

基于图注意力网络预测人类微生物与药物关联



史赛如¹, 孔舒², 张冀¹

1. 河南科技大学数学与统计学院 (河南洛阳 471023)

2. 北京建筑大学理学院 (北京 102616)

【摘要】目的 采用图注意力网络 (graph attention network, GAT) 预测人类微生物与药物之间的潜在关联。**方法** 选取三个常用的微生物-药物关联 (microbe-drug associations, MDA) 数据集 (MDAD、aBiofilm 和 Drug Virus), 基于数据集中丰富的生物信息构建一个异构网络, 并提出一种基于 GAT 框架预测 MDA 的模型——GATMDA 模型, 用于预测微生物与药物间的关联。**结果** 与现有的 8 种预测方法相比, GATMDA 通过三种交叉验证方法在三个数据集上具有较好的预测效果。在 5 折交叉验证的性能评估中, 在三个数据集上的受试者工作特征曲线下的面积 (area under the curve, AUC) 分别为 0.988 6、0.994 1 和 0.983 6, 精确率-召回率曲线下的面积 (area under the precision-recall curve, AUPR) 分别为 0.966 7、0.986 9 和 0.879 5。通过病例研究进一步验证了 GATMDA 在预测 MDA 方面的有效性。**结论** 基于 GAT, GATMDA 模型可以通过构建的异构网络对微生物-药物进行有效的关联预测。

【关键词】 微生物-药物关联; 多核融合; 图注意力网络; 异构网络; 交叉验证

Predicting human microbe-drug associations based on graph attention network

SHI Sairu¹, KONG Shu², ZHANG Ji¹

1. School of Mathematics and Statistics, Henan University of Science and Technology, Luoyang 471023, Henan Province, China

2. School of Science, Beijing University of Civil Engineering and Architecture, Beijing 102616, China

Corresponding author: ZHANG Ji, Email: zhangji@haust.edu.cn

【Abstract】Objective Graph attention network (GAT) was used to predict the potential association between human microbes and drugs. **Methods** Three commonly used microbe-drug associations (MDA) datasets including MDAD, aBiofilm and Drug Virus were selected. Based on the rich biological information in the datasets, a heterogeneous network was constructed and a GAT-based framework for predicting MDA called GATMDA was proposed, which was used to predict the association between microbes and drugs. **Results** Compared with the existing eight prediction methods, GATMDA had better prediction effect on three datasets through three cross-validation methods. During the performance evaluation of 5-fold cross-validation on three datasets, the area under the curve (AUC) were 0.988 6, 0.994 1, and 0.983 6, respectively, and the area under the precision-recall curve (AUPR) were 0.966 7, 0.986 9 and 0.879 5. The effectiveness of GATMDA in predicting MDA was further validated through

DOI: [10.12173/j.issn.1004-4337.202309229](https://doi.org/10.12173/j.issn.1004-4337.202309229)

通信作者: 张冀, 硕士研究生导师, Email: zhangji@haust.edu.cn

<https://slyyx.whuznhmedj.com/>

case studies. **Conclusion** Based on GAT, the GATMDA model can effectively predict MDA through the constructed heterogeneous network.

【Keywords】 Microbe-drug associations; Multiple kernel fusion; Graph attention network; Heterogeneous network; Cross validation

人类微生物是一个复杂而多样的群落，对人类健康有重要影响。成人体内大约有 1 014 个细菌，相当于人类细胞总数的 10 倍，这些细菌可以产生大量的基因产物支持人体内的各种生化或代谢活动^[1]。微生物对人类健康起着重要作用，可能导致转录组、蛋白质组和代谢组的变化，从而进一步损害人体组织，最终导致各种疾病，如肥胖、癌症和糖尿病等^[2-4]。研究表明，微生物参与药物的吸收和代谢，从而调节药物的疗效和毒性^[5-7]。因此，微生物-药物关联 (microbe-drug association, MDA) 研究受到越来越多的关注。

大量潜在的微生物-药物关系已被既往研究证实。Kovac 等证明了粪肠球菌和白色念珠菌菌株对环丙沙星轻微敏感^[6]。Szczuka 等研究发现，环丙沙星可抑制表皮葡萄球菌生物膜的形成^[7]。然而，传统的湿实验室实验用于揭示微生物与药物之间的关联是费力且昂贵的。因此，有效和准确地预测 MDA 的计算方法是对实验方法的补充^[8-11]。Sharma 等于 2017 年开发了一种预测代谢酶和肠道细菌种类的计算方法，可用于药物分子生物转化^[8]。Zhu 等提出了一种基于 KATZ 测量的微生物-药物预测模型^[10]。Long 提出了一种基于图卷积网络的 MDA 预测框架 (GCNMDA)，该框架应用条件随机场可以确保相似的节点具有相似表示^[11]。虽然已经提出了许多预测 MDA 的计算方法，但在特征提取过程中无法保留微生物和药物的综合特征。为了解决上述问题，本研究基于图注意力网络 (graph attention network, GAT) 提出了一种新模型 GATMDA，用于二分网络中的 MDA 预测。

1 资料与方法

1.1 数据来源

本研究数据来源于三个已知微生物与药物关联的数据集^[11]。第一个数据集是 MDAD 数据集，去除冗余信息后，包含 1 373 种药物和 173 种微生物之间的 2 470 种已知关联；第二个数据集是 aBiofilm，存储了抗生物膜制剂的资源及其在抗生

素耐药性方面的潜在影响，其中挑选了 2 884 种微生物-药物对进行研究；第三个数据集是 Drug Virus，记录了多种人类病毒的相关化合物的活性及其发展，包含 95 种病毒和 175 种药物间的 933 种关联。上述三个数据集的详细信息见表 1。

表1 三个数据集的详细数据

Table 1. Detailed data for three datasets

数据集名称	微生物	药物	关联
MDAD	173	1 373	2 470
aBiofilm	140	1 720	2 884
Drug Virus	95	175	933

