

0. 摘要

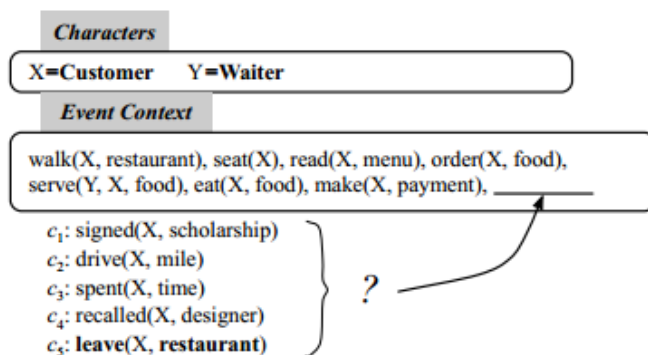
脚本事件预测需要一个模型来预测给定现有事件上下文的后续事件。以前基于事件对或事件链的模型不能充分利用密集的事件连接，这可能会限制它们的事件预测能力。为了弥补这一缺陷，文章构建一个事件图，以便更好地利用事件网络信息来预测脚本事件。

首先从大量新闻语料中提取叙事事件链，然后构建叙事事件进化图 (NEEG) 基于提取的链。NEEG可以被视为描述事件进化原理和模式的知识库。为了解决NEEG上的推理问题，我们提出了一种可扩展图神经网络 (SGNN) 来模拟事件交互并学习更好的事件表示。

SGNN 不是计算整个图的表示，而是每次只处理相关节点，这使得我们的模型对于大规模图是可行的。通过比较输入上下文事件表示和候选事件表示之间的相似性，我们可以选择最合理的后续事件。在广泛使用的《纽约时报》语料库上的实验结果表明通过使用经典的挖词填空式的评估标准来评估模型，该模型明显优于最先进的方法。

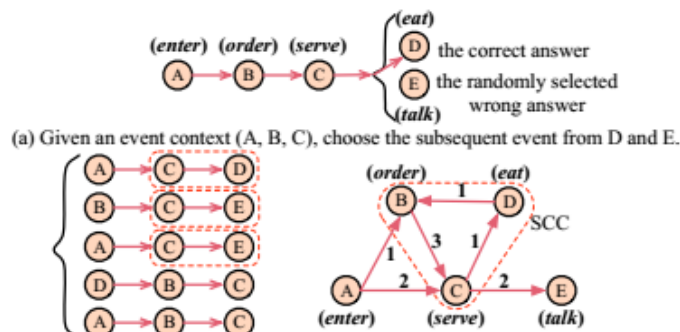
1. 简介

理解文本中描述的事件对许多人工智能 (AI) 应用至关重要，例如话语理解、意图识别和对话生成。脚本事件预测是这项工作中最具挑战性的任务。这项任务是Chambers 和 Jurafsky 在2008年首次提出的，他们将它定义为给出一个现有的事件背景，需要从候选人列表中选择最合理的后续事件(如图1所示——典型的餐厅脚本)



前人在脚本事件预测任务上应用的典型方法包括基于事件对的PMI [3]和EventComp [4]，以及基于事件链条的PairLSTM [5]。尽管这些方法取得了一定的成功，但是事件之间丰富的连接信息仍没有被充分利用。为了更好地利用事件之间的稠密连接信息，本文提出构建一个叙事事理图谱 (Narrative Event Evolutionary Graph, NEEG)，然后在该图谱上进行网络表示学习，进一步预测后续事件。

下图的例子对本文想要利用更丰富事件连接信息 (事件图结构) 的动机进行了很好的说明。给定事件上下文A(enter), B(order), C(serve), 我们想要从D(eat)和E(talk)中挑选出正确的后续事件。其中，D(eat)是正确的后续事件而E(talk)是一个随机挑选的高频混淆事件。基于事件对和事件链的方法很容易挑选出错误的E(talk)事件，因为图3(b)中的事件链条显示C和E比C和D有更强的关联性。然而，基于(b)中的链条构建 (c) 中的事件网络之后，我们发现B，C和D形成了一个强连通分量。如果能够学习到B，C和D形成的这种强连接图结构信息，则我们更容易预测出正确的后续事件D。



