

引用格式:孟祥铭,赵文英,孙晓岩,等. 基于图神经网络(GNN)的交互作用模型参数预测模型[J]. 北京化工大学学报(自然科学版),2026,53(5):11-22.

MENG XiangMing, ZHAO WenYing, SUN XiaoYan, et al. Prediction model of interaction model parameters based on graph neural networks (GNN)[J]. Journal of Beijing University of Chemical Technology (Natural Science), 2026, 53(5): 11-22.

基于图神经网络(GNN)的交互作用模型参数预测模型

孟祥铭¹ 赵文英¹ 孙晓岩¹ 陶少辉^{1*} 陈玉石² 李 峰²

(1. 青岛科技大学 化工学院 过程系统工程研究所, 青岛 266042; 2. 石化盈科信息技术有限责任公司, 上海 200040)

摘 要: 交互作用模型参数是 Wilson、非随机双流体(NRTL)等活度系数模型中的关键热力学参数,其准确性直接决定了汽液相平衡及分离过程计算结果的可靠性。针对传统参数获取方法依赖实验数据回归以及通用准化学官能团活度系数(UNIFAC)模型估算存在数据覆盖受限、非线性方程组求解不稳定及适用范围受约束等核心瓶颈,构建了一种基于分子结构表征的双分支图神经网络(GNN)模型,实现了对 Wilson 交互作用模型参数的高效直接预测。该模型以分子图为输入,分别提取溶质与溶剂分子特征,并通过消息传递机制深入学习分子间潜在的结构关联,同时引入并联残差策略以降低预测误差。基于 4 068 组二元体系交互作用模型参数数据进行训练与测试,结果表明,所提出的并联残差 GNN 模型在多项误差指标上均显著优于 UNIFAC 方法,并且相较于单一 GNN 模型,其预测精度得到进一步提高。将预测的 Wilson 交互作用模型参数 A_{ij} 、 A_{ji} 、 B_{ij} 、 B_{ji} 代入 Wilson 方程计算无限稀释活度系数 γ_i^∞ ,所得结果与文献计算值高度一致,验证了该模型预测参数的热力学合理性与工程适用性。

关键词: 交互作用模型参数; 图神经网络(GNN); Wilson 模型; UNIFAC 模型; 无限稀释活度系数

中图分类号: TQ013. 1; TP183 **DOI:** 10. 13543/j. bhexb. 2026. 05. 002

引 言

交互作用模型参数是 Wilson、非随机双流体(NRTL)等活度系数模型中的核心变量,用于表征分子间非理想相互作用,其准确性很大程度上影响汽液相平衡(VLE)计算以及化工过程模拟结果的可靠性。在工程实践中,交互作用模型参数通常通过两种方法确定。一种方法是基于汽液相平衡实验数据进行参数回归,该方法具有明确的物理依据,预测精度较高,是目前最为可靠的参数获取方式,但

该方法成本高昂且数据覆盖有限。根据 Dortmund 数据库统计,在最重要的 1 000 种化合物中,仅有约 1. 7% 的二元体系具有可用于参数调整的实验数据,其余系统仍然主要依赖预测模型^[1-2]。另一种方法是在缺乏实验数据的情况下,借助基于基团贡献法(GCM)的通用准化学官能团活度系数(UNIFAC)预测模型进行估算。UNIFAC 模型是目前最常用的交互作用模型参数预测方法,该方法通过 UNIFAC 计算组分在不同温度下的无限稀释活度系数,并基于无限稀释活度系数与交互作用参数之间的函数关系,通过求解非线性方程组获得不同温度下的交互参数,进而拟合构建二元作用模型参数。此外,UNIFAC 模型也可以估算状态方程的交互作用参数。最初的 PSRK(predictive Soave-Redlich-Kwong)模型巧妙地结合 UNIFAC 和 SRK 方程式,并使用零压力的混合规则,能够预测含超临界组分体系的高压 VLE^[3]。通过将 NRTL 方程整合到硬球状态方程,

收稿日期:2026-03-04

基金项目:国家自然科学基金(22178190/22408195);山东省重点研发计划(2024CXGC010405);山东省化工过程模拟与优化工业软件重点实验室项目(PKL2024F23)

第一作者:女,2001 年生,硕士生

* 通信联系人

E-mail:tau@qust. edu. cn

基团贡献状态方程(GC-EOS)模型的适用范围已拓展至高达 30 MPa 压力以及超临界萃取计算^[4]。针对立方型状态方程交互参数难以确定的问题,PPR78 模型提出了通过基团贡献法计算温度 T 依赖的二元交互作用参数 $\kappa_{ij}(T)$ 的理论表达式,该模型仅需纯组分的临界性质即可实现饱和和混合物的高精度预测^[4]。现有的基团贡献模型仍受制于 UNIFAC 参数缺失的问题,同时在 Wilson 等模型的参数获取上,习惯“先利用 UNIFAC 计算无限稀释活度系数,再反解交互参数”这一方式,但是该反演过程严重依赖数值求解器,易受初始值影响而出现不收敛或多解的问题^[5-6],在面对部分互溶体系及活度系数极值情况时,其适用性和可靠性难以保证^[3,7]。

图神经网络(GNN)是一种专门用于处理非欧几里得空间图结构数据的算法框架^[8-13],能够直接解析分子拓扑结构信息,这一特性使其近期在关于热力学参数预测的研究中备受关注。GNN 通过引入消息传递机制捕捉节点(原子)与边(化学键)之间的关联关系,在处理非欧几里得数据时表现出极强的跨领域泛化潜力^[14]。针对分子图的特征提取,作为 GNN 的一种经典且高效的变体,图卷积网络(GCN)因无需复杂的特征分解过程且具有高度的灵活性,目前正逐步确立其主导地位^[15]。GNN 的应用已延伸至化工领域工艺系统设计。例如,在软测量建模中,已有研究尝试将单元操作与物流网络概括为过程拓扑图,借助 GNN 挖掘其时空关联,显著提高了模型的鲁棒性与数据利用效率^[16]。这说明 GNN 等深度学习模型在处理分子拓扑特征时,能够自动识别并量化影响分子间相互作用的关键结构片段,为微观结构特征与宏观热力学模型参数之间的关联提供了有效的计算途径。然而,目前的热力学应用主要停留在对活度系数等表观性质的“端到端”拟合上,针对活度系数模型交互作用模型参数结构化预测的相关研究相对匮乏。

本文提出了一种从分子拓扑结构直接预测热力学交互作用模型参数的新路径。利用 RDKit^[17]解析 SMILES^[18]序列构建分子图,设计了双分支 GNN 分别独立提取溶质和溶剂分子的结构特征,并通过消息传递机制深入学习分子间潜在的结构关联;同时,引入并联残差学习策略构建二次误差修正模型,以进一步增强预测鲁棒性。这种“端到端”的预测模式避免了传统非线性拟合中复杂的迭代计算与解的不稳定性。更重要的是,由于预测对象为

