

粒子群算法优化 RBF-SVM 沙尘暴预报模型参数

路志英, 李艳英, 陆 洁, 赵智超
(天津大学电气与自动化工程学院, 天津 300072)

摘 要: 为提高沙尘暴的预报准确率, 针对目前已出现的 RBF-SVM 沙尘暴预报模型中的参数优化进行研究. 利用基本粒子群优化算法 (SPSO 算法) 中粒子速度及其位置与 RBF-SVM 模型中参数相对应, 对沙尘暴进行预报, 为解决 SPSO 算法易陷入局部解的缺陷, 提出了惯性权值自适应调节的改进粒子群算法 (WPSO 算法), 并对沙尘暴 RBF-SVM 模型参数进行了优化. 仿真结果表明, 无论是 SPSO 算法, 还是 WPSO 算法, 在优化 RBF-SVM 沙尘暴预报模型参数方面都表现出了良好的性能, 分别比已有的 SVM 方法的预报准确率提高了 22.3% 和 45.3%.

关键词: 支持向量机; 参数优化; 粒子群优化; 沙尘暴预报

中图分类号: TP301.6

文献标志码: A

文章编号: 0493-2137 (2008) 04-0413-06

Parameters Optimization of RBF-SVM Sand-Dust Storm Forecasting Model Based on PSO

LU Zhi-ying, LI Yan-ying, LU Jie, ZHAO Zhi-chao

(School of Electrical Engineering and Automation, Tianjin University, Tianjin 300072, China)

Abstract: To improve the accuracy of sand-dust storm forecasting, an RBF-SVM method with automatic parameter selection was presented in this paper. The proposed method used the simple particle swarm optimization (SPSO) algorithm to get the optimal parameter, in which the velocity and position of each particle correspond a group of RBF-SVM parameters. However, since the PSO tends to get into local optimal solutions, a weight particle swarm optimization (WPSO) algorithm was proposed, in which the weights changed dynamically with a liner rule, to optimize the parameters of RBF-SVM. The simulation results show that both PSO-RBF-SVM and WPSO-RBF-SVM can get high recognition accuracy and efficiency. And the accuracy ratios of two kinds of sand-dust storm forecasting are improved by 22.3% and 45.3% compared with the previous SVM, respectively.

Keywords: support vector machine; parameters optimization; particle swarm optimization; sand-dust storm forecasting

SVM 是基于结构风险最小化原则的分类技术, 它将最大分界面分类器思想和基于核的方法结合在一起, 表现出了很好的泛化能力, 但其性能依赖于学习算法的参数, 然而 SVM 参数选择没有固定的方法, 只能通过实验比较选择. 因此 SVM 的参数确定问题一直是研究的热点问题. 目前已有的方法^[1-6]: ①试凑法, 是比较常用且非常行之有效的方法, 但缺乏

足够的理论依据, 仅凭经验调整, 带有一定的盲目性; ②网格搜索法, 可使各核函数参数相互解耦, 便于并行计算, 提高了运行效率, 但其计算量很大; ③通过最速下降法求 LOO 上界的极小点确定参数方法易陷入局部最优解; ④最小最大化优化方法; ⑤变尺度混沌优化算法等. 上述这些方法都属于估计方法和近似方法, 存在精度差、计算量大等问题. 因此, 寻

收稿日期: 2007-10-30; 修回日期: 2008-01-09.

作者简介: 路志英 (1964—), 女, 副教授.

通讯作者: 路志英, luzhy@tju.edu.cn.

找一个行之有效且计算简单的参数选择方法,依然是需要研究工作者继续深入研究的难点课题。

智能优化算法具有分布性、自组织性、鲁棒性强等优点,虽然对它的研究还处于初级阶段,但已经被成功地应用于函数优化、神经网络等领域。其中粒子群算法作为群智能算法中的代表,引起了研究者的广泛关注,其理论研究在不断深入,其应用领域也在不断拓展,这充分说明了群智能发展所蕴藏的巨大潜力。

笔者针对基于 RBF 核函数的 SVM 沙尘暴预报模型^[7] (RBF-SVM) 中惩罚参数 C 和核参数 σ 对预报结果准确率的影响进行参数优化方法研究,因此对 RBF-SVM 模型的参数进行自动优化研究具有重要的意义。

