

◎图形、图像、模式识别◎

改进的 PSO 在说话人辨识中的应用

骆瑞玲^{1,2}, 李明², 李睿²LUO Rui-ling^{1,2}, LI Ming², LI Rui²

1.石河子大学 信息科学与技术学院, 新疆 石河子 832000

2.兰州理工大学 计算机与通信学院, 兰州 730050

1.College of Information Science and Technology, Shihezi University, Shihezi, Xinjiang 832000, China

2.School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

E-mail: lrl_inf@shzu.edu.cn

LUO Rui-ling, LI Ming, LI Rui. Application of improved PSO in speaker identification. *Computer Engineering and Applications*, 2010, 46(2): 135-137.

Abstract: Aiming at the shortcoming of Particle Swarm Optimization(PSO) which is easily relapsing into local extremum, an improved PSO is proposed in this paper. This approach applies the evolution speed factor as the Trigger conditions to stochastically disturb the local optimal solution. The improved PSO algorithm can not only improve extraordinarily the convergence velocity in the evolutionary optimization, but also can adjust the balance between global and local exploration suitably. Then a speaker identification approach using this improved algorithm to train SVM is presented. The SVM can receive the optimal hyper plane with less support vectors by the improved PSO, and then the training samples are reduced and the identification speed is improved.

Key words: speaker identification; Particle Swarm Optimization(PSO); evolution speed factor; extremum disturbance

摘 要:针对 PSO 算法容易陷于局部极值的缺点,提出了一种改进的 PSO 优化算法(IPSO)。该算法根据粒子进化速度对粒子个体极值进行自适应扰动,使粒子及时跳出局部极值点而继续优化,从而扩大粒子搜索范围。改进后的 PSO 算法加快了收敛速度,能够很好地调整算法的全局与局部搜索能力之间的平衡。同时,给出了应用 IPSO 算法训练支持向量机的方法,并将其应用于说话人辨识。改进后的 PSO 可以使 SVM 用较少的 SV 取得最优分类面,从而减少 SVM 的训练量,提高了说话人辨识速度。

关键词:说话人辨识;粒子群优化算法;速度进化因子;极值扰动

DOI: 10.3778/j.issn.1002-8331.2010.02.041 文章编号: 1002-8331(2010)02-0135-03 文献标识码: A 中图分类号: TP391

1 引言

说话人辨识就是判定待测说话人的语音属于多个参考说话人之中的某一个,它可以应用于声音拨号、刑侦破案、机要保密、玩具和家用电器等领域。

支持向量机(Support Vector Machine, SVM)^[1]是一种基于结构风险最小化的统计学习方法,在解决小样本、非线性及高维数模式识别问题表现出特有的优势,由于其出色的学习性能和较强的推广能力,已成为机器学习领域的一个研究热点,在说话人识别中也得到广泛的应用^[2-3]。SVM 训练的实质为求解一个带有界约束和线性不等式约束的凸二次规划问题,对大规模的数据,其计算量大运算速度慢。粒子群优化算法^[4-5](Particle Swarm Optimization, PSO)是由 J.Kennedy 和 Eberhart 等人受鸟群和鱼群觅食行为的启发而提出的一种智能优化算法,它实质上是一种迭代优化算法。与遗传算法不同,PSO 不是通过

遗传算子进化,而是通过群体之间的协作与竞争来实现对解空间的快速搜索。PSO 算法简单、需要调节的参数少、在寻优过程中计算量大大减少,节省了寻优的时间。目前已广泛应用于函数优化、神经网络训练等应用领域。但是,粒子群优化算法根据全体粒子和粒子自身的搜索经验向着最优解的方向发展,在进化后期收敛速度变慢,算法收敛精度不高。为此,许多学者提出了不同的改进算法,如文献[4]提出以收敛速度的快慢自适应地对惯性权重进行调节的 PSO 算法;Berhart 和 Shi^[5]提出了惯性权重线性递减的 PSO 算法,在优化方程的性能上有明显的效果;文献[6]和文献[7]从不同角度提出了动态改变惯性权重的自适应粒子群算法,改善了粒子摆脱局部极值的能力,提高了收敛精度。但这些策略主要是调整参数 ω 来提高收敛速度和精度,都不是从 PSO 收敛于局部极值的根本原因上采取改进策略。为了增强全局搜索能力、提高收敛速度,从 PSO 算法陷入局部

