

药物靶标作用关系预测结果评价及查询验证

余冬华¹ 郭茂祖^{1,2,3} 刘晓燕¹ 程 爽⁴
¹(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)
²(北京建筑大学电气与信息工程学院 北京 100044)
³(建筑大数据智能处理方法研究北京市重点实验室(北京建筑大学) 北京 100044)
⁴(中国工程物理研究院材料所 四川绵阳 621900)
(donghuayu@hit.edu.cn)

Predicted Results Evaluation and Query Verification of Drug-Target Interaction

Yu Donghua¹, Guo Maozu^{1,2,3}, Liu Xiaoyan¹, and Cheng Shuang⁴
¹(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)
²(School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044)
³(Beijing Key Laboratory of Intelligent Processing for Building Big Data, Beijing 100044)
⁴(Institute of Materials, China Academy of Engineering Physics, Mianyang, Sichuan 621900)

Abstract The drug-target interaction prediction is one of the important assistant approaches in drug discovery and design, however, experimental identification and validation of potential drug-target encoded by the human genome is both costly and time-consuming. Therefore, querying and validating the predicted drug-target interaction in databases is an important assessment of prediction methods. In this paper, the query and validation method of drug-target interaction named as DTcheck (drug-target check) is developed and designed with Web spider based on KEGG, DrugBank, ChEMBL databases, which realizes efficient query and validation function for drug-target pair providing both KEGG DRUG ID and KEGG GENES ID. ID mapping function is also designed in DTcheck, which can map Uniprot ID from DrugBank and ChEMBL into KEGG GENE ID. DTcheck expands 907, 766, 458, 40 pairs of new drug-target interaction for Enzyme, IC (ion channel), GPCR (G-protein-coupled receptor), NR (nuclear receptor) standard datasets, respectively. Moreover, combined with query and validation result, the analysis of the prediction results of the BLM (bipartite local models) method shows that evaluation of Top *N* is more reasonable than AUC (area under curve) value for the prediction method of drug-target interaction. It also shows that the BLMD method with low AUC value is superior to the BLMmax method with high AUC value in predicting the drug-target interaction.

Key words drug-target interaction prediction; query and validation; drug-target data; AUC evaluation; Top *N* evaluation

收稿日期:2018-12-18;修回日期:2019-01-28
基金项目:国家自然科学基金项目(61871020,61571164,61671189);北京市教委科技计划重点项目(KZ201810016019);北京建筑大学市属高校基本科研业务费专项资金(X18197,X18198,X18203)
This work was supported by the National Natural Science Foundation of China (61871020, 61571164, 61671189), the Scientific Research Key Project of Beijing Educational Committee (KZ201810016019), and the Special Funds for Basic Research Business in Municipal Universities of Beijing Architectural University (X18197, X18198, X18203).
通信作者:郭茂祖(guomaozu@bucea.edu.cn)

摘要 药物靶标作用关系预测是一种重要的辅助药物研发手段,而生物实验验证药物靶标作用关系耗钱耗时,因此,在数据库中查询验证预测的药物靶标作用关系是对预测方法的重要评价.基于 KEGG, DrugBank, ChEMBL 这 3 个数据库,利用爬虫获取信息的方式设计开发了药物靶标作用关系查询验证方法 DTcheck(drug-target check),实现了对于提供 KEGG DRUG ID 及 KEGG GENES ID 的药物靶标对的高效查询验证功能,并利用 DTcheck 分别为 Enzyme, IC(ion channel), GPCR(G-protein-coupled receptor), NR(nuclear receptor)四个标准数据集扩充新增药物靶标作用关系 907,766,458,40 对.此外,结合 DTcheck 查询验证,以 BLM(bipartite local models)方法为例分析了预测结果的评价问题,结果表明,采用 AUC(area under curve)值评价药物靶标作用关系预测方法没有 Top N 评价合理,且 AUC 值低的 BLMD 方法在预测新的药物靶标作用关系时优于 AUC 值高的 BLMmax 方法.

关键词 药物靶标作用关系预测;查询验证;药物靶标数据集;AUC 评价;Top N 评价

中图法分类号 TP391

药物研发是医药界的驱动力,也引起了学术界的研究热潮,计算机辅助药物设计是对传统的实验型药物发现的补充,加速了新药开发进程,缩短药物研发时间,降低药物研发成本,同时提升药物研发成功率.

