(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116189760 A
(43)申请公布日 2023.05.30

(21)申请号 202310418206.5

(22)申请日 2023.04.19

(71)申请人 中国人民解放军总医院
地址 100853 北京市海淀区复兴路28号

(72)发明人 王珊 汤永 李顺飞 刘建超 刘丽华

(74)专利代理机构 成都东恒知盛知识产权代理事务所(特殊普通合伙) 51304
专利代理师 何健雄

(51)Int.Cl.
G16B 15/30(2019.01)
G16B 30/10(2019.01)
G16B 40/00(2019.01)
G16C 20/50(2019.01)
G16C 10/00(2019.01)
G16C 20/90(2019.01)
G16H 20/10(2018.01)

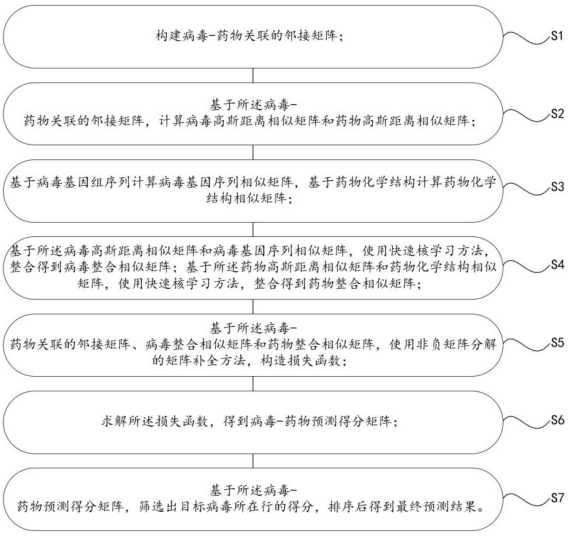
权利要求书4页 说明书11页 附图3页

(54)发明名称

基于矩阵补全的抗病毒药物筛选方法、系统及存储介质

(57)摘要

本发明提供了基于矩阵补全的抗病毒药物筛选方法、系统及存储介质，属于生物信息学、计算生物学与人工智能交叉技术领域，方法通过系统实现，方法包括：S1.构建病毒-药物关联的邻接矩阵；S2.计算病毒高斯距离相似矩阵和药物高斯距离相似矩阵；S3.计算病毒基因序列相似矩阵和药物化学结构相似矩阵；S4.使用快速核学习方法，整合得到病毒整合相似矩阵和药物整合相似矩阵；S5.使用非负矩阵分解的矩阵补全方法，构造损失函数；S6.求解损失函数，得到病毒-药物预测得分矩阵；S7.基于所述病毒-药物预测得分矩阵，筛选、排序后得到最终预测结果。本发明能高效地筛选出病毒有效治疗药物，为特定情况下应急解决方案提供思路。



CN 116189760 A

1.基于矩阵补全的抗病毒药物筛选方法,其特征在于,包括如下步骤:

S1.构建病毒-药物关联的邻接矩阵;

S2.基于所述病毒-药物关联的邻接矩阵,计算病毒高斯距离相似矩阵和药物高斯距离相似矩阵;

S3.基于病毒基因组序列计算病毒基因序列相似矩阵,基于药物化学结构计算药物化学结构相似矩阵;

S4.基于所述病毒高斯距离相似矩阵和病毒基因序列相似矩阵,使用快速核学习方法,整合得到病毒整合相似矩阵;基于所述药物高斯距离相似矩阵和药物化学结构相似矩阵,使用快速核学习方法,整合得到药物整合相似矩阵;

S5.基于所述病毒-药物关联的邻接矩阵、病毒整合相似矩阵和药物整合相似矩阵,使用非负矩阵分解的矩阵补全方法,构造损失函数;

S6.求解所述损失函数,得到病毒-药物预测得分矩阵;

S7.基于所述病毒-药物预测得分矩阵,筛选出目标病毒所在行的得分,排序后得到最终预测结果。

2.根据权利要求1所述的基于矩阵补全的抗病毒药物筛选方法,其特征在于,在S1中:

输入已知的病毒-药物关联对,构建病毒-药物关联的邻接矩阵A;

若为已知关联对,则对应位置为1,否则为0;

所述邻接矩阵A的行数为病毒数量nv,列数为药物数量nd。

3.根据权利要求1所述的基于矩阵补全的抗病毒药物筛选方法,其特征在于,在S2中:

若药物d(i)与某个病毒之间存在关联,则对应位置记为1,否则记为0,形成一个1×nv大小的0或1构成的向量,记之为药物d(i)的向量谱IP(d(i)),nv为病毒数量;然后计算药物d(i)和d(j)之间的高斯距离相似性:

$$S^{d}(d(i),d(j))=\exp(-\gamma_{d}\left\|IP(d(i))-IP(d(j))\right\|^{2});$$

上式中,IP(d(j))为药物d(j)的向量谱;参数γ_{d}用于控制核带宽,通过归一化新带宽参数γ'_{d}获得:

$$\gamma_{d}=\gamma'_{d}/\left(\frac{1}{nd}\sum_{i=1}^{nd}\left\|IP(d(i))\right\|^{2}\right);$$

其中,nd为药物数量;以类似的方式定义病毒v(i)和v(j)之间的高斯距离相似性,得到1×nd大小的0或1构成的向量,记之为病毒v(i)的向量谱IP(v(i)),计算病毒v(i)和v(j)之间的高斯距离相似性:

$$S^{v}(v(i),v(j))=\exp(-\gamma_{v}\left\|IP(v(i))-IP(v(j))\right\|^{2});$$

参数γ_{v}用于控制核带宽,通过归一化新带宽参数γ'_{v}获得:

$$\gamma_{v}=\gamma'_{v}/\left(\frac{1}{nv}\sum_{i=1}^{nv}\left\|IP(v(i))\right\|^{2}\right);$$

以上γ'_{d}和γ'_{v}都是常数;IP(v(j))为病毒v(j)的向量谱。

4.根据权利要求1所述的基于矩阵补全的抗病毒药物筛选方法,其特征在于,在S3中:

基于病毒基因组序列，使用多序列比方法计算病毒基因序列相似矩阵；

基于药物的化学结构，得到药物MACCS指纹，采用谷本系数计算药物化学结构相似矩阵。

5.根据权利要求1所述的基于矩阵补全的抗病毒药物筛选方法，其特征在于，在S4中：

所述快速核学习方法的半正定规划式为：

$$\min_{\lambda^v}\left\|\sum_{j=1}^{2}\lambda_j^v S_j^v - AA^T\right\|_F^2 + \mu^v\left\|\lambda^v\right\|^2;$$

