

doi:10.3969/j.issn.1673-9833.2020.06.005

基于一致性学习预测药物-靶标相互作用

彭利红, 田雄飞, 周立前

(湖南工业大学 计算机学院, 湖南 株洲 412007)

摘要: 提出了一种基于局部全局一致性 (LLGC) 学习的药物-靶标相互作用预测模型。该模型基于邻近结点及流形结构或聚类中的结点更有可能有相同标签这一结论, 综合考虑靶标和药物数据的全局和局部特征, 融合靶标的序列相似性和药物-靶标网络的拓扑结构信息, 提出药物-靶标相互作用预测方法, 挖掘来自标准数据集中的药物-靶标相互作用数据。为了分析局部全局一致性方法的性能, 在酶、离子通道、GPCR 与核受体 4 个数据集中对此方法与 SBGI、KBMF2K、NetCBP 和 WNN-GIP 进行了比较, 实验结果表明, 除了在核受体数据中 LLGC 的 AUC 值比 NetCBP 和 WNN-GIP 中的略低外, 在其他 3 个数据中, LLGC 的性能都优于其他方法。确定模型性能后, 将其用于药物-靶标相互作用数据预测, 给出了得分最高的 5 个药物-靶标相互作用数据, 且得知标准数据集中已知的药物-靶标相互作用数据绝大部分出现在预测集的前 20% 中, 91% 以上出现在预测集的前 50% 中。这个结果表明, LLGC 能有效预测药物与靶标之间的潜在关联。

关键词: 局部全局一致性; 药物-靶标相互作用; 药物重定位; 靶标序列相似性; 机器学习

中图分类号: TM769

文献标志码: A

文章编号: 1673-9833(2020)06-0027-07

引用格式: 彭利红, 田雄飞, 周立前. 基于一致性学习预测药物-靶标相互作用 [J]. 湖南工业大学学报, 2020, 34(6): 27-33.

Prediction of Drug-Target Interactions Based on Consistency Learning

PENG Lihong, TIAN Xiongfei, ZHOU Liqian

(College of Computer Science, Hunan University of Technology, Zhuzhou Hunan 412007, China)

Abstract: A prediction model has been proposed of drug target interaction based on local and global consistency (LLGC) learning. Based on the conclusion that adjacent nodes and nodes in manifold structures or clusters are more likely to have the same label, a comprehensive consideration is given to the global and local characteristics of target and drug data, with the sequence similarities of targets and topological structure information of drug-target network integrated together, and a consistency learning method is proposed to mine the drug target interaction data from standard data set according to standard data set. In order to analyze the performance of LLGC, a comparison has thus been made between the proposed method and SBGI, KBMF2K, NetCBP and WNN-GIP in four datasets of enzyme, ion channel, GPCR and nuclear receptor. The experimental results show that the performance of LLGC is better than other methods except that the AUC value of LLGC in nuclear receptor data is slightly lower than that of NetCBP and WNN-GIP. After determining the performance of the LLGC model, the model is applied for the drug-target interaction prediction, with the five drug-target interaction data with the highest score thus given. The results show that most of the known drug-

收稿日期: 2020-06-29

基金项目: 国家自然科学基金青年基金资助项目 (61803151)

作者简介: 彭利红 (1978-), 女, 湖南岳阳人, 湖南工业大学副教授, 博士, 主要从事生物信息学方面的教学与研究,

E-mail: plhhnu@163.com

通信作者: 周立前 (1970-), 男, 湖南涟源人, 湖南工业大学教授, 博士, 主要从事生物信息学方面的教学与研究,

E-mail: zhoulq11@163.com

target interactions in the four data sets appear in the first 20% of the prediction results, with more than 91% appearing in the first 50% of the prediction set. This result show that LLGC can effectively predict potential associations between drugs and targets.

