

基于离散微粒群优化算法的 SVM 参数选择

马 驰,阮秋琦

(北京交通大学 计算机学院 信息所,北京 100044)

摘 要:支持向量机(SVM)的学习性能和泛化能力主要取决于参数选择,然而传统的优化算法难以解决此问题。文中通过支持向量的个数建立优化目标函数,采用微粒群优化(PSO)算法对其优化,寻找最优参数。PSO 是一种新兴的基于群体智慧的进化算法。实验表明,微粒群优化算法是支持向量机参数选择的有效方法。

关键词:支持向量机;微粒群优化算法;参数选择;人脸识别

中图分类号:TP181

文献标识码:A

文章编号:1673-629X(2007)12-0020-04

Parameter Selection for SVM Based on Discrete PSO

MA Chi, RUAN Qiu-qi

(Information Institute of Sciences, School of Computer, Beijing Jiaotong University, Beijing 100044, China)

Abstract:Support vector machine (SVM) has been proved a powerful technique for solving problems in pattern classification and regression, but its learning capacity and generalization capacity mainly depends on the parameters selection of it. Parameters selection for SVM is very complex in nature and quite hard to solve by conventional optimization techniques. So in this paper, a new methodology, based on PSO, for parameters selection of support vector machine is proposed. The simulation (face recognition) result assures the validity of the methodology on time compared with leave-one-out method.

Key words:support vector machine; discrete particle swarm optimization; parameter selection; face recognition

0 引 言

支持向量机(Support Vector Machines, SVM)是 V. Vapnik 等根据统计学习理论提出的一种新的机器学习方法,其优势在于小样本问题、推广能力和非线性处理能力,尤其在处理高维数据时,有效地解决了“维数灾难”问题。支持向量机从本质上讲是一种前向神经网络,根据结构风险最小化准则,在使训练样本分类误差极小化的前提下,尽量提高分类器的泛化推广能力。这种学习算法已经在手写数字识别、三维目标识别、人脸识别、文本图像分类、时间序列预测、主成分分析等实际问题中,表现出了良好的学习能力^[1]。

受到鸟群觅食的启发, J. Kennedy 和 R. C. Eberhart 于 1995 年提出了微粒群算法(PSO),该算法是一种进化计算技术,经过众多学者十余年的研究,它已被广泛应用于解决多个领域的优化问题,如神经网络训练^[2]、电磁学^[3]等等。与传统的优化算法相比,微粒群

算法能搜索非线性多峰的复杂空间,同时能处理参数空间变化的优化问题。与同是进化算法的遗传算法相比较,微粒群算法能更快达到全局最优,且基本不受问题峰数和维数的限制^[2]。

SVM 性能依赖于参数的选择。文中提出了基于 DPSO 的参数选择方法,并以人脸识别为例,与网格搜索法进行比较。实验结果表明,在同等搜索精度的条件下,本方法在运算时间上有明显优势,方便了核函数参数的选择,对于核函数的进一步研究有重要意义。

1 SVM 分类器

1.1 SVM 基本原理

SVM 包括分类和回归两种基本形式^[4,5],这里以 SVM 分类为例来说明参数的选取。SVM 分类本质上是两类问题,训练样本为 (x_i, y_i) , $x \in R^n$, $y \in \{-1, 1\}$, $i = 1, 2, \dots, l$, 其中 y 标志样本类别。SVM 的目标是根据风险最小的原理,构造一个目标函数,寻找满足分类要求的分割超平面:

$$w^T x + b = 0 \quad (1)$$

使得训练集中的点尽可能地远离该超平面,是使两类数据间的最大间隔 $2 / \|w\|$ 尽可能大(见图 1),其

收稿日期:2007-03-06

基金项目:国家自然科学基金资助项目(60472033;60472062)

作者简介:马 驰(1982-),男,山西太原人,硕士研究生,研究方向为模式识别与智能系统;阮秋琦,教授,博士生导师,研究方向为图像处理、计算机视觉多媒体信息处理、虚拟现实、CSW。