网络药理学(network pharmacology)^[1-2]和药物重定位(drug repositioning)^[3-4]颠覆了传统的药物研发理念,完善了药物靶标作用关系(drug-target interaction)预测的理论基础,为计算机辅助药物设计提供了新思路,加速了药物发现与设计.随着公开数据库 KEGG^[5], DrugBank^[6], ChEMBL^[7] 等中的可用数据增多,很多预测药物靶标作用关系的计算方法也被提出^[8-13].其主要思路为:建立合适的药物靶标数据,构建作用关系预测方法并做出合理的评价,然后进行实际作用关系预测.这些计算方法弥补了实验方法中的时间消耗和高昂成本投入的缺陷.

Yamanishi 等人^[14]在 2008 年基于 KEGG 数据库建立了 Enzyme, IC(ion channel), GPCR(G-protein-coupled receptor), NR(nuclear receptor)四个药物靶标作用关系数据集(包括药物靶标作用关系矩阵、药物-药物相似度矩阵、靶标-靶标相似度矩阵),其中,已经验证的药物靶标作用关系用 1 表示,未知作用关系的药物靶标对(unknown interaction)用 0 表示,受到广泛接受及认同,Google 学术显示引用次数为 467 次,被认为是药物靶标作用关系预测模型的标准检验数据集.2009 年, Bleakley 等人^[15]在上述 4 个数据集的基础上,提出了经典的二分局部模型(bipartite local models, BLM)药物靶标作用关系预测方法,并把其当作二分类模型,采用 ROC(receiver operating characteristic curve)曲线及曲线下方面积(area under curve, AUC)作为该方法的评价标准.

该数据集的建立促使了很多预测方法的诞生,而 AUC 评价标准也对后续方法产生了重要影响.然而,该评价需要将未知作用关系的药物靶标对视为负例,这是一种不合理的假设,因为很多未知的作用关系可能实际上存在相互作用关系,只是到目前为止,仍没有被实验所证实.况且,预测方法的目标就是需要将潜在的作用关系预测出来,即将作用关系 0 预测成 1.尽管如此,这种不合理的评价方式也一直在延续.作为一种弥补手段,会将预测结果中极有可能的药物靶标对(比如 Top N,即按照预测的药物靶标对作用关系可能性从大到小排序,取前 N 对)在数据库中查询验证. Laarhoven 等人^[16]提出 GIP-RLS(Gaussian interaction profile-regularized least squares)方法后,用 AUC 及 AUPR(area under precision-recall curve)进行了评价.对于需要预测的药物靶标作用关系,仅对每个数据集 Top 20 查验过 KEGG, DrugBank, ChEMBL 数据库. Gonen^[17]提出的 KBMF2K(kernelized Bayesian matrix factorization with twin kernels)预测方法,用 AUC 进行评价,且仅对每个数据集 Top 5 查验过 KEGG, DrugBank, ChEMBL 数据库. Mei 等人^[18]提出 BLMNII(bipartite local models with neighbor-based interaction-profile)预测方法后,用 AUC 及 AUPR 进行了评价,而没有对预测的药物靶标对进行数据库的查询验证,也就没有验证所预测的作用关系中是否存在已经被实验验证的药物靶标对. Hao 等人^[19]提出的 RLS-KF(regularized least squares with kernel fusing)预测方法,用 AUC 及 AUPR 进行评价,仅对 NR 数据集中预测 Top 100 对药物靶标进行数据库查询验证,虽然其 AUC 值接近于 1,对于二分类模型来说,接近于 1 的 AUC 值表示该模型

