



# 基于矩阵填充算法的抗癌药物敏感性预测

徐 慧, 贺平安

(浙江理工大学理学院, 杭州 310018)

**摘 要:** 在已有矩阵填充模型的基础上, 融入癌细胞系的相似性信息与抗癌药物的相似性信息, 提出了一种低秩矩阵填充模型, 用于预测因受试验条件等影响而缺失的抗癌药物敏感性数据。将模型应用在 CGP 数据库中有缺失值的抗癌药物敏感性数据中, 利用已有数据训练新模型的参数, 并利用训练得到的最优参数对 CGP 数据库中的数据进行预测, 得到最优近似值。结果表明, 相较于其他模型, 该模型能有效提高对缺失数据的填充效果。

**关键词:** 低秩矩阵填充; 癌症细胞系; 抗癌药物敏感性; 相似性; 抗癌药物; 预测

中图分类号: O29

文献标志码: A

文章编号: 1673-3851 (2020) 03-0277-06

## Prediction of anticancer drug sensitivity based on matrix completion algorithm

XU Hui, HE Ping'an

(School of sciences, Zhejiang Sci-Tech University, Hangzhou 310018, China)

**Abstract:** The similarity information of cancer cell lines and the similarity information of anticancer drugs were incorporated into the existing matrix completion model. A low rank matrix completion model was proposed to predict the missing data of susceptibility to anticancer drugs due to the influence of experimental conditions and other factors. The model was applied to the data of anticancer drug sensitivity with missing values in CGP database. The parameters of the new model were trained by the existing data. The optimal parameters obtained by training were adopted to predict the data in the CGP database so as to gain the optimal approximate value. The results show that, compared to other models, the model can effectively improve the filling effect of missing data.

**Key words:** low rank matrix completion; cancer cell line; anticancer drug sensitivity; similarity; anticancer drug; prediction

## 0 引 言

肿瘤是威胁人类健康的常见病和多发病, 病死率极高。肿瘤的异质性是指同一种肿瘤在不同患者身上的某些生物特征表现不同, 进而对于相同抗癌药物的反应也会有所不同。因此对不同的肿瘤患者要采用不同的治疗方案, 即肿瘤的个性化医疗。目前的肿瘤药物均会对患者产生一些毒副作用, 因此

通常进行抗肿瘤药物敏感性的体外试验, 根据试验结果选择最佳治疗药物, 减少对病人产生的毒副作用。目前, 功能性的肿瘤药物敏感性体外试验主要有 ATP 含量测定<sup>[1]</sup>、四甲基偶氮唑盐比色法<sup>[2]</sup>、三维立体组织培养法 (Histoculture drug response assay, HDRA)、胶原凝胶包埋培养法 (CD-DST 法)、端粒酶活性法 (Telomerase activity assay)、人源肿瘤组织异种移植 (Patient-derived tumor

收稿日期: 2019-04-17 网络出版日期: 2019-07-01

基金项目: 国家自然科学基金项目 (61772027)

作者简介: 徐 慧 (1994—), 女, 河南驻马店人, 硕士研究生, 主要从事生物信息学方面的研究。

通信作者: 贺平安, E-mail: pinganhe@zstu.edu.cn

xenograft model, PDX)、裸鼠肾包膜下移植法、动态BH3分析(Dynamic BH3 profiling, DBP)等<sup>[3]</sup>。各种抗癌药物敏感性试验对于试验材料、试验条件要求较高,且耗费较大。例如三维立体组织培养法中的检测材料为固体,因而不能检测一些类似胸腹水等液体样本。因此,在药物敏感性数据中经常有很多缺失的数据,这给试验结果分析带来很大的困扰。

数学中的矩阵填充问题是通过真实矩阵观测到的元素来估计矩阵中未观测到的其他元素<sup>[4]</sup>。目前,矩阵填充理论已在图像修复与图像分类<sup>[5]</sup>、人脸识别<sup>[6]</sup>、半监督聚类与推荐系统<sup>[7]</sup>等机器学习、模式识别以及计算机视觉领域得到了应用。2009年,Candès等<sup>[8]</sup>证明,在适当条件(数据具有低秩性)下,通过求解凸优化问题,可以计算得到不完整矩阵的近似矩阵来完成对不完整矩阵的填充。在此之后,一些新方法陆续被提出<sup>[9-10]</sup>。2010年Cai等<sup>[11]</sup>在Keshavan等<sup>[12]</sup>的基础上,对矩阵填充算法进行改进,在不考虑噪音影响的情况下,利用改进算法对血凝素抑制试验中的缺失数据进行了预测。黄莉等<sup>[13]</sup>又对Cai等<sup>[11]</sup>提出的算法进行改进,把模型应用在癌细胞系百科全书(CCLE)数据库中有部分数据缺失的抗癌药物敏感性数据中,结果显示误差降低了22.7%。