中 w 为超平面的法向量, b 为超平面的偏移量。

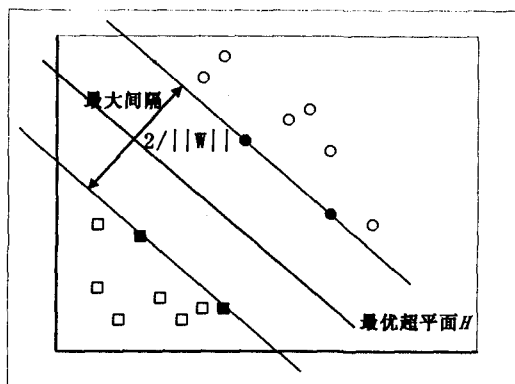


图 1 最优分类超平面

对于线性不可分的情形,必须“软化”对间隔的要求,要解决的问题是求解下面的最优化问题:

$$\begin{aligned} \min J(w, \xi) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t. } y_i((w \cdot x_i) + b) &\geq 1 - \xi_i, i = 1, \dots, l \\ \xi_i &\geq 0, i = 1, \dots, l \end{aligned} \quad (2)$$

其中 $C > 0$ 是惩罚参数。这是一个二次规划问题,二次规划问题的 Wolfe 对偶问题的约束远远比原始问题的约束简单。所以,为了求解 SVM, 常常通过引入 Lagrange 函数,求解式(2)问题的 Wolfe 对偶问题:

$$\begin{aligned} \min & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) - \sum_{j=1}^l \alpha_j \\ \text{s.t. } & \sum_{i=1}^l y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, \dots, l \end{aligned} \quad (3)$$

其中 α 为 Lagrange 乘子,根据最优化理论中的 KKT 条件,只有少量样本的 α 值不为零,这些样本即为支持向量。由于问题(3)中只出现两个向量的内积运算, Vapnik 等人用满足 Mercer 条件的核函数 $K(x, y)$ 来代替内积运算 (x, y) , 来实现线性算法的非线性化。其基本思想是作非线性映射 $\varphi(x)$, 将输入数据空间映射到一个高维的内积空间 F (也称特征空间), 然后在 F 中实现线性分划。此时判别函数是:

$$f(x) = \text{sign} \left\{ \sum_{i=1}^n \alpha_i y_i K(x_i, x) + b \right\} \quad (4)$$

1.2 核函数

核函数^[1]在支持向量机中是至关重要的,采用不同的核函数将会导致不同的 SVM 算法,然而核函数的选择至今没有一个完善的理论指导。目前主要使用的核函数主要有以下三种:

(1) 多项式形式核函数:

$$K(x, y) = ((x, y) + 1)^d$$

此时得到的支持向量机就是一个 d 阶多项式分类器。特别当 $d = 1$ 时,此时就是线性核函数,也成为线

性支持向量机。

(2) 径向基核函数:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma}\right)$$

得到的支持向量机是一种径向基函数分类器。它与传统的神经网络中的 RBF 网络的基本区别是:这里的每一个基函数的中心都对应于一个支持向量,它们以及输出权值都是由算法自动确定的。

(3) Sigmoid 核函数:

$$K(x, y) = \tanh(v(x, y) + c)$$

其中 v 和 c 是常数。则支持向量机实现的就是一个两层的感知器神经网络,不过这里网络的权值、隐层结点数都是由算法自动确定的。

1.3 多类问题

标准的支持向量机是针对两类问题的,对于多类问题,通常的处理方法有^[6]:层次分类法(也叫二叉树法),一对多,一对一。文中采用分类效果最佳的一对一方法,即对于 m 类中的任意两类样本 i 和 $j, i \neq j$, 构建一个标准的支持向量机 $f_{i,j}$ 将它们分开。由于 $f_{i,j}$ 是对称的,即: $f_{i,j} = -f_{j,i}$, 因此一共需要 $C_m^2 = m(m-1)/2$ 个支持向量机。

2 基于 DPSO 的 SVM 参数选择

2.1 离散微粒群算法原理

微粒群算法将群体内的每个个体看作没有体积和重量的微粒,微粒在 n 维搜索空间内以一定的速度飞行搜索^[3]。飞行速率和方向根据微粒自身的飞行经验和整个微粒群的飞行经验动态调整。微粒 i 的当前位置为 $X_i = (x_{i1}, x_{i2}, \dots, x_{in})$, 当前飞行速度为 $V_i = (v_{i1}, v_{i2}, \dots, v_{in})$, 所经历的最好位置即 i 所经历过的具有最好适应值的位置为 $P_i = (p_{i1}, p_{i2}, \dots, p_{in})$ 。群体中所有微粒所经历的最好位置为 P_{best} 。带惯性权重的微粒群算法的飞行方程为:

$$v_{ij}(t+1) = w(t)v_{ij}(t) + c_1(t)r_{1j}(t)(p_{ij}(t) - x_{ij}(t)) + c_2(t)r_{2j}(t)(p_{\text{best}j}(t) - x_{ij}(t)) \quad (5)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (6)$$

