

基于鸟群算法的 SVM 参数选择

肖海军, 卢常景*, 何凡

(中国地质大学 数学与物理学院, 武汉 430074)

摘要 针对支持向量机(SVM)分类器参数选择问题,提出了基于鸟群算法(BSA)的SVM参数选择方法(BSA-SVM),以优化SVM惩罚参数和核参数.鸟群算法具有优化精度高、鲁棒性好等特点,将SVM参数作为鸟群算法目标函数的优化参数,在搜索到最优值的同时得到最优参数.通过8个UCI标准数据集的MATLAB仿真对比实验,验证了BSA-SVM能有效提高分类准确性.实验结果表明:BSA-SVM能更加准确地找到SVM最优参数,从而加强SVM学习与泛化能力,是一种有效的SVM参数优化方法.

关键词 鸟群算法;支持向量机;参数选择

中图分类号 O235 文献标识码 A 文章编号 1672-4321(2017)03-0090-05

Parameter Optimization of Support Vector Machine Based on Bird Swarm Algorithm

Xiao Haijun, Lu Changjing, He Fan

(School of Mathematics & Physics, China University of Geosciences, Wuhan 430074, China)

Abstract According to the problem of selecting parameters in support vector machine(SVM), a parameter selection method(BSA-SVM) based on bird swarm algorithm(BSA) for optimizing the SVM penalty parameter and kernel parameter is proposed. BSA is a new optimization algorithm which possesses the characteristic of high precision and good robustness. Hence, parameters selection problem of SVM can be regarded as a combinatorial optimization problem by setting the objective function in BSA, which can obtain the best value and the optimum parameters at the same time. The simulation comparison experiments on 8 UCI datasets illustrate that the BSA-SVM effectively improves the classification accuracy of SVM. It is shown that the BSA-SVM can find the optimal parameters more accurately and strengthen the learning and generalization ability of SVM, so it's an effective parameter optimization method of SVM.

Keywords bird swarm algorithm(BSA); support vector machine(SVM); parameter selection

1995年,Vapnik等在基于统计学VC维理论和结构风险最小化原理的基础上提出了支持向量机^[1](SVM).此后,SVM在解决小样本、非线性和高维识别问题方面展现出了较强的优势,现今已被广泛应用于模式识别、回归估计等领域^[2-4].

然而,SVM作为一种新颖的机器学习算法,它还有许多需要改进的地方.其中如何合理地选择参数便是关键问题之一.因为参数取值将直接影响到预测性能的好坏.由于缺乏一定的理论指导,早期的参数选择往往依据网格搜索法^[5]、梯度下降法^[6]等,这些方法的特点是原理简单,实现容易,但存在着搜索精度

不高、计算量大等问题.最近十几年来利用优化算法优化SVM参数也成为一个研究热点,例如Peng提出了基于粒子群算法(PSO)的SVM参数选择方法^[7],庄严提出了基于蚁群算法(ACO)的SVM参数选择方法^[8],魏峻提出了基于蝙蝠算法(BA)的SVM参数选择方法^[9],它们都能有效地寻找到合适的参数,从而提高SVM分类预测性能.这几种优化算法都存在着相应的特点:PSO算法收敛速度快,全局性能优良,但是局部性能较差,很容易陷入局部最优;ACO算法具有较强的鲁棒性,但收敛速度慢,搜索过程中容易出现停滞;BA全局性能好,但搜索后期收敛精度差、收

收稿日期 2017-05-06 * 通讯作者 卢常景 研究方向:数据挖掘 E-mail: luchangjing@cug.edu.cn

作者简介 肖海军(1965-)男,教授,博士,研究方向:数据挖掘 E-mail: Xiaohj@cug.edu.cn

基金项目 国家自然科学基金资助项目(11301492);中国地质大学(武汉)基础研究基金项目(CUGL140420)

收敛速度慢且容易陷入局部极值. 鸟群算法(BSA)作为一种全新的仿生优化算法, 相对于其他优化算法具有优化精度高、稳定性强、收敛速度快等优点, 能较好地克服传统优化算法容易陷入局部最优的缺陷^[10]. 因此, 采用鸟群算法来进行 SVM 参数选择也将会是一种行之有效的办法.