1 支持向量机

1.1 支持向量机模型

支持向量机的基本思想是通过图 1 所示的最优分类超平面 H 将实心点和空心点两类线性可分样本进行分类,其中 H_1 、 H_2 分别为过各类样本中离分类超平面最近的样本且平行于最优分类超平面的平面。所谓最优分类面就是要求分类面不但能将两类训练样本正确分开(训练错误率为 0),而且使分类间隔最大。

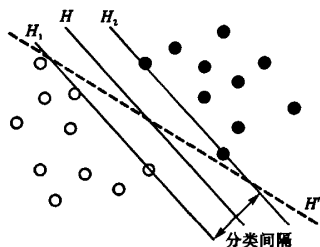


图 1 线性可分 SVM

Fig. 1 Linear separable SVM

对于线性不可分样本(图 2(a))通过映射 ϕ 将样本映射到新的空间(图 2(b)),这样就使线性不可分问题转换为线性可分问题。因此,设分类样本集 $\{(x_i, y_i)\} (i=1, \dots, k, y_i \in \{+1, -1\})$ 可由最优分类函数

$$f(x) = (w, \phi(x)) + b \quad (1)$$

实现分类, w 、 b 根据

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k \xi_i, \quad C > 0 \quad (2)$$

初始条件 $(w, \phi(x_i)) + b \geq 1 - \xi_i, \xi_i \geq 0$

得到。通过拉格朗日乘数法,可将分类函数(1)变为

$$f(x) = \sum_{i=1}^{N_s} \alpha_i \langle \phi(x), \phi(x_i) \rangle + b \quad (3)$$

式中 $K(x, x_i) = \langle \phi(x), \phi(x_i) \rangle$ 为核函数。

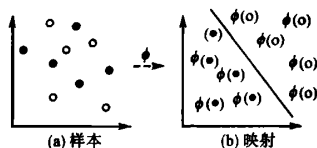


图 2 线性不可分 SVM

Fig. 2 Non-linear separable SVM

1.2 支持向量机核函数的选择及其参数对分类的影响

沙尘暴预报是二分类问题,解决沙尘暴预报的支持向量机模型是一对一分类情况。在 56 维特征空间中无法找到一个合适的线性分类超平面将两类数据完全分开,属线性不可分问题,故需要选择一个合适的核函数解决沙尘暴预报问题。

根据泛函的有关理论,只要满足 Mercer 条件的函数就可以作为核函数,常用的核函数有如下 3 种。

(1) Sigmoid 核函数

$$K(x, x_i) = \tanh[v(x^T \cdot x_i) + c]$$

(2) RBF 核函数

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

(3) Polynomial 核函数

$$K(x, x_i) = [\gamma(x \cdot x_i) + r]^q, \quad r > 0$$

Vapnik^[8]的研究表明支持向量机的参数对其分类效果会产生很大的影响。Polynomial 核函数和 Sigmoid 核函数有 2 个参数,影响参数优化的复杂度,而且实验证明特定参数下,Polynomial 核函数和 Sigmoid 核函数与 RBF 核函数性能相当,所以本文选用 RBF 核函数,因此,影响预报模型准确率的主要原因是 RBF 核函数参数 σ^2 和 SVM 模型中惩罚因子 C 。核参数 σ^2 主要影响样本数据在高维特征空间中分布的复杂程度,而惩罚因子 C 则是在确定的特征空间中调节支持向量机置信范围和经验风险的比例。

对 SVM 核函数参数和惩罚因子 C 同时进行优化具有重要的意义。

2 粒子群算法理论

粒子群优化 (particle swarm optimization, PSO) 算法是一种进化计算技术 (evolutionary computa-

tion), 由 Eberhart 和 Kennedy^[9]于 1995 年提出的. 该算法的基本思想源于对鸟群捕食的行为研究.