Keywords: learning with local and global consistency (LLGC); drug-target interaction; drug relocation; target sequence similarity; machine learning

1 研究背景

药物研发是现代社会中的一个重要课题^[1]。通常情况下, 新药研发从确定思路到投入市场大约需要花费 12~15 a 的时间, 成本约 10 亿美元, 且成功率低于 10%^[2]。据统计, 美国食品药品监督管理局 (food and drug administration, FDA) 每年批准能作为新药的分子大约只有 20 种^[3]。诺贝尔奖获得者 James Black 认为, 从旧药研发新药可能是研发新药的有效途径^[4]。

药物重定位又称“老药新用”, 是一种比较好的研发新药方法, 可节约研发时间, 降低研发成本, 减少研发失败带来的风险, 因而受到越来越多的关注。药物重定位是指发现现有药物的新用途, 其生物学基础是一药多靶和一靶多药^[5]。基于计算方法对药物-靶标相互作用进行预测, 以发现某个药物的多个潜在靶标或某个靶标的多个药物, 从而实现对现有药物的重定位, 可以显著减少后期过程 (如临床试验) 的负荷, 节省大量资源和时间。

药物-靶标相互作用预测方法可以分为基于网络的方法和基于机器学习的方法。其中, 基于网络的方法把药物-靶标相互作用矩阵看成是一个网络, 基于网络模型预测其潜在关联。Y. Yamanishi 等^[6]把药物和靶标整合成统一的“药理空间”, 提出一种二分图学习模型, 用于推测靶标与药物之间的相互作用; M. Campillos 等^[7]在“不相关药物的相似脱靶效应可导致药物的相似副作用”这一理论基础上, 根据一体化医学语言系统 (unified medical language system, UMLS) 术语对药物的副作用进行了分类, 提出衡量药物副作用相似性的方法, 预测共享相同靶标的药物; K. Bleakley 等^[8]基于药物的化学结构预测药物的药理相似性, 然后利用局部二分图模型预测未知的药物-靶标相互作用; M. Gönen^[9]使用全概率模型进行预测, 把药物和靶标投影进统一空间, 在此基础上整合贝叶斯模型, 得到了 KBMF2K (Kernelized Bayesian matrix factorization), 再利用协方差计算中的变分近似进行推断。Cheng F. X. 等^[10]提出

了 3 种基于网络的推断方法, 即基于靶标相似性的推断方法 (target-based similarity inference, TBSI)、基于药物相似性的推断方法 (drug-based similarity inference, DBSI) 和基于网络的推断方法 (network-based inference, NBI), 以预测可能的药物-靶标相互作用。

基于机器学习的方法提出机器学习相关模型, 在确定模型性能后对药物与靶标之间潜在的关联进行预测。Y. Yamanishi 等^[11]基于药物的理化性质和药物-靶标相互作用网络的拓扑性质, 提出了基于支持向量机的方法 (supervised bipartite graph ingerence, SBGI) 挖掘可能的药物-靶标相互作用。S. Mizutani 等^[12]基于药物-靶标相互作用谱和副作用谱, 定义稀疏典型相关分析进行预测。L. Jacob 等^[13]利用核函数和支持向量机 (support vector machine, SVM), 研究了药物-靶标的相互作用。Xia Z. 等^[14]提出了一种基于流形正则化的半监督学习模型, 以挖掘药物与靶标之间的潜在关联。T. Van Laarhoven 等^[15]首先使用药物核函数和靶标核函数对相互作用数据进行稀疏化, 然后定义药物相互作用谱和靶标相互作用谱, 根据这些相互作用谱向量定义高斯核函数, 并把高斯相互作用谱核与药物核和靶标核结合起来, 利用正则化最小二乘法寻找能作用于已知靶标的新药。Chen H. L. 等^[16]根据与靶标的相关性对药物进行排序打分, 提出采用网络一致性方法 (network consistency-based prediction method, NetCBP), 推测药物-靶标的相互作用。T. VAN Laarhoven 等^[17]首先定义了药物和靶标的相似性核, 然后在正则化最小二乘法的基础上提出带权最近邻方法 (weighted nearest neighbor-gaussian interaction profile, WNN-GIP)。M. A. Heiskanen 等^[18]对目前的预测方法进行了评估, 发现基于排序的打分方法效果较好。Wang L. 等^[19]利用深度学习中的堆叠自动编码器, 充分提取药物分子的化学子结构和靶标的蛋白序列特征, 以识别药物-靶标相互作用半监督学习模型, 有效识别了药物-靶标相互作用数据。机器学习方法有效预测了药物与靶标之间存在的关联, 然而, 由

于缺少负样本数据,使得监督学习方法预测效果不够理想,而且大部分有监督预测方法的时间复杂度较高^[13];同时,这些方法都严重依赖于药物的化学结构相似性和靶标的序列相似性^[19-24]。