其中 j 表示微粒的第 j 维, t 表示第 t 代, w 为惯性权重,通常取 0.8 到 1.2 之间, c_1, c_2 为通常在 0 到 2 间取值, w, c_1, c_2 都是随时间随机变化的数, r_1, r_2 为两个相互独立的随机函数,服从 $U(0, 1)$ 分布。为了减少飞行过程中微粒离开搜索空间的可能性,通常设定 $v_{ij} \in [-v_{\text{max}}, v_{\text{max}}]$ 。

为保证离散微粒群算法在整数空间内进行搜索,对算法进行如下修改:

$$\text{令 } \phi_1 = c_1(t)r_{1j}(t)(p_{ij}(t) - x_{ij}(t))$$

$$\phi_2 = c_2(t)r_{2j}(t)(p_{bestj}(t) - x_{ij}(t))$$

其中 $\phi_1 \in [a_1, b_1]$, $\phi_2 \in [a_2, b_2]$, ϕ_1 和 ϕ_2 为在各自区间内等概率分布的整数。区间范围确定如下:

$$a_1 = \begin{cases} 0, & p_{ij}(t) > x_{ij}(t) \\ c_1(p_{ij}(t) - x_{ij}(t)), & p_{ij}(t) \leq x_{ij}(t) \end{cases}$$

$$b_1 = \begin{cases} c_1(p_{ij}(t) - x_{ij}(t)), & p_{ij}(t) > x_{ij}(t) \\ 0, & p_{ij}(t) \leq x_{ij}(t) \end{cases}$$

$$a_2 = \begin{cases} 0, & p_{bestj}(t) > x_{ij}(t) \\ c_2(p_{bestj}(t) - x_{ij}(t)), & p_{bestj}(t) \leq x_{ij}(t) \end{cases}$$

$$b_2 = \begin{cases} c_2(p_{bestj}(t) - x_{ij}(t)), & p_{bestj}(t) > x_{ij}(t) \\ 0, & p_{bestj}(t) \leq x_{ij}(t) \end{cases}$$

则式(5)修改成:

$$v_{ij}(t+1) = \text{int}(w(t)v_{ij}(t)) + \phi_1 + \phi_2 \quad (7)$$

其中 $\text{int}(w(t)v_{ij}(t))$ 是对 $w(t)v_{ij}(t)$ 取整。将 X_i 初始化为整数,式(7)保证了算法在整数空间内搜索。

2.2 基于 DPSO 的 SVM 参数选择

惠康华、李春利提出 W_2^1 再生核函数^[7]:

$$K(x, y) = \frac{\cosh(x + y - a - b) + \cosh(|x - y| - b + a)}{2\sinh(b - a)} \quad (8)$$

其中 $a = 0, b = 1$, 鉴于 a, b 确定后该核函数分母为常数,文中对其改进为:

$$K(x, y) = \lambda \cosh\left(\frac{x + y}{w_1} - a - b\right) + (1 - \lambda) \cosh\left(\frac{|x - y|}{w_2} - b + a\right) \quad (9)$$