Wilson 模型参数而非单一物性值,因此研究结果能够直接集成至流程模拟软件中进行全浓度范围的相平衡计算,使得所预测的模型参数具有更强的物理外推性与工程应用价值。

1 模型构建

1.1 Wilson 模型

在处理液相非理想溶液及相平衡计算时,Wilson 模型是一类被广泛采用的经典模型。该模型的理论基础在于引入了局部组成这一假设,即受分子间相互作用能差异的影响,溶液内部微观区域的分子分布并非随机,而是呈现出与宏观总体浓度不同的局部聚集特征,该模型通过引入交互作用参数来量化表征组分间的能量差异。

针对二元混合体系,Wilson 模型通过式(1)描述组分 i 的活度系数 γ_i^∞ 。

$$\ln \gamma_i = -\ln(x_i + x_j \Lambda_{ij}) + x_j \left[\frac{\Lambda_{ij}}{x_i + x_j \Lambda_{ij}} - \frac{\Lambda_{ji}}{x_j + x_i \Lambda_{ji}} \right] \quad (1)$$

式中: x_i 和 x_j 分别表示组分 i 和 j 的摩尔分数, Λ_{ij} 表示组分 i 对组分 j 的 Wilson 交互作用参数, Λ_{ji} 表示组分 j 对组分 i 的 Wilson 交互作用参数。从热力学本质上看,参数 Λ_{ij} 由液相摩尔体积比 V_i^L 与能量参数($g_{ij} - g_{ii}$)共同决定(式(2))。

$$\ln \Lambda_{ij} = \ln \frac{V_j^L}{V_i^L} - \frac{(g_{ij} - g_{ii})}{RT} \quad (2)$$

式中: V_i^L 和 V_j^L 分别表示组分 i 和组分 j 的饱和液相摩尔体积,L/mol;能量项 $\frac{(g_{ij} - g_{ii})}{RT}$ 呈现显式的 $1/T$ 温度依赖关系; R 表示摩尔气体常数,J/(mol·K); T 表示绝对温度,K。为了便于工程上的参数回归与温度外推,通常将其简化为如下经验式(3)。

$$\ln \Lambda_{ij} = A_{ij} + \frac{B_{ij}}{T} \quad (3)$$

式中: A_{ij} 和 B_{ij} 表示组分 i 和 j 之间的二元交互作用模型参数。参数 A_{ij} 主要反映分子间体积差异的常数项贡献,而 B_{ij} 刻画能量对相互作用强度的调制作用。

对于无限稀释条件($x_i \rightarrow 0$ 或 $x_j \rightarrow 0$),可推导出无限稀释活度系数 γ_i^∞ 的表达式。该条件下体系对分子间相互作用极为敏感, $\ln \gamma_i^\infty$ 常被用作评估模型参数可靠性的重要指标。

1.2 分子表示与模型输入

为了将分子信息导入计算框架,利用SMILES^[19]编码对溶质和溶剂分子进行表征。针对每一个分子,借助化学信息学软件包RDKit^[17]提取原子特征和键特征,构建分子的图结构。该分子图可表示为式(4)。

$$G = (V, E) \quad (4)$$

式中:节点集合 V 和边集合 E 分别对应分子中的原子和化学键。每一个节点和边均由一个特征向量描述,用来表征分子的相关结构信息。

采用的节点特征和边特征及其维度分别汇总于表1和表2中。选择这些特征旨在区分同一分子中原子和化学键之间的基本结构差异,其依据是先前采用类似方法得到的研究结果^[20-21]。每个特征向量的维度由该特征在研究数据集中可取的不同类别数量决定。节点特征主要包括原子类型、是否成环、芳香性、杂化方式、成键数、形式电荷以及连接氢原子数(表1),用于表征分子结构中原子所处的局部化学环境。边特征包括键类型、共轭性以及是否成环(表2),用于描述原子间的键合关系及其结构特性。其中,节点特征与边特征中的“是否成环”并非重复。节点特征中的成环侧重于刻画原子受限的局部空间与电子环境,而边特征中的成环界定了化学键的刚性与旋转自由度。例如在联苯类结构中,两个处于环中的原子可由一根非环化学键相连。在GNN的消息传递过程中,两者分别从状态属性和信息传递权重的角度,共同协助模型精准区分分子的拓扑构象。

上述所有特征均采用独热编码方式(one-hot)^[22]表示,使得不同类别的化学信息能够以统一的向量形式输入到图神经网络模型中。而对于仅

表1 分子图中节点(原子)特征及其维度

Table 1 Characteristics and dimensions of the nodes (atoms) in the molecular graph

特征	含义	维度
原子类型	原子的元素类型(C、Br、Cl、N、O、I、S、F、P、Si、Li)	11
是否成环	该原子是否位于环结构中	1
芳香性	该原子是否属于芳香体系	1
杂化方式	原子的杂化状态(sp、sp ² 、sp ³)	3
成键数	该原子参与形成的化学键数量	5
形式电荷	原子的形式电荷	3
连接氢原子数	与该原子相连的氢原子数量	4

表2 分子图中边(化学键)特征及其维度

Table 2 Characteristics and dimensions of the edges (chemical bonds) in the molecular graph

特征	含义	维度
键类型	化学键类型(单键、双键、三键、芳香键)	4
共轭性	该化学键是否为共轭键	1
是否成环	该化学键是否位于环结构中	1

包含两类逻辑取值的特征(“是”或“否”),例如“是否成环”“芳香性”及“共轭性”,为避免维度冗余,未采用二维的独热编码,而是直接将其编码为单个二元变量,取值为1代表“是”,0代表“否”。该特征构建方式在保证分子结构信息完整表达的同时,有效避免了显式引入氢原子节点,从而降低了分子图的规模和计算复杂度。

本模型目前的输入特征主要基于分子的原子-键拓扑结构。需要说明的是,虽然模型未显式地将氢键能、缔合度等非共价相互作用参数作为独立特征输入,但GNN模型的核心优势在于其多层消息传递机制。通过对包含强极性官能团(如—OH、—NH₂等)及其邻域拓扑环境的迭代聚合,能够在高维隐空间中捕捉到这些官能团引起分子间强相互作用的特征模式。这种隐式建模方式使模型在一定程度上具有识别氢键及团聚效应的能力,从而能够捕捉复杂的非线性热力学响应。

作为示例,图1展示了碳酸二乙酯(CCOC(=O)OCC)从“SMILES表示”到“分子图表示”的编码过程。利用RDKit可以对分子中每一个原子计算one-hot编码的特征向量,并将其作为矩阵 $A \in \mathbb{R}^{n_a \times 28}$ 中的一行,其中 n_a 表示分子中原子的数量。类似地,通过将分子中每一条化学键对应的特征向量按行堆叠,可以构建化学键特征矩阵 $B \in \mathbb{R}^{n_b \times 6}$,其中 n_b 表示分子中化学键的数量。需要指出的是,为了直观呈现特征向量的结构,图1中的矩阵示意图采用从1开始的自然数对特征维度进行序号排列,每个序号对应表1中的一种特征类型,键特征按照相同方法处理。分子图的连通性由邻接矩阵 $C \in \mathbb{R}^{n_a \times n_a}$ 进行描述。最终,借助PyTorch Geometric^[23]将上述3个矩阵统一存储为张量集合,用来表示单个分子的图结构。

