

基于变精度粗糙集的粗集神经网络

张东波^{①②} 王耀南^② 黄辉先^①

^①(湘潭大学信息工程学院 湘潭 411105)

^②(湖南大学电气与信息工程学院 长沙 410082)

摘要: 本文研究了基于变精度粗糙集模型下的粗集神经网络设计,对 β 近似约简条件进行了弱化推广,同时提出了 β 近似约简的选取原则。在对 Brodatz 纹理图像的分类实验中,比较了经典粗集神经网络 RNN 和变精度粗集神经网络 VPRNN 的性能,VPRNN 不仅具有更为精简的结构和更短的训练时间,而且具有更强的近似决策和泛化能力。

关键词: 变精度粗糙集;粗集神经网络; β 近似约简;纹理分类

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2008)08-1913-05

Rough Neural Network Based on Variable Precision Rough Set

Zhang Dong-bo^{①②} Wang Yao-nan^② Huang Hui-xian^①

^①(Institute of Information Engineering, Xiangtan University, Xiangtan 411105, China)

^②(College of Electrical and Information Engineering, Hunan University, Changsha 410082, China)

Abstract: The design of the rough neural network based on variable precision rough set model is studied. The condition of β -approximation reduction is generalized and the criteria for selecting a β -approximation reduction are introduced. In the experiment of the Brodatz texture image classification, the performance of conventional RNN(Rough Neural Network) and VPRNN(Variable Precision Rough set Neural Network) is compared. The results indicate that VPRNN not only has more simplify structure and less training time, but also, has better approximation decision-making ability and generalization ability than RNN.

Key words: Variable Precision Rough Set (VPRS); Rough Neural Network (RNN); β -approximation reduction; Texture classification

1 引言

近年来,粗集神经网络集成技术的研究引起研究人员的广泛关注,其中,“强耦合”集成方式^[1]的粗集神经网络得到了广泛的应用^[2-5]。在该集成方式下,粗糙集通过对原始数据的分析,在保留原始信息不受损的基础上,提取精简的知识规则,并进而指导神经网络设计。该类粗集神经网络,可以在保证性能的前提下,去除网络冗余,网络具有明确的语义和较快的收敛速度。但在以往的文献中,粗糙集分析往往基于传统的 Pawlak 粗糙集模型,所处理的分类必须是完全正确或肯定的,而没有某种程度的“包含”或“属于”,因此对误分类非常敏感,缺乏对噪声数据的适应能力。Pawlak 粗糙集模型另一局限性是它所处理的对象是已知的且从模型中得到的所有结论仅仅适用于这些对象集。而在实际应用中,往往要将一些小规模对象集中得到的结论应用到大规模的对象集中去,这也限制了依据 Pawlak 模型设计的应用系统的推广泛化能力。Ziarko 提出的变精度粗糙集(VPRS)模

型^[6],通过引入“多数包含”关系,允许一定的误分类,避免了对小的分类误差过度敏感,这一方面完善了近似空间的概念,另一方面也有利于粗糙集理论从认为不相关的数据中发现相关的数据,因此变精度粗糙集成为处理噪声数据,提高泛化能力的有力工具并得到有效应用^[7-9]。本文研究了基于变精度粗糙集的粗集神经网络的设计,并将其用于 Brodatz 纹理图像^[10]的分类。

2 变精度粗糙集模型的 β 近似约简分析和选取原则

2.1 变精度粗糙集模型的 β 近似约简

基于变精度粗糙集的约简称为 β 近似约简,为求取 β 近似约简,定义分类品质指标^[6]:

$$\gamma^\beta(P, D) = \text{card}(\text{POS}^\beta(P, D)) / \text{card}(U) \quad (1)$$