本文在黄莉等<sup>[13]</sup>的模型基础上,融合了抗癌药物的相似性信息,提出了一种新的预测抗癌药物敏感性缺失数据的方法,并利用该算法对癌症基因组计划(Cancer Genome Project, CGP)<sup>[14]</sup>数据库中的954个细胞系与192个抗癌药物之间的敏感性数据的缺失值进行预测。通过10倍交叉验证调试参数,

获得模型中参数的最优取值,并与已有模型的预测效果作比较,以验证本文模型的有效性。

## 1 数据与方法

### 1.1 数据

癌症基因组计划数据是癌症药物敏感性基因组学(Genomics of Drug Sensitivity in Cancer, GDSC)计划中的数据。GDSC数据库是由英国桑格研究院维护,包含了肿瘤细胞对抗癌药物的敏感度信息、癌基因组突变信息。

GDSC中的数据来自75000个实验,包括约200个抗癌药物在1000多种肿瘤细胞中的反应。GDSC中的癌基因组突变信息来自COSMIC数据库,包含了癌基因点突变信息、基因扩增与丢失信息、组织类型信息以及表达谱信息等。用户可以从化合物(Compound)、癌症特征(Cancer feature)和细胞系(Cell line)这3个方面对数据库进行检索,检索结果中详细列出了癌症特征或细胞系对不同药物的反应,且以图形化的方式呈现,并附有相关文献。

在CGP数据库中药物敏感性指标通过荧光细胞活力分析获取,抗癌药物敏感性数据测量指标有AUC、IC50,本文采用AUC数据。

本文使用的数据包括两部分,一部分为954个癌细胞系关于16445个基因的基因表达量数据,另一部分为954个癌细胞系与192个抗癌药物的有缺失的敏感性数据。由于数据矩阵很大,表1仅给出了954个癌细胞系对于192个抗癌药物的敏感性数据的部分数据。

表1 原始抗癌药物的敏感性数据

癌细胞系	抗癌药物				
	Sunitinib	CGP-082996	BMS-509744	PAC-1	FR-180204
MC-CAR	0.9036	0.9292	0.9631	0.8116	0.9445
PFSK-1	NA	NA	NA	0.9613	0.9802
A673	NA	NA	NA	0.9893	0.9819
ES3	0.9399	0.8970	0.8495	0.8852	0.9616
ES5	0.9342	0.9779	0.9151	0.8406	0.9160
ES7	0.9351	0.9380	0.9203	0.8696	0.9474
NCI-H2227	NA	NA	NA	0.9829	0.9684

注:NA表示缺失数据。

观察表1可以看到,抗癌药物敏感性数据的范围均在0到1之间。与黄莉等<sup>[13]</sup>使用的CCLE数据库中的数据不同,CCLE数据库中的药物敏感性数据在0到10之间,其数据中的元素为IC50值。本文中药物敏感性数据为AUC数据,尽管AUC数据和IC50数据均可以作为衡量药物敏感性的标准,

但AUC数据与IC50数据中的数值意义不同。对于IC50数据,IC50值越大,表示癌细胞系对于药物的敏感性越差;IC50值越小,表示癌细胞系对于药物的敏感性越强。AUC数据的意义与IC50数据刚好相反。

## 1.2 低秩矩阵填充模型

填充矩阵的目标是填充矩阵的缺失数据,使数据完整。可以把抗癌药物敏感性数据看作一个矩阵  $A$ ,那么  $A$  就有与行对应的  $m$  个癌细胞系和与列对应的  $n$  个抗癌药物。填充矩阵就是利用矩阵  $A$  的已有数据对其缺失数据进行预测。假设  $A$  是一个低秩矩阵,且  $A$  的秩  $r \ll \min\{m, n\}$ 。由奇异值分解定理可知:存在矩阵  $U_{m \times r}$ 、 $H_{r \times r}$  和  $V_{n \times r}$ ,使  $X = U_{m \times r} H_{r \times r} (V_{n \times r})^T$ 。Cai 等<sup>[11]</sup>提出的低秩矩阵填充模型可以用公式表示为:

$$\begin{cases} \min_x \left( \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (A_{ij}^E - X_{ij}^E)^2 + \lambda g(X) \right) \\ \text{s.t. } X = U_{m \times r} H_{r \times r} (V_{n \times r})^T \end{cases} \quad (1)$$