为了缓解这个问题,本文基于局部全局一致性学习方法(learning with local and global consistency, LLGC)^[25],综合考虑靶标蛋白的序列特征和药物-靶标相互作用网络的拓扑性质,提出药物-靶标相互作用挖掘方法的LLGC,通过分析比较,发现所提LLGC方法获得了较好的预测性能。

2 数据和模型

2.1 数据

本文使用的数据由 Y. Yamanishi 等^[6]提供,可从 <http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/> 下载,具体包括药物相似性矩阵、靶标蛋白相似性矩阵及药物-靶标相互作用矩阵。通过 KEGG GENES 数据库^[26]提取靶标的序列信息,基于正则化打分函数以度量其序列相似性:对于两个靶标 t_1 与 t_2 ,假设 $SW(t_1, t_2)$ 表示其 Smith-Waterman 分值,则其序列相似性可用以下公式计算:

$$S_p(t_1, t_2) = \frac{SW(t_1, t_2)}{\sqrt{SW(t_1, t_1)}\sqrt{SW(t_2, t_2)}},$$

从而得到靶标之间的相似性矩阵 S_{sp} 。

Y. Yamanishi 等^[6]分析 DrugBank^[27]、KEGG^[28]和 SuperTarget^[29]等数据库发现,已知药物靶标的人类核受体、G 蛋白偶联受体(GPCRs)、离子通道和酶的数量分别是 54, 223, 210, 445 个,相对应的靶标蛋白分别为 26, 95, 204, 664 个,它们已知的相互作用数分别为 90, 635, 1 476, 2 926^[6, 10, 13-18],详细情况如表 1 所示。在本实验中,使用 Y. Yamanishi 等^[6]提供的数据作为标准数据集,以评估模型的优劣。

表 1 药物-靶标相互作用数据集

Table 1 Drug-target interaction data set

项 目	核受体	GPCRs	离子通道	酶
药物数量	54	223	210	445
靶标数量	26	95	204	664
药物与靶标相互作用的数量	90	635	1 476	2 926

2.2 模型

给定 n 个靶标和 m 个药物,若靶标 i 靶向药物 j ,则 $y_{ij}=1$,否则 $y_{ij}=0$,用矩阵 $Y=[y_1, y_2, \dots, y_n]^T$ 表示靶标靶向药物的情况,即药物-靶标相互作用网络。如果多个靶标蛋白与药物相互作用,则靶标相似性程度可以由共享药物数量来度量。因此,可以基于药物-靶标相互作用网络中靶标共享药物数,来度量

其相似性:

$$S_{np}(i, j) = \frac{Y_i Y_j^T}{\|Y_i\| \|Y_j\|} (i, j=1, 2, \dots, n). \quad (1)$$

综合序列相似性和药物-靶标网络的拓扑结构相似性,可计算如下靶标相似性矩阵:

$$S_p' = S_{sp} + \mu S_{np}. \quad (2)$$

式(1)(2)中: S_{np} 为网络拓扑结构相似性矩阵; S_{sp} 为序列相似性矩阵;参数 μ 用于衡量两者间的重要程度,且

$$\mu = \frac{\sum_{i=1}^n \sum_{j=1}^n S_{np_{ij}}}{\sum_{i=1}^n \sum_{j=1}^n S_{sp_{ij}}}.$$

对 S_p' 进行归一化后得到:

$$S_p = S_{p_{ij}}' / \sum_{k=1}^n S_{p_{ik}}'. \quad (3)$$

用 $F \in \{0, 1\}^{n \times m}$ 表示预测到的药物-靶标相互作用矩阵。 $f_{ij} \in \{0, 1\}$ 表示经过学习后,模型预测到的靶标 i 与药物 j 的关联情况,若值为 1,表示靶标 i 靶向药物 j ,否则靶标 i 与药物 j 没有关联。如果药物 i 与药物 j 非常相似,且靶标靶向药物 i ,则此靶标更有可能靶向药物 j , $F=[f_1, f_2, \dots, f_n]^T$ 。

LLGC 学习认为:邻近的结点更有可能有相同的标签,这是数据的局部性问题;在流形结构或聚类中的结点更有可能具有相同的标签,这是数据的全局性问题。为探究新的药物-靶标相互作用,综合考虑靶标和药物数据的全局和局部特征,提出一致性药物-靶标相互作用预测算法 LLGC,把已知的药物-靶标相互作用信息传播至其邻接点结点,从而获得新的药物-靶标相互作用。