式中，第一项为重构损失范数项，表示相似矩阵的整合误差大小；第二项为正则化项，作用是避免过拟合；其中A为病毒-药物关联邻接矩阵，S_j^v(j=1,2)分别表示病毒高斯距离相似矩阵和病毒基因序列相似矩阵，μ^v为正则化参数，$\lambda^v\in R^{1\times 2}$为待求解的系数，通过$\lambda^v$得到病毒整合相似矩阵$S_v$：

$$S_v=\sum_{j=1}^{2}\lambda_j^v S_j^v;$$

同理，按照上述可获得药物化学结构相似矩阵与药物高斯距离相似矩阵集成参数$\lambda^d\in R^{1\times 2}$，然后计算药物整合相似矩阵$S_d$：

$$S_d=\sum_{j=1}^{2}\lambda_j^d S_j^d;$$

其中S_j^d(j=1,2)分别表示药物高斯距离相似矩阵和药物化学结构相似矩阵。

6.根据权利要求1所述的基于矩阵补全的抗病毒药物筛选方法，其特征在于，在S5中：

使用非负矩阵分解的矩阵补全方法构造的损失函数如下：

$$\min_{W,H,C}\|A-WH+(I-A)\odot C\|_F^2+\lambda_C\|C\|_F^2$$

$$+\lambda_v\sum_{i,j=1}^{nv}\|(WH)^i-(WH)^j\|^2S_v(i,j)\quad;$$

$$+\lambda_d\sum_{p,q=1}^{nd}\|(WH)_p-(WH)_q\|^2S_d(p,q)$$

$$s.t.\ \ W\geqslant 0,\ H\geqslant 0,\ C\geqslant 0;$$

式中$\|A-WH+(I-A)\odot C\|_F^2$部分是重建损失项，⊙表示两个矩阵的Hadamard乘积；其中$A\in R^{nv\times nd}$是已知的病毒-药物关联的邻接矩阵，nv和nd分别为病毒的数量和药物的数量；矩阵I为全1矩阵，C表示矩阵A待补全部分，W和H为待求解的基矩阵和系数矩阵，$\lambda_c\|C\|_F^2$部分是F范数项，约束矩阵C防止过拟合；其余部分是流形约束项，$(WH)^i$和$(WH)^j$分别代表WH的第i和j行，$(WH)_p$和$(WH)_q$分别代表WH的第p和q列；$S_v(i,j)$表示病毒整合相似矩阵的第(i,

j)个元素，$S_d(p,q)$表示病毒整合相似矩阵的第(p,q)个元素；λ_c、λ_v和λ_d是正则化参数。

7.根据权利要求6所述的基于矩阵补全的抗病毒药物筛选方法，其特征在于，在S6中：

求解所述构造损失函数时，分别固定其中一个变量，然后求导数令之为0再反解，得矩阵P、Q、C、W和H的迭代求解公式，具体如下：

$P=HH^T$；

$Q=WH$；

$C^*=(I-A)\odot C$；

$$W_{ik}=W_{ik}\frac{\left(\left(A+C^*\right)H^T+\lambda_d S_d WP+\lambda_v WHS_v H^T\right)_{ik}}{\left(WP+\lambda_d D_d WP+\lambda_v WHD_v H^T\right)_{ik}}$$

$$H_{jk}=H_{jk}\frac{\left(W^T\left(A+C^*\right)+\lambda_d W^T S_d Q+\lambda_v W^T QS_v\right)_{jk}}{\left(W^T Q+\lambda_d W^T D_d Q+\lambda_v W^T QD_v\right)_{jk}}$$；

$$C(i,j)=\frac{-\tilde{Y}(i,j)Y_{wh}(i,j)}{\tilde{Y}(i,j)^2+\lambda_c}$$

其中$A\in R^{nv\times nd}$是已知的病毒-药物关联的邻接矩阵，矩阵I为全1矩阵，C表示矩阵A待补全部分，W和H为待求解的基矩阵和系数矩阵，W_{ik}、H_{jk}分别代表矩阵W的第(i,k)个、矩阵H的第(j,k)个元素；λ_c、λ_v和λ_d是正则化参数；$\odot$表示两个矩阵的Hadamard乘积；$Y_{wh}=A-WH$，$\tilde{Y}=I-A$，D_d或D_v为对角矩阵，其元素为矩阵S_d或S_v按列求和再对角化；更新以上矩阵直到收敛。

8.基于矩阵补全的抗病毒药物筛选系统，其特征在于，包括：

邻接矩阵构建模块，用于构建病毒-药物关联的邻接矩阵；

高斯距离相似矩阵计算模块，用于基于所述病毒-药物关联的邻接矩阵，计算病毒高斯距离相似矩阵和药物高斯距离相似矩阵；

病毒基因序列相似矩阵与药物化学结构相似矩阵计算模块，用于基于病毒基因组序列计算病毒基因序列相似矩阵，基于药物化学结构计算药物化学结构相似矩阵；

整合相似矩阵计算模块，用于基于所述病毒高斯距离相似矩阵和病毒基因序列相似矩阵，使用快速核学习方法，整合得到病毒整合相似矩阵；基于所述药物高斯距离相似矩阵和药物化学结构相似矩阵，使用快速核学习方法，整合得到药物整合相似矩阵；

损失函数构造模块，用于基于所述病毒-药物关联的邻接矩阵、病毒整合相似矩阵和药物整合相似矩阵，使用非负矩阵分解的矩阵补全方法，构造损失函数；

损失函数求解模块，用于求解所述损失函数，得到病毒-药物预测得分矩阵；

预测模块，用于基于所述病毒-药物预测得分矩阵，筛选出目标病毒所在行的得分，排序后得到最终预测结果。

9.根据权利要求8所述的基于矩阵补全的抗病毒药物筛选系统，其特征在于，还包括：

处理器，分别与所述邻接矩阵构建模块、高斯距离相似矩阵计算模块、病毒基因序列相

似矩阵与药物化学结构相似矩阵计算模块、整合相似矩阵计算模块、损失函数构造模块、损失函数求解模块和预测模块连接；

存储器，与所述处理器连接，并存储有可在所述处理器上运行的计算机程序；

其中，当所述处理器执行所述计算机程序时，所述处理器控制所述邻接矩阵构建模块、高斯距离相似矩阵计算模块、病毒基因序列相似矩阵与药物化学结构相似矩阵计算模块、整合相似矩阵计算模块、损失函数构造模块、损失函数求解模块和预测模块工作，以实现如权利要求1～7中任意一项所述的基于矩阵补全的抗病毒药物筛选方法。

10.一种计算机可读存储介质，其特征在于，所述存储介质存储计算机指令，当计算机读取所述计算机指令时，所述计算机执行如权利要求1～7中任意一项所述的基于矩阵补全的抗病毒药物筛选方法。

基于矩阵补全的抗病毒药物筛选方法、系统及存储介质

技术领域

[0001] 本发明涉及生物信息学、计算生物学与人工智能交叉的技术领域，尤其是涉及基于矩阵补全的抗病毒药物筛选方法、系统及存储介质。

背景技术

[0002] 按常规方法研发药物可能需要耗时十多年、耗资数十亿美元，在短时间内开发出一种有效抗病毒药物是极为困难的。考虑到已成熟的药品，其有效性、安全性和毒性都是经过测试的，于是“老药新用”，从已经应用的药品中寻找有效方案是应对突发疫情的一种高效解决方法。

[0003] 抗病毒药物筛选方法已有报道，其中一类是基于结构的虚拟筛选方法，如使用动力学模拟技术，计算潜在药物和靶标间的结合能力，通过分子动力学模拟计算药物的吸收、分布、代谢、排泄和毒性等。此类方法通常存在模拟过程复杂、对使用者经验要求高等不足。国防科技大学天河超算团队提出了基于自由能微扰-绝对结合自由能方法的新冠药物虚拟筛选技术，但这种基于自由能的大规模筛选，对算力要求较高，需要借助超级计算机平台，且耗时以周计算。

发明内容

[0004] 本发明提供基于矩阵补全的抗病毒药物筛选方法、系统及存储介质，可以根据病毒-药物关联数据，准确高效地预测抗病毒相关药物。

[0005] 本说明书实施例的第一方面公开了基于矩阵补全的抗病毒药物筛选方法，包括如下步骤：

S1.构建病毒-药物关联的邻接矩阵；

S2.基于所述病毒-药物关联的邻接矩阵，计算病毒高斯距离相似矩阵和药物高斯距离相似矩阵；

S3.基于病毒基因组序列计算病毒基因序列相似矩阵，基于药物化学结构计算药物化学结构相似矩阵；

S4.基于所述病毒高斯距离相似矩阵和病毒基因序列相似矩阵，使用快速核学习方法，整合得到病毒整合相似矩阵；基于所述药物高斯距离相似矩阵和药物化学结构相似矩阵，使用快速核学习方法，整合得到药物整合相似矩阵；