1.2 方法

1.2.1 人类微生物与药物的关联

为了推测微生物-药物网络中的新关联，本研究将问题框架视为一个生物二分网络的关联预测任务。在此网络中，微生物和药物分别被表示为两类不同的节点。定义药物 N_d 的节点集为 $A = \{a_1, a_2, \dots, a_{N_d}\}$ ，微生物 N_m 的节点集则被定义为 $B = \{b_1, b_2, \dots, b_{N_m}\}$ 。网络中的边是微生物与药物之间的关联，可以表示为邻接矩阵 $Y \in R^{N_d \times N_m}$ 。当 $Y_{ij}=1$ 时，表示一个微生物 b_j ($1 \leq j \leq N_m$) 对应一个药物 a_i ($1 \leq i \leq N_d$)。相反， $Y_{ij}=0$ 表示关联未知。研究目标是生成一个与 Y 相同维度的预测矩阵 F^* 来预测未知的关联。图 1 展示了 GATMDA 算法流程。

1.2.2 构建异构网络

为了将网络信息纳入数据整合，根据邻接矩阵 Y 构建了一个包括微生物网络 S_m 、药物网络 S_d 和 MDA 网络的异构网络^[12]：

$$M = \begin{pmatrix} S_d & Y \\ Y^T & S_m \end{pmatrix} \quad (\text{式 1})$$

1.2.3 图注意力网络

GAT 是一个基于空间的图卷积网络，核心在于聚合邻居特征的过程中聚焦更多重要邻居的特征贡献^[13-14]。GAT 在本研究中被用于提取微生物和药物特征。具体而言，对于上面定义的二进制网络的邻接矩阵，GAT 定义如下：

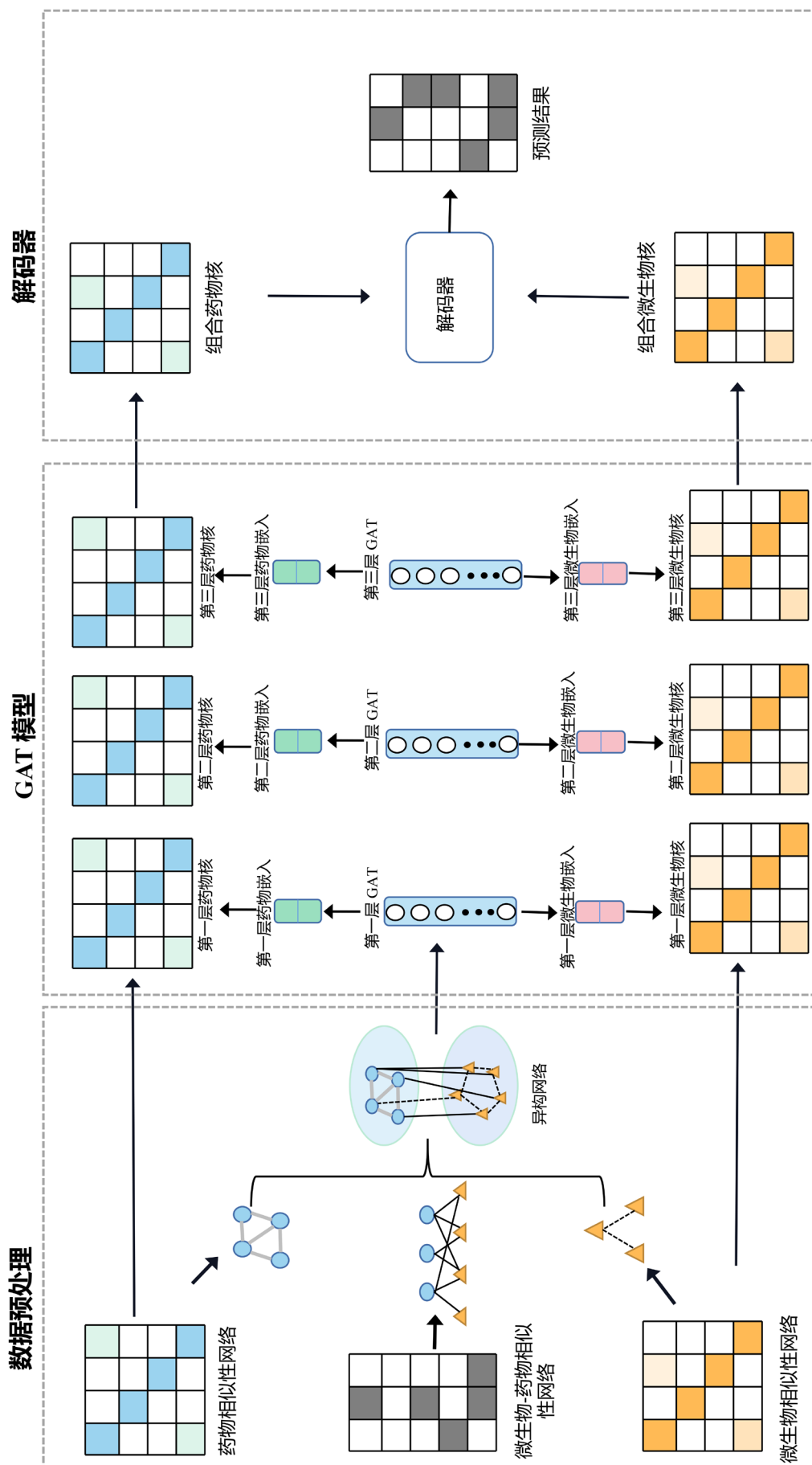


图1 GATMDA算法的原理示意图
Figure 1. The schematic diagram of the GATMDA algorithm

$$H^{(l)} = f(H^{(l-1)}, M) = \sigma(GAT(H^{(l-1)}, M)) \quad (式 2)$$

其中 $H^{(l)}$ 是节点的 l 层嵌入, $l=1, \dots, L$, $\sigma(\cdot)$ 是非线性激活函数 (ReLU), GAT 表示单个图注意力层, 整个 L 层 GAT 架构由多个图注意力层堆叠。初始输入是一组节点特征 $h = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_n\}$, $\vec{h}_i \in R^F$, 其中 n 是节点的数量, F 是每个节点中特征的数量。该层生成一组新的节点特征 $\{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_n\}$, $\vec{h}_i \in R^F$, 并且通过将权重矩阵 $W \in R^{F \times F}$ 应用于每个节点。注意力系数为:

$$e_{ij} = a(W\vec{h}_i, W\vec{h}_j) \quad (式 3)$$

在通过 softmax 函数进行归一化之后, 将系数变为:

$$\alpha_{ij} = \text{soft max}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})} \quad (式 4)$$

