

# 一种混沌人工鱼群算法对 SVM 参数的优化及应用

朱文静,白 静

(太原理工大学 信息工程学院,山西 太原 030024)

**摘 要:** 通过结合混沌模型实现对人工鱼群算法中各行为的改进,提出了一种混沌人工鱼群算法(CAFSA)优化 SVM 参数的方法,并采用测试函数进行测试和比较,再将寻优的参数运用到一个非特定人、孤立词的语音识别系统中.实验结果表明,在不同低信噪比和不同词汇量的条件下,基于混沌人工鱼群算法的 SVM 模型与基于基本人工鱼群算法的 SVM 模型相比,收敛速度明显加快,语音识别率也有不同程度的提高.

**关键词:** 混沌人工鱼群算法;支持向量机;参数优化;语音识别

中图分类号: TP31

文献标识码: A

文章编号: 1000-7180(2016)03-0089-05

## A Chaos Artificial Fish Swarm Algorithm for Parameters Optimization and Application of Support Vector Machine

ZHU Wen-jing, BAI Jing

(College of Information Engineering, Taiyuan University of Technology, Taiyuan 030024, China)

**Abstract:** This paper presents parameters of SVM were optimized by a chaos artificial fish swarm algorithm. By using some test functions, it is proved that combining with chaotic model can improve the behavior of artificial fish swarm algorithm. The optimized parameters are used into a speech recognition system of non-specific persons and isolated words. The experimental results show that convergence rates and speech recognition correct rates based on SVM using chaos artificial fish swarm algorithm are better than those by artificial fish swarm algorithm optimization parameters under different low SNRs and different words.

**Key words:** chaos artificial fish swarm algorithm; support vector machine; parameters optimization; speech recognition

### 1 引言

支持向量机<sup>[1]</sup>(SVM)算法是一种研究有限样本预测的特殊机器学习方法,具有泛化能力强、维数不敏感、收敛到全局最优等特点,可以有效地解决过学习、欠学习及“维数灾难”等问题<sup>[2]</sup>.但影响支持向量机性能的一个重要因素是参数的选择问题.因此,研究支持向量机参数的优化方法,对其发展具有重要的意义.

人工鱼群算法<sup>[3]</sup>(AFSA)作为一种分布式的群体智能优化算法,自 2002 年由李晓磊等人提出后就一直备受关注.该算法具有并行处理能力,鲁棒性强,对初始参数值设定要求低,进行全局寻优时收敛

性能好等特点<sup>[4]</sup>.然而,人工鱼群算法在后期存在收敛精度低和收敛速度慢的缺陷<sup>[5]</sup>.因此,本文在人工鱼群优化算法中引入混沌思想,利用混沌序列的随机性、遍历性和规律性来提高算法的效率,得到一种性能更好的混沌人工鱼群算法(CAFSA),并将通过该算法优化后的支持向量机的核参数与惩罚因子应用于语音识别系统中.仿真实验结果表明,混沌人工鱼群算法能够有效地避免早熟问题,加快收敛速度,同时语音识别率也有相应的提高.

### 2 支持向量机

支持向量机的主要目标是构建一个能有效将待分类输入样本无错误分开的最优分类面,且使距离

该平面两侧最近的两类样本间的分类间隔最大,从而实现样本的分类问题.图1所示为二分类问题线性可分的情况,图中使用圆形点和菱形点分别表示两类样本, $H$ 为有效将两类样本区分开的分类面, $H_1, H_2$ 均平行于分类面,且分别距分类面最近的两类样本点, $H_1$ 和 $H_2$ 之间的距离(margin)称为最大分类间隔距离.

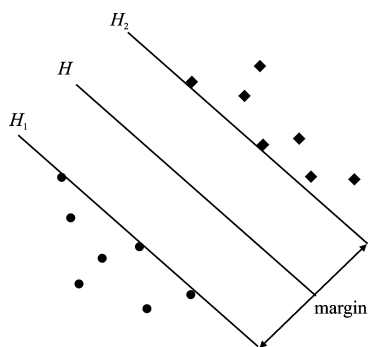


图1 二维线性可分情况

设线性可分样本集 $(x_i, y_i), x_i \in R^d, y_i \in \{+1, -1\}, i \in \{1, 2, \dots, l\}$ ,其中 $d$ 为输入维数, $l$ 为样本数量, $y_i$ 表示向量 $x_i$ 所属类别号.如果训练样本线性不可分,可引入松弛变量 $\xi_i \geq 0$ ,分类超平面方程为 $w \cdot x + b = 0$ ,则将求解最优超平面转化为凸二次规划问题:

$$\begin{aligned} \min \quad & |w|^2/2 + C(\sum_i \xi_i) \\ \text{s. t. } & y_i(w \cdot x_i + b) \geq 1 - \xi_i. \end{aligned}$$

式中,常数 $C > 0$ 是惩罚因子, $C$ 越大表示对错误分类的惩罚越大.由此求得决策函数为

$$f(x) = \text{sgn}(\sum_{i=1}^l \alpha_i^* y_i K(x_i, x) + b^*).$$

式中, $\text{sgn}()$ 是符号函数, $K(,)$ 为核函数, $\alpha^*$ 为最优解, $b^*$ 为分类阈值.

