

• 计算机工程与应用 •

融合动态图表示和自注意力机制的级联预测模型



张凤荔, 王雪婷*, 王瑞锦, 汤启友, 韩英军

(电子科技大学信息与软件工程学院 成都 610054)

【摘要】传统的级联预测模型不考虑信息传播过程中的动态性且极大依赖于人工标记特征, 推广性差, 预测准确性低。为此, 该文提出一种融合动态图表示和自注意力机制的级联预测模型 (DySatCas)。该模型采用端到端的方式, 避免了人工标记特征造成级联图表示困难的问题; 通过子图采样捕获级联图的动态演化过程, 引入自注意力机制, 更好地融合在观测窗口中学到的信息级联图的动态结构变化和时序特征, 为网络合理地分配权重值, 减少了信息的损失, 提升了预测性能。实验结果表明, DySatCas 与现有的基线预测模型相比, 预测准确性有明显提升。

关键词 级联预测; 深度学习; 动态图表示; 信息传播; 自注意力机制

中图分类号 TP183 文献标志码 A doi:10.12178/1001-0548.2021100

Cascade Prediction model based on Dynamic Graph Representation and Self-Attention

ZHANG Fengli, WANG Xueting*, WANG Ruijin, TANG Qiyu, and HAN Yingjun

(School of Information and Software Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract The traditional cascade prediction models do not consider the dynamics features in the process of information diffusion and rely on artificial marking features heavily, which have the problems of poor generalization and low prediction accuracy. This paper proposes a cascade prediction model (information cascade with dynamic graphs representation and self-attention, DySatCas) that combines dynamic graph representation and self-attention mechanism. The model adopts an end-to-end approach, avoiding the difficult problem of cascade graph representation caused by artificial labeling features, capturing the dynamic evolution process of cascade graphs through sub-graph sampling. And it introduces a self-attention mechanism to better integrate in the dynamic structure changes and temporal characteristics of the information cascade graph learned in the observation window, which can allocate weight values to the network reasonably, reduce the loss of information, and improve the prediction performance. Experimental results show that DySatCas has significantly improved prediction accuracy compared with the existing baseline prediction model.

Key words cascades prediction; deep learning; dynamic graph representation; information diffusion; self-attention mechanism

如今, Twitter、微博、Facebook、YouTube 和微信等在线社交平台已成为热点事件传播的主要途径。某一事件很快就能在社交网络中通过新闻报道、论坛帖子^[1] 和视频动态等各种形式被广泛传播。在事件传播过程中, 信息级联是很重要的组成部分。信息级联是指博客、电子邮件以及社交网站

中识别到的用户级联转发序列与网络关系图^[2]。通过对信息级联模型的研究, 可以了解信息的传播规律和特性, 为事件演化下一步研究提供技术支持。

现今, 在信息级联预测领域中, 基于特征工程方法的模型只适用于特定场景^[3-4]; 依赖于假设的点过程 (如泊松过程和霍克斯过程) 的生成式方法的

收稿日期: 2021-04-12; 修回日期: 2021-09-03

基金项目: 国家自然科学基金 (61802033, 61472064, 61602096); 四川省区域创新合作项目 (2020YFQ0018); 四川省科技计划重点研发项目 (2021YFG0027, 2020YFG0475, 2018GZ0087, 2019YJ0543); 博士后基金项目 (2018M643453); 广东省国家重点实验室项目 (2017B030314131); 网络与数据安全四川省重点实验室开放课题 (NDSMS201606)

作者简介: 张凤荔 (1963-), 女, 教授, 博导, 主要从事网络安全和网络工程、云计算以及大数据和机器学习方面的研究。

*通信作者: 王雪婷, E-mail: chloewongxt@qq.com

模型无法同时捕获级联图的基础结构和扩散的不确定性^[4]；基于深度学习技术的模型，可以对时间序列和静态级联图结构进行建模。但现实生活中，信息级联图的结构是随着时间动态变化的。在建模过程中，考虑级联图的动态变化性，捕获级联图时序及结构特征，对于刻画信息传播过程有重要作用。

本文提出一个基于级联图的动态演化过程的模型。该模型考虑信息传播过程中的不确定性，将图的动态演化过程和自注意力机制相结合，实现对级联图的时序及结构特征的同时捕获；且仅依赖级联图的结构和时序信息，利用深度学习技术以端到端的方式，实现级联的潜在语义捕获。实验在新浪微博和 APS 实际数据集上进行，相比传统的级联预测模型，提升了预测的准确性。