将式 3 代入式 4, 可以表示注意力机制的系数如下:

$$\alpha_{ij} = \frac{\exp(\text{Leaky ReLU}(\bar{a}^T [W\vec{h}_i \parallel W\vec{h}_j]))}{\sum_{k \in N_i} \exp(\text{Leaky ReLU}(\bar{a}^T [W\vec{h}_i \parallel W\vec{h}_k]))} \quad (式 5)$$

其中, a 是注意力系数, $\bar{a}^T \in R^{2F}$ 表示参数化权重向量, LeakyReLU 表示激活函数, T 表示矩阵转置, \parallel 是连接运算, N_i 是节点 i 的邻居集合。在计算归一化注意力系数后, 每个节点的最终输出特征可以计算为:

$$\vec{h}_i = \alpha_{ii}W\vec{h}_i + \sum_{k \in N_i} \alpha_{ik}W\vec{h}_k \quad (式 6)$$

对于第一层构造初始嵌入 $H^{(0)}$, 如下所示:

$$H^{(0)} = \begin{pmatrix} 0 & Y \\ Y^T & 0 \end{pmatrix} \quad (式 7)$$

1.2.4 多核融合

多层 GAT 模型可以计算表示具有不同图结构的信息的多个嵌入。由于不同的嵌入表示不同的结构信息, 因此由不同嵌入组成的核将表示不同角度节点之间的相似性。结合现有的相似性矩阵, 可以得到药物空间 $S^D = \{S_d, K_{h_1}^d, \dots, K_{h_L}^d\}$ 和微生物空间 $S^M = \{S_m, K_{h_1}^m, \dots, K_{h_L}^m\}$ 的核集。 $K_{h_i}^d$ 和 $K_{h_i}^m$ 分别是药物和微生物嵌入的核矩阵。为了提高预测性能, 分别在两个空间中对上述核进行了多核融合, 通过加权方法组合多个核矩阵。组合内核定义如下:

$$K_d = \sum_{i=1}^{L+1} a_i^d S_i^d \quad (式 8)$$

$$K_m = \sum_{i=1}^{L+1} b_i^m S_i^m \quad (式 9)$$

其中 S_i^d 和 S_i^m 是药物和微生物核集中的第 i 个核, a_i 和 b_i 是每个核对应的权值, L 是对应的层数。

1.2.5 解码器

最后, 通过应用一个改进的对偶拉普拉斯正

则化最小二乘 (DLapRLS) 框架来预测关联, 提高预测性能。DLapRLS 是基于两个特征空间的核矩阵模型。在这项工作中, 基于 DLapRLS, 将药物 - 微生物特征空间组合纳入最小二乘框架来构建一个新的目标函数。目标函数的定义如下:

$$\min J = \|K_d \alpha_d + (K_m \alpha_m)^T - 2Y_{train}\|_F^2 + \varphi(\|\alpha_d\|_F^2 + \|\alpha_m\|_F^2) \quad (式 10)$$

DLapRLS 方法的目的是通过最小化以上目标函数寻找最优的预测结果。其中, $\|\cdot\|_F$ 是 Frobenius 范数, $Y_{train} \in R^{N_d \times N_m}$ 是训练集中的 MDA 的邻接矩阵; $K_d \in R^{N_d \times N_d}$ 是可训练矩阵; $\alpha_d, \alpha_m^T \in R^{N_d \times N_m}$ 和 $K_m \in R^{N_m \times N_m}$ 分别是两个特征空间中的融合核即微生物和药物之间的相似性度量。 $K_d \alpha_d$ 代表在药物空间下, 药物与微生物的关联预测结果; $K_m \alpha_m$ 代表微生物空间下, 药物与微生物的关联预测结果。 φ 是用来平衡正则化项的衰减因子 (Decay factor)。

由上述可得, $K_d \alpha_d$ 和 $K_m \alpha_m$ 可以实现不同空间下的微生物和药物关联性的表示。故对两个空间的预测结果进行平均融合作为最终预测结果, 进而有效整合微生物 - 药物的空间信息。因此, 基于 DLapRLS, 来自两个特征空间的 MDA 的最终预测 F^* 组合如下:

$$F^* = \frac{K_d \alpha_d + (K_m \alpha_m)^T}{2} \quad (式 11)$$

上述 F^* 也作为模型 GATMDA 的最终输出结果, 根据 F^* 可以得到微生物和药物的关联预测矩阵, 基于该矩阵的得分可以衡量微生物 - 药物关联性。

1.3 统计分析

采用 Python 3.7 软件进行数据分析。本研究建立 GATMDA 模型以预测微生物与药物的关联性, K 折交叉验证用于评估预测性能。在交叉验证期间, 所有关联平均分为 K 个部分。在每一次折叠中, 选择其中一个作为测试集, 其余用作训练集, 用于训练和验证模型, 总共 K 个折叠。模型的性能评估采用受试者工作特征 (receiver operating characteristic, ROC) 曲线和精确率 - 召回率 (precision-recall, PR) 曲线, 评价指标为 ROC 曲线下面积 (area under the curve, AUC) 和精确率 - 召回率曲线下的面积 (area under the precision-recall curve, AUPR)。GATMDA 中重要的参数包括衰减因子 (decay factor) φ 、迭代时间 (iteration time) N 、学习率 (learning rate)。

