

崔子文,王欢,李函,等.基于图神经网络的植物间相互作用预测方法[J].华中农业大学学报,2025,44(2):301-310.
DOI:10.13300/j.cnki.hnlkxb.2025.02.031

基于图神经网络的植物间相互作用预测方法

崔子文,王欢,李函,韦乐

华中农业大学信息学院,武汉 430070

摘要 针对如何精准地根据先验知识预测植物间的抑制与促进相互作用问题,构建基于植物间相互作用的异质网络,提出一种基于图神经网络的相互作用预测方法。该方法主要由表征学习器、作用识别器和类型分类器三大模块构成。表征学习器负责提取植物间相互作用特征,作用识别器与表征学习器共同预测植物节点之间是否产生相互作用,并对类型分类器施加干扰,以最大程度地获得在不同类型相互作用间可迁移的特征。类型分类器旨在准确区分植物间相互作用的类型,以避免受到作用识别器的干扰。最后,基于三大模块之间的相互对抗关系来增强对植物间抑制与促进相互作用类型差异的鲁棒性,从而有效应对目标植物间相互作用的预测问题。该方法在伴生植物数据集上的 AUC、精确率(precision)和准确率(accuracy)的表现相较于现有的 SEAL、GATNE、HeGAN、PME、SVM 和 RF 方法中的最优方法 SVM 分别提高了 7.74、1.61 和 8.62 百分点,分别达到了 92.00%、80.12% 和 86.21%。结果表明,该方法通过降低目标类型差异的干扰,可以精准地预测植物间的相互作用,可以应用于优化农业生产实践。

关键词 植物间相互作用;鲁棒性;图神经网络;抑制与促进

中图分类号 TP18 **文献标识码** A **文章编号** 1000-2421(2025)02-0301-10

植物间相互作用描述的是在特定生态环境中 2 种或多种植物之间的互动关系^[1]。例如,黑桦植物会与其他低矮的植物,如蓝莓,形成共生关系,黑桦提供阴凉和保护,而蓝莓则通过其叶片分解提供养分^[2]。但农业环境中的互动并非总是正面的,如稗草释放的化感物质可能抑制水稻的生长^[3]。这些植物之间的互动平衡对生态稳定性、物种的适应能力以及资源的分布起到了至关重要的作用。植物间相互作用预测在农学中具有重要意义,它既可以为研究者提供洞察植物社区结构与动态的途径,同时又能为农业实践者提供更加高效和生态友好的田间管理策略。随着农业向现代化和精准化发展,精确地预测植物间的交互作用为新型种植策略和技术的出现提供了基础^[4],并在农业可持续发展中起到关键作用。

近年来,随着植物间相互作用预测研究的持续深入,深度学习技术已被认为是植物间相互作用预测研究的一种创新方法,并展现出广泛的应用前景。

深度学习,源于机器学习技术,其设计灵感来源于模拟人脑神经网络结构,旨在处理和分析庞大的数据,进而识别其内在的复杂模式和关系^[5]。在农业发展实际应用中,图神经网络(graph neural network, GNN)已成功应用于农作物生长环境的分析^[6-7]。在植物间的相互作用研究中,GNN 被用于建模和预测不同植物之间的复杂关系^[8]。此外,GNN 也被广泛应用于农业病虫害的监测与预测,通过对农作物与环境因素构建图结构,从而准确识别和预测潜在的病害爆发^[9]。基于 GNN 的模型还可以模拟不同作物在多样化种植条件下的生长表现,从而优化作物组合以实现最佳产量和健康状况^[10-11]。深度学习应用于植物间相互作用预测研究实际意义有以下几点。首先,该技术使研究者更深入地探索植物间复杂交互关系,并进一步解读其背后的生物规律与机制^[12-13]。其次,该技术还能指导农作物种植方案的优化,最大化农作物的产出^[14]。此外,该技术可以辅助减少对化学肥料和农药的依赖^[15],进而维护土壤

收稿日期:2024-08-19

基金项目:教育部“春晖计划”合作科研项目(202201700);“一带一路”创新人才交流外国专家项目(DL2023157004L);农作物育种数据融合共享与支撑体系建设项目(2023ZD0406101-3)

崔子文,E-mail:cuiziwen@webmail.hzau.edu.cn

通信作者:王欢,E-mail:hwang@mail.hzau.edu.cn

健康和生物多样性,推进农业生态的持续健康发展。

以上研究表明,深度学习缺乏对不同植物间相互作用的分析。因此,针对如何精准预测目标的类型植物间相互作用的问题,本研究从深度学习的角度提出了一种基于图神经网络的植物间相互作用预测方法(graph neural network-driven method for predicting plant-plant interactions, GNN-PPI)。该方法旨在为农业研究者在植物间的相互作用预测方面提供重要的参考依据,进而为农业的可持续发展提供新的动力。

1 预测方法及预测涉及到的数据集

1.1 模型总体概述

本研究收集了现实世界中植物间的相互作用信息,并将其建模为基于植物间相互作用的异质网络,其中节点代表植物实体,节点之间的链接代表植物间相互作用^[16]。相比传统的同质网络,这种基于植物间相互作用的异质网络融合了多种植物间的互动

模式。这种融合所带来的多元信息不仅强调了相互作用的多样性,还增加了互动过程中的不确定性和复杂性。为了制定一种能广泛应用于预测各种植物间相互作用的通用策略,本研究提出了GNN-PPI。该方法通过利用植物间相互作用类型的可转移性,为预测目标植物间相互作用提供了一种更为泛化的方法。如图1所示,为了实现这一目标,GNN-PPI中的表征学习器、作用识别器和类型分类器通过一种三者博弈的方式实现植物间相互作用预测。作用识别器与表征学习器协同工作,旨在学习在不同植物间相互作用类型中可迁移的特征表示,进而干扰类型分类器,而类型分类器不断提高区分植物间相互作用类型的能力。经过反复迭代,最终,GNN-PPI将表征学习器、作用识别器和类型分类器整合在一起,学习不同植物间相互作用类型之间的可迁移特征表示,以精准地预测目标植物间相互作用存在的可能性。

