

结合事件链与事理图谱的脚本事件预测模型

孙盼, 王琪, 万怀宇

(北京交通大学 计算机与信息技术学院 交通数据分析与挖掘北京市重点实验室, 北京 100044)

摘要: 现有脚本事件预测模型在事件表示时未充分考虑各个元素之间的相关性,且不能同时利用事件链和事理图谱中的信息进行事件预测。针对事件表示不全面和信息融合不充分的问题,提出一种结合事件链和事理图谱的脚本事件预测模型ECGNet。将每个事件的各个元素构造成一个短句,使用Transformer编码器捕获元素之间的序列信息,从而获得更准确的事件表示。在此基础上,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)捕获隐藏在事理图谱中的演化模式,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。基于纽约时报和新浪新闻两个数据集的实验结果表明,ECGNet能够有效融合事件链和事理图谱的信息进行脚本事件预测,与PMI、Bigram、SAM-Net、SGNN等模型相比,其准确率较最优值取得了3%以上的提升。
关键词: 脚本事件预测;事件表示;事件链;事理图谱;注意力机制

开放科学(资源服务)标志码(OSID):



中文引用格式: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.

英文引用格式:SUN P, WANG Q, WAN H Y. Script event prediction model combining event chains and event evolutionary graphs[J]. Computer Engineering, 2022, 48(4): 119-125.

Script Event Prediction Model Combining Event Chains and Event Evolutionary Graphs

SUN Pan, WANG Qi, WAN Huaiyu

(Beijing Key Laboratory of Traffic Data Analysis and Mining, School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

[Abstract] Given a sequence of events that have occurred, the goal of script event prediction is to infer what happens next. Previous approaches ignore the semantic correlations between the words describing the event and are thus unable to simultaneously consider the information in the event chains and event evolutionary graphs for event prediction. To solve the problem of incomplete event representation and insufficient fusion, this study proposes a script event prediction model combining event chains and event evolutionary graphs, namely ECGNet. First, the words describing an event are treated as a short sentence and the transformer encoder is utilized to capture the order between the words to obtain an accurate event representation. Specifically, a Long Range Temporal Orders (LRTO) module is designed to learn the strong order information in contextual event chains, and a Global Event Evolutionary Patterns (GEEP) module is constructed to capture the global patterns in event evolutionary graphs. Finally, a gated multi-attention mechanism is proposed to dynamically integrate the orders and evolutionary patterns to predict what happens next. Experimental results based on two datasets of New York Times and Sina News show that ECGNet can effectively integrate the information of event chains and event maps for script event prediction. Compared with PMI, Bigram, Semantic Attribute Matching Networks (SAM-Net), Self-Generating Neural Network (SGNN), and other models, the event prediction accuracy is improved by more than 3% compared with the optimal value.

[Key words] script event prediction; event representation; event chain; event evolutionary graph; attention mechanism
DOI: 10.19678/j.issn.1000-3428.0060781

0 概述

脚本事件预测任务对于许多自然语言理解应用至

关重要,例如文本理解^[1]、意图识别^[2]和问答系统^[3]。脚本事件预测任务定义为:给定特定场景中已经发生的一系列上下文事件,预测后续可能发生的事件^[4]。

基金项目:国家重点研发计划(2018YFC0830200)。

作者简介:孙盼(1996—),女,硕士研究生,主研方向为数据挖掘、信息抽取;王琪,硕士研究生;万怀宇(通信作者),副教授、博士。

收稿日期:2021-02-02 修回日期:2021-03-30 E-mail: sunpan@bjtu.edu.cn

早期关于脚本事件预测的研究都是基于统计的方法,例如 PMI(Pointwise Mutual Information)^[4]和 Bigram^[5]方法。但是,基于统计的方法无法对训练集中没有出现过的事件进行预测,缺乏泛化性能,极大地限制了模型的预测能力。近年来,通过神经网络把事件映射至低维稠密的向量空间这一做法可以有效上述问题,取得了较好的预测效果^[6]。目前,基于神经网络的脚本事件预测方法大体上可以分为基于事件对、基于事件链和基于事理图谱3类方法。虽然基于神经网络的方法取得了较好的效果,但仍难以有效解决脚本事件预测任务的两大难点,即如何进行准确的事件表示以及如何充分利用事件之间复杂的相关性进行后续预测。

脚本事件预测的一个难点是如何进行事件表示,通常每个事件用四元组(主语、谓语、宾语及介词宾语)的形式表示。现有的方法把表示事件的4个元素的嵌入进行拼接或者平均得到整个事件的表示,即“加性”模型^[7]。“加性”模型对表示事件的词嵌入有很强的依赖,只能学习事件的浅层语义信息,难以区分两个事件之间细微的区别。以“X eat apple”和“X establish apple”为例,两个事件都包含“apple”,如果仅用“加性”模型把表示事件的词向量进行拼接,会导致这两个事件的表示非常接近,然而这两个事件并非相似事件。因此,如何进行事件的“非加性”表示是本研究的一个难点。得到事件的表示之后,现有方法单独利用事件链的时序信息或事理图谱中的事件演化模式进行后续事件预测,没有有效整合这两类信息对后续事件进行预测。因此,如何捕获时序信息和事件演化模式并有效融合两者进行事件预测,是本研究的另一个难点。

