



基于遗传算法的 SVM 参数优化研究

芮铭敏¹

(东南大学计算机科学与工程学院, 南京, 211189)

摘要: 支持向量机(SVM)是机器学习领域的一种比较新颖的方法, 能够克服维数灾难等传统统计学问题, 能够较好地处理小样本学习问题, 而其中的核函数参数和惩罚因子至关重要。作者利用文献[6]所提供的实验结果进行比较性的实验, 首先验证遗传算法(GA)的 RFB 核函数稳定性以及与网格法比较实验结果的优劣, 再利用 GA 对 SVM 的参数进行进一步的优化实验, 然后设计了改进后的自适应 GA(AGA), 并进行同样的实验。通过实验的结果对比, GA 和 AGA 都表现出有更好的效果, 为更多基于改进的 GA 的 SVM 参数优化提供参考。

关键词: 支持向量机; 分类; 遗传算法; 自适应;

Study on optimizing the SVM parameters based on GA

RUI Ming-min¹

(School of Computer Science and Engineering, Southeast University, Nanjing, 211189)

Abstract: Support Vector Machine is an emerging method in the field of Machine Learning. It could overcome some traditional difficulties like Curse of Dimensionality, and deal with the problem of small sample learning. Kernel function parameter and cost parameter are the most important factors which affect the result of SVM. We first use GA and AGA methods to show the stability of RFB kernel function, and compare with the grid searching method in reference [6] with the same data set, then do a further test. The results provide a reference for optimizing the kernel function parameters by methods based on GA.

Key words: Support Vector Machine; Classification; Genetic Algorithm; Adaptive

分类(Classification)是数据挖掘(Data Mining)的一种重要的分析方法, 对于分类本身的研究方法很多, 其中有一种较新的方法——支持向量机(Support Vector Machine, SVM)^[1]。SVM 是由 Vapnik 等人在 1995 年首先提出的, 有着严格的统计学习理论基础, 针对小样本、非线性、高维度模式识别问题有着很大的优势。由于有着很好的泛化能力和学习能力, 21 世纪以来, SVM 理论被应用到各个领域, 包括函数逼近、回归分析、模式识别等。SVM 算法将问题最终转化为一个求凸二次规划的对偶问题, 而其中的惩罚因子 c 和核参数 g 的选择直接影响最终的结果^[2]。目前优化 SVM 参数的方法有很多, 比如网格法、粒子群算法、神经网络, 遗传算法等, 而本文要解决的问题是离散的函数优化问题, 所以选择了较为经典遗传算法。遗传算法(Genetic Algorithm, GA)是由美国 Michigan 大学 Holland 教授等人创立的一种概率搜索算法, 基于 Darwin 的进化论

和 Mendel 的遗传学, 利用计算机模拟生物的进化过程, 最基本的操作是选择(Select)、交叉(Crossover)和变异(Mutation)。GA 算法不容易陷入局部最优解, 具有可扩展性和并行计算能力等优点, 应用于函数优化、组合优化、机器学习等领域^[3]。

本文首先介绍了 SVM 的理论知识, 并提出本文需要解决的问题: 优化核函数参数 g 和惩罚因子 c ; 然后叙述了 GA 的基本原理和算法流程, 并提出改进的 GA 设计方法; 接着利用文献[6]提供的 SVM 的参数优化结果, 利用 GA 和改进后的 GA 进行对比性实验, 并总结 GA 和 AGA 相对于文献[6]所用方法的优势和不足。

1 支持向量机 SVM

SVM 是统计学理论中比较新的方法, 它不像传统的统计学那样需要大量的样本才能接近理论值。SVM 有可靠的理论基础和简洁的数学形式, 目前正处于广泛的应用和发展阶段, 本文主要研究的是利用遗传算法对 SVM 的 RFB 核函数参数进行优化。

作者简介: 芮铭敏, (1988-), 男, 硕士研究生, E-mail: wise_1001@163.com.



