

引用格式:闫效莺,吴莹,李润洲.基于流形正则化非负矩阵分解预测药物-靶蛋白作用关系[J].科学技术与工程,2019,19(33):325-329  
Yan Xiaoying, Wu ying, Li Runzhou. Prediction of drug-target interaction with manifold regularized non-negative matrix factorization[J]. Science Technology and Engineering, 2019, 19(33): 325-329

# 基于流形正则化非负矩阵分解预测 药物-靶蛋白作用关系

闫效莺<sup>1,2</sup> 吴莹<sup>1</sup> 李润洲<sup>1</sup>

(西安石油大学计算机学院<sup>1</sup>,西安 710065;西北工业大学自动化学院<sup>2</sup>,西安 710072)

**摘 要** 识别药物-靶蛋白作用关系是当前药物研究的重要内容,其可帮助识别已有药物的新功能,发现药物的“偏靶蛋白”等。现有预测算法对新药物的作用靶蛋白,及新靶蛋白的作用药物预测存在困难,由此提出一种新奇的基于流形正则化非负矩阵分解的新药物/新靶蛋白作用关系预测算法,该方法首先通过聚类算法构建新药物/新靶蛋白的初始作用标签,然后设计引入流形学习正则化约束的非负矩阵分解算法预测药物-靶蛋白作用关系,最后在四个经典数据集中测试,并与最新预测算法 BLM-NII、RLS-WNN 和 WKNKN + WGRMF 算法进行比较,证明本文算法可获取较高的预测精度。  
**关键词** 药物 靶蛋白 聚类 流形学习正则化 非负矩阵分解  
**中图法分类号** TP391.45; **文献标志码** A

药物重定位是识别已有药物新功能的重要手段,但是通过生物实验方法识别药物-靶蛋白作用关系不仅费钱、费力,而且费时,因此需要开发计算方法,高效预测药物-靶蛋白作用关系<sup>[1]</sup>。

药物-靶蛋白作用关系预测可分为四大场景<sup>[2]</sup>: S1 场景是预测已知药物和已知靶蛋白之间新的作用关系;S2 场景是预测新药物与已知靶蛋白之间是否存在作用关系;S3 场景是预测新靶蛋白与已知药物之间是否存在作用关系;S4 场景则是预测新药物与新靶蛋白之间是否存在作用关系。这里的“已知药物/靶蛋白”指在当前已知的药物-靶蛋白作用关系中,存在与该药物/靶蛋白有作用关系的靶蛋白/药物;“新药物/靶蛋白”指在当前已知的药物-靶蛋白作用关系中,不存在该药物/靶蛋白的任何作用关系。S1 场景中,由于药物和靶蛋白都存在其他已知作用关系,故较容易实现。S2 和 S3 场景分别对应新药物、新靶蛋白的作用关系预测,所涉及的药物和靶蛋白没有任何已知作用关系,只能借助药物特征(或相似性)矩阵、靶蛋白特征(或相似性)矩阵进行预测。S4 场景可通过组合 S2 和 S3 场景的结果得到<sup>[3]</sup>。

当前,大多数基于机器学习的药物-靶蛋白作用

关系预测方法都可以很好地实现 S1 场景预测,如基于分类器模型的 BLM 算法<sup>[4]</sup>,将已知药物-靶蛋白对作为正样本,其他未知的作用关系作为负样本,采用 SVM 分类器预测药物-靶蛋白作用关系。RLS-Kron 算法<sup>[5]</sup>基于高斯谱核矩阵,采用正则化最小二乘算法预测药物-靶蛋白作用关系。这里提出的基于异构网络交互信息标签传播预测 LPMIHN 算法<sup>[6]</sup>和基于回归模型的网络排序算法<sup>[7]</sup>都是基于网络相关理论的预测方法。这些算法都可以很好地实现已知药物和已知靶蛋白之间新潜在作用关系预测。但是,对新药物/新靶蛋白的靶蛋白/药物作用关系预测困难。