2 结果

2.1 参数敏感性分析

首先, 在 MDAD 数据集上使用 5 折交叉验证 (5-fold cross-validation, 5-CV) 选择模型参数, 设置默认参数层数 $L=3$, 嵌入维度分别为 $K_1=256, K_2=64, K_3=32$ 。衰减因子 φ 被用于调节公式 (10) 中正则化项的影响, φ 的取值范围为 0.000 005 至 0.5, 步长设定为 10。如图 2-A 和图 2-D 所示, 参数 φ 对模型性能的影响较为有限, 表明模型具备一定的鲁棒性。当 φ 值为 0.000 5 时, 模型达到最优性能。学习率也是一个非常重要的参数, 当学习率太大时, 模型很难收敛, 较小的学习率可能会导致一个较长的训练过程, 一个合理的学习率可以使模型收敛到局部最小值。因此, 学习率被设定在 $\{1e-1, 1e-2, 1e-3, 5e-3, 1e-4, 5e-4\}$, 并对 GATMDA 在各个学习率下的性能进行了评估。如图 2-B 和图 2-E 所示, 学习率 $1e-1$ 从提升至 $5e-4$, GATMDA 的性能先提高后略有下降, 当学习率为 $1e-3$ 时, 模型展现出最佳效果。迭代次数对模型同样至关重要, 其决定了可训练参数的更新频率。本研究将 N 的取值范围设定为 1 至 12, 以 1 为步长。图 2-C 和图 2-F

分别显示了不同迭代次数下的 AUC 值和 AUPR 值, 当迭代次数为 5 时, AUPR 值开始趋于稳定。为了使模型完全收敛, 确定迭代次数为 10。

2.2 与现有预测方法的比较

将 GATMDA 与现有的 8 种生物二分网络预测方法进行比较, 如表 2 所示, 在 MDAD 数据集上, GATMDA 模型的预测性能最好 (AUC=0.988 6, AUPR=0.966 7), 优于其他 8 个模型; 在 aBiofilm 数据集上, GATMDA 模型的 AUC 值和 AUPR 值均最高 (AUC= 0.994 1, AUPR=0.986 9); 在 Drug Virus 数据集上, GATMDA 模型的 AUC 值最高 (AUC=0.983 6), AUPR 值排名第二 (AUPR=0.879 5)。

本研究在 2 折交叉验证 (2-fold cross-validation, 2-CV) 和 10 折交叉验证 (10-fold cross-validation, 10-CV) 的设置下在三个数据集上对所有方法进行了比较, 结果见表 3 和表 4。在 2-CV 的条件下, GATMDA 模型在 MDAD 数据集 (AUC=0.982 6, AUPR=0.948 2)、aBiofilm 数据集 (AUC=0.984 1, AUPR=0.957 1) 和 Drug Virus 数据集 (AUC=0.955 0) 上展现出优于其他方法的预测性能。在 10-CV 下, GATMDA 模型展示了最佳的预测性能, 在 MDAD 数据集上 AUC

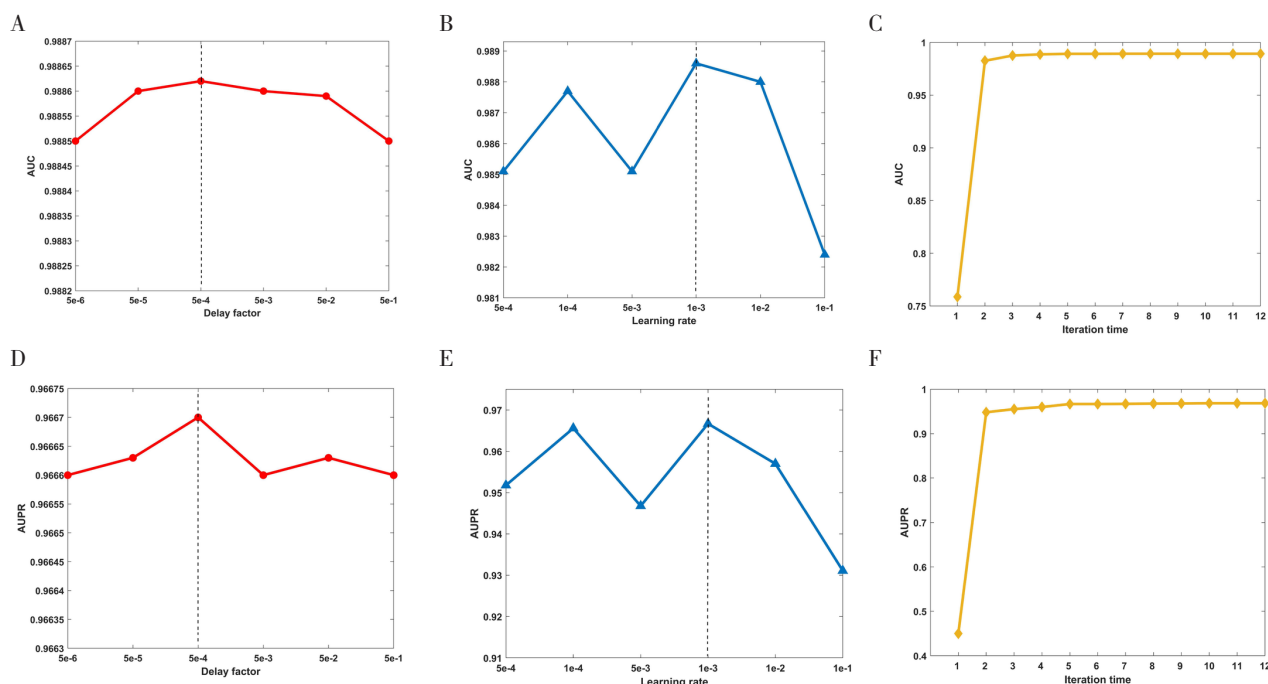


图2 GATMDA的参数灵敏度

Figure 2. Parameter sensitivity of GATMDA

注: A. 衰减因子在不同参数下的AUC得分折线图; B. 学习率在不同参数下的AUC得分折线图; C. 迭代时间在不同参数下的AUC得分折线图; D. 衰减因子在不同参数下的AUPR得分折线图; E. 学习率在不同参数下的AUPR得分折线图; F. 迭代时间在不同参数下的AUPR得分折线图。

