(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 107506591 A
(43)申请公布日 2017.12.22

(21)申请号 201710748221.0

(22)申请日 2017.08.28

(71)申请人 中南大学
地址 410083 湖南省长沙市岳麓区麓山南路932号

(72)发明人 王建新 罗慧敏 李敏 蒋辉 卢诚谦

(74)专利代理机构 长沙市融智专利事务所 43114
代理人 杨萍

(51)Int.Cl.
*G06F 19/00*(2011.01)

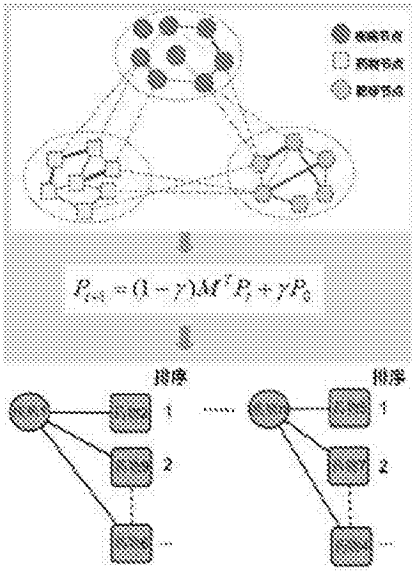
权利要求书4页 说明书10页 附图6页

(54)发明名称

一种基于多元信息融合和随机游走模型的药物重定位方法

(57)摘要

本发明公开了一种基于多元信息融合和随机游走模型的药物重定位方法。通过集成已有的疾病数据、药物数据、靶标数据、疾病-药物关联数据、疾病-基因关联数据和药物-靶标关联数据，构建疾病-靶标-药物异构网络。扩展基本的随机游走模型到所构建的异构网络上，通过有效的利用全局网络信息，为疾病推荐候选治疗药物。本发明简单有效，通过与其他方法比较，及在标准数据集上测试表明，该发明在药物重定位方面具有较好的预测性能。




CN 107506591 A

1.一种基于多元信息融合和随机游走模型的药物重定位方法，其特征在于，包括以下步骤：

1)构建疾病-靶标-药物异构网络：利用已知的疾病数据、药物数据、靶标数据、疾病-药物关联数据、疾病-靶标关联数据和药物-靶标关联数据，构建疾病网络、药物网络、靶标网络、疾病-药物关联网络、疾病-靶标关联网络和药物-靶标关联网络；通过关联网络连接疾病网络、药物网络和靶标网络，得到疾病-靶标-药物异构网络；

2)扩展基本随机游走模型到该异构网络：首先根据已知的疾病-药物关联数据和疾病-靶标关联数据构建随机游走的初始概率矩阵；然后利用已知的药物相似性、疾病相似性、靶标相似性、疾病-药物关联数据、疾病-靶标关联数据和药物-靶标关联数据，构建随机游走的转移矩阵；

3)预测新的药物-疾病关联：对于给定的疾病，依据所构建的初始概率矩阵和转移矩阵，迭代地在异构网络中进行随机游走，执行直到游走结果达到收敛状态；根据游走结果，得到给定疾病与所有药物存在关联的概率值，概率值越大，表明疾病与药物之间存在关联的可能性越大；按照概率值的大小，把与给定疾病不存在已知关联的所有药物进行排序，从而为给定疾病推荐新的治疗药物。

2.根据权利要求1所述的基于多元信息融合和随机游走模型的药物重定位方法，所述步骤1)包括以下步骤：

1.1)基于疾病的表型信息，计算疾病之间的相似性值，构建疾病网络；在疾病网络中，顶点集合$D=\{d_1,d_2,\cdots,d_n\}$表示n种疾病，顶点$d_i$和顶点$d_j$之间有边相连接，疾病i和疾病j之间的相似性值即为该条边的权值；

1.2)基于药物的化学结构信息，计算药物之间的相似性值，构建药物网络；在药物网络中，顶点集合$R=\{r_1,r_2,\cdots,r_m\}$表示m种药物，顶点$r_i$和顶点$r_j$之间有边相连接，药物i和药物j之间的相似性值即为该条边的权值；

1.3)基于靶标的序列信息，计算靶标之间的相似性值，构建靶标网络；在靶标网络中，顶点集合$T=\{t_1,t_2,\cdots,t_p\}$表示p种靶标，顶点$t_i$和顶点$t_j$之间有边相连接，靶标i和靶标j之间的相似性值即为该条边的权值；

1.4)基于已知的疾病-药物关联数据，构建疾病-药物关联网络；将疾病-药物关联网络建模为一个二分图$G_{dr}(D,R,E)$，其中$E(G)\subseteq R\times D$，$E(G)=\{e_{ij},d_i与r_j之间的边\}$，如果疾病$d_i$与药物$r_j$之间存在已知关联，则$d_i$与$r_j$之间的边权重设置为1，否则设置为0；

1.5)基于已知的疾病-靶标关联数据，构建疾病-靶标关联网络；将疾病-靶标关联网络建模为一个二分图$G_{dt}(D,T,E)$，其中$E(G)\subseteq D\times T$，$E(G)=\{e_{ij},d_i与t_j之间的边\}$，如果疾病$d_i$与靶标$t_j$之间存在已知关联，则$d_i$与$t_j$之间的边权重设置为1，否则设置为0；

1.6)基于已知的药物-靶标关联数据，构建药物-靶标关联网络；将药物-靶标关联网络建模为一个二分图$G_{rt}(R,T,E)$，其中$E(G)\subseteq R\times T$，$E(G)=\{e_{ij},r_i与t_j之间的边\}$，如果药物$r_i$与靶标$t_j$之间存在已知关联，则$r_i$与$t_j$之间的边权重设置为1，否则设置为0；

1.7)构建疾病-靶标-药物异构网络，该网络包括疾病网络、药物网络、靶标网络、疾病-药物关联网络、疾病-靶标关联网络和药物-靶标关联网络，其中疾病网络、药物网络和靶标网络通过对应的关联网络连接。

3.根据权利要求2所述的基于多元信息融合和随机游走模型的药物重定位方法，所述

步骤2)包括以下步骤：

第一步：构建初始概率矩阵$P_0$；

对于给定疾病为d，预测d的候选药物，则将给定疾病d作为疾病网络中的种子节点，将与给定疾病d存在已知关联的所有药物节点作为药物网络中的种子节点，将与给定疾病d存在已知关联的所有靶标节点作为靶标网络中的种子节点；根据这三个网络中的种子节点定义，将异构网络的初始概率矩阵$P_0$表示为：

$$P_0=\begin{bmatrix}\lambda_R Pr_0\\ \lambda_T Pt_0\\ (1-\lambda_R-\lambda_T)Pd_0\end{bmatrix} \quad (1)$$

其中，$Pr_0$、$Pt_0$和$Pd_0$分别表示药物网络、靶标网络和疾病网络的初始概率向量；$Pr_0$包含m个元素，分别对应m个药物的初始概率；如果第j个药物与给定疾病d存在关联，则$Pr_0$中的第j个元素值为1/(与给定疾病d存在关联的药物个数)，否则$Pr_0$中的第j个元素值为0；$Pt_0$包含p个元素，分别对应p个靶标的初始概率；如果第j个靶标与给定疾病d存在关联，则$Pt_0$中的第j个元素值为1/(与给定疾病d存在关联的靶标个数)，否则$Pt_0$中的第j个元素值为0；$Pd_0$包含n个元素，分别对应n个疾病的初始概率；$Pd_0$中与给定疾病d相应的元素的元素值为1，其他元素值为0；

参数$\lambda_R$，$\lambda_T$和$1-\lambda_R-\lambda_T$对应药物网络、靶标网络和疾病网络的重要性，$\lambda_R$，$\lambda_T$，$1-\lambda_R-\lambda_T\in[0,1]$，通过交叉验证实验选择最优参数值；

第二步：构建转移概率矩阵M：

$$M=\begin{bmatrix}M_{RR} & M_{RT} & M_{RD}\\ M_{TR} & M_{TT} & M_{TD}\\ M_{DR} & M_{DT} & M_{DD}\end{bmatrix} \quad (2)$$