为应对上述挑战,本文提出一种结合事件链和事理图谱的脚本事件预测方法 ECGNet。将表示事件的4个元素视为一个长度为4的短句输入到 Transformer 编码器^[8],以学习事件的深层语义信息得到更准确的事件表示,并利用基于长短期记忆(Long Short-Term Memory, LSTM)网络^[9]的长程时序(Long Range Temporal Orders module, LRTO)模块对叙事事件链进行学习,捕获事件链内部的时序信息。同时,设计基于图神经网络(Graph Neural Network, GNN)^[10]的全局事件演化(Global Event Evolutionary Patterns, GEEP)模块对事理图谱进行学习,捕获事件发展的演化模式。此外,本文还设计一种门控注意力机制来动态融合两个模块学习到的信息,从而实现最终的预测。

1 相关工作

脚本是指被存放在知识库中用来进行推理的事件序列^[11]。2008年,CHAMBERS等^[4]提出一种可以从新闻文本中无监督地提取大规模叙事事件链的方

法。2016年,GRANROTH-WILDING等^[6]在前人的研究基础上,提出了多选完形填空(Multiple Choice Narrative Cloze, MCNC)方法来评估脚本事件预测模型的效果,如图1所示。

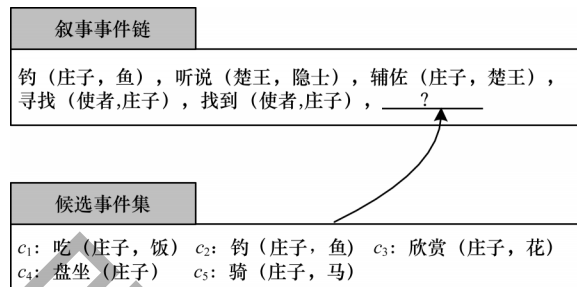


图1 多选完形填空样例

Fig.1 Example of MCNC

文献[4,6]的工作极大地推动了脚本事件预测领域的发展,下面对相关工作进行简要介绍。

1.1 脚本事件表示

脚本事件表示是脚本事件预测任务中一个比较重要的子任务。文献[4]在最初的任务定义中用<predicate; dependency>来表示一个事件,这种表示方式被称作 predicate-GR。其中, predicate表示事件的核心动词, dependency表示事件的主语和核心动词之间的依赖关系,例如“subject”和“object”。predicate-GR方式有时可以包括事件的核心信息,例如<逮捕; obj>,但有时也会丢失重要的信息,难以理解事件的真实含义,例如<去; subj>,此时“去工作”或者是“去度假”不得而知。为了解决这个问题,文献[12]提出采用<Arg1; Relation; Arg2>三元组的形式来表示事件,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。为了更准确地表示事件,文献[13]提出使用(p, a_0, a_1, a_2)四元组的形式来表示事件,其中 p 表示事件的核心动词, a_0 表示事件的主语, a_1 表示事件的宾语, a_2 表示事件的介词宾语。四元组形式包含的信息更全面,因此,本文也采用这一形式进行事件表示。

1.2 脚本事件预测

脚本事件预测任务从二十世纪七十年代提出至今先后经历了人工编码阶段、基于统计的阶段和基于神经网络的阶段。

第一阶段:人工编码阶段。早期的脚本事件预测主要依赖于人工从故事文本中学习事件的规则,然后对学习到的事件进行清洗和处理,构造专家系统。这类人工编码模型的缺点显而易见:1)人工编码具有不全面性,对专家没有标注过的故事文本无法进行后续的预测;2)人工标注无法处理大规模的数据集,无法适应信息爆炸增长的当代社会。

第二阶段:基于统计的阶段。这一阶段的方法借鉴了语言模型的思想,通过学习各个事件之间的概率分布进行脚本事件预测。PMI模型^[4]通过统计

训练集中两个事件同时发生的频率来计算这两个事件同时发生的概率,进而进行后续事件的预测。Bigram模型^[5]则采用二元条件概率来表示两个事件的关联强度。这类基于统计的模型仍然存在表示能力不全面、泛化性能差等问题。