将分子表示为图结构的方法允许GNN等模型基于分子结构信息及其在目标预测中的相对重要性,自动学习并优化用于描述二元体系的分子指

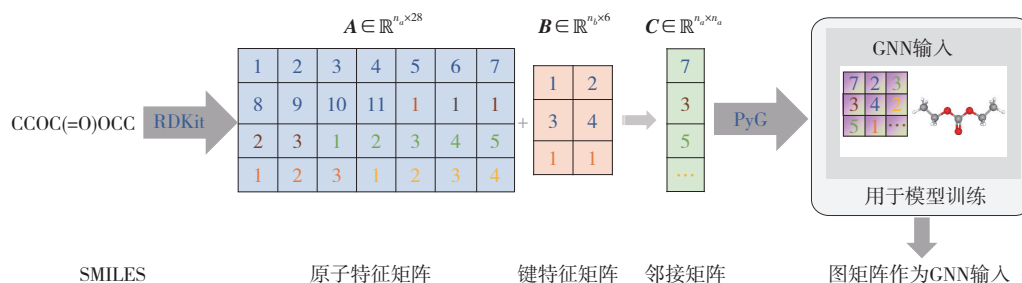


图1 从SMILES序列构建初始分子图

Fig. 1 Construction of the initial molecular graph from a SMILES string

纹,从而更有效地预测交互作用模型参数。

1.3 双分支GNN构建

GNN通过对图结构数据执行一系列数学变换,将输入图映射为更新后的输出图,即实现图到图的表示学习。图2展示了基于GNN的交互作用模型参数预测模型。图2中每一个二元体系由两个分子图组成,分别对应溶质和溶剂分子,这两个分子图共同映射至交互作用模型参数值。为此,分别采用两个结构相同但参数独立的GNN对溶质和溶剂分子的分子图进行特征学习。对于每个卷积层 l ,通过

式(5)的消息传递方案获得具有更新节点特征的图^[21]。

$$\mathbf{h}_v^{(l+1)} = \mathbf{W}^{(l)\text{T}} \mathbf{h}_v^{(l)} + \sum_{w \in \mathcal{N}(v)} \mathbf{h}_w^{(l)} \phi_E(\mathbf{e}_{vw}) \quad (5)$$

式中: $\mathbf{h}_v^{(l+1)}$ 表示节点 v 在第 $l+1$ 层中的 d 维特征向量, $\mathbf{W}^{(l)} \in \mathbb{R}^{d \times d}$ 表示可学习的权重矩阵, $\mathcal{N}(v)$ 表示节点 v 的邻域节点集合, $\mathbf{e}_{vw}$ 表示与中心节点相连的化学键特征向量, $\phi_E(\cdot)$ 表示将节点间的化学键特征映射为边条件权重矩阵 $\mathbb{R}^{d \times d}$ 的函数, $\phi_E(\cdot)$ 通过1个带有ReLU激活函数的单隐层神经网络实现。

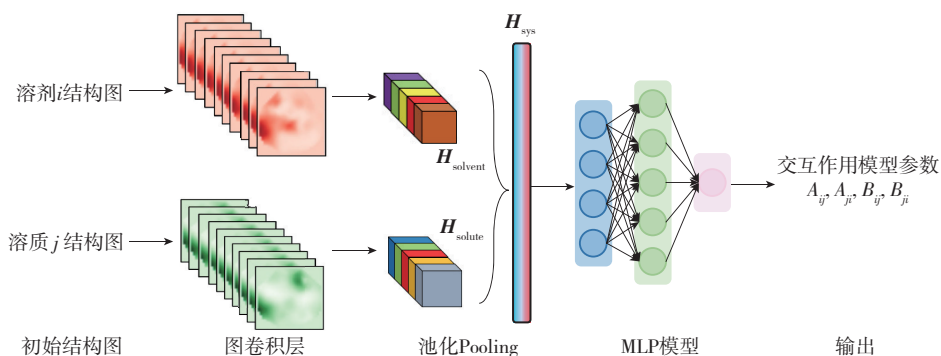


图2 基于GNN的交互作用模型参数预测模型

Fig. 2 Prediction model of the interaction model parameter based on GNN

消息传递过程如下。首先,化学键特征向量 $\mathbf{e}_{vw}$ 经由函数 $\phi_E(\cdot)$ 处理后,与相邻节点的特征向量 $\mathbf{h}_w^{(l)}$ 进行加和,形成邻域消息。随后,该邻域消息与中心节点自身特征向量 $\mathbf{h}_v^{(l)}$ 经线性变换后的结果相加,得到更新后的节点表示 $\mathbf{h}_v^{(l+1)}$,该向量定义了下一层的分子图中对应节点的特征表示。经过 L 层图卷积操作后,得到 L 个逐层更新的分子图。尽管部分研究仅使用最后一层生成的分子图作为最终分子表示,但已有研究表明,综合利用多层更新图中所包含的信息有助于提升GNN模型的预测性能^[20-21],这一思想被称为“跳跃知识”。通过对 L 个更新分子图取平均的方式构建最终分子图表示。随后,将该分

子图输入至Set2Set全局池化层,得到一个固定维度的向量,用于表示对应的分子物种,即分子指纹。Set2Set池化方法具有置换不变性,并且相较于简单的求和或平均池化方式具有更强的表达能力。

在获得溶质分子指纹 H_{solute} 和溶剂分子指纹 H_{solvent} 后,将二者进行拼接,形成二元体系指纹 H_{sys} ,可以将其表示为二进制系统指纹,如式(6)所示。

$$\mathbf{H}_{\text{sys}} = \mathbf{H}_{\text{solvent}} // \mathbf{H}_{\text{solute}} \quad (6)$$

式中, $//$ 表示串联。该指纹可视为对溶质-溶剂体系结构特征的综合表征。最后,将该二元体系指纹输入至多层感知机(MLP)模型中,用于交互作用模型参数预测。

本研究设计的双分支GNN框架在逻辑上分为3个阶段。首先是独立编码阶段,通过两个结构相同的图卷积网络分别对溶质和溶剂分子图进行局部化学环境的特征提取;其次是信息融合阶段,通过池化操作生成代表独立分子的指纹后,将其拼接为二元体系的全局状态向量;最后是参数映射阶段,利用MLP将融合后的体系特征非线性映射为宏观的Wilson交互参数。

该方法的关键优势是整个模型框架能够基于指定的误差度量,通过反向传播实现“端到端”训练。一方面,通过图卷积操作,模型能够自动识别并学习与溶质分子和溶剂分子在分子层面非理想性相关的关键结构特征(可能与分子形状对非理想性的贡献有关);另一方面,基于二元体系指纹的MLP能够进一步捕捉溶质-溶剂相互作用对体系非理想行为的影响。这一过程无需人工选择官能团或经验拟合二元基团相互作用参数,明显区别于传统的基团贡献方法中依赖经验进行参数构建的建模方式。

