

K插值单纯形法核极限学习机的研究

苏一丹 李若愚 覃华* 陈琴

(广西大学计算机与电子信息学院 南宁 530004)

摘要: 针对核极限学习机高斯核函数参数选优难, 影响学习机训练收敛速度和分类精度的问题, 该文提出一种K插值单纯形法的核极限学习机算法。把核极限学习机的训练看作一个无约束优化问题, 在训练迭代过程中, 用Nelder-Mead单纯形法搜索高斯核函数的最优核参数, 提高所提算法的分类精度。引入K插值为Nelder-Mead单纯形法提供合适的初值, 减少单纯形法的迭代次数, 提高了新算法的训练收敛效率。通过在UCI数据集上的仿真实验并与其它算法比较, 新算法具有更快的收敛速度和更高的分类精度。

关键词: 核极限学习机; 核参数; Nelder-Mead单纯形法; K插值法

中图分类号: TP181

文献标识码: A

文章编号: 1009-5896(2018)08-1860-07

DOI: 10.11999/JEIT171104

Kernel Extreme Learning Machine Based on K Interpolation Simplex Method

SU Yidan LI Ruoyu QIN Hua CHEN Qin

(College of Computer and Electronic Information, Guangxi University, Nanning 530004, China)

Abstract: The kernel Extreme Learning Machine (ELM) has a problem that the kernel parameter of the Gauss kernel function is hard to be optimized. As a result, training speed and classification accuracy of kernel ELM are negatively affected. To deal with that problem, a novel kernel ELM based on K interpolation simplex method is proposed. The training process of kernel ELM is considered as an unconstrained optimal problem. Then, the Nelder-Mead Simplex Method (NMSM) is used as an optimal method to search the optimized kernel parameter, which improves the classification accuracy of kernel ELM. Furthermore, the K interpolation method is used to provide appropriate initial values for the Nelder-Mead simplex to reduce the number of iterations, and as a result, the training speed of ELM is improved. Comparative results on UCI dataset demonstrate that the novel ELM algorithm has better training speed and higher classification accuracy.

Key words: Kernel Extreme Learning Machine (KELM); Kernel parameter; Nelder-Mead Simplex Method (NMSM); K Interpolation method

1 引言

极限学习机^[1](Extreme Learning Machine, ELM)是近年出现的神经网络分类算法, 已广泛应用于天文学、生物学、机械工程等领域^[2-5], 其训练过程关键是解一线性方程组, 获得分类参数。传统ELM用神经网络构造线性方程组的系数矩阵, 所构造的系数矩阵存在病态问题, 导致极限学习机不稳定、分类精度下降。针对此问题, Huang等人^[6]提出了核极限学习机(Kernel Extreme Learning

Machine, KELM), 用支持向量机的核函数构造满秩的系数矩阵^[7-10]。KELM使用最广泛的核函数是高斯核函数, 它的核参数与数据集有关^[11-13], 核参数的取值不当, 会直接影响KELM的训练收敛和分类性能, 如何为数据集确定一个合适的核参数, 目前还没有成熟的理论作指导, 值得深入研究。Nelder-Mead单纯形法(Nelder-Mead Simplex Method, NMSM)是一种解无约束优化问题的方法^[14], 文献^[15,16]应用NMSM搜索参数的最优值取得了较好的效果。借鉴此思想, 本文提出一种K插值单纯形法的核极限学习机(K -interpolation Simplex KELM, KS-KELM), 把KELM的训练看作一个无约束优化问题, 核参数是待优化变量, 在KELM的训练迭代过程中, 通过Nelder-Mead单纯形法的反射、扩展、压缩、收缩等操作搜索核参数的最优

收稿日期: 2017-11-24; 改回日期: 2018-04-23; 网络出版: 2018-06-07

*通信作者: 覃华 1694520545@qq.com

基金项目: 国家自然科学基金(61762009)

Foundation Item: The National Natural Science Foundation of China (61762009)

值, 使KELM的训练快速收敛; 引入K插值为NMSM提供合适的核参数初值, 解决NMSM对参数初值敏感的问题; 最后在测试数据集上验证本文算法的可行性和有效性。

2 核极限学习机

对于含有 L 个隐藏节点的ELM, 训练数据集

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \mathbf{x}_1 + b_1) & g(\mathbf{w}_2 \mathbf{x}_1 + b_2) & \cdots & g(\mathbf{w}_L \mathbf{x}_1 + b_L) \\ g(\mathbf{w}_1 \mathbf{x}_2 + b_1) & g(\mathbf{w}_2 \mathbf{x}_2 + b_2) & \cdots & g(\mathbf{w}_L \mathbf{x}_2 + b_L) \\ \vdots & \vdots & \ddots & \vdots \\ g(\mathbf{w}_1 \mathbf{x}_N + b_1) & g(\mathbf{w}_2 \mathbf{x}_N + b_2) & \cdots & g(\mathbf{w}_L \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L} \quad (2)$$