第三阶段:基于神经网络的阶段。这一阶段方法主要分为:基于事件对的方法,例如 Word2Vec^[14]和 EventComp^[6];基于事件链的方法,例如 PairLSTM^[15]和 SAM-Net^[16];基于事理图谱图的方法,例如 SGNN^[17]。该阶段的方法都是将事件映射成低维稠密的向量进行后续预测。具体而言,基于事件对的方法重点关注叙事事件链中的事件和候选集中的事件之间的相关性,利用事件对的相关性进行后续事件预测,这类方法完全忽略了叙事事件链中各个事件之间的时序关系。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,采用循环神经网络(Recurrent Neural Network, RNN)捕获叙事事件链的时序信息,进而对后续事件进行预测,这类方法对类似于图2中展示的网络购票观影这种链式场景下的事件预测非常有效。



图2 链式事件示例

Fig.2 Example of chain events

目前基于事理图谱的方法研究较少,文献[17]提出根据训练集中所有的叙事事件链构建事理图谱,通过学习事理图谱中事件之间的稠密连接关系进行后续事件预测,这类方法更适用于类似于图3所示的金融货币超发这类环状场景下的预测。

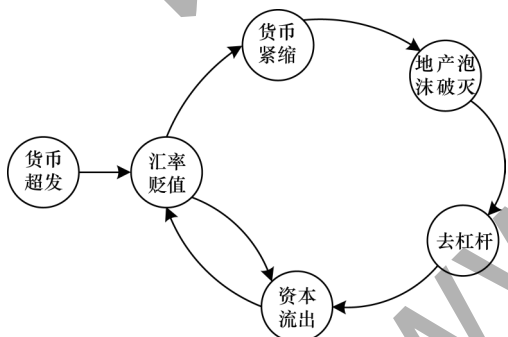


图3 环式事件示例

Fig.3 Example of ring events

2 问题定义

遵循前人的工作,本文采用四元组 (p, a_0, a_1, a_2) 形式来表示脚本事件,其中, p 表示事件的核心动词, a_0 表示事件的主语, a_1 表示事件的宾语, a_2 表示事件的介词宾语, a_0, a_1 和 a_2 称为事件的参数。

脚本事件预测的任务可以描述为:已知特定场景下已经发生的一系列事件 $\{e_1, e_2, \dots, e_n\}$, 从候选事件集 $\{e_{c1}, e_{c2}, \dots, e_{cm}\}$ 中选择接下来最有可能发生的

事件 e_c^* 。其中每个已经发生的事件 $e_i (i \in [1, n])$ 和每个候选事件 $e_{c_j} (j \in [1, m])$ 都是 (p, a_0, a_1, a_2) 的形式。本文把已经发生的事件序列称为上下文叙事事件链,其中每个事件称为上下文事件。

本文采用 MCNC 作为评估方法。MCNC 对本事件预测任务进行简化,大幅缩小候选事件集的范围,把候选事件集从所有的事件集缩小至有限事件集,即除了正确的后续事件外,其余的候选事件从事务集中随机抽取,并用正确后续事件的主语替换其他候选事件的主语,因此,候选事件集中有且仅有一个是正确答案,且所有候选事件共享主语。最后,采用预测准确率评估模型的有效性。

3 ECGNet模型

本文提出的 ECGNet 模型将事件视为短句,使用 Transformer 学习事件内部的语义关联从而得到全面的事件表示。同时,设计基于 LSTM 的长程时序模块来捕获事件链内部的时序信息,构建基于 GNN 的全局演化模块来捕获事理图谱中事件的发展规律和演化模式。最后,提出一种门控注意力机制来动态融合时序信息和演化信息进行后续事件预测。

具体而言,ECGNet 模型共分为 4 个模块(如图 4 所示):

- 1) 事件表示层,把每个脚本事件映射至低维稠密的向量空间。
- 2) 长程时序模块(LRTO),通过 LSTM 对叙事事件链中的上下文事件进行建模,得到融合时序信息的事件表示。
- 3) 全局事件演化模块(GEEP),根据训练集中所有的叙事事件链构建事理图谱,然后利用 GNN 学习融合演化模式的事件表示。
- 4) 门控注意力层,通过门控注意力机制动态融合 LRTO 模块学习到的事件表示和 GEEP 模块学习到的事件表示,进行后续事件预测。

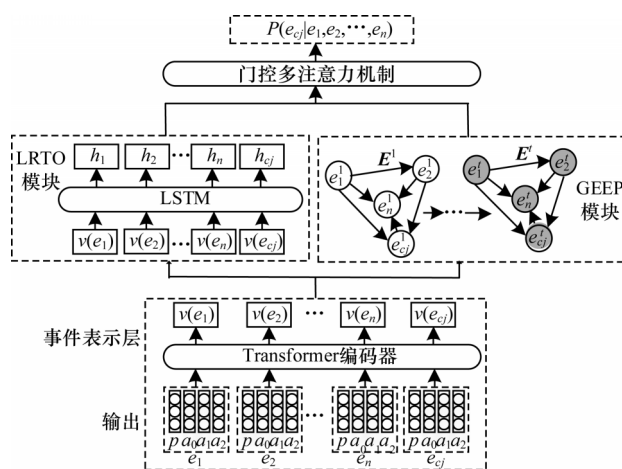


图4 ECGNet模型的整体架构

Fig.4 Overall architecture of ECGNet model

3.1 事件表示层

事件表示层的作用是把每个输入事件编码为一个低维稠密的向量。输入是表示每个事件的4个元素的初始词嵌入(通过 GloVe 方法^[18]预训练得到的 d 维词向量),初始词嵌入在后续模型的训练过程中可以通过学习微调。对于少于3个参数的事件,用“NULL”填充缺失的参数。例如对事件“Tom orders food”,其中, p 为“order”, a_0 为“Tom”, a_1 为“food”, a_2 为“NULL”。简单起见,使用零向量来表示“NULL”和不在词表中的词。