本文选用径向基(RBF)核函数:

$$K(x_i, x_j) = \exp\{-\gamma |x_i - x_j|^2\}.$$

式中, $\gamma$ 为非零常数.该函数具有较好的收敛域,可实现非线性映射,适用于语音训练样本特征参数多、维数高的特点.因此,对于核函数为RBF的支持向量机来说,参数组 $(C, \gamma)$ 的选择就直接决定了其性能的好坏.

### 3 混沌人工鱼群算法优化SVM参数的提出

#### 3.1 混沌模型

混沌是指在确定的非线性系统中出现的一种貌似无规则的随机现象.混沌优化算法具有对初值敏

感、易跳出局部极小、搜索速度快、计算精度高、全局接近收敛的特点. Logistic映射就是一种典型的混沌模型,其表达式为 $z_{k+1} = \mu z_k(1 - z_k)$ ,其中, $z_k \in (0, 1)$ 的随机数, $\mu$ 是控制参数,当 $\mu = 4$ 时,系统处于完全混沌状态.

#### 3.2 混沌优化AFSA的改进方法

人工鱼群算法通过模仿鱼群生存方式中的觅食、聚群、追尾等主要行为来解决实际的优化问题.其思路是先针对每个人工鱼个体实现局部寻优,再不断地进行迭代更新,最终找出鱼群中的最优值实现全局寻优<sup>[6]</sup>.但是,人工鱼群算法后期易陷入局部极值,存在收敛精度低和收敛效率差的问题.为了弥补这些不足,本文利用混沌序列随机性、遍历性等特点,对人工鱼群算法采用一种新的优化方法,其具体改进方法如下所示.

(1) 利用Logistic映射产生初始鱼群.首先将人工鱼群活动区域(即寻优参数)的取值范围设为 $(a, b)$ ,并将混沌序列按式 $X = a + z(b - a)$ 的方法生成原始鱼群,再经过多次混沌迭代后,选用得到适应值最大(小)的一组作为最终初始种群,此时生成的人工鱼群质量更佳,分布也更均匀.

(2) 对人工鱼群的觅食、聚群、追尾行为进行混沌优化.在觅食行为中,如果当前人工鱼 $X_i$ 随机选择的下一状态为 $X_j$ ,则需要按照式 $X_{\text{best}} = X_j + m \cdot X_i$ 进行一次混沌搜索,其中 $m = 4 \times \text{step}(1 - \text{step})$ ,如果 $X_{\text{best}}$ 位置优于 $X_j$ ,则替换 $X_j$ 执行 $X_{\text{best}}$ ,否则执行 $X_j$ .同时在执行每次行为时,所需的随机数均由Logistic映射经过多次混沌迭代后产生.

改进后的计算步骤如下所示.

Step1:设置人工鱼群规模、最大迭代次数、视野范围、移动步长、最大尝试次数、拥挤度因子等固定参数值.

Step2:利用混沌模型,对人工鱼群进行初始化.计算每个人工鱼个体相对应的适应值,并将最优个体计入公告板.

Step3:执行混沌优化的聚群、追尾和觅食等行为,对每个人工鱼个体的位置进行更新,并与公告板上的记录进行比较替换.

Step4:判断是否达到最大迭代次数或满足其它设定的结束条件,如果满足则执行step5;否则返回step3,重新进行下一次迭代寻优.

Step5:算法终止,输出最优值.

#### 3.3 混沌人工鱼群算法的测试

本文选用以下两个典型函数对混沌人工鱼群算

法的精准性和可行性进行测试.

测试函数 1:

$$f_1(x) = \frac{\sin(x_1)}{x_1} \frac{\sin(x_2)}{x_2}$$
$$-10 \leq x_i \leq 10, i = 1, 2.$$

该函数在定义域范围内的全局最大值为 1, 最优值为(0,0).