为解决此问题,Mei 等<sup>[8]</sup>对 BLM 算法进行改进,提出基于邻居谱(neighbor-based interaction-profile inferring, NII)的 BLM-NII 算法;Van 等<sup>[9]</sup>提出加权邻居谱(weighted nearest neighbor, WNN)的思路对 RLS-Kron 算法<sup>[5]</sup>进行改进;基于此, Ezzat 等<sup>[10]</sup>在图正则化矩阵分解算法模型中引入加权 K 近邻(WKNKN)思想,用于解决新药物/新靶蛋白的作用关系预测。然而,这里基于邻居谱 NII、基于加权邻居谱 WNN,及基于加权 K 近邻 WKNKN 算法的核心思想都是采用加权邻居谱构建新药物/新靶蛋白的先验信息。近年来有研究表明<sup>[11,12]</sup>:位于同一个聚类社团内的药物和靶蛋白具有较高相似性,融入更多异构数据源特征,如药物-药物相互作用网络(drug-drug interaction, DDI),蛋白作用关系网络(protein-protein interaction, PPI)可进一步提高药物-

2019 年 3 月 24 日收到 国家自然科学基金项目(51707158)、陕西省教育厅项目(17JK0603)和西安石油大学博士科研启动基金(134890003)资助  
第一作者简介:闫效莺,山西阳泉人,博士,讲师。研究方向:生物信息学、机器学习等。E-mail: xiaoying\_yan@126.com。

靶蛋白作用关系预测精度。

基于此,提出一种融合聚类社团结果的流形正则化非负矩阵分解算法(MRNMFC-DT)用于药物-靶蛋白作用关系预测,首先基于 DDI 和 PPI 网络聚类结果修正药物相似性矩阵和靶蛋白相似性矩阵;之后基于修正后的药物相似性矩阵和靶蛋白相似性矩阵分别进行社团聚类,并由聚类结果构建新药物/新靶蛋白的初始交互作用谱;最后采用流形正则化非负矩阵分解算法预测药物-靶蛋白作用关系。

1 数据来源

现采用 Yamanishi 等<sup>[4]</sup>收集整理的包含人类四大蛋白质家族的已知药物-靶蛋白作用关系数据集。这四个数据集分别为酶(EN)、离子通道(IC)、G 蛋白偶联受体(GPCR)和核受体(NR)(可以从 <http://web.kuicr.kyoto-u.ac.jp/supp/yoshi/drugtarget/> 中下载得到)。

每个数据集都包含三个矩阵,药物相似性矩阵、靶蛋白相似性矩阵和药物-靶蛋白作用关系矩阵。其中,由  $n$  个药物、 $m$  个靶蛋白组成的作用关系矩阵  $Y_{n \times m}$  是一个 0/1 矩阵,若  $d_i$  与  $t_j$  之间存在已知作用关系,则  $y(i, j) = 1$ ; 否则,  $y(i, j) = 0$ 。此外,  $Y_{n \times m}$  矩阵的第  $i$  行是药物  $d_i$  的作用谱,记为  $Y(d_i)$ ;  $Y_{n \times m}$  矩阵的第  $j$  列是靶蛋白  $t_j$  的作用谱,记为  $Y(t_j)$ 。表 1 列出了这四个数据集的统计信息,其中核受体( $N_R$ )数据集中,药物( $N_d$ )有 54 个,靶蛋白( $N_t$ )有 26 个,已知的药物-靶蛋白作用关系( $E_{dt}$ )有 90 个。

表 1 四个药物-靶点关系数据集的统计特性  
Table 1 Statistical characteristics of four datasets

| Dataset | $N_d$ | $N_t$ | $E_{dt}$ |
|---------|-------|-------|----------|
| EN      | 445   | 664   | 2 926    |
| IC      | 210   | 204   | 1 476    |
| GPCR    | 223   | 95    | 635      |
| NR      | 54    | 26    | 90       |

2 药物/靶蛋白相似性矩阵构建及调整

药物相似性矩阵  $S^d_{n \times n}$  是基于药物化合物结构数据计算产生的;靶蛋白相似性矩阵  $S^t_{m \times m}$  是基于靶蛋白序列相似性信息计算产生的,详细计算过程可参考文献[7]。考虑融入更多异构数据源特征(如药物-药物相互作用网络 DDI, 蛋白作用关系网络 PPI)可进一步提高药物-靶蛋白作用关系预测精度<sup>[12]</sup>, 现使用基于 DDI 网络和 PPI 网络的聚类结果对相似性矩阵进行修正,具体过程如下:

基于 DrugBank (V3)<sup>[13]</sup> 中下载整理的药物-药物相互作用网络 DDI 数据,构建药物-药物权重网络  $N_{DDI}$ ; 基于 HPRD 数据集<sup>[14]</sup> 中下载整理的蛋白作用关系网络 PPI 数据,构建靶蛋白-靶蛋白权重网络  $N_{PPI}$ , 网络中的节点分别代表药物和靶蛋白,边权重则是在 DDI 和 PPI 网络中两个节点之间共同拥有的药物/蛋白数目。

之后,采用 ClusterONE 算法<sup>[15]</sup>, 基于  $N_{DDI}$  和  $N_{PPI}$  网络分别进行社团挖掘, ClusterONE 是一种最早用于解决蛋白质复合物类别识别问题而提出的基于内聚力引导搜索的聚类算法,目前已成功地应用于药物重定位等相关研究中。经 ClusterONE 算法聚类后得到的社团内聚力值定义为

$$Q(M) = \frac{w^{in}(M)}{w^{in}(M) + w^{bound}(M) + p(M)} \tag{1}$$

式(1)中,  $w^{in}(M)$  指社团  $M$  内各节点之间边的连接权重和;  $w^{bound}(M)$  指社团  $M$  内部节点与  $M$  之外节点之间边的连接权重和;  $p(M)$  为惩罚项,用来平衡数据本身的误差影响。

通常,属于同一个聚类团的药物/靶蛋白,其行为、功能较相似<sup>[11,12]</sup>, 由此,基于药物-药物权重网络  $N_{DDI}$  和靶蛋白-靶蛋白权重网络  $N_{PPI}$ , 使用 ClusterONE 算法分别对药物和靶蛋白进行聚类分析,并使用聚类团的内聚力值对药物结构相似性矩阵和靶蛋白序列相似性矩阵分别进行如下调整:

若药物  $d_i$  和  $d_j$  属于同一个聚类团  $M^1_d$ , 则将  $d_i$  和  $d_j$  之间的相似性值  $S_{d_{ij}}$  按式(2)进行调整:

$$S^1_{d_{ij}} = [1 + Q_d(M^1_d)] S_{d_{ij}} \tag{2}$$

式(2)中,  $Q_d(M^1_d)$  指聚类团  $M^1_d$  的内聚力值。与此类似,若靶蛋白  $t_i$  和  $t_j$  属于同一个聚类社团  $M^1_t$ , 则将  $t_i$  和  $t_j$  的相似性值  $S_{t_{ij}}$  进行调整,  $S^1_{t_{ij}} = [1 + Q_t(M^1_t)] S_{t_{ij}}$ 。由此,得到调整后的药物相似性矩阵  $S^1_d(d^1_{ij})$  和靶蛋白相似性矩阵  $S^1_t(t^1_{ij})$ 。

3 融合聚类社团结果的流形正则化非负矩阵分解算法 MRNMFC-DT

3.1 新药物/新靶蛋白初始标签构建

对新药物  $d_k$ , 由于当前没有任何靶蛋白与之有作用关系,所以其作用谱  $Y(d_k)$  中元素全为 0; 同样地,新靶蛋白  $t_l$  的作用谱  $Y(t_l)$  也都是 0。本文提出采用 ClusterONE<sup>[15]</sup> 聚类算法,基于调整后的药物相似性矩阵  $S^1_d(d^1_{ij})$  和调整后的靶蛋白相似性矩阵  $S^1_t(t^1_{ij})$  分别对药物和靶蛋白进行聚类分析,并根据社团聚类结果初始化药物和靶蛋白的当前作用谱。

对药物  $d_k$  而言,设与其属于同一个聚类社团

$M_d^2$  的其他已知药物集合为  $\{d_i\}$ , 则通过式(3) 可得到药物  $d_k$  的新作用谱数据:

$$Y'_d(d_k) = \frac{1}{Q_d} \sum_{d_i \in M_d^2} w_i Y_d(d_i) \tag{3}$$

式(3) 中,  $d_i$  是与药物  $d_k$  同属于一个聚类社团  $M_d^2$  的已知药物;  $i$  为对  $M_d^2$  中已知药物, 根据其与  $d_k$  的相似性值由大到小排序后得到的序号; 加权系数  $w_i$ , 定义为  $w_i = \mu^{i-1} S_d^1(d_k, d_i)$ ,  $\mu \in [0, 1]$ ;  $Q_d$  用于对加权和进行归一化处理, 定义为  $Q_d = \sum_{i \in M_d^2} S_d^1(d_k, d_i)$ 。

