

基于多头注意力 GAN 的药物靶点亲和力预测研究进展与展望

徐璐洋, 谢跃生, 岳中乐, 陈彦君, 孟凡爱, 陈 果

华北理工大学 河北唐山

【摘要】近年来深度学习方法在小分子药物与蛋白质靶标的亲和力 (DTA) 预测领域取得了显著进展, 生成对抗网络 (GAN) 已被应用于该领域, 但当前主流方法多局限于浅层 GAN 架构。本文综述了基于深度学习的 DTA 预测方法, 探讨了其在提高预测精度和模型可解释性方面的潜力。文章首先回顾了传统方法和深度学习方法的优缺点, 随后重点介绍了多头注意力机制和 GAN 的结合应用, 最后总结了当前研究的挑战和未来发展方向。

【关键词】DTA 预测; 生成对抗网络; Attention 机制; SHAP

【收稿日期】2025 年 1 月 22 日

【出刊日期】2025 年 2 月 26 日

【DOI】10.12208/j.ijmd.20250017

Research progress and prospects of drug target affinity prediction based on multi head attention GAN

Luyang Xu, Yuesheng Xie, Zhongle Yue, Yanjun Chen, Fanai Meng, Guo Chen

North China University of Technology Tangshan, Hebei

【Abstract】Recent advances in deep learning have significantly improved small molecule-drug target affinity (DTA) prediction, with generative adversarial networks (GANs) emerging as a promising approach. However, current GAN-based methods primarily use shallow architectures. This paper reviews deep learning-based DTA prediction methods, examining their potential to enhance accuracy and interpretability. We compare traditional and deep learning approaches, highlight the integration of multi-head attention with GANs, and discuss current challenges and future directions.

【Keywords】DTA prediction; Generate adversarial networks; Attention mechanism; SHAP

1 研究背景

药物-靶点亲和力 (DTA) 预测是计算机辅助药物设计的关键环节, 其预测长期以来都是该领域的重要挑战。传统方法依赖生物实验或计算模拟虽具备一定可靠性, 但普遍面临耗时长、成本高等问题。近年来深度学习技术的突破为 DTA 预测提供了新范式, 尤其是生成对抗网络 (GAN) 与多头注意力机制的结合, 可自动挖掘药物分子与靶标蛋白的深层特征, 显著提升了预测性能与泛化能力。然而现有模型仍存在数据依赖性高、可解释性不足等挑战。

本文系统综述了基于多头注意力 GAN 的 DTA 预测研究进展, 分析其技术优势与局限性, 并展望未来发展方向。

2 DTA 预测的传统方法

2.1 实验方法

传统 DTA 预测主要依赖蛋白质微阵列和亲和层析的生物实验方法。以蛋白质微阵列为例, 该技术借助微型高密度阵列固定蛋白质, 随后采用荧光基团等亲和标签标记待测分子, 通过追踪矩阵中标记的信号来确定药物分子的靶标蛋白。此方法的主要优势在于其高通量性, 可同时鉴定蛋白质组中的靶蛋白和非靶蛋白。但为筛选出对特定蛋白质靶标兼具有效性和安全性的药物分子, 研发人员需对数千种化合物进行测试, 这一过程存在效率低、成本高且周期长的缺陷。

2.2 计算方法

为突破实验方法的局限, 多种计算手段被引入 DTA 预测领域。其中分子对接技术已成为结构生物学与药物化学领域的重要工具, 该技术通过拟合化合物与靶标的三维结构预测结合位点及结合能, 可

高效筛选虚拟化合物库以发现新型配体并为结构导向的配体设计提供关键切入点。随着技术演进, 融合物理模型、机器学习算法及实验验证的混合策略显著提升了预测效能。

除分子对接, 分子动力学模拟^[1]也在虚拟筛选中发挥重要作用。例如 Adnan 团队通过整合网络药理学、分子对接及分子动力学技术, 揭示了凤凰枣抗念珠菌作用的分子靶标及潜在机制。此外, 自由能扰动及 MM/GBSA 等方法为结合自由能/亲和力计算提供了有效解决方案。