假设在一个 d 维解的目标搜索空间中, 有 n 个代表潜在问题解的粒子组成一个种群, 该种群可表示为

$$S=\{X_1, X_2, \cdots, X_n\}$$

式中, $X_i=(X_{i1}, X_{i2}, \cdots, X_{id})^T$, ($i=1, 2, \cdots, n$), 表示第 i 个粒子在 d 维空间中的位置. 这 n 个粒子通过与求解问题相关的目标函数求得相应的适应度值, 获得第 i 个粒子自身的最优解 P_i (即个体极值) 以及所有粒子中的最优解 P_g (即全局极值). 每个粒子由以下式子^[10-11]

$$v_i^{k+1}=\omega v_i^k+c_1\text{rand}(1)(P_i^k-X_i^k)+c_2\text{rand}(2)(P_g^k-X_i^k)\tag{4}$$

$$X_i^{k+1}=X_i^k+v_i^{k+1}\tag{5}$$

共同决定其下一步追随当前最优粒子在解空间中搜索的位置. 因此, PSO 算法是基于个体间的协作与竞争来完成复杂搜索空间中最优解的搜索, 是一种启发式全局优化算法.

式 (4) 通过 3 部分来计算第 i 个粒子的新速度, 其中: ωv_i^k 为第 i 个粒子 k 时刻的速度; $c_1\text{rand}(1) \cdot$

$(P_i^k-X_i^k)$ 为 k 时刻第 i 个粒子的当前位置与其历史最好位置 (个体极值) 之间的距离; $c_2\text{rand}(2) \cdot (P_g^k-X_i^k)$ 为 k 时刻第 i 个粒子的当前位置与群体历史最好位置 (群体极值) 之间的距离. 在上述各部分中: ω 为惯性因子, 权衡全局和局部搜索能力; c_1 、 c_2 为学习因子, 是两个正常数; $\text{rand}(1)$ 、 $\text{rand}(2)$ 为均匀分布于 $[0, 1]$ 之间的两个随机数.

式 (5) 用于计算第 i 个粒子 $k+1$ 时刻的新位置.

3 基于粒子群算法的 SVM 参数优化

以灾害性天气——沙尘暴为背景, 通过粒子群算法解决沙尘暴 SVM 预报模型^[7]的参数优化问题.

3.1 SVM 沙尘暴预报模型样本

根据我国出现沙尘暴的区域、时间及成因, 本文研究所用的沙尘暴样本集选自美国环境气象中心 (NCEP) 提供的 1981—1997 年 2 月~6 月东经 70°~115°、北纬 35°~55°的 3 个物理场数据, 样本数据的结构为 2.5×2.5 (每格 2.5°) 格点场 (经度和纬度的交点) 数据阵, 如表 1 所示.

表 1 沙尘暴样本及历史数据源

Tab. 1 Samples and original data of the sand-dust storm

时 间		地 点		样本数	源场(物理场)		
年	月	东 经	北 纬		50 kPa 高度值 (H5)	85 kPa 位温值 (SE)	70 kPa 风速值 (UV)
1981—1997	2—6	70°~115° (均分 18 格)	35°~55° (均分 8 格)	2 027			

注: 风速为投影到东西和南北两个垂直方向上的值.

由于高度值场、高度场、位温场的数据维数均为 $19 \times 9=171$, 风形场为 $19 \times 9 \times 2=342$, 所以, 沙尘暴样本数据的维数 (属性) 高达 855. 对源场数据利用主成分分析法 (PCA) 进行数据预处理, 得到 56 维的特征向量作为预报模型的实验样本^[7], 如表 2 所示.

表 2 实验样本

Tab.2 Samples of the experiment

样本采集 时间/年	总个 数	非沙尘 暴日	沙尘 暴日	沙尘暴 日频数/ %	单样 本维 数
1981—1997 (初春~初夏)	2 027	1 454	573	28.27	56

3.2 SVM 参数对编码

PSO 算法是一种启发式全局搜索算法^[12], 要使其能搜索到最优解, 首先给每一个需要优化的参数指定搜索空间范围, 设定空间越大, 搜索的精度就会越

低, 但若指定的搜索空间不够大, 则会导致只能寻得局部最优解的情况, 故指定搜索空间需综合考虑决定. 本文将粒子群优化算法中粒子速度及其位置与 RBF-SVM 模型中的参数相对应. 为了提高搜索效率先采用大步长网格搜索法对参数对 (C, σ) 在其搜索空间进行搜索. 设 C 和 σ 分别取 N 个和 M 个, 则分别训练由 (C, σ) 参数对所构成的 $N \times M$ 个 SVM 模型, 从而在 $N \times M$ 个 (C, σ) 组合中得到使支持向量机分类效果最好的一个组合作为最优参数对. 本文设定 (C, σ) 参数对的取值范围分别为 $[0, 20]$ 和 $[50, 100]$, 并分别对参数 σ 、 C 进行二进制编码.

