

B 题：药物属性预测

近年来，随着网络技术的快速发展和大数据挖掘技术的成熟，人们的数据分析能力也在逐步提升，可以采集的数据规模越来越大。尤其是伴随着电商和短视频媒体的发展，产生了大量的图结构数据。图结构数据的研究非常有价值，由于图结构数据的复杂性，使得这方面的研究工作十分具有挑战性。

现实世界中有很多问题都可以使用图来建模，图数据是目前模式识别与机器学习领域重要的研究对象。例如，网上购物软件采用基于图深度学习的方法可以精准地向用户推荐喜欢的商品，图在推荐系统上的建模能力比较强；在生物医疗上，可以图深度学习技术设计新药物或者探究药物间的相互作用；而在引文网络中，论文通过引用关系被相互连接，并可通过分析这些关系将它们分组，正如图 1 所示。这些例子展示了基于图的学习系统在不同领域的应用潜力和价值。图是一种功能强大的结构，可以用来建模几乎任何类型的数据。社交网络、文本文档、万维网、化学化合物和蛋白质-蛋白质相互作用网络，通常都是用图表表示的数据的例子^[2]。由于图形结构的丰富数据，图上的机器学习最近成为一项非常重要的任务。近年来，越来越多的学者关注图表示学习的研究工作，图表示学习主要应用在图分类、节点分类和链路预测等任务中。

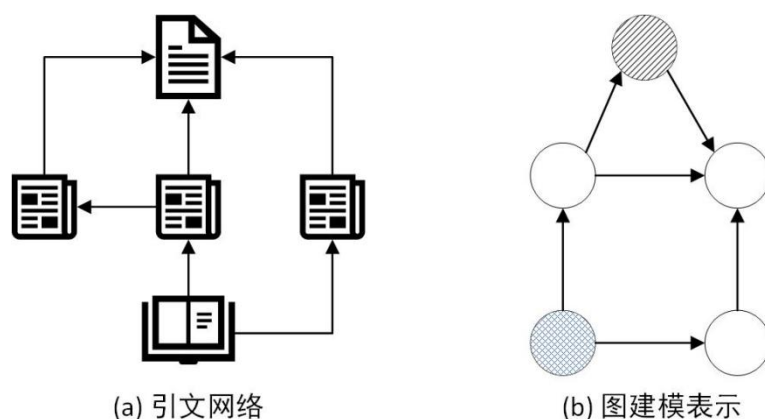


图 1 引文网络的建模示例图

1. 附件是药物分子的数据（图数据），请您利用传统方法建立药物分子的分类模型，并给出分类精度及其结果分析。

2. 传统药物分子分类方法依赖于复杂的化学属性分析和生物实验，不仅耗时耗力，而且难以处理大规模的分子数据。因此，发展一种高效、准确的分子分类方法成为了当前科研的一个热点。与此同时，一些研究人员将神经网络应用到药物分子挖掘中，提出图神经网络，这种方法能够端到端进行模型的优化学习，在图分类准确度有较大提升。请您给出一种图神经网络模型对附件中的数据进行分类，并给出分类精度及其结果分析。

3. 现有图神经网络模型在处理具有节点特征稀疏性和信息冗余的图结构数据时面临挑战，这限制了模型在复杂网络分析中的应用效果。请您尝试给出一种新的药物分子分类方法突破这种限制，给出试验结果，并进行分析讨论。

附件：

数据集是 MMM_data 数据集是图结构数据集，包含 188 个化合物分子，并有两个类别，分别代表这个分子是诱变芳香族或杂芳香族。

数据是 Matlab 格式，其中，am 是每个图的邻接矩阵，al 是每个节点的标签，nl 和 el 是图的其他标签。