

# 基于深度学习的信息级联预测方法综述



张志扬 张凤荔 谭琪 王瑞锦

电子科技大学信息与软件工程学院 成都 610054

(13980044734@163.com)

**摘要** 在线社交媒体极大地促进了信息的产生和传递,加速了海量信息之间的传播与交互,使预测信息级联的重要性逐渐突显。近年来,深度学习已经被广泛用于信息级联预测(Information Cascade Prediction)领域。文中主要对基于深度学习的信息级联预测方法的研究现状与经典算法进行分类、梳理与总结。根据信息级联特征刻画的侧重点不同,将基于深度学习的信息级联预测方法分为时序信息级联预测方法与拓扑信息级联预测方法,并进一步将时序信息级联预测方法分为基于随机游走(Random Walk)的方法与基于扩散路径的方法,将拓扑信息级联预测方法分为基于全局拓扑结构的方法与基于邻域聚合的方法;并对每类方法进行详细的原理阐述与优缺点介绍,介绍了信息级联预测领域常用的数据集与评价指标,在宏观与微观两种信息级联预测场景下对基于深度学习的信息级联预测算法进行实验对比,并讨论了一些信息级联预测算法中常用的算法实现细节。最后,总结了该领域未来可能的研究方向与发展趋势。

**关键词:** 在线社交媒体;深度学习;信息级联预测;级联增量预测;流行度预测;节点预测

**中图法分类号** TP183

## Review of Information Cascade Prediction Methods Based on Deep Learning

ZHANG Zhi-yang, ZHANG Feng-li, TAN Qi and WANG Rui-jin

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

**Abstract** Online social media greatly promotes the generation and transmission of information, exacerbates the communication and interaction between massive amounts of information, and highlights the importance of predicting information cascades. In recent years, deep learning has been widely used in the field of information cascade prediction. This paper mainly classifies, sorts, and summarizes the current research status of deep learning-based information cascade prediction methods and classic algorithms. According to the different emphasis of information cascade feature characterization, the information cascade prediction method based on deep learning is divided into time series information cascade prediction method and topology information cascade prediction method. The time series information cascade prediction method is further divided into methods based on random walks and methods based on diffusion paths, and the topology information cascade prediction method is divided into methods based on global topological structure and methods based on neighborhood aggregation. This paper details the principles and advantages and disadvantages of each type of method, and introduces the data sets and evaluation indicators commonly used in the field of information cascade prediction, and compares the information cascade prediction algorithms based on deep learning in the macro and micro information cascade prediction scenarios, and discusses some technical details commonly used in information cascade prediction algorithms. Finally, this paper summarizes the field possible future research directions and development trends.

**Keywords** Online social media, Deep learning, Information cascade prediction, Cascade size prediction, Popularity prediction, Node prediction

## 1 引言

当前,在 Twitter, Instagram 和微博等在线社交平台的驱

动下,信息发布与过滤的方式已发生了极大的变化,在这些社交媒体中,信息不仅可以通过人群之间的社交关系进行共享与传播,还会在此过程中衍生出一系列新的共享关系与模式。

收到日期:2020-03-23 返修日期:2020-05-23 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61802033, 61472064, 61602096);四川省科技计划(2018GZ0087, 2019YJ0543);四川省区域创新合作项目(2020YFQ0018);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)

This work was supported by the National Natural Science Foundation of China(61802033, 61472064, 61602096), Sichuan Science and Technology Program(2018GZ0087, 2019YJ0543), Sichuan Regional Innovation Cooperation Project(2020YFQ0018), Chinese Postdoctoral Science Foundation(2018M643453), Guangdong Provincial Key Laboratory Project(2017B030314131) and Network and Data Security Key Laboratory of Sichuan Province Open Issue(NDSMS201606).

通信作者:张凤荔(fzhang@uestc.edu.cn)

在促进信息快速传播的同时,要保障信息传播的安全,就需要研究新信息级联传播模型下的级联动力学(Cascade Dynamics Modeling),它不但能对信息级联的传播路径进行预测,而且有助于了解级联信息传播的基本机理,并衍生出信息级联预测的各种下游应用,例如虚假新闻检测和病毒式营销识别。从应用点的角度来看,现有关于信息级联建模的研究主要集中在两个方面:1)宏观任务侧重于估算级联增长和预测爆发,主要包括级联增量预测以及信息流行度预测,这是关于信息级联粗略的估计,因此不适合一些细微的预测任务;2)微观任务研究局部的社会影响模式,因此这些任务更关注用户级别的建模,常见应用包括推断用户的行为状态或预测感染的下一系列节点。

最近,深度学习在不同领域的最新应用催生了基于深度学习的级联预测(Information Cascade Prediction Based on Deep Learning, ICPbDL)的相关研究,早期的研究主要集中在利用顺序网络框架对信息级联的时序结构进行建模,并辅以拓扑图采样方法来捕捉信息级联的局部拓扑结构。然而,这类方法在建模时忽略了信息级联的完整拓扑结构,导致其在复杂网络中的预测效果并不理想。后来,研究关注点集中于利用深度学习框架对信息级联的拓扑结构进行建模,以学习级联拓扑结构中的深层特征表示。本文分析总结了相关研究,将 ICPbDL 分为时序信息级联预测方法(ICP Based on Timing Structure, ICPbTS)与拓扑信息级联预测方法(ICP Based on Topology, ICPbTO)两类,并详细介绍与比较了这两类预测方法所包含的典型算法。如图 1 所示,ICPbTS 根据级联采样方法的不同可分为基于随机游走的方法与基于扩散路径的方法;ICPbTO 根据拓扑结构的学习方式不同可分为基于全局拓扑结构的方法与基于邻域聚合的方法。

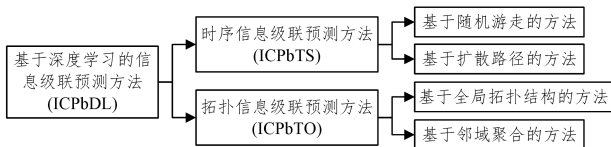


图 1 基于深度学习的级联预测方法分类

Fig. 1 Classification of cascaded prediction methods based on deep learning

本文第 2 节介绍信息级联预测的背景与问题定义,并简要介绍传统信息级联预测算法的原理与优缺点;第 3 节介绍 ICPbTS 的相关算法;第 4 节介绍 ICPbTO 的相关算法;第 5 节讨论一些信息级联预测领域常用的算法实现细节;第 6 节对信息级联预测常用的公开数据集以及常用评价指标进行总结,并对一些典型算法进行分析对比;最后总结全文。

## 2 相关背景与传统算法介绍

### 2.1 相关背景及概念定义

信息级联预测问题源于对博客<sup>[1-3]</sup>、电子邮件<sup>[4-5]</sup>、产品推荐<sup>[6]</sup>以及社交网站(例如 Facebook 和 Twitter<sup>[7-8]</sup>)中识别到的用户级联扩散序列与网络关系图进行分析。根据信息级联预测方法所预测的目标不同,可将其分为宏观信息级联预测与微观信息级联预测。宏观信息级联预测主要预测一段时间后的级联增量,以此衍生出流行度预测与病毒级联分类等下游应用。微观信息级联预测更侧重于用户节点级别的建模,

通常应用于信息级联中后续网络节点的传播行为预测<sup>[9-10]</sup>。

早期有较多的研究致力于表征信息级联的结构特性和内容信息<sup>[11-12]</sup>,但这些研究进行的级联预测均在级联结构信息采集方面表现不足<sup>[13-14]</sup>。深度学习模型有望自动学习信息级联的特征表示,这种表示在神经网络训练过程中将会从具体到抽象,较低级别的表示形式将根据较高级别的反馈进行更新,因此深度神经网络特别擅长学习将这些表示映射到预测结果的非线性函数。深度学习的这种特性使其在社交网络分析中的节点推荐与序列检测等领域得到了广泛的应用,其中循环神经网络(Recurrent Neural Network, RNN)及其变种已经成为了绝大多数复杂网络分析与预测的核心算法,它将复杂网络的建模视为序列建模问题,从而避免了扩散模型和特征设计所强加的先验知识,同时灵活地捕获了信息级联序列的顺序依赖性。基于循环神经网络的 ICPbTS 方法虽然有上述优点,但其在捕捉网络的拓扑结构信息方面有明显不足,这类方法只能依靠随机游走等采样方法来将信息级联的局部拓扑结构传入模型,导致其对信息级联拓扑结构的表示缺乏完整性。后来,研究关注点集中于利用深度神经网络捕捉信息级联拓扑结构的 ICPbTO 方法。

基于信息级联拓扑结构的 ICPbTO 不再只将级联图用于采样,而是直接将级联图作为模型的输入,利用深度学习框架来学习其深层特征表示。相比 ICPbTS, ICPbTO 的优势是在复杂的社交网络上建模信息级联中的用户影响力与扩散状态,提升了模型的可解释性。ICPbTO 可通过两种形式实现:1)将全局拓扑结构作为节点嵌入信息,随节点特征一起传入循环神经网络框架进行训练;2)基于邻域聚合的原理,使用图神经网络递归聚合级联图中邻域节点的特征向量,来学习节点的拓扑结构信息。一些 ICPbTO 方法将图神经网络与循环神经网络结合,因此这些 ICPbTO 方法也可以被视为 ICPbTS 的改进策略。

本文统一使用如下符号定义级联与级联图:给定  $\mathcal{M}$  条信息,表示为  $\mathcal{M} = \{m^i\} (1 \leq i \leq M)$ 。对于每条信息  $m^i$ ,使用级联  $C = \{(u_j^i, v_j^i, t_j^i)\}$  记录信息  $m^i$  的扩散过程。在不同的数据集中,信息的扩散方式不同,本文将扩散方式统一称为转发,元组  $(u_j^i, v_j^i, t_j^i)$  对应于第  $j$  次转发,这意味着节点  $v_j^i$  转发了节点  $u_j^i$  的信息  $m^i$ ,信息的发布时间与第  $j$  次转发之间的间隔时间为  $t_j^i$ 。级联在扩散网络中的拓扑结构表示称为级联图,一张级联图可以对应一条或者若干条信息扩散路径,本文将级联图定义为  $G = (V, E)$ ,其中  $V$  是节点集,  $E \in V \times V$  是边集。在一些数据集中,级联图的拓扑结构信息需要人工提取,例如 Sina Weibo 数据集中一个信息级联样本可以对应提取一张拓扑图表示(即邻接矩阵);也有数据集直接给出级联图的拓扑结构,例如 Twitter 数据集会给出由数据集内所有节点与扩散路径组成的拓扑图结构,这类拓扑图一般被称为全球网络,它包含了整个数据集中所有信息级联序列的拓扑结构关系。