1.1 基本原理

在 n 维欧氏空间 R^n 上, 有训练集 $T=\{(x_1, y_1), \dots, (x_l, y_l)\} \in (R^n \times Y)$ [4], 其中 $x_i \in R^n$ 为输入指标向量, 有 n 个特征, $y_i \in \{-1, +1\}$ 为输出指标, 或称样本标签, $i=1, \dots, l$ 。用于分离的超平面方程为 $w \cdot x + b = 0$, 其中 w 为空间中的权向量, b 为阈值。如图 1, 满足条件的分割线并不唯一, 而最优的 w 和 b 会使 H_1 到 H_2 的间隔最大, 最大间隔的法向量即为向量 w 。

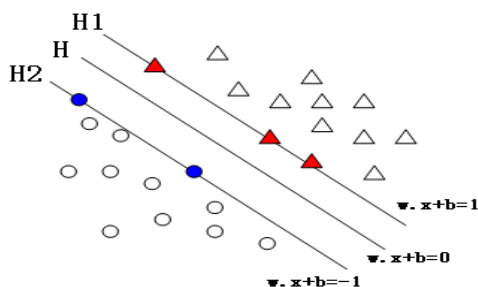


图 1 最优超平面的几何意义

至此, 该问题转化为求 H_1 和 H_2 距离 $2/\|w\|$ 的最大值, 最终变成求式(1), 其中 $\varepsilon = (\varepsilon_1, \dots, \varepsilon_l)^T$, ε_i 为松弛变量, c 为惩罚因子。

$$\begin{cases} \min_{w, b, \varepsilon} & \frac{1}{2} \|w\|^2 + c \sum_{i=1}^l \varepsilon_i, c > 0, \\ \text{s.t.} & y_i((w \cdot x_i) + b) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0, i=1 \dots l \end{cases} \quad (1)$$

并不是所有的问题都是线性可分类的, 对于非线性分类问题, 需要通过选择适当的核函数将其转为线性可分类问题。举一个最简单的例子, 分割线是 $x_2 + y_2 = 1$, 可以令 $u = x_2$, $v = y_2$, 问题就转为了线性问题 $u + v = 1$, 这里的核函数为

$$K(x, y) = u(x) \cdot v(y) \quad (2)$$

其中 $u = u(x)$, $v = v(y)$ 。而此时空间也从 2 维变成了 4 维。所以对于非线性问题, 最主要的就是找到合适的核函数 $K(x, y)$, 将输入样本的空间 R 映射到更高维度的空间 T , 是问题转化成线性问题。

1.2 参数的选择

核函数是对最后效果有决定性影响的因素之一, 所以核函数的选择自然就成了一件重要的事情。目前最常用核函数主要有 4 类:

(1) 线性核函数 $K(x, y) = x \cdot y$;

(2) 多项式核函数 $K(x, y) = [(x, y) + 1]^q$ [5], 得到 q 阶多项式分类器;

(3) 径向基函数 RBF $K(x, y) = \exp\{-\frac{\|x - y\|^2}{\sigma^2}\}$;

(4) Sigmoid 核函数 $K(x, y) = \tanh(v(x, y) + c)$ 。

作者将采用平均效果较好的 RBF 核函数, 而本实验的主要目的是讨论 RBF 核函数参数 g 以及式(1)中 c 的优化选择。

SVM 的参数选择方法很多, 最简单的有经验选择法、实验试凑法, 还有比较常用的梯度下降法、交叉验证法、网格法, 另外还有遗传算法、粒子群算法、免疫算法、神经网络等智能算法[6], 本文重点研究的是利用遗传算法进行 SVM 的参数优化。

2 遗传算法 GA

GA 有模式定理和积木块假设作为数学理论基础, 前者证明 GA 存在找到全局最优解的可能, 是遗传算法的主要定理; 后者则证明 GA 具有找到最优解的能力。

2.1 GA 概述

任何问题想要用 GA 解决, 首要问题是编码, 它直接影响后面的流程及最终解的优劣, 一般方法有二进制码、格雷码等, 不论哪种, 编码必须有完备性。接着就是适应度函数的选择, 适应度是衡量某一代种群所能达到或接近最优解的程度, 它能抑制欺骗问题的产生, 是整个算法的驱动力。然后就是选择、交叉和变异三个重要环节, 其中方法很多, 是研究的热点所在, 图 2 是 GA 的基本流程图。