1.4 模型实验设置

在数据集训练时,本研究采用随机划分的方式将数据集按8:2的比例划分为训练集和测试集,训练集内部再进一步划分出10%用于模型验证及超参数调优。与分子骨架划分等其他数据划分方式相比,所选方法采用分层嵌套式随机划分+集成学习(Ensemble)策略,优先关注模型在已知化学空间中的插值能力,而非盲目追求对未知结构的外推能力。统计分析显示,划分后的各子集在交互作用参数分布上保持了一致性,确保了训练的可靠性^[10]。

模型的超参数依据验证集上的性能进行选择,报告的结果均基于测试集。模型采用Python3.9环境并基于PyTorch Geometric^[23]实现。为提高模型训练的稳定性,在每一层图卷积之后引入批归一化层,并采用0.1的dropout概率以降低过拟合风险。除最后一层图卷积外,其余各层均采用Leaky ReLU作为激活函数。模型训练过程中采用Adam优化算法,Adam初始学习率为0.001,对均方误差(MSE)损失函数进行最小化,批大小设为32个二元体系,训练轮数为200个epoch。当验证集误差在连续3个epoch内未出现下降(阈值设为 10^{-4})时,学习率按0.8的比例进行衰减。上述最优超参数通过基于高斯过程的贝叶斯优化方法确定,其中不同交互作用模型参数的最优超参数如表3所示。所有数值实验

均在Intel(R) Xeon(R) Silver 4208 CPU 2.10 GHz (16 GB显存)上完成。

表3 交互作用模型参数的最优超参数

Table 3 Optimal hyperparameters of the interaction model parameters

超参数类型	超参数搜索区间	最优值			
		A_{ij}	A_{ji}	B_{ij}	B_{ji}
GNN 卷积层数 L	(3, 6)	4	5	5	5
dropout 概率	(0.0, 0.3)	0.1	0.1	0.1	0.1
卷积层隐层维度	(20, 50)	50	30	30	30
中间层神经元数	(32, 128)	39	64	64	64
MLP 层数	(2, 4)	3	3	3	3
MLP 头 1	(20, 80)	27	50	50	50
MLP 头 2	(10, 50)	40	25	25	25

为进一步提升模型的预测精度并提升稳定性,本研究引入一种集成学习方法——聚合(bagging)。该集成策略通过构建并训练多个独立的单一GNN模型^[8],在网络结构与超参数设置保持一致的前提下,仅通过不同的随机数据划分与参数初始化引入差异性。综合计算成本与性能提升幅度,最终选取3个子模型进行集成,其预测结果通过简单平均得到。该集成策略能够在一定程度上抵消单个模型的预测误差,有助于降低整体预测误差。实验表明,在当前数据规模下,少量子模型的集成已能够有效降低预测误差并改善泛化性能。

1.5 双分支并联残差GNN模型

并联残差模型是一种通过引入误差修正机制来提高预测精度的建模方法。在传统化工领域,该思想常用于将机理模型与数据驱动模型相结合,相较于单独使用任一模型,这两种模型的结合在预测精度、可解释性和模型复杂度等方面具有显著优势^[24-26]。从更广义的机器学习角度来看,残差学习也可以在纯数据驱动模型之间实施,即通过构建两阶段预测结构,使第二阶段模型专门学习第一阶段模型的预测残差分布,从而降低整体误差。

近年来,将贝叶斯矩阵补全方法^[27]与UNIFAC-Dortmund模型^[3, 28]相结合的预测框架被提出,该框架可用于推导无限稀释活度系数^[29]。该方法基于实验测得的溶剂-溶质 $\ln\gamma_i^\infty$ 数据构建稀疏二元体系矩阵,其中矩阵的行和列分别对应溶剂和溶质,矩阵元素为实验活度系数值。尽管该方法在训练集

已涵盖的已知溶质-溶剂组分对体系中表现良好,但其性能仍受到实验数据覆盖范围及矩阵稀疏性的限制,在独立测试集上的系统对比研究仍相对不足。相比之下,混合残差架构在化学体系中的应用更为广泛,其最典型的应用模式为并联残差模型^[30-31],具体的双分支并联残差 GNN 模型如图 3 所示(以交互作用模型参数 A_{ij} 为例),其优势在于实现方式灵活。在该类模型中,第二个机器学习模型用于学习初始 GNN 模型预测误差的分布特征,并将其预测残差与第一阶段模型的输出相加,从而实现对原始预测结果的修正与提升。这种逐步校正误差的建模思想在机器学习中也被称为提升方法(boosting)^[32-33]。

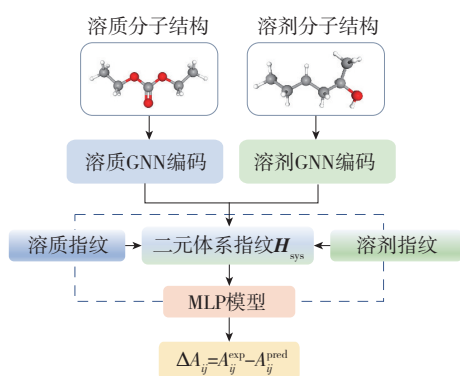


图3 双分支并联残差 GNN 模型示意图

Fig. 3 Schematic diagram of the dual-branch parallel residual GNN model

为了考察并联残差模型在交互作用模型参数预测中的应用潜力,将所提出的基于 GNN 的模型与机理模型进行耦合,构建并联残差预测框架。首先通过 GNN 将模型残差的预测结果与模型本身的预测值相加,得到交互作用模型参数预测结果。与单一 GNN 模型相比,并联残差模型在多个体系上的预测误差更低。此处采用的 GNN 模型以由 3 个独立 GNN 组成的集成模型形式进行训练,其训练目标不再是直接预测交互作用模型参数,而是学习模型的交互作用模型参数残差,其定义如式(7)所示(以 ΔA_{ij} 为例)。

$$\Delta A_{ij} = A_{ij}^{\text{exp}} - A_{ij}^{\text{pred}} \quad (7)$$

式中: A_{ij}^{exp} 和 A_{ij}^{pred} 分别表示 A_{ij} 的实验值和模型预测值。

1.6 数据集与模型性能评价

1.6.1 数据集

本研究采用的交互作用模型参数数据集来自

化工流程模拟软件 Aspen^[34],对应的无限稀释活度系数数据来源于 Wilson^[35]的研究。数据集中给出了 Wilson 模型下的交互作用模型参数 (A_{ij} 、 A_{ji} 、 B_{ij} 、 B_{ji}) (式(2))、无限稀释活度系数 ($\ln \gamma_1^\infty$ 、 $\ln \gamma_2^\infty$) 以及基于 UNIFAC 方法计算得到的 $\ln \gamma_i^\infty$ 。由于 Wilson 与 UNIFAC 模型在理论假设和参数来源方面存在差异,因此本研究同时保留这两类结果以便从不同热力学框架下评估数据的一致性与合理性。