其中，$M_{RR}$是药物网络的网内转移矩阵，包括任一药物节点到其它药物节点的转移概率；$M_{TT}$是靶标网络的网内转移矩阵，包括任一靶标节点到其它靶标节点的转移概率；$M_{DD}$是疾病网络的网内转移矩阵，包括任一疾病节点到其它疾病节点的转移概率；$M_{RD}$是药物网络和疾病网络的网间转移矩阵，包括药物节点到疾病节点的转移概率；$M_{RT}$是药物网络和靶标网络的网间转移矩阵，包括药物节点到靶标节点的转移概率；$M_{DR}$是疾病网络和药物网络的网间转移矩阵，包括疾病节点到药物节点的转移概率；$M_{DT}$是疾病网络和靶标网络的网间转移矩阵，包括疾病节点到靶标节点的转移概率；$M_{TR}$是靶标网络和药物网络的网间转移矩阵，包括靶标节点到药物节点的转移概率；$M_{TD}$是靶标网络和疾病网络的网间转移矩阵，包括靶标节点到疾病节点的转移概率；各个网内转移矩阵和网间转移矩阵中的元素计算方法如下：

$$M_{DD}(i,j)=\begin{cases}A_{DD}(i,j)/\sum_j A_{DD}(i,j) & if\ \sum_j A_{DR}(i,j)=0,\ \sum_j A_{DT}(i,j)=0\\ (1-\lambda_{DR})A_{DD}(i,j)/\sum_j A_{DD}(i,j) & if\ \sum_j A_{DR}(i,j)\neq 0,\ \sum_j A_{DT}(i,j)=0\\ (1-\lambda_{DT})A_{DD}(i,j)/\sum_j A_{DD}(i,j) & if\ \sum_j A_{DR}(i,j)=0,\ \sum_j A_{DT}(i,j)\neq 0\\ (1-\lambda_{DR}-\lambda_{DT})A_{DD}(i,j)/\sum_j A_{DD}(i,j) & if\ \sum_j A_{DR}(i,j)\neq 0,\ \sum_j A_{DT}(i,j)\neq 0\end{cases} \quad (3)$$

$$M_{RR}(i,j)=\begin{cases}A_{RR}(i,j)/\sum_j A_{RR}(i,j) & if\ \sum_j A_{RD}(i,j)=0,\ \sum_j A_{RT}(i,j)=0\\ (1-\lambda_{RD})A_{RR}(i,j)/\sum_j A_{RR}(i,j) & if\ \sum_j A_{RD}(i,j)\neq 0,\ \sum_j A_{RT}(i,j)=0\\ (1-\lambda_{RT})A_{RR}(i,j)/\sum_j A_{RR}(i,j) & if\ \sum_j A_{RD}(i,j)=0,\ \sum_j A_{RT}(i,j)\neq 0\\ (1-\lambda_{RD}-\lambda_{RT})A_{RR}(i,j)/\sum_j A_{RR}(i,j) & if\ \sum_j A_{RD}(i,j)\neq 0,\ \sum_j A_{RT}(i,j)\neq 0\end{cases} \quad (4)$$

$$M_{TT}(i,j)=\begin{cases}A_{TT}(i,j)/\sum_j A_{TT}(i,j) & if\ \sum_j A_{TD}(i,j)=0,\ \sum_j A_{TR}(i,j)=0\\ (1-\lambda_{TD})A_{TT}(i,j)/\sum_j A_{TT}(i,j) & if\ \sum_j A_{TD}(i,j)\neq 0,\ \sum_j A_{TR}(i,j)=0\\ (1-\lambda_{TR})A_{TT}(i,j)/\sum_j A_{TT}(i,j) & if\ \sum_j A_{TD}(i,j)=0,\ \sum_j A_{TR}(i,j)\neq 0\\ (1-\lambda_{TD}-\lambda_{TR})A_{TT}(i,j)/\sum_j A_{TT}(i,j) & if\ \sum_j A_{TD}(i,j)\neq 0,\ \sum_j A_{TR}(i,j)\neq 0\end{cases} \quad (5)$$

$$M_{DR}(i,j)=\begin{cases}\lambda_{DR}A_{DR}(i,j)/\sum_j A_{DR}(i,j) & if\ \sum_j A_{DR}(i,j)\neq 0\\ 0 & otherwise\end{cases} \quad (6)$$

$$M_{RD}(i,j)=\begin{cases}\lambda_{RD}A_{RD}(i,j)/\sum_j A_{RD}(i,j) & if\ \sum_j A_{RD}(i,j)\neq 0\\ 0 & otherwise\end{cases} \quad (7)$$

$$M_{RT}(i,j)=\begin{cases}\lambda_{RT}A_{RT}(i,j)/\sum_j A_{RT}(i,j) & if\ \sum_j A_{RT}(i,j)\neq 0\\ 0 & otherwise\end{cases} \quad (8)$$

$$M_{DT}(i,j)=\begin{cases}\lambda_{DT}A_{DT}(i,j)/\sum_j A_{DT}(i,j) & if\ \sum_j A_{DT}(i,j)\neq 0\\ 0 & otherwise\end{cases} \quad (9)$$

$$M_{TR}(i,j)=\begin{cases}\lambda_{TR}A_{TR}(i,j)/\sum_j A_{TR}(i,j) & if\ \sum_j A_{TR}(i,j)\neq 0\\ 0 & otherwise\end{cases} \quad (10)$$

$$M_{TD}(i,j)=\begin{cases}\lambda_{TD}A_{TD}(i,j)/\sum_j A_{TD}(i,j) & if\ \sum_j A_{TD}(i,j)\neq 0\\ 0 & otherwise\end{cases} \quad (11)$$

其中，参数$\lambda_{DR}$表示从疾病网络到药物网络的跳转概率，$\lambda_{RD}$表示从药物网络到疾病网络的跳转概率，$\lambda_{DT}$表示从疾病网络到靶标网络的跳转概率，$\lambda_{TD}$表示从靶标网络到疾病网络的跳转概率，$\lambda_{RT}$表示从药物网络到靶标网络的跳转概率，$\lambda_{TR}$表示从靶标网络到药物网络的跳转概率；其中A为疾病-靶标-药物异构网络的邻接矩阵：

$$A=\begin{bmatrix}A_{RR} & A_{RT} & A_{RD}\\ A_{RT}^{T} & A_{TT} & A_{TD}\\ A_{RD}^{T} & A_{TD}^{T} & A_{DD}\end{bmatrix} \quad (12)$$

其中，A的主对角线上的三个子矩阵$A_{RR}$、$A_{TT}$、$A_{DD}$对应的是药物网络、靶标网络和疾病网络的邻接矩阵；$A_{RT}$、$A_{RD}$、$A_{TD}$对应的是药物-靶标网络、药物-疾病网络和靶标-疾病网络的邻接矩阵，$A_{RT}^{T}$、$A_{RD}^{T}$、$A_{TD}^{T}$分别是$A_{RT}$、$A_{RD}$、$A_{TD}$的转置矩阵。

4.根据权利要求3所述的基于多元信息融合和随机游走模型的药物重定位方法，所述步骤3)对于给定疾病d，预测其候选药物包括以下步骤：

基于所构建的疾病-靶标-药物异构网络，以及在第一步和第二步分别定义的初始概率矩阵$P_0$和转移概率矩阵M，迭代地在异构网络中进行随机游走；

迭代到第t+1步时的概率矩阵$P_{t+1}$为：

$$P_{t+1}=(1-\gamma)M^{T}P_{t}+\gamma P_{0} \quad (13)$$

其中，$\gamma$为重启概率，取值范围为[0,1]；

当$P_{t+1}$与$P_t$之间的差别小于某个很小的阈值时(比如$10^{-10}$)，认为游走达到稳定状态，结束迭代；

将最终的概率矩阵记为P，P中的每个元素表示游走者到达相应节点的最终概率；最终的概率矩阵P包含三部分：Pr，Pt和Pd；其中Pr中的第i个元素表示疾病d与药物$r_i$之间存在关联的概率；Pt中的第i个元素表示疾病d与靶标$t_i$之间存在关联的概率；Pd中的第i个元素表示疾病d与疾病$d_i$之间存在关联的概率；

如果药物$r_i$与疾病d之间不存在已知关联，则药物$r_i$称为疾病d的候选药物；根据Pr中存放的所有候选药物与疾病d之间存在关联的概率值大小为给定疾病推荐候选药物。

# 一种基于多元信息融合和随机游走模型的药物重定位方法

## 技术领域

[0001] 本发明涉及生物信息学领域，具体涉及一种基于多元信息融合和随机游走模型的药物重定位方法，为疾病推荐候选治疗药物。

## 背景技术

[0002] 当前，尽管在药物研发中的投资不断增长，但是每年被美国食品药品监督管理局FDA(Food and Drug Administration)批准上市的新药数量很少。新药研发依然是一个周期较长、耗资巨大，而且存在较高的风险和较低的成功率。统计表明，一个新药从研发到上市，大约需要15年的时间，花费超过8亿美元。目前，很多制药公司试图通过计算机分子辅助设计、高通量筛选、组合化学等创新技术来提高开发新药的速率，但销售额仍远远不及新药研究和开发所需费用。此外，新药研发过程中，大多数候选药物分子因不能通过早期实验和毒性评估而终止，这是药物研发成本高、时间长的原因。数据表明，从临床I期到最后通过批准上市的总成功率仅为9.6％，10个进入临床的药物，仅有1个能最终上市。

[0003] 针对这个问题，药物重定位(Drug Repositioning Or Drug Repurposing)技术正成为药物研发的重要策略。药物重定位，又称之为“老药新用”，“开发药物的新疗效”，指的是利用相关的技术方法对已有药物进行筛选、组合或改造，从而发掘已有药物新适应症的过程。由于开展重定位研究的药物通常已通过了临床试验的几个阶段或已上市，因此这些药物的新用途更容易获得药品监管部门的批准，可以大大降低药物研发成本、缩短研发周期，不仅能够为病人提供帮助，也具有更高的投入产出效率，能够为药企带来可观的经济效益。近年来，政府机关、学术机构和医药企业在药物重定向研究方面的投入日益增大。

