基于因果机制的分子属性预测

蔡瑞初[1]，许遵鸿[1]，陈道鑫[1]，杨振辉[1]，李梓健[1]，郝志峰[2]
（1.广东工业大学计算机学院，广东 广州 510006；2.汕头大学工学院，广东 汕头 515063）

摘 要：在量子化学领域，分子性质预测是一项基础而关键的任务，广泛应用于药物发现、化学合成预测等多个领域。随着人工智能的发展，深度学习方法在该领域得到了广泛应用。然而，当前的方法往往采用微观和宏观视图两种极端的抽象层次来对分子性质进行建模，导致难以推广到分布之外样本的挑战。化学的介观视图提供了一个有益的中间层次，通过包含与性质相关的功能基团的介观成分来描述分子性质。通过考虑这些介观成分，并从因果关系的角度对其进行建模，可以更加关注与性质相关的功能基团。为了实现该目标，提出一种介观成分识别模型。该模型基于分子数据的介观因果生成过程和变分自编码器的框架，通过学习与分子性质相关的介观成分的表示，实现对分子性质的预测。首先假设原子隐变量遵循高斯分布和语义隐子结构遵循多元伯努利分布，将分子数据输入神经网络来识别原子隐变量和语义隐子结构。接着利用识别出来的原子隐变量和语义隐子结构来预测分子性质。为了能够识别出原子隐变量和语义隐子结构，利用变分下界和稀疏项来构造模型的损失函数。实验结果表明，该模型不仅在性能上取得先进的结果，而且提供了深入的解释，为模型预测提供了更全面的理解，提高分子性质预测的准确性和泛化能力。

关键词：分子属性预测；因果；分布外泛化；图表征；图神经网络
中图分类号：TP391 **文献标志码：**A **DOI：**10.19678/j.issn.1000-3428.0068937

Causal Mechanism-Based Molecular Property Prediction

CAI Ruichu[1], XU Zunhong[1], CHEN Daoxin[1], YANG Zhenhui[1], LI Zijian[1], HAO Zhifeng[2]
(1. School of Computer, Guangdong University of Technology, Guangzhou 510006, Guangdong, China;
2. College of Engineering, Shantou University, Shantou 515063, Guangdong, China)

【Abstract】In the field of quantum chemistry, molecular property prediction is a fundamental and critical task, which is widely used in many fields such as drug discovery and chemical synthesis prediction. With the development of artificial intelligence, deep learning methods have been widely used in this field. However, current methods often adopt two extreme levels of abstraction, namely micro- and macro-views, to model molecular properties, posing challenges in generalizing to out-of-distribution samples. The mesoscopic view of chemistry provides a beneficial intermediate level for describing molecular properties through mesoscopic components containing functional groups associated with these properties. By considering these mesoscopic components and modeling them from a causal perspective, more attention can be paid to the functional groups related to these properties. To achieve this goal, this study proposes a Mesoscopic Component Identification(MCI) model. This model is based on a mesoscopic causal generative process that uses molecular data and a framework of variational autoencoders. The proposed model predicts molecular properties by learning the representation of mesoscopic components related to molecular properties. Initially, the model assumes that the atomic latent variables and semantic latent substructure follow Gaussian and multivariate Bernoulli distributions, respectively. Molecular data are then input into a neural network to identify the atomic latent variables and semantic latent substructure. Next, the identified atomic latent variables and semantic latent substructures are used to predict molecular properties. To identify the substructures of the atomic and semantic latent variables, variational lower bounds and sparse terms are used to construct the loss function of the model. Experiments demonstrate that our model not only achieves state-of-the-art performance but also offers in-depth explanations that provide a more comprehensive understanding of model predictions and improve the accuracy and generalization ability of molecular property predictions.

【Key words】molecular property prediction; causal; out-of-distribution generalization; graph representation; Graph Neural Network(GNN)

收稿日期：2023-12-01 **修回日期：**2024-02-01
基金项目：科技创新 2030-“新一代人工智能”重大项目(2021ZD0111501)；国家优秀青年科学基金 (62122022)；国家自然科学基金(61876043, 61976052, 62206061)。
通信作者 E-mail：cairuichu@gmail.com

0 引言

人工智能通过高效的计算算法和富有洞察力的实验结果，在科学领域展现出了深远的影响[1-3]。其中，分子性质预测在药物发现[4-6]、化学合成[7-8]等多个领域都取得了开创性的应用。由于利用密度泛函理论(DFT)[9]进行分子预测需要大量计算资源，一些基于图神经网络(GNN)[10-11]的方法被提出，将分子视为图结构化数据。这些方法可以根据不同的抽象尺度被分类为微观或宏观方法。

在微观角度上，研究人员通过利用低阶信息进行深入探索。例如，MGCN[12]模拟了分子性质预测中来自不同角度的局部相互作用，为理解分子内部结构提供了关键见解。另一方面，GeomGCL[13]采用了分子的几何结构，例如角度和距离，作为学习分子表示的重要因素。而在宏观角度上，研究人员则致力于利用高阶结构信息。HU 等[14]通过引入监督和结构相似性的限制，成功地获得了稳健的图嵌入，进一步拓展了对分子图的抽象表示。此外，RONG 等[15]进行进一步研究，考虑图级别的自监督任务，为图神经网络的发展提供了更为全面的视角。

尽管这些方法取得了出色的性能，但因为分别在两个极端的抽象尺度上考虑，当测试分布与训练分布发生未知变化时，对分布外(OOD)泛化的能力不能得到保障。

化学中的介观视角[16]揭示了分子性质是由与性质相关的功能团组成的介观成分描述的。受此启发，本文提出一个因果生成过程，利用与性质相关的介观成分进行稳健的分子性质预测。在此基础上，通过提取介观成分增强分布外泛化能力，并基于随机变分推断提出了介观成分识别(MCI)模型。