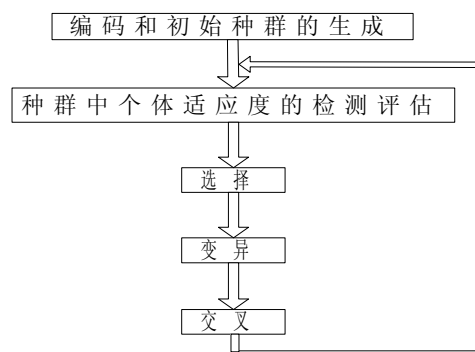


图 2 遗传算法 GA 的基本流程图



2.2 自适应遗传算法 AGA

GA 中的交叉概率 pc 和变异概率 pm 的选择是影响算法性能的关键所在。为了保持种群的多样性,在种群集中或者趋于局部最优解时, pc 和 pm 要增加;而当种群比较分散时, pc 和 pm 就减少。而为了保护最优个体,对于适应度高于平均适应度的个体, pc 和 pm 就减少;反之增加。

我们所说的 GA 又称简单遗传算法(Simple GA, SGA),它不能根据种群的优良来调整 pc 和 pm 的大小,而找到一个固定的最佳 pc 和 pm 值很困难。自适应遗传算法(Adaptive GA,AGA)在 SGA 改进算法的一种,它根据种群的优良适当调整 pc 和 pm 的大小,能够更好地保证 GA 的收敛性,实验者可以根据实验的研究目的进行相应的改进。

3 AGA_SVM 算法设计

3.1 GA_SVM 算法

在寻找 SVM 参数最优解时,有时会有试探法、经验法等偶然性很强的方法,更多的是用网格搜索法。如果对结果要求不是很高,或者更强调时间更少,可以通过网格搜索找到不错的解。然而网格搜索主要的缺点是:当搜索区间和步长确定后,就只能在确定的一些离散点中找最优值,而很可能错过更优的解。

因为求 SVM 最优参数的实质是二次凸规划问题,即求解函数最优值,而 GA 的最主要的用处之一就是函数最优化,并且效果一般而言都不错。因为 GA 是一种高效的随机搜索算法,不会将搜索范围限定在某一些解中,克服了网格搜索法的缺点,可以通过多次的尝试,找到比网格搜索法更好的解。GA_SVM 的算法流程可以参照图 3 的步骤 1-4, 8-10, 12 和 14, GA 的一些参数设定为:最大遗传代数 MAXGEN=200,种群数量 NIND=50,代沟 GAP=0.7,个体二进制编码位数 PRECI=30,交叉率 $pc=0.7$,变异率 $pm=0.01$ 。

3.2 AGA_SVM 算法

由于保持种群多样性和保护最优个体本身就存在着矛盾,本文做出适当的权衡,在保护种群最优值的前提下追求种群多样性。图 3 为 AGA_SVM 对训练集进行参数优化的具体设计流程。

