

基于多相似度融合的药物重定位推荐算法

陈鹏, 鲍天嘉智, 余肖生

(三峡大学计算机与信息学院, 湖北宜昌 443002)

摘要: 药物重定位算法可以为发现药物的新用途提供理论上的支持。针对传统药物重定位推荐算法的不足, 提出一种基于多相似度融合的药物重定位推荐算法(MSF)。首先通过药物-疾病数据源计算出疾病相似度, 再通过药物-化学结构、药物-靶蛋白以及药物-副作用数据源计算出三种相似度并融合为药物相似度, 最后利用两种相似度计算药物与疾病对应关系的预测值, 并通过预测值融合方法融合为最终预测值。实验结果表明, 与 SLAMS 算法和 DRCFFS 算法相比, MSF 算法在精确率和召回率上有较大的提高; 与针对单个数据源的药物重定位算法相比, 通过融合多种数据源, 预测值的可靠性和精确性都有进一步的提升。案例分析显示, MSF 算法可以有效地预测出有治疗效果的药物-疾病组合。

关键词: 药物重定位; 协同过滤; 数据稀疏性; 相似度融合; 预测值融合

中图分类号: TP399

文献标识码: A

文章编号: 1673-629X(2021)01-0168-07

doi: 10.3969/j.issn.1673-629X.2021.01.030

Recommendation Algorithm for Drug Repositioning Based on Multi-similarity Fusion

CHEN Peng, BAO Tian-jiazhi, YU Xiao-sheng

(School of Computer and Information, Three Gorges University, Yichang 443002, China)

Abstract: Drug repositioning algorithms can provide theoretical support for discovering new uses of drugs. To overcome the shortcomings of traditional drug relocation recommendation algorithm, we propose a drug repositioning recommendation (MSF) based on multi-similarity fusion. Disease similarity is calculated by drug-disease data sources, and then three similarities are fused into drug similarity by drug-chemical structure, drug-target protein and drug-side effect data sources. Finally, two similarities are used to calculate the predicted value and fuse it into the final predicted value. The experiment shows that compared with SLAMS and DRCFFS, MSF has a greater improvement in accuracy and recall rate. Compared with collaborative filtering algorithm for a single data source, the reliability and accuracy of predicted values are further improved by fusing multiple data sources. Case analysis shows that MSF can effectively predict drug-disease combination with therapeutic effect.

Key words: drug repositioning; collaborative filtering; data sparsity; similarity fusion; predictive value fusion

0 引言

药物重定位 (drug repurposing), 俗称“老药新用”, 是指通过现有的技术手段将已经产生适应症的药物重新定位, 寻找其新的适应症^[1]。药物重定位这一概念自被提出以来, 国内外学者对该方向算法的研究投入了巨大的精力^[2-3]。张永祥等认为药物重定位是网络药理学的重要应用领域, 介绍了基于小分子特征、基于蛋白靶点特征等方法^[4]。谢达菲等提出了利用药物-靶标关系、药物-药物关系和药物-疾病关系等方法进行药物重定位的计算预测^[5]。文献[6]将深度学习的方法应用于药物重定位。Luo 等人提出

了一种基于综合相似度和随机游走的药物重定位算法^[7-8], 通过将药物和疾病特征信息与已知的药物-疾病关系相结合, 首次提出了药物与疾病相似性的综合相似度计算方法。文献[9-11]提出了基于多数据源融合的药物重定位算法。文献[12-15]提出了一系列基于协同过滤的药物重定位算法研究, 从多源数据的角度出发, 通过协同过滤来计算药物-疾病对应关系预测值。

由于数据稀疏性对协同过滤的影响较大, 所以现阶段基于协同过滤的药物重定位算法大多通过融合多种数据源进行计算的方法来减小影响, 但在计算过程

收稿日期: 2020-03-12

修回日期: 2020-07-15

基金项目: 国家重点研究发展计划资助项目(2016YFC0802500)

作者简介: 陈鹏(1973-), 男, 博士, 教授, CCF 会员(73012M), 从事大数据分析技术研究; 通讯作者: 余肖生(1973-), 男, 博士, 副教授, CCF 会员(98980M), 从事健康医疗大数据分析研究。

中仅使用了药物相似度,忽略了疾病相似度的作用。

综上所述,传统的基于协同过滤的药物重定位算法虽然有一定的效果,但仍有很大的进步空间。该文提出了一种基于多相似度融合的药物重定位推荐算法(MSF),首先由药物-疾病数据源计算出疾病相似度,其他三种数据源计算出的三种相似度融合为药物相似度,再分别利用基于项目和基于用户的协同过滤进行药物与疾病对应关系的预测值计算,并通过融合方法将两种预测值融合为最终的药物-疾病关系预测值。MSF算法在充分利用多源数据的前提下,将计算出的相似度和预测值融合,降低了数据稀疏性对协同过滤的影响。