其中, $g(\cdot)$ 是隐藏节点的激活函数, $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{im}]^T \in R^n$ 为第 i 个隐藏节点的权值向量, b_i 表示第 i 个隐藏节点的偏置。

式(1)中, $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \dots, \boldsymbol{\beta}_L^T]^T_{L \times m}$, $\boldsymbol{\beta}_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T \in R^m$ 是第 i 个隐藏节点到 m 个输出节点之间的权重向量。

式(1)中, 右端项 $\mathbf{T} = [\mathbf{t}_1^T, \mathbf{t}_2^T, \dots, \mathbf{t}_N^T]^T_{N \times m}$, 其中 $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$ 是训练样本 \mathbf{x}_i 的类别列向量, 若训练样本 \mathbf{x}_i 属于第 q 类, 则 \mathbf{x}_i 的类别向量 \mathbf{t}_i 的第 q 个元素为1其余元素为0, 可表示为 $\mathbf{t}_i = [0, \dots, 0, \overset{q}{1}, 0, \dots, 0]^T$ 。

ELM的训练, 关键是解式(1)线性方程组, 得ELM的分类参数 $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = \mathbf{H}^{-1} \mathbf{T} \quad (3)$$

对于待分类的数据样本 \mathbf{x}_p , 其分类函数为

$$\mathbf{f}(\mathbf{x}_p) = \mathbf{h}(\mathbf{x}_p) \boldsymbol{\beta} \quad (4)$$

其中, $\mathbf{h}(\mathbf{x}_p) = [g(\mathbf{w}_1 \mathbf{x}_p + b_1) \ g(\mathbf{w}_2 \mathbf{x}_p + b_2) \ \cdots \ g(\mathbf{w}_L \mathbf{x}_p + b_L)]$ 。

在式(3)中, 由于用神经网络构造的系数矩阵 \mathbf{H} 存在病态问题, 文献[6]根据BARTLET理论和KKT条件, 对ELM训练模型进行优化, 优化后分类参数改写为

$$\boldsymbol{\beta} = \mathbf{H}^T \left(\frac{\mathbf{I}}{C} + \mathbf{H} \mathbf{H}^T \right)^{-1} \mathbf{T} \quad (5)$$

分类函数改写为

$$\mathbf{f}(\mathbf{x}_p) = \mathbf{h}(\mathbf{x}_p) \boldsymbol{\beta} = \mathbf{h}(\mathbf{x}_p) \mathbf{H}^T \left(\frac{\mathbf{I}}{C} + \mathbf{H} \mathbf{H}^T \right)^{-1} \mathbf{T} \quad (6)$$

其中, C 为惩罚参数, $\mathbf{h}(\mathbf{x}_p) \mathbf{H}^T$ 和 $\mathbf{H} \mathbf{H}^T$ 均为矩阵内积形式, 用满足Mercer's条件的核矩阵代替内积, 有

$$\boldsymbol{\Omega}_N = \mathbf{H} \mathbf{H}^T; \ \boldsymbol{\Omega}_{N_{ij}} = \mathbf{h}(\mathbf{x}_i) \mathbf{h}(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) \quad (7)$$

\mathbf{x} 有 N 个训练样本, m 个分类类别, 一个训练样本记为 \mathbf{x}_i , $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$, 其中 $i=1, \dots, N$ 。ELM在此数据集上的训练, 可用一个线性方程组表示为^[1]

$$\mathbf{H} \boldsymbol{\beta} = \mathbf{T} \quad (1)$$

其中, 系数矩阵 \mathbf{H} 为

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \mathbf{x}_1 + b_1) & g(\mathbf{w}_2 \mathbf{x}_1 + b_2) & \cdots & g(\mathbf{w}_L \mathbf{x}_1 + b_L) \\ g(\mathbf{w}_1 \mathbf{x}_2 + b_1) & g(\mathbf{w}_2 \mathbf{x}_2 + b_2) & \cdots & g(\mathbf{w}_L \mathbf{x}_2 + b_L) \\ \vdots & \vdots & \ddots & \vdots \\ g(\mathbf{w}_1 \mathbf{x}_N + b_1) & g(\mathbf{w}_2 \mathbf{x}_N + b_2) & \cdots & g(\mathbf{w}_L \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L} \quad (2)$$

其中, $K(\cdot)$ 是核函数。 $K(\mathbf{x}_i, \mathbf{x}_j)$ 是核矩阵 $\boldsymbol{\Omega}_N$ 中第 i 行、第 j 列的元素。由此, 核ELM分类参数写为

$$\tilde{\boldsymbol{\beta}} = \left(\frac{\mathbf{I}}{C} + \boldsymbol{\Omega}_N \right)^{-1} \mathbf{T} \quad (8)$$

分类函数写为

$$\begin{aligned} \tilde{\mathbf{f}}(\mathbf{x}_p) &= \mathbf{h}(\mathbf{x}_p) \mathbf{H}^T \tilde{\boldsymbol{\beta}} \\ &= \begin{bmatrix} K(\mathbf{x}_p, \mathbf{x}_1) \\ K(\mathbf{x}_p, \mathbf{x}_2) \\ \vdots \\ K(\mathbf{x}_p, \mathbf{x}_N) \end{bmatrix}^T \tilde{\boldsymbol{\beta}} \end{aligned} \quad (9)$$