1 SVM 分类器

1.1 SVM 分类原理

在二分类问题中, 设样本集为 $\{(x_i, y_i) \mid i = 1, 2, \dots, l\}$, 其中 $x_i \in R^l$ 表示输入变量, $y_i \in \{+1, -1\}$ 表示输出变量, l 为样本集个数. 通过引入非线性映射函数 $\varphi(x)$ 将处于低维空间的输入变量映射到高维空间, 可以构造出一个最优分类超平面:

$$f(x) = \omega \cdot \varphi(x) + b = 0, \quad (1)$$

其中 ω 表示权重向量, b 表示最优超平面位移.

为了最小化结构风险, 构造出的最优超平面需满足以下约束条件:

$$y_i \cdot (\omega \cdot \varphi(x_i) + b) \geq 1 \quad i = 1, 2, \dots, l. \quad (2)$$

若构造的最优分类超平面仍存在少量分类错误, 那么可以引入非负松弛变量来量 ζ_i , 化分类器的这种分类误差. 此时分类超平面最优化问题可表述为:

$$\min \frac{1}{2} \omega \cdot \omega + C \sum_{i=1}^l y_i (\omega \cdot \varphi(x_i) + b) \geq 1 - \zeta_i, \quad (3)$$

$$\zeta_i \geq 0 \quad i = 1, 2, \dots, l,$$

其中 C 为惩罚系数, 表示对错误分类的惩罚代价, C 越大, 惩罚代价越大.

对于该二次规划问题, 采用拉格朗日乘子法将其转化为对偶形式:

$$\min \sum_{i=1}^l \alpha_i + \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j (\varphi(x_i) \cdot \varphi(x_j)), \quad (4)$$

$$0 \leq \alpha_i \leq C, \sum_{i=1}^l \alpha_i y_i = 0.$$

在(4)中引入核函数, 可转化为:

$$\min \sum_{i=1}^l \alpha_i + \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j), \quad (5)$$

其中 $K(x_i, x_j)$ 称为核函数 $K(x_i, x_j) = \varphi(x_i) \cdot \varphi(x_j)$, 最终得到分类超平面:

$$f(x) = \text{sgn} \left(\sum_{i=1}^l \alpha_i y_i K(x_i, x) + b \right), \quad (6)$$

采用不同的核函数会生成不同的 SVM 分类器, 目前 SVM 中常见的核函数主要有: 线性函数 $K(x_i, x_j) = x_i \cdot x_j$, 多项式函数 $K(x_i, x_j) = (x_i \cdot x_j + 1)^d$, Sigmoid 函数 $K(x_i, x_j) = \tanh(\gamma x_i \cdot x_j + r)$, $\gamma > 0$, 径向基函数

$K(x_i, x_j) = \exp(-\|x - x_i\| / 2\sigma^2)$. 其中径向基函数只需要确定一个核函数, 并且相关研究表明, 它是一种在分类问题上表现优异的核函数^[11], 因而本文选择径向基函数作为 SVM 核函数.

1.2 SVM 参数

SVM 参数的选择对分类器模型的学习能力和预测能力都有重要影响, 因此参数选择是 SVM 一个关键性问题. 对于选用径向基核函数的 SVM 分类器, 其参数的选取包括惩罚参数 C 与核宽度 σ . 惩罚参数 C 用于调节 SVM 分类器置信范围与经验风险之间的比例, 以提高其推广能力. C 取值过小时称为“欠学习”, 表示对经验误差的惩罚小, 此时数据拟合程度低. C 取值过大时称为“过学习”, 表示对经验误差的惩罚大, 此时数据拟合程度高, 但泛化能力会降低. 核宽度 σ 的取值大小跟学习样本的输入空间成正相关关系, 它体现了特征子空间分布的情况. 研究显示, 较小的 σ 值将导致学习样本过度拟合, 而较大的值将导致学习样本欠拟合, 而这两者都会影响 SVM 分类器泛化能力^[12]. 因此, 惩罚参数 C 与核宽度 σ 的取值过大或过小都会严重影响 SVM 分类器的性能.