1 方法分析

信息级联预测方法主要有以下 3 类：

1) 基于特征建模的预测方法。特征建模方法基于特征工程，利用经典机器学习技术建模进行级联预测。如文献 [5] 发现加入与用户相关的特征可以有效地预测用户是否会参与级联；文献 [6] 验证了在微博数据集中结构特征和时序特征的重要性。这些方法都需要某一特定领域的大量相关特征，具有特殊性，不能够迁移至其他应用场景；且用户个人信息和私有社交平台特征信息具有私密性，难以得到。

2) 基于生成式的预测方法。生成式方法主要从转发到达速率方面入手，对每条消息的到达强度函数独立建模。如文献 [7] 通过借鉴生存分析的思想来预测在线内容的流行程度；文献 [8] 提出了一个生成概率模型，用增强泊松过程来预测科学论文的行程度。文献 [9] 将级联转换成一组扩散路径，用自激励霍克斯点过程对级联进行建模。此类方法所观测的数据是观测时间窗口内的每个事件的发生，而不是未来要预测的增量流行度。

3) 基于深度学习的预测方法^[10]。在深度学习级联预测模型中，节点的嵌入方法对节点表示有着关键性的作用。节点表示的准确性，将直接影响模型预测结果。DeepCas^[11] 是基于图表示学习的信息级联预测方法，它在借鉴了 DeepWalk^[12] 的基础上，采用随机游走对级联图采样，然后将采样后的级联序列结合注意力机制传入双向门控递归单元 (Bi-GRU)，最终得到节点的嵌入。基于注意力机制的预测模型，采用 GloVe^[13] 对消息内容进行嵌入，并使用 node2vec 对用户图进行编码。基于深度学习

的级联预测方法，不需要在级联图上人工标记特征，预测准确性高。

现有的级联模型都是针对拥有固定节点集和边集的静态图。在现实生活中，级联图的本质是动态的，它的结构会随着时间的变化而呈现一个动态演化的过程。这就要求在学习级联图节点表示时，不仅要保留结构上的临近性，还要同时捕获其时序演变特征^[1]。文献 [14] 将自注意力机制应用于动态图表示；文献 [15] 同时考虑到图的异构性和动态性的特点，通过神经网络结合自注意力机制实现节点动态嵌入；文献 [16] 提出了一种能够捕捉动态图演化的动力学特征方法。上述方法针对的是广泛的动态图，在信息级联领域还没有相关应用。

因此本文提出 DySatCas 模型，主要将级联图的动态性定义为一系列的快照子图，融合自注意力机制捕获级联图的结构和时序特征，从而学习到节点的动态表示。

2 问题定义及模型设计

2.1 问题定义

现有 M 条消息，每条消息转发记录对应一张级联图，如图 1 所示，深灰色节点为消息发送者，白色节点为表示观测时间内转发该消息的用户，浅灰色节点为将会新增的转发用户，带箭头实线表示用户间的转发关系，虚线表示将会出现的转发关系。

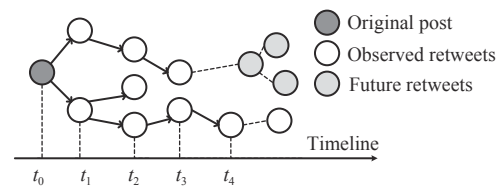


图 1 消息 C_i 的级联图

级联图的动态性定义为其 N 个演变级联子图，表示为 $G_i^T = \{\zeta_i(t_1), \zeta_i(t_2), \dots, \zeta_i(t_N)\}$, $\zeta_i(t_j) = (v_i^{t_j}, \varepsilon_i^{t_j}, t_j)$ ($1 \leq j \leq N$) 是级联图 G_i 在时间 t_j 的有向带权快照图^[17]。带权邻接矩阵为 A^j , $v_i^{t_j}$ 和 $\varepsilon_i^{t_j}$ 分别是图 $\zeta_i(t_j)$ 在时间 t_j 时的节点集合和边集合，具体级联演化示例如图 2 所示。

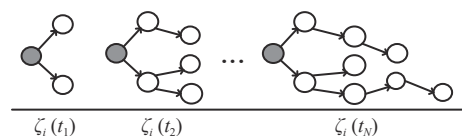


图 2 消息 C_i 的级联演化图 G_i^T

将信息级联预测问题定义为: 给定消息 C_i 的级联图 G_i 和在观测时间 $[0, T]$ 内的级联演化序列, G_i^T 预测此条消息在观测窗口结束时, 信息级联的增长量 $P_i = |v_i^{T+\Delta t}| + |v_i^T|$, $|v_i^*|$ 代表级联图的大小。