测试函数 2:

$$f_2(x) = x_1^2 + x_2^2 - \cos(18x_1) - \cos(18x_2)$$
$$-2 \leq x_i \leq 2, i = 1, 2.$$

该函数在定义域范围内的全局最小值为 -2, 最优值为(0,0).

本文采用粒子群算法、基本人工鱼群算法与混沌人工鱼群算法作比较. 表 1 所示为以上优化算法均进行 30 次实验后的实验对比结果. 图 2 和图 3 分别表示将两个测试函数通过三种优化算法寻优得到的一次对比截图.

表 1 基于 PSO、AFSA、CAFSA 的优化结果比较

| 优化算法  | 计算值   | $f_1(x)$             | $f_2(x)$            |
|-------|-------|----------------------|---------------------|
| PSO   | 最优值   | (0.048 9, -0.096 2)  | (0.344 7, -0.013 0) |
|       | 最优值   | 0.998 1              | -1.850 6            |
|       | 平均最优值 | 0.966 5              | -1.597 6            |
| AFSA  | 最优值   | (-0.014 7, 0.016 0)  | (0.004 4, -0.014 6) |
|       | 最优值   | 0.999 9              | -1.962 4            |
|       | 平均最优值 | 0.976 7              | -1.778 6            |
| CAFSA | 最优值   | (-0.002 8, -0.002 0) | (-0.001 0, 0.001 2) |
|       | 最优值   | 1.000 0              | -1.999 6            |
|       | 平均最优值 | 0.998 8              | -1.901 4            |

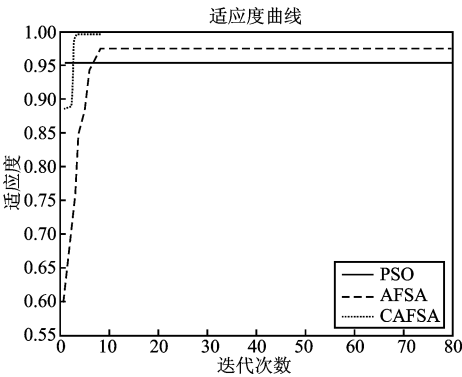


图 2 CAFSA、AFSA 与 PSO 对  $f_1(x)$  函数的优化过程

影响 SVM 分类性能的主要因素有惩罚因子  $C$  和核参数  $\gamma$ , 因此选取最佳的参数组合  $(C, \gamma)$  是 SVM 参数优化的关键. 在由参数  $C$  和  $\gamma$  构建的二维平面上, 运用混沌人工鱼群算法寻优得到最大适应

如图 2 和图 3 所示, CAFSA 算法与 AFSA 和 PSO 相比, 具有更好的收敛性和搜索效率, 能以较少的迭代次数搜索出全局最优值. 表 1 表明: CAFSA 比 AFSA 和 PSO 的寻优效果更好, 实验最优值更接近理论最优值.

3.4 用混沌人工鱼群算法优化 SVM 参数的实现

实验所用语音库由训练集和测试集两部分组成. 在算法寻优过程中, 先利用混沌初始化的人工鱼群作为 SVM 模型的初始参数值, 对语音库中的训练集和测试集进行训练和测试且得到样本的识别率, 再将混沌优化的觅食、追尾、聚群等行为产生的下一代鱼群作为新参数代入模型中, 重复上述过程, 经过持续的迭代寻优, 直到满足混沌人工鱼群算法设定的结束条件, 然后利用最终产生的最优参数组构建训练模型, 对测试样本集进行测试实验. 图 4 为混沌人工鱼群算法优化 SVM 参数的流程图.

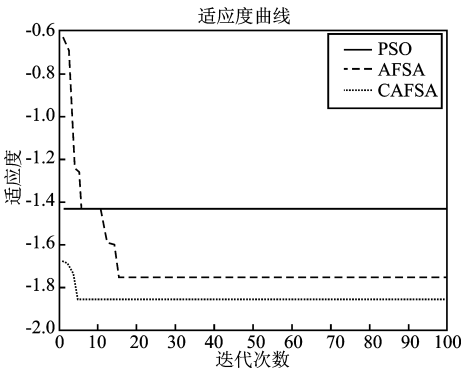


图 3 CAFSA、AFSA 与 PSO 对函数  $f_2(x)$  的优化过程值的的过程如下所示.

Step1: 根据经验对 SVM 参数组合  $(C, \gamma)$  的取值范围及算法中其他参数初值进行设定, 并利用混沌搜索方法对人工鱼群算法进行初始化, 产生初始种群.

