

基于 EDA-DP 的 LS-SVM 参数优化选择

熊 杨^{1,2}, 肖怀铁¹, 王 伟²

(1. 国防科学技术大学电子科学与工程学院, 长沙 410073; 2. 中国人民解放军 63636 部队, 兰州 732750)

摘 要: 通过分析最小二乘支持向量机(Least Square Support Vector Machine, LS-SVM)模型的超参数选择对分类器的影响, 提出一种采用多样性保持的分布估计算法(EDA-DP)优化选择 LS-SVM 模型参数的方法。使用基于 EDA-DP 的 LS-SVM 分类器模型对基准数据集和雷达目标高分辨距离像数据集进行仿真实验, 结果表明, 该模型相比基于网格法的分类器模型, 平均识别率分别提高了 4.2% 和 1.76%, 具有更好的分类性能和泛化能力。

关键词: 最小二乘支持向量机; 模型参数优化选择; 多样性保持的分布估计算法; 目标识别; 径向基函数

Optimal Selection of LS-SVM Parameter Based on EDA-DP

XIONG Yang^{1,2}, XIAO Huai-tie¹, WANG Wei²

(1. School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China;

2. Unit 63636 of PLA, Lanzhou 732750, China)

【Abstract】 By analyzing the impacts of Least Square Support Vector Machine(Least Square Support Vector Machine, LS-SVM) model hyperparameter selection on the classifier, this paper proposes a method using estimation of distribution algorithms with diversity preservation named EDA-DP to optimally select model parameters of LS-SVM. Experiments are operated to recognize the benchmarks and radar High Range Resolution Profile(HRRP) datasets by using LS-SVM classifier. Compared to the grid-based method, the average recognition rate of LS-SVM classifier based on EDA-DP are increased by 4.2% and 1.76%. Experimental results demonstrate that the classifier model with EDA-DP achieves better classification ability and generalization capacity.

【Key words】 Least Squares Support Vector Machine(Least Square Support Vector Machine, LS-SVM); optimal selection of model parameters; Estimation of Distribution Algorithm with Diversity Preservation(EDA-DP); target recognition; Radial Basis Function(RBF)

DOI: 10.3969/j.issn.1000-3428.2011.14.048

1 概述

支持向量机(Support Vector Machine, SVM)是一种基于间隔最大化原理及结构风险最小化原理的特殊分类器, 其着重研究小样本情况下的分类问题, 具有优越的泛化性能^[1]。SVM 使用核方法解决非线性问题, 通过映射函数 ϕ 将训练样本非线性地映射到一个更高维的特征空间, 从而建立能最大化分类间隔的分类超平面。

最小二乘支持向量机(Least Squares Support Vector Machine, LS-SVM)^[2]是 SVM 的扩展。LS-SVM 采用等式约束条件代替标准 SVM 的不等式约束, 从而将标准 SVM 的二次规划问题转换为一个线性问题, 得到的分类模型既简化了 SVM 的表达式, 又保持了 SVM 原有的优点。

LS-SVM 同样基于间隔最大化原理及结构风险最小化原理, 与标准 SVM 一样具有良好的泛化性能。其性能不仅与给定的训练数据及测试数据有关, 还取决于模型超参数(规范参数与核参数)的精确选择。因此, 在基于 LS-SVM 的目标识别中, 需要通过优化选择模型参数来提高分类器的性能。传统凭经验选择模型参数时, 要求对所研究的问题有很好的把握, 而且其具有很大的随意性, 获得合理的参数难度较大。网格法^[3]和交叉验证法^[4]是应用较多的传统 SVM 参数选择算法。算法从给出的初始条件出发, 在给定的参数选择范围内采用网格式搜索方式寻找所有网格点形成的参数矩阵中的最优参数对。网格法容易实现, 但当待选参数在 2 个以上或解决大样本问题时, 处理难度较大, 计算量很大^[5]。文献[4]提出一种自动选择 SVM 超参数的方法, 使用特定的标准估计

