

doi:10.3969/j.issn.1001-4616.2019.03.019

基于聚类核的核极速学习机

王丽娟^{1,2}, 丁世飞¹

(1.中国矿业大学计算机科学与技术学院,江苏 徐州 221116)
(2.徐州工业职业技术学院信息与电气工程学院,江苏 徐州 221140)

[摘要] 传统的神经网络学习算法(如 BP 算法)需要调整大量的网络参数,例如输入权值以及隐层单元的偏置,而极速学习机只需要设置网络的隐层节点个数,在算法执行过程中不需要调整网络的输入权值,便可以产生唯一的最优解,因此它具有学习速度快且泛化性能好的优点.随着极速学习机的研究发展,核极速学习机的相关理论被提出.核极速学习机是将核函数引入到极速学习机中,可以得到最小二乘解,具有更稳定的泛化性能.本文在核极速学习机的基础上提出了一种基于 Bagged 聚类核的核极速学习机的分类方法,首先对已有的标记样本和所有的无标记样本采用多次 k 均值聚类,去构造 Bagged 聚类核,然后对 Bagged 聚类核和径向基核进行求和,最终用于核极速学习机的训练中.与传统核极速学习机相比,本文提出的方法可以使用所有的无标记样本,从而尽可能地提高分类的准确率.最后本文通过实验验证了方法的可行性.

[关键词] 极速学习机, k 均值聚类, Bagged 聚类核, RBF 核函数

[中图分类号] TP3 [文献标志码] A [文章编号] 1001-4616(2019)03-0145-06

A Novel Extreme Learning Machine Based on Bagged Kernel

Wang Lijuan^{1,2}, Ding Shifei¹

(1.School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)
(2.School of Information and Electrical Engineering, Xuzhou College of Industrial Technology, Xuzhou 221140, China)

Abstract: The traditional neural network learning algorithm (BP algorithm) need to set a large amount of network training parameter, and prone to local optimal solution. Extreme learning machine (ELM) need to set the number of hidden layer nodes of networks, while execution of the algorithm does not need to adjust the network weights of the input and hidden element bias, and can produce the optimal solution, thus it has advantages of fast learning speed and good generalization capability. Extreme learning machine as a kind of machine learning method, with simple and easy to use, and effective single hidden layer feed forward neural network learning algorithm, caught the attention of more and more researchers. With the research and development of extreme learning machine, the theory of nuclear extreme learning machine has been continually raised. Nuclear ultimate learning machine is introduced to limit the kernel learning machine, with which you can get a least-squares optimization solution, a more stable, better generalization performance. We now put forward a novel extreme learning machine based on bagged kernel classification method. First of all, the existing tag samples and all unmarked samples use k -means clustering algorithm for many times to construct the bagged clustering nucleus. Then, bagged clustering nucleus and the radial basis calculate the sum, and eventually it is used in training and classification of extreme learning machine. Compared with the traditional extreme learning machine, the new algorithm can use all unmarked sample information, as much as possible to improve the classification accuracy, and further improve the running speed. Through the experimental data set, we verify the feasibility of the method.

Key words: ELM, k -means clustering, Bagged kernel, RBF kernel function

ELM 是一种简单有效的单隐层前馈神经网络,2004 年由南洋理工大学 Huang G B 教授^[1-2]提出. ELM 基于随机特征映射理论^[3],只需要人为设定网络的隐层节点数目,在算法运行期间不需要调整网络的输

收稿日期:2019-07-05.

基金项目:国家自然科学基金项目(61672522、61379501).