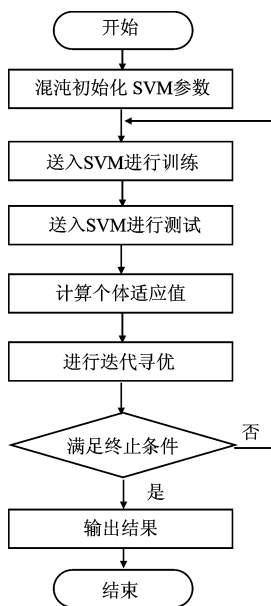


图 4 混沌人工鱼群算法优化 SVM 参数流程图

Step2:将固定值和产生的初始种群均作为参数代入 SVM 模型实现对训练集的训练,并得到相应的训练模型,然后对测试集进行测试,此时该模型的适应值就相当于样本生成的识别率。

Step3:通过混沌优化的觅食、追尾、聚群行为对参数 $(C,\gamma)$ 进行寻优,产生下一代参数 $(C,\gamma)$ 种群。

Step4:再次将获得的新一代参数 $(C,\gamma)$ 种群代入 SVM 模型,并训练和测试样本集,从而得到相应的适应值<sup>[7]</sup>。

Step5:此时与混沌人工鱼群算法设定的终止条件进行比较,如果达到条件要求则执行步骤 6;否则需要返回步骤 3。

Step6:算法结束,为 SVM 模型寻得最优参数组 $(C,\gamma)$ 并得到样本的识别率。

4 实验结果与分析

实验基于 MATLAB R2010b 平台,在一个非特定人、孤立词的语音识别系统中进行.实验选用 16 个人的孤立词发音作为语音库,每个人需要分别对 10 词、20 词、30 词、40 词和 50 词这些不同的词汇量发音 3 次.在该语音库中人为的添加了高斯白噪声,所得信噪比分别为 0 dB、5 dB、10 dB.选用其中 9 人的语音数据作为训练样本,而另外 7 人在相应的 SNR 和词汇量下则构成了测试样本。

表 2~4 表明:在不同词汇量和不同低信噪比的条件下,基于 CAFSA-SVM 模型与基本 AFSA-SVM 模型相比,语音识别率均有不同程度的提高,同时

SVM 训练和识别的时间平均可缩短 18.52%,最多可降低 38.96%,表明收敛速度也明显加快.综上可得本文提出的混沌人工鱼群算法是一种较有效的 SVM 参数优化方法,可使 SVM 有更强的抗噪性能和更好的收敛性。

表 2 SNR = 0 dB 时 AFSA 和 CAFSA 的实验结果对比

| 词汇量 | 实验方法  | 识别率/% | 错分样本数/个 | 时间/s  |
|-----|-------|-------|---------|-------|
| 10  | AFSA  | 86.19 | 29      | 259   |
|     | CAFSA | 88.57 | 24      | 206   |
| 20  | AFSA  | 85.48 | 61      | 943   |
|     | CAFSA | 86.90 | 55      | 773   |
| 30  | AFSA  | 86.67 | 84      | 2 212 |
|     | CAFSA | 86.83 | 83      | 2 055 |
| 40  | AFSA  | 87.14 | 108     | 4 198 |
|     | CAFSA | 87.14 | 108     | 3 616 |
| 50  | AFSA  | 85.81 | 149     | 6 104 |
|     | CAFSA | 85.90 | 148     | 5 096 |

表 3 SNR = 5 dB 时 AFSA 和 CAFSA 的实验结果对比

| 词汇量 | 实验方法  | 识别率/% | 错分样本数/个 | 时间/s  |
|-----|-------|-------|---------|-------|
| 10  | AFSA  | 91.90 | 17      | 228   |
|     | CAFSA | 92.38 | 16      | 205   |
| 20  | AFSA  | 92.38 | 32      | 912   |
|     | CAFSA | 92.62 | 31      | 904   |
| 30  | AFSA  | 91.75 | 52      | 2 358 |
|     | CAFSA | 92.22 | 49      | 2 067 |
| 40  | AFSA  | 91.07 | 75      | 4 846 |
|     | CAFSA | 91.31 | 73      | 3 892 |
| 50  | AFSA  | 90.95 | 95      | 7 855 |
|     | CAFSA | 90.95 | 95      | 5 383 |

表 4 SNR = 10 dB 时 AFSA 和 CAFSA 的实验结果对比

| 词汇量 | 实验方法  | 识别率/% | 错分样本数/个 | 时间/s  |
|-----|-------|-------|---------|-------|
| 10  | AFSA  | 91.43 | 18      | 287   |
|     | CAFSA | 93.33 | 14      | 208   |
| 20  | AFSA  | 92.62 | 31      | 1 121 |
|     | CAFSA | 92.86 | 30      | 851   |
| 30  | AFSA  | 92.22 | 49      | 3 468 |
|     | CAFSA | 94.13 | 37      | 2 117 |
| 40  | AFSA  | 93.81 | 52      | 4 676 |
|     | CAFSA | 93.81 | 52      | 3 493 |
| 50  | AFSA  | 93.24 | 71      | 6 626 |
|     | CAFSA | 93.33 | 70      | 5 392 |