1 相关理论

1.1 药物重定位

药物重定位已经成为医学领域研究的一大热门。现阶段,由于药物和疾病数据的大量增长,从不同角度进行药物重定位研究的案例也越来越多。为证明本文对药物重定位的研究角度是切实可行的,从药物与疾病的关系、药物化学结构、药物靶蛋白和药物副作用四个方面进行讨论。

Chiang 等人^[16]提出了一种从疾病的角度看待药物重定位的观点,当两种疾病可以被多种相同的药物治疗时,认为两种疾病是相似的。如果存在一种药物只对其中一种疾病有治疗效果,则认为该药对另一种疾病也存在潜在的治疗关系,可以作为治疗该疾病的候选药物。药物的化学结构被认为可以用来度量药物间相似性,Dudley 等人^[17]提出药物的化学性质与其治疗效果有密切的关系,药物的化学结构和生物活性之间存在定量关系,所以药物的化学结构可以作为药物重定位的研究方向。药物靶蛋白是药物治疗疾病的关键因素,含有相似靶蛋白的药物也可能会有类似的作用效果^[18],因此靶蛋白可以作为药物重定位中度量药物相似性的一个研究角度。同样,与药物对疾病的治疗效果类似,药物产生的副作用提供了人类的表型特征,因此从药物的副作用角度进行药物重定位的研究也是可行的^[19]。

通过上述讨论证明从药物与疾病的关系、药物化学结构、药物靶蛋白和药物副作用这四个角度进行药物重定位研究是可行的。但是上述文献提供的方法都是从多个角度中的一类或两类进行药物重定位研究,可能会导致预测值的偏差,因此,该文从四个角度综合考虑并进行研究,减小了预测结果的有偏性,更具有实际应用价值。

1.2 协同过滤

协同过滤算法在近几年被广泛应用于各大电商行

业以及影视平台的应用推荐中^[20-21]。传统基于邻域的协同过滤算法分为两种,即基于用户的和基于项目的,两种算法的本质都是通过选取 k 个最相似的邻居进行评分预测。

协同过滤的基本流程为:先通过余弦相似度、皮尔逊相关系数、谷本系数等相似度度量方法计算用户或项目的相似度。再利用计算出的相似度找到相似度最高的 k 个邻居,由对应的两种公式求出预测分数:

(1)定义 r_{ui} 为用户 u 对项目 i 的评分, D 为用户 u 的邻居集合, \bar{r}_u 为用户 u 对所有项目的平均评分, sim_{uv} 为用户 u 和用户 v 的相似度, r_{vi} 为用户 v 对项目 i 的评分, \bar{r}_v 表示用户 v 对所有项目的平均评分,基于用户的预测值计算如式(1):

$$r_{ui} = \bar{r}_u + \frac{\sum_{v \in D} \text{sim}_{uv} \cdot (r_{vi} - \bar{r}_v)}{\sum_{v \in D} \text{sim}_{uv}} \quad (1)$$

(2)定义 r_{ui} 为用户 u 对项目 i 的评分, D 为项目 i 的邻居集合, sim_{ij} 表示项目 i 和 j 的相似度, r_{uj} 为用户 u 对项目 j 的评分,基于项目的预测值计算如式(2):

$$r_{ui} = \frac{\sum_{j \in D} \text{sim}_{ij} \cdot r_{uj}}{\sum_{j \in D} \text{sim}_{ij}} \quad (2)$$

2 基于多相似度融合的药物重定位推荐算法

传统的基于协同过滤的药物重定位算法仅使用药物相似度进行预测值计算,往往忽略了疾病相似度在药物重定位中的作用且研究角度比较单一,而由于单个数据源的稀疏问题,导致计算出的有效药物、疾病相似度较少,许多项目无法找到合适的邻居,因而计算出的预测值偏差较大^[22]。该文提出了 MSF 算法,即:首先由药物-疾病数据源计算出疾病相似度,再通过药物-化学结构、药物-靶蛋白以及药物-副作用数据源计算出三种相似度融合为药物相似度,使用疾病相似度和药物相似度计算两种预测值并融合为最终的预测值,整体流程如图 1 所示。

