



(12)发明专利

(10)授权公告号 CN 104781458 B

(45)授权公告日 2017.09.22

(21)申请号 201380050955.1

北野宏明 河冈义裕

(22)申请日 2013.09.27

(74)专利代理机构 中科专利商标代理有限责任  
公司 11021

(65)同一申请的已公布的文献号  
申请公布号 CN 104781458 A

代理人 张莉

(43)申请公布日 2015.07.15

(51)Int.Cl.

(30)优先权数据  
2012-219730 2012.10.01 JP

C40B 30/02(2006.01)

G06F 17/18(2006.01)

G06F 17/30(2006.01)

G06F 19/16(2006.01)

(85)PCT国际申请进入国家阶段日  
2015.03.30

(56)对比文件

(86)PCT国际申请的申请数据  
PCT/JP2013/076248 2013.09.27

JP 3722420 B2,2005.11.30,全文.

US 20060122787 A1,2006.06.08,全文.

(87)PCT国际申请的公布数据  
W02014/054526 JA 2014.04.10

Michael J. Keiser et al.Predicting  
new molecular targets for known drugs.  
《Nature》.2009,第462卷第175-182页.

(73)专利权人 独立行政法人科学技术振兴机构  
地址 日本埼玉县

审查员 刘春杰

(72)发明人 蒂亚哥·乔斯·达席尔瓦洛佩斯

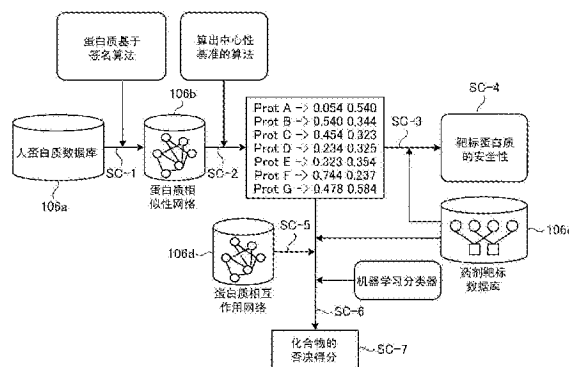
权利要求书5页 说明书16页 附图10页

(54)发明名称

批准预测装置、批准预测方法及计算机可读  
记录介质

(57)摘要

本发明的批准预测装置、批准预测方法计算  
构成蛋白质相似性网络的蛋白质的中心性尺度  
即相似性中心性尺度,计算构成蛋白质间相互  
作用网络的蛋白质的中心性尺度即相互作用中  
心性尺度,使用将各药剂的批准属性、相似性中  
心性尺度的各药剂的每个靶标的合计值以及平均  
值、和相互作用中心性尺度的各药剂的每个靶标  
的合计值以及平均值作为训练数据的分类器,来  
计算将验证对象的化合物分类为不批准的药剂  
的概率即否决得分,输出否决得分。



1. 一种批准预测装置,至少具备输出部、存储部和控制部,上述批准预测装置的特征在于,

上述存储部具备:

蛋白质排列信息存储单元,其存储与蛋白质的氨基酸排列相关的排列信息;

相似性网络信息存储单元,其存储与基于蛋白质彼此的相似性而构成的蛋白质相似性网络相关的相似性网络信息;

药剂靶标存储单元,其将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的上述蛋白质相关的蛋白质信息建立对应并存储;和

相互作用网络信息存储单元,其存储与基于上述蛋白质间的相互作用而构成的蛋白质间相互作用网络相关的相互作用网络信息,

上述控制部具备:

相似性网络信息保存单元,其在基于存储在上述蛋白质排列信息存储单元中的上述排列信息,使用基于签名算法相互检测到上述蛋白质彼此的上述相似性的情况下,建立由相互检测到上述相似性的上述蛋白质彼此构成的上述蛋白质相似性网络,将与该蛋白质相似性网络相关的上述相似性网络信息保存在上述相似性网络信息存储单元;

相似性中心性尺度计算单元,其基于存储在上述相似性网络信息存储单元中的上述相似性网络信息,计算构成上述蛋白质相似性网络的上述蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度;

相互作用中心性尺度计算单元,其基于存储在上述相互作用网络信息存储单元中的上述相互作用网络信息,计算构成上述蛋白质间相互作用网络的上述蛋白质的包含上述度中心性、上述介数中心性、上述距离的中心性、以及上述Burt的约束值的中心性尺度即相互作用中心性尺度;

否决得分计算单元,其使用将存储在上述药剂靶标存储单元中的各药剂的上述批准属性、由上述相似性中心性尺度计算单元计算出的上述相似性中心性尺度的上述各药剂的每个上述靶标的合计值以及平均值、和由上述相互作用中心性尺度计算单元计算出的上述相互作用中心性尺度的上述各药剂的每个上述靶标的合计值以及平均值作为训练数据的分类器,来计算将验证对象的化合物分类为上述不批准的上述药剂的概率即否决得分;和

否决得分输出单元,其经由上述输出部使由上述否决得分计算单元计算出的上述否决得分输出。

2. 一种批准预测装置,至少具备输出部、存储部和控制部,上述批准预测装置的特征在于,

上述存储部具备:

蛋白质排列信息存储单元,其存储与蛋白质的氨基酸排列相关的排列信息;

相似性网络信息存储单元,其存储与由具有相似性的蛋白质彼此构成的蛋白质相似性网络相关的相似性网络信息;和

药剂靶标存储单元,其将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的上述蛋白质相关的蛋白质信息建立对应并存储,

上述控制部具备:

相似性网络信息保存单元,其在基于存储在上述蛋白质排列信息存储单元中的上述排

列信息,使用基于签名算法相互检测到上述蛋白质彼此的上述相似性的情况下,建立由相互检测到上述相似性的上述蛋白质彼此构成的上述蛋白质相似性网络,将与该蛋白质相似性网络相关的上述相似性网络信息保存在上述相似性网络信息存储单元;

相似性中心性尺度计算单元,其基于存储在上述相似性网络信息存储单元中的上述相似性网络信息,计算构成上述蛋白质相似性网络的上述蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度;

批准判定单元,其使用由上述相似性中心性尺度计算单元计算出的构成该蛋白质相似性网络的验证对象的蛋白质的上述相似性中心性尺度来取得以下判定结果,所述判定结果是根据将构成上述蛋白质相似性网络的与存储在上述药剂靶标存储单元中的上述蛋白质信息相应的上述蛋白质作为上述靶标的上述药剂的上述批准属性,来判定该验证对象的上述蛋白质是否包含在上述批准的上述药剂或上述不批准的上述药剂的上述靶标的范围内的判定结果;和

判定结果输出单元,其经由上述输出部使由上述批准判定单元取得的上述判定结果输出。

3. 根据权利要求2所述的批准预测装置,其特征在于,

上述批准判定单元,其根据将构成上述蛋白质相似性网络的与存储在上述药剂靶标存储单元中的上述蛋白质信息相应的上述蛋白质作为上述靶标的上述药剂的上述批准属性,在由上述相似性中心性尺度计算单元计算出的上述验证对象的蛋白质的上述相似性中心性尺度中包含的上述度中心性高、上述距离的中心性低、上述Burt的约束值非常低的情况下,生成上述验证对象的上述蛋白质包含在上述不批准的上述药剂的上述靶标的范围内的判定结果。

4. 一种批准预测方法,在至少具备输出部、存储部和控制部的批准预测装置中执行,其特征在于,

上述存储部具备:

蛋白质排列信息存储单元,其存储与蛋白质的氨基酸排列相关的排列信息;

相似性网络信息存储单元,其存储与基于蛋白质彼此的相似性而构成的蛋白质相似性网络相关的相似性网络信息;

药剂靶标存储单元,其将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的上述蛋白质相关的蛋白质信息建立对应并存储;和

相互作用网络信息存储单元,其存储与基于上述蛋白质间的相互作用而构成的蛋白质间相互作用网络相关的相互作用网络信息,

上述批准预测方法包括在上述控制部中执行的如下步骤:

相似性网络信息保存步骤,在基于存储在上述蛋白质排列信息存储单元中的上述排列信息,使用基于签名算法相互检测到上述蛋白质彼此的上述相似性的情况下,建立由相互检测到上述相似性的上述蛋白质彼此构成的上述蛋白质相似性网络,将与该蛋白质相似性网络相关的上述相似性网络信息保存在上述相似性网络信息存储单元;

相似性中心性尺度算出步骤,基于存储在上述相似性网络信息存储单元中的上述相似性网络信息,计算构成上述蛋白质相似性网络的上述蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度;

相互作用中心性尺度算出步骤,基于存储在上述相互作用网络信息存储单元的上述相互作用网络信息,计算构成上述蛋白质间相互作用网络的上述蛋白质的包含上述度中心性、上述介数中心性、上述距离的中心性、以及上述Burt的约束值的中心性尺度即相互作用中心性尺度;

否决得分算出步骤,使用将存储在上述药剂靶标存储单元中的各药剂的上述批准属性、由上述相似性中心性尺度算出步骤计算出的上述相似性中心性尺度的上述各药剂的每个上述靶标的合计值以及平均值、和由上述相互作用中心性尺度算出步骤计算出的上述相互作用中心性尺度的上述各药剂的每个上述靶标的合计值以及平均值作为训练数据的分类器,来计算将验证对象的化合物分类为上述不批准的上述药剂的概率即否决得分;和

否决得分输出步骤,经由上述输出部使由上述否决得分算出步骤计算出的上述否决得分输出。

5. 一种批准预测方法,在至少具备输出部、存储部和控制部的批准预测装置中执行,其特征在于,

上述存储部具备:

蛋白质排列信息存储单元,其存储与蛋白质的氨基酸排列相关的排列信息;

相似性网络信息存储单元,其存储与由具有相似性的蛋白质彼此构成的蛋白质相似性网络相关的相似性网络信息;和

药剂靶标存储单元,其将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的上述蛋白质相关的蛋白质信息建立对应来存储,

上述批准预测方法包括在上述控制部中执行的如下步骤:

相似性网络信息保存步骤,在基于存储在上述蛋白质排列信息存储单元中的上述排列信息,使用基于签名算法相互检测到上述蛋白质彼此的上述相似性的情况下,建立由相互检测到上述相似性的上述蛋白质彼此构成的上述蛋白质相似性网络,将与该蛋白质相似性网络相关的上述相似性网络信息保存在上述相似性网络信息存储单元;

相似性中心性尺度算出步骤,基于存储在上述相似性网络信息存储单元中的上述相似性网络信息,计算构成上述蛋白质相似性网络的上述蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度;

批准判定步骤,使用由上述相似性中心性尺度算出步骤计算出的构成该蛋白质相似性网络的验证对象的蛋白质的上述相似性中心性尺度来取得以下判定结果,所述判定结果是根据将构成上述蛋白质相似性网络的与存储在上述药剂靶标存储单元中的上述蛋白质信息相应的上述蛋白质作为上述靶标的上述药剂的上述批准属性,判定该验证对象的上述蛋白质是否包含在上述批准的上述药剂或上述不批准的上述药剂的上述靶标的范围内的判定结果;和

判定结果输出步骤,经由上述输出部使在上述批准判定步骤取得的上述判定结果输出。

6. 一种存储有程序的计算机可读记录介质,该程序在至少具备输出部、存储部和控制部的批准预测装置中执行,其特征在于,

上述存储部具备:

蛋白质排列信息存储单元,其存储与蛋白质的氨基酸排列相关的排列信息;

相似性网络信息存储单元,其存储与基于蛋白质彼此的相似性而构成的蛋白质相似性网络相关的相似性网络信息;

药剂靶标存储单元,其将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的上述蛋白质相关的蛋白质信息建立对应并存储;和

相互作用网络信息存储单元,其存储与基于上述蛋白质间的相互作用而构成的蛋白质间相互作用网络相关的相互作用网络信息,

上述程序包括在上述控制部中执行的如下步骤:

相似性网络信息保存步骤,在基于存储在上述蛋白质排列信息存储单元中的上述排列信息,使用基于签名算法相互检测到上述蛋白质彼此的上述相似性的情况下,建立由相互检测到上述相似性的上述蛋白质彼此构成的上述蛋白质相似性网络,将与该蛋白质相似性网络相关的上述相似性网络信息保存在上述相似性网络信息存储单元;

相似性中心性尺度算出步骤,基于存储在上述相似性网络信息存储单元中的上述相似性网络信息,计算构成上述蛋白质相似性网络的上述蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度;

相互作用中心性尺度算出步骤,基于存储在上述相互作用网络信息存储单元的上述相互作用网络信息,计算构成上述蛋白质间相互作用网络的上述蛋白质的包含上述度中心性、上述介数中心性、上述距离的中心性、以及上述Burt的约束值的中心性尺度即相互作用中心性尺度;

否决得分算出步骤,使用将存储在上述药剂靶标存储单元中的各药剂的上述批准属性、由上述相似性中心性尺度算出步骤计算出的上述相似性中心性尺度的上述各药剂的每个上述靶标的合计值以及平均值、和由上述相互作用中心性尺度算出步骤计算出的上述相互作用中心性尺度的上述各药剂的每个上述靶标的合计值以及平均值作为训练数据的分类器,来计算将验证对象的化合物分类为上述不批准的上述药剂的概率即否决得分;和

否决得分输出步骤,经由上述输出部使由上述否决得分算出步骤计算出的上述否决得分输出。

7.一种存储有程序的计算机可读记录介质,该程序在至少具备输出部、存储部和控制部的批准预测装置中执行,其特征在于,

上述存储部具备:

蛋白质排列信息存储单元,其存储与蛋白质的氨基酸排列相关的排列信息;

相似性网络信息存储单元,其存储与由具有相似性的蛋白质彼此构成的蛋白质相似性网络相关的相似性网络信息;和

药剂靶标存储单元,其将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的上述蛋白质相关的蛋白质信息建立对应来存储,

上述程序包括在上述控制部中执行的如下步骤:

相似性网络信息保存步骤,在基于存储在上述蛋白质排列信息存储单元中的上述排列信息,使用基于签名算法相互检测到上述蛋白质彼此的上述相似性的情况下,建立由相互检测到上述相似性的上述蛋白质彼此构成的上述蛋白质相似性网络,将与该蛋白质相似性网络相关的上述相似性网络信息保存在上述相似性网络信息存储单元;

相似性中心性尺度算出步骤,基于存储在上述相似性网络信息存储单元中的上述相似

性网络信息,计算构成上述蛋白质相似性网络的上述蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度;

批准判定步骤,使用由上述相似性中心性尺度算出步骤计算出的构成该蛋白质相似性网络的验证对象的蛋白质的上述相似性中心性尺度来取得以下判定结果,所述判定结果是根据将构成上述蛋白质相似性网络的与存储在上述药剂靶标存储单元中的上述蛋白质信息相应的上述蛋白质作为上述靶标的上述药剂的上述批准属性,判定该验证对象的上述蛋白质是否包含在上述批准的上述药剂或上述不批准的上述药剂的上述靶标的范围内的判定结果;和

判定结果输出步骤,经由上述输出部使在上述批准判定步骤取得的上述判定结果输出。

## 批准预测装置、批准预测方法及计算机可读记录介质

### 技术领域

[0001] 本发明涉及批准预测装置、批准预测方法以及程序。

### 背景技术

[0002] 过去以来,公开了进行已有的化合物的脱靶或副作用的预测的技术。

[0003] 在非专利文献1所记载的蛋白质功能识别中,公开了基于配位体的相似性来检测蛋白质的分组化所引起的药物的脱靶的技术,在以前未在文献中报告的与受体对抗的方面,在美沙酮、依米丁以及洛哌丁胺(loperamid)这样的药剂中发现未预料的关系。

[0004] 另外,在非专利文献2所记载的药剂靶标识别中,公开了如下技术:作为出发点使用由市售医药品引起的副作用来研讨脱靶效应,通过根据副作用将药分组化来将具有适应和结构的药剂分组化,决定以前未知的药剂的追加靶标蛋白质。

[0005] 另外,在非专利文献3所记载的已知药剂的新的分子靶标预测中,公开了如下技术:基于配位体的相似性将蛋白质分组化,调查脱靶效应,由此来发现已知的药剂当中报告的靶标以外的靶标。

[0006] 另外,在非专利文献4所记载的药剂靶标相互作用网络预测中,公开了如下技术:将与蛋白质排列以及药剂靶标相关的信息建立关联来新作成称作药理学空间(pharmacological space)的资源,使用该资源来使已知的药物的追加靶标变得明确,将药剂靶标分类为酶、离子通道、G蛋白质耦联受体、以及核内受体这4个类。

[0007] 另外,在非专利文献5所记载的药剂活性大规模预测中,公开了如下技术:作成在市售药的副作用的预测以及说明中使用的药剂靶标副作用网络,根据医药品和特定的蛋白质之间无意识的相互作用的各种关联性来找出以前不能说明的副作用。

[0008] 另外,非专利文献6所记载的药剂诱导肝脏损伤预测系统是用于识别引起肝脏的损伤的可能性高的化合物的预测系统,公开了如下技术:将预测对象限定在肝脏,基于科学文献的调查来预测某种化合物易于引起对肝脏的损伤这样的特性。在此,在药剂诱导肝脏损伤预测系统中,预测有给肝脏带来有害的影响的可能性的几种蛋白质以及路径。

[0009] 先行技术文献

[0010] 非专利文献

[0011] 非专利文献1:Keiser MJ,Roth BL,Armbruster BN,Ernsberger P,Irwin JJ,Shoichet BK.(2007)Relating protein pharmacology by ligand chemistry,Nature Biotechnology,25,197-206.

[0012] 非专利文献2:Campillos M,Kuhn M,Gavin AC,Jensen LJ,Bork P.(2008)Drug Target Identification Using Side-Effect Similarity,Science,321,263-266.

[0013] 非专利文献3:Keiser MJ,Setola V,Irwin JJ,Laggner C,Abbas AI,Hufeisen SJ,Jensen NH,Kuijjer MB,Matos RC,Tran TB,Whaley R,Glennon RA,Hert J,Thomas KL,Edwards DD,Shoichet BK,Roth BL.(2009)Predicting new molecular targets for known drugs,Nature,462,175-181.

[0014] 非专利文献4:Yamanishi Y,Araki M,Gutteridge A,Honda W,Kanehisa M. (2008)Prediction of drug-target interaction networks from the integration of chemical and genomic spaces,Bioinformatics,24,i232-i240.

[0015] 非专利文献5:Lounkine E,Keiser MJ,Whitebread S,Mikhailov D,Hamon J,Jenkins JL,Lavan P,Weber E,Doak AK,Cote S,Shoichet BK,Urban L. (2012)Large-scale prediction and testing of drug activity on side-effect targets,Nature,486,361-367.

[0016] 非专利文献6:Liu Z,Shi Q,Ding D,Kelly R,Fang H,et al. (2011)Translating Clinical Findings into Knowledge in Drug Safety Evaluation-Drug Induced Liver Injury Prediction System(DILIps).PLoS Comput Biol 7(12):e1002310.

[0017] 发明的概要

[0018] 发明要解决的课题

[0019] 但是,在非专利文献1到6所记载的现有的药剂靶标预测技术中,有不能基于靶标蛋白质的特性来量化医药品的批准的可能性的问题。

## 发明内容

[0020] 本发明鉴于上述问题点而提出,目的在于提供一种能够量化医药品的批准或否决的可能性的批准预测装置、批准预测方法、以及程序。

[0021] 用于解决课题的手段

[0022] 为了达成这样的目的,本发明的批准预测装置是至少具备输出部、存储部和控制部的批准预测装置,上述批准预测装置的特征在于,上述存储部具备:相似性网络信息存储单元,其存储与基于蛋白质彼此的相似性而构成的蛋白质相似性网络相关的相似性网络信息;药剂靶标存储单元,其将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的上述蛋白质相关的蛋白质信息建立对应并存储;和相互作用网络信息存储单元,其存储与基于上述蛋白质间的相互作用而构成的蛋白质间相互作用网络相关的相互作用网络信息,上述控制部具备:相似性中心性尺度计算单元,其基于存储在上述相似性网络信息存储单元中的上述相似性网络信息,计算构成上述蛋白质相似性网络的上述蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度;相互作用中心性尺度计算单元,其基于存储在上述相互作用网络信息存储单元中的上述相互作用网络信息,计算构成上述蛋白质间相互作用网络的上述蛋白质的包含上述度中心性、上述介数中心性、上述距离的中心性、以及上述Burt的约束值的中心性尺度即相互作用中心性尺度;否决得分计算单元,其使用将存储在上述药剂靶标存储单元中的上述各药剂的上述批准属性、由上述相似性中心性尺度计算单元计算出的上述相似性中心性尺度的上述各药剂的每个上述靶标的合计值以及平均值、和由上述相互作用中心性尺度计算单元计算出的上述相互作用中心性尺度的上述各药剂的每个上述靶标的合计值以及平均值作为训练数据的分类器,来计算将验证对象的化合物分类为上述不批准的上述药剂的概率即否决得分;和否决得分输出单元,其经由上述输出部使由上述否决得分计算单元计算出的上述否决得分输出。

[0023] 另外,本发明的批准预测装置是至少具备输出部、存储部和控制部的批准预测装

置,上述批准预测装置的特征在于,上述存储部具备:相似性网络信息存储单元,其存储与由具有相似性的蛋白质彼此构成的蛋白质相似性网络相关的相似性网络信息;和药剂靶标存储单元,其将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的上述蛋白质相关的蛋白质信息建立对应并存储,上述控制部具备:相似性中心性尺度计算单元,其基于存储在上述相似性网络信息存储单元中的上述相似性网络信息,计算构成上述蛋白质相似性网络的上述蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度;批准判定单元,使用由上述相似性中心性尺度计算单元计算出的构成该蛋白质相似性网络的验证对象的蛋白质的上述相似性中心性尺度来取得以下判定结果,所述判定结果是根据将构成上述蛋白质相似性网络的与存储在上述药剂靶标存储单元中的上述蛋白质信息相应的上述蛋白质作为上述靶标的上述药剂的上述批准属性,来判定构成该验证对象的上述蛋白质是否包含在上述被批准的上述药剂或上述不批准的上述药剂的上述靶标的范围内的判定结果;和判定结果输出单元,其经由上述输出部使由上述批准判定单元取得的上述判定结果输出。

[0024] 另外,本发明的批准预测装置在上述记载的批准预测装置的基础上,特征在于,上述存储部还具备:蛋白质排列信息存储单元,其存储与上述蛋白质的氨基酸排列相关的排列信息,上述控制部还具备:相似性网络信息保存单元,其在基于存储在上述蛋白质排列信息存储单元中的上述排列信息,使用基于签名算法相互检测到上述蛋白质彼此的上述相似性的情况下,建立由该相互检测到上述相似性的上述蛋白质彼此构成的上述蛋白质相似性网络,将与该蛋白质相似性网络相关的上述相似性网络信息保存在上述相似性网络信息存储单元。

[0025] 另外,本发明的批准预测装置在上述记载的批准预测装置的基础上,特征在于,上述批准判定单元,其根据将构成上述蛋白质相似性网络的与存储在上述药剂靶标存储单元中的上述蛋白质信息相应的上述蛋白质作为上述靶标的上述药剂的上述批准属性,在由上述相似性中心性尺度计算单元计算出的上述验证对象的蛋白质的上述相似性中心性尺度中包含的上述度中心性高、上述距离的中心性低、上述Burt的约束值非常低的情况下,生成上述验证对象的上述蛋白质包含在上述不批准的上述药剂的上述靶标的范围内的判定结果。

