

基于探测粒子群的小波核极限学习机算法

陈晓青¹ 陆慧娟¹ 关 伟² 郑文斌¹

(中国计量学院信息工程学院 杭州 310018)¹ (中国计量学院现代科技学院 杭州 310018)²

摘 要 在分析核极限学习机原理的基础上,将小波函数作为核函数运用于极限学习机中,形成小波核极限学习机(WKELM)。实验表明,该算法提高了分类性能,增加了鲁棒性。在此基础上利用探测粒子群(Detecting Particle Swarm Optimization,DPSO)对WKELM参数优化,最终得到分类效果较优的DPSO-WKELM分类器。通过采用UCI基因数据进行仿真,将该分类结果与径向基核极限学习机(KELM)、WKELM等算法结果进行比较,得出所提算法具有较高的分类精度。

关键词 核极限学习机,探测粒子群,算法优化,分类精度

中图法分类号 TP181 文献标识码 A

Wavelet Kernel Extreme Learning Machine Algorithm Based on Detecting Particle Swarm Optimization

CHEN Xiao-qing¹ LU Hui-juan¹ GUAN Wei² ZHENG Wen-bin¹

(College of Information Engineering, China Jiliang University, Hangzhou 310018, China)¹

(College of Modern Science and Technology, China Jiliang University, Hangzhou 310018, China)²

Abstract In this paper, the principle of the kernel extreme machine was studied. Wavelet function was chosen to be the extreme learning machine's kernel function. Experiments show that this algorithm improves the classification accuracy and increases the robustness. Based on this method, we used detecting particle swarm optimization(DPSO) to optimize and set the initial parameters of WKELM in order to obtain the optimal WKELM classifier DPSO-WKELM. We used UCI gene data for simulation. The classification results are compared with the results of radial basis kernel extreme learning machine (KELM) and WKELM. The comparison shows that the proposed algorithm has higher classification accuracy.

Keywords KELM, DPSO, Algorithm optimization, Classification accuracy

1 引言

虽然基于梯度下降的神经网络广泛地应用于多层前馈神经网络的训练中,但是由于该方法存在学习速度慢、容易陷入局部最优解和在不同场合下其参数难以调解等复杂问题^[1],Huang在2006年提出了极限学习机(Extreme Learning Machine,ELM),这一算法成为机器学习领域的重要成果,ELM学习速度快,泛化性能好。但是,在ELM训练过程中,算法本身每次随机产生的输入层权值和隐层节点偏置并不能保证模型有稳定最优的正确率,一般需要人工进行多次调整,这就涉及人为经验化的操作,对分类精度会产生较大影响^[2]。2008年,G. B. Huang等^[3]提出了基于随机搜索的增量型极限学习机EI-ELM,减少了增量型极限学习机的迭代次数,并克服了输出过拟合的问题。2012年,邓万宇等^[4]提出一种正则极限学习机(Regularized Extreme Learning Machine,RELIM)的理论,该方法是将结构风险最小化理论和加权最小二乘法引入到ELM中,同年Huang等^[5]通过对比ELM与

向量机的建模和求解过程,提出了核极限学习机(Kernel Extreme Learning Machine,KELM)。2013年,王等^[10]提出了小波核极限学习机(WKELM),并通过证明得出该分类器具有较高的分类性能。

探测粒子群(DPSO)是对PSO的一种改进算法,该算法选定少数粒子,让其不同于普通粒子的搜索路径,以螺旋折线路径进行搜索,这些粒子称为探测粒子。DPSO算法是利用探测粒子与普通粒子联合进行更高效率的搜索,这种方法可以有效地避免PSO算法的早熟收敛,提高算法的收敛速度和收敛精度,其缺点是增加了一定的额外计算开销。本文通过分析WKELM存在的缺点,提出一种新的DPSO-WKELM分类器,在该算法中,首先通过WKELM对给定数据进行初始化,产生一组输入权值和隐层偏置,选择其中部分作为探测粒子进行路径搜索,将寻得的较优路径给剩下的普通粒子作为起始位置进行寻优,最后将得到的最优输入权值和隐层偏置带入WKELM中训练。