为学习表示事件的4个元素之间丰富的语义关联,本文提出把每个事件视为长度为4的短句,并采用 Transformer 编码器学习每个元素的嵌入。每个事件的四元素的初始词嵌入 $\{v_p, v_{a_0}, v_{a_1}, v_{a_2}\}$ 经过 Transformer 编码之后得到 $\{z_p, z_{a_0}, z_{a_1}, z_{a_2}\}$, 然后通过式(1)把4个元素的词嵌入进行拼接作为整个事件表示:

$$v(e)=[z_p, z_{a_0}, z_{a_1}, z_{a_2}] \quad (1)$$

其中: $z_p, z_{a_0}, z_{a_1}, z_{a_2} \in \mathbb{R}^d$; $[]$ 表示拼接操作。因此,该层对于每个事件 e 的输出是 $v(e) \in \mathbb{R}^{4d}$ 。

3.2 长程时序模块

长程时序模块(LRTO)采用 LSTM 来学习上下文事件链中的时序信息。输入是上下文叙事事件链 $\{v(e_1), v(e_2), \dots, v(e_n) | v(e_i) \in \mathbb{R}^{4d}, i \in [1, n]\}$ 和对应的候选集 $\{v(e_{c_1}), v(e_{c_2}), \dots, v(e_{c_m}) | v(e_{c_j}) \in \mathbb{R}^{4d}, j \in [1, m]\}$ 。按照顺序把上下文事件序列 $\{v(e_1), v(e_2), \dots, v(e_n)\}$ 输入 LSTM 中,根据式(2)得到每个事件 e_i 的隐藏表示 h_i :

$$h_i = \text{LSTM}(v(e_i), h_{i-1}) \quad (2)$$

其中: $h_i \in \mathbb{R}^u$, 其包括从事件 e_1 到事件 e_i 的长程时序语义信息; u 是 LSTM 层的隐藏状态维度。初始隐藏状态 h_0 和所有的 LSTM 参数通过随机初始化产生。通过式(2)可以得到每个上下文事件链的隐藏表示序列 $\{h_1, h_2, \dots, h_n\}$ 。

对于每个事件链所对应的候选事件集,把其中每个候选事件拼接至叙事事件链的末端。具体而言,对于每个候选事件 e_{c_j} , LSTM 的输入为当前候选事件的表示 $v(e_{c_j})$ 和叙事事件链末端事件 e_n 的隐藏表示 h_n , 根据式(3)得到其对应的隐藏表示 h_{c_j} :

$$h_{c_j} = \text{LSTM}(v(e_{c_j}), h_n) \quad (3)$$

其中: $h_{c_j} \in \mathbb{R}^u$, 其包含上下文事件链的信息和当前候选事件的信息。通过式(3)可以得到所有候选事件的表示 $\{h_{c_1}, h_{c_2}, \dots, h_{c_m}\}$ 。

最后,把 n 个上下文事件的隐藏表示和 m 个候选事件的隐藏表示根据式(4)表示为 $H \in \mathbb{R}^{(m+n) \times u}$:

$$H = (h_1, h_2, \dots, h_n, h_{c_1}, h_{c_2}, \dots, h_{c_m}) \quad (4)$$

3.3 全局事件演化模块

全局事件演化模块(GEEP)旨在通过构造事理图谱来学习事件演化模式,以指导后续事件的预测。GEEP 模块分为两个步骤:1)根据训练集中的叙事事

件链构建事理图谱;2)使用可缩放的图神经网络(Scaled Graph Neural Network, SGNN)^[16]学习隐藏在事理图谱中的事件演化模式。

3.3.1 事理图谱构建

本文把事理图谱(Event Evolutionary Graph, EEG)形式化表示为 $\text{EEG} = \{V, Q\}$, 其中, V 表示节点集, Q 表示边集。为了缓解稀疏性问题,节点集中的节点用每个事件的谓语动词 p 表示。节点 p_i 和 p_j 之间的边权重 $w(p_i, p_j)$ 表示在事件 p_i 发生的条件下, p_j 发生的可能性,其计算过程如式(5)所示:

$$w(p_i, p_j) = \frac{c_{\text{count}}(p_i, p_j)}{\sum_k c_{\text{count}}(p_i, p_k)} \quad (5)$$