S5.基于所述病毒-药物关联的邻接矩阵、病毒整合相似矩阵和药物整合相似矩阵，使用非负矩阵分解的矩阵补全方法，构造损失函数；

S6.求解所述损失函数，得到病毒-药物预测得分矩阵；

S7.基于所述病毒-药物预测得分矩阵，筛选出目标病毒所在行的得分，排序后得到最终预测结果。

[0006] 在本说明书公开的实施例中，在S1中：

输入已知的病毒-药物关联对，构建病毒-药物关联的邻接矩阵A；

若为已知关联对，则对应位置为1，否则为0；

所述邻接矩阵A的行数为病毒数量nv，列数为药物数量nd。

[0007] 在本说明书公开的实施例中，在S2中：

若药物d(i)与某个病毒之间存在关联，则对应位置记为1，否则记为0，形成一个1×nv大小的0或1构成的向量，记之为药物d(i)的向量谱IP(d(i))，然后计算药物d(i)和d(j)之间的高斯距离相似性：

$$S^d(d(i),d(j))=\exp(-\gamma_d\|IP(d(i))-IP(d(j))\|^2)；$$

上式中，参数γ_d用于控制核带宽，通过归一化新带宽参数γ'_d获得：

$$\gamma_d=\gamma'_d/\left(\frac{1}{nd}\sum_{i=1}^{nd}\|IP(d(i))\|^2\right)；$$

以类似的方式定义病毒v(i)和v(j)之间的高斯距离相似性，得到1×nd大小的0或1构成的向量，记之为病毒v(i)的向量谱IP(v(i))，计算病毒v(i)和v(j)之间的高斯距离相似性：

$$S^v(v(i),v(j))=\exp(-\gamma_v\|IP(v(i))-IP(v(j))\|^2)；$$

参数γ_v用于控制核带宽，通过归一化新带宽参数γ'_v获得：

$$\gamma_v=\gamma'_v/\left(\frac{1}{nv}\sum_{i=1}^{nv}\|IP(v(i))\|^2\right)；$$

以上γ'_d和γ'_v都是常数。

[0008] 在本说明书公开的实施例中，在S3中：

基于病毒基因组序列，使用多序列比方法计算病毒基因序列相似矩阵；

基于药物的化学结构，得到药物MACCS指纹，采用谷本系数(Tanimoto Coefficient，即Jaccard相似度)计算药物化学结构相似矩阵。

[0009] 在本说明书公开的实施例中，在S4中：

所述快速核学习方法的半正定规划式为：

$$\min_{\lambda^v}\left\|\sum_{j=1}^{2}\lambda_j^vS_j^v-AA^T\right\|_F^2+\mu^v\|\lambda^v\|^2；$$

式中，第一项为重构损失范数项，表示相似矩阵的整合误差大小；第二项为正则化项，作用是避免过拟合；其中A为病毒-药物关联邻接矩阵，S_j^v(j=1,2)分别表示病毒高斯距离相似矩阵和病毒基因序列相似矩阵，μ^v为正则化参数，$\lambda^v\in R^{1\times2}$为待求解的系数，通过$\lambda^v$得到病毒整合相似矩阵：

$$S_v=\sum_{j=1}^{2}\lambda_j^vS_j^v；$$

同理，按照上述可获得药物化学结构相似矩阵与药物高斯距离相似矩阵集成参数$\lambda^d\in R^{1\times2}$，然后计算药物整合相似矩阵：

$$S_d = \sum_{j=1}^{2} \lambda_j^d S_j^d ;$$

其中S_j^d(j=1,2)分别表示药物高斯距离相似矩阵和药物化学结构相似矩阵。

[0010] 在本说明书公开的实施例中,在S5中:

使用非负矩阵分解的矩阵补全方法构造的损失函数如下:

$$\min_{W,H,C} \| A - WH + (I - A) \odot C \|_F^2 + \lambda_C \|C\|_F^2$$

$$+\lambda_v \sum_{i,j=1}^{nv} \|(WH)^i - (WH)^j\|^2 S_v(i,j)$$

;

$$+\lambda_d \sum_{p,q=1}^{nd} \|(WH)_p - (WH)_q\|^2 S_d(p,q)$$

s.t. $W \geqslant 0$, $H \geqslant 0$, $C \geqslant 0$;

式中 $\|A - WH + (I - A) \odot C\|_F^2$ 部分是重建损失项,其中$A \in R^{nv \times nd}$是已知的病毒-药物关联的邻接矩阵,nv和nd分别为病毒的数量和药物的数量;矩阵I为全1矩阵,C表示矩阵A待补全部分,W和H为待求解的基矩阵和系数矩阵,$\lambda_c \|C\|_F^2$部分是F范数项,约束矩阵C防止过拟合;其余部分是流形约束项,$(WH)^i$和$(WH)^j$分别代表WH的第i和j行,$(WH)_p$和$(WH)_q$分别代表WH的第p和q列;$S_v(i,j)$表示病毒整合相似矩阵的第(i,j)个元素,$S_d(p,q)$表示病毒整合相似矩阵的第(p,q)个元素;λ_c、λ_v和λ_d是正则化参数。

[0011] 在本说明书公开的实施例中,在S6中:

求解所述构造损失函数时,分别固定其中一个变量,然后求导数令之为0再反解,得矩阵P、Q、C、W和H的迭代求解公式,具体如下:

$P=HH^T$;

$Q=WH$;

$C^*=(I-A) \odot C$;

$$W_{ik} = W_{ik} \frac{\left(\left(A + C^* \right) H^T + \lambda_d S_d W P + \lambda_v W H S_v H^T \right)_{ik}}{\left(W P + \lambda_d D_d W P + \lambda_v W H D_v H^T \right)_{ik}} ;$$

$$H_{jk} = H_{jk} \frac{\left(W^T \left(A + C^* \right) + \lambda_d W^T S_d Q + \lambda_v W^T Q S_v \right)_{jk}}{\left(W^T Q + \lambda_d W^T D_d Q + \lambda_v W^T Q D_v \right)_{jk}} ;$$

$$C(i,j) = \frac{-\tilde{Y}(i,j) Y_{wh}(i,j)}{\tilde{Y}(i,j)^2 + \lambda_c} ;$$

其中⊙表示两个矩阵的Hadamard乘积；$Y_{wh}=A-WH$，$\tilde{Y}=I-A$，D_d或D_v为对角矩阵，其元素为矩阵S_d或S_v按列求和再对角化；更新以上矩阵直到收敛。

[0012] 本发明实施例的第二方面公开了基于矩阵补全的抗病毒药物筛选系统，包括：

邻接矩阵构建模块，用于构建病毒-药物关联的邻接矩阵；

高斯距离相似矩阵计算模块，用于基于所述病毒-药物关联的邻接矩阵，计算病毒高斯距离相似矩阵和药物高斯距离相似矩阵；

病毒基因序列相似矩阵与药物化学结构相似矩阵计算模块，用于基于病毒基因组序列计算病毒基因序列相似矩阵，基于药物化学结构计算药物化学结构相似矩阵；

整合相似矩阵计算模块，用于基于所述病毒高斯距离相似矩阵和病毒基因序列相似矩阵，使用快速核学习方法，整合得到病毒整合相似矩阵；基于所述药物高斯距离相似矩阵和药物化学结构相似矩阵，使用快速核学习方法，整合得到药物整合相似矩阵；