本文使用UCI数据集上的Breast、Brain和Colon 3组数

本文受国家自然科学基金资助项目(61272315,60842009),浙江省自然科学基金(Y1110342,Y1080950)资助。

陈晓青(1990—),女,硕士生,CCF会员,主要研究方向为机器学习、数据挖掘,E-mail:769668323@qq.com;陆慧娟 女,博士,教授,CCF常务理事,主要研究方向为机器学习、模式识别和生物信息学等,E-mail:hjlu@cjl.edu.cn(通信作者);关 伟 男,讲师,主要研究方向为机器学习、生物信息学;郑文斌 男,博士,副教授,主要研究方向为模式识别。

据的分类实验验证该算法的分类性能、泛化性能以及有效性,最后给出结论。

2 极限学习机

ELM 在提出后其学习速度快、泛化能力强的特点得到了各种理论与应用的验证,同时,ELM 避免了梯度下降法带来的许多问题,ELM 的输入偏置和输出权值是一个随机产生的线性系统,并且在训练过程中只需求解输出权值最小化而不用调整。

假设矩阵 (x_i, t_i) 含有 N 个不同样本,其中, $x_i = [x_{i1}, x_{i2}, \dots, x_{iN}]^T \in R^N$, $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$, $i=1, 2, \dots, N$, 含有 L 个单隐层节点数目, $L \leq N$, $g(x)$ 为 Sigmoid 类型的激活函数,则 ELM 的数学模型公式如式(1)所示:

$$t_i = \sum_{j=1}^L \beta_j g_i(x_i) = \sum_{j=1}^L \beta_j g(w_j x_i + b_j) \quad (1)$$

其中, $j=1, 2, \dots, N$, $w_i = [w_1, w_2, \dots, w_N]$ 表示输入的第 i 个隐层神经元的权值向量, $\beta_i = [\beta_1, \beta_2, \dots, \beta_m]^T$ 表示第 i 个隐层节点到输出层的链接权值, $b_i = [b_1, b_2, \dots, b_m]^T$ 表示第 i 个隐层节点的偏置,式(1)表示的 N 个方程可用矩阵的形式表示如下:

$$H\beta = T \quad (2)$$

其中

$$H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_L \cdot x_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_L \cdot x_N + b_L) \end{bmatrix}_{N \times L}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, T = \begin{bmatrix} t_1^T \\ \vdots \\ t_m^T \end{bmatrix}_{N \times m} \quad (3)$$

H 是 ELM 的隐层输出矩阵,由式(3)可看出,该矩阵共有 N 行 L 列,随机产生的偏置和输入权值不需要调节,可由 Moore-Penrose 广义逆定理通过矩阵分解求值,目前应用最多的求解 H^* 方法是奇异值分解。

$$\beta = H^* T \quad (4)$$

3 基于小波核函数的极限学习机

3.1 核极限学习机

根据 BARTLET 理论,前馈网络中权值越小,其泛化性能越好。在 ELM 中,传统单隐层前馈神经网络训练问题已被转化为求解线性方程组问题,将求得的最小范数最小二乘解作为网络输出权值。因此,ELM 将训练误差最小化和输出权重范数作为训练目标:

$$\min: \|\beta\| \text{ and } \sum \|\beta \cdot h(x_i) - t_i\|^2 \quad (5)$$

其中, β 是连接隐含节点的权重向量, $h(x_i)$ 为隐层核映射。

从标准优化理论的观点看,上述的优化问题可采用简化的约束优化问题求解,则上述目标可重新改写为:

$$\min: L_P = \frac{1}{2} \|\beta\|^2 + \frac{1}{2} C \sum_{i=1}^N \xi_i^2 \quad (6)$$