核ELM的训练过程, 最关键处是用核函数构造核矩阵 $\boldsymbol{\Omega}_N$, 核ELM最常用的核函数是高斯核函数:

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\delta^2}} \quad (10)$$

其中, δ 是高斯核的核参数。高斯核参数 δ 取值不当, 会影响核ELM对训练样本特征的识别, 降低式(8)分类参数 $\tilde{\boldsymbol{\beta}}$ 和式(9)分类器 $\tilde{\mathbf{f}}(\cdot)$ 的准确性。因此, 要提高核ELM的分类性能, 必须对高斯核参数 δ 进行寻优, 为训练数据集确定一个合适的 δ 值, 本文使用文献[14]的Nelder-Mead单纯形法寻优核ELM的 δ 值。

3 用单纯形法优化高斯核参数

本文所用的1维单纯形子算法流程如表1所示。表1算法的一些关键点和细节说明如下。

3.1 步骤1构造初始化单纯形

通常直接使用 δ_0 构造顶点A, 即: $\delta_A = \delta_0$ 。构造顶点B的方法是: 如果 δ_0 无限小, 即 $\delta_0 < eps$, 则 $\delta_B = \delta_0 + 0.005$; 如果 δ_0 不是无限小, 则按比例扩展, 即 $\delta_B = \tau \delta_0$, τ 为比例系数, 一般取0.75。

3.2 步骤3至步骤6中计算公式与参数说明

(1) 步骤3中反射点的核参数值为 $\delta_R = (1 + \mu)\delta_{\text{best}} - \mu\delta_{\text{worst}}$, 其中 μ 称为反射系数($\mu > 0$)。

表1 1维单纯形子算法

输入: 初始搜索点 δ_0 , 数据集 X , 目标函数的计算公式 $F(\delta)$, 最大迭代次数 MT , 单纯形的最小半径 SR 。

输出: 最优核参数 δ^* 。

步骤1 根据给定的初始点 δ_0 构造初始单纯形;

步骤2 对于1维单纯形的两个顶点 A 和 B , 计算出目标函数值 $F(\delta_A)$ 和 $F(\delta_B)$, 函数值较小的为 $best$ 点, 较大的为 $worst$ 点;

步骤3 计算反射点 R 和反射点处的目标函数值 $F(\delta_R)$, 如果反射点的目标函数值优于 $best$ 点, 则进入步骤4; 如果差于 $best$ 点, 则转步骤5; 如果反射点与 $best$ 点处目标值相等, 则转步骤6;

步骤4 计算扩展点 E 和扩展点处的目标函数值 $F(\delta_E)$, 比较反射点 R 处和扩展点 E 的目标函数值, 选择目标函数值较小的点替换单纯形中的 $worst$ 点, 然后转步骤7;

步骤5 比较反射点 R 与 $worst$ 点处的目标函数值, 根据比较结果确定压缩点 M 的位置, 若压缩点 M 的目标函数值 $F(\delta_M)$ 与 $worst$ 点相比更小, 则用 M 点替换 $worst$ 点, 然后转步骤7; 否则, 转步骤6;

步骤6 执行单纯形收缩操作;

步骤7 若达到最大迭代次数 MT , 或单纯形半径小于 SR , 则迭代结束, $best$ 顶点对应的 δ 值即为最优核参数值 δ^* ; 否则, 转步骤2进行下一次迭代。

(2) 步骤4中扩展点的核参数值为 $\delta_E = \pi\delta_R + (1 - \pi)\delta_{best}$, 其中 π 称为扩张系数($\pi > 1$)。

(3) 步骤5中, 若 $F(\delta_R) < F(\delta_{worst})$, 压缩点的核参数值为 $\sigma_M = \varphi\sigma_R + (1 - \varphi)\sigma_{best}$, 若 $F(\delta_R) \geq F(\delta_{worst})$, 则压缩点的核参数值为 $\sigma_M = \varphi\sigma_{worst} + (1 - \varphi)\sigma_{best}$ 。其中 φ 是压缩系数($0 < \varphi < 1$)。

(4) 步骤6中, 若从步骤5进入步骤6, 则向 $best$ 点收缩, $worst$ 点核参数值更新为 $\sigma_{worst} = \gamma\sigma_{worst} + (1 - \gamma)\sigma_{best}$, 若从步骤3进入步骤6, 则向 $worst$ 点收缩, $best$ 点核参数值更新为 $\sigma_{best} = \gamma\sigma_{best} + (1 - \gamma)\sigma_{worst}$, 其中 γ 为收缩系数($0 < \gamma < 1$)。

