



(12)发明专利申请

(10)申请公布号 CN 110957002 A

(43)申请公布日 2020.04.03

(21)申请号 201911301346.4

(22)申请日 2019.12.17

(71)申请人 电子科技大学

地址 610000 四川省成都市高新区(西区)  
西源大道2006号

(72)发明人 刘勇国 李杨 李巧勤 杨尚明

(74)专利代理机构 成都行之专利代理事务所  
(普通合伙) 51220

代理人 李朝虎

(51) Int. Cl.

G16B 15/30(2019.01)

G16B 40/00(2019.01)

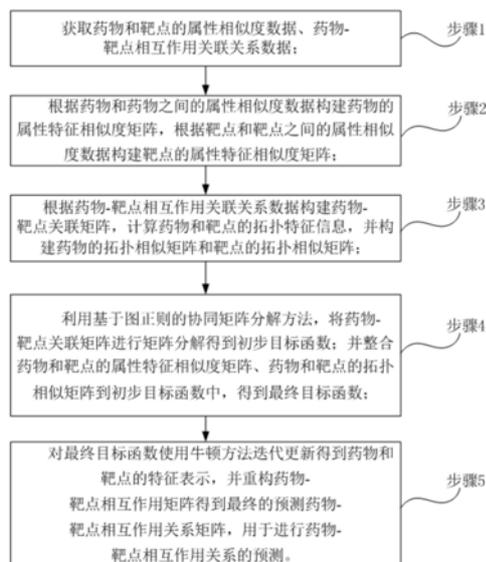
权利要求书3页 说明书12页 附图3页

(54)发明名称

一种基于协同矩阵分解的药物靶点相互作用关系预测方法

(57)摘要

本发明公开了一种基于协同矩阵分解的药物靶点相互作用关系预测方法,不仅考虑药物属性相似性和靶点属性相似性,同时结合药物拓扑结构相似性和靶点拓扑结构相似性,提高药物靶点相互作用关系预测准确率。包括:获取药物属性相似度和靶点属性相似度数据、药物靶点相互作用关系数据;构建药物属性特征相似度矩阵、靶点属性特征相似度矩阵、药物拓扑相似矩阵和靶点拓扑相似矩阵;对药物靶点关联矩阵进行基于图正则的协同矩阵分解,并整合药物和靶点的属性特征相似度和拓扑结构相似度信息,得到最终目标函数;用牛顿方法迭代更新得到药物和靶点的特征表示,重构药物-靶点相互作用关系矩阵,进行药物靶点相互作用关系的预测。



CN 110957002 A

1. 一种基于协同矩阵分解的药物靶点相互作用关系预测方法,其特征在于,该方法包括以下步骤:

步骤1: 获取药物和靶点的属性相似度数据、药物-靶点相互作用关联关系数据;

步骤2: 根据药物和药物之间的属性相似度数据构建药物的属性特征相似度矩阵,根据靶点和靶点之间的属性相似度数据构建靶点的属性特征相似度矩阵;

步骤3: 根据药物-靶点相互作用关联关系数据构建药物-靶点关联矩阵,计算药物和靶点的拓扑特征信息,并构建药物的拓扑相似矩阵和靶点的拓扑相似矩阵;

步骤4: 利用基于图正则的协同矩阵分解方法,将药物-靶点关联矩阵进行矩阵分解得到初步目标函数;并整合药物和靶点的属性特征相似度矩阵、药物和靶点的拓扑相似矩阵到初步目标函数中,得到最终目标函数;

步骤5: 对最终目标函数使用牛顿方法迭代更新得到药物和靶点的特征表示,并重构药物-靶点相互作用矩阵得到最终的预测药物-靶点相互作用关系矩阵,用于进行药物-靶点相互作用关系的预测。

2. 根据权利要求1所述的一种基于协同矩阵分解的药物靶点相互作用关系预测方法,其特征在于,步骤2中根据药物和药物之间的属性相似度数据构建药物的属性特征相似度矩阵,具体地,药物-药物属性相似度矩阵中的两个药物之间的相似度计算公式为:

$$Sim_{stru}(d_i, d_j) = \frac{|d_i \cap d_j|}{|d_i \cup d_j|} = \frac{|d_i \cap d_j|}{|d_i| + |d_j| - |d_i \cap d_j|}$$

式中,  $|d_i \cap d_j|$  表示两个药物  $d_i$  和  $d_j$  之间的最大公共子图中原子的数量,  $|d_i \cup d_j|$  表示  $d_i$  和  $d_j$  数量之和减去  $d_i$  和  $d_j$  之间的最大公共子图中原子的数量;

将  $N_d$  个药物之间的相似度全部计算出来,使用药物-药物属性相似度矩阵  $S_d \in R^{N_d * N_d}$  表示,  $S_d(i, j) \in [0, 1)$  表示第  $i$  个药物和第  $j$  个药物之间的结构属性相似度。

3. 根据权利要求1或2所述的一种基于协同矩阵分解的药物靶点相互作用关系预测方法,其特征在于,步骤2中根据靶点和靶点之间的属性相似度数据构建靶点的属性特征相似度矩阵,具体地,靶点-靶点属性相似度矩阵中的两个靶点之间的相似度计算公式为:

$$Sim_{seq}(t_i, t_j) = \frac{SW(t_i, t_j)}{\sqrt{SW(t_i, t_i)SW(t_j, t_j)}}$$

式中,  $t_i$  和  $t_j$  表示靶点列表中第  $i$  个和第  $j$  个靶点,  $SW(\cdot, \cdot)$  表示原始的 Smith-Waterman 分数;

将  $N_t$  个靶点之间的相似度全部计算出来,使用靶点-靶点属性相似度矩阵  $S_t \in R^{N_t * N_t}$  表示,  $S_t(i, j)$  表示第  $i$  个靶点和第  $j$  个靶点之间的序列属性相似度。

4. 根据权利要求1所述的一种基于协同矩阵分解的药物靶点相互作用关系预测方法,其特征在于,步骤3中药物和靶点的拓扑特征信息的提取均采用 node2vec 方法,具体地,药物拓扑特征计算包括:

(1) 已知 DTI 矩阵  $Y$ ,  $Y \in R^{N_d * N_t}$ ,  $N_d$  表示药物的总数,  $N_t$  表示靶点的总数,  $Y(i, j)$  表示第  $i$  个药物和第  $j$  个靶点之间的相互作用关系,当第  $i$  个药物和第  $j$  个靶点之间存在相互作用关系,  $Y(i, j) = 1$ ; 当第  $i$  个药物和第  $j$  个靶点之间不存在相互作用关系,  $Y(i, j) = 0$ ;

(2) 根据Y矩阵构建无权值无向的网络图 $G=(V,E)$ ,  $V$ 是节点集合, 其中 $|V|=N_d+N_t$ ,  $|V|$ 表示节点的数量;  $E$ 是边集合, 其中 $|E|=\sum_{i \in N_d} \sum_{j \in N_t} Y_{ij}$ ,  $|E|$ 表示边的数量; 当 $Y(i,j)=1$ 时, 存在边 $e_{ij}$ , 使得 $v_i$ 与 $v_j$ 相连; 为了获取药物的拓扑特征, 在网络图 $G$ 上进行二阶随机游走, 给定源节点 $u$ , 随机游走的固定长度 $l$ ,  $c_i$ 表示游走过程中第 $i$ 个节点, 游走的起点是 $c_0=u$ ; 节点 $c_i$ 由如下公式产生:

$$P(c_i = x | c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z} & \text{当}(v,x) \in E \text{时} \\ 0 & \text{其他} \end{cases}$$

式中,  $Z$ 是归一化常数,  $\pi_{vx}$ 是节点 $v$ 和 $x$ 之间的转移概率, 指的是节点 $v$ 下一步访问节点 $x$ 的概率;  $\pi_{vx}$ 的取值情况:

$$\pi_{vx}(v,x) = \begin{cases} \frac{1}{p} & \text{当}d_{vx} = 0 \text{时} \\ 1 & \text{当}d_{vx} = 1 \text{时} \\ \frac{1}{q} & \text{当}d_{vx} = 2 \text{时} \end{cases}$$

式中,  $d_{vx}$ 表示节点 $v$ 和节点 $x$ 之间的最短路径, 参数 $p$ 和 $q$ 用来控制随机游走的速度和距初始节点 $u$ 的距离;

(3) 在存在特征表达 $f$ 的条件下, 最大化发现节点 $u$ 的网络邻居 $N_s(u)$ 的 $\log$ 概率, node2vec的目标函数为:

$$\max_f \sum_{u \in V} \log Pr(N_s(u) | f(u))$$

用 $f:V \rightarrow R^d$ 表示节点到拓扑特征表达的映射函数,  $d$ 是拓扑特征空间的维度; 对于任意一个节点 $u \in V$ ,  $N_s(u) \subset V$ 表示节点 $u$ 的在网络图 $G$ 上的邻居节点,  $f(u)$ 表示节点 $u$ 的拓扑特征;