留一法误差值, 采用梯度下降法优化选择 SVM 参数。另外, 文献[5-6]将遗传算法引入 SVM 的参数选择中, 取得了较好的效果。本文通过分析 LS-SVM 规范参数 γ 和核参数的选择对分类器的影响, 提出一种利用多样性保持的分布估计算法 EDA-DP(Estimation of Distribution Algorithm with Diversity Preservation)优化选择 LS-SVM 模型参数的方法。

2 模型参数选择对 LS-SVM 的影响

LS-SVM 有 2 类超参数: 规范参数 γ 及核参数^[7]。在给定的训练样本及测试样本的情况下, 模型参数的选择极大地影响着 LS-SVM 的分类性能及泛化能力。

考虑二分类问题, 设训练样本集为:

$$\{x_k, y_k\}_{k=1}^N \in \mathbf{R}^m \times \{+1, -1\}$$

根据 LS-SVM 分类器原始优化问题, 可定义 Lagrange 函数为^[2]:

$$L(w, b, e, \alpha) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^N e_k^2 - \sum_{k=1}^N \alpha_k \{y_k [w^T \phi(x_k) + b] - 1 + e_k\} \quad (1)$$

其中, $\phi(x)$ 为原始空间到特征空间的映射; γ 为规范系数; e_k 为松弛变量; α_k 为 Lagrange 算子。

根据 KKT(Karush-Kuhn-Tucker)条件可得:

基金项目: 国防科学技术大学重点实验室基金资助项目

作者简介: 熊 杨(1981—), 男, 工程师、硕士研究生, 主研方向: 自动目标识别, 模糊工程; 肖怀铁, 教授、博士生导师; 王 伟, 工程师、硕士

收稿日期: 2010-12-19 **E-mail:** xyang007@126.com

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{k=1}^N \alpha_k y_k \varphi(x_k) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{k=1}^N \alpha_k y_k = 0 \\ \frac{\partial L}{\partial e_k} = 0 \Rightarrow \alpha_k = \gamma e_k \\ \frac{\partial L}{\partial \alpha_k} = 0 \Rightarrow y_k [w^T \varphi(x_k) + b] - 1 + e_k = 0 \end{cases} \quad (2)$$

将式(2)代入式(1), Lagrange 函数可以表示为:

$$L(w, b, e, \alpha) = \sum_{k=1}^N \alpha_k - \frac{1}{2} \sum_{k,l=1}^N \alpha_k \alpha_l y_k y_l \langle \varphi(x_k) \cdot \varphi(x_l) \rangle - \frac{1}{2\gamma} \langle \alpha \cdot \alpha \rangle = \sum_{k=1}^N \alpha_k - \frac{1}{2} \sum_{k,l=1}^N \alpha_k \alpha_l y_k y_l K(x_k, x_l) - \frac{1}{2\gamma} \langle \alpha \cdot \alpha \rangle \quad (3)$$

因此, 在 α 上最大化上述目标函数等价于最大化:

$$W(\alpha) = \sum_{k=1}^N \alpha_k - \frac{1}{2} \sum_{k,l=1}^N \alpha_k \alpha_l y_k y_l \left(K(x_k, x_l) + \frac{1}{\gamma} \delta_{kl} \right) \quad (4)$$

其中, $\delta_{kl} = \begin{cases} 1 & \text{if } k=l \\ 0 & \text{otherwise} \end{cases}$ 。令:

$$K'(x, z) = K(x, z) + \frac{1}{\gamma} I \quad (5)$$

其中, I 为单位矩阵, 则可以将参数 γ 当成核函数的一个参数, 式(4)可以表示为:

$$W(\alpha) = \sum_{k=1}^N \alpha_k - \frac{1}{2} \sum_{k,l=1}^N \alpha_k \alpha_l y_k y_l K'(x_k, x_l) \quad (6)$$