算法 1 用迭代法挖掘药物-靶标相互作用

输入:靶标序列相似性矩阵,药物-靶标相互作用网络。

输出:新的药物-靶标相互作用矩阵 F^* 。

Step 1 由公式(1)~(3)获得靶标相似性矩阵 S_p ,并令 $S_{p_{ii}}=0$,以避免在标签学习过程中对结点进行了自增强处理。

Step 2 对靶标相似性矩阵进行正则化,得到拉普拉斯矩阵 $L=D-S_p$,其中 D 是对角矩阵,第 i 个对角元素值为 $d_i=\sum_{j=1}^n S_{p_{ij}}$ 。

Step 3 对 $F(t+1)=\alpha L F(t)+(1-\alpha)Y$ 迭代求解,直至收敛。 α 为介于 0 和 1 之间的参数。

Step 4 假设 F^* 表示解集 $\{F(t)\}$ 的极限,则 F^* 为所求问题的解。

Step 5 输出预测的药物-靶标关联网络, 预测药物-靶标相互作用。

考虑算法 1 中第 t 步的迭代模型

$$F(t+1)=\alpha L F(t)+(1-\alpha) Y \quad (4)$$

的收敛性。模型第一部分接收来处邻接点的标签信息, 第二部分用于保存初始标签信息。 $L=D-S_p$, 参数 $\alpha \in (0, 1)$, 用于平衡来自邻接点的标签信息和初始标签信息之间的重要程度。

由公式 (4) 得

$$F(t)=(\alpha L)^{t-1} Y+(1-\alpha) \sum_{i=0}^{t-1}(\alpha L)^i Y。 \quad (5)$$

由 $0<\alpha<1$, L 与矩阵 $M=D^{-1} S_p=D^{-\frac{1}{2}} L D^{\frac{1}{2}}$ 相似, 因而 L 的特征值在 $[-1, 1]$ 。从而

$$\lim_{t \rightarrow \infty}(\alpha L)^{t-1}=0, \quad (6)$$

$$\lim_{t \rightarrow \infty} \sum_{i=0}^{t-1}(\alpha L)^i=(I-\alpha L)^{-1}。 \quad (7)$$

假设 $F(t)$ 收敛于 F^* , 对于药物-靶标相互作用预测问题, 可以对其求极限, $F^*=\lim F(t)$, 综合公式 (5)~(7), 可得:

$$F^*=\lim_{t \rightarrow \infty} F(t)=(1-\alpha)(I-\alpha L)^{-1} Y。 \quad (8)$$

其中: I 为 $n \times n$ 的单位矩阵;

$$F(0)=Y。 \quad (9)$$

不失一般性, 公式 (8) 等价于下式:

$$F^*=(I-\alpha L)^{-1} Y。 \quad (10)$$

根据公式 (10), 无需迭代, 可直接计算出药物-靶标相互作用矩阵为

$$F_1^*=(I-\alpha L)^{-1} Y。 \quad (11)$$

若用 M^T 代替 L , 则可得到如下公式 (11) 的另一种形式:

$$F_2^*=(I-\alpha M^T)^{-1} Y。 \quad (12)$$

结合 M 的定义, 可得到另一药物-靶标相互作用矩阵:

$$F_2^*=(D-\alpha S_p)^{-1} Y。 \quad (13)$$

综合公式 (11) 和 (13), 可得到如下发现药物-靶标相互作用的模型

$$F_{\text{fin}}^*=\frac{1}{2} F_1^*+\frac{1}{2} F_2^*=\frac{1}{2}(I-\alpha L)^{-1} Y+\frac{1}{2}(D-\alpha S_p)^{-1} Y。 \quad (14)$$

从而在算法 1 的基础上进行改进, 得到算法 2。

算法 2 基于 LLGC 发现药物-靶标相互作用

输入: 靶标序列相似性矩阵, 药物-靶标相互作用网络。

输出: 药物-靶标相互作用矩阵 F_{fin}^* 。

Step 1 根据公式 (1)~(3), 计算靶标相似

性矩阵 S_p , 并令 $S_{p_{ii}}=0$, 以避免在标签学习过程中对结点进行了自增强处理。

Step 2 对靶标相似性矩阵进行正则化, 得到拉普拉斯矩阵 $L=D-S_p$, 其中 D 是对角矩阵, 第 i 个对角元素值为 $d_i=\sum_{j=1}^n S_{p_{ij}}$ 。

