

基于机器学习的 COX 抑制剂预测模型研究

聂长森¹, 白 勇¹, 柳贤德²

(1. 海南大学 信息科学技术学院, 海南 海口 570228;

2. 海南大学 农学院, 海南 海口 570228)

摘 要:针对目前 COX(环氧合酶)抑制剂较少且抑制效果差的问题,以及传统的化学实验筛选 COX 抑制剂分子的方法中成本高且效率低的问题,基于机器学习算法,提出并建立了一种 COX 抑制剂的预测模型。该模型可高效且准确地找到 COX 抑制剂,通过大量搜集文献中的数据建立数据集,使用 Mold2 软件计算化合物分子描述符,利用自组织特征映射神经网络(SOM)划分训练集和测试集,应用随机森林(RF)和支持向量机(SVM)等机器学习算法分别建立了 COX 抑制剂预测模型。实验对比发现,SOM 结合 RF 算法较传统化学实验方法具有更好的预测精度,且预测效率也有大幅提升。实验研究表明,基于自组织神经网络和随机森林的机器学习方法建立的 COX 抑制剂预测模型,具有很好的分类预测效果,可以为 COX 抑制剂的分析与预测提供有力的研究工具。

关键词:COX 抑制剂;机器学习方法;自组织特征神经网络;随机森林;支持向量机

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2017)10-0074-04

doi:10.3969/j.issn.1673-629X.2017.10.016

Study on COX Inhibitor Prediction Model Based on Machine Learning

NIE Chang-sen¹, BAI Yong¹, LIU Xian-de²

(1. College of Information Science & Technology, Hainan University, Haikou 570228, China;

2. College of Agriculture, Hainan University, Haikou 570228, China)

Abstract: In allusion of the lack in COX (Cyclooxygenase) inhibitor and its poor inhibition effect, moreover for the reason that the traditional COX inhibitor screening must be performed through chemical experiment in high cost and low efficiency, a forecast model of COX inhibitors based on machine learning algorithm is proposed and established. It can find COX inhibitor efficiently and accurately. In the establishing process the data set with huge collection of data in the literature has been built up and then the molecular descriptors with the software of Mold2 has been calculated and divided into training set and testing set with the method of SOM. However, two ML methods, Support Vector Machine (SVM) and Random Forest (RF), are employed to develop a prediction method for searching inhibitors and non-inhibitors of COX from the literature. The verification experiments show that the algorithm of SOM and RF has a better prediction accuracy, which also has a higher efficiency compared with the traditional chemical methods. The results of investigation demonstrate that the COX inhibitor prediction models based on SOM and RF has a good classification prediction effect and provides powerful instrument for analysis and prediction of COX inhibitor.

Key words: COX inhibitors; machine learning; SOM; random forests; support vector machines

0 引 言

如何高效地筛选具有高活性的 COX 抑制剂对于探索人体炎症的治疗具有非常重要的意义。环氧合酶是体内花生四烯酸代谢过程中最主要的限速酶,存在两种亚型:原生型(COX-1)和诱导型(COX-2)^[1-3]。

COX-1 存在于血管、肾脏和胃,具有生理保护作用,如维持胃肠道黏膜的完整性,调节肾血流量和血小板功能;COX-2 是一种诱导酶,在组织损伤、炎症时,细胞因子和其他炎性介质诱导激活炎症部位 COX-2,由此产生 PGG₂/PGH₂,从而出现炎症反应。人类对环氧

收稿日期:2016-11-22

修回日期:2017-03-13

网络出版时间:2017-07-19

基金项目:国家自然科学基金资助项目(31660733);海南省应用技术研发与示范推广专项(ZDXM2015065);海南省社会发展科技专项(SF201421)

作者简介:聂长森(1994-),男,硕士研究生,研究方向为移动通信与智能信息处理;白 勇,教授,博士,研究方向为移动通信与智能信息处理;柳贤德,副教授,博士,通讯作者,研究方向为兽医公共卫生学。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20170719.1113.082.html>