本文将常见的信息级联预测任务定义为如下 3 类。

(1)级联增量预测。给定级联  $C$  在  $[0, T)$  时间段(一般称为观察窗口)内的转发路径  $\{(u_j, v_j, t_j) \mid t_j \leq T\}$ ,则其对应的级联图为  $G^T = (V^T, E^T)$ 。级联增量预测指在给定的时间间隔  $\Delta t$  之后预测级联  $C$  节点增加数的回归问题,记为  $\Delta s_c = |V_{c, \Delta t}^+| - |V_c^-|$ 。级联增量预测的一个特例是流行度预测,它预测观察到的节点数和级联的最终节点数之间的

增量  $\Delta s_c^c = |V_c^c| - |V_c^T|$ 。

(2)病毒级联预测。病毒级联预测与级联增量预测类似,但病毒级联预测是分类问题。给定阈值  $Thr$ ,对于级联  $C$ ,如果它的  $\Delta s_c \geq Thr$ ,则将其定义为病毒样本,样本标签设为 1;否则,将其定义为非病毒样本,标签设为 0。病毒级联预测的目的是将级联样本归类为病毒样本或非病毒样本。

(3)节点预测。给定级联  $C$  在  $[0, T)$  时间段(一般称为观察窗口)内的转发路径  $\{(u_j, v_j, t_j) | t_j \leq T\}$ ,节点预测的任务是预测将被感染的用户  $u_{j+1} \in V$ ,因此节点预测的目标是建立一个能够学习条件概率  $p(u_{j+1} | \{(u_1, t_1), \dots, (u_j, t_j)\})$  的模型。

### 2.2 传统信息级联预测方法

传统的信息级联预测方法主要分为以下 3 类。

#### (1)基于特征建模的方法<sup>[15]</sup>

基于特征的方法从原始数据中提取各种手工特征,通常包括信息内容特征、用户属性特征、级联的结构和时间特征,然后将它们输入到机器学习算法中进行预测<sup>[16-20]</sup>。基于特征的方法需要结合多种人工特征,且所有这些特征的设计在很大程度上取决于特定的领域知识,使得这种方法难以得到推广。

#### (2)基于生成建模的方法<sup>[21]</sup>

基于生成建模的方法通常将在线社交媒体的级联增量或流行度视为转发的到达过程,并针对每个消息分别对到达过程的强度函数进行建模<sup>[22-29]</sup>。这些方法通常不会直接对未来的流行度进行优化,也不会独立学习每个消息的参数。因此,这些方法无法充分利用信息级联中隐含的动态流行度信息,预测模型的可解释性较差。

#### (3)基于扩散模型的方法

基于扩散模型的方法通常做出一些信息在扩散过程中的理论假设,并简化了现实中的复杂网络关系。经典的扩散模型包括独立的级联模型和线性阈值模型<sup>[30-32]</sup>。这些方法在表征信息的扩散过程中获得了一定的成功,但它们在很大程度上取决于基础的扩散模型,在复杂的信息级联网络中表现较差。

### 3 时序信息级联预测方法

ICPbTS 将传统的信息级联序列处理算法与循环神经网络结合,放弃使用级联图  $G$  作为神经网络模型的输入,仅将其用于数据预处理的过程中,通过构建基于循环神经网络的信息级联扩散模型来重点捕捉信息级联的时序结构,具体的模型框架如图 2 所示。ICPbTS 简化了信息级联预测模型的输入,并在学习信息级联的时序结构的同时,利用注意力机制或基于拓扑图的采样方法来建模信息级联扩散过程中的空间结构特征以及信息级联在扩散过程中的时间衰减效应。

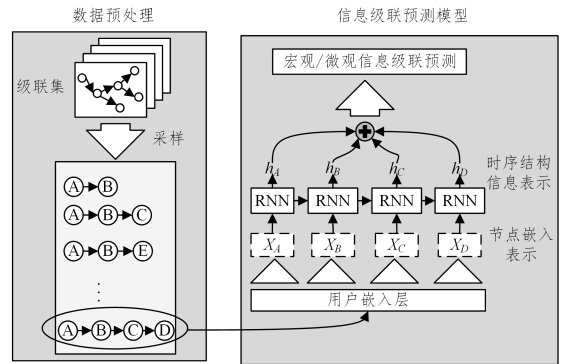


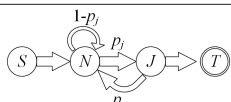
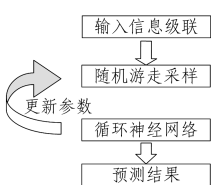
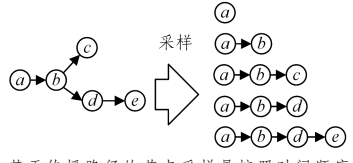
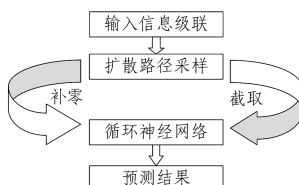
图 2 时序信息级联预测模型

Fig. 2 Time series information cascade prediction model

复杂的信息级联往往包含大量节点、边以及复杂的分支结构,直接考虑完整信息级联的代价将十分高昂,如何利用高效率的采样方法表示原始信息级联是信息级联研究的重心之一。为了弥补循环神经网络模型在处理级联拓扑结构方面的不足,ICPbTS 的采样方法还需同时兼顾信息级联的空间结构信息。我们根据模型采用的不同采样方法将 ICPbTS 分为两类:基于随机游走的方法与基于扩散路径的方法。表 1 列出了两种方法的代表算法、采样原理、算法流程与各自的优缺点。

表 1 时序信息级联预测方法的对比

Table 1 Comparison of ICPbTS methods

方法类别	代表算法	采样原理	算法流程	优/缺点
基于随机游走的方法	DeepCas <sup>[33]</sup> , MPR-SNE <sup>[34]</sup>	 <p>每个扩散图的随机游动始于起始状态 <math>S</math>, 终于状态 <math>T</math>. <math>N</math> 和 <math>J</math> 分别为扩散和跳跃状态. <math>p_j</math> 表示执行随即跳转的概率, 其大小本质上决定了采样的信息级联序列的长度; <math>p_o</math> 表示扩散到邻居节点的概率, 其本质上决定了采样的信息级联序列的预期数量. 随机游走采样完成的等长子序列将传入 RNN 进行学习, 同时更新随机游走的采样概率参数。</p>		<p>优点： 随机游走能够适应图的局部变化, 在图结构的演化过程中, 不需要每次都对整个图的随机游走重新计算, 计算效率高。</p> <p>缺点： 采样序列不是按照信息传播的真实过程进行采样的, 难以捕捉信息级联中的拓扑结构信息。</p>
基于扩散路径的方法	DeepHawkes <sup>[35]</sup> , LDA-DeepHawkes <sup>[36]</sup> , DeepDiffuse <sup>[37]</sup>	 <p>基于传播路径的节点采样是按照时间顺序的, 用一组传播路径表示信息级联, 每条路径表示级联中一个真实的传播过程分支. 传播路径采样完成的子序列通常不等长, 可以对较短的序列进行补零后直接传入 RNN 进行学习, 也可以按固定的窗口长度进行逐步截取后再传入 RNN。</p>		<p>优点： 可以捕捉原始信息级联中所蕴含的局部结构信息, 能够有效表示信息传播过程中的扩散动态。</p> <p>缺点： 采样灵活性较差, 采样效率较低, 仅能捕捉局部的空间结构信息, 无法捕捉全局的结构信息, 存在一定的信息损失。</p>

### 3.1 基于随机游走的方法

随机游走是图论中的重要算法之一,在自然语言处理(NLP)、节点分类与链接预测等领域具有广泛应用。随机游走首先给定当前访问的起始节点,从其邻居节点中随机选取节点作为下一个访问节点,重复此过程,直到序列长度满足预设条件。因此,随机游走是一种可重复访问已访问节点的深度优先遍历(Depth First Search, DFS)算法。在信息级联预测领域,文献[38]首先将网络中的节点与自然语言中的单词进行类比,并使用固定长度的随机行走路径来模拟信息级联扩散过程中节点的“扩散上下文”,从而可以使用与自然语言处理中相同的方法来学习节点的特征表示。基于随机游走的信息级联采样过程可以概括为将信息级联序列  $C$  中的节点与转发路径描述为拓扑图  $G$ ,然后在  $G$  上执行随机游走过程,且满足  $C$  中的时序关系。假设现在随机游走采样至节点  $v$ ,并处于马尔可夫链中的状态  $N$ ,则它扩散至  $v$  的邻居节点  $u \in N_c(v)$  的转移概率为:

$$p(u \in N_c(v) | v) = \frac{sc_i(u) + \alpha}{\sum_{s \in N_c(v)} (sc_i(s) + \alpha)} \quad (1)$$

其中,  $N_c(v)$  表示节点  $v$  在状态  $N$  时的邻居节点集合;  $sc_i(u)$  是节点  $u$  的转移评分函数,一般用节点  $u$  在级联图上的度或  $u$  与  $v$  之间的权重来计算;  $\alpha$  是平滑参数。类似地,在级联图的节点集  $V_c$  上可以使用节点  $u$  的跳跃评分函数  $sc_j(u)$  来定义其跳跃概率:

$$p(u) = \frac{sc_j(u) + \alpha}{\sum_{s \in V_c} (sc_j(s) + \alpha)} \quad (2)$$

随机游走中节点的跳跃与转移概率决定了采样的执行结果。具体来说,是否执行节点在拓扑图上向邻居节点转移或直接进入终止状态的概率  $p_o$  本质上决定了采样序列的数量,而是否执行节点的随机跳跃的概率  $p_j$  决定了采样序列的长度,这两个概率在确定级联图的随机游走采样表示中起着关键作用。传统的随机游走采样中的节点转移概率是人为设置的超参数,这导致传统的随机游走无法适应不同的信息级联序列之间普遍存在的差异性。针对这一问题,文献[33]通过将随机游走采样纳入深度学习框架中以自适应的方式学习这两个概率,使用注意力机制来学习如何将采样得到的若干子序列组合为原始的级联图,在这一过程中,信息级联的局部空间结构信息也被纳入信息级联的整体表示中。