尽管传统方法具有较强的解释性, 但其计算资源消耗大、高通量处理能力不足的局限, 难以应对大规模药物-靶标对的筛选需求, 需进一步探索更高效的预测方法。

3 DTA 预测的机器学习方法

3.1 传统机器学习方法

在药学领域, 计算机辅助药物设计等范式兴起, 加速了药物研发进程并提高了预测准确性。当前基于机器学习的 DTA 预测方法分为传统机器学习方法和深度学习方法。传统机器学习方法如随机森林回归、支持向量机回归等通过提取药物和靶标的特征进行预测。例如 Pahikkala 等人提出的 KronRLS 模型^[2]利用 S-W 算法计算药物-靶标对的相似性, 具有高效的数学基础和编程简便性但其仅能捕捉线性依赖关系。为此 Heidemeyer 等人开发了 SimBoost 框架, 通过梯度增强机集成药物和靶点的网络特征实现 DTA 预测及置信区间估计。此外 Pair 等人提出的 Pred-binding 方法融合蛋白质序列、化合物结构及分子指纹信息, 采用 SVM 与随机森林集成学习策略实现了高精度预测。

尽管这些方法在 DTA 预测中表现良好, 但依赖人工特征工程及浅层模式识别机制, 在提取药物和靶标的高级隐藏特征方面存在不足, 可能导致预测精度受限, 难以全面反映药物与靶点之间的真实相互作用。

3.2 深度学习方法

近年来, 深度学习在图像处理和自然语言处理等领域的突破性进展, 推动了 DTA 预测范式的革新。与传统机器学习方法相比, 深度学习通过自动挖掘药物分子结构与靶标序列中的深层隐含特征, 展现出更优的预测性能。例如 Öztürk 等人提出的

DeepDTA 框架, 采用双卷积神经网络分别学习药物 SMILES 编码与蛋白质序列的表征, 经多层感知机融合后实现亲和力预测。同多数前沿的 DTA 预测模型类似, DeepDTA 基于带有标注数据的监督学习框架进行模型训练与预测。相反, 赵玲玲等人开创的 GANsDTA 以无监督的方式实现蛋白质序列和化合物 SMILES 的特征提取^[3], 拓展了弱标签场景下的预测能力。

表 1 基于 Davis 和 KIBA 数据集的多种先进方法的性能比较

数据集	模型	MSE	CI	R ^{m2} 指标
Davis	DeepDTA	0.261	0.878	0.603
	GraphDTA	0.254	0.880	—
	DeepAffinity	0.253	0.900	—
	KronRLS	0.379	0.871	0.407
	SimBoost	0.282	0.872	0.655
	GANs	0.276	0.881	0.653
KIBA	DeepDTA	0.194	0.863	0.673
	GraphDTA	0.139	0.889	—
	DeepAffinity	0.188	0.842	—
	KronRLS	0.411	0.782	0.342
	SimBoost	0.222	0.836	0.629
	GANs	0.224	0.866	0.675

在药物结构表征中常使用 RDKit 等工具将 SMILES 字符串转换为分子图, 利用图神经网络 (GNN) 提取结构特征, 通过卷积神经网络 (CNN) 处理靶标序列以学习高级特征。例如 Nguyen 团队开发的 GraphDTA 模型结合 GCN 和 CNN, 分别从药物分子图和靶标序列中提取特征, 并通过全连接层预测 DTA, 其性能优于多数深度学习模型^[4]。此类模型主要依赖分子图表示, 未能充分捕捉单个子结构间的相互作用。为此朱志琴团队提出一种基于扩散模型的图优化模块, 将转换器和扩散器引入 TDGraphDTA 模型, 通过多尺度信息交互和特征提取, 优化分子图表征并提升模型的可解释性。

尽管基于深度学习的 DTA 预测模型在准确性和泛化能力上不断进步, 但其“黑盒”特性导致可解释性不足, 如何增强其可解释性有望成为未来研究趋势。

4 引入解释性算法的 DTA 模型

4.1 引入注意力机制

机器学习的“黑盒”特性限制了其在药物研究中的应用, 因此可解释人工智能方法备受关注, 注意力机制被广泛应用于特征捕捉。

具体而言, 赵启昌等人提出的 AttentionDTA 模型, 设计两个独立 1D-CNNs 分别提取药物 SMILES 序列与蛋白质序列信息, 引入双侧多头注意力机制揭示药物与蛋白质的潜在关联。张丽团队开发了分子表示块 DTA 预测模型 (MRBDTA), 整合嵌入与位置编码、分子表示块及交互学习模块, 利用多头注意力机制定位蛋白质与药物的相互作用位点并进行了可解释性分析。邓磊等人通过深度残差网络中的多头自注意力机制, 开发了 DeepMHADTA 模型, 该方法在特征提取层应用残差网络, 并通过多头注意力机制层促进特征关联, 将蛋白质与药物特征融合后输入全连接层预测, 展现出卓越性能。彭杨团队提出基于注意力感知和模态融合的 DTI 方法, 在药物分支结合 Graph Transformer 与 GCN 以全面表征药物分子的全局结构与生化信息; 在蛋白质分支中则采用独立编码模块提取全局与局部特征^[5]。