2.1 相似度计算

2.1.1 疾病相似度

疾病相似度通过药物-疾病数据进行计算。药物-疾病治疗关系数据从 UMLS^[23]中的 NDF-RT(national drug file-reference terminology)进行采样,其中药物与疾病有作用关系则值为 1,若无关系则为 0。这种药物-疾病数据集在文献[11]中被认为是药物重定位的“金”标准数据集。基于文献[16]的思想,该文对于疾病相似度的计算只考虑药物与疾病是否有对应的治疗

关系,而不考虑治疗效果的优劣,所以疾病相似度通过谷本系数进行,与使用余弦相似度等方法相比,在不影响相似度准确性的情况下,简化了计算的复杂程度。疾病相似度的计算公式如式(3):

$$\text{sim}(i_a, i_b) = \frac{|I_{ab}|}{I_a + I_b - |I_{ab}|} \quad (3)$$

其中, $\text{sim}(i_a, i_b)$ 表示疾病 a 和疾病 b 的相似度, I_a 表示可以治疗疾病 a 的药物数量, I_b 表示可以治疗疾病 b 的药物数量, $|I_{ab}|$ 表示可以同时治疗疾病 a 和疾病 b 的药物数量。 $\text{sim}(i_a, i_b)$ 的值应该在区间 $[0, 1]$ 之间。

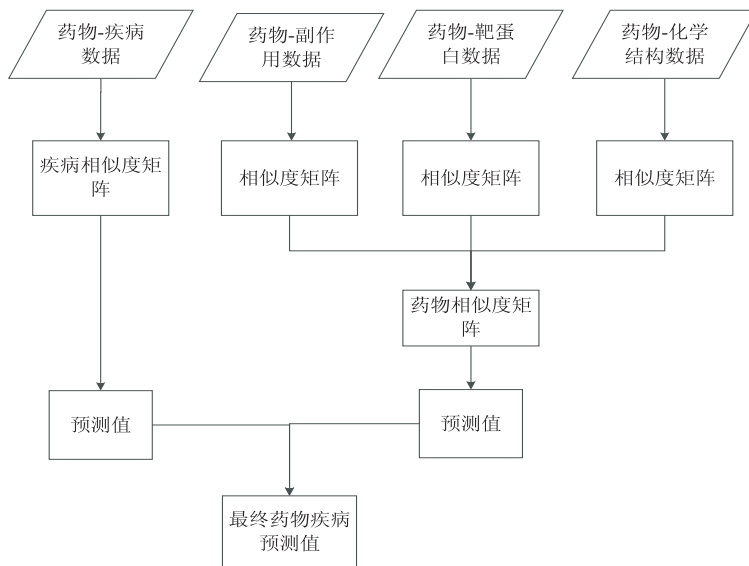


图 1 算法流程

2.1.2 药物相似度

在文献[17-19]的基础上,该文认为通过药物-化学结构、药物-靶蛋白和药物-副作用三种数据源计算出的药物相似度可以通过一定的权重融合为最终的药物相似度,其中药物-化学结构数据从 PubChem^[24] 进行采样,药物-靶蛋白数据从 UniPort Knowledgebase^[25] 中采样,药物-副作用数据从 SIDER^[26] 数据库中采样。融合相似度的方法有效缓解了单个数据源因为数据稀疏而导致计算出的有效相似度较少的问题,并且通过融合减少了计算最终预测值所需要的步骤。由于采用相似度融合的方法,所以三种数据源应该选用同一种相似度计算方法。与计算疾病相似度类似,用于药物相似度计算的三种数据源中只考虑药物与三种属性的对应关系,所以同样采用谷本系数进行相似度的计算,如式(4):

$$\text{sim}(d_a, d_b) = \frac{|D_{ab}|}{D_a + D_b - |D_{ab}|} \quad (4)$$

其中, $\text{sim}(d_a, d_b)$ 表示药物 a 和药物 b 的相似度,(a) 在使用药物-化学结构数据源的计算过程中, D_a 表示药物 a 包含的化学结构数量, D_b 表示药物 b 包含的化学结构数量, $|D_{ab}|$ 表示药物 a 和药物 b 包含的相同化学结构数量;(b) 在使用药物-靶蛋白数据源的计算过程中, D_a 表示药物 a 对应的靶蛋白数, D_b 表示药物 b 对应的靶蛋白数, $|D_{ab}|$ 表示药物 a 和药物 b 对应的相同靶蛋白数;(c) 在使用药物-副作用数据源的计算