$$\text{s. t. } h(x) \beta = t_i - \xi_i, i=1, 2, \dots, N$$

其中, ξ_i 为训练误差, C 为惩罚参数。

根据 KKT 条件,可以定义 Lagrange 函数求解以上问题,即可以等效为求解如式(7):

$$L_{PELM} = \frac{1}{2} \|\beta\|^2 + \frac{1}{2} C \sum_{i=1}^N \xi_i^2 - \sum_{i=1}^N \alpha_i (h(x_i) \beta - t_i + \xi_i) \quad (7)$$

其中,每个拉格朗日算子 α_i 均对应于第 i 样本,可得到如下的 KKT 优化条件:

$$\frac{\partial L_{PELM}}{\partial \beta} = 0 \Rightarrow \beta = \sum_{i=1}^N \alpha_i h(x_i)^T = H^T \alpha$$

$$\frac{\partial L_{PELM}}{\partial \xi_i} = 0 \Rightarrow \alpha_i \xi_i = 0, i=1, 2, \dots, N \quad (8)$$

$$\frac{\partial L_{PELM}}{\partial \alpha_i} = 0 \Rightarrow h(x_i) \beta - t_i + \xi_i = 0, i=1, 2, \dots, N$$

其中 $\alpha = [\alpha_1, \dots, \alpha_N]^T$ 。 H 是隐层输出矩阵,只与样本的个数和隐层节点数有关,与输出节点个数无关,分类中与样本的类别数目也无关。

将式(8)整理得:

$$h(x_i) H^T C \xi_i - t_i^T + \xi_i^T = 0$$

$$\vdots$$

$$h(x_N) H^T C \xi_N - t_N^T + \xi_N^T = 0 \quad (9)$$

令

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \cdots & \vdots \\ t_{N1} & \cdots & t_{Nm} \end{bmatrix} \quad (10)$$

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} \quad (11)$$

再将式(9) - 式(11)整理合并得:

$$\left(\frac{1}{C} + HH^T\right) \alpha = T \quad (12)$$

最终推导得到:

$$\beta = H^T (HH^T + \frac{1}{C})^{-1} T \quad (13)$$

于是 KELM 的逼近函数可以写成:

$$f(x) = h(x) H^T (HH^T + \frac{1}{C})^{-1} T \quad (14)$$

应用 Mercer's 条件定义 KELM 的核矩阵为:

$$\Omega_N = H_N H_N^T: \Omega_{N_{ij}} = h(x_i) h(x_j) = K(x_i, x_j) \quad (15)$$

其中, $i, j \in (1, 2, \dots, N)$, $K(x_i, x_j)$ 为核函数,是核矩阵 Ω_N 位于第 i 行、第 j 列的元素。

于是 KELM 的求解公式为:

$$f(x_p) = \begin{bmatrix} K(x_p, x_1) \\ \vdots \\ K(x_p, x_N) \end{bmatrix}^T \alpha_N \quad (16)$$

$$\alpha_N = \left(\frac{I_N}{C} + \Omega_N\right)^{-1} T_N \quad (17)$$

其中,式(17)中 α_N 为 KELM 网络的输出权值。

3.2 小波核函数

同支持向量机一样,若一个函数能够满足 Mercer 定理,则该函数便可作为极限学习机的核函数,常见的核函数有 RBF 核函数、线性核函数、多项式核函数等,本文将小波函数的概念引入核函数,构造小波核函数。

已知一个母小波函数 $h(x)$,其小波基函数可以表示如下:

$$h(x) = \sqrt{a} \Phi\left(\frac{x - c_1}{a}\right) \quad (18)$$

其中, a 为伸缩因子, c_1 为平移因子。根据张量积的理论,一

个多维的小波函数可以写成多个一维小波函数的张量积[7]。

$$h(x) = \prod_{i=1}^n h(x_i) \quad (19)$$

根据式(19)可以构造平移不变的核函数,如式(20):

$$K(x, x') = K(Kx - x') = \prod_{i=1}^n \Phi\left(\frac{x_i - x'_i}{a}\right) \quad (20)$$