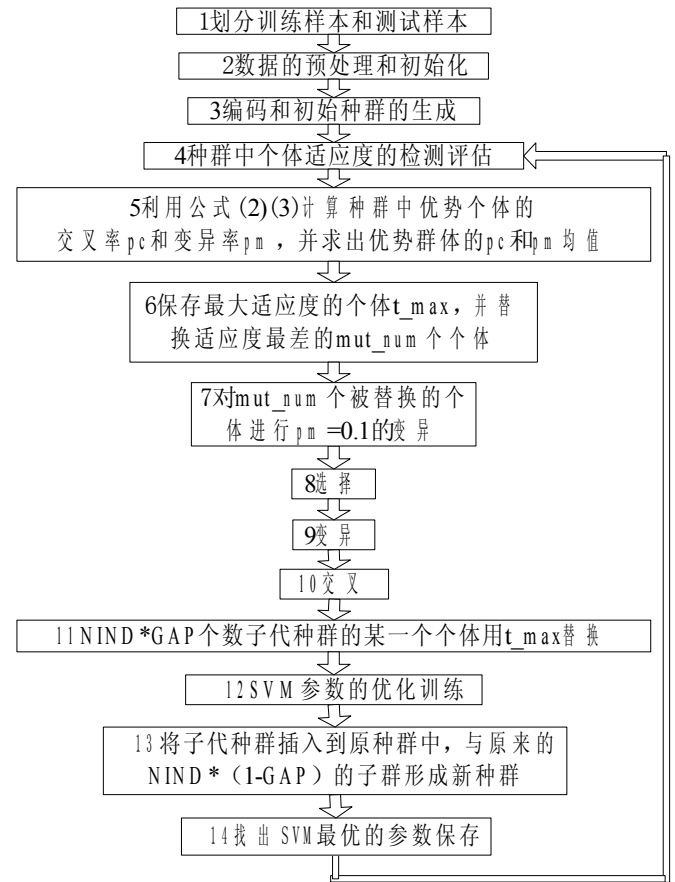


图 3 AGA_SVM 对训练集参数优化的设计流程图

(1)在处理最重要的交叉率 pc 和变异率 pm 时,本文借鉴文献[10]所提供的公式计算优势个体(适应度不小于平均值的个体) pc 和 pm ,即公式(3)和(4)。其中 f_i 表示当前个体的适应度, f_{avg} 表示当前种群的平均适应度, f_{max} 表示最好适应度, NULL 表示不计算当前群体的 pc 和 pm 。

$$pc_i = \begin{cases} k1 - k2 \times (f_i - f_{avg}) / (f_{max} - f_{avg}) & f_i \geq f_{avg} \\ NULL & f_i < f_{avg} \end{cases} \quad (3)$$

$$pm_i = \begin{cases} k3 - k4 \times (f_i - f_{avg}) / (f_{max} - f_{avg}) & f_i \geq f_{avg} \\ NULL & f_i < f_{avg} \end{cases} \quad (4)$$

pc 和 pm 取值一般为 0.4-0.99 和 0.001-0.1,作者取 $k1=0.9, k2=0.6, k3=0.08, k4=0.05$,将 pc 和 pm 控制在中间段,即 pc 为 0.6-0.9, pm 为 0.03-0.08。

(2)作者设计的 AGA 算法中,没有过多强调个体的 pc 和 pm ,而是以种群为整体,即各代种群的 pc 和 pm 不同,而同一代优势个体的 pc 和 pm 为这一代优势群体的均值(非优势个体不考虑)。利用



公式(5)和(6)求出当前种群的优势群体的平均 pc 和 pm ，其中 $NIND$ 为种群的数量，其余变量与公式(3)(4)意义相同。

$$pc = \sum_{i=1}^{NIND} pc_i (f_i \geq f_{avg}) \quad (5)$$

$$pm = \sum_{i=1}^{NIND} pm_i (f_i \geq f_{avg}) \quad (6)$$

(3)本 AGA 算法最主要的是目的保证最优个体传到下一代，所以将某代适应度最好的一个个体暂存为 t_max ，在进行选择、交叉、变异之后用 t_max 取代某一个个体，再进行 SVM 参数训练。

(4)为了更好地传承父代的基因，本 AGA 算法将适应度最差的 mut_num 个个体用 t_max 替换，这里取 $mut_num = NIND * 0.1$ ，为了在交叉时能够有更优的父代基因。仅仅这样处理，可能会出现最优基因统治种群的现象，即出现局部最优解，而先对这 mut_num 个个体进行变异后再交叉是一个不错的方法，本 AGA 算法是在选择之前对这个特殊的群体进行了 pm 为 0.1 的变异（当然也可以在种群选择、交叉、变异之后，SVM 参数训练之前进行这样的处理）。

(5)因为 AGA 在保护最优个体的前提下，还要体现更好的种群多样性，所以种群数量选择稍微大一些，取 $NIND=80$ 。

3 实验研究分析

3.1 实验步骤

本实验是基于 MATLAB R2011a 平台、开源软件 LIBSVM^[7] MATLAB 工具箱进行的，LIBSVM 是台湾大学林智仁(Chih-Jen Lin)教授等开发设计的一个简单、易于使用和快速有效的 SVM 模式识别与回归软件包。该软件可以解决 C-SVM、v-SVM、 ϵ -SVR 和 v-SVR 等问题，目前已经有 Matlab、Java、Python 等几十种语言版本。本实验的基本步骤如下所示：

