了部分基线模型,其性能对比如表 5 所示。

表 5 常见基准方法对比

Tab. 5 Comparison between Common benchmarking methods /%

方法	Acc	方法	Acc
PMI <sup>[1]</sup>	30.52	NEEG <sup>[34]</sup>	52.37
Bigram <sup>[23]</sup>	29.67	Mcer <sup>[31]</sup>	56.64
Eventcomp <sup>[15]</sup>	49.57	ECGNet <sup>[37]</sup>	56.11
LSTM <sup>[25]</sup>	50.86	NEEG <sup>[34]</sup> + Eventcomp <sup>[15]</sup> + PairLSTM <sup>[36]</sup>	54.93
CH-LSTM <sup>[16]</sup>	50.91	Mcer + Mcer-LSTM <sup>[31]</sup>	60.76
SAM-Net <sup>[29]</sup>	52.68	GraphBERT <sup>[39]</sup>	60.72
PairLSTM <sup>[36]</sup>	50.91		

#### 3.4 结果分析

从表 5 的实验结果可以看到,相比基于统计学习的方法 PMI 和 Bigram,基于深度学习方法的准确度有了超过 19.05% 的提升,极大提高了预测事件的效率。原因在于,基于统计学习的方法鲁棒性差且,模型预测能力不足,而深度学习学习方法具备很强的学习能力。

通过对比基于深度学习的各种方法发现,从基于事件对的方法 Eventcomp 到基于事件链的方法 LSTM 到基于事件图的方法 NEEG 再到结合型建模的方法 ECGNet,预测性能正在逐步提升。与基于统计学习的方法相比,基于事件对的方法缓解了事件稀疏性,但它忽略了事件之间的时序关系,因此表现不如基于事件链和事件图的方法。基于事件链的方法和基于事件图的方法由于捕获了事件之间的序列特征均取得不错的预测效果,但由于事件之间的演化模式对于脚本事件预测而言也很重要,而基于事件图的方法正可以以构建图谱的方式获取事件的发展规律,所以基于事件图的方法预测效果优于基于事件链的方法。从表 5 可以看出,结合型建模的方法性能整体表现最佳,Mcer + Mcer-LSTM 取得了最好的实验效果。ECGNet 相比单纯使用基于事件链或是事件图的模型而言,准确度至少提升了 3.74%,NEEG + EventComp + PairLSTM 相比这三个模型准确度提升超过 2.56%,Mcer + Mcer-LSTM 相比 Mcer 将准确度从 56.64% 提升到了 60.76%,GraphBERT 采用预训练语言模型 BERT 自动构建事件图,同样表现出很好的预测效果。由于不同模型之间具有互补效应,存在各自的优势,所以可以相

互受益,从而表现出最佳的性能。

## 4 结束语

### 4.1 脚本事件预测总结

本文对脚本事件预测的已有研究进行了总结,主要是从事件表征、脚本建模和模型分析三个角度进行叙述。

在事件表征部分,事件的要素构成从单个论元演化至围绕多个论元主角,事件的表示方式也从最初的(predicate; dependency)发展至更丰富的事件表示方法。事件表示学习的主要任务是组合事件元素向量,嵌入的向量保留了事件中的语义信息,但存在建模不足的问题,因此进一步考虑如情感信息等外部知识,丰富事件的语义信息。

在脚本建模方面,主要分为基于统计学习的方法和基于深度学习的方法。以 PMI 和 Bigram 为代表的基于统计学习的方法,模型根据训练集中事件出现的次数确定相关性,缺乏泛化性能。通过使用适当的神经网络架构以及足够大的数据,深度学习网络可以学习从一个向量空间到另一个向量空间的任何映射,因而深度学习的应用使得脚本事件预测性能获得明显提升。在基于深度学习的各项方法中,结合型建模表现最佳,将不同的模型进行组合,模型之间互相补充,产生了更优的预测性能。