已知 $x, x' \in R$, 本文选用 morlet 小波函数 $\Phi(x) = \cos(1.75x) \exp(-\frac{x^2}{2})$ 构造相应的小波核函数,核函数形式如下:

$$\text{Wavelet kernel}(x, x') = \prod_{i=1}^n [\cos(1.75 \times (\frac{x_i - x'_i}{a})) \exp(-(\frac{x_i - x'_i}{2})^2)] \quad (21)$$

对于小波核是一种被允许的极限学习机核,文献[10]已经给出证明,本文在此不予证明。

4 探测粒子群算法

1995 年 James Kennedy 和 Russel Eberhart 提出 PSO 算法,该方法是基于群体演化的随机全局优化的一种智能优化算法,其中心思想是对鸟群或鱼群合作捕食行为的研究[8,9]。PSO 中每个个体称为一个“粒子”,每个粒子代表一个问题的潜在解。假设在一个 d 维搜索空间中,有 m 个粒子构成的种群,粒子 $i(i=1, 2, \dots, m)$ 在看 k 次迭代时的状态属性可由两个向量描述:位置属性 $x_i^k = [x_{i1}^k, x_{i2}^k, \dots, x_{id}^k]$ 和速度向量 $v_i^k = [v_{i1}^k, v_{i2}^k, \dots, v_{id}^k]$,速度限制在区间 $[-v_{\max,d}, v_{\max,d}]$ 内,为确定到 k 次迭代时粒子 i 的个体最优位置 $pbest_{id}$,以及粒子 i 在种群搜索到的全局最优位置 $gbest_{id}$,需要对粒子的适应度大小进行评价,所以第 i 个粒子的速度和位置的更新公式如下:

$$v_{id}^{k+1} = v_{id}^k + c_1 r_1 (pbest_{id}^k - x_{id}^k) + c_2 r_2 (gbest_{id}^k - x_{id}^k) \quad (22)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (23)$$

其中, c_1, c_2 是学习因子(又称加速系数),是粒子具有自我总结和向群体中优秀个体学习的能力; r_1, r_2 是区间 $[0, 1]$ 内均匀分布的随机数。

针对 PSO 在处理多峰函数问题时容易出现早熟收敛的情况,一种新的粒子群优化算法探测粒子群算法(DPSO)被提出,该方法的主要思想来源于文献[7],该算法的思想源于两点:(1)蜜蜂觅食时会派出个别蜜蜂做探测;(2)在鸟群寻找目标时,它的飞行路径不仅做折线还做盘旋。本文算法主要从 m 个粒子群中选择 m_s 个粒子(一般取 $m_s=1$)作为探测粒子,让其进行螺旋折线收敛搜索以试图发现更多更好的潜在解。

在 DPSO 算法中,其实现过程如下:

- 1) 按式(22)、式(23)更新所有粒子的速度和位置;
- 2) 选取种群中第 j 个粒子($m_s=1$)作为探测粒子,并将粒子 j 的当前位置 x_{jd}^k 作为探测粒子的初始位置 $x_{sd}^{k'}(k'=0)$;
- 3) 选取目标点位置;
- 4) 反复迭代计算探测粒子的位置 $x_{sd}^{k'}$ 和速度 $v_{sd}^{k'}$,将探测粒子的最优位置作为下一代普通粒子的起始位置用标准 PSO 更新公式继续迭代。