其中: $X$  是预测得到的  $A$  的近似完整矩阵; $E$  表示矩阵  $A$  中数据位置的集合,当  $(i, j) \in E, A_{ij}^E = A_{ij}$ ,  $X_{ij}^E = X_{ij}$ , 否则  $A_{ij}^E = 0, X_{ij}^E = 0$ ;函数  $g(X)$  是一个正则化项,可以用公式表示为:

$$g(X) = \sum_{i=1}^m g_1\left(\frac{\|U^i\|^2}{3\mu_0 r}\right) + \sum_{i=1}^n g_1\left(\frac{\|V^i\|^2}{3\mu_0 r}\right),$$

且当  $Z \geq 1$  时,  $g_1(Z) = e^{(Z-1)^2} - 1$ , 否则  $g_1(Z) = 0$ 。 $U^i$  表示  $U$  的第  $i$  行,  $V^i$  表示  $V$  的第  $i$  行。 $\mu_0 = \max\{m, n\}$ 。

黄莉等<sup>[13]</sup>认为,如果两个癌细胞系的相似性越高,癌症药物敏感性越相似,并根据该观点提出了以下改进模型:

$$\begin{cases} \min_x \left( \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (A_{ij}^E - X_{ij}^E)^2 + \lambda_1 g(X) + \lambda_2 \sum_{i=1}^{m-1} \sum_{j=i+1}^m S_{ij} \|X_i - X_j\|^2 \right) \\ \text{s.t. } X = U_{m \times r} H_{r \times r} (V_{n \times r})^T \end{cases} \quad (2)$$

其中: $S_{ij}$  表示癌细胞系基因表达量之间的 Pearson 相关性系数; $X_i$  表示  $X$  的第  $i$  行。

研究表明,某些药物对于相同癌症类型的患者具有不同的疗效,这是因为基因突变越多,对癌症药物就越敏感<sup>[15]</sup>,于是本文假设:基因表达量相似性越高的癌细胞系与药物的敏感性数据相似性越高。因为抗癌药物敏感性是癌细胞系对癌症药物的敏感性,于是本文考虑不同癌症细胞系对于不同药物的敏感性是否与药物的相关性有关,因此假设:越高相似性的药物有着越高相似性的敏感性数据。基于以上假设,本文将药物相似性信息融合进模型 2,提出以下新的矩阵填充模型:

$$\begin{cases} \min_x \left( \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n (A_{ij}^E - X_{ij}^E)^2 + \lambda_1 g(X) + \lambda_2 \sum_{i=1}^{m-1} \sum_{j=i+1}^m S_{ij} \|X_i - X_j\|^2 + \lambda_3 \sum_{i=1}^{n-1} \sum_{j=i+1}^n D_{ij} \|X_i^T - X_j^T\|^2 \right) \\ \text{s.t. } X = U_{m \times r} H_{r \times r} (V_{n \times r})^T \end{cases} \quad (3)$$

其中:通过计算每个癌细胞系的基因表达数据之间的 Spearman 相关性系数得到  $S_{ij}$ ;  $D_{ij}$  表示药物  $i$  与药物  $j$  的药物相似性,通过不同药物的敏感性数据计算不同药物之间的 Pearson 相关性系数得到;  $X_i^T$  表示  $X$  的第  $i$  列。

## 1.3 交替梯度下降法

本文使用交替梯度下降法对模型 3 进行优化。交替梯度下降法是对梯度下降法的改进,算法的步骤为:

a) 设置初值。对矩阵  $A$  的空缺部分用 0 代替得到矩阵  $A^E$ , 并对  $A^E$  进行奇异值分解,得到  $A^E = UH(V)^T$ , 分别设置  $U$  的初值为  $U^{(0)} = U_0 \sqrt{m}$ 、 $V$  的初值为  $V^{(0)} = V_0 \sqrt{n}$ , 其中:  $U_0$  表示  $U$  的前  $r$  列;  $V_0$  表示  $V$  的前  $r$  列。

b) 固定  $U^{(k)}$  和  $V^{(k)}$ , 通过最小化(3)式迭代计算得到相应的  $H^{(k)}$ 。

$$H_{ij}^{(k,a+1)} = H_{ij}^{(k,a)} \frac{((U^{(k)})^T A V^{(k)})_{ij}}{((U^{(k)})^T U^{(k)} H^{(k,a)} (V^{(k)})^T V^{(k)})_{ij} + 2\lambda_2 M_{ij} + 2\lambda_3 N_{ij}}.$$