上述的反射系数 μ 、扩展系数 π 、压缩系数 φ 、收缩系数 γ 的取值, 采用文献[14]标准Nelder-

Mead单纯形法的默认值, 分别为 $\mu = 1$, $\pi = 2$, $\varphi = 0.5$, $\gamma = 0.5$ 。

4 K插值单纯形法核极限学习机

K插值单纯形法核极限学习机的具体训练算法流程如表2所示。

表2算法中的若干要点说明如下。

4.1 步骤1中的初始化参数说明

算法初始化参数设置如下: 插值参数 K 一般设为12, 正则化参数 C 一般取1, 最大迭代次数 MT 设为100, 半径 SR 设为0.001。

4.2 步骤3的训练精度计算方法

核ELM在一个核参数 δ_i 下的训练精度 P_{δ_i} 定义为: $P_{\delta_i} = \eta_{\delta_i}/N$ 。其中 η_{δ_i} 为在核参数 δ_i 下核ELM正确分类的样本数, N 为训练样本总数。

表2 K插值单纯形法核极限学习机训练算法

输入: 训练数据集 X , 测试数据集 Y 。

输出: 最优核参数 δ^* 、分类函数 $\tilde{f}(\cdot)$, 测试数据集的分类结果。

步骤1 初始化参数, 给定插值数目 K , 正则化参数 C , 最大迭代次数 MT , 单纯形最小半径 SR ;

步骤2 数据预处理;

步骤3 计算 K 个插值核参数在训练数据集 X 上的训练精度, 并从中找到合适的初始搜索点 δ_0 :

(1) 构造核参数 δ 的 K 个插值点 δ_i 。 $\delta_i = 2^{i-1}$, 其中 $i = 1, 2, \dots, K$;

(2) 用式(8)分别计算 K 个插值点 δ_i 下相应的分类参数 $\tilde{\beta}_i$, 其中 $i = 1, 2, \dots, K$;

(3) 用式(9)计算 K 个插值点 δ_i 下的分类结果和训练精度 P_{δ_i} , 其中 $i = 1, 2, \dots, K$;

(4) 构造训练精度矩阵 $S = \begin{bmatrix} \delta_1 & P_{\delta_1} \\ \delta_2 & P_{\delta_2} \\ \vdots & \vdots \\ \delta_k & P_{\delta_k} \end{bmatrix}$;

(5) 删除训练精度矩阵 S 中 $P_{\delta_i} \geq 0.999$ 的行, 得到 S' ;

(6) 取 S' 矩阵中的 $S'(1, 1)$ 元素作为初始搜索点 δ_0 ;

步骤4 调用单纯形法优化核参数子算法, 进行迭代搜索, 传入搜索初值 δ_0 , 数据集 X , 目标函数计算公式 $F(\delta)$, 最大迭代次数 MT 和单纯形最小半径 SR , 得到最优核参数 δ^* ;

步骤5 根据最优核参数 δ^* , 用式(8)计算核ELM的分类参数 $\tilde{\beta}$;

步骤6 对于测试数据集 Y 中的待分类样本 x_p , 根据最优核参数和步骤5的 $\tilde{\beta}$, 用式(9)计算 x_p 的类别。

4.3 步骤4单纯形目标函数的定义

在第3节的单纯形子算法中, 需要传入目标函数的公式表达式 $F(\delta)$, 单纯形法在迭代过程中调用此公式计算各单纯形顶点的目标函数值。首先, 核ELM的训练误差定义为 $|1 - P_{\delta}|$, 但由于核ELM的训练误差为零时, 存在过学习(过拟合)问题, 故本文的目标函数设计为: $F(\delta) = |0.999 - P_{\delta}|$ 。

使用高斯核函数的SVM, 核参数 δ 有两个特性^[17,18], 当 $\delta \rightarrow 0$ 时, 所有训练样本均能正确分类, 即训练误差为0, 此时SVM存在过学习; 当 $\delta \rightarrow \infty$ 时, 高斯核对样本点判断能力为0, 即训练误差趋于常数, SVM存在欠学习。在高斯核参数 δ 从 ∞ 逐渐趋于0的过程中, 高斯核SVM的学习推广能力经历了由低到高又变低的过程。即随着高斯核参数 δ 从 $\infty \rightarrow 0$, 测试精度会随着训练精度的提高而提高, 当训练精度达到100%(训练误差为0)之后, 继续缩小核参数 δ , 训练精度依然维持100%, 但是会出现过学习现象。如图1所示。

故应该选择合适的核参数使训练误差足够小, 但又不能陷入训练误差为0的区域。因此, 在单纯形目标函数设计中, 用 $|0.999 - P_{\delta}|$ 代替训练误差 $|1 - P_{\delta}|$ 。

4.4 算法时间复杂度分析

算法的步骤1执行初始化操作, 计算时间复杂度是 $O(1)$ 。步骤2数据预处理, 计算量与数据集的大小 N 成正比, 计算时间复杂度是 $O(N)$ 。步骤3用

