

粗糙 one-class 支持向量机

王磊^{1,2} 杨一帆¹ 周启海¹

(西南财经大学经济信息工程学院 成都 610074)¹ (西南财经大学中国支付体系研究中心 成都 610074)²

摘要 粗糙集理论是处理不确定性和不完备信息的重要方法之一。通过将粗糙集理论引入到 one-class 支持向量机,提出了一种新颖的粗糙 one-class 支持向量机。通过定义上近似超平面和下近似超平面,使得训练样本能根据在粗糙间隔中的位置,自适应地对决策超平面产生影响。并且,outlier 样本由于距离上近似超平面较近并产生较小的间隔误差,不会导致决策超平面对它们产生明显的过拟合。实验结果表明,粗糙 one-class 支持向量机的泛化性能优异,识别率和误识率均优于传统的 one-class 支持向量机。

关键词 粗糙集, one-class, 支持向量机

中图分类号 TP391.4 **文献标识码** A

Rough Set-based One-class Support Vector Machine

WANG Lei^{1,2} YANG Yi-fan¹ ZHOU Qi-hai¹

(School of Economics Information Engineering, Southwestern University of Finance & Economics, Chengdu 610074, China)¹

(Research Center of China Payment System, Southwestern University of Finance & Economics, Chengdu 610074, China)²

Abstract The rough set theory is an important mathematical tool to deal with uncertainty and incompleteness. This paper proposed a novel rough one-class support vector machine by introducing rough margin into one-class support vector machine. With the definitions of upper approximation and lower approximation hyperplanes, the influences of training samples on the decision hyperplane are determined adaptively by their position within the rough margin. Moreover, outlier samples are prone to produce small margin errors since they lie close to the upper approximation hyperplane, so that the overfitting problem of decision hyperplane can be avoided. Experimental results on UCI datasets show the superior generalization performance of rough one-class support vector machine.

Keywords Rough set, One-class, Support vector machine

在故障诊断、人脸识别、网络异常检测和文本分类等应用领域,常常遇到如下的 one-class 问题:容易获取大量的目标类模式(或正常数据),但获取少量的非目标类模式(或异常数据)非常困难或代价很高;分类的任务是寻找一个最优决策函数以便描述目标类模式所属的区域并能够将目标类与所有非目标类分割开^[1,2]。

支持向量机(support vector machine, SVM)是基于结构风险最小化原则的机器学习方法,具有收敛到全局最优、维数不敏感、泛化能力强的优点^[3]。Schölkopf 等人将 SVM 推广到 one-class 问题,巧妙地构造最优决策超平面将目标类样本和坐标原点分割开^[4]。此外,文献[5,6]提出了类似思想,构造非线性映射空间中的超球体来覆盖目标类样本。

然而如文献[7]所指出, SVM 和 one-class SVM 采用的 Hinge 损失函数使得它们容易对 outlier 产生过拟合现象;文献[5]进一步指出,当 SVM 的模型参数选择不适当的时候,这种现象更加明显。文献[7,8]分别通过为训练样本设置非线性自适应罚损失和模糊隶属度来抑制 SVM 的过拟合问题,但针对 one-class 支持向量机过拟合问题的研究却很少出

现。

本文将粗糙集理论引入到 one-class 支持向量机,提出了一种新颖的粗糙 one-class 支持向量机,它能够有效抑制由于 outlier 引起的过拟合问题,并取得优异的泛化性能。

1 one-class 支持向量机

Schölkopf 等人提出的 one-class 支持向量机的核心思想是:利用非线性映射 ϕ 在映射特征空间中寻找一个决策超平面,尽可能地将目标类样本和坐标原点分开,其中, l 个目标类训练样本为 $S = \{ | x_i | x_i \in R^d, i = 1, 2, \dots, l \}$ 。这等价于求解如下的二次规划问题:

$$\begin{aligned} \min_{\omega, \xi, \rho} & \frac{1}{2} \| \omega \|^2 + \frac{1}{\nu} \sum_{i=1}^l \xi_i - \rho \\ \text{s. t. } & \omega \cdot \phi(x_i) \geq \rho - \xi_i, \xi_i \geq 0, i = 1, 2, \dots, l \end{aligned} \quad (1)$$

