

图神经网络在社交网络虚假信息传播预测中的应用

芮嘉阳

南京市江原安迪科正电子研究发展有限公司 江苏南京 211100

摘要：随着社交网络的发展与普及，虚假信息的传播越来越严重，给社会带来了很大负面影响。传统的信息传播机制在应对社交网络中复杂的数据流动时显得力不从心，存在明显局限性。而图神经网络（graph neural networks, GNN）作为当下预测假新闻传播的关键方法，正逐渐展现出其独特优势。本文就探讨图神经网络对社交网络中虚假新闻的预测过程，分析图神经网络的基本理论和社会网络的基本属性之间的关系，提出图神经网络对社会网络中假新闻的优势以及面临的问题，对未来发展方向进行展望。

关键词：图神经网络；社交网络；虚假信息；传播预测；深度学习

引言

伴随着社交媒体的普及应用，谣言信息通过社交媒体的传播速度与传播范围在不断的扩大，对社会产生了极其严重的不良影响。传统的信息传播方式无法准确把握社交网络中的信息传播规律，尤其是虚假信息的广泛传播。近来，基于图形的学习方法——图神经网络（GNN）已经在社交网络方面进行了研究，尤其是在虚假信息传播预测方面具有了巨大的潜力。利用图神经网络可以学习社交网络中节点之间的连接关系以及信息的传播特点，进而准确地定位虚假信息的传播路径与传播影响范围。虽然基于GNN方面的研究取得了一定进展，但仍然有许多问题需要解决，比如数据不平衡性、标签噪声以及与时有关的困难等。本文主要介绍了基于图神经网络在虚假信息传播预测问题方面的研究，以及它的优势、实际应用以及未来的发展。

一、研究背景与意义

1. 图神经网络的研究进展

随着社交网络和互联网社区的兴起，信息传输方式及速度发生了翻天覆地的变化。在这样的背景中，假新闻的传播已经成为全球性的问题，给社会和人们的感情以及互联网安全带来了极大的影响。不过，传统数据统计方式和机器学习的策略在应对复杂、变动和大体量的社交网络数据时经常会出现诸如数据构型复杂度高、节点间的关联种类多以及信息转发机理不清楚等问题。针对此问题，GNN被引入进来，作为一种新兴的深度学习工具，GNN可以高效率模拟图数据、把握节点的关联程

度，因此对于社交网络信息流的分析，特别是假新闻的预测领域有着很广泛的适用性。

2. 研究目的

本文主要任务是探讨图神经网络在社交网络中对假消息的预测，揭示图神经网络的基本原理、优势以及问题，并结合证据证明了其在上述方面的优势。首先从基础阐述了图神经网络的基础知识，包括图形的数据表征方式、结点的嵌入方法及图形卷积网络等。然后侧重于介绍了图神经网络在社交网络中的应用情况，尤其是假消息传播这方面，接下来用实例和实验结果证明了图神经网络对解决假消息传播问题的强大潜力，并展望了图神经网络未来的趋势。

二、图神经网络概述

1. 图神经网络的基本原理

GNN是一种基于图形数据而开发的深度神经网络模型，与传统的数据类型比如序列或图像不同，它是由节点（点）和链接（边）构成的一个新型数据格式，主要用于描述对象之间关系或交互。GNN基于多级神经网络搭建，基于节点间关联信息不断更新节点表达，因此可以检测到图中的复杂结构和节点相关性信息。通俗来讲，GNN的基本思想是通过图形拓扑结构进行信息传输，并通过节点聚集过程获得每个节点的性质描述。

1.1 图数据表示与节点嵌入

对于图形数据的表示方式，在GNN当中，通常可以通过邻接矩阵或者是邻接列表来表示图形结构，并将每一条边连接，表示节点的关联。对于节点嵌入表示，可以将每一个节点转成低维度空间，如此图形结构以及节

点属性即可以被保持，目的是通过训练得到各节点的向量表示，这些向量表示可以呈现出节点属性，以及节点周围邻域节点的表示信息。在设计GNN模型的过程中，可以将节点的嵌入向量变化为邻域节点参数属性来完成优化，实现更加多样化的表征效果。