在数据构建过程中,最初共收集到 4 640 组溶质-溶剂二元体系数据。对原始数据集进行规范的清洗与筛选:首先剔除无法获取有效 SMILES 表达式的结构缺失样本 23 组,随后采用中位数聚合方法^[36]合并重复体系 495 组,然后结合参数分布特征剔除异常值 54 组,最终获得 4 068 组可靠数据用于后续模型训练与性能评价。

所用数据集的交互作用模型参数的概率密度分布如图 4 所示。在原始数据收集中,各参数的取值范围较宽,分别为: A_{ij} (-2.5, 2.6), A_{ji} (-9.7, 2.4), B_{ij} (-2 173, 778), B_{ji} (-2 598, 2 720)。经数据清洗和异常极端值处理后,参数范围调整如下: A_{ij} (-2.06, 2.06), A_{ji} (-2.06, 2.06), B_{ij} (-1 513, 565), B_{ji} (-1 835, 799)。清洗后的参数分布整体呈现非均匀特征,多数样本集中于中等数值区间,而尾部区域样本数量相对较少,该分布形态客观反映了当前公开的热力学数据库中二元交互作用参数的实际统计特征。在数据预处理过程中基于分布特征与误差分析结果进行样本筛选,在保证物理合理性的前提下优化数据质量。最终数据集在样本规模与数据可靠性之间实现平衡,为后续基于 GNN 的交互作用模型参数预测模型构建提供了稳定的数据基础。

1.6.2 模型性能评价

对模型性能的评价并非依赖单一维度,而是围绕多个评估维度展开,具体包括平均绝对误差(MAE)、预测误差的标准偏差(SDEP)、均方误差(MSE)、均方根误差(RMSE)、决定系数(R^2)及平均绝对相对偏差(AARD),各指标的计算公式如式(8)~(13)所示。其中,MAE、MSE 和 RMSE 主要衡量模型预测值与实验值之间的绝对误差与整体误差水平,SDEP 进一步考察误差分布的集中程度, R^2 和 AARD 分别从拟合优度和相对误差角度对结果进行评估。上述指标从多角度剖析比较了单一 GNN 模型与并联残差 GNN 模型在预测交互作用模

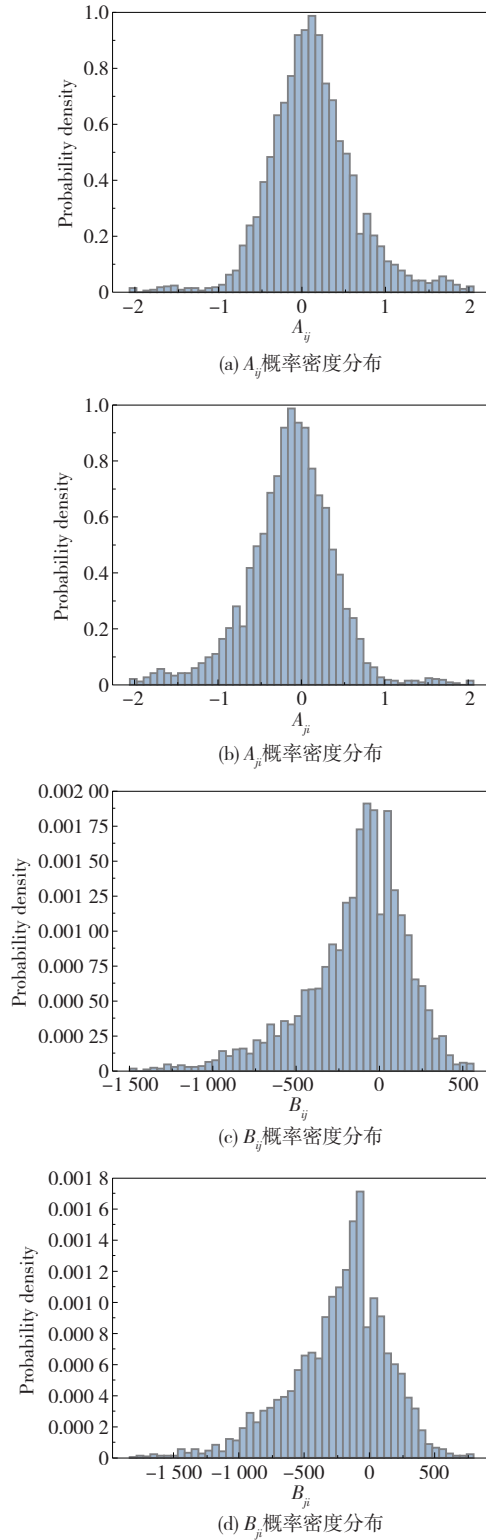


图 4 交互作用模型参数的概率密度分布

Fig. 4 Probability density distribution of the interaction model parameters

型参数时的性能差异。

$$E_{\text{MAE}} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (8)$$

$$P_{\text{SDEP}} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left[(y_i - \hat{y}_i) - \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i) \right]^2} \quad (9)$$

$$E_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (10)$$

$$E_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (12)$$

$$D_{\text{AARD}} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (13)$$

式中: y_i 与 $\hat{y}_i$ 分别表示第 i 个样本的实验基准值与模型预测值, $\bar{y}$ 为实验基准值的平均值, N 为样本总数。

2 结果与讨论

2.1 不同模型的预测性能比较

UNIFAC 模型、单一 GNN 模型和并联残差 GNN 模型的预测性能汇总于表 4 中。其中, 所有性能指标均基于未缩放的 A_{ij} 、 A_{ji} 、 B_{ij} 、 B_{ji} 数值计算得到。引入 UNIFAC 模型作为对比基准, 该模型基于基团贡献理论, 通过分子中各功能基团的组成与基团间相互作用参数计算 $\ln \gamma_i^\infty$, 然后将其代入 Wilson 方程中, 通过求解非线性方程组反向解析出对应的二元交互作用模型参数 A_{ij} 、 A_{ji} 、 B_{ij} 、 B_{ji} 。

表 4 中, 单一 GNN 模型及并联残差 GNN 模型的各项评价指标均对应于在不同随机训练/验证集划分条件下独立训练 3 个独立子模型所得预测结果的集成平均值, 这种集成策略旨在通过平滑单一网络的随机波动来提升整体预测的鲁棒性。GNN 模型与 UNIFAC 模型的关键区别在于 GNN 模型能够覆盖整个数据集 (100%), 而 UNIFAC 模型通常仅限于其适用的化学空间范围。UNIFAC 系列模型仅适用于具有相应 UNIFAC 基团二元相互作用参数的体系, 已有研究表明, 这种对参数可用性的依赖会将模型的适用范围限制在特定区域内^[37]。UNIFAC 模型的性能指标基于其可预测的完整可行数据集进行统计, 而 GNN 模型的结果分别针对训练集、验证集和测试集单独给出。此外, 为便于比较, 报告了仅在训练集上训练得到的 GNN 模型在完整数据集上的预测性能, 并在表 4 中以“complete”标注。从测试集结果来看, 单一 GNN 模型的 MAE 指标远低于 UNIFAC 模型, 在其他评价指标上也表现出更优的预测性能, 因此可以认为单一 GNN 模型在高度非理