过程中, D_a 表示药物 a 会产生的副作用数, D_b 表示药物 b 会产生的副作用数, $|D_{ab}|$ 表示药物 a 和药物 b 产生的相同副作用数。同样, $\text{sim}(d_a, d_b)$ 的值也应该在区间 $[0, 1]$ 之间。

sim_d 表示由三种相似度融合得到的最终药物相似度, sim_s 表示由药物-化学结构算出的相似度, sim_p 表示由药物-靶蛋白算出的相似度, sim_f 表示由药物-副作用算出的相似度。使用式(5)进行计算:

$$\text{sim}_d = \alpha \text{sim}_s + \beta \text{sim}_p + \gamma \text{sim}_f \quad (5)$$

其中, $\alpha + \beta + \gamma = 1$ 。

相似度融合的权值计算采用试探法,以 0.1 为步长进行试探,通过多次实验确定一组能使效果最优的权值。试探法的优点是使用简单且较为准确,缺点是运行效率较低,但在设置合适的步长时可以在一定的时间范围内得到预期的结果。通过试探法得出当 $\alpha = 0.2, \beta = 0.4, \gamma = 0.4$ 时,效果达到最优。

2.2 预测值计算

通过两种相似度计算方法得到疾病相似度和药物相似度后,为了融合两类相似度计算出的预测结果,提出预测值融合的方法,如式(6):

$$P_{aq} = \omega_1 P_{aq}^i + \omega_2 P_{aq}^d \quad (6)$$

其中, P_{aq} 为融合后药物 a 与疾病 q 对应关系的预测值, P_{aq}^i 为通过疾病相似度算出的药物 a 与疾病 q 的对应关系预测值, P_{aq}^d 为由药物相似度算出的药物 a 与疾病 q 的对应关系预测值, ω_1 和 ω_2 分别表示两种类型预

测值的权重,且需要满足: $\omega_1 + \omega_2 = 1$ 。

为了计算 P_{aq}^i ,需要通过疾病相似度选取 k 个最相似的邻居。在选取好 k 个邻居后,通过式(7)进行 P_{aq}^i 的计算:

$$P_{aq}^i = \frac{\sum_{y \in T} \text{sim}(I_q, I_y) \times s_{a,y}}{\sum_{y \in T} \text{sim}(I_q, I_y)} \quad (7)$$

其中, $\text{sim}(I_q, I_y)$ 表示疾病 q 和疾病 y 的相似度, T 为疾病 q 的邻居集合, $s_{a,y}$ 表示药物 a 与疾病 y 的对应关系值。以上计算公式是基于项目的,所以只能通过疾病相似度计算预测值,通过药物相似度计算预测值 P_{aq}^d 则需要使用式(8):

$$P_{aq}^d = \bar{s}_a + \frac{\sum_{b \in T} \text{sim}(D_a, D_b) \times (s_{b,q} - \bar{s}_b)}{\sum_{b \in T} \text{sim}(D_a, D_b)} \quad (8)$$

其中, $\text{sim}(D_a, D_b)$ 表示药物 a 和药物 b 的相似度, T 为药物 a 的邻居集合, \bar{s}_a 和 \bar{s}_b 表示基于药物-疾病数据源计算出的药物 a 和药物 b 与所有疾病对应关系值的平均数, $s_{b,q}$ 表示药物 b 与疾病 q 的对应关系值。

由于两种预测值也要进行融合,为了保持数据的一致性,两类算法中选取的邻居数应该为相同值。根据多次实验结果,设置 $\omega_1 = 0.6, \omega_2 = 0.4$ 。

2.3 算法流程

Step1:将四种数据源转换成对应的矩阵,药物-疾病矩阵通过式(3)计算疾病相似度,其他三种数据源得出的矩阵通过式(4)计算出相似度。

Step2:由 Step1 算出四种相似度,将通过式(4)算出的相似度用式(5)融合成药物相似度,通过式(3)得出的相似度作为疾病相似度。

Step3:得出的疾病相似度通过式(7)计算出药物与疾病对应关系的预测值,药物相似度则通过式(8)计算出预测值。

Step4:用式(6)将两类预测值融合得到最终的药物与疾病对应关系预测值。

3 实验结果及分析

3.1 数据集

为了验证算法的有效性,该文采用的数据集是文献[9]中的数据,该数据集是通过 UMLS、PubChem、UniPort Knowledgebase 和 SIDER 中的数据进行人工处理和清洗获得的,其中包括药物-疾病数据、药物-化学结构数据、药物-靶蛋白数据和药物-副作用数据。四种数据均采用二进制表示对应的关系:0 表示无相关对应关系,1 表示有对应关系。数据集包括 536 种药物,以及对应的 578 种疾病、881 种药物化学结构、1 385 种副作用和 775 种对应靶蛋白。其对应的数