1.2 图卷积网络 (GCN) 与其变种

图卷积网络 (GCN) 是图神经网络的代表性方法之一，它是将卷积运算从传统的网格型数据 (如图像) 扩展到图形化数据上。GCN 通过将每个节点的属性及其周围的节点的属性加权融合来进行节点属性更新。确切而言，GCN 是利用节点的连通信息来完成特征传播，使得任一点的表示能包含与其有关的其他点的特征。其核心过程即图卷积步骤，即用加权平均的方式处理节点周围的特征以捕捉节点之间的结构信息。还有许多其他基于 GNN 的衍生算法如图注意力网络 (GAT)、图自动编码器 (GAE) 等。

2. 图神经网络在社交网络中的应用

2.1 社交网络的图结构特征

通常，Socialmedia 中的社交网络由其使用者结点以及用户节点之间的社会联系 (如朋友、粉丝或评分等) 构成。在这个社交网络中，每个节点都是社交媒体中的一个使用者，边表示两个使用者之间的关系交互。鉴于这个社交媒体网络的图形结构具有非常明显的社会特征，如小世界现象、社区结构以及度分布不平衡等。小世界现象就是大部分的使用者通过较少的中介人即可和网络中的其他使用者建立起联系，说明信息的传播通常是快速而有效的。而社区结构主要体现社交网络中的大部分使用者形成较紧密的社区并且这些用户社区中用户之间的交流通常比较高，社区之间的交互较小。

2.2 图神经网络对社交网络的适应性分析

将 GNN 应用到社交网络中的优势在于社交网络的图形特点很容易被 GNN 刻画，特别是社交网络中的节点间依赖关系对信息流动的影响，有效捕捉二者之间的关系，能够帮助我们理解社交网络的信息流；其次社交网络中的信息流动具有时空关联，传统图神经网络的方式也可以通过对 GNN 采用时间窗等手段获取流过程中的信息。但面对社交网络图的规模较大和数据流动的快速性问题，如何能够更加有效处理大规模图形的计算以及如何解决噪声和信息不平等仍是对于 GNN 的挑战。

三、虚假信息传播的特点与挑战

1. 虚假信息传播的机制

1.1 社交网络中的信息流动模型

在社交媒体中的假新闻传播，信息传播是重要的动力来源。通常而言，信息会在用户们的交流中被分享和传输，会形成一定的网状关系，每个结点代表的是一个人，而连接线上代表的是两个人之间的关系 (例如好友、关注、转发等)。但是，信息在传播过程中也会受到一定的影响，例如人们之间的社交连接，人们对于信息的理解程度，个人兴趣等。因此，社交媒体中的信息传播类型可以分为：扩散效应、感染效果以及流行病等。这些分类都是以图形学传播动力为基础，分析该如何将信息从一端的结点传递到另一端的结点。通俗地说，信息在传播的时候，很大程度上也是依靠各结点 (人) 之间的影响力与辐射力。

1.2 虚假信息的扩散特征

谣言的传播行为存在较为明显的非线性传播特征，其传播方式和传播速度等，都和其可靠性截然不同。首先，谣言往往在社交平台中呈现出选择性传播特征，即更容易被特定群体、社群或小圈子内的成员接收和转发，同时谣言的情感信息、情节片段、恐惧化要素会引起人们的心理反应而推进信息传播。此外，谣言往往表现出突然的传播特征。有时候谣言可能经过较少的关键节点后快速到达庞大的网络节点，类似流行病学传播的传播规律，在社交平台的环境中甚至会引起信息瞬间到达大量使用者，平台内的内容被虚假内容占满。

2. 虚假信息传播的挑战

2.1 数据不平衡与标签噪声

对于虚假信息传播最主要的挑战便是数据不平衡以及标签噪声。对于社交媒体环境来说，真实的言论会远大于虚假的言论，因此，在建立起辨识真假的机器学习模型过程中经常会出现数据不平衡的问题。数据不平衡会导致机器学习模型在对真实消息和假消息作出处理时不偏不倚，从而无法准确识别这些较少类别的信息 (也就是假消息)。此外，由于在社交媒体上获得的标注信息中往往会包含噪声。这是因为，社交媒体中的用户在上传自己的帖子时往往会受到诱导、夸张、误标的影响而误置标签。

2.2 网络异质性与多样性

因社交媒体的多样性和特殊性，使得其很难准确预测假消息的传播。社交媒体上人的行为、社交方式、网络结构极其复杂，而每个人又有着不同的社交需求和新闻偏好，比如年轻人和老年人在获取和传播新闻行为上的差异，又或有的人偏爱娱乐新闻而对政治新闻兴趣较