表2 三个数据集上不同预测方法在5折交叉验证下的性能比较

Table 2. Performance comparison between different prediction methods on three datasets under 5-CV

方法	MDAD		aBiofilm		Drug Virus	
	AUC	AUPR	AUC	AUPR	AUC	AUPR
KATZHMDA	0.872 3	0.838 4	0.901 3	0.902 0	0.780 9	0.755 4
WMGHMDA	0.865 4	0.838 1	0.845 1	0.890 3	0.723 0	0.768 7
NTSHMDA	0.830 2	0.792 4	0.821 3	0.763 9	0.738 9	0.697 3
IMCMDA	0.746 6	0.777 3	0.775 0	0.857 2	0.623 5	0.696 2
GCMDR	0.848 5	0.850 9	0.877 2	0.884 7	0.824 3	0.820 6
BLM-NII	0.923 1	0.926 3	0.925 6	0.933 8	0.891 3	0.872 2
MKGCN	0.987 6	0.960 3	0.992 6	0.977 7	0.982 1	0.904 8
SCSMDA	0.961 7	0.947 3	0.961 9	0.953 9	0.886 5	0.863 0
GATMDA	0.988 6	0.966 7	0.994 1	0.986 9	0.983 6	0.879 5

注：KATZHMDA、WMGHMDA和NTSHMDA用于预测微生物疾病的相关性^[15]；IMCMDA和GCMDR分别用于microRNA-disease关联的预测和microRNA-drug关系的鉴定；BLM-NII用于解决药物-靶点相互作用^[16]；MKGCN^[17]和SCSMDA^[18]用于预测微生物与药物之间的关联；SCSMDA是基于结构增强对比学习和自适应负采样策略预测微生物-药物关联的最新方法。

表3 三个数据集上不同预测方法在2折交叉验证下的性能比较

Table 3. Performance comparison between different prediction methods on three datasets under 2-CV

方法	MDAD		aBiofilm		Drug Virus	
	AUC	AUPR	AUC	AUPR	AUC	AUPR
KATZHMDA	0.836 5	0.825 9	0.888 7	0.890 9	0.722 1	0.712 4
WMGHMDA	0.851 9	0.820 1	0.830 9	0.880 4	0.688 2	0.729 4
NTSHMDA	0.781 6	0.731 4	0.800 5	0.728 4	0.697 6	0.654 3
IMCMDA	0.728 5	0.761 8	0.760 8	0.840 4	0.573 9	0.657 9
GCMDR	0.843 2	0.842 3	0.876 4	0.882 4	0.802 9	0.806 2
BLM-NII	0.858 3	0.846 9	0.902 3	0.925 9	0.757 0	0.750 2
MKGCN	0.982 0	0.940 0	0.983 1	0.952 0	0.954 0	0.834 7
SCSMDA	-	-	-	-	-	-
GATMDA	0.982 6	0.948 2	0.984 1	0.957 1	0.955 0	0.753 0

表4 三个数据集上不同预测方法在10折交叉验证下的性能比较

Table 4. Performance comparison between different prediction methods on three datasets under 10-CV

方法	MDAD		aBiofilm		Drug Virus	
	AUC	AUPR	AUC	AUPR	AUC	AUPR
KATZHMDA	0.892 9	0.840 3	0.907 6	0.909 8	0.789 5	0.767 0
WMGHMDA	0.872 9	0.845 2	0.844 8	0.883 9	0.723 2	0.762 1
NTSHMDA	0.854 7	0.829 9	0.821 1	0.776 7	0.747 3	0.705 8
IMCMDA	0.746 6	0.772 7	0.774 7	0.857 8	0.637 8	0.718 6
GCMDR	0.859 0	0.863 4	0.873 2	0.878 6	0.828 0	0.827 1
BLM-NII	0.864 4	0.879 2	0.942 5	0.946 3	0.901 2	0.902 4
MKGCN	0.988 3	0.966 2	0.995 1	0.983 0	0.984 3	0.915 1
SCSMDA	-	-	-	-	-	-
GATMDA	0.989 3	0.968 5	0.996 3	0.984 4	0.986 3	0.904 0

值为 0.989 3、AUPR 值为 0.968 5；在 aBiofilm 数据集上 AUC 值为 0.996 3、AUPR 值为 0.984 4；在 Drug Virus 数据集上 AUC 值为 0.986 3、AUPR 值为 0.904 0。GATMDA 在前两个数据集上超越了其他 8 种评估方法，证明 GATMDA 是一个预测 MDA 的高效且强大的模型。GATMDA 模型基于 3 种交叉验证方法在三个数据集上的 ROC 曲线和 PR 曲线见图 3。

2.3 案例研究

通过 Drug Virus 数据集的案例研究，进一步测试 GATMDA 的预测效果。案例研究选择 HIV-1 测试模型的预测性能，并预测了可能有效治疗的药物。HIV 是一种逆转录病毒，可破坏 CD4 T 细胞，是获得性免疫缺陷综合征的病原体。艾滋病毒分为两种类型：HIV-1，引起全球流行病；HIV-2，致病性较弱，主要局限于西非。因

此，选择 HIV-1 为案例进行试验。在试验中，Griffith 等测量了 15 名 HIV 感染患者中 stavudine (2',3'-didehydro-3'-deoxythymine) 对 HIV-1 的抗病毒功效，试验结果显示，stavudine 具有显著而持久的抗病毒作用^[19]。Enfuvirtide 是一种新型 HIV-1 融合抑制剂，在体外和体内均具有针对 HIV-1 的有效抗病毒活性^[20]。如表 5 所示，预测 HIV-1 相关药物中，前 10 名药物全部在文献中得到支持；在预测的前 20 种和 30 种药物中，95% 和 93% 的药物得到了文献的支持，并被证明可以治疗或预防 HIV-1。选择 175 种预测药物中和 HIV-1 关联的前 30 种药物测试 GATMDA 的有效性，见图 4-A。通过绘制条形图和散点图可视化前 30 种预测 HIV-1 的药物，见图 4-B 和图 4-C，这些预测结果证明了 GATMDA 模型预测微生物-药物网络中潜在关联的能力。

