華東理工大學学报 自然科学版 Natural Science Edition
JOURNAL OF EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY

基于注意力等变几何扩散模型的完整分子生成

张百川 谢博 张立国 罗健旭

Complete Molecular Generation Based on Attention-Equivariant Geometric Diffusion Models

ZHANG Baichuan, XIE Bo, ZHANG Liguo, LUO Jianxu

在线阅读 View online: http://doi.org/10.14135/j.cnki.1006-3080.20251024001

您可能感兴趣的其他文章

Articles you may be interested in

一种基于冷扩散模型的复杂反应流场建模方法

A Modeling Method for Complex Turbulent Reactive Flow Field Based on Cold Diffusion Model

华东理工大学学报（自然科学版）. 2024, 50(5): 720-731 http://doi.org/10.14135/j.cnki.1006-3080.20231010001

基于多通道图神经网络和CNN-BiLSTM的漏洞检测方法

Vulnerability Detection Method Based on Multi-Channel Graph Neural Network and CNN-BiLSTM

华东理工大学学报（自然科学版）. 2025, 51(6): 835-842 http://doi.org/10.14135/j.cnki.1006-3080.20250317001

融合建模的图神经网络会话推荐模型

Graph Neural Network Session Recommendation Model with Fusion Modeling

华东理工大学学报（自然科学版）. 2025, 51(6): 827-834 http://doi.org/10.14135/j.cnki.1006-3080.20250225002

基于单模态生理信号无监督特征学习的驾驶压力识别

Driving Stress Detection Based on Unsupervised Feature Learning of Single Module Physiological Signal

华东理工大学学报（自然科学版）. 2021, 47(4): 475-482 http://doi.org/10.14135/j.cnki.1006-3080.20200728001

基于多源域深度域自适应的脑力负荷识别

Mental Workload Recognition Based on Multi-Source Deep Domain Adaptation

华东理工大学学报（自然科学版）. 2023, 49(5): 744-753 http://doi.org/10.14135/j.cnki.1006-3080.20220723002

基于双层自适应集成残差主成分分析的复杂非线性过程监测

Two-Layer Adaptive Ensemble Residual Principal Component Analysis for Complex Nonlinear Process Monitoring

华东理工大学学报（自然科学版）. 2024, 50(1): 88-96 http://doi.org/10.14135/j.cnki.1006-3080.20221210001

扫码关注公众号，获取更多信息！

文章编号：1006-3080(2026)04-0478-11 DOI：10.14135/j.cnki.1006-3080.20251024001

基于注意力等变几何扩散模型的完整分子生成

张百川[1]， 谢 博[1]， 张立国[2]， 罗健旭[1]
（华东理工大学 1. 信息科学与工程学院； 2. 化学与分子工程学院, 上海 200237）

摘要：针对分子生成任务，本文提出了一种注意力等变几何扩散模型(AEGDM)，该模型应用离散扩散与连续扩散混合生成扩散框架，实现原子坐标、原子种类及化学键的完整生成。在AEGDM模型中，注意力等变图神经网络(Attention Equivariant Graph Neural Network，AEGNN)与残差增强神经网络(Residual Boosting Neural Network，RBNN)分别用以提高模型的表征能力及加快模型训练速度。在QM9数据集上的实验结果表明，该模型在生成分子的唯一性与新颖性方面均取得显著提升，唯一性达到98.2%，新颖性达到70.9%。这一性能提升表明AEGDM在化学空间探索方面具有较强的表征能力，能够有效推动创新性候选分子的发展。

关键词：扩散模型；图神经网络；分子生成；变压器；集成学习

中图分类号：TP391.41 **文献标志码：**A

分子构象作为分子在三维空间中的存在形式，其几何构型直接决定了物质的物理特性、化学性质及生物活性特征[1]。分子构象研究不仅是解析分子行为机制的基础科学问题，更是药物设计、材料开发及计算生物学等领域的关键技术支撑[2]。蒙特卡罗(Markov Chain Monte Carlo, MCMC)或分子动力学(Molecular Dynamics, MD)[3]的传统方法计算成本高昂，存在周期冗长且探索空间受限等固有瓶颈。为此，研究者引入深度学习技术，借助其强大的表示学习与高效推理能力，显著提升了分子模拟与采样任务的工作效率与可扩展性[4]。

为了有效学习局部化学环境与全局拓扑模式，在深度学习中，2D分子拓扑特征通常通过图神经网络(Graph Neural Network, GNN)进行编码，将分子抽象为由原子节点与化学键边构成的图结构[5]。3D空间信息则依托几何深度学习架构进行建模，显式引入原子坐标、键角、二面角等几何约束，确保模型在欧几里得群变换下保持严格的等变或不变性质。代表性方法包括SchNet[6]、DimeNet[7]及SE(3)-Transformer[8]等，这些工作推动了分子表征范式的根本性变革。随着人工智能与计算化学的深度融合，传统依赖单一维度分子表征的范式已难以满足复杂生物系统的多尺度建模需求。融合2D信息与3D构象的多模态表征方法，凭借其化学可解释性与物理真实性的协同优势，正成为精准预测分子性质与相互作用的前沿方向。

近年来，生成式人工智能的突破性进展为分子设计范式革新提供了全新机遇[9]。在分子生成领域，常见的生成范式有变分自编码器(Variational Autoencoder, VAE)[10]、扩散模型(Denoising Diffusion Probabilistic Models, DDPM)[11]以及流模型(Flow Model)[12]等。VAE与Flow Model虽具备较快的采样速度，但在生成多样性和保真度方面存在一定局限；相比之下，DDPM凭借其强大的分布建模与渐进式去噪能力，在复杂分子构型空间的学习与高质量分子生成任务中展现出显著优势。

Xu等[13]首次将扩散框架引入分子构象生成，通过引入等变图神经网络(Equivariant Graph Neural Network, EGNN)[14]保持旋转等变性，并将坐标约束于零质心子空间以实现平移不变性，成功构建了

收稿日期：2025-10-24

作者简介：张百川(2000—)，男，安徽定远人，硕士生，主要研究方向为基于深度学习的分子生成。E-mail: z2655065175@163.com

通信联系人：罗健旭，E-mail: Jxluo@ecust.edu.cn

引用本文：张百川, 谢 博, 张立国, 等. 基于注意力等变几何扩散模型的完整分子生成 [J]. 华东理工大学学报(自然科学版), 2026, 52(4): 478-488.

Citation: ZHANG Baichuan, XIE Bo, ZHANG Liguo, *et al*. Complete Molecular Generation Based on Attention-Equivariant Geometric Diffusion Models[J]. Journal of East China University of Science and Technology, 2026, 52(4): 478-488.

首个几何扩散框架, 后续研究普遍沿袭其核心设计原则, 即严格保持等变性与不变性约束以提升生成质量。然而, 此开创性工作存在本质局限: 仅能生成分子坐标空间结构, 未涉及原子种类预测这一关键要素。

针对上述局限性, Hoogeboom 等[15] 提出了等变扩散模型(Equivariant Diffusion Model, EDM), 将原子类型的联合预测与三维坐标生成过程统一于一个旋转等变的生成框架中。该方法通过在扩散过程中显式建模原子种类与空间坐标的协同演化, 实现了对分子图结构与几何构型的联合优化, 从而有效提升了生成分子的化学合理性与三维构象的真实性。Xu 等[16] 提出了联合生成原子类型与三维坐标的扩散框架, 通过编码器将原子类型特征与空间坐标映射至潜在空间进行扩散处理, 再利用解码器重构原子种类与坐标, 其中编码器、解码器以及去噪模型均为等变图神经网络, 以此保持旋转等变性。该方法通过变分自编码器有效提升了离散数据的生成能力, 在有效性等指标上取得突破性进展。但该框架未能显式建模原子间的成键关系, 仅能通过距离阈值推断连接性, 导致生成的分子拓扑结构完整性不足。Xu 等[17] 在此基础上提出双轨注意力模型(DTN), 通过改进噪声预测模型的特征提取机制, 在保持生成质量的同时提升了模型鲁棒性。