2.2 级联预测模型 DySatCas

融合动态图表示和自注意力机制的级联预测模型 DySatCas 如图3所示。整个模型分为4个部分, 分别是节点嵌入层、结构自注意力层、时序自注意力层、级联增量预测层。以某一级联图 G_i 及其演变子图 G_i^T 为例:

1) 节点嵌入层: 输入级联图 G_i 和演变子图 G_i^T , 本层将采样得到的演变子图分别进行编码, 得

到节点的向量表示矩阵 x_v 。

2) 结构自注意力层: 将节点嵌入层得到的向量矩阵 x_v 输入该层, 计算节点及其邻居节点的注意力权重, 并将该节点的邻居节点的特征聚合到该节点, 得到该节点新的向量表示 h_v^j 。

3) 时序自注意力层: 将结构自注意力层输出的节点向量表示 h_v^j 和该节点的位置嵌入 p^j 求和后, 传入时序自注意力层。该层对每个节点在其历史子图上的状态进行学习, 得到每个节点的表示 \tilde{h}_v^j 。

4) 级联增量预测层: 将时间 T 得到的每个节点的表示 \tilde{h}_v^N 输入到多层感知器中, 最后得到级联增量的预测结果 P_i 。

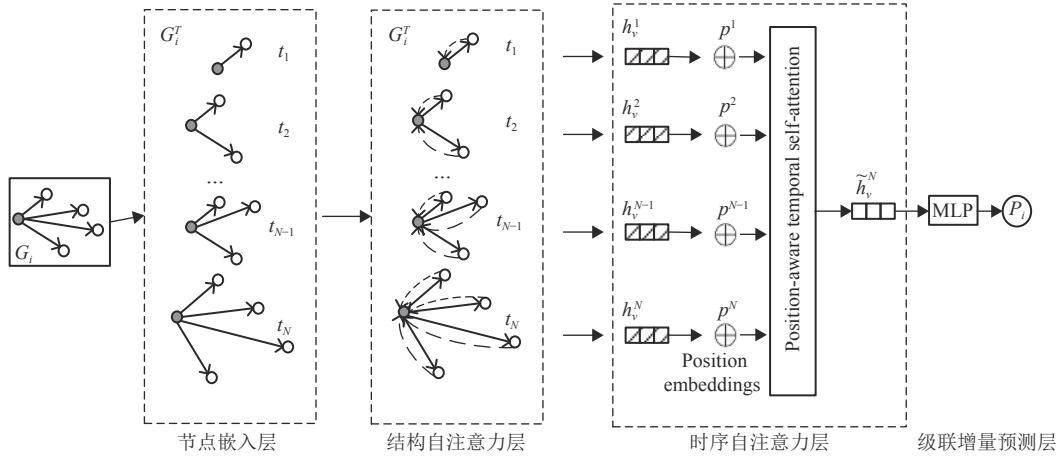


图3 融合动态图表示和自注意力机制的级联预测模型

2.2.1 节点嵌入层

在级联图 G_i 的基础上, 以相同时间间隔 T/N 作为标准得到演变级联子图 G_i^T , $G_i^T = \{\zeta_i(t_1), \zeta_i(t_2), \dots, \zeta_i(t_N)\}$ 。然后将 $\zeta_i(t_j)$ ($1 \leq j \leq N$)中的节点嵌入到向量空间 $\{x_v^j \in \mathbb{R}^D, \forall v \in V\}$, D 是节点嵌入的维度, 可调整。

2.2.2 结构自注意力层

结构自注意力层的主要作用是通过多层堆叠的自注意力机制对节点及其邻居节点的特征进行学习, 将学习到的特征聚合到该节点上, 得到对该节点的空间结构特征表示。该层的输入是演变级联子图 $\zeta_i(t_j)$ 中每个节点的 D 维向量表示 $\{x_v^j \in \mathbb{R}^D, \forall v \in V\}$ 。通过该层可捕获到节点在级联子图 $\zeta_i(t_j)$ 的局部属性, 输出为新的可调整的 F 维向量表示 $\{h_v^j \in \mathbb{R}^F, \forall v \in V, j \in [1, N]\}$ 。