表4 不同模型的预测性能比较

Table 4 Comparison of prediction performance for different models

模型	参数	数据集	AARD/%	MAE	R^2	SDEP	MSE	RMSE
UNIFAC	A_{ij}	complete	410.29	0.51	0.05	1.69	2.85	1.69
	A_{ji}	complete	493.48	0.61	0.03	2.59	6.73	2.60
	B_{ij}	complete	315.99	282.03	-0.04	591.43	349 974.00	591.59
	B_{ji}	complete	245.02	331.53	0.00	829.38	689 002.70	830.06
单一 GNN	A_{ij}	train	30.22	0.05	0.98	0.05	0.01	0.07
		valid	36.73	0.06	0.96	0.08	0.01	0.09
		test	35.32	0.06	0.96	0.08	0.01	0.10
		complete	31.77	0.05	0.98	0.06	0.01	0.08
	A_{ji}	train	28.85	0.05	0.98	0.05	0.01	0.07
		valid	34.84	0.06	0.97	0.07	0.01	0.09
		test	36.26	0.06	0.96	0.08	0.01	0.10
		complete	30.81	0.05	0.98	0.06	0.01	0.08
	B_{ij}	train	96.10	146.82	0.58	132.30	39 067.34	197.63
		valid	114.01	171.81	0.43	158.76	54 731.46	233.94
		test	117.02	177.33	0.39	163.66	58 230.86	241.31
		complete	101.72	154.92	0.53	141.96	44 157.22	210.13
	B_{ji}	train	87.55	165.48	0.63	151.90	50 478.69	224.63
		valid	107.04	213.17	0.41	201.63	86 173.27	293.42
		test	113.05	216.61	0.39	206.70	89 647.69	299.41
		complete	94.22	179.54	0.56	170.12	61 176.82	247.33
并联残差 GNN	A_{ij}	train	25.18	0.04	0.98	0.05	0.00	0.07
		test	29.60	0.05	0.97	0.08	0.01	0.10
		complete	26.06	0.05	0.98	0.06	0.01	0.07
	A_{ji}	train	25.70	0.05	0.98	0.05	0.00	0.07
		test	32.86	0.06	0.97	0.08	0.01	0.10
		complete	27.13	0.05	0.98	0.06	0.01	0.08
	B_{ij}	train	94.46	146.46	0.58	133.45	39 257.67	198.14
		test	112.53	174.08	0.40	161.90	56 516.71	237.73
		complete	98.08	151.99	0.55	140.04	42 711.17	206.67
	B_{ji}	train	86.61	166.15	0.62	155.91	51 913.25	227.84
		test	110.07	212.14	0.41	204.10	86 660.83	294.38
		complete	91.30	175.35	0.58	167.68	58 866.18	242.62

表中数据均为各模型预测值与 Wilson 交互作用模型参数实验值的比较结果。

想体系上的预测能力高于 UNIFAC 模型。在这 3 种模型中,并联残差 GNN 模型表现出最优的预测性能,其不仅在 R^2 和 AARD 指标上表现最佳,而且在 MAE、MSE 等绝对误差指标上也优于 UNIFAC 和单一 GNN 模型。

表 4 中,各模型对 B_{ij} 和 B_{ji} 预测的 MSE 显著大于 A_{ij} 和 A_{ji} ,这主要是由参数的物理量级与数据分布跨度决定的。Wilson 方程中 B_{ij} 参数代表能量相互作用项,其绝对数值通常高达数百至上千(跨度为 -1 513 至 799),远大于量级为个位数的 A_{ij} 参数。由

于MSE对绝对误差具有平方放大效应,因此 B_{ij} 参数的MSE指标呈现出较大数值。由于所涵盖的二元体系极性差异巨大,并联残差GNN模型在进行全局统一拟合时,边缘极值样本也会贡献较大的绝对残差。然而,后续反算 γ_i^∞ 的结果表明,这种因参数量级带来的绝对误差并未显著影响最终热力学性质计算的可靠性。

图5给出了并联残差GNN模型与UNIFAC模型在所有可行体系上的对比一致性(以 A_{ij} 为例)。图5显示,UNIFAC模型在部分体系中存在明显的系统性偏差,相对于理想对角线呈现出一定程度的偏移,尤其在参数取值范围较宽的区域偏移更为明显,这一现象与UNIFAC模型对基团参数和相互作用参数依赖较强、适用范围受限的特点密切相关。而并联残差GNN模型的预测结果在整体范围内更紧密地分布于理想预测线(对角线)附近,离群点数量明显减少,并且在高值与低值区间均保持较为一致的预测精度,其预测结果相对于理想模型的整体离散程度明显低于UNIFAC模型。并联残差建模并未显著改变模型对参数整体趋势的拟合能力,而是主要体现在误差分布的收敛性提升方面。这说明在交互作用模型参数预测任务中,主模型已能够有效学习分子结构与参数之间的主导映射关系,而残差网络主要用于修正局部偏差与非均匀误差结构。因此,并联残差GNN模型在Wilson交互作用模型参数预测任务中相比单一GNN模型提升了预测精度与稳定性,尤其在误差分布收敛性与极端样本预测方面表现更为突出。

在交互作用模型参数预测中,基于分子结构的

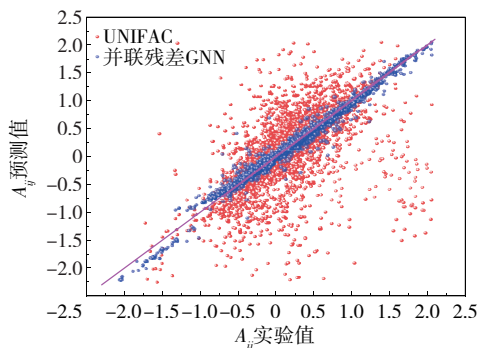


图5 UNIFAC模型与并联残差GNN模型的 A_{ij} 实验值与预测值对比

Fig. 5 Comparison of experimental and predicted A_{ij} values for the UNIFAC model and the parallel residual GNN model

GNN模型能够直接建立溶质-溶剂结构与Wilson参数之间的关联性,而无需依赖基团划分与经验相互作用参数。与UNIFAC-Wilson路径相比,单一GNN模型在整体预测精度和参数分布稳定性方面均具有显著优势。引入并联残差模型后,GNN模型对局部误差和极端样本的预测精度显著提高,这表明在当前数据条件下,参数预测误差具有一定的结构性特征,可以通过逐步建模进行有效修正。