### 4.2 脚本事件预测挑战

尽管近年来关于脚本事件预测的研究不断涌现,但相关研究仍面临一些挑战<sup>[43-46]</sup>。

a) 标准的评价体系。脚本事件预测领域目前没有标准评价体系,这使得不同模型之间的比较更加困难。如果没有标准的评价体系,就不能准确地判断一个模型是否真的有效。因此,建立一个标准的评价体系对于脚本事件预测的发展尤为重要。

b) 结合事理图谱。用图知识来捕获事件间丰富的连接关系的研究相对较少,由于现实中很多元素的连接关系很复杂,而图包含节点和连接,它可以更充分地表示事件,例如两个事件之间的因果关系等,这样就可以获取更多信息用来推理后续事件。在实际工作中,由于事件包含的信息较多,所以事理图谱的体量也会很大,这就给训练带来一定的难度。为了提高运行效率,在后续的研究中可以继续沿用子图的处理方法,探索出更高效的算法。

c) 进行更深层次的事件表示研究。在根据上下文事件预测后续的事件的情况中,事件对后续事件的影响程度不同,所以可以考虑给不同部分的事件赋予不同的权重以辅助事件预测。例如如果尾部事件与前续事件发生了明显的转折,在这种情况下,显然尾部事件对预测结果的影响更大,因此就可以给尾部事件赋予更高的权重。针对事件表示研究来说,目前的脚本事件预测工作仍然需要提取与事件相关的事件参与者,如果事件的参与者是未知的话,这将会使预测的效率降低,所以在未来如何才能更高效地使用事件脚本仍是值得探索的。另外,为了充分利用事件文本来获得更多有效信息辅助事件推理工作探索,可以通过挖掘事件之间更复杂的交互的方法,并尝试引入如知识图谱等外部背景信息。

d) 其他挑战。脚本事件预测在未来很有研究前景,但是现在脚本事件预测的下游应用较少,所以随着脚本事件预测的发展,在未来的工作中,可以逐步将脚本事件预测方法用到更多的下游应用中去。现在脚本事件预测主要面向的是英文文本,中文的研究较少,加上中文文本的语法较为复杂,所以如何从中文文本中抽取相关信息来进行预测也是值得研究的。

### 4.3 脚本事件预测展望

通过对当前脚本事件预测研究进展的梳理,可以展望未来脚本事件预测的发展方向<sup>[47-50]</sup>:

a) 在数据方面。由于大多数现有的脚本事件预测方法的处理范式都是建模成为多项选择填空的形式,而这种数据集很少,所以构建新的数据集的成本很高。那么如何花费最少的成本达到最大的训练效果是未来值得关注的研究方向。

b) 针对脚本事件预测的评测,提出更有效的评测方法。虽然 MCNC 目前被广泛用作任务评测,但完型填空的方法在实际应用当中不便于验证结果的正确性。因此,随着神经网络的发展,研究出更高效的评测方法也是未来研究的发展趋势。

c) 针对事件表示来说。近年来,图神经网络在各个研究领域大放异彩,但是如何才能更好地利用事件表示构建图?已有研究者尝试用对比学习<sup>[51,52]</sup>的方法,考虑多个正例和负例,以此来更好地利用事件的共现信息来学习事件表示。所以,如何更好地进行事件表示学习,采用何种形式学习到更多相关知识,这也是今后的一个研究方向。

d) 基于其落地使用来看。目前脚本事件预测在工业界的落地场景中是很少见的。以金融场景的推理为例,如何才能确定一个推理模型是否正确,找到这样的训练数据本身都是很难的。因此,在未来的研究中,将脚本事件预测以何种方式应用到更多的现实落地场景中也是值得探索的。

## 5 结束语

脚本事件预测是通过给定已发生的事件上下文,预测后续最有可能发生的事件,它对于事件风险预测、事件演化挖掘等现实落地场景,能够起到一定的参考意义。本文详细阐述了对目前脚本事件预测的研究进展。首先介绍了脚本事件预测的概念;然后以事件表征、脚本建模和任务评测为脉络梳理了现有研究;最后分析了目前脚本事件预测面临的挑战,对其前景进行展望。希望综述内容能够为脚本事件预测领域的研究提供理论指导和创新思路,在未来能够应用到更多的落地场景中。

### 参考文献:

- [1] Schank R C, Abelson R P. Scripts, plans, and knowledge [C]// Proc of the 4th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers, 1975: 151-157.
- [2] Chambers N, Jurafsky D. Unsupervised learning of narrative event chains [C]// Proc of the 46th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL Press, 2008: 789-797.
- [3] Balasubramanian N, Soderland S, Etzioni O. Generating coherent event schemas at scale [R]. Stroudsburg, PA: ACL Press, 2013.
- [4] Pichotta K, Mooney R. Statistical script learning with multi-argument events [C]// Proc of the 14th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg, PA: ACL Press, 2014: 220-229.
- [5] Ahrendt S, Demberg V. Improving event prediction by representing script participants [C]// Proc of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL Press, 2016: 546-551.
- [6] Pichotta K, Mooney R. Learning statistical scripts with LSTM recurrent neural networks [C]// Proc of AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2016.
- [7] Lee I T, Goldwasser D. Feel: featured event embedding learning [C]// Proc of AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018: 4840-4847.
- [8] Bai Long, Guan Saiping, Guo Jiafeng, et al. Integrating deep event-level and script-level information for script event prediction [C]// Proc of Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL Press, 2021: 9869-9878.
- [9] Mikolov T, Sutskever I, Chen Kai, et al. Distributed representations of words and phrases and their compositionality [J]. *Advances in Neural Information Processing Systems*, 2013, 26(1): 3111-3119.
- [10] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space [EB/OL]. (2013). <https://arxiv.org/abs/1301.3781>.
- [11] Weber N, Balasubramanian N, Chambers N. Event representations with