1 相关工作

本节主要关注分子性质预测和图分类中已有的技术。

1.1 分子性质预测

分子性质预测[17-19]是物理和化学领域中的一个重要研究问题[7]。密度泛函理论[9]是最重要的方法之一，但存在计算耗时的问题。近年来，由于图神经网络[20]在结构数据上取得了成功，因此也被应用来对分子进行建模[21]。这些方法可以根据不同的抽象尺度分为微观方法和宏观方法。微观方法主要利用低阶属性，如原子或化学键。ROGERS 等[22]从分子描述符或化学指纹中学习表示。LU 等[12]提出了 MGCN 方法，该方法利用原子、成对和三重交互作用来学习分子表示。考虑到单向消息传递可能导致表示不足，SONG 等[23]提出了交际消息传递神经网络。最近，LI 等[13]提出了 GeomGCL，以节点-边交互方式学习带有几何信息的分子表示。宏观方法主要利用高阶信息，如 ZHANG 等[24]考虑到分子的分层结构，开发了面向片段的多尺度图注意力网络用于分子性质预测。为了获得高质量的嵌入，HU 等[14]引入了图级监督和结构相似性限制来预训练图神经网络，从而提高了分子性质预测的性能。RONG 等[15]将图级信息整合到基于 Transformer 的框架中。然而，由于过于极端的抽象尺度，这些方法可能会受到分布偏移的影响。此外，研究人员通过预定义的图案[25]与领域知识[26]来运用与性质相关的亚结构，但它们通常需要大量手动标记，并且很难推广到具有未知图案或功能基团的分子。本文采用中间的介观视角，并提取具有理论保证的与性质相关的功能基团。

1.2 非独立同分布的图分类

本文的工作也与非独立同分布的图分类问题[27-28]相关。现有的非独立同分布[29]研究主要集中在计算机视觉[30]和自然语言处理领域[31]，但对于图结构数据的非独立同分布的挑战尚未得到充分探索，并且越来越受到关注。由于现有的图神经网络[32]缺乏分布外的泛化能力，导致性能不佳，LI 等[33]提出的 OOD-GNN 方法通过消除相关和无关图表示之间的统计依赖关系来解决非独立同分布的挑战。由于虚假相关性导致图神经网络泛化能力差，FAN 等[34]提出了 StableGNN 方法，通过稳定学习来提取图神经网络的因果表示。WU 等[35]提出的 DIR 策略通过干预来提取不变的因果理由，从而减轻图结构数据背后的选择偏差。

分子性质预测可以被视为图的非独立同分布泛化问题的特例。本文通过介观因果生成过程来解决这个问题，与化学中的介观视角相一致。

2 分子数据的介观因果生成过程

微观和宏观方法无法保障分布外(OOD)泛化的能力。图 1 展示了一个分子性质二元预测任务(彩色效果见《计算机工程》官网 HTML 版，下同)，其中正性属性由“酰胺”(-CONH2)官能团决定。如图 1 的第 1 行所示，从微观角度看，现有方法大多只考虑了原子之间的化学键结构而忽略了局部分子结构，导致了模型容易产生假阳性结果。如图 1 的第 3 行所示，从宏观角度看，现有的宏观方法主要关注分子全局结构，因为全局分子结构中容易包含和类

别不相关的子结构，例如图 1 的第 3 行中的苯环所示，所以这类方法容易产生假阴性的结果。

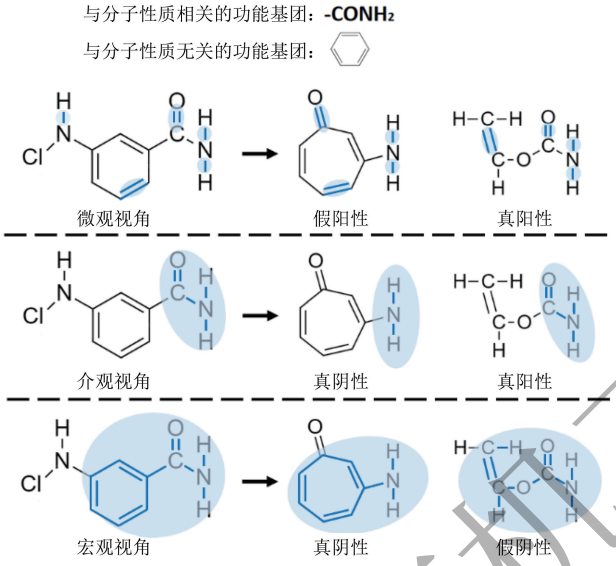

图 1 分子预测的示例图

Fig. 1 Example diagram of molecular prediction

在量子化学领域，分子性质可以由介观成分描述，如图 1 的第 2 行所示，通过介观分析可以提取出和语义相关的分子子结构，从而使得分子预测模型更加鲁棒。

本文提出了一种通过介观视角连接宏观和微观视角的介观因果生成过程，如图 2 所示。从微观视图到介观视图，首先让 $\boldsymbol{x}$ 表示观察到的原子属性，比如原子手性、形式电荷以及原子是否在环中等。进一步让 $\boldsymbol{G}_n$、$\boldsymbol{G}_h$ 表示噪声子结构隐变量（即与性质无关的功能基团的邻接矩阵）和语义子结构隐变量（即与性质相关的功能基团的邻接矩阵）。$\boldsymbol{z}$ 表示编码低阶信息的原子隐变量。让 $\boldsymbol{x}\rightarrow\boldsymbol{G}_n$、$\boldsymbol{G}_h$、$\boldsymbol{z}$ 表示 $\boldsymbol{G}_n$、$\boldsymbol{G}_h$、$\boldsymbol{z}$ 如何从 $\boldsymbol{x}$ 中解耦出来的过程。具体来说，给定函数 f_z、f_h、f_n，分别使 $\boldsymbol{z}=f_z(\boldsymbol{x})$，$\boldsymbol{G}_h=f_h(\boldsymbol{x})$，$\boldsymbol{G}_n=f_n(\boldsymbol{x})$。从介观视图到宏观视图，使 $\boldsymbol{A}$ 和 y 分别表示观察到的分子结构和分子性质。原子序数 $\boldsymbol{k}$ 是 $\boldsymbol{s}$ 的监督信号。由于观察到的分子结构由噪声隐子结构和语义隐子结构组成，本文假设观察到的分子结构 $\boldsymbol{A}$ 由 $\boldsymbol{G}_n$ 和 $\boldsymbol{G}_h$ 生成，用 g_A 表示，即 $\boldsymbol{A}=g_A(\boldsymbol{G}_n,\boldsymbol{G}_h)$。由于分子性质通常由包含与性质相关的功能基团的介观成分决定，本文假设分子性质 y 由语义隐子结构 $\boldsymbol{G}_h$ 和原子隐变量 $\boldsymbol{z}$ 控制，用 g_y 表示，即 $y=g_y(\boldsymbol{G}_h,\boldsymbol{z})$。

