
研究提出基于双视图化学语言对齐的分子表示学习框架

作者：writer 来源：中国科学院

本文原地址：<https://www.iikx.com/news/progress/41063.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

研究提出基于双视图化学语言对齐的分子表示学习框架。

分子表示学习是人工智能理解化学结构和开展分子性质预测的重要基础，其表征质量直接影响模型的预测能力。基于单一SMILES的序列建模虽然具有良好的计算效率和可扩展性，但在充分表征复杂分子结构信息方面仍存在一定局限。

中国科学院成都生物研究所团队提出了一种基于SMILES – InChI双视图化学语言对齐的分子表示学习框架ClipMol。该方法无需显式三维构象和外部语料，仅利用同一分子的SMILES和InChI两种互补化学表示，通过动态SMILES随机化构建拓扑保持的多视图，并采用双编码器双向对比学习，使模型能够学习不同化学表达方式之间一致的结构信息。同时，研究设计了结合长短期记忆网络全局特征提取、多尺度卷积神经网络局部特征提取和门控融合机制的混合编码器，自适应整合分子的全局拓扑信息与局部化学结构特征。

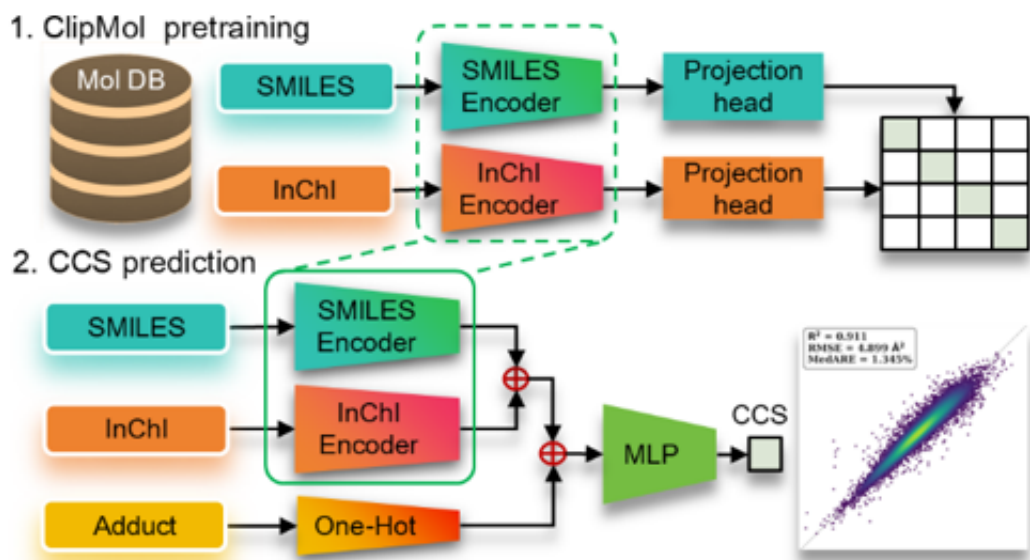
研究人员在MoleculeNet十项分子分类和回归任务上对ClipMol进行了评价。结果表明，其大规模预训练版本ClipMol-XL在多项任务中取得最优或具有竞争力的性能，显示出良好的分子表示迁移能力。进一步将ClipMol应用于离子迁移—质谱中的碰撞截面（CCS）预测，在METLIN-CCS和ALLCCS两个独立数据集上均表现出稳定的预测性能，且在不同加合离子和多种化学类别中保持较

²仍达到0.876，表明预训练获得的分子表示本身已包含丰富的CCS相关结构信息。

研究表明，SMILES与InChI双视图化学语言对齐为分子表示学习提供了一种兼顾结构保真性、计算可扩展性和任务迁移能力的新策略，并为离子迁移—质谱相关CCS预测及其他分析化学任务提供了有效的分子表示基础。

相关研究成果发表在《分析化学》（*Analytical Chemistry*）上。研究工作得到国家自然科学基金等的支持。

[论文链接](#)



分子表示学习与碰撞截面预测研究获进展

研究团队单位：成都生物研究所

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发