损失函数构造模块，用于基于所述病毒-药物关联的邻接矩阵、病毒整合相似矩阵和药物整合相似矩阵，使用非负矩阵分解的矩阵补全方法，构造损失函数；

损失函数求解模块，用于求解所述损失函数，得到病毒-药物预测得分矩阵；

预测模块，用于基于所述病毒-药物预测得分矩阵，筛选出目标病毒所在行的得分，排序后得到最终预测结果。

[0013] 在本说明书公开的实施例中，所述基于矩阵补全的抗病毒药物筛选系统还包括：

处理器，分别与所述邻接矩阵构建模块、高斯距离相似矩阵计算模块、病毒基因序列相似矩阵与药物化学结构相似矩阵计算模块、整合相似矩阵计算模块、损失函数构造模块、损失函数求解模块和预测模块连接；

存储器，与所述处理器连接，并存储有可在所述处理器上运行的计算机程序；

其中，当所述处理器执行所述计算机程序时，所述处理器控制所述邻接矩阵构建模块、高斯距离相似矩阵计算模块、病毒基因序列相似矩阵与药物化学结构相似矩阵计算模块、整合相似矩阵计算模块、损失函数构造模块、损失函数求解模块和预测模块工作，以实现上述中任意一项所述的基于矩阵补全的抗病毒药物筛选方法。

[0014] 本发明实施例的第三方面公开了一种计算机可读存储介质，所述存储介质存储计算机指令，当计算机读取所述计算机指令时，所述计算机执行上述中任意一项所述的基于矩阵补全的抗病毒药物筛选方法。

[0015] 综上所述，本发明至少具有以下有益效果：

本发明构通过构建病毒-药物关联的邻接矩阵，分别计算病毒高斯距离相似矩阵和药物高斯距离相似矩阵；使用病毒基因组序列计算病毒基因序列相似矩阵，使用药物的化学结构信息计算药物化学结构相似矩阵；使用快速核学习法计算病毒整合相似矩阵、药物整合相似矩阵；结合非负矩阵分解、矩阵补全与图正则化方法构建损失函数，迭代求解得到病毒-药物关联预测得分矩阵，筛选、排序得到最终结果。本发明能快速、高效地筛选出病毒有效治疗药物，弥补生物医学实验方法耗时长、成本高的不足，为特定情况下应急解决方案提供了思路。

附图说明

[0016] 为了更清楚地说明本发明实施例技术方案，下面将对实施例描述中所需要使用的

附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0017] 图1为本发明中所涉及的基于矩阵补全的抗病毒药物筛选方法的步骤示意图。

[0018] 图2为本发明中所涉及的基于矩阵补全的抗病毒药物筛选方法的流程示意图。

[0019] 图3为本发明中所涉及的基于矩阵补全的抗病毒药物筛选方法与基线方法五折交叉验证的结果比较图。

[0020] 图4为本发明中所涉及的基于矩阵补全的抗病毒药物筛选系统的示意图。

具体实施方式

[0021] 在下文中，仅简单地描述了某些示例性实施例。正如本领域技术人员可认识到的那样，在不脱离本发明实施例的精神或范围的情况下，可通过各种不同方式修改所描述的实施例。因此，附图和描述被认为本质上是示例性的而非限制性的。

[0022] 下文的公开提供了许多不同的实施方式或例子用来实现本发明实施例的不同结构。为了简化本发明实施例的公开，下文中对特定例子的部件和设置进行描述。当然，它们仅仅为示例，并且目的不在于限制本发明实施例。此外，本发明实施例可以在不同例子中重复参考数字和/或参考字母，这种重复是为了简化和清楚的目的，其本身不指示所讨论各种实施方式和/或设置之间的关系。

[0023] 下面结合附图对本发明的实施例进行详细说明。

[0024] 需要注意的是，本说明书的实施例中所使用的已知人类药物-病毒关联数据是从有关文献中收集的，先使用文本挖掘技术对文献报道的经过实验验证的药物-病毒相互作用对进行整理后，获得455个已证实的人类病毒-药物相互作用，涉及34种病毒与219种药物(文献DOI:10.1016/j.asoc.2021.107135)；药物化学结构从DrugBank数据库下载，病毒基因组核苷酸序列从美国国家生物技术信息中心NCBI数据库获得。

[0025] 如图1和图2所示，本说明书实施例的第一方面公开了基于矩阵补全的抗病毒药物筛选方法，包括如下步骤：

S1.构建病毒-药物关联的邻接矩阵。

[0026] 输入已知的病毒-药物关联对，构建病毒-药物关联的邻接矩阵A；

$$\begin{cases} A(v(i),d(j))=1 & \text{病毒}v(i)\text{与药物}d(j)\text{存在关联} \\ A(v(i),d(j))=0 & \text{病毒}v(i)\text{与药物}d(j)\text{尚未发现关联} \end{cases};$$

得到的邻接矩阵A元素为0或1，大小为34行×219列，i与j的取值范围满足1≤i≤34，1≤j≤219。

[0027] S2.基于病毒-药物关联的邻接矩阵，计算病毒高斯距离相似矩阵和药物高斯距离相似矩阵。

[0028] 若药物d(i)与某个病毒之间存在关联，则对应位置记为1，否则记为0，形成一个1×34大小的0或1构成的向量，记之为药物d(i)的向量谱IP(d(i))，然后计算药物d(i)和d(j)之间的高斯距离相似性：

$$S^{d}(d(i),d(j))=\exp(-\gamma_{d}\left\|IP(d(i))-IP(d(j))\right\|^{2});$$

上式中，IP(d(j))为药物d(j)的向量谱；参数γ_{d}用于控制核带宽，通过归一化新带宽参数γ'_{d}获得：

$$\gamma_{d}=\gamma'_{d}/\left(\frac{1}{nd}\sum_{i=1}^{nd}\left\|IP(d(i))\right\|^{2}\right);$$

以类似的方式定义病毒v(i)和v(j)之间的高斯距离相似性，若某一个病毒v(i)与某药物之间存在关联，则对应位置记为1，否则记为0，形成一个1×219大小的0或1构成的向量，记之为病毒v(i)的向量谱IP(v(i))，然后计算病毒v(i)和v(j)之间的高斯距离相似性：

$$S^{v}(v(i),v(j))=\exp(-\gamma_{v}\left\|IP(v(i))-IP(v(j))\right\|^{2});$$

上式中，参数γ_{v}用于控制核带宽，通过归一化新带宽参数γ'_{v}获得：

$$\gamma_{v}=\gamma'_{v}/\left(\frac{1}{nv}\sum_{i=1}^{nv}\left\|IP(v(i))\right\|^{2}\right);$$

以上γ'_{d}和γ'_{v}都是常数，取$\gamma'_{d}=\gamma'_{v}=1$；IP(v(j))为病毒v(j)的向量谱。

[0029] 其中nv表示病毒的数量，此例中为34，nd表示药物的数量，此例中为219，此步计算后得到大小为34×34的对称矩阵S_{1}^{v}(病毒高斯距离相似矩阵)和大小为219×219的对称矩阵S_{1}^{d}(药物高斯距离相似矩阵)，且这两个矩阵元素值全都在0到1之间。