具体来说, 结构自注意力层通过计算在 $\zeta_i(t_j)$ 中节点 v 直接邻居的注意力权重, 来作为函数的输入节点嵌入。具体运算定义为:

$$h_v^j = \sigma \left(\sum_{u \in N_v} \alpha_{uv} W^S x_u^j \right) \quad \alpha_{uv} = \frac{e^{e_{uv}}}{\sum_{w \in N_v} e^{e_{vw}}} \quad (1)$$

$$e_{uv} = \sigma \left(A_{uv}^j \cdot a^T \left[W^S x_u^j \parallel W^S x_v^j \right] \right) \quad \forall (u, v) \in \varepsilon$$

式中, $N_v = \{u \in V : (u, v) \in \varepsilon\}$ 为节点 v 在图 $\zeta_i(t_j)$ 中直接邻居的集合; $W^S \in \mathbb{R}^{F \times D}$ 是应用于图中每个节点变换的共享权重; $a \in \mathbb{R}^{2D}$ 是注意力函数的加权向量参数化; \parallel 表示拼接操作; $\sigma(\cdot)$ 是非线性激活函数; A_{uv}^j 是当前图中边 (u, v) 的权重; α_{uv} 是通过 softmax 函数对在 V 中每个节点的邻居节点进行归一化操作后的学习参数, 表示在当前图中节点 u 对于节点 v 的影响力; 同时使用 LeakyRELU 非线性激活函数来计算注意力权重, 然后用指数线性单元 (exponential linear unit, ELU) 来激活, 并将激活结果作为输出表示。当 α_{uv} 为0时, 节点 u 和 v 在图 $\zeta_i(t_j)$ 中是没有连接的, 使用稀疏矩阵可以有效的实现其邻居节点所

隐藏的自我注意力。

2.2.3 时序自注意力层

时序自注意力层同结构自注意力层类似,采用多个堆叠的时序自注意力模块,充分考虑级联图的动态演化过程,捕捉多个时间段的节点的时序变化。该层的输入是特定节点 v 在不同时间步长的一系列表示。采用文献 [18] 所提出的位置嵌入方法,得到节点各子图中绝对时序位置的嵌入序列 $\{p^1, p^2, \dots, p^N\}, p^N \in \mathbb{R}^F$ 然后结合结构注意力层的节点 v 在不同的时间步长下的向量表示得到该层的输入 $\{h_v^1 + p^1, h_v^2 + p^2, \dots, h_v^N + p^N\}$, N 是总时间步长。该层输出是 $\{\tilde{h}_v^1, \tilde{h}_v^2, \dots, \tilde{h}_v^N\}, \tilde{h}_v^n \in \mathbb{R}^{F'}$ 。节点的输入及输出分别结合时间步长合并表示为 $X_v \in \mathbb{R}^{N \times F}$ 和 $\tilde{H}_v \in \mathbb{R}^{N \times F'}$ 。节点 v 在时间步长 j 的输入表示为 x_v^j , 且用 x_v^j 作为该节点时间步长小于 j 的历史状态的查询。和结构自注意力机制在其邻居节点进行特征学习表示不同,时序自注意力机制更关注每个节点的历史状态。

通过对注意力按比例缩放的点积的形式来计算节点 v 在时间步长 j 的输出表示,其中查询、键和值的集合是节点的输入表示。查询、键和值首先分别通过线性投影矩阵 $W_q \in \mathbb{R}^{F \times F'}, W_k \in \mathbb{R}^{F \times F'}$ 和 $W_v \in \mathbb{R}^{F \times F'}$ 转换到不同空间。在计算时考虑所有小于 j 的时间步长,直到时间步长到达 j , 来保持其自回归性。时序自注意力函数定义为:

$$\tilde{H}_v = \beta_v (X_v W_v) \quad \beta_v^{ij} = \frac{e^{e_v^{ij}}}{\sum_{k=1}^N e^{e_v^{kj}}} \quad (2)$$

$$e_v^{ij} = \left(\frac{((X_v W_q)(X_v W_k)^T)_{ij}}{\sqrt{F'}} + S_{ij} \right)$$

式中, $\beta_v \in \mathbb{R}^{N \times N}$ 是通过乘法注意力函数得到的注意力权重矩阵; $S \in \mathbb{R}^{N \times N}$ 是其每个子项 $S_{ij} \in \{-\infty, 0\}$ 拥有强自回归属性的掩码矩阵。为了对时序进行编码,将 S 定义为:

$$S_{ij} = \begin{cases} 0 & i \leq j \\ -\infty & \text{其他} \end{cases}$$

当 $S_{ij} = -\infty$ 时,通过 softmax 函数所得到的注意力权重为 0,例如当 $\beta_v^{ij} = 0$ 时,时间步长 i 到 j 不