2 基于鸟群算法的 SVM 参数选择

2.1 鸟群算法原理

鸟群算法(BSA)是 Meng 等^[10]于 2015 年提出的一种全新仿生优化算法. 它的灵感来源于鸟群的三大行为: 觅食行为、警戒行为和飞行行为. 为了形成该算法 5 个理想化的规则简化如下.

规则 1: 每只鸟都可以在警戒行为和觅食行为间随意切换, 是否搜寻食物或保持警戒被描述为一个随机策略.

规则 2: 当觅食时, 每只鸟都会及时记录并更新自身当前最佳经验和鸟群当前最佳经验, 并且这些信息能及时共享.

规则 3: 当处于警惕时, 每只鸟都会尝试移向群体中间, 这种行为会引发群体竞争而使自身受到影响. 拥有高储量食物的鸟比拥有低储量食物的鸟更可能靠近群体中心.

规则 4: 鸟群会定期飞往另一个地点. 当飞到另一个地点时, 每只鸟会在生产者与乞讨者之间切换身份. 拥有最高储量食物的鸟会成为生产者, 而拥有最低储量食物的鸟成为乞讨者, 食物储量介于这两者之间的鸟随机选择成为生产者或乞讨者.

规则 5: 生产者积极寻找食物, 而乞讨者会随机

跟从生产者去寻找食物。

假设鸟群种群规模为 N , 在时间点 t 时可由位置 $x_j^t (j = 1, 2, \dots, N)$ 来描述, 所处维度空间为 D 维。BSA 原理可描述如下。

觅食行为: 规则 1 可表示为一个随机策略。随机生成一个服从均匀分布的值, 如果该值小于 $P (P \in (0, 1))$, 其中 P 为恒定值, 那么那只鸟将会觅食, 否则它将保持警戒。规则 2 可以表述如下:

$$x_{ij}^{t+1} = x_{ij}^t + (p_{ij} - x_{ij}^t) \times T \times \text{rand}(0, 1) + (g_j - x_{ij}^t) \times S \times \text{rand}(0, 1), \quad (7)$$

其中 $j \in [1, \dots, D]$, $\text{rand}(0, 1)$ 表示随机生成一个服从 $(0, 1)$ 均匀分布的值。 T 和 S 是两个正数, 可分别称为认知加数系数和社会加速系数。 p_{ij} 是 i 第只鸟当前最佳位置, g_j 是鸟群当前最佳位置。

警戒行为: 规则 3 下, 由于相互竞争每只鸟都不会直接移向群体中间, 其位置更新公式可以表述如下:

$$x_{ij}^{t+1} = x_{ij}^t + A_1(\text{mean}_j - x_{ij}^t \times \text{rand}(0, 1)) + A_2(p_{ij} - x_{ij}^t \times \text{rand}(-1, 1)), \quad (8)$$

$$A_1 = \alpha_1 \times \exp\left(-\frac{pFit_i}{\text{sumFit} + \varepsilon} \times N\right), \quad (9)$$

$$A_2 = \alpha_2 \times \exp\left(\frac{pFit_i - pFit_k}{|pFit_k - pFit_i| + \varepsilon}\right) \times N \times \frac{pFit_k}{\text{sumFit} + \varepsilon}, \quad (10)$$

其中 $k (k \neq i)$ 为正整数, 是从 1 到 N 间随机选择的。 α_1 与 α_2 是介于 $(0, 2)$ 之间的两个常数, $pFit_i$ 表示第 i 只鸟最佳适应度值, sumFit 表示鸟群个体适应度值的总和。 ε 是最小常数, 以避免除 0 错误。 mean_j 表示鸟群在 j 维的平均位置。

飞行行为: 由规则 4, 鸟群个体会分为生产者和乞讨者。生产者和乞讨者的位置更新公式可表述如下:

$$x_{ij}^{t+1} = x_{ij}^t + \text{randn}(0, 1) \times x_{ij}^t, \quad (11)$$

$$x_{ij}^{t+1} = x_{ij}^t + (x_{ij}^t - x_{ij}^t) \times FL \times \text{randn}(0, 1), \quad (12)$$