通讯联系人:丁世飞,博士,教授,博导,研究方向:人工智能、机器学习、支持向量机等. E-mail:dingsf@cumt.edu.cn

入层和隐层之间的参数,并且一次学习就可得最优解,具有较快的学习速度.此后,ELM 被广泛地研究和应用^[4-7].在 ELM 的基础上,Huang 提出核极速学习机的概念^[8-9],核方法是一种非常有效的机器学习方法^[10].相比传统的算法,核 ELM(KELM)具有以下优势,首先,相比传统的极限学习算法,KELM 可以解决传统 ELM 算法随机初始化的问题,并且对模型参数具有较好的鲁棒性.其次,稀疏 KELM 能用较少的样本表示出较多的样本特性,能提高时空效率^[11-12].KELM 分为单核和多核算法,多核 ELM 在数据异构、数据不规则、样本不平坦等情况下能较好地应用^[13-14].最后,KELM 可以较好地应用在图像处理中,如人脸识别、图像分类等,同时在故障诊断中也有广泛的应用^[15-17].

传统的 ELM 与 KELM 模型更多的是应用于有监督的学习训练,然而大量标记数据的获取通常比较困难,若只使用少量标记数据,有监督学习训练所得学习模型不具有良好的泛化能力,同时浪费了大量未标记数据所包含的信息.在 ELM 模型的研究中,目前关于无监督和半监督的 ELM 算法相对较少,传统的方法大多基于拉普拉斯矩阵和流形理论,构造关于数据的无监督先验信息来作为正则化和初始化的手段,且这些传统的方法很少用在 KELM 模型中.因此,如何有效地利用少量标记数据和大量未标记数据共同改善学习性能的半监督学习成为当前 KELM 模型研究的热点问题之一^[18-19].并且在数据分布不平稳的时候,使用传统的 RBF 核不能很好地建模数据的分布特征.根据以上问题,本文提出了一种基于聚类核(依据聚类假设思想的半监督核)极速学习机算法,聚类核可以有效地利用样本的无监督特征,尝试在原数据分布的基础上,对数据进行无监督特征映射,得到更加平稳的数据分布,因此,该算法可以同时用少量标注样本和大量未标记样本进行训练,在原有 RBF 核的基础上引入 Bagged 聚类核,设置并调整 RBF 核与聚类核的权重系数,在不造成过高成本消耗的前提下,可以得到比较高的学习精度,达到令人满意的分类效果.

1 极速学习机的基本原理

ELM 是 Huang 等提出的一种高效简洁的单隐层前馈神经网络的学习算法.ELM 模型由输入层、隐层和输出层组成,其模型结构如图 1 所示,该模型具有 j 个输入层神经元、 n 个隐层神经元、 m 个输出层神经元,在隐层上的激活函数是 $g(x)$.

对于 N 个不同的学习样本 $(x_i, y_i) \in \mathbf{R}^n \times \mathbf{R}^m (i = 1, 2, \dots, N)$,如图 1 所示,隐层神经元数目为 n ,隐层激活函数为 $g(x)$,那么隐层的输出可以表示为式(1),隐层的输出和输出神经元的输出值关系可以表示为式(2).

$$h = g(ax + b), \tag{1}$$

$$h(x_i) V = y_i, \quad i = 1, 2, \dots, N, \tag{2}$$

式(2)又可以写成:

$$HV = Y, \tag{3}$$

式中,

$$H = \begin{bmatrix} g(a_1, b_1, x_1) & g(a_1, b_1, x_2) & \cdots & g(a_n, b_n, x_N) \\ g(a_2, b_2, x_1) & g(a_2, b_2, x_2) & \cdots & g(a_n, b_n, x_N) \\ \vdots & \vdots & \ddots & \vdots \\ g(a_n, b_n, x_1) & g(a_n, b_n, x_2) & \cdots & g(a_n, b_n, x_N) \end{bmatrix}^T,$$

$$V = \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix}_{n \times m}, \quad Y = \begin{bmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times m},$$

