

基于 IFOA 算法的 SVM 参数优化及其应用^①

赵 伟

(陕西省行政学院 电子设备与信息管理处, 西安 710068)

摘 要: 为了提高果蝇优化算法的种群多样性和果蝇搜索的遍历性, 有效提高算法的收敛精度, 提出一种改进的果蝇算法(Improving fruit fly optimization algorithm, IFOA), 仿真实验表明, IFOA 算法保持了搜索过程中的搜索尺度变化, 平衡了算法的全局与局部搜索能力. 在此基础上, 为了改善支持向量机模型参数选择的随机性和盲目性, 提高模式分类的准确率, 提出并建立了一种 IFOA-SVM 模式分类模型. 该方法将 IFOA 算法引入到支持向量机模型参数优化中, 建立性能最优的支持向量机模型. 应用该模型对 UCI 机器学习数据库中 wine 数据集进行模式分类研究, 通过算法对比分析, 结果表明: 提出的改进果蝇优化算法在收敛速度和寻优效率上均有一定的提高, 依此而建立的 IFOA-SVM 模式分类模型具有较准确的分类准确率, 从而也验证了该模式分类方法在 wine 数据集分类应用中的有效性.

关键词: 果蝇算法; 支持向量机; 参数优化; 分类模型

SVM Parameters Optimizing Based on Improved Diminishing Step Fruit Fly Optimization Algorithm and Its Application

ZHAO Wei

(Department of Electronic equipment and information management, Shaanxi Academy of Government, Xi'an 710068, China)

Abstract: In order to advance population diversity and ergodic property for fruit fly optimization algorithm, enhance its convergence precision effectively, an algorithm named improving fruit fly optimization algorithm (abbreviated as IFOA) is proposed in this paper. The simulation experiment shows that this algorithm maintains changing in scale and balances the overall and local searching capability. In order to improve the randomness and blindness in choosing SVM model parameter artificially, enhancing accuracy for pattern classification at the same time. A method using IFOA in the field of SVM model parameter optimization is put forward and established. In this method, IFOA is applied into penalty factor and kernel function parameters optimization for SVM, with which the optimal model parameters will be chosen and the optimal SVM model can be established. This model is used in pattern classification research for wine data in UCI machine study database, different algorithms were used for comparison, the result shows that, the improved FOA has a fast speed in convergence and high efficiency in optimization, a better classification accuracy could be reached for IFOA-SVM model. The effectiveness for IFOA in wine database classification is proved thereby.

Key words: diminishing step; fruit fly; support vector machine(SVM); parameter optimization; classification model

支持向量机(support vector machine, SVM)^[1,2]是一种基于统计学习理论的机器学习算法, 该算法具有理论完备、适应性强、全局优化、泛化性能好等优点. 在 SVM 分类器的建立中, 其核函数和惩罚因子的选取十

分关键. 目前该问题仍缺乏理论指导^[3], 因此相关科研人员常将智能优化算法引入到该领域研究中, 常见的算法包括遗传算法和粒子群算法等^[4], 但在实际应用中, 均存在一定的不足: 遗传算法收敛速度慢, 且容

^① 基金项目:国家自然科学基金(61272509);陕西省“百人计划”和国家自然科学基金委员会重大国际(地区)合作研究项目(61120106010)

收稿时间:2014-10-29;收到修改稿时间:2014-12-05

易陷入局部最优值^[5,6]；粒子群优化算法收敛速度快，但在进化后期收敛速度降低，易容易陷入局部最优，局部搜索能力比较差^[7]。

果蝇优化算法(fruit fly optimization algorithm, FOA)^[8]于2011年由台湾学者Wen-Tsao Pan提出，是一种基于果蝇觅食行为推演出寻求全局优化的群智能方法。算法通过嗅觉搜索行为提高果蝇个体的多样性，增大果蝇个体的搜索空间；通过视觉搜索使果蝇个体快速收敛。因而具有参数少、计算速度快、全局寻优能力强，易于实现等特点。