[0026] 另外,本发明的批准预测方法是在至少具备输出部、存储部和控制部的批准预测装置中执行的批准预测方法,其特征在于,上述存储部具备:相似性网络信息存储单元,其存储与基于蛋白质彼此的相似性而构成的蛋白质相似性网络相关的相似性网络信息;药剂靶标存储单元,其将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的上述蛋白质相关的蛋白质信息建立对应并存储;和相互作用网络信息存储单元,其存储与基于上述蛋白质间的相互作用而构成的蛋白质间相互作用网络相关的相互作用网络信息,上述批准预测方法包括在上述控制部中执行的如下步骤:相似性中心性尺度算出步骤,基于存储在上述相似性网络信息存储单元中的上述相似性网络信息,计算构成上述蛋白质相似性网络的上述蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度;相互作用中心性尺度算出步骤,基于存储在上述相互作用网络信息存储单元的上述相互作用网络信息,计算构成上述蛋白质间相互作用网络的上述蛋白质的包含上述度中心性、上述介数中心性、上述距离的中心性、

以及上述Burt的约束值的中心性尺度即相互作用中心性尺度；否决得分算出步骤，使用将存储在上述药剂靶标存储单元中的上述各药剂的上述批准属性、由上述相似性中心性尺度算出步骤计算出的上述相似性中心性尺度的上述各药剂的每个上述靶标的合计值以及平均值、和由上述相互作用中心性尺度算出步骤计算出的上述相互作用中心性尺度的上述各药剂的每个上述靶标的合计值以及平均值作为训练数据的分类器，来计算将验证对象的化合物分类为上述不批准的上述药剂的概率即否决得分；和否决得分输出步骤，其经由上述输出部使由上述否决得分算出步骤计算出的上述否决得分输出。

[0027] 另外，本发明的批准预测方法是在至少具备输出部、存储部和控制部的批准预测装置中执行的批准预测方法，上述批准预测方法的特征在于，上述存储部具备：相似性网络信息存储单元，其存储与由具有相似性的蛋白质彼此构成的蛋白质相似性网络相关的相似性网络信息；和药剂靶标存储单元，其将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的上述蛋白质相关的蛋白质信息建立对应来存储，上述批准预测方法包括在上述控制部中执行的如下步骤：相似性中心性尺度算出步骤，基于存储在上述相似性网络信息存储单元中的上述相似性网络信息，计算构成上述蛋白质相似性网络的上述蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度；批准判定步骤，使用由上述相似性中心性尺度算出步骤计算出的构成该蛋白质相似性网络的验证对象的蛋白质的上述相似性中心性尺度来取得以下判定结果，所述判定结果是根据将构成上述蛋白质相似性网络的与存储在上述药剂靶标存储单元中的上述蛋白质信息相应的上述蛋白质作为上述靶标的上述药剂的上述批准属性，判定该验证对象的上述蛋白质是否包含在上述被批准的上述药剂或上述不批准的上述药剂的上述靶标的范围内的判定结果；和判定结果输出步骤，经由上述输出部使在上述批准判定步骤取得的上述判定结果输出。

[0028] 另外，本发明的程序是用于使至少具备输出部、存储部和控制部的批准预测装置执行的程序，上述程序的特征在于，上述存储部具备：相似性网络信息存储单元，其存储与基于蛋白质彼此的相似性而构成的蛋白质相似性网络相关的相似性网络信息；药剂靶标存储单元，其将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的上述蛋白质相关的蛋白质信息建立对应来存储；和相互作用网络信息存储单元，其存储与基于上述蛋白质间的相互作用而构成的蛋白质间相互作用网络相关的相互作用网络信息，在上述控制部中执行如下步骤：相似性中心性尺度算出步骤，基于存储在上述相似性网络信息存储单元中的上述相似性网络信息来计算构成上述蛋白质相似性网络的上述蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度；相互作用中心性尺度算出步骤，基于存储在上述相互作用网络信息存储单元的上述相互作用网络信息来计算构成上述蛋白质间相互作用网络的上述蛋白质的包含度中心性、上述介数中心性、上述距离的中心性、以及上述Burt的约束值的中心性尺度即相互作用中心性尺度；和否决得分算出步骤，使用将存储在上述药剂靶标存储单元中的上述各药剂的上述批准属性、由上述相似性中心性尺度算出步骤计算出的上述相似性中心性尺度的上述各药剂的每个上述靶标的合计值以及平均值、和由上述相互作用中心性尺度算出步骤计算出的上述相互作用中心性尺度的上述各药剂的每个上述靶标的合计值以及平均值作为训练数据的分类器，计算将验证对象的化合物分类为上述不批准的上

述药剂的概率即否决得分;和否决得分输出步骤,经由上述输出部使由上述否决得分算出步骤计算出的上述否决得分输出。

[0029] 另外,本发明的程序是用于使至少具备输出部、存储部和控制部的批准预测装置执行的程序,上述存储部具备:相似性网络信息存储单元,其存储与由有相似性的蛋白质彼此构成的蛋白质相似性网络相关的相似性网络信息;和药剂靶标存储单元,其将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的上述蛋白质相关的蛋白质信息建立对应并存储,在上述控制部中执行如下步骤:相似性中心性尺度算出步骤,基于存储在上述相似性网络信息存储单元中的上述相似性网络信息,计算构成上述蛋白质相似性网络的上述蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度;批准判定步骤,根据将构成上述蛋白质相似性网络的与存储在上述药剂靶标存储单元中的上述蛋白质信息相应的上述蛋白质作为上述靶标的上述药剂的上述批准属性来判定构成该蛋白质相似性网络的验证对象的上述蛋白质是否包含在上述批准的上述药剂或上述不批准的上述药剂的上述靶标的范围内,使用在上述相似性中心性尺度算出步骤计算出的该验证对象的蛋白质的上述相似性中心性尺度来取得判定结果;和判定结果输出步骤,经由上述输出部输出在上述批准判定步骤取得的上述判定结果。

[0030] 发明效果

[0031] 另外,根据本发明,由于计算构成蛋白质相似性网络的蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度,计算构成蛋白质间相互作用网络的蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相互作用中心性尺度,使用将各药剂的批准属性、所计算出的相似性中心性尺度的各药剂的每个靶标的合计值以及平均值、和所计算出的相互作用中心性尺度的各药剂的每个靶标的合计值以及平均值作为训练数据的分类器,计算将验证对象的化合物分类为不批准的药剂的概率即否决得分,并经由输出部输出所计算出的否决得分,因此,通过考虑作为1个化合物的靶标的所有蛋白质的特性,实现能利用于多个靶标化合物的批准/不批准的预测中的效果。另外,根据本发明,通过进行利用了机器学习的分类的表示候补化合物引起不希望的副作用的可能性的评分,从而实现帮助决定在医药品开发的初始阶段被使用的、具有较高的被批准的可能性的化合物的优先顺位的效果。

[0032] 根据本发明,由于计算构成蛋白质相似性网络的蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值在内的中心性尺度即相似性中心性尺度,根据将构成蛋白质相似性网络的蛋白质作为靶标的药剂的批准属性来判定构成该蛋白质相似性网络的验证对象的蛋白质是否包含在被批准的药剂或不批准的药剂的靶标的范围内,使用被计算出的该验证对象的蛋白质的相似性中心性尺度来取得判定结果,经由输出部输出所取得的判定结果,因此能够确定各个蛋白质的特性,实现能够判断是否不可能带来有害的影响的效果。另外,根据本发明,起到能利用于siRNA治疗或单靶标化合物(aka magic bullets)等的单独靶标的评价、或调节单一的特定的蛋白质的活性的技术中的效果。

[0033] 另外,根据本发明,由于在使用基于签名算法相互检测到蛋白质彼此的相似性的情况下,制作由该相互检测到相似性的蛋白质彼此构成的蛋白质相似性网络,保存与该蛋白质相似性网络相关的相似性网络信息,因此实现能够提供一种比过去公开的网络数据具

有更高的相似性的网络数据的效果。

[0034] 另外,根据本发明,由于根据将构成蛋白质相似性网络的蛋白质作为靶标的药剂的批准属性,在所计算出的验证对象的蛋白质的相似性中心性尺度中所含的度中心性高、距离的中心性低、Burt的约束值非常低的情况下,生成验证对象的蛋白质包含在不批准的药剂的靶标的范围内的判定结果,因此实现能准确地确定易于进行非特异性结合、易于引起副作用的蛋白质。

## 附图说明

[0035] 图1是表示本实施方式的基本原理的流程图。

[0036] 图2是表示本实施方式的基本原理的流程图。

[0037] 图3是表示本实施方式中的批准预测装置的构成的一例的框图。

[0038] 图4是表示本实施方式的批准预测装置的处理的一例的流程图。

[0039] 图5是表示本实施方式中的排列信息的一例的图。

[0040] 图6是表示本实施方式中的相似性网络信息的一例的图。

[0041] 图7是表示本实施方式中的Burt的约束的一例的图。

[0042] 图8是表示本实施方式中的蛋白质的中心性尺度的一例的图。

[0043] 图9是表示本实施方式中的存储在药剂靶标数据库的信息的一例的图。

[0044] 图10是表示本实施方式中的批准或不批准的药剂的靶标的中心性尺度的一例的图。

[0045] 图11是表示本实施方式中的相互作用网络信息的一例的图。

[0046] 图12是表示本实施方式中的分类器的性能改善的一例的图。

[0047] 图13是表示本实施方式中的分类器的分类精度的一例的图。

[0048] 图14是表示本实施方式中的分类器的一例的图。

[0049] 图15是表示本实施方式中的输出信息的一例的图。

## 具体实施方式

[0050] 以下基于附图来详细说明本发明所涉及的批准预测装置、批准预测方法、以及程序的实施方式。另外,并不由本实施方式限定本发明。

[0051] [本发明的实施方式的概要]

[0052] 以下参考图1以及图2说明本发明的实施方式的概要,之后详细说明本实施方式的构成以及处理等。

[0053] [概要(其一)]

[0054] 首先参考图1来说明本发明的实施方式的概要的一例。图1是表示本实施方式的基本原理的流程图。本实施方式概略地具有以下的基本特征。

[0055] 即,本实施方式的批准预测装置的控制部如图1所示那样,算出构成蛋白质相似性网络的蛋白质的包含度中心性、介数中心性、距离的中心性 以及Burt的约束值的中心性尺度即相似性中心性尺度(步骤SA-1)。

[0056] 然后,批准预测装置的控制部基于构成蛋白质相似性网络的将蛋白质作为靶标的药剂的批准属性,使用在步骤SA-1计算出的该验证对象的蛋白质的相似性中心性尺度来取

得构成该蛋白质相似性网络的验证对象的蛋白质是否包含在被批准的药剂、或不批准的药剂的靶标的范围内的判定结果(步骤SA-2)。

[0057] 然后,批准预测装置的控制部经由输出部输出在步骤SA-2取得的判定结果(步骤SA-3),结束处理。

[0058] 以上结束概要(其一)的说明。

[0059] [概要(其二)]

[0060] 另外,参考图2来说明本发明的实施方式的概要的一例。图2是表示本实施方式的基本原理的流程图。

[0061] 本实施方式的批准预测装置的控制部如图2所示那样算出构成蛋白质相似性网络的蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值的中心性尺度即相似性中心性尺度(步骤SB-1)。

[0062] 然后,批准预测装置的控制部算出构成蛋白质间相互作用网络的蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值的中心性尺度即相互作用中心性尺度(步骤SB-2)。

[0063] 然后,批准预测装置的控制部使用将各药剂的批准属性、在步骤SB-1计算出的相似性中心性尺度的各药剂的每个靶标的合计值以及平均值、和在步骤SB-2计算出的相互作用中心性尺度的各药剂的每个靶标的合计值以及平均值作为训练数据的分类器,来计算将验证对象的化合物分类在不批准的药剂的概率即否决得分(步骤SB-3)。