预测接近于完全正确,但对于 NR 数据集来说,直到 Top 100 时,才有 31 对药物靶标已经获得了生物化学实验验证,对于目前可以查询到的已验证的药物靶标对来说,仍然有 9 对不包含在内,这也表明,如此之高的预测精度,在实际预测中远未达到理想结果.Luo 等人^[20]提出 DTINet(network integration pipeline for drug-target interaction)预测方法,也用 AUC 及 AUPR 进行了模型评价,并对预测结果的 Top 150 药物靶标对进行了数据库查询验证,特别地,选择了极有可能具有作用关系的 COX inhibitory 进行了生物化学实验验证,实验结果证实了 DTINet 模型的预测准确.Hao 等人^[21]提出 DNILMF(dual-network integrated logistic matrix factorization)预测方法,用 AUC 及 AUPR 评价该模型,采用与 Yamanishi 等人^[14]相似的方式收集了一个新数据集,在最后的预测结果中,仅对 Top 5 的药物靶标对进行了查询验证.Olayan 等人^[22]提出 DDR(novel method that improves the drug-target interaction prediction accuracy)预测方法,在方法的评价方面,舍弃了 AUC 评价,而选择加权平均 AUPR 评价,并且对预测中的 Top 25 进行了查询验证.Peng 等人^[23]提出 PreNNDS (drug-target identification mode by integrating neighbor interaction profiles, nonnegative matrix factorization, discriminative low-rank representation, and sparse representation classification into a unified framework)预测方法,仅采用 AUPR 进行评价,与其他文章按照预测得分(或概率)选择 Top N 个药物靶标对进行查询验证不同,Peng 等人选择了 2 个靶标(ID 分别为 hsa1132 及 hsa1124)所对应的 Top 20 的药物进行查询验证,同样地,选择了 2 个药物(ID 分别为 D00255 及 D00195)所对应的 Top 20 的靶标进行查询验证.Liu 等人^[24]提出 NRLMF(neighbor-hood regularized logistic matrix factorization)预测方法后,用 AUC 及 AUPR 进行评价,该文在 KEGG, DrugBank,ChEMBL,Matador 数据库中查询验证了 4 个数据集中 Top 1000 的药物靶标对,因此,其给出了 Top 10,Top 30,Top 50 的评价结果.然而,其并未提供可以快速高效地对如此大量的未知作用关系药物靶标对查询验证的方法,无法减轻后续学者研究该问题的工作量.实际上,对于 Enzyme,IC, GPCR,NR 四个数据集来说,Top 1000 仍然是未知药物靶标作用关系中很小的一部分.多种方法的查询验证 Top N 值总结见表 1:

Table 1 The Checked Top N Values of Various Methods
表 1 各种方法查询验证 Top N 值

Method	Top N	Reference
GIP-RLS	20	Ref [16]
KBMF2K	5	Ref [17]
RLS-KF	100	Ref [19]
DTINet	150	Ref [20]
DNILMF	5	Ref [21]
DDR	25	Ref [22]
PreNNDS	20	Ref [23]
NRLMF	1 000	Ref [24]

因此,如何验证这些预测结果的准确性是研究者面临的一个挑战,在目前有限条件下,考虑到时间消耗及实验成本,不可能对所有未知的药物靶标作用进行生物化学实验验证.综合现有的各种评价方法,一种可行的方式就是用 AUC 及 AUPR 验证预测方法可行性及预测精度,并利用数据共享,如公开数据库 DrugBank,KEGG,ChEMBL,提供的很多药物靶标之间的作用关系数据,将预测的作用关系逐对在数据库中进行查询验证,确认该未知的作用关系是否已经被相关实验验证.这种方式得到了学术界普遍认可.然而,现有的方法也仅仅查询验证 Top N 药物靶标对,因为查询验证是一个重复且耗时的工作,且未知作用关系的药物靶标对占预测总数的绝大多数,而更为合理的验证预测结果的方式为:对所有未知作用关系的药物靶标对在数据库上查询验证一遍.

针对上述问题,本文不仅为药物靶标作用关系验证提供一种高效、便捷、可行的查询验证方法 DTcheck(drug-target check),并利用 DTcheck 查询的结果对 4 个标准数据集及 AUC 评价做了分析,主要贡献归纳为

- 1) 结合网络爬虫特性,设计 DTcheck 查询验证方法并实现.
- 2) 基于 KEGG,DrugBank,ChEMBL 数据库,通过 DTcheck 将 4 个数据集中已经验证的药物靶标作用关系补充完整.
- 3) 对于(批量)提供 KEGG DRUG ID 及 KEGG GENES ID 的药物靶标对,DTcheck 可以在线查询并返回其在 KEGG,DrugBank,ChEMBL 数据库中的验证结果.
- 4) 利用 DTcheck 查询的结果,重新分析了 BLM 方法的预测结果,探讨了 ROC 曲线评价药物

靶标作用关系的不合理性,并引入 Top N 评价指标,实际结果表明 AUC 值低的 BLMD 在预测新的药物靶标作用关系时要优于 AUC 值高的 BLMDmax,给 AUC 评价的不合理性提供了依据。