(4) 使用随机梯度下降优化node2vec的目标函数, 最后得到药物的 $d$ 维拓扑特征,  $N_d$ 个药物的 $d$ 维拓扑特征矩阵表示为 $P_d \in R^{N_d * d}$ 。

5. 根据权利要求4所述的一种基于协同矩阵分解的药物靶点相互作用关系预测方法, 其特征在于, 采用node2vec方法来计算靶点拓扑特征, 得到靶点的 $d$ 维拓扑特征,  $N_t$ 个靶点的 $d$ 维拓扑特征矩阵表示为 $P_t \in R^{N_t * d}$ 。

6. 根据权利要求1所述的一种基于协同矩阵分解的药物靶点相互作用关系预测方法, 其特征在于, 步骤3中构建药物和靶点的拓扑相似矩阵, 计算药物-药物拓扑特征相似性、靶点-靶点拓扑特征相似性; 具体地, 药物-药物拓扑特征相似性计算是结合药物的拓扑特征信息, 两个药物向量在拓扑特征空间下的相似度可通过余弦相似度表示; 给定两个药物 $d_i$ 和 $d_j$ 的拓扑特征向量表示 $x_i$ 和 $x_j$ , 使用余弦相似度计算药物-药物拓扑特征相似性:

$$Sim_{Drugtp}(d_i, d_j) = \frac{x_i x_j^T}{\|x_i\| \cdot \|x_j\|}$$

将 $N_d$ 个药物之间的相似度全部计算出来, 使用药物-药物拓扑特征相似度矩阵 $P_d \in R^{N_d * N_d}$ 表示,  $P_d(i,j)$ 表示第 $i$ 个药物和第 $j$ 个药物之间的拓扑特征相似度。

7. 根据权利要求1所述的一种基于协同矩阵分解的药物靶点相互作用关系预测方法, 其特征在于, 步骤4中的最终目标函数的公式表示如下:

$$\begin{aligned}
 L = & \sum_{i=1}^{N_d} \sum_{j=1}^{N_t} W_{ij} (Y_{ij} - a_i b_j^T)^2 + \lambda_l \left( \sum_{i=1}^{N_d} \|a_i\|^2 + \sum_{j=1}^{N_t} \|b_j\|^2 \right) + \lambda_d \sum_{i=1}^{N_d} \sum_{p=1}^{N_d} (S_d(d_i, d_p) - a_i a_p^T)^2 \\
 & + \lambda_t \sum_{j=1}^{N_t} \sum_{q=1}^{N_t} (S_t(t_j, t_q) - b_j b_q^T)^2 + \lambda_m \sum_{i=1}^{N_d} \sum_{p=1}^{N_d} \|a_i - a_p\|^2 P_d \\
 & + \lambda_n \sum_{j=1}^{N_t} \sum_{q=1}^{N_t} \|b_j - b_q\|^2 P_t
 \end{aligned}$$

式中,  $N_d$ 表示药物的数量,  $N_t$ 表示靶点的数量;  $W$ 是 $N_d$ 和 $N_t$ 的权重矩阵,  $W_{ij}=1$ 表示药物-靶点相互作用关系是已知的, 否则表示是未知的;  $Y$ 表示已知药物-靶点相互作用关系;  $\lambda_l$ 、 $\lambda_d$ 、 $\lambda_t$ 、 $\lambda_m$ 、 $\lambda_n$ 是正则化参数;  $a$ 和 $b$ 分别表矩阵分解得到的药物和靶点的特征向量;  $S_d$ 矩阵表示药物之间的属性特征相似度矩阵,  $S_t$ 矩阵表示靶点之间的属性特征相似度矩阵;  $P_d$ 矩阵描述了药物之间的结构相似关系,  $P_t$ 矩阵描述了靶点之间的结构相似关系。

8. 根据权利要求1所述的一种基于协同矩阵分解的药物靶点相互作用关系预测方法, 其特征在于, 步骤5中的重构药物-靶点相互作用矩阵得到的最终的预测药物-靶点相互作用关系矩阵 $F$ 为:

$$F = AB^T$$

式中,  $A$ 和 $B$ 为将已知药物-靶点相互作用关系矩阵分解成两个矩阵,  $A$ 和 $B$ 分别是 $N_d \times K$ ,  $N_t \times K$ ,  $K$ 是特征空间的维度。

## 一种基于协同矩阵分解的药物靶点相互作用关系预测方法

### 技术领域

[0001] 本发明涉及药物靶点相互作用关系预测技术领域,尤其涉及一种基于协同矩阵分解的药物靶点相互作用关系预测方法。

### 背景技术

[0002] 药物靶点是药物与人体作用的结合位点,包括基因、受体、酶、离子通道、转运体、核酸等,通过药物与位点的结合影响生物学事件的改变,从而实现药物的治疗效果。药物-靶点相互作用(drug-target interactions,DTI)的鉴定是现代药物发现和开发的基础。药物-靶点相互作用预测对药物发现、药物副作用预测、药物重定位,以及发现与现有药物相互作用的新靶点的过程有重要的作用。传统用于鉴定新的DTI的生物化学实验方法所需设备极其昂贵且耗费时间较长。随着公开的化学生物数据库的出现,近些年出现的药物靶点数据库有PhID、Therapeutic Target Database (TTD)、DrugBank、Binding DB、PharmGKB、ChEMBL等,这些数据资源为DTI的预测提供了基础。近年来,结合计算机技术可以快速低成本地识别潜在的DTI。

[0003] 目前的研究方法中,药物-靶点相互作用关系预测可以分为三种:基于配体的方法、基于分子对接的方法和基于化学基因组学的方法。基于配体的方法假设相似的药物具有相似的性质,也会结合相似的靶点蛋白,通过使用配体之间的相似性来预测新的DTI。基于分子对接方法通过利用药物和靶点的3D结构预测药物和靶标的相互作用,广泛用于生物学。化学基因组学是一门通过结合基因组学与化学资源来研究活性化合物与细胞内靶点的潜在关系,并应用于药物和靶点发现的新兴学科。基于化学基因组学的方法可以在短时间内进行大规模预测,为实验筛选提供候选药物或者靶点,根据采用的方法又可分为三类:基于分类的方法、基于网络扩散的方法和基于矩阵分解的方法。基于分类的方法一般使用药物-药物和靶点-靶点之间的相似性、药物和靶点的结构特征,利用机器学习的方法预测药物-靶点相互作用关系的有无;基于网络扩散的方法将药物和靶点看作节点,药物-靶点之间的关系看作边,增加药物-药物、靶点-靶点之间的相似性关系等构建网络,在网络上使用随机游走等网络传播方法预测未知的药物-靶点相互作用关系。

[0004] 基于矩阵分解的方法利用矩阵分解将药物靶点关联关系矩阵分解为两个低秩矩阵,对应于药物和靶点的特征空间。SPLCMF模型[L.Y.Xia,Z.Y.Yang,H.Zhang,et al.Improved Prediction of Drug-Target Interactions Using Self-Paced Learning with Collaborative Matrix Factorization[J].Journal of Chemical Information and Modeling,2019]将自步学习集成到基于协同过滤的矩阵分解中,整合药物相似度网络、靶点相似度网络和已知DTI网络到正则最小二乘中,进一步提高模型的预测能力,并且在存在重噪声和丢失数据的情况下可以有效地避免不良的局部最小值。

[0005] 上述的药物-靶点相互作用预测取得一定进展,但仍存在以下问题:1、基于配体的方法对于拥有少数已知配体的靶点效果不好;2、基于分子对接方法存在模拟过程耗时和许多靶点的3D结构未知的问题;3、目前存在的基于矩阵分解的方法只考虑到药物和靶点的属

性特征,没有考虑DTI网络中药物和靶点节点的网络潜在特征,即忽略了药物和靶点的拓扑特征,不能准确的表示药物和靶点的特征,从而导致预测结果不准确。

## 发明内容

[0006] 本发明所要解决的技术问题是现有的基于矩阵分解的方法药物和靶点的表示信息有限,并且没有考虑药物或者靶点的局部几何结构信息,导致很多情况下精确度不高的问题。本发明提供了解决上述问题的一种基于协同矩阵分解的药物靶点相互作用关系预测方法,提出图正则化约束来保持药物-靶点相互作用网络中的拓扑结构信息,使得在原始的数据空间中相似的药物或者靶点,在矩阵分解重构后也保持相似的特性,有效提高药物靶点相互作用关系预测精度。

[0007] 本发明通过下述技术方案实现:

[0008] 一种基于协同矩阵分解的药物靶点相互作用关系预测方法,该方法包括以下步骤:

[0009] 步骤1:获取药物和靶点的属性相似度数据、药物-靶点相互作用关联关系数据;

[0010] 步骤2:根据药物和药物之间的属性相似度数据构建药物的属性特征相似度矩阵,根据靶点和靶点之间的属性相似度数据构建靶点的属性特征相似度矩阵;

[0011] 步骤3:根据药物-靶点相互作用关联关系数据构建药物-靶点关联矩阵,计算药物和靶点的拓扑特征信息,并构建药物的拓扑相似矩阵和靶点的拓扑相似矩阵;

[0012] 步骤4:利用基于图正则的协同矩阵分解方法,将药物-靶点关联矩阵进行矩阵分解得到初步目标函数;并整合药物和靶点的属性特征相似度矩阵、药物和靶点的拓扑相似矩阵到初步目标函数中,得到最终目标函数;

[0013] 步骤5:对最终目标函数使用牛顿方法迭代更新得到药物和靶点的特征表示,并重构药物-靶点相互作用矩阵得到最终的预测药物-靶点相互作用关系矩阵,用于进行药物-靶点相互作用关系的预测。

[0014] 工作原理是:现有的基于矩阵分解的方法药物和靶点的表示信息有限,并且没有考虑药物或者靶点的局部几何结构信息,导致很多情况下精确度不高的问题。本发明方法不仅考虑药物-药物之间和靶点-靶点之间的属性相似性,同时结合药物-药物之间和靶点-靶点之间的拓扑结构相似性;本发明方法分别计算药物-药物之间和靶点-靶点之间的属性相似性和拓扑结构相似性,基于图正则的协同矩阵分解模型,不仅可以学习到药物和靶点的属性特征,还可以学习到DTI(即药物-靶点相互作用)网络中药物和靶点的拓扑特征,能够提高预测药物和靶点之间的相互作用关系的准确率。

[0015] 进一步地,步骤2中根据药物和药物之间的属性相似度数据构建药物的属性特征相似度矩阵,具体地,所述药物-药物属性相似度矩阵中的两个药物之间的相似度计算公式为:

$$[0016] \quad Sim_{Stru}(d_i, d_j) = \frac{|d_i \cap d_j|}{|d_i \cup d_j|} = \frac{|d_i \cap d_j|}{|d_i| + |d_j| - |d_i \cap d_j|}$$

[0017] 式中,  $|d_i \cap d_j|$  表示两个药物  $d_i$  和  $d_j$  之间的最大公共子图中原子的数量,  $|d_i \cup d_j|$  表示  $d_i$  和  $d_j$  数量之和减去  $d_i$  和  $d_j$  之间的最大公共子图中原子的数量,即  $d_i$  和  $d_j$  的并集;

[0018] 为了便于计算机输入,将  $N_d$  个药物之间的相似度全部计算出来,使用药物-药物属

性相似度矩阵  $S_d \in R^{N_d * N_d}$  表示,  $S_d(i, j) \in [0, 1)$  表示第  $i$  个药物和第  $j$  个药物之间的结构属性相似度, 其值越高, 表示两个药物之间的相似程度越大。

[0019] 进一步地, 步骤2中根据靶点和靶点之间的属性相似度数据构建靶点的属性特征相似度矩阵, 具体地, 靶点-靶点属性相似度矩阵中的两个靶点之间的相似度计算公式为:

$$[0020] \quad Sim_{Seq}(t_i, t_j) = \frac{SW(t_i, t_j)}{\sqrt{SW(t_i, t_i)SW(t_j, t_j)}}$$

[0021] 式中,  $t_i$  和  $t_j$  表示靶点列表中第  $i$  个和第  $j$  个靶点,  $SW(\cdot, \cdot)$  表示原始的Smith-Waterman分数;

[0022] 将  $N_t$  个靶点之间的相似度全部计算出来, 使用靶点-靶点属性相似度矩阵  $S_t \in R^{N_t * N_t}$  表示,  $S_t(i, j)$  表示第  $i$  个靶点和第  $j$  个靶点之间的序列属性相似度。

[0023] 进一步地, 由于目前仅利用药物的属性信息计算药物相似性, 忽略了药物在药物-靶点网络中的拓扑信息的相似性, 不能更准确的计算药物-药物的相似性。为了提取药物在药物-靶点相互作用网络中的拓扑特征, 步骤3中药物和靶点的拓扑特征信息的提取均采用 node2vec 方法, 具体地, 药物拓扑特征计算包括:

[0024] (1) 已知DTI矩阵  $Y, Y \in R^{N_d * N_t}$ ,  $N_d$  表示药物的总数,  $N_t$  表示靶点的总数,  $Y(i, j)$  表示第  $i$  个药物和第  $j$  个靶点之间的相互作用关系, 当第  $i$  个药物和第  $j$  个靶点之间存在相互作用关系,  $Y(i, j) = 1$ ; 当第  $i$  个药物和第  $j$  个靶点之间不存在相互作用关系,  $Y(i, j) = 0$ ;

[0025] (2) 根据  $Y$  矩阵构建无权值无向的网络图  $G = (V, E)$ ,  $V$  是节点集合, 其中  $|V| = N_d + N_t$ ,  $|V|$  表示节点的数量;  $E$  是边集合, 其中  $|E| = \sum_{i \in N_d} \sum_{j \in N_t} Y_{ij}$ ,  $|E|$  表示边的数量; 当  $Y(i, j) = 1$  时, 存在边  $e_{ij}$ , 使得  $v_i$  与  $v_j$  相连; 为了获取药物的拓扑特征, 在网络图  $G$  上进行二阶随机游走, 给定源节点  $u$ , 随机游走的固定长度  $l$ ,  $c_i$  表示游走过程中第  $i$  个节点, 游走的起点是  $c_0 = u$ ; 节点  $c_i$  由如下公式产生:

$$[0026] \quad P(c_i = x | c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z} & \text{当 } (v, x) \in E \text{ 时} \\ 0 & \text{其他} \end{cases}$$

[0027] 式中,  $Z$  是归一化常数,  $\pi_{vx}$  是节点  $v$  和  $x$  之间的转移概率, 指的是节点  $v$  下一步访问节点  $x$  的概率;  $\pi_{vx}$  的取值情况:

$$[0028] \quad \pi_{vx}(v, x) = \begin{cases} \frac{1}{p} & \text{当 } d_{vx} = 0 \text{ 时} \\ 1 & \text{当 } d_{vx} = 1 \text{ 时} \\ \frac{1}{q} & \text{当 } d_{vx} = 2 \text{ 时} \end{cases}$$

[0029] 式中,  $d_{vx}$  表示节点  $v$  和节点  $x$  之间的最短路径, 参数  $p$  和  $q$  用来控制随机游走的速度和距初始节点  $u$  的距离;

[0030] (3) 目标函数表达的含义是, 在存在特征表达  $f$  的条件下, 最大化发现节点  $u$  的网络邻居  $N_s(u)$  的  $\log$  概率, node2vec 的目标函数为:

$$[0031] \quad \max_f \sum_{u \in V} \log Pr(N_s(u)|f(u))$$

[0032] 用  $f: V \rightarrow R^d$  表示节点到拓扑特征表达的映射函数,  $d$  是拓扑特征空间的维度; 对于任意一个节点  $u \in V$ ,  $N_s(u) \subset V$  表示节点  $u$  的在网络图  $G$  上的邻居节点,  $f(u)$  表示节点  $u$  的拓扑特征;

[0033] (4) 使用随机梯度下降优化 node2vec 的目标函数, 最后得到药物的  $d$  维拓扑特征,  $N_d$  个药物的  $d$  维拓扑特征矩阵表示为  $P_d \in R^{N_d * d}$ 。

[0034] 进一步地, 和计算药物的拓扑特征相同, 采用 node2vec 方法来计算靶点拓扑特征, 得到靶点的  $d$  维拓扑特征,  $N_t$  个靶点的  $d$  维拓扑特征矩阵表示为  $P_t \in R^{N_t * d}$ 。

[0035] 进一步地, 步骤3中构建药物和靶点的拓扑相似矩阵, 计算药物-药物拓扑特征相似性、靶点-靶点拓扑特征相似性; 具体地, 药物-药物拓扑特征相似性计算是结合药物的拓扑特征信息, 两个药物向量在拓扑特征空间下的相似度可通过余弦相似度表示; 给定两个药物  $d_i$  和  $d_j$  的拓扑特征向量表示  $x_i$  和  $x_j$ , 使用余弦相似度计算药物-药物拓扑特征相似性:

$$[0036] \quad Sim_{Drugtp}(d_i, d_j) = \frac{x_i x_j^T}{\|x_i\| \cdot \|x_j\|}$$