在构造探测粒子的速度更新公式中,抽取标准 PSO 速度更新式(22)中的一部分,并加以改造给出了探测粒子的速度更新公式如下:

$$v_{sd}^{k+1} = c_{s1} v_{sd}^k + c_{s2} (Target - x_{sd}^k) \quad (24)$$

由上式可见,探测粒子的速度更新公式主要是从标准 PSO 的速度更新公式中提炼而来,在式(22)中,删除其中间项,在第一项 v_{sd}^k 前加系数 c_{s1} ,同时将第三项 $c_2 r_2 (gbest_{id}^k - x_{id}^k)$ 前的系数 $c_2 r_2$ 替换为 c_{s2} ,再将 $(gbest_{id}^k - x_{id}^k)$ 中的 $gbest_{id}^k$ 替换为 $Target$ 。 c_{s1}, c_{s2} 可取为 $[0, 1]$ 之间的值,本文都取为 0.8, $Target$ 取可行域内值,并配上标准 PSO 位置更新式(23)就可以得到一个以 $Target$ 为中心的收敛的搜索路径。

5 算法描述及实验结果

本文提出 DPSO-WKELM 算法是为了得到分类精度更高、泛化能力更好的模型。DPSO-WKELM 算法结合了 DPSO 算法的高稳定性和 WKELM 的分类精度高、泛化性能好的特点,该算法的流程描述如下:

- 1) 给定训练与测试数据集,在训练前,对数据进行归一化处理;
- 2) 建立基于 DPSO 的小波核极限学习机的神经网络拓扑结构,设置隐层神经元数目,选择激活函数。
- 3) 产生种群,设定粒子数 m ,产生 m 个在 $[-1, 1]$ 范围内的随机数向量作为粒子的个体,每个个体有 $D = N(n+1)$ 个元素, N 为隐层节点数, n 为输入层神经元个数。
- 4) 初始化 DPSO 的速度、惯性权重、加速度因子以及最大迭代次数等。
- 5) 计算每个粒子适应度值。根据式(1)~式(3)计算得到数据集的实际输出,并求出期望输出与实际输出的均方根误差,即得到每个粒子的适应度,找到每个粒子的个体极值和种群的群体极值。
- 6) 根据式(21)~式(23)更新 DPSO 的速度及位置。
- 7) 判断是否达到最大迭代次数或者最小误差,若达到,则停止迭代,否则,转到步骤 5),继续迭代。

为评估该算法的性能,本文对算法进行了实验分析及仿真,从 UCI 标准数据集中选择肿瘤数据中的 Breast、Brain、Colon 等基因数据进行实验训练与测试,其中每个数据集如表 1 所列。