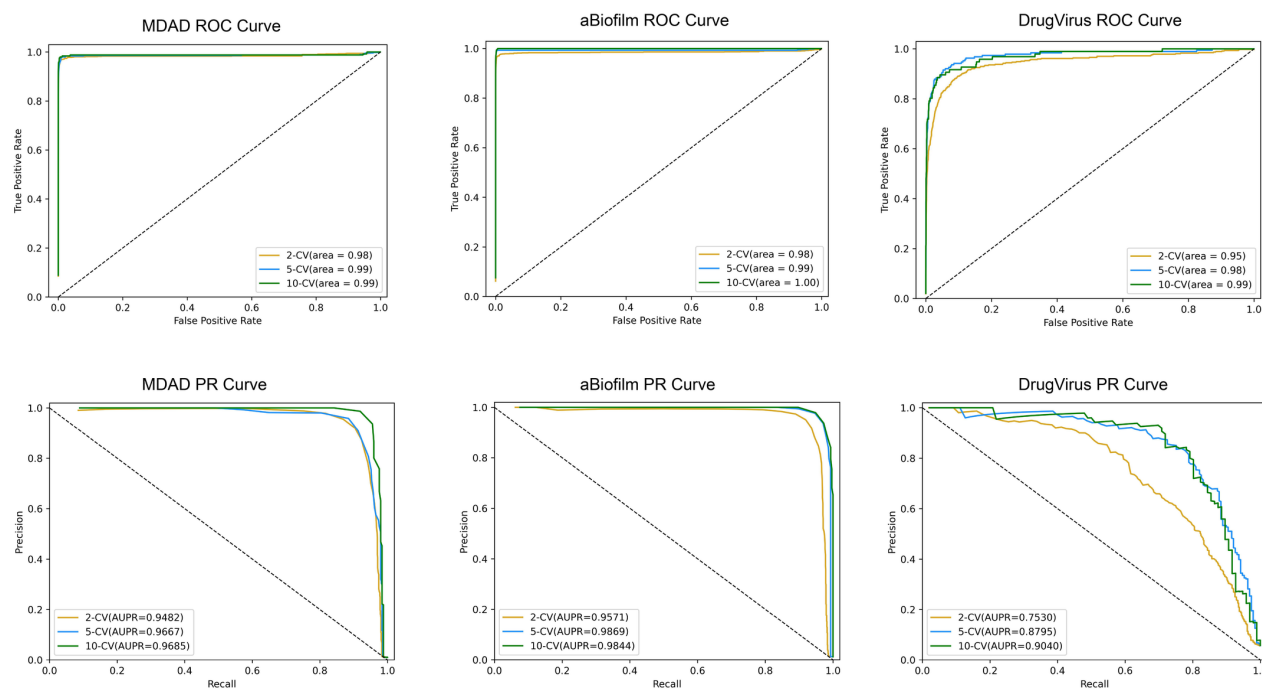


图3 GATMDA在MDAD、aBiofilm和Drug Virus数据集上的ROC和PR曲线

Figure 3. The ROC and PR curves of GATMDA on the MDAD, aBiofilm and Drug Virus datasets

表5 预测HIV-1的前30种药物

Table 5. Top 30 predicted HIV-1-associated drugs

排名	药物	证据	排名	药物	证据
1	Stavudine	PMID: 8627080	16	Bevirimat	PMID: 26482309
2	Enfuvirtide	PMID: 16160172	17	Rilpivirine	PMID: 33257955
3	Abacavir	PMID: 26601161	18	Amprenavir	PMID: 20695887
4	Didanosine	PMID: 11322269	19	Saquinavir	PMID: 22664974

续表5

排名	药物	证据	排名	药物	证据
5	BrecaNavir	PMID: 17261626	20	Darunavir	PMID: 24338166
6	Calanolide A	PMID: 8930167	21	Tenofovir disoproxil	PMID: 12965939
7	Zidovudine	PMID: 7856987	22	Efavirenz	PMID: 15482231
8	Zalcitabine	PMID: 1281077	23	Dolutegravir	PMID: 24694091
9	Tipranavir	PMID: 23015723	24	Fosamprenavir	PMID: 17547501
10	Maraviroc	PMID: 29444937	25	Atazanavir	PMID: 16266202
11	Nevirapine	PMID: 1371691	26	CYT107	-
12	Emtricitabine	PMID: 14609348	27	Thymalfasin	PMID: 28106477
13	Raltegravir	PMID: 21434946	28	Alpha-lipoic acid	PMID: 1724477
14	Delavirdine	-	29	Indinavir	PMID: 8971147
15	Inosine pranobex	PMID: 1703281	30	Trametinib	PMID: 29197575

注：PMID (PubMed Unique Identifier, PubMed唯一标识码)，是PubMed搜索引擎中收录的生命科学和医学等领域的文献编号，使用与国际标准书号(ISBN)和DOI类似。

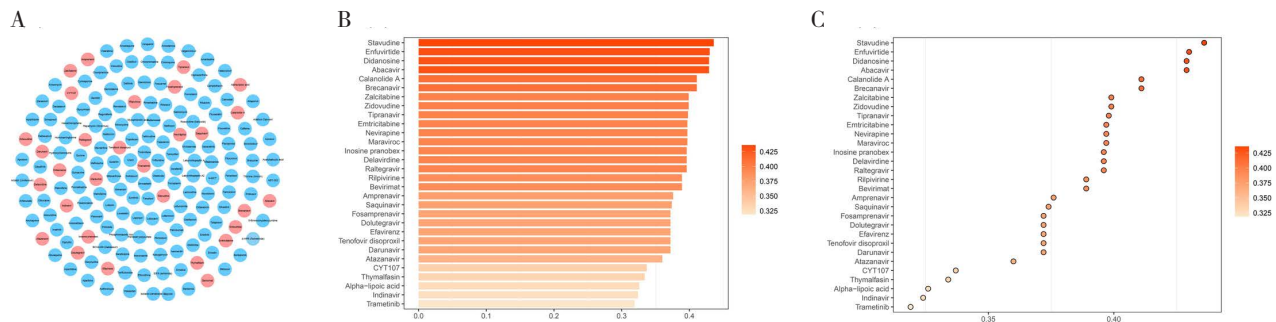


图4 前30种HIV-1预测的相关药物可视化

Figure 4. Visualization of the top 30 predicted HIV-1-related drugs

注：A. Drug Virus数据集的175种预测药物。红色表示与HIV-1关联的前30种药物，其余用蓝色表示；B. 与HIV-1关联的前30种药物的条形图，颜色深浅表示得分的大小，颜色越深，关联性越强；C. 与HIV-1关联的前30种药物的散点图。