使用注意力机制。

2.2.4 级联增量预测层

在级联增量预测层,通过将 \tilde{h}_v^N 输入到多层感知器 (multilayer perceptron, MLP) 中,计算得到最终级联增量的预测值 P_i :

$$P_i = f(\tilde{h}_v^N) = \text{MLP}(\tilde{h}_v^N) \quad (3)$$

2.3 模型训练及算法描述

2.3.1 模型训练

该模型的最终目标是对级联增量大小进行预测,采用式 (10) 作为损失函数来对模型进行优化,在迭代中损失连续 10 次不再减小时结束训练:

$$L(P_i, \hat{P}_i) = \frac{1}{M} \sum_{i=1}^M (\log_2 P_i - \log_2 \hat{P}_i)^2 \quad (4)$$

式中, P 是所有消息的总量; P_i 是消息 C_i 的预测级联增量; \hat{P}_i 是实际增量。

2.3.2 算法描述

给定信息级联图 G_i , 经基于不同时间步的子图算法采样后,得到原始信息级联的子图快照集合 $\zeta_i(t_j)$ 并得到其对应的序列表示 G_i^T 。将得到的序列表示输入到结构自注意力机制层,对每个节点及其邻居结构进行学习,得到 h_v^j 。然后将 h_v^j 与 p^j 相加输入到时序自注意力机制层得到 \tilde{h}_v^N 。最后将各节点融合了动态演化的结构特征和时序特征的向量表示依次通过多层感知机,并输出最终流行度增量预测值 P_i 。

整个算法流程与图 3 模型架构图一致,主要分为节点嵌入、结构自注意力、时序自注意力和级联增量预测 4 部分。

算法: DySatCas

输入: 级联图 G_i 及其演化序列 G_i^T 的集合、级联图数量 M 、时间步长 N

输出: 级联增量预测值 P_i 的集合

模型初始化, 参数设置

Repeat

For $1 \leq i \leq M$ Do

For $j \in [1, N]$

Obtain $\zeta_i(t_j)$ by sampling G_i at time step t_j

/*节点嵌入*/

Compute h_v^j by Eq.(1) for $v \in V_i^j$ /*结构自注

意力*/

Compute $h_v^j + p^j$


```

    Compute  $\tilde{H}_v$  by Eq.(2) /*时序自注意力*/
  End For
  Obtain  $\tilde{h}_v^N$ 
   $P_i = \text{MLP}(\tilde{h}_v^N)$  /*增量预测*/
  Compute  $L(P_i, \hat{P}_i)$  by Eq.(4)
  Update Parameters
End For
Until convergence
Return a set of  $P_i$ 

```

3 实验与分析

3.1 实验数据集

本文主要基于两个实际场景的公开数据集来对所提出模型的有效性进行评估, 分别是预测新浪微博的转发量和在数据集 APS 上预测论文的引用量, 如表 1 所示。

Sina Weibo: 每一个样本对应一条微博消息的转发路径, 其中包含转发用户的 id 以及转发的时间戳。

APS(American physical society): 该数据集包含了美国物理学会的期刊从 1893–2009 年所发表的论文及其引用, 每一篇论文及其引文都形成了一个引用级联。

表 1 实验数据集的详细统计

Dataset	Sina Weibo	APS
Number of Cascades	119313	207685
Number of Nodes	6738040	616316
Number of Edges	455412321	247319593
Avg.Popularity	240	51
Avg.Observed Popularity	54	19
Avg.Sequence Length	2.237	3.999

3.2 超参数设定

时间自注意力层单元数、结构自注意力层单元数和嵌入层单元数为 128, 多层感知层单元数为 128×64 , Adam 优化器的初始学习率为 0.0005, 迭代轮数 2000 轮, 时间步长度 N 为 12。

3.3 评价指标

和文献 [17] 相同, 使用均方对数误差 MSLE 和平均绝对百分比误差 MAPE 来对所提出的模型进行评估, 具体评估公式为:

$$\text{MSLE} = \frac{1}{M} \sum_{i=0}^{M-1} (\log_2 \hat{P}_i - \log_2 P_i)^2 \quad (5)$$

$$\text{MAPE} = \frac{1}{M} \sum_{i=0}^{M-1} \frac{|\log_2 \hat{P}_i - \log_2 P_i|}{\log_2 \hat{P}_i} \quad (6)$$