Vignac 等[18] 首次实现了原子类型-3D 坐标-化学键的联合生成, 构建了首个端到端完整分子扩散系统, 在完整分子生成领域取得重要突破。同期, 文献 [19-20] 分别提出类似的完整分子生成方法, 共同推进了完整分子生成技术的发展。但完整分子生成仍面临双重挑战: 首先, 现有模型大部分仅针对节点特征与坐标特征进行提取, 而化学键的特征提取与生成对模型表征能力提出了更高要求; 其次, 生成任务的扩展显著增加了模型的计算复杂度——模型需额外处理化学键特征的提取与生成, 导致训练时间大幅增加, 亟需开发高效的训练加速方法。

针对上述两个问题, 本文提出了注意力等变几何扩散模型(Attention Equivariant Geometric Diffusion Model, AEGDM), 该模型由注意力等变图神经网络(Attention Equivariant Graph Neural Network, AEGNN)和残差增强神经网络(Residual Boosting Neural Network, RBNN)组成。其中, AEGNN 用于提升模型的表征能力, 而 RBNN 通过增强训练过程的稳定性、加快收敛速度, 从而提升模型的训练效率。

1　完整分子生成

1.1　分子生成

分子生成是指利用计算模型自动设计具有特定化学结构与功能属性的分子的过程。其核心目标是通过算法探索化学空间, 产生新颖、稳定且符合需求的分子构型, 为药物设计、材料发现等领域提供候选化合物。分子生成分为 3 种: (1)仅生成坐标; (2)同时生成坐标与原子种类; (3)同时生成坐标、原子种类以及化学键。本文将第 3 种称为完整分子生成, 并选择其作为所进行的生成任务。分子图 $\boldsymbol{G}=(\boldsymbol{X},\boldsymbol{C},\boldsymbol{E})$, 其中, $\boldsymbol{X}$ 与 $\boldsymbol{C}$ 为节点特征, $\boldsymbol{E}$ 为边特征; $\boldsymbol{X}\in R^{N\times 3}$, N 为原子数目, $\boldsymbol{X}_i$ 为第 i 个原子的坐标; $\boldsymbol{C}\in R^{N\times d}$, d 为特征数, $\boldsymbol{C}_i$ 为第 i 个原子的类型特征; $\boldsymbol{E}$ 为化学键, $\boldsymbol{E}_{ij}\in R^{N\times N}$, 代表第 i 个原子与第 j 个原子是否相连以及相连的键类型。分子图如图 1 所示, 分子坐标由 N 个原子三维坐标组成, 原子类型数据由原子种类的 one-hot 编码与原子电荷特征组成, 化学键表示为对应原子是否相连以及相连的键类型。

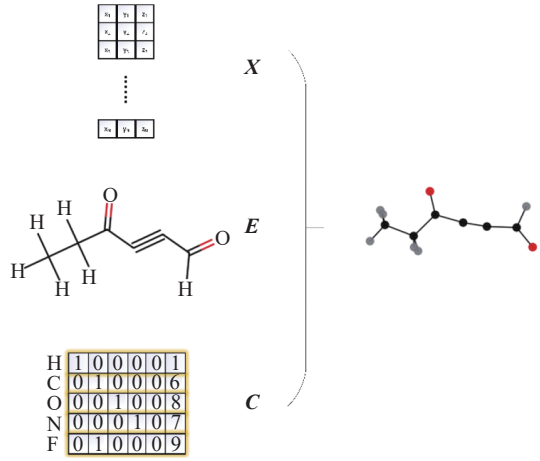


图 1　分子图表示

Fig. 1　Molecular generation representation

1.2　扩散模型

扩散模型(Denoising Diffusion Probabilistic Models, DDPM)是一种基于多步渐进式分布变换的生成框架。其核心机制在于通过马尔可夫链逐步扰动数据分布, 并学习逆向重构目标分布的参数化过程。该模型包含两个阶段: 前向扩散过程系统性地注入噪声以破坏数据结构, 反向生成过程则通过训练好的神经网络迭代恢复数据的内在分布特性。

前向生成过程中, 模型会逐步向原始数据中添加高斯噪声, 经过一系列步骤后, 原始数据的信息被完全打乱, 变成纯粹的高斯噪声。这个过程按照特

定的公式进行添加,随着步骤的增加,添加噪声的比例越来越大,前向过程如式(1)~式(5)所示。

$$q\left(\boldsymbol{G}^{1:T}|\boldsymbol{G}^0\right)=\prod_{t=1}^{T}q\left(\boldsymbol{G}^t|\boldsymbol{G}^{t-1}\right) \tag{1}$$

$$q\left(\boldsymbol{G}^t|\boldsymbol{G}^{t-1}\right)=N\left(\boldsymbol{G}^t;\sqrt{\alpha_t}\boldsymbol{G}^{t-1},\beta_t\boldsymbol{I}\right). \tag{2}$$

$$q\left(\boldsymbol{G}^t|\boldsymbol{G}^0\right)=N\left(\boldsymbol{G}^t;\sqrt{\overline{\alpha}_t}\boldsymbol{G}^0,(1-\overline{\alpha}_t)\boldsymbol{I}\right) \tag{3}$$

$$\alpha_t=1-\beta_t \tag{4}$$

$$\overline{\alpha}_t=\prod_{i=1}^{T}\alpha_i \tag{5}$$

其中, $\boldsymbol{G}^t$ 表示第 t 时刻加噪后的分子图数据, $\boldsymbol{I}$ 为单位矩阵, β_t 代表噪声调度,由时间 1~T 逐渐递增。式(1)描述了从初始状态到第 T 步的完整前向扩散过程;式(2)定义了单步马尔可夫转移,即从第 t–1 步到第 t 步的加噪过程;式(3)则为闭式解形式,实现了从初始状态直接加噪至第 t 步的等效计算。式(4)是对单步加噪过程中参数的定义,其中 β_t 是第 t 步的噪声调度(预定义的、随时间递增的噪声强度),可理解为第 t 步状态中“保留原始信息”的比例(β_t 越大, α_t 越小,加噪程度越高)。式(5)是对多步累积保留比例的定义, $\overline{\alpha}_t$ 代表经过 t 步加噪后,初始状态信息在第 t 步状态中“剩余的累积比例”。

在反向生成阶段,扩散模型通过迭代去噪实现数据重构。该过程的核心在于利用训练好的噪声预测网络,根据当前时刻的 $\boldsymbol{G}^t$ 和 t,估计注入的噪声分量。将此预测噪声代入预设的去噪变换规则,即可计算出前一时刻的数据状态 $\boldsymbol{\mu}_\theta(\boldsymbol{G}^t,t)$(即 $\boldsymbol{G}^{t-1}$)。从完全噪声化的初始状态 $\boldsymbol{G}^t$ 出发,经过 t 次连续去噪迭代,模型逐步将各向同性高斯分布转化为目标数据分布,最终生成符合物理规律的结构化输出,反向生成过程如式(6)~式(8)所示。式(6)描述了扩散模型的反向生成过程,从第 T 步到第 0 步的完成过程;式(7)定义了单步去噪过程,即从第 t 步到第 t–1 步的去噪过程;式(8)为式(7)的等效计算。

$$p_\theta\left(\boldsymbol{G}^{0:T-1}|\boldsymbol{G}^T\right)=\prod_{t=1}^{T}p_\theta\left(\boldsymbol{G}^{t-1}|\boldsymbol{G}^t\right) \tag{6}$$

$$p_\theta\left(\boldsymbol{G}^{t-1}|\boldsymbol{G}^t\right)=N\left(\boldsymbol{G}^{t-1};\mu_\theta(\boldsymbol{G}^t,t),\sigma_t^2\boldsymbol{I}\right) \tag{7}$$

$$\boldsymbol{\mu}_\theta(\boldsymbol{G}^t,t)=\frac{1}{\sqrt{\alpha_t}}\left(\boldsymbol{G}^t-\frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(\boldsymbol{G}^t,t)\right) \tag{8}$$