据稀疏程度分别为:0.992 0,0.859 2,0.945 5,0.995 4。稀疏程度主要用于观察数据集中无作用的数据占总数据的比例,从得出的数据可以看出四种数据源的稀疏程度都比较高,无作用数据占总数据的比例较大。

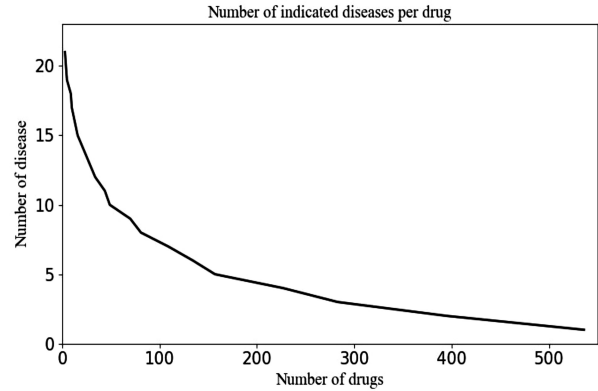


图 2 药物-疾病对应数量关系

从药物-疾病数据源中可以算出每一种药物对应可以治疗的疾病数,如图 2 所示,可以看出大约只有 50 种(10%)药物可以治疗 10 种以上疾病,大部分药物(75%)只能治疗 5 种以下的疾病。

3.2 实验结果

文中将药物重定位算法当作一个二分类问题来看待,对于每种药物,如果疾病可以治疗则为 1,反之则为 0,为了保证实验的准确性和健壮性,使用十折交叉验证法进行实验。该文采用准确率、召回率、F-score 和 ROC 曲线来进行算法效果好坏的判断。为了给出这四种分类指标的准确定义,首先定义一个二分类问题的混淆矩阵,如表 1 所示。

表 1 混淆矩阵

	预测值	
	真实值	TP(真正例)
	FP(假正例)	TN(真反例)

通过混淆矩阵可以对精确率、召回率、F-score 进行定义,如下:

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

$$F = \frac{2 \times P \times R}{P + R} \quad (11)$$

为了确定计算中最相似的邻居数 k 使预测值最精确,提供了每种数据源在不同邻居数情况下的精确率表现,如图 3 所示。通过观察每种数据源各自的精确率变化,确定一个使算法效果最好的 k 值。通过观察可以看出当 $k = 25$ 时,各数据源保持在精确率比较高的稳定状态,虽然在 $k > 25$ 时,药物-化学结构和药物-疾病数据源的精确率会有小幅度的提升,但是药物-

靶蛋白和药物-副作用数据源的精确率会有大幅的下降,所以该文将邻居数选取为 25。

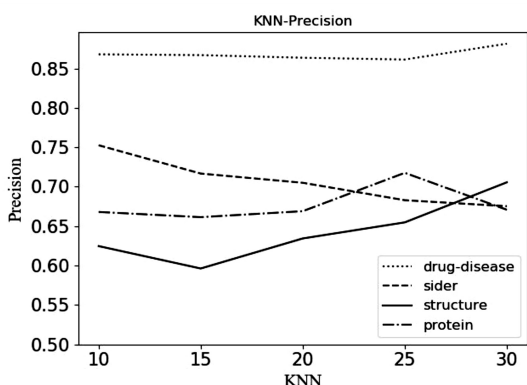


图 3 各数据源在不同邻居数下的精确率

在确定邻居数 k 后,通过图 4、图 5 对比 MSF 算法和其他算法的效果。在观察图 4 (ROC 曲线) 时发现,所有数据源的 AUC 值都很高,这是由于药物重定位是一个高度不平衡问题,所有的数据源都是稀疏的,即使假正例的数量变化很大,也会因为真反例的基数大,而假正率只发生极小的变化^[9]。所以在 ROC 曲线和 AUC 值对于区分算法效果表现较差的情况下,采用了 P-R 曲线(见图 5)和精确率、召回率、F-score(见表 2)来进一步证明 MSF 算法优于传统药物重定位算法。