2.2 无限稀释活度系数计算结果比较

为了进一步验证所预测参数在热力学计算中的有效性与一致性,将并联残差GNN模型预测的交互作用模型参数代入Wilson方程,重新计算 $\ln\gamma_i^\infty$,并与实验值及传统UNIFAC方法进行对比。将上述基于GNN-Wilson路径与传统UNIFAC方法计算得到的两组 $\ln\gamma_i^\infty$ 分别与基于Wilson原始模型参数计算得到的 $\ln\gamma_i^\infty$ 实验值进行系统对比,以评估不同方法在最终热力学性质预测层面的表现,结果见表5。结果表明,基于GNN预测参数并经Wilson方程反求解得到的 $\ln\gamma_i^\infty$ 与Wilson交互作用模型参数实验值在整体趋势和数值水平上保持较好的一致性,其AARD为240.22%($\ln\gamma_1^\infty$)和216.56%($\ln\gamma_2^\infty$)。相比之下,UNIFAC方法在部分体系中表现出较大偏差,尤其是在参数分布较为极端或体系非理想性较强的区域,其预测误差更为明显,AARD为250.65%($\ln\gamma_1^\infty$)和267.08%($\ln\gamma_2^\infty$)。这与前文中交互作用模型参数预测结果的分析一致。

从误差分布特征来看,基于GNN-Wilson路径获得的 $\ln\gamma_i^\infty$ 在整体误差范围和离群点数量上均优于UNIFAC结果,说明通过分子结构直接预测交互作用模型参数,再借助Wilson模型进行热力学计算,是一条可行且稳定的计算路径。这同时表明,GNN所学习到的参数映射关系并非仅在数值层面成立,而是能够在后续热力学模型计算中保持物理一致性。基于GNN模型的核心优势是避免了传统Wilson参数反演过程中对非线性方程组求解的依赖,从而显著降低了数值不稳定的风险。同时,这种方法不受UNIFAC模型中基团参数和基团交互参数完整性的限制,因此其提供了一种更灵活的方法来预测无限稀释活度系数。

3 结论

(1)本研究将双分支GNN与并联残差二次建模

表5 并联残差GNN模型和UNIFAC模型对 $\ln\gamma_i^\infty$ 的计算结果比较Table 5 Comparison of the calculation results of the parallel residual GNN model and UNIFAC model for $\ln\gamma_i^\infty$

模型	$\ln\gamma_i^\infty$	AARD/%	MAE	R^2	SDEP	MSE	RMSE
并联残差 GNN	$\ln\gamma_1^\infty$	240.22	0.46	0.58	0.65	0.47	0.68
	$\ln\gamma_2^\infty$	216.56	0.55	0.63	1.00	0.67	0.82
UNIFAC	$\ln\gamma_1^\infty$	250.65	0.49	0.49	0.83	0.70	0.84
	$\ln\gamma_2^\infty$	267.08	0.62	0.27	1.00	1.39	1.18

表中数据均为各模型预测值与 Wilson 交互作用模型参数实验值的比较结果。

有机结合,实现了对二元交互作用模型参数的高效、稳健预测。其中,并联残差架构能够显式捕捉局部偏差,显著增强参数预测在分布边缘区域的鲁棒性。

(2)将模型预测的交互作用参数代入 Wilson 方程反向解析出无限稀释活度系数 γ_i^∞ ,其整体计算精度和跨体系泛化能力优于传统的 UNIFAC 估算方法,验证了智能预测参数在热力学性质计算中的物理合理性与可靠性。

本研究提出的双分支 GNN 结合并联残差机制的 Wilson 交互作用模型参数预测框架,一方面打破了传统基团贡献法(如 UNIFAC)对未知功能基团或缺失相互作用参数的依赖,极大地拓宽了复杂体系的物性预测边界;另一方面完全规避了传统方法依靠汽液相平衡实验数据进行反向回归时,求解器易陷入初值敏感、不收敛及多解性的数值不稳定瓶颈。

本模型主要适用于常规有机二元体系,该数据驱动模型的预测精度高度依赖训练集对特定组分的覆盖程度。对于训练集中未涵盖的稀有元素体系、极复杂大分子或具有强离子相互作用的特殊体系,其预测结果的可靠性需结合实验数据进一步进行评估。此外,当前模型尚未显式考虑温度变化的影响以及严格的热力学一致性约束,后续将重点探索在网络结构中嵌入物理先验约束,同时挖掘模型在特定化学体系中的深层功能。未来本模型将逐步推广至更复杂的多组分及电解质体系,以进一步提升智能热力学模型的工业应用价值。

参考文献:

- [1] LI Z, CONSTANTINOU L, BAUR R, et al. Review of group contribution methods for prediction of thermodynamic properties of long-chain hydrocarbons[J]. Molecular Physics, 2025, 123(21-22): e2563020.
- [2] GMEHLING J, CONSTANTINESCU D, SCHMID B. Group contribution methods for phase equilibrium calculations[J]. Annual Review of Chemical and Biomolecular Engineering, 2015, 6: 267-292.
- [3] HAYER N, WENDEL T, MANDT S, et al. Advancing thermodynamic group-contribution methods by machine learning: UNIFAC 2.0[J]. Chemical Engineering Journal, 2025, 504: 158667.
- [4] CLETO P B, DURAU V G, KANDA L R S, et al. Phase equilibrium measurements and thermodynamic modeling of CO₂+guaicol, CO₂+guaicol and ethanol, and CO₂+guaicol, ethanol, and water[J]. Fluid Phase Equilibria, 2025, 591: 114319.
- [5] MARCILLA A, REYES-LABARTA J A, OLAYA M M. Should we trust all the published LLE correlation parameters in phase equilibria? Necessity of their assessment prior to publication[J]. Fluid Phase Equilibria, 2017, 433: 243-252.
- [6] DA CRUZ SILVA K, ABREU C S, VIEIRA A W, et al. Aqueous two-phase systems formed by different phase-forming components: equilibrium diagrams and dye partitioning study[J]. Fluid Phase Equilibria, 2020, 520: 112664.
- [7] WINTER B, WINTER C, ESPER T, et al. SPT-NRTL: a physics-guided machine learning model to predict thermodynamically consistent activity coefficients[J]. Fluid Phase Equilibria, 2023, 568: 113731.
- [8] MEDINA E I S, LINKE S, STOLL M, et al. Graph neural networks for the prediction of infinite dilution activity coefficients[J]. Digital Discovery, 2022, 1(3): 216-225.
- [9] WAIKHOM L, PATGIRI R. Graph neural networks: methods, applications, and opportunities[PP/OL]. V2. arXiv (2021-09-08) [2026-03-04]. <https://doi.org/10.48550/arXiv.2108.10733>.
- [10] RITTIG J G, GAO Q, DAHMEN M, et al. Graph neural networks for the prediction of molecular structure-prop-