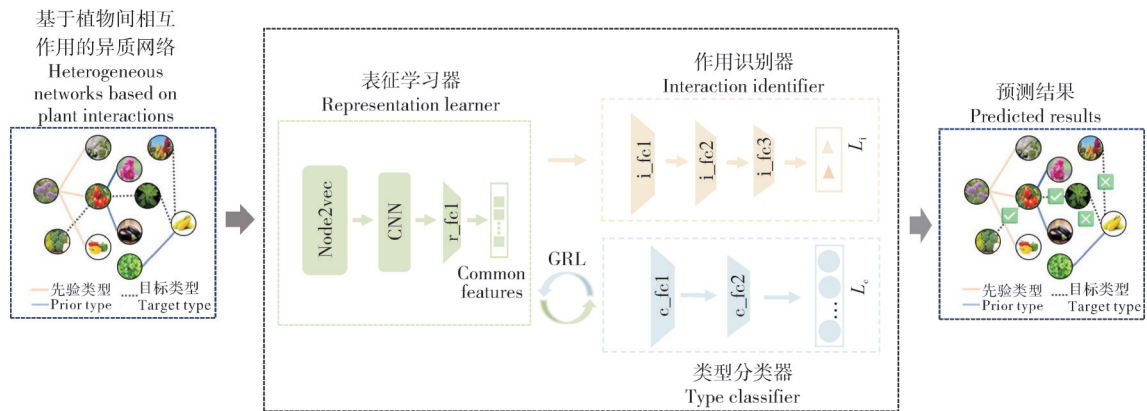


图1 基于图神经网络的植物间相互作用预测方法的整体架构图

Fig. 1 Overall architecture of a graph neural network-driven method for predicting plant-plant interactions

1.2 问题定义

本研究将植物间相互作用的异质网络形式化表示为 $G=(V, E, T)$ 。其中, V 表示植物节点集合, E 表示植物间相互作用的集合, T 为植物间相互作用的类型集合,包含促进、被促进和抑制3类。每条相互作用 $e \in E$ 对应1个类型 $T_e \in T$ 。本研究关注于植物间促进、被促进与抑制作用的预测问题,并将其定义如下:在给定的植物间相互作用的异质网络 $G=(V, E, T)$ 中,依据相互作用类型的不同,将植物间的促进作用和被促进作用划分为先验类型 T^h ,将抑制作用划分为目标类型 T^n 。本研究将利用先验类型 T^h 中的植物间相互作用数据进行模型训练,使其学习植物间相互作用的潜在模式,并利用训练好的模

型对目标类型 T^n 进行预测。

1.3 表征学习器

本研究将表征学习器表示为 $G_o(e; \theta_o)$,其中, θ_o 表示该模块需要优化的参数。表征学习器的核心目标是基于网络表征学习,提取植物节点间相互作用的特征表示^[17]。网络表征学习通过将异质网络中的植物节点的原始信息映射到低维向量空间,使其能够有效捕捉节点间的结构信息和语义关系^[18]。本研究采用Node2vec算法^[19]作为网络表征学习的具体实现方法,来提取植物间相互作用的特征表示,具体步骤如下。

首先,Node2vec通过有偏随机游走将基于植物间相互作用的异质网络 G 转换为一系列植物节点序

列。本研究将该随机游走过程定义为采样策略 S , 并用 $N_S(u) \subset V$ 表示植物节点 u 的网络邻域节点集合。受 Skip-gram 模型^[20]启发, 本研究利用这些节点序列作为模型输入, 以学习从植物节点原始信息到低维特征表示的映射函数 $f: V \rightarrow \mathbb{R}^d$ (其中, d 表示特征维数, \mathbb{R} 表示实数空间, \mathbb{R}^d 即表示 d 维的实数向量空间)。特征学习的优化目标基于最大似然估计, 定义如下:

$$\max_f \sum_{u \in V} \sum_{v \in N_S(u)} \log P_r(v|f(u)) \quad (1)$$

其中, $f(u)$ 为植物节点 u 的特征表示函数, $P_r(v|f(u))$ 代表在给定节点 u 的特征表示 $f(u)$ 时, 邻域节点 v 发生的共现概率:

$$P_r(v|f(u)) = \frac{\exp(f(u) \cdot f(v))}{\sum_{v' \in V} \exp(f(u) \cdot f(v'))} \quad (2)$$

节点 u 与其邻域中的节点 v 之间的相似度越高 (即 $f(u) \cdot f(v)$ 的值越大), 其共现概率 $P_r(v|f(u))$ 也就越大。通过优化该目标函数, 可得到一个能够最大化植物节点 u 在其网络邻域 $N_S(u)$ 内共现概率的映射函数 f 。最终, 使得具有相互作用的植物节点在低维向量空间中的特征表示更为接近, 从而提升模型对植物间相互作用的预测能力。

在获得节点特征表示后, 对于植物节点间的相互作用 $e=(u, v)$, 首先分别提取植物节点 u 和 v 的特征表示 $f(u)$ 和 $f(v)$ 。接着, 利用点积特征融合函数 $g(\bullet)$ 将每对植物节点的特征进行融合, 得到相互作用的特征表示。最终, 所有训练集 E_{train} 中植物节点间相互作用的特征集合表示为:

$$F = \{g(f(u), f(v)) | e=(u, v) \in E_{\text{train}}\} \quad (3)$$

基于上述特征集合 F , 本研究进一步引入一维卷积神经网络^[21]和全连接层, 以挖掘潜在可迁移的类型特征, 进而将已知的促进、被促进特征迁移至目标任务的抑制特征。在表征学习器、作用识别器和类型分类器三者之间的迭代优化过程中, F 不断被调整优化为可迁移的特征表示。