表 1 数据集信息表

数据集	样本总数	特征维数	类别数量
Breast	768	8	2
Brain	105	8	5
Colon	84	70	1

本实验是在 Core i7-4790 CPU 3.6GHz, 16GB 内存的硬件环境与操作系统为 Windows7 环境下,通过 Matlab 编程实现仿真。

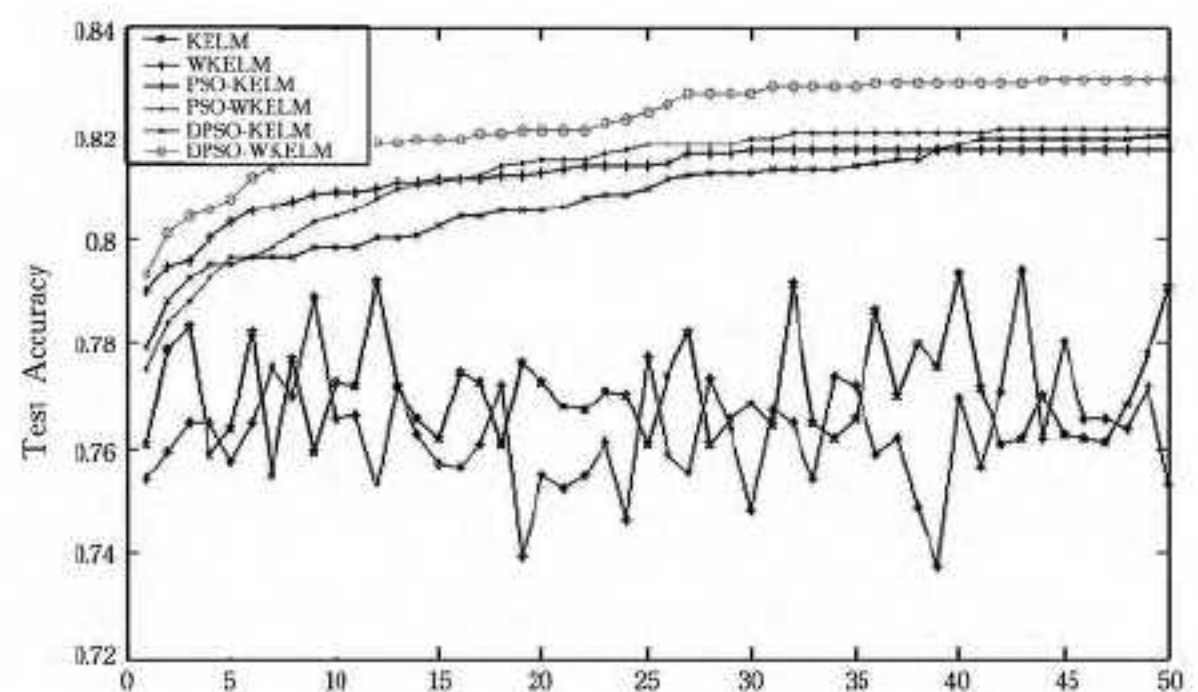


图 1 Breast 数据集上几种算法的测试精度对比

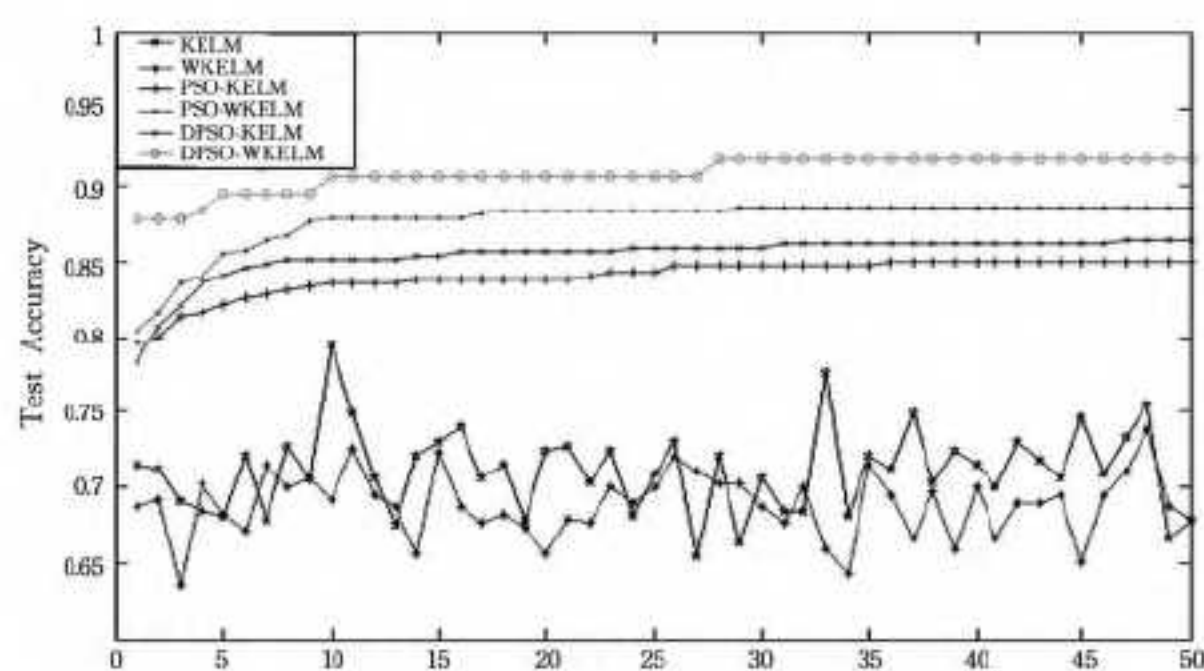


图2 Brain数据集上几种算法的测试精度对比

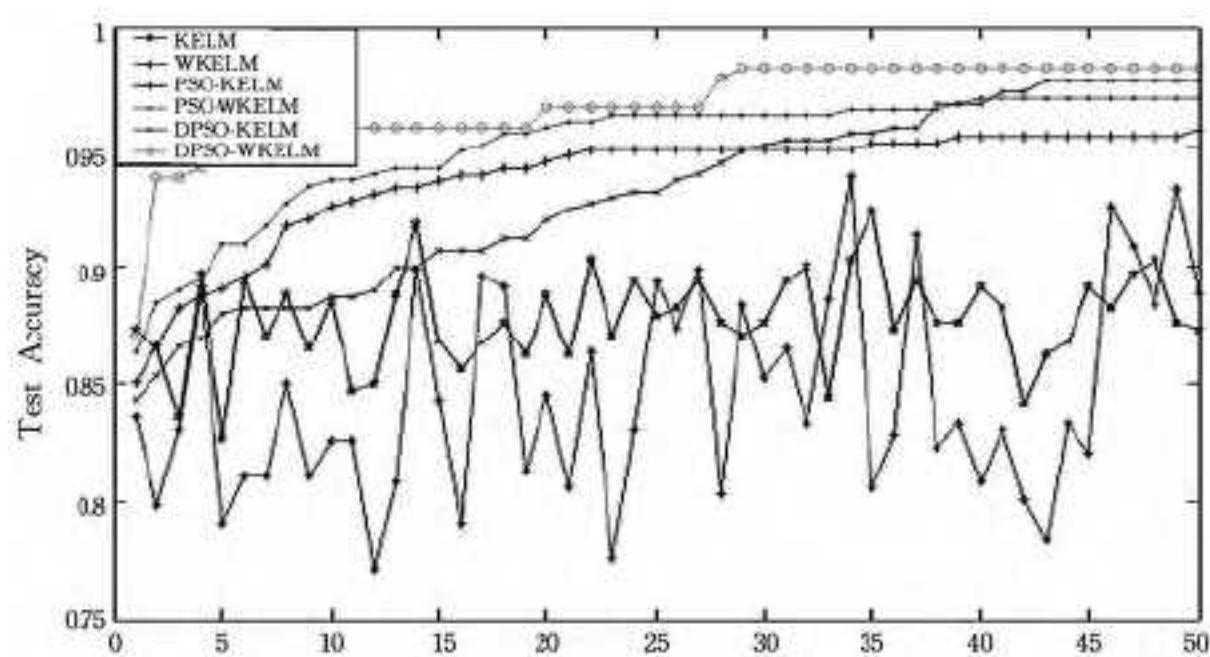


图3 Colon数据集上几种算法的测试精度对比

表2 DPSO-WKELM与其他算法的均方差比较

数据集	KELM	WKELM	PSO-KELM	PSO-WKELM	DPSO-KELM	DPSO-WKELM
Breast	0.0014	0.0015	0.0016	0.0013	0.0012	0.0011
Brain	0.0039	0.0030	0.0018	0.0025	0.0021	0.0016
Colon	0.0055	0.0065	0.0036	0.0037	0.0033	0.0028

结束语 本文用小波核的概念取代传统核函数中的径向基核函数,从实验结果图对比可以得出分类精度得到了显著的提高。同时将WKELM算法与改进的PSO算法进行结合,形成DPSO-WKELM算法。用DPSO对WKELM算法进行参数优化,实现了比传统ELM算法更好的分类效果和泛化性能。