其中 ϵ_θ 为噪声预测网络。

扩散模型的训练通过最大化数据对数似然的变分下界来优化参数,简化得到公式(9),即训练目标为预测高斯噪声。人们研究发现,这种训练目标简洁,并且能得到很好的训练效果。其中, ϵ 为高斯噪声。

$$L_{\text{simple}}=\mathbb{E}_{t,G^t,\epsilon}\left(\left\|\epsilon-\epsilon_\theta(\boldsymbol{G}^t,t)\right\|^2\right) \tag{9}$$

以上为连续扩散,除了连续领域的扩散,还有适合于离散数据扩散的离散扩散[21]。在分子生成任务中,坐标为连续数据,应用连续扩散进行生成;原子类型和化学键为离散类型的数据,应用离散扩散进行生成。

在离散情况下,前向转移矩阵 $\boldsymbol{Q}_t$ 可以表示为 $[\boldsymbol{Q}_t]_{ij}$,如式(10)所示。

$$[\boldsymbol{Q}_t]_{ij}=q(\boldsymbol{G}^t=j|\boldsymbol{G}^{t-1}=i) \tag{10}$$

$$q\left(\boldsymbol{G}^t|\boldsymbol{G}^{t-1}\right)=\text{cat}\left(\boldsymbol{G}^t;p=\boldsymbol{G}^{t-1}\boldsymbol{Q}_t\right) \tag{11}$$

其中, $[\boldsymbol{Q}_t]_{ij}$ 代表第 t–1 时刻 i 状态转移到第 t 时刻 j 状态的概率。从第 t–1 时刻转移到第 t 时刻的转移如式 (11) 所示,cat 表示离散时的概率分布。

从第 0 时刻到第 t 时刻的转移如式 (12) 所示,其中, $\bar{Q}_t$ 如式 (13) 所示,代表第 0 时刻到第 t–1 时刻转移矩阵的乘积。

$$q\left(\boldsymbol{G}^t|\boldsymbol{G}^0\right)=\text{cat}\left(\boldsymbol{G}^t;p=\boldsymbol{G}^0\bar{\boldsymbol{Q}}_t\right) \tag{12}$$

$$\bar{\boldsymbol{Q}}_t=\boldsymbol{Q}_1\boldsymbol{Q}_2\ldots\boldsymbol{Q}_t \tag{13}$$

反向过程中,由第 t 时刻到第 t–1 时刻的转移公式如式(14)所示,其中, $\boldsymbol{Q}_t^{\mathrm{T}}$ 代表 $\boldsymbol{Q}_t$ 的转置, $\odot$ 代表逐点乘积。

$$q\left(\boldsymbol{G}^{t-1}|\boldsymbol{G}^t,\boldsymbol{G}^0\right)=\text{cat}(\boldsymbol{G}^{t-1};\boldsymbol{G}^t\boldsymbol{Q}_t^{\mathrm{T}}\odot\boldsymbol{G}^0\bar{\boldsymbol{Q}}_{t-1}) \tag{14}$$

1.3 注意力等变图神经网络

给定输入空间 x 和输出空间 y,对于任意群元素 g 属于变化群 M(如 SE(3) 群),网络函数 f: $x\to y$ 满足式(15):

$$f(g\cdot x)=g\cdot f(x) \tag{15}$$

则该网络函数具有等变性,满足等变性的网络函数为等变神经网络。

等变图神经网络(EGNN)由多个等变图卷积层(Equivariant Graph Convolutional Layer, EGCL)组成,单个等变图卷积层的公式如式(16)~式(18)所示。

$$\boldsymbol{m}_{ij}=\phi_e\left(\boldsymbol{h}_i^l,\boldsymbol{h}_j^l,\boldsymbol{d}_{ij}^2,e_{ij}\right) \tag{16}$$

式中, ϕ_e 为前馈神经网络, $\boldsymbol{h}_i^l$ 、 $\boldsymbol{h}_j^l$ 分别代表经过 l 层更新后的第 i、j 个原子的类型特征, $\boldsymbol{d}_{ij}^2$ 为第 i 个节点(原子)与第 j 个节点(原子)距离的平方, e_{ij} 表示第 i 个节点(原子)与第 j 个节点(原子)的连接情况。式(16)表示以这些特征作为输入,利用 ϕ_e 提取特征 $\boldsymbol{m}_{ij}$ 。

式(17)中,前馈神经网络 ϕ_h ,以 $\boldsymbol{h}_i^l$ 和经过消息

传递的 $\boldsymbol{m}_{ij}$ 为输入, 对节点特征 $\boldsymbol{h}_i^l$ 进行更新, 得到 $\boldsymbol{h}_i^{l+1}$ 。

$$\boldsymbol{h}_i^{l+1}=\phi_h\left(\boldsymbol{h}_i^l,\sum_{j\neq i}\boldsymbol{m}_{ij}\right) \tag{17}$$

式(18)中, $\boldsymbol{x}_i^l$ 、$\boldsymbol{x}_j^l$ 分别代表经过 l 层更新后的第 i、j 个原子的坐标, 利用前馈神经网络 ϕ_x 对坐标进行更新, 得到 $\boldsymbol{x}_i^{l+1}$ 。

$$\boldsymbol{x}_i^{l+1}=\boldsymbol{x}_i^l+\sum_{j\neq i}\frac{\boldsymbol{x}_i^l-\boldsymbol{x}_j^l}{\boldsymbol{d}_{ij}+1}\phi_x\left(\boldsymbol{h}_i^l,\boldsymbol{h}_j^l,\boldsymbol{d}_{ij}^2,e_{ij}\right) \tag{18}$$

本文提出的 AEGNN 基于 EGNN, 利用注意力对于原子类型进行特征提取, 所提取的特征用于原子的类型、三维坐标以及化学键的更新, 其中坐标更新时采用了 EGNN 等变更新的思想, AEGNN 的结构如图 2 所示。