3.3 适应度函数的确定

气象上, 沙尘暴预报的准确率由成功界限指数 CSI 来衡量预报模型的性能. CSI 的定义为

$$CSI=\frac{c_i}{c_i+f_i+f_2}\times 100\% \tag{6}$$

式中: c_i 为正确报出的沙尘暴日数; f_1 为沙尘暴漏报

日数; f_2 为空报沙尘暴日数。

由成功界限指数的定义可知, CSI 的值为非负实数, CSI 值越大预报准确率越高, 支持向量机预报模型的分类效果越好。由此可见, 支持向量机的参数对组合越理想, CSI 的值就越大。本文设定 CSI 为适应度函数。

3.4 运行参数的设定

在 PSO 算法中, 惯性权值决定着粒子前一时刻的速度对当前时刻位置的影响 (参见式 (5)), 较大的 ω 取值会扩大粒子的搜索空间, 有利于使其跳出局部极小点; 较小的 ω 取值会增强粒子的局部搜索能力, 有利于算法收敛^[13-14]。大量的实验研究表明: 面对不同的具体问题, ω 的取值不同, 但一般情况下, ω 的取值只要在 (0, 1.4) 之间, PSO 算法就会收敛^[13]。在本文优化过程中, ω 的取值为 0.8 (多次实验的最佳效果值)。

3.5 运行过程

PSO 算法的流程如图 3 所示。

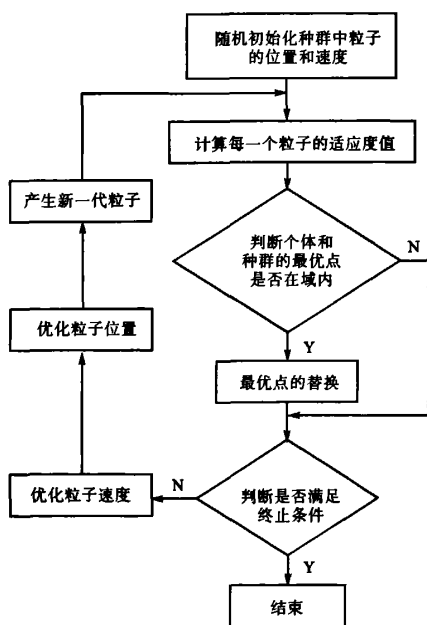


图 3 基本粒子群优化算法流程

Fig. 3 Flow chart of the simple PSO algorithm

实现 RBF-SVM 优化参数的具体步骤如下。

步骤 1 随机初始化粒子种群, 即随机初始化种群中每一个粒子的位置和速度。

步骤 2 根据适应度函数 CSI, 计算出所有粒子的适应度值。

步骤 3 根据计算出的适应度值确定种群中每

一个粒子的最优历史位置。

步骤 4 根据计算出的适应度值确定种群中所有粒子中的最优历史位置。

步骤 5 根据迭代式 (4) 和式 (5) 对种群中的所有粒子进行速度和位置的迭代更新。

步骤 6 重复步骤 2~5 直到满足算法的迭代条件终止。

4 粒子群算法的改进

PSO 算法作为一种新出现的群智能算法, 因其算法简单、容易实现等优点在优化领域得到了广泛应用, 但基本 PSO 算法也存在容易陷入局部极小值点等缺点, 为了提高算法的性能, 目前已出现了多种对粒子群的改进算法, 多数是集中在对其参数惯性权重的改进上。

惯性权值 ω 控制粒子的全局搜索能力和局部搜索能力之间的平衡, 较大的 ω 有利于跳出局部极小点, 而较小的 ω 有利于算法收敛。本文引入 Shi 和 Eberhart^[13] 的研究成果, 采用了自适应调整 ω 的控制策略。算法的基本思想是: 在运行的初期使用较大的惯性权值, 使算法具有较强的全局搜索能力; 运行后期则使用较小的惯性权值, 以提高算法的局部搜索能力, 即随着迭代次数的增大, 线性地减小 ω 值。本文自适应调节惯性权值的计算式为

$$\omega = (\omega_1 - \omega_2) \times \frac{\max(\alpha) - \alpha}{\max(\alpha)} + \omega_2 \quad (7)$$