基金项目:甘肃省自然科学基金(the Natural Science Foundation of Gansu Province of China under Grant No.2007GS04782)。

作者简介:骆瑞玲(1980-),女,助教,硕士研究生,主要研究方向:生物身份识别、模式识别;李明(1959-),男,教授,主要研究方向为数据挖掘和人工智能、智能信息处理;李睿(1971-),女,副教授,主要研究方向为模式识别、智能信息处理。

收稿日期:2009-01-13 修回日期:2009-06-16

极值的根本原因入手,提出一种根据进化速度对粒子的个体极值进行自适应扰动的 PSO 算法,改进后的 PSO 使粒子能够及时跳出局部极值而增强其全局搜索能力,同时提高其收敛速度。并将 IPSO 训练支持向量机的方法用于说话人辨识,实验表明提出的方法提高了说话人识别的速度。

2 支持向量机

支持向量机^[1]的核心思想是在特征空间寻找最优超平面将两类样本无误的分开,且分类间隔最大。从某种程度上说,是样本中少量的支持向量决定了最优超平面。

对于给定的数据集, $\{x_i, y_i\}, i=1, 2, \dots, N, y_i \in \{-1, +1\}, x_i \in \mathcal{R}^n$, 超平面由以下方程描述:

$$\omega \cdot x + b = 0 \quad \omega \in R^n, b \in R \quad (1)$$

求最优分类面的问题可以归结为下面的优化问题:

$$\min Q(\omega, \xi) = \frac{\|\omega\|^2}{2} + C \sum_{i=1}^n \xi_i \quad (2)$$

$$y_i(\omega \cdot x_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad i=1, 2, \dots, N \quad (3)$$

其中,参数 ξ_i 为松弛变量, C 是惩罚系数。若样本点不是线性可分的,则首先通过一个非线性映射 $\phi: R^d \rightarrow F$, 将其投影到一个高维空间使之线性可分,则原空间的点积在高维核空间 F 可以用 Mercer 核^[2]来表示。引入 Lagrange 乘子 a_i , 上述方程的求解可以转化为下面的二次规划问题:

$$\begin{aligned} \max W(a) &= \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j a_i a_j K(x_i, x_j) \\ \text{s.t.} \quad \sum_{i=1}^n y_i a_i &= 0, 0 \leq a_i \leq C, i=1, 2, \dots, N \end{aligned} \quad (4)$$

其中 $K(x_i, x_j)$ 是核函数,其中与 $a_i > 0$ 对应的样本点 x_i 即为支持向量。最后的分类的决策函数定义为:

$$f(x) = \text{sgn} \left[\sum_{i=1}^n a_i y_i k(x_i \cdot x) + b \right] \quad (5)$$

3 基本 PSO 优化算法

PSO 算法^[3-6]建立的基础是将优化问题的一个可能的候选解假设为一个无体积和质量的粒子,在 d 维空间粒子通过自身和其他粒子经验来调整自己的位置。有 n 个粒子组成的粒子群,其中第 i 个粒子在 d 维解空间的速度为 $v_i = (v_{i1}, v_{i2}, \dots, v_{id})^T, x_i = (x_{i1}, x_{i2}, \dots, x_{id})^T$ 表示粒子的位置,可根据它来计算相应的适应值,根据适应值可衡量粒子 i 的优劣。每个粒子当前自身找到的个体最优值 $pbest_{id}$ 和整个种群找到的全局最优值 $gbest_d$,在迭代计算过程中将被记录下,粒子通过跟踪两个历史最优位置并根据:

$$v_{id}^{(t+1)} = \omega v_{id}^t + c_1 \text{rand}_1() (pbest_{id}^t - x_{id}^t) + c_2 \text{rand}_2() (gbest_{id}^t - x_{id}^t) \quad (6)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \quad i=1, 2, \dots, n \quad (7)$$