## 5 结束语

本文通过测试函数的测试和比较,验证了混沌人工鱼群算法(CAFSA)的准确性和有效性.同时运用该方法可以实现对 RBF 核函数的 SVM 参数组合( $C, \gamma$ )的优化,并适用于语音识别系统.实验结果表明混沌人工鱼群算法比基本人工鱼群算法更易跳出局部极小,收敛速度更快,系统识别率更高.但这种方法仍然存在一些不足之处,在某些条件下(如 0 dB40 词、5 dB50 词和 10 dB40 词)识别率并没有显著的提高,且总体最佳参数的寻优时间较长,所以该问题有待进一步地研究和解决.

### 参考文献:

- [1] 邓乃扬,田英杰.数据挖掘中的新方法:支持向量机[M].北京:科学出版社,2004.
- [2] 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.
- [3] 李晓磊,邵之江,钱积新.一种基于动物自治体的寻优

模式:鱼群算法[J].系统工程理论与实践,2002,22(11):32-38.

- [4] 高雷阜,赵世杰,高晶.人工鱼群算法在 SVM 参数优化选择中的应用[J].计算机工程与应用,2013,49(23):86-90.
- [5] Jiang Mingyan, Cheng Yongming. Simulated annealing artificial fish swarm algorithm [C] // Proceedings of the 8th World Congress on Intelligent Control and Automation. Jinan, China:IEEE, 2010: 1590-1593.
- [6] 李晓磊,薛云灿,路飞,等.基于人工鱼群算法的参数估计方法[J].山东大学学报,2004,34(3):84-87.
- [7] 杨利红.基于 AFSA 的 SVM 参数优化及其应用[D].太原:太原理工大学,2012.

### 作者简介:

朱文静 女,(1990-),硕士研究生.研究方向为语音信号处理.

白 静(通讯作者) 女,(1965-),教授,硕士生导师.研究方向为语音信号处理. E-mail: bj613@126.com.

(上接第 88 页)

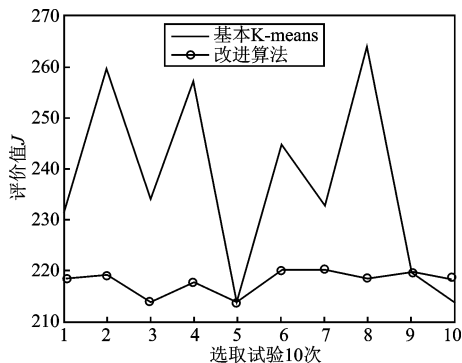


图4 Glass 数据集

在初始聚类中心随机选择导致聚类结果波动较大和聚类数目确定的缺点进行改进.但在确定聚类数目值的时候,采用了与经验值比较的方法,影响了准确率.下一步工作是提出一种合理判定聚类结束条件的方法,减少人为因素影响,以便能更加准确地获得聚类数目.

### 参考文献:

- [1] 雷小锋,谢昆青,林帆,等.一种基于 K-Means 局部最优的

高效聚类算法[J].软件学报,2008,19(7):1683-1692.

- [2] 贺玲,吴玲达,蔡益朝.数据挖掘中的聚类算法综述[J].计算机应用研究,2007,23(1):10-13.
- [3] 孙吉贵,刘杰,赵连宇.聚类算法研究[J].软件学报,2008,19(1):48-61.
- [4] 胡伟.改进的层次 K 均值聚类算法[J].计算机工程与应用,2013,49(2):157-159.
- [5] 徐沁,罗斌.结合 means-shift 与 MST 的 K-means 聚类算法[J].计算机工程,2013,39(12):214-210.
- [6] Galluccio L, Michel O, Comon P, et al, Graph based kmeans clustering[J]. Signal Processing, 2012, 92(9): 1970-1984.
- [7] 李翔宇,王开军,郭躬德.基于网格最小生成树的聚类算法选择[J].模式识别与人工智能,2013,26(1):34-41.

### 作者简介:

贾瑞玉 女,(1965-),副教授,硕士生导师.研究方向为智能计算与数据挖掘. E-mail: 764702886@qq.com.

李 振 男,(1991-),硕士生.研究方向为数据挖掘.