针对果蝇优化算法中固定步长对于搜索效率敏感且不易选取的问题，本文提出一种改进果蝇算法(Improving fly optimization algorithm, IFOA)，将步长由固定值变为依觅食进程而递减，采用模拟实验验证了该方法的有效性。在此基础上，为了改善支持向量机模型参数选择的随机性和盲目性，运用该算法优化支持向量机分类模型的惩罚因子 c 与核函数参数 g ，并建立基于IFOA-SVM的模式分类模型，对UCI数据库中的wine数据进行模式分类验证了该方法的有效性。

1 改进果蝇优化算法及模拟验证

1.1 果蝇优化算法基本原理

作者Wen-Tsao Pan从果蝇的觅食行为得到启发，提出一种基于果蝇觅食行为而寻求全局优化的新方法，果蝇本身在感官知觉上优于其它物种，尤其在嗅觉和视觉上。果蝇的嗅觉器官能够很好地搜集漂浮在空气中的各种气味，然后飞近食物位置后亦可以用敏锐的视觉发现食物与同伴聚集的位置，并向该方向飞去。

依照果蝇搜寻食物的特性，将其归纳为几个必要的步骤：

- 1) 随机初始果蝇群体位置坐标。

$$\text{Init } X_axis; \text{Init } Y_axis.$$

- 2) 根据果蝇群体位置坐标，随机初始化群体中果蝇个体的位置坐标，使其利用嗅觉搜寻食物。

$$X_i = \text{Init } X_axis + \text{Random Value}.$$

$$Y_i = \text{Init } Y_axis + \text{Random Value}.$$

- 3) 由于果蝇群体无法得知食物位置，因此首先估计每个果蝇个体与坐标原点的距离(Dist)，并将其倒数定义为味道浓度判定值(Si)。

$$D_{ist_i} = \sqrt{X_i^2 + Y_i^2}; S_i = 1 / D_{ist_i}.$$

- 4) 将味道浓度判定值(Si)代入到味道浓度判定函数(Fitness function)中，从而得出该果蝇个体利用嗅觉得到的味道浓度(Smelli)。

$$\text{Smell}_i = \text{Function}(S_i).$$

- 5) 找出此果蝇群体中的味道浓度最高的果蝇个体。

$$[\text{bestSmell } \text{bestIndex}] = \max(\text{Smell}).$$

- 6) 判断味道浓度是否优于前一代味道浓度，若是则执行7)。否则，重复执行2)~6)。

- 7) 保留最佳味道浓度值与最佳果蝇个体的位置坐标，果蝇群体利用视觉向该位置飞去。

$$\text{Smellbest} = \text{bestSmell}.$$

$$X_axis = X(\text{bestIndex}).$$

$$Y_axis = Y(\text{bestIndex}).$$

1.2 果蝇优化算法基本原理

在传统的果蝇算法FOA^[9,10]中，其搜索步长为定值。根据介值理论，在连续搜索空间中，如果极值点附近存在一个邻域，则该邻域内的点的目标函数值正比于该点到极值点的距离。根据相关理论，在实际优化问题中，较优解的周围可能存在更优的解。体现在SOA算法中，即可表现为，当个体处于较优位置时，应该在较小的邻域搜索；当个体处于较差位置时，应该在较大邻域内搜索。

因此，为了改进果蝇算法由于步长固定取值带来的全局搜索与局部搜索能力的不足，提出一种改进的果蝇算法(Improving FOA)，该算法在果蝇算法的基础上，变固定步长为改进。步长值计算公式如下：

$$\text{step} = \text{step0} - \frac{\text{step0}}{\max \text{gen}} \times (\text{gen} - 1) \quad (1)$$

其中，step0为初始步长值，一般取较大值。maxgen为最大觅食迭代数。gen为当前觅食代数。当第一代觅食时，step=step0，果蝇个体步长为最大值step0。随着觅食迭代次数的增加，步长按照maxgen的差值递减，直至最后一代减小至maxgen。