[0037] 将  $N_d$  个药物之间的相似度全部计算出来, 使用药物-药物拓扑特征相似度矩阵  $P_d \in R^{N_d * N_d}$  表示,  $P_d(i, j)$  表示第  $i$  个药物和第  $j$  个药物之间的拓扑特征相似度。

[0038] 同时: 和计算药物-药物拓扑特征相似性相同, 最后得到靶点-靶点的拓扑特征相似度矩阵  $P_t \in R^{N_t * N_t}$ 。

[0039] 进一步地, 步骤4详细步骤如下:

[0040] 为了充分利用药物-药物和靶点-靶点的相似信息, 得到药物和靶点的潜在特征表示, 进行协同矩阵分解, 将已知药物-靶点相互作用关系矩阵  $Y$  分解成两个低秩矩阵  $A$  和  $B$ , 分别对应药物和靶点的特征空间。用  $A$  和  $B$  的内积近似药物-靶点相互作用关系矩阵  $Y$ :

$$[0041] \quad Y \approx AB^T$$

[0042] 式中,  $A$  和  $B$  分别是  $N_d \times K$ ,  $N_t \times K$ ,  $K$  是特征空间的维度。

[0043] 基于已知的药物-靶点相互作用关联关系, 使用最小化重构误差来估计矩阵分解中的低秩矩阵  $A$  和  $B$ , 矩阵分解初步目标函数为:

$$[0044] \quad \begin{aligned} & \arg \min \|W \odot (Y - AB^T)\|_F^2 + \lambda_l (\|A\|_F^2 + \|B\|_F^2) \\ & = \sum_{i=1}^{N_d} \sum_{j=1}^{N_t} W_{ij} (Y_{ij} - a_i b_j^T)^2 + \lambda_l \left( \sum_{i=1}^{N_d} \|a_i\|^2 + \sum_{j=1}^{N_t} \|b_j\|^2 \right) \end{aligned}$$

[0045] 式中,  $\|\cdot\|_F^2$  表示 Frobenius 范数。 $W$  是  $N_d$  和  $N_t$  的权重矩阵,  $W_{ij} = 1$  表示药物-靶点相互作用关系是已知的, 否则表示是未知的。 $\odot$  表示元素积, 为了区分已知药物-靶点对和未知药物-靶点对, 对  $W$  和  $(Y - AB^T)$  求元素积。 $\lambda_l$  是正则化参数, 防止训练数据过拟合。

[0046] 整合药物-药物属性特征相似度、靶点-靶点属性特征相似度到目标函数中, 得到:

$$\begin{aligned}
& \arg \min \|W \odot (Y - AB^T)\|_F^2 + \lambda_l (\|A\|_F^2 + \|B\|_F^2) + \lambda_d \|S_d - AA^T\|_F^2 + \lambda_t \|S_t - BB^T\|_F^2 \\
[0047] \quad & = \sum_{i=1}^{N_d} \sum_{j=1}^{N_t} W_{ij} (Y_{ij} - a_i b_j^T)^2 + \lambda_l \left( \sum_{i=1}^{N_d} \|a_i\|^2 + \sum_{j=1}^{N_t} \|b_j\|^2 \right) \\
& \quad + \lambda_d \sum_{i=1}^{N_d} \sum_{p=1}^{N_d} (S_d(d_i, d_p) - a_i a_p^T)^2 + \lambda_t \sum_{j=1}^{N_t} \sum_{q=1}^{N_t} (S_t(t_j, t_q) - b_j b_q^T)^2
\end{aligned}$$

[0048] CMF (协同矩阵分解) 的原理是让已知药物-靶点相互作用关系Y矩阵近似于A和B的内积。 $\lambda_1$ 、 $\lambda_d$ 、 $\lambda_t$ 、 $\lambda_m$ 、 $\lambda_n$ 是正则化参数。属性特征相似度矩阵 $S_d$ 和 $S_t$ 的矩阵分解,其过程公式表示如下:

$$[0049] \quad S_d \approx AA^T$$

$$[0050] \quad S_t \approx BB^T$$

[0051]  $P_d$ 矩阵描述了药物之间的结构相似关系,为了整合药物信息,在低维空间中保留药物图形结构,引入基于药物图形拓扑的正则化术语:

$$[0052] \quad \lambda_m \sum_{i=1}^{N_d} \sum_{p=1}^{N_d} \|a_i - a_p\|^2 P_d$$

[0053] 同理,  $P_t$ 矩阵描述了靶点之间的结构相似关系,引入基于靶点图形拓扑的正则化术语:

$$[0054] \quad \lambda_n \sum_{j=1}^{N_t} \sum_{q=1}^{N_t} \|b_j - b_q\|^2 P_t$$

[0055] 最终目标函数为:

$$\begin{aligned}
[0056] \quad L = & \sum_{i=1}^{N_d} \sum_{j=1}^{N_t} W_{ij} (Y_{ij} - a_i b_j^T)^2 + \lambda_l \left( \sum_{i=1}^{N_d} \|a_i\|^2 + \sum_{j=1}^{N_t} \|b_j\|^2 \right) + \lambda_d \sum_{i=1}^{N_d} \sum_{p=1}^{N_d} (S_d(d_i, d_p) - a_i a_p^T)^2 \\
& + \lambda_t \sum_{j=1}^{N_t} \sum_{q=1}^{N_t} (S_t(t_j, t_q) - b_j b_q^T)^2 + \lambda_m \sum_{i=1}^{N_d} \sum_{p=1}^{N_d} \|a_i - a_p\|^2 P_d \\
& + \lambda_n \sum_{j=1}^{N_t} \sum_{q=1}^{N_t} \|b_j - b_q\|^2 P_t
\end{aligned}$$

[0057] 式中,  $N_d$ 表示药物的数量,  $N_t$ 表示靶点的数量;  $W$ 是 $N_d$ 和 $N_t$ 的权重矩阵,  $W_{ij}=1$ 表示药物-靶点相互作用关系是已知的, 否则表示是未知的;  $Y$ 表示已知药物-靶点相互作用关系;  $\lambda_1$ 、 $\lambda_d$ 、 $\lambda_t$ 、 $\lambda_m$ 、 $\lambda_n$ 是正则化参数;  $a$ 和 $b$ 分别表矩阵分解得到的药物和靶点的特征向量;  $S_d$ 矩阵表示药物之间的属性特征相似度矩阵,  $S_t$ 矩阵表示靶点之间的属性特征相似度矩阵;  $P_d$ 矩阵描述了药物之间的结构相似关系,  $P_t$ 矩阵描述了靶点之间的结构相似关系。

[0058] 进一步地, 步骤5中对最终目标函数使用牛顿方法迭代更新得到药物和靶点的特征表示, 具体包括:

[0059] 由于药物潜在特征向量 $a_i$ 和靶点潜在特征向量 $b_j$ 耦合在一起, 不容易求解, 因此使用牛顿方法更新 $a_i$ 和 $b_j$ 。

[0060] 首先,计算L相对于 $a_i$ 和 $b_j$ 的偏导数:

$$[0061] \quad \nabla_{a_i} L = a_i \left( \sum_{j=1}^{N_t} W_{ij} b_j^T b_j + \lambda_l I_k + \lambda_d \sum_{p=1}^{N_d} a_p^T a_p + \lambda_m \left( \sum_{i=1}^{N_d} P_{dij} + \sum_{i=1}^{N_d} P_{dji} \right) I_k \right) \\ - \sum_{j=1}^{N_t} W_{ij} Y_{ij} b_j - \lambda_d \sum_{p=1}^{N_d} S_d(d_i, d_p) a_p - \lambda_m \sum_{i=1}^{N_d} (P_{dij} + P_{dji}) a_p$$

$$[0062] \quad \nabla_{b_j} L = b_j \left( \sum_{i=1}^{N_d} W_{ij} a_i^T a_i + \lambda_l I_k + \lambda_t \sum_{q=1}^{N_t} b_q^T b_q + \lambda_n \left( \sum_{j=1}^{N_t} P_{tij} + \sum_{j=1}^{N_t} P_{tji} \right) I_k \right) \\ - \sum_{i=1}^{N_d} W_{ij} Y_{ij} a_i - \lambda_t \sum_{q=1}^{N_t} S_t(t_j, t_q) b_q - \lambda_n \sum_{j=1}^{N_t} (P_{tij} + P_{tji}) b_q$$