1.4 作用识别器

本研究将作用识别器表示为 $G_i(F; \theta_o, \theta_i)$, 其中, θ_i 为作用识别器需要优化的参数, F 和 θ_o 来自表征学习器。作用识别器的输入是上述表征学习器的输出特征 F , 其核心功能是判断两植物节点之间是否存在相互作用。对于给定的植物间相互作用 $e \in E_{\text{train}}$, 其特征表示为 F_e , 作用识别器 $G_i(F_e; \theta_o, \theta_i)$ 的输出即为植物节点之间产生相互作用 e 的存在似

然值。 $G_i(F_e; \theta_o, \theta_i)$ 的值越大, 说明给定的植物间相互作用 e 真实存在的概率越高。在测试集 E_{test} 中, 真实存在的植物间相互作用 e 的存在似然值应大于不存在的相互作用的似然值。

为了训练作用识别器, 使其能够有效地区分不同的植物间相互作用关系, 本研究使用交叉熵损失函数^[22]作为优化目标, 定义预测损失 (Loss_i , 公式中以 L_i 表示) 如下:

$$L_i(\theta_o, \theta_i) = -\mathbb{E}_{e \sim E_{\text{test}}} [m_e \log(G_i(F_e; \theta_o, \theta_i)) + (1 - m_e) \log(1 - G_i(F_e; \theta_o, \theta_i))] \quad (4)$$

其中, m_e 的值取 1 或者 0。当 $m_e = 1$ 时, 表示作用识别器对植物间相互作用 e 的预测正确, 即植物间存在相互作用 e ; 当 $m_e = 0$ 时, 则表示预测错误, 即植物间不存在相互作用 e 。在模型训练过程中, 通过优化参数 θ_o 和 θ_i , 最小化预测损失函数 $\text{Loss}_i(\theta_o, \theta_i)$ 。训练过程旨在提高模型在识别植物间相互作用方面的鲁棒性和判别能力^[23], 从而使其更适用于植物间相互作用的预测任务。

鉴于植物间相互作用类型的多样性, 本研究采用启发式学习策略, 旨在通过模型在先验类型 T^h 与目标类型 T^n 之间的有效迁移, 进一步实现其对未见类型相互作用的预测能力。这一策略已在社会网络链路预测研究中得到验证^[24]。为此, 本模型设计如下类型分类器。

1.5 类型分类器

本研究将类型分类器表示为 $G_c(F; \theta_o, \theta_c)$ 。其中, θ_c 为类型分类器的待学习参数, F 和 θ_o 来自表征学习器。类型分类器的核心任务是基于特征 F 判断植物间相互作用的类型。为了训练类型分类器, 使其能够有效评估植物间相互作用特征的可转移性, 本研究采用交叉熵损失函数作为优化目标, 定义类型分类损失 (Loss_c , 公式中以 L_c 表示) 如下:

$$L_c(\theta_o, \theta_c) = -\mathbb{E}_{e \sim E_{\text{train}}} [\sum_{t \in T^n} \alpha_e \log(G_c^t(F_e; \theta_o, \theta_c))] \quad (5)$$

其中, T^h 为训练集 E_{train} 中植物间促进、被促进相互作用的类型集合, $G_c^t(\{F_e\}; \theta_o, \theta_c)$ 表示类型分类器针对植物间相互作用 e 的类型预测正确的概率。当预测类别与真实类别一致时, $\alpha_e = 1$; 否则, $\alpha_e = 0$ 。较小的分类损失 $\text{Loss}_c(\theta_o, \theta_c)$ 表明模型对相互作用的类型分类性能较好。为了使 F 中学习到的特征表示具备可转移性, 训练过程中通过优化参数 θ_o 来最大化 $\text{Loss}_c(\theta_o, \theta_c)$, 确保 F 表示的特征满足可转移性要求, 能够适用于不同类型的植物间相互作用预测

任务。

1.6 模型的整体构建

在植物异质网络中植物间相互作用预测问题中,本研究将网络中存在的植物间相互作用 E 作为正样本,并随机选择相同类型且数量为 $|E|$ 的不存在的样本作为负样本,共同构成训练集 E_{train} ,并配对相应的标签集 Y_L 和 Z_L ,其中 Y_L 表示相互作用是否存在的标签, Z_L 表示其类型标签。模型通过表征学习器 $G_o(e; \theta_o)$ 、作用识别器 $G_i(F_i; \theta_o, \theta_i)$ 和类型分类器 $G_c(F_c; \theta_o, \theta_c)$ 之间的三者博弈进行训练。表征学习器和作用识别器会共同优化特征表示 F ,以使得 F 中的可迁移特征能够更好地捕获植物间不同类型相互作用的共性,从而干扰类型分类器的学习。而类型分类器则通过发现特定类型特征,精准地识别植物间相互作用的类型,避免被表征学习器和作用识别器干扰。基于这种对抗关系,本模型最终损失函数($\text{Loss}_{\text{final}}$,公式中以 L_{final} 表示)的定义如下:

$$L_{\text{final}}(\theta_o, \theta_i, \theta_c) = \lambda_i L_i(\theta_o, \theta_i) - \lambda_c L_c(\theta_o, \theta_c) \quad (6)$$

其中, λ_i 和 λ_c 是用于平衡预测损失 Loss_i 和类型分类损失 Loss_c 的超参数。通过最小化 $\text{Loss}_{\text{final}}$,模型能够在增强相互作用预测准确性的同时,确保学习到的特征对于未见相互作用类型具备良好的可转移性。

上述训练过程的目标是寻找模型最终损失函数的鞍点,参数更新过程如下:

$$(\hat{\theta}_o, \hat{\theta}_i) = \arg \min_{\theta_o, \theta_i} L_{\text{final}}(\theta_o, \theta_i, \hat{\theta}_c) \quad (7)$$

$$\hat{\theta}_c = \arg \max_{\theta_c} L_{\text{final}}(\hat{\theta}_o, \hat{\theta}_i, \theta_c) \quad (8)$$

其中,公式(7)表示在参数 θ_c 固定的条件下,找到使最终损失函数 $\text{Loss}_{\text{final}}$ 最小化的表征学习器和作用识别器的参数 θ_o, θ_i ;公式(8)表示在参数 θ_o 固定的条件下,找到使 $\text{Loss}_{\text{final}}$ 最大化的类型分类器的参数 θ_c 。 $\text{Loss}_{\text{final}}$ 的最大-最小优化问题即为鞍点问题。