[0030] S3.基于病毒基因组序列计算病毒基因序列相似矩阵，基于药物化学结构计算药物化学结构相似矩阵。

[0031] 输入病毒基因组序列，使用多序列比对工具MAFFT计算得到病毒基因序列相似矩阵S_{2}^{v}；输入SMILES编码表示的药物化学结构，然后用化学信息学软件RDKit或Open Babel获得药物的分子访问系统指纹(MACCS)，再使用R包RxnSim计算Tanimoto相似度，得到药物化学结构相似矩阵S_{2}^{d}，具体计算方法是，对d(i)和d(j)两种药物，将此两种药物的MACCS片段二进制表示的字符串集分别记为D(i)和D(j)，d(i)和d(j)间的相似度S^{d}_{ij}值可以用下面公式计算：

$$S_{ij}^{d}=\frac{D(i)\cap D(j)}{D(i)\cup D(j)}。$$

[0032] S4.基于病毒高斯距离相似矩阵和病毒基因序列相似矩阵，使用快速核学习方法，整合得到病毒整合相似矩阵；基于药物高斯距离相似矩阵和药物化学结构相似矩阵，使用快速核学习方法，整合得到药物整合相似矩阵。

[0033] 使用快速核学习方法整合病毒基因序列相似矩阵和病毒高斯距离相似矩阵，具体是通过求解下面的半正定规划式：

$$\min_{\lambda^{v}}\left\|\sum_{j=1}^{2}\lambda_{j}^{v}S_{j}^{v}-AA^{T}\right\|_{F}^{2}+\mu^{v}\left\|\lambda^{v}\right\|^{2};$$

式中，第一项为重构损失范数项，表示相似矩阵的整合误差大小；第二项为正则化

项，作用是避免过拟合；其中A为病毒-药物关联邻接矩阵，S_j^v(j=1,2)分别表示病毒高斯距离相似矩阵和病毒基因序列相似矩阵，μ^v为正则化参数，$\lambda^v \in R^{1\times 2}$为待求解的系数，使用Matlab软件中的CVX工具箱求解得到病毒整合相似矩阵：

$$S_v = \sum_{j=1}^{2} \lambda_j^v S_j^v ;$$

同理，按照上述可获得药物化学结构相似矩阵与药物高斯距离相似矩阵集成参数$\lambda^d \in R^{1\times 2}$，然后计算药物整合相似矩阵：

$$S_d = \sum_{j=1}^{2} \lambda_j^d S_j^d ;$$

其中S_j^d(j=1,2)分别表示药物高斯距离相似矩阵和药物化学结构相似矩阵。

[0034] S5.基于病毒-药物关联的邻接矩阵、病毒整合相似矩阵和药物整合相似矩阵，使用非负矩阵分解的矩阵补全方法，构造损失函数。

[0035] 使用非负矩阵分解的矩阵补全方法构造的损失函数如下：

$$\min_{W,H,C} \|A - WH + (I - A) \odot C\|_F^2 + \lambda_c \|C\|_F^2$$

$$+\lambda_v \sum_{i,j=1}^{nv} \|(WH)^i - (WH)^j\|^2 S_v(i,j) \quad ;$$

$$+\lambda_d \sum_{p,q=1}^{nd} \|(WH)_p - (WH)_q\|^2 S_d(p,q)$$

$$s.t.\ W\geqslant 0,\ H\geqslant 0,\ C\geqslant 0;$$

式中 $\|A - WH + (I - A) \odot C\|_F^2$ 部分是重建损失项，其中$A \in R^{nv\times nd}$是已知的病毒-药物关联的邻接矩阵，nv和nd分别为病毒的数量和药物的数量，即矩阵大小为34行×219列；矩阵I为全1矩阵，C表示矩阵A待补全部分，W和H为待求解的基矩阵和系数矩阵，$\lambda_c \|C\|_F^2$部分是F范数项，约束矩阵C防止过拟合；其余部分是流形约束项，$(WH)^i$和$(WH)^j$分别代表WH的第i和j行，$(WH)_p$和$(WH)_q$分别代表WH的第p和q列；$S_v(i,j)$表示病毒整合相似矩阵的第(i,j)个元素，$S_d(p,q)$表示病毒整合相似矩阵的第(p,q)个元素；λ_c、λ_v和λ_d是正则化参数。

[0036] S6.求解损失函数，得到病毒-药物预测得分矩阵。

[0037] 求解构造损失函数时，使用拉格朗日乘子法，依据Karush-Kuhn-Tucker(KKT)条件分别固定其中一个变量，然后求导数令之为0再反解，得矩阵P、Q、C、W和H的迭代求解公式，具体如下：

$P=HH^T$；

$Q=WH$；

$C^*=(I-A)\odot C$；

$$W_{ik}=W_{ik}\frac{\left(\left(A+C^{*}\right)H^{T}+\lambda_{d}S_{d}WP+\lambda_{v}WHS_{v}H^{T}\right)_{ik}}{\left(WP+\lambda_{d}D_{d}WP+\lambda_{v}WHD_{v}H^{T}\right)_{ik}};$$

$$H_{jk}=H_{jk}\frac{\left(W^{T}\left(A+C^{*}\right)+\lambda_{d}W^{T}S_{d}Q+\lambda_{v}W^{T}QS_{v}\right)_{jk}}{\left(W^{T}Q+\lambda_{d}W^{T}D_{d}Q+\lambda_{v}W^{T}QD_{v}\right)_{jk}};$$

$$C(i,j)=\frac{-\tilde{Y}(i,j)Y_{wh}(i,j)}{\tilde{Y}(i,j)^{2}+\lambda_{c}};$$

其中⊙表示两个矩阵的Hadamard乘积;W_{ik}、H_{jk}分别代表矩阵W的第(i,k)个、矩阵H第(j,k)个元素;$Y_{wh}=A-WH$,$\tilde{Y}=1-A$,D_d或D_v为对角矩阵,其元素为矩阵S_d或S_v按列求和再对角化;更新以上矩阵直到收敛。

[0038] S7.基于病毒-药物预测得分矩阵,计算病毒-药物关联对预测分数$Q_{final}=WH$,并根据病毒-药物关联对预测分数Q_{final},筛选出目标病毒所在行的得分,排序后得到最终预测结果。

[0039] 在上述中,经初步优化后选取正则化参数$\lambda_c=1$、$\lambda_v=\lambda_d=0.0001$;使用Matlab编程实现上述算法时,将矩阵W初始化为34行×100列的随机矩阵,H初始化为100行×219列的随机矩阵,W和H的全部元素都在(0,1)区间范围,矩阵C初始化为0矩阵,大小同矩阵A即34行×219列;设定次数为800时退出迭代,循环运行结束后得到矩阵W、H;计算预测分数矩阵$Q_{final}=WH$,得到最终预测结果,预测结束。

[0040] 本发明的有效性验证:

如图1和图2所示的基于矩阵补全的抗病毒药物筛选方法,采用五重交叉验证进行预测性能评估,具体实施方式为:先将所有已知的药物-病毒关联随机平均分成5组,再将5组中的每一组依次设为测试样本,其他组作为训练样本(测试样本选取情况不同时,依赖测试样本计算所得的高斯距离相似矩阵亦随之改变)。使用训练样本作为本方法的输入得到预测结果,最后将该组中每个测试样本的预测分数与候选样本的分数进行比较。为了减少生成测试样本的过程中随机划分对结果造成的影响,进行了100次五折交叉验证。

[0041] 使用Matlab编程计算后获得了如下数据,如图3所示为本方法NMFMCVDA与现已报道的几种病毒-药物筛选模型之间的AUROC(ROC曲线下面积)值比较。本方法在五折交叉验证中取得了0.8544±0.0047的AUROC值,表现出了比几种经典模型更加出色的预测性能。

[0042] 另外一方面,对具体某种病毒,如新型冠状病毒(SARS-CoV-2)使用本方法来做预测,筛选评分矩阵Q_{final}中SARS-CoV-2对应的行即取得新冠相关药物的预测得分,将其降序排列后中前20个药物有16个能够得到已报道文献的支持。

[0043] 下表展示了预测结果前20个药物名称和支持的文献PMID、引文格式或DOI号。

[0044]

排序	药名	支持文献
1	Chloroquine	PMID:32074550
2	Ribavirin	PMID:22555152
3	Nitazoxanide	Chem.Biol.Lett.,2020,7(3),192-196.