3 讨论

居住在人体上的微生物在人类健康中起着关键作用^[21]。预测MDA可以促进个性化药物的有效开发，并了解微生物和药物之间的联系。与传统方法相比，计算方法能够在全球范围内识别靶向现有药物或针对具有已知微生物的新药的靶向微生物^[22]。值得注意的是，MDA预测也是生物二分网络中的一个链接预测问题^[23-24]。本研究提出了一个计算框架GATMDA，用于预测微生物-药物的关联。GATMDA由两部分组成，第一部分使用GAT进行特征提取，实验表明，利用该机制可以生成更可靠的推理信息；另一部分是利用改进的DLapRLS进行预测，充分利用了微生物-药物空间的信息进行预测。与传统的多核学习不同，本研究通过多层GAT提取各种嵌入特征来构建核

矩阵，可以提供不同的核矩阵，并实现使用多种信息的目的^[25]。与现有的生物二分网络检测模型相比，GATMDA模型在三个MDA数据集上表现出了较好的预测性能。此外，关于HIV-1的案例研究表明，GATMDA可以准确地发现新的MDA。

虽然GATMDA具有良好的预测性能，但对于不同密度的数据集仍存在一定的偏差。如GATMDA在Drug Virus数据集上的表现弱于在MDAD和aBiofilm数据集，这表明GATMDA模型的泛化性能仍有改进的空间。微生物对药物治疗过程的影响包括激活、钝化和毒性，准确识别药物上未知微生物的类型是药物开发和精准医学的基本要求，但GATMDA无法预测微生物-药物的类型。因此，为了更准确地了解微生物在药物治疗过程中的作用机制，建立一个有效的深度学习模型来预测微生物与药物之间的关系有待进一

步研究。

综上,本研究通过构建 GATMDA 模型,重点探讨了人类 MDA 预测。GATMDA 模型侧重于结合 GAT 和多核融合来探索微生物-药物的空间信息,进而有效地整合微生物与药物之间丰富的生物学信息,该模型具有捕捉微生物和药物之间复杂关联的能力,为预测微生物与药物关联提供了有效的新方法。

参考文献

- 1 Sommer F, Bäckhed F. The gut microbiota--masters of host development and physiology[J]. *Nat Rev Microbiol*, 2013, 11(4): 227-238. DOI: [10.1038/nrmicro2974](https://doi.org/10.1038/nrmicro2974).
- 2 David LA, Maurice CF, Carmody RN, et al. Diet rapidly and reproducibly alters the human gut microbiome[J]. *Nature*, 2014, 505(7584): 559-563. DOI: [10.1038/nature12820](https://doi.org/10.1038/nature12820).
- 3 Davenport ER, Mizrahi-Man O, Michelini K, et al. Seasonal variation in human gut microbiome composition[J]. *PloS One*, 2014, 9(3): e90731. DOI: [10.1371/journal.pone.0090731](https://doi.org/10.1371/journal.pone.0090731).
- 4 Donia MS, Cimermanic P, Schulze CJ, et al. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics[J]. *Cell*, 2014, 158(6): 1402-1414. DOI: [10.1016/j.cell.2014.08.032](https://doi.org/10.1016/j.cell.2014.08.032).
- 5 Williamson R, Hakenbeck R, Tomasz A. In vivo interaction of beta-lactam antibiotics with the penicillin-binding proteins of streptococcus pneumoniae[J]. *Antimicrob Agents Chemother*, 1980, 18(4): 629-637. DOI: [10.1128/AAC.18.4.629](https://doi.org/10.1128/AAC.18.4.629).
- 6 Kovac J, Kovac D, Slobodnikova L, et al. Enterococcus faecalis and Candida albicans in the dental root canal and periapical infections[J]. *Bratisl Lek Listy*, 2013, 114(12): 716-720. DOI: [10.4149/bl_2013_151](https://doi.org/10.4149/bl_2013_151).
- 7 Szczuka E, Jabłońska L, Kaznowski A. Effect of subinhibitory concentrations of tigecycline and ciprofloxacin on the expression of biofilm-associated genes and biofilm structure of staphylococcus epidermidis[J]. *Microbiology (Reading)*, 2017, 163(5): 712-718. DOI: [10.1099/mic.0.000453](https://doi.org/10.1099/mic.0.000453).
- 8 Sharma AK, Jaiswal SK, Chaudhary N, et al. A novel approach for the prediction of species-specific biotransformation of xenobiotic/drug molecules by the human gut microbiota[J]. *Sci Rep*, 2017, 7(1): 9751. DOI: [10.1038/s41598-017-10203-6](https://doi.org/10.1038/s41598-017-10203-6).
- 9 杨煜清. 基于层次贝叶斯模型的微生物关联网络推断方法研究[D]. 北京: 清华大学, 2019. [Yang YQ. Study of methods of microbial association network inference based on hierarchical Bayesian model[D]. Beijing: Tsinghua University, 2019.] DOI: [10.27266/d.cnki.gqhau.2019.000405](https://doi.org/10.27266/d.cnki.gqhau.2019.000405).
- 10 Zhu LZ, Duan GH, Yan C, et al. Prediction of microbe-drug associations based on chemical structures and the KATZ measure[J]. *Current Bioinformatics*, 2021, 16(6): 807-819. DOI: [10.2174/1574893616666210204144721](https://doi.org/10.2174/1574893616666210204144721).
- 11 Long Y, Wu M, Kwoh CK, et al. Predicting human microbe-drug associations via graph convolutional network with conditional random field[J]. *Bioinformatics*, 2020, 36(19): 4918-4927. DOI: [10.1093/bioinformatics/btaa598](https://doi.org/10.1093/bioinformatics/btaa598).
- 12 于诗睿, 李爱花, 林紫洛, 等. 基于异构网络的相关数据挖掘任务研究综述[J]. *医学信息学杂志*, 2023, 44(4): 28-34. [Yu SR, Li AH, Lin ZL, et al. A review of related data mining tasks based on heterogeneous networks[J]. *Journal of Medical Intelligence*, 2023, 44(4): 28-34.] DOI: [10.3969/j.issn.1673-6036.2023.04.005](https://doi.org/10.3969/j.issn.1673-6036.2023.04.005).
- 13 Wang W, Chen H. Predicting miRNA-disease associations based on graph attention networks and dual Laplacian regularized least squares[J]. *Brief Bioinform*, 2022, 23(5): bbac292. DOI: [10.1093/bib/bbac292](https://doi.org/10.1093/bib/bbac292).
- 14 龙亚辉. 基于图机器学习的微生物网络关系预测算法研究[D]. 长沙: 湖南大学, 2022. [Long YH. Graph-based machine learning algorithms for microbe network prediction[D]. Changsha: Hunan University, 2022.] DOI: [10.27135/d.cnki.ghudu.2021.001004](https://doi.org/10.27135/d.cnki.ghudu.2021.001004).
- 15 Chen X, Huang YA, You ZH, et al. A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases[J]. *Bioinformatics*, 2017, 33(5): 733-739. DOI: [10.1093/bioinformatics/btw715](https://doi.org/10.1093/bioinformatics/btw715).
- 16 Mei JP, Kwoh CK, Yang P, et al. Drug-target interaction prediction by learning from local information and neighbors[J]. *Bioinformatics*, 2013, 29(2): 238-245. DOI: [10.1093/bioinformatics/bts670](https://doi.org/10.1093/bioinformatics/bts670).
- 17 Cui XH, Qu XL, Li DM, et al. MKGCN: multi-modal knowledge graph convolutional network for music