其中,参数 $\nu \in (0, 1)$ 不仅控制原点到超平面的最大间隔,也与训练样本中的 outlier 比率相关。 $\rho / \| \omega \|$ 为超平面距离坐标原点的距离,也即分类间隔。通过引入 Lagrange 乘子 α_i , 可以推导出规划(1)的对偶问题为:

到稿日期:2008-10-27 返修日期:2009-01-09 本文受国家自然科学基金青年项目(60803106),西南财经大学科学研究基金(QN0806)资助。
王磊(1978-),男,博士,讲师,主要研究方向为机器学习、模式识别等, E-mail: wanglei_t@swufe.edu.cn; 杨一帆 博士生,主要研究方向为电子支付、数据挖掘;周启海 教授,博士生导师,主要研究方向为计算几何、算法分析与设计。

$$\min_a \frac{1}{2} \sum_{i,j=1}^l a_i a_j k(x_i, x_j) \quad (2)$$

s. t. $0 \leq a_i \leq 1/ul, \sum_i a_i = 1, i, j = 1, 2, \dots, l$

其中, $k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ 为满足 Mercer 条件的核函数。显然, 对偶问题(2)也是典型二次规划, 求解出最优 Lagrange 乘子 a_i^* 后, one-class 支持向量机的决策函数为:

$$f(x) = \text{sgn}(\sum_{i=1}^l a_i^* k(x_i, x) - \rho) \quad (3)$$

其中, $0 < a_i^* < 1/ul$ 对应的训练样本称为“边界支持向量”, $a_i^* = 1/ul$ 对应的样本称为“非边界支持向量”, 剩余的样本是非支持向量机。

从图1(a)可以看出, 如果训练样本中存在少量 outlier, 它们由于拥有较大的间隔误差 $\xi_i / \|\omega\|$ 而成为非边界支持向量。根据文献[5]的分析, 决策超平面将向 outlier 样本靠近, 从而产生过拟合现象并偏离最优位置。此时, 位于超平面上的样本的性质具有不确定性, 可能属于边界支持向量、非支持向量和 outlier 的任何一种。

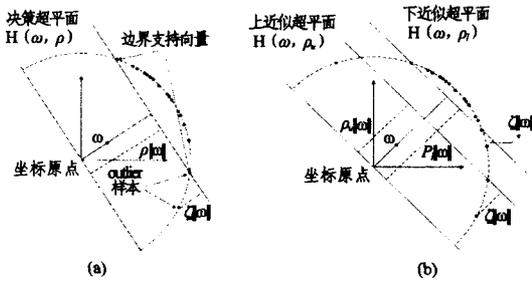


图1 one-class 支持向量机和粗糙 one-class 支持向量机的决策超平面(均采用 RBF 核函数)

2 粗糙 one-class 支持向量机

2.1 粗糙集简介

粗糙集理论是新近出现的一种刻画不确定性和不完整信息的数学工具[9]。

令 $U = \{x_1, x_2, \dots, x_n\}$ 表示任意非空的有限集合, 即论域; A 为非空的有限属性集合。则任意属性子集 $B \subseteq A$ 确定论域 U 上的一个等价关系 $IND(B)$:

$$IND(B) = \{(x_i, x_j) \in U \times U \mid \forall a \in B, a(x_i) = a(x_j)\} \quad (4)$$

其中, $a(x_i)$ 是样本 x_i 的属性 a 的值。

显然, 若 $(x_i, x_j) \in IND(B)$, 则 x_i 和 x_j 在属性子集 B 下相互不可分辨。因此, 等价关系 $IND(B)$ 将论域 U 分割成了若干等价类, 记为 $U/IND(B)$, 其中包含样本 x_i 的等价类记为 $[x_i]_{IND(B)}$ 。

