基于核函数的非线性口袋算法

许建华,张学工,李衍达
(清华大学自动化系,智能技术与系统国家重点实验室,北京 100084)

摘 要: 应用满足 Mercer 条件的核函数设计非线性算法已经成为机器学习领域一项新的非线性技术. 核感知器算法利用核思想非线性地推广了线性感知器算法,使其可以处理原始输入空间中的非线性分类问题和高维特征空间中的线性问题. 线性口袋算法改进了线性感知器算法,能够直接处理线性不可分问题. 为了进一步改进线性口袋算法和核感知器算法,本文提出基于核函数的非线性口袋算法,即核口袋算法,其目标是找到一个使错分样本数最小的非线性判别函数,并证明了其收敛性. 核口袋算法的特点是用简单的迭代过程和核函数来实现非线性分类器的设计. 基准数据集的实验结果证明核口袋算法的性能优于线性口袋算法和核感知器算法.

关键词: 核函数; 感知器算法; 非线性; 口袋算法; 支持向量机
中图分类号: TP18 **文献标识码**: A **文章编号**: 0372-2112 (2003) 04-0612-04

Nonlinear Pocket Algorithm with Kernels

XU Jian-hua, ZHANG Xue-gong, LI Yan-da
(Department of Automation, Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, Beijing 100084, China)

Abstract: Designing nonlinear algorithms with kernel functions satisfying the Mercer condition, has become a novel nonlinear technique in the machine learning. By using kernel idea the kernel perceptron algorithm nonlinearly generalizes the linear perceptron algorithm. It can handle the linearly non-separable classification problems in the original input space and the linearly separable ones in the feature space. The linear pocket algorithm improves the perceptron algorithm and can deal with the linearly non-separable problems directly. In order to improve the linear pocket algorithm and kernel perceptron algorithm, in this paper the nonlinear pocket algorithm based on kernels (i. e. kernel pocket algorithm) is proposed, whose objective is to find a nonlinear discriminant function that can minimize the number of misclassified training samples. Its convergence is also proved. Its advantage is to implement a nonlinear classifier using a simply iterative procedure and kernel functions. The experiment results from some benchmark data sets show that the performance of our kernel technique is prior to that of the linear pocket algorithm and kernel perceptron algorithm.

Key words: kernel function; perceptron algorithm; nonlinear; pocket algorithm; support vector machines

1 引言

原始的 Rosenblatt 感知器算法(Perceptron Algorithm)只能处理线性可分问题[1,2]. 对线性不可分问题,算法在计算过程中会发生振荡现象,即算法无法停止. 在实际工作中很多学者设计出各种准则使算法终止(例如设置最大迭代次数、学习步长递减等),但是最终解的性质却是不确定的. Gallant[3,4]在感知器算法的迭代过程中引入一个口袋权向量来存放正确运行次数最多的感知器权向量,并称这一感知器的改进算法为口袋算法(Pocket Algorithm),其目标是找到一个借错分样本最少的解(最优解),并且证明了对于整数和有理数输入情况下口袋算法的收敛性(即口袋算法收敛定理). 因此口袋算法可以处理不可分问题,噪音数据和矛盾数据. Muselli[5]证明对于实数输入口袋算法收敛定理仍然成立,并且进一步证明了口袋算法的一个变换(带棘齿的口袋算法, Pocket Algorthm with Ratchet)在有限步迭代后能够在概率 1 找到一个最优解.

Vapnik 等人提出的支持向量机(Support Vector Machine, SVM)是近几年机器学习领域最有影响力的研究成果[6~8]. 在 SVM 中,一个引人注目的特点是核的思想,即应用满足 Mercer 条件的核函数代替两向量间抽人积运算来实现非线性变换,而不需要非变换的具体形式[8]. 尽管这一思想可以追溯到二十世纪六十年代的势函数方法[9],SVM 的成楞促使众多学者用核的思想改造经典的线性算法,提出相应的基于核的非线性算法,例如核主成分分析(KPCA)[10]、核 Fisher 判别分析(KFD)[11]等. 文[12]提出了一种基于核函数的非线性感知器算法(即核感知器算法),使之可以处理非线性分类问题,但是它也只能处理特征空间的线性可分问题.