式中, a_{ij} 表示第 i 个输入层的节点与第 j 个隐层上的节点的连接矩阵上的值, b_j 表示隐层上的第 j 个神经元

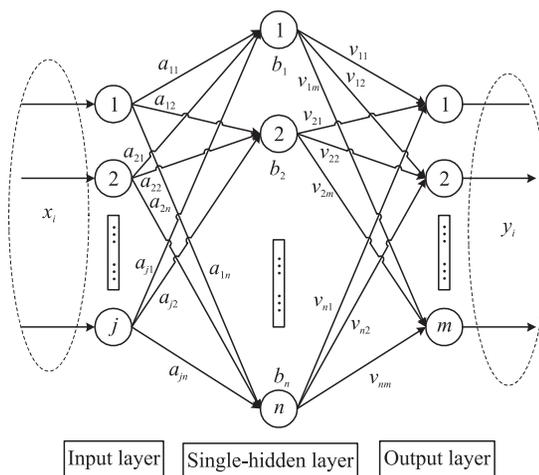


图 1 极速学习机的模型结构

Fig. 1 The model structure of ELM

节点的偏置值, u_{jk} 表示隐层上的第 j 个神经元节点与输出层第 k 个输出神经元节点的连接矩阵的值, \mathbf{H} 表示输出矩阵. 可以证明, 如果激活函数 $g(x)$ 无限可微, 那么在算法运行的期间不需要调整网络的输入权值以及隐层的偏置, 输出权值可以通过输入矩阵 \mathbf{a}_j , 隐层的偏置值 \mathbf{b}_j 和输出矩阵 \mathbf{H} 经过一系列的线性变换求得.

综上所述, 可知用极速学习机求隐层和输出层的输出矩阵可分为 3 步:

- (1) 任意选择 0~1 之间的数值输入输入权值 \mathbf{a}_j 和隐层的偏置值 \mathbf{b}_j ;
- (2) 计算网络输出层的输出矩阵 \mathbf{H} ;
- (3) 计算网络隐层和输出层之间的输出权值 \mathbf{V} :

$$\mathbf{V} = \mathbf{H}^\dagger \mathbf{Y}, \quad (4)$$

式中, \mathbf{H}^\dagger 表示隐层输出矩阵 \mathbf{H} 的广义逆矩阵.

2 基于 Bagged 聚类核的极速学习机

2.1 聚类核

在早期工作中, Chapelle 等提出了一种通过修改核矩阵的特征谱去构造聚类核的框架的方法^[20], 包括随机游走核和谱聚类核. 随机游走核是一个规范化和对称的过渡矩阵, 其对应于一个 t 步随机游走. 可以将 RBF 核看作是一个在含有结点 x_i 的图上的进行随机游走所得到的过渡矩阵, 随后可以通过对规范化的过渡矩阵进行谱分解计算 t 步随机游走核. 随机游走核依据的是马尔可夫随机游走的思想, 将所有的样本点视为一个全连通图上的结点, 而两两结点之间的距离等于随机游走过程中结点之间的状态转移概率, 因而可以计算得到一个 $(m+n) * (m+n)$ 的转移矩阵, 其中 $(m+n)$ 是样本点的数目, 可以设想对于每个结点, 都存在一个可以运动的粒子, 它们按某种转移概率随机游走到其他结点. 那么 t 阶转移矩阵的第 i 行 p'_i 就是某个粒子 x_i 在经过 t 步跳转后的概率分布. 则经过 t 步转移之后, 出发位置在 x_i 的粒子跳转到顶点 x_j 的概率是 p'_{ij} , 由于转移概率矩阵可以识别出具有高的类内转移概率和低的类间转移概率, 从而可以有效避免样本点在初始状态下分布的不确定性, 进而更好地反映数据间的相似度. 因而在随机游走核中, 根据样本点之间的近邻关系建立相似度矩阵, 计算 t 步转移概率矩阵, 并将其用于核矩阵的计算中. 其中 t 主要用于控制在新的特征空间中样本之间分布的密度. 谱聚类核则使用由谱聚类算法所得到的新的特征向量, 即在规范化的相似度矩阵中使用前 k 个特征向量. 对样本点在特征空间中进行重新表示, 使得位于同一聚类或高密度区域中的样本点在新的空间中分布得更紧凑. 这些方法的主要缺陷是必须对角化一个大小为 $(m+n)$ 的矩阵, $(m+n)$ 是标记样本和非标记样本之和.