合酶抑制剂的研究一直是药物研究的热点,自 1999 年第一代特异性 COX-2 抑制剂,即昔布类药物—塞来昔布和罗非昔布先后在国外和国内上市以来,专家学者们对 COX 抑制剂的研究从未停止。现已证明,COX-1 不仅参与炎症并且有加重炎症的作用,而 COX-2 似乎主要参与早期炎症,而在慢性炎症阶段反而有抗炎作用。那么如何寻找一种 COX 抑制剂,对 COX-1 和 COX-2 都具有抑制作用,科学家们对此进行了大量研究。基本思想是基于分子描述符和机器学习算法,对前人实验的数据进行分析,利用计算机高效提取 COX 抑制剂的特征,建立 COX 抑制剂的预测模型,并利用现有的 COX 抑制剂进行验证。

设计一个高效的 COX 抑制剂预测模型的任务非常艰巨,虽然近年来科学家们对 COX 抑制剂的研究较多,但是已知的抑制剂化合物非常有限,所采用的化合物来源于文献搜集,一共 54 个,对 COX-1 和 COX-2 具有有效的抑制作用。由于 COX 抑制剂的数据库非常有限,给预测模型的建立带来了很大的挑战,但是机器学习算法凭借其优良的数据筛选特性,一直以来在化和物结构预测、药代动力学、药效动力学等方面均有非常好的效果。为此,在实验中分别建立了随机森林模型和支持向量机的 COX 抑制剂预测模型,并将自组织神经网络(SOM)^[4-8]分别与这两种算法相结合,建立了四种预测模型,并进行了对比验证。

1 材料与方法

分别采用 SOM 结合随机森林(RF)及支持向量机(SVM)等机器学习算法建立了 COX 抑制剂的分析和预测模型。采用的实验数据来源于文献检索的 COX 抑制剂,利用 ChemBioDraw 软件绘制得到其二维(2D)结构,然后使用化合物分子格式转换软件 openbabel 将分子结构转换为 sdf 格式进行保存,使用著名的 Mold2 软件进行分子描述符的计算^[9-13]。化合物量化处理后,利用 SOM 进行训练集和测试集的划分,然后使用随机森林算法对训练集进行学习,对测试集进行预测分析,通过与 SVM 算法的预测结果对比后发现,SOM 结合 RF 算法的预测正确率较高。

1.1 分子描述符

采用国家毒理学研究中心(NCTR)设计的 Mold2 软件进行分子描述符的计算。每个化合物的分子描述符有 777 个数据,分别代表化合物的不同结构和属性,由于许多文献和书籍中都有详细的描述^[14],故只做简单介绍。化合物的分子描述符可以分为经验描述符和理论描述符。

经验描述符来源于物质的实验数据,如熔点、沸点等,因此有其自身的缺点,例如当化合物缺少相应的实

验值时,则不能进行 QSAR 研究。为确保实验的有效性和预测的正确率,不采用经验描述符。

理论描述符是指单从分子结构本身就能计算出来的描述符。采用著名的 Mold2 软件来计算理论描述符,当然还有一些开源工具包或在线服务器也可以计算理论描述符,这里不做赘述。

1.2 数据集

采用的数据集为 54 个 COX 抑制剂,全部从文献中搜集,并且来源于同一个实验室。这些抑制剂的 IC₅₀ 值范围从小于 0.05 μM 到大于 50 μM 。因为这是对化合物进行分类预测,所以根据 IC₅₀ 值将所有抑制剂划分为两类:8 个高活性类(IC₅₀ 值低于 1 μM)和 46 个低活性类(IC₅₀ 值高于 1 μM)。图 1 列出了两个代表性 COX 抑制剂的母本结构。

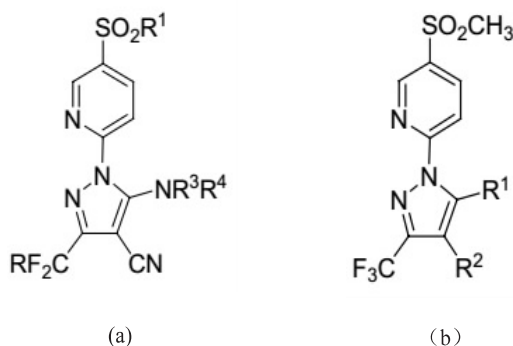


图 1 代表性 COX 抑制剂母本结构

1.3 训练集和测试集的划分

利用计算机进行化合物的特征提取,并进行药性预测,一个重要任务是要进行训练集和测试集的划分。训练集作为输入,由随机森林等智能算法进行特征提取^[15],并建立训练模型。测试集利用建立好的模型进行预测,也就是进行特征匹配,以期得到好的预测效果。测试集是检验训练模型预测结果与实验结果是否一致的重要一步。所以训练集和测试集的划分是否合理,直接决定实验结果的有效性。

