

文章编号: 1001-0920(2007)12-1385-05

## 基于改进 PSO 算法的 LS-SVM 油层识别模型

夏克文, 董 瑶, 杜红斌

(河北工业大学 信息工程学院, 天津 300401)

**摘 要:** 为解决常规油层识别方法因其本身缺陷而无法取得理想效果的缺点, 提出一种基于改进 PSO 算法的 LS-SVM 油层识别模型, 即综合已有改进的 PSO 模型提出一种新的改进形式, 并用此算法迭代求解 LS-SVM 中出现的矩阵方程, 从而避免矩阵求逆, 加快 LS-SVM 算法的训练速度, 节省内存, 而且求得最优解. 实际应用表明, 所提出的识别模型优于 BP 模型和经典 SVM 模型, 识别精度高、收敛速度快、效果显著.

**关键词:** 最小二乘支持向量机; 粒子群优化算法; 油层识别

**中图分类号:** TP18      **文献标识码:** A

### Oil layer recognition model of LS-SVM based on improved PSO algorithm

XIA Ke-wen, DONG Yao, DU Hong-bin

(School of Information Engineering, Hebei University of Technology, Tianjin 300401, China. Correspondent: DONG Yao, E-mail: dy198209@163.com)

**Abstract:** To solve the intrinsic shortcomings in general oil layer recognition methods, which is hard to obtain an ideal effect in application, the oil layer recognition model of LS-SVM based on the improved PSO algorithm is presented to iteratively solve the linear system of equations in LS-SVM algorithm, which is a new improved form by synthesized the existing model of PSO. By using the improved LS-SVM algorithm, the problem of solving inverse matrix is resolved and the training velocity of LS-SVM algorithm is quickened. Memory is saved and the least square solution is always got. The actual application shows that the improved LS-SVM model is superior to BP and SVM model in oil layer recognition, which not only has much greater accuracy, but also improves the velocity of convergence.

**Key words:** Least squares support vector machines (LS-SVM); Particle swarm optimization (PSO); Oil layer recognition

### 1 引 言

油层识别是目前石油勘探与开发的重要环节, 测井分析专家和地质学家都十分关心油层的测井解释及其流体的性质识别方法的研究. 面对油田勘探中愈加复杂的岩性、裂缝、低孔隙度、低渗透率和低电阻率等类型储集层的油层识别评价问题, 常规线性、经验性测井解释技术<sup>[1]</sup>常常表现得无能为力.

随着计算智能技术的迅猛发展, 神经计算、进化计算和粒计算等计算智能技术在许多领域得到成功应用, 已成为油层识别的一种主要技术. 比如在小样本情形下, 基于结构风险最小化 (SRM) 准则<sup>[2]</sup>的支持向量机 (SVM) 能够有效地避免经典学习方法中存在的过学习、维数灾难、局部极小等问题<sup>[3]</sup>, 具有

良好的泛化能力, 在各种分类问题中得到了成功应用<sup>[4,5]</sup>. 根据实际需要, Suyken 提出的最小二乘支持向量机 (LS-SVM)<sup>[6]</sup>能够克服大规模问题存在训练速度较慢的缺点, 是因为 LS-SVM 将二次优化问题转化为一个线性方程组的求解问题. 但 LS-SVM 算法在求解过程中总会出现矩阵的求逆, 对于实际工程, 大规模问题在微机上难以实现的. 为此, 研究一种新的求解方法, 即采用再现群智能的粒子群优化算法 (PSO) 来求解任意线性方程组, 不仅加快了计算速度, 而且避免了矩阵求逆.

本文在研究 PSO, LS-SVM 及其改进算法的基础上, 提出一种基于改进 PSO 的 LS-SVM 油层识别模型, 并采用基于粗糙集相似度的属性约简算法

收稿日期: 2006-10-12; 修回日期: 2006-12-25.

基金项目: 国家自然科学基金项目 (60173058, 60377020).

作者简介: 夏克文 (1965—), 男, 湖南邵阳人, 教授, 博士生导师, 博士, 从事智能信息处理技术的研究; 董瑶 (1982—), 女, 河北唐山人, 硕士生, 从事计算智能与数据挖掘的研究.