4	Camostat	PMID:22496216
5	N4-Hydroxycytidine	暂未确认
6	Niclosamide	PMID:32125140
7	Remdesivir	PMID:32020029
8	Gemcitabine	PMID:24841273
9	Mycophenolic Acid	PMID:5799033
10	Umifenovir	PMID:32360480
11	Mizoribine	PMID:32886002DOI:10.1152/ajpheart.00506.2020.
12	Alisporivir	PMID:32376613
13	Memantine	PMID:32554535
14	Betulinic Acid	暂未确认
15	Disulfiram	暂未确认
16	6-Azauridine	暂未确认
17	Lopinavir	PMID:32251767
18	Hydroxychloroquine	PMID:32150618
19	Tacrolimus	PMID:33495742
20	Amodiaquine	PMID:32246834,32834612,32631083,32317408

综上,本发明的优点:

1、通过引入范数约束项能有效防止过拟合,减轻训练数据集中存在的内在噪声的影响,进而使得病毒-药物关联预测结果更具有鲁棒性、更准确;

2、通过将重构损失项划分为已知与待补全两部分构建,本方法具有较好的可扩展性与健壮性,能获得较佳预测结果;

3、方法借鉴了流形学习理论,通过融合拉普拉斯项,刻画局部流形结构,能够高效利用阴性样本信息,提升预测性能。

[0045] 如图4所示,本发明实施例的第二方面公开了基于矩阵补全的抗病毒药物筛选系统,包括:

邻接矩阵构建模块,用于构建病毒-药物关联的邻接矩阵;

高斯距离相似矩阵计算模块,用于基于病毒-药物关联的邻接矩阵,计算病毒高斯距离相似矩阵和药物高斯距离相似矩阵;

病毒基因序列相似矩阵与药物化学结构相似矩阵计算模块,用于基于病毒基因组序列计算病毒基因序列相似矩阵,基于药物化学结构计算药物化学结构相似矩阵;

整合相似矩阵计算模块,用于基于病毒高斯距离相似矩阵和病毒基因序列相似矩阵,使用快速核学习方法,整合得到病毒整合相似矩阵;基于药物高斯距离相似矩阵和药物化学结构相似矩阵,使用快速核学习方法,整合得到药物整合相似矩阵;

损失函数构造模块,用于基于病毒-药物关联的邻接矩阵、病毒整合相似矩阵和药物整合相似矩阵,使用非负矩阵分解的矩阵补全方法,构造损失函数;

损失函数求解模块,用于求解损失函数,得到病毒-药物预测得分矩阵;

预测模块,用于基于病毒-药物预测得分矩阵,筛选出目标病毒所在行的得分,排序后得到最终预测结果。

[0046] 在本说明书公开的实施例中,基于矩阵补全的抗病毒药物筛选系统还包括:

处理器，分别与邻接矩阵构建模块、高斯距离相似矩阵计算模块、病毒基因序列相似矩阵与药物化学结构相似矩阵计算模块、整合相似矩阵计算模块、损失函数构造模块、损失函数求解模块和预测模块连接；

存储器，与处理器连接，并存储有可在处理器上运行的计算机程序；

其中，当处理器执行计算机程序时，处理器控制邻接矩阵构建模块、高斯距离相似矩阵计算模块、病毒基因序列相似矩阵与药物化学结构相似矩阵计算模块、整合相似矩阵计算模块、损失函数构造模块、损失函数求解模块和预测模块工作，以实现上述中任意一项的基于矩阵补全的抗病毒药物筛选方法。

[0047] 本发明实施例的第三方面公开了一种计算机可读存储介质，存储介质存储计算机指令，当计算机读取计算机指令时，计算机执行上述中任意一项的基于矩阵补全的抗病毒药物筛选方法。

[0048] 以上所述实施例是用以说明本发明，并非用以限制本发明，所以举例数值的变更或等效元件的置换仍应隶属本发明的范畴。

[0049] 由以上详细说明，可使本领域普通技术人员明了本发明的确可达成前述目的，实已符合专利法的规定。

[0050] 尽管已描述了本发明的优选实施例，但本领域内的技术人员一旦得知了基本创造性概念，则可对这些实施例作出另外的变更和修改。所以，所附权利要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。以上所述仅为本发明的较佳实施例而已，并不用以限制本发明，应当指出的是，凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等，均应包含在本发明的保护范围之内。

[0051] 应当注意的是，上述有关流程的描述仅仅是为了示例和说明，而不限定本说明书的适用范围。对于本领域技术人员来说，在本说明书的指导下可以对流程进行各种修正和改变。然而，这些修正和改变仍在本说明书的范围之内。

[0052] 上文已对基本概念做了描述，显然，对于阅读此申请后的本领域的普通技术人员来说，上述发明披露仅作为示例，并不构成对本申请的限制。虽然此处并未明确说明，但本领域的普通技术人员可能会对本申请进行各种修改、改进和修正。该类修改、改进和修正在本申请中被建议，所以该类修改、改进、修正仍属于本申请示范实施例的精神和范围。

[0053] 同时，本申请使用了特定词语来描述本申请的实施例。例如“一个实施例”、“一实施例”、和/或“一些实施例”意指与本申请至少一个实施例有关的某一特征、结构或特性。因此，应当强调并注意的是，本说明书中在不同位置两次或以上提及的“一实施例”或“一个实施例”或“一替代性实施例”并不一定是指同一实施例。此外，本申请的一个或多个实施例中的某些特征、结构或特点可以进行适当的组合。

[0054] 此外，本领域的普通技术人员可以理解，本申请的各方面可以通过若干具有可专利性的种类或情况进行说明和描述，包括任何新的和有用的过程、机器、产品或物质的组合，或对其任何新的和有用的改进。因此，本申请的各个方面可以完全由硬件实施、可以完全由软件（包括固件、常驻软件、微代码等）实施、也可以由硬件和软件组合实施。以上硬件或软件均可被称为“单元”、“模块”或“系统”。此外，本申请的各方面可以采取体现在一个或多个计算机可读介质中的计算机程序产品的形式，其中计算机可读程序代码包含在其中。

[0055] 本申请各部分操作所需的计算机程序代码可以用任意一种或以上程序设计语言

编写，包括如Java、Scala、Smalltalk、Eiffel、JADE、Emerald、C++、C#、VB.NET、Python等的面向对象程序设计语言、如C程序设计语言、VisualBasic、Fortran2103、Perl、COBOL2102、PHP、ABAP的常规程序化程序设计语言、如Python、Ruby和Groovy的动态程序设计语言或其它程序设计语言等。该程序代码可以完全在用户计算机上运行、或作为独立的软件包在用户计算机上运行、或部分在用户计算机上运行部分在远程计算机运行、或完全在远程计算机或服务器上运行。在后种情况下，远程计算机可以通过任何网络形式与用户计算机连接，比如局域网(LAN)或广域网(WAN)，或连接至外部计算机(例如通过因特网)，或在云计算环境中，或作为服务使用如软件即服务(SaaS)。