在神经网络与随机游走采样的基础上,文献[34]提出了一种保留多路径关系的社交网络嵌入方法 MPR-SNE,以学习社交网络中的多路径或多关系网络结构和节点属性(例如性别、职业等节点标签)。具体而言,网络结构形成节点之间社交关系路径的各种长度,一些节点信息使这些路径具有丰富的语义。然而,节点属性信息通常是稀疏且不完整的,甚至与网络结构不一致,为了更好地学习节点表示,MPR-SNE 首先利用随机游走探索节点之间的多个社交关系路径,然后利用双向多路径关系单元(Bidirectional Multipath Relationship Unit, BMRU)来捕获多路径关系的性质。其中, BMRU 可以理解成一种结合了节点的属性信息的双向循环神经

网络单元。为了保留节点的属性信息,该文将节点属性的学习建模为监督学习问题,使用 BMRU 的输出来预测当前节点的属性,使用加权交叉熵来衡量预测概率误差。最后,该文优化循环神经网络与节点属性监督学习的联合目标函数,以获得节点的嵌入表示。

### 3.2 基于扩散路径的方法

与基于随机游走的方法不同,基于扩散路径的方法用一组真实的传播路径表示信息级联。对于一个信息级联而言,基于传播路径采样后得到的序列数量与该级联所包含的节点数相等,最大序列长度与级联中最长的传播路径的长度相等。对于整个样本数据集而言,最大序列长度与样本中最长的传播路径的长度相等。扩散路径采样可以在遵守信息级联中节点的时序关系的同时,完全还原级联图的局部拓扑结构,因此其结构捕捉的完整性高于随机游走采样。2017年, Cao 等<sup>[35]</sup>针对常见的生成建模方法解释性不足的缺点,对 ICPbTS 的基础模型进行了改进,提出了基于霍克斯过程<sup>[29]</sup>的循环神经网络模型,为信息级联预测模型加入了霍克斯过程的 3 个可解释因素,即在端到端监督框架下的用户影响力、自激机制和时间衰减效应。霍克斯过程的关键是对新转发行为的发生概率进行建模,可以将其写为:

$$\rho_t^i = \sum_{j: t_j^i \leq t} \mu_j^i \phi(t - t_j^i) \quad (3)$$

其中,  $\rho$  是消息  $m^i$  在时间  $t$  的新转发发生概率,  $t_j^i$  是起始转发到第  $j$  个转发之间经过的时间,  $\mu_j^i$  是第  $j$  个转发直接影响的潜在用户数,  $\phi(t)$  是时间衰减函数。文献[35]直接通过用户的 id 来学习用户的自身特征表示,将转发路径输入循环神经网络,并汇总所有子路径的最后一个隐藏状态表示,以模拟每个转发的自激机制,最后使用非参数的方式学习时间衰减效应。基于霍克斯过程的信息级联预测模型框架如图 3 所示。

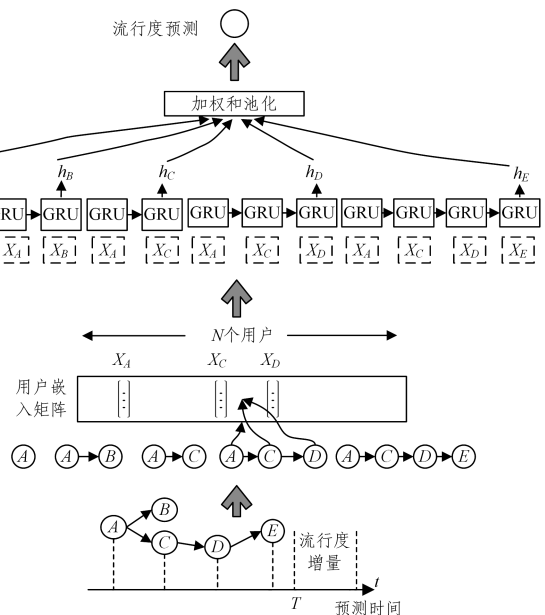


图 3 基于霍克斯过程的信息级联预测模型

Fig. 3 Information cascade prediction model based on Hawkes process

基于霍克斯过程的建模方法提高了预测模型的可解释性,却在一定程度上简化了信息扩散的内容复杂性与时序复杂性,其内容复杂性缺失体现在:霍克斯过程只对信息的传播过程进行建模,忽略了信息本身的属性对传播的影响。为此,文献[36]提出了考虑节点内容信息的 LDA-DeepHawkes 模型,利用 LDA 主题分类模型抽取消息文本内容的主题,使用深度模型学习主题的表达向量,模拟主题之间的自激机制。LDA 模型已经被广泛应用于社交网络分析(主要包括主题挖掘与演化)<sup>[39-43]</sup>,LDA 以  $M$  条消息的文本内容为输入,以主题-词频矩阵和文档-主题矩阵为输出。主题-词频矩阵描述每个主题生成不同词的概率,文档-主题矩阵描述每条消息对应到不同主题的概率。模型通过加权和池化的方式将只考虑信息级联的 DeepHawkes 模型与 LDA 主题分类模型融合,更加全面地建模了信息扩散过程。

霍克斯过程的时序复杂性缺失体现在其对时间衰竭效应的建模过于简单,因为普通的循环神经网络在对序列数据进行建模时会假定单调的时间依赖性。在流行度预测中,相邻的节点转发可能具有不同的时间间隔。通常,较短的时间间隔表示有关社交媒体内容的讨论更加激烈,普通的循环神经网络很难捕捉这种差异。此外,社交媒体中的用户可能在不同时间具有不同的活动级别,例如,用户在休息时间可能更活跃,而在工作时间则不太活跃,普通的霍克斯过程并不足以捕捉到这种差异。文献[44]提出了一种时间嵌入增强型循环神经网络,为了捕获复杂的时间信息,该网络向 RNN 的每个输入添加了时间嵌入向量。时间嵌入向量由 3 部分组成:1) 每小时的时间嵌入向量;2) 每个工作日的时间嵌入向量;3) 第  $i$  次转发的时间嵌入向量,表示为  $s(t_0 - t_i)$ ,其中  $s$  是嵌入函数,该文选择了有效表示能力较好的余弦函数, $t_0$  是观察时间, $t_i$  是第  $i$  次转发的时间。因此,在每个步骤中,RNN 的状态更新将同时考虑输入和 3 种时间信息,本文采用双向门控循环单元(Bidirectional Recurrent Neural Network, BiGRU) 替换基础的 RNN 单元。模型通过文本内容编码器、用户编码器和时间序列编码器分别从数据中学习文本内容、用户和时间序列表示。最后,融合层将这 3 个编码器产生的特征进行组合,输出节点的嵌入表示。

受限于扩散路径采样的样本稀疏性,注意力机制对于扩散路径采样数据的计算代价十分昂贵。为了解决这一缺点,Islam 等<sup>[37]</sup>提出了增强 DeepDiffuse 模型,该模型尽管也使用了扩散路径采样,但一次只采样并分析级联序列的一部分,并根据其评估移至扩散路径的下一个位置,这样可以有效地分析较长的信息级联序列。该模型具有两个主要组件:级联分析器网络(Cascade Analyzer Network, CAN)和级联预测器网络(Cascade Predictor Network, CPN)。CAN 生成级联子序列的内部矢量表示  $r_j$ ,该子序列由 CPN 在上个时间步计算得出。CPN 获取当前时间的子序列表示  $r_j$  并将其输入循环神经网络,计算级联序列的下一个位置  $l_j$ ,并计算下一个节点的时序信息。CPN 的最后部分是输出层,该层计算下一个扩散节点的概率分布。

总的来说,由于信息级联的规模一般分布不均,导致扩散路径采样后的序列数据具有一定的稀疏性,这在一定程度上限制了模型的灵活性与预测能力;同时,扩散路径采样用一组子序列表示信息级联,尽管该方法保留了序列之间的相关性,能够捕捉局部结构信息,但却忽视了信息级联的整体结构,与随机游走采样一样无法捕捉全局结构信息。

#### 4 拓扑信息级联预测方法

ISPBTS 主要通过不同的信息级联采样方法来捕捉节点的拓扑结构信息,这样虽然有一定的效果,但仍受限于顺序网络框架无法捕捉复杂扩散信息的先天不足,且采样过程中所保留的级联拓扑结构信息容易在模型学习过程中流失。随着图嵌入算法的研究逐渐过渡到神经网络时代,基于拓扑结构的 ISPBTO 不再仅将节点的拓扑表示用于数据预处理,而是将节点的邻接信息同节点本身的特征一起纳入模型,使整个拓扑表示学习从静态学习向动态的归纳式学习演变,无论是模型拟合能力还是泛化能力都有较大的提升<sup>[45]</sup>。

针对时序信息级联预测方法的一些普遍问题,ISPBTO 不仅学习级联序列中每个节点的时序表示,而且尝试结合深度学习框架来学习级联图的拓扑结构表示。ISPBTO 的大部分灵感来源于表示学习和深度学习在文本<sup>[46]</sup>、图像<sup>[47]</sup>等各个领域的应用。

如图 4 所示,这类方法在时序信息级联预测方法的基础上,使用局部连接<sup>[48]</sup>、光谱方法<sup>[49]</sup>和热核签名<sup>[50]</sup>等方法来采样网络的拓扑图,然后通过注意力机制和图神经网络等框架进行结构特征信息的传递与学习。这些方法具有为同构图生成相同表示的优点,但图神经网络的缺点是对节点的时间顺序不敏感。因此,在利用图表示方法学习信息级联拓扑结构的同时,还需利用循环神经网络来捕捉节点的时序结构特征。本文根据节点拓扑结构嵌入的学习方式不同,将拓扑信息级联预测方法分为两类:基于全局拓扑结构的方法与基于邻域聚合的方法。表 2 列出了从原理、代表方法和优缺点 3 方面对这两种方法进行对比的结果。