[0064] 然后,批准预测装置的控制部经由输出部输出在步骤SB-3计算出的否决得分(步骤SB-4),结束处理。

[0065] 以上结束本实施方式的概要的说明。

[0066] [批准预测装置100的构成]

[0067] 接下来参考图3来说明本实施方式中的批准预测装置100的构成的详细。图3是表示本实施方式中的批准预测装置100的构成的一例的框图,概念性地仅示出该构成当中与本发明有关系的部分。在此,在本实施方式中的批准预测装置100中,将在一个筐体内具备所有各构成、单独进行处理的构成(独立型)设为批准预测装置100来进行说明,但并不限于该实施例,也可以构成在分离的筐体内具备各构成、经由网络300等连接来作为1个概念的装置(例如云计算等)。

[0068] 在图3中,外部系统200经由网络300与批准预测装置100相互连接,也可以具有提供与蛋白质排列信息、药剂信息、药剂靶标信息、以及/或者蛋白质间相互作用信息等相关的外部数据库、和/或执行用户界面等的WEB站点的功能等。

[0069] 在此,外部系统200也可以构成为WEB服务器或ASP服务器等。另外,外部系统200的硬件构成可以由一般市售的工作站、个人计算机等的信息处理装置及其附属装置构成。另外,外部系统200的各功能也可以由外部系统200的硬件构成中的CPU、盘装置、存储器装置、输入装置、输出装置、通信控制装置等以及控制它们的程序等实现。

[0070] 另外,网络300具有将批准预测装置100和外部系统200相互连接的功能,例如是因特网等。

[0071] 另外,批准预测装置100概略地具备:控制部102、通信控制接口部104、存储部106和输入输出控制接口部108。在此,批准预测装置100也可以还具备:至少包含显示部112的

输出部、以及输入部114。另外，输出部也可以还包含声音输出部、以及印刷输出部等。在此，控制部102是统一控制批准预测装置100整体的CPU等。另外，通信控制接口部104是连接到与通信线路等连接的路由器等的通信装置(未图示)的接口，输入输出控制接口部108是与输出部以及输入部114连接的接口。另外，存储部106是保存各种数据库或表等的装置。这些批准预测装置100的各部经由任意的通信路以能通信的方式连接。进而，该批准预测装置100经由路由器等的通信装置以及专用线等有线或无线的通信线路以能通信的方式与网络300连接。

[0072] 保存在存储部106的各种数据库或表(蛋白质排列信息数据库106a、相似性网络信息数据库106b、药剂靶标数据库106c、以及相互作用网络信息数据库106d)是固定盘装置等的贮存单元。例如，存储部106保存各种处理中所用的各种程序、表、文件、数据库、以及WEB页面等。

[0073] 这些存储部106的各构成要素当中的蛋白质排列信息数据库106a是存储与蛋白质的氨基酸排列相关的排列信息的蛋白质排列信息存储单元。在此，氨基酸排列也可以是人类蛋白质的氨基酸排列。另外，排列信息也可以是FASTA格式。这些排列信息预先存储在蛋白质排列信息数据库106a中，批准预测装置100的控制部102定期以及/或者根据控制部102进行的处理经由网络300从外部系统200(例如NCBI或UNIPROT等)下载最新的数据，并上传存储在蛋白质排列信息数据库106a中的排列信息。

[0074] 另外，相似性网络信息数据库106b是存储与由具有相似性的蛋白质彼此构成的蛋白质相似性网络(PSIN)相关的相似性网络信息的相似性网络信息存储单元。

[0075] 另外，药剂靶标数据库106c是将包含与药剂的批准或不批准相关的批准属性的药剂信息、和与成为该药剂的靶标的蛋白质相关的蛋白质信息建立对应来存储的药剂靶标存储单元。在此，所谓不批准的药剂，在本实施方式中可以是将在药剂批准中撤回(Withdraw)或违法(Illicit)的药剂作为有问题的药剂而视作1个分组的药剂。即，所谓有问题的药剂，可能是由于带来有害的影响而从必须从市场撤回的药剂、或者社会所不允许的违法药剂(例如兴奋剂或致幻剂等)，是需要与被批准的药剂区别的药剂。这些与药剂批准相关的药剂信息以及蛋白质信息也可以预先存储在药剂靶标数据库106c中，批准预测装置100的控制部102定期以及/或者根据控制部102进行的处理经由网络300从外部系统200(例如Drugbank(<http://www.drugbank.ca/>)等)下载最新的数据，并将与存储在药剂靶标数据库106c的药剂批准相关的药剂信息以及蛋白质信息上传。

[0076] 另外，相互作用网络信息数据库106d是存储与基于蛋白质间的相互作用而构成的蛋白质间相互作用网络(PPI)相关的相互作用网络信息的相互作用网络信息存储单元。也可以是这些相互作用网络信息预先存储在相互作用网络信息数据库106d中，批准预测装置100的控制部102定期以及/或者根据控制部102进行的处理经由网络300从外部系统200(例如HIPPIE(<http://cbdm.mdc-berlin.de/tools/hippie/>)等)下载最新的数据，并将存储在相互作用网络信息数据库106d的相互作用网络信息上传。

[0077] 另外，通信控制接口部104进行批准预测装置100与网络300(或路由器等的通信装置)间的通信控制。即，通信控制接口部104具有经由通信线路与外部系统200以及其它终端等通信数据的功能。

[0078] 另外，输入输出控制接口部108进行输出部(显示部112)以及输入部114的控制。

[0079] 在此,作为显示部112,可以是显示应用等的显示画面的显示单元(例如由液晶或有机EL等构成的显示器、监视器或者触控面板等)。另外,输入部114可以是例如键输入部、触控面板、控制板(例如触控鼠标以及游戏摇杆等)、鼠标、键盘、或麦克风等。另外,作为声音输出部,例如也可以是扬声器等。另外,作为印刷输出部,例如可以是打印机等。

[0080] 另外,在图3中,控制部102具有用于保存OS(Operating System,操作系统)等的控制程序、规定各种处理次序等的程序、以及所需要数据的内部存储器。并且,控制部102通过这些程序等来进行用于执行各种处理的信息处理。控制部102在功能概念上具备:相似性网络信息保存部102a、相似性中心性尺度算出部102b、批准判定部102c、判定结果输出部102d、相互作用中心性尺度算出部102e、否决得分算出部102f、以及否决得分输出部102g。

[0081] 这当中,相似性网络信息保存部102a是如下那样的相似性网络信息保存部:在基于存储在蛋白质排列信息数据库106a的排列信息使用基于签名算法而相互检测到蛋白质彼此的相似性的情况下,作成由该相互检测到相似性的蛋白质彼此构成的蛋白质相似性网络(PSIN),将与该蛋白质相似性网络相关的相似性网络信息保存在相似性网络信息数据库106b中。

[0082] 另外,相似性中心性尺度算出部102b是基于存储在相似性网络信息数据库106b的相似性网络信息来计算构成蛋白质相似性网络的蛋白质的包含度中心性(Degree)、介数中心性(Betweenness)、距离的中心性(Closeness Centrality)、以及Burt的约束值(Burt's constraint)的中心性尺度即相似性中心性尺度的相似性中心性尺度计算单元。在此,度中心性是表示在网络内与其它节点直接相连到何种程度(几个)的指标。另外,介数中心性是通过对于为了与网络内的其它节点连接而必须通过的最短路径的数量进行计数来测量蛋白质的网络的中心性的指标。另外,距离的中心性是测定为了到达网络内其它全部节点而需要的步骤数的指标。另外,Burt的约束值是根据学习分组内的个人的位置关系和优点的社会学的脉络来提出的指标。

[0083] 另外,批准判定部102c是如下那样的批准判定单元:根据将构成蛋白质相似性网络的基于存储在药剂靶标数据库106c的蛋白质信息的蛋白质作为靶标的药剂的批准属性来判定构成该蛋白质相似性网络的验证对象的蛋白质来是否包含在批准的药剂或不批准的药剂的靶标的范围内,使用由相似性中心性尺度算出部102b计算出的该验证对象的蛋白质的相似性中心性尺度来取得其判定结果。在此,也可以在由相似性中心性尺度算出部102b计算出的验证对象的蛋白质的相似性中心性尺度中所含的度中心性高、距离的中心性低、Burt的约束值非常低的情况下,批准判定部102c,根据将构成蛋白质相似性网络的基于存储在药剂靶标数据库106c的蛋白质信息的蛋白质作为靶标的药剂的批准属性,生成验证对象的蛋白质包含在不批准的药剂的靶标的范围内的判定结果。另外,验证对象的蛋白质可以是基于由用户经由输入部114输入的蛋白质信息的蛋白质。

[0084] 另外,判定结果输出部102d是经由输出部输出由批准判定部102c取得的判定结果的判定结果输出单元。在此,判定结果输出部102d可以使判定结果显示在显示部112。另外,判定结果输出部102d也可以经由印刷输出部输出判定结果。

[0085] 另外,相互作用中心性尺度算出部102e是基于存储在相互作用网络信息数据库106d的相互作用网络信息来计算构成蛋白质间相互作用网络的蛋白质的包含度中心性、介数中心性、距离的中心性、以及Burt的约束值的中心性尺度即相互作用中心性尺度的相互

作用中心性尺度计算单元。

[0086] 另外, 否决得分算出部102f是如下那样的否决得分计算单元: 使用将存储在药剂靶标数据库106c的各药剂的批准属性、由相似性中心性尺度算出部102b计算出的相似性中心性尺度的各药剂的每个靶标的合计值以及平均值、和由相互作用中心性尺度算出部102e计算出的相互作用中心性尺度的各药剂的每个靶标的合计值以及平均值作为训练数据的分类器, 来计算将验证对象的化合物分类为不批准的药剂的概率即否决得分。在此, 验证对象的化合物(药剂)也可以是基于由用户经由输入部114输入的化合物信息的化合物。

[0087] 另外, 否决得分输出部102g是经由输出部输出由否决得分算出部102f计算出的否决得分的否决得分输出单元。在此, 否决得分输出部102g也可以使显示部112显示否决得分。另外, 否决得分输出部102g也可以经由印刷输出部输出否决得分。

[0088] 以上结束本实施方式中的批准预测装置100的构成的一例的说明。

[0089] [批准预测装置100的处理]

[0090] 接下来, 以下参考图4乃至图15来详细说明这样构成的本实施方式中的批准预测装置100的处理的详细情况。图4是表示本实施方式中的批准预测装置100的处理的一例的流程图。

[0091] 如图4所示那样, 相似性网络信息保存部102a, 在基于存储在人类蛋白质数据库(蛋白质排列信息数据库) 106a中的排列信息, 使用搜索蛋白质的同族体间的相似点的蛋白质基于签名算法而相互检测到蛋白质彼此的相似性的情况下, 作成由该相互检测到相似性的蛋白质彼此构成的蛋白质相似性网络(PSIN), 将与该蛋白质相似性网络相关的相似性网络信息保存在相似性网络信息数据库106b(步骤SC-1)。在此, 在为了找到相似的蛋白质而使用从NCBI人类蛋白质数据库中查询、比较22000个人类蛋白质的各自的PSI-BLAST工具(Schaffer, et al., 2001)的情况下, 与现有的研究(Atkinson, et al., 2009; Camoglu, et al., 2006; Rattei, et al., 2010; Valavanis, et al., 2010; Weston, et al., 2004; Zhang and Grigorov, 2006)不同, 得到未示出相互关联性(即, 在查询蛋白质A而鉴别出蛋白质B是相似的情况下, 查询其相反的蛋白质B而将蛋白质A鉴别为相似)的结果。为此, 接受该结果, 相似性网络信息保存部102a制作使用了图表逻辑表现的新的蛋白质相似性网络(PSIN)。在此, 在该蛋白质相似性网络(PSIN)中, 节点表征蛋白质, 仅在鉴别出蛋白质的排列具有高的相似性且是双向的命中(即, 在鉴别到蛋白质A与蛋白质B相似的情况下, 鉴别出其相反的蛋白质B与蛋白质A相似)时, 才用边缘将2个节点连接。由此相似性网络信息保存部102a制作包含19721个节点以及776598个边缘的蛋白质相似性网络(PSIN)。