[0063] 然后,计算L相对于 $a_i$ 和 $b_j$ 的二阶导数:

$$[0064] \quad \nabla_{a_i}^2 L = \sum_{j=1}^{N_t} W_{ij} b_j^T b_j + \lambda_l I_k + \lambda_d \sum_{p=1}^{N_d} a_p^T a_p + \lambda_m \left( \sum_{i=1}^{N_d} P_{dij} + \sum_{i=1}^{N_d} P_{dji} \right) I_k$$

$$[0065] \quad \nabla_{b_j}^2 L = \sum_{i=1}^{N_d} W_{ij} a_i^T a_i + \lambda_l I_k + \lambda_t \sum_{q=1}^{N_t} b_q^T b_q + \lambda_n \left( \sum_{j=1}^{N_t} P_{tij} + \sum_{j=1}^{N_t} P_{tji} \right) I_k$$

[0066] 由于矩阵 $\nabla_{a_i}^2 L$ 和 $\nabla_{b_j}^2 L$ 是对称和正定的,这意味着 $\nabla_{a_i}^2 L$ 和 $\nabla_{b_j}^2 L$ 是可逆的。因此,通过使用偏导数和二阶导数,可以根据牛顿方法更新 $a_i$ 和 $b_j$ :

$$[0067] \quad a_i \leftarrow a_i - \nabla_{a_i} L (\nabla_{a_i}^2 L)^{-1}$$

$$[0068] \quad b_j \leftarrow b_j - \nabla_{b_j} L (\nabla_{b_j}^2 L)^{-1}$$

[0069] 进一步地,步骤5中的DTI矩阵(即药物-靶点相互作用矩阵)得到的最终的预测药物-靶点相互作用关系矩阵F为:

$$[0070] \quad F = AB^T$$

[0071] 式中,A和B为将已知药物-靶点相互作用关系矩阵分解成两个矩阵,A和B分别是 $N_d \times K$ , $N_t \times K$ ,K是特征空间的维度。

[0072] 本发明与现有技术相比,具有如下的优点和有益效果:

[0073] 1、本发明一种基于协同矩阵分解的药物靶点相互作用关系预测方法,本发明给出了一种药物-靶点预测的框架,该方法不仅考虑了药物和靶点的属性特征,还考虑了药物和靶点的拓扑结构特征,通过提出的正则项对相关性进行约束;

[0074] 2、本发明一种基于协同矩阵分解的药物靶点相互作用关系预测方法,本发明基于图正则协同矩阵分解模型,提出图正则化约束来保持药物-靶点相互作用网络中的拓扑结构信息,从而提高预测药物-靶点相互作用关系的精确度。

## 附图说明

[0075] 此处所说明的附图用来提供对本发明实施例的进一步理解,构成本申请的一部分,并不构成对本发明实施例的限定。在附图中:

[0076] 图1为本发明一种基于协同矩阵分解的药物靶点相互作用关系预测方法流程图。

[0077] 图2为本发明方法的基于图正则的协同矩阵分解模型图。

[0078] 图3为本发明方法的基于图正则的协同矩阵分解过程图。

[0079] 图4为本发明方法的属性特征相似度矩阵分解过程图。

## 具体实施方式

[0080] 为使本发明的目的、技术方案和优点更加清楚明白,下面结合实施例和附图,对本发明作进一步的详细说明,本发明的示意性实施方式及其说明仅用于解释本发明,并不作为对本发明的限定。

[0081] 实施例1

[0082] 如图1至图4所示,本发明一种基于协同矩阵分解的药物靶点相互作用关系预测方法,该方法包括以下步骤:

[0083] 步骤1:获取药物和靶点的属性相似度数据、药物-靶点相互作用关联关系数据;

[0084] 步骤2:根据药物和药物之间的属性相似度数据构建药物的属性特征相似度矩阵,根据靶点和靶点之间的属性相似度数据构建靶点的属性特征相似度矩阵;

[0085] 步骤3:根据药物-靶点相互作用关联关系数据构建药物-靶点关联矩阵,计算药物和靶点的拓扑特征信息,并构建药物的拓扑相似矩阵和靶点的拓扑相似矩阵;

[0086] 步骤4:利用基于图正则的协同矩阵分解方法,将药物-靶点关联矩阵进行矩阵分解得到初步目标函数;并整合药物和靶点的属性特征相似度矩阵、药物和靶点的拓扑相似矩阵到初步目标函数中,得到最终目标函数;

[0087] 步骤5:对最终目标函数使用牛顿方法迭代更新得到药物和靶点的特征表示,并重构药物-靶点相互作用矩阵得到最终的预测药物-靶点相互作用关系矩阵,用于进行药物-靶点相互作用关系的预测。

[0088] 本发明技术方案使用的数据来自于Yamanishi的数据集[Yamanishi, Y.; Araki, M.; Gutteridge, A.; Honda, W.; Kanehisa, M. Prediction of Drug-Target Interaction Networks from the Integration of Chemical and Genomic Spaces. *Bioinformatics* 2008, 24, i232-i240.],其中根据靶点类型分为核受体(NR)、G蛋白偶联受体(GPCR)、离子通道(IC)和酶(E)四个数据集。NR数据集包括54个药物、26个靶点和90条药物-靶点相互作用关联关系;GPCR数据集包括223个药物、95个靶点和635条药物-靶点相互作用关联关系;IC数据集包括210个药物、204个靶点和1476条药物-靶点相互作用关联关系;E数据集包括445个药物、664个靶点和2926条药物-靶点相互作用关联关系;

[0089] 本发明技术方案的主流程图如图1所示,主要包括以下步骤:

[0090] (1) 计算药物-药物属性相似性

[0091] 从Yamanishi数据集中获取药物-药物属性相似度数据,Yamanishi数据集中药物-药物属性相似度数据使用SIMCOMP工具计算两个药物的相似性。SIMCOMP工具提供基于公共子结构的全局相似性分数计算,其原理是:

[0092] 给定两个药物 $d_i$ 和 $d_j$ ,分别表示药物列表中第 $i$ 和第 $j$ 个药物的化学结构,由原子作为顶点和共价键作为边缘的2D图形。首先从Yamanishi的数据集获得药物列表,即药物的名称列表;然后,在KEGG LIGAND数据库中查询药物的化合物编号;最后,在SIMCOMP工具中输入其中一个药物的化合物编号,该工具根据该化合物的2D图形计算与其他药物的最大公共子结构,最后按相似程度由高到低排列药物。

[0093] 具体相似度计算方法是使用Jaccard相似度进行计算两个药物的相似度:

$$[0094] \quad Sim_{Stru}(d_i, d_j) = \frac{|d_i \cap d_j|}{|d_i \cup d_j|} = \frac{|d_i \cap d_j|}{|d_i| + |d_j| - |d_i \cap d_j|}$$

[0095] 式中, $|d_i \cap d_j|$ 表示 $d_i$ 和 $d_j$ 之间的最大公共子图中原子的数量, $|d_i \cup d_j|$ 表示 $d_i$ 和 $d_j$ 数量之和减去 $d_i$ 和 $d_j$ 之间的最大公共子图中原子的数量,即 $d_i$ 和 $d_j$ 的并集。为了便于计算机输入,将 $N_d$ 个药物之间的相似度全部计算出来,使用药物-药物属性相似度矩阵 $S_d \in R^{N_d * N_d}$ 表示, $S_d(i, j) \in [0, 1]$ 表示第 $i$ 个药物和第 $j$ 个药物之间的结构属性相似度,其值越高,表示两个药物之间的相似程度越大。

[0096] (2) 计算靶点-靶点属性相似性

[0097] 从Yamanishi数据集中获取靶点-靶点属性相似度数据,由于原始的Smith-Waterman分数受到靶点序列长度的影响,导致原始的Smith-Waterman分数不能准确衡量靶点之间的属性相似度,因此使用归一化Smith-Waterman分数计算两个靶点之间的相似度。归一化Smith-Waterman分数的计算如下:

$$[0098] \quad Sim_{Seq}(t_i, t_j) = \frac{SW(t_i, t_j)}{\sqrt{SW(t_i, t_i)SW(t_j, t_j)}}$$

[0099] 式中, $t_i$ 和 $t_j$ 表示靶点列表中第 $i$ 个和第 $j$ 个靶点, $SW(\cdot, \cdot)$ 表示原始的Smith-Waterman分数。为了便于计算机输入,将 $N_t$ 个靶点之间的相似度全部计算出来,使用靶点-靶点属性相似度矩阵 $S_t \in R^{N_t * N_t}$ 表示, $S_t(i, j)$ 表示第 $i$ 个靶点和第 $j$ 个靶点之间的序列属性相似度。

[0100] (3) 计算药物拓扑特征

[0101] 由于目前仅利用药物的属性信息计算药物相似性,忽略了药物在药物-靶点网络中的拓扑信息的相似性,不能更准确的计算药物-药物的相似性。为了提取药物在药物-靶点相互作用网络中的拓扑特征,本发明使用node2vec方法。