其中: $c_{\text{count}}(p_i, p_j)$ 表示事件 p_i 和 p_j 在训练集的事件链中同时出现的次数。构建完事理图谱后,采用 SGNN 对事理图谱进行学习。

3.3.2 基于 SGNN 的事理图谱学习

为了节省计算资源和提高效率,把每个事件链视为一个子图输入门控图神经网络(Gated Graph Neural Network, GGNN)^[19]学习事件链中每个事件的表示,这种把子图作为输入的图神经网络被称为可缩放的门控图神经网络(SGNN)^[17]。SGNN 的输入分为两部分:一部分是叙事事件链的所有事件表示 E , 另一部分是子图的邻接矩阵 A 。事件表示 E 如式(6)所示,子图的邻接矩阵 A 从全局事理图谱中抽取,抽取过程如式(7)所示:

$$E = (v(e_1), v(e_2), \dots, v(e_n), v(e_{c_1}), v(e_{c_2}), \dots, v(e_{c_m})) \quad (6)$$

$$A[i, j] = \begin{cases} w(p_i, p_j), & (p_i, p_j) \in Q \\ 0, & \text{其他} \end{cases} \quad (7)$$

SGNN 的计算过程如式(8)~式(12)所示:

$$a^{(t)} = A^T \cdot E^{(t-1)} + b \quad (8)$$

$$z^{(t)} = \sigma(W^z \cdot a^{(t)} + U^z \cdot E^{(t-1)}) \quad (9)$$

$$r^{(t)} = \sigma(W^r \cdot a^{(t)} + U^r \cdot E^{(t-1)}) \quad (10)$$

$$c^{(t)} = \tanh(W \cdot a^{(t)} + U(r^{(t)} \odot E^{(t-1)})) \quad (11)$$

$$E^{(t)} = (1 - z^{(t)}) \odot E^{(t-1)} + z^{(t)} \odot c^{(t)} \quad (12)$$

式(8)是指子图的节点通过邻接矩阵进行消息的传递,式(9)~式(12)和 GRU 类似,通过 $t-1$ 时刻的信息和其他节点来更新 t 时刻节点的状态。 $z^{(t)}$ 是更新门, $r^{(t)}$ 是重置门, σ 是 sigmoid 激活函数, \odot 是按位相乘。GEEP 模块的输出为 t 时刻的节点表示 $E^{(t)}$, 为了下文表述方便,把 GEEP 模块的输出简单记为 E , E 和 LRTO 模块的输出 H 具有相同的维度。

3.4 门控注意力层

门控注意力层的输入是 LRTO 模块的输出 $H \in \mathbb{R}^{(m+n) \times u}$ 和 GEEP 模块的输出 $E \in \mathbb{R}^{(m+n) \times u}$ 。为了解决时序信息和图信息对不同的样本预测的重要程度不同这一问题^[20],本文提出一种门控机制 G 来表示时序信息和演化模式的重要程度,计算过程如式(13)所示:

$$G = \sigma([H, E] \cdot W_g + b_g) \quad (13)$$

其中: $G \in \mathbb{R}^{(m+n) \times u}$; $[H, E] \in \mathbb{R}^{(m+n) \times 2u}$ 表示 H 和 E 的拼接; $W_g \in \mathbb{R}^{2u \times u}$ 是权重矩阵; b_g 是偏置项。原始的 H 和 E 根据式(14)和式(15)得到新的 H_g 和 E_g :

$$H_g = H \odot G \quad (14)$$

$$E_g = E \odot (1 - G) \quad (15)$$

其中: \odot 指按位相乘; H_g 和 E_g 与原来的维度相同。

得到 H_g 之后,把 H_g 分为两部分:上下文事件 $H_g^1 \in \mathbb{R}^{n \times u}$ 和候选事件 $H_g^2 \in \mathbb{R}^{m \times u}$ 。根据式(16)得到上下文事件和候选事件的权重矩阵 $\alpha \in \mathbb{R}^{n \times m}$ 。根据式(17)得到上下文事件链对于候选事件的代表向量 $v_h \in \mathbb{R}^{m \times u}$,最后把该代表向量与候选事件之间的余弦相似性作为候选事件的得分 s_h ,如式(18)所示:

$$\alpha = \text{softmax}(H_g^1 \cdot (H_g^2)^T) \quad (16)$$

$$v_h = \text{relu}(\alpha^T \cdot H_g^1) \quad (17)$$

$$s_h = \text{cosine}(H_g^2 \cdot v_h) \quad (18)$$

得到 E_g 后,GEPP模块对每个候选事件的打分 s_g 根据同样的方式计算得到,计算过程如式(19)~式(21)所示:

$$\beta = \text{softmax}(E_g^1 \cdot (E_g^2)^T) \quad (19)$$

$$v_g = \text{relu}(\beta^T \cdot E_g^1) \quad (20)$$

$$s_g = \text{cosine}(E_g^2 \cdot v_g) \quad (21)$$

在此基础上,通过式(22)计算候选事件 e_{cj} ($j \in [1, m]$) 的最终得分 $s_{\text{score}}(e_{cj})$:

$$s_{\text{score}}(e_{cj}) = w \cdot s_h(e_{cj}) + (1 - w) \cdot s_g(e_{cj}) \quad (22)$$