1 DTcheck 查询验证关键问题及解决思路

Yamanishi 等人^[14]提供的标准数据集中采用 KEGG ID 标注药物及靶标,而 DrugBank, ChEMBL 有其自己的 ID 体系,如药物 D02441(KEGG DRUG ID)在 DrugBank 对应的 ID 为 DB00311,在 ChEMBL 中对应的 Compound ID 为 CHEMBL18,

而靶标 hsa766(KEGG GENE ID)需要与 UniProt 数据库^[25]中的 P43166(UniProt ID)对应起来,通过 DrugBank 中的药物或者 ChEMBL 中的药物(或化合物)查询到的对应靶标,会提供对应的 UniProt ID.因此,其中一个关键问题就是需要将 3 个数据库中提供的不同 ID 合理对应起来。

药物靶标作用关系是成对出现的,如 D02441~ hsa766,因此,在查询验证过程中,可通过 D02441 在 KEGG DRUG 中查找到对应靶标 hsa766,也可在 DrugBank 中通过 DB00311 查找到对应靶标 P43166(UniProt ID),而 hsa766 与 P43166 代表同一个靶标,即不同的 ID 代表同一个对象.相对应地,

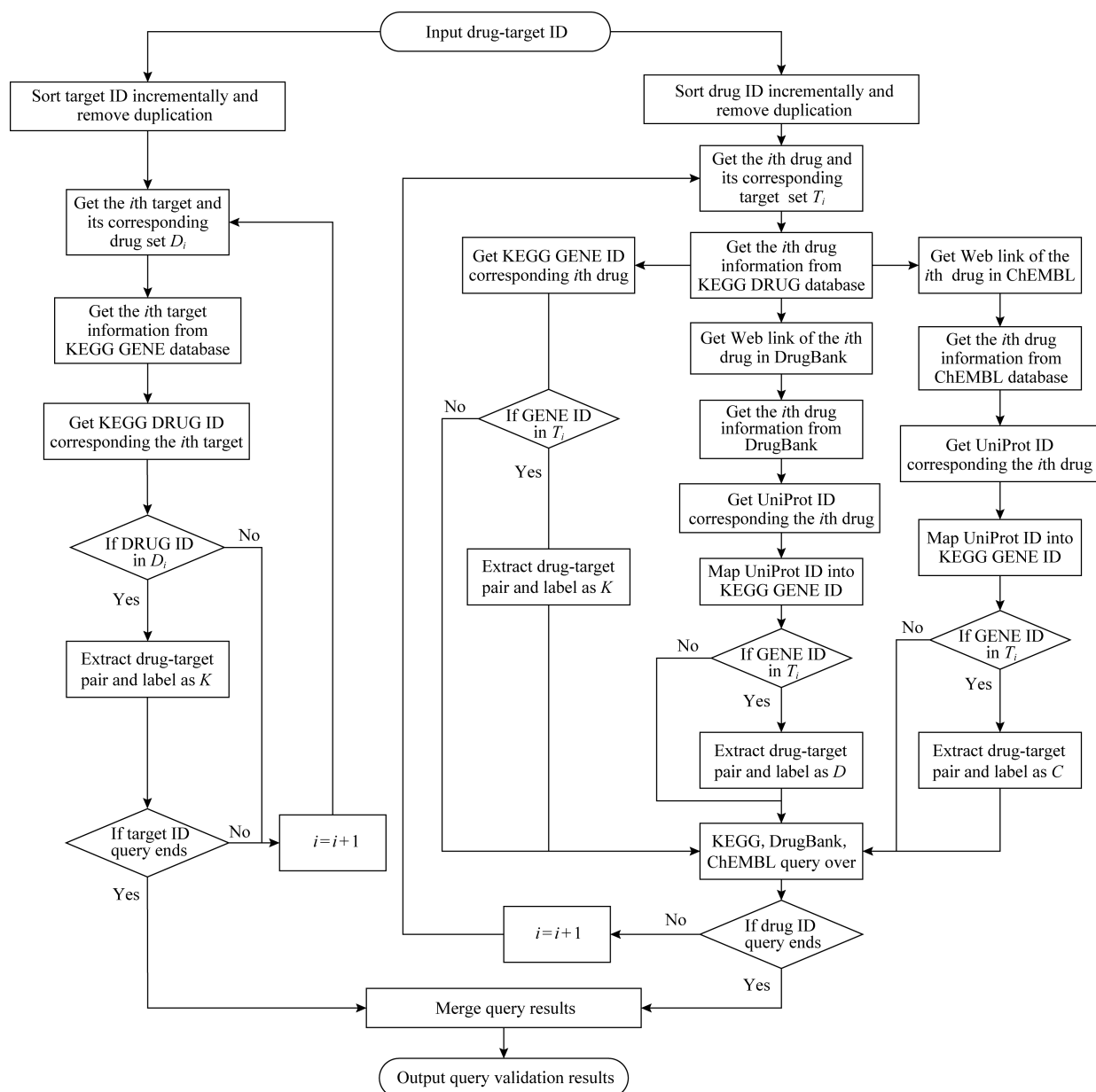


Fig. 1 Technical route and flow of DTcheck

图 1 DTcheck 技术路线及流程

也可以从靶标 hsa766 出发,从 KEGG GENE 中查找到药物 D02441.因此,另一个关键问题就是在查询验证中,需要对药物靶标对进行双向查询验证.

由于 KEGG DRUG 中提供了该药物在 ChEMBL 及 DrugBank 中的对应 ID 和相应的网页链接,因此,利用爬虫抓取数据的特性,读取该药物在 KEGG DRUG 中所提供的对应靶标 ID 后,需要额外读取该药物在其他 2 个数据库中的 ID 及链接.而在 KEGG GENE 中查询的数据仅是该靶标是否有对应的药物即可.因此,在药物靶标对的双向查询验证中,可以优先查询靶标,然后查询药物.

从 KEGG DRUG 中读取该药物在 ChEMBL, DrugBank 中的链接后,可以独立地进行查询,然后将查询结果合并,最后核对待验证的药物靶标对. KEGG DRUG 提供的 ChEMBL 链接可能会有多个,需要遍历所有可能链接,读取该链接返回的 ChEMBL Target 信息及其跳转链接,就可以查询到该药物对应靶标的 UniProt ID.从 KEGG DRUG 还可以读取该药物的 DrugBank 跳转链接,利用爬虫访问该链接,获取返回信息,从其返回信息中读取 Targets, Enzymes, Carriers, Transporters 四个版块信息所提供的 UniProt ID 即可.这样就可以获取该药物在 ChEMBL, DrugBank 两个数据库中所对应靶标的 UniProt ID.