Step 3 根据 Step 2 的计算结果求解 F_1^* 和 F_2^* 。

Step 4 根据公式 (14) 得到靶标的标签矩阵 F_{fin}^* 。

Step 5 输出药物-靶标相互作用矩阵 F_{fin}^* 。

3 实验结果和讨论

实验中, 把药物-靶标相互作用数据以随机形式分成大小相等的 5 份, 依次选择 1 份用作测试集, 剩下的 4 份用作训练集, 实验重复 20 次, 取 20 次实验的平均值作为最终结果。参数 α 取值 0.9。

3.1 与其他方法分析比较

把预测方法 LLGC 与 SBGI、KBMF2K、NetCBP 和 WNN-GIP 进行分析比较, 所得结果如表 2 所示。

表 2 各方法的预测结果汇总表

Table 2 Prediction performance of different DTI prediction methods

数据集	预 测 方 法				
	SBGI	KBMF2K	NetCBP	WNN-GIP	LLGC
酶	0.821	0.832	0.825	0.861	0.876
离子通道	0.692	0.799	0.803	0.775	0.813
GPCRs	0.811	0.857	0.824	0.872	0.883
核受体	0.814	0.824	0.839	0.839	0.827

分析表 2 所示预测结果可以得知, LLGC 方法基于 AUC (area under curve) 的性能要好于 SBGI 和 KBMF2K 方法的。与 SBGI 方法相比, 4 个数据集中 AUC 值分别提高了 6.70%, 17.49%, 8.88%, 1.60%; 与 KBMF2K 方法相比, 4 个数据集中 AUC 值分别提高了 5.29%, 1.75%, 3.03%, 0.36%; 与 WNN-GIP 和 NetCBP 相比, 除在核受体中 AUC 值比 WNN-GIP 和 NetCBP 稍低 (低 1.43%) 之外, 在 AUC 上均好于其他 4 种方法或与其他方法基本持平。

3.2 预测药物-靶标相互作用

课题组基于一致性学习, 考虑靶标相似性和药物-靶标相互作用网络的拓扑性质, 提出一种靶标相似性度量方法。在此基础上, 考虑药物和靶标数据的全局一致性和局部一致性特征, 提出 LLGC 方法寻找药物-靶标相互作用。LLGC 方法将对每个药物与靶标之间的关系进行打分, 并且根据得分对每一个药物-靶标对进行排序, 从而预测药物-靶标的相互作用; 然后, 在 DrugBank 数据库^[27]、KEGG 数据

库^[28]和 SuperTarget^[29]数据库中,对预测到的药物 – 靶标相互作用进行分析验证。

表 3~6 分别列出了核受体中、GPCRs 中、离子通道中和酶中预测到的分值最高的 5 个药物 – 靶标相互作用对。

表 3 核受体中预测到的排序前 5 的
药物 – 靶标相互作用对

Table 3 Predicted top 5 drug-target interaction pairs in nuclear receptor		
排序	药物 – 靶标相互作用	注 释
1	hsa2099	雌激素受体
	D00182	炔诺酮
2	hsa2100	雌激素受体 β
	D00067	雌酮
3	hsa5915	维甲酸受体 β
	D00348	异维 A 酸
4	hsa2100	雌激素受体 β
	D00898	双烯雌酚
5	hsa2100	维甲酸受体 γ
	D00312	硫酸钠雌酮

表 3 所示为核受体中预测到的排序前 5 位的药物 – 靶标相互作用对,以靶标 hsa2099 为例,预测到核受体中,靶标 hsa2099 靶向药物 D00182。在 UniProt 数据库^[30]中,靶标 hsa2099 为核激素受体,调节真核细胞的基因表达,影响细胞的繁殖和目标组织的分化,调节涉及各种激酶的雌激素的信号传导。在 SuperTarget、DrugBank 和 KEGG 数据库中,检索到 hsa2099 靶向药物 D00066、D00067、D00105 和 D00554,其中 D00066 是主要的孕激素类固醇,主要由黄体 and 胎盘分泌。孕激素作用于子宫、乳腺和大脑,有助于胚胎移植、妊娠维护和乳腺组织的发展。D00067 是哺乳动物的主要雌激素,雌激素酮是一种类固醇,主要产生于卵巢、胎盘和外围组织。D00105 是哺乳动物雌激素类固醇的一种形式,主要由卵巢和胎盘产生,可以合成各种同分异构体。D00554 也是一种具有较高的雌激素效能的口服药物。而 D00182 可以合成孕激素,可用于治疗月经不调与子宫内膜异位等疾病。这说明靶标 hsa2099 靶向药物 D00182 有一定的根据。