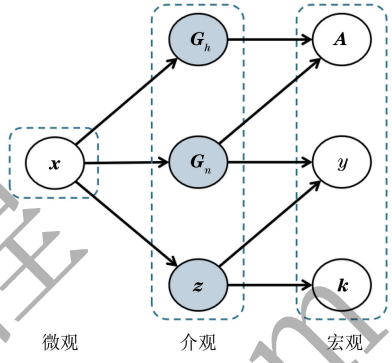

图 2 分子数据中的介观因果生成过程图示

Fig. 2 Illustration of mesoscopic causal generation process for molecule data

最后，进一步使用 $\boldsymbol{k}=g_k(\boldsymbol{z})$ 来表示原子序数如何由原子隐变量决定。因果机制可以形式化表示如下：

$$\boldsymbol{z}=f_z(\boldsymbol{x}),\boldsymbol{G}_h=f_h(\boldsymbol{x}),\boldsymbol{G}_n=f_n(\boldsymbol{x})$$
$$\boldsymbol{A}=g_A(\boldsymbol{G}_n,\boldsymbol{G}_h),\ y=g_y(\boldsymbol{G}_h,\boldsymbol{z}),\ \boldsymbol{k}=g_k(\boldsymbol{z}) \tag{1}$$

基于上述因果生成过程，本文的目标是利用训练数据集学习一个强大的分子性质预测模型。换句话说，本文的目标是在给定训练数据集的情况下估计条件分布 $P(y\mid\boldsymbol{x},\boldsymbol{A},\boldsymbol{k})$。

3 介观成分识别模型

基于上文提到的理论结果，本文提出了介观成分识别（MCI）模型，它通过介观变分自编码器（M-VAE）实现。介观变分自编码器通过建模所提出的因果生成过程恢复原子隐变量 $\boldsymbol{z}$、语义隐子结构 $\boldsymbol{G}_h$ 以及噪声隐子结构 $\boldsymbol{G}_n$，并利用 $\boldsymbol{z}$ 和 $\boldsymbol{G}_h$ 预测分子性质 y。

3.1 介观变分自编码器

本文通过建模提出的因果生成过程来恢复原子隐变量 $\boldsymbol{z}$ 的分布。为了实现这一点，本文提出了介观变分自编码器（M-VAE）。首先使用随机变分推断来建模 $\boldsymbol{x}$、y、$\boldsymbol{A}$ 和 $\boldsymbol{k}$ 的联合分布，并推导出如式（2）所示的证据下界（ELBO）：

$$\begin{aligned}L_{\mathrm{ELBO}}=&-D_{\mathrm{KL}}(Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})\parallel P(\boldsymbol{G}_h))-D_{\mathrm{KL}}(Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})\parallel P(\boldsymbol{G}_n\mid\boldsymbol{G}_h,\boldsymbol{z}))-\\&D_{\mathrm{KL}}(Q(\boldsymbol{z}\mid\boldsymbol{x})\parallel P(\boldsymbol{z}\mid\boldsymbol{G}_h))+E_{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})}E_{Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})}\ln P(\boldsymbol{A}\mid\boldsymbol{G}_h,\boldsymbol{G}_n)+\\&E_{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})}E_{Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})}E_{Q(\boldsymbol{z}\mid\boldsymbol{x})}\ln P(\boldsymbol{x}\mid\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})+E_{Q(\boldsymbol{z}\mid\boldsymbol{x})}\ln P(\boldsymbol{k}\mid\boldsymbol{z})+E_{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})}E_{Q(\boldsymbol{z}\mid\boldsymbol{x})}\ln P(y\mid\boldsymbol{G}_h,\boldsymbol{z})\end{aligned} \tag{2}$$

式中：$D_{\mathrm{KL}}(\cdot\mid\cdot)$ 表示 Kullback-Leibler（KL）散度；$Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})$、$Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})$ 和 $Q(\boldsymbol{z}\mid\boldsymbol{x})$ 用于近似 $\boldsymbol{G}_h$、$\boldsymbol{G}_n$ 和 $\boldsymbol{z}$ 的分布；$P(\boldsymbol{A}\mid\boldsymbol{G}_h,\boldsymbol{G}_n)$、$P(\boldsymbol{x}\mid\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})$、$P(\boldsymbol{k}\mid\boldsymbol{z})$ 和 $P(y\mid\boldsymbol{G}_h,\boldsymbol{z})$ 分别表示图结构解码器、原子特征解码器、原子标签解码器和分子属性预测器。

ELBO 具体推导过程如下：

首先根据贝叶斯理论分解条件概率，如式（3）所示：

$$\ln P(\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y)=\ln\frac{P(\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y,\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})}{P(\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z}\mid\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y)}=\ln\frac{P(\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y,\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})}{P(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y)P(\boldsymbol{G}_n,\boldsymbol{z}\mid\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y,\boldsymbol{G}_h)}=$$
$$\ln\frac{P(\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y,\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})}{P(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y)P(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A},\boldsymbol{G}_h)P(\boldsymbol{z}\mid\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y,\boldsymbol{G}_h)}\tag{3}$$