其中:  $a$  表示迭代次数,当第  $a$  次迭代与第  $a+1$  次迭代得到的  $H^{(k)}$  值之间的差小于  $10^{-7}$  时,迭代停止;  $M = (U^{(k)})^T B^{(k)} U^{(k)} H^{(k,a)} (V^{(k)})^T V^{(k)}$ ,  $N = (U^{(k)})^T U^{(k)} H^{(k,a)} (V^{(k)})^T C^{(k)} V^{(k)}$ ,  $B$ 、 $C$  是对称矩阵,  $B = [b_{ij}]_{m \times m}$ ,  $C = [c_{ij}]_{n \times n}$ , 且

$$b_{ij} = \begin{cases} \sum_{k \neq i} S_{ik}, & i = j \\ -S_{ik}, & i \neq j \end{cases}$$

$$c_{ij} = \begin{cases} \sum_{k \neq i} D_{ik}, & i = j \\ -D_{ik}, & i \neq j \end{cases}.$$

c) 根据梯度下降法更新  $U^{(k+1)}$ 、 $V^{(k+1)}$ , 即

$$U^{(k+1)} = U^{(k)} + t \nabla U^{(k)}, V^{(k+1)} = V^{(k)} + t \nabla V^{(k)}.$$

$\nabla U$  表示矩阵  $U$  的梯度、 $\nabla V$  表示矩阵  $V$  的梯度,分别为

$$\nabla U = ((UHV^T) - A) V H^T + U Q_U + \lambda_1 f(U, Q_U) + \lambda_2 B U H (V^T V) H^T +$$

$$4\lambda_3(UHV^T)CVH^T, \nabla V = ((UHV^T) - A)^T UH + VQ_V + \lambda_1 f(V, 2e^{(Q_{V1} - I_2)^2} (Q_{V1} - I_2)) + 4\lambda_2 VH^T U^T BUH + 4\lambda_3 CVH^T (U^T U) H.$$

其中:

$$Q_U = \frac{1}{m} U^T (A - (UHV^T)) VH^T,$$

$$Q_V = \frac{1}{n} V^T (A - (UHV^T))^T UH,$$

$$I_1 = (1, 1, \dots, 1)_{m \times 1}^T, I_2 = (1, 1, \dots, 1)_{n \times 1}^T,$$

$$Q_{U1} = \frac{1}{2\alpha r} \begin{bmatrix} \sum_{j=1}^r U_{1j}^2 \\ \sum_{j=1}^r U_{2j}^2 \\ \vdots \\ \sum_{j=1}^r U_{mj}^2 \end{bmatrix}, Q_{V1} = \frac{1}{2\alpha r} \begin{bmatrix} \sum_{j=1}^r V_{1j}^2 \\ \sum_{j=1}^r V_{2j}^2 \\ \vdots \\ \sum_{j=1}^r V_{nj}^2 \end{bmatrix},$$

其中:  $\alpha = \max\{m, n\}$ ,  $f(A_{m \times r}, B_{m \times 1}) = Z_{m \times r}, Z_{ij} =$

$$\begin{cases} \frac{1}{\alpha r} A_{ij} B_i, & B_i > 0 \\ 0, & \text{其他} \end{cases}.$$

梯度的计算与已有文献[11-13]的计算相似。

d)重复第 b)步和第 c)步,当算法收敛(迭代误差  $< 10^{-8}$ )或达到给定的迭代次数(本文设置 50 次)时停止。

#### 1.4 模型评估

本文采用均方根误差(RMSE)对模型进行评价。一般来说, RMSE 值越小,表示模型越好,预测

结果的准确性就越高。假设有两个向量  $p$  和  $q$ ,  $p = (p_1, p_2, \dots, p_k)$ ,  $q = (q_1, q_2, \dots, q_k)$ 。在本文中  $p$  表示原始值,  $q$  表示预测值,那么对应的 RMSE 为:

$$RMSE = \sqrt{\frac{\sum_{i=1}^k (p_i - q_i)^2}{k}}.$$

#### 1.5 交叉验证

10 倍交叉验证是把样本随机均分为 10 份,其中 1 份作为测试集,其余 9 份作为训练集。重复 10 次,直至每一份都曾作为测试集,得到 10 个均方根误差值,然后取平均值作为最后的均方根误差值。