训练集和测试集的划分方法主要有以下几种:

(1) 随机划分方法。

该方法是最简单、最易实现也是使用最多的一种训练集和测试集的划分方法。在实现该算法时,只需利用编程语言中的随机函数进行分类即可,但值得注意的是,由于训练集和测试集的划分,对实验结果至关重要。使用该方法一般需要对训练集和测试集进行多次划分,然后取平均结果作为划分模型。

(2) 主成分分析(PCA)划分方法。

主成分分析是用于数值分类研究的一种重要方法,目前的应用也较为广泛。其基本思想是实现多维问题低维化,用二维或者三维欧氏空间的直观散点图来刻划类群或作其他分析。但是实验最重要的是根据

化合物的分子描述符提取特征,并根据数据分布,使训练集和测试集具有最相似的分布,以达到预期效果,所以对目前的实验并不十分合适。

(3)SOM 划分方法^[16]。

该网络是一个由全连接的神经元阵列组成的无教师、自组织、自学习网络。该网络空间中不同区域的神经元具有各自不同的分工,可以根据输入空间中的输入向量进行学习和分类。SOM 神经网络是一种发展较为成熟,经过实验验证的人工智能算法,对训练集和测试集的划分取得了非常好的效果。

1.4 机器学习方法

采用了两种最新的机器学习算法—RF^[17-22] 和 SVM^[23-24],通过 RF 和 SVM 方法将 COX 抑制剂的预测转化为一个二元的分类问题,即通过训练集训练后预测一种化合物是 COX 抑制剂或者不是 COX 抑制剂。对于 RF 和 SVM 算法的具体实现原理,这里不做详述,只对其思想做简单介绍。

RF 是一种决策树自然生长且很多个决策树预测器组合在一起的分类方法^[25]。每棵决策树依赖于对输入向量进行随机独立抽样所获得的数值,且森林中的所有决策树都具有相同的分布。每棵树都不受干涉地自然生长到最大规模,然后对一个新的数据点给出自己的预测。也就是说,这颗树投票决定这一新数据点的类别。当大量的决策树生成以后,整个森林就选择最多数的投票结果作为对这个数据点类别的判定。

SVM 是一种基于统计学习理论中结构风险最小化(SRM)原则的方法^[26],而统计学习理论是一种著名的与核函数相关的机器学习方法。SVM 方法通过使用核函数,把输入变量投射到高维特征空间,然后从输入向量中选择一个所谓支持向量的小的子集。在变换后的空间中,通过最大间隔的原则构建一个最优化的分类超平面,从而把这些输入向量分成了两种不同的类别。

1.5 特征选择方法与模型建立

采用 SOM、RF、SVM 三种算法建立模型。其中自组织特征映射神经网络是一种发展比较成熟的特征分类算法,主要作为训练集和测试集的划分模型。因为训练集和测试集划分的主要目的就是确保训练集的点占据整个数据集空间,测试集的点接近训练集的点。使用 SOM 神经网络的方法对于独立预测集进行选择,这种方法是基于化合物的化学空间来选择测试集分子。使用 3×3 的 SOM 神经网络,把所有化合物映射到 9 个位置。相似的对象映射到相似的位置。在这个网络中按照训练集和测试集 3 : 2 的比例进行选择。训练集用来建立分类模型,而测试集用来评估模型的预测能力。其中,训练集包含 33 个分子(29 个低活性,4

个高活性),测试集包含 21 个分子(17 个低活性,4 个高活性)。然后分别建立随机森林和支持向量机模型进行训练和测试,并对两个模型的预测结果进行对比。

2 实验结果与分析

2.1 模型的预测性能

对 COX 抑制剂进行预测,根据划分的训练集进行训练,测试集的数据根据预测结果与实际化合物的活性是否相同,来对预测的正确率进行判断。将预测正确率记为 P ,测试集高活性化合物数量为 M ,低活性化合物数量为 N ,预测正确的高活性化合物数量为 m ,预测正确的低活性化合物数量为 n 。则总预测正确率为:

$$P = \frac{m + n}{M + N} \times 100\%$$
 (1)

对高活性化合物的预测正确率为:

$$P_1 = \frac{m}{M} \times 100\%$$
 (2)

对低活性化合物的预测正确率为:

$$P_2 = \frac{n}{N} \times 100\%$$
 (3)