[0056] 此外，除非权利要求中明确说明，本申请所述处理元素和序列的顺序、数字字母的使用、或其他名称的使用，并非用于限定本申请流程和方法的顺序。尽管上述披露中通过各种示例讨论了一些目前认为有用的发明实施例，但应当理解的是，该类细节仅起到说明的目的，附加的权利要求并不仅限于披露的实施例，相反，权利要求旨在覆盖所有符合本申请实施例实质和范围的修正和等价组合。例如，尽管上述各种组件的实现可以体现在硬件设备中，但是它也可以实现为纯软件解决方案，例如，在现有服务器或移动设备上的安装。

[0057] 同理，应当注意的是，为了简化本申请披露的表述，从而帮助对一个或多个发明实施例的理解，前文对本申请的实施例的描述中，有时会将多种特征归并至一个实施例、附图或对其的描述中。然而，本申请的该方法不应被解释为反映所申明的客体需要比每个权利要求中明确记载的更多特征的意图。相反，发明的主体应具备比上述单一实施例更少的特征。

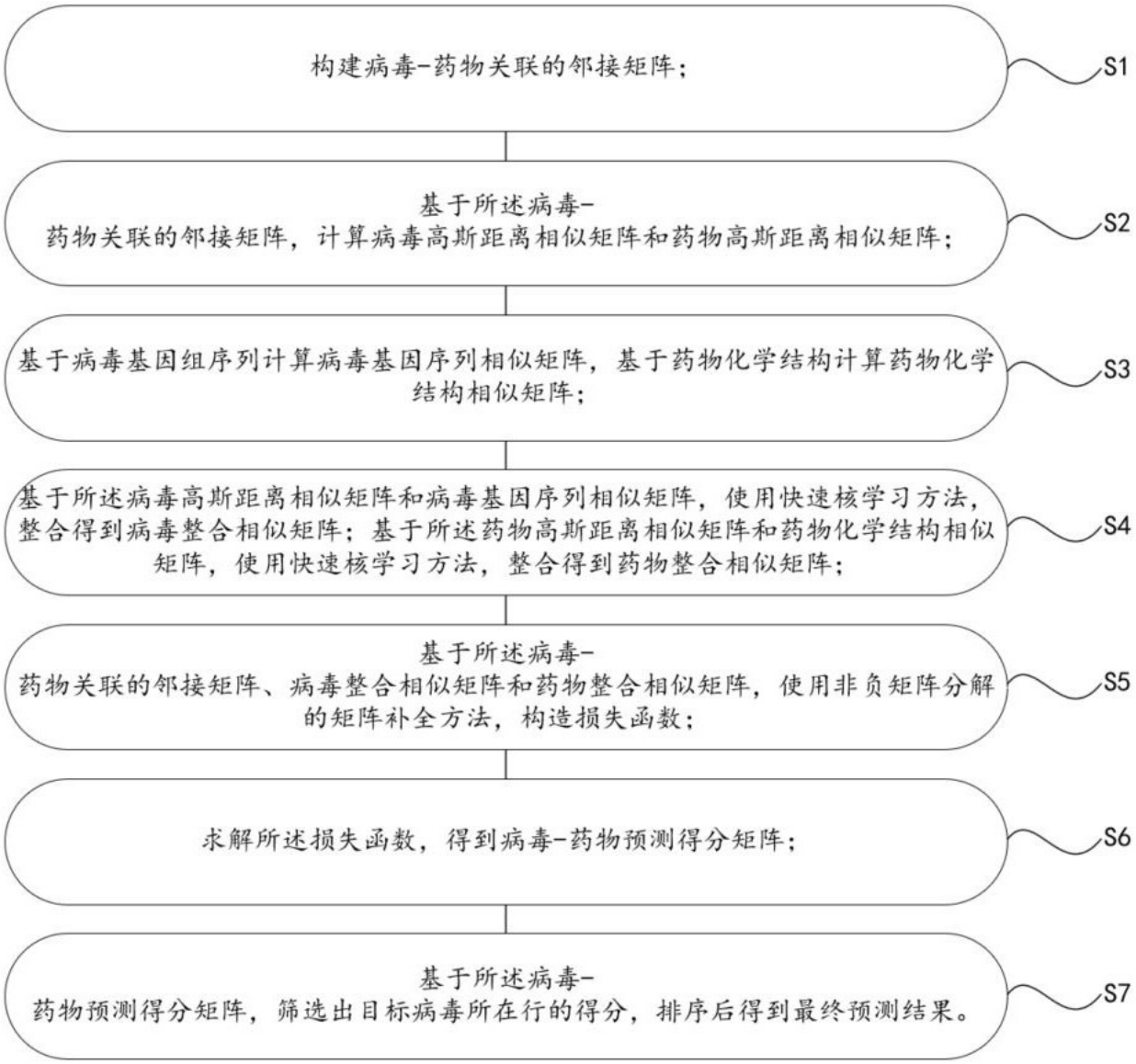

构建病毒-药物关联的邻接矩阵；
S1
基于所述病毒-药物关联的邻接矩阵，计算病毒高斯距离相似矩阵和药物高斯距离相似矩阵；
S2
基于病毒基因组序列计算病毒基因序列相似矩阵，基于药物化学结构计算药物化学结构相似矩阵；
S3
基于所述病毒高斯距离相似矩阵和病毒基因序列相似矩阵，使用快速核学习方法，整合得到病毒整合相似矩阵；基于所述药物高斯距离相似矩阵和药物化学结构相似矩阵，使用快速核学习方法，整合得到药物整合相似矩阵；
S4
基于所述病毒-药物关联的邻接矩阵、病毒整合相似矩阵和药物整合相似矩阵，使用非负矩阵分解的矩阵补全方法，构造损失函数；
S5
求解所述损失函数，得到病毒-药物预测得分矩阵；
S6
基于所述病毒-药物预测得分矩阵，筛选出目标病毒所在行的得分，排序后得到最终预测结果。
S7