运用 LLGC 方法对核受体中的标准数据集进行预测,发现在 90 个已知的药物 – 靶标相互作用对中,有 48 个药物 – 靶标相互作用对出现在预测集的前 5% 中,约占有已知数据对的 53.33%; 69 个药物 – 靶标相互作用对在预测集的前 10% 中,约占有已知数据对的 76.67%; 75 个药物 – 靶标相互作用对在预测集的前 20% 中,约占有已知数据对的 83.33%; 所有 90 个已知药物 – 靶标相互作用对在预测集的前 50% 中。

表 4 GPCRs 中预测到的排序前 5 的
药物 – 靶标相互作用对

Table 4 Predicted top 5 drug-target interaction pairs in GPCRs		
排序	药物 – 靶标相互作用	注 释
1	hsa154	肾上腺素能受体 β2
	D00358	美托洛尔
2	hsa154	肾上腺素能受体 α-2C
	D04625	异他林
3	hsa153	肾上腺素能受体 β2
	D01386	盐酸麻黄碱
4	hsa1814	多巴胺受体 D3
	D02354	硫乙拉嗪
5	hsa1814	多巴胺受体 D3
	D00283	氯氮平

表 4 为 GPCRs 中预测到的排序前 5 的药物 – 靶标相互作用对,同时在运用 LLGC 方法对 GPCRs 中的标准数据集进行预测过程中,发现在 223 个已知药物 – 靶标相互作用对中,有 93 个药物 – 靶标相互作用对出现在预测集的前 5% 中,约占有已知数据对的 41.70%; 153 个药物 – 靶标相互作用对在预测集的前 10% 中,约占有已知数据对的 68.61%; 172 个药物 – 靶标相互作用对在预测集的前 20% 中,约占有已知数据对的 77.13%; 203 个药物 – 靶标相互作用对在预测集中的前 50%,约占有已知数据对的 91.03%。

表 5 离子通道中预测到的排序前 5 的
药物 – 靶标相互作用对

Table 5 Predicted top 5 drug-target interaction pairs in ion channel		
排序	药物 – 靶标相互作用	注 释
1	hsa6331	钠通道 V 型亚单位 α
	D00538	唑尼沙胺
2	hsa6331	钠通道 V 型亚单位 α
	D00537	托吡酯
3	hsa6331	钠通道 V 型亚单位 α
	D00552	氨基苯甲酸乙酯
4	hsa6328	钠通道蛋白 3 型亚单位 α
	D00733	地布卡因
5	hsa6328	钠通道蛋白 3 型亚单位 α
	D06172	妥卡尼

表 5 为离子通道中预测到的排序前 5 的药物 – 靶标相互作用对,同时在运用 LLGC 方法对离子通道中的标准数据集进行预测过程中,发现在 1 476 个已知药物 – 靶标相互作用对中,有 514 个药物 – 靶标相互作用对出现在预测集的前 5% 中,约占有已知数据对的 34.82%; 826 个药物 – 靶标相互作用对在预测集的前 10% 中,约占有已知数据对的 55.96%; 1 059 个药物 – 靶标相互作用对在预测集的前 20% 中,约占有已知数据对的 71.75%; 1 344

个药物-靶标相互作用对在预测集的前 50% 中, 约占有已知数据对的 91.06%。

表 6 酶中预测到的排序前 5 的药物-靶标相互作用对

Table 6 Predicted top 5 drug-target interaction pairs in enzyme

排序	药物-靶标相互作用	注 释
1	hsa5742	前列腺素 G/H 合成酶 1
	D00418	米诺地尔
2	hsa5743	前列腺素内过氧化物合酶 1
	D05868	苯丁酸钠
3	hsa5743	前列腺素内过氧化物合酶 2
	D00528	无水咖啡因
4	hsa5742	前列腺素 G/H 合成酶 1
	D00448	吡啶
5	hsa5743	前列腺素内过氧化物合酶 2
	D00097	水杨酸