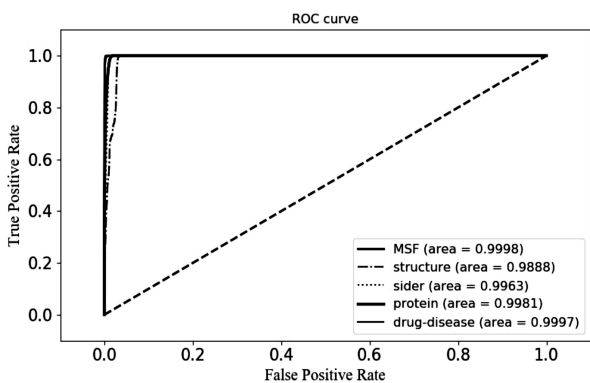


图 4 基于单个数据源的传统药物重定位算法和 MSF 算法的 ROC 曲线比较

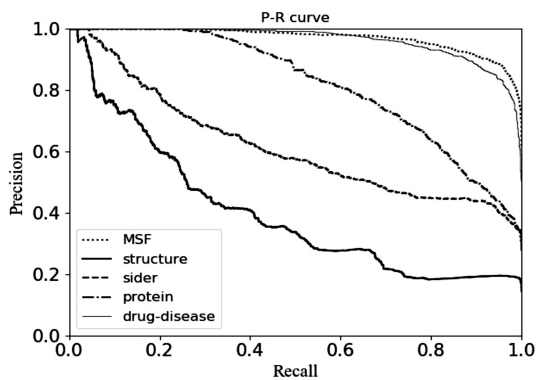


图 5 单个数据源的传统药物重定位算法和 MSF 算法的 P-R 图比较

表 2 单个数据源与 MSF 算法指标对比

Datasets	Precision	Recall	F-score
药物-疾病	0.891 0	0.922 8	0.891 0
药物-化学结构	0.654 8	0.173 6	0.274 4
药物-靶蛋白	0.717 6	0.726 3	0.721 9
药物-副作用	0.682 9	0.295 6	0.412 6
MSF(融合后)	0.912 3	0.896 3	0.904 2

通过 P-R 曲线可以看出,MSF 算法的 P-R 曲线包裹了所有单个数据源的 P-R 曲线。通过表 2 可以看出,只有通过药物-疾病数据源得出的召回率比 MSF 算法略高,但是在精确率和 F-score 方面 MSF 算法都表现得更好,据此可以得出 MSF 优于基于单个数据源的药物重定位推荐算法。

表 3 展示了 MSF 算法与其他两种算法 (SLAMS 算法^[9] 和 DRCFFS 算法^[14]) 的比较。可以看出与 SLAMS 算法和 DRCFFS 相比,MSF 算法在三种指标上都有较大的提升。

表 3 算法效果比较

算法	Precision	Recall	F-score
SLAMS	0.345 2	0.650 5	0.451 0
DRCFFS	0.870 0	0.756 1	0.808 5
MSF	0.912 3	0.896 3	0.904 2

通过上述几种不同指标的对比表明,MSF 算法在 P-R 曲线以及各项指标上都有良好的表现,可以得到更精确的药物重定位预测结果。

4 案例分析

药物重定位算法的目的是预测旧药物的新用途,为了证明提出的算法是有实效的和现实意义的,表 4 给出了利用 MSF 算法预测的一些药物-疾病组合,其中头孢西丁临床证明可以治疗由敏感菌导致的呼吸道感染等疾病,这与可以治疗支气管炎的预测相符合。厄他培南对肺炎的治疗作用在文献[27]中已经得到证实。文献[28]中进行了头孢克洛治疗鼻窦炎的临床实验,治愈率达到 15.5%。文献[29]中给出了利用美罗培南治疗皮肤软组织感染的实例。

表 5 给出了甲状腺肿大治疗预测值前五的药物,其中甲硫咪唑和丙硫氧嘧啶已经得到临床验证^[30],剩下三种药物尚未得到临床证明,但在文献[31]中双硫仑被证明对甲状腺治疗有一定作用。

综上所述,MSF 算法可以预测出一些已经得到临床认证的药物-疾病组合,并且也能发现一些尚未证明有临床意义的药物-疾病组合,但是药物的治疗效果与疾病相关症状对应。通过一系列的案例证明,MSF 算法具有可行性,对于临床药物重定位具有一定

的辅助作用。

表4 预测药物-疾病治疗

药物名称	预测治疗疾病	实际效果
头孢西丁	支气管炎	半合成的广谱头孢抗生素,用于静脉注射,主要治疗呼吸道感染等疾病
头孢克洛	鼻窦炎	主要用于治疗对该药品敏感菌产生的急性咽炎、急性扁桃体炎
厄他培南	肺炎,变形杆菌属感染	临床上主要用于治疗复杂性腹腔感染、复杂性皮肤软组织感染等疾病
洛美沙星	嗜血杆菌感染	是一种氟喹诺酮类抗生素,用于治疗细菌感染,包括支气管炎和尿路感染
美罗培南	软组织感染	用于治疗多种不同的感染,包括脑膜炎以及肺炎