3.4 对比分析

3.4.1 基线模型

1) **Feature-Linear:** 基于特征的信息级联预测模型是应用最广泛的。本文基线模型提取如消息发送者/转发者的特征、结构和时间特征, 将此类特征输入到线性回归模型和 MLP 模型中对信息级联增量进行预测。

2) **DeepCas^[11]:** 将深度学习应用于信息级联预测的模型, 级联图采样方式为随机游走, 并采用双向 GRU 和注意力机制来预测端到端的信息级联增量。

3) **Topo-LSTM^[19]:** 使用 LSTM 来对图中节点间关系进行建模。根据上一个节点状态, 得到下一个节点的预测结果。

4) **DeepHawkes^[9]:** 结合了深度学习和自激励过程来进行信息级联预测。

3.4.2 模型评价指标对比

为了检验 DySatCas 模型的效果, 分别从现在流行的 3 类级联预测方法中, 选择其中具有代表性的模型来进行比较, 具体结果如表 2 所示。通过对比有以下发现:

1) **DySatCas** 的预测效果优于基线。在微博数据集上, 观测时间为 0.5 h, MSLE 和 MAPE 的结果分别比第一最佳基线 (DeepHawkes) 高出 2.5% 和 3.4%; 观测时间为 1 h, MSLE 和 MAPE 均比其他基线效果好。在数据集 APS 上, 观测时间为 5 年, MSLE 和 MAPE 的结果比第一最佳基线分别高出 10% 和 4%, 除观测时间 3 年, MAPE 略低于 DeepHawkes 外, 其他数据均优于基线, 且随着观测时间增加, 预测误差逐渐减少, 这表明了 DySatCas 模型的对级联图的动态演化过程学习的有效性。

2) 从实验结果不难看出, 在两个数据集中, 基于特征的基线模型的预测效果并不比基于深度学习的基线模型差。在微博数据集中, 基于特征的基线模型预测准确性甚至高于传统的基于深度学习的基线模型 (DeepCas、Topo-LSTM)。但该方法极大地依赖于人工提取特征, 存在很大的不稳定性 and 不确定性。并且通过人工的方法, 很难提取到较复杂信息中所隐含的特征, 适用场景相对固定, 推广性差。

表 2 两种数据集下使用 MSLE 和 MAPE 两种评价指标对不同模型及变体的评估表现

模型	Sina Weibo				APS			
	0.5 h		1 h		3years		5years	
	MSLE	MAPE	MSLE	MAPE	MSLE	MAPE	MSLE	MAPE
Feature-Linear	2.892	0.377	2.805	0.378	2.178	0.282	2.141	0.286
DeepCas	3.073	0.387	2.957	0.396	2.194	0.301	1.987	0.298
Topo-LSTM	2.923	0.372	2.741	0.387	2.043	0.275	1.926	0.285
DeepHawkes	2.587	0.323	2.475	0.303	1.942	0.267	1.873	0.282
DySatCas	2.523	0.312	2.467	0.298	1.789	0.273	1.675	0.270
DySatCas-Temporal	2.954	0.367	2.876	0.381	2.042	0.350	1.973	0.348
DySatCas-Structural	2.791	0.348	2.589	0.379	1.872	0.312	1.798	0.327

3) DeepCas 是基于图表示学习的级联预测模型。其主要使用随机游走的方法对级联图进行采样,但没有考虑信息的实际传播顺序,同时忽略了传播过程中的时序依赖性,使得预测效果较其他深度学习方法较差。这说明了在信息级联预测中,级联图的结构特征和时序特征的重要性。

4) Topo-LSTM 是基于 RNN 的级联预测模型,采用扩散拓扑结构,但没有考虑时间因素以及不确定的级联效应,使得预测效果与 DeepHawkes 及 DySatCas 有一定的差距。

5) DeepHawkes 模型是一种结合了生成式方法和深度学习的模型。在级联预测过程中,不考虑级联的拓扑信息,依赖于时间序列进行建模,主要关注点在于自我激励机制对级联增量的影响。虽然相比 DeepCas 和 Topo-LSTM 预测准确性更高,但因为忽略了级联的结构特征,且忽略了级联的动态性,使得预测效果与 DySatCas 相比不够理想。从而说明了在级联预测中,级联图的结构特征、时序特征及其传播过程中的动态过程都是影响预测效果的关键因素。