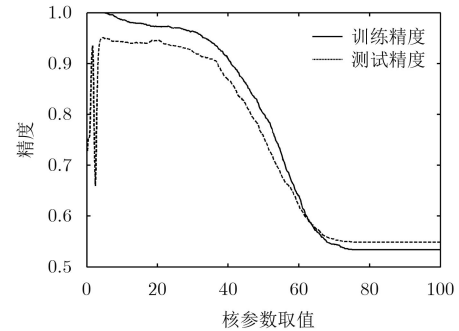


图1 高斯核参数取值对训练精度和测试精度的影响

插值法找出核参数的初始值, 计算时间复杂度是 $O(N^2k)$, 其中 k 是插值点的数目。步骤4用单纯形法搜索最优核参数, 计算时间复杂度是 $O(N^2t)$, 其中 t 为单纯形子算法的迭代次数。步骤5计算极限学习机的分类参数, 计算时间复杂度是 $O(N^2)$ 。步骤6计算待分类样本的类别, 计算时间复杂度是 $O(N)$ 。综上分析, 本文所提算法的计算时间复杂度为 $O(1+N+N^2k+N^2t+N^2+N) = O(N^2(k+t+1)+2N+1)$, 由此可见, 本文算法的计算时间复杂度为多项式时间。

5 仿真实验与分析

实验所用测试数据集如表3所示。其中前11个为UCI数据集, 最后1个是人造数据集。

将本文算法与传统核ELM算法作比较, 表4是两算法的分类精度比较结果。对于传统核ELM,

表3 实验数据集列表

数据集	样本数目	属性数目	簇的数目	数据集	样本数目	属性数目	簇的数目
DNA	1967	180	3	Segment	2310	19	7
Letter	20000	16	26	Satellite	6435	36	6
Msplice	3044	240	3	Vowel	990	14	11
Musk	5692	166	2	Liver	345	7	2
Cnae	1080	857	9	Wilt	4839	5	2
Chess	28056	6	18	D31	3100	2	31

表4 两种核极限学习机算法的分类精度比较

数据集	传统核ELM(%)						KS-KELM				
	$\delta=0.01$	$\delta=0.1$	$\delta=1$	$\delta=10$	$\delta=50$	$\delta=100$	最好精度	最差精度	平均精度	δ 值	精度(%)
DNA	21.41	24.63	80.94	94.65	66.38	55.03	94.65	21.41	57.17	4.5	95.72
Letter	4.19	83.32	89.04	77.88	54.22	49.69	89.04	4.19	59.72	2.4521	96.92
Msplice	54.86	59.59	79.53	94.24	75.52	54.86	94.24	54.86	69.77	6	94.82
Musk	84.08	86.09	89.96	95.22	95.94	96.13	96.13	84.08	91.24	224	96.99
Cnae	9.27	80.39	91.38	88.58	71.55	49.35	91.38	9.27	65.09	1.6631	92.67
Chess	0.17	58.55	62.10	32.82	24.35	22.08	62.10	0.17	33.35	0.832	63.21
D31	89.33	95.67	96.67	92.33	23.50	16.33	96.67	16.33	68.97	1.7471	97.17

由于事先不知道各数据集核参数 δ 的最佳取值, δ 分别取0.01, 0.1, 1, 10, 50, 100。由表4可知, 在7个测试数据集上, 本文算法KS-KELM的分类精度比传统KELM平均分类精度提高41.73%, 并且均优于传统核ELM的最好精度。以Cane数据集为例, 传统核ELM的平均精度65.09%, 本文算法的分类精度是92.67%, 本文算法分类精度优于传统核ELM; 本文算法所获的最优 δ 值是1.6631, 而传统核ELM最好精度所用的 δ 值为1, 两者较为接近, 由此可知, 本文用 K 插值单纯形法搜索核ELM的最优核参数 δ 是可行的。对于高维数据集DNA, Msplice, Musk, Cnae, 本文算法的分类精度较传统核ELM平均提高了38%, 对于多类别数据集Letter, Chess, D31, 本文算法的分类精度较传统核ELM平均提高了60.17%, 由此可知, 本文算法对复杂数据集具有较好的自适应性。表4的实验结果表明: 本文用 K 插值单纯形法搜索核ELM的最优核参数是可行的, 此方法对各类数据集有较好的自适应性, 从而使本文算法能应用于更广泛的场合。

由于单纯形法对待搜索参数的初始值较为敏感, 参数初值取值不当, 可能会导致单纯形法搜索失败。本文提出用 K 插值为单纯形法提供适当的核参数初值, 图2是6个测试数据集上, K 插值对单纯形法核ELM训练过程的影响。在图2中, 用“Simplex”表示普通单纯形法, 其初始核参数 $\delta_0=100$ 。图2中一个数据点代表一次训练迭代, 由图2可知, 使用 K 插值法后, 单纯形法的核ELM在6个测试集