有了一个叙事事理图谱之后，为了解决大规模图结构上的推理问题，文章中提出了一个可扩展的图神经网络模型在该图谱上进行网络表示学习。

近期，许多网络表示学习方法涌现出来，典型的浅层方法如DeepWalk、LINE、Node2Vec，最早提出的用于图结构上的神经网络方法如GNN以及改进方法GGNN，近期基于卷积的方法如GCN，GraphSAGE，GAT等。然而，这些方法存在这样那样的问题，不能够直接适用于本文的任务。有的是无监督方法，不能有效利用有监督信息，性能会打折扣；有的只能应用于小规模图结构，扩展性较差；有的只能用于无向图，而本文的叙事事理图谱是有向有环加权图；有的只能学习到较短距离的信息传递。

通过比较，文章选择了更为合适的GGNN模型。但是它不能直接用于本文的大规模叙事事理图谱图结构上。本文通过在训练中每次只在一个小规模相关子图上进行计算，来将GGNN扩展到大规模有向有环加权图上，改进后的模型叫做Scaled Graph Neural Network (SGNN)。该模型经过修改也可以应用于其他网络表示学习任务中。

2. 模型（方法）

如下图所示，模型包括两步。第一步是构建一个基于叙述事件链的事理图谱；第二步是提出了一个比例图神经网络来解决在构造的事件图上的推理问题。

2.1 叙事事理图谱的构建

NEEG的构建包含两步：(1)从新闻通讯社语料库中提取叙事事件链；(2)构建基于已提取的事件链的NEEG

为了与前人工作对比，我们使用了相似的语料库和事件链提取方法（2016 Granroth-Wilding and Clark）。提取出了一系列叙事事件链 $S = \{s_1, s_2, s_3, \dots, s_N\}$ 举个例子，假设

$s_i = \{walk(T, restaurant, -), seat(T, -, -), read(T, menu, -), order(T, food, -), serve(waiter, food, T), eat(T, food, fork)\}$

T是这个链条中所有事件共享的主角实体， e_i 是一个四元组 $\{p(a_0, a_1, a_2)\}$ 组成的事件，其中 p 是预测动词， a_0, a_1, a_2 是动词的主语、宾语和间接宾语。

NEEG也可以形式化的表示为 $G = \{V, E\}$ ，其中 $V = \{v_1, v_2, v_3, \dots, v_P\}$ 是节点的集合， $E = \{l_1, l_2, l_3, \dots, l_Q\}$ 是边的集合。为了克服事件的稀疏性问题，我们用抽象形式 (v_i, r_i) 来表示事件 e_i ，其中 v_i 由非引理谓词的动词表示， r_i 是 v_i 与链实体T的语法依存关系，如 $e_i = (eats, subt)$ 。这种事件表示被称为表语。（Granroth-Wilding and Clark, 2016）

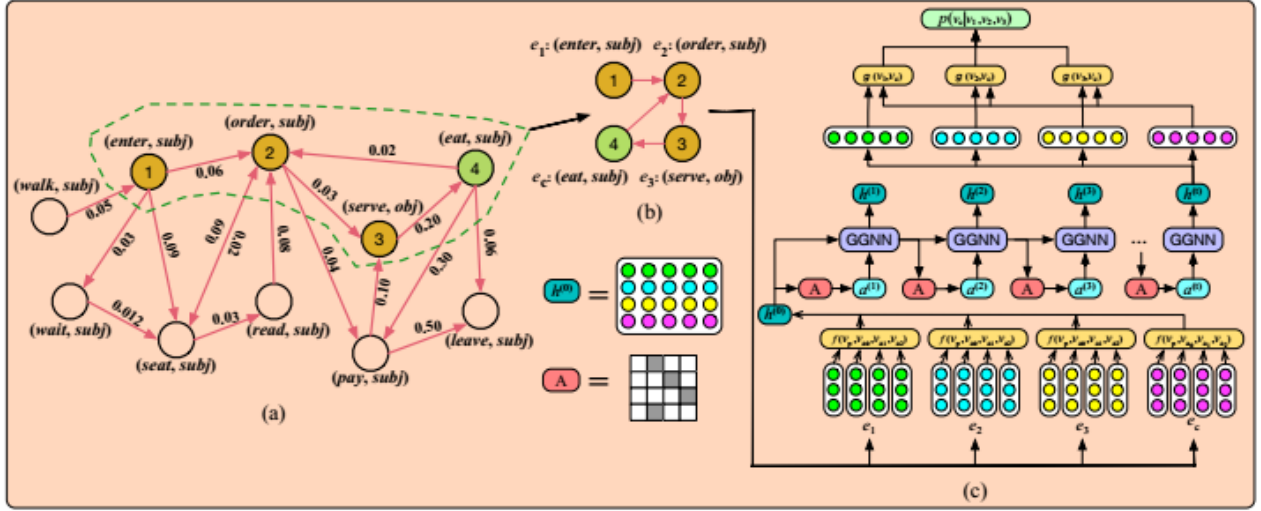
对训练事件链中的所有谓词GR二元图进行计算，并将每个谓词GR二元图视为E中的边 l_i 。每个 l_i 是一个有向边 $v_i \rightarrow v_j$ 以及一个权重 w ，可以通过以下公式计算：