去除冗余属性,使其在实际油层应用中识别精度高且收敛速度快。

## 2 算法改进

### 2.1 改进 PSO 算法

PSO 算法是一种新兴的再现群智能的进化计算技术. Kennedy 等提出的 PSO 算法<sup>[7]</sup>采用下列公式对粒子进行操作:

$$v_{i+1} = v_i + \omega \cdot v_i + r_1 \cdot (p_{best} - x_i) + r_2 \cdot (g_{best} - x_i), \quad (1)$$

$$x_{i+1} = x_i + v_{i+1}. \quad (2)$$

很多学者在研究上述基本模型时发现,若加速常数  $\omega$  和  $r_1$  以及最大速度等参数太大,粒子群可能错过最优解,导致算法不收敛;而在收敛的情况下,所有的粒子都向最优解的方向飞去,所有粒子趋向同一化,使得后期收敛速度明显变慢,同时算法收敛到一定精度时无法继续优化,所能达到的精度也较低.为解决此问题,可采取增加收敛因子,惯性权重和约束因子等措施.作者在综合分析惯性权重模型<sup>[8]</sup>和收敛因子模型<sup>[9]</sup>等几种改进模型的基础上,得到如下新的改进形式:

$$v_{i+1} = \omega \cdot [ \omega \cdot v_i + r_1 \cdot (p_{best} - x_i) + r_2 \cdot (g_{best} - x_i) ], \quad (3)$$

$$x_{i+1} = x_i + \omega \cdot v_{i+1}. \quad (4)$$

其中

$$\omega = \frac{2}{|2 - \sqrt{2 - 4}|}, \quad (5)$$

$$\omega = \frac{(\omega_{max} - \omega_{min}) \cdot T_{max}}{T}. \quad (6)$$

这种改进 PSO 算法,既保证了算法的收敛性,加快了收敛速度,又提高了求解精度。

### 2.2 基于改进 PSO 的 LS-SVM 算法

设训练集

$$S = \{ (x_k, y_k) / k = 1, 2, \dots, N \},$$

其中  $x_k \in R^n$  和  $y_k \in R$  分别为输入和输出数据.与经典 SVM 不同,LS-SVM 利用 SRM 准则构造如下最小化目标函数及其约束条件:

$$\min_{b, e} J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \sum_{k=1}^N e_k^2;$$

$$\text{s. t. } y_k = w^T \Phi(x_k) + b + e_k, \quad k = 1, 2, \dots, N. \quad (7)$$

为求解式(7)的优化问题,转化为求解如下线性方程组<sup>[6]</sup>:

$$\begin{bmatrix} 0 & 1^T \\ 1 & K + \lambda^{-1} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix}. \quad (8)$$

其中:  $\mathbf{1} = [1, \dots, 1]^T$ ,  $K = \Phi(x)^T \Phi(x_k)$  是满足

Mercer 条件的核函数,  $Y = [y_1, \dots, y_N]^T$ ,  $a = [a_1, \dots, a_N]^T$ .

将方程(8)改写为如下矩阵方程的形式:

$$AX = z(A \quad R^{m \times n}, z \quad R^m). \quad (9)$$

LS-SVM 算法是用最小二乘法来求解  $X = (x_1, x_2, \dots, x_n)^T$ .然而,对于实际工程的大规模问题,由于  $A^T A$  维数较大,在微机上难以实现矩阵的求逆过程.为此,可采用迭代计算的方法求解矩阵方程.

本文提出一种基于改进 PSO 的 LS-SVM 算法,主要创新之处在于将矩阵方程(9)的最小二乘求解转化为采用改进 PSO 算法来进行迭代求解.这样,可以避免出现矩阵求逆,而且加快了计算速度,其流程如下:

1) 初始化粒子群.设定粒子群参数,在定义空间  $R^n$  中随机产生  $n$  个粒子  $x_1, x_2, \dots, x_n$ ,组成初始种群  $X(t)$ ;随机产生各粒子的初始速度  $v_1, v_2, \dots, v_n$ ,组成速度矩阵  $V(t)$ ;每个粒子的个体最优解  $p_{best_i}$  初始值为  $x_i$  的初始值.