上平均迭代次数为12.33次, 而普通单纯形法核ELM的平均迭代次数为19.56次, 因此, 使用 K 插值后, 单纯形核ELM的训练迭代次数减少了36.96%, 有效地提高了核ELM的训练效率。以Letter数据集为例, K 插值单纯形法的迭代次数为14次, 而普通单纯形法的迭代次数为24, 使用 K 插值后, 核ELM的训练迭代次数减少了41.67%。尤其在Msplice数据集上, 迭代次数减少50%。此外, 由图2还可以看出, K 插值还能避免单纯形法陷入局部最优解, 以DNA和Musk数据集为例, 普通单纯形法在搜索核参数过程会陷入局部最优解, 在图2上表现为训练误差曲线近似水平并且停留在一个大于零的值上, 即: 训练无法跳出局部最优, 训练不收敛; 在使用 K 插值后, 训练误差曲线能趋近于零, 表明单纯形法能避免陷入局部最优, 训练快速收敛。图2的实验结果表明: K 插值能为单纯形法提供一个适合的搜索初值, 有效地解决了单纯形法对搜索参数初值敏感的问题, 加速了核ELM的训练收敛。

表5给出本文算法与近两年4种相关算法的分类精度比较。4种相比较的极限学习机算法分别是: 带权重变化和决策融合的极限学习机(WMOS-2017)^[19]、基于文化基因算法的极限学习机(M-2016)^[20]、实例克隆极限学习机(IC-2017)^[21]、基于多层感知机的极限学习机(H-2016)^[22]。表5相关比较算法的分类精度取自相应的文献, 表中的“—”符号表示相比较算法文献未给出该数据集的

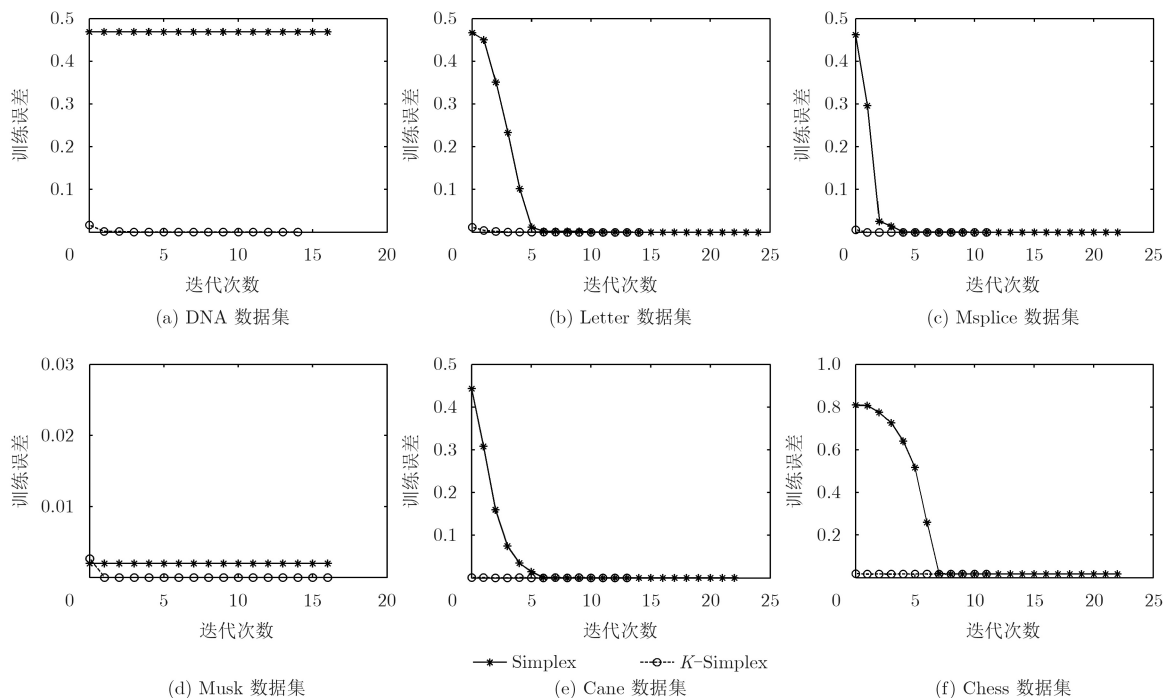


图2 K 插值法对单纯形核ELM训练收敛的影响

表5 5种极限学习机算法分类精度的比较

数据集	WMOS-2017 ^[19]	M-2016 ^[20]	IC-2017 ^[21]	H-2016 ^[22]	KS-KELM
Segment	94.12	95.56	—	—	95.12
Satellite	88.54	88.25	80.03	90.9	91.30
Letter	86.48	72.30	76.84	95.82	96.92
Vowel	88.78	59.74	—	64.94	92.12
Liver	—	—	66.56	72.17	80.26
Wilt	96.18	97.35	96.52	—	98.06

分类精度。由表5可知, 在高维数据集(Satellite, Letter, Vowel), 以及低维数据集(Liver, Wilt)上, 本文算法的分类精度均优于相比较的4种算法, 分类精度平均高出8.76%。在表3中, 本文算法的分类精度在Segment数据集上较M-2016低0.46%, 但这个分类精度偏差很小; 在其余的数据集上, 本文算法的分类精度均优于相比较的算法, 这说明本文引入 K 插值单纯形法优化核ELM的参数是可行的、有效的。