靶标的 UniProt ID 与 KEGG GENE ID 并不一样,后者也是 4 个标准数据集所提供的 ID,因此,需要先将 UniProt ID 转换成 KEGG GENE ID 才能核查待验证的药物靶标对的作用关系是否已经被证实.此时可以借助 UniProt 数据库提供的 ID mapping 功能,将靶标的 UniProt ID 转换成 KEGG GENE ID,这一步的实现需要借助爬虫发送数据请求并获取返回信息的特性.

2 DTcheck 查询验证技术路线

解决了 DTcheck 查询验证中遇到的关键问题,本文设计了如图 1 所示的技术路线流程,其代码链接为 <https://github.com/Yu123456/DTcheck>.从图 1 中可以看出,双向查询分成 2 个部分独立查询,最终需要将 2 部分查询结果合并,而 3 个数据库之间,需要先从 KEGG DRUG 中获取 DrugBank, ChEMBL 数据库的跳转链接,然后依次获取相应数据库中的靶标信息,而从这 2 个数据库中返回的靶

标是 UniProt ID,为了与待验证的药物靶标对核对,需要增加一步 ID 转换,把 UniProt ID 转换成 KEGG GENE ID.

3 预测结果评价及查询验证分析

Yamanishi 等人^[14]于 2008 年建立了 Enzyme, IC, GPCR, NR 四个标准数据集,如表 2 所示.其中, drug-target interactions 代表当时已经验证的药物靶标作用关系;unknown interaction 代表当时未知的药物靶标作用关系;从最后 2 行已知与未知药物靶标作用关系具体数值来看,未知作用关系远多于已知作用关系.该数据集从建立至今已有 10 年之久,在这期间有很多作用关系得到生物实验验证,本文采用 DTcheck 查询 KEGG, DrugBank, ChEMBL 三个数据库,将新增的作用关系补充完整.伴随这 4 个数据集, Bleakley 等人^[15]于 2009 年提出了经典的 BLM 预测方法,并用 AUC 进行评价,本文将依据 DTcheck 查询到的结果,重新探讨 AUC 评价的不合理性,并采用 Top N 中已经被验证的药物靶标对的数目重新分析预测结果.