来更新自己的速度和位置,搜索解空间中的最优解。其中: ω 为惯性权值, ω 较大时算法具有较强的全局搜索能力, ω 较小则算法倾向于局部搜索。 c_1 和 c_2 都为正常数,称为加速系数,通常 $c_1=c_2=2$; $\text{rand}_1()$ 和 $\text{rand}_2()$ 是两个 $[0, 1]$ 之间的随机数, $pbest_{id}$ 是当前粒子的历史最优位置, $gbest_{id}$ 是整个粒子群的历史最优位置, $v_{id} \in [-v_{\max}, v_{\max}]$ 。

4 IPSO 及在说话人辨识中的应用

对与文本无关的说话人识别而言,提取的语音特征矢量较

多,用标准的 SVM 训练计算复杂度较大,训练和识别速度慢。 PSO 算法是一种新型的具有高效全局搜索能力的进化算法,调整参数少,容易实现,收敛能力强,实现时间短。将 PSO 算法应用于训练支持向量机,可以很好解决二次规划问题。

4.1 改进的 PSO 优化算法(IPSO)

文献[9]中的定理 1 证明了 PSO 算法可以没有粒子速度的概念,避免了人为确定参数 $[-v_{\max}, v_{\max}]$ 而影响粒子的收敛速度和收敛精度。简化粒子群优化方程可以表示为:

$$x_{id}^{t+1} = \omega x_{id}^t + c_1 r_1 (p_{id}^t - x_{id}^t) + c_2 r_2 (p_{gd}^t - x_{id}^t) \quad (8)$$

式(8)右边的第 1 项为“历史(history)”部分,表示过去对现在的影响;第 2 项为“认知(cognition)”部分,表示粒子对自身的思考;第 3 项为“社会(social)”部分,表示与相邻粒子的比较和模仿,实现粒子间的信息共享与合作。这样在优化的过程中不必考虑速度这一项,可以减少计算量,加快寻优速度。以进化速度因子为触发条件,对简化的粒子群优化算法进行改进。受文献[10]的启发,将上式写为:

$$x_i^{t+1} = \omega x_i^t + c_1 r_1 (pbest_i^t - x_i^t) + c_2 r_2 (p_g - x_i^t) \quad (9)$$

从上式可以看出, r_1, r_2 随着粒子群数目和迭代次数的变化而变化。首先给出全局进化速度因子 H ^[6]和个体进化速度因子 H' ^[7]分别为:

$$H = \frac{F(gbest_i^{t-1})}{F(gbest_i^t)} \quad (10)$$

$$H' = \frac{F(pbest_i^{t-1})}{F(pbest_i^t)} \quad (11)$$

其中, $F(\cdot)$ 为适应度函数。显然, $0 < H \leq 1, 0 < H' \leq 1$ 。有前面的分析知,基本 PSO 通过惯性权重的线性减少来调节全局和局部搜索能力,变化比较单一,使 PSO 不能很好地适应复杂情况,在迭代后期则因全局搜索能力变弱而易陷入局部最优。为了改善式(9)的全局搜索能力,通过粒子进化速度的快慢对个体极值随机扰动来改变粒子搜索空间的位置,使其及时跳出局部最优搜索区域而继续搜索,从而提高粒子的全局搜索能力。而其他的粒子仍然按式(9)进化。在简化 PSO 算法中,采用进化速度因子作为触发条件,对个体极值进行随机扰动,极值扰动算子定义为:

$$pbest_i^t = pbest_i^t * (1 + \gamma H') \quad (12)$$

则改进后的粒子群优化方程为:

$$x_i^{t+1} = \omega x_i^t + c_1 r_1 \{pbest_i^t * (1 + \gamma H') - x_i^t\} + c_2 r_2 (p_g - x_i^t) \quad (13)$$

其中 γ 的取值如下:

$$\gamma = \begin{cases} 0 & 0 < H' < 1 \\ \text{rand}(0, 1) & H' = 1 \end{cases} \quad (14)$$