其中 $\text{randn}(0, 1)$ 表示均值为 0, 标准差为 1 的高斯分布随机数, $k \in (1, \dots, N)$, $k \neq i$ 。 $FL (FL \in (0, 2))$ 表示乞讨者将跟从生产者去寻找食物。另外假设每只鸟会在 FQ 时间间隔内飞往另一个地方, FQ 为正整数。

2.2 BSA-SVM 模型

BSA-SVM 模型的基本思想为: 将惩罚参数 C 与核参数 σ 作为鸟群个体位置变量, 在这个二维空间上, 通过适应度函数来评价个体所处位置的好坏, 由鸟群活动时的三大行为: 觅食行为、警戒行为和飞行行为来更新位置信息, 直到迭代结束时获得鸟群最佳

位置信息, 最终输出最佳惩罚参数 C_{best} 与核参数 σ_{best} 。BSA-SVM 模型实现步骤可以描述如下。

步骤 1: 数据标准化。为消除量纲影响, 通过 $x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$ 对数据集进行标准化处理。其中 x 是原始值, x_{\min} 与 x_{\max} 分别为相应特征的最小值与最大值, x^* 为归一化后的值。

步骤 2: 初始化 BSA 相关参数。当 $t = 0$ 时, 在一定范围内随机生成鸟群个体初始位置 $x(i, 1)$, 范围为 (C_{\min}, C_{\max}) , $x(i, 2)$ 范围为 $(\sigma_{\min}, \sigma_{\max})$ 。同时设置鸟群种群规模 N 、迭代次数 M 、飞行频率 FQ 、觅食概率 P 和常量 $T, S, \alpha_1, \alpha_2, FL$ 。

步骤 3: 计算适应度值。根据适应度函数计算每只鸟的适应度值, 由适应度值确定个体自身最佳位置和群体最佳位置。适应度值计算是根据 $x(i, 1)$ 与 $x(i, 2)$ 进行得到相应值。计算方法为:

$$\begin{cases} \text{cmd} = [-v \cdot \text{num2str}(v), \quad \rho - c \cdot \text{num2str}(x(i, 1)) \\ \quad \quad \quad \rho - g \cdot \text{num2str}(x(i, 2))] \\ \text{fitness}(i) = \text{svmtrain}(\text{train_label}, \text{train_cmd}) \end{cases},$$

其中 v 表示交叉验证的次数, ρ 与 g 分别代表惩罚参数 C 与核参数 σ , train_label 为训练数据集标签, train 是训练数据集, $\text{fitness}(i)$ 表示第 i 只鸟当前适应度值。

步骤 4: 迭代寻优。迭代开始, 判断 $t\%FQ$ 是否有余数。

1) 若有余数, 则鸟群个体只进行觅食行为或警戒行为。采用规则 1 来判别鸟群个体进行哪种行为, 随机生成一个服从均匀分布的值, 若该值小于觅食概率 P , 则该个体发生觅食行为, 用公式 (7) 更新位置; 否则该个体保持警戒, 用公式 (8) 更新位置。