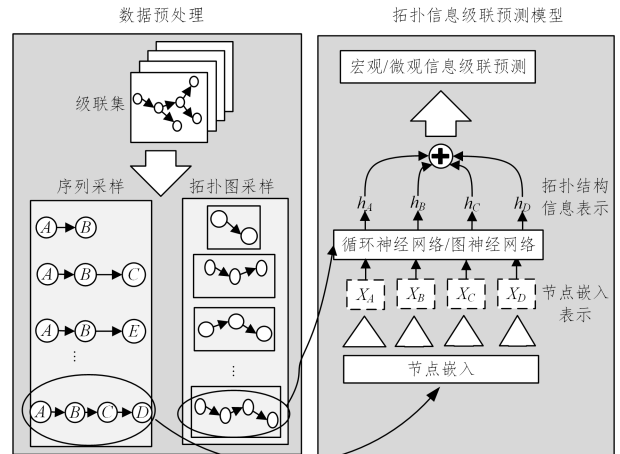
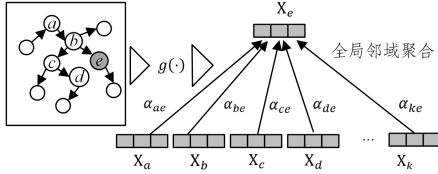
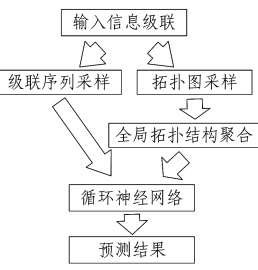
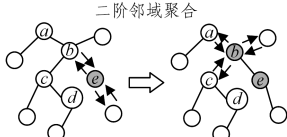
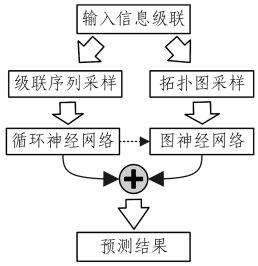


图 4 拓扑信息级联预测模型

Fig. 4 Topological information cascade prediction model

表 2 拓扑信息级联预测方法的对比  
Table 2 Comparison of ISPbTS methods

分类	代表方法	嵌入原理	算法流程	优/缺点
基于全局拓扑结构的方法	SNIDSA <sup>[51]</sup> 、 Topo-LSTM <sup>[52]</sup>	 <p>基于全局拓扑结构的方法将信息级联序列的全局拓扑结构 <math>g(\cdot)</math> 作为模型的输入, 利用注意力机制等框架对级联图进行全局邻域聚合之后, 再通过循环神经网络进行学习, 为每个节点生成融合了结构特征的向量表示。</p>		<p>优点: 能够完整地捕捉结构信息, 可以与时序信息级联预测方法中的一些采样方法进行结合, 例如文献[52]将扩散路径采样与拓扑结构结合。</p> <p>缺点: 模型需要传入全局的拓扑结构表示, 在大型稀疏图上的计算代价较高, 训练时间较长, 且模型的扩展性较差。</p>
基于邻域聚合的方法	CasCN <sup>[53]</sup> 、 DMT-LIC <sup>[54]</sup> 、 FOREST <sup>[55]</sup>	 <p>基于邻域聚合的方法通过神经网络中的邻域采样、图卷积操作和图注意力机制等算法聚合节点的局部邻域嵌入(一般为一阶邻域或二阶邻域), 从而获得节点的结构特征向量表示, 将其与节点的时序表示结合之后进行预测。</p>		<p>优点: 邻域聚合方法通过迭代的方式使拓扑结构特征学习的效率更高, 模型的扩展性与灵活性较好, 训练时间也较短。</p> <p>缺点: 迭代聚合过程可能造成结构信息的部分丢失, 全局拓扑结构的捕捉能力不如基于全局拓扑结构的方法。</p>

4.1 基于全局拓扑结构的方法

基于全局拓扑结构的方法以嵌入信息级联的全局拓扑结构来进行节点结构特征学习, 这类方法仍然以循环神经网络为节点时序信息的学习框架, 并不直接对得到的节点时序表示进行监督学习, 而是与信息级联中节点的拓扑结构嵌入表示共同学习。文献[52]在基于扩散路径的时序信息级联预测方法的基础上, 设计了新的动态扩散拓扑模型, 其拓扑结构的表示方式与在拓扑图上进行扩散路径采样类似。该文还为每个节点设计了两种不同“用户角色”的节点嵌入向量, 即发送者嵌入向量与接收者嵌入向量, 并尝试将每个活动节点与发送者嵌入向量一起嵌入, 将每个非活动节点与接收者嵌入向量一起嵌入, 且不仅对节点的静态趋势进行了编码, 而且对扩散拓扑的动态上下文进行了编码。另外, 由于现有的循环神经网络模型主要关注序列结构输入, 虽然有少数神经网络模型对不同的应用领域设计的静态有向无环图(Directed Acyclic Graph, DAG)进行建模, 如 DAG-RNN<sup>[56-57]</sup>将每个 2D 图像建模为场景标记的 DAG, RNN-LE<sup>[58]</sup>将每个蛋白质氨基酸的接触图建模为拓扑图来预测蛋白质的结构, 但这类循环神经网络模型仍基于纯 RNN 体系结构, 无法捕获动态扩散过程的复杂性。因此, 文献[52]提出了用于扩散拓扑 DAG 结构的循环神经网络模型 Topo-LSTM, 它以动态 DAG 为输入, 并将 DAG 中的每个节点生成的拓扑感知的嵌入为输出。Topo-LSTM 通过节点  $v_i$  的发送者嵌入可以得出目前为止还激活了其他哪些节点以及拓扑图如何扩散到  $v_i$ , 通过学习每个非活动节点的接收者嵌入, 并根据其与活动节点的发送者嵌入的接近程度来预测下一个扩散节点。Topo-LSTM 的模型框架如图 5 所示。

利用全局拓扑结构捕捉级联扩散上下文的另一种方法是引入结构注意模块 (Structure Attention Module, SAM)<sup>[51]</sup>, 该模块在注意力机制的基础上为每个受感染用户的结构扩散上下文进行建模, 通过信息级联的拓扑图表示构造了图的结构上下文向量:

$$a_{ij} = g(v_i, v_j) \cdot w^T[x_i; x_j] \tag{4}$$

其中,  $v_j$  是直接或间接连接到  $v_i$  的结构上下文用户;  $g(\cdot)$  是在拓扑图上定义的用于测量图中两个用户的连接紧密程度的函数;  $w^T[x_i; x_j]$  是基于串联的注意力机制, 它测量  $v_i$  对每个  $v_j$  的注意力得分;  $[\cdot]$  表示向量串联。基于 SAM, 文献[51]开发了一种具有结构关注度的新型顺序神经信息扩散模型 SNIDSA, 采用基于循环神经网络的框架对历史顺序扩散进行建模。为了有效地整合顺序信息和结构信息, SNIDSA 还设计了一种有效的门控机制来更新神经隐藏状态, 以整合在隐藏状态下压缩的历史扩散信息、当前受感染的节点及其扩散上下文。所提出的门控单元与 SRU<sup>[59]</sup> 的体系结构相似。基于门控单元与结构注意模块的信息级联预测框架如图 6 所示。

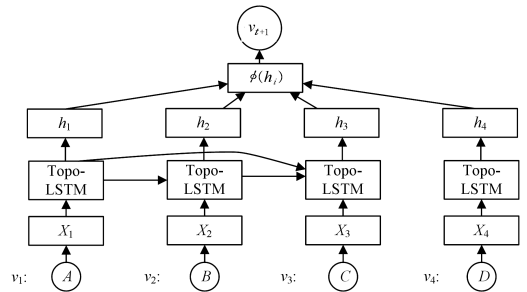


图 5 Topo-LSTM 模型框架

Fig. 5 Topo-LSTM model framework

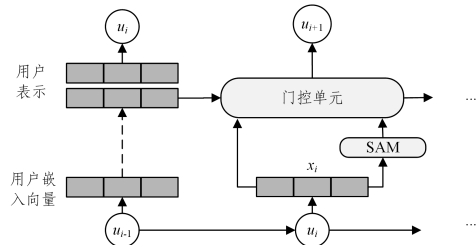


图 6 基于门控单元与结构注意模块的信息级联预测框架

Fig. 6 Information cascade prediction based on gating unit and SAM

## 4.2 基于邻域聚合的方法

基于全局拓扑结构的方法以信息级联的完整拓扑结构为输入,这类方法对于社交媒体中常见的大型稀疏图来说是一种计算代价很高且不适用的方法。基于邻域聚合的方法使用图神经网络迭代地聚合了节点的邻域嵌入,并使用在前一次迭代中获得的嵌入及其嵌入函数来获得新的嵌入,模型的多次迭代允许学习嵌入一个节点来描述全局邻域。仅局部邻域的聚合嵌入使这类方法比基于全局拓扑结构的方法更具扩展性。在信息级联预测领域,常用的图神经网络框架包括 GraphSAGE<sup>[60]</sup>、图卷积网络(Graph Convolutional Network, GCN)<sup>[61]</sup>与图注意力网络(Graph Attention Network, GAT)<sup>[62]</sup>等。文献[55]提出了采用与 GraphSAGE 类似的邻域采样策略来聚合节点邻域特征,以更新信息级联节点的结构特征向量。

$$f_v = \text{relu}(\mathbf{W} \cdot \frac{1}{Z} \sum_{k=1}^Z f_{u_k} + \mathbf{b}) \quad (5)$$

其中,  $\{u_k, k=1, 2, \dots, Z\}$  是从集合  $\{v\} \cup N(v)$  中采样得到的节点集,  $\mathbf{W}$  和  $\mathbf{b}$  分别是节点结构特征向量的权重矩阵与偏置值向量, 激活函数  $\text{relu}(\cdot) = \max(\cdot, 0)$ 。假设级联观察窗口  $[0, T)$  内有  $j$  个节点, 则该级联序列的拓扑结构信息可建模为  $j$  个节点的结构特征向量的平均池化。

$$s_k = \text{mean}(f_{v_1}, f_{v_2}, \dots, f_{v_j}), t_j \leq T \quad (6)$$

对于拓扑结构信息较为稀疏的级联图, 还可以递归地处理式(6)的步骤, 以探索节点  $v$  的较大邻域。根据经验, 两阶的邻域探索已经足以提供较好的结构特征刻画效果。最后, 基于邻域聚合的方法, 通过将级联序列的拓扑结构信息  $s_k$  与循环神经网络输出的时序结构信息  $h_k$  进行拼接, 最后通过 softmax 层进行预测。

$$p_k = \text{softmax}(\mathbf{W} \cdot \text{concat}(s_k, h_k) + \mathbf{b}) \quad (7)$$