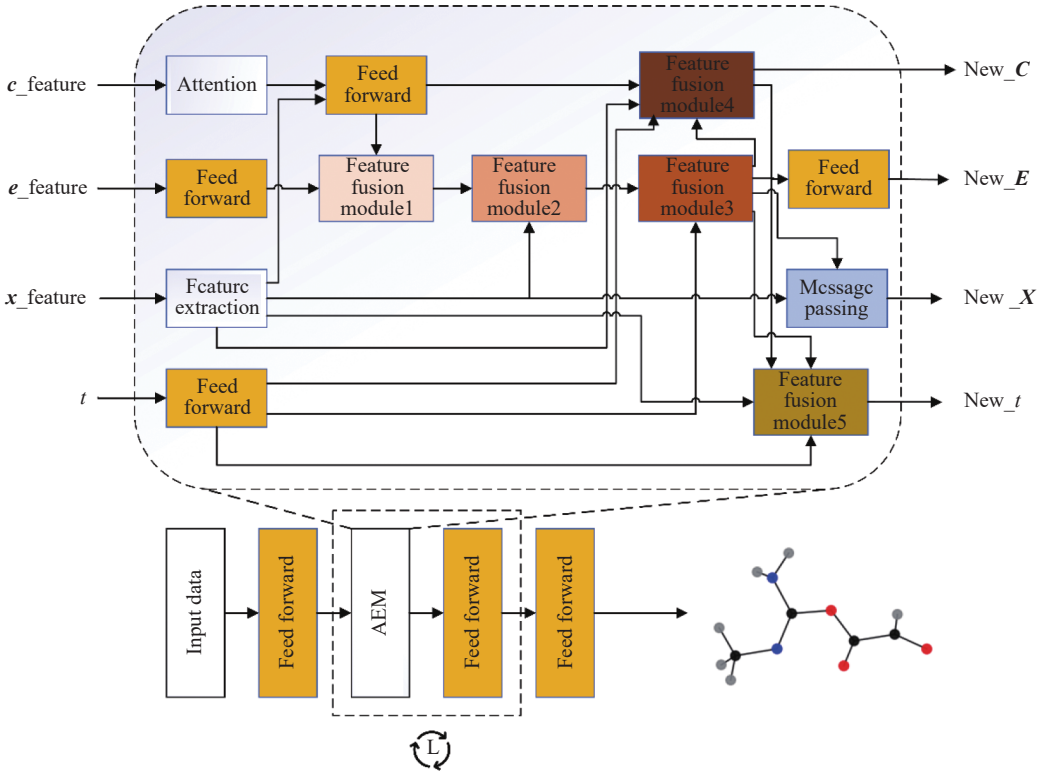


图 2　注意力等变图神经网络

Fig. 2　Attention equivariant graph neural network

在 AEGNN 中, 输入数据包含 4 种特征, 原子类型($\boldsymbol{c}$_feature), 化学键($\boldsymbol{e}$_feature), 坐标($\boldsymbol{x}$_feature), 以及时间(t_feature), 各自经过对应的前馈神经网络, 再连续经过 L 个注意力等变模块(AEM)和前馈神经网络进行更新特征, 最后经过各自的全连接神经网络得到分子图 $\boldsymbol{G}$。

在 AEM 层中, 原子类型首先通过多头自注意力机制进行特征提取, 化学键特征经由全连接层进行编码, 时间(t_feature)通过另一全连接层实现嵌入更新。坐标特征通过特征提取模块(Feature Extraction Module)得到距离特征以及余弦角特征, 两者合称为几何特征, 如图 3 所示; 上述几何特征与自注意力层输出的节点特征经全连接层融合后, 生成统一的空间信息特征。此后, 将自注意力节点特征、空间信息特征与时间特征连续经过 3 个特征融合模块, 分别用于化学键特征的更新, 并通过全连接层输出修正后的化学键表示。基于更新后的化学键特征, 经过特征融合模块 4 通过特征聚合更新节点状态。同时, 融合更新后的化学键特征与时间特征, 通过特征融合模块 5 对时间编码进行更新。最后, 结合距离特征、化学键特征与时间特征, 通过消息传递机制对坐标特征进行迭代更新, 输出新的原子坐标。

在 AEM 层中, 特征融合模块 1、2、3 虽为独立结构, 但其计算流程具有高度相似性, 具体操作如图 4 所示。以特征融合模块 1(Feature Fusion Module 1)为例, 该过程以化学键特征作为输入 1, 原子类型特征作为输入 2。输入 1 同时经由两个前馈神经网络处理: 其中一个网络的输出直接与另一网络输出和输入 2 的乘积相加, 最终生成融合后的特征表示。该特征融合模块的设计蕴含明确的物理意义, 其核心本质在于实现特征空间中概率分布的转移。扩散模型的核心机理正是通过逐步迭代完成从初始高斯分

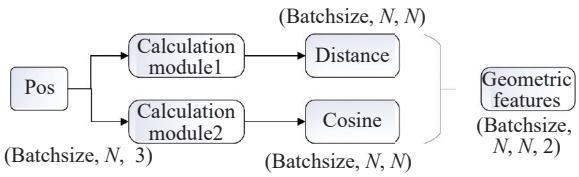


图 3 特征提取模块

Fig. 3 Feature extraction module

布向目标数据分布的动态转变，本模块通过两个结构独立的前馈神经网络，分别对原子类型特征分布的均值与方差进行拟合，进而实现该特征分布的自适应转移与重构；同时，模块引入残差连接结构，有效缓解了深度模型训练过程中的梯度消失问题，显著增强了模型的训练稳定性与特征学习能力。

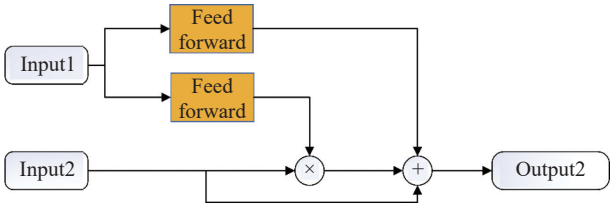


图 4 特征融合模块 1、2、3

Fig. 4 Feature fusion module 1、2、3

在特征融合模块 4(Feature Fusion Module 4)中，化学键特征与几何特征分别经由前馈神经网络进行处理，所得结果与注意力特征进行相乘操作，进而与注意力特征相加，生成中间特征 1。随后，将时间特征 t 与特征 1 输入至特征融合模块 1 进行处理，其输出依次经由主邻域聚合(PNA)[22] 模块整合与前馈神经网络映射，最终获得新的原子类型特征，如图 5 所示。

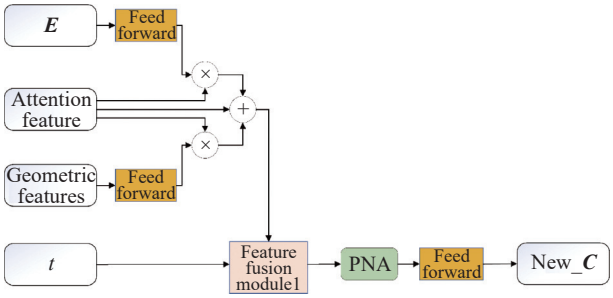


图 5 特征融合模块 4

Fig. 5 Feature fusion module 4

在特征融合模块 5 中，本文采用 PNA 模块对 3 种输入特征进行降维处理，并将降维后的特征与时间特征 t 进行逐元素相加，从而生成增强的时间特征 t，如图 6 所示。

在消息传递模块中，本文用神经网络对节点特征以及化学键特征进行映射，对坐标进行等变更新，如式(19)所示。

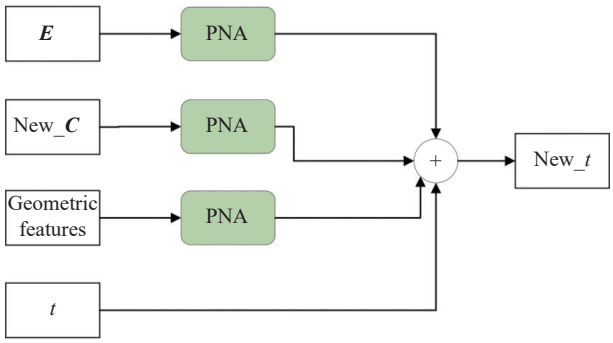


图 6 特征融合模块 5

Fig. 6 Feature fusion module 5

$$\boldsymbol{X}_i^{l+1}=\boldsymbol{X}_i^l+\sum_{j\neq i}\frac{\left(\boldsymbol{X}_i^l-\boldsymbol{X}_j^l\right)}{d_{ij}}\phi_C\left(\boldsymbol{E}_{ij},\boldsymbol{T}_{ij}\right) \tag{19}$$

其中，$\boldsymbol{E}_{ij}$ 为化学键特征，$\boldsymbol{T}_{ij}$ 为几何特征，d_{ij} 为两个原子间距离，$\boldsymbol{X}_i^l$ 代表第 l 层第 i 原子的坐标。

1.4 残差增强神经网络

针对完整分子生成训练时间过长的问题，本文提出了残差增强神经网络(Residual Boosting Neural Network, RBNN)。其结构属于集成学习范畴，借鉴了 Boosting 算法的思想，采用串行集成策略。该结构训练两个模型：核心生成器 (Kernel Generator, KG) 和残差优化器 (Residual Refiner, RR)。两个模型由相同的基本模块堆叠构成(本文所用的基本模块为 AEM)，其中 KG 堆叠较深，用于捕捉复杂的非线性关系以生成高精度初始结果，用以拟合真实结果；RR 则采用较浅的堆叠结构，专注于拟合 KG 输出与真实结果间的残差。这种设计在降低计算负担的同时，增强了模型的校正能力。两个模型以级联方式连接：KG 的输出作为 RR 的输入，RR 的残差校正结果与 KG 初始输出相加得到最终预测。通过这种集成学习机制，模型训练能够在较短时间内收敛，并生成更符合物理规律的分子结构，残差增强神经网络(RBNN)如图 7 所示。