其次在等式两边求期望，可以得到如式(4)所示的等式：

$$\ln P(\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y)=D_{\mathrm{KL}}(Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})\parallel P(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y))+$$
$$D_{\mathrm{KL}}(Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})\parallel P(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A},\boldsymbol{G}_h))+D_{\mathrm{KL}}(Q(\boldsymbol{z}\mid\boldsymbol{x})\parallel P(\boldsymbol{z}\mid\boldsymbol{x},\boldsymbol{k},y,\boldsymbol{G}_h))+$$
$$\ln\frac{P(\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y,\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})}{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{z}\mid\boldsymbol{x})}\tag{4}$$

最后，由 $D_{\mathrm{KL}}(\cdot\mid\cdot)\geqslant 0$，得到如式(5)所示的等式：

$$\ln P(\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y)\geqslant\ln\frac{P(\boldsymbol{x},\boldsymbol{k},\boldsymbol{A},y,\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})}{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{z}\mid\boldsymbol{x})}=$$
$$\ln\frac{P(\boldsymbol{A}\mid\boldsymbol{G}_h,\boldsymbol{G}_n)P(\boldsymbol{x},\boldsymbol{k},y,\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})}{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{z}\mid\boldsymbol{x})}=$$
$$\ln\frac{P(\boldsymbol{A}\mid\boldsymbol{G}_h,\boldsymbol{G}_n)P(\boldsymbol{x}\mid\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})P(\boldsymbol{k},y,\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})}{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{z}\mid\boldsymbol{x})}=$$
$$\ln\frac{P(\boldsymbol{A}\mid\boldsymbol{G}_h,\boldsymbol{G}_n)P(\boldsymbol{x}\mid\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})P(\boldsymbol{k}\mid\boldsymbol{z})P(y,\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})}{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{z}\mid\boldsymbol{x})}=$$
$$\ln\frac{P(\boldsymbol{A}\mid\boldsymbol{G}_h,\boldsymbol{G}_n)P(\boldsymbol{x}\mid\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})P(\boldsymbol{k}\mid\boldsymbol{z})P(y\mid\boldsymbol{G}_h,\boldsymbol{z})}{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})}+\ln\frac{P(\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})}{Q(\boldsymbol{z}\mid\boldsymbol{x})}=$$
$$\ln\frac{P(\boldsymbol{A}\mid\boldsymbol{G}_h,\boldsymbol{G}_n)P(\boldsymbol{x}\mid\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})P(\boldsymbol{k}\mid\boldsymbol{z})P(y\mid\boldsymbol{G}_h,\boldsymbol{z})}{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})}+\ln\frac{P(\boldsymbol{G}_n\mid\boldsymbol{G}_h,\boldsymbol{z})P(\boldsymbol{G}_h,\boldsymbol{z})}{Q(\boldsymbol{z}\mid\boldsymbol{x})}=$$
$$\ln\frac{P(\boldsymbol{A}\mid\boldsymbol{G}_h,\boldsymbol{G}_n)P(\boldsymbol{x}\mid\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})P(\boldsymbol{k}\mid\boldsymbol{z})P(y\mid\boldsymbol{G}_h,\boldsymbol{z})}{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})}+\ln\frac{P(\boldsymbol{G}_n\mid\boldsymbol{G}_h,\boldsymbol{z})P(\boldsymbol{z}\mid\boldsymbol{G}_h)P(\boldsymbol{G}_h)}{Q(\boldsymbol{z}\mid\boldsymbol{x})}=$$
$$-D_{\mathrm{KL}}(Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})\parallel P(\boldsymbol{G}_h))-D_{\mathrm{KL}}(Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})\parallel P(\boldsymbol{G}_n\mid\boldsymbol{G}_h,\boldsymbol{z}))-$$
$$D_{\mathrm{KL}}(Q(\boldsymbol{z}\mid\boldsymbol{x})\parallel P(\boldsymbol{z}\mid\boldsymbol{G}_h))+E_{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})}E_{Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})}\ln P(\boldsymbol{A}\mid\boldsymbol{G}_h,\boldsymbol{G}_n)+$$
$$E_{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})}E_{Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})}E_{Q(\boldsymbol{z}\mid\boldsymbol{x})}\ln P(\boldsymbol{x}\mid\boldsymbol{G}_h,\boldsymbol{G}_n,\boldsymbol{z})+E_{Q(\boldsymbol{z}\mid\boldsymbol{x})}\ln P(\boldsymbol{k}\mid\boldsymbol{z})+E_{Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})}E_{Q(\boldsymbol{z}\mid\boldsymbol{x})}\ln P(y\mid\boldsymbol{G}_h,\boldsymbol{z})\tag{5}$$

实现细节如下所示：

1)实现 $Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})$ 和 $Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})$：在这一部分，本文的目标是通过从 $Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})$ 和 $Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})$ 中采样来获得 $\boldsymbol{G}_h$ 和 $\boldsymbol{G}_n$。本文首先假设 $\boldsymbol{G}_h$ 和 $\boldsymbol{G}_n$ 遵循多元伯努利分布，其参数为 $\hat{\boldsymbol{B}}_h$ 和 $\hat{\boldsymbol{B}}_n$，即 $Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A}):=P(\boldsymbol{G}_h;\hat{\boldsymbol{B}}_h)$ 和 $Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A}):=P(\boldsymbol{G}_n;\hat{\boldsymbol{B}}_n)$。本文通过 3 个步骤计算 $Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})$ 和 $Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})$ 的参数。首先使用 2 个基于图神经网络的架构分别提取节点嵌入 $\boldsymbol{Z}_h$ 和 $\boldsymbol{Z}_n$。其次计算参数矩阵 $\hat{\boldsymbol{B}}_h$、$\hat{\boldsymbol{B}}_n$，它们表示了 $\boldsymbol{G}_h$ 和 $\boldsymbol{G}_n$ 每条边存在的概率。然后本文从估计的分布中采样 $\boldsymbol{G}_h$ 和 $\boldsymbol{G}_n$。总之，上述 2 个步骤可以形式化表示如下：

$$\boldsymbol{Z}_h=E_h(\boldsymbol{x},\boldsymbol{A},\theta_h),\hat{\boldsymbol{B}}_h=\sigma(\boldsymbol{Z}_h\boldsymbol{Z}_h^{\mathrm{T}}),$$
$$\boldsymbol{G}_h\sim P(\boldsymbol{G}_h;\hat{\boldsymbol{B}}_h)$$
$$\boldsymbol{Z}_n=E_n(\boldsymbol{x},\boldsymbol{A},\theta_n),\hat{\boldsymbol{B}}_n=\sigma(\boldsymbol{Z}_n\boldsymbol{Z}_n^{\mathrm{T}}),$$
$$\boldsymbol{G}_n\sim P(\boldsymbol{G}_n;\hat{\boldsymbol{B}}_n)\tag{6}$$