分别采用四种方法进行预测,正确率如表 1 所示。

表 1 四种预测模型正确率

预测	随机划分训	随机划分训	SOM 划分	SOM 划分
正确	练集结合随	练集结合支持	训练集结合	训练集结合
率	机森林方法	向量机方法	随机森林	合支持向
			方法	量机方法
P	76.2% ~ 85.7%	76.1% ~ 90.5%	80.9%	85.7%
P_1	25% ~ 75%	25% ~ 50%	75%	25%
P_2	71% ~ 92%	94.1% ~ 94.7%	82.4%	100%

2.2 模型的分析

通过表 1 发现,训练集和测试集的划分方法,对实验结果有着至关重要的作用。运用随机划分的方法,明显出现了预测结果不稳定的现象,这主要是由于每次随机划分的训练集和测试集不能各自完整地代表整个数据集的特征,并且每次实验受划分的高活性化合物和低活性化合物的数量影响较大,导致结果非常不稳定,预测正确率波动较大。而采用自组织特征映射神经网络方法划分训练集和测试集,实验结果比较稳定,而且 SOM 结合 RF 算法^[27-29] 的整体预测正确率在 80.9% 左右,SOM 结合 SVM 算法的正确率稳定在 85.7% 左右。

可以发现,虽然 SOM 结合 SVM 算法的整体预测正确率较高^[30-34],但是对于高活性化合物的预测正确率却相当低,所以对于今后的研究帮助不是很大。SOM 结合 RF 算法的预测正确率相对比较理想。

2.3 模型的验证

通过文献调研查询了 15 个 COX 抑制剂药物,对研究建立的 COX 抑制剂预测模型进行验证。首先使用 ChemBioDraw 软件绘制这 15 个药物的二维结构,然后使用 openbabel 软件进行格式转换,并使用 Mold2 软件计算分子描述符。将这 15 个化合物的分子描述符输入研究建立的基于 SOM 和随机森林算法的预测模型进行预测,预测结果如表 2 所示。其中,预测结果‘1’表示该药物为高活性,预测结果‘2’表示该药物为低活性。

表 2 COX 抑制剂药物预测结果

序号	产品名称	类别	预测结果
1	右旋酮基布洛芬	COX 抑制剂	1
2	呱氨托美丁	COX 抑制剂	1
3	罗非昔布	COX 抑制剂	1
4	Valdecoxib	COX 抑制剂	1
5	Parecoxib	COX 抑制剂	1
6	美洛西康	COX 抑制剂	2
7	奥芬酸钠	COX 抑制剂	2
8	米氟米特	COX 抑制剂	2
9	塞来昔布	COX 抑制剂	1
10	Eltenac	COX 抑制剂	2
11	Indomethacin	COX 抑制剂	1
12	Curcumin	COX 抑制剂	1
13	Tepoxalin	COX 抑制剂	1
14	resveratrol	COX 抑制剂	1
15	Arthroicine	COX 抑制剂	1

由预测结果可知,该模型将这 15 个化合物预测为高活性 COX 抑制剂,由于这些药物已经上市并投入使用,多数是 COX 的高活性抑制剂。预测结果表明有 73.3% 的药物为高活性,符合实际情况,表明该研究在 COX 抑制剂的预测方面具有重要的现实意义。

3 结束语

将分子描述符和机器学习方法应用于 COX 抑制剂,提出并建立了 COX 抑制剂的分类和预测模型。研究结果表明,相对于 SOM 结合 SVM 算法、随机划分训练集结合 RF 算法、随机划分训练集结合 SVM 算法,SOM 结合 RF 的机器学习算法预测准确率高且效果好,同时还可节省大量时间和资源成本。

参考文献:

[1] Sakya S M, DeMello K M L, Minich M L, et al. 5-heteroatom substituted pyrazoles as canine COX-2 inhibitors. Part 1: structure-activity relationship studies of 5-alkylamino pyrazoles and discovery of a potent, selective, and orally active analog[J]. Bioorganic & Medicinal Chemistry Letters, 2006, 16(2):288-292.

[2] Sakya S M, Cheng H, DeMello K M L, et al. 5-heteroatom-substituted pyrazoles as canine COX-2 inhibitors. Part 2: structure-activity relationship studies of 5-alkylethers and 5-thioethers[J]. Bioorganic & medicinal Chemistry Letters, 2006, 16(5):1202-1206.