根据算法的改进，IFOA算法在进化初期搜索步长大，具有最大的搜索空间，全局寻优能力最强。随着进化代数的增加，其局部搜索能力逐步增强。实现了全局搜索能力和局部寻优能力的动态平衡。

1.3 IFOA 算法模拟验证

为了验证IFOA算法的寻优效果，计算以下函数的最小值，如式(5)所示。函数曲线如图1所示。

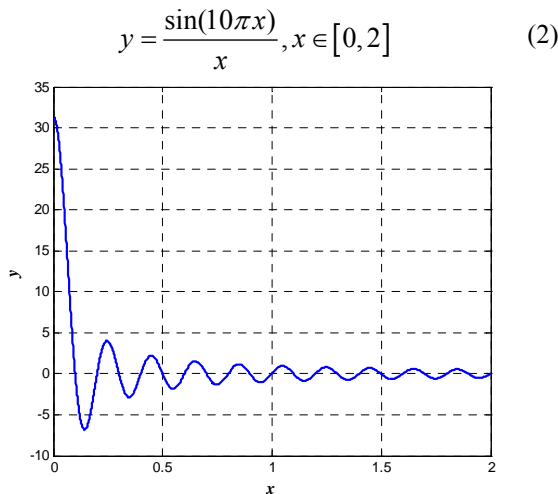


图 1 仿真函数图形

分别选用 FOA 算法(步长 30, 步长 3)和 IFOA 算法(初始步长 30)进行最小值寻优, 经觅食迭代之后, 三种对比算法的最优值觅食曲线分别如图 2、图 3、图 4 所示。