2) 评价各粒子适应度(fitness).在具体矩阵方程求解中,可按残差  $(z - AX)$  的均方差来定义适应度函数

$$f(x) = \frac{1}{mn} \sum_{i=1}^m \left( \sum_{j=1}^n z_i - A_{ij} x_j \right)^2. \quad (10)$$

3) 对每个粒子,比较当前适应度  $f(x_i)$  和历史最好位置适应度  $f(p_{best_i})$ , if  $f(x_i) < f(p_{best_i})$ , then  $p_{best_i} = x_i$ ;比较群体所有粒子当前适应度  $f(x_i)$  和群体最好位置适应度  $f(g_{best_i})$ , if  $f(x_i) < f(g_{best_i})$ , then 全局最优解  $g_{best_i} = x_i$ .

4) 根据改进 PSO 模型(3)和(4),更新粒子的速度和位置,产生新种群  $X(t+1)$ ,速度调整规则如下:

$$v_i = \begin{cases} V_{max}, & v_i > V_{max}; \\ -V_{max}, & v_i < -V_{max}. \end{cases} \quad (11)$$

5) 检查结束条件.若满足,则结束寻优,返回当前最优个体为结果;否则  $T = T + 1$ ,转至步骤2).设定结束条件为寻优达到最大迭代次数  $T_{max}$  或评价价值小于给定精度.

6) 输出结果,得到矩阵方程的最小二乘解,即对应式(8)中最优参数  $b$  和  $\{a_i\}_{i=1}^N$ .

7) 将  $\{a_i\}_{i=1}^N$  和  $b$  代入

$$y(x) = \sum_{k=1}^N a_k K(x, x_k) + b, \quad (12)$$

得到识别函数式,再输入待识别样本进行识别.本文选取径向基函数为核函数

$$K(x, x_k) = \exp(-\|x - x_k\|^2 / 2\sigma^2). \quad (13)$$

**例 1** 为了验证上述改进 PSO 算法求解矩阵方程的有效性,下面用改进 PSO 算法和基本 PSO 算法分别求解重力勘探中一个密度反演典型例子<sup>[10]</sup>:

$$Ax = z.$$

其中

$$A =$$

2.2	3.7	1.5	0.7	0.4	6.1	5.7	2.1
4.5	6.8	2.2	1.0	0.5	8.6	8.3	2.8
4.5	12.4	3.7	1.5	0.7	8.6	11.7	3.9
2.2	16.0	6.8	2.2	1.0	6.1	14.1	5.7
1.0	12.4	12.4	3.7	1.4	3.9	14.1	8.3
0.6	6.8	16.0	6.8	2.2	2.5	11.7	11.7
0.3	3.7	12.4	12.4	3.6	1.7	8.3	14.1
0.2	2.2	6.8	16.0	6.5	1.2	5.7	14.1
0.1	1.5	3.7	12.4	10.8	0.9	3.9	11.7
0.1	1.0	2.2	6.8	10.8	0.7	2.8	8.3
0.1	0.7	1.5	3.7	6.5	0.5	2.1	5.7

$$Z = [7.5, 11.3, 15.2, 17.5, 18.0, 17.8, 18.1, 18.3, 17.6, 14.0, 8.9]^T.$$

采用基本 PSO 算法和改进 PSO 算法求解上述矩阵方程.设粒子数目  $N = 10$ ,最大迭代次数  $T_{max} = 1000$ ,其迭代误差曲线如图 1 所示.由图 1 可以看出,基本 PSO 算法迭代 584 次误差精度达到  $6.5767 \times 10^{-2}$ ,而改进 PSO 算法在迭代 42 次便达到这一精度,迭代 176 次后误差精度达到  $2.4034 \times 10^{-3}$ .输出  $x$  的结果对比如表 1 所示.

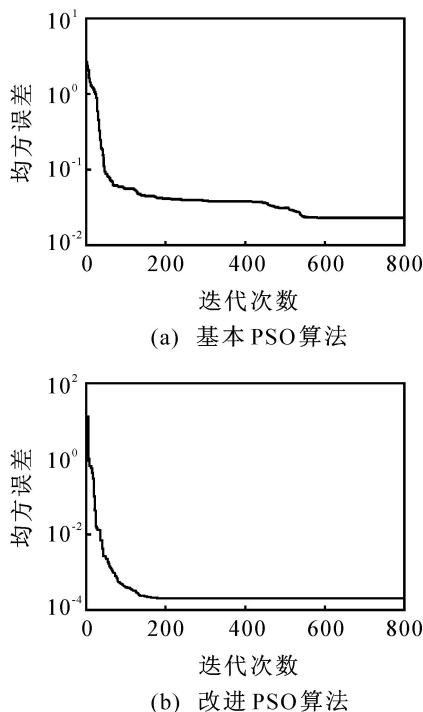


图 1 PSO 算法训练网络的误差演化曲线

由此可见,采用改进 PSO 算法求解矩阵方程,与基本 PSO 算法相比,其收敛速度快,计算精度高,效果令人满意.