表 6 给出了酶中预测到的排序前 5 的药物-靶标相互作用对, 同时在运用 LLGC 方法对酶中的标准数据集进行预测过程中, 发现在 2 926 个已知药物-靶标相互作用对中, 有 1 624 个药物-靶标相互作用对出现在预测集的前 5% 中, 约占有已知数据对的 55.50%; 有 2 117 个药物-靶标相互作用对在预测集的前 10% 中, 约占有已知数据对的 72.35%; 2 489 个药物-靶标相互作用对在预测集的前 20% 中, 约占有已知数据对的 85.06%; 2 926 个药物-靶标相互作用对在预测集的前 50% 中, 即所有已知药物-靶标相互作用对都出现在预测集的前 50% 中。

4 结 语

药物-靶标相互作用预测是现代药物研发中的一个重要课题。基于湿实验挖掘药物与靶标之间的相互作用耗时耗力, 故需要开发相应的计算方法以发现这种潜在关联。因而本文基于一致性学习, 融合靶标序列相似性和药物-靶标相互作用网络的拓扑性质, 考虑数据的局部特性和全局特性, 提出一致性药物-靶标相互作用预测方法 LLGC, 以预测来自人类核受体、GPCRs、离子通道和酶中的药物-靶标相互作用。与其他方法相比, LLGC 在 AUC 上优于其他方法或与其他方法持平。同时, 标准数据集中绝大部分已知的药物-靶标相互作用预测出现在预测集的前 20% 中, 91% 以上出现在预测集的前 50% 中。今后, 课题组将结合深度学习相关模型和各种生物特征, 以改进药物-靶标相互作用的预测性能。

参考文献:

[1] MADHUKAR N S, KHADE P K, HUANG L D,

et al. A Bayesian Machine Learning Approach for Drug Target Identification Using Diverse Data Types[J]. Nature Communications, 2019, 10(1): 5221.

- [2] SUN M Y, ZHAO S D, GILVARY C, et al. Graph Convolutional Networks for Computational Drug Development and Discovery[J]. Briefings in Bioinformatics, 2020, 21(3): 919-935.
- [3] LIN H H, ZHANG L L, YAN R, et al. Network Analysis of Drug-Target Interactions: a Study on FDA-Approved New Molecular Entities Between 2000 to 2015[J]. Scientific Reports, 2017, 7(1): 12230.
- [4] TURANLI B, ALTAY O, BORÉN J, et al. Systems Biology Based Drug Repositioning for Development of Cancer Therapy[J]. Seminars in Cancer Biology, 2019, 9(20): 1-8. doi:10.1016/j.semcancer.2019.09.020.
- [5] ANDRADE C H, NEVES B J, MELO-FILHO C C, et al. In Silico Chemogenomics Drug Repositioning Strategies for Neglected Tropical Diseases[J]. Current Medicinal Chemistry, 2019, 26(23): 4355-4379.
- [6] YAMANISHI Y, ARAKI M, GUTTERIDGE A, et al. Prediction of Drug-Target Interaction Networks from the Integration of Chemical and Genomic Spaces[J]. Bioinformatics, 2008, 24(13): i232-i240.
- [7] CAMPILLOS M, KUHN M, GAVIN A C, et al. Drug Target Identification Using Side-Effect Similarity[J]. Science, 2008, 321: 263-266.
- [8] BLEAKLEY K, YAMANISHI Y. Supervised Prediction of Drug-Target Interactions Using Bipartite Local Models[J]. Bioinformatics, 2009, 25(18): 2397-2403.
- [9] GÖNEN M. Predicting Drug-Target Interactions from Chemical and Genomic Kernels Using Bayesian Matrix Factorization[J]. Bioinformatics, 2012, 28(18): 2304-2310.
- [10] CHENG F X, LIU C, JIANG J, et al. Prediction of Drug-Target Interactions and Drug Repositioning via Network-Based Inference[J]. PLoS Computational Biology, 2012, 8(5): e1002503.
- [11] YAMANISHI Y, KOTERA M, KANEHISA M, et al. Drug-Target Interaction Prediction from Chemical, Genomic and Pharmacological Data in an Integrated Framework[J]. Bioinformatics (Oxford, England), 2010, 26(12): i246-i254.
- [12] MIZUTANI S, PAUWELS E, STOVEN V, et al. Relating Drug-Protein Interaction Network with Drug Side Effects[J]. Bioinformatics, 2012, 28(18): i522-i528.
- [13] JACOB L, VERT J P. Protein-Ligand Interaction Prediction: an Improved Chemogenomics Approach[J]. Bioinformatics, 2008, 24(19): 2149-2156.
- [14] XIA Z, WU L Y, ZHOU X B, et al. Semi-Supervised Drug-Protein Interaction Prediction from Heterogeneous