下面将对 LS-SVM 规范参数 γ 及核参数进行讨论, 研究它们对模型分类性能的影响。

2.1 参数 γ 对分类器的影响

在标准 SVM 中, 惩罚系数用来控制训练误差最小化和分类间隔最大化之间的平衡, 即实现错分样本比例与算法复杂度之间的折中, 在特征空间中调节学习机器经验风险和置信范围, 使学习机器的泛化能力达到最好^[8]。LS-SVM 中的规范参数 γ 与标准 SVM 中的惩罚系数具有相似的作用。由式(1)可以看出, γ 值越小, 表示对样本错分的惩罚越小, 同时, 式(4)中最大化 $W(\alpha)$ 得到的值也越小, 此时学习器的算法复杂度较小而经验风险较大; 当 γ 足够大时, $(1/\gamma)\delta_{ij} \rightarrow 0$, 分类器变为硬间隔分类器, γ 对分类的影响基本恒定, 经验风险和推广能力几乎不再变化。

2.2 核参数对分类器的影响

为了解决非线性问题, LS-SVM 和 SVM 引入了核函数, 利用核函数 $K(x, x_k)$ 代替原空间中的内积, 将数据通过映射函数 φ 映射到高维特征空间 F 中, 将原空间中的非线性问题转换为特征空间中的线性可分问题。核函数不同, 确定的映射函数和特征空间也不同, 改变其类型或参数值将会改变映射函数及特征空间的维数。特征空间的维数决定了能在此空间中构造的线性分类面的最大 VC(Vapnik & Chervonenkis) 维, 也就决定了线性分类面能达到的最小经验误差^[8]。VC 维越大, 可能得到的最小经验误差就越小, 而同时, 特征空间对应的最优分类超平面就越复杂, 模型置信范围也越大, 因此, 实际风险可能仍较大; 反之, 当 VC 维较小时, 置信范围小而经验误差大, 实际风险同样较大。上述 2 种情况得到的 SVM 模型泛化能力可能都不高。因此, 只有选择合适的核函数及核参数值, 才能得到最小的经验误差和置信范围组合, 从而得到泛化能力良好的分类器。

常用的核函数有多项式核函数、径向基函数(Radial Basis Function, RBF)及 Sigmoid 核函数。RBF 核函数是一个普适的

核函数, 具有广泛的代表性。本文选择 RBF 函数^[9]:

$$K(x, x_k) = \exp\left\{-\|x - x_k\|_2^2 / \sigma^2\right\}$$

作为研究对象, 研究其参数 σ 对 LS-SVM 分类器的影响。

图 1、图 2 分别为 LS-SVM 分类器识别错误率随规范参数 γ 和 RBF 核参数 σ^2 的变化情况。可以看出, 只有选择适当的 γ 、 σ 参数组合时, LS-SVM 才能得到最优的分类性能和泛化能力。仿真采用基准数据集的 Waveform 数据, 分别为 100 个训练样本及 300 个测试样本。图 1 中参数 γ 及图 2 中参数 σ^2 的取值均采用初始值为 0.01、步长为 1.5 的等比数列组成的数据矩阵。