为了改进 Gallant 口袋算法和核感知器算法,本文提出了

收稿日期:2001-09-15;修回日期:2002-06-20
基金项目:国家自然科学基金(No. 60275007)

基于核函数的非线性算法,即核口袋算法,其目标是找到一组系数使错分训练样本数为最少,并且证明了核口袋算法的收敛性. 实验结果表明核口袋算法的性能明显优于线性口袋算法和核感知器算法.

2　感知器算法的线性和核形式

假设现有来自二类(ω_1,ω_2)的训练样本集:

$$\{(x_1,y_1),(x_2,y_2),\cdots,(x_l,y_l)\} \tag{1}$$

其中 $x_i\in\mathbf{R}^n$; $y_i=+1$, $\forall x_i\in\omega_1$, $y_i=-1$, $\forall x_i\in\omega_2$; l 是训练集的总样本数.

如果二类训练样本线性可分,则存在一超平面:

$$f(x)=<w,x>+b=0 \tag{2}$$

可将二类样本正确地分开,其中 $<\cdot,\cdot>$ 为内积运算, $w\in\mathbf{R}^n$, $b\in\mathbf{R}$; $f(x)>0\ \forall x\in\omega_1$, $f(x)\leq 0\ \forall x\in\omega_2$. 任取一初始值,线性感知器算法的迭代更新公式为:

$$w\Leftarrow w+y_jx_j,\ b\Leftarrow b+y_j \quad \text{if} \quad f(x_j)\,y_j\leq 0 \tag{3}$$

其中 j 是随机或顺序选取的某一训练样本. 感知器收敛定理证明:对线性可分问题,从任意的初值出发,感知器算法都能在有限步迭代后收敛到一个解.

为了提高感知器算法处理复杂分类问题的能力,文[12]提出了基于核函数的非线性感知器算法,即核感知器算法,其基本思路是:将模式向量非线性地映射到一高维特征空间,在新的空间中构造一线性感知器算法,并使算法中只出现两向量间内积运算,最后用核函数来代替内积运算,从而实现算法的非线性化.

在核感知器算法中,基于核的非线性决策函数为:

$$f^{\Phi}(x)=\sum_{i=1}^{l}\alpha_iy_iK(x_i,x)+\beta \tag{4}$$

相应地,核感知器算法的迭代更新公式为:

$$\begin{aligned}&\alpha_i\Leftarrow\alpha_i+k(x_i,x_j)\,y_iy_j\\&\beta\Leftarrow\beta+y_j\end{aligned} \quad \text{if } f^{\Phi}(x_j)\,y_j\leq 0 \tag{5}$$

其中 α_i, $i=1,\cdots,l$, $k(x,y)$ 是满足 Mercer 条件的核函数.

目前常用的核函数有:多项式核 $k(x,y)=(<x,y>+c)^d$;径向基函数核 $k(x,y)=e^{\frac{-\|x-y\|^2}{2\sigma^2}}$;二层神经网:$k(x,y)=\tanh(\gamma<x,y>+c)$. 值得注意的是前二核函数所有参数都满足 Mercer 条件,而二层神经网只有一些参数的满足 Mercer 条件. 现在的决策函数在特征空间中是线性的,但在原始空间中是非线性的,决策规则为:当 $f^{\Phi}(x)>0$,则 $x\in\omega_1$;否则 $x\in\omega_2$.