[0102] 已知DTI矩阵 $Y, Y \in R^{N_d * N_t}, N_d$ 表示药物的总数, $N_t$ 表示靶点的总数。 $Y(i, j)$ 表示第 $i$ 个药物和第 $j$ 个靶点之间的相互作用关系。当第 $i$ 个药物和第 $j$ 个靶点之间存在相互作用关系, $Y(i, j) = 1$ ;当第 $i$ 个药物和第 $j$ 个靶点之间不存在相互作用关系, $Y(i, j) = 0$ 。

[0103] 根据 $Y$ 矩阵构建无权值无向的网络图 $G = (V, E), V$ 是节点集合,其中 $|V| = N_d + N_t, |V|$ 表示节点的数量。 $E$ 是边集合,其中 $|E| = \sum_{i \in N_d} \sum_{j \in N_t} Y_{ij}, |E|$ 表示边的数量。当 $Y(i, j) = 1$ 时,存在边 $e_{ij}$ ,使得 $v_i$ 与 $v_j$ 相连。

[0104] 为了获取药物的拓扑特征,在网络图 $G$ 上进行二阶随机游走,给定源节点 $u$ ,随机游走的固定长度 $l, c_i$ 表示游走过程中第 $i$ 个节点,游走的起点是 $c_0 = u$ 。节点 $c_i$ 由如下公式产生:

$$[0105] \quad P(c_i = x | c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z} & \text{当}(v, x) \in E \text{时} \\ 0 & \text{其他} \end{cases}$$

[0106] 式中,  $Z$  是归一化常数,  $\pi_{vx}$  是节点  $v$  和  $x$  之间的转移概率, 指的是节点  $v$  下一步访问节点  $x$  的概率。 $\pi_{vx}$  的取值情况:

$$[0107] \quad \pi_{vx}(v, x) = \begin{cases} \frac{1}{p} & \text{当} d_{vx} = 0 \text{ 时} \\ 1 & \text{当} d_{vx} = 1 \text{ 时} \\ \frac{1}{q} & \text{当} d_{vx} = 2 \text{ 时} \end{cases}$$

[0108] 式中,  $d_{vx}$  表示节点  $v$  和节点  $x$  之间的最短路径。参数  $p$  和  $q$  用来控制随机游走的速度和距初始节点  $u$  的距离。

[0109] node2vec 的目标函数为:

$$[0110] \quad \max_f \sum_{u \in V} \log Pr(N_s(u) | f(u))$$

[0111] 用  $f: V \rightarrow \mathbb{R}^d$  表示节点到拓扑特征表达的映射函数,  $d$  是拓扑特征空间的维度。对于任意一个节点  $u \in V$ ,  $N_s(u) \subset V$  表示节点  $u$  的在网络图  $G$  上的邻居节点,  $f(u)$  表示节点  $u$  的拓扑特征。目标函数表达的含义是, 在存在特征表达  $f$  的条件下, 最大化发现节点  $u$  的网络邻居  $N_s(u)$  的  $\log$  概率。使用随机梯度下降优化 node2vec 的目标函数。

[0112] 最后得到药物的  $d$  维拓扑特征,  $N_d$  个药物的  $d$  维拓扑特征矩阵表示为  $P_d \in \mathbb{R}^{N_d * d}$ 。

[0113] (4) 计算靶点拓扑特征

[0114] 和计算药物的拓扑特征相同, 最后得到靶点的  $d$  维拓扑特征,  $N_t$  个靶点的  $d$  维拓扑特征矩阵表示为  $P_t \in \mathbb{R}^{N_t * d}$ 。

[0115] (5) 计算药物-药物拓扑特征相似性

[0116] 结合药物的拓扑特征信息, 两个药物向量在拓扑特征空间下的相似性可通过余弦相似性表示。给定两个药物  $d_i$  和  $d_j$  的拓扑特征向量表示  $x_i$  和  $x_j$ , 使用余弦相似性计算药物-药物拓扑特征相似性:

$$[0117] \quad Sim_{Drugtp}(d_i, d_j) = \frac{x_i x_j^T}{\|x_i\| \cdot \|x_j\|}$$

[0118] 将  $N_d$  个药物之间的相似性全部计算出来, 使用药物-药物拓扑特征相似性矩阵  $P_d \in \mathbb{R}^{N_d * N_d}$  表示,  $P_d(i, j)$  表示第  $i$  个药物和第  $j$  个药物之间的拓扑特征相似性。

[0119] (6) 计算靶点-靶点拓扑特征相似性

[0120] 和计算药物-药物拓扑特征相似性相同, 最后得到靶点-靶点的拓扑特征相似性矩阵  $P_t \in \mathbb{R}^{N_t * N_t}$ 。

[0121] (7) 基于图正则的协同矩阵分解

[0122] 为了充分利用药物-药物和靶点-靶点的相似信息, 得到药物和靶点的潜在特征表示, 进行协同矩阵分解, 将已知药物-靶点相互作用关系矩阵  $Y$  分解成两个低秩矩阵  $A$  和  $B$ , 分

别对应药物和靶点的特征空间。如图2所示,用A和B的内积近似药物-靶点相互作用关系矩阵Y:

$$[0123] \quad Y \approx AB^T$$

[0124] 式中,A和B分别是 $N_d \times K, N_t \times K, K$ 是特征空间的维度。

[0125] 基于已知的药物-靶点相互作用关联关系,使用最小化重构误差来估计矩阵分解中的低秩矩阵A和B,矩阵分解初步目标函数为:

$$[0126] \quad \arg \min \|W \odot (Y - AB^T)\|_F^2 + \lambda_l (\|A\|_F^2 + \|B\|_F^2) \\ = \sum_{i=1}^{N_d} \sum_{j=1}^{N_t} W_{ij} (Y_{ij} - a_i b_j^T)^2 + \lambda_l \left( \sum_{i=1}^{N_d} \|a_i\|^2 + \sum_{j=1}^{N_t} \|b_j\|^2 \right)$$

[0127] 式中, $\|\cdot\|_F^2$ 表示Frobenius范数。 $W$ 是 $N_d$ 和 $N_t$ 的权重矩阵, $W_{ij}=1$ 表示药物-靶点相互作用关系是已知的,否则表示是未知的。 $\odot$ 表示元素积,为了区分已知药物-靶点对和未知药物-靶点对,对 $W$ 和 $(Y-AB^T)$ 求元素积。 $\lambda_l$ 是正则化参数,防止训练数据过拟合。

[0128] 整合药物-药物属性特征相似度、靶点-靶点属性特征相似度到目标函数中,得到:

$$[0129] \quad \arg \min \|W \odot (Y - AB^T)\|_F^2 + \lambda_l (\|A\|_F^2 + \|B\|_F^2) + \lambda_d \|S_d - AA^T\|_F^2 + \lambda_t \|S_t - BB^T\|_F^2 \\ = \sum_{i=1}^{N_d} \sum_{j=1}^{N_t} W_{ij} (Y_{ij} - a_i b_j^T)^2 + \lambda_l \left( \sum_{i=1}^{N_d} \|a_i\|^2 + \sum_{j=1}^{N_t} \|b_j\|^2 \right) \\ + \lambda_d \sum_{i=1}^{N_d} \sum_{p=1}^{N_d} (S_d(d_i, d_p) - a_i a_p^T)^2 + \lambda_t \sum_{j=1}^{N_t} \sum_{q=1}^{N_t} (S_t(t_j, t_q) - b_j b_q^T)^2$$

[0130] CMF(协同矩阵分解)的原理是让已知药物-靶点相互作用关系Y矩阵近似于A和B的内积。 $\lambda_l, \lambda_d, \lambda_t, \lambda_m, \lambda_n$ 是正则化参数。属性特征相似度矩阵 $S_d$ 和 $S_t$ 的矩阵分解如图4所示,其过程公式表示如下:

$$[0131] \quad S_d \approx AA^T$$

$$[0132] \quad S_t \approx BB^T$$

[0133]  $P_d$ 矩阵描述了药物之间的结构相似关系,为了整合药物信息,在低维空间中保留药物图形结构,引入基于药物图形拓扑的正则化术语:

$$[0134] \quad \lambda_m \sum_{i=1}^{N_d} \sum_{p=1}^{N_d} \|a_i - a_p\|^2 P_d$$

[0135] 同理, $P_t$ 矩阵描述了靶点之间的结构相似关系,引入基于靶点图形拓扑的正则化术语:

$$[0136] \quad \lambda_n \sum_{j=1}^{N_t} \sum_{q=1}^{N_t} \|b_j - b_q\|^2 P_t$$

[0137] 最终目标函数为:

$$\begin{aligned}
L = & \sum_{i=1}^{N_d} \sum_{j=1}^{N_t} W_{ij} (Y_{ij} - a_i b_j^T)^2 + \lambda_l \left( \sum_{i=1}^{N_d} \|a_i\|^2 + \sum_{j=1}^{N_t} \|b_j\|^2 \right) + \lambda_d \sum_{i=1}^{N_d} \sum_{p=1}^{N_d} (S_d(d_i, d_p) - a_i a_p^T)^2 \\
[0138] \quad & + \lambda_t \sum_{j=1}^{N_t} \sum_{q=1}^{N_t} (S_t(t_j, t_q) - b_j b_q^T)^2 + \lambda_m \sum_{i=1}^{N_d} \sum_{p=1}^{N_d} \|a_i - a_p\|^2 P_d \\
& + \lambda_n \sum_{j=1}^{N_t} \sum_{q=1}^{N_t} \|b_j - b_q\|^2 P_t
\end{aligned}$$