本研究通过在表征学习器和类型分类器之间增加梯度反转层^[25],实现特征学习与类型判别之间的动态平衡。在前向传播阶段,梯度反转层作为恒等函数传递信号;而在反向传播过程中,梯度反转层将梯度信号反转,使得表征学习器学习到的特征能够最大程度地干扰类型分类器的分类能力,从而提升模型对未见类型相互作用的预测能力。

1.7 植物间相互作用预测涉及到的数据集

伴生植物数据集来源于维基百科页面“伴生植物列表”上的表格。数据集的格式考虑了基于图的

任务的应用,比如推荐系统中常用的链路预测^[26]和节点分类^[27]。数据集一共有996行和4列,每列分别是:源节点、链接、目标节点以及源节点类别。源节点表示链接开始的植物名称。数据集中的链接指明了链接的类型,包括促进、被促进和避免。目标节点是指链接结束处的植物名称。源节点类别分为蔬菜、香草、花卉和水果。尽管原始数据集中存在3种链接类型,但被促进只是促进的逆向关系(仅在逻辑上如此)。此外,某些目标节点可能在源节点中不存在。通过这个数据集,研究人员可以探索不同植物间相互作用关系。数据集提供了关于植物之间的链接类型和源节点属性的信息,可用于进行基于图的任务,如预测植物之间的链接关系或对植物进行分类。这些信息对于研究植物之间的相互作用、构建基于植物间相互作用的异质网络以及探索伴生种植的实践意义具有重要价值。该数据集的官方网址为: <https://www.kaggle.com/datasets/aramacus/companion-plants>。

本研究对植物间相互作用构建的异质网络进行了细致的统计分析,记录了异质网络中3种主要的相互作用类型——促进、被促进以及抑制类型,并详细统计了各类型链接的数量。3种类型的具体数量分别为:促进作用类型的链接452条、被促进作用类型的链接416条、抑制作用类型的链接127条。在实验设计中,促进作用及被促进作用的链接总计868条,被归类为训练集数据;而抑制作用的链接,总计127条,被选作测试集数据。原始的伴生植物数据是以植物名字命名的植物节点对,不能被直接用于构建基于植物间相互作用的异质网络。本研究对它们进行了3步预处理步骤。第1步,为了准确定位和识别各节点,本研究将植物节点的名称映射到其独特的序号;第2步,为了更准确地描述节点之间的相互关系,本研究以植物节点序号为基础建立了植物节点对之间的联系;第3步,为了理解植物间相互作用的结构,本研究根据节点对构建了基于植物间相互作用的异质网络。

1.8 基线方法

为客观地评价本研究提出的模型的优势,采用SEAL^[28]、GATNE^[29]、HeGAN^[30]、PME^[31]、SVM^[32]和RF^[33]6种模型作为对比方法,这些方法涵盖了基于机器学习的分类模型、网络结构的方法以及基于拓扑结构相似性的方法。

1) SEAL。SEAL是一种启发式的植物相互作

用预测算法,它利用图神经网络从局部子图中学习启发式信息,以自动学习适合当前网络的“启发式”。

2)GATNE。GATNE是一种针对基于植物间相互作用的异质网络的表示学习方法。它支持直推式和归纳式学习,可以适用于多种类型的数十亿植物节点和植物间相互作用组成的网络,每个植物节点关联不同的属性。

3)HeGAN。HeGAN是一种用于基于植物间相互作用的异质网络嵌入的新框架,它利用生成式对抗网络进行训练。在博弈过程中,鉴别器和生成器被同时训练,以提高植物节点嵌入的质量。

4)PME。PME是一种基于植物间相互作用的异质网络表征模型,旨在将基于植物间相互作用的异质网络嵌入到低维空间中,同时保留原始空间中的全局和局部网络结构。

5)SVM。SVM是常见的机器学习方法,特别适用于处理具有多个特征和属性的复杂数据集。在传统的SVM中,输入数据被表示为高维特征空间中的向量集。

6)RF。RF是一种决策树的集成模型,它的每棵树都是从输入数据的一个自助样本中构建出来的。RF的一个实用优点是,它可以自动预测植物间相互作用是否存在的概率。

1.9 参数设置

本研究中使用的模型在随机梯度下降中设置学习率 $\eta=0.001$ 。每次迭代使用32的批次大小。将所有方法的嵌入维数 d 设置为64。对于模型中基于随机游走的方法,将游走次数设置为 $w=10$,游走长度设置为 $l=5$,窗口大小设置为 $\tau=10$ 。将 λ 设置为1。为了更好地对比SEAL、GATNE、HeGAN、PME、SVM和RF这几种方法,SEAL、GATNE、HeGAN和PME在试验中遵照原来的方法^[28-33]设置,SVM采用线性核函数并设置惩罚系数为50,RF中设置 $n_estimators=50$ 。

1.10 试验设置

在进行试验设置时,从基于植物间相互作用的异质网络数据中选取了一部分植物间相互作用类型作为先验类型,而将剩余的植物间相互作用类型作为目标类型。具体而言,基于植物间相互作用的异质网络数据中包含了3种类型的植物间相互作用,即促进、被促进和抑制作用。为了将植物间相互作用名称数据化,本研究将这3种类型分别标记为类型0、1和2。试验分别以类型0、1和2轮流作为目标类型,

以研究不同目标类型对模型潜在的影响因素。

2 结果与分析

2.1 试验对比

为了验证GNN-PPI的预测性能,将GNN-PPI与其他6种方法进行对比。本研究使用3个典型的评价指标AUC^[34]、准确率(accuracy)^[35]和精确率(precision)^[36]来验证不同预测方法的性能。这3个指标分别从整体预测准确性、区分正负样本以及对正类样本识别能力的角度对模型的表现进行验证。GNN-PPI与其他6种方法对比得到的AUC、precision和accuracy如表1所示。