表5 甲状腺肿大治疗预测值前五的药物

药物名称	实际效果
甲硫咪唑	用于治疗甲状腺功能亢进症状,甲状腺危象等
硫酸羟脲	临床效果用于恶性黑色素瘤、胃癌、恶性淋巴瘤等
乙胺丁醇	影响结核杆菌菌体核糖核酸的合成而具有杀抑作用
双硫仑	是一种戒酒的药物,服用后即使饮用少量酒精也会产生不适
丙硫氧嘧啶	药效主要通过抑制甲状腺内过氧化物酶系统,阻止甲状腺内酪氨酸的碘化及碘化酪氨酸的缩合,从而抑制甲状腺激素的合成

5 结束语

提出了一种基于多相似度融合的药物重定位推荐算法(MSF算法),实验结果显示MSF算法与SLAMS算法和DRCFFS相比,在三种指标(精确率、召回率和F-score)上的表现更好。综上所述,提出的MSF算法优于传统的药物重定位算法,可以预测有治疗效果的药物-疾病组合,为更好地发挥药物治疗效果、提升药物利用价值起到一定的作用。但是,MSF算法也有一定的缺陷,比如:通过试探法计算相似度融合的权值耗时较长,计算疾病相似度所采用的数据源较少。在后续的研究过程中应当对相似度融合的方法进行改进并且在疾病相似度的计算上使用多种数据源。

参考文献:

- [1] 徐燕,郭颖. 药物重定位的研究策略[J]. 中国新药杂志,2017,26(11):1253-1258.
- [2] 王可鉴,石乐明,贺林,等. 中国药物研发的新机遇:基于医药大数据的系统性药物重定位[J]. 科学通报,2014,59(18):1790-1796.
- [3] 汪浩,王海平,吴信东,等. 药物-疾病关系预测:一种推荐系统模型[J]. 中国药理学通报,2015,31(12):1770-1774.
- [4] 张永祥,程肖蕊,周文霞. 药物重定位——网络药理学的重要应用领域[J]. 中国药理学与毒理学杂志,2012,26(6):779-786.
- [5] 谢达菲,李鹏,李非,等. 药物重定位的计算分析方法[J]. 生物化学与生物物理进展,2012,39(11):1029-1036.
- [6] 马浚. 基于深度学习的药物重定位方法研究[D]. 兰州:兰州大学,2019.
- [7] LUO H, WANG J, LI M, et al. Drug repositioning based on comprehensive similarity measures and Bi-random walk algorithm[J]. Bioinformatics,2016,32(17):2664-2671.
- [8] LUO H, LI M, WANG S, et al. Computational drug repositioning using low-rank matrix approximation and randomized algorithms[J]. Bioinformatics,2018,34(11):1904-1912.
- [9] ZHANG P, AGARWAL P, OBRADOVIC Z. Computational drug repositioning by ranking and integrating multiple data sources[C]//Joint European conference on machine learning and knowledge discovery in databases. Prague, Czech Republic; Springer,2013:579-594.
- [10] 朱桥. 基于多元信息融合的药物重定位算法研究[D]. 长沙:湖南大学,2018.
- [11] LI J, LU Z. A new method for computational drug repositioning using drug pairwise similarity[C]//IEEE international conference on bioinformatics & biomedicine. Philadelphia, PA, USA; IEEE,2012:1-4.
- [12] 章啸. 协同过滤算法在药物重定位中的研究与应用[D]. 上海:东华大学,2017.
- [13] 林耀进,张佳,林梦雷,等. 基于协同过滤的药物重定位算法[J]. 南京大学学报:自然科学版,2015,51(4):834-841.
- [14] ZHANG J, LI C, LIN Y, et al. Computational drug repositioning using collaborative filtering via multi-source fusion[J]. Expert Systems with Applications,2017,84:281-289.
- [15] 刘杰,金柳,景波. 基于药物和疾病特征关联的药物重定位混合推荐算法[J]. 计算机应用研究,2020,37(3):672-675.
- [16] CHIANG A P, BUTTE A J. Systematic evaluation of drug-disease relationships to identify leads for novel drug uses[J]. Clinical Pharmacology & Therapeutics,2009,86(5):507-510.
- [17] DUDLEY J T, DESHPANDE T, BUTTE A J. Exploiting