Table 2 Drug-Target and Their Interactions

表 2 药物靶标及其作用关系信息

Items	Enzyme	IC	GPCR	NR
Number of Drugs	445	210	223	54
Number of Targets	664	204	95	26
Drug-Target Interactions	2 926	1 476	635	90
Unknown Interactions	292 554	41 364	20 550	1 314

3.1 数据集已验证作用关系补充

经过 DTcheck 的查询,每个数据集都可以找到新验证的药物靶标对,如表 3 所示.从增长率 (increased ratio) 来看,新药物靶标关系增加最快的数据集为 GPCR;新发现的药物靶标对数最多的却是 Enzyme 数据集,达到 907 对;虽然 NR 药物靶标对的数目最少,但也将近增加一半,达到 40 对.

Table 3 Drug-Target and Their Increased Interactions

表 3 药物靶标数据集新增作用关系

Items	Enzyme	IC	GPCR	NR
Number of Drugs	445	210	223	54
Number of Targets	664	204	95	26
Drug-Target Interactions	2 926	1 476	635	90
Increased Interactions	907	766	458	40
Increased Ratio	0.310	0.519	0.721	0.444

表 4 给出了 NR 数据集的 DTcheck 查询到的 40 对已经验证的药物靶标作用关系,其中,DrugID, TargetID 分别表示药物、靶标的 ID;C,D 分别表示在 ChEMBL,DrugBank 数据库中查询到该药物靶标对作用关系;K 需要做一个区分,在 KEGG 栏中,表示通过药物找到了对应的靶标,在 KEGGh 栏中,则表示通过靶标找到了对应的药物;空白表示对应数据库中未查询到相关信息.表 4 中出现大量空白,表明 3 个数据库之间的信息并没有完全同步,因此,必须查询每一个数据库,避免遗漏.对于 NR 数据集来说,DrugBank 提供的已验证作用关系最多,其次是 KEGG,最后才是 ChEMBL.

Table 4 New Increased Drug-Target Interactions of NR
表 4 NR 数据集新增药物靶标关系

DrugID	TargetID	ChEMBL	DrugBank	KEGG	KEGGh
D00348	hsa5916	C		K	K
D00348	hsa5915	C		K	K
D00690	hsa2908	C	D	K	K
D00040	hsa9970		D		
D00040	hsa7421		D		
D00066	hsa2908		D		
D00066	hsa367		D		
D00066	hsa8856		D		
D00066	hsa4306		D		
D00067	hsa2100		D	K	K
D00067	hsa367		D		
D00075	hsa4306		D		
D00075	hsa2099		D		
D00105	hsa2104		D		
D00105	hsa8856		D		
D00312	hsa2100		D	K	K
D00312	hsa367		D		
D00327	hsa2908		D		
D00443	hsa2908		D		
D00443	hsa367		D		
D00443	hsa8856		D		
D00443	hsa5241		D		
D00506	hsa8856		D		
D00554	hsa8856		D		
D00565	hsa5468		D		
D00565	hsa8856		D		
D00577	hsa367		D		
D00577	hsa8856		D		
D00585	hsa8856		D		

Continued (Table 4)					
DrugID	TargetID	ChEMBL	DrugBank	KEGG	KEGGh
D00586	hsa8856		D		
D00596	hsa6258		D		
D00596	hsa6257		D		
D00596	hsa6256		D		
D00596	hsa5465		D		
D00950	hsa367		D		
D00951	hsa2099		D		
D00954	hsa367		D		
D05341	hsa5465		D		
D00898	hsa2100			K	K
D00962	hsa2100			K	K

Note: The blank indicates that the interaction relationship has not been queried yet.

由于 Enzyme, GPCR, IC 这 3 个数据集查询到的新增作用关系数目较多,限于篇幅,不宜在文中直接给出,其查询的相应结果见 <https://github.com/Yu123456/Drug-targetChecked>.

3.2 BLM 方法预测结果查询验证及评价分析

BLM 可以细分为 3 个子方法,即 BLMd, BLMt, BLMmax,其中 BLMmax 是取 BLMd, BLMt 二者预测的较大值.Bleakley 等人^[15]依据 AUC(如表 5 所示,较优者黑体表示)评价指出,在 3 个方法中, BLMmax 性能最好, BLMd 性能最差.