2.1.1 Bagged 聚类核的基本思想

Jason Weston, Christina Leslie 等人提出了通过多次 k 均值聚类去构造 Bagged 聚类核^[16], 该方法的思路是通过 k 均值聚类算法估计两两样本位于同一聚类的概率, 其时间复杂度为 $O(rk(m+n)d)$ 的样本数据的维数为 d , k 均值聚类运行次数为 r . Bagged 聚类核主要依据聚类假设, 即位于高密度区域的两个点应该有相同的标记, 对两两样本之间的距离或相似度进行调整. 使得样本间的相似度更准确地符合样本的实际分布.

2.1.2 Bagged 聚类核的构造过程

(1) 运行 N 次 k 均值聚类运算, 对于样本点, 每次的聚类结果为 $C_j(x_i)$, 其中 j 为聚类运算的次数. $j=1, 2, 3, \dots, N$, 其中 x 是所有的标记样本和部分非标记样本. 每次的 k 均值聚类中心都是随机产生的.

(2) 利用如下公式, 建立 Bagged 聚类核:

$$K_{\text{bag}}(x, x') = \frac{\sum_j [c_j(x) = c_j(x')]}{N}, \quad (5)$$

式中, 在第 j 次迭代过程中, 如 x 和 x' 在同一类中则 $[C(x_i) = C(x_j)]$ 返回值为 1, 否则返回值为 0.

$K_{\text{bag}}(x, x')$ 是一个满足 Mercer 定理的核, 它等价于一个在 N_k 维空间 $\phi(x_i) = \langle [c_j(x) = q] : j=1, \dots, k \rangle$ 之间的内积.

2.2 基于 Bagged 聚类核的极速学习机分类方法

基于聚类核的 ELM 分类方法首先对样本中的多数样本进行聚类构建一个 Bagged 聚类核,随后与径向基核进行相加. 以下为具体的构造算法:

- (1) 设标记样本的数目为 m , 非标记样本的数目为 n . 构建一个 $m * (m+n)$ 的零矩阵.
- (2) 使用 RBF 核函数, 计算标记样本集的核矩阵 K_{train} , 核矩阵 K_{train} 为 $m * m$ 的矩阵. 计算标记样本和非标记样本所组成的测试核矩阵, 测试核矩阵 K_{test} 为 $m * n$ 的矩阵.
- (3) 假设样本集合 X 由标记样本和非标记样本组成. 对样本集合 X 运行 N 次 k 均值聚类.
- (4) 根据聚类结果建立 Bagged kernel, 具体的计算公式为:

$$K_{bag}(x_i, x_j) = \frac{[C(x_i) = C(x_j)]}{N}, \tag{6}$$

式中, $i = \{1, 2, \dots, m\}$, $j = \{1, 2, \dots, m+n\}$. 如果 x_i 和 x_j 在同一类中, 则 $[C(x_i) = C(x_j)]$ 返回值为 1, 否则返回值为 0.

- (5) 对于 K_{bag} , 求得一个 $m * m$ 的矩阵 K'_{bag} 和一个 $m * n$ 的矩阵 $K_{testbag}$.

- (6) K_{train} 和 K'_{bag} 进行求和, a 为权重因子. 得到训练核矩阵 $K(x_i, x_j)$:

$$K(x_i, x_j) = aK'_{bag}(x_i, x_j) + (1-a)K_{train}(x_i, x_j) \tag{7}$$

式中, K'_{bag} 是由第(5)步中得到的 K_{bag} 对应标记样本构成的矩阵, x_i 和 x_j 为标记样本 $i = \{1, 2, \dots, m\}$, $j = \{1, 2, \dots, m\}$.

- (7) 采用核矩阵 $K(x_i, x_j)$ 训练 KELM 模型.
- (8) 采用第(6)步得到的极速学习机和第(7)步得到的测试核矩阵, 完成对测试样本的分类.