[0092] 在此, 参考图5来说明本实施方式中的排列信息的一例。图5是表示本实施方式中的排列信息的一例的图。

[0093] 如图5所示那样, 存储在蛋白质排列信息数据库106a的排列信息可以是P63261、以及P49281等人类蛋白质的FASTA格式的蛋白质排列信息。

[0094] 另外, 参考图6来说明本实施方式中的相似性网络信息的一例。图6是表示本实施方式中的相似性网络信息的一例的图。

[0095] 如图6所示那样, 在本实施方式中的相似性网络信息中可以包含蛋白质的名称、与该蛋白质相似的蛋白质(Neighbours)的名称、排列得分、以及2个蛋白质相似的区域的信息。在图6中, 将与Q3MI94和Q9Y473的相似性相关的相似性网络信息、以及与Q9P2V4和

Q8NOV4的相似性相关的相似性网络信息作为一例而示出。

[0096] 返回图4,相似性中心性尺度算出部102b基于存储在相似性网络信息数据库106b的相似性网络信息,使用算出中心性基准的算法来计算构成蛋白质相似性网络(PSIN)的蛋白质的度中心性(Degree)、介数中心性(Betweenness)、距离的中心性(Closeness Centrality)、以及Burt的约束值(Burt's constraint)的(步骤SC-2)。

[0097] 在此,说明本实施方式中的构成PSIN的蛋白质的中心性尺度。首先,相似性中心性尺度算出部102b,在1(相连最少的指标)到441(相连最多的指标)的范围内计算表示与PSIN内的节点直接相连到何种程度的指标即度中心性。

[0098] 另外,相似性中心性尺度算出部102b使用由节点i与节点j间的最短路径的数量 $s_{ij}$ 、以及通过节点v的最短路径的比例 $s_{ij}(v)$ 构成的以下的数式(1)来计算介数中心性B(v)。

[0099] [数式1]

$$[0100] \quad B(v) = \sum_{i \neq j, v \neq i \text{ and } v \neq j} \frac{s_{ij}(v)}{s_{ij}} \quad \dots(1)$$

[0101] 另外,相似性中心性尺度算出部102b使用由以节点v与节点i间的步长表征的距离 $d(v, i)$ 构成的以下的数式(2)来计算距离的中心性C(v)。

[0102] [数式2]

$$[0103] \quad C(v) = \frac{1}{\sum_{i \neq v} d(v, i)} \quad \dots(2)$$

[0104] 另外,相似性中心性尺度算出部102b使用由节点j与节点i的关系的比例强度、以及节点j与节点q的关系的比例强度之积 $p_{iq}p_{qj}$ 构成的以下的数式(3)来计算Burt的约束值C(i)。

[0105] [数式3]

$$[0106] \quad C(i) = \sum_j (p_{ij} + \sum_q p_{iq}p_{qj})^2, \quad \text{with } q \neq i, j, \text{ and } j \neq i \quad \dots(3)$$

[0107] 在此,参考图7来说明本实施方式中的Burt的约束值。图7是表示本实施方式中的Burt的约束的一例的图。

[0108] Burt的约束值是根据学习分组内的个人的位置关系和优点的社会学的脉络而提出的手法,若在图7中将节点设为是个人,则在图7的左图中,由于在全部节点都有代替连接,因此能进行与他人的交涉以及交易。另一方面,在如图7的右图那样存在结构性的孔(Structural hole)的情况下,由于在节点2与节点3间不能认识彼此的存在,因此节点1变成位于有利于交涉的位置。在本实施方式中,通过将此运用在作为蛋白质的节点的相似上下文中,具有小的Burt的约束值的蛋白质(节点)成为具有一般位于不同的蛋白质家族间的多个领域的蛋白质,具有大的Burt的约束值的蛋白质(节点)成为表示少量的邻接蛋白质和排列相似性的蛋白质。

[0109] 另外,参考图8来说明本实施方式中的蛋白质的中心性尺度的一例。图8是表示本实施方式中的蛋白质的中心性尺度的一例的图。

[0110] 如图8所示那样,相似性中心性尺度算出部102b,也可以算出度中心性(Degree)、介数中心性(Betweenness)、距离的中心性(Closeness Centrality)、以及Burt的约束值

(Burt's constraint), 作为构成PSIN的蛋白质 (P14784、P14854、P14859、P14867、P14868、P14902、以及P14920) 的中心性尺度, 并输出该中心性尺度的列表。

[0111] 返回图4, 批准判定部102c根据将构成蛋白质相似性网络的基于存储在药剂靶标数据库106c的蛋白质信息的蛋白质作为靶标的药剂的批准属性, 来判定构成该蛋白质相似性网络的验证对象的蛋白质是否包含在被批准的药剂或不批准的药剂的靶标的范围内, 使用在步骤SC-2由相似性中心性尺度算出部102b计算出的该验证对象的蛋白质的度中心性、介数中心性、距离的中心性、以及Burt的约束值来取得判定结果(靶标蛋白质的安全性)(步骤SC-3)。即, 批准判定部102c可以将构成蛋白质相似性网络的蛋白质的中心性尺度、以及存储在药剂靶标数据库106c的列表作为需要, 来决定能假定为药剂靶标的批准或不批准(撤回(Withdraw)、以及驳回(Illicit))的值的范围。在此, 在该步骤中, 仅考虑各个蛋白质, 不考虑化合物能作为靶标的蛋白质全集合。另外, 先决定各个药剂靶标的特性是因为, 由于单一靶标的化合物(魔术弹(magic bullet))、或基于siRNA的治疗方法等被设计为仅阻碍单一靶标, 因此需要在靶标的治疗抑制为安全的前提下进行靶标的选择。

[0112] 另外, 批准判定部102c根据将构成蛋白质相似性网络的基于存储在药剂靶标数据库106c的蛋白质信息的蛋白质作为靶标的药剂的批准属性, 来生成如下这样的判定结果: 即, 在步骤SC-2中由相似性中心性尺度算出部102b计算出的验证对象的蛋白质的相似性中心性尺度中包含的度中心性高、距离的中心性低、Burt的约束值非常低的情况下, 验证对象的蛋白质包含在不批准的药剂的靶标的范围内。

[0113] 在此, 参考图9来说明本实施方式中的存储在药剂靶标数据库106c中的信息的一例。图9是表示本实施方式中的存储在药剂靶标数据库106c的信息的一例的图。

[0114] 如图9所示那样, 在本实施方式中的存储在药剂靶标数据库106c中的信息中, 可以包含药剂名称(Drug)、成为该药剂的靶标的蛋白质的名称(Targets)、以及与(由日本的厚生劳动省或美国的FDA等进行的)该药剂的批准或者不批准相关的批准属性(Status)。

[0115] 在此, 参考图10来说明本实施方式中的批准或不批准的药剂的靶标的中心性尺度的一例。图10是表示本实施方式中的批准或不批准的药剂的靶标的中心性尺度的一例的图。

[0116] 如图10所示那样, 不批准(有问题的)药剂的靶标蛋白质与批准药剂的靶标蛋白质相比, 示出更高的度中心性(Degree)、非常低的Burt的约束值(Burt's constraint)、以及在负的对数刻度更低距离的中心性(Closeness Centrality)。如图10所示那样, 可知在批准药剂的靶标中, 与其它众多的蛋白质的结构共享度低(low-degree), 而在不批准的靶标中, 多个蛋白质之间的结构共享度高, 因此不批准药物的靶标具有易于引起非特异性结合以及副作用这样的特征。

[0117] 返回图4, 判定结果输出部102d使显示部112显示由批准判定部102c取得的靶标蛋白质的安全性(步骤SC-4)。在此, 判定结果输出部102d也可以经由印刷输出部输出判定结果。另外, 判定结果输出部102d也可以为了确认用户感兴趣的蛋白质是处于安全的药剂靶标的范围内还是处于危险的药剂靶标的范围内而输出能查询的列表。

[0118] 另一方面, 相互作用中心性尺度算出部102e基于存储在相互作用网络信息数据库106d中的相互作用网络信息来计算构成蛋白质间相互作用网络(PPI)的蛋白质的度中心性、介数中心性、距离的中心性、以及Burt的约束值(步骤SC-5)。

[0119] 在此,参考图11来说明本实施方式中的相互作用网络信息的一例。图11是表示本实施方式中的相互作用网络信息的一例的图。

[0120] 如图11所示那样,本实施方式中的相互作用网络信息也可以包含将物理上相互作用的蛋白质彼此设为组的蛋白质列表。

[0121] 返回图4,否决得分算出部102f使用将存储在药剂靶标数据库106c中的各药剂的批准属性、步骤SC-2中由相似性中心性尺度算出部102b计算出的度中心性、介数中心性、距离的中心性、以及Burt的约束值的各药剂的每个靶标的合计值以及平均值、和步骤SC-5中由相互作用中心性尺度算出部102e计算出的度中心性、介数中心性、距离的中心性以及Burt的约束值的各药剂的每个靶标的合计值以及平均值作为训练数据的机器学习的分类器,来计算将验证对象的化合物分类为不批准的药剂的概率即否决得分(步骤SC-6)。在此,在药剂靶标数据库106c中,由于几乎所有的已有药剂(化合物)都曾经与多个蛋白质结合而阻碍其活性,即报告了多个药剂靶标,因此,需要针对各化合物考虑成为靶标的全部蛋白质的中心性尺度。为此,否决得分算出部102f使用蛋白质相似性网络(PSIN)、以及蛋白质相互作用网络(PPI),针对各药剂算出每个靶标的度中心性、介数中心性、距离的中心性、以及Burt的约束值的合计以及平均,作为从PSIN将8个属性、从PPI将8个属性以及表示化合物(批准或不批准)的类的1个属性输入到分类器的最终数据集使用。在此,机器学习的分类器也可以是WEKA等已有的数据包(Wishart,2006)等的机器学习的分类器的集合。

[0122] 另外,在本实施方式中,作为训练、以及预测步骤的向导,使用机器学习的分类和药剂分类(批准或不批准),并对最终数据集使用10倍交叉验证法来进行处理。进而,在本实施方式中,使用几个不同的分类算法来执行该次序,确认到前处理技术的使用、以及对相同的数据集合使用来自蛋白质相似性网络(PSIN)的中心性尺度和来自蛋白质相互作用网络(PPI)的中心性尺度的2种情形下强化了预测的性能。

[0123] 在此,本实施方式中的前处理可以用下述的3个步骤进行。起初,可以以合成的数据集合的其它实例的手段以及模式填充缺失值;第二,扩大更小的类的实例的数量;最后,对数据集合进行采样。在本实施方式中的数据集合中,由于由批准的药剂靶标类和最大300例的不批准(有问题的)药剂靶标类的多个实例构成,因此需要从更小的类对应的标本中收集更多的样本。为此,鉴于新化合物的开发成本,由于将批准的药剂误分类为有问题的药剂而产生的麻烦,小于将有问题的药剂误分类为批准的药剂所引起麻烦,因此在本实施方式中,为了较多地收集小的类,较少地采样大规模的类,因而也可以使用SMOTE算法。这样的战略提升了包含各种尺寸的数据集合的分类器的性能。另外,为了执行第2个步骤即重采样,也可以从数据集合随机选择实例,即两次选择相同的实例。进而,在新的数据集合中,既可以是与原始的数据集合相同的实例数以及属性的数量,也可以是50到60的唯一的实例。