其中: w 是可学习的参数。

最后,对于每个叙事事件链对应的 m 个候选事件,采用 softmax 计算每个候选事件最终的发生概率,如式(23)所示,选择发生概率最高的事件作为模型输出,如式(24)所示:

$$p(e_{cj} | e_1, e_2, \dots, e_n) = \frac{\exp(s_{\text{score}}(e_{cj}))}{\sum_{k \in [1, m]} \exp(s_{\text{score}}(e_{ck}))} \quad (23)$$

$$e_c^* = \underset{k}{\operatorname{argmax}} p(e_{ck} | e_1, e_2, \dots, e_n) \quad (24)$$

3.5 模型训练

给定一系列的叙事事件链和候选事件集,训练目标是使最小化式(25)所示的边界损失函数:

$$L(\theta) = \sum_i \sum_j (\max(0, \gamma - s_{ij} + s_{ij})) + \frac{\lambda}{2} (\|\theta\|)^2 \quad (25)$$

其中: N 是事件链的条数; m 是叙事事件链对应的候选集的大小; s_{ij} 表示第 i 个叙事事件链对第 j 个候选事件的打分; y 表示正确后续事件的标签; γ 是边界损失函数的参数,在本文中设置为 0.015; λ 是 L_2 正则项的惩罚因子,在本文中设置为 $1e-8$; θ 表示模型的所有参数。在验证集上调整所有的超参数。每轮训练样本数为 1 000,丢弃率为 0.4,初始词嵌入为 128 维,SGNN 的层数设置为 2 层,学习率为 0.000 1,在训练过程中使用 RMS 优化算法。

4 实验与结果分析

本文采用 MCNC 评估方法,将 ECGNet 模型和多个当前最新的基线模型进行对比,验证其有效性。

4.1 数据集

本文使用两个数据集进行实验。一个是文献[17]公开的 NYT 英文数据集,该数据集从 Gigaword 语料库的纽约时报部分中进行提取。NYT 数据集共包括 160 331 个样本,本文按照 140 331:10 000:10 000 的比例划分训练集、验证集和测试集。

由于目前为止脚本事件预测领域仅有一个公开的英文数据集,因此本文基于文献[21]收集的新浪新闻语料库处理了一个中文数据集,命名为 SinaNews。原始新闻语料库由不同系列主题的新闻组成,每个系列包含了同一主题的众多新闻文章。本文把同一主题下的新闻按照时间顺序视为叙事事件链,每则新闻被视为一个事件,由于新闻的标题包含了该则新闻的主要内容,因此本文仅使用新闻标题作为事件。具体处理流程按照以下步骤进行:

1) 划分叙事事件链。将每个主题下的新闻按照时间顺序划分,每 5 则新闻为一组,一组新闻被视为一个事件链。其中前 4 则新闻作为上下文事件,最后一则新闻作为要预测的新闻事件,即预测标签。经过划分,得到 147 622 个来自不同主题的新闻事件链,按照 127 622:10 000:10 000 的比例划分训练集、验证集和测试集。

2) 抽取新闻事件。通过依赖解析从每则新闻标题中抽取事件的相关要素(主语、谓语、宾语、介词宾语)。依赖解析工具为 HanLP[22]。

3) 构造候选事件集。遵循 MCNC 标准,本文对每个叙事事件链构造候选事件集,把每组新闻中的最后一则新闻作为标签,从事件集中随机抽取 $m-1$ 则新闻,用标签事件的主语替换随机选取的新闻事件的主语。因此,候选集中有且仅有一个正确的后续事件,且候选集中的所有的事件共享主语。

4.2 基线模型

为了对 ECGNet 模型的效果进行全面评估,本文分别选择了经典的基于统计学和事件对的脚本事件预测方法,以及最新的基于事件链和事理图谱的方法作为基线模型:PMI[4]和 Bigram[5]是基于统计的模型;Word2Vec[14]和 EventComp[6]是基于事件对的模型,它们通过学习上下文事件和候选事件对之间的关系对后续事件预测的影响;PairLSTM[15]、HierLSTM[21]和 SAM-Net[16]是基于事件链的方法,它们通过循环神经网络学习叙事事件链的时序信息进行后续事件预测;SGNN[17]是一种基于事理图谱的方法,其构造叙事事理图谱描述事件之间的发展规律,并提出采用可缩放的图神经网络学习事理图谱中的事件表示进行后续事件预测。