[0139] (8) 优化目标函数

[0140] 由于药物潜在特征向量 $a_i$ 和靶点潜在特征向量 $b_j$ 耦合在一起,不容易求解,因此使用牛顿方法更新 $a_i$ 和 $b_j$ 。

[0141] 首先,计算L相对于 $a_i$ 和 $b_j$ 的偏导数:

$$\begin{aligned}
\nabla_{a_i} L = & a_i \left( \sum_{j=1}^{N_t} W_{ij} b_j^T b_j + \lambda_l I_k + \lambda_d \sum_{p=1}^{N_d} a_p^T a_p + \lambda_m \left( \sum_{i=1}^{N_d} P_{dij} + \sum_{i=1}^{N_d} P_{dji} \right) I_k \right) \\
& - \sum_{j=1}^{N_t} W_{ij} Y_{ij} b_j - \lambda_d \sum_{p=1}^{N_d} S_d(d_i, d_p) a_p - \lambda_m \sum_{i=1}^{N_d} (P_{dij} + P_{dji}) a_p \\
[0142] \quad \nabla_{b_j} L = & b_j \left( \sum_{i=1}^{N_d} W_{ij} a_i^T a_i + \lambda_l I_k + \lambda_t \sum_{q=1}^{N_t} b_q^T b_q + \lambda_n \left( \sum_{j=1}^{N_t} P_{tij} + \sum_{j=1}^{N_t} P_{tji} \right) I_k \right) \\
& - \sum_{i=1}^{N_d} W_{ij} Y_{ij} a_i - \lambda_t \sum_{q=1}^{N_t} S_t(t_j, t_q) b_q - \lambda_n \sum_{j=1}^{N_t} (P_{tij} + P_{tji}) b_q
\end{aligned}$$

[0143] 然后,计算L相对于 $a_i$ 和 $b_j$ 的二阶导数:

$$\nabla_{a_i}^2 L = \sum_{j=1}^{N_t} W_{ij} b_j^T b_j + \lambda_l I_k + \lambda_d \sum_{p=1}^{N_d} a_p^T a_p + \lambda_m \left( \sum_{i=1}^{N_d} P_{dij} + \sum_{i=1}^{N_d} P_{dji} \right) I_k$$

$$\nabla_{b_j}^2 L = \sum_{i=1}^{N_d} W_{ij} a_i^T a_i + \lambda_l I_k + \lambda_t \sum_{q=1}^{N_t} b_q^T b_q + \lambda_n \left( \sum_{j=1}^{N_t} P_{tij} + \sum_{j=1}^{N_t} P_{tji} \right) I_k$$

[0146] 由于矩阵 $\nabla_{a_i}^2 L$ 和 $\nabla_{b_j}^2 L$ 是对称和正定的,这意味着 $\nabla_{a_i}^2 L$ 和 $\nabla_{b_j}^2 L$ 是可逆的。因此,通过使用偏导数和二阶导数,可以根据牛顿方法更新 $a_i$ 和 $b_j$ :

$$[0147] \quad a_i \leftarrow a_i - \nabla_{a_i} L (\nabla_{a_i}^2 L)^{-1}$$

$$[0148] \quad b_j \leftarrow b_j - \nabla_{b_j} L (\nabla_{b_j}^2 L)^{-1}$$

[0149] (9) 药物-靶点相互作用关系预测

[0150] 重构DTI矩阵得到最终的预测药物-靶点相互作用关系矩阵:

$$[0151] \quad F = AB^T$$

[0152] 根据以上步骤,实施本发明方法,本发明方法不仅考虑药物-药物之间和靶点-靶点之间的属性相似性,同时结合药物-药物之间和靶点-靶点之间的拓扑结构相似性;本发明方法分别计算药物-药物之间和靶点-靶点之间的属性相似性和拓扑结构相似性,基于图正则的协同矩阵分解模型,考虑将网络的拓扑结构作为正则化项融入协同矩阵分解过程中,不仅可以学习到药物和靶点的属性特征,还可以学习到DTI(即药物-靶点相互作用)网络中药物和靶点的拓扑特征,充分利用了节点属性和网络拓扑互为补充的优势,能够提高预测药物和靶点之间的相互作用关系的准确率。

[0153] 现有背景技术方案通常只使用药物和靶点属性相似度矩阵进行药物靶点相互作用关系预测,这种方法可以获得节点的潜在空间表示,但由于没有考虑网络的拓扑结构信息,影响预测精度。根据“guilt by association”假设,两种类似的药物往往具有共同的靶点,或者说两个具有类似拓扑结构的节点,往往体现着相似的功能。因此本发明将网络的拓扑结构也融入到模型中。这样,即使两个药物的属性不同,但如果它们在网络中共享相似的连接,它们也可能拥有共同的靶点;反之,药物和靶点之间即使缺乏连接,但如果两个药物有相似的属性,则也可能拥有共同的靶点。双方信息互补,从而提高药物和靶点之间的相互作用关系预测精度。

[0154] 以上所述的具体实施方式,对本发明的目的、技术方案和有益效果进行了进一步详细说明,所应理解的是,以上所述仅为本发明的具体实施方式而已,并不用于限定本发明的保护范围,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