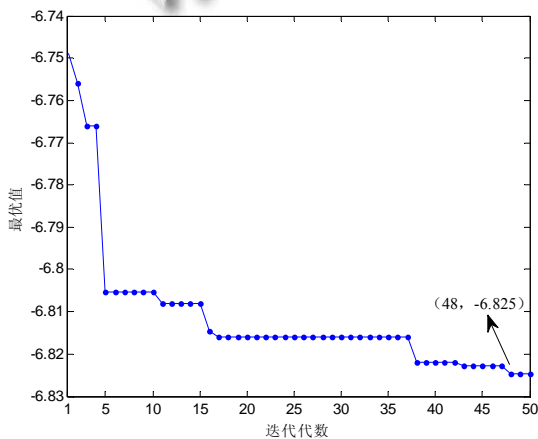


图 2 IFOA 算法效果

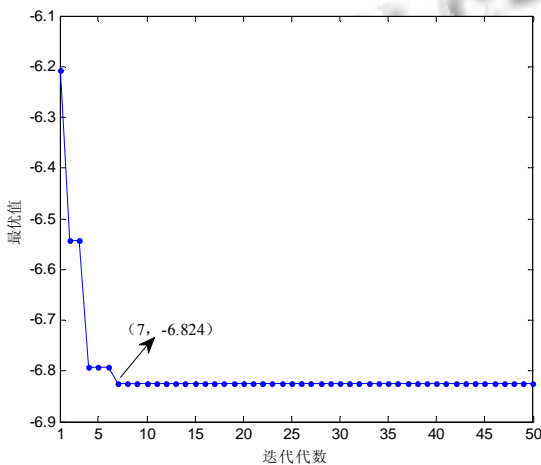


图 3 FOA(步长 30)算法效果

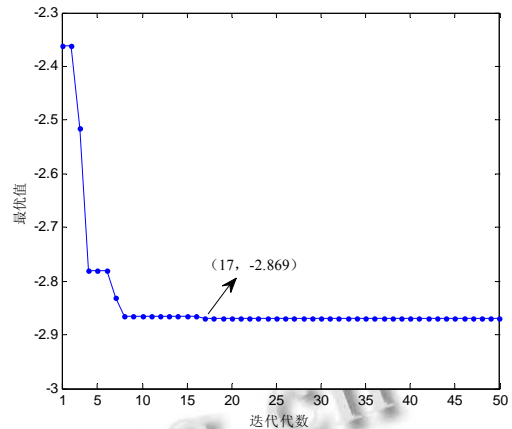


图 4 FOA(步长 3)算法效果

综合分析, IFOA 算法在觅食初期, 由于搜索范围的最大化, 全局寻优能力强, 第 1 代即达到较优值-6.748, 随着觅食迭代次数的增加逐渐接近最优值, 觅食末期, 随着局部搜索能力的增强, 还有轻微的最优值调整, 最后在第 48 代处达到最优值-6.825; L=30 时 FOA 的全局搜索能力强, 在第 7 代即达到最优值-6.824, 但之后一直保持不变, 可见, 固定大小的搜索尺度导致大步长 FOA 在迭代后期局部搜索能力不强, 全局搜索能力强, 而局部搜索能力不够; 当 L=3 时, 果蝇在第 17 代达到最优值-2.869, 可见, 由于小步长导致算法局部搜索能力强, 而全局搜索能力弱, 使算法过早地陷入局部最优值. 无法跳出局部最优解. 综上可见, IFOA 保持了搜索过程中的搜索尺度变化, 平衡了算法的全局与局部搜索能力. 对于寻优问题可取得较好效果。

2 基于 IFOA-SVM 的分类识别研究

2.1 数据来源与预处理

为了验证本文提出的 IFOA-SVM 模式分类方法的性能, 选取 UCI 及其学习数据库^[11]中的 wine 数据集, 该数据库记录了在意大利同一区域上三种不同品种的葡萄酒的化学成分分析, 数据集规模为 178×13, 属于多变量的模式分类问题. 实验的硬件环境为 M450@2.40GHz, 内存为 2.00G, 仿真工具为 Matlab2010.

将 178 个样本的 50%作为训练集, 另 50%作为测试集, 用于测试基于 IFOA 优化的 SVM 的分类性能.

2.2 基于 IFOA 的支持向量机参数优化

首先以 89 个样本为训练集, 对 SVM 模型进行训练, 运用改进的 FOA 算法对 SVM 惩罚因子 c 和核函数参数 g 进行优化. 以训练样本的分类准确率为味道判别函数,

果蝇群体迭代次数 GM=50, 种群规模 Ps=20. 图 6

为 50 次果蝇群体觅食后适应度函数的演化情况, 显然, 训练样本的准确率在第 29 代时达到最优值 98.8764% 并保持稳定, 此时惩罚因子 $c=2.55029$, 核函数参数 $g=0.89366$.

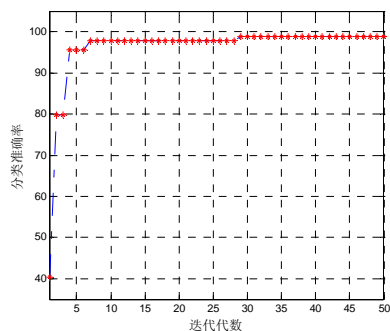


图 5 基于 IFOA 的适应度函数演化曲线

为了比较说明 IFOA 算法在性能优化方面的优势, 分别采用 IFOA、FOA、PSO、GA 四种算法对 SVM 的参数进行优化, 由于优化的结果具有一定的随机性, 对每种算法分别进行 20 组运算, 对其运算指标取均值进行对比分析. 计算结果如表 1 所示.