表1 GNN-PPI和6种对比方法的对比结果

Table 1 Comparison results of the GNN-PPI and six comparison methods			
方法 Methods	曲线下面积 AUC	精确率 Precision	准确率 Accuracy
GNN-PPI	92.00	80.12	86.21
SEAL	70.25	68.73	50.59
GATNE	57.14	62.63	58.62
HeGAN	62.46	70.61	72.35
PME	61.76	73.52	74.14
SVM	84.26	78.51	77.59
RF	69.87	70.62	66.38

从表1结果可知,与SEAL、GATNE、HeGAN、PME、SVM和RF相比,GNN-PPI在伴生植物数据集上获得的AUC、precision和accuracy最高,比次优的SVM方法分别提高7.74、1.61和8.62百分点,分别达到了92.00%、80.12%和86.21%。这是因为基于植物间相互作用的异质网络包含了“促进”“被促进”和“抑制”3种相互作用类型,而其他对比方法关注的是在训练集中先验类型的植物间相互作用预测损失最小化,忽略了植物间相互作用类型的不确定性。这直接导致了其他对比方法得到的结果相对较差。具体而言,其他6种方法倾向于捕捉特定类型的特征,无法很好地泛化到不同的植物间相互作用类型之间。此外,测试集中植物间相互作用类型的不确定性和多样性限制了对比方法在基于植物间相互作用的异质网络中的预测性能。相比之下GNN-PPI利用对抗训练学习可转移的植物间相互作用的特征,以抵抗测试集中植物间相互作用类型不确定性的干扰。因此,在基于植物间相互作用的异质网络中,GNN-PPI展现出实质性的预测改进效果。

2.2 消融实验

本研究提出的GNN-PPI模型在对抗学习过程中学习到不同类型植物间相互作用的可转移特征表示。为了使GNN-PPI模型能够学习可转移的特征表示,模型设计了一个类型分类器,用于判断提取的特征是否符合可转移特征的标准。通过最大化类型分类器的损失函数,模型有效地优化了提取特征的可转移性。然而,即使没有这种最大化过程,实验仍然可以根据植物间相互作用的初步特征表示来预测植物间相互作用。因此,为验证在训练阶段学习可转移特征表示的必要性,本研究设计了GNN-PPI的一个变体GNN-PPI⁻用于比较分析。GNN-PPI与GNN-PPI⁻的唯一区别在于,GNN-PPI⁻在植物间相互作用类型分类器中不考虑最大化过程。实验使用GNN-PPI和GNN-PPI⁻模型对基于植物间相互作用的异质网络进行相互作用预测以突出最大化类型分类器的损失函数的重要性。实验结果基于40次独立运行的平均值,GNN-PPI与GNN-PPI⁻的性能比较结果如图2所示。

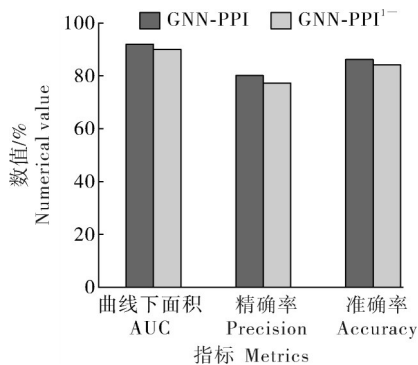


图2 GNN-PPI和GNN-PPI⁻在伴生植物数据集上的性能比较

Fig. 2 Performance comparison of GNN-PPI and GNN-PPI⁻ on the companion plant dataset

根据图2的结果,与GNN-PPI⁻相比,GNN-PPI的性能得到了稳定的提升。这有力地表明,学习可转移的特征表示对于提高基于植物间相互作用的异质网络中植物间相互作用的预测性能是非常重要的。在伴生植物数据集上,GNN-PPI在AUC、Precision和Accuracy上总是比GNN-PPI⁻得到更大的值。这种表现说明,基于表征学习器,作用识别器和类型分类器之间的三者博弈,GNN-PPI能够学习一般植物间相互作用的特征表示,这些特征可以从一种植物间相互作用类型转移到其他植物间相互作用类型。表征学习器与作用识别器协同工作以试图捕

捉植物间相互作用类型之间的可迁移特征以干扰类型分类器,而类型分类器试图区分植物间相互作用类型以避免被干扰。相比之下,由于缺少最大化过程,GNN-PPI⁻侧重于学习不可转移的类型特定特征,从而失去了在预测过程中抵抗植物间相互作用类型不确定性的泛化能力。因此,得益于通过对抗训练中不断对可转移特征表示的优化,GNN-PPI比GNN-PPI⁻具有更好的预测性能。

2.3 植物间相互作用类型试验

根据图3的结果,可以观察到目标类型的植物间相互作用数量与试验结果呈线性关系。结果显示,选择植物间相互作用数较少的类型(如抑制类型)作为预测对象可以获得更好的预测性能,而选择植物间相互作用数较多的类型(如促进类型)作为预测对象则会导致预测性能下降。这表明先验类型的植物间相互作用构成的网络结构对模型的预测性能有显著影响。这个观察结果可以引发更深入的分析。预测性能的差异可能与先验类型和目标类型的植物间相互作用之间的复杂关系有关。如果目标类型的植物间相互作用较少,模型能够更容易准确地预测,因为较少的植物间相互作用数量减少了网络中的噪声和干扰。相比之下,选择较多植物间相互作用的目标层作为预测对象可能增加了预测的困难,因为更多的植物间相互作用数量可能导致更多的变化和不确定性。

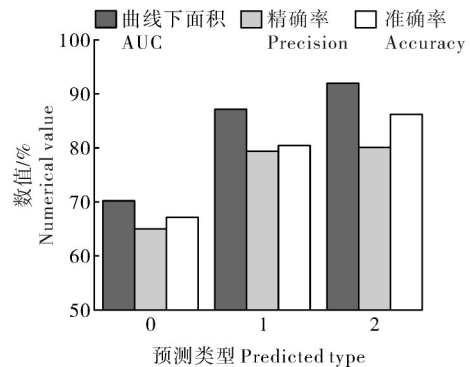


图3 不同类型植物间相互作用被选择为预测对象的试验结果

Fig. 3 The plant-plant interaction of different types was selected to predict the results of the experiment