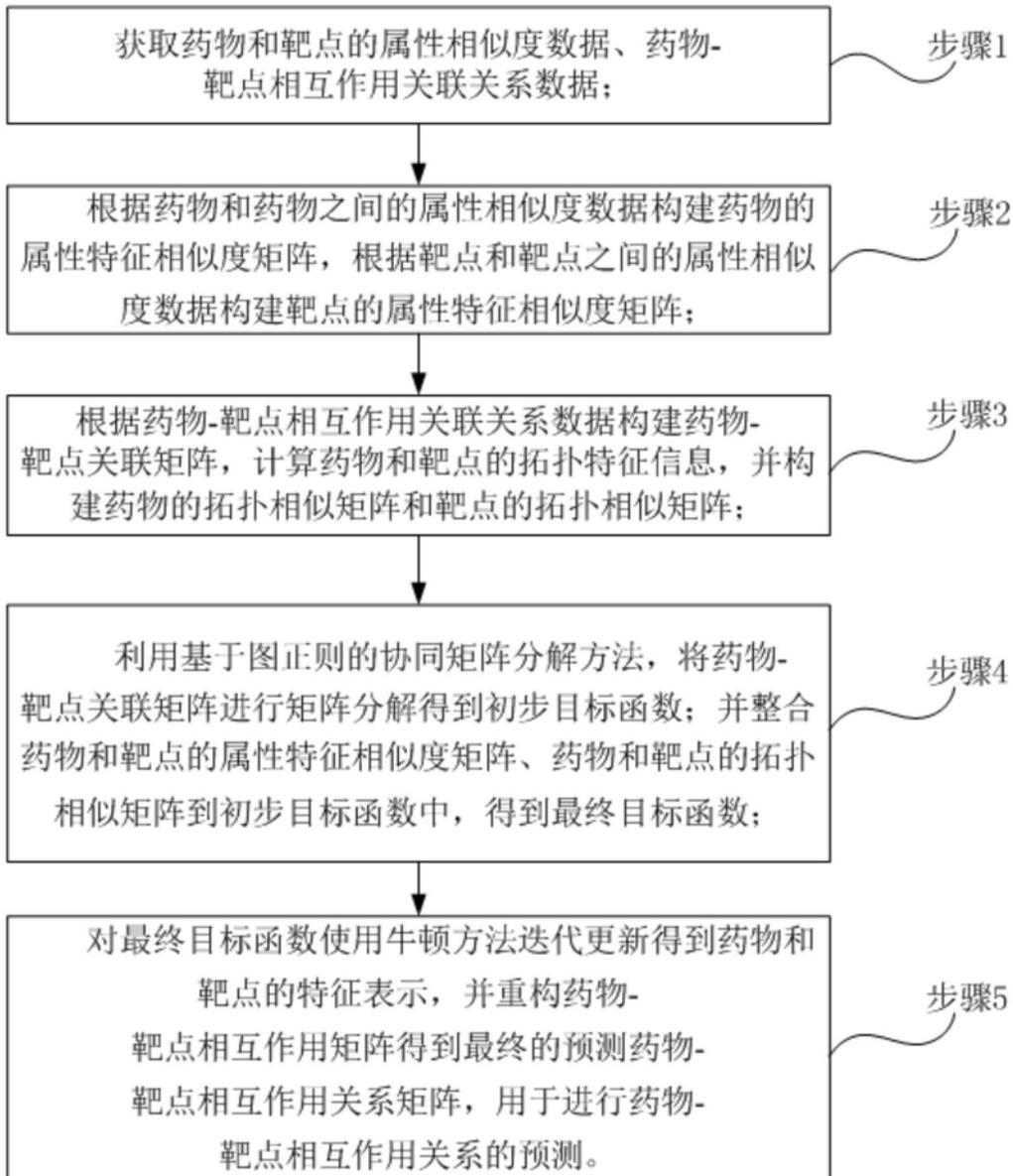


图1

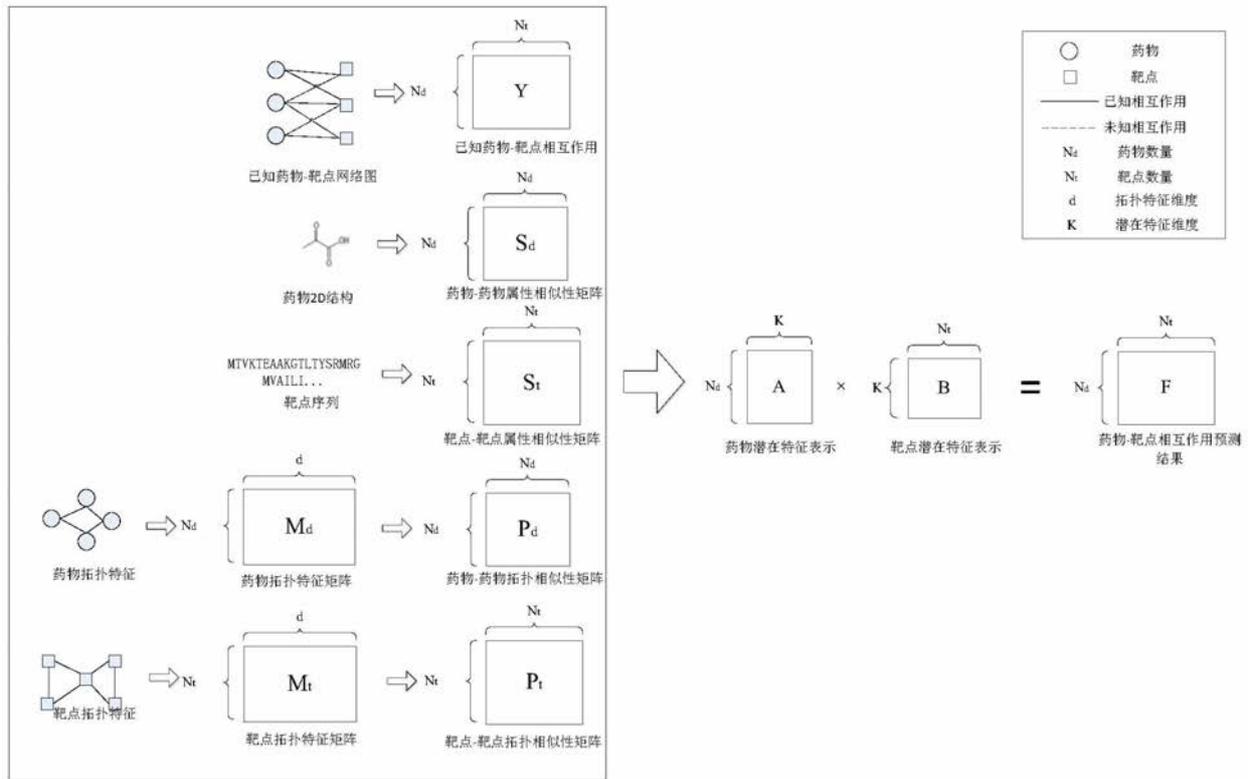


图2

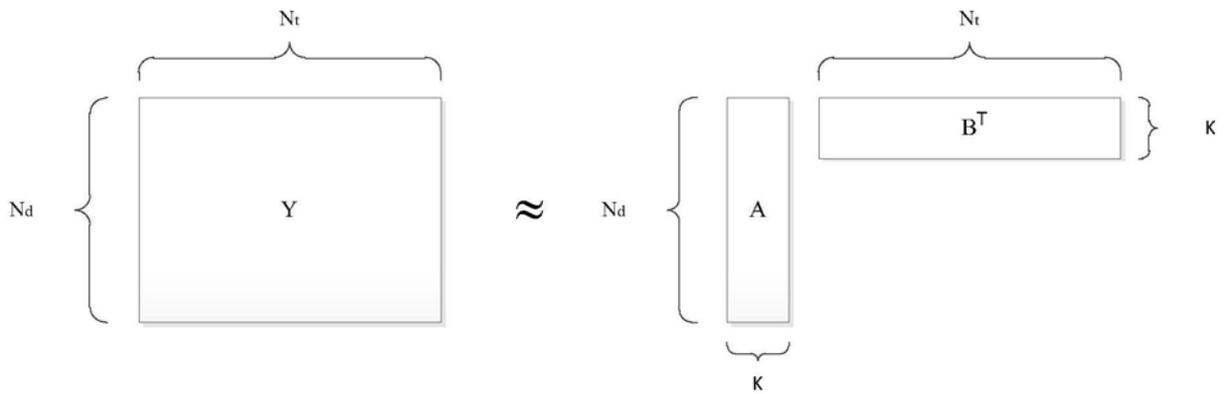


图3

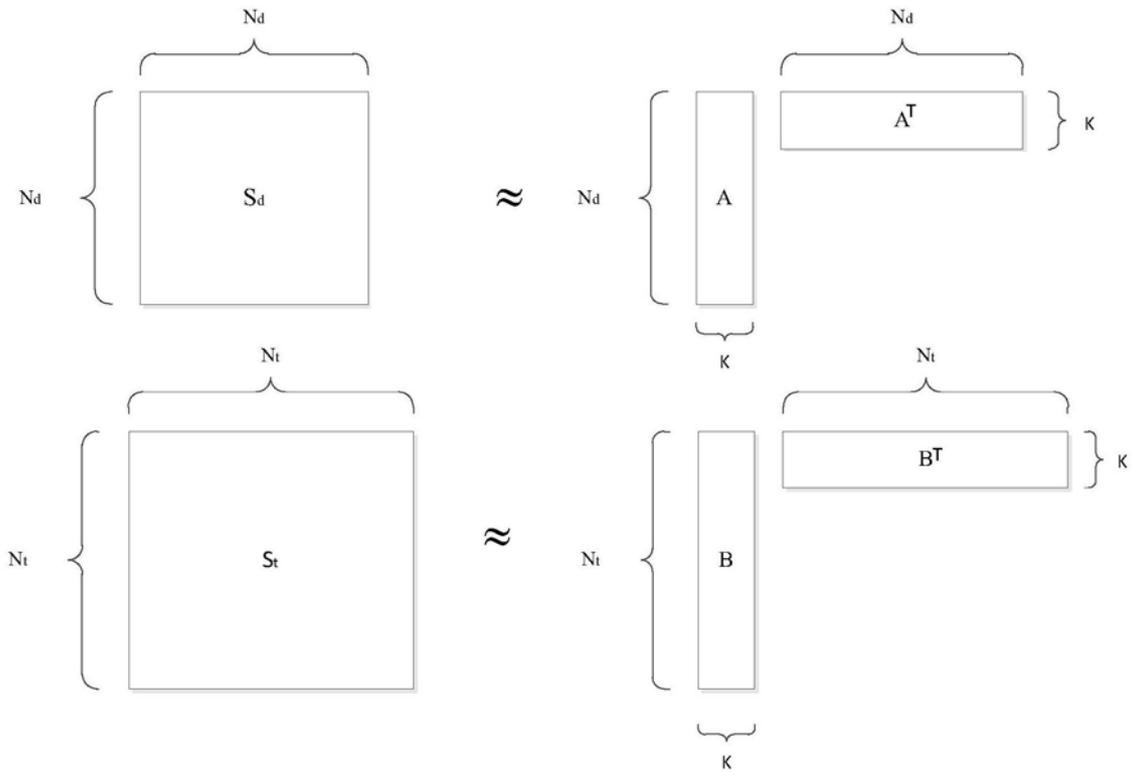


图4