Table 5 AUC of 4 Datasets
表 5 AUC 评价值

Method	Enzyme	IC	GPCR	NR
BLMd	83.1	74.5	82.3	81.2
BLMt	94.2	93.5	87.2	53.6
BLMmax	97.3	97.0	95.3	85.8

Note: The boldface represents the best one of the three methods.

从 AUC 评价来看, BLMmax 非常显著地优于 BLMd,因此, BLMmax 对新药物靶标作用关系的预测结果也应该优于 BLMd,然而,表 6 却展现出截然相反的结论.在 Top N 评价中,从相同的 N 中找到更多验证关系者性能更优,在表 6 中较优者已经加粗表示.在 Top 50 之前的每一个 Top N 中, BLMd 预测正确的药物靶标对均多于 BLMmax;在 Top 50 至 Top 1 000 之间,才出现后一种方法优于前一种方法的趋势,如 Top 1 000 时, BLMmax 预测正确的药物靶标对的数目多于 BLMd.然而,开发药物靶标作用关系预测方法的初衷是为实验生物化学家提供

有实质指导作用的建议,即尽可能少的实验下找到真正的作用关系,因为用生物实验验证药物靶标作用关系既耗时又费钱,因此,Top 50 评价会是

一个更优的选择.从这个方面来说,BLMd 优于 BLMmax,这也表明仅用 AUC 评价药物靶标预测模型不合理.

Table 6 Comparing New Interaction of Top N Between BLMmax and BLMd
表 6 BLMmax 与 BLMd 中 Top N 中新作用关系比较

Top N	Enzyme		IC		GPCR		NR	
	BLMd	BLMmax	BLMd	BLMmax	BLMd	BLMmax	BLMd	BLMmax
Top 10	4	2	10	3	8	6	4	3
Top 20	6	6	15	7	13	11	6	4
Top 30	9	8	20	13	18	17	7	6
Top 50	20	11	25	22	34	23	9	10
Top 100	27	18	42	52	50	42	12	13
Top 200	41	38	68	95	75	75	18	22
Top 1 000	90	96	165	250	163	186	36	38

Note: The boldface represents the better interaction between BLMd and BLMmax.

4 总 结

本文利用爬虫特性,设计并实现了 DTcheck 在 KEGG,DrugBank,ChEBML 数据库上高效、便捷的药物靶标作用关系查询验证,利用 DTcheck 实现对标准数据集新增作用关系的扩展补充,基于 BLM 方法的预测结果,说明了 AUC 评价药物靶标作用关系预测的不合理性,及 Top N 可以给出更加合理的评价.本文提供的药物靶标对在线查询验证思路及实现,大大节省研究者查询验证时间,使得在方法的验证中不仅仅限于每个数据集仅查询预测排序靠前的少数药物靶标对,有利于更准确地评价药物靶标预测方法.

DTcheck 也有待于继续扩展,目前仅仅查询 KEGG,DrugBank,ChEMBL 三个数据库,实际上,Matador,SuperTarget 等数据库也可以查询到一些相关作用关系.我们将会继续对此工作扩展,以期 DTcheck 更具有广泛性.

参 考 文 献

[1] Hopkins A. Network pharmacology: The next paradigm in drug discovery [J]. Nature Chemical Biology, 2008, 4(11): 682-690

[2] Liu Ailin, Du Guanhua. Network pharmacology: New guidelines for drug discovery [J]. Acta Pharmaceutica Sinica, 2010, 45(12): 1472-1477 (in Chinese)
(刘艾林,杜冠华. 网络药理学: 药物发现的新思[J].药学报, 2010, 45(12): 1472-1477)

[3] Ashburn T, Thor K. Drug repositioning: Identifying and developing new uses for existing drugs [J]. Nature Reviews Drug Discovery, 2004, 3(8): 673-683

[4] Chong C, Sullivan D. New uses for old drugs [J]. Nature, 2007, 448: 645-646

[5] Kanehisa M, Goto S, Sato Y, et al. KEGG for integration and interpretation of large-scale molecular data sets [J]. Nucleic Acids Research, 2012, 40(D1): D109-D114

[6] Wishart D, Knox C, Guo Anchi, et al. DrugBank: A knowledgebase for drugs, drug actions and drug targets [J]. Nucleic Acids Research, 2008, 36(Suppl1): D901-D906

[7] Gaulton A, Bellis L, Bento A, et al. ChEMBL: A large-scale bioactivity database for drug discovery [J]. Nucleic Acids Research, 2012, 40(D1): D1100-D1107