以坐标的损失函数为例，图中，KG 的损失函数(KGLoss)的定义如式(20) 所示；RR 的损失函数(RRLoss)由式(21)给出，总的损失见式(22)。

$$L_{\mathrm{KG}}=\mathbb{E}_{t,G^t,\epsilon}\left(\left\|\boldsymbol{G}-\boldsymbol{G}_\theta^0\left(\boldsymbol{G}^t\right)\right\|^2\right) \tag{20}$$

$$L_{\mathrm{RR}}=\mathbb{E}_{t,G^t,\epsilon}\left(\left\|\left(\boldsymbol{G}-\boldsymbol{G}_\theta^0\left(\boldsymbol{G}^t\right)\right)-\epsilon'_\theta\left(\boldsymbol{G}_\theta^0\left(\boldsymbol{G}^t\right)\right)\right\|^2\right) \tag{21}$$

$$L_{\mathrm{total}}=L_{\mathrm{KG}}+L_{\mathrm{RR}} \tag{22}$$

式中，$\boldsymbol{G}_\theta^0$ 为 KG_AEGNN，ϵ'_θ 为 RR_AEGNN，$\boldsymbol{G}$ 为原始数据。

1.5 注意力等变几何扩散模型

本文的方法用到了两种扩散模型，连续扩散与离散扩散，连续扩散用于生成坐标，离散扩散用于生成原子类型和化学键(文献 [20] 发现原子类型和化

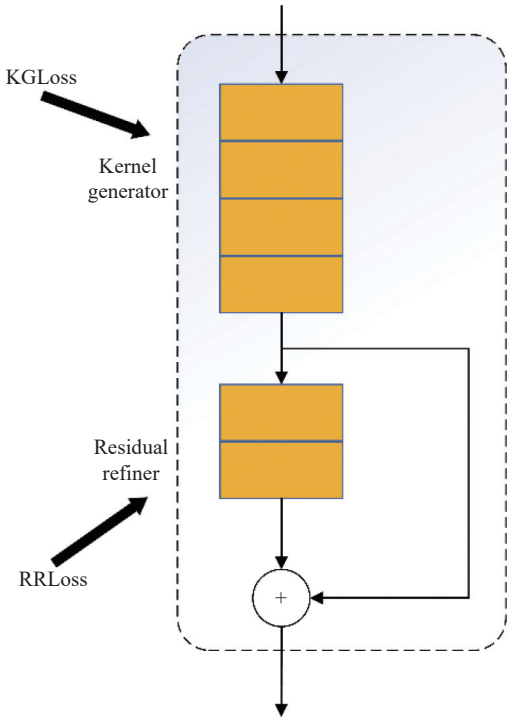


图 7　残差增强神经网络结构

Fig. 7　Residual boosting network structure

学键在离散空间进行扩散的性能更好)。此外, 本文中所用的扩散训练目标未选择预测噪声, 而是选择预测原始数据。式(23)为式(8)利用原始数据预测的等价替换式, 式(24)为预测原始数据的损失函数。

$$\boldsymbol{\mu}_\theta\left(\boldsymbol{G}^t,\boldsymbol{G}^0\right)=\frac{\sqrt{\overline{\alpha}_{t-1}}\beta_t}{1-\overline{\alpha}_t}\boldsymbol{G}_\theta^0(\boldsymbol{G}^t,t)+\frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\boldsymbol{G}^t \tag{23}$$

$$L=\mathbb{E}_{t,G^t,\boldsymbol{\epsilon}}\left(\left\|\boldsymbol{G}^{\boldsymbol{0}}-\boldsymbol{G}_\theta^0(\boldsymbol{G}^t,t)\right\|^2\right) \tag{24}$$

其中, $\boldsymbol{G}_\theta^0$ 为预测原始数据的神经网络, $\boldsymbol{G}^{\boldsymbol{0}}$ 为原始数据。

神经网络采用 AEGNN, 并应用了残差增强神经网络。对于损失函数的选择, 因为采用了两种扩散模式, 要分别对对应的数据采取相应的损失函数。对于坐标这一连续数据, 它属于回归任务, 其损失函数选择为 MSE; 对于原子类型和化学键, 它属于分类任务, 损失函数采用交叉熵函数。

DDPM 的核心在于通过参数化神经网络实现概率分布变换, 本文的扩散模型命名为注意力等变几何扩散模型(Attention Equivariant Geometric Diffusion Model, AEGDM), 如图 8 所示。在训练阶段(前向过程), 通过随机采样时间步和高斯噪声向量, 将坐标噪声向量进行零质心化处理, 以保持分子 3D 结构的平移不变性, 将原始数据点经过一步加噪变换生成扰动样本, 随后利用神经网络预测该时间点的噪声分量, 并通过优化目标函数更新模型参数。在采样阶段(反向过程), 从零质心化的高斯分布初始状态出发, 利用去噪变换公式(23)逐步重构数据: 每一步基于当前状态预测噪声分量, 对其零质心化, 计算前一时刻的样本状态, 最终通过多步迭代获得生成样本。表 1 所示为 AEGDM 的连续数据训练算法(连续数据与离散数据同步训练, 这里以连续数据为例), 表 2 所示为 AEGDM 的采样算法。

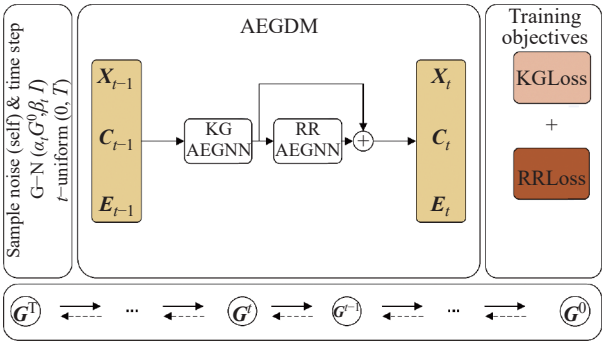


图 8　注意力等变几何扩散模型

Fig. 8　Attention equivariant geometric diffusion model

表 1　AEGDM 训练过程

Table 1　Training process of AEGDM

Algorithm 1: Training process of AEGDM

1: Input: Molecular data $\boldsymbol{G}=(\boldsymbol{X},\boldsymbol{C},\boldsymbol{E})$

2: Initial: Kernel Generator $\boldsymbol{G}_\theta^0$, Residual Refiner ϵ_θ

3: repeat

4:　$\boldsymbol{G}^0\sim q(\boldsymbol{G}^0)$

5:　$t\sim$ Uniform($\{1,\cdots,T\}$)

6:　$\epsilon\sim N(0,\boldsymbol{I})$

7:　Take gradient descent step on

8:　$\nabla_\theta\|\boldsymbol{G}^0-\boldsymbol{G}_\theta^0(\sqrt{\overline{\alpha}_t}\boldsymbol{G}^0+\sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon},t)\|^2$

9:　$\nabla_\theta\|\varepsilon_\theta(\boldsymbol{G}_\theta^0)-(\boldsymbol{G}^0-\boldsymbol{G}_\theta^0(\sqrt{\overline{\alpha}_t}\boldsymbol{G}^0+\sqrt{1-\overline{\alpha}_t}\boldsymbol{\epsilon},t))\|^2$