式中：E_h 和 E_n 表示具有对应训练参数 θ_h、θ_n 的特征提取器；$\boldsymbol{Z}_h$、$\boldsymbol{Z}_n$ 表示特征矩阵；$\sigma(\cdot)$ 是 Sigmoid 函数。在这里本文使用 Gumbel-Softmax 来分别对 $\boldsymbol{G}_h$ 和 $\boldsymbol{G}_n$ 进行采样。因此，式(2)中的 $\boldsymbol{D}_{\mathrm{KL}}(Q(\boldsymbol{G}_n\mid\boldsymbol{x},\boldsymbol{A})\parallel P(\boldsymbol{G}_n\mid\boldsymbol{G}_h,\boldsymbol{z}))$，$D_{\mathrm{KL}}(Q(\boldsymbol{G}_h\mid\boldsymbol{x},\boldsymbol{A})\parallel P(\boldsymbol{G}_h))$ 表示了伯努利分布的 Kullback-Leibler(KL)散度。

2)实现 $Q(\boldsymbol{z}\mid\boldsymbol{x})$：本文的目标是通过从 $Q(\boldsymbol{z}\mid\boldsymbol{x})$ 中采样来获得 $\boldsymbol{z}$。因此，首先假设 $\boldsymbol{z}$ 遵循具有参数 μ 和 σ 的高斯分布，即 $Q(\boldsymbol{z}\mid\boldsymbol{x})=N(\mu,\sigma)$。随后使用 2 个多层感知器(MLP)架构来估计 μ 和 σ，可以形式化表示如下：

$$\mu=E_\mu(\boldsymbol{x};\theta_\mu)$$
$$\sigma=E_\sigma(\boldsymbol{x};\theta_\sigma)\tag{7}$$

式中：θ_μ 和 θ_σ 是训练参数。因此，式(2)中的 $D_{\mathrm{KL}}(Q(\boldsymbol{z}\mid\boldsymbol{x})\parallel P(\boldsymbol{z}\mid\boldsymbol{G}_h))$ 表示了高斯分布的 Kullback-Leibler(KL)散度。

3)实现 $P(\boldsymbol{A}\mid\boldsymbol{G}_h,\boldsymbol{G}_n)$：$P(\boldsymbol{A}\mid\boldsymbol{G}_h,\boldsymbol{G}_n)$ 用于利

用 $\boldsymbol{G}_h$ 和 $\boldsymbol{G}_n$ 生成分子结构。在形式上有：

$$\hat{\boldsymbol{A}} = D_A(\boldsymbol{G}_h, \boldsymbol{G}_n; \omega_A) \tag{8}$$

式中：$\hat{\boldsymbol{A}}$ 是重构的结构；ω_A 表示训练参数。

4)实现 $P(\boldsymbol{x} \mid \boldsymbol{G}_h, \boldsymbol{G}_n, \boldsymbol{z})$：$P(\boldsymbol{x} \mid \boldsymbol{G}_h, \boldsymbol{G}_n, \boldsymbol{z})$ 用于重构观察到的原子特征，有：

$$\hat{\boldsymbol{x}} = D_x(\boldsymbol{G}_h, \boldsymbol{G}_n, \boldsymbol{z}; \omega_x) \tag{9}$$

式中：$\hat{\boldsymbol{x}}$ 是重构的特征；ω_x 表示训练参数。

5)实现 $P(\boldsymbol{k} \mid \boldsymbol{z})$：$P(\boldsymbol{k} \mid \boldsymbol{z})$ 用于重构每个原子的标签(即分子序数)，使用一个 MLP 来预测 $\boldsymbol{k}$，可以表示为：

$$\hat{\boldsymbol{k}} = D_k(\boldsymbol{z}; \omega_k) \tag{10}$$

式中：$\hat{\boldsymbol{k}}$ 是预测的标签；ω_k 是训练参数。

6)实现 $P(y \mid \boldsymbol{G}_h, \boldsymbol{z})$：$P(y \mid \boldsymbol{G}_h, \boldsymbol{z})$ 用于利用 $\boldsymbol{G}_h$ 和 $\boldsymbol{z}$ 来预测分子性质，本文使用一个基于图神经网络的架构来生成 $\hat{y}$，可以形式化表示如下：

$$\hat{y} = D_y(\boldsymbol{G}_h, \boldsymbol{z}; \omega_y) \tag{11}$$

式中：$\hat{y}$ 是预测的标签；ω_y 是 D_y 的训练参数。

3.2 模型概述

本文基于变分自编码器的框架，同时为了进一步提高模型的性能，在损失函数中加入了稀疏正则化，并通过超参数来控制不同的损失项对模型的影响，将提出的 MCI 方法的总损失形式化表示如下：

$$L_{\text{total}} = -E_{\text{ELBO}} + \alpha L_h + \beta L_n \tag{12}$$

式中：L_h 和 L_n 分别表示 $\boldsymbol{G}_h$ 和 $\boldsymbol{G}_n$ 的 L1 范数稀疏正则化；α、β 表示超参数。