为了描述样本分类的不确定性, 粗糙集理论采用了集合的上近似集和下近似集的概念, 对于任意集合 $X \subseteq U$, 它们分别定义为:

$$B_+(X) = \{Y \in U/IND(B) \mid Y \subseteq X\} \quad (5)$$

$$B_-(X) = \{Y \in U/IND(B) \mid Y \cap X \neq \emptyset\} \quad (6)$$

如图2所示, 上近似集和下近似集将论域 U 分割成3个区域: 正域 $POS(X)$, 负域 $NEG(X)$ 和边界域 $BND(X)$ 。

于是, 对于集合 X 如果存在 $B_-(X) \neq B_+(X)$, 则称它是一个粗糙集, 并且满足 $BND(X) \neq \emptyset$ 。

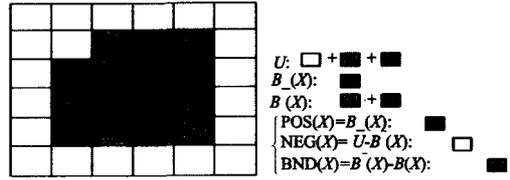


图2 粗糙集的示意图

2.2 粗糙 one-class 支持向量机

借用粗糙集的概念, 定义决策超平面的上近似超平面 $H(\omega, \rho_u)$ 和下近似超平面 $H(\omega, \rho_l)$, 并将坐标原点与它们之间的距离定义为粗糙间隔, 如图1(b)所示。显然, 位于下近似超平面之内的样本明确地属于目标类, 位于上下近似超平面之间的样本可能是 outlier, 也可能属于目标类, 而位于上近似超平面之外的样本确定地视为 outlier 样本。

为了寻找最大的粗糙间隔, 求解如下的二次规划:

$$\min_{\omega, \xi, \xi', \rho_l, \rho_u} \frac{1}{2} \|\omega\|^2 - \rho_l - \rho_u + \frac{1}{ul} \sum_{i=1}^l \xi_i + \frac{\delta}{ul} \sum_{i=1}^l \xi_i' \quad (7)$$

s. t. $\omega \cdot \phi(x_i) \geq \rho_l - \xi_i - \xi_i'$
 $0 \leq \xi_i \leq \rho_l - \rho_u, \xi_i' \geq 0$
 $\rho_l \geq 0, \rho_u \geq 0$

其中, 参数 $\delta > 1$ 。显然, 式(7)是 one-class 支持向量机原始优化问题的扩展, 称为“粗糙 one-class 支持向量机”。

引入 Lagrange 乘子 $\alpha_i \geq 0, \beta_i \geq 0, \lambda_i \geq 0, \eta \geq 0, \mu_1 \geq 0$ 和 $\mu_2 \geq 0$, 构造如下的 Lagrange 函数。

$$L = \frac{1}{2} \|\omega\|^2 - \rho_l - \rho_u + \frac{1}{ul} \sum_{i=1}^l \xi_i + \frac{\delta}{ul} \sum_{i=1}^l \xi_i' - \sum_{i=1}^l \alpha_i [\omega \cdot \phi(x_i) - \rho_l + \xi_i + \xi_i'] - \sum_{i=1}^l \beta_i \xi_i + \sum_{i=1}^l \lambda_i (\xi_i - \rho_l + \rho_u) - \sum_{i=1}^l \eta \xi_i' - \mu_1 \rho_l - \mu_2 \rho_u \quad (8)$$