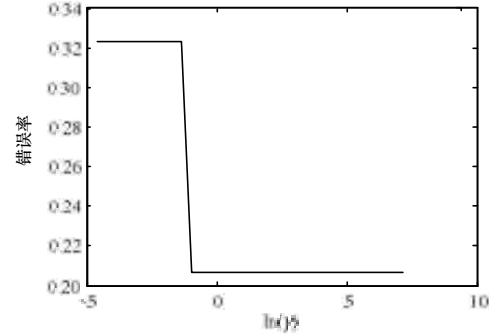


图 1 错误率随 γ 的变化情况

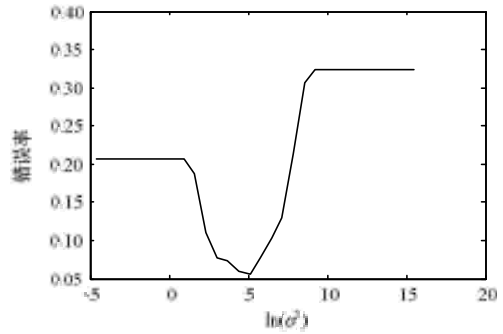


图 2 错误率随 σ^2 的变化情况

3 基于 EDA-DP 算法的 LS-SVM 参数优化

EDA-DP 算法^[10]是传统分布估计算法的扩展。EDA 是一种新型优化算法, 其在生物进化的宏观层面上进行数学建模, 具有良好的全局搜索能力, 相对于遗传算法能更有效地解决高维问题, 降低时间复杂性; 但 EDA 局部优化能力较差, 容易出现早熟现象^[11]。EDA-DP 在 EDA 中加入混沌变异操作, 利用混沌运动的随机性、遍历性、初值敏感性和规律性等特性, 在解空间内进行优化搜索, 使之快速跳出局部极值区, 找到全局最优解。EDA-DP 优化算法^[10]流程如图 3 所示。

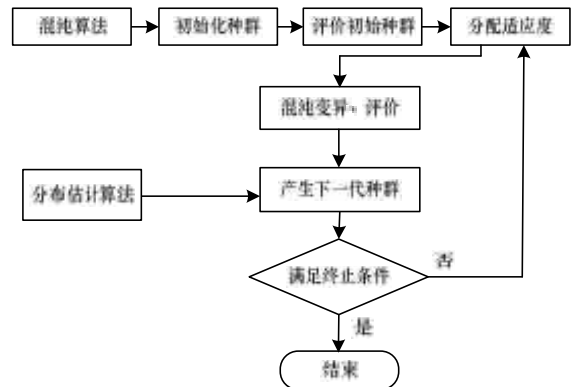


图 3 EDA-DP 算法流程

本文利用 EDA-DP, 在合适的评价函数及适应度函数的基础上,对 LS-SVM 分类器参数进行优化选择,步骤如下:

(1)确定 LS-SVM 参数 (γ, σ^2) 的搜索空间,选定目标识别的训练数据集与测试数据集,设置算法参数。

(2)利用 Logistic 混沌模型初始化种群 $X=[\gamma, \sigma^2]$ 。

(3)评价初始种群。训练种群中的每个个体(即每组参数)得到其 LS-SVM 模型,以错误率作为评价函数:

$$f = N_e / N_t$$

其中, N_e 为错误样本数; N_t 为总样本数。

(4)分配个体适应度。

(5)对种群执行混沌变异。

(6)采用第(3)步中的评价函数评价变异后的种群,用评价函数值较小的个体代替旧个体。

(7)用分布估计算法产生下一代个体。

(8)判断是否满足收敛条件,若满足,结束优化算法,选择最优个体作为 LS-SVM 模型参数;若不满足,转第(4)步继续执行。

(9)利用算法优化选择的 γ 、 σ^2 参数值训练得到 LS-SVM 模型,形成最优分类器。

4 实验及结果分析

本文使用基准数据集及雷达多目标高分辨距离像数据集 HRRP(High Range Resolution Profile)的多组数据进行实验,验证将 EDA-DP 算法应用于 LS-SVM 参数选择的有效性 & 实用性。仿真采用 LS_SVMlab 工具箱,在 Matlab R2007a 环境下实现。

4.1 基准数据集实验

基准数据集包括多种不同类别的数据,本文选择其中的 7 种数据用来测试基于 EDA-DP 算法的 LS-SVM 模型参数优化选择方法的性能。表 1 显示了采用 EDA-DP 算法及网格法在基准数据集上得到的模型选择结果,包括 2 种算法优化得到的 γ 和 σ^2 值以及目标正确识别率。可以看出,采用 EDA-DP 方法得到的 LS-SVM 训练模型对基准数据集的平均识别率高于网格法,验证了本文方法在分类器模型参数选择方面的有效性。

表 1 基准数据集的 LS-SVM 参数优化选择结果

数据	训练 数据 数目	测试 数据 数目	EDA-DP			网格法		
			γ	σ^2	识别率 /(%)	γ	σ^2	识别率 /(%)
Banana	400	600	374.83	27.74	84.50	106.62	1.04	75.17
Breast cancer	150	113	461.67	50.89	72.57	0.54	16.96	71.68
Diabetis	468	300	274.80	87.91	72.00	1.89	31.72	69.67
German	300	200	307.70	36.11	62.50	0.55	13.48	63.50
Heart	170	100	487.48	9.71	79.00	15.79	14.72	77.00
Thyroid	140	75	234.33	37.03	100.00	1.68	2.31	86.67
Waveform	400	600	79.98	92.49	87.17	129.70	20.17	84.67
平均识别率/(%)			79.68			75.48		