10: until converged

表 2　AEGDM 采样过程

Table 2　Sampling process of AEGDM

Algorithm 2: Sampling process of AEGDM

1: Input: Kernel Generator z_θ^0, Residual Refiner ϵ_θ

2: $\boldsymbol{G}^T\sim N(0,\boldsymbol{I})$

3: for t in T, $T-1$, $\cdots$, 0 do

4:　$\boldsymbol{G}_X^{t-1}=\frac{\sqrt{\overline{\alpha}_{t-1}}\beta_t}{1-\overline{\alpha}_t}\boldsymbol{G}_\theta^0(\boldsymbol{G}_X^t,t)+\frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\boldsymbol{G}^t$

5:　$\boldsymbol{G}_{C,\,E}^{t-1}=\boldsymbol{G}_{C,\,E}^t\boldsymbol{Q}_t^t\odot\boldsymbol{G}_{\theta C,\,E}^0\bar{\boldsymbol{Q}}_{t-1}$

6:　$\boldsymbol{G}^{t-1}=\boldsymbol{G}^{t-1}+\epsilon_\theta(\boldsymbol{G}^{t-1})$

7: end for

8: return z^0

2 实验结果与分析

2.1 数据集

本文所用的数据集是QM9数据集，由C、H、O、F、N组成的133 885个有机小分子及其构象组成，其中存在同分异构体，重原子至多9个，单个分子的原子数至多29个，其中分子平均有18个原子。数据集中的特征有原子种类、原子数目、3D坐标、原子连接情况、相应的谐波频率、偶极矩、极化率、能量、焓以及原子化的自由能。本文中所用的特征主要为前4种，其他特征并未纳入到模型输入。数据集被划分为训练集、测试集以及验证集，训练集有100 000个分子，测试集有13 388个分子，验证集有20 497个分子。

2.2 评价指标

本文采用分子生成研究中广泛采用的评价指标，包括分子稳定性(Molecular Stability, MS)、原子稳定性(Atom Stability, AS)、有效性(Validity)、唯一性(Uniqueness)与新颖性(Novelty)。以上指标取值越高，表明生成质量越好。

原子稳定性定义为生成分子中满足标准化合价规则的原子数占总原子数的比例。分子稳定性指全部原子均满足化合价约束的生成分子数目占分子总数目的百分比。原子有效性依据生成分子的原子坐标与类型信息构建分子图，并借助Rdkit化学信息学工具包尝试将其转化为对应的分子对象及SMILES字符串，成功转换则判定为有效分子，有效性以有效分子数占生成总数的比例计算。原子唯一性通过对全部有效分子的SMILES字符串进行去重，得到独特分子集合，进而计算独特分子数占有效分子数的比例。原子新颖性则评估独特分子中未出现在训练数据集中的SMILES字符串比例，用于衡量模型生成新分子结构的能力。

除精度指标外，模型的推理效率同样是衡量其实际应用价值的核心要素。为全面评估模型的推理性能，本文选取4项关键指标展开量化分析，包括模型参数量(Params)、CPU单分子采样耗时(Sample-1 (CPU))、GPU批量采样500个分子的耗时(Sample-500 (GPU))以及扩散过程步数(Diff_steps)。

2.3 实验设置

实验在配置了3.00 GHz的Intel i9-10980XE CPU和NVIDIA GeForce RTX 3090的GPU上运行，基于python3.9语言开发，使用Pytorch库搭建网络模型、Pytorch_lightning库训练模型、wandb库实时观测训练指标。

本论文以文献[23]提出的噪声调度作为基本噪声调度，对其进行了改进，以适应3种数据(坐标，原子类型以及化学键)的生成，如式(25)~式(27)所示。

$$f(t)=\cos\left(\frac{\left(\frac{t}{T}\right)^{v}+s}{1+s}\cdot\frac{\pi}{2}\right)^{2} \tag{25}$$

$$\overline{\alpha}_t=\frac{f(t)}{f(0)} \tag{26}$$

$$\beta_t=1-\frac{\overline{\alpha}_t}{\overline{\alpha}_{t-1}} \tag{27}$$

式(25)定义了噪声调度的核心函数，其中t表示扩散过程的时间步，T为总时间步数，s是用于控制初始阶段噪声强度的偏移参数，新增的参数v则是调度特性的调节项(通过调整v的取值，可适配不同类型的数据分布)。该函数通过余弦函数的平方形式，实现了噪声强度随时间步的平滑变化。式(26)为$\overline{\alpha}_t$的计算方式，式(27)为β_t的计算方式。在这里，用v_x、v_e、v_c代表原子坐标、化学键、原子类型的v值，其中v_x=2.5, v_e=1, v_c=1.5, s=0.008, 3种数据的噪声调度如图9所示。扩散步数采用文献[18]的步数为500。

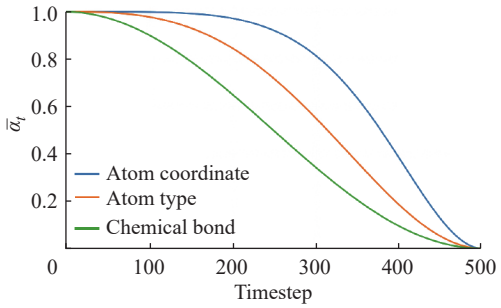


图9 噪声调度

Fig. 9 Noise Schedule

本文为多任务联合生成，总损失函数为各个损失函数乘以对应权重，如式(28)所示。

$$\begin{aligned}\mathrm{Loss}_{\mathrm{total}}=&\lambda_X\left\|\widehat{\boldsymbol{X}}-\boldsymbol{X}\right\|^2+\lambda_C\mathrm{CrossEntropy}\left(\widehat{\boldsymbol{C}},\boldsymbol{C}\right)+\\&\lambda_E\mathrm{CrossEntropy}\left(\widehat{\boldsymbol{E}},\boldsymbol{E}\right)\end{aligned} \tag{28}$$

其中，$\widehat{\boldsymbol{X}}$为模型预测的坐标，$\boldsymbol{X}$为原始数据的坐标，采用MSE损失；$\widehat{\boldsymbol{C}}$为模型预测的原子类型，$\boldsymbol{C}$为原始数据的原子类型，$\widehat{\boldsymbol{E}}$为模型预测的化学键，$\boldsymbol{E}$为原始数据的化学键，采用交叉熵损失。对于权重的选择，本文采用λ_X=3, λ_C=2, λ_E=1。

在本文的实验中，扩散模型采用AdamW优化器进行训练，学习率设置为2×10^{-4}，权重衰减系数为1×10^{-12}，批量大小为96，训练轮次为1 000。在模型结构

方面, 采用的 AEGNN 基础模块的层数为 KG_AEGNN 堆叠 8 层, RR_AEGNN堆叠 4 层。此外, 各类输入数据经前馈神经网络映射后的隐藏维度为坐标特征 64 维, 原子类型特征 256 维, 化学键特征 64 维, 时间步特征 256 维。

2.4　实验结果及分析

2.4.1　AEGDM 实验结果　基于 QM9 数据集, 将本文方法与 MiDi[18] 以及 MolDiff[20] 进行了比较。本研究采用每种方法生成 10 000 个分子样本, 其中 AEGDM 以运行 3 次评估, 以评估其在分子生成任务中的性能, 相关结果汇总于表 3。从实验结果可以看出, 本文提出的方法 AEGDM 在多个关键指标上展现出良好的表现。

表 3　不同方法实验结果

Table 3　Experimental results of different methods

Metric	MS/%	AS/%	Validity/%	Uniqueness/%	Novelty/%
DATA	98.7	99.8	98.9	99.9	—
MiDi	**97.5**	**99.8**	**97.9**	97.6	67.5
MolDiff	94.2	99.5	93.8	—	—
AEGDM	95.6±0.1	99.7±0.0	96.4±0.2	**98.2±0.2**	**70.9±0.4**