4.3 总体实验结果

ECGNet 和基线模型在测试集上的结果如表 1 所示。根据表中的结果能够得到以下结论:

1) EventComp、PairLSTM 和其他的神经网络模型效果优于统计模型。这是因为基于统计的模型无法应对稀疏性和泛化性的问题。

2)基于事件链和事理图谱的方法与基于事件对的方法相比效果有所提升,证明了仅考虑上下文事件和候选事件对之间的关系进行后续事件预测是不够的。基于事件链的方法和基于事理图谱的方法取得了较好的实验效果。这证明了时序信息和事件演化模式对后续事件的预测的重要性。

3)本文提出的 ECGNet 模型取得的实验效果优于所有基线模型且提升明显(在两个数据集上均比次优模型提升预测准确率 3% 以上),这充分证明了同时利用时序信息和演化模式进行后续事件预测是非常重要的。此外,ECGNet 模型在中英文数据集上均取得了最好的实验效果,也验证了模型的鲁棒性。

表 1 不同模型的准确率

Table 1 Accuracy of different models %		
模型	NYT	SinaNews
PMI	31.25	36.57
Bigram	28.76	34.64
Word2Vec	42.10	43.37
EventComp	49.66	58.49
PairLSTM	50.34	61.64
HierLSTM	50.86	61.82
SAM-Net	52.68	60.38
SGNN	52.32	61.58
ECGNet	56.11	65.26

4.4 事件表示层探究实验

本节采用一系列对比实验验证事件表示方法对预测结果的影响。不同方法的输入均为一个事件的 4 个元素的词嵌入,输出为该事件的表示。

“Concat”方法把事件的 4 个元素的词嵌入进行拼接作为事件表示。“Average”方法把 4 个元素的词嵌入进行平均作为事件表示。“Comp”方法遵循前人的研究,采用 tanh 层把 4 个元素进行组合得到事件的表示,计算过程如式(26)所示:

$$v(e_i) = \tanh(W_p v_p + W_0 v_{a0} + W_1 v_{a1} + W_2 v_{a2}) \quad (26)$$

“LSTM”方法把 4 个元素输入 LSTM,然后把每个元素的隐藏状态进行拼接作为事件表示。“Transformer”是 ECGNet 的事件表示层。

不同的事件表示方法的实验结果如表 2 所示。根据表中的结果可以发现:“Average”方法在两个数据集上的表现最差,这是因为把各个元素的词嵌入进行简单的平均会丢失一些特征;“LSTM”方法考虑到了表示一个事件的 4 个元素之间的时序信息,因此取得了比“Concat”和“Comp”更好的实验效果;ECGNet 模型采用 Transformer 编码器进行事件表示,在两个数据集上都取得了最好的效果。这也验证了本文的假设:事件内部的 4 个元素之间的时序语义信息对于事件表示和后续预测是非常重要的,Transformer 编码器能够较好地捕获这些时序和语义信息,从而获得更好的预测效果。

表 2 不同事件表示方法的准确率

Table 2 Accuracy of different event representation methods %

方法	NYT	SinaNews
Concat	52.34	62.00
Average	48.98	58.26
Comp	52.65	60.97
LSTM	53.25	62.16
Transformer	56.11	65.26

4.5 消融实验

本节通过消融实验,验证 ECGNet 各个模块的效果,结果如表 3 所示。

表 3 消融实验结果

Table 3 Ablation experiment results %		
方法	NYT	SinaNews
-ER	52.34	62.00
-LRTO	55.33	64.07
-GEEP	55.36	64.58
-Gated M-att	55.01	64.37
Transformer+att	46.13	51.16
ECGNet	56.11	65.26

对表 3 中的结果进行分析,具体如下:

“-ER”表示去掉事件表示层,简单地把事件的 4 个元素的词嵌入进行拼接作为事件表示进行后续预测。去掉事件表示层之后在 NYT 和 SinaNews 上准确率分别下降了 3.77 和 3.26 个百分点,验证了事件表示对于后续事件预测有着至关重要的作用。

“-LRTO”表示去掉 LRTO 模块,只考虑事理图谱中的事件发展规律和演化模式,使用 GEEP 模块得到的事件表示进行后续预测。去掉 LRTO 模块之后,实验效果有明显的下降,这证明了叙事事件链中上下文事件之间的时序特征不可忽视。

“-GEEP”表示去掉 GEEP 模块,只考虑叙事事件链中的时序信息,使用 LRTO 模块得到的事件表示进行后续预测。从实验结果可以看到,去掉 GEEP 模块后,预测效果明显下降,验证了事理图谱中的事件演化模式对于脚本事件预测任务是非常重要的。

“-Gated M-att”表示去掉门控注意力层,在 LRTO 模块和 GEEP 模块分别使用注意力机制对候选事件进行打分,将两个模块对候选事件的打分进行相加作为该候选事件的得分。实验结果证明了本文提出的门控注意力机制能够动态融合时序信息和演化模式,进行后续事件预测。