文献[53]提出了直接基于子图结构进行数据采样与特征学习的 CasCN 模型, 将图谱理论桥接到深度神经网络中, 利用图卷积网络提高了信息级联预测的拓扑结构信息捕捉能力。经典的图卷积网络通过对其邻居的表示取未加权平均值来计算其隐藏表示, 但不能直接用于级联建模, 因为其只关注无向图, 而无向图又不能考虑级联演化的时间信息。为此, CasCN 将单个级联节点  $X_i$  的卷积运算建模, 并依靠 Diplacian<sup>[63]</sup>解决无法捕获不同级联上随机游走的特征差异问题。在获得每个级联图的子级联图序列  $A_i^T$  和拉普拉斯矩阵  $\Delta c$  的邻接表示之后, CasCN 结合经典 LSTM 和 GCN 学习节点的结构与时序表示, 其核心是将图卷积操作与循环神经网络结合, 以合并时序结构信息与拓扑结构信息。模型通过将现有存储单元替换为新单元  $c_i$  来更新存储单元, 如式(8)所示:

$$\begin{aligned} i_i &= \sigma(\mathbf{W}_i * \mathbf{G}X_i + \mathbf{U}_i * \mathbf{G}h_{i-1} + \mathbf{V}_i \odot c_{i-1} + \mathbf{b}_i) \\ f_i &= \sigma(\mathbf{W}_f * \mathbf{G}X_i + \mathbf{U}_f * \mathbf{G}h_{i-1} + \mathbf{V}_f \odot c_{i-1} + \mathbf{b}_f) \\ o_i &= \sigma(\mathbf{W}_o * \mathbf{G}X_i + \mathbf{U}_o * \mathbf{G}h_{i-1} + \mathbf{V}_o \odot c_i + \mathbf{b}_o) \\ c_i &= f_i \odot c_{i-1} + i_i \odot \tanh(\mathbf{W}_c * \mathbf{G}X_i + \mathbf{U}_c * \mathbf{G}h_{i-1} + \mathbf{b}_c) \end{aligned} \quad (8)$$

其中,  $*$  表示图卷积操作,  $\sigma(\cdot)$  是逻辑 sigmoid 函数, 它的

$i, f, o, b$  分别代表输入门、忘记门、输出门和偏置向量。 $\mathbf{W}, \mathbf{U}$  和  $\mathbf{V}$  是权重矩阵。CasCN 的框架如图 7 所示, 可以作为图卷积网络在 ISPbDL 中的典型框架进行使用与改进。

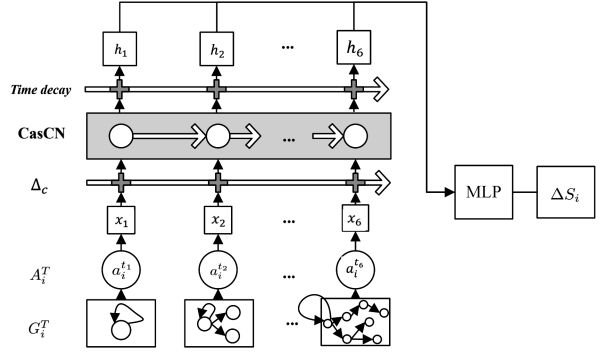


图 7 基于 GCN 的信息级联预测模型 CasCN

Fig. 7 CasCN: GCN-based information cascade prediction model

在复杂的拓扑网络结构中, 同构假设可能不正确, 例如在社交网络的扩散动力学中, 活跃的邻居明显比不活跃的邻居更重要, 而 GCN 通过对邻居赋予相同的权重进行平均求和, 无法为用户节点赋予不同的扩散依赖。因此, 使用更加灵活的注意力机制来区别对待节点邻居是在较为复杂的拓扑网络中的一个更好的选择。文献[54]提出多任务级联预测框架 DMT-LIC, 并在其中加入了多头注意机制的图注意力网络, 其逐层传播规则为:

$$\mathbf{H}^{(l+1)} = \sigma\left(\sum_{k=1}^K \mathbf{A}^k \mathbf{W}^k \mathbf{h}^{(l)}\right) \quad (9)$$

其中,  $\mathbf{W}^k$  是一组独立的可训练权重矩阵;  $K$  是注意力的头数;  $\sigma(\cdot)$  表示激活函数, 即  $\text{ReLU}(\cdot)$ ,  $\mathbf{H}^{(l)} \in \mathbb{R}^{N \times F}$  是第  $l$  层的激活矩阵, 其中  $N$  是级联图中的节点数,  $F$  是特征数。第一层输入  $\mathbf{H}^{(0)} = \text{Adj} + \mathbf{I}_N$ , 其中  $\text{Adj} \in \mathbb{R}^{N \times N}$  是级联子图的邻接矩阵,  $\mathbf{I}_N$  是单位矩阵,  $\mathbf{A}^k = [a_{ij}]_{N \times N}$  是自注意力机制的注意力矩阵, 其定义如下:

$$a_{ij} = \frac{\exp(\text{LeakyReLU}(c^T \mathbf{W}_{hi} \parallel \mathbf{W}_{hj}))}{\sum_{k \in N} \text{LeakyReLU}(c^T \mathbf{W}_{hi} \parallel \mathbf{W}_{hk})} \quad (10)$$

在图注意力网络层之后, 级联图  $G$  可表示为向量矩阵  $\mathbf{H}^{\text{cas}} \in \mathbb{R}^{N \times d_{\text{cas}}}$ , 模型通过不同的密集层将表示连接起来以预测结果, 即预测级联增量大小的宏观任务与下一个受感染的节点的微观任务。

## 5 实现细节

为了提高 ICPbDL 的预测性能, 丰富预测算法的功能, 除了设计出更好的算法流程框架之外, 本节将介绍在信息级联预测算法设计实现时的一些常用技术细节, 将这些实现细节与本文提到的信息级联预测算法结合, 可以在一定程度上提升算法的性能, 或拓展算法的功能应用。近几年的主流信息级联预测算法均不同程度地采用了如下 6 种实现细节。

(1) 社交网络数据预处理。为了避免模型在学习时受到社交网络扩散过程中的昼夜规律(即夜晚的转发数通常远小于白天)的影响, 一般仅考虑发布时间在 8:00 和 18:00 之间的信息级联; 分别在流行度达到最终流行度 50%, 60% 和

70%时进行观察窗口的截取,对于每个观察窗口,仅考虑在其中具有不少于10条转发和不超过1000条转发的级联,以减弱无效信息级联的影响。

(2)结合循环神经网络与图神经网络。从一些最新的信息级联预测方向的研究<sup>[53-54]</sup>可以看出,将用于捕捉时序结构的循环神经网络与用于捕捉拓扑结构的图神经网络结合是完善信息级联结构表示的必要手段,现有的方法主要利用信息级联的拓扑结构表示(如邻接矩阵)来提升循环神经网络的结构特征刻画能力,或使用RNN学习得到的节点隐藏状态表示来提升图神经网络的特征刻画能力。在这两种方法中,两种神经网络框架均以串联的形式结合,我们推荐以残差的形式将两种神经网络的特征表示输出再次聚合,以减小模型在参数传递过程中的信息损失。

(3)使用平均池化作为拓扑结构特征的聚合策略。邻域结构特征的聚合策略包括加权和池化、平均池化、卷积、注意力加权等。在实践中发现,使用平均池化作为聚合策略可以使模型更加简单有效。相比加权和池化,平均池化可以在保证模型性能的同时有效防止模型过拟合<sup>[55]</sup>。

(4)使用多头注意力机制<sup>[64]</sup>。多头注意力机制可以将用户嵌入投影到具有不同线性投影的多个子空间中,在每个子空间上独立执行注意力机制,最后将所有注意力权重进行拼接,提高了模型的可表达性。多头注意力的优势在于可以使模型独立地从不同角度(即子空间)“分割和表示”信息级联的结构信息,因此比传统的注意力机制更强大。实践表明,无论是在循环神经网络还是在图注意力网络上,使用多头注意力机制均能对模型的性能有一定提升。在常见的信息级联预测算法中,注意力机制的头数 $K$ 一般取值为4或8。

(5)以非参数的形式建模时间衰竭效应。时间衰竭效应即一个节点对其他节点的影响会随着时间而产生不同程度的减小。在不同的领域均有已定义的时间衰竭函数,这些函数的选择因不同的情况而异,且都具有参数假设的局限性,在很大程度上受到假设的先验分布的影响。我们推荐两种以非参数形式学习时间衰竭效应的方法。1)注意力机制:注意力机制在无法获取信息级联具体转发时间的情况下也可以用于建模时间衰竭效应。一般的做法是将注意力机制与循环神经网络结合,如采用Att-LSTM和Att-GRU等神经网络框架。文献[26]采用覆盖策略来调整注意力的错配,使注意力的分配模式能够更好地反映传播的真实结构;文献[72]采用了两级

注意力机制,用节点级别的注意力表征用户之间的扩散依赖,用序列级别的注意力表征级联的时间衰竭效应与用户在级联中的重要性。2)离散变量表示:假设观察到的信息级联的时间窗为 $[0, T]$ ,即 $\{[t_0, t_1), [t_1, t_2), \dots, [t_{l-1}, t_l = T)\}$ ,将窗口分成 $l$ 个不相交的时间间隔,学习每个时间间隔对应的时间衰减变量 $\lambda = \{\lambda_m, m \in (1, 2, \dots, l)\}$ ,其中 $m$ 是时间间隔函数, $t$ 时刻转发的时间间隔 $m$ 可表示为:

$$m = \frac{t - t_0}{T/l} \quad (11)$$

(6)将微观与宏观的信息级联预测任务结合。多任务学习是深度学习框架的发展趋势,在信息级联预测领域中,将宏观与微观信息级联预测相结合的方法有以下两种。1)基于共享表示<sup>[54]</sup>:即设计多任务层,在捕获学习级联图的基础结构和扩散过程中的节点序列之后,将两种不同的信息级联表示与特定任务的嵌入连接在一起,形成针对不同任务的新表示;然后通过不同的神经网络模型进行特征表示学习,最后将不同的表示合并后进行不同的预测任务,包括级联增量大小与下一个受感染的用户节点的预测。2)基于强化学习<sup>[55]</sup>:强化学习通过策略梯度等算法来解决微观预测的参数不可微问题,将宏观的流行度预测纳入微观的节点预测模型中。这类方法通过微观信息级联预测模型对观察到的级联序列进行编码,通过策略梯度等强化学习算法<sup>[65]</sup>更新宏观信息级联预测的学习参数。