[0124] 在此,参考图12来说明本实施方式中的分类器的性能改善的一例。图12是表示本实施方式中的分类器的性能改善的一例的图。

[0125] 如图12所示那样,示出在本实施方式中的分类器中,通过进行前处理技术的使用、以及对相同的数据集合使用来自PSIN的中心性尺度和来自PPI的中心性尺度,从而能相当提升对有问题的药剂类的分类器的灵敏度。

[0126] 进而,在本实施方式中,比较使用3个不同战略的15个机器学习的分类器的预测力。在最初的方法中,使用10倍交叉验证来进行比较。另外,在第二方法中,将原始的数据

集合分割为用实例的70/100和35/100分别设定的训练集和测试集来进行比较。在本实施方式中,从原始的数据集合中500次以上随机选择药剂,没有偏倚地进行调整。另外,在将数据集合分割为训练集和测试集时,仅对训练集进行事前处理。

[0127] 在此,参考图13来说明本实施方式中的分类器的分类精度的一例。图13是表示本实施方式中的分类器的分类精度的一例的图。

[0128] 如图13所示那样,在本实施方式中的分类器的现实的精度测定中,使用了针对药剂的批准类或有问题的类的真阳性率的调和平均(harmonic mean of the true positive rates)。在此,如图13所示那样,由于(通过参数的最优化以及前处理技术的使用)几乎全部的分类因子(分类器)都具有同样的性能,因此在本实施方式中,为了进一步的医药品的安全性预测,以对所有算法必然具有的偏置进行补正为目的,使用了利用不同原理构筑的7个发挥最高的性能算法(KSTAR、IBK、Decorate、END ClassBalancedND、JRip、或RotationForest等)。

[0129] 另外,参考图14来说明本实施方式中的分类器的一例。图14是表示本实施方式中的分类器的一例的图。

[0130] 如图14所示那样,不管有无原始的数据集合的调整,都确认到KStar、Decorate、RotationForest、以及Random Forest表示最高的性能,因此在本实施方式中,将这些最佳4个算法用于进一步的分析中。在测试阶段中,在分类器对到目前为止未检测到的实例进行分类时,这7个最佳的算法算出各药剂属于有问题的类的概率,使用计算出的概率来制作否决得分(RS)这样的名字的索引。在此,在本实施方式中,也可以使用反调和平均(contra harmonic mean)来将对这些概率进行平均后的值作为RS。RS的值也可以暗示预测化合物是安全(RS接近于0.0)、还是有害(RS接近于1.0)。

[0131] 返回图4,否决得分输出部102g使显示部112显示由否决得分算出部102f计算出的化合物的否决得分(步骤SC-7),结束处理。在此,否决得分输出部102g经由印刷输出部输出化合物的否决得分。

[0132] 在此,参考图15来说明本实施方式中的输出信息的一例。图15是表示本实施方式中的输出信息的一例的图。

[0133] 如图15所示那样,否决得分输出部102g也可以输出药剂和各自的否决得分(0.00-1.00间的值)的列表。在此,有问题的药剂具有接近于1.00的得分值,批准药剂具有接近于0.00的得分。另外,图15表示输入从Drugbank数据库得到的已有的药剂而得出的一例。如此,通过输入成为用户感兴趣的药剂候补的化合物,从而能确认靶标蛋白质和化合物的否决得分。另外,通过使用本实施方式的手法,成功地在已有的1000种批准药和不批准药之间正确地区别,确认了本方法的有效性。

[0134] 以上结束本实施方式中的批准预测装置100的处理的一例的说明。

[0135] [其它实施方式]

[0136] 到此为止对本发明的实施方式进行了说明,但本发明除了上述的实施方式以外,还能在权利要求书的技术思想范围内以各种不同的实施方式实施。

[0137] 例如,将批准预测装置100以独立的形态进行处理的情况作为一例进行了说明,但批准预测装置100也可以根据来自客户终端(与批准预测装置100分开的框体)的请求进行处理,将其处理结果返还给该客户机终端。

[0138] 另外,还能将实施方式中说明的各处理当中作为自动进行的处理而说明的處理的全部或一部分手动进行,或者也能将作为手动进行的处理而说明的處理的全部或一部分用公知的方法自动进行。

[0139] 此外,关于上述文献中和附图中示出的处理次序、控制次序、包含具体的名称、各处理的登录数据或检索条件等参数的信息、画面例、数据库构成,除了特别标记的情况以外都能任意变更。

[0140] 另外,关于批准预测装置100,图示的各构成要素是功能概念的要素,并不一定非要物理上如图示那样构成。

[0141] 例如,关于批准预测装置100的各装置所具备的处理功能、特别是由控制部102进行的各处理功能,能将其全部或任意一部分用CPU(Central Processing Unit)以及由该CPU解释执行的程序实现,另外,也可以作为基于布线逻辑的硬件来实现。另外,程序包含后述的用使计算机执行本发明所涉及的方法的程序化的命令,记录在非临时的计算机可读的记录介质中,根据需要由批准预测装置100进行机械式读取。即,在ROM或HDD(Hard Disk Drive)等的存储部106等中,记录了用于与OS(Operating System,操作系统)联动来对CPU发出命令、并进行各种处理的计算机程序。该计算机程序通过载入到RAM而被执行,与CPU联动来构成控制部。

[0142] 另外,该计算机程序可以存储在经由任意的网络300与批准预测装置100连接的应用程序服务器中,还可根据需要下载其全部或一部分。

[0143] 另外,也可以将本发明所涉及的程序保存在计算机可读的记录介质中,另外,还能构建为程序产品。在此,所谓该“记录介质”,包含存储卡、USB存储器、SD卡、软盘、光磁盘、ROM、EPROM、EEPROM、CD-ROM、MO、DVD、以及Blu-ray Disc等的任意的“可移动的物理介质”。

[0144] 另外,所谓「程序」,是以任意的语言或记述方法记述的数据处理方法,不依赖于源代码或二进制码等的格式。另外,“程序”并不一定限于单一的构成,包含分散为多个模块或库而构成的程序、与以OS(Operating System,操作系统)为代表的另外的程序联动来达成其功能的程序。另外,关于实施方式所示的各装置中用于读取记录介质的具体的构成、读取次序或者读取后的安装次序等,能使用周知的构成和次序。

[0145] 保存在存储部106中的各种数据库等(蛋白质排列信息数据库106a、相似性网络信息数据库106b、药剂靶标数据库106c、以及相互作用网络信息数据库106d)是RAM、ROM等的存储器装置、硬盘等的固定盘装置、软盘、以及光盘等的贮存单元,保存用于提供各种处理或WEB站点各种程序、表、数据库、以及WEB页面用文件等。

[0146] 另外,批准预测装置100可以作为已知的台式或笔记本式个人计算机、便携式电话、智能手机、PHS、以及PDA等的便携式终端装置、和 workstation 等的信息处理装置而构成,另外,也可以在该信息处理装置连接任意的周边装置而构成。另外,批准预测装置100也可以通过在该信息处理装置安装实现本发明的方法的软件(包含程序、数据等)来实现。

[0147] 进而,装置的分散、合并的具体系统并不限于图示,能根据各种附加等或根据功能负载,以任意的单位功能或物理上分散、合并其全部或一部分来构成。即,可以任意组合实施上述的实施方式,也可以选择性地实施实施方式。

[0148] 产业上的可利用性

[0149] 如以上详述说明的那样,根据本发明,能够提供一种能够量化医药品的批准或否

决的可能性的批准预测装置、批准预测方法、以及程序,因此特别在医疗、制药、药物研发、以及生物学研究等各种领域中极其有用。

- [0150] 符号说明
- [0151] 100 批准预测装置
- [0152] 102 控制部
- [0153] 102a 相似性网络信息保存部
- [0154] 102b 相似性中心性尺度算出部
- [0155] 102c 批准判定部
- [0156] 102d 判定结果输出部
- [0157] 102e 相互作用中心性尺度算出部
- [0158] 102f 否决得分算出部
- [0159] 102g 否决得分输出部
- [0160] 104 通信控制接口部
- [0161] 106 存储部
- [0162] 106a 蛋白质排列信息数据库
- [0163] 106b 相似性网络信息数据库
- [0164] 106c 药剂靶标数据库
- [0165] 106d 相互作用网络信息数据库
- [0166] 108 输入输出控制接口部
- [0167] 112 显示部
- [0168] 114 输入部
- [0169] 200 外部系统
- [0170] 300 网络