根据约束规划问题的 KKT 条件[3], 得到:

$$\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^l \alpha_i \phi(x_i) = 0 \quad (9)$$

$$\frac{\partial L}{\partial \xi_i} = \frac{1}{ul} - (\alpha_i + \beta_i - \lambda_i) = 0 \quad (10)$$

$$\frac{\partial L}{\partial \xi_i'} = \frac{\delta}{ul} - (\alpha_i + \eta) = 0 \quad (11)$$

$$\frac{\partial L}{\partial \rho_l} = -1 + \sum_{i=1}^l \alpha_i - \sum_{i=1}^l \lambda_i - \mu_1 = 0 \quad (12)$$

$$\frac{\partial L}{\partial \rho_u} = -1 + \sum_{i=1}^l \lambda_i - \mu_2 = 0 \quad (13)$$

$$\alpha_i [\omega \cdot \phi(x_i) - \rho_l + \xi_i + \xi_i'] = 0, \beta_i \xi_i = 0, \eta \xi_i' = 0 \quad (14)$$

$$\lambda_i (\xi_i - \rho_l + \rho_u) = 0, \mu_1 \rho_l = 0, \mu_2 \rho_u = 0 \quad (15)$$

将式(9)至式(13)代入 Lagrange 函数 L , 经过整理得到

$$\min_a \frac{1}{2} \sum_{i,j=1}^l a_i a_j k(x_i, x_j) \quad (16)$$

s. t. $0 \leq a_i \leq \frac{\delta}{ul}, \sum_{i=1}^l a_i \geq 2$

当对偶规划的最优解 a^* 求出后, 可以根据 KKT 互补条件(14)和(15)对训练样本 x_i 相对于粗糙间隔的位置进行分析:

(a) $a_i^* = 0, x_i$ 位于 $H(\omega, \rho_l)$ 之内, 满足 $\omega \cdot \phi(x_i) > \rho_l$;

(b) $0 < a_i^* < 1/ul, x_i$ 位于 $H(\omega, \rho_l)$ 上, 满足 $\omega \cdot \phi(x_i) = \rho_l$, 称为“下近似界支持向量”;

(c) $a_i^* = 1/ul, x_i$ 位于 $H(\omega, \rho_l)$ 和 $H(\omega, \rho_u)$ 之间, 满足 $\omega \cdot \phi(x_i) = \rho_l - \xi_i$, 称为“粗糙间隔支持向量”;

表1 PHD数据集的构成

	类别									
	0	1	2	3	4	5	6	7	8	9
训练集	780	779	780	719	780	720	720	778	719	719
测试集	363	364	364	336	364	335	336	364	336	336

(d) $1/vl < \alpha_i^* < \delta/vl$, x_i 位于 $H(\omega, \rho_u)$ 之上, 满足 $\omega \cdot \phi(x_i) = \rho_u$, 称为“上近似界支持向量”;

(e) $\alpha_i^* = \delta/vl$, x_i 位于 $H(\omega, \rho_u)$ 之外, 满足 $\omega \cdot \phi(x_i) = \rho_u - \xi_i'$, 它们是“非边界支持向量”, 一般与 outlier 相关。

此外, 假设样本 x_j 位于下近似超平面上, 也即 $0 < \alpha_j^* < 1/vl$, 则由式(14)可以立即得到 $\omega \cdot \phi(x_j) - \rho_l = 0$ 。因此, ρ_l 可以由下式求得:

$$\rho_l = \omega \cdot \phi(x_j) = \sum_{i=1}^l \alpha_i^* k(x_i, x_j) \quad (17)$$

考虑到下近似超平面内的样本属于非支持向量, 并确切地属于目标类。因此, 将粗糙 one-class 的决策函数定义为:

$$f(x) = \text{sgn}(\sum_i \alpha_i^* k(x_i, x) - \rho_l) \quad (18)$$

显然, 决策超平面完全由 ω 和 ρ_l 决定, 根据式(9)和式(17), 它们又完全由最优 Lagrange 乘子 α^* 决定。依据前文的分析, 训练样本按照它们在粗糙间隔中的相对位置获得了不同的 α_i^* 值(可以分成5种情况), 从而自适应地对决策超平面产生了影响, 并共同决定了它在特征空间中的位置。