- (1)导入数据，将样本分为训练集和测试集
- (2)数据预处理，先将数据归一化，然后转成适合 LIBSVM 处理的形式。
- (3)利用所研究的算法，通过训练集进行参数 c

和 g 的寻优，核函数选择 LIBSVM 默认的 RFB 核函数。

(4)利用最优参数 c 和 g 对整个训练集进行训练，获取支持向量机模型，这一步至关重要，直接关系到最后的结果。

(5)利用获取的最优参数进行测试，可以得出测试样本的正确率（也可以将全体样本用做训练，这样就没有测试环节了）。

(6)对实验结果的分析 and 评价

(7)结束。

3.2 对比性实验研究

3.2.1 数据集 1

数据集 1 是 Breast Cancer Wisconsin，输入样本包含 10 个属性，2 个类别，共 699 个样本，其中 16 个样本存在 1 个特征值的缺失，为了保证和文献[6]有相同的数据集，本实验也将其删除。未改进的 GA 的数据分析结果见表 1，作者一共进行了 10 次实验， $Bestc$ 和 $Bestg$ 为每一次测试的最优参数。

表 1 基于 GA 的 RFB 核函数对数据集 1 的最优化参数表

次	正确率	训练集		测试集	
数	(%)	$Bestc$	$Bestg$	(%)	(%)
1	95.9004	97.9206	2.03868	100.0	98.3607
2	97.3646	1.0525	0.0251712	96.4000	100.0
3	97.2182	35.6199	0.00068601	96.4000	100.0
4	97.2182	47.1245	0.00041602	96.4000	100.0
5	96.1933	1.00073	1.78414	98.8000	98.3607
6	95.9004	95.9683	2.05031	100.0	98.3607
7	97.3646	1.1496	0.0212728	96.4000	100.0
8	96.0469	0.83655	1.79499	97.8000	97.8142
9	96.3397	8.83999	0.0596849	97.0000	99.4536
10	96.1933	5.60186	0.151889	97.2000	99.4536
avg	96.5740	—	—	97.6400	99.1804

将数据集 1 分为训练集和测试集两部分，500 个训练样本中两类的样本分别为 300 和 200，余下的 183 个测试样本则分别是 144 和 39。在实验之前，作者先用文献[6]提供的参数 $c=0.125, g=0.125$ 进行检验，得出的结果与文献[6]中相同，说明这样的样本分类可以进行对比性实验，实验的结果见表 1 的训练集和测试集，表 2 做了同样的实验，只是将 GA 换成了改进后的 AGA。



表 2 基于 AGA 的 RFB 核函数对数据集 1 的最优化参数表

次 数	正确率 (%)	Bestc	Bestg	训练集 (%)	测试集 (%)
1	97.2182	0.180855	0.123088	96.4000	99.4536
2	97.2182	0.533182	0.0817347	96.6000	99.4536
3	97.2182	0.345354	0.0898074	96.4000	99.4536
4	97.3646	0.725975	0.0291078	96.4000	100.0
5	97.3646	1.2437	0.0224894	96.4000	100.0
6	97.2182	0.531215	0.0413118	96.4000	100.0
7	97.2182	0.364224	0.0983961	96.4000	99.4536
8	97.3646	0.934079	0.0254082	96.4000	100.0
9	97.2182	0.133687	0.148924	96.4000	99.4536
10	97.3646	0.657629	0.0346951	96.4000	100.0
avg	97.2768	---	---	96.4200	99.7268

表 4 基于 AGA 的 RFB 核函数对数据集 2 的最优化参数表

次 数	正确率 (%)	Bestc	Bestg	训练集 (%)	测试集 (%)
1	84.8148	10.772	0.0120472	87.0000	84.2857
2	84.8148	13.7149	0.0100888	87.0000	84.2857
3	84.8148	11.2744	0.0175654	87.5000	85.7143
4	84.8148	9.58937	0.0198672	87.5000	85.7143
5	84.4444	24.1296	0.0078236	87.5000	82.8571
6	84.8148	13.8612	0.0113107	87.0000	84.2857
7	84.8148	10.4261	0.017063	87.5000	82.8571
8	84.8148	13.3891	0.012493	86.5000	82.8571
9	84.4444	21.536	0.0083163	87.0000	84.2857
10	84.8148	10.7685	0.0162152	87.0000	85.7143
avg	84.7407	---	---	87.1500	84.2857