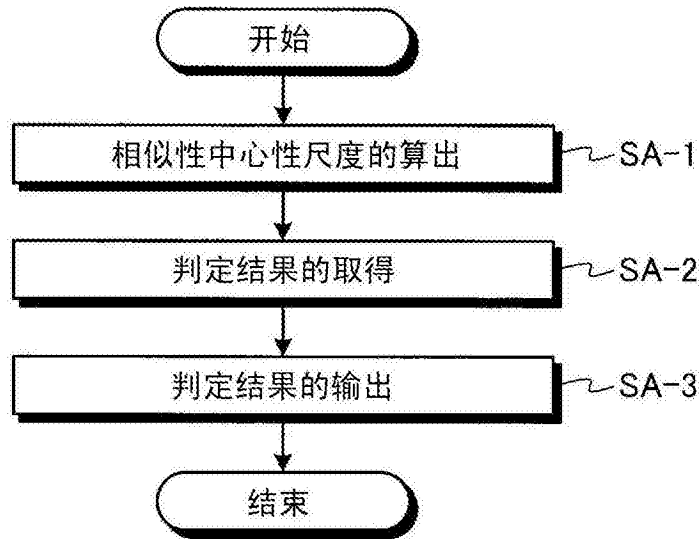


图1

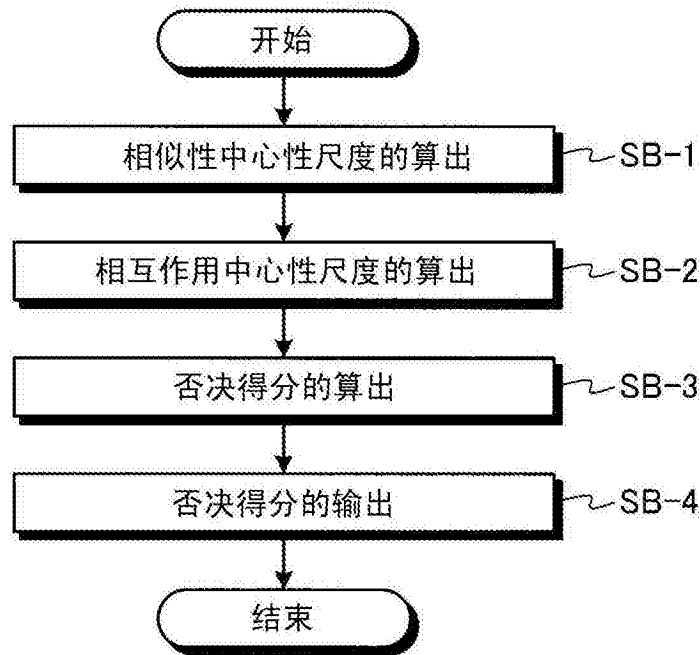


图2

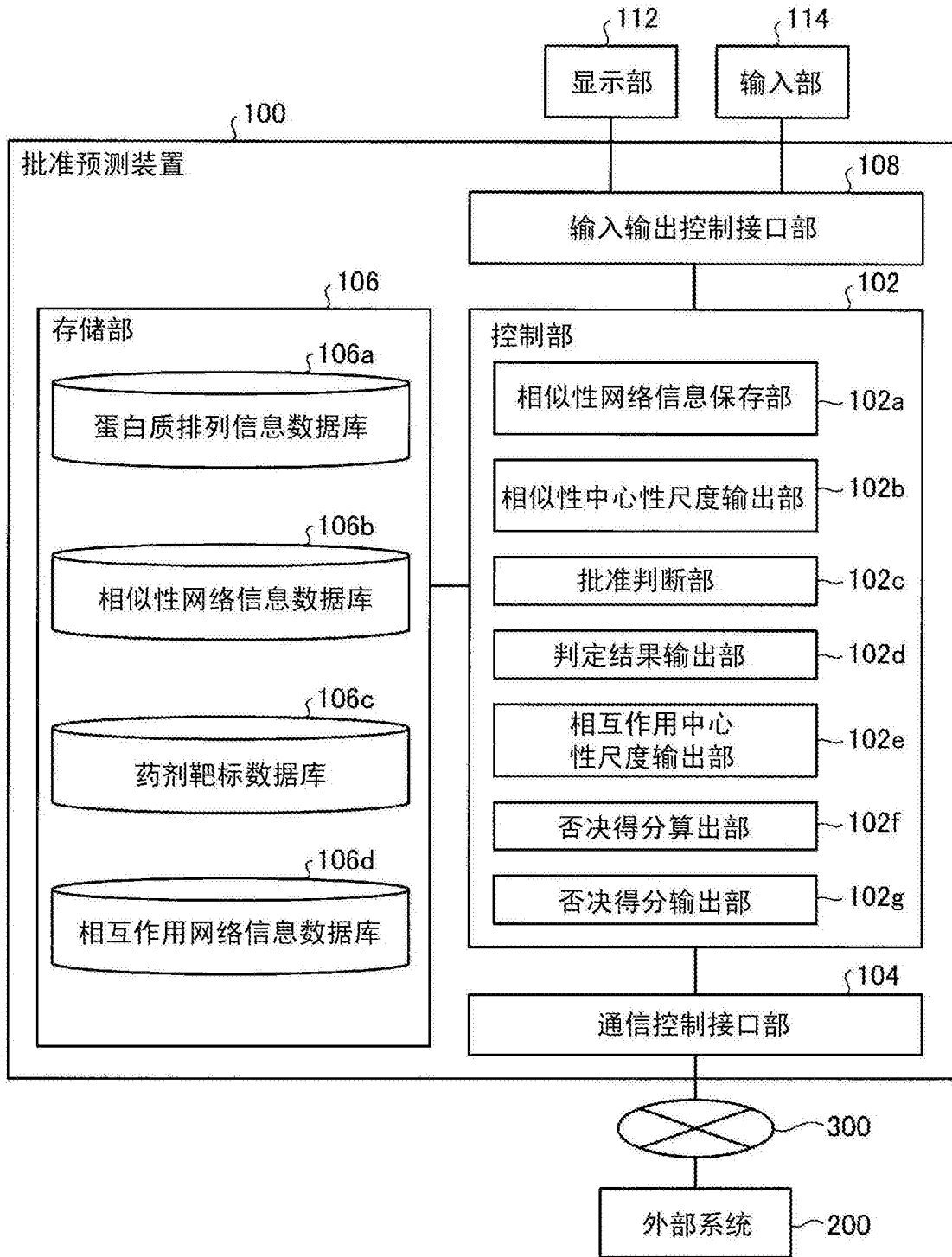


图3

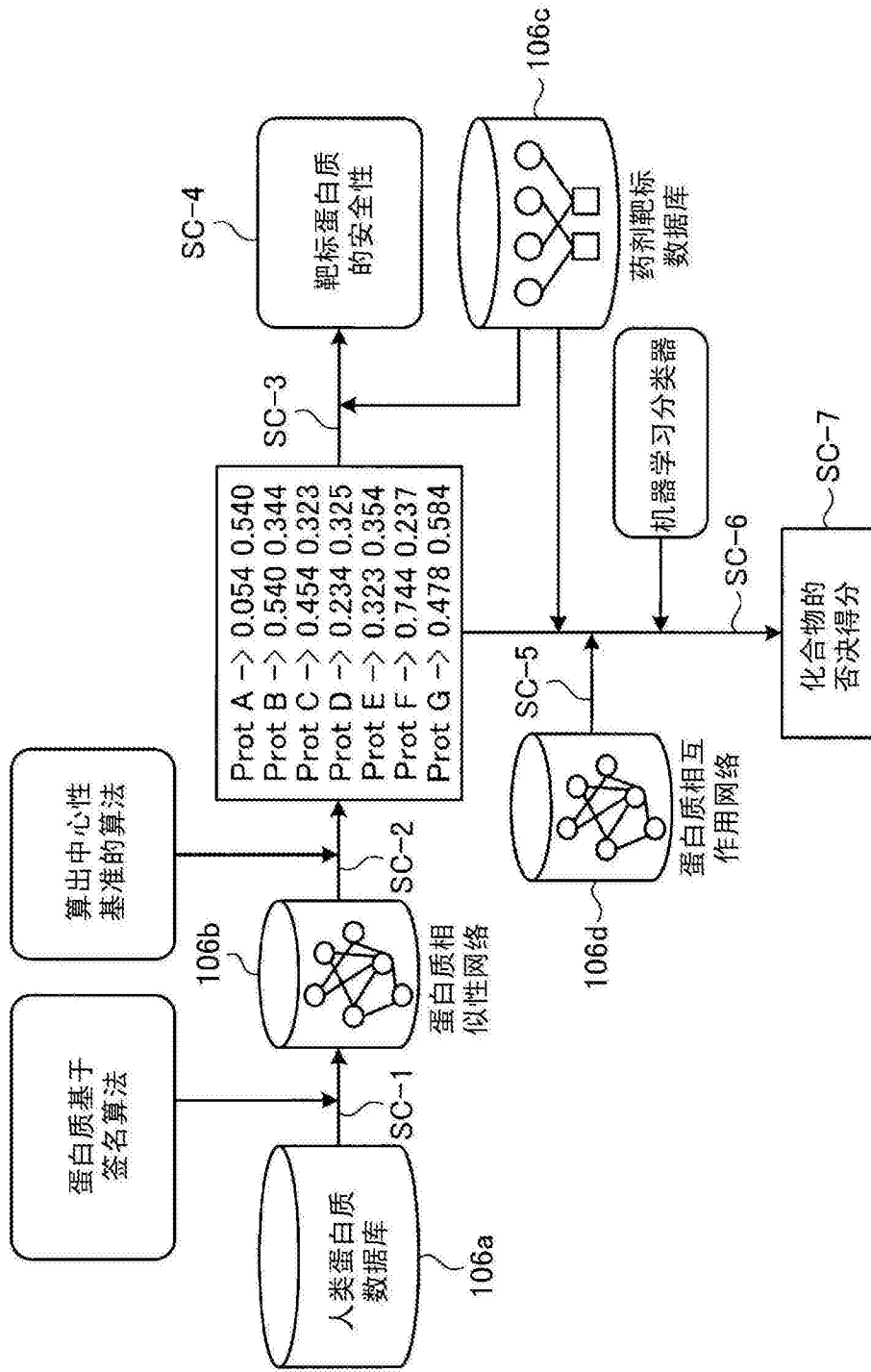


图4

## &gt;P63261

MEEEIAALVIDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQS  
KRGILTLYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKMT  
QIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVTHTVPIYEGYALPHAILRLDL  
AGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEMATAASSSSLEKSY  
ELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMKCDVDIRKDLANTVLS  
GGTTMYPGIADRMQKEITALAPSTMKIKIIPPERKYSVWIGGSILASLSTFQQMWISKQ  
EYDESGPSIVHRKCF

## &gt;P49281

MVLGPEQKMSDDSVSGDHGESASLGNINPAYSNPSLSQSPGDSEEFATYFNEKISIPEE  
EYSCFSFRKLWAF TGPGFLMSIAYLDPGNIESDLQSGAVAGFKLLWILLLATLVGLLLQR  
LAARLGVV TGLHLAEVCHRQYPKVP RVILWLMVELAIIGSDMQEVI GSAIAINLLSVGRI  
PLWGGVLIT IADTFVFLFLDKYGLR KLEAFFGFLITIMALTFGYEYVTVKPSQSQVLKGM  
FVPSCSGCRTPQIEQAVGIVGAVIMPHNMYLHSALVKSRQVNRNNKQEVREANKYFFIES  
CIALFVSFIINVFVVSVFAEAFFGKTNEQVVEVCTNTSSPHAGLFPKDNSTLAVDIYKGG  
VVLGCYFGPAALYIWAVGILAAGQSSTMTGTYSGGQFVMEGFLNLKWSRFARVVLTRSAI  
IPTLLVAVFQDVEHLTGMNDFLNLVLSLQLPALIPILTFTSLRPVMSDFANGLGWRIAG  
GILVLIIC SINMYFVVVYVRDLGHVALYVVAVVSVAYLGFV FYLGWQC LIALGMSFLDC  
GHTCHLGLTAQPELYLLNTMDADSLVSR

图5

Q3M194 Q9Y473 210.901 536 1.67035e-54 82 119 1 214 377  
 GKMFGQNSTLVYHKAJHTGEKPYKCNECGKAFNQSHLSRHHRLHTGEKPYKCNDGCGKAFIHQSSLARHHRLHTGEKSYKCEECDFVFS  
 QKSNLERHKIHTGEKPYKCNECHKTFSHRSSLPCHRRHLSGKPYKCNECGKTFNVQSHLSRHHRLHTGEKPYCKVKVCDKAFMCHSYLA  
 NHTRIHSGEKPKYKCN ECGKAHNH LIDSSIKP C M S S 5 9 0 2 3 0  
 GQLFSSSDACSKNIHTGETFCGKGNQCRKVCQGHKQSLKQHQ-IHTQKKPDGCGSECGSFTQKSHLFAQQRIHSVGNLHECGKCGKAFM  
 POLKLSVYLDHTGDIKCEGKVFQIRSELLTHQKTHTRKPKYKCHDCCGKAFFQMLSLFRHQHRSREKLYECSECGKGFSONSTLIH  
 QKIHTGERQYACSECGKAFTOKSTLSLHQRIHS 442 G++F +S+ K IHTGE K N+C K + L +H +HT +KP C++CG +F  
 +S L R+H+ ++C +C + F + L + HTG+ P C E C K F R S L H++ H+ +KPYKC++CGK F L R H R H+ EK Y+C C K F  
 +S L H +IH+GE+ Y C+ECGKA S+ + S

Q9P2V4 Q8N0V4 94.6147 232 1.58178e-19 47 82 23 291 23  
 CPSQCSCSLHIMGDGSKARTVVCNDPDMTLPASIPPDTSRLRLERTAIRRVPGEAFRPLGRLEQLWLWPYNALSELNALMLRGLRRLREL  
 LPGNRLAAFPWAALRDAPKLRLLDLQANRLSAPAEARFLENLTFDLSSNQLMRLPQELIVSWAHLETGIFPPGHPRRVLGLQDNPWA  
 CDCRLYDLVHLLDGWAPNLAFIETELRCASPRSLAGVAFSQLELRKCGQPELHPGVASIRSLGTTALLRCGATGVPPEMSWRRANGRPL  
 N G T V H Q E V S S D G T S W T L L 3 1 3 3 6  
 GPATCSCTKESIICVGS-----SWVPRIVPGDISSLSLVNGTFSEIKDRMFSLPSLQLLLLNSNSFTIIRDDAFAGLFHLEYLFIEGNKIETISR  
 NAFRGLRDLTHLSLANNHIKALPRDVFSDLSIELDLRGNKFECDCAKAWLYLWKMTNS-----TVSDVLCIGPPEYQEKKLNVDVTSFD  
 YECTTTDFVHQTLPYQSVSDTFNSKNDVYVAIAQPSMENCVMLEWDHIEMNFRSYDNITGQSVGCKAILDDQVVFVVAQLFGGSHIY  
 303 CP+ CSC+ + + P +P D S L L + F L L + L L N+ + + GL L L + GN++  
 A R L L L N + A+P + L++L LDL N+ + + T + Y L D + + T+ V+ +  
 + +L + + S ++ V V+ +