- recommender systems[J]. *Electronics*, 2023, 12(12): 2688. DOI: [10.3390/electronics12122688](https://doi.org/10.3390/electronics12122688).
- 18 Tian Z, Yu Y, Fang H, et al. Predicting microbe–drug associations with structure–enhanced contrastive learning and self–paced negative sampling strategy[J]. *Brief Bioinform*, 2023, 24(2): bba634. DOI: [10.1093/bib/bbac634](https://doi.org/10.1093/bib/bbac634).
- 19 Griffith BP, Brett–Smith H, Kim G, et al. Effect of stavudine on human immunodeficiency virus type 1 virus load as measured by quantitative mononuclear cell culture, plasma RNA, and immune complex–dissociated antigenemia[J]. *J Infect Dis*, 1996, 173(5): 1252–1255. DOI: [10.1093/infdis/173.5.1252](https://doi.org/10.1093/infdis/173.5.1252).
- 20 Mink M, Mosier SM, Janumpalli S, et al. Impact of human immunodeficiency virus type 1 gp41 amino acid substitutions selected during enfuvirtide treatment on gp41 binding and antiviral potency of enfuvirtide in vitro[J]. *J Virol*, 2005, 79(19): 12447–12454. DOI: [10.1128/JVI.79.19.12447-12454.2005](https://doi.org/10.1128/JVI.79.19.12447-12454.2005).
- 21 Aggarwal N, Kitano S, Puah GRY, et al. Microbiome and human health: current understanding, engineering, and enabling technologies[J]. *Chem Rev*, 2023, 123(1): 31–72. DOI: [10.1021/acs.chemrev.2c00431](https://doi.org/10.1021/acs.chemrev.2c00431).
- 22 Madhukar NS, Khade PK, Huang L, et al. A Bayesian machine learning approach for drug target identification using diverse data types[J]. *Nat Commun*, 2019, 10(1): 5221. DOI: [10.1038/s41467-019-12928-6](https://doi.org/10.1038/s41467-019-12928-6).
- 23 Zhang ZC, Zhang XF, Wu M, et al. A graph regularized generalized matrix factorization model for predicting links in biomedical bipartite networks[J]. *Bioinformatics*, 2020, 36(11): 3474–3481. DOI: [10.1093/bioinformatics/btaa157](https://doi.org/10.1093/bioinformatics/btaa157).
- 24 Li Z, Wang X, Li J, et al. Deep attributed network representation learning of complex coupling and interaction[J]. *Knowl. Based Syst*, 2021, 212: 106618. DOI: [10.1016/j.knosys.2020.106618](https://doi.org/10.1016/j.knosys.2020.106618).
- 25 Gönen M, Alpaydin E. Multiple kernel learning algorithms[J]. *Journal of Machine Learning Research*, 2011, 12(64): 2211–2268. DOI: [10.5555/1953048.2021071](https://doi.org/10.5555/1953048.2021071).

收稿日期: 2023 年 09 月 22 日 修回日期: 2023 年 12 月 20 日
本文编辑: 张 苗 黄 笛

引用本文: 史赛如, 孔舒, 张冀. 基于图注意力网络预测人类微生物与药物关联[J]. 数理医药学杂志, 2024, 37(2): 81–90. DOI: [10.12173/j.issn.1004-4337.202309229](https://doi.org/10.12173/j.issn.1004-4337.202309229).
Shi SR, Kong S, Zhang J. Predicting human microbe–drug associations based on graph attention network[J]. *Journal of Mathematical Medicine*, 2024, 37(2): 81–90. DOI: [10.12173/j.issn.1004-4337.202309229](https://doi.org/10.12173/j.issn.1004-4337.202309229).