对靶蛋白  $t_l$ , 其作用谱的构建过程同上, 采用如下公式计算:

$$Y'_t(t_l) = \frac{1}{Q_t} \sum_{t_j \in M_t^2} w_j Y_t(t_j) \tag{4}$$

$$w_j = \mu^{j-1} S_t^1(t_l, t_j), \quad \mu \in [0, 1] \tag{5}$$

$$Q_t = \sum_{t_j \in M_t^2} S_t^1(t_l, t_j) \tag{6}$$

所有药物和靶蛋白的新作用谱向量按序排列后, 分别形成新作用谱矩阵  $Y'_d$  和  $Y'_t$ , 之后对这两个矩阵加权平均, 形成新的作用谱矩阵:

$$Y_{\text{new}} = \frac{a_1 Y_d + a_2 Y_t}{a_1 + a_2} \tag{7}$$

为方便计算, 设置  $a_1 = a_2 = 1$ 。之后通过将原作用谱矩阵  $Y_{n \times m}$  中值为 0 的元素, 使用新作用谱矩阵  $Y_{\text{new}}$  中对应元素值代替, 得到当前药物-靶蛋白作用关系矩阵  $Y'$ 。

3.2 流形正则化非负矩阵分解算法

对非负矩阵  $Y'_{n \times m}$ , 采用标准 NMF 可将其分解为两个低秩非负矩阵  $A_{n \times k}$  和  $B_{m \times k} [k \ll \min(n, m)]$ 。经迭代运算, 不断减小  $Y'$  和  $AB^T$  之间的重构误差, 对应目标函数为

$$\min_{A,B} \|Y' - AB\|_F^2, \quad A \geq 0, B \geq 0 \tag{8}$$

考虑标准 NMF 只是工作在欧氏空间, 不能发现原始数据空间内在的几何结构。为了进一步增强学习能力, Cai 等<sup>[16]</sup> 提出一种图正则化的非负矩阵分解算法, 其将流形学习的思想引入 NMF 中, 该算法已成功地应用于 microRNA-疾病作用关系预测<sup>[17]</sup>, 及高光谱数据降维研究中<sup>[18]</sup>。更新后的目标函数定义<sup>[17]</sup> 为

$$\begin{aligned} \min_{A,B} & \|Y' - AB\|_F^2 + \alpha (\|A\|_F^2 + \|B\|_F^2) + \\ & \beta \sum_{i,p=1}^n \|a_i - a_p\|^2 S_d^*(d_i, d_p) + \\ & \gamma \sum_{j,q=1}^m \|b_j - b_q\|^2 S_t^*(t_j, t_q); \\ & A \geq 0, B \geq 0 \end{aligned} \tag{9}$$

式(9) 中,  $\alpha, \beta$  和  $\gamma$  均为正则化系数;  $a_i, b_j$  分别是  $A$

和  $B$  矩阵的第  $i$  行和第  $j$  行;  $n$  和  $m$  分别为药物和靶蛋白数目; 式中的第一项对应  $Y'$  的 NMF 分解; 第二项是对低维表示进行平滑的 L2 正则化; 最后两项是流形学习正则项, 其认为如果两个药物/靶蛋白在其自身相似性空间很接近, 那么在映射空间也应当很接近<sup>[18]</sup>;  $S_d^*$  和  $S_t^*$  是稀疏的药物和靶蛋白相似性矩阵。

3.3 相似性矩阵稀疏化

引入流形学习正则项的 NMF 算法需要进行相似性矩阵稀疏化处理<sup>[17]</sup>, 本文仍然使用 ClusterONE 算法的聚类结果对药物和靶蛋白相似性矩阵进行稀疏化。