小。这样多样的特性使得假消息的传播并不仅依赖节点特征（个人兴趣爱好、人脉关系），还受到网络结构（人脉关系紧密程度、信息流通路等）的影响。

四、图神经网络在虚假信息传播预测中的应用

1. 图神经网络在虚假信息传播预测中的优势

1.1 充分利用社交网络图结构

GNN在假消息扩散预测方面优势明显，主要是基于社交网的图特征，人们在社交网络中并非仅由自身信息进行决策，还受到社交伙伴及其自身意见的影响，这种现象能体现在图形式中。GNN通过聚合相邻节点信息，有分析节点之间关系的能力，可把握每个节点在全球社交网络中的位置和功能。对假信息扩散预测而言，GNN可以区分哪些用户是信息流的关键节点，哪些信息会由关键节点及时扩散出去，通过深度分析社会网络的图结构，能够向节点间传播信息并准确推测假信息传输的路径和速度，有利于真伪鉴别和控制。

1.2 自动化特征学习与表示

传统的深度学习模式需要人类人工定义特征，而图神经网络可以自动将这个过程完成，简化了对人工特征的制定；GNN能够在无监督的情况下从图形数据中自动学习节点的低维表示（embedding），该表示包含了节点的属性和特性，位置和连接等等信息。假新闻预测不仅仅是单个节点的消息属性，还要考虑它们之间的连接强度、传播路线等等复杂的网络因素，由于自动特征学习技术GNN能够捕捉到并抓住假新闻传播过程中重要的特征，例如消息的传播路径、传播速度等，从而增强预测性能。

2. 应用案例分析

2.1 基于GCN的虚假信息传播预测模型

已有较多研究验证了采用图卷积网络（GCN）设计的假消息扩散预测架构具备可行性。GCN采用图形卷积操作提取社交网络中连接矩阵，得到每颗节点的表达形式。在该假消息扩散预测问题中，GCN能够动态更新各节点的特征，让其包含自身的特征同时包含来自其它节点的特征值。叠加上来的若干层GCN能够表示出复杂节点间的联系，从而预测假消息扩散的路径。实验证明了基于GCN的此方法能够准确识别假消息的传播路线，且达到了较高的准确率。

2.2 基于图卷积注意力网络（GAT）的虚假信息识别

在假新闻识别任务中，图神经网络GAT表现亮眼。相比于一般的GCN，GAT所使用的注意力机制赋予每个节点根据其邻域节点的重要性自主改变传播信息的权重值。具体地，GAT是学习每条边所附着的注意力权重来决定如何以邻近节点的重要性程度分配信息流动比例的大小。这对于假新闻识别很重要，因为在社交媒体上不同的用户由于对信息产生影响的强弱程度和可靠性的不同，可能存在一些核心的推手，可能就是假新闻的传播者。

结论

本文基于GNN模型在复杂网络社交平台上的应用情况，研究得出该方法在应对社交网络复杂网络数据时的效果。GNN可以有效还原社交媒体中各个节点间以及节点与边间的关系，并且可以主动探索消息转发的深层模式。这就为我们提供了另外一种在社交平台上预知防范新闻假消息的方案。然而在当前的GNN模型下，仍旧存在数据隐私以及数据实时更新和GNN模型可解释性等问题。在未来，相信随着计算机的发展，GNN将会结合多种数据融合、实时调整和多维度多领域联合应用等方式来提升假新闻的准确度和预知速度。

参考文献

- [1] 龚泽林, 郭彦宏. 图神经网络在社交网络信息传播预测中的应用研究[J]. 计算机科学, 2023 (12): 45-53.
- [2] 秦思涵, 曹昊天. 基于图卷积网络的虚假信息检测方法[J]. 软件学报, 2024 (3): 78-87.
- [3] 鲁明哲, 施紫莹. 社交网络中虚假信息传播模型与图神经网络分析[J]. 系统仿真学报, 2023 (10): 102-110.
- [4] 庞宜轩, 纪子轩. 图注意力网络在虚假信息识别中的应用[J]. 计算机应用研究, 2024 (5): 60-69.
- [5] 童景辰, 卢雨薇. 社交网络虚假信息传播预测的深度学习方法综述[J]. 电子学报, 2023 (11): 121-130.
- [6] 周峤, 林兴澎, 周赵斌, 许力. 在线社交网络中基于双向动态图注意力网络的异质图谣言检测方法[J]. 小型微型计算机系统, 2024, 45 (11): 2609-2617.
- [7] 邵云飞, 宋友, 王宝会. 基于社交网络图节点度的神经网络个性化传播算法研究[J]. 计算机科学, 2023, 50 (4): 6.