4.2 引入 SHAP 机制

除 Attention 机制外, 还可整合其他可解释性算法如 SHAP。SHAP 是一种受博弈论启发的方法, 旨在解释机器学习模型的预测, 它为每个输入特征生成一个 SHAP 值, 该值指示该特征如何有助于指定数据点的预测。Lamens 等学者构建了一个多类化合物活性预测模型的合理化测试系统, 该系统融合反事实 (CF) 与 Shapley 附加解释两种 XAI 技术以解析特征关联性。通过识别单靶点和双靶点活性化合物的微小结构修饰并结合特征映射, 利用 CF 与 SHAP 值计算, 为模型决策提供化学层面阐释。同年刘璐团队开发了结合机器学习、SHAP 和分子动力学模拟的通用管道来预测活性化合物和片段, 以发现 B 细胞淋巴瘤治疗的多靶点潜在抑制剂。

5 结论与展望

在 DTA 预测领域, 传统实验方法虽然准确性高但受限于耗时、成本高昂和通量不足, 难以满足大规模药物筛选的需求。计算方法的出现如分子对接和分子动力学模拟, 为虚拟筛选提供了有力工具, 但仍面临计算资源消耗大和高通量处理能力不足的

挑战。机器学习尤其深度学习方法的引入, 通过自动挖掘药物和靶标的高级隐藏特征, 显著提升了预测的准确性和稳健性。DeepDTA、GANsDTA 等模型的创新, 展示了深度学习在 DTA 预测中的巨大潜力。

为增强模型的可解释性, 注意力机制和 SHAP 等可解释性算法被应用于 DTA 预测模型中。注意力机制揭示药物与蛋白质间的潜在关联, 为模型的决策提供了直观的解释。SHAP 值基于博弈论原理, 定量分析了各特征对预测结果的贡献度, 进一步加深了对模型决策机制的理解。

尽管基于深度学习的 DTA 预测模型取得了显著进展, 但仍存在一些挑战和问题。例如, 模型的可解释性有待提高, 跨模态信息融合不充分影响了模型对复杂相互作用的建模能力。

综上所述, 未来 DTA 预测方法的研究可从以下几个关键方面展开深入探讨:

(1) 合理融合 CNN、GAN 等深度学习技术, 从药物与靶标的序列信息、结构特征中提取出 DTA 预测的重要隐含特征。

(2) 尝试采用多种类型的深度学习深入挖掘药物与靶点的特性, 提升 DTA 预测的精准度及模型的泛化性能。

(3) 当前多数基于深度学习的 DTA 预测方法在可解释性方面存在不足, 因此增强模型的可解释性有望成为未来研究的重要切入点。

参考文献

- [1] Zhang D, Zhang T, Zhu J, et al. Discovery of novel targets and mechanisms of MEK inhibitor Selumetinib for LGG treatment based on molecular docking and molecular dynamics simulation[J]. *Journal of Molecular Modeling*, 2022, 28(6): 138.
- [2] Pahikkala T, Airola A, Pietilä S, et al. Toward more realistic drug-target interaction predictions[J]. *Briefings in bioinformatics*, 2015, 16(2): 325-337.
- [3] Zhu Z, Yao Z, Zheng X, Qi G, Li Y, Mazur N, et al. Drug-target affinity prediction method based on multi-scale information interaction and graph optimization. *Comput[J]. Biol*, 2023, Med. 167.

- [4] Thin Nguyen, Hang Le, Thomas P Quinn, Tri Nguyen, Thuc Duy Le, Svetha Venkatesh, GraphDTA: predicting drug-target binding affinity with graph neural networks, Bioinformatics, Volume 37, Issue 8, March 2021, Pages 1140–1147.

- [5] 彭杨,朱小飞,胡冬冬.基于注意力感知和模态融合的药物-

靶点相互作用预测方法[J/OL].太原理工大学学报:1-10[2024-04-23].

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

<https://creativecommons.org/licenses/by/4.0/>



OPEN ACCESS