[0004] 随着高通量筛选、基因组测序等技术的发展，已经搜集了大量药物以及疾病方面的相关数据，为药物重定位的研究和发展奠定了基础。目前针对药物重定位的方法主要分为基于机器学习、基于网络、基于文本挖掘和语义推理三大类别。其中，基于网络的药物重定位方法随着各种生物数据(如基因组学、药物基因组学、临床数据等)的不断积累而日益受到关注。例如，Chiang和Butte根据关联推定的原理，假设两个疾病共享相似的治疗，那么用于治疗其中一种疾病的药物也可能治疗另一种疾病。在此基础上，提出了一种新的药物重定位方法。Wang等人构建了加权的疾病-药物，应用图聚类算法识别关联紧密的疾病和药物模块，然后将每个模块内的疾病-药物关联作为对应疾病的候选药物。陈等人将引入社交网络领域中的推荐模型思想，把药物看作用户，疾病看作商品，并假设结构相似的药物可能治疗相似的疾病，进而提出一种面向药物重定位的推荐模型。基于所构建的药物-疾病二分图预测潜在的药物疾病关联关系。Luo等人提出了一种基于集成的相似性度量和双向随机游走的药物重定位方法。在计算药物相似性、疾病相似性时，除了分别利用药物特性信息与疾病特性信息，还充分考虑了当前数据集中已知药物-疾病关联信息对相似性度量的作用，使得所计算的相似性值能更好的反映药物间的相似度和疾病间的相似度。在此基础上，构建了药物-疾病异构网络，基于该异构网络，采用双向随机游走算法为所有的药物预测候选疾病。然而，这些基于网络的药物重定位方法仅仅使用了疾病、药物构建的关联网络。而生

物信息学技术的迅速发展已经积累了多种用于刻画生物分子关联的信息，可用于构建各种生物信息网络，如蛋白质交互网络、药物-靶标网络等，为药物重定位提供了新的发展机遇。

[0005] 在药物重定位研究方面，已有一些集成多源生物网络的方法被成功应用到疾病-药物关联预测中。比如，Wang等人集成了疾病、药物和靶标三种生物信息构建了异构网络模型，提出了一种计算重定位框架TL_HGBI。Martinez等人提出了一种基于网络的候选药物预测方法，DrugNet，该方法同时集成了疾病、药物和靶标网络。这两种方法的实验结果证实集成多源生物信息可以提高药物重定位的预测效果。然而，相对于现有可用的生物信息来说，如何集成并构建多源生物信息网络并进行有效预测的研究仍处于初级阶段。对于TL_HGBI方法，该方法没有集成已被实验验证的疾病-基因关联信息；而DrugNet完成从药物网络到疾病网络的直接或间接扩散，但是没有有效的利用从疾病网络到药物网络的信息扩散。因此，有必要设计一种融合多种生物信息并能充分利用这些生物信息进行药物发现的重定位方法。

## 发明内容

[0006] 本发明所解决的技术问题是，针对现有技术的不足，提出一种基于多元信息融合和随机游走模型的药物重定位方法，本发明能充分利用全局网络信息，提高预测性能；简单有效，易于实施，

[0007] 本发明所提供的技术方案为：

[0008] 一种基于多元信息融合和随机游走模型的药物重定位方法，包括以下步骤：

[0009] 1)构建疾病-靶标-药物异构网络：利用已知的疾病数据、药物数据、靶标数据、疾病-药物关联数据、疾病-靶标关联数据和药物-靶标关联数据，构建疾病网络、药物网络、靶标网络、疾病-药物关联网络、疾病-靶标关联网络和药物-靶标关联网络；通过关联网络连接疾病网络、药物网络和靶标网络，得到疾病-靶标-药物异构网络；

[0010] 2)扩展基本随机游走模型到该异构网络：首先根据已知的疾病-药物关联数据和疾病-靶标关联数据构建随机游走的初始概率矩阵；然后利用已知的药物相似性、疾病相似性、靶标相似性、疾病-药物关联数据、疾病-靶标关联数据和药物-靶标关联数据，构建随机游走的转移矩阵；

[0011] 3)预测新的药物-疾病关联：对于给定的疾病，依据所构建的初始概率矩阵和转移矩阵，迭代地在异构网络中进行随机游走，执行直到游走结果达到收敛状态；根据游走结果，得到给定疾病与所有药物存在关联的概率值，概率值越大，表明疾病与药物之间存在关联的可能性越大；按照概率值的大小，把与给定疾病不存在已知关联的所有药物进行排序，从而为给定疾病推荐新的治疗药物。

[0012] 类似地，通过本步骤还可以对于给定的药物，预测其新的适用的疾病，即预测给定药物的新适应症。

[0013] 以下对本发明方法进行详细说明。

[0014] 一、计算疾病相似性、药物相似性和靶标相似性，构建疾病-靶标-药物异构网络

[0015] 1.1)基于疾病的表型信息，计算疾病之间的相似性值，构建疾病网络；在疾病网络中，顶点集合$D=\{d_1,d_2,\cdots,d_n\}$表示n种疾病，顶点$d_i$和顶点$d_j$之间有边相连接，疾病i和疾病j之间的相似性值即为该条边的权值；

[0016] 1.2)基于药物的化学结构信息，计算药物之间的相似性值，构建药物网络；在药物网络中，顶点集合R＝{$r_1$,$r_2$,…,$r_m$}表示m种药物，顶点$r_i$和顶点$r_j$之间有边相连接，药物i和药物j之间的相似性值即为该条边的权值；

[0017] 1.3)基于靶标的序列信息，计算靶标之间的相似性值，构建靶标网络；在靶标网络中，顶点集合T＝{$t_1$,$t_2$,…,$t_p$}表示p种靶标，顶点$t_i$和顶点$t_j$之间有边相连接，靶标i和靶标j之间的相似性值即为该条边的权值；

[0018] 1.4)基于已知的疾病-药物关联数据，构建疾病-药物关联网络；将疾病-药物关联网络建模为一个二分图$G_{dr}$(D,R,E)，其中E(G)⊆R×DE(G)＝{$e_{ij}$,$d_i$与$r_j$之间的边}，如果疾病$d_i$与药物$r_j$之间存在已知关联，则$d_i$与$r_j$之间的边权重设置为1，否则设置为0；

[0019] 1.5)基于已知的疾病-靶标关联数据，构建疾病-靶标关联网络；将疾病-靶标关联网络建模为一个二分图$G_{dt}$(D,T,E)，其中E(G)⊆D×TE(G)＝{$e_{ij}$,$d_i$与$t_j$之间的边}，如果疾病$d_i$与靶标$t_j$之间存在已知关联，则$d_i$与$t_j$之间的边权重设置为1，否则设置为0；

[0020] 1.6)基于已知的药物-靶标关联数据，构建药物-靶标关联网络；将药物-靶标关联网络建模为一个二分图$G_{rt}$(R,T,E)，其中E(G)⊆R×TE(G)＝{$e_{ij}$,$r_i$与$t_j$之间的边}，如果药物$r_i$与靶标$t_j$之间存在已知关联，则$r_i$与$t_j$之间的边权重设置为1，否则设置为0；

[0021] 1.7)构建疾病-靶标-药物异构网络，该网络包括疾病网络、药物网络、靶标网络、疾病-药物关联网络、疾病-靶标关联网络和药物-靶标关联网络，其中疾病网络、药物网络和靶标网络通过对应的关联网络连接。

[0022] 该异构网络对应的邻接矩阵A可以表示为：

[0023] $$A=\begin{bmatrix} A_{RR} & A_{RT} & A_{RD} \\ A_{RT}^{T} & A_{TT} & A_{TD} \\ A_{RD}^{T} & A_{TD}^{T} & A_{DD} \end{bmatrix} \quad (1)$$

[0024] 其中，A的主对角线上的三个子矩阵$A_{RR}$、$A_{TT}$、$A_{DD}$对应的是药物网络、靶标网络和疾病网络的邻接矩阵；$A_{RT}$、$A_{RD}$、$A_{TD}$对应的是药物-靶标网络、药物-疾病网络和靶标-疾病网络的邻接矩阵，$A_{RT}^{T}$、$A_{RD}^{T}$、$A_{TD}^{T}$分别是$A_{RT}$、$A_{RD}$、$A_{TD}$的转置矩阵。

[0025] 二、扩展基本随机游走模型到该异构网络

[0026] 基于所构建的异构网络，本发明模拟在异构网络中进行随机游走的过程，实现为特定疾病推荐候选治疗药物。本发明基于扩展的随机游走模型(RWR)。RWR描述随机游走者从种子节点开始，随机选择转移到其中一个邻居节点的过程。在经过多次游走迭代之后，到达网络中所有节点的概率达到收敛状态，然后对所有候选节点依照到达该节点的概率大小进行排序。RWR数学表示如下：