4 实验设计与结果分析

4.1 实验数据集

为了评估本文方法的性能，本文在 Open Graph Benchmark(OGB)[36] 的 6 个分子性质预测数据集上进行实验。

所有这些数据集中的分子都经过 RDKit[37] 的预处理。将每个分子视为一个图，其中节点是原子，边是化学键。观察到的原子是 9 维向量属性，包含原子手性和其他附加的原子特征，如形式电荷以及原子是否在环中。根据分子的性质，这些数据集可以分为 2 个子任务：二元分类和多标签分类。在预处理过程中，这些数据集使用一个骨架分割过程，根据它们的二维结构框架将分子分割开来。骨架分割试图将结构不同的分子分隔到不同的子集中，从而更真实地估计模型在实验设置中的性能。由于训练集的选择偏差，这个预处理过程将不可避免地引入功能团之间的虚假相关性。

4.2 评价指标

本文使用 AUC-ROC 作为二元分类和多标签分类的评估指标，这是在 OGB[36] 基准数据集中推荐的。本文选择具有最佳验证性能的模型，并在测试集上评估所选模型。对于每种方法，本文使用多个不同的随机种子，并报告均值和标准误差。

4.3 对比方法

除了传统的图神经网络，如图卷积网络[38]、图注意力网络[11]和 GraphSage[39] 之外，本文还考虑了用于图分类的最新算法，如 GIN[40] 和 SGC[41]。至于分子性质预测方法，首先考虑 CMPNN[42]，通过一种沟通内核加强了节点和边之间的消息交互。另外还考虑了 JKNet[43] 和 DiffPool[44]，利用了高级别的图信息。此外，本文还考虑了像 StableGNN[34] 和 DIR[35] 这样的最新方法，解决分子性质预测任务中的分布外挑战。

为了评估稀疏性正则化的有效性，本文进一步提出了分别去除 MCI-h 和 MCI-n 的正则化的消融模型。

4.4 实验结果

表 1 展示了在 6 个分子性质预测数据集上的实验结果，其中，表中的数值是在使用不同的随机种子进行的 4 次重复实验中取平均得到的，括号中的值表示标准差。根据实验结果可以发现：

1)所提出的 MCI 模型在大多数数据集上优于所有其他对比模型，这归因于所提出的介观因果生成过程和介观变分自编码器的双重作用，实验结果表明了 MCI 模型具有分布外的泛化能力。

2)一些基于图神经网络的图分类方法，如 SGC 和 GIN，在分子性质预测任务中表现不佳，这意味着这些方法具有较差的泛化能力，因为它们考虑了整个分子图，而整个分子图中包含了一些与预测的性质无关的子结构，这些与性质无关的结构信息会干扰模型的预测。DIFFPOOL 取得了更好的结果，这是因为它学习了分层表示，并可能过滤与性质无关的结构信息。

3)在 moltoxcast 数据集中，所提出的模型并没有达到最佳性能。造成这些结果的原因有两个，首先，moltoxcast 数据集包含 600 多种分子性质，非常复杂，所有模型在该数据集上都取得了接近的结果。其次，所提出的模型需要基于对所提出的因果机制进行良好建模的假设，但这个数据集的规模太小，提出的模型无法学习到准确的因果机制。

表 1 在 6 个分子性质预测数据集上各方法的 AUC-ROC 值

Table 1 AUC-ROC values of various methods on six molecular property prediction datasets

模型	molbace	molbbbp	molclintox	moltox21	molsider	moltoxcast
GCN	0.758 8(0.019)	0.664 7(0.009)	0.862 3(0.028)	0.677 5(0.007)	0.584 3(0.003)	0.624 4(0.006)
GAT	0.811 7(0.008)	0.671 7(0.005)	0.834 7(0.014)	0.688 1(0.005)	0.595 6(0.010)	0.614 1(0.006)
GraphSAGE	0.785 1(0.017)	0.661 6(0.010)	0.886 0(0.024)	0.688 8(0.006)	0.605 9(0.002)	0.628 2(0.007)
GIN	0.745 0(0.028)	0.677 2(0.019)	0.868 6(0.038)	0.642 0(0.002)	0.564 7(0.012)	0.633 5(0.003)
GIN0	0.743 6(0.035)	0.666 5(0.013)	0.893 1(0.021)	0.646 2(0.009)	0.596 8(0.015)	0.628 9(0.002)
MoNet	0.769 2(0.009)	0.695 2(0.005)	0.867 5(0.0122)	0.670 2(0.003)	0.600 3(0.004)	0.624 8(0.005)
SGC	0.712 8(0.018)	0.611 7(0.030)	0.777 6(0.019)	0.664 9(0.011)	0.590 6(0.003)	0.628 3(0.001)
JKNet	0.789 9(0.134)	0.656 2(0.008)	0.816 3(0.028)	0.659 8(0.005)	0.581 8(0.016)	0.635 7(0.006)
DIFFPOOL	0.746 9(0.111)	0.633 5(0.022)	0.904 8(0.024)	0.690 5(0.009)	0.575 8(0.015)	0.621 7(0.005)
CMPNN	0.721 5(0.049)	0.640 3(0.017)	0.794 7(0.046)	0.704 8(0.011)	0.579 9(0.008)	**0.639 4(0.011)**
StableGNN-GCN	0.769 5(0.033)	0.688 2(0.039)	0.879 8(0.024)	0.708 0(0.003)	0.591 5(0.012)	0.632 9(0.007)
StableGNN-SAGE	0.807 3(0.040)	0.684 7(0.025)	0.909 6(0.020)	0.691 4(0.002)	0.521 5(0.019)	0.625 1(0.003)
DIR	0.783 4(0.015)	0.646 7(0.017)	0.812 9(0.031)	0.696 6(0.029)	0.579 4(0.011)	0.619 6(0.014)
MCI	**0.820 8(0.005)**	**0.730 0(0.029)**	**0.920 3(0.008)**	**0.736 4(0.004)**	**0.625 2(0.006)**	0.633 1(0.003)
p-value	**0.003**	**0.028**	**0.026**	**0.0**	**0.0**	**0.844**