2.4 结果验证

本试验主要对GNN-PPI预测的结果进行了验证。GNN-PPI预测的得分前十的植物节点之间相互作用的预测结果如表2所示。表2中的每一行显示植物节点对及其相应的模型预测分数 Predic-

tion score $\in[0,1]$ 。预测分数越接近于1表示本研究提出的模型越确认2种植物之间越会产生抑制作用。这说明本研究提出的GNN-PPI认为表2所展示的植物对确实存在抑制的相互作用。在使用原始数据库(<https://www.kaggle.com/datasets/aramacus/companion-plants>)验证结果后,预测结果的准确性得到了确认。

表2 GNN-PPI预测的植物之间产生相互作用的分数量表

Table 2 A score table of plant-plant interactions predicted by the GNN-PPI

植物 Plant	植物 Plant	预测分数 Prediction score
草莓 Strawberries	十字花科蔬菜 Brassicas	0.999 99
草莓 Strawberries	秋葵 Okra	0.999 99
草莓 Strawberries	茄子 Eggplant	0.999 79
百香果 Passion fruit	秋葵 Okra	0.999 78
豌豆 Peas	洋葱 Onion	0.999 70
草莓 Strawberries	胡椒 Peppers	0.999 66
草莓 Strawberries	马铃薯 Potatoes	0.999 63
大蒜 Garlic	南瓜 Squash	0.999 58
葡萄 Grapes	萝卜 Radishes	0.999 55
咖啡豆 Beans	洋葱 Onion	0.999 52

由表2可见,GNN-PPI预测结果得分排名第一的植物对为草莓(strawberries)和十字花科蔬菜(brassicas),得分0.999 99。即,草莓和十字花科蔬菜之间会产生抑制作用。文献[37]也证实了该预测结果。图4展示了这2种植物间产生抑制作用的示意图。

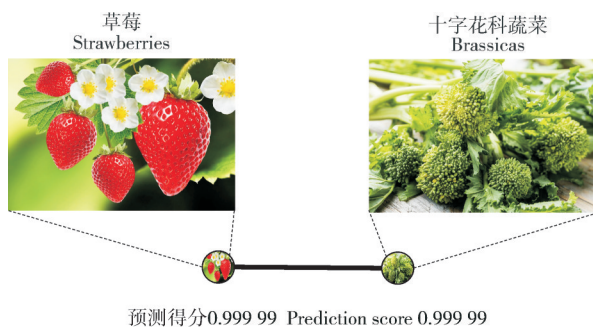


图4 GNN-PPI预测产生抑制作用的1对植物种类

Fig. 4 GNN-PPI predicts inhibitory interaction on 1 pair of plant species

3 讨论

本研究通过基于深度学习的方法对植物间相互作用的异质网络进行分析,以揭示其中的关联模式和规律,并为优化植物的配套种植方案提供指导和

决策支持。为了实现这一目标,本研究通过将现实中不同植物间的相互作用建模成基于植物间相互作用的异质网络,提出一种基于图神经网络的植物间相互作用预测方法(GNN-PPI)。本研究提出的GNN-PPI通过深度学习技术精准预测植物间的相互作用,从而优化配套种植方案并提升农业效率。该模型为实现精准农业提供了关键的数据支持,有助于实现农业资源的高效利用。

GNN-PPI能精准预测植物间存在抑制作用的原因有以下几个因素。首先,这2种植物能释放出具有抑制效果的化学物质,称为化感作用^[38]。例如,植物可能通过根系释放化合物,如苯酚类物质、挥发性有机物或化感物质,这些物质对周围植物的生长产生抑制作用^[39]。具体释放的化学物质类型和浓度因植物物种而异。其次,草莓和十字花科蔬菜在土壤中争夺相同的营养元素。由于这2种植物对某种特定营养元素的需求相似且生长速度相近,它们会相互竞争,导致生长抑制^[40]。这种竞争可能使其中一种植物无法获得足够的营养来支持其生长。最后,草莓和十字花科蔬菜都受到某些共同的病虫害问题的困扰。如果其中一种植物感染了病原体或受到虫害侵袭,它们成为传播病害或害虫的来源^[41],从而对周围的植物产生抑制作用。

此外,GNN-PPI方法在目标类型的植物间相互作用预测方面展示出了较好的效果。这一点对于农业研究具有实际意义。在农业生态系统中,对基于植物间相互作用的异质网络的深入理解和预测能力是至关重要的。例如,为了高效地利用资源,农民需要了解植物之间存在相互促进或抑制的关系。通过使用GNN-PPI,研究人员可以更准确地预测植物之间的相互作用,即使这些作用之前从未被观察到。这可以帮助农民优化种植策略,决定哪些植物应该种植在一起以获得最佳的生长效果或预防害虫。总的来说,GNN-PPI方法能够更好地理解和预测植物间的复杂相互作用,从而实现可持续的农业生产。

虽然当前的研究成果已经为这一领域带来了新的认识,但为了进一步提高预测精度并增加应用范围,以下几个方向值得探索。

1)交叉学科合作,推进生态农业的实践。GNN-PPI提供了对植物间相互作用的深度理解,这为生态农业提供了重要的理论基础。未来,可以与生态学、土壤学、农业经济学等相关学科合作,共同研究如何将GNN-PPI的预测结果与其他学科知识相结合,以

推动生态农业的全面实践。

2)传感器数据与模型整合的深度优化。随着现代传感技术在农业中的广泛应用,如土壤湿度、气象条件和植物生长状态的实时监测成为可能。未来的研究可以着重于如何将这些实时数据与GNN-PPI进行深度整合,实现对植物间相互作用的实时监控和预测。这种结合可以使得种植策略更加灵活,对环境变化有更快的响应。