- erty relationships [PP/OL]. V1. arXiv (2022-07-25) [2026-03-04]. <https://doi.org/10.48550/arXiv.2208.04852>.
- [11] SHEN X, PAN S, CHOI K S, et al. Corrigendum to “domain-adaptive message passing graph neural network” [Neural Netw. 164 (2023) 439–454] [J]. Neural Networks, 2023, 168: 337–338.
- [12] WANG Y, LI Z, FARIMANI A B. Graph neural networks for molecules [PP/OL]. V2. arXiv (2023-02-06) [2026-03-04]. <https://doi.org/10.48550/arXiv.2209.05582>.
- [13] QU C, SCHNEIDER B I, KEARSLEY A J, et al. Applying graph neural network models to molecular property prediction using high-quality experimental data [J]. Artificial Intelligence Chemistry, 2024, 2(1): 100050.
- [14] MURGOD T R, REDDY P S, GADDAM S, et al. A survey on graph neural networks and its applications in various domains [J]. SN Computer Science, 2025, 6: 26.
- [15] ZHANG S, TONG H, XU J, et al. Graph convolutional networks: a comprehensive review [J]. Computational Social Networks, 2019, 6: 11.
- [16] THEISEN M F, MEESTERS G M H, SCHWEIDTMANN A M. Graph neural networks for soft sensors: learning from process topology and operational data [J]. Computers and Chemical Engineering, 2026, 206: 109532.
- [17] BENTO A P, HERSEY A, FÉLIX E, et al. An open source chemical structure curation pipeline using RDKit [J]. Journal of Cheminformatics, 2020, 12: 51.
- [18] WU J N, WANG T, CHEN Y, et al. t-SMILES: a fragment - based molecular representation framework for de novo ligand design [J]. Nature Communications, 2024, 15: 4993.
- [19] HASHEMIAN S, KHAN Z, KALHAN P, et al. Learning SMILES semantics: Word2Vec and transformer embeddings for molecular property prediction [J]. Algorithms, 2025, 18: 547.
- [20] SCHWEIDTMANN A M, RITTIG J G, KÖNIG A, et al. Graph neural networks for prediction of fuel ignition quality [J]. Energy & Fuels, 2020, 34(9): 11395–11407.
- [21] GILMER J, SCHOENHOLZ S S, RILEY P F, et al. Neural message passing for quantum chemistry [C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, 2017: 1263–1272.
- [22] SHEN C, WANG Q, PRIEBE C E. One - hot graph encoder embedding [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(6): 7933–7938.
- [23] FEY M, LENSSEN J E. Fast graph representation learning with PyTorch Geometric [PP/OL]. V3. arXiv (2019-04-25) [2026-03-04]. <https://doi.org/10.48550/arXiv.1903.02428>.
- [24] VON STOSCH M, OLIVEIRA R, PERES J, et al. Hybrid semi-parametric modeling in process systems engineering: past, present and future [J]. Computers and Chemical Engineering, 2014, 60: 86–101.
- [25] GLASSEY J, VON STOSCH M. Hybrid modeling in process industries [M]. Boca Raton: CRC Press, 2018: 19–44.
- [26] MCBRIDE K, MEDINA E I S, SUNDMACHER K. Hybrid semi - parametric modeling in separation processes: a review [J]. Chemie Ingenieur Technik, 2020, 92(7): 842–855.
- [27] YUCHI H S, MAK S, XIE Y. Bayesian uncertainty quantification for low - rank matrix completion [PP/OL]. V4. arXiv (2022-03-25) [2026-03-04]. <https://doi.org/10.48550/arXiv.2101.01299>.
- [28] CONSTANTINESCU D, GMEHLING J. Further development of modified UNIFAC (Dortmund): revision and extension 6 [J]. Journal of Chemical & Engineering Data, 2016, 61(8): 2738–2748.
- [29] JIRASEK F, BAMLER R, MANDT S. Hybridizing physical and data-driven prediction methods for physico-chemical properties [J]. Chemical Communications, 2020, 56(82): 12407–12410.
- [30] LV J, WANG L. Hybrid modeling of adsorption process using mass transfer and machine learning techniques for concentration prediction [J]. Journal of Saudi Chemical Society, 2025, 29: 12.
- [31] GARGALO C L, MALANCA A A, AOUICHAOU A R N, et al. Navigating industry 4.0 and 5.0: the role of hybrid modelling in (bio)chemical engineering’s digital transition [J]. Frontiers in Chemical Engineering, 2024, 6: 1494244.
- [32] MIENYE I D, SUN Y. A survey of ensemble learning: concepts, algorithms, applications, and prospects [J]. IEEE Access, 2022, 10: 99129–99149.
- [33] ZHAO H K, LIU W H, WANG Y X, et al. Comparative analysis of algorithmic approaches in ensemble learning: bagging vs. boosting [J]. Scientific Reports, 2025, 15: 34218.
- [34] WOOLEY R J, PUTSCHE V. Development of an ASPEN PLUS physical property database for biofuels components [R]. Golden: National Renewable Energy Laboratory

- (NREL), 1996.
- [35] WILSON G M. Vapor-liquid equilibrium. XI. A new expression for the excess free energy of mixing[J]. Journal of the American Chemical Society, 1964, 86(2): 127-130.
- [36] DOMINGO-FERRER J, TORRA V. Median-based aggregation operators for prototype construction in ordinal scales[J]. International Journal of Intelligent Systems, 2003, 18(6): 633-655.
- [37] BROUWER T, SCHUUR B. Model performances evaluated for infinite dilution activity coefficients prediction at 298.15 K [J]. Industrial & Engineering Chemistry Research, 2019, 58(20): 8903-8914.

Prediction model of interaction model parameters based on graph neural networks (GNN)

MENG XiangMing¹ ZHAO WenYing¹ SUN XiaoYan¹ TAO ShaoHui^{1*}
CHEN YuShi² LI Feng²

(1. Institute of Process System Engineering, College of Chemical Engineering, Qingdao University of Science and Technology, Qingdao 266042;
2. Petro-Cyber Works Information Technology Co., Ltd., Shanghai 200040, China)

Abstract: The interaction model parameters are the key thermodynamic parameters in activity coefficient models such as Wilson and non-random two-liquid (NRTL) models. Their accuracy directly determines the reliability of the results of vapor - liquid phase equilibrium and separation process calculations. The core bottlenecks of traditional parameter acquisition methods relying on experimental data regression and universal functional activity coefficient (UNIFAC) model estimation include limited data coverage, unstable solution of nonlinear equations and limited scope of application. To address these issues, we constructed a dual-branch graph neural network (GNN) model based on molecular structure characterization enabling efficient and direct prediction of the parameters of the Wilson interaction model. This model takes molecular diagrams as input, separately extracts the molecular features of solutes and solvents, and, through the message-passing mechanism, deeply learns the potential structural correlations between molecules. At the same time, we introduce a parallel residual strategy to reduce the prediction error. Using 4 068 binary system interaction model parameter data for training and testing, our proposed parallel residual GNN model significantly outperforms the UNIFAC method across multiple error indicators, and compared with the single GNN model, the prediction accuracy is further improved. The predicted parameters A_{ij} , A_{ji} , B_{ij} , B_{ji} of the Wilson interaction model were substituted into the Wilson equation to calculate the infinite dilution activity coefficient γ_i^∞ . The results were highly consistent with the calculated value in the literature, verifying the thermodynamic rationality and engineering applicability of our model.

Key words: interaction model parameters; graph neural networks (GNN); Wilson model; UNIFAC model; infinite dilution activity coefficient

(责任编辑:于少云)