此外, 从图1(b)可以看出, 训练样本中的 outlier 产生的间隔误差为 $\xi_i' / \|\omega\|$ 且距离上近似超平面较近(取值较小), 不会对决策超平面(即下近似超平面)产生显著影响。因此, 粗糙 one-class 支持向量机能够有效避免训练样本过拟合现象的产生, 后续的实验结果也说明了这一点。

3 实验与分析

为了检验本文提出的粗糙 one-class 支持向量机(以下简称 ROSVM)在处理分类问题时的泛化性能, 在 UCI 数据库的 Pen-based Handwritten Digits(PHD)数据集^[10]上进行了数值实验。基本情况如表1所列, 实验共包含10类数据样本, 其中每个样本由16个特征描述。

表2 ROSVM和OCSVM在PHD数据集上的测试结果($\times 100\%$)

类别	s=0				s=0.5				s=2.0			
	ROSVM		OCSVM		ROSVM		OCSVM		ROSVM		OCSVM	
	PR	FR	PR	FR	PR	FR	PR	FR	PR	FR	PR	FR
0类	98.62	2.14	98.62	2.46	97.25	2.71	92.56	11.17	92.56	8.32	84.57	26.83
1类	98.63	2.04	98.35	1.69	96.98	2.90	92.03	9.95	93.13	7.81	82.42	21.20
2类	99.18	1.44	99.18	1.44	97.25	2.14	93.13	9.60	93.96	4.27	86.26	17.38
3类	99.11	1.90	99.11	1.90	97.02	2.59	92.26	10.34	92.26	6.17	82.14	23.60
4类	99.45	1.18	99.18	1.31	97.25	1.91	94.24	8.17	93.68	6.51	87.36	15.79
5类	98.21	2.78	98.51	2.66	96.12	3.70	91.04	12.33	93.13	7.43	80.60	27.76
6类	98.21	2.31	98.21	2.34	96.43	3.13	91.40	9.14	93.45	6.23	83.33	19.89
7类	98.35	2.39	98.08	2.87	96.43	2.84	92.31	10.49	92.58	7.21	84.07	22.30
8类	97.92	3.16	98.21	2.88	96.13	4.62	90.77	12.14	91.96	7.56	78.49	26.31
9类	98.51	2.25	98.51	2.06	97.02	2.69	92.56	9.68	93.15	5.79	83.62	18.18

当进一步增加训练集中 outlier 样本的比率时($s=2.0\%$, 或者添加134~136个), OCSVM的性能下降更为明显, 在第5类样本上的识别率仅为80.60%, 而误识率高达27.76%。相反, 粗糙 one-class 支持向量机的性能并没有迅速下降, 体现了对 outlier 较强的不敏感性。

因此, 实验结果验证了本文提出的粗糙 one-class 支持向量机是一种非常有效的求解单类问题的方法, 能够获得优异的泛化性能。特别是在训练集中存在 outlier 样本的情况下, 取得的识别率和误识率均显著优于传统的 OCSVM 方法。

结束语 one-class 支持向量机对训练样本中的 outlier 比较敏感, 容易产生过拟合问题。本文结合粗糙间隔理论和 one-class 支持向量机方法, 提出一种新颖的粗糙 one-class 支

持向量机, 它采用由上、下近似超平面定义的粗糙间隔代替原有的间隔, 从而抑制由于 outlier 引起的决策超平面过拟合的问题。理论分析和实验结果充分说明了新方法的优异泛化性能。

实验过程中, 针对每一类样本都采用 ROSVM 训练一个 one-class 分类器, 从而将该类样本与剩余所有类样本分割开。并且为了模拟 outlier 样本的影响, 针对每一个训练集, 都按 $s\%$ 的概率从剩余9个类的训练集中随机抽取样本作为 outlier 添加到该训练集。所有的 one-class 分类器, 均由全部测试样本进行验证, 并统计它们的识别率 PR 和误识率 FR。