3)模型的实际应用与农业决策整合。技术的最终目标应是转化为实际应用,为农业生产实践带来价值。未来的研究方向可以探索如何将本模型整合进实时的农业管理系统中,为农民提供即时、科学的决策建议。这不仅要求模型的预测结果简单明了、易于理解,还要求系统具有友好的用户界面,使得农民能够轻松地获取和使用。

参考文献 References

- [1] JACQUET F, JEUFFROY M H, JOUAN J, et al. Pesticide-free agriculture as a new paradigm for research[J/OL]. *Agronomy for sustainable development*, 2022, 42(1): 8 [2024-08-19]. <https://doi.org/10.1007/s13593-021-00742-8>.
- [2] DELUCA T H, ZACKRISSON O, NILSSON M C, et al. Quantifying nitrogen-fixation in feather moss carpets of boreal forests[J]. *Nature*, 2002, 419(6910): 917-920.
- [3] 金殿玉, 谢立勇, 赵洪亮, 等. 大气CO₂浓度升高条件下稻稗共生系统中稗草对水稻光合生理的影响[J]. *中国农业气象*, 2022, 43(3): 204-214. JIN D Y, XIE L Y, ZHAO H L, et al. Impacts of barnyard grass on photosynthesis and physiology of rice under elevated atmospheric CO₂ concentration[J]. *Chinese journal of agrometeorology*, 2022, 43(3): 204-214 (in Chinese with English abstract).
- [4] 王海光. 智慧植保及其发展建议[J]. *中国农业大学学报*, 2022, 27(10): 1-21. WANG H G. Smart phytoprotection and suggestions for its development[J]. *Journal of China Agricultural University*, 2022, 27(10): 1-21 (in Chinese with English abstract).
- [5] LIU W B, WANG Z D, LIU X H, et al. A survey of deep neural network architectures and their applications[J]. *Neurocomputing*, 2017, 234: 11-26.
- [6] LI P, HAO H, ZHANG Z, et al. A field study to estimate heavy metal concentrations in a soil-rice system: application of graph neural networks[J/OL]. *Science of the total environment*, 2022, 832: 155099 [2024-08-19]. <https://doi.org/10.1016/j.scitotenv.2022.155099>.
- [7] YANG F, ZHANG D, ZHANG Y, et al. Prediction of corn variety yield with attribute-missing data via graph neural network[J/OL]. *Computers and electronics in agriculture*, 2023, 211: 108046 [2024-08-19]. <https://doi.org/10.1016/j.compag.2023.108046>.
- [8] EL-SHAWA S. Exploring and predicting plant-arthropod interactions: hierarchical modelling of species communities and graph neural network approaches[D]. Guelph: University of Guelph, 2023.
- [9] KUSHARKI M B, MUHAMMAD-BELLO B L. AIoT-enabled precision agriculture for sustainable crop disease management: advancing SDGs through graph attention neural networks[M]//MOHIUDDIN A, ABU S M, RAJU C, et al. *Artificial intelligence of things for achieving sustainable development goals*. Cham: Springer Nature Switzerland, 2024: 143-160.
- [10] LEI L, YANG Q, YANG L, et al. Deep learning implementation of image segmentation in agricultural applications: a comprehensive review[J/OL]. *Artificial intelligence review*, 2024, 57(6): 149 [2024-08-19]. <https://doi.org/10.1007/s10462-024-10775-6>.
- [11] FANG Y. Smart agriculture: advancements on plant disease detection with big data and deep learning[D]. Melbourne: Deakin University, 2022.
- [12] SERIN E A R, NIJVEEN H, HILHORST H W M, et al. Learning from co-expression networks: possibilities and challenges[J/OL]. *Frontiers in plant science*, 2016, 7: 444 [2024-08-19]. <https://doi.org/10.3389/fpls.2016.00444>.
- [13] DENG Z Y, ZHANG J M, LI J Y, et al. Application of deep learning in plant-microbiota association analysis[J/OL]. *Frontiers in genetics*, 2021, 12: 697090 [2024-08-19]. <https://doi.org/10.3389/fgene.2021.697090>.
- [14] AYOUB SHAIKH T, RASOOL T, RASHEED LONE F. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming[J/OL]. *Computers and electronics in agriculture*, 2022, 198: 107119 [2024-08-19]. <https://doi.org/10.1016/j.compag.2022.107119>.
- [15] SHAH T M, NASIKA D P B, OTTERPOHL R. Plant and weed identifier robot as an agroecological tool using artificial neural networks for image identification[J/OL]. *Agriculture*, 2021, 11(3): 222 [2024-08-19]. <https://doi.org/10.3390/agriculture11030222>.
- [16] LOSAPIO G, MONTESINOS-NAVARRO A, SAIZ H. Perspectives for ecological networks in plant ecology[J]. *Plant ecology & diversity*, 2019, 12(2): 87-102.
- [17] LI B T, PI D C. Network representation learning: a systematic literature review[J]. *Neural computing and applications*, 2020, 32(21): 16647-16679.
- [18] 侯祥英, 崔运鹏, 刘娟. 深度学习在植物基因组学与应用育种中的应用现状与展望[J]. *农业图书情报学报*, 2022, 34(8): 4-18. HOU X Y, CUI Y P, LIU J. Applications and prospect analysis of deep learning in plant genomics and crop breeding[J]. *Journal of library and information science in agriculture*,