3.2.2 数据集 2

数据集 2 是 Heart Disease, 输入样本包含 13 个属性, 2 个类别, 共 270 个样本。未改进的 GA 的数据分析结果见表 3, 作者一共进行了 10 次实验, Bestc 和 Bestg 为每一次测试的最优参数。

表 3 基于 GA 的 RFB 核函数对数据集 2 的最优化参数表

次 数	正确率 (%)	Bestc	Bestg	训练集 (%)	测试集 (%)
1	84.8148	16.2505	0.010183	87.0000	84.2857
2	84.8148	19.6164	0.0091022	87.0000	82.8571
3	84.8148	13.9911	0.0109977	87.0000	84.2857
4	84.8148	14.1051	0.0117294	86.5000	84.2857
5	84.8148	13.4759	0.0115994	87.0000	84.2857
6	84.8148	16.5207	0.0101188	87.0000	84.2857
7	84.8148	10.0941	0.0178133	87.5000	84.2857
8	84.8148	15.6819	0.0107422	86.5000	84.2857
9	84.8148	15.4457	0.011026	86.5000	82.8571
10	84.8148	14.0196	0.0101512	87.0000	84.2857
avg	84.8148	---	---	86.9000	84.00

将数据集 2 分为训练集和测试集两部分, 200 个训练样本中两类的样本分别为 110 和 90, 余下的 70 个测试样本则分别是 40 和 30。在实验之前, 作者先用文献[6]提供的参数 $c=256, g=0.000976563$ 进行检验, 得出的结果与文献[6]中相同, 说明这样的样本分类可以进行对比性实验, 实验的结果见表 3 的训练集和测试集, 表 4 做了同样的实验, 只是将 GA 换成了改进后的 AGA。

3.3 结果分析

为了避免实验结果带来的偶然性, 作者对每个数据集都做了 10 次实验, 将 10 次实验所得结果的平均值与文献[6]进行对比, 具体的对比结果见表 5 和表 6。

表 5 数据集 1 的对比结果

	网格法(%)	GA(%)	AGA(%)
全体正确率(%)	97.2128	96.5740	97.2768
训练集正确率(%)	96.4000	97.6400	96.4200
测试集正确率(%)	99.4536	99.1804	99.7268

表 5 结果显示, 对于数据集 1, GA 的效果没有文献[6]网格好, AGA 的效果与网格基本持平, 而在正确率上略胜一筹。

表 6 数据集 2 的对比结果

	网格法(%)	GA(%)	AGA(%)
全体正确率(%)	84.0741	84.8148	84.7407
训练集正确率(%)	87.5000	86.9000	87.1500
测试集正确率(%)	84.2857	84.0000	84.2857

表 6 结果显示, 对于数据集 2, GA 和 AGA 的效果都比文献[6]网格好, 而在正确率上也略有优势。值得一提的是, 惩罚因子 c 的取值一般是越小越好, 而文献[6]中的 c 选择过大。作者做了 10 组实验, c 都在 25 以内, 是文献[6]的 10%不到。其中最好的一组是 GA 中的第 7 组, c 不到文献[6]的 4%, 具体的对比结果如表 7 所示。



表 7 网格最优参数与 GA 最优参数的对比

	网格法($c=256$, $g=0.000976563$)	GA($c=10.0941$, $g=0.0178133$)
全体正确率(%)	84.0741	84.8148
训练集正确率(%)	87.5000	87.5000
测试集正确率(%)	84.2857	84.2857

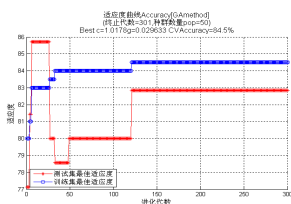
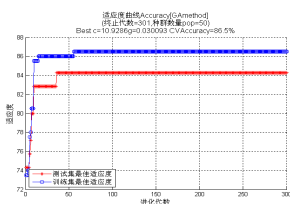
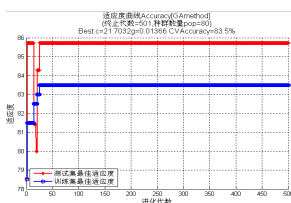
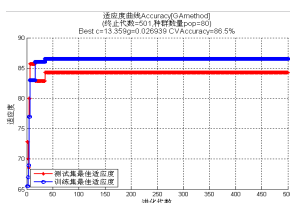
从表 7 中可以很明显地看出, GA 的最优解比网格的最优解更稳定并且参数更优, 所以 GA 是值得考虑的一种方法。