对药物相似性矩阵而言, 首先使用基于加权网络  $N_{\text{DDI}}$  的 ClusterONE 聚类结果  $M_d^1$  和基于  $S_d^1(d_{ij}^1)$  的 ClusterONE 聚类结果  $M_d^2$ , 通过如下公式构建一个权重矩阵  $W^d$ , 基本思路: 如果药物  $d_i$  和  $d_j$  同时属于两个聚类社团, 则  $W^d$  中相应元素值为 1, 若两个药物只属于其中一个聚类社团, 则  $W^d$  中相应元素值为 0.5, 否则元素值为 0。

$$W_{ij}^d = \begin{cases} 1, & d_i, d_j \in M_d^1 \ \& \ d_i, d_j \in M_d^2 \\ 0.5, & d_i, d_j \in M_d^1 \mid d_i, d_j \in M_d^2 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

类似地, 对靶蛋白相似性矩阵的稀疏化, 则是首先使用基于加权网络  $N_{\text{ppi}}$  的 ClusterONE 聚类结果  $M_t^1$  和基于  $S_t^1(t_{ij}^1)$  的 ClusterONE 聚类结果  $M_t^2$ , 通过式(11) 构建一个权重矩阵  $W^t$ 。

$$W_{ij}^t = \begin{cases} 1, & t_i, t_j \in M_t^1 \ \& \ t_i, t_j \in M_t^2 \\ 0.5, & t_i, t_j \in M_t^1 \mid t_i, t_j \in M_t^2 \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

之后, 使用  $S_{ij}^{d*} = W_{ij}^d S_{ij}^{d1}$  和  $S_{ij}^{t*} = W_{ij}^t S_{ij}^{t1}$  分别产生稀疏的药物相似性矩阵  $S^{d*}$  和稀疏的靶蛋白相似性矩阵  $S^{t*}$ 。

3.4 优化求解过程

流形正则化非负矩阵分解算法目标函数写为矩阵形式<sup>[18]</sup>:

$$\begin{aligned} \min_{A,B} & \|Y - AB^T\|_F^2 + \alpha (\|A\|_F^2 + \|B\|_F^2) + \\ & \beta \text{Tr}(A^T L_d A) + \gamma \text{Tr}(B^T L_t B), \quad A \geq 0, B \geq 0 \end{aligned} \tag{12}$$

式(12) 中,  $\text{Tr}(\cdot)$  是矩阵的迹;  $L_d = D_d - S^{d*}$  和  $L_t = D_t - S^{t*}$  是  $S^{d*}$  和  $S^{t*}$  的拉普拉斯矩阵;  $D_d = \sum_k S_{ik}^{d*}$  和  $D_t = \sum_l S_{jl}^{t*}$  是对角矩阵, 元素值分别为  $S^{d*}$  和  $S^{t*}$  的行之和。

该目标函数是一个多变量非凸函数, 没有封闭解, 采用迭代更新求解, 即先固定  $A$  求解  $B$ , 然后固定  $B$  求解  $A$ , 详细计算过程可参考文献[17], 迭代更

新见式(13)、式(14)。

$$a_{ik} = a_{ik} \frac{(YB + \beta S^{d*} A)_{ik}}{(AB^T B + \alpha A + \beta D_d A)_{ik}}$$

(13)

$$b_{jk} = b_{jk} \frac{(Y^T A + \gamma S^{t*} B)_{jk}}{(BA^T A + \alpha B + \gamma D_t B)_{jk}}$$

(14)

迭代收敛获得  $A$  和  $B$  后,通过  $Y^* = AB^T$  计算药物-靶蛋白作用关系预测得分矩阵,其中,  $Y^*$  中每一行中得分排名靠前的靶蛋白,其与该药物存在作用关系的概率值较大。

4 实验结果及讨论

考虑主要关注新药物作用靶蛋白、新靶蛋白作用药物两种预测场景,因此评价 MRNMFC-DT 算法性能时,在四个标准数据集中,主要与 BLM-NII<sup>[8]</sup>、RLS-WNN<sup>[9]</sup> 和 WKNKN + WGRMF<sup>[10]</sup> 算法进行比较,这三种算法适用于新药物和新靶蛋白作用关系预测。此外,在测试过程中,采用两类 10-CV: 10-CV<sub>d</sub> 和 10-CV<sub>t</sub>,其中,对新药物的作用靶蛋白预测,采用 10-CV<sub>d</sub>,将所有药物近似平均地分为 10 份,取其中一份为测试集,其余 9 份为训练集,对测试集中药物的所有作用关系都删除。同样地,对新靶蛋白的作用药物预测,采用 10-CV<sub>t</sub>,将所有靶蛋白近似平均地分为 10 份,其中一份为测试集,9 份为训练集,将测试集中靶蛋白的所有作用关系都删除。算法参数设置为  $K=20$ ,  $\mu=0.8$ ,  $\alpha=\beta=\gamma=1$ 。

