

科技新闻

·卓越亮点·

负样本也能教AI找材料

近年来,机器学习正在改变材料研发方式。模型可以从已有材料的组成、结构和性能数据中学习规律,再预测未知材料的性质,从而加快候选材料筛选。但材料数据并不像理想教材那样均匀分布:高性能材料更容易被报道,失败实验和低性能材料往往被忽略;同时,实验和计算成本又使可用数据长期处于“小样本”状态。对于AI模型来说,问题不只是“数据够不够多”,还包括“数据分布是否合理”。如果训练集只集中在少数高性能区域,模型可能会误以为许多未知材料都值得期待;如果只做随机均匀采样,又可能漏掉真正决定性能边界的关键区域。

中国科学院长春应用化学研究所彭博、王大鹏团队围绕“数据分布如何影响模型泛化能力”这一问题,提出并验证了有意识地协同利用正、负样本数据的策略。研究团队先构建了6个10维高维基准函数,用来模拟材料性质空间中常见的复杂形貌,包括单峰、多峰、周期性起伏和尖锐局部极值等情形。随后,团队比较了7种采样策略,包括均匀随机采样、马尔可夫链蒙特卡洛采样等自适应采样(只采取最好的区域)方法。结果显示,传统上被认为可靠的“全空间均匀覆盖”并不总是最优;真正有助于模型泛化的训练集,既要覆盖整体空间,也要有

意识地包含正、负2个极值区域的信息。

这意味着,“失败数据”并不是噪声或负担,而是帮助模型理解材料性能边界的参照物。正样本告诉模型“什么样的材料可能好”,负样本则告诉模型“什么样的材料不该被高估”。二者共同出现时,模型更容易划清高性能区、普通区和低性能区之间的边界,从而在未知材料上给出更稳健的预测。

(来源于中国科技期刊卓越行动计划入选期刊:《物理化学学报》, 2026,42(X):100291)

植物借“蛋白招募”重塑根系,实现盐碱突围

土壤盐渍化是限制农林业生产的主要非生物胁迫之一。为在盐碱环境中生存,植物需要对根系构型进行适应性重塑,以便更有效地获取水分和养分。在这一过程中,植物激素细胞分裂素(CK)发挥着核心调控作用。研究发现,A类响应因子(Type-ARRs)是调控CK信号的关键负调控元件,与植物的耐盐性密切相关。然而,A类响应因子自身缺乏DNA结合结构域,它们究竟是如何跨越这一结构障碍,实现对下游基因的精准转录调控尚不清楚。

东北林业大学杨静莉团队报

道了杨树(*Populus ussuriensis*)中A类响应因子PuRR9通过磷酸化介导的“蛋白招募”策略,成功抑制CK生物合成,进而调控根系可塑性并增强植物耐盐性的全新路径。

盐胁迫会促使PuRR9发生磷酸化修饰。磷酸化后的PuRR9巧妙地充当了“招募者”,增强与转录抑制因子PuZFP1的相互作用,将PuZFP1精准锚定至细胞分裂素合成关键基因PuIPT3的启动子区域。该复合物增强了PuZFP1对PuIPT3的转录抑制,导致根部细胞分裂素合成显著减少。这种激素水平的下降有效解除了对根系生长的束缚,促使根系的伸长,从而显著增强了杨树的耐盐能力。

该研究创制了PuRR9新种质,为盐碱地造林及退化生态系统恢复提供优良的先锋树种材料储备。明确CK依赖于磷酸化的负反馈调节环路的耐盐调控网络,为培育抗盐林木新品种提供了重要的理论支撑与育种靶点。

(来源于中国科技期刊卓越行动计划入选期刊:《Journal of Integrative Plant Biology》, 2026,10.1111/jipb.70290)



新疆哈密伊吾县树龄约3000年的原始胡杨林

(图片来源:国家林业和草原局)