AEGDM 在唯一性(Uniqueness, 98.2%)和新颖性(Novelty, 70.9%) 两个指标上优于所有对比的基线方法, 表明该方法能够生成更多样化且与训练数据差异显著的新颖分子结构, 体现出较强的分布外生成能力。这说明 AEGDM 在探索化学空间方面具有更强的潜力, 有助于发现具有创新性的候选分子。在稳定性和有效性方面, MiDi 表现优于 AEGDM。针对于该结果进行分析, MiDi 可同步生成分子的 4 类核心数据, 包括原子坐标、原子类型、电荷状态与化学键信息。其中, 电荷状态采用三元离散化表示(即−1、0、1), 分别对应负电荷、电中性与正电荷 3 种状态, 该表示方式与原子化合价具有直接关联。原子稳定性与分子稳定性的定义本身以化合价规则为核心依据, 而 MiDi 在生成过程中显式编码电荷特征, 这一设计使其在原子稳定性与分子稳定性这两项指标上的表现优于本研究提出的 AEGDM。进一步地, 分子生成的有效性指标与稳定性存在内在相关性, 受稳定性优势的传导作用, MiDi 在有效性指标上同样展现出相对 AEGDM 的性能优势。

表 4 呈现了不同方法的推理效率对比结果。为了遵循等量的原则, 本文将 MolDiff 的扩散步数设置为 500(原论文中扩散步数为 1 000), 比较其推理效率。上述结果表明, AEGDM 在批量分子生成任务中展现出最优性能(350.8 s), MolDiff(500 步)在单分子采样中展现最优性能(41.3 s)。而 MolDiff 尽管具有最小的模型参数量(11.9×10^6), 但其批量采样效率却显著落后(589.3 s)。该现象的成因在于 MolDiff 的采样函数引入第 2 个模型 bond_predictor(表中 MolDiff 的参数由两个模型共同组成, MolDiff 模型参数量为 5.7×10^6, bond_predictor 为 6.2×10^6), bond_predictor 在迭代过程的每一步前向传播并反向传播求梯度, 大大减慢了运行速度。

表 4　不同方法推理效率实验结果

Table 4　Experimental results on the inference efficiency of different methods

Metric	Params/10^6	Sample-1(CPU)/s	Sample-500(GPU)/s	Diff_step
MiDi	24.1	45.3	440.6	**500**
MolDiff	**11.9**	83.4	1 201.8	1 000
MolDiff	**11.9**	**41.3**	589.3	**500**
AEGDM	20.2	43.2	**350.8**	**500**

2.4.2　RBNN 结构对比实验　RBNN 可以设计为并行和串行两种结构, 如图 10 所示。针对两种策略, 本文开展了系统的对比实验, 实验结果如表 5 所示。实验结果表明, 串行(SI)效果优于并行(PI), 因此本文采用了方案(b)。

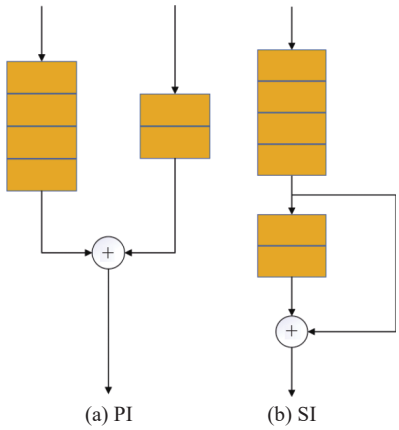


图 10　RBNN 结构

Fig. 10　Structure of RBNN

针对上述实验的结果进行分析, 串行集成方法的核心设计动机, 在于利用基学习器间的相关性实现误差的迭代传递与修正; 而并行集成方法的核心动机, 则是依托基学习器的独立性, 通过“群体决策”抵消单个模型的波动[24]。本研究中采用的两个基学习器(KG_AEGNN 与 RR_AEGNN)本质上属于同质化结构, 且训练数据完全一致, 其输出天然具有较高相关性; 若采用并行结构, 将直接违背并行集成对“基

学习器独立性”的基本要求,无法通过并行融合实现性能增益。且 RBNN 结构本质上属于串行集成所用的方法。

表 5 RBNN 对比实验

Table 5 Comparative experiment of RBNN

Metric	MS/%	AS/%	Validity/%	Uniqueness/%	Novelty/%
DATA	98.7	99.8	98.9	99.9	—
PI	94.3	99.6	95.2	97.9	70.5
SI	**95.6**	**99.7**	**96.4**	**98.2**	**70.9**

前向训练过程中,在串行结构上的输入为加噪后的原始数据,首先通过 KG_AEGNN 完成“加噪数据→原始数据”的初步预测,随后将该预测结果输入 RR_AEGNN,由后者聚焦于“预测原始数据与真实数据的残差”的拟合;而在并行结构中,两个基学习器的输入均为加噪后的原始数据,此时 RR_AEGNN 需同时完成“加噪数据→原始数据”与“预测值→真实值的残差”拟合,任务复杂度显著提升。此外,RR_AEGNN 属于弱学习器(堆叠层数较少),其表征能力不足以拟合上述复杂分布,对于结果可能产生负影响。而串行结构所产生的影响大概率为正面,因此并行结构在各项实验指标中全面劣于串行结构。

2.4.3 RBNN 分子稳定性对比实验 为评估 RBNN 结构的有效性,本文设计了对比实验。在实验设置中,模型共进行 100 轮训练,每 5 轮进行一次采样生成,每次生成 1 000 个分子,未采取 RBNN 结构的模型 12 层;采用 RBNN 结构的模型,KG_AEGNN 为 8 层,RR_AEGNN 为 4 层。由于原子稳定性和有效性与分子稳定性高度相关,图 11 中仅展示分子稳定性的变化趋势;同时,由于采样规模较小(1 000 个分子),唯一性与新颖性指标均接近 100%,因此未予呈现。最终,图中绘制了分子稳定性随训练轮次变化的折线图,以直观比较有 RBNN 与无 RBNN 时模型的性能差异。

如图 11 所示,引入 RBNN 模块后,模型在分子稳定性方面呈现出显著改善:其平均性能明显优于未引入该结构的模型,且所达到的最高性能指标也优于原模型。实验结果表明,RBNN 结构不仅有效提升了模型训练的稳定性(使整体的突然急剧变化减少),还进一步优化了生成质量(整体以及最高的分子稳定性高于原模型),从而加速实验收敛过程并增强整体性能表现。

为验证基于 RBNN 结构的模型在性能上是否具有统计学意义的提升,本研究采用配对 t 检验分析,

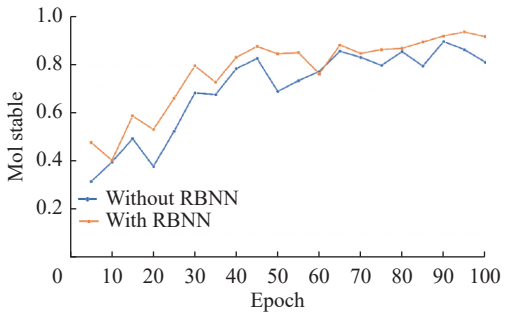


图 11 RBNN 分子稳定性对比实验

Fig. 11 Comparison experiment of RBNN molecular stability

对比引入 RBNN 结构与未引入该结构的模型在分子稳定性预测任务上的表现差异,并通过显著性检验确保性能改进的稳健性。

实验设计如下,零假设为采用 RBNN 结构的模型与未采用该结构的模型在性能上无显著差异,备择假设为采取 RBNN 结构的模型显著优于未采取 RBNN 结构的模型。本实验中,两种模型采取不同随机种子均独立运行 8 次(模型采用 RBNN分子稳定性实验的设置),每次训练 100 轮,并选取各次最优结果进行对比。检验自由度设为 7,显著性水平设定为 $\alpha = 0.05$ 。t 检验计算得到 t =2.567 0, p = 0.018 6。由于 p 值小于 0.05,可在 $\alpha = 0.05$ 水平上拒绝零假设,接受备择假设。