表 1 两种算法的输出结果对比

期望输出 $x$	基本 PSO 算法的输出	改进 PSO 算法的输出
0.15	0.456 6	0.140 4
0.3	0.141 8	0.242 1
0.1	0.166 4	0.087 9
0.2	0.424 5	0.231 5
0.6	0.727 6	0.619 4
0.35	0.118 7	0.335 3
0.4	0.718 8	0.500 6
0.5	0.074 7	0.437 3
均方差	$6.5767 \times 10^{-2}$	$2.4034 \times 10^{-3}$

### 3 油层识别模型的建立

基于改进 PSO 的 LS-SVM 油层识别模型如图 2 所示.

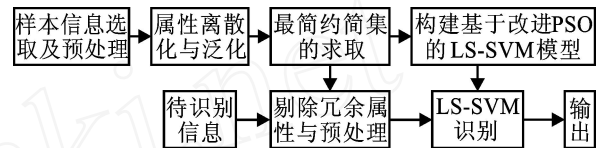


图 2 基于改进 PSO 的 LS-SVM 油层识别模型

识别模型主要包括如下几个步骤:

1) 样本信息选取与预处理. 样本集资料的选取要完备、全面,应与油层评价密切相关,尽量保证所选取的资料不重叠.将样本分为训练样本和测试样本两部分,为避免出现计算饱和现象,要对样本数据进行归一化处理,使输入的样本数据在  $[0, 1]$  之间,归一化公式如下:

$$x = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{14}$$

这里:  $x \in [x_{min}, x_{max}]$ ,  $x_{min}$  和  $x_{max}$  分别为最小值和最大值.

2) 属性离散化和泛化.为实现样本信息属性约简,先将所提取的样本信息进行决策属性泛化,并采用黄金分割优选法实现连续属性离散化.

3) 样本信息属性约简.石油测井数据包括声、电、核等各种测井信息.较为完善的测井系列一般都提供 10 多种测井属性信息,但有用的测井属性并不同等重要,常常存在冗余属性.为剔除样本集中的冗余属性,必须进行属性约简.这里采用基于相似度的属性约简算法<sup>[11]</sup>,求出样本信息的最简约集,从而得到反映油层特性的最简条件属性集.

4) LS-SVM 建模.在 LS-SVM 模型中,输入经属性约简后的样本信息,采用改进 PSO 算法迭代寻找 LS-SVM 中的最优参数  $\{a_i\}_{i=1}^N$  和  $b$ ,从而得到训练好的 LS-SVM 预测模型,即识别模型.

5) 识别输出.用训练好的 LS-SVM 模型对整个井段进行油层识别,并输出结果.

### 4 实例分析

为了检验本文提出的基于改进 PSO 的 LS-SVM 油层识别模型的应用效果,下面举例进行详细分析.

#### 1) 样本信息选取与预处理

样本取自某油田一关键井储层的测井数据,在井段1 200 ~ 1 320 m处提取265个样本点测井数据作为训练样本,其中油层 112 个,干层 153 个. 样本信息的条件属性有 11 个,即:AC(声波时差),CALI(井径),GR(自然伽马),NG(中子伽马),RA2(视电阻率 2),RA4(视电阻率 4),RI(侵入带电阻率),RM(泥浆电阻率),RT(地层真电阻率),RXO(冲洗带地层电阻率),SP(自然电位). 样本信息的决策属性为{干层,油层}. 令决策属性  $D = \{d\}$ ,  $d = \{d_i = i, i = -1, 1\}$ ,其中 -1 和 1 分别代表干层和油层.