式中: ω_1 为惯性权值的初值; ω_2 为惯性权值的终值; $\max(\alpha)$ 为最大迭代次数。

5 实验结果及分析

5.1 基本粒子群算法及其改进后算法的训练结果

由于 SVM 能较好地解决小样本问题, 因此, 本文将沙尘暴样本采用随机抽样的方式选取 600 个作为训练集, 选取 1 000 个作为测试集。

由于粒子群算法是全局寻优的概率搜索算法, 其种群初始化结果均决定着每次的寻优过程, 尽管同一种算法针对同一组训练集, 由于初始种群的不同, 每次训练的过程也有所不同, 但最终都能稳定到一个最优平衡点。图 4 和图 5 所示的仿真结果是出现频率最高的情况, 具有代表意义。

由图 4 和图 5 仿真结果可以看出, 改进后的粒子群算法无论是准确率, 还是运行速度都有了很大的提

高。尽管寻优过程不太相同,但都能在 10 代以内找到最优解,在不降低训练质量的同时大大地提高了寻优的效率。需要说明的是:本文模型是对未来 24 h 天气进行预测的,而且表明沙尘暴天气的站点数至少要在 3 个以上。

由此可见,改进算法的性能优于基本粒子群算法。

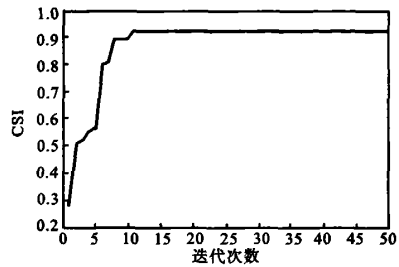


图 4 基本粒子群算法优化参数情况((C,σ)=(83.519 7, 1.493 8), CSI=0.930 4)

Fig. 4 Parameters of SPSO((C,σ)=(83.519 7, 1.493 8), CSI=0.930 4)

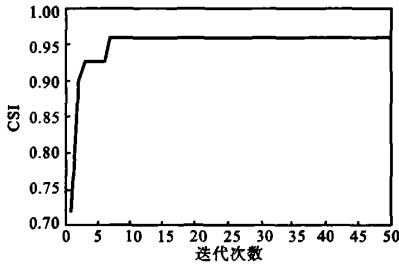


图 5 改进后的粒子群算法优化参数情况((C,σ)=(87.472 7, 1.042 3), CSI=0.959 0)

Fig. 5 Parmeters of WPSO((C,σ)=(87.472 7, 1.042 3), CSI=0.959 0)

5.2 预报测试结果

沙尘暴预报测试结果如表 3 所示,由表 3 不难看出,改进后的粒子群算法所找到的最优参数对于沙

尘暴预报模型中,所得预报结果的准确率远大于基本粒子群算法和文献[7]的手动调节参数方法。另外,10 次测试结果相同,表明支持向量机分类器具有很好的稳定性。

6 结 论

(1) 成功地将粒子群优化算法应用于沙尘暴预报模型中,拓展了粒子群算法的应用领域。

(2) 粒子群算法 PSO 有效地优化了沙尘暴 RBF-SVM 预报模型中的参数对。无论是基本 PSO 算法,还是自适应调节权值的改进 PSO 算法,在优化支持向量机沙尘暴预报模型参数方面都表现出了良好的性能,分别比文献[7]的沙尘暴预报准确率提高了 22.3%和 45.3%。

(4) 改进后的粒子群算法(WPSO)在不降低训练质量的同时大大地提高了寻优效率,其快速性高于基本粒子群算法(SPSO)。

(5) 实验测试次数结果再次表明:在参数对及核函数一定的情况下,支持向量机分类器具有很好的稳定性。

参考文献:

[1] 王 鹏,朱小燕.基于 RBF 核的 SVM 的模型选择及应用[J].计算机工程与应用,2003,39(24):72-73.
Wang Peng, Zhu Xiaoyan. Model selection of SVM with RBF kernel and its application[J]. Computer Engineering and Applications, 2003, 39(24): 72-73 (in Chinese).