考虑在药物-靶蛋白作用关系预测问题中,正样本数目较少,负样本数目较多,通过 PR (precision-recall) 曲线下面积(即 AUPR 值)来衡量。其中,精确率(Precision)和召回率(Recall)定义分别为: Precision = TP/(TP + FP) 和 Recall = TP/(TP + FN)。AUPR 性能可较好地惩罚负样本被判断为正样本<sup>[5]</sup>,两种场景的预测结果见表 2 和表 3。可见,提出的 MRNMFC-DT 具有较高的 AUPR 性能,证明本文算法可以较高精度用于新药物的作用靶蛋白预测和新靶蛋白的作用药物预测。

表 2 10-CV<sub>d</sub> 下, BLM-NII、RLS-WNN、WKNKN + WGRMF 和 MRNMFC-DT 算法的 AUPR 预测结果  
Table 2 AUPR results of BLM-NII, RLS-WNN, WKNKN + WGRMF and MRNMFC-DT in 10-CV<sub>d</sub> test

| 方法            | Es           | ICs          | GPCRs        | NRs          |
|---------------|--------------|--------------|--------------|--------------|
| BLM-NII       | 0.167(0.011) | 0.201(0.009) | 0.233(0.010) | 0.410(0.043) |
| RLS-WNN       | 0.386(0.014) | 0.319(0.011) | 0.363(0.008) | 0.519(0.025) |
| WKNKN + WGRMF | 0.401(0.013) | 0.369(0.017) | 0.401(0.012) | 0.528(0.033) |
| MRNMFC-DT     | 0.430(0.015) | 0.378(0.015) | 0.435(0.015) | 0.538(0.020) |

表 3 10-CV<sub>t</sub> 下, BLM-NII、RLS-WNN、WKNKN + WGRMF 和 MRNMFC-DT 算法的 AUPR 预测结果  
Table 3 AUPR results of BLM-NII, RLS-WNN, WKNKN + WGRMF and MRNMFC-DT in 10-CV<sub>t</sub> test

| 方法            | Es           | ICs          | GPCRs        | NRs          |
|---------------|--------------|--------------|--------------|--------------|
| BLM-NII       | 0.583(0.021) | 0.634(0.008) | 0.447(0.024) | 0.418(0.022) |
| RLS-WNN       | 0.761(0.015) | 0.746(0.004) | 0.547(0.025) | 0.468(0.030) |
| WKNKN + WGRMF | 0.798(0.018) | 0.799(0.007) | 0.585(0.027) | 0.446(0.015) |
| MRNMFC-DT     | 0.834(0.018) | 0.805(0.008) | 0.608(0.023) | 0.460(0.017) |

5 结论

在非负矩阵分解算法基础上,提出一种适用于药物-靶蛋白作用关系预测的融合社团聚类结果的流形正则化非负矩阵分解算法 MRNMFC-DT。得到主要结论如下:

- (1)引入更多异构源数据,基于药物-药物相互作用网络和蛋白作用关系网络分别对药物、靶蛋白相似性网络进行修正。
- (2)基于相似性网络分别进行药物、靶蛋白聚类分析,并利用模块特性构建新药物、新靶蛋白的初始标签。
- (3)提出的算法能以较高精度预测新药物的作用靶蛋白、新靶蛋白的作用药物,对药物研究具有重要指导意义。
- (4)考虑相似性空间与药物-靶蛋白映射空间的内在几何结构,引入流形学习算法,并作为正则项加入到非负矩阵分解算法的目标函数中进行组合优化,提高了算法预测性能。

参 考 文 献

1 Zhang S W, Yan X Y. Some remarks on prediction of drug-target interaction with network models [J]. Current Topics in Medicinal Chemistry, 2017, 17 (21): 2456-2468

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 Yan X Y, Zhang S W. Identifying drug-target interactions with decision template [J]. Current Protein & Peptide Science, 2018, 19 (5): 498-506