- 2022, 34(8):4-18 (in Chinese with English abstract).
- [19] GROVER A, LESKOVEC J. Node2vec: scalable feature learning for networks[J]. KDD: proceedings international conference on knowledge discovery & data mining, 2016, 2016: 855-864.
- [20] PENG H, LI J X, YAN H, et al. Dynamic network embedding via incremental skip-gram with negative sampling[J/OL]. Science China information sciences, 2020, 63(10):202103[2024-08-19]. <https://doi.org/10.1007/s11432-018-9943-9>.
- [21] O'SHEA K, NASH R. An introduction to convolutional neural networks [DB/OL]. arXiv, 2015: 1511.08458 [2024-08-19]. <https://doi.org/10.48550/arXiv.1511.08458>.
- [22] LI L, DOROSLOVAČKI M, LOEW M H. Approximating the gradient of cross-entropy loss function[J]. IEEE access, 2020, 8:111626-111635.
- [23] BARRON J T. A general and adaptive robust loss function [C]//LARRY D, PHILIP T, ZHU S C. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach, CA, USA: IEEE, 2019:4331-4339.
- [24] WANG H, CUI Z W, LIU R G, et al. A multi-type transferable method for missing link prediction in heterogeneous social networks[J]. IEEE transactions on knowledge and data engineering, 2023, 35(11):10981-10991.
- [25] HE Z W, ZHANG L. Multi-adversarial faster-RCNN for unrestricted object detection[C]//JINWOO S, SANG U L, IN S K. 2019 IEEE/CVF international conference on computer vision (ICCV). Seoul, Korea (South): IEEE, 2019:6667-6676.
- [26] AI J, LIU Y Y, SU Z, et al. Link prediction in recommender systems based on multi-factor network modeling and community detection [J/OL]. EPL (europhysics letters), 2019, 126(3): 38003 [2024-08-19]. <https://doi.org/10.1209/0295-5075/126/38003>.
- [27] WU Q, ZHAO W, LI Z, et al. Nodeformer: a scalable graph structure learning transformer for node classification [J]. Advances in neural information processing systems, 2022, 35: 27387-27401.
- [28] ZHANG M, CHEN Y. Link prediction based on graph neural networks [J]. Advances in neural information processing systems, 2018: 31 [2024-08-19]. <https://dl.acm.org/doi/10.5555/3327345.3327423>.
- [29] CEN Y K, ZOU X, ZHANG J W, et al. Representation learning for attributed multiplex heterogeneous network [C]//ANKUR T, VIPIN K, LIRÓMER R, et al. Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. Anchorage AK USA: ACM, 2019: 1358-1368.
- [30] HU B B, FANG Y, SHI C. Adversarial learning on heterogeneous information networks [C]//ANKUR T, VIPIN K, LIRÓMER R, et al. Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. Anchorage AK USA: ACM, 2019:120-129.
- [31] CHEN H X, YIN H Z, WANG W Q, et al. PME: projected metric embedding on heterogeneous networks for link prediction [C]//Proceedings of the 24th ACM SIGKDD International conference on knowledge discovery & data mining. London: ACM, 2018:1177-1186.
- [32] NOBLE W S. What is a support vector machine? [J]. Nature biotechnology, 2006, 24(12):1565-1567.
- [33] BIAU G, SCORNET E. A random forest guided tour [J]. Test, 2016, 25:197-227.
- [34] ZHAO Z L, GOU Z Y, DU Y H, et al. A novel link prediction algorithm based on inductive matrix completion [J/OL]. Expert systems with applications, 2022, 188: 116033 [2024-08-19]. <https://doi.org/10.1016/j.eswa.2021.116033>.
- [35] NASIRI E, BERAHMAND K, LI Y F. Robust graph regularization nonnegative matrix factorization for link prediction in attributed networks [J]. Multimedia tools and applications, 2023, 82(3):3745-3768.
- [36] SHABAZ M, GARG U. Predicting future diseases based on existing health status using link prediction [J]. World journal of engineering, 2022, 19(1):29-32.
- [37] UGOLINI L, MARTINI C, LAZZERI L, et al. Control of postharvest grey mould on strawberries by glucosinolate-derived allyl-isothiocyanate treatments [J]. Postharvest biology and technology, 2014, 90:34-39.
- [38] MACÍAS F A, DURÁN A G, MOLINILLO J M G. Allelopathy: the chemical language of plants [J]. Progress in the chemistry of organic natural products, 2020, 112:1-84.
- [39] XIE Y Q, TIAN L B, HAN X, et al. Research advances in allelopathy of volatile organic compounds (VOCs) of plants [J/OL]. Horticulturae, 2021, 7(9): 278 [2024-08-19]. <https://doi.org/10.3390/horticulturae7090278>.
- [40] ZHANG D Q, YAN D D, CHENG H Y, et al. Effects of multi-year biofumigation on soil bacterial and fungal communities and strawberry yield [J/OL]. Environmental pollution, 2020, 256: 113415 [2024-08-19]. <https://doi.org/10.1016/j.envpol.2019.113415>.
- [41] KAZAK C, KARUT K, DÖKER İ. Important pests in vegetable crops [M]//HUI Y H. Handbook of vegetable preservation and processing, Boca Raton: CRC Press, 2015:134-157.

A graph neural network-based method for predicting interactions between plants

CUI Ziwen, WANG Huan, LI Han, WEI Le

College of Informatics, Huazhong Agricultural University, Wuhan 430070, China

Abstract A heterogeneous network-based on interactions between plants was constructed, and a graph neural network-based method for predicting interactions between plants was proposed to solve the problem of accurately predicting inhibitory and promotive interactions between plants based on prior knowledge. The method is primarily composed of three modules including a representation learner, an interaction identifier, and a type classifier. The representation learner is responsible for extracting representations of interactions. The interaction identifier collaborates with the representation learner to predict whether an interaction exists on types of targeted plant and applies perturbations to the type classifier to maximally acquire transferable features across different types of interaction. The type classifier is designed to accurately distinguish types of interactions between plants to avoid being affected by the interaction identifier. The inter adversarial relationship among the three modules is leveraged to enhance robustness against the differences in the types of inhibitory and promotive interactions between plants, thereby effectively solving the problem of accurately predicting targeted interactions between plants. The AUC, precision, and accuracy of this method on the dataset of companion plant improved by 7.74, 1.61, and 8.62 per cent, respectively compared with the best method SVM that of existing methods including SEAL, GATNE, HeGAN, PME, SVM, and RF, reaching 92.00%, 80.12%, and 86.21%, respectively. The results indicate that the proposed method effectively mitigates the interference caused by target interaction type differences, enabling accurate prediction of plant interactions. This approach can be applied to optimize agricultural production practices.

Keywords interactions between plants; robustness; graph neural networks; inhibition and promotion

(责任编辑:边书京)