$$w(v_j|v_i) = \frac{count(v_i, v_j)}{\sum_k count(v_i, v_k)}$$

其中 $count(v_i, v_j)$ 表示二元组 (v_i, v_j) 在训练事件链中出现的频率。

上面的简单来说就是 所有的事件二元图都形成一条事件演化有向边， $v_i \rightarrow v_j$ 边上的权重 w 可以用上面的公式计算

构建完成的叙事事理图谱G有104940个预测GR节点和6187046个直接和权重边。图表3展示了G中的一个本地子图，描述了一个在餐馆场景可能出现的事件。不同于事件对或事件链，事件图有紧密的联系并且包含更丰富的事件交互信息。



(图表3 本文提出的SGNN模型的框架, 假设a是构建好的NEEG, 子图可能在每时每刻包含了所有的情景信息 (节点1,2,3) 和所有的被选择的候选通信节点(图中仅仅画出了一个候选事件4))

2.2 可扩展图神经网络模型

2005年 最早提出图神经网络 (GNN)

Li等人改进 ——> 将门控循环单元 (GRU) 应用到GNN上形成了GGNN

缺陷: 多用于小规模, 几十上百个节点, 每次都要对整个图进行处理, 不适用于本文的大规模图结构

解决上述的问题, 引入“分治”的思想

每次处理涉及到的一个子图 (包括一条上下文和候选事件节点, 以及他们的边, 如上图中的b)

如上图c所示, SGNN模型主要包含3个部分: 事件表示层、GGNN计算层、挑选后续事件层

2.2.1 学习初始事件表示

NEEG中的每个抽象事件节点实际上表示了多个具体的四元组事件。为了将4个事件元素组合成一个事件表示 向量, 本文对比了如下几种方法:

- 平均法: 使用所有4个事件元素词向量的平均值来作为整个事件的表示;
- 非线性变换: $v_e = \tanh(W_p \cdot v_p + W_0 \cdot v_{a_0} + W_1 \cdot v_{a_1} + W_2 \cdot v_{a_2} + b)$
- 拼接: 直接将4个事件元素词向量拼接起来作为整个事件的表示。

2.2.2 更新子图中的事件表示

GGNN被用来学习事件之间的交互作用并更新每个子图上的事件表示。每次输入到SGNN的为两个矩阵 $h^{(0)}$ and A 。其中 $h^{(0)}$ 包含了上下文加所有候选事件的向量表示; A 是对应的子图邻接矩阵, 该矩阵决定了子图 中的事件如何交互作用。如下面的公式所示:

$$A[i, j] = \begin{cases} w(v_j | v_i) & \text{if } v_i \rightarrow v_j \in E \\ 0 & \text{others.} \end{cases}$$

GGNN模型运作方式类似广泛应用的GRU模型, 如下面公式所示。不同的是每个循环GGNN都会更新子图中所有节点的事件表示, 每次循环节点之间的信息传递给一度相邻节点。多个循环得以让各个节点的信息在子图 结构上充分流动交互。最终模型的输出为 $h^{(t)}$, 包含了学习到的子图中的事件表示。