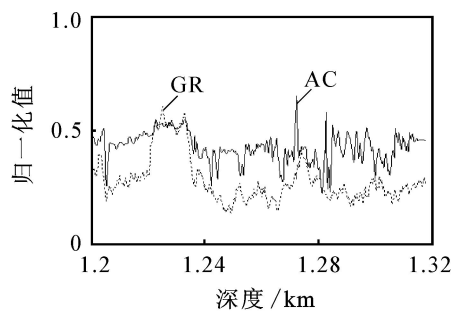
#### 2) 样本信息属性约简

将 11 个条件属性经预处理和离散化处理后,采用基于相似度的属性约简算法,得到最简约简集为 {AC,GR,RT,RXO,SP}. 从测井学原理的角度看,这几个属性是最能反映油层特性的测井信息. 表 2 所示为约简后的 5 个样本属性的数值范围.

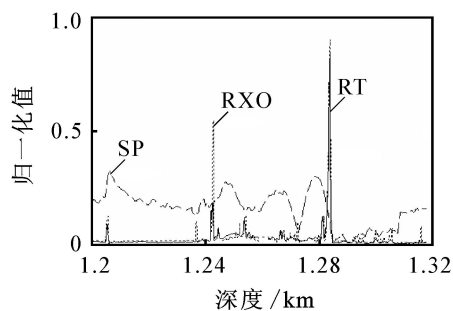
将这 5 个属性在井段1200 ~ 1320m之间进

表 2 约简后的样本属性数值范围

属性	AC	GR	RT	RXO	SP
最小值	100	4	0.0	0.0	- 50
最大值	600	20	500	300	30



(a) 属性 GR,AC



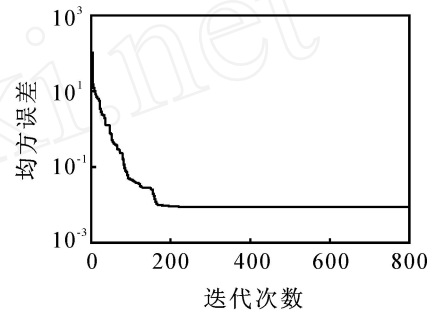
(b) 属性 SP,RXO,RT

图 3 经过归一化后的 5 个属性的曲线

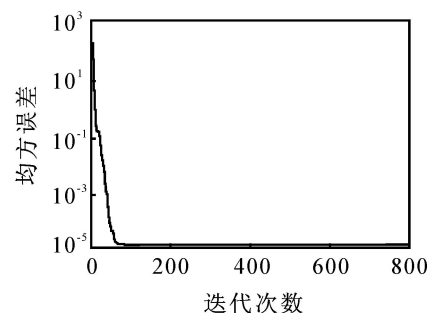
行归一化处理,如图 3 所示. 其中:横轴表示深度,纵轴表示归一化值.

#### 3) 迭代寻优

设粒子群规模为 20,解空间为 266 维(即寻优参数的个数),最大迭代次数为 1 000,加速常数  $c_1 = c_2 = 2.05$ ,则收敛因子  $\omega = 0.729$ ,初始惯性权重  $\omega = 0.9$ . LS-SVM 中正规化参数  $\lambda$  选取 1 000,径向基核函数的宽度参数  $\sigma^2$  选取 0.125. 利用基本 PSO 和改进 PSO 求解 LS-SVM 算法中的式(8)所得到的误差曲线如图 4 所示. 基本 PSO 算法迭代 204 次误差精度达到  $1.15 \times 10^{-2}$ ,而改进 PSO 算法达到同一精度仅需迭代 19 次,迭代 81 次后误差精度便达到  $5.7547 \times 10^{-5}$ ,可见改进 PSO 在收敛速度和精度上均优于基本 PSO.



(a) 基本 PSO 算法



(b) 改进 PSO 算法

图 4 PSO 算法训练网络的误差演化曲线

#### 4) 识别结果比较

将迭代寻优后得到的预测模型对全部井段 944 个样本进行油层识别,分别建立 BP 网络识别模型和经典 SVM 模型,并将识别结果与本文提出的识别模型相比较.