4.5 消融实验

为了评估稀疏性正则化的有效性，本文设计了 MCI-h 和 MCI-n。图 3 展示了在 molbace、molbbbp 和 molclintox 数据集上的实验结果。

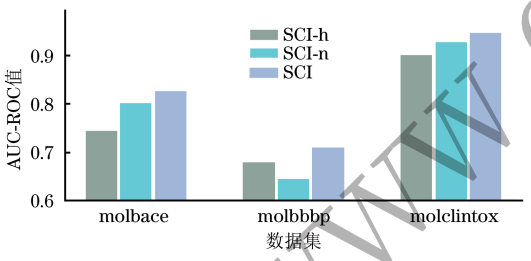

图 3 在 molbace、molbbbp 和 molclintox 数据集上的消融实验结果

Fig.3 Experimental results of ablation on the molbach, molbbbp, and molclintox datasets

根据图 3 中呈现的实验结果可以得出：

1)标准 MCI 优于 MCI-h，这是因为稀疏性正则化有助于提取语义隐子结构，与语义隐子结构的先验稀疏性一致。

2)标准 MCI 也优于 MCI-n，由于限制 $\boldsymbol{G}_n$ 的稀疏性限制系数性可以有效减少冗余信息，从而提高了模型泛化能力。

5 结束语

本文提出了一种用于稳健分子性质预测的介观成分识别模型。该模型基于分子数据的介观因果生成过程和变分自编码器的框架，通过学习与分子性质相关的介观成分的表示，实现对分子性质的预测。实验结果表明，所提出模型不仅在分子性质预测中提供了有效的解决方案，而且在生物化学的解释和参考方面也提供了一些有参考价值的结果。

参考文献

[1] ATZ K, GRISONI F, SCHNEIDER G. Geometric deep learning on molecular representations[J]. Nature Machine Intelligence, 2021, 3: 1023-1032.

[2] DE ALMEIDA A F, MOREIRA R, RODRIGUES T. Synthetic organic chemistry driven by artificial intelligence [J]. Nature Reviews Chemistry, 2019, 3: 589-604.

[3] JUMPER J, EVANS R, PRITZEL A, et al. Highly accurate protein structure prediction with AlphaFold[J]. Nature, 2021, 596(7873): 583-589.

[4] VAMATHEVAN J, CLARK D, CZODROWSKI P, et al. Applications of machine learning in drug discovery and development[J]. Nature Reviews Drug Discovery, 2019, 18: 463-477.

[5] 朱洪翔，傅钰江，李雪，等. 基于神经网络的分子性质预测算法研究进展[J]. 科学技术与工程，2023，23(19)：8061-8070.
ZHU H X, FU Y J, LI X, et al. Overview of molecular property prediction algorithms with neural network[J]. Science Technology and Engineering, 2023, 23(19): 8061-8070. (in Chinese)

[6] 李林洁. 基于机器学习的分子性质预测与生成技术研究[D]. 成都：电子科技大学，2023.
LI L J. Research on molecular property prediction and generation technology based on machine learning [D]. Chengdu: University of Electronic Science and Technology of China, 2023. (in Chinese)

[7] BUTLER K T, DAVIES D W, CARTWRIGHT H, et al. Machine learning for molecular and materials science[J]. Nature, 2018, 559: 547-555.

[8] 于佳卉. 基于深度学习的有机化合物合成可行性预测[D]. 杭州：浙江大学，2022.
YU J H. Feasibility prediction of organic compound synthesis