其中 $a = 0, b = 1, w_1, w_2$ 为归一化参数, λ 为加权系数。参数变化导致核函数变化见图 2。

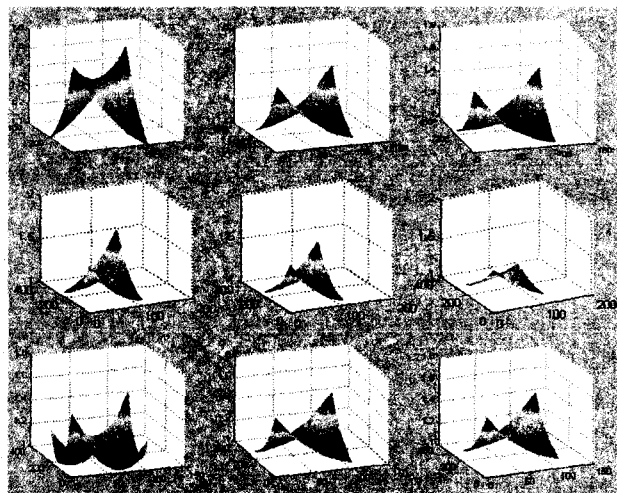


图 2 核函数随参数变化图

图 2 第一行中三幅核函数图像从左到右分别是 $\lambda = 0.2, 0.5, 0.7$; 第二行中三幅核函数图像从左到右分别是 $w_1 = 0.5, 1, 1.5$; 第三行中三幅核函数图像从左到右分别是 $w_2 = 0.5, 2, 2.5$ 。

推广能力是指学习机对未知数据进行测试时的分类性能。对 SVM 推广能力的估计还没有简单、准确的方法,而好的推广估计方法是实现 SVM 参数选择的基础。已经证明,留一法对分类性能的估计是无偏的^[6,8],但其计算量比较大,因此提出以下定理。

定理 1^[9]: 如果一组训练样本能够被一个最优分类面和支持向量机分开,则对于测试样本分类错误率的期望的上界是训练样本中平均的支持向量占总训练样本数的比例,即:

$$E(P(\text{error})) \leq \frac{E[\text{支持向量个数}]}{\text{训练样本数} - 1} \quad (10)$$

因此只要能够选择适当的参数,构造一个支持向量数相对较少的最优分类面,便可以得到较好的分类。

优化过程就是要找到使 $E(P(\text{error}))$ 最小的 C 和 λ, w_1, w_2 , 通过对径向基核函数 SVM 的参数 C 研究可知, C 是在结构风险和样本误差之间作出折衷,取值与可容忍的误差相关,较大的 C 值允许较小的误差,较小的 C 值则允许较大的 C 误差。如图 3 所示,只要 C 取值足够大,就可得到较好的分类能力。于是文中将利用微粒群算法选取最优参数 λ, w_1 和 w_2 。图 3 证实了定理 1 的结论。

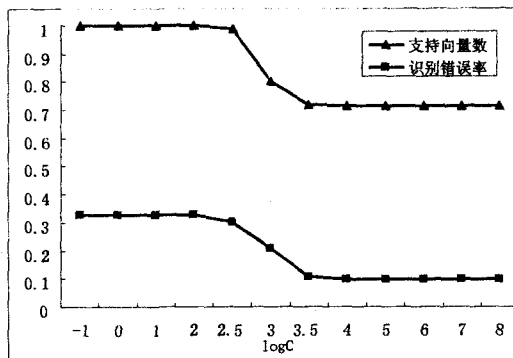


图 3 支持向量数与识别错误率

图 3 中偏上的曲线表示支持向量数与(训练样本数 - 1)的比值,偏下的曲线表示识别错误率。

选取适应函数

$$d_0 = \frac{\text{支持向量个数}}{\text{训练样本数} - 1} \quad (11)$$

需要搜索的参数 λ, w_1 和 w_2 , 从而确定微粒的搜索空间为三维 $X_i = (x_{i1}, x_{i2}, x_{i3})$, 搜索优化过程如下^[3]:

Step1 初始化设置微粒群的规模、惯性权值、加速系数、最大允许迭代次数或适应值误差限,及各微粒的初始位置和初始速度等。

Step2 评价各微粒的初始适应值。

Step3 对于每个微粒,将其适应值与所经历过的最好位置 P_i 的适应值进行比较,若较小,则将其作为当前的最好位置。

Step4 对于每个微粒,将其 P_i 与全局所经历的

最好位置 P_{best} 的适应值进行比较,若较小,则将其作为当前全局的最好位置。

Step5 确定惯性权重 w ;

Step6 根据式(5)、(6)对微粒的速度和位置进行进化;

Step7 如果 d_0 没有达到迭代终止条件,则返回 Step2;否则结束,此时为最好位置。

3 实验分析

数据库为 orl 人脸库,共 40 类,每类 10 个样本。每一类的前五个样本作为训练样本,后五个样本作为检测样本。通过 LDA^[5],DCT^[10]和 ICA^[11]提取人脸图像的特征,组成人脸特征的组合特征,用于 SVM 的训练与识别。