[8] Mousavian Z, Masoudi N. Drug-target interaction prediction via chemogenomic space: Learning-based methods [J]. Expert Opinion on Drug Metabolism & Toxicology, 2014, 10(9): 1273-1287

[9] Chen Xing, Yan Chenggang, Zhang Xiaotian, et al. Drug-target interaction prediction: Databases, Web servers and computational models [J]. Briefings in Bioinformatics, 2016, 17(4): 696-712

[10] Cheng Tiejun, Hao Ming, Takeda T, et al. Large-scale prediction of drug-target interaction: A data-centric review [J]. The AAPS Journal, 2017, 19(5): 1264-1275

[11] Anusuya S, Keshherwani M, Priya K, et al. Drug-target interactions: Prediction methods and applications [J]. Current Protein & Peptide Science, 2018, 19(6): 537-561

[12] Hao Ming, Bryant S, Wang Yanli. Open-source chemogenomic data-driven algorithms for predicting drug-target interactions [J/OL]. Briefings in Bioinformatics, 2018 [2018-07-30]. <https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bby010/4840612>

- [13] Guo Quan, Wang Xicheng, Li Chunlian. Research and development of molecule docking design based on grid [J]. Journal of Computer Research and Development, 2004, 41(12): 2054-2059 (in Chinese)
(郭权, 王希诚, 李纯莲. 药物分子对接中应用网格的研究与进展[J]. 计算机研究与发展, 2004, 41(12): 2054-2059)
- [14] Yamanishi Y, Araki M, Gutteridge A, et al. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces [J]. Bioinformatics, 2008, 24(13): 1232-1240
- [15] Bleakley K, Yamanishi Y. Supervised prediction of drug-target interactions using bipartite local models [J]. Bioinformatics, 2009, 25(18): 2397-2403
- [16] Laarhoven T, Nabuurs S, Marchiori E. Gaussian interaction profile kernels for predicting drug-target interaction [J]. Bioinformatics, 2011, 27(21): 3036-3043
- [17] Gonen M. Predicting drug-target interactions from chemical and genomic kernels using Bayesian matrix factorization [J]. Bioinformatics, 2012, 28(18): 2304-2310
- [18] Mei Jianping, Kwoh C, Yang Peng, et al. Drug-target interaction prediction by learning from local information and neighbors [J]. Bioinformatics, 2013, 29(2): 238-245
- [19] Hao Ming, Wang Yanli, Bryant S. Improved prediction of drug-target interactions using regularized least squares integrating with kernel fusion technique [J]. Analytica Chimica Acta, 2016, 909: 41-50
- [20] Luo Yunan, Zhao Xinbin, Zhou Jingtian, et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information [J]. Nature Communications, 2017, 8: 573
- [21] Hao Ming, Bryant S, Wang Yanli. Predicting drug-target interactions by dual-network integrated logistic matrix factorization [J]. Scientific Reports, 2017, 7: 40376
- [22] Olayan R, Ashoor H, Bajic V. DDR: Efficient computational method to predict drug-target interactions using graph mining and machine learning approaches [J]. Bioinformatics, 2018, 34(7): 1164-1173

- [23] Peng Lihong, Liao Bo, Zhu Wen, et al. Predicting drug-target interactions with neighbor interaction information and discriminative low-rank representation [J]. Current Protein & Peptide Science, 2018, 19(5): 455-467
- [24] Liu Yong, Wu Ming, Miao Chunyan, et al. Neighborhood regularized logistic matrix factorization for drug-target interaction prediction [J]. PLOS Computational Biology, 2016, 12(2): e1004760
- [25] Apweiler R, Bairoch A, Wu C, et al. UniProt: The universal protein knowledgebase [J]. Nucleic Acids Research, 2004, 32(Suppl1): D115-D119



Yu Donghua, born in 1988. PhD candidate. Student member of CCF. His main research interests include machine learning and bioinformatics.



Guo Maozu, born in 1966. Professor. PhD supervisor. Member of CCF. His main research interests include machine learning, smart city and bioinformatics.



Liu Xiaoyan, born in 1963. PhD. Associate professor and master supervisor. Her main research interests include bioinformatics and data mining. (liuxiaoyan@hit.edu.cn)



Cheng Shuang, born in 1985. PhD. Her main research interest is bioinformatics. (chengs777@163.com)