[3] Sakya S M, Hou X, Minich M L, et al. 5-heteroatom substituted pyrazoles as canine COX-2 inhibitors. Part III: molecular modeling studies on binding contribution of 1-(5-methylsulfonyl) pyrid-2-yl and 4-nitrile[J]. Bioorganic & Medicinal Chemistry Letters, 2007, 17(4):1067-1072.

[4] Kobuchi Y, Tanoue M. Learning and forgetting-how they should be balanced in SOM algorithm[C]//IEEE international conference on neural networks. San Francisco, CA, USA: IEEE, 2004:745-749.

[5] 姚登举,杨静,詹晓娟. 基于随机森林的特征选择算法[J]. 吉林大学学报:工学版, 2014, 44(1):137-141.

[6] 黄亚捷,叶回春,张世文,等. 基于自组织特征映射神经网络的中国耕地生产力分区[J]. 中国农业科学, 2015, 48(6):1136-1150.

[7] 谢倩倩,李汀芳,章文. 基于集成学习的离子通道药物靶点预测[J]. 计算机科学, 2015, 42(4):177-180.

[8] 聂斌,郝竹林,桂宝,等. 基于随机森林的中药寒、热药性代谢组学判别方法研究[J]. 江西中医药大学学报, 2015(2):82-86.

[9] 闫树英,陈志宏,惠娜,等. 基于 RF 和 KNN 的三种肝炎分类模型的建立[J]. 宁夏医学杂志, 2015, 37(6):496-498.

[10] 刘建伟,刘媛,罗雄麟. 半监督学习方法[J]. 计算机学报, 2015, 38(8):1592-1617.

[11] 何冰,罗勇,李秉轲,等. 基于分子描述符和机器学习方法预测和虚拟筛选乳腺癌靶向蛋白 HEC1 抑制剂[J]. 物理化学学报, 2015, 31(9):1795-1802.

[12] 任伟,孔德信. 定量构效关系研究中分子描述符的相关性[J]. 计算机与应用化学, 2009, 26(11):1455-1458.

[13] 郝明. 基于化学信息学方法的药物分子计算研究[D]. 大连:大连理工大学, 2012.

[14] 俞书浩. 功能基因组学和化学信息学协同的药物研发数据挖掘方法[D]. 上海:上海交通大学, 2013.

[15] 曹正凤. 随机森林算法优化研究[D]. 北京:首都经济贸易大学, 2014.

[16] 白耀辉,陈明. 利用自组织特征映射神经网络进行可视化聚类[J]. 计算机仿真, 2006, 23(1):180-183.

[17] 张华伟,王明文,甘丽新. 基于随机森林的文本分类模型研究[J]. 山东大学学报:理学版, 2006, 41(3):5-9.

[18] 袁芳娟. 基于随机森林的年龄估计[D]. 天津:河北工业大学, 2012.

[19] 雍凯. 随机森林的特征选择和模型优化算法研究[D]. 哈尔滨:哈尔滨工业大学, 2008.

[20] 李欣海. 随机森林模型在分类与回归分析中的应用[J]. 应

查阅文献,根据最小二乘法的性质可知,采用最小二乘法得到的线性回归方程的估计误差最小^[15]。新分类器在对无监督样本处理的过程中,以标准误差为分类依据,保证了每一类样本标准误差最小,即新分类器相比较经典分类器的准确率较高。

4 结束语

大数据的涌现使人们在处理复杂的数据对象时面临巨大的挑战。数据的多源异构、质量的良莠不齐使得传统的机器学习法不能有效地处理。对于模式识别中无监督数据的处理问题,单从主观性选取数据特征对数据做分类处理无法保证分类模型的准确度。因此立足于研究分析的目的,选择合适的量化指标才能有效地建立以数据为中心的分类模型。

现有的模式识别中处理分类问题的方法通常会选择样本均值、样本众数、样本中位数作为分类特征,并计算最终样本分类结果的错判率来判断该分类特征的优劣性,然而这种方法并不能保证样本分类结果的错判率最小,人为主观选择分类特征会降低模型的拟合效果。针对这一不足,提出了一种利用标准误差作为分类特征的线性分类器。分类模型的标准误差的大小反映了分类结果的错判率高低,保证分类模型中标准误差最小也就是保证样本分类模型的错判率最小,该方法可以高效地对样本进行分类。实验结果表明,该分类器得到的分类模型准确率较高。

参考文献:

- [1] 徐 鹏,林 森.基于 C4.5 决策树的流量分类方法[J].软件学报,2009,20(10):2692-2704.
- [2] 张建萍,刘希玉.基于聚类分析的 K-means 算法研究及应用[J].计算机应用研究,2007,24(5):166-168.