2) 若没有余数, 则将鸟群个体分为生产者和乞讨者。若为生产者, 则用公式 (11) 更新位置; 若为乞讨者, 则用公式 (12) 更新位置。

步骤 5: 更新个体最佳位置和群体最佳位置。若第 i 只鸟当前位置优于先前自身最佳位置, 则将当前位置更新为自身最佳位置, 同时更新鸟群当前最佳位置。

步骤 6: 判断是否达到最大迭代次数, 若迭代结束则转至步骤 7; 否则令 $t = t + 1$, 转回步骤 4。

步骤 7: 输出鸟群最佳适应度值及鸟群最佳位置所包含的惩罚参数 C_{best} 与核参数 σ_{best} 。

BSA-SVM 模型具体流程可归纳为图 1。

3 仿真实验

3.1 数据源

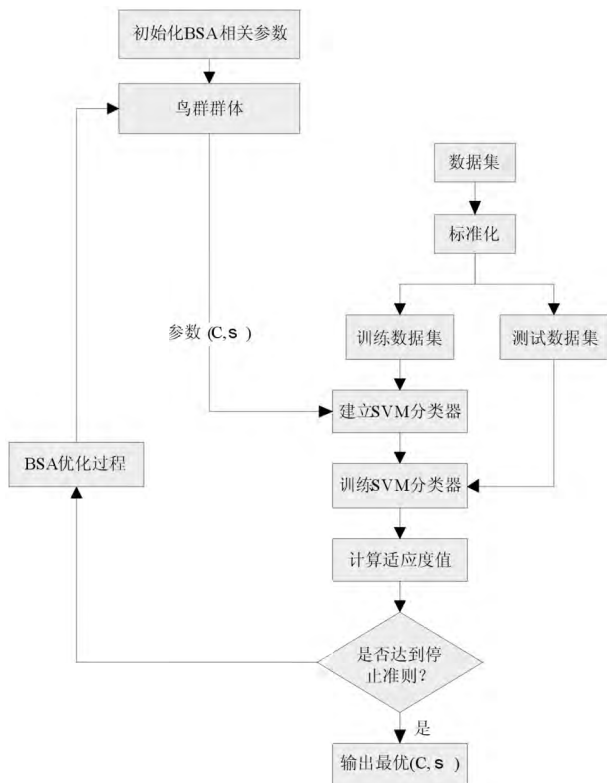


图 1 BSA-SVM 模型流程图

Fig.1 Flow chart of BSA-SVM model

为了验证 BSA-SVM 模型性能, 本文从 UCI 标准数据库(<http://archive.ics.uci.edu/ml/>) 选用了 8 个数据集, 与不做参数选择的 SVM 方法、基于粒子群算法的 SVM 参数选择方法^[9](PSO-SVM) 进行 MATLAB 仿真对比实验. 关于 8 个数据集的具体说明可见表 1.

3.2 实验设置

SVM 惩罚参数 C 搜索范围为 $[0.1, 1000]$ 核参

表 1 UCI 数据集

Tab.1 UCI dataset

数据集	类别数	特征数	训练样本数	测试样本数
Sonar	2	60	155	53
Heart	2	13	220	50
Wine	3	13	89	89
Glass	6	9	140	74
Seeds	3	7	150	60
Iris	3	4	90	60
Ionosphere	2	34	251	100
Zoo	7	16	73	28

说明: 训练样本和测试样本在实验时均为随机选取

数 σ 搜索范围为 $[0.01, 100]$. BSA-SVM 模型初始参数设置如下: 种群规模 $N = 30$, 迭代次数 $M = 200$, 飞行频率 $FQ = 10$, 交叉验证次数 $v = 10$, $T = S = 1.5$, $\alpha_1 = \alpha_2 = 1$. PSO-SVM 模型初始参数设置为: 学习因子 $c_1 = c_2 = 1.5$, 交叉验证次数、种群规模和迭代次数与鸟群算法均相同. 当采用不做参数选择的 SVM 方法时, 惩罚参数 C 与核参数 σ 为默认值.

本文采用了台湾林智仁教授开发的 Libsvm3.22 工具箱训练和测试 SVM 分类器. 实验平台为 Intel 奔腾双核处理器(2.5GHz) 2GB 内存, Windows 7 操作系统, 所用软件为 Matlab 2014a. 由于模型计算结果具有一定的波动性, 为了公平性与准确性, 所有模型在每个数据集上得到 C_{best} 与 σ_{best} 后, 都会对相应测试集进行 12 次预测分类, 在剔除 1 次最高分类准确率和 1 次最低分类准确率后, 取剩余 10 次分类准确率的平均值作为最终平均分类准确率.

3.3 实验结果分析

SVM、PSO-SVM 与 BSA-SVM 分类测试结果如表 2 所示, 其中准确率指测试样本的平均分类准确率.