由于DPSO-WKELM算法是分别在PSO和ELM中引入探测粒子和小波核函数,因此其计算复杂性增加,但是与其他算法比较发现DPSO-WKELM算法均方差最小,稳定性最好,且分类精度最高,说明DPSO-WKELM算法是一种可靠有效的分类算法。

参考文献

- [1] He Zhi-bin, Wen Xiao-hu, Liu Hu, et al. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region[J]. Journal of Hydrology, 2014, 509(13):379-386
- [2] 陆慧娟, 安春霖, 马小平, 等. 基于输出不一致测度的极限学习机集成的基因表达数据分类[J]. 计算机学报, 2013, 36(2):341-348
- [3] Huang G. B, Chen L, Siew C K. Enhanced random search based incremental extreme learning machine [J]. Neurocomputing, 2008, 71(18):3460-3468
- [4] Deng W Y, Chen L. Regularized extreme learning machine in Data mining[J]. Proc IEEE Symp Comput Intell, 2009, 10(11): 385-389

为了验证小波核函数在粒子群算以及带活跃算子的粒子群算法中的性能,用KELM、WKELM、PSO-KELM、PSO-WKELM、DPSO-KELM、DPSO-WKELM在Breast、Brain、Colon这3个数据集上分别采用10次5折交叉验证,将数据集分成5份,轮流将其中4份作为训练数据集,1份作为测试数据集进行实验,取10次结果精度的平均值作为算法的精度,对应的精度图分别如图1—图3所示。

从图1—图3可以看出,对比PSO-KELM和PSO-WKELM这两个算法的结果,加入基核函数的极限学习机的分类精度高,同样DPSO-KELM和DPSO-WKELM算法比较,DPSO-WKELM算法的分类精度也有明显的提高,相较于DPSO-KELM,分类精度在数据集Breast和Brain上提高较明显。

本文将DPSO-WKELM算法与其他算法的均方差进行比较,如表2所列。

由以上6种算法的均方差对比可知,KELM算法与WKELM算法最不稳定,随着迭代次数的增加一直上下波动,且不易收敛,均方差最高,经过PSO或DPSO进行参数优化后,均方差明显下降,稳定性提高,随着迭代次数的增加逐渐进入收敛,说明参数优化的必要性,但是PSO易进入早熟收敛,DPSO算法效果更好。

- [5] Huang Guang-bin, Zhou Hong-ming, Ding Xiao-jian, 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
- [6] Han Fei, Yao Hai-fen, Ling Qing-hua. An improved Extreme Learning Machine Based on Particle Swarm Optimization[J]. Bio-Inspired Computing and Applications Lecture Notes in Computer Science, 2013, 16(20):87-93
- [7] Zhang Ying-nan, Teng Hong-fei. Detecting particle swarm optimization. [J]. Concurrency & Computation Practice & Experience, 2009, 21(4):449-473
- [8] Shi Y, Eberhart R C. A modified swarm optimizer [C]// IEEE International Conference of Evolutionary Computation. Anchorage, Alaska: IEEE Press, May, 1998
- [9] Kennedy J, Eberhart R. Particle swarm optimization [C]// Proceeding of IEEE International Conference on Neural Networks. Piscataway, NJ: IEEE Press, 1995:1942-1948
- [10] 王杰, 郭晨龙. 小波核极限学习机分类器[J]. 微电子学与计算机, 2013, 30(10):73-80
- [11] 周有, 侯铁双. 基于小波核函数支持向量算法的信号检测[J]. 计算机仿真, 2013, 30(1):263-267
- [12] Zhang Li, Zhou Wei-da, Jiao Li-cheng. Wavelet Support Vector Machine[J]. IEEE Transactions on Cybernetics, 2004, 34(1): 33-39
- [13] 陈忠杰, 蔡勇, 蒋刚. 复高斯小波核函数的支持向量机研究[J]. 计算机应用研究, 2012, 29(9):3263-3265