3　口袋算法的线性和核形式

在线性和核感知器算法中,对不可分问题,常用的策略是设置最大迭代次数来终止计算. 这时解的性质是不确定的,甚至是最坏的解,即正确分类的样本数最少[2]. Gallant[3]改进了线性感知器算法,提出了口袋算法,其目标是找到一个使错分样本最少的解,即最优解. 口袋算法的基本思想是在感知器算法的计算过程中增加一个口袋权向量和阀值,用于存放正确运行次数最多的感知器权向量和阀值. 对于正数和有理数输入,Gallant[3]证明口袋算法的收敛定理. 1990 年 Gallant[4]提出了口袋算法的若干变型,其中带棘齿的口袋算法需要检查能否正确分类更多的训练样本. Muselli[5]进一步证明口袋算法收敛定理对所有实数输入仍然成立,并且指出带棘齿的口袋算法能以概率 1 在有限步迭代内找到最优解. 本文主要研究带棘齿的口袋算法[4],并简单地称之为口袋算法.

与感知器算法相比,口袋算法有二个特点:第一,样本必须随机选取;第二,增加了适当的检查,以防止解的性质在迭代过程变坏(棘齿是一种机械装置,防止齿轮倒转,这里的含义是防止迭代过程中的解变坏). 尽管口袋算法仍是线性学习机器,在理论和实际计算上都证明其性能优于线性感知器算法. 核感知器算法的分类能力强于线性感知器算法,但仍只能处理特征空间中线性可分问题,对不可分问题面临同样的问题,强制算法终止,解的性质不确定.

为了改进口袋算法[2~4]和核感知器算法[12],本文用口袋算法的思想改进核感知器算法(或者说用核思想改造线性口袋算法),提出基于核函数的非线性口袋算法,简称核口袋算法,如下所述,其中 $\alpha=[\alpha_1,\alpha_2,\cdots,\alpha_l]$. 类似于线性口袋算法,核口袋算法通过在核感知器算法中增加适当的检查,来改善核感知器算法的性能.

算法(核口袋算法):

输入:训练样本集$\{(x_1,y_1),\cdots,(x_l,y_l)\}$,最大迭代次数.

输出:系数向量和阀值 α^*, β^*. 临时变量: α, β:迭代过程的系数向量和阀值; n: α, β能够连续正确分类样本的次数; n^*: α^*, β^*能够连续正确分类样本的次数; m: α, β能够正确分类的样本数; m^*: α^*, β^*能够正确分类的样本数.

第一步　置所有的临时变量为零.

第二步　随机地选取一个样本(x_j,y_j).

第三步　如果 α, β能够正确分类这一样本,则,

3a: $n=n+1$

3b:如果 $n>n^*$,则

3ba:计算 α, β能够正确分类的样本数 m.

3bb:如果 $m>m^*$,则

3bba: $\alpha^*\Leftarrow\alpha$, $\beta^*\Leftarrow\beta$

$n^*\Leftarrow n$, $m^*\Leftarrow m$

3bbb:如果所有的样本都被正确分类,算法结束.

否则

3c:按下列公式更新权向量和阀值 α, β:

$\alpha_i=\alpha_i+y_jk(x_i,x_j)$

$\beta=\beta+y_j$

3d:置 $n=0$

第四步　如果没有达到最大迭代次数,转向第二步. 否则结束算法.

4　核口袋算法的收敛性

线性口袋算法的拓扑结构与线性感知器相同,如图 1 所示. 关于线性棘齿口袋算法有下列收敛定理.

口袋算法收敛定理(带棘齿)　如果训练样本集是有限的,则口袋算法是有限最优的,即口袋算法在有限步迭代后以

概率 1 找到最优解(错分训练样本最少的解)[5].

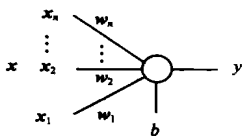

图 1 感知器和口袋算法的拓扑结构图

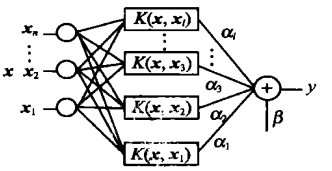

图 2 核感知器和核口袋算法的拓扑结构图

核感知器和核口袋算法也具有同样的拓扑结构(如图 2 所示). 比较图 1 与图 2 后,可以发现图 2 的后半部分就相当于一个感知器,其输入样本是一个 l 维向量 $[k(x, x_1), k(x, x_2), \cdots, k(x, x_l)]^T$,权向量和阈值为 α, β. 这样,对应于原始样本集(1),变换后的训练样本集为:

$$\{(z_1, y_1), \cdots, (z_l, y_l)\} \tag{6}$$

其中 $z_i = [k(x_i, x_1), k(x_i, x_2), \cdots, k(x_i, x_l)]^T$

核口袋算法收敛定理(带棘齿) 如果训练样本集是有限的,则核口袋算法是有限最优的,即核口袋算法在有限步迭代后以概率 1 找到最优系数向量和阈值 α, β,使错分样本最少.