表 2 实验结果对比

Tab.2 Comparison of experimental results

数据集	SVM		PSO-SVM		BSA-SVM		
	准确率/%	(c σ)	准确率/%	训练时间/s	(c σ)	准确率/%	训练时间/s
Sonar	57.74	(365.48 0.23)	86.22	427.97	(989.91 1.66)	88.68	429.97
Heart	79.6	(15.8394 0.01)	82.8	617.24	(38.54 0.01)	85.8	583.95
Wine	93.03	(500.65 0.01)	94.38	108.29	(670.93 1.52)	95.62	105.84
Glass	50.49	(4.82 20.38)	62.03	308.37	(896.68 0.11)	66.62	300.83
Seeds	88.67	(722.98 3.23)	91.67	217.26	(595.09 2.15)	93.67	203.73
Iris	93.33	(15.05 12.54)	93.71	66.2	(729.65 28.12)	94.83	67.33
Ionosphere	91.8	(62.19 0.01)	92.4	936.74	(913.66 0.4)	94.8	947.12
Zoo	90.71	(12.85 1.25)	91.43	52.87	(525.8 0.48)	93.93	56.74

由表 2 可看出, 经过 PSO 算法或 BSA 优化参数后的 SVM 分类器显然比未做参数选择的 SVM 分类器好, 参数优化后的 SVM 分类器能显著提高分类准

准确率. 此外, BSA-SVM 模型与 PSO-SVM 模型在时间效率上差不多, 但 BSA-SVM 模型在 8 个数据集上的分类准确率相对 PSO-SVM 模型均有不同程度的

提高.

4 结语

针对传统方法优化 SVM 参数存在的问题,本文提出使用鸟群算法来优化 SVM 参数.基于 8 个 UCI 数据集的 MATLAB 仿真对比实验表明,BSA-SVM 模型加强了分类器学习能力和泛化能力,能较好地提高分类准确率,是一种有效和可靠的 SVM 参数选择方法.鸟群算法作为一种新颖的优化算法,其在 SVM 参数选择中的应用才刚刚开始,未来的研究工作除了如何提高优化准确率外,还可以集中于如何改善寻优的时间效率.

参 考 文 献

- [1] Vapnik V N. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1995: 988-999.
- [2] 刘洛霞. 基于 SVM 的多变量函数回归分析研究[J]. 电光与控制, 2013, 20(6): 50-57.
- [3] Huanrui H. New mixed kernel functions of SVM used in pattern recognition [J]. Cybernetics & Information Technologies, 2016, 16(5): 5-14.
- [4] Korpela J, Miyaji R, Maekawa T, et al. Toothbrushing performance evaluation using smartphone audio based on hybrid HMM-recognition/SVM-regression model [J]. Journal of Information Processing, 2016, 24(2): 302-313.
- [5] Chapelle O, Vapnik V, Bousquet O, et al. Choosing multiple parameters for support vector machines [J]. Machine Learning, 2002, 46(1): 131-159.
- [6] Keerthi S S. Efficient tuning of SVM hyperparameters using radius/margin bound and iterative algorithms [J]. Neural Networks IEEE Transactions, 2002, 13(5): 1225-1229.
- [7] Peng Xiyuan, Hong Xing. Parameter selection method for SVM with PSO [J]. Chinese Journal of Electronics, 2006, 15(4): 638-642.
- [8] 庄严, 白振林, 许云峰. 基于蚁群算法的支持向量机参数选择方法研究[J]. 计算机仿真, 2011, 28(5): 216-219.
- [9] 魏峻. 基于蝙蝠算法的支持向量机参数优化[J]. 宝鸡文理学院学报(自然科学版), 2015, 35(3): 1-6.
- [10] Meng X B, Gao X Z, Lu L, et al. A new bio-inspired optimisation algorithm: Bird Swarm Algorithm [J]. Journal of Experimental & Theoretical Artificial Intelligence, 2015, 28(4): 1-15.
- [11] 宋晖, 薛云, 张良均. 基于 SVM 分类问题的核函数选择仿真研究[J]. 计算机与现代化, 2011, 2011(8): 133-136.
- [12] Han S, Cao Q, Han M. Parameter selection in SVM with RBF kernel function [J]. Berichte Der Bunsengesellschaft für Physikalische Chemie, 2012, 85(4): 1-4.