[0027] $$P_{t+1}=(1-\gamma)M^{T}P_t+\gamma P_0 \quad (2)$$

[0028] 其中，γ表示重启概率，在游走过程中，在某节点的游走者以概率γ直接返回到种子节点，或者以概率1-γ随机地选择与该节点相邻的边，沿这条边移动到下一个节点；γ的取值范围为[0,1]，可以根据交叉验证实验选取最优值；M是转移矩阵，其中元素$M_{ij}$表示从节点i转移到节点j的概率；$M^{T}$是M的转置矩阵；$P_0$是初始概率矩阵，其中每个种子节点被赋予等同的概率[1/(种子节点数)]。$P_t$是在迭代到第t步时的概率向量，其中第i个元素表示游走者到达第i个节点的概率。多次迭代之后，当$P_{t+1}$与$P_t$之间的差别小于某个很小的阈值时(比如$10^{-10}$)，可以认为游走达到稳定状态P。本发明基于所构建的疾病-靶标-药物异构网络，扩

展随机游走模型，为所有的疾病预测候选药物。该算法的过程描述如下：

[0029] 第一步：构建初始概率矩阵$P_0$；

[0030] 随机游走者在游走过程的每一步，可以概率γ重新回到种子节点开始游走。比如，给定疾病为d，预测d的候选药物，将给定疾病d作为疾病网络中的种子节点，将与给定疾病d存在已知关联的所有药物节点作为药物网络中的种子节点，将与给定疾病d存在已知关联的所有靶标节点作为靶标网络中的种子节点；根据这三个网络中的种子节点定义，异构网络的初始概率矩阵$P_0$包括$Pr_0$，$Pt_0$和$Pd_0$三部分，分别表示药物网络、靶标网络和疾病网络的初始概率向量；其中$Pr_0$包含m个元素，分别对应m个药物的初始概率；如果第j个药物与给定疾病d存在关联，则$Pr_0$中的第j个元素值为1/(与给定疾病d存在关联的药物个数)，否则$Pr_0$中的第j个元素值为0；$Pt_0$包含p个元素，分别对应p个靶标的初始概率；如果第j个靶标与给定疾病d存在关联，则$Pt_0$中的第j个元素值为1/(与给定疾病d存在关联的靶标个数)，否则$Pt_0$中的第j个元素值为0；$Pd_0$包含n个元素，分别对应n个疾病的初始概率；$Pd_0$中与给定疾病d相应的元素的元素值为1，其他元素值为0；所创建异构网络的初始概率矩阵表示为：

[0031] $$P_0=\begin{bmatrix}\lambda_R Pr_0\\ \lambda_T Pt_0\\ (1-\lambda_R-\lambda_T)Pd_0\end{bmatrix} \quad (3)$$

[0032] 其中，参数$\lambda_R$，$\lambda_T$和$1-\lambda_R-\lambda_T$对应药物网络、靶标网络和疾病网络的重要性，$\lambda_R$，$\lambda_T$，$1-\lambda_R-\lambda_T\in[0,1]$，通过交叉验证实验选择最优参数值。如果参数$\lambda_R$比$\lambda_T$和$1-\lambda_R-\lambda_T$大，则表示药物网络比靶标网络、疾病网络重要，在游走过程中的每一步选择重新从种子节点开始游走时，游走者更易于选择药物网络的种子节点。

[0033] 第二步：构建转移概率矩阵M；

[0034] 在所构建的疾病-靶标-药物异构网络，随机游走者首先基于初始概率选择从种子节点开始游走，然后以一定概率选择转移到当前节点的邻居节点，或者重新从种子节点开始游走。因此，需要计算每个节点到其邻居节点的转移概率。异构网络的转移概率矩阵定义如下：

[0035] $$M=\begin{bmatrix}M_{RR} & M_{RT} & M_{RD}\\ M_{TR} & M_{TT} & M_{TD}\\ M_{DR} & M_{DT} & M_{DD}\end{bmatrix} \quad (4)$$

[0036] 矩阵M中包含九个子矩阵，包含三个网内转移矩阵和六个网间转移矩阵；其中，$M_{RR}$是药物网络的网内转移矩阵，包括任一药物节点到其它药物节点的转移概率；$M_{TT}$是靶标网络的网内转移矩阵，包括任一靶标节点到其它靶标节点的转移概率；$M_{DD}$是疾病网络的网内转移矩阵，包括任一疾病节点到其它疾病节点的转移概率；$M_{RD}$是药物网络和疾病网络的网间转移矩阵，包括药物节点到疾病节点的转移概率；$M_{RT}$是药物网络和靶标网络的网间转移矩阵，包括药物节点到靶标节点的转移概率；$M_{DR}$是疾病网络和药物网络的网间转移矩阵，包括疾病节点到药物节点的转移概率；$M_{DT}$是疾病网络和靶标网络的网间转移矩阵，包括疾病节点到靶标节点的转移概率；$M_{TR}$是靶标网络和药物网络的网间转移矩阵，包括靶标节点到药物节点的转移概率；$M_{TD}$是靶标网络和疾病网络的网间转移矩阵，包括靶标节点到疾病节点的转移概率。

[0037] 在异构网络上进行随机游走的过程中，游走者可以选择转移到当前网络内其他节

点或者其他网络中的节点。比如，当游走者位于疾病网络中的某节点，他可以游走到其他疾病节点，或者跳转到药物网络、靶标网络。所以需要定义不同网络之间的跳转概率，并通过交叉验证实验选择最优参数值。定义参数$\lambda_{DR}$，表示从疾病网络(D)到药物网络(R)的跳转概率；$\lambda_{RD}$表示从药物网络(R)到疾病网络(D)的跳转概率；$\lambda_{DT}$表示从疾病网络(D)到靶标网络(T)的跳转概率；$\lambda_{TD}$表示从靶标网络(T)到疾病网络(D)的跳转概率；$\lambda_{RT}$表示从药物网络(R)到靶标网络(T)的跳转概率；$\lambda_{TR}$表示从靶标网络(T)到药物网络(R)的跳转概率。如果游走者在某个疾病节点，该疾病节点与某些药物节点和靶标节点关联，则他跳转到药物网络的概率是$\lambda_{DR}$，跳转到靶标网络的概率是$\lambda_{DT}$，在当前网络内转移的概率是$1-\lambda_{DR}-\lambda_{DT}$。

[0038] 基于公式(1)中定义的矩阵A，可以计算公式(4)中的每个子矩阵。基于对应网络的相似性数据和关联信息，可以构建公式(4)中的网内转移矩阵。比如，疾病网络的网内转移矩阵$M_{DD}$的定义如下：

[0039]
$$M_{DD}(i,j)=\begin{cases} A_{DD}(i,j)/\sum_j A_{DD}(i,j) & if \sum_j A_{DR}(i,j)=0,\ \sum_j A_{DT}(i,j)=0 \\ (1-\lambda_{DR})A_{DD}(i,j)/\sum_j A_{DD}(i,j) & if \sum_j A_{DR}(i,j)\neq 0,\ \sum_j A_{DT}(i,j)=0 \\ (1-\lambda_{DT})A_{DD}(i,j)/\sum_j A_{DD}(i,j) & if \sum_j A_{DR}(i,j)=0,\ \sum_j A_{DT}(i,j)\neq 0 \\ (1-\lambda_{DR}-\lambda_{DT})A_{DD}(i,j)/\sum_j A_{DD}(i,j) & if \sum_j A_{DR}(i,j)\neq 0,\ \sum_j A_{DT}(i,j)\neq 0 \end{cases} \quad (5)$$

[0040] 在等式(5)中，$A_{DD}$对应的是疾病网络的邻接矩阵。当随机游走者位于疾病网络的某节点，如果该节点在药物网络和靶标网络中没有关联节点，则他只能在疾病内部游走；如果该节点在药物网络中有关联节点，但是在靶标网络中没有关联节点，则他在疾病内部游走的概率是$1-\lambda_{DR}$；如果该节点在药物网络中没有关联节点，但是在靶标网络中有关联节点，则他在疾病内部游走的概率是$1-\lambda_{DT}$；如果该节点在药物网络和靶标网络中都有已知关联节点，则他在疾病内部游走的概率是$1-\lambda_{DR}-\lambda_{DT}$。

[0041] 类似的，药物网络的网内转移矩阵$M_{RR}$和靶标网络的网内转移矩阵$M_{TT}$的定义如下：

[0042]
$$M_{RR}(i,j)=\begin{cases} A_{RR}(i,j)/\sum_j A_{RR}(i,j) & if \sum_j A_{RD}(i,j)=0,\ \sum_j A_{RT}(i,j)=0 \\ (1-\lambda_{RD})A_{RR}(i,j)/\sum_j A_{RR}(i,j) & if \sum_j A_{RD}(i,j)\neq 0,\ \sum_j A_{RT}(i,j)=0 \\ (1-\lambda_{RT})A_{RR}(i,j)/\sum_j A_{RR}(i,j) & if \sum_j A_{RD}(i,j)=0,\ \sum_j A_{RT}(i,j)\neq 0 \\ (1-\lambda_{RD}-\lambda_{RT})A_{RR}(i,j)/\sum_j A_{RR}(i,j) & if \sum_j A_{RD}(i,j)\neq 0,\ \sum_j A_{RT}(i,j)\neq 0 \end{cases} \quad (6)$$