图 1

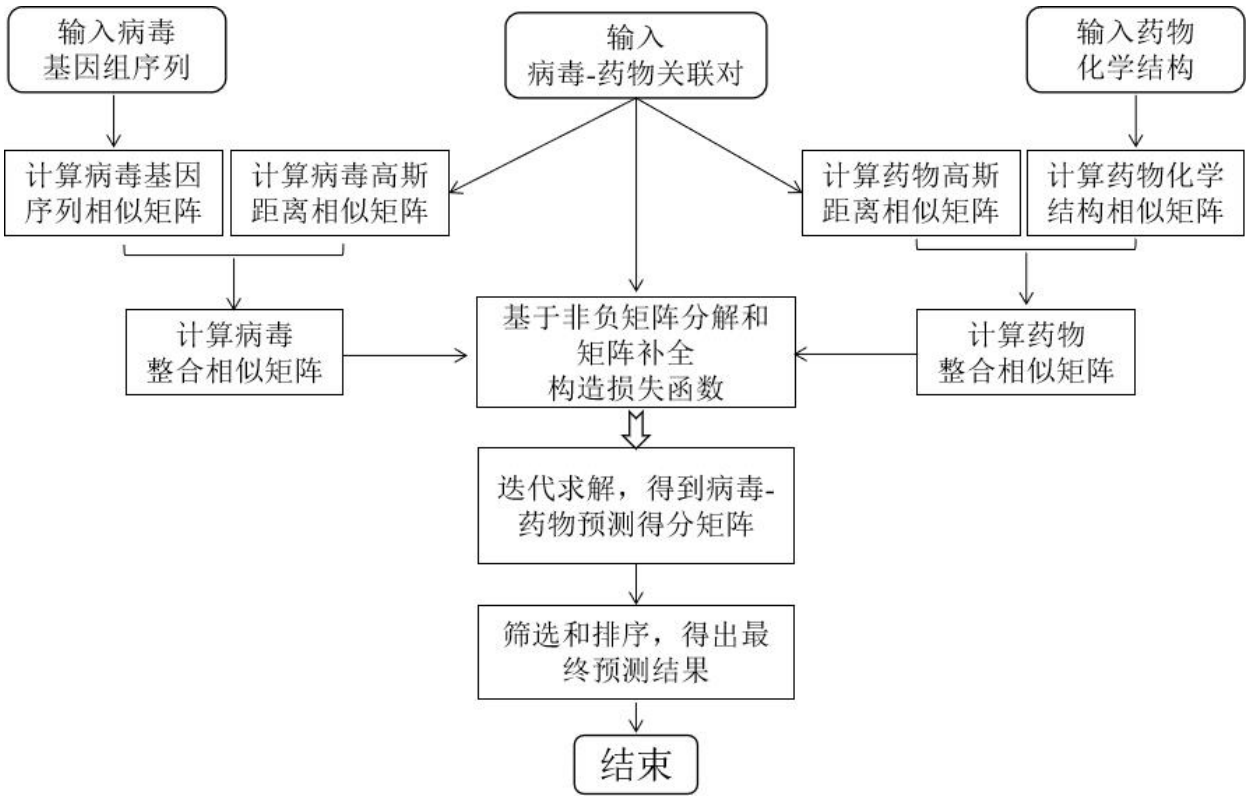

输入病毒
基因组序列
输入
病毒-药物关联对
输入药物
化学结构
计算病毒基因
序列相似矩阵
计算病毒高斯
距离相似矩阵
计算药物高斯
距离相似矩阵
计算药物化学
结构相似矩阵
计算病毒
整合相似矩阵
基于非负矩阵分解和
矩阵补全
构造损失函数
计算药物
整合相似矩阵
迭代求解，得到病毒-
药物预测得分矩阵
筛选和排序，得出最
终预测结果
结束

图 2

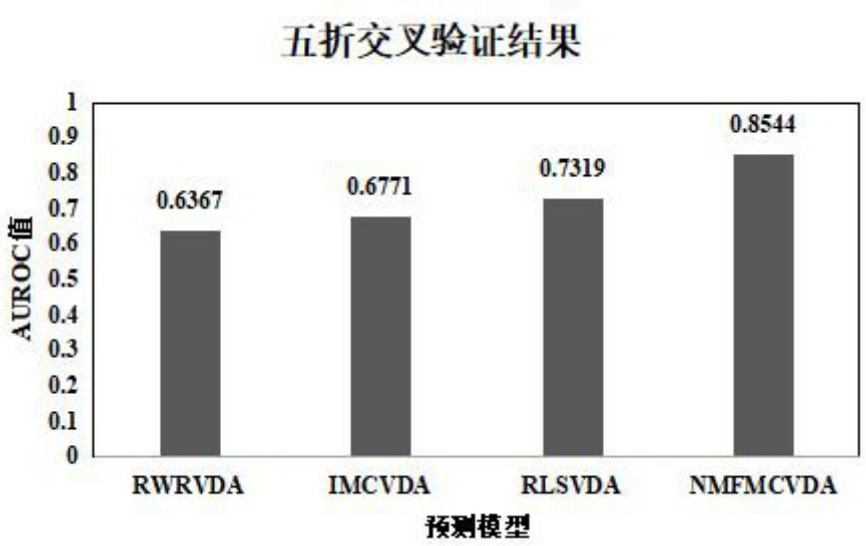

五折交叉验证结果
AUROC值
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.6367
0.6771
0.7319
0.8544
RWRVDA
IMCVDA
RLSVDA
NMFMCVDA
预测模型

图 3

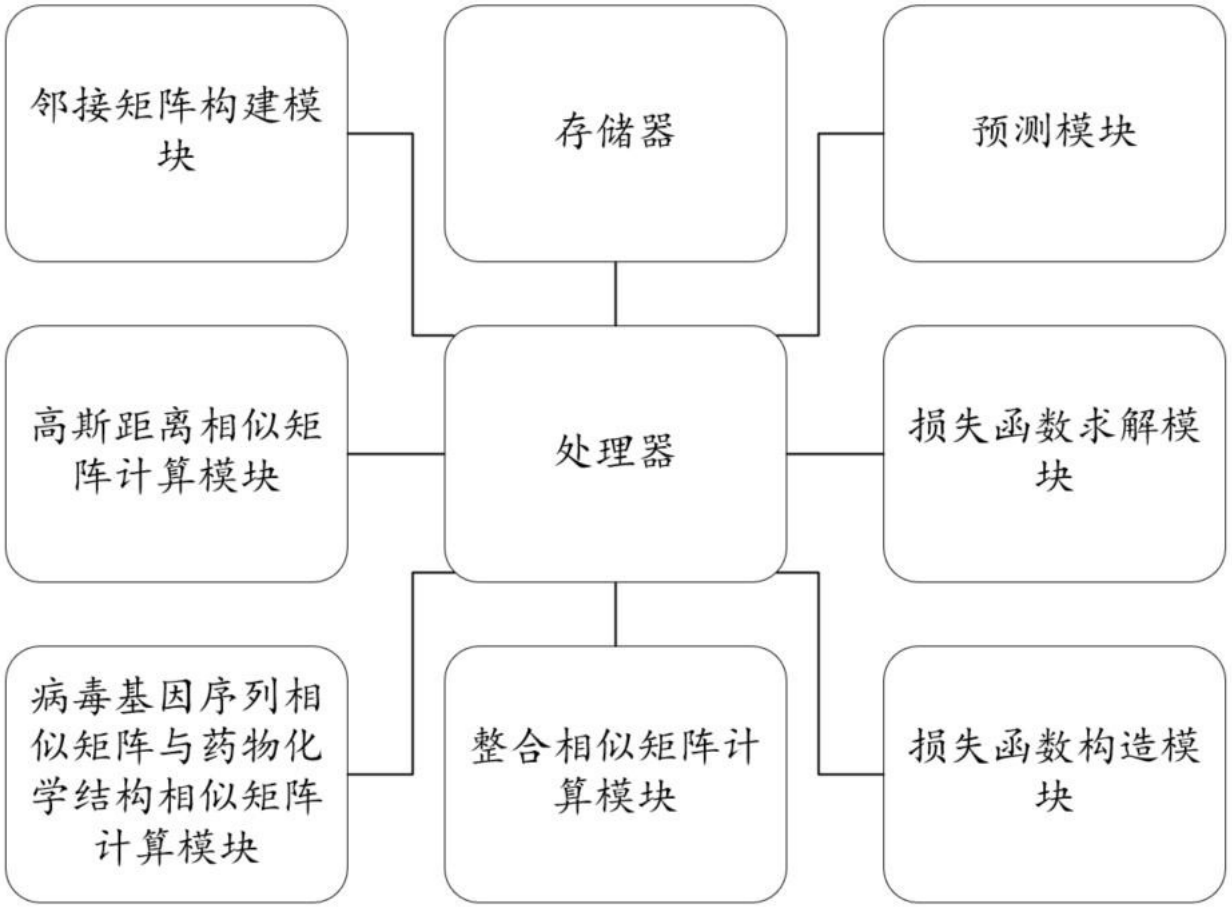

邻接矩阵构建模块
存储器
预测模块
高斯距离相似矩阵计算模块
处理器
损失函数求解模块
病毒基因序列相似矩阵与药物化学结构相似矩阵计算模块
整合相似矩阵计算模块
损失函数构造模块

图 4