3 实验结果与分析

3.1 实验数据集

为了验证提出算法的有效性, 本文采用 UCI 数据集 (Abalone、Hill_valley、Musk(version1)). 在实验中, 我们对原始数据集中的数据做了处理, 随机选取小部分的样本作为训练样本, 剩余的大量样本则作为测试样本, 训练样本与测试样本的比例为 1:2. 每类的标签分别设定为 5、10. 本文实验数据集的详细信息见表 1.

表 1 UCI 数据集的信息

Table 1 The information of UCI data set

数据集		样本数目		属性信息		类别
		训练样本	测试样本	属性数目	属性特征	
Abalone	Abalone_5	15	4 162	8	Categorical Integer Real	3
	Abalone_10	30	4 147			
Hill_valley	hv_with_5	10	1 202	100	Real	2
	hv_with_10	20	1 192			
	hv_without_5	10	1 202			
	hv_without_10	20	1 192			
Musk(version1)	Clean1_5	10	466	166	Integer Real	2
	Clean1_10	20	456			

3.2 实验结果

实验中需要确定的控制参数有 Bagged 核参数 j , 核矩阵参数 i . 表 2 展示了基于 Bagged 聚类核的 ELM 和 RBF 核极速学习机的实验结果, 本实验的执行环境是 MATLAB 2012a, 鉴于 ELM 的性能不太稳定, 在表 2 中给出的是运行 10 次的平均性能.

从实验结果可以看出基于 Bagged 聚类核的核极速学习机算法的性能相比于传统的核 ELM 性能得到进一步提升. 传统的 RBF 核可能存在过于平滑的问题, 导致数据分布中有些相近的模式(mode)难以区分. 而我们提出的算法综合了传统的 RBF 核和聚类核, 希望使用 RBF 核建模平稳分布, 使用 Bagged 聚类核来实现非平稳分布的特征映射, 从而得到特征空间下的平稳分布, 以便更好地完成分类任务. 此外我们还可以从聚类核权重系数 a 与测试数据集精度的二维图中更直观地感受 RBF 核与聚类核比例的不同所带来的精度变化. 测试精度随聚类核权重系数 a 的变化曲线可以表示为图 2~5.

表 2 在 UCI 数据集上分类问题的性能比较
Table 2 Performance comparison on data sets

数据集	算法	训练样本		测试样本	
		训练时间/s	训练精度	测试时间/s	测试精度
Abalone_5	基于 Bagged 聚类核的核极速学习机	3.400 6	0.666 7	0.010 4	0.541 6
	RBF 核极速学习机	$3.446 7 \times 10^{-4}$	0.800 0	0.004 9	0.541 3
Abalone_10	基于 Bagged 聚类核的核极速学习机	3.384 0	0.633 3	0.011 2	0.554 9
	RBF 核极速学习机	$5.435 5 \times 10^{-4}$	0.700 0	0.008 8	0.551 2
hv_with_5	基于 Bagged 聚类核的核极速学习机	1.202 4	0.700 0	0.004 5	0.520 8
	RBF 核极速学习机	$3.484 3 \times 10^{-4}$	0.700 0	0.004 1	0.518 3
hv_with_10	基于 Bagged 聚类核的核极速学习机	1.246 2	0.550 0	0.006 1	0.521 0
	RBF 核极速学习机	$3.837 4 \times 10^{-4}$	0.550 0	0.004 9	0.519 3
Clean1_5	基于 Bagged 聚类核的核极速学习机	0.494 4	1	0.003 3	0.585 8
	RBF 核极速学习机	$3.740 3 \times 10^{-4}$	1	0.003 4	0.575 1
Clean1_10	基于 Bagged 聚类核的核极速学习机	0.525 0	1	0.003 2	0.721 5
	RBF 核极速学习机	$4.785 5 \times 10^{-4}$	1	0.002 5	0.688 6
hv_without_5	基于 Bagged 聚类核的核极速学习机	1.248 8	0.900 0	0.010 9	0.537 4
	RBF 核极速学习机	$2.768 2 \times 10^{-4}$	1	0.004 1	0.533 3
hv_without_10	基于 Bagged 聚类核的核极速学习机	1.209 6	0.800 0	0.005 6	0.510 9
	RBF 核极速学习机	$3.682 3 \times 10^{-4}$	0.850 0	0.005 1	0.505 9