[0043]
$$M_{TT}(i,j)=\begin{cases} A_{TT}(i,j)/\sum_j A_{TT}(i,j) & if \sum_j A_{TD}(i,j)=0,\ \sum_j A_{TR}(i,j)=0 \\ (1-\lambda_{TD})A_{TT}(i,j)/\sum_j A_{TT}(i,j) & if \sum_j A_{TD}(i,j)\neq 0,\ \sum_j A_{TR}(i,j)=0 \\ (1-\lambda_{TR})A_{TT}(i,j)/\sum_j A_{TT}(i,j) & if \sum_j A_{TD}(i,j)=0,\ \sum_j A_{TR}(i,j)\neq 0 \\ (1-\lambda_{TD}-\lambda_{TR})A_{TT}(i,j)/\sum_j A_{TT}(i,j) & if \sum_j A_{TD}(i,j)\neq 0,\ \sum_j A_{TR}(i,j)\neq 0 \end{cases} \quad (7)$$

[0044] 根据已知的关联数据，可以构建M中的六个网间转移矩阵。比如，疾病网络和药物网络的网间转移矩阵$M_{DR}$定义如下：

[0045]
$$M_{DR}(i,j)=\begin{cases} \lambda_{DR}A_{DR}(i,j)/\sum_j A_{DR}(i,j) & if \sum_j A_{DR}(i,j)\neq 0 \\ 0 & otherwise \end{cases} \quad (8)$$

[0046] 当随机游走者位于疾病网络的某节点，如果该节点在药物网络中有关联节点，则他以概率$\lambda_{DR}$跳转到药物网络；否则，他不能跳转到药物网络。类似的，其他的网间转移矩阵

$M_{RD}$、$M_{RT}$、$M_{DT}$、$M_{TR}$和$M_{TD}$定义如下：

[0047] $$M_{RD}(i,j)=\begin{cases}\lambda_{RD}A_{RD}(i,j)/\sum_j A_{RD}(i,j) & if\sum_j A_{RD}(i,j)\neq 0\\ 0 & otherwise\end{cases} \quad (9)$$

[0048] $$M_{RT}(i,j)=\begin{cases}\lambda_{RT}A_{RT}(i,j)/\sum_j A_{RT}(i,j) & if\sum_j A_{RT}(i,j)\neq 0\\ 0 & otherwise\end{cases} \quad (10)$$

[0049] $$M_{DT}(i,j)=\begin{cases}\lambda_{DT}A_{DT}(i,j)/\sum_j A_{DT}(i,j) & if\sum_j A_{DT}(i,j)\neq 0\\ 0 & otherwise\end{cases} \quad (11)$$

[0050] $$M_{TR}(i,j)=\begin{cases}\lambda_{TR}A_{TR}(i,j)/\sum_j A_{TR}(i,j) & if\sum_j A_{TR}(i,j)\neq 0\\ 0 & otherwise\end{cases} \quad (12)$$

[0051] $$M_{TD}(i,j)=\begin{cases}\lambda_{TD}A_{TD}(i,j)/\sum_j A_{TD}(i,j) & if\sum_j A_{TD}(i,j)\neq 0\\ 0 & otherwise\end{cases} \quad (13)$$

[0052] 三、实现在异构网络中的随机游走，预测新的药物-疾病关联；

[0053] 给定疾病d，预测候选治疗药物，基于所构建的疾病-靶标-药物异构网络，以及在第一步和第二步分别定义的初始概率矩阵$P_0$和转移概率矩阵M，在异构网络中进行随机游走，经过若干次游走之后，达到稳定状态，对应的概率矩阵记为P，P中的每个元素表示游走者到达相应节点的最终概率。

[0054] 概率矩阵P包含三部分：Pr，Pt和Pd；其中Pr中的第i个元素表示疾病d与药物$r_i$之间存在关联的概率；Pt中的第i个元素表示疾病d与靶标$t_i$之间存在关联的概率；Pd中的第i个元素表示疾病d与疾病$d_i$之间存在关联的概率。如果药物$r_i$与疾病d之间不存在已知关联，则药物$r_i$称为疾病d的候选药物。Pr中存放有所有候选药物与疾病d之间存在关联的概率，概率值越大，表示该药物与疾病d存在关联的可能性越大，根据概率值为给定疾病推荐候选药物(新的治疗药物)。

[0055] 有益效果：

[0056] 本发明基于相似的药物更易于关联相似的疾病、相似的疾病更易于关联相似的药物的假设，融合已知的多元生物信息构建疾病-靶标-药物异构网络，然后基于该异构网络，采用扩展随机游走算法，迭代地在所构建的异构网络上执行随机游走，预测潜在的、新的疾病-药物关联，识别疾病的新的治疗药物和已存在药物的新适应症。本发明能充分利用全局网络信息，提高预测性能。该药物重定位方法能有效地挖掘已知药物的新的潜在适应症。本发明简单有效，易于实施，通过与其他方法比较，及在标准数据集上测试表明，该发明在药物重定位方面具有较好的预测性能。

## 附图说明

[0057] 图1本发明(RWHNDR)流程图；

[0058] 图2对于预测已知疾病的候选药物，基于留一交叉验证评价本发明(RWHNDR)与所比较方法TL_HGBI、DrugNet的性能；图2(a)不同方法预测结果对应的ROC曲线，图2(b)不同的Top阈值下正确检索到的关联数。

[0059] 图3对于预测新疾病的候选药物，基于留一交叉验证评价本发明(RWHNDR)与所比

较方法TL_HGBI、DrugNet的性能;图3(a)不同方法预测结果对应的ROC曲线,图3(b)不同的Top阈值下正确检索到的关联数。

[0060] 图4评价集成靶标信息对于预测性能的影响;图4(a)为已知疾病预测候选药物,DR_RWRH与本发明(RWHNDR)预测结果对应的ROC曲线,图4(b)为新疾病预测候选药物,DR_RWRH与本发明(RWHNDR)预测结果对应的ROC曲线。

[0061] 图5在新数据集上的留一交叉验证,为已知疾病预测候选药物。图5(a)不同方法预测结果对应的ROC曲线。图5(b)不同的Top阈值下正确检索到的关联数。

[0062] 图6在新数据集上的留一交叉验证,为新疾病预测候选药物。图6(a)不同方法预测结果对应的ROC曲线。图6(b)不同的Top阈值下正确检索到的关联数。

## 具体实施方式

[0063] 如图1所示,本发明具体实现过程如下:

[0064] 一、计算疾病、药物和靶标相似性、构建药物-疾病异构网络疾病-靶标-药物异构网络;

[0065] 本方法所应用的数据集包括疾病集合、药物集合、靶标集合、疾病-药物关联数据、疾病-靶标关联数据与药物-靶标关联数据。

[0066] 首先,计算疾病、药物和靶标相似性:

[0067] 1.药物相似性计算

[0068] 基于药物的SMILES化学结构信息,利用CDK(Chemical development kit)计算药物之间的化学结构相似性,也称为分子相似性。根据所有的药物对相似性,构建药物相似性矩阵。

[0069] 2.疾病相似性计算

[0070] 疾病相似性是通过工具MinMiner计算得到的,该工具基于疾病的表型信息计算疾病间的相似性。根据所有的疾病对的相似性,构建疾病相似性矩阵。

[0071] 3.靶标相似性计算

[0072] 基于靶标蛋白的氨基酸序列信息计算靶标之间的相似性。从Uniprot数据库中获取靶标蛋白的序列信息,然后利用R包(Rcpi,基于序列比对计算蛋白序列相似性)计算靶标的序列相似性。根据所有的靶标对相似性,构建靶标相似性矩阵。

[0073] 然后,基于疾病相似性矩阵、药物相似性矩阵和靶标相似性矩阵,构建疾病网络、药物网络和靶标网络。

[0074] 最后,构建疾病-靶标-药物异构网络,该网络包括疾病网络、药物网络、靶标网络、疾病-药物关联网络、疾病-靶标关联网络和药物-靶标关联网络,其中疾病网络、药物网络和靶标网络通过对应的关联网络连接。

[0075] 二、扩展基本随机游走模型到该异构网络;

[0076] 三、实现在异构网络中的随机游走,预测新的药物-疾病关联;

[0077] 给定疾病d,预测候选治疗药物,基于所构建的疾病-靶标-药物异构网络,以及在第一步和第二步分别定义的初始概率矩阵$P_0$和转移概率矩阵M,在异构网络中进行随机游走,经过若干次游走之后,达到稳定状态,对应的概率矩阵记为P,P中的每个元素表示游走者到达相应节点的最终概率。概率矩阵P包含三部分:Pr,Pt和Pd;其中Pr中的第i个元素表