## 6 信息级联预测算法分析与对比

在对信息级联预测进行实验时,若要对每种方法进行公平、客观的评价,就必须使用权威的数据集和统一的评价指标。根据信息级联预测的下游应用,我们将数据集的应用范围分为宏观级联预测与微观级联预测,对应的典型应用场景分别为流行度预测与节点预测。本节将介绍多种信息级联预测实验中常用的一些公共数据集以及衡量算法性能的指标,并对一些经典方法的实验结果进行系统地分析和对比。由于信息级联序列的级联图既可以由数据集直接给出,也可以由级联序列中的节点关系得到,因此所有的数据集均可执行ICPbTS与ICPbTO方法。表3列出了信息级联预测的常用公开数据集,我们关注的是每个数据集的应用场景、全球网络中的级联图数、节点数与边数,以及各自的数据集划分情况。

表3 常用的信息级联预测数据集

Table 3 Commonly used information cascaded prediction dataset

Dataset	Literature	Mission	V	E	Training set	Validation set	Test set
Twitter	文献[66]	多种应用	137093	3589811	26871	1574	6663
Sina Weibo	文献[35]	多种应用	10077	11956	38576	8266	8266
AMINER	文献[33]	多种应用	131415	842542	34347	7428	7337
APS	文献[27]	流行度预测	13945	15508	24658	5254	5279
HEP-PH	文献[67]	流行度预测	34546	421578	3478	848	848
Memetracker	文献[68]	节点预测	5000	313669	N/A	N/A	N/A
Digg	文献[69]	节点预测	279632	2617993	N/A	N/A	N/A
BlogCatalog	文献[70]	节点预测	10312	333983	N/A	N/A	N/A
Flickr	文献[70]	节点预测	80513	5899882	N/A	N/A	N/A

注:“多种应用”表示该数据集在流行度预测与节点预测中均可使用,N/A表示相关论文未提及

## 6.1 常用公开数据集

(1)Twitter<sup>[66]</sup>。由 Twitter 公共 API 收集,其中每个样本对应一个推文。样本内容包括推文的编号、在观察时间内该推文的转发用户与转发时间戳。

(2)Sina Weibo<sup>[35]</sup>。收集 2016 年 6 月 1 日产生的所有原始消息,并在接下来的 24 小时内跟踪每条消息的转发。文献<sup>[35]</sup>过滤掉转发少于 10 条的消息,剩余 119313 条消息。

(3)AMINER<sup>[33]</sup>。基于 1992 年至 2002 年间的 DBLP 引用构建了全球网络  $G$ ,训练集中包括了 2003 年至 2007 年间发表的论文,验证集和测试集分别使用 2008 年和 2009 年发表的论文。

(4)APS<sup>[27]</sup>。包括 1893 年至 2009 年间 11 种 APS 期刊发表的所有论文以及其中的引文。一篇论文的所有引文都形成一个级联,级联的流行度标签是被引用的次数。论文的引文级联公开了潜在的级联路径,即在转发文本中扮演“// @username:”的角色序列。

(5)HEP-PH<sup>[67]</sup>。涵盖了 1993 年 1 月至 2003 年 4 月(124 个月)期间被发表的论文。如果文章  $i$  引用文章  $j$ ,则级联图包含从  $i$  到  $j$  的有向边。

(6)Memetracker<sup>[68]</sup>。收集了 4403 个新闻网站上发布的 54847 个受欢迎的 meme(流行短语),用于研究人群中 meme 的传播模式。每个 meme 都被视为信息项,网站或博客的每个 URL 被视为用户节点,共享时间以及网站名称构成了级联序列,平均级联长度为 17.0。

(7)Digg<sup>[69]</sup>。包含了 2009 年上升到 Digg 网站头条(一个月内)的所有新闻数据。转发的时间戳以及匿名的用户 ID 构成了级联。数据集共有 279632 个用户节点,平均级联长度为 30.0。

(8)BlogCatalog。博客社交关系网络,网络中的标签代表博客作者的组成员身份。BlogCatalog 上有 10312 个节点、333983 条边和 39 个不同的标签。文献<sup>[34]</sup>随机选择 31 个标签作为训练嵌入的配置文件信息,选择 3 个标签进行验证,其余的标签用于评估节点分类任务的性能。

(9)Flickr。照片共享社交网络,标签代表了用户照片风格的兴趣集。Flickr 上有 80513 个节点、5899882 条边和 195 个不同的标签。文献<sup>[34]</sup>随机选择 167 个标签作为节点属性信息进行训练,选择 18 个标签进行验证,其余的标签用于评估。

## 6.2 评价指标

根据实验预测结果的不同,我们将信息级联预测的性能评价指标分为两类:宏观预测评价指标与微观预测评价指标。

### 6.2.1 宏观信息级联预测评价指标

在宏观信息级联预测方法中,常用的评价指标包括均方误差(Mean Square Error, MSE)、均方对数转换误差(Mean Square Log Error, MSLE)、中位数平方对数转换误差(median squared log error, mSLE)、均方根误差(Root Mean Square Error, RMSE)、平均相对平方误差(Mean Relative Squared Error, MRSE)和平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)。其中,MSLE 是级联增量预测领域使用频率最高的评价指标。在结果评价时,一般选取 MSLE 和

mSLE 作为评价指标,MSLE 和 mSLE 的计算公式如式(12)和式(13)所示。

(1)均方对数转换误差(MSLE)常用来评估预测的准确性,是回归任务的常见选择。

$$MSLE = \frac{1}{M} \sum_{i=1}^M SLE^i \quad (12)$$

其中,  $SLE^i = (\log \Delta \hat{s}_i - \log \Delta s_i)^2$ 。

(2)中位数平方对数转换误差(mSLE)是  $SLE^i, i=1, 2, 3, \dots, M$  的中位数,因此能够有效减轻离群点的影响。

$$mSLE = \text{medain}(SLE^1, \dots, SLE^M) \quad (13)$$

### 6.2.2 微观信息级联预测评价指标

在微观信息级联预测方法中,通常会用到一些基础的性能评价指标,包括曲线下面积(Area Under Curve, AUC)<sup>[71]</sup>、精度(Precision)、召回率(Recall)和 F1 指标(F1 Score)。除此以外,一般将下一个激活节点的预测任务视为排名问题,并将用户节点的转发概率作为其得分进行排序,常用的基于排名的性能评价指标包括以下 4 个。

(1)Acc@k<sup>[26]</sup>:前  $k$  个排名的精度。

(2)MRR<sup>[26]</sup>:平均倒数排名。

(3)Hits@k<sup>[37]</sup>:前  $k$  个节点中包含下一个激活节点的比率。

(4)MAP@k<sup>[37]</sup>:即平均 AP(Average Precision)值。

## 6.3 实验结果与对比

随着计算机性能的提升,ICPbDL 方法主要以“提高预测准确度”为目的进行算法改进,因为预测类算法的时间性能并不是研究重点,本文仅考虑各种算法的“预测准确度”。按照预测任务的不同,将本文介绍的信息级联预测算法分为两类进行对比:宏观信息级联预测方法与微观信息级联预测算法。本文实验将用户节点嵌入层单元数设为 128,循环神经网络层单元数设为 256(双向循环神经网络则设为 128),多头图注意力网络单元数为  $8 \times 8$ ,多层感知器单元数为  $32 \times 64$ ,drop-out 率设为 0.25,Adam 优化器初始学习率为 0.0001。另外,在基于扩散路径的方法中,我们将序列采样长度设为 20,以减小注意力机制的计算成本。为了区分信息级联预测的宏观与微观应用场景,在本文的对比实验中,宏观信息级联预测方法采用 Sina Weibo 作为数据集,流行度增量作为标签;微观信息级联预测方法采用 Twitter 数据集,下一个扩散节点的 id 作为标签。

### 6.3.1 宏观信息级联预测方法实验对比

宏观信息级联预测方法的实验对比如表 4 所列,主要比较因素包括关键技术、实验数据集和评价指标等。

从表 4 可以看出,基于随机游走和扩散路径的 ICPbTS 方法预测效果均低于 ICPbTO 方法,其中 DeepHawkes<sup>[35]</sup>与 LDA-DeepHawkes<sup>[36]</sup>因加入了霍克斯过程的原理,使得模型的可解释性大大提高,且模型结构较为简单,因此成为了广泛使用的基线模型。宏观信息级联预测方法的最新进展主要集中于使用图神经网络(GNN)中的邻域采样聚合原理对级联网络进行拓扑结构的建模与学习,图卷积网络与图注意力网络在此领域效果显著,其论文成果包括文献<sup>[53]</sup>和文献<sup>[54]</sup>。

表4 宏观信息级联预测方法实验对比

Table 4 Experimental comparison of macro information cascade prediction methods

分类方法	方法名称	文献	发表年份	关键技术	数据集	MSLE
基于随机游走的方法	DeepCas	文献[33]	2017	GRU, 注意力机制	Sina Weibo	3.107
	DeepHawkes	文献[35]	2017	霍克斯过程	Sina Weibo	2.854
基于扩散路径的方法	LDA-DeepHawkes	文献[36]	2019	LDA 模型	Sina Weibo	2.832
	NPP	文献[44]	2018	BiGRU	Sina Weibo	N/A
基于邻域聚合的方法	CasCN	文献[53]	2019	GCN, 注意力机制	Sina Weibo	2.817
	DMT-LIC(宏观)	文献[54]	2019	GAT	Sina Weibo	2.820
	FOREST(宏观)	文献[55]	2019	策略梯度算法	Sina Weibo	2.833

注: N/A 表示相关论文未提及或无法复现该项

### 6.3.2 微观信息级联预测方法实验对比

微观信息级联预测方法的实验对比如表5所列, 主要比较因素包括关键技术、实验数据集和评价指标等。

从表5可以看出, 预测性能较好的微观信息级联预测方法以ICPbTO为主, 这主要是因为在使用用户节点级别的信息级联建模中, 级联的拓扑结构信息的重要性会大大增加, 而ICPbTS方法更注重信息级联的时序结构信息, 因此并不适用于微观的节点预测。最新的微观信息级联预测方法的主要