由于去掉事件表示层后预测效果下降明显,本文设置了“Transformer+att”实验,即通过 Transformer 学习到事件表示之后,使用注意力机制和 softmax 层进行后续事件预测。实验结果表明,即使 Transformer 具有强大的表示学习能力,但仍然需要结合叙事事件链中

的时序信息和事理图谱中的全局事件演化模式,才能取得更好的预测效果。

5 结束语

本文针对脚本事件预测事件表示不全面、信息融合不充分等问题,提出一种动态融合事件链和事理图谱的模型ECGNet。在NYT英文数据集和SinaNews中文数据集上,使用预测准确率作为评估标准,ECGNet模型取得了较好的效果。此外,本文还进行了事件表示探究实验,实验结果证明使用Transformer编码器学习事件表示的方法非常有效。同时,消融实验也验证了ECGNet模型各个模块的有效性。本文模型只对事件链中的时序信息和事理图谱信息进行了融合,下一步将通过深度整合知识图谱和事理图谱,进一步提升脚本事件预测的准确率。

参考文献

- [1] KANJIRANGAT V, MELLACE S, ANTONUCCI A. Temporal embeddings and transformer models for narrative text understanding[C]//Proceedings of the 3rd Workshop on Narrative Extraction from Texts Co-located with the 42nd European Conference on Information Retrieval. Lisbon, Portugal:CEUR-WS, 2020:71-77.
- [2] TAKATSU H, YOKOYAMA K, MATSUYAMA Y, et al. Recognition of intentions of users' short responses for conversational news delivery system[C]//Proceedings of the 20th Annual Conference of the International Speech Communication Association. Graz, Austria:ISCA, 2019:1193-1197.
- [3] PICKERING T, JORDANOUS A. Applying narrative theory to aid unexpectedness in a story generation system[C]//Proceedings of the 8th International Conference on Computational Creativity. Georgia, USA:ACC, 2017:213-220.
- [4] CHAMBERS N, JURAFSKY D. Unsupervised learning of narrative event chains[C]//Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA:ACL, 2008:789-797.
- [5] JANS B, BETHARD S, VULIC I, et al. Skip n-grams and ranking functions for predicting script events[C]//Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg, USA:ACL, 2012:336-344.
- [6] GRANROTH-WILDING M, CLARK S. What happens next? Event prediction using a compositional neural network model[C]//Proceedings of the 13th AAAI Conference on Artificial Intelligence. Palo Alto, USA:AAAI, 2016:2727-2733.
- [7] DING X, LIAO K, LIU T, et al. Event representation learning enhanced with external commonsense knowledge[C]//Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg, USA:ACL, 2019:4893-4902.
- [8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of NIPS'17. [S.l.]:NIPS, 2017:5998-6008.
- [9] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [10] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2009, 20(1):61-80.
- [11] SCHANK R C, ABELSON R P. Scripts, plans, goals, and understanding[M]. New York, USA:Psychology Press, 2013.
- [12] BALASUBRAMANIAN N, SODERLAND S, MAUSAM, et al. Generating coherent event schemas at scale[C]//Proceedings of 2013 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA:ACL, 2013:1721-1731.
- [13] PICHOTTA K, MOONEY R J. Statistical script learning with multi-argument events[C]//Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg, USA:ACL, 2014:220-229.
- [14] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]//Proceedings of NIPS'13. [S.l.]:NIPS, 2013:3111-3119.
- [15] WANG Z Q, ZHANG Y, CHANG C Y. Integrating order information and event relation for script event prediction[C]//Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA:ACL, 2017:57-67.
- [16] LÜ S W, QIAN W H, HUANG L T, et al. SAM-Net: integrating event-level and chain-level attentions to predict what happens next[C]//Proceedings of the 33th AAAI Conference on Artificial Intelligence. Palo Alto, USA:AAAI, 2019:6802-6809.
- [17] 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. Stockholm, Sweden:IJCAI, 2018:4201-4207.
- [18] PENNINGTON J, SOCHER R, MANNING C. GloVe: global vectors for word representation[C]//Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA:ACL, 2014:1532-1543.
- [19] LI Y J, TARLOW D, BROCKSCHMIDT M, et al. Gated graph sequence neural networks[C]//Proceedings of the 4th International Conference on Learning Representations. Stroudsburg, USA:ACL, 2018:273-283.
- [20] CHEN Y B, 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 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA:ACL, 2018:1267-1276.
- [21] HU L M, LI J Z, NIE L Q, et al. What happens next? Future subevent prediction using contextual hierarchical LSTM[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. Palo Alto, USA:AAAI, 2017:3450-3456.
- [22] HE H. HanLP: Han language processing[EB/OL]. [2020-10-25]. <https://github.com/hankcs/HanLP>.