3.5 模型分析

3.5.1 模型变体

为了更好地了解 DySatCas 模型各组成部分对于模型整体预测效果的影响,引入两种变体进行对比。

1) DySatCas-Temporal: 在该变体中,忽略时序注意力层,直接将通过结构自注意力层的向量表示输入到多层感知器进行预测。

2) DySatCas-Structural: 在该变体中,忽略结构注意力层,直接将通过子图采样后的向量编码输入到时序自注意力层,然后将时序自注意力机层的表示输入到多层感知器中。

从表 2 中可以看出:与 DySatCas 相比,在观

测时间为 0.5 h 时,移除结构自注意力层的 DySatCas-Structural 和移除时序自注意力层的 DySatCas-Temporal 预测误差分别增加了 10% 和 17%,其中移除了时序自注意力层的变体误差更大,说明了级联预测过程中,时序特征对预测效果有更大的影响力,同时也证明了本文模型各组成部分的有效性。

3.5.2 时间步长

为了验证时间步长在动态图演化过程中的影响,在微博数据集上,分别将时间步长设置为 2、3、4、5、6、7、8、9、10、11、12,得到预测结果如图 4 所示。由图可知,随着时间步长越长级联预测效果越好,后期误差出现变化不大的原因主要是时间步划分过细,导致后期级联图变化不大,所学习到的节点表示相近。

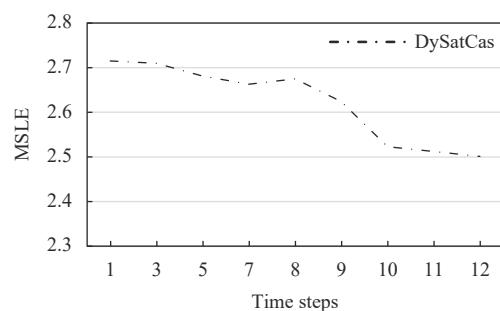


图 4 不同步长性能预测比较

3.5.3 观测窗口

观测时间窗口也是 DySatCas 模型中最重要的参数,如图 5 所示,在微博数据集和 APS 数据集上,分别把观测时间窗口设置为 0.5、1、2、3、4 h 和 3、4、5、6、7、8 年,由实验结果可知观察时间窗口越长, DySatCas 的训练数据越多,从而就能做出更准确的预测。

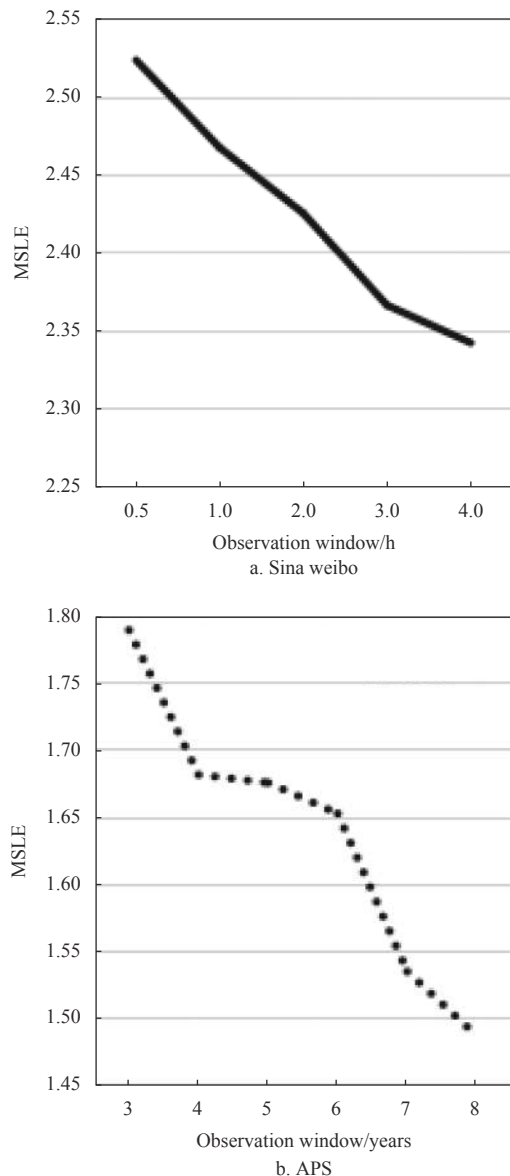


图 5 不同观测窗口在不同数据集上的表现

4 结束语

本文融合动态图学习和自注意力机制, 提出了级联预测模型 DySatCas。该模型不严重依赖人工标记特征, 充分考虑级联传播过程中的动态性和不确定性, 根据级联图的结构特征和时序特征来进行级联预测。更好地帮助人们了解信息传播机制和规律, 也为后续事件演化研究提供有力的技术支撑。未来考虑结合多任务、多视图等概念, 进一步对信息传播特点进行研究。