图3是本文算法的时间复杂度特征曲线图。将6个测试数据集的数据规模以2,4,8,16倍复制扩展, 生成更大规模的数据集。然后在这些新数据集上, 应用本文算法学习新数据集, 绘制出算法的训练时间随数据集规模变化的趋势图。由图3可以看出, 当测试集的数据规模呈指数增长时, 本文算法的训练时间呈多项式曲线变化趋势, 实验结果与本文算法计算时间复杂度分析的结论符合。

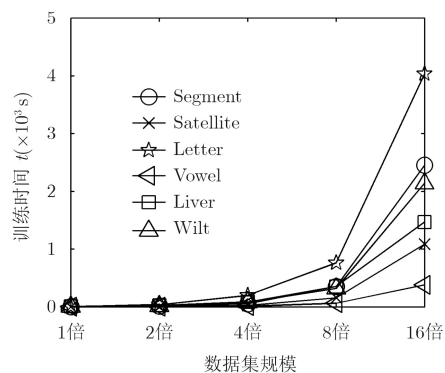


图3 本文算法的时间复杂度特征曲线

6 结束语

本文提出用 K 插值单纯形法寻优极限学习机的核参数, 提高极限学习机的训练效率和分类精度, 由测试数据集上的实验结果分析可得到两个结论: 一是用Nelder-Mead单纯形法优化核极限学习机的核参数是可行的, 所提算法的分类精度优于相比较的其它算法; 二是用 K 插值为Nelder-Mead单纯形法提供核参数搜索初值, 能有效地减少单纯形法的

搜索迭代次数, 从而提高了核极限学习机的训练效率, 并且此方法对数据集具有良好地自适应性, 能使所提算法适用于更广泛的场合。

参考文献

- [1] HUANG Guangbing, ZHU Qinyu, and SIEW C K. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2006, 70(1): 489–501. doi: 10.1016/j.neucom.2005.12.126.
- [2] SALCEDO S S, CASANOVA M C, PASTOR S A, et al. Daily global solar radiation prediction based on a hybrid coral reefs optimization – extreme learning machine approach[J]. *Solar Energy*, 2014, 105: 91–98. doi: 10.1016/j.solener.2014.04.009.
- [3] SONG Yuedong and ZHANG Jiexiang. Discriminating preictal and interictal brain states in intracranial EEG by sample entropy and extreme learning machine[J]. *Journal of Neuroscience Methods*, 2016, 257: 45–54. doi: 10.1016/j.jneumeth.2015.08.026.
- [4] SURI M and PARMAR V. Exploiting intrinsic variability of filamentary resistive memory for extreme learning machine architectures[J]. *IEEE Transactions on Nanotechnology*, 2015, 14(6): 963–968. doi: 10.1109/TNANO.2015.2441112.
- [5] DING Shifei, ZHAO Han, ZHANG Yanan, et al. Extreme learning machine: Algorithm, theory and applications[J]. *Artificial Intelligence Review*, 2015, 44(1): 103–115. doi: 10.1007/s10462-013-9405-z.
- [6] HUANG Guangbing, ZHOU Hongming, DING Xiaojian, et al. Extreme learning machine for regression and multiclass classification[J]. *IEEE Transactions on Systems Man & Cybernetics Part B*, 2012, 42(2): 513–529. doi: 10.1109/TSMCB.2011.2168604.
- [7] HUANG Guangbing, DING Xiaojian, and ZHOU Hongming. Optimization method based extreme learning machine for classification[J]. *Neurocomputing*, 2010, 74(1): 155–163. doi: 10.1016/j.neucom.2010.02.019.
- [8] VAN G T, SUYKENS J A, LANCKRIET G, et al. Bayesian framework for least-squares support vector machine classifiers, gaussian processes, and kernel Fisher discriminant analysis[J]. *Neural Computation*, 2014, 14(5):