based on deep learning[D]. Hangzhou: Zhejiang University, 2022. (in Chinese)
[9] NEESE F. Prediction of molecular properties and molecular spectroscopy with density functional theory: from fundamental theory to exchange-coupling[J]. Coordination Chemistry Reviews, 2009, 253(5/6): 526-563.
[10] BATTAGLIA P W, HAMRICK J B, BAPST V, et al. Relational inductive biases, deep learning, and graph networks[EB/OL]. [2023-10-30]. https://arxiv.org/abs/1806.01261v3.
[11] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. [2023-10-30]. https://arxiv.org/abs/1710.10903.
[12] LU C Q, LIU Q, WANG C, et al. Molecular property prediction: a multilevel quantum interactions modeling perspective[J]. Artificial Intelligence, 2019, 33(1): 1052-1060.
[13] LI S L, ZHOU J B, XU T, et al. GeomGCL: geometric graph contrastive learning for molecular property prediction [J]. Artificial Intelligence, 2022, 36(4): 4541-4549.
[14] HU W, LIU B, GOMES J, et al. Strategies for pre-training graph neural networks[C]// Proceedings of International Conference on Learning Representations. Washington D. C., USA: IEEE Press, 2020:451-462.
[15] RONG Y, BIAN Y, XU T, et al. Self-supervised graph transformer on large-scale molecular data[C]//Proceedings of Advances in Neural Information Processing Systems. Cambridge, USA: MIT Press, 2020: 12559-12571.
[16] BUENO P R, BENITES T A, DAVIS J J. The mesoscopic electrochemistry of molecular junctions [J]. Scientific Reports, 2016, 6: 18400.
[17] LIU Z T, LIN L Q, JIA Q Q, et al. Transferable multilevel attention neural network for accurate prediction of quantum chemistry properties via multitask learning[J]. Journal of Chemical Information and Modeling, 2021, 61(3): 1066-1082.
[18] 周彪. 基于图对比学习的分子性质预测研究[D]. 合肥: 合肥学院, 2023.
ZHOU B. Research on molecular property prediction based on graph contrastive learning[D]. Hefei: Hefei University, 2023. (in Chinese)
[19] 卫平柱. 基于邻域交互图卷积神经网络的化学分子性质预测研究[D]. 合肥: 合肥学院, 2023.
WEI P Z. Prediction of chemical molecular properties based on convolution neural network of neighborhood interactive graph[D]. Hefei: Hefei University, 2023. (in Chinese)
[20] 汪维泰, 王晓强, 李雷孝, 等. 时空图神经网络在交通流预测研究中的构建与应用综述[J]. 计算机工程与应用, 2024, 64(8): 31-45.
WANG W T, WANG X Q, LI L X, et al. Overview of the construction and application of spatiotemporal graph neural networks in traffic flow prediction research[J]. Computer Engineering and Applications, 2024, 64(8): 31-45. (in Chinese)
[21] WIEDER O, KOHLBACHER S, KUENEMANN M, et al. A compact review of molecular property prediction with graph neural networks[J]. Drug Discovery Today Technologies, 2020, 37: 1-12.
[22] ROGERS D, HAHN M. Extended-connectivity fingerprints [J]. Journal of Chemical Information and Modeling, 2010, 50(5): 742-754.
[23] SONG Y, ZHENG S J, NIU Z M, et al. Communicative representation learning on attributed molecular graphs[C]// Proceedings of the 29th International Joint Conference on Artificial Intelligence. Washington D. C., USA: IEEE Press, 2020: 331-332.
[24] ZHANG Z Q, GUAN J H, ZHOU S G. FraGAT: a fragment-oriented multi-scale graph attention model for molecular property prediction [J]. Bioinformatics, 2021, 37(18): 2981-2987.
[25] ZHANG Z, LIU Q, WANG H, et al. Motif-based graph self-supervised learning for molecular property prediction [C]// Proceedings of Advances in Neural Information Processing Systems. Cambridge, USA: MIT Press, 2021: 15870-15882.
[26] SUN M Y, XING J, WANG H J, et al. MoCL: data-driven molecular fingerprint via knowledge-aware contrastive learning from molecular graph[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021: 3585-3594.
[27] 丁婧娴, 李翔, 孙纪舟, 等. 融合多特征和双向图分类的专家推荐方法[J]. 数据采集与处理, 2023, 38(5): 1214-1225.
DING J X, LI X, SUN J Z, et al. Expert recommendation method combining multi-features and Bi-directional graph classification[J]. Journal of Data Acquisition and Processing, 2023, 38(5): 1214-1225. (in Chinese)
[28] 吕超, 孟相浩, 崔格格, 等. 基于图分类的智能车辆复杂场景风险等级评估与建模[J]. 北京理工大学学报, 2023, 43(7): 726-733.
LÜ C, MENG X H, CUI G G, et al. Risk level estimating and modeling of complex scenarios for intelligent vehicles based on graph classification[J]. Transactions of Beijing Institute of Technology, 2023, 43(7): 726-733. (in Chinese)
[29] SHEN Z Y, LIU J S, HE Y, et al. Towards out-of-distribution generalization: a survey[EB/OL]. [2023-10-30]. https://arxiv.org/abs/2108.13624.
[30] ZHANG X X, CUI P, XU R Z, et al. Deep stable learning for out-of-distribution generalization[C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE Press, 2021: 5372-5382.
[31] CHEN J A, SHEN D H, CHEN W Z, et al. HiddenCut: simple data augmentation for natural language understanding with better generalizability[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2021: 1211-1223.
[32] 卢敏, 原子婷. 结合图对比学习的多图神经网络会话推荐方法[J]. 计算机科学, 2024, 51(5): 54-61.
LU M, YUAN Z T. Multi graph neural network session recommendation method based on graph contrastive learning [J]. Computer Science, 2024, 51(5): 54-61. (in Chinese)
[33] LI H Y, WANG X, ZHANG Z W, et al. OOD-GNN: out-of-distribution generalized graph neural network[EB/OL]. [2023-10-30]. https://arxiv.org/abs/2112.03806.
[34] FAN S H, WANG X, SHI C, et al. Generalizing graph neural networks on out-of-distribution graphs[EB/OL]. [2023-10-30]. https://arxiv.org/abs/2111.10657.
[35] WU Y X, WANG X, ZHANG A, et al. Discovering invariant rationales for graph neural networks[EB/OL]. [2023-10-30]. https://arxiv.org/abs/2201.12872.
[36] HU W, FEY M, ZITNIK M, et al. Open graph benchmark: datasets for machine learning on graphs[C]// Proceedings of Advances in Neural Information Processing Systems. Cambridge, USA: MIT Press, 2021:22118-22133.
[37] LANDRUM G. RDKit: open-source chem-informatics[EB/OL]. [2023-10-30]. https://github.com/rdkit/rdkit.
[38] KIPF T, WELLING M. Semi-supervised classification with graph convolutional networks[EB/OL]. [2023-10-30]. https://arxiv.org/abs/1609.02907.
[39] HAMILTON W L, YING Z, LESKOVEC J. Inductive representation learning on large graphs[C]//Proceedings of Advances in Neural Information Processing Systems.

Cambridge, USA: MIT Press, 2017:578-587.
[40] XU K, HU W, LESKOVEC J. How powerful are graph neural networks? [EB/OL]. [2023-10-30]. https://arxiv.org/abs/1810.00826.
[41] WU F, SOUZA A, ZHANG T, et al. Simplifying graph convolutional networks [C]// Proceedings of International Conference on Machine Learning. Washington D. C., USA: IEEE Press, 2000: 6861-6871.
[42] SONG Y, ZHENG S J, NIU Z M, et al. Communicative representation learning on attributed molecular graphs[C]// Proceedings of 20th International Joint Conference on Artificial Intelligence and 17th Pacific Rim International Conference on Artificial Intelligence. Washington D. C., USA: IEEE Press, 2020: 2831-2838.
[43] XU K, LI C, TIAN Y, et al. Representation learning on graphs with jumping knowledge networks[C]//Proceedings of International Conference on Machine Learning. Washington D. C., USA: IEEE Press, 2018: 5453-5462.
[44] YING R, YOU J, MORRIS C, et al. Hierarchical graph representation learning with differentiable pooling [C]// Proceedings of Advances in Neural Information Processing Systems. Cambridge, USA: MIT Press, 2018: 31-45.

编辑 索书志