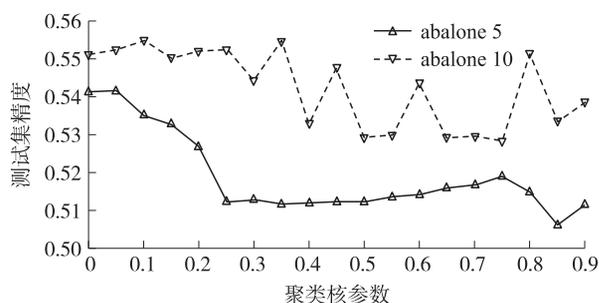


图 2 数据集 abalone 实验结果

Fig. 2 The experimental results of abalone dataset

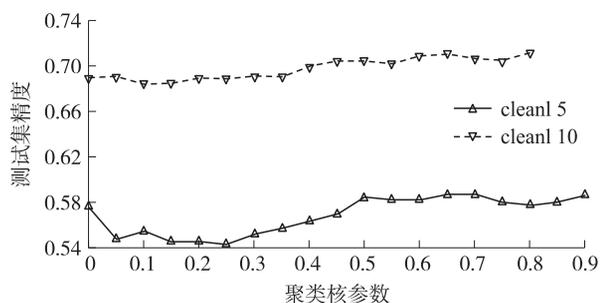


图 3 数据集 clean1 实验结果

Fig. 3 The experimental results of Clean1 dataset

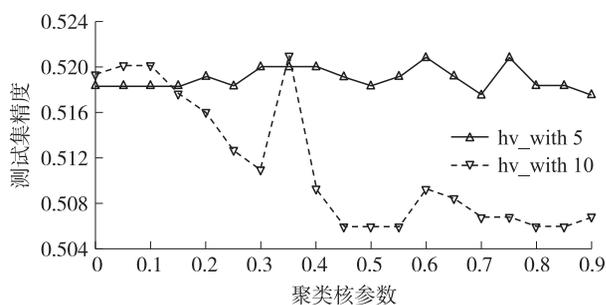


图 4 数据集 hv_with 实验结果

Fig. 4 The experimental results of hv_with dataset

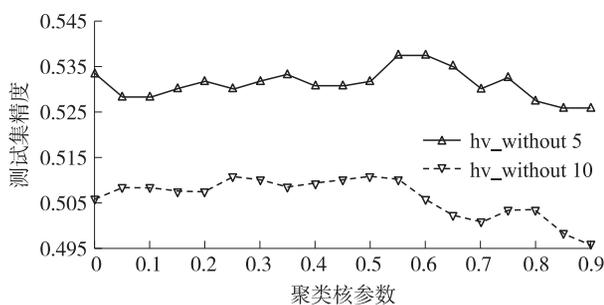


图 5 数据集 hv_without 实验结果

Fig. 5 The experimental results of hv_without dataset

当聚类核系数 a 为 0 时,表示无聚类核的 KELM 的结果,反之,当 a 为 1 时,表示由聚类核构成的 KELM 的实验结果. 从实验结果总体来看,随着聚类核参数的不断选择,各个数据集中都出现多个高于核极速学习机结果的点,这说明恰当地选择聚类核参数能够提升传统的 RBF 核 ELM 算法的性能. 通过观察,图 2、图 3 中每一类别中标记数目为 10 的数据集实验结果要优于标记数目为 5 的数据集;而图 4、图 5 中标记个数少的数据集的测试实验结果要优于标记数目多的数据集. 究其原因,这是由于各个数据集有着不同的结构特点,因此算法在不同的数据集上表现的性能是不一样的.