示疾病d与药物$r_i$之间存在关联的概率;Pt中的第i个元素表示疾病d与靶标$t_i$之间存在关联的概率;Pd中的第i个元素表示疾病d与疾病$d_i$之间存在关联的概率。如果药物$r_i$与疾病d之间不存在已知关联,则药物$r_i$称为疾病d的候选药物。Pr中存放有所有候选药物与疾病d之间存在关联的概率,概率值越大,表示该药物与疾病d存在关联的可能性越大,根据概率值为给定疾病推荐候选药物(新的治疗药物)。

[0078] 四、实验验证

[0079] 1.评价指标

[0080] 本发明(RWHNDR)不能同时为所有的疾病预测候选治疗药物,也就是每次预测只能为给定疾病预测候选药物。另外,在标准数据集中,每个疾病平均有6.18个已知的关联药物,所以留一交叉验证适用于评价RWRHDR的预测性能。

[0081] 数据集中所有未知的药物-疾病关联,作为候选药物-疾病关联。数据集中每条已知的药物-疾病关联轮流作为测试数据集,剩余的已知关联作为测试数据集,进行实验。其中,测试集中的药物-疾病关联所包含的药物称为测试药物,疾病称为测试疾病。测试疾病作为疾病网络中的种子节点;与测试疾病存在已知关联的药物(不包含测试药物)作为药物网络中的种子节点;与测试疾病存在已知关联的靶标作为靶标网络中的种子节点。与测试疾病不存在已知关联的药物,以及测试药物,被称为候选药物。根据预测得到的概率值,所有的候选药物按降序排列。对特定的阈值,如果测试药物的关联大于这个阈值,这个关联被认为是一个true positive(TP);如果小于这个阈值,则是一个false negative(FN)。另外,如果候选药物的关联大于这个阈值,这个关联被认为是一个false positive(FP);如果小于这个阈值,则是一个true negative(TN)。通过变换不同的阈值,可以计算不同的真阳性率TPR(True Positive Rate)和假阳性率FPR(False Positive Rate),从而可以得到ROC曲线,计算该曲线下方的面积可以得到AUC值,AUC值被用来评测算法性能。

[0082] 除了AUC值,算法预测结果中,排在前面的关联在实际应用中也非常重要。因此,我们还用所预测的排在前面的关联来评价方法。比如,排在前10的预测结果中,被正确预测到的测试集中的关联数。一般,排在预测结果靠前部分的已知关联越多,该预测方法越具有实用性。

[0083] 2.与其它方法的比较

[0084] 为了评价本发明所提出的预测方法的有效性,将本发明(RWHNDR)与其他两种方法进行比较(TL_HGBI和DrugNet)。TL_HGBI是基于关联推定(guilt-by-association)的三层异构网络图模型,能够识别疾病、药物和靶标之间的关联关系;DrugNet是基于网络的药物重定位方法,通过在网络之间扩散信息,完成药物-疾病关联关系的预测。

[0085] 本发明应用到两种预测问题中,一种是为已知疾病识别候选药物,另一种是为新疾病识别候选药物。这里,已知疾病就是已经有治疗药物的疾病,新疾病是没有任何治疗药物的疾病。很明显,在为已知疾病识别候选药物的预测问题中,包含更多的已知信息。

[0086] (1)为已知疾病预测候选药物

[0087] 标准数据集中,有216个疾病至少关联了两个药物,这些疾病涉及1836条已知的疾病-药物关联。在留一交叉验证中,这些测试疾病的一条已知药物关联被删除后,还包含有其它的已知药物关联。这种情况下,测试疾病及它所关联的药物和靶标集合作为种子节点,为已知疾病预测候选药物。

[0088] 留一交叉验证实验结果如图2所示，从结果可以看出，本发明(RWHNDR)方法的AUC值为0.926，而其它两种方法TL_HGBI和DrugNet的AUC值分别为0.881和0.771。另外，从预测的Top-ranked结果来看，1836条已知疾病-药物关联中，有1079条关联被排在预测结果中的前1％中，优于其他预测方法。Top-ranked结果在实际应用中特别重要，所以本发明优于其他方法。

[0089] (2)为新疾病预测候选药物

[0090] 标准数据集中，有97个疾病只关联了一个药物。在留一交叉验证中，给定测试疾病的一条已知药物关联被删除后，这个测试疾病成为没有任何药物关联的新疾病。因此，在这种情况下，只有测试疾病和它所关联的靶标集合作为种子节点，为新疾病预测候选药物。

[0091] 所有方法的留一交叉验证结果如图3所示，从结果可以看出，本发明(RWHNDR)方法的AUC值为0.841，而其它两种方法TL_HGBI和DrugNet的AUC值分别为0.625和0.822。另外，从预测的Top-ranked结果来看，97条已知疾病-药物关联中，有45条关联被排在预测结果中的前1％中，而其他方法预测得到的关联数少于本发明方法。

[0092] (3)集成target信息对预测的影响

[0093] 为评价集成target信息对预测性能的影响，本发明提出DR_RWRH方法，该方法实现在药物-疾病异构网络中的随机游走，从而为特定疾病推荐候选药物。与本发明(RWHNDR)方法的区别是，DR_RWRH方法没有利用target信息。这里分析为已知疾病和新疾病推荐药物的两种情况，采用留一交叉验证的实验结果如图4所示。实验结果表明，在为新疾病预测候选药物时，本发明方法明显优于DR_RWRH方法。因此，集成target信息能在一定程度上提高预测的准确性。

[0094] (4)案例分析

[0095] 前面已经通过交叉验证实验说明了本发明在预测疾病-药物关联方面的有效性，基于标准数据集，将该发明应用到未知药物-疾病关系的预测中。在预测过程中，用标准数据集中的所有已知关联作为训练集，本发明(RWHNDR)在该数据集上进行预测，按照预测结果对未知的疾病-药物关联进行排序，得分越高的疾病-药物对之间存在关联的可能性越大。主要关注排序靠前的预测结果，通过查找文献，验证为每个疾病推荐的排名前5位的候选药物的准确性。本发明中选取了神经障碍及癌症疾病作案例分析，包括4个疾病Huntington disease(OMIM:143100)、Parkinson disease(OMIM:168600)、Breast cancer(OMIM:114480)和Lung cancer(OMIM:211980)。

[0096] 为这4个疾病预测的Top-5ranked药物及文献支撑结果如表1所示。比如，Huntington disease是一种遗传性中枢神经系统疾病，在所预测的排在前5的药物中，有两个药物对Huntington disease的治疗研究在相关文献中得到验证。其中，药物Carbamazepine最初用于治疗三叉神经痛相关的癫痫和疼痛，对治疗Huntington disease中的排尿障碍、抑郁偏执等已有相关的研究报道。另外，药物Dantrolene已在相关研究中证实可以Huntington disease的潜在治疗药物。案例分析结果表明本发明方法预测的结果将对生物学实验具有一定的指导作用。

[0097] 表1.案例分析结果

[0098]

| Disease (OMIM IDs) | Top 5 candidate drugs (DrugBank IDs) | References |
|---|---|---|
| Huntington disease (143100) | Carbamazepine (DB00564) | confirmed |
| | Dantrolene (DB01219) | confirmed |
| | Vigabatrin (DB01080) | |
| | LORAZEPAM (DB00186) | |
| | Tizanidine (DB00697) | |
| Parkinson disease (168600) | Paclitaxel (DB01229) | |
| | Docetaxel (DB01248) | |
| | Biperiden (DB00810) | |
| | Rivastigmine (DB00989) | confirmed |
| | Levodopa (DB01235) | confirmed |
| Breast cancer (114800) | Caffeine (DB00201) | |
| | Ethinyl Estradiol (DB00977) | confirmed |
| | Aspirin (DB00945) | confirmed |
| | Arsenic trioxide (DB01169) | confirmed |
| | Estramustine (DB01196) | |
| Lung cancer (211980) | Vincristine (DB00541) | |
| | Methotrexate (DB00563) | |
| | Sorafenib (DB00398) | confirmed |
| | Cisplatin (DB00515) | confirmed |
| | Daunorubicin (DB00694) | |

[0099] (5)在其他数据集上的验证

[0100] 对于药物-疾病关联预测方法的评估，很多研究都是通过采用交叉验证实验来分析方法的准确性，且基本上只在单一的数据集上做验证。而本发明除了在标准数据集上做评价之外，还在所收集的新的数据集上评价预测性能。

[0101] 在这新的数据集上，通过留一交叉验证，分析本发明对于已知疾病和新疾病的推荐候选药物的准确性，并完成与其他两种最新方法的比较。相关的实验结果如图5和图6所示，从AUC值、Top-ranked指标，可以看到本发明方法的结果优于其他方法。