这一定理的证明与线性情况类似,只需将变换后的样本集作为线性算法定理的输入.

5 实验结果与分析

为了评价和比较感知器、核感知器、口袋和核口袋算法的性能,本文从网上下载的八个数据集(www. first. gmd. de/ ～raetsch). 表 1 说明八个数据集的基本情况,包括样本的维数、训练样本集的样本数、测试样本集的样本数、样本集的数目.

表 1 八个数据集的基本情况

数据集名称	样本维数	训练集的样本数	测试集的样本数	样本集的数目
Banana	2	400	4900	100
Breast Cancer	9	200	77	100
Diabetis	8	468	300	100
German	20	700	300	100
Heart	13	170	100	100
Splice	60	1000	2175	20
Thyroid	5	140	75	100
Titanic	3	150	2051	100

表 2 感知器、口袋算法、核感知器、核口袋算法及 SVM 算法在测试集上的平均错误率及方差

数据集名称	感知器	口袋算法	核感知器	核口袋算法	SVM 算法
Banana	48.7 ±5.7	40.4 ±1.4	14.2 ±3.7	11.2 ±0.7	11.5 ±0.7
Breast Cancer	35.3 ±8.1	28.8 ±4.1	32.0 ±11.4	26.5 ±5.0	26.0 ±4.7
Diabetis	31.2 ±4.8	24.5 ±1.9	31.6 ±8.7	24.2 ±1.8	23.5 ±1.7
German	31.5 ±3.4	26.4 ±2.6	32.6 ±11.0	24.1 ±2.3	23.6 ±2.1
Heart	22.4 ±5.4	18.4 ±3.6	22.8 ±7.6	17.4 ±3.3	16.0 ±3.3
Splice	21.4 ±4.5	17.4 ±0.9	15.2 ±4.1	12.1 ±0.8	10.9 ±0.7
Thyroid	14.2 ±5.7	8.9 ±2.9	4.6 ±2.5	4.6 ±2.1	4.8 ±2.2
Titanic	34.8 ±15.6	22.9 ±1.3	32.2 ±14.5	22.4 ±1.1	22.4 ±1.0

本文用这八个数据集来评价和比较感知器、核感知器、口袋和核口袋算法的性能,其指标是测试集错误率的平均值和方差,计算结果如表 2 所示,其中 RBF 核函数参数的选取参考了 Mika 等人 SVM 算法的参数(www. first. gmd. de/ ～raetsch). 表 2 同时给出了 SVM 方法的结果[11]. 从表 2 可以看出,对不可分问题,核感知器算法并不能保证其性能一定优于感知器算法,表中三个数据集(Diabetis、German 和 Heart)的错误率反而有所增大;在所有的八个数据集上,线性口袋算法的错误率明显小于线性感知器算法;核口袋算法的错误率明显小于口袋算法和核感知器算法;在四种算法中,核口袋算法的性能是最好的.

因为本文在 RBF 核的参数选取中参考了 SVM 算法的 RBF 参数,重点比较核口袋算法与 SVM 算法的结果,有三个数据集(Banana, Thyroid, Titanic)(表中黑体)的错误率,核口袋算法小于等于 SVM;二者之间的错误率差小于 1 %,有三个数据集(Breast Cancer, Diabetis, German);还有二样本集错误率的差为 1.2 %(Splice)和 1.4 %(Heart). 这些实验结果说明核口袋算法的性能非常接近于 SVM.