此外, 所有的分类器均采用 RBF 核函数。ROSVM 的参数 v 和 δ , 以及核参数 σ 均采用 5-fold 交叉验证的方法在单独的验证数据集上进行选取。表2给出了 ROSVM 以及传统的 one-class 支持向量机(OCSVM)在 PHD 数据集上的主要实验结果(参数经选取为 $v=0.2, \delta=1.25$ 和 $\sigma=10$)。

实验结果表明, 当训练集不包含 outlier 样本时, 粗糙 one-class 支持向量机和传统的 OCSVM 在所有情况下的性能均相差不多。但是, 当按照 $s=0.5\%$ 的概率向训练集中添加 outlier 样本(实际添加33~34个), 则 OCSVM 训练产生的分类器的识别率 PR 明显降低(平均降低6.37%), 而误识率明显增加(平均增加8.19%); 相比较而言, 提出的粗糙 one-class 支持向量机的识别率和误识率都没有显著变化, 仅仅是分别下降和增加了1.62%和0.76%。这种现象表明 OCSVM 对于 outlier 样本产生了明显的过拟合问题, 使得训练的决策超平面显著偏离最优, 而本文提出的粗糙 one-class 支持向量机方法采用粗糙间隔限制了 outlier 对决策超平面的过拟合, 因而能够获得更优的识别率和更低的误识率。

参考文献

- [1] Manevitz L, Yousef M. One-class SVMs for Document Classification [J]. Journal of Machine Learning Research, 2001, 2: 139-154
- [2] 徐图, 罗瑜, 何大可. 超球体单类支持向量机的 SMO 训练算法 [J]. 计算机科学, 2008, 35(6): 178-180
- [3] Vapnik V N. 统计学习理论的本质 [M]. 张学工, 译. 北京: 清华

[4] Schölkopf B, Platt J C, Shawe-Taylor J, et al. Estimating the Support of a High-Dimensional Distribution[J]. Neural Computation, 2001, 13:1443-1471

[5] Tax D M J, Duin R P W. Support Vector Data Description[J]. Machine Learning, 2004, 54(1):45-66

[6] Campbell C, Bennett P. A Linear Programming Approach to Novelty Detection[M]. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2001

[7] Wang L, Jia H D, Li J. Training Robust Support Vector Machine with Smooth Ramp Loss in the Primal[J]. Neurocomputing, 2008, 71:3020-3025

[8] Lin C-F, Wang S-D. Fuzzy Support Vector Machine[J]. IEEE Transactions on Neural Networks, 2002, 13(2):464-471

[9] Pawlak Z. Rough Sets [J]. International Journal of Computer and Information Sciences, 1982, 11:341-356

[10] <http://www.ics.uci.edu/~mlearn/MLRepository.html>

(上接第 226 页)

表 1 训练输出

序号	故障模式	样本输出				
		Y5	Y4	y3	Y2	y1
1	正常	8.487351e-007	4.14521e-008	5.221232-e007	0.0022568923	0.99985628
2	2 测点轴承外圈	4.427036e-006	0.00036523	0.112956823	0.9912356	0.00452356
3	保持架断裂	0.009320653	8.214523e-007	0.997653	0.0053568923	0.000568926
4	齿面磨损	0.001425621	0.989526325	0.006532561	0.00752658	2.562356e-005
5	5 测点轴承外圈	0.9902356	0.004003562	0.001720345	0.000387452	0.99562358
6	正常	0.002002365	0.001203502	0.001725622	0.000425623	0.00865239
7	2 测点轴承外圈	0.002325	2.366526e-005	0.00124512	0.9942563	5.36528e-005
8	保持架断裂	1.23052e-006	0.002100223	0.9912965	0.00359862	0.28501e-005
9	齿面磨损	0.0085623	0.9990001	0.001525124	0.00979025	0.35623e-006
10	5 测点轴承外圈	0.9967121	0.005856923	0.00986592	3.15489e-005	0.12032591