(上接第 77 页)

- 用昆虫学报,2013,50(4):1190-1197.
- [21] 游 伟,李树涛,谭明奎.基于 SVM-RFE-SFS 的基因选择方法[J].中国生物医学工程学报,2010,29(1):93-99.
- [22] 马景义,谢邦昌.用于分类的随机森林和 Bagging 分类树比较[J].统计与信息论坛,2010,25(10):18-22.
- [23] 董 婷.支持向量机分类算法在 MATLAB 环境下的实现[J].榆林学院学报,2008,18(4):94-96.
- [24] 曹东升.化学生物信息学新方法及其在医药研究中的应用[D].长沙:中南大学,2013.
- [25] 刘孝良.基于半监督学习的随机森林算法研究与应用[D].青岛:中国海洋大学,2013.
- [26] 廖明桥.基于支持向量机的半监督式分类学习方法[D].哈尔滨:哈尔滨工程大学,2013.
- [27] 刘晓东.基于组合策略的随机森林方法研究[D].大连:大

- [3] 张高胤,谭成翔,汪海航.基于 K-近邻算法的网页自动分类系统的研究及实现[J].计算机技术与发展,2007,17(1):21-23.
- [4] 平 源.基于支持向量机的聚类及文本分类研究[D].北京:北京邮电大学,2012.
- [5] 张 婷,戴 芳,郭文艳.基于 ISODATA 聚类的词汇树图像检索算法[J].计算机科学,2014,41(11A):123-127.
- [6] 李 飞,薛 彬,黄亚楼.初始中心优化的 K-Means 聚类算法[J].计算机科学,2002,29(7):94-96.
- [7] 钱夕元,邵志清.模糊 ISODATA 聚类分析算法的实现及其应用研究[J].计算机工程与应用,2004,40(15):70-71.
- [8] Ng R T,Han J. CLARANS:a method for clustering objects for spatial data mining[J].IEEE Transactions on Knowledge & Data Engineering,2002,14(5):1003-1016.
- [9] 黄 韬,刘胜辉,谭艳娜.基于 k-means 聚类算法的研究[J].计算机技术与发展,2011,21(7):54-57.
- [10] 高学军,王振友.多元统计回归模型在医疗保障基金数额分配中的应用[J].统计与决策,2009(9):145-146.
- [11] Everitt B. Cluster analysis[J]. Quality & Quantity,1980,14(1):75-100.
- [12] Tufekci P,Kaya H. Combined cycle power plant data set[DB/OL].(2014-03-26)[2016-05-02].<http://archive.ics.uci.edu/ml/datasets/Combined%20Cycle%20Power%20Plant>.
- [13] 朱佳贤.无指导学习环境下基于属性相关性分析和聚类算法的属性选择问题研究[J].管理学报,2005,2(S):162-165.
- [14] German D M,Adams B,Hassan A E. The evolution of the R software ecosystem[C]//17th European conference on software maintenance and reengineering. [s. l.]:IEEE,2013:243-252.
- [15] 丁克良,沈云中,欧吉坤.整体最小二乘法直线拟合[J].辽宁工程技术大学学报:自然科学版,2010,29(1):44-47.

连理工大学,2013.

- [28] 方匡南,吴见彬,朱建平,等.随机森林方法研究综述[J].统计与信息论坛,2011,26(3):32-38.
- [29] 王爱平,万国伟,程志全,等.支持在线学习的增量式极端随机森林分类器[J].软件学报,2011,22(9):2059-2074.
- [30] 陈永健.半监督支持向量机分类方法研究[D].西安:陕西师范大学,2014.
- [31] 张灿淋.基于支持向量机的半监督式增量学习研究[D].杭州:浙江工业大学,2014.
- [32] 赵 莹.半监督支持向量机学习算法研究[D].哈尔滨:哈尔滨工程大学,2010.
- [33] 周志华.基于分歧的半监督学习[J].自动化学报,2013,39(11):1871-1878.
- [34] 杨南海,黄明明,赫 然,等.基于最大相关熵准则的鲁棒半监督学习算法[J].软件学报,2012,23(2):279-288.