- Biological Spaces[J]. BMC Systems Biology, 2010, 4(2): 1-16.
- [15] VAN LAARHOVEN T, NABUURS S B, MARCHIORI E. Gaussian Interaction Profile Kernels for Predicting Drug-Target Interaction[J]. Bioinformatics, 2011, 27(21): 3036-3043.
- [16] CHEN H L, ZHANG Z P. A Semi-Supervised Method for Drug-Target Interaction Prediction with Consistency in Networks[J]. PLoS One, 2013, 8(5): e62975.
- [17] VAN LAARHOVEN T, MARCHIORI E. Predicting Drug-Target Interactions for New Drug Compounds Using a Weighted Nearest Neighbor Profile[J]. PLoS One, 2013, 8(6): 1-6.
- [18] HEISKANEN M A, AITTOKALLIO T. Predicting Drug-Target Interactions Through Integrative Analysis of Chemogenetic Assays in Yeast[J]. Molecular BioSystems, 2013, 9(4): 768-779.
- [19] WANG L, YOU Z H, CHEN X, et al. A Computational-Based Method for Predicting Drug-Target Interactions by Using Stacked Autoencoder Deep Neural Network[J]. Journal of Computational Biology, 2018, 25(3): 361-373.
- [20] TANG J, TANOLI Z U R, RAVIKUMAR B, et al. Drug Target Commons: a Community Effort to Build a Consensus Knowledge Base for Drug-Target Interactions[J]. Cell Chemical Biology, 2018, 25(2): 224-229.
- [21] EZZAT A, WU M, LI X L, et al. Computational Prediction of Drug-Target Interactions Using Chemogenomic Approaches: an Empirical Survey[J]. Briefings in Bioinformatics, 2019, 20(4): 1337-1357.
- [22] CHEN X, YAN C C, ZHANG X T, et al. Drug-Target Interaction Prediction: Databases, Web Servers and Computational Models[J]. Briefings in Bioinformatics, 2016, 17(4): 696-712.
- [23] MENG F R, YOU Z H, CHEN X, et al. Prediction of Drug-Target Interaction Networks from the Integration of Protein Sequences and Drug Chemical Structures[J]. Molecules, 2017, 22(7): 1119.
- [24] ZHOU L Q, LI Z J, YANG J L, et al. Revealing Drug-Target Interactions with Computational Models and Algorithms[J]. Molecules, 2019, 24(9): 1714.
- [25] ZHOU D Y, BOUSQUET O, LAL T N, et al. Learning with Local and Global Consistency[J]. Advances in Neural Information Processing Systems, 2003, 16(3). doi: 10.1007/s11222-007-9033-z.
- [26] KANEHISA M, FURUMICHI M, TANABE M, et al. KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs[J]. Nucleic Acids Research, 2017, 45(D1): D353-D361.
- [27] WISHART D S, FEUNANG Y D, GUO A C, et al. DrugBank 5.0: a Major Update to the DrugBank Database for 2018[J]. Nucleic Acids Research, 2018, 46(D1): D1074-D1082.
- [28] KANEHISA M, SATO Y, FURUMICHI M, et al. New Approach for Understanding Genome Variations in KEGG[J]. Nucleic Acids Research, 2019, 47(D1): D590-D595.
- [29] HECKER N, AHMED J, VON EICHBORN J, et al. SuperTarget Goes Quantitative: Update on Drug-Target Interactions[J]. Nucleic Acids Research, 2012, 40(D1): D1113-D1117.
- [30] UNIPROT CONSORTIUM T. UniProt: The Universal Protein Knowledgebase[J]. Nucleic Acids Research, 2018, 46(5): 2699.

(责任编辑: 廖友媛)