- drug-disease relationships for computational drug repositioning[J]. *Briefings in Bioinformatics*, 2011, 12(4):303-311.
- [18] LI J, ZHU X, CHEN J Y, et al. Building disease-specific drug-protein connectivity maps from molecular interaction networks and pubmed abstracts[J]. *PLoS Computational Biology*, 2009, 5(7):e1000450.
- [19] CAMPILLOS M, KUHN M, GAVIN A C, et al. Drug target identification using side-effect similarity[J]. *Science*, 2008, 321(5886):263-266.
- [20] 徐蕾, 杨成, 姜春晓, 等. 协同过滤推荐系统中的用户博弈[J]. *计算机学报*, 2016, 39(6):1176-1189.
- [21] 杨佳莉, 李直旭, 许佳捷, 等. 一种自适应的混合协同过滤推荐算法[J]. *计算机工程*, 2019, 45(7):222-228.
- [22] 冷亚军, 陆青, 梁昌勇. 协同过滤推荐技术综述[J]. *模式识别与人工智能*, 2014, 27(8):720-734.
- [23] OLIVIER B. The unified medical language system (UMLS): integrating biomedical terminology[J]. *Nucleic Acids Research*, 2004, 32(Database issue):D267-D270.
- [24] WANG Y, XIAO J, SUZEK T O, et al. PubChem: a public information system for analyzing bioactivities of small molecules[J]. *Nucleic Acids Research*, 2009, 37(Web Server issue):W623-W633.
- [25] ROLF A, AMOS B, WU C H, et al. UniProt: the universal protein knowledgebase[J]. *Nucleic Acids Research*, 2004, 32(Database Issue):D115-D119.
- [26] KUHN M, CAMPILLOS M, LETUNIC I, et al. A side effect resource to capture phenotypic effects of drugs[J]. *Molecular Systems Biology*, 2010, 6(1):343.
- [27] BADER J C, LAKOTA E A, DALE G E, et al. Pharmacokinetic-pharmacodynamic evaluation of ertapenem for patients with hospital-acquired or ventilator-associated bacterial pneumonia[J]. *Antimicrobial Agents and Chemotherapy*, 2019, 63(6):318-319.
- [28] 头孢克洛缓释片多中心研究小组. 头孢克洛缓释片治疗成人急性鼻窦炎临床研究[J]. *中国耳鼻咽喉头颈外科*, 2004, 11(4):233-236.
- [29] HIRAKI Y, YOSHIDA M, MASUDA Y, et al. Successful treatment of skin and soft tissue infection due to carbapenem-resistant *Acinetobacter baumannii* by ampicillin-sulbactam and meropenem combination therapy[J]. *International Journal of Infectious Diseases*, 2013, 17(12):e1234-e1236.
- [30] 杨海辰. 甲硫咪唑联合丙基硫氧嘧啶治疗甲状腺功能亢进症临床疗效研究[J]. *临床合理用药杂志*, 2014, 7(35):92.
- [31] RAMIREZ G, ABLES M F, BUTCHER D E, et al. Evaluation of the hypothalamic-hypophysial, thyroid, and gonadal axes before and after disulfiram administration in patients with chronic alcoholism[J]. *Southern Medical Journal*, 1988, 81(11):1407-1411.
-
- (上接第 76 页)
- [6] BASSIL Y, ALWANI M. OCR post-processing error correction algorithm using Google online spelling suggestion[J]. *Journal of Emerging Trends in Computing & Information Sciences*, 2012, 3(1):90-99.
- [7] GUPTA M R, JACOBSON N P, GARCIA E K. OCR binarization and image pre-processing for searching historical documents[J]. *Pattern Recognition*, 2007, 40(2):389-397.
- [8] 柴晨阳. 印刷体汉字识别处理技术综述[J]. *计算机光盘软件与应用*, 2014(24):141-142.
- [9] 弓耀辉. 图像文字识别中的预处理技术研究综述[J]. *信息通信*, 2017(9):291-292.
- [10] 杨丽娟, 李利. 基于双线性插值的内容感知图像缩放算法仿真[J]. *计算机仿真*, 2019, 36(12):244-248.
- [11] LI Jiang, WAN Heyang, SHANG Songhao. Comparison of interpolation methods for mapping layered soil particle-size fractions and texture in an arid oasis[J]. *CATENA*, 2019, 190:104514.
- [12] 闫三虎, 胡卫东, 罗小平. 改进二值化算法在 QR 码识别中的应用[J]. *计算机系统应用*, 2011, 20(6):165-168.
- [13] YAGI K, SUGIURA Y, HASEGAWA K, et al. Gait measurement at home using a single RGB camera[J]. *Gait & Posture*, 2019, 76:136-140.
- [14] 陈海峰, 丁丽丽. 二值化图像的灰度处理算法研究[J]. *电脑与电信*, 2019(7):34-38.
- [15] 白元明, 孔令成, 张志华, 等. 基于改进 OTSU 算法的快速作物图像分割[J]. *江苏农业科学*, 2019, 47(24):231-236.