6 讨论和结论

线性感知器算法在模式识别和神经网络的发展历史中具有重要的地位,它的局限性是只能处理线性可分问题. 口袋算法改善了感知器算法的性能,可以处理线性不可分问题,其目标是找到一个使错分样本最少的解. 根据 SVM 等方法利用核函数构造非线性算法的思想,核感知器算法构造了基于核函数的非线性感知器算法,使其能够处理特征空间的可分、原始属性空间的非线性问题,提高了感知器的处理分类问题的能力. 但是,对线性不可分问题,任何终止策略都无法确保感知器算法的性能. 本文提出了核口袋算法,其目标是找到一个使错分样本最少的解,从而改善了口袋算法和核感知器算法的性能. 八个数据集的计算结果表明:核口袋算法的性能优于感知器算法、口袋算法和核感知器算法,接近于 SVM 的性能.

对所有的基于核函数的算法,选取合适的核函数所面临的一个问题,我们正在研究如何根据实际样本数据确定核函数的参数或直接确定核函数矩阵. 另一项工作是将 SVM 中大间隔的思想融合到核口袋算法中,进一步提高其性能.

参考文献:

[1] Rosenblatt F. The perceptron: probabilistic model for information storage and organization in the brain [J]. Psychological Review, 1958, 65 (6): 386 - 408.

[2] Gallant, S I. Neural Networks Learning and Expert Systems [M]. Cambridge, MA: MIT Press, 1993.

[3] Gallant, S I. Opitmal linear discriminant [A]. Proc. Eighth Int Conf Pattern Recognition [C]. Paris France: EICPR, 1986. 849 - 853.

[4] Gallant, S I. Perceptron-based learning algorithm [J]. IEEE Transactions on Neural Networks, 1990, 1 (2): 179 - 191.

[5] Muselli M. On convergence properties of pocket algorithm [J]. IEEE Transactions on Neural Networks, 1997, 8 (3): 623 - 629.

[6] Cortes, C, & Vapnik, V N. Support vector networks [J]. Machine

Learning ,1995 ,20(3) :273 - 297.

[7] Vapnik ,V N. The Nature of statistical learning theory [M]. New York: Springer-Verlag ,1995.

[8] 张学工. 关于统计学习理论与支持向量机 [J]. 自动化学报, 2000 ,26(1) :32 - 44.

[9] Aizerman ,M A ,Braverman , E M , & Rozonoer ,L I. Theoretical foundations of the potential function method in pattern recognition learning [J]. Automation and Remote Control ,1964 ,25 :821 - 837.

[10] Scholkopf ,B ,Smola ,A , & Muller , K - R. Nonlinear component analysis as a kernel eigenvalue problem [J]. Neural Computation ,1998 ,10 (5) :1299 - 1319.

[11] Mika ,S , Ratsch , G, Weston ,J , Scholkopf ,B , & Muller , K - R. Fisher discriminant analysis with kernels [A]. Neural Networks for Signal Processing IX [C]. New York: IEEE Press ,1999. 41 - 48.

[12] 许建华,张学工,李衍达. 一种基于核函数的非线性感知器算法 [J]. 计算机学报,2002 ,25(7) :1 - 7.

作者简介:

许建华　男,1962 年生于浙江省长兴县,1985 年毕业于成都地质学院应用地球物理系,获得工学学士学位,1987 年毕业于中国科技大学地球和空间科学系,获得理学硕士学位,现为清华大学自动化系博士研究生,主要从事信号处理、模式识别、神经网络、机器学习等领域的理论和应用研究.

张学工　1965 年生于山东青州,1994 毕业于清华大学自动化系,获得工学博士学位. 现为清华大学自动化系教授,清华大学自动化系信息处理研究所所长,主要从事生物信息学、模式识别、神经网络、统计学习理论研究.

李衍达　男,1936 年出生于广东东莞,1959 年毕业于清华大学自动控制系,中国科学院院士,清华大学自动化系教授,清华大学信息科学与技术学院院长,清华大学生物信息学研究所所长,校学术委员会主任,主要从事信息处理与生物信息学研究.