[2] 杨 旭,纪玉波,田 雪.基于遗传算法的 SVM 参数选取[J].辽宁石油化工大学学报,2004,24(1):54-58.
Yang Xu, Ji Yubo, Tian Xue. Parameters selection of SVM based on genetic algorithm[J]. Journal of Liaoning University of Petroleum and Chemical Technology, 2004, 24(1): 54-58 (in Chinese).

[3] 王兴玲,李占斌.基于网格搜索的支持向量机核函数参数的确定[J].中国海洋大学学报,2005,35(5):859-862
Wang Xingling, Li Zhanbin. Identifying the parameters of the kernel function in support vector machines based on the grid-search method[J]. Periodical of Ocean University of China, 2005, 35(5): 859-862 (in Chinese).

[4] 齐志泉,田英杰,徐志洁.支持向量机中的核参数选择问题[J].控制工程,2005,12(4):379-381.
Qi Zhiquan, Tian Yingjie, Xu Zhijie. Kernel-parameter

表 3 自动选取参数的测试结果

Tab. 3 Results of the automatic selecting parameters

训练方法	训练样本数	参数对(C,σ)的训练结果	测试样本数	预报准确率 CSI/%	测试次数
SPSO	600	(83.519 7, 1.493 8)	1 000	71.2	10
WPSO	600	(87.472 7, 1.042 3)	1 000	84.6	10

注:文献[7]的预报准确率为 58.2%.

- selection problem in support vector machine [J]. *Control Engineering of China*, 2005, 12 (4): 379-381 (in Chinese) .
- [5] 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.
- Dong Chunxi, Rao Xian, Yang Shaoquan, et al. Method for selecting the parameters of support vector machines [J]. *Systems Engineering and Electronics*, 2004, 26 (8): 1117-1120 (in Chinese) .
- [6] 袁小芳, 王耀南. 基于混沌优化算法的支持向量机参数选取方法 [J]. *控制与决策*, 2006, 21 (1): 111-113, 117.
- Yuan Xiaofang, Wang Yaonan. Selection of SVM parameters using chaos optimization algorithms [J]. *Control and Decision*, 2006, 21 (1): 111-113, 117 (in Chinese) .
- [7] 路志英, 张启孟, 赵智超. 基于 SVM 的沙尘暴预测模型的研究 [J]. *天津大学学报*, 2006, 39 (9): 1110-1114.
- Lu Zhiying, Zhang Qimeng, Zhao Zhichao. Sand-dust storm forecasting model based on SVM [J]. *Journal of Tianjin University*, 2006, 39 (9): 1110-1114 (in Chinese) .
- [8] Vapnik C. *Statistical Learning Theory* [M]. New York: John Wiley and Sons, 1998.
- [9] Kennedy J, Eberhart R C. Particle swarm optimization [C] // *Proceedings of IEEE International Conference on Neural Networks*. Perth, 1995: 1942-1948.
- [10] Shi Y, Eberhart R C. A modified particle swarm optimizer [C] // *Proceedings of IEEE International Conference on Evolutionary Computation*. Anchorage, 1998: 69-73.
- [11] Shi Y, Eberhart R C. Empirical study of particle swarm optimization [C] // *Proceedings of the World Multi-Conference on Systemics, Cybernetics and Informatics*. Orlando, FL, 2000: 1945-1950.
- [12] Parsopoulos K E, Vrahatis M N. Modification of the particle swarm optimizer for locating all the global minima [J]. *Artificial Neural Networks and Genetic Algorithms*, Springer, 2001: 324-327.
- [13] Shi Y, Eberhart R C. Parameter selection in particle swarm optimization [C] // *The 7th Annual Conference on Evolutionary Programming*. New York, Springer-Verlag, 1998: 591-600.
- [14] El-Gallad A, El-hawary M, Sallam A, et al. Enhancing the particle swarm optimizer via proper parameters selection [C] // *Proceedings of the IEEE Conference on Electrical and Computer Engineering*. Canada, 2002: 792-797.