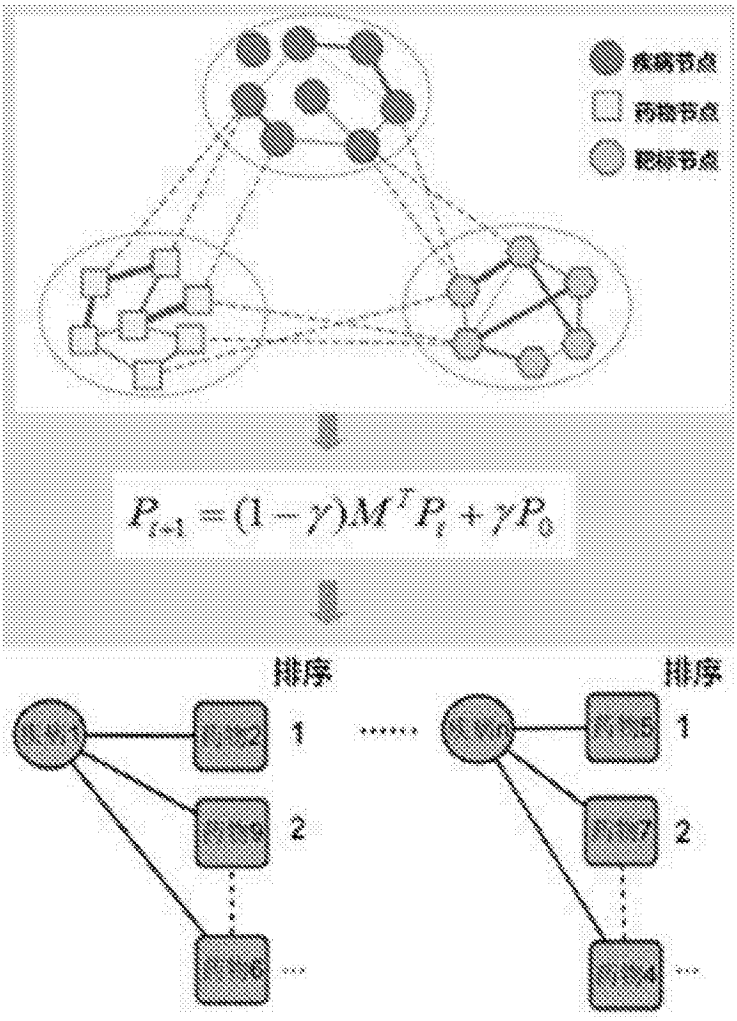
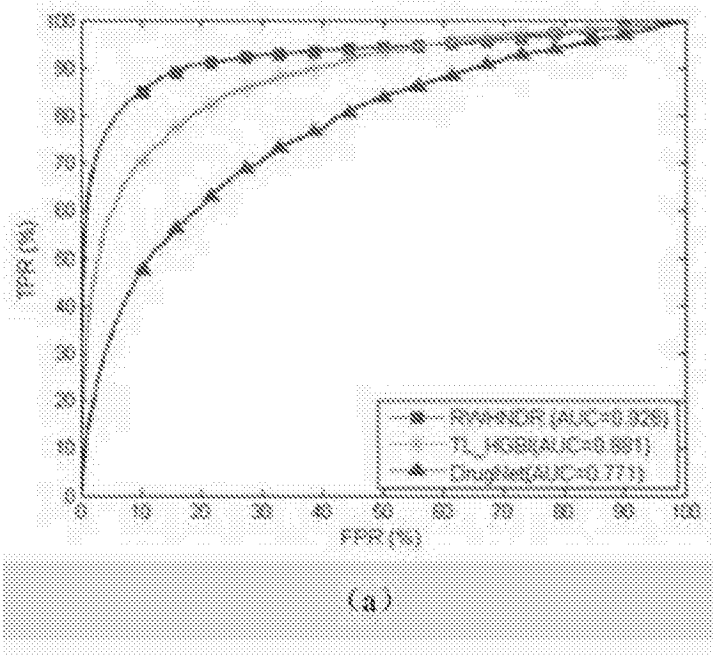
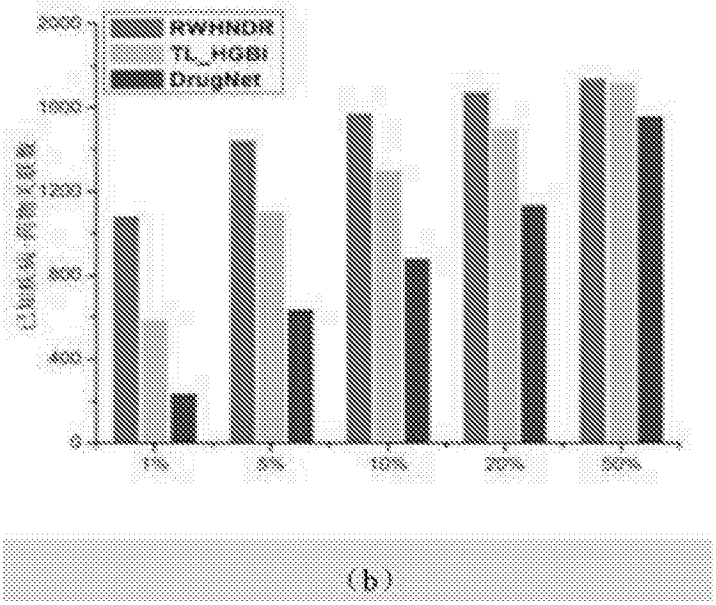



图1


TPR (%)
FPR (%)
RWHNDR (AUC=0.826)
TL_HGBI(AUC=0.881)
DrugNet(AUC=0.771)

(a)


RWHNDR
TL_HGBI
DrugNet
1%
5%
10%
20%
50%

(b)

图2

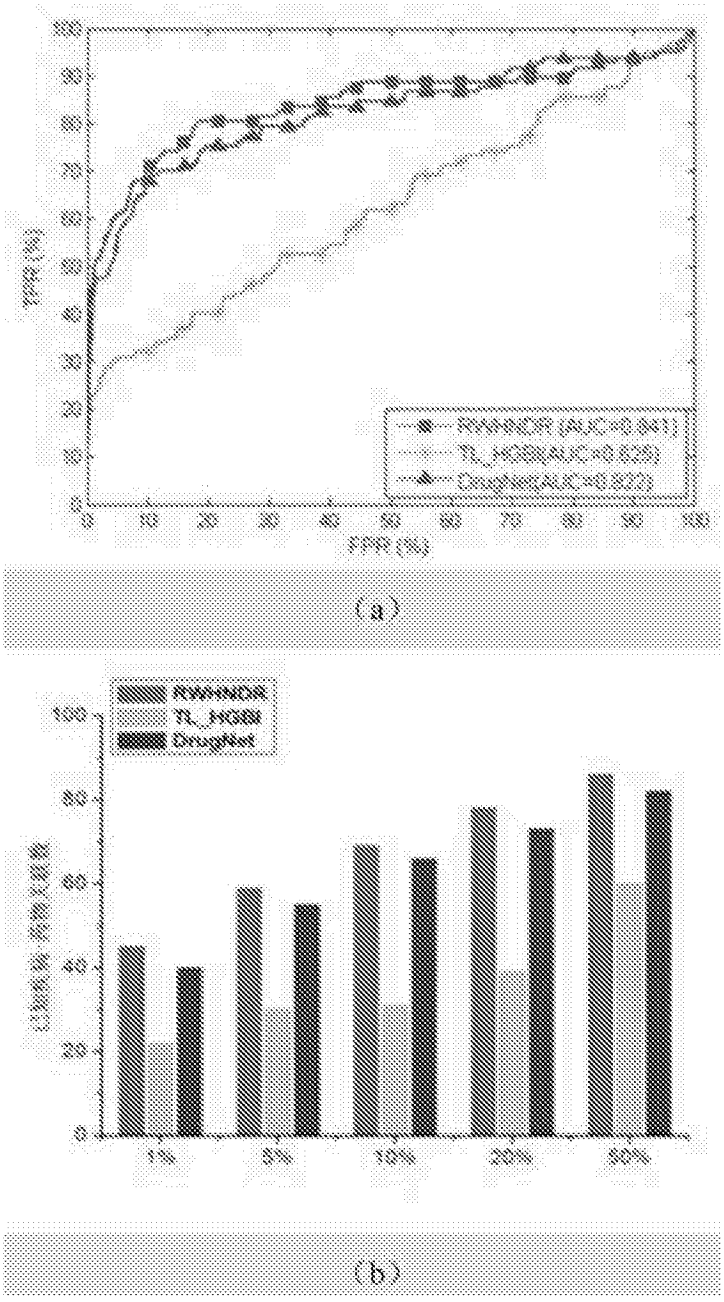

TPR (%)
FPR (%)
RWHNOR (AUC=0.841)
TL_HGBI(AUC=0.629)
DrugNet(AUC=0.872)
(a)
RWHNOR
TL_HGBI
DrugNet
1%
5%
10%
20%
50%
(b)