本文把药物敏感性数据中的已有数据随机均分为 10 份,选取其中的 9 份作为训练集,把另外 1 份的值设置为 0。通过训练参数  $r, \lambda_1, \lambda_2, \lambda_3$  得到一个新的完整矩阵,取完整矩阵中与测试集对应的部分值,计算得到两者的均方根误差,运行 10 次,得到平均的 RMSE 值。本文采用 5 次 10 倍交叉验证,既把 10 倍交叉验证运行 5 次得到 5 个平均的 RMSE 值,再取平均,就得到本文最终的 RMSE 值。

## 2 结果与讨论

根据 CGP 数据库中的癌细胞系基因表达量数据计算得到癌细胞系之间的 Pearson 相似性数据,部分癌细胞系相似性数据见表 2。根据表 1 中的已有抗癌药物敏感性数据计算得到抗癌药物之间的 Spearman 相似性数据,部分抗癌药物相似性数据见表 3。

表 2 部分癌细胞系相似性数据

癌细胞系	BB30-HNC	BB49-HNC	MC-CAR	H9	KS-1
BB30-HNC	1.0000	-0.0056	0.0053	0.0062	-0.0150
BB49-HNC	-0.0056	1.0000	-0.0101	0.0067	0.0069
MC-CAR	0.0053	-0.0101	1.0000	0.0122	-0.0031
H9	0.0062	0.0067	0.0122	1.0000	0.0023
KS-1	-0.0150	0.0069	-0.0031	0.0023	1.0000

表 3 部分抗癌药物相似性数据

抗癌药物	Erlotinib	Rapamycin	Sunitinib	PHA-66575	MG-132
Erlotinib	1.0000	0.6593	0.7625	0.7713	0.7910
Rapamycin	0.6593	1.0000	0.7411	0.7432	0.7173
Sunitinib	0.7625	0.7411	1.0000	0.8374	0.7779
PHA-66575	0.7713	0.7432	0.8374	1.0000	0.7894
MG-132	0.7910	0.7173	0.7779	0.7894	1.0000

对比表 2 与表 3 中的相似性数值,可以看到:癌细胞系相似性与抗癌药物相似性数据对应的数量级差别较大,因此本文在参数的设置中也取不同大小的数值做训练。本文参数  $\lambda_1$  在  $[0, 1]$  区间取值,间距设为 0.1000;参数  $\lambda_2$  在  $[0, 0.1]$  取值,间距设为 0.0100;

参数  $\lambda_3$  在  $[0, 0.001]$  区间取值,间距设为 0.0001;矩阵的秩  $r$  在  $[5, 30]$ ,间距设为 5。通过实验发现,当  $r=10$ ,  $\lambda_1=0.1000$ ,  $\lambda_2=0.0100$ ,  $\lambda_3=0.0001$  时,表 1 的填充效果达到最佳,此时最小 RMSE 为 0.1153。部分填充后的抗癌药物敏感性数据见表 4。

表4 填充后的部分抗癌药物敏感性数据

癌细胞系	抗癌药物				
	Sunitinib	CGP-082996	BMS-509744	PAC-1	FR-180204
MC-CAR	0.8708	0.9645	0.9329	0.8875	0.9324
PFSK-1	0.5751	0.6054	0.5891	0.9673	0.9504
A673	0.5759	0.6042	0.5831	0.9671	0.9456
ES3	0.9095	0.9805	0.9601	0.9379	0.9610
ES5	0.8997	0.9665	0.9290	0.8707	0.8954
ES7	0.8898	0.9152	0.8877	0.8946	0.9161
NCI-H2227	0.7309	0.8146	0.7884	0.9646	0.9993

为了与已有模型的填充效果作比较,采用本文模型对黄莉等<sup>[13]</sup>的 IC<sub>50</sub> 数据进行计算,其中仍采用参数  $r = 10$ ,  $\lambda_1 = 0.1000$ ,  $\lambda_2 = 0.0100$ ,  $\lambda_3 =$

0.0001,得到的 RMSE 值为 0.6414,较原文献的结果 0.6688 降低了 4.10%。利用本文模型填充后的部分 IC<sub>50</sub> 数据见表 5。