4 Bleakley K, Yamanishi Y. Supervised prediction of drug-target interactions using bipartite local models [J]. Bioinformatics, 2009, 25 (18): 2397-2403

5 Van Laarhoven T, Nabuurs S B, Marchiori E. Gaussian interaction profile kernels for predicting drug-target interaction [J]. Bioinformatics, 2011, 27(21): 3036-3043

6 闫效莺, 康 磊. 基于回归模型的网络排序算法预测药物靶点关系 [J]. 科学技术与工程, 2016, 16(22): 36-40

- Yan Xiaoying, Kang Lei. A regression-based network ranking prediction model for drug-target interaction [J]. Science Technology and Engineering, 2016, 16(22): 36-40
- 7 Yan X Y, Zhang S W, Zhang S Y. Prediction of drug-target interaction by label propagation with mutual interaction information derived from heterogeneous network [J]. Molecular Biosystems, 2016, 12(2): 520-531
- 8 Mei J P, Kwok C K, Yang P, et al. Drug-target interaction prediction by learning from local information and neighbors[J]. Bioinformatics, 2013, 29(2): 238-245
- 9 Van L T, Marchiori E. Predicting drug-target interactions for new drug compounds using a weighted nearest neighbor profile[J]. PloS one, 2013, 8(6): e66952
- 10 Ezzat A, Zhao P, Wu M, et al. Drug-target interaction prediction with graph regularized matrix factorization[J]. IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2017, 14(3): 646-656
- 11 Zhao S, Shao L. A co-module approach for elucidating drug-disease associations and revealing their molecular basis[J]. Bioinformatics, 2012, 28(7): 955-961
- 12 Luo Y, Zhao X, Zhou J, et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information [J]. Nature Communications, 2017, 8(1): 573-586
- 13 Knox C, Law V, Jewison T, et al. DrugBank 3.0: a comprehensive resource for ãomics research on drugs [J]. Nucleic Acids Research, 2011, 39(S1): D1035-D1041
- 14 Prasad T K, Goel R, Kandasamy K, et al. Human protein reference database—2009 update [J]. Nucleic Acids Research, 2009, 37(S1): D767-D772
- 15 Nepusz T, Yu H, Paccanaro A. Detecting overlapping protein complexes in protein-protein interaction networks[J]. Nature Methods, 2015, 9(5): 471-472
- 16 Cai D, He X, Han J, et al. Graph regularized nonnegative matrix factorization for data representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1548-1560
- 17 Xiao Q, Luo J, Liang C, et al. A graph regularized non-negative matrix factorization method for identifying microRNA-disease associations[J]. Bioinformatics, 2017, 34(2): 239-248
- 18 彭艳斌, 苏先创, 邱薇薇, 等. 基于流形正则化非负矩阵分解的高光谱数据降维[J]. 光电子·激光, 2018, 29(2): 224-228
- Peng Yanbin, Su Xianchuang, Qiu Weiwei, et al. Manifold regularized NMF based hyperspectral data dimensionality reduction [J]. Journal of Optoelectronics Laser, 2018, 29(2): 224-228

## Prediction of Drug-target Interaction with Manifold Regularized Non-negative Matrix Factorization

YAN Xiao-ying<sup>1,2</sup>, WU Ying<sup>1</sup>, LI Run-zhou<sup>1</sup>

(College of Computer Science, Xi'an Shiyou University<sup>1</sup>, Xi'an 710065, China; School of Automation, Northwestern Polytechnical University<sup>2</sup>, Xi'an 710072, China)

[**Abstract**] Identification of drug-target interactions (DTIs) is very important for drug research, which can help to find the new uses for old drugs or to discover the off-target of a given drug. Currently, the prediction algorithms have difficulty in finding interactions for new drugs and new targets. A novel method that uses manifold regularized nonnegative matrix factorization framework is proposed to predict potential targets/drugs for new drugs/targets. Firstly, it used clustering approaches to construct interaction profiles for new drugs/targets, then adopted the manifold regularized nonnegative matrix factorization algorithm to predict the drug-target interaction. Finally, extensively testing was applied on four datasets. Through comparison with other recently proposed BLM-NII, RLS-WNN and WKNKN + WGRMF, the algorithm attains high prediction performance in terms of AUPR.

[**Key words**] drug target clustering manifold learning regularized nonnegative matrix factorization