创新点包括两点: 1) 设计了复杂网络的特殊社交网络嵌入方法, 例如文献[26]针对交叉依赖关系网络设计了加入覆盖策略的注意力RNN, 文献[34]针对多路径关系网络设计了BM-RU循环单元来提取多路径关系; 2) 为模型加入了注意力机制及其变种, 已被证实可以更加有效地建模节点的复杂依赖, 也可以帮助模型更好地提取级联扩散过程中的信息<sup>[37,72]</sup>, 其中文献[72]分别建模用户级别的注意力与级联级别的注意力, 这两级注意力机制是此方向的最新进展。

表5 微观信息级联预测方法实验对比

Table 5 Experimental comparison of micro information cascade prediction methods

分类方法	方法名称	文献	发表年份	关键技术	数据集	评价指标	MAP@50/%
基于随机游走的方法	DeepDiffuse	文献[37]	2018	双级联网络, 注意力机制	Twitter	MAP@50	16.54
基于扩散路径的方法	MPR-SNE	文献[34]	2019	BM-RU 循环单元	Twitter	MAP@50	N/A
基于全局拓扑结构的方法	SNIDSA	文献[51]	2018	注意结构模块	Twitter	MAP@50	16.10
	Topo-LSTM	文献[52]	2017	扩散拓扑模型	Twitter	MAP@50	16.18
基于邻域聚合的方法	DMT-LIC(微观)	文献[54]	2019	GAT	Twitter	MAP@50	16.97
	FOREST(微观)	文献[55]	2019	策略梯度算法	Twitter	MAP@50	17.31
	HiDAN	文献[72]	2019	两级注意力机制	Twitter	MAP@50	N/A

注: N/A 表示相关论文未提及或无法复现该项

**结束语** 深度学习技术在信息级联预测领域的应用已经成为一大研究热点, 本文汇总了截至目前为止的所有ICPb-DL方法, 分别进行了真实问题的陈述与相关概念的定义, 并从时序结构捕捉与拓扑结构捕捉两方面对信息级联预测方法进行分类。本文重点讨论了每种方法的技术特点以及优缺点, 讨论了一些该领域常用的可以提高模型性能的实现细节, 最后使用全面的评估指标对一些典型的信息级联预测方法进行对比与分析。

根据对信息级联预测这一领域的重点问题与发展趋势的探讨, 我们认为该领域还存在以下具有挑战性的研究方向。

#### (1) 基于多模态/多视图的信息级联预测

多模态/多视图指信息的不同描述方式, 移动社交媒体中的多媒体数据, 如文本信息、图像信息、视频信息往往是多种信息级联的传递媒介(例如一次人物行为状态往往会同时使用多种信息同时进行分享与传播)<sup>[73-74]</sup>。当信息级联中的多个模态共存时, 需要同时从多个异质信息源提取特征, 并且考虑多个模态信息的一致性和互补性。多视图信息级联预测的最新进展主要集中于对文本内容与短视频内容的转发进行预测, 如文献[44]对文本内容和用户的嵌入向量进行联合学习, 并采用注意力机制和融合层来捕获不同类型信息之间的相互作用, 以学习不同主题的内容中隐含的不同的流行度演变模式和不同的影响用户表示; 文献[75]将低秩的多视图节点嵌入信息和回归分析集成到一个统一的深度学习框架中, 将多视图特征映射到一个公共子空间, 使所有视图共享的最低秩

表示能够捕获所有视图的全局结构, 实现了多种信息模态的联合流行度预测。整体来看, 现有的信息级联预测方法大多仍针对单一模态进行向量嵌入与模型构建, 极少考虑数据在不同模态直接的映射关系, 如何提取与表征这种映射关系, 并在此基础上构建多模态信息级联预测模型, 是当前急需突破的一大研究方向。

#### (2) 结合时空与社区结构信息的信息级联预测

目前的社交媒介信息是包含地理位置信息、复杂社区结构信息等数据的集合。这些数据含有丰富的语义信息, 并且数据结构复杂多样。传统的数据特征表示模式如词袋模式、One-Hot表示、基于概率生成模式在这类语义信息抽取任务中的特征表达能力受限, 特别是在深度神经网络节点嵌入中带来特征表示稀疏而引起梯度消失等问题。文献[76]针对基于位置的社交网络(Location-based Social Network, LBSN)实现了具有地理特征的用户社区结构建模, 体现了用户属性与地理标签簇间的关联性, 是基于地空协作的社交网络建模在复杂网络分析上的成功应用。但是, 由于人物的社会性活动产生的地理空间信息存在地理空间拓扑限制、空间关联性、时间周期模式性和位置偏好性问题, 并且其产生的社交属性也存在社交偏好和社交群体属性偏好, 带有地理位置标签的信息级联在特征联合表达方面仍然缺乏良好的适配能力, 这是信息级联预测亟待解决的问题。

#### (3) 利用生成模型提高信息级联特征的刻画能力

生成式模型通过理解数据的生成模式, 并通过VAE<sup>[77-78]</sup>

和 GAN<sup>[79-81]</sup> 等方法实现离散数据生成模型的构建。在信息级联预测领域,生成模型可以通过实现面向社交媒体数据的数据生成框架,来探索非高斯分布的数据生成模式,其与深度神经网络的结合有效弥补了传统的基于生成建模方法无法捕捉隐含信息特征的不足。目前,生成模型还较少应用于信息级联预测领域,而 VAE 和 GAN 等模型已经广泛应用于社交网络分析的轨迹识别与序列检测等领域。文献[82]提出一种基于 VAE 的半监督模型,通过学习捕捉大量的未标签数据特征,提升了社交网络媒体中轨迹移动性特征的识别效果。基于生成建模的信息级联预测可能成为该领域的下一个研究热点。

#### (4) 基于元学习的信息级联预测

元学习是集成学习中用于基分类器融合的一种方法,以基分类器的结果为中间特征再次学习一个分类器,再次学习的分类器被称为元分类器,元分类器中使用的算法称为元分类算法。在数据挖掘和文本分类中,大量的研究已经证明元学习是一种非常有效的集成学习方法<sup>[83-84]</sup>。元算法选用深度信念网络(Deep Belief Network, DBN),通过非线性映射自动地学习输入结点的特征,有很强的特征学习能力,能够更精确地对输入样本进行分类。深度信念网络已被成功应用于书写体识别、3D 目标识别、信息检索等领域。然而,如果将深度信念网络直接应用于信息级联预测,则会在输入结点和隐层数较多的情况下面临计算量大、运行时间较长的问题,如何将元学习模式迁移到社交网络分析领域中,并将其应用于信息级联预测仍是一大挑战。

## 参 考 文 献

- [1] YANG J, COUNTS S. Predicting the speed, scale, and range of information diffusion in twitter [C] // Fourth International AAAI Conference on Weblogs and Social Media, 2010.
- [2] GRUHL D, GUHA R, LIBEN-NOWELL D, et al. Information diffusion through blogspace [C] // Proceedings of the 13th international conference on World Wide Web. ACM, 2004: 491-501.
- [3] LESKOVEC J, MCGLOHON M, FALOUTSOS C, et al. Patterns of cascading behavior in large blog graphs [C] // Proceedings of the 2007 SIAM international conference on data mining. Society for Industrial and Applied Mathematics, 2007: 551-556.
- [4] GOLUB B, JACKSON M O. Using selection bias to explain the observed structure of internet diffusions [J]. Proceedings of the National Academy of Sciences, 2010, 107(24): 10833-10836.
- [5] LIBEN-NOWELL D, KLEINBERG J. Tracing information flow on a global scale using Internet chain-letter data [J]. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105(12): 4633-4638.
- [6] LESKOVEC J, ADAMIC L A, HUBERMAN B A. The dynamics of viral marketing [J]. ACM Transactions on the Web (TWEB), 2007, 1(1): 5.
- [7] DOW P A, ADAMIC L A, FRIGGERI A. The anatomy of large facebook cascades [C] // Seventh international AAAI conference on weblogs and social media, 2013.
- [8] KUMAR R, MAHDIAN M, MCGLOHON M. Dynamics of conversations [C] // Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2010: 553-562.
- [9] QIU J, TANG J, MA H, et al. Deepinf: Social influence prediction with deep learning [C] // Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018: 2110-2119.
- [10] GE Y, CHEN S C. Graph Convolutional Network for Recommender Systems [J/OL]. [2020-02-20]. <http://www.jos.org.cn/1000-9825/5928.htm>.
- [11] KUPAVSKII A, OSTROUMOVA L, UMNNOV A, et al. Prediction of retweet cascade size over time [C] // Proceedings of the 21st ACM International Conference on Information and Knowledge Management. ACM, 2012: 2335-2338.
- [12] MA Z, SUN A, CONG G. On predicting the popularity of newly emerging hashtags in Twitter [J]. Journal of the American Society for Information Science and Technology, 2013, 64(7): 1399-1410.
- [13] PETROVIC S, OSBORNE M, LAVRENKO V. Rt to win! predicting message propagation in twitter [C] // Fifth International AAAI Conference on Weblogs and Social Media, 2011.
- [14] SZABO G, HUBERMAN B A. Predicting the popularity of online content [J]. Communications of the ACM, 2010, 53(8): 80-88.
- [15] GUO R, SHAKARIAN P. A comparison of methods for cascade prediction [C] // Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE Press, 2016: 591-598.
- [16] CHENG J, ADAMIC L, DOW P A, et al. Can cascades be predicted? [C] // Proceedings of the 23rd International Conference on World Wide Web. ACM, 2014: 925-936.
- [17] WENG L, MENCZER F, AHN Y Y. Virality prediction and community structure in social networks [J]. Scientific reports, 2013, 3: 2522.
- [18] TSUR O, RAPPOPORT A. What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities [C] // Proceedings of the fifth ACM International Conference on Web Search and Data Mining. ACM, 2012: 643-652.
- [19] BAKSHY E, HOFMAN J M, MASON W A, et al. Everyone's an influencer: quantifying influence on twitter [C] // Proceedings of the 4th ACM International Conference on Web Search and Data Mining. ACM, 2011: 65-74.
- [20] MARTIN T, HOFMAN J M, SHARMA A, et al. Exploring limits to prediction in complex social systems [C] // Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2016: 683-694.
- [21] BAO P, SHEN H W, JIN X, et al. Modeling and predicting popularity dynamics of microblogs using self-excited hawkes processes [C] // Proceedings of the 24th International Conference on World Wide Web. ACM, 2015: 9-10.
- [22] ZHAO Q, ERDOGDU M A, HE H Y, et al. Seismic: A self-exciting point process model for predicting tweet popularity [C] // Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015: 1513-1522.