图3

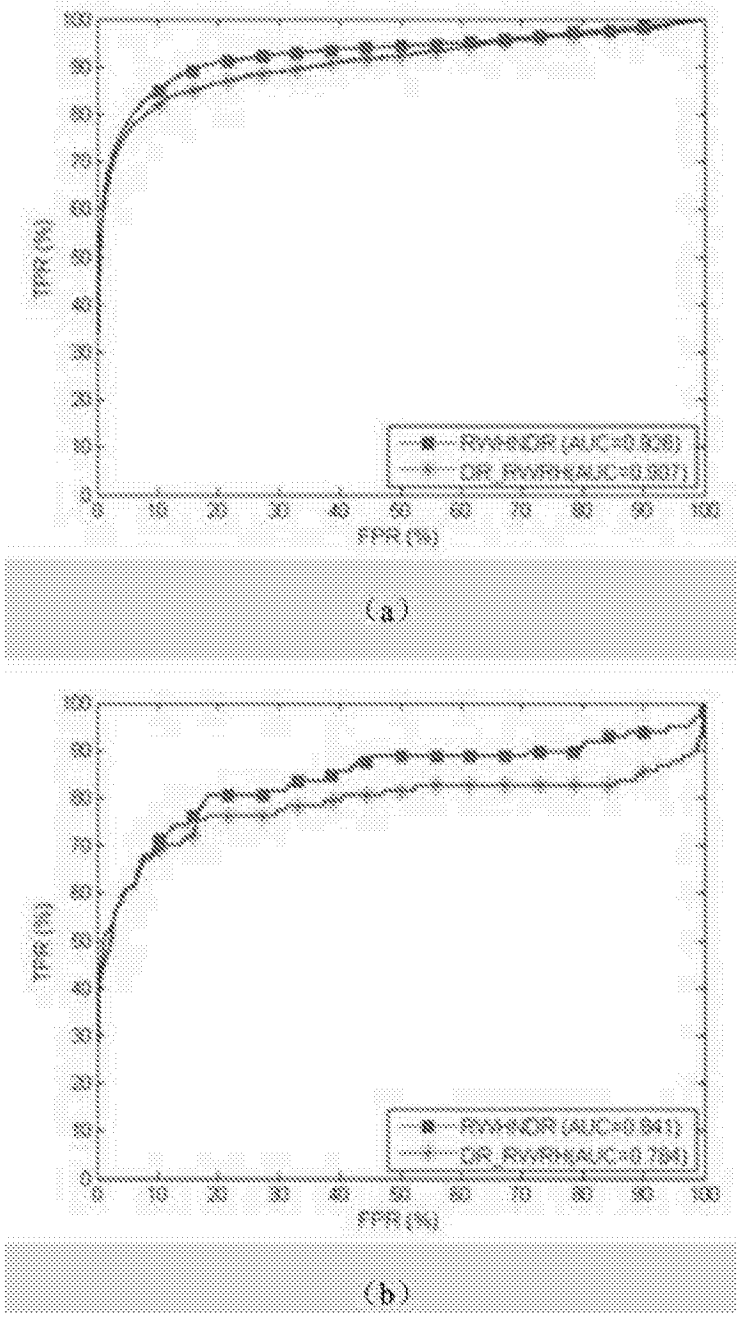
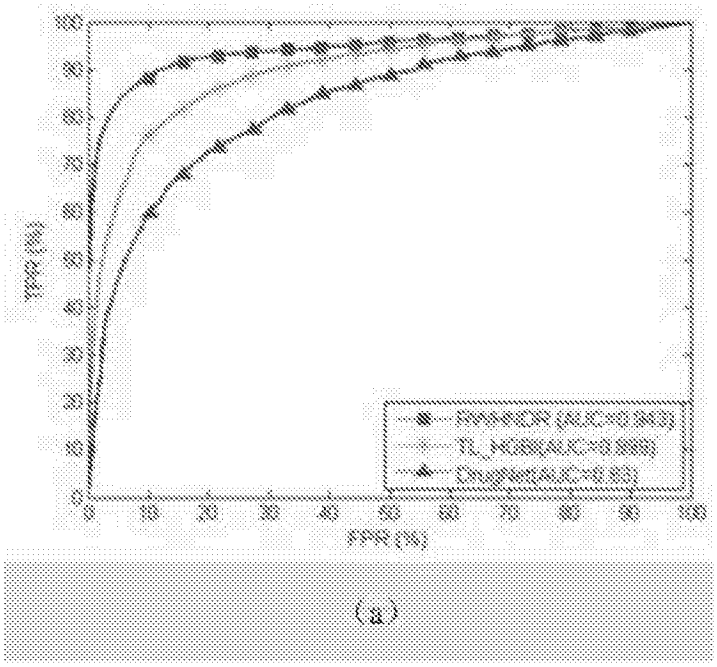
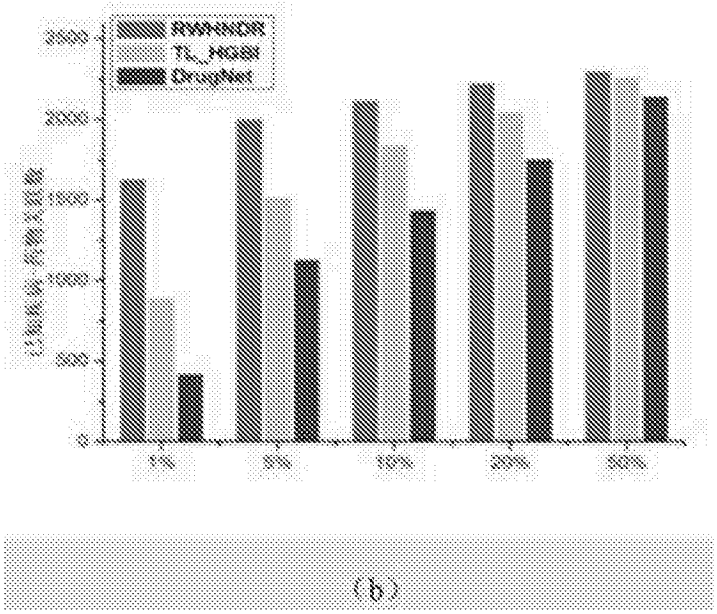



图4

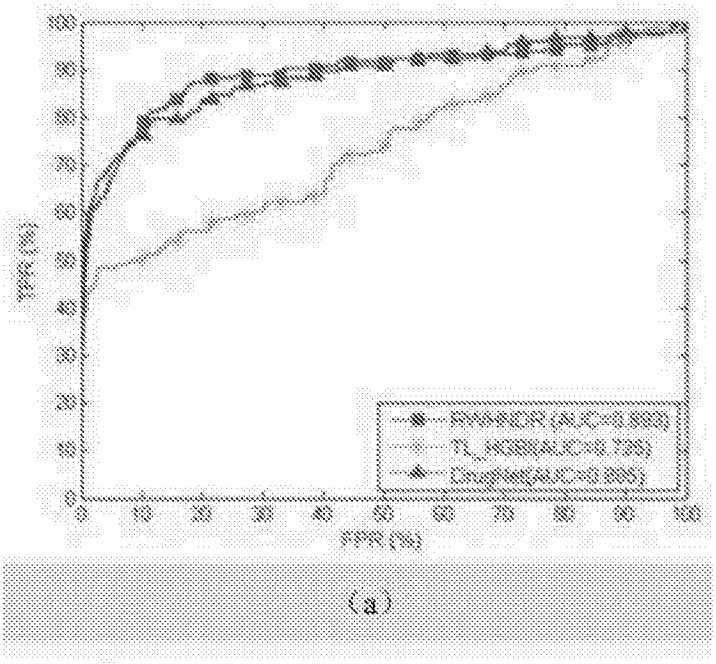

TPR (%)
FPR (%)
0
10
20
30
40
50
60
70
80
90
100

(a)

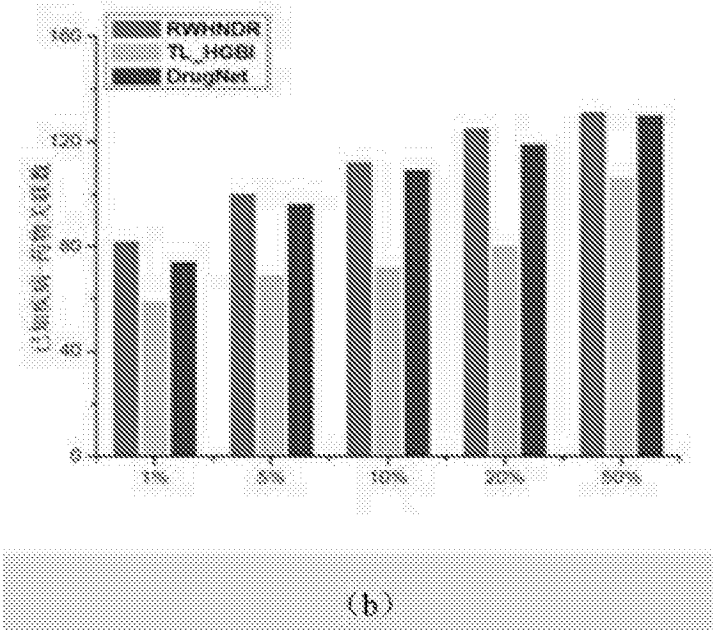

RWHVDR
TL_HGBI
DrugNet
2500
2000
1500
1000
500
0
1%
5%
10%
20%
50%

(b)

图5


TPR (%)
FPR (%)
0
10
20
30
40
50
60
70
80
90
100

(a)


RWNNDR
TL_HGBI
DrugNet
1%
5%
10%
20%
50%

(b)

图6