为了衡量识别模型的性能,定义以下几个性能指标:均方根误差

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2}; \quad (15)$$

最大正误差

$$MAXPE = \max\{e_i, 0\}; \quad (16)$$

最大负误差

$$MAXNE = \min\{e_i, 0\}. \quad (17)$$

式中:  $e_i = y_i - \hat{y}_i$ ,  $\hat{y}_i$  和  $y_i$  分别为识别输出值和期望输出值。

BP 网络识别模型采用 5-13-1 的网络结构,学习因子取为 0.05,隐含层和输出层的传输函数均选取 Tansig 函数 ( $f(x) = 2 / (1 + e^{-2x}) - 1$ ), 经典 SVM 模型也采用径向基核函数,识别结果见表 3。

表 3 识别结果比较

识别模型	RMSE	MAXPE	MAXNE	识别率 / %
BP 网络	0.267 9	1	- 0.997 1	90.6
经典 SVM	0.251 2	1	- 1	92.3
改进 LS-SVM	0.119 0	0.836 2	- 0.803 6	98.8

由表 3 可知,本文提出的改进 LS-SVM 识别模型在识别率等性能指标上明显优于 BP 模型和经典 SVM 模型,在识别精度、泛化能力上优于 BP 网络模型与 SVM 识别模型。此外,本文方法在油层识别率上未达到 100%,即有误判点,这与样本的选取不完备或构建的模型的泛化能力有限等因素有关。总之,本文提出的基于改进 PSO 的 LS-SVM 识别模型是切实可行且行之有效的,不但有效地避免了 BP 模型中存在的过学习、维数灾难、局部极小等问题,而且解决了 SVM 模型对于大规模样本所出现的计算复杂度高、计算速度慢等问题。

### 5 结 论

基于结构风险最小化 (SRM) 准则的 LS-SVM 是一种良好的分类方法。本文提出的基于改进 PSO 迭代优化的 LS-SVM 算法不仅避免了矩阵求逆计算,而且总能求得最优解,其训练速度和求解精度均得到提高。实际应用表明,本文提出的基于改进 PSO 的 LS-SVM 识别模型适用于解决油层识别等大规模样本问题,效果显著,具有很好的应用前景。

### 参考文献 (References)

[1] 《测井学》编写组. 测井学 [M]. 北京: 石油工业出版社, 1998.  
(《Well logging》Compile Group. Well logging [M]. Beijing: Petroleum Industry Press, 1998.)

[2] Vladimir N Vapnik. The nature of statistical learning theory [M]. New York: Springer, 2000.

[3] Senf Alexander, Chen Xue-wen, Zhang Anne. Comparison of one-class SVM and two-class SVM for fold recognition [C]. Neural Information Processing—13th Int Conf ICONIP 2006, Proc. Hong Kong, 2006: 140-149.

[4] Shih Frank Y, Zhang Kai. Support vector machine networks for multi-class classification [J]. Int J of Pattern Recognition and Artificial Intelligence, 2005, 19 (6): 775-786.

[5] Cortes C, Vapnik V. Support vector machine [J]. Machine Learning, 1995, 20(3): 273-297.

[6] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers [J]. Neural Processing Letters, 1999, 9(3): 293-300.

[7] Kennedy J, Eberhart R C. Particle swarm optimization [C]. Proc IEEE Int Conf Neural Networks. Piscataway: IEEE Press, 1995: 1942-1948.

[8] Shi Y, Eberhart R. A modified particle swarm optimizer [C]. IEEE World Congress on Computational Intelligence. Anchorage, 1998: 69-73.

[9] Clerc M. The swarm and the queen: Towards a deterministic and adaptive particle swarm optimization [C]. Proc of the Congress of Evolutionary Computation. Washington, 1999: 1951-1957.

[10] 焦荣昌. 线性方程组的正则化解法及其应用 [J]. 物探与化探, 1994, 18(3): 200-207.  
(Jiao Rong-chang. The regularized solution of linear equations set and its application [J]. Geophysical and Geochemical Exploration, 1994, 18(3): 200-207.)

[11] 夏克文, 刘明霄, 张志伟, 等. 基于属性相似度的属性约简算法 [J]. 河北工业大学学报, 2005, 34(4): 20-23.  
(Xia Ke-wen, Liu Ming-xiao, Zhang Zhi-wei, et al. An approach to attribute reduction based on attribute similarity [J]. J of Hebei University of Technology, 2005, 34(4): 20-23.)

(上接第 1380 页)

[10] Fahlman S E. An empirical study of learning speed in back-propagation networks [R]. CMU-CS-88-162, Carnegie-Mellon University, 1988.

[11] Scholkopf B, Smola A, Williamson R C, et al. New support vector algorithms [J]. Neural Computation, 2000, 12(5): 1207-1245.