$$a^{(t)} = A^T h^{(t-1)} + b$$

$$z^{(t)} = \sigma(W^z a^{(t)} + U^z h^{(t-1)})$$

$$r^{(t)} = \sigma(W^r a^{(t)} + U^r h^{(t-1)})$$

$$c^{(t)} = \tanh(W a^{(t)} + U(r^{(t)} \odot h^{(t-1)}))$$

$$h^{(t)} = (1 - z^{(t)}) \odot h^{(t-1)} + z^{(t)} \odot c^{(t)}$$

2.2.3 挑选正确的后续事件

得到事件表示 $h^{(t)}$ 之后，文章中采用了与前人类似的方法进行后续事件的挑选。

第 i 个上下文事件与第 j 个候选事件的相关性分数可以通过 $s_{ij} = g(h_i^{(t)}, h_{c_j}^{(t)})$ 进行计算，整个上下文与第 j 个候选事件的相关性用 $s_j = \frac{1}{n} \sum_{i=1}^n s_{ij}$ 进行计算，基于公式 $c = \max_j s_j$ 来挑选出正确的那个候选事件作为最终的预测结果。

文章中还比较了多个相似性计算函数 g ，比如Manhattan相似度、cosine相似度、内积相似度以及欧几里得相似度

而且在 s_{ij} 的计算过程中引入了注意力（attention）机制： $s_{ij} = \alpha_{ij} g(h_i^{(t)}, h_{c_j}^{(t)})$

2.2.4 损失函数

包括一个大间隔损失和一个正则项。其中 s_{Iy} 表示第 y 个上下文和第 y 个候选事件的相关分数， I_y 是对应的正确候选事件的索引。

$$L(\Theta) = \sum_{I=1}^N \sum_{j=1}^k (\max(0, \text{margin} - s_{Iy} + s_{Ij})) + \frac{\lambda}{2} \|\Theta\|^2$$

3. 实验

3.1 数据集

3.2 实验结果

	Training	Development	Test
#Documents	830,643	103,583	103,805
#Chains for NEEG	5,997,385	-	-
#Chains for SGNN	140,331	10,000	10,000

Table 1: Statistics of our datasets.

Methods	Accuracy
Random	20.00
PMI [Chambers and Jurafsky, 2008]	30.52
Bigram [Jans <i>et al.</i> , 2012]	29.67
Word2vec [Mikolov <i>et al.</i> , 2013]	42.23
DeepWalk [Perozzi <i>et al.</i> , 2014]	43.01
EventComp [Granroth-Wilding and Clark, 2016]	49.57
PairLSTM [Wang <i>et al.</i> , 2017]	50.83
SGNN-attention (without attention)	51.56
SGNN (ours)	52.45
SGNN+PairLSTM	52.71
SGNN+EventComp	54.15
SGNN+EventComp+PairLSTM	54.93

Table 2: Results of script event prediction accuracy (%) on the test set. SGNN-attention is the SGNN model without attention mechanism. Differences between SGNN and all baseline methods are significant ($p < 0.01$) using t-test, except SGNN and PairLSTM ($p = 0.246$).

- 通过比较PMI、Bigram和其他方法，可以发现基于向量表示学习的方法显著超过了基于频率统计的方法；
- 用过比较Word2vec和DeepWalk，以及EventComp、PairLSTM和SGNN，可以发现基于图结构的方法超过了基于事件对或者事件链的方法；
- SGNN模型取得了单模型最好的预测结果52.45%；
- 在预测阶段使用注意力机制可以提升SGNN的性能，说明上下文中的事件对挑选后续事件有不同的重要性。此外，本文也进行了不同单模型的融合实验。结果表明不同模型之间可以互补，三个模型SGNN,EventComp和PairLSTM的组合取得了最好的实验结果54.93%。

4. 结论

本文简要介绍了我实验室提出的事理图谱概念及初步成果。在IJCAI的论文中，我们提出构建叙事事理图谱并基于网络表示学习来解决脚本事件预测的任务。具体地，为了更好地利用事件之间的稠密连接信息，本文首先从大规模无结构化语料中抽取脚本事件链条，然后基于这些链条构建了一个叙事事理图谱（NEEG）。为了解决NEEG大规模图结构上的推断问题，基于前人提出的图神经网络GGNN，本文提出了一个可扩展的图神经网络（SGNN）来建模事件之间的交互作用并学习到更好的事件表示，这些学习到的事件表示可以用于脚本事件预测任务。实验结果表明事件图结构比事件对或者事件链更加有效，可以显著提高脚本事件预测的准确率。