参 考 文 献

- [1] CHEN X, ZHOU F, ZHANG K, et al. Information diffusion prediction via recurrent cascades convolution[C]//2019 IEEE 35th International Conference on Data Engineering (ICDE). Macao, China: IEEE, 2019: 770-781.
- [2] ZHANG Z, ZHANG F, TAN Q, et al. review of information cascade prediction methods based on deep learning[J]. Computer Science, 2020, 47(7): 141-153.
- [3] ZHOU F, XU X, TRAJCEVSKI G, et al. A survey of information cascade analysis: Models, predictions, and recent advances[J]. ACM Computing Surveys (CSUR), 2021, 54(2): 1-36.
- [4] ZHOU F, JING X, XU X, et al. Continual information cascade learning[C]//GLOBECOM 2020-2020 IEEE Global Communications Conference. [S.l.]: IEEE, 2020: 1-6.
- [5] ZHANG Q, GONG Y, WU J, et al. Retweet prediction with attention-based deep neural network[C]//Proceedings of the 25th ACM international on Conference on Information and Knowledge Management. [S.l.]: ACM, 2016: 75-84.
- [6] GAO S, MA J, CHEN Z. Effective and effortless features for popularity prediction in microblogging network[C]// Proceedings of the 23rd International Conference on World Wide Web. Seoul: ACM, 2014: 269-270.
- [7] LEE J G, MOON S, SALAMATIAN K. Modeling and predicting the popularity of online contents with Cox proportional hazard regression model[J]. Neurocomputing, 2012, 76(1): 134-145.
- [8] SHEN H, WANG D, SONG C, et al. Modeling and predicting popularity dynamics via reinforced poisson processes[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2014, 28(1): 291-297.
- [9] CAO Q, SHEN H, CEN K, et al. Deephawk: Bridging the gap between prediction and understanding of information cascades[C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. Singapore: ACM, 2017: 1149-1158.
- [10] FENG X, ZHAO Q, LIU Z. Prediction of information cascades via content and structure proximity preserved graph level embedding[J]. Information Sciences, 2021, 560: 424-440.
- [11] LI C, MA J, GUO X, et al. Deepcas: An end-to-end predictor of information cascades[C]//Proceedings of the 26th international conference on World Wide Web. New York: ACM, 2017: 577-586.
- [12] PEROZZI B, AL-RFOU R, SKIENA S. Deepwalk: Online learning of social representations[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. [S.l.]: ACM, 2014: 701-710.
- [13] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). [S.l.]: Association for Computational Linguistics, 2014: 1532-1543.
- [14] SANKAR A, WU Y, GOU L, et al. Dysat: Deep neural representation learning on dynamic graphs via self-attention networks[C]//Proceedings of the 13th ACM

- International Conference on Web Search and Data Mining (WSDM). Houston: ACM, 2020: 519-527.
- [15] XUE H, YANG L, JIANG W, et al. Modeling dynamic heterogeneous network for link prediction using hierarchical attention with temporal RNN[EB/OL]. [2020-11-24]. <https://arxiv.org/abs/2004.01024>.
- [16] GOYAL P, CHHETRI S R, CANEDO A M. Capturing network dynamics using dynamic graph representation learning: U. S. Patent Application 16/550, 771[P]. 2020-3-5.
- [17] ZHOU F, XU X, ZHANG K, et al. Variational information diffusion for probabilistic cascades prediction[C]//IEEE INFOCOM 2020-IEEE Conference on Computer Communications. Toronto: IEEE, 2020: 1618-1627.
- [18] GEHRING J, AULI M, GRANGIER D, et al. Convolutional sequence to sequence learning[C]//International Conference on Machine Learning. [S.l.]: PMLR, 2017: 1243-1252.
- [19] WANG J, ZHENG V W, LIU Z, et al. Topological recurrent neural network for diffusion prediction[C]//2017 IEEE International Conference on Data Mining (ICDM). New Orleans: IEEE, 2017: 475-484.

编辑 叶芳