- 1115–1147. doi: [10.1162/089976602753633411](https://doi.org/10.1162/089976602753633411).
- [9] ADANKON M M, CHERIEL M, and AYAT N F. Optimizing resources in model selection for support vector machines[J]. *Pattern Recognition*, 2007, 40(3): 953–963. doi: [10.1016/j.patcog.2006.06.012](https://doi.org/10.1016/j.patcog.2006.06.012).
- [10] XU Yitian, PAN Xianli, ZHOU Zhijian, *et al.* Structural least square twin support vector machine for classification[J]. *Applied Intelligence*, 2015, 42(3): 527–536. doi: [10.1007/s10489-014-0611-4](https://doi.org/10.1007/s10489-014-0611-4).
- [11] XIAO Yingchao, WANG Huangang, ZHANG Lin, *et al.* Two methods of selecting Gaussian kernel parameters for one-class SVM and their application to fault detection[J]. *Knowledge-Based Systems*, 2014, 59(2): 75–84. doi: [10.1016/j.knsys.2014.01.020](https://doi.org/10.1016/j.knsys.2014.01.020).
- [12] XIAO Yingchao, WANG Huangang, and XU Weili. Parameter selection of gaussian kernel for one-class SVM[J]. *IEEE Transactions on Cybernetics*, 2015, 45(5): 941–953. doi: [10.1109/TCYB.2014.2340433](https://doi.org/10.1109/TCYB.2014.2340433).
- [13] TIAN Meng and WANG Wenjian. An efficient Gaussian kernel optimization based on centered kernel polarization criterion[J]. *Information Sciences*, 2015, 322: 133–149. doi: [10.1016/j.ins.2015.06.010](https://doi.org/10.1016/j.ins.2015.06.010).
- [14] LAGARIAS J C, WRIGHT M H, WRIGHT P E, *et al.* Convergence properties of the Nelder-Mead simplex method in low dimensions[J]. *SIAM Journal on Optimization*, 1998, 9(1): 112–147. doi: [10.1137/S1052623496303470](https://doi.org/10.1137/S1052623496303470).
- [15] 冯增喜, 任庆昌, 彭彦平, 等. 基于单纯形法的MFAC参数寻优[J]. *控制工程*, 2016, 23(3): 405–410. doi: [10.14107/j.cnki.kzgc.150456](https://doi.org/10.14107/j.cnki.kzgc.150456).
- FENG Zengxi, REN Qingchang, PENG Yanping, *et al.* Optimizing the parameters of MFAC based on the simplex method[J]. *Control Engineering of China*, 2016, 23(3): 405–410. doi: [10.14107/j.cnki.kzgc.150456](https://doi.org/10.14107/j.cnki.kzgc.150456).
- [16] 郑伟博, 张纪会. 基于Nelder-Mead单纯形法的改进量子行为粒子群算法[J]. *复杂系统与复杂性科学*, 2016, 13(2): 97–104. doi: [10.13306/j.1672-3813.2016.02.012](https://doi.org/10.13306/j.1672-3813.2016.02.012).
- ZHENG Weibo and ZHANG Jihui. A improved quantum behaved particle swarm optimization algorithm using nelder and mead's simplex algorithm[J]. *Complex Systems & Complexity Science*, 2016, 13(2): 97–104. doi: [10.13306/j.1672-3813.2016.02.012](https://doi.org/10.13306/j.1672-3813.2016.02.012).
- [17] KEERTHI S S and LIN C J. Asymptotic behaviors of support vector machines with gaussian kernel[J]. *Neural Computation*, 2003, 15(7): 1667–1689. doi: [10.1162/089976603321891855](https://doi.org/10.1162/089976603321891855).
- [18] 张小云, 刘允才. 高斯核支撑向量机的性能分析[J]. *计算机工程*, 2003, 29(8): 22–25. doi: [10.3969/j.issn.1000-3428.2003.08.009](https://doi.org/10.3969/j.issn.1000-3428.2003.08.009).
- ZHANG Xiaoyun and LIU Yuncai. Performance analysis of support vector machines with Gauss kernel[J]. *Computer Engineering*, 2003, 29(8): 22–25. doi: [10.3969/j.issn.1000-3428.2003.08.009](https://doi.org/10.3969/j.issn.1000-3428.2003.08.009).
- [19] 罗家祥, 罗丹, 胡跃明. 带权重变化和决策融合的ELM在线故障检测[J]. *控制与决策*, 2017. doi: [10.13195/j.kzyjc.2017.0274](https://doi.org/10.13195/j.kzyjc.2017.0274).
- LUO Jiexiang, LUO Dan, and HU Yueming. A new online extreme learning machine with varying weights and decision level fusion for fault detection[J]. *Control and Decision*, 2017. doi: [10.13195/j.kzyjc.2017.0274](https://doi.org/10.13195/j.kzyjc.2017.0274).
- [20] ZHANG Yongshan, WU Jia, CAI Zhihua, *et al.* Memetic extreme learning machine[J]. *Pattern Recognition*, 2016, 58: 135–148. doi: [10.1016/j.patcog.2016.04.003](https://doi.org/10.1016/j.patcog.2016.04.003).
- [21] ZHANG Yongshan, WU Jia, Zhou Chuan, *et al.* Instance cloned extreme learning machine[J]. *Pattern Recognition*, 2017, 68: 52–65. doi: [10.1016/j.patcog.2017.02.036](https://doi.org/10.1016/j.patcog.2017.02.036).
- [22] TANG Jiexiong, DENG Chenwei, and HUANG Guangbing. Extreme learning machine for multilayer perceptron[J]. *IEEE Transactions on Neural Networks & Learning Systems*, 2016, 27(4): 809–821. doi: [10.1109/TNNLS.2015.2424995](https://doi.org/10.1109/TNNLS.2015.2424995).
- 苏一丹: 男, 1962年生, 教授, 研究方向为自然计算、数据挖掘。
李若愚: 男, 1989年生, 硕士生, 研究方向为智能优化及在数据挖掘中的应用。
覃 华: 男, 1972年生, 教授, 研究方向为量子计算理论与近似动态规划最优化方法、数据挖掘。
陈 琴: 女, 1974年生, 副教授, 研究方向为数据挖掘。