- tensor-based compositions [C]// Proc of AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018; 4946-4953.
- [12] Socher R, Huval B, Manning C D, *et al.* Semantic compositionality through recursive matrix-vector spaces [C]// Proc of Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Stroudsburg, PA: ACL Press, 2012; 1201-1211.
- [13] Modi A, Titov I. Inducing neural models of script knowledge [C]// Proc of the 18th Conference on Computational Natural Language Learning. Stroudsburg, PA: ACL Press, 2014; 49-57.
- [14] Modi A, Titov I. Learning semantic script knowledge with event embeddings [EB/OL]. (2013). <https://arxiv.org/abs/1312.5198>.
- [15] Granroth-Wilding M, Clark S. What happens next? Event prediction using a compositional neural network model [C]// Proc of AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2016; 2727-2733.
- [16] Hu Linmei, Li Juanzi, Nie Liqiang, *et al.* What happens next? Future subevent prediction using contextual hierarchical LSTM [C]// Proc of the 31st AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2017; 3450-3456.
- [17] Ding Xiao, Liao Kuo, Liu Ting, *et al.* Event representation learning enhanced with external commonsense knowledge [EB/OL]. (2019). <https://arxiv.org/abs/1909.05190>.
- [18] Ding Xiao, Zhang Yue, Liu Ting, *et al.* Deep learning for event-driven stock prediction [C]// Proc of the 24th International Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2015; 2327-2333.
- [19] Zhao Sendong, Wang Quan, Massung S, *et al.* Constructing and embedding abstract event causality networks from text snippets [C]// Proc of the 10th ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2017; 335-344.
- [20] Wang Ruize, Tang Duyu, Duan Nan, *et al.* K-adapter: infusing knowledge into pre-trained models with adapters [EB/OL]. (2020). <https://arxiv.org/abs/2002.01808>.
- [21] Lyu Shangwen, Zhu Fuqing, Hu Songlin. Integrating external event knowledge for script learning [C]// Proc of the 28th International Conference on Computational Linguistics. 2020; 306-315.
- [22] Du Li, Ding Xiao, Liu Ting, *et al.* Learning event graph knowledge for abductive reasoning [C]// Proc of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. 2021; 5181-5190.
- [23] Jans B, Bethard S, Vulic I, *et al.* Skip n-grams and ranking functions for predicting script events [C]// Proc of the 13th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg, PA: ACL Press, 2012; 336-344.
- [24] Hochreiter S, Schmidhuber J. Long short-term memory [J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [25] Pichotta K, Mooney R. Learning statistical scripts with LSTM recurrent neural networks [C]// Proc of AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2016; 2800-2806.
- [26] Pichotta K, Mooney R J. Using sentence-level LSTM language models for script inference [C]// Proc of the 54th Annual Meeting of the Association for Computational Linguistics. 2016; 279-289.
- [27] Kiros R, Zhu Yukun, Salakhutdinov R R, *et al.* Skip-thought vectors [C]// Proc of the 28th International Conference on Neural Information Processing Systems. 2015; 3294-3302.
- [28] Lin Zhouhan, Feng Minwei, Santos C N, *et al.* A structured self-attentive sentence embedding [EB/OL]. (2017). <https://arxiv.org/abs/1703.03130>.
- [29] Lyu Shangwen, Qian Wanhui, Huang Longtao, *et al.* SAM-Net: integrating event-level and chain-level attentions to predict what happens next [C]// Proc of the 33rd AAAI Conference on Artificial Intelligence and the 31st Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019; 6802-6809.
- [30] Huang Gao, Liu Zhuang, Van Der Maaten L, *et al.* Densely connected convolutional networks [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2017; 2261-2269.
- [31] Wang Lihong, Yue Juwei, Guo Shu, *et al.* Multi-level connection enhanced representation learning for script event prediction [C]// Proc of Web Conference 2021. New York: ACM Press, 2021; 3524-3533.