同时利用有限元软件 ANSYS 产生 900 组样本,并分为训练样本、测试样本和验证样本。其神经网络结构和本文的结构一致,通过计算得到均方误差,如表 2 中列出的 ANSYS。

表 2 本文与文献[2]方法比较

网络名称	结构形式各层神经元	均方误差/%
本文	8-15-4	0.0170
文献[2]	8-17-5	0.2153
ANSYS	8-15-4	0.2858

从本文与文献[2]比较结果来看,本文的算法优化得到的齿轮箱体积比原方案(文献[2])结构减轻了 29.41%,取得了比较好的优化效果。同时与文献[3]计算结果相当接近,证明神经网络具有较高的精度,均方误差比文献[2]提高了 0.1983%,比 ANSYS 提高了 0.2688%,整个优化过程产生样本需要约 2.5h,训练神经网络约 35min,最终粒子群优化不到 5min,整个过程只需要近 150min(计算机性能 Intel P4, CPU: 2.66GHz, 内存: 1G)。虽然计算时间稍长,但获得了较好的优化结果。此外,在优化过程中,基本不需要人为干涉,对使用人员的建模水平要求不高,优化结果客观可信,且处理较复杂的优化问题不会需要更长的时间。

结束语 本文基于对粒子群算法和神经网络各自特性的研究,提出了一种新的优化算法。它不依赖具体结构形状,仅需要一定的样本就可以自适应地进行优化问题的处理,具有广泛的适用性。采用自适应 T-S 模糊模型与 PSO 相结合的算法来优化 BP 神经网络结构,既使得现有的 BP 算法优势得到进一步发挥,又使得神经网络的结构能适合具体的工程研究对象,提高了算法的普适性。扩展模糊模型的 PSO 神经网络算法在对齿轮箱这种形状不规则的问题应用中显示出良好的效果,较之传统优化方法具有更好的智能性。对于复杂工程结构优化问题,由于高度的非线性、多变量和多约束,往往存在优化过程建模困难、计算量大、迭代缓慢等问题,本文所

提出的方法对有效解决此类复杂工程优化问题提供了新途径。

参考文献

[1] Kennedy J, Eberhart R C. Particle swarm optimization[C]// Proc. of the IEEE Conf. on Neural Networks. IV. Perth: IEEE Press, 1995: 1942-1948. <http://www.engr.iupui.edu/~shi/Conference/psopap4.html>

[2] Lin C J, Hong S J. The design of neuro-fuzzy networks using particle swarm optimization and recursive singular value decomposition[J]. Neurocomputing, 2007, 71(1-3):297-310

[3] Takagi T, Sugeno M. Fuzzy identification of systems and its applications to modeling and control [J]. IEEE Trans System Man and Cyb, 1985, 15(1):116 - 132

[4] Mao Hong-wei, Pan Hong-xia, Lin Wen-li. Mavelet neural network based on particle swarm optimization algorithm and its application in fault diagnos of gear-box[J]. Journal of Vibration and Shock, 2007, 26(5):133-137

[5] Lu Z S, Hou Z R. Particle swarm optimization with adaptive mutation. Acta Electronica Sinica, 2004, 32(3):416-420

[6] Mukherjee V, Ghoshal S P. Intelligent particle swarm optimized fuzzy PID controller for AVR system [J]. Electric Power Systems Reasearch, 2007, 77(12):1689-1698

[7] 郭海丁,路志峰. 基于 BP 神经网络和遗传算法的结构优化设计[J]. 航空动力学报, 2003, 18(2):216-220

[8] Chatterjee A, Siarry P. Nonlinear Inertia Weight Variation for Dynamic Adaptation in Particle Swarm Optimization[J]. Computers and Operations Research, 2006, 33(3):859-871

[9] 赵学智,邹春华,等. 小波神经网络的参数初始化研究[J]. 华南理工大学学报:自然科学版, 2003, 31(2):77-84