表5 利用本文模型填充后的 IC<sub>50</sub> 部分数据

癌细胞系	抗癌药物				
	L.685458	ZD.6474	Sorafenib	Irinotecan	PD.0325901
LN18_CENTRAL_NERVOUS_SYSTEM	0.5067	1.1446	0.5893	3.2395	0.6940
A2780_OVARY	0.5794	1.5220	0.7529	3.7418	3.3080
OVTOKO_OVARY	0.7056	1.1198	0.7194	2.7513	0.5519
SJSA1_BONE	0.4734	1.1807	0.5843	3.2168	1.4344
SKMES1_LUNG	1.1919	1.8364	1.2209	4.0512	2.0942
JHH5_LIVER	0.4170	1.2158	0.5661	3.1742	2.3803
MFE280_ENDOMETRIUM	0.2932	0.9335	0.4100	2.6309	1.4110
MFE319_ENDOMETRIUM	0.5219	1.3106	0.6456	3.5884	1.5613
MFE296_ENDOMETRIUM	0.5837	1.2716	0.6635	3.6786	0.3841
GI1_CENTRAL_NERVOUS_SYSTEM	0.2998	1.0510	0.4322	3.2200	0.9968

### 3 结 论

本文考虑与抗癌药物敏感性数据相关的抗癌药物和癌细胞系的信息,运用低秩填充算法,提出了一个融入癌细胞系与抗癌药物相似性信息的低秩矩阵填充模型。该模型利用现有实验数据分析数据内容,将癌细胞基因表达数据与癌症药物数据融入模型中,对抗癌药物敏感性缺失数据进行预测,得到最优 RMSE 值为 0.6414,提高了模型对缺失数据预测的准确性。对抗癌药物敏感性缺失数据进行预测得到完整的数据,有利于研究者们利用数据更加准确地得到肿瘤治疗的生物标记物,为肿瘤患者的“个性化”治疗提供帮助。

#### 参考文献:

- [1] Kurbacher C M, Cree I A, Bruckner H W, et al. Use of an ex vivo ATP luminescence assay to direct chemotherapy for recurrent ovarian cancer[J]. Anti-cancer Drugs, 1998, 9(1): 51-57.
- [2] Xu J M, Song S T, Tang Z M, et al. Predictive

chemotherapy of advanced breast cancer directed by MTT assay in vitro[J]. Breast Cancer Research and Treatment, 1999, 53(1): 77-85.

- [3] 卢毅卓,罗琪. 肿瘤药物敏感性试验方法研究进展[J]. 医学综述, 2007, 13(7): 551-553.
- [4] Farahat H K, Ledermann W. Matrices with prescribed characteristic polynomials [J]. Proceedings of the Edinburgh Mathematical Society, 1959, 11(3): 143-146.
- [5] Cabral R, La Torre F D, Costeira J P, et al. Matrix completion for weakly-supervised multi-label image classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1): 121-135.
- [6] 刘青山,卢汉清,马颂德. 综述人脸识别中的子空间方法[J]. 自动化学报, 2003, 29(6): 900-911.
- [7] Ramlatchan A, Yang M Y, Liu Q, et al. A survey of matrix completion methods for recommendation systems [J]. Big Data Mining and Analytics, 2018, 1(4): 308-323.
- [8] Candès E J, Recht B. Exact matrix completion via convex optimization[J]. Foundations of Computational Mathematics, 2009, 9(6): 717-772.

- [9] Candes E J, Tao T. The power of convex relaxation; Near-optimal matrix completion[J]. IEEE Transactions on Information Theory, 2010, 56(5): 2053-2080.
- [10] Hastie T, Mazumder R, Lee J. Matrix completion and low-rank SVD via fast alternating least squares[J]. Journal of Machine Learning Research, 2014, 16(1): 3367-3402.
- [11] Cai Z P, Zhang T, Wan X F. A computational framework for influenza antigenic cartography [J]. PLoS Computational Biology, 2010, 6(10): e1000949.
- [12] Keshavan R H, Montanari A, Oh S. Matrix completion from a few entries[J]. IEEE Transactions on Information Theory, 2010, 56(6): 2980-2998.
- [13] 黄莉, 贺平安. 矩阵填充算法在抗癌药物敏感性研究中的运用[J]. 浙江理工大学学报, 2017, 37(6): 881-887.
- [14] Garnett M J, Edelman E J, Heidorn S J, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells [J]. Nature, 2012, 483(7391): 570-575.
- [15] Rizvi N A, Hellmann M D, Snyder A, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer [J]. Science, 2015, 348(6230): 124-128.

(责任编辑:康 锋)