3 结 论

本文提出了注意力等变扩散模型用于分子生成任务。AEGNN 借助多头自注意力提取节点特征联合更新分子图中的边特征、原子坐标与节点特征,在严格保持旋转与平移等变性的条件下,通过逆向扩散过程逐步重构分子中的原子类型、三维结构及化学键。RBNN 通过协同训练两个模型进一步提升生成性能。在 QM9 数据集上的实验表明,该方法在生成分子的唯一性与新颖性方面均获得显著提高,证明 AEGDM 能够有效探索化学空间,并促进具有创新结构的候选分子发现。此外,RBNN 机制可加速实验迭代、增强模型表现,为分子生成模型的迭代优化提供了重要技术支撑。

传统的扩散模型方法对于噪声进行预测,因为原始数据分布过于复杂,模型的生成能力不足以拟合如此复杂的分布。本文中,采取了预测原始数据的扩散方法,利用 RBNN 结构去缓解模型的生成能力不足的问题,利用 KG_AEGNN 与 RR_AEGNN 进一步提升 AEGDM 的总体生成能力。此外,自注意力机制提高了模型的全局数据观测能力,使模型能

捕捉到各个节点之间联系, 提升了模型对分子内部长程依赖与协同效应的推理能力。模型内在的等变性保证了所有生成构象的物理合理性, 其输出本身就是能量友好的, 无需依赖后续力场优化来‘修正’。这提升了模型的可靠度, 也意味着其内部表示更贴合真实的分子几何空间。三者的协同作用显著提升了 AEGDM 的整体生成性能与结构合理性。

参考文献:

[1] CORSO G, STÄRK H, JING B, *et al*. Diffdock: Diffusion steps, twists, and turns for molecular docking[EB/OL]. (2022-10-04). https://arxiv.org/abs/2210.01776.

[2] 仝建波, 李云飞, 刘淑玲, 等. 一种新型拓扑分子表征方法的应用研究 [J]. 环境化学, 2010(6): 1167-1171.

[3] DE VIVO M, MASETTI M, BOTTEGONI G, *et al*. Role of molecular dynamics and related methods in drug discovery[J]. Journal of Medicinal Chemistry, 2016, 59(9): 4035-4061.

[4] MANSIMOV E, MAHMOOD O, KANG S, *et al*. Molecular geometry prediction using a deep generative graph neural network[J]. Scientific Reports, 2019, 9(1): 20381.

[5] GILMER J, SCHOENHOLZ S S, RILEY P F, *et al*. Neural message passing for quantum chemistry[C]//International conference on machine learning. Sydney, AUS: ICML, 2017: 1263-1272.

[6] SCHÜTT K, KINDERMANS P J, SAUCEDA FELIX H E, *et al*. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions[C]//Advances in Neural Information Processing Systems. Long Beach, USA: [s.n.], 2017: 992-1002.

[7] GASTEIGER J, GROSS J, GUENNEMANN S. Directional Message Passing for Molecular Graphs[EB/OL]. (2020-03-06). https://arxiv.org/abs/2003.03123

[8] FUCHS F, WORRALL D, FISCHER V, *et al*. SE (3)-transformers: 3D roto-translation equivariant attention networks[J]. Advances in Neural Information Processing Systems, 2020, 33: 1970-1981.

[9] BILODEAU C, JIN W, JAAKKOLA T, *et al*. Generative models for molecular discovery: Recent advances and challenges[J]. Wiley Interdisciplinary Reviews: Computational Molecular Science, 2022, 12(5): e1608.

[10] KINGMA D P, WELLING M. Auto-encoding variational bayes[EB/OL]. (2013-12-20). https://arxiv.org/abs/1312.6114.

[11] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models[J]. Advances in Neural Information Processing Systems, 2020, 33: 6840-6851.

[12] DINH L, SOHL-DICKSTEIN J, BENGIO S. Density estimation using real NVP[EB/OL]. (2016-05-27). https://arxiv.org/abs/1605.08803.

[13] XU M, YU L, SONG Y, *et al*. Geodiff: A geometric diffusion model for molecular conformation generation[EB/OL]. (2022-03-06). https://arxiv.org/abs/2203.02923

[14] SATORRAS V G, HOOGEBOOM E, WELLING M. E(n) equivariant graph neural networks[C]//International Conference on Machine Learning. Vienna, AUT: ICML, 2021: 9323-9332.

[15] HOOGEBOOM E, SATORRAS V G, VIGNAC C, *et al*. Equivariant diffusion for molecule generation in 3d[C]//International Conference on Machine Learning. Maryland, USA: ICML, 2022: 8867-8887.

[16] XU M, POWERS A S, DROR R O, *et al*. Geometric latent diffusion models for 3D molecule generation[C]//International Conference on Machine Learning. Hawaii, USA: ICML, 2023: 38592-38610.

[17] XU C, WANG H, WANG W, *et al*. Geometric-facilitated denoising diffusion model for 3D molecule generation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(1): 338-346.

[18] VIGNAC C, OSMAN N, TONI L, *et al*. MiDi: Mixed graph and 3D denoising diffusion for molecule generation [C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Cham: Springer Nature Switzerland, 2023: 560-576.

[19] HUA C, LUAN S, XU M, *et al*. Mudiff: Unified diffusion for complete molecule generation[EB/OL]. (2023-04-28). https://arxiv.org/abs/2304.14621.

[20] PENG X, GUAN J, LIU Q, *et al*. MolDiff: Addressing the atom-bond inconsistency problem in 3D molecule diffusion generation[C]//International Conference on Machine Learning. Hawaii, USA: ICML, 2023: 27611-27629.

[21] AUSTIN J, JOHNSON D D, HO J, *et al*. Structured denoising diffusion models in discrete state-spaces[J]. Advances in Neural Information Processing Systems, 2021, 34: 17981-17993.

[22] CORSO G, CAVALLERI L, BEAINI D, *et al*. Principal neighbourhood aggregation for graph nets[J]. Advances in Neural Information Processing Systems, 2020, 33: 13260-13271.

[23] NICHOL A Q, DHARIWAL P. Improved denoising diffusion probabilistic models[C]//International conference on machine learning. Vienna, AUT: ICML, 2021: 8162-8171.

[24] ZHOU Z H. Ensemble Methods: Foundations and Algorithms[M]. 2nd ed. New York: Chapman and Hall/CRC, 2025.

Complete Molecular Generation Based on Attention-Equivariant Geometric Diffusion Models

ZHANG Baichuan[1], XIE Bo[1], ZHANG Liguo[2], LUO Jianxu[1]
(1. School of Information Science and Engineering; 2. School of Chemistry and Molecular Engineering, East China University of Science and Technology, Shanghai 200237, China)

Abstract: This paper proposes an attention-equivariant diffusion model (comprising AEGNN and RBNN) for molecular generation tasks. The AEGNN leverages multi-head self-attention to jointly update edge features, atomic coordinates, and node features within a molecular graph. Subject to strict rotational and translational equivariance, it progressively reconstructs atom types, three-dimensional structures, and chemical bonds via a reverse diffusion process. The RBNN further improves generative performance through co-training two submodels, both built by stacking identical basic modules (denoted as AEM in this paper). The Knowledge Generator (KG) adopts a deeper layer stack to capture complex nonlinear relationships, producing high-precision initial predictions that align with the ground truth. By contrast, the Residual Refiner (RR) adopts a shallower architecture, focusing on fitting the residual between the KG's output and the true target values. This design cuts computational overhead while boosting the model's error correction capacity. The two modules are connected in a cascade structure: the output of the KG acts as the input to the RR, and the final prediction is obtained by adding the residual correction from the RR to the initial output of the KG. Experiments conducted on the QM9 dataset show substantial improvements in the uniqueness and novelty of generated molecules, demonstrating that the AEGDM can efficiently explore chemical space and aid the discovery of structurally innovative candidate molecules. Moreover, the RBNN mechanism speeds up experimental iteration and elevates overall model performance, offering critical technical support for the iterative optimization of molecular generation frameworks.

Key words: diffusion model; graph neural network; molecular generation; transformer; ensemble learning

（责任编辑：王晓丽）