H' 考虑了每个粒子运行的历史,也反映了每个粒子的进化速度,即 H' 值越小,粒子进化速度越快。在进化初期,粒子群很快收敛到最佳粒子附近,当经过了一定的迭代次数之后,粒子的速度矢量趋于 0,粒子有可能陷入停滞不前的状态, H' 值保持为 1。此时对局部最优值进行随机扰动,使其向新的方向扩展,但仍保留“生前”的记忆,即它还记得全局最优值和个体最优值。这样可以扩大粒子的搜索范围,增强其全局搜索能力。而 H 反映了粒子群进化速度,即 H 值越小,进化速度越快。若 $H=1$,可以断定离子群停止进化或者找到了最优解。在该算法中惯性权重仍然采用线性递减策略^[5]。

改进的 PSO 算法步骤如下:

步骤 1 初始化粒子群 (a_1, a_2, \dots, a_n) , 确定群体规模 n , 最初权重 ω_{init} 和最终权重 ω_{end} , 设定算法总的迭代次数 $iter_{总}$ 。

步骤 2 将每个粒子的个体极值 $pbest_i$ 设置为当前位置, 利用适应函数即式(15)计算每个粒子的适应度, 取适应度最好的粒子所对应的个体极值作为最初的全局极值 p_g ;

步骤 3 根据式(10)计算 H ;

步骤 4 如果 $H=1$ 或达到最大迭代次数, 执行步骤 10, 否则执行步骤 5;

步骤 5 根据式(11)计算 H' , 若 $H'=1$, 产生随机数 γ 并根据(13)更新粒子的位置; 反之, 若 $0 < H' < 1$ 粒子仍然按式(9)进化; 更新粒子的惯性权重;

步骤 6 由式(15)评价每个粒子的适应值;

步骤 7 将每个粒子的适应值与其 $pbest_i$ 对应的适应值比较, 若优, 更新 $pbest_i$, 否则保留原值;

步骤 8 将更新后的每个粒子的 $pbest_i$ 与全局极值 p_g 比较, 若优, 更新 p_g , 否则保留原值;

步骤 9 将迭代次数加 1, 并返回步骤 3;

步骤 10 输出 p_g , 算法结束。

4.2 IPSO 训练 SVM

由 SVM 的数学模型可知, 需要求解的是支持向量系数 a_i ($i=1, 2, \dots, m$), 粒子编码为 (a_1, a_2, \dots, a_m) , 自适应函数取为:

$$f(p) = \sum_{i=1}^m a_i - \frac{1}{2} \sum_{i,j=1}^m a_i a_j y_i y_j K(x_i, x_j) \quad (15)$$

根据 PSO 算法和 SVM 的要求, 粒子的初始位置应满足条件 $\sum_{i=1}^m a_i y_i = 0$ 和 $0 \leq a_i \leq C$ ($i=1, 2, \dots, m$)。PSO 在优化前期收敛速度很快, 但在优化后期收敛速度变得缓慢, 从而使搜索空间受到极大的限制, 直接影响到 SVM 的训练速度和性能。这主要是粒子群难以摆脱局部极值的原因。在简化粒子群算法中, 以粒子的进化速度因子为触发条件, 对个体极值进行随机扰动, 来增强其全局搜索能力, 找出最优的支持向量系数, 从而提高 SVM 的训练速度和分类性能。采用改进 PSO 训练 SVM 的过程如下:

(1) 初始化 PSO 的参数, 并按照文献[11]提出的方法对粒子群进行初始化, 设置当前迭代数 $iter=1$;

(2) 对每个粒子执行下述操作: 按式(15)计算各粒子的适应度, 更新相应的 $pbest_i$ 和 p_g ; 按式(11)计算 H' 根据 4.1 节中的步骤 5 更新粒子的惯性权重和位置矢量;

(3) $iter=iter+1$, 如果 $iter > iter_{总}$, 则 p_g 即为 SVM 的训练结果, 否则返回第(2)步重复上述过程;