- [32] Regneri M, Koller A, Pinkal M. Learning script knowledge with web experiments [C]// Proc of the 48th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL Press, 2010; 979-988.
- [33] Glavaš G, Šnajder J. Construction and evaluation of event graphs [J]. *Natural Language Engineering*, 2015, 21(4): 607-652.
- [34] Li Zhongyang, Ding Xiao, Liu Ting. Constructing narrative event evolutionary graph for script event prediction [C]// Proc of the 27th International Joint Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018; 4201-4207.
- [35] Xiong Kai, Ding Xiao, Du Li, *et al.* Heterogeneous graph knowledge enhanced stock market prediction [J]. *AI Open*, 2021, 2012(2): 168-174.
- [36] Wang Zhongqing, Zhang Yue, Chang Chingyun. Integrating order information and event relation for script event prediction [C]// Proc of Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL Press, 2017; 57-67.
- [37] 孙盼, 王琪, 万怀宇. 结合事件链与事理图谱的脚本事件预测模型 [J]. *计算机工程*, 2022, 48(4): 119-125. (Sun Pan, Wang Qi, Wan Huaiyu. Event chains and graphs combined neural network for script event prediction [J]. *Computer Engineering*, 2022, 48(4): 119-125.)
- [38] Huang Zhenyu, Wang Yongjun Xu Hongzuo, *et al.* Script event prediction based on pre-trained model with tail event enhancement [C]// Proc of the 5th International Conference on Computer Science and Artificial Intelligence. New York: ACM Press, 2021; 242-248.
- [39] Du Li, Ding Xiao, Zhang Yue, *et al.* A graph enhanced BERT model for event prediction [M] Findings of the Association for Computational Linguistics. Stroudsburg, PA: ACL Press, 2022; 2628-2638.
- [40] Modi A. Event embeddings for semantic script modeling [C]// Proc of the 20th SIGNLL Conference on Computational Natural Language Learning. Stroudsburg, PA: ACL Press, 2016; 75-83.
- [41] Mostafazadeh N, Chambers N, He X, *et al.* A corpus and cloze evaluation for deeper understanding of commonsense stories [C]// Proc of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL Press, 2016; 839-849.
- [42] Lee I T, Goldwasser D. Multi-relational script learning for discourse relations [C]// Proc of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL Press, 2019; 4214-4226.
- [43] Taylor W L. Cloze procedure: a new tool for measuring readability [J]. *Journalism quarterly*, 1953, 30(4): 415-433.
- [44] Han Yi, Qiao Linbo, Zheng Jianming, *et al.* A survey of script learning [J]. *Frontiers of Information Technology & Electronic Engineering*, 2021, 22(3): 341-373.
- [45] 胡志磊, 靳小龙, 陈剑赞, 等. 事件图谱的构建、推理与应用 [J]. *大数据*, 2021, 7(3): 80-96. (Hu Zhilei, Jin Xiaolong, Chen Jianyun, *et al.* Construction, reasoning and applications of event graphs [J]. *Big Data*, 2021, 7(3): 80-96.)
- [46] 吴雨钊. 脚本事件预测综述 [J]. *现代计算机*, 2021(15): 113-116. (Wu Yuzhao. Overview of script event prediction [J]. *Modern Computer*, 2021(15): 113-116.)
- [47] 李忠阳. 面向文本事件预测的事理图谱构建及应用方法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2021. (Li Zhongyang. Research on the construction and application method of affection atlas for text event prediction [D]. Harbin: Harbin Institute of Technology, 2021.)
- [48] 孙盼. 基于事理图谱的脚本事件预测方法研究 [D]. 北京: 北京交通大学, 2021. (Sun Pan. Research on scripted event prediction method based on affair map [D]. Beijing: Beijing Jiaotong University, 2021.)
- [49] Rudinger R, Rastogi P, Ferraro F, *et al.* Script induction as language modeling [C]// Proc of Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL Press, 2015; 1681-1686.
- [50] Ding Xiao, Li Zhongyang, Liu Ting, *et al.* ELG: an event logic graph [EB/OL]. (2019). <https://arxiv.org/abs/1907.08015>.
- [51] Chen Ting, Kornblith S, Norouzi M, *et al.* A simple framework for contrastive learning of visual representations [C]// Proc of the 37th International Conference on Machine Learning. Ithaca: Cornell University Press, 2020; 1597-1607.
- [52] Zheng Jianming, Cai Fei, Liu Jun, *et al.* Multistructure contrastive learning for pretraining event representation [J]. *IEEE Trans on Neural Networks and Learning Systems*, 2022; 1-13.