4 结论

本文基于半监督思想,在传统极速学习机的理论上,提出一种基于 Bagged 聚类核的核极速学习机算法. 从实验结果上看,本文提出的算法相比于传统的 KELM 算法,在分类精度上有一定的提高. 聚类

核虽也是一种半监督核,但本文中提出的算法,还不是完全意义上的半监督极速学习机算法.本文目前只是综合了部分监督学习的特征和聚类特征,如何更好地将分类特征融合到ELM的核函数中仍是接下来要研究的问题.此外,由于ELM本身的不稳定性,会给算法带来一定的精度波动,如何利用核函数增加算法的稳定性,也是接下来研究的重点.

[参考文献]

- [1] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine; a new learning scheme of feed forward neural networks[C]//International Joint Conference on Neural Networks. Budapest, 2004: 985-990.
- [2] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70: 489-501.
- [3] VAPNIK V N. The nature of statistical learning theory[M]. New York: Springer, 1995.
- [4] 孔怡青. 半监督学习及其应用研究[D]. 无锡: 江南大学, 2009.
- [5] 李小冬. 核极速学习机的理论与算法及其在图像处理中的应用[D]. 杭州: 浙江大学, 2014.
- [6] 邓万宇, 郑庆华, 陈琳, 等. 神经网络极速学习方法研究[J]. 计算机学报, 2010, 33(2): 279-287.
- [7] ZHU Q Y, QIN A K, SUGANTHAN P N, et al. Evolutionary extreme learning machine[J]. Pattern recognition, 2005, 38(10): 1759-1763.
- [8] HUANG G B, ZHOU H M, DING X J, et al. Extreme learning machine for regression and multiclass classification[J]. IEEE transactions on systems, man, and cybernetics, part B: cybernetics, 2012, 42(2): 513-529.
- [9] HUANG G B, SIEW C K. Extreme learning machine with randomly assigned RBF kernels[J]. International journal of information technology, 2005, 11(1): 16-24.
- [10] 丁世飞. 孪生支持向量机: 算法、理论与拓展[M]. 北京: 科学出版社, 2017.
- [11] LÜ F, HAN M. Hyperspectral image classification based on multiple reduced kernel extreme learning machine[J]. International journal of machine learning and cybernetics, 2019, <https://doi.org/10.1007/s13042-019-00926-5>.
- [12] ZHANG J, DING S F, ZHANG N, et al. An incremental extreme learning machine based on deep feature embedded[J]. International journal of machine learning and cybernetics, 2016, 7(1): 111-120.
- [13] 梁吉业, 高嘉伟, 常瑜. 半监督学习研究进展[J]. 山西大学学报(自然科学版), 2009, 32(4): 528-534.
- [14] DING S F, ZHANG J, XU X Z, et al. A wavelet extreme learning machine[J]. Neural computing and applications, 2016, 27(4): 1033-1040.
- [15] WANG H B, LIU X, SONG P, et al. Sensitive time series prediction using extreme learning machine[J]. International journal of machine learning and cybernetics, 2019. (<https://doi.org/10.1007/s13042-019-00924-7>)
- [16] WESTON J, LESLIE C, ZHOU D, et al. Semi-supervised protein classification using cluster kernels[C]//Advances in Neural Information Processing System 17(NIPS 2004). Vancouver, 2004: 595-602.
- [17] DING S F, ZHANG Y N, XU X Z. A novel extreme learning machine based on hybrid kernel function[J]. Journal of computers, 2013, 8(8): 2110-2117.
- [18] HUANG G, HUANG G B, SONG S G, et al. Trends in extreme learning machines: a review[J]. Neural networks, 2015, 61: 32-48.
- [19] HUANG G, SONG S G, JATINDER N D. Semi-supervised and unsupervised extreme learning machines[J]. IEEE transactions on cybernetics, 2014, 44(12): 2405-2417.
- [20] CHAPELLE O, WESTON J, SCHOLKOPF B. Cluster kernels for semi-supervised learning[C]//Advances in Neural Information Processing System 16(NIPS 2003). Vancouver, 2003: 601-608.

[责任编辑: 顾晓天]