图6

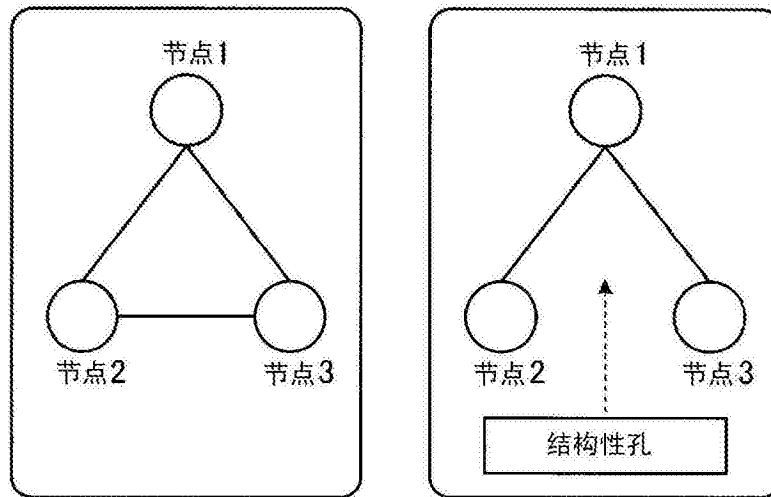


图7

蛋白质 ID	度中心性	介数中心性	距离的中心性	Burt 的约束值
P14784	55	12671.535	0.000	0.043
P14854	1	0.000	5.0709939148073e-05	1.000
P14859	224	544.657	0.000	0.016
P14867	59	0.000	0.000	0.066
P14868	4	0.000	5.07176548156413e-05	0.766
P14902	1	0.000	5.0709939148073e-05	1.000
P14920	1	0.000	5.0709939148073e-05	1.000

图8

药剂	靶标	状态
乙酰水杨酸	P23219	批准
格列噻啉	P48048, Q09428	批准
非那西汀	P23219	撤回

图9

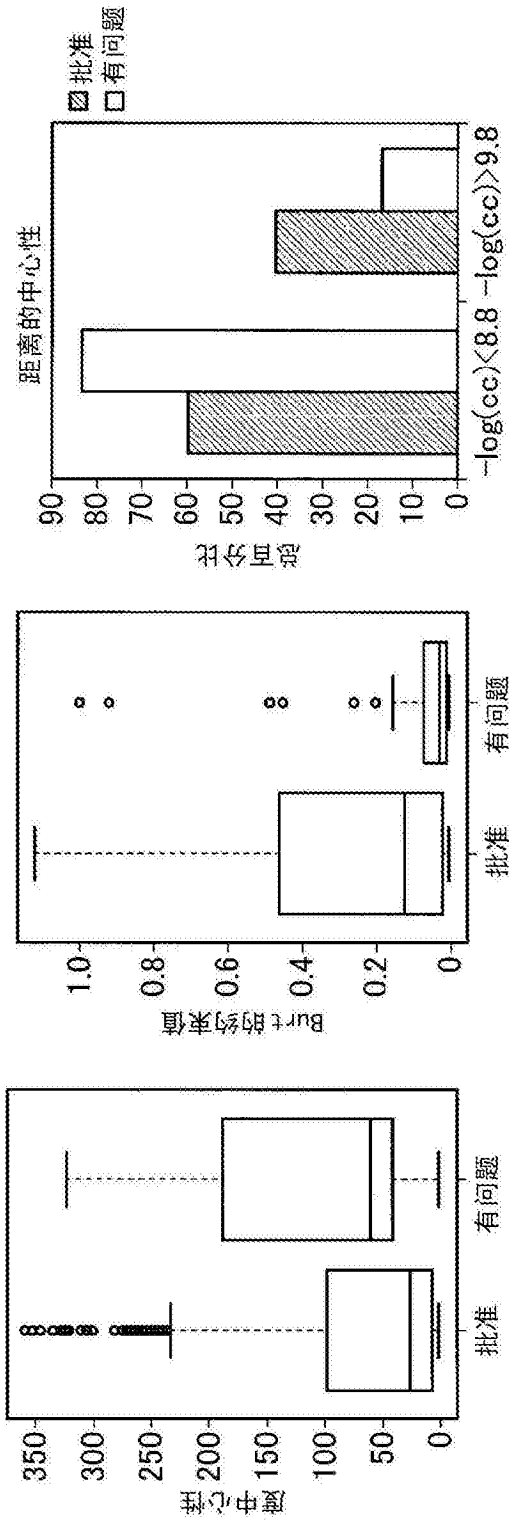


图10

蛋白质 1	蛋白质 2
AL1A1_HUMAN	AL1A1_HUMAN
ITA7_HUMAN	ACHA_HUMAN
NEB1_HUMAN	ACTG_HUMAN
SRGN_HUMAN	CD44_HUMAN
GRB7_HUMAN	ERBB2_HUMAN
PAK1_HUMAN	ERBB2_HUMAN
DLG4_HUMAN	ERBB2_HUMAN
P85B_HUMAN	ERBB2_HUMAN
PTN18_HUMAN	ERBB2_HUMAN

图11

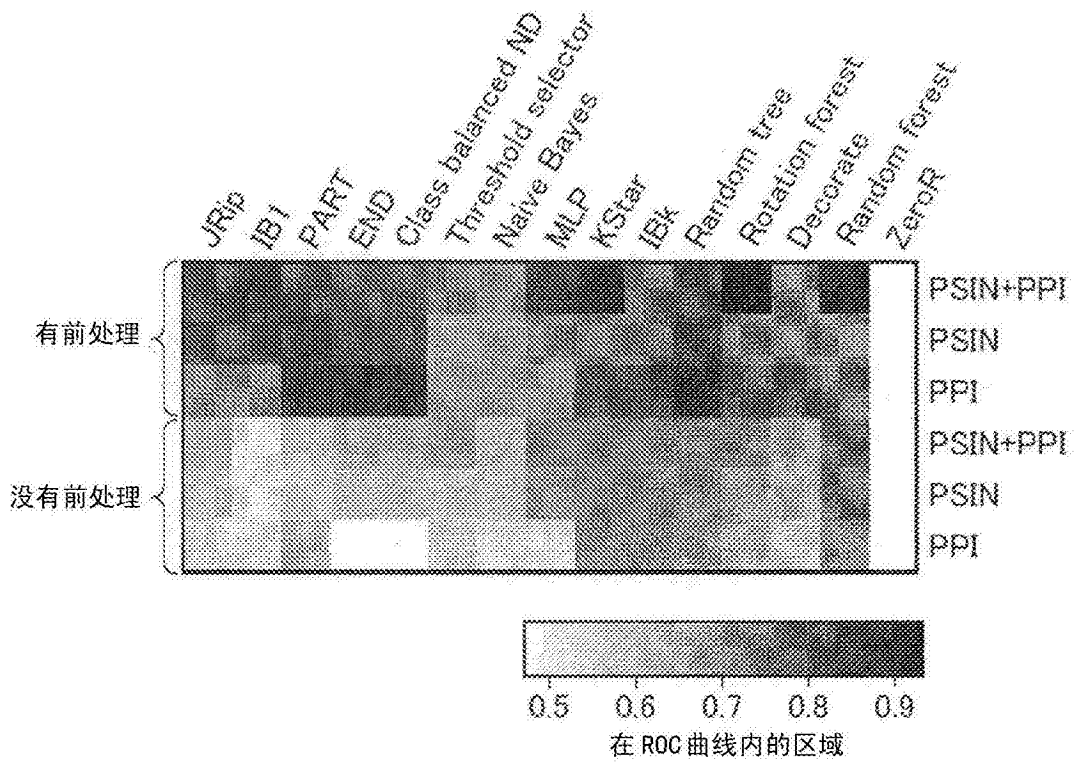


图12

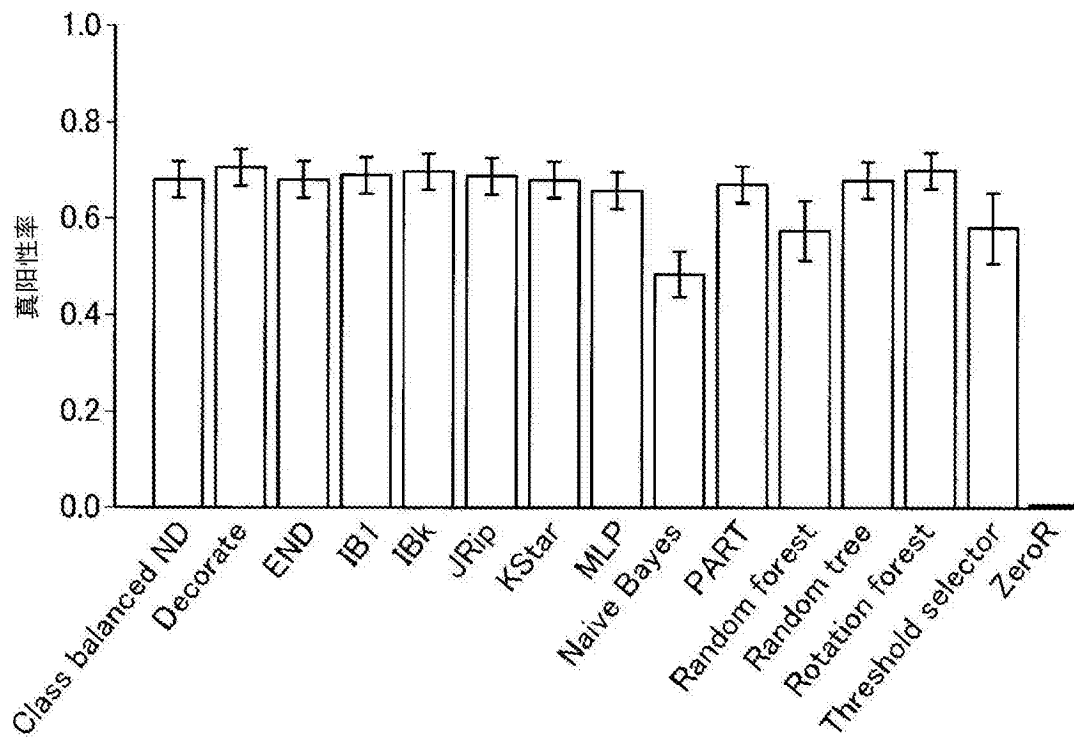


图13

分类器	参考资料
ZeroR	(Witten, et al., 2011)
Naive Bayes	(John and Langley, 1995)
Threshold Selector	(Witten, et al., 2011)
MLP	(Cybenko, 1992)
JRip	(Cohen, 1995)
IB1	(Aha, et al., 1991)
PART	(Frank and Witten, 1998)
END	(Dong, et al., 2005)
Class Balanced ND	(Dong, et al., 2005)
IBk	(Aha, et al., 1991)
Random Tree	(Witten, et al., 2011)
Rotation Forest	(Rodriguez and Kuncheva, 2006)
Random Forest	(Breiman, 2001)
Decorate	(Melville and Mooney, 2003)
KStar	(Cleary and Trigg, 1995)

图14

药物	真实类	算法 1	算法 2	算法 3	算法 4	否决得分
AZD8075	有问题	1.00	1.00	1.00	1.00	0.98
佐美酸	有问题	1.00	1.00	1.00	1.00	0.98
BMS 275291	有问题	1.00	1.00	0.99	1.00	0.98
育享宾碱	批准	0.00	0.00	0.00	0.01	0.01
多虑平	批准	0.00	0.00	0.00	0.01	0.00
利福布汀	批准	0.00	0.00	0.00	0.00	0.00

图15