4.2 雷达目标 HRRP 数据实验

本文采用 4 种飞机目标的 HRRP 进行识别实验,4 种飞机分别为 B737、B747、F16 和 F18。每一种飞机数据有 171 个弹道点即 171 幅 HRRP,每幅 HRRP 有 151 个特征维

数。实验中抽取每一种数据的 1/2 作为训练数据,剩下的 1/2 作为测试数据。采用 EDA-DP 与网格法优化选择模型参数,所得 LS-SVM 模型的识别效果如表 2 所示。可以看出,采用本文算法优化选择分类器模型参数,得到的 LS-SVM 模型识别效果更为理想。

表 2 雷达目标 HRRP 数据的 LS-SVM 分类器识别效果

目标 类型	EDA-DP			网格法		
	γ	σ^2	识别率/(%)	γ	σ^2	识别率/(%)
B737	269.15	1.46	87.06	69.37	1.46	87.06
B747	53.48	51.06	97.64	8.74	43.38	97.64
F16	55.44	9.87	92.94	48.17	3.77	89.41
F18	86.24	37.97	95.29	41.57	1.70	91.76
平均识别率/(%)			93.23	91.47		

5 结束语

基于支持向量机的目标识别性能不仅与给定的训练数据及测试数据有关,还取决于模型参数的精确选择。本文讨论了选择不同的规范参数和核参数对 LS-SVM 模型分类性能的影响,证明只有选择合适的参数组合,才能得到最优的模型分类性能和泛化能力;通过对基准数据集及雷达目标 HRRP 数据的识别实验,验证基于 EDA-DP 算法得到的 LS-SVM 模型具有良好的分类性能,得到了较高的目标正确识别率,说明本文基于 EDA-DP 的 LS-SVM 参数选择方法具有有效性 & 实用性。

参考文献

[1] Vapnik V. The Nature of Statistical Learning Theory[M]. New York, USA: Springer, 1995.

[2] Suykens J A K, Vandewalle J. Least Squares Support Vector Machine Classifiers[J]. Neural Processing Letters, 1999, 9(3): 293-300.

[3] 李京华, 张聪颖, 倪 宁. 基于参数优化的支持向量机战场多目标声识别[J]. 探测与控制学报, 2010, 32(1): 1-5.

[4] Chapelle O, Vapnik V. Choosing Multiple Parameters for Support Vector Machines[J]. Machine Learning, 2002, 46(1-3): 131-159.

[5] 王克奇, 杨少春, 戴天虹, 等. 采用遗传算法优化最小二乘支持向量机参数的方法[J]. 计算机应用与软件, 2009, 26(7): 109-111.

[6] Yang Hong, Lou Fei, Xu Yuge, et al. GA Based LS-SVM Classifier for Waste Water Treatment Process[C]//Proc. of the 27th Chinese Control Conference. Kunming, China: [s. n.], 2008: 436-439.

[7] Adankon M M, Cheriet M. Model Selection for the LS-SVM. Application to Handwriting Recognition[J]. Pattern Recognition, 2009, 42(12): 3264-3270.

[8] 刘东辉, 卞建鹏, 付 平, 等. 支持向量机最优参数选择的研究[J]. 河北科技大学学报, 2009, 30(1): 58-61.

[9] 刘 胜, 傅荟璇, 王宇超. 基于分割区间 LS-SVM 的摄像机标定[J]. 计算机工程, 2009, 35(24): 179-181.

[10] 王雪松, 程玉虎. 机器学习理论、方法及应用[M]. 北京: 科学出版社, 2009: 142-158.

[11] 周树德, 孙增圻. 分布估计算法综述[J]. 自动化学报, 2007, 33(2): 113-124.

编辑 张 帆