搜索范围 $0.1 \leq \lambda \leq 0.9, 0.5 \leq w_1 \leq 3, 0.5 \leq w_2 \leq 3$ 。当格点数逐渐增加,即搜索精度逐渐提高,由(11)式得到的错误率上界见表 1。

表 1 错误率上界

格点数	64	125	196	343	5625
λ	0.63	0.7	0.74	0.77	0.8
w_1	0.5	0.5	0.5	0.5	0.5
w_2	2.16	2.38	2	2.17	2.2
E	0.7977	0.7914	0.7881	0.7863	0.7838

其中 $E = \text{支持向量数} / (\text{训练样本数} - 1)$ 。

将获得的最优参数用于人脸识别,两种参数搜索方法比较见表 2,3。说明本方法搜索的最优参数与网格搜索法完全一致,但搜索速度大大提高,耗时比较见图 4。

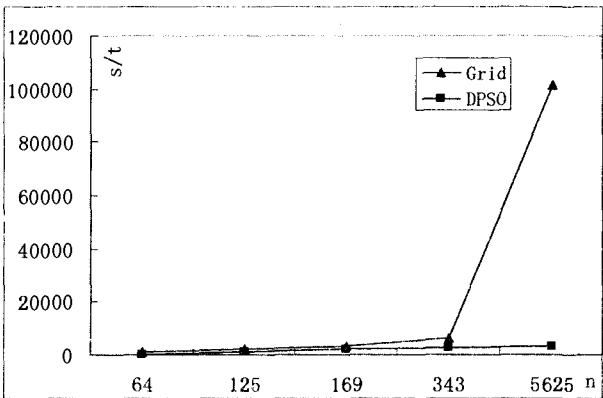


图 4 耗时比较

图 4 横坐标为网格点数,纵坐标为搜索时间,随着网格点数的增加,网格法的耗时大幅度增加,而 DPSO 的耗时则增加很慢,说明在同等搜索精度下,DPSO 方法在搜索速度上有明显的优势。

表 2 两种参数搜索法比较(1)

方法	核函数	λ	w_1	w_2
Grid	rbf	0.8	0.5	2
DPSO	rbf	0.8	0.5	2

表 3 两种参数搜索法比较(2)

方法	E	识别正确率	搜索时间(s)
Grid	78.38%	98%	101250
DPSO	78.38%	98%	3110

4 结 论

提出了基于离散微粒群的支持向量机参数选择法。并将离散位离群方法用于改进的再生核函数的参数选择,最终实现了人脸识别,识别率达到了 98%。结果证明,同常用的网格搜索法相比,离散微粒群方法大大提高了核函数最优参数搜索的速度,是一种有效的参数选择方法,并为进一步研究改进的再生核函数提供了实验支持。

参考文献:

- [1] 邓乃扬,田英杰.数据挖掘中的新方法:支持向量机[M].北京:科学出版社,2004.
- [2] Eberhart R C, Kennedy J. A new optimizer using particle swarm theory [C]//Proceedings of the Sixth International Symposium on Micro Machine and Human Science. New York, NY, USA: IEEE, 1995: 39-43.
- [3] Shi Y, Eberhart R C. A modified particle swarm optimizer [C]//Proceedings 1998 IEEE International Conference on Evolutionary Computation. New York, NY, USA: IEEE, 1998: 69-73.
- [4] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Knowledge Discovery and Data Mining, 1998, 2(2): 121-167.
- [5] Vapnik. 统计学习理论的本质[M]. 张学工译. 北京:清华大学出版社, 2000.
- [6] 边肇祺,张学工. 模式识别[M]. 第 2 版. 北京:清华大学出版社, 2000.
- [7] 惠康华,李春利. 基于 W_2 再生核支持向量机的模式分类研究[J]. 计算机工程, 2005, 31(21): 128-129.
- [8] 董春曦,饶 鲜,杨绍全,等. 支持向量机参数选择方法研究[J]. 系统工程与电子技术, 2004, 26(8): 1117-1120.
- [9] 张燕昆. 基于核方法的人脸识别技术的研究[D]. 上海:上海交通大学, 2003.
- [10] 阮秋琦. 数字图像处理[M]. 北京:电子工业出版社, 2004.
- [11] Hyvarinen A, Oja E. Independent Component Analysis: a survey[J]. Neural Computing Surveys, 1999, 2(4): 94-128.