(4) 根据更新的各粒子的适应值, 计算支持向量系数 a_i ($i=1, 2, \dots, m$);

(5) 根据公式(5)计算决策函数。

5 实验结果及分析

采用 WINDOWS 自带的录音软件录音, 采样频率 11 025 Hz, 16 bit 量化, 普通环境下, 对 30 个(20 男 10 女)说话人进行录音, 每名录音人员均用普通话, 每个人录制 8 遍, 前 5 遍用于训练, 后 3 遍用于识别, 且每人录音长度不少于 2 分钟。考虑到说话人语音特征会随着时间改变, 故数据分三次采样, 每次间隔半个月左右。对得到的语音数据去静音, 预加重 $1-0.97z^{-1}$, 加

窗(Hamming 窗, 帧长 32 ms, 帧移 16 ms), 每帧提取 12 维的 MFCC 倒谱系数(不包括 0 阶)及其一阶差分系数, 形成 24 维的语音特征矢量序列。

在实验中, 分别用 PSO 和 IPSO 算法训练 SVM, 粒子数都取为 20, 实验独立运行 20 次, 最大迭代次数为 400, $c_1=c_2=2$ 。在 PSO 算法中, 惯性权重 ω 设置为 0.7, 在 IPSO 算法中 ω 的取值从 0.7 减小至 0.4。对每两个人建立一个 SVM, 对 30 个人的训练语音将建立 435 个 SVM, 并且将训练结果作为说话人模板存储下来。在实验中选择径向基函数 $k(x, x_i) = \exp(-\frac{\|x-x_i\|^2}{\sigma^2})$ 和

多项式函数进行了比较, 在 RBF 中惩罚系数 C 为 100, 对参数 σ 选择不同的值进行比较, 结果见表 1。表 2 和表 3 分别给出了 IPSO 和 PSO 训练 SVM 时在不同说话人数(N)下对识别率的影响及不同方法对说话人识别性能的影响。

表 1 不同核函数对识别效果的影响

方法	RBF		多项式($d=2$)
	$\sigma=1.5$	$\sigma=2.5$	
SVM	94.05%	89.32%	90.21%
PSO-SVM	96.20%	92.10%	91.70%
IPSO-SVM	96.30%	93.50%	94.31%

表 2 用不同方法在不同说话人数下识别率

	RR/(%)			
	$N=10$	$N=15$	$N=25$	$N=30$
SVM	95.7	95.1	94.8	94.6
PSO-SVM	96.9	96.7	96.4	96.2
IPSO-SVM	97.3	97.2	96.9	96.7

表 3 不同方法的识别性能比较

	SVM	PSO-SVM	IPSO-SVM
RR/(%)	94.6	96.2	96.7
T/s	3.46	2.35	1.21
SV	897	473	224

从表 1 可以看出, 用 RBF 核函数, 各方法识别率都高于多项式核函数。用 RBF 核函数当参数 σ 取 1.5 时, 各方法取得了较好的识别效果。且随 σ 的增大各算法的识别效果变差, 显然核函数及参数的选择对实验结果有显著影响。因此, 在后续实验中采用 RBF 核函数且取 $\sigma=1.3$ 。

比较表 2 和表 3 可以看出, 用 IPSO 算法训练 SVM 进行说话人识别时, 虽然随着说话人数的增加识别率有所下降, 总体上用 IPSO 算法训练 SVM 时识别效果仍然好于 PSO 算法。且 IPSO 通过粒子寻优用较少的支持向量数取得了最优的分类平面, 由于真正对分类起作用的支持向量数目较少, 其计算量也小, 从而提高了说话人辨识的速度。

6 结论

从 PSO 容易陷于局部极值的原因入手, 提出了一种新的 IPSO-SVM 的说话人辨识方法。IPSO 算法以进化速度因子为触发条件, 对粒子群的局部最优值随机扰动。改进后的 PSO 算法在收敛速度、收敛精度和全局搜索能力方面均有一定的改善。实验表明, 该方法相对标准的 SVM 和 PSO-SVM 方法在辨识时间上有明显的优势。