表 1 算法 20 组实验平均结果对比

算 法	训练准确率	平均训练时间/s	平均收敛代数
IFOA	98.0213	4.01	25
FOA	96.3721	4.45	32
PSO	97.9655	5.23	18
GA	95.3468	5.68	39

可以看出, 与其他 3 种优化算法相比, IFOA 具有最高的寻优精度且运行时间最短, 因此, 搜索效率最高; FOA 运行时间上与 IFOA 相当, 但由于局部寻优能力较低, 导致寻优精度略低; 相比而言, PSO 与 GA 优化算法在运行速率上要低于 IFOA 与 FOA. 二者相比, PSO 取得了较高的寻优精度, 而 GA 则由于过早收敛而陷入局部最优, 寻优精度最低. 可见, IFOA 在保持 FOA 的优势前提下, 进一步提高了算法的寻优精度和搜索效率.

2.3 基于 SVM 的模式分类

以经过 IFOA 优化得到的 SVM 参数为基础, 建立 SVM 模型. 对测试样本进行模式分类, 分类效果如图 7 所示. 可以看出, 在 89 个测试样本中, 除第三类中有一个样本被识别为第二类外, 其余测试样本全部得到正确的分类, 准确率达到 98.88%. 可以看出, 基于 IFOA-SVM 的模式分类方法在对 wine 数据集的分类应用中是有效的.

3 结语

针对实际问题中果蝇优化算法步长参数对寻优效率影响权重重大, 不易选取的不足, 提出改进的果蝇优

化算法. 该算法实现了全局搜索能力和局部寻优能力的动态平衡, 使果蝇群体伴随觅食行为进程而实现全局搜索能力和局部寻优能力的此消彼长. 将该算法应用到 SVM 模型的参数优化中, 搜索效率、预测精度、运算速度均取得了较好效果.

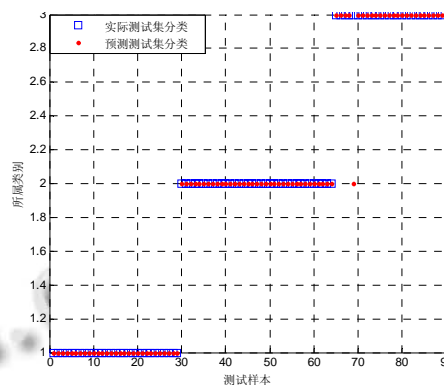


图 6 模式分类结果

参考文献

- 1 石东源, 卢炎生, 王星华, 段献忠. SVG 及其在电力系统软件图形化中的应用初探. 继电器, 2004, 32(16): 37-40.
- 2 Vapnik VN. The nature of statistical learning theory. New York(USA): Springer-Verlag, 1995.
- 3 Cortes C, Vapnik VN. Supporter vector networks. Machine Learning, 1995, 20(3): 273-297.
- 4 Chapelle O, Vapnik V, Bousquet O, et al. Choosing multiple parameters for support vector machines. Machine Learning, 2002, 46(1): 131-159.
- 5 朱凤明, 樊明龙. 混沌粒子群算法对支持向量机模型参数的优化. 计算机仿真, 2010, 27(11): 183-186.
- 6 周晓剑. 基于梯度信息的最小二乘支持回归机. 南京理工大学学报(自然科学版), 2011, 35(1): 138-143.
- 7 陈涛. 基于差分进化算法的支持向量机参数选择. 计算机工程与应用, 2011, 47(5): 24-26.
- 8 潘丽姣, 吴红英. 混沌逃逸粒子群优化算法在 WSN 覆盖优化中的应用. 重庆邮电大学学报(自然科学版), 2014, 26(2): 177-181.
- 9 Pan WT. A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example. Knowledge-Based Systems, 2012, (26): 69-74.
- 10 周平, 白广忱. 基于神经网络与果蝇优化算法的涡轮叶片低循环疲劳寿命健壮性设计. 航空动力学报, 2013, 28(5): 1013-1018.
- 11 张翔, 陈林. 基于果蝇优化算法的支持向量机故障诊断. 电子设计工程, 2013, 21(16): 90-93.
- 12 Forina M, et al. PARVUS-An Extendible Package for Data Exploration. <http://archive.ics.uci.edu/html>. [2014-05-15].