3.4 进一步实验

上面的实验证明了基于 GA 的 RFB 核函数同样有很好的稳定性, 并且 GA 在整体上优于网格法。但是很多时候我们并不能够用所有的样本去训练, 然后在测试其中的部分样本(例如数据集 1, 用 683 个样本训练, 然后分别测试 500 个训练集和 183 个测试集)。这个只能去验证算法的稳定性。但如果真的要测试一些样本, 是不能够将它们作为训练集的(因为我们还不知道它们的属于哪一类)。

由于篇幅的限制, 本小节只针对数据集 2 进行简单的讨论。作者用 200 个训练集训练, 另外 70 个测试集测试, 旨在分析 GA 预测的正确率以及收敛的代数。

作者用 $v=5$ 和 $v=10$ 进行交叉检验, 测试的结果显示: (1)GA 测试的正确率有 82.8571% 和 84.2857%, AGA 测试的正确率有 85.7143% 和 84.2857%。(2)GA 的收敛代数分别为 121 代和 56 代, AGA 的收敛代数分别为 26 代和 36 代。作者做实验时选择的 GA 和 AGA 终止代数分别为 200 和 500, 实验结果如图 4-1 至 4-4 所示。

图 4-1 $v=5$ 的 GA 测试结果图 4-2 $v=10$ 的 GA 测试结果图 4-3 $v=5$ 的 AGA 测试结果图 4-4 $v=10$ 的 AGA 测试结果

4 小结与展望

为了确保实验不是偶然现象, 本文首先利用 GA 和 AGA 对数据集的 10 次测试, 通过实验的数据分析, 证明基于 GA 的 RFB 函数有很好的稳定性, 而且稳定性和正确率在整体上优于文献[6]的网格法, 尤其是数据集 2, 这为利用基于 GA 的 RFB 函数参数优化提供了理论依据。

针对 GA 进行改进的方法非常多, 而且很多都有着广泛的应用, 作者通过简单的 AGA_SVM 算法设计, 为试图通过改进 GA 以提高测试样本的正确率提供参考。

在进一步的实验中, 作者发现未改进的 GA 和作者设计的 AGA(特别是 AGA)都可能存在不同程度的早熟, 作者下一步将会对自己设计的 AGA 进行改进, 以避免可能存在的早熟现象。

参考文献

- [1] Vapnik V N. The Nature of Statistical Learning Theory[M]. NY: Springer Verlag, 1995
- [2] Chapelle O, Vapnik V, Mukherjee S, et al. Choosing Multiple Parameters for Support Vector machines[J]. Machine Learning, 2002, 46: 131-159.
- [3] 雷英杰等编著. MATLAB 遗传算法工具箱及应用[M]. 西安: 西安电子科技大学出版社, 2005.4.
- [4] 邓乃扬, 田英杰著. 支持向量机: 论、算法与拓展. 北京: 科学出版社, 2009.
- [5] 张学工. 关于统计学理论与支持向量机[J]. 自动化学报, 2000.1, 26(1): 32-42.
- [6] 丰国和. SVM 分类核函数及参数选择比较[J]. Computer Engineering and Applications 计算机工程与应用, 2011, 47(3): 123-124, 128.
- [7] LIBSVM: Library for Support Vector Machines[EB/OL]. [2010-04-15]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [8] 吴建斌, 李太全, 田茂. 改进的遗传算法在白细胞识别中的应用研究[J]. Computer Engineering and Applications 计算机工程与应用, 2007, 43(27): 243-245.
- [9] 方瑞明著. 支持向量机理论及其应用分析[M]. 北京: 中国电力出版社, 2007.
- [10] 刘胜, 李妍妍. 自适应 GA-SVM 参数选择算法研究. 哈尔滨工程大学学报, 2007.4, 28(4): 398-402.