- [23] CRANE R, SORNETTE D. Robust dynamic classes revealed by measuring the response function of a social system[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(41): 15649-15653.
- [24] RIZOIU M A, XIE L, SANNER S, et al. Expecting to be hip: Hawkes intensity processes for social media popularity[C]// *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017: 735-744.
- [25] XIAO S, YAN J, YANG X, et al. Modeling the intensity function of point process via recurrent neural networks[C]// *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.
- [26] WANG Y, SHEN H, LIU S, et al. Cascade Dynamics Modeling with Attention-based Recurrent Neural Network[C]// *IJCAI*. 2017: 2985-2991.
- [27] SHEN H, WANG D, SONG C, et al. Modeling and predicting popularity dynamics via reinforced poisson processes [C] // *Twenty-eighth AAAI Conference on Artificial Intelligence*. 2014.
- [28] GAO J, SHEN H, LIU S, et al. Modeling and predicting retweeting dynamics via a mixture process[C]// *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016: 33-34.
- [29] GAO S, MA J, CHEN Z. Modeling and predicting retweeting dynamics on microblogging platforms [C] // *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*. ACM, 2015: 107-116.
- [30] GOMEZ-RODRIGUEZ M, LESKOVEC J, SCHÖLKOPF B. Modeling information propagation with survival theory[C]// *International Conference on Machine Learning*. 2013: 666-674.
- [31] KEMPE D, KLEINBERG J, TARDOS É. Maximizing the spread of influence through a social network [C] // *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2003: 137-146.
- [32] WANG Y, SHEN H W, LIU S, et al. Learning user-specific latent influence and susceptibility from information cascades[J]. *arXiv preprint arXiv:1310.3911*, 2013.
- [33] 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*. International World Wide Web Conferences Steering Committee, 2017: 577-586.
- [34] LIN J, ZHANG L, HE M, et al. Multi-path relationship preserved social network embedding [J]. *IEEE Access*, 2019, 7: 26507-26518.
- [35] CAO Q, SHEN H, CEN K, et al. Deephawkes: Bridging the gap between prediction and understanding of information cascades [C]// *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 2017: 1149-1158.
- [36] WANG S J, ZHOU L H, KONG B, et al. LDA-DeepHawkes model for predicting information cascade [J/OL]. *Journal of Frontiers of Computer Science and Technology*: 1-21. [2019-10-14]. <http://kns.cnki.net/kcms/detail/11.5602.TP.20190628.1726.006.html>.
- [37] ISLAM M R, MUTHIAH S, ADHIKARI B, et al. DeepDiffuse: Predicting the Who and When in Cascades[C]// *2018 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2018: 1055-1060.
- [38] 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*. ACM, 2014: 701-710.
- [39] RAMAGE D, DUMAIS S, LIEBLING D. Characterizing microblogs with topic models[C]// *4th International AAAI Conference on Weblogs and Social Media*. 2010.
- [40] ZHANG C Y, SUN J L, DING Y Q. Topic mining for microblog based on MB-LDA model[J]. *Journal of Computer Research and Development*, 2011, 48(10): 1795-1802.
- [41] KANG J H, LERMAN K, GETOOR L. LA-LDA: a limited attention topic model for social recommendation[C]// *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. Heidelberg: Springer 2013: 211-220.
- [42] LIU Y, WANG J, JIANG Y. PT-LDA: A latent variable model to predict personality traits of social network users[J]. *Neurocomputing*, 2016, 210: 155-163.
- [43] NI L P, LIU X J, MA C Y. Topic Evolution Analysis Based on LDA Model and AP Clustering[J]. *Computer Technology and Development*, 2016, 26(12): 6-11.
- [44] CHEN G, KONG Q, XU N, et al. NPP: A neural popularity prediction model for social media content [J]. *Neurocomputing*, 2019, 333: 221-230.
- [45] WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J]. *arXiv:1901.00596*, 2019.
- [46] BENGIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model[J]. *Journal of machine learning research*, 2003, 3(Feb): 1137-1155.
- [47] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]// *Advances in Neural Information Processing Systems*. 2012: 1097-1105.
- [48] NIEPERT M, AHMED M, KUTZKOV K. Learning convolutional neural networks for graphs[C]// *International Conference on Machine Learning*. 2016: 2014-2023.
- [49] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[C]// *Advances in Neural Information Processing Systems*. 2016: 3844-3852.
- [50] LI C, GUO X, MEI Q. Deepgraph: Graph structure predicts network growth[J]. *arXiv preprint arXiv:1610.06251*, 2016.
- [51] WANG Z, CHEN C, LI W. A Sequential Neural Information Diffusion Model with Structure Attention[C]// *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, 2018: 1795-1798.
- [52] 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)*. IEEE, 2017: 475-484.
- [53] 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)*. IEEE, 2019: 770-781.

- [54] CHEN X,ZHANG K,ZHOU F,et al. Information Cascades Modeling via Deep Multi-Task Learning[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM,2019:885-888.
- [55] YANG C,TANG J,SUN M,et al. Multi-scale information diffusion prediction with reinforced recurrent networks[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. AAAI Press,2019:4033-4039.
- [56] SHUAI B,ZUO Z,WANG B,et al. Dag-recurrent neural networks for scene labeling[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:3620-3629.
- [57] BALDI P,POLLASTRI G. The principled design of large-scale recursive neural network architectures--dag-rnns and the protein structure prediction problem[J]. Journal of Machine Learning Research,2003,4(Sep):575-602.
- [58] BIANCHINI M,MAGGINI M,SARTI L,et al. Recursive neural networks for processing graphs with labelled edges:Theory and applications[J]. Neural Networks,2005,18(8):1040-1050.
- [59] LEI T,ZHANG Y,ARTZI Y. Training rnns as fast as cnns[J]. arXiv:1709.02755,2017.
- [60] HAMILTON W,YING Z,LESKOVEC J. Inductive representation learning on large graphs[C]//Advances in Neural Information Processing Systems. 2017:1024-1034.
- [61] KIPF T N,WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907,2016.
- [62] VELIKOVI P,CUCURULL G,CASANOVA A,et al. Graph attention networks[J]. arXiv preprint arXiv:1710.10903,2017.
- [63] LI Y,ZHANG Z L. Digraph laplacian and the degree of asymmetry[J]. Internet Mathematics,2012,8(4):381-401.
- [64] VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017:5998-6008.
- [65] SUTTON R S,MCALLESTER D A,SINGH S P,et al. Policy gradient methods for reinforcement learning with function approximation[C]//Advances in Neural Information Processing Systems. 2000:1057-1063.
- [66] HODAS N O,LERMAN K. The simple rules of social contagion [J]. Scientific Reports,2014,4:4343.
- [67] GEHRKE J,GINSPARG P,KLEINBERG J. Overview of the 2003 KDD Cup[J]. Acm SIGKDD Explorations Newsletter, 2003,5(2):149-151.
- [68] LESKOVEC J,BACKSTROM L,KLEINBERG J. Meme-tracking and the dynamics of the news cycle[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM,2009:497-506.
- [69] HOGG T,LERMAN K. Social dynamics of digg[J]. EPJ Data Science,2012,1(1):5.
- [70] TANG L,LIU H. Relational learning via latent social dimensions[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2009: 817-826.
- [71] BUCKLEY C,VOORHEES E M. Retrieval evaluation with incomplete information[C]//Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM,2004:25-32.
- [72] WANG Z,LI W. Hierarchical diffusion attention network[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. AAAI Press,2019:3828-3834.
- [73] SUN S. Multi-view Laplacian support vector machines[C]// International Conference on Advanced Data Mining and Applications. Heidelberg:Springer,2011:209-222.
- [74] SUN S,SHAWE-TAYLOR J,MAO L. PAC-Bayes analysis of multi-view learning[J]. Information Fusion,2017,35:117-131.
- [75] JING P,SU Y,NIE L,et al. Low-rank multi-view embedding learning for micro-video popularity prediction[J]. IEEE Transactions on Knowledge and Data Engineering,2017,30(8):1519-1532.
- [76] GONG W H,CHEN Y Q,PEI X B,et al. Community detection combined with multi-dimensional relationships in location-based social networks[J/OL]. Journal of Software, 1-16. [2019-10-08]. <http://jos.org.cn/1000-9825/5269.htm>.
- [77] BODIN E,MALIK I,EK C H,et al. Nonparametric inference for auto-encoding variational Bayes[J]. arXiv:1712.06536,2017.
- [78] MASCI J,MEIER U,CIREŞAN D,et al. Stacked convolutional auto-encoders for hierarchical feature extraction[C]// International Conference on Artificial Neural Networks. Heidelberg: Springer,2011:52-59.
- [79] CHEN X,DUAN Y,HOUTHOOFT R,et al. Infogan:Interpretable representation learning by information maximizing generative adversarial nets[C]//Advances in Neural Information Processing Systems. 2016:2172-2180.
- [80] GANIN Y,USTINOVA E,AJAKAN H,et al. Domain-adversarial training of neural networks[J]. The Journal of Machine Learning Research,2016,17(1):2096-2030.
- [81] REZAEI A,XIAO C,GAO J,et al. Protecting Sensitive Attributes via Generative Adversarial Networks [J]. arXiv:1812.10193,2018.
- [82] ZHOU F,GAO Q,TRAJCEVSKI G,et al. Trajectory-User Linking via Variational AutoEncoder[C]//IJCAI. 2018:3212-3218.
- [83] DIAO W,SUN X,DOU F,et al. Object recognition in remote sensing images using sparse deep belief networks[J]. Remote Sensing Letters,2015,6(10):745-754.
- [84] CUI Z,CAO Z,YANG J,et al. Hierarchical recognition system for target recognition from sparse representations[J]. Mathematical Problems in Engineering,2015.



**ZHANG Zhi-yang**, born in 1997, post-graduate, is a member of China Computer Federation. His main research interests include machine learning, data mining and cascade prediction.



**ZHANG Feng-li**, born in 1963, Ph.D. professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include network security and network engineering, cloud computing and big data and machine learning.