$\text{card}(\cdot)$ 为集合的基数, $\text{POS}^\beta(P, D)$ 为关于条件属性 P 的 β 正域, P 是条件属性 C 的子集, D 为决策属性, $\gamma^\beta(P, D)$ 衡量了论域 U 的元素能够依据概率 β 被归入到一定决策类的元素比例。 $\beta \in (0.5, 1]$,可根据实际需要来确定。任何一个 β 近似约简必须满足条件:

$$(1) \gamma^\beta(C, D) = \gamma^\beta(\text{RED}^\beta(C, D), D);$$

2007-01-19 收到, 2007-09-26 改回

国家自然科学基金(60775047),湖南省自然科学基金(06JJ5112)和湘潭大学跨学科交叉科研项目(06IND12)资助课题

(2)从 $RED^\beta(C,D)$ 中去掉任何属性, 都不能再使上式成立。

$RED^\beta(C,D)$ 是条件属性 C 关于决策属性 D 的一个 β 近似约简。在实际应用中, 条件(1)仍是一个比较苛刻的条件, 本文在后面的纹理图像分类应用中, 将条件(1)弱化为

$$|\gamma^\beta(C,D) - \gamma^\beta(RED^\beta(C,D),D)| < \varepsilon \quad (2)$$

为保证分类性能, ε 通常取得较小, 且可依据实际情况进行调整, 本文取 $\varepsilon = 0.02$ 。

2.2 β 近似约简异常分析

最优 β 值的选择是非常困难的, 在相同分类品质指标 γ^β 下, 满足约简条件的 β 值通常为一个区间范围。 β 近似约简也可能出现异常, 主要表现为 β 区间和正区域的异常变化, 导致约简系统不能完全正确反映原信息系统的基本特征。

表 1 是一给定的信息系统, 可以求得 β 在 (0.5, 1] 区间连续取值时的全部 β 约简, 见图 1。图 1 用图示的方法给出了所有 β 约简的品质指标 γ^β 及对应的 β 区间。

表 1 信息系统

对象	C_1	C_2	C_3	C_4	C_5	C_6	D
O_1	1	1	1	1	1	1	M
O_2	1	0	1	0	1	1	M
O_3	0	0	1	1	0	0	M
O_4	1	1	1	0	0	1	F
O_5	1	0	1	0	1	1	F
O_6	0	0	0	1	1	0	F
O_7	1	0	1	0	1	1	F

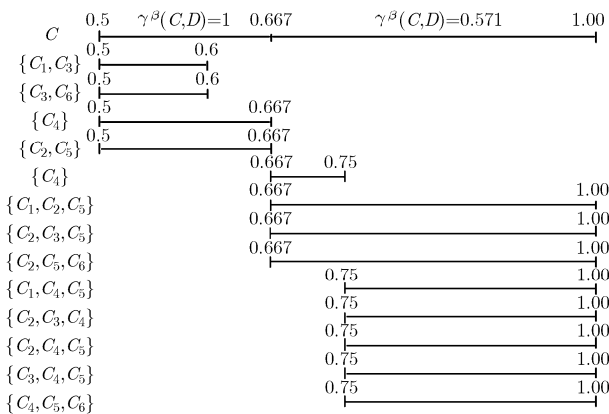


图 1 信息系统表 1 的全部 β 近似约简

在无属性约简且 $\gamma^\beta = 1$ 的情况下, β 取值区间为 (0.5, 0.667], 求得 4 个 β 约简, 其中 $\{C_1, C_3\}$ 和 $\{C_3, C_6\}$ 的 β 取值区间收缩为 (0.5, 0.6]; 而在 $\gamma^\beta = 0.571$ 时, 初始 β 取值区间为 (0.667, 1], 求得 9 个 β 约简, 其中 $\{C_4\}$ 的 β 取值

区间变为 (0.667, 0.75], 出现上限收缩, 而约简 $\{C_1, C_4, C_5\}$, $\{C_2, C_3, C_4\}$, $\{C_2, C_4, C_5\}$, $\{C_3, C_4, C_5\}$, $\{C_4, C_5, C_6\}$ 的 β 取值区间出现了下限收缩, 为 (0.75, 1]。由于约简前后的信息系统应根据相同的 γ^β 和 β 来比较信息的一致性, 在相同 γ^β 前提下, β 取值区间出现异常, 约简属性不能提供与全部属性 C 完全一致的信息。可以求得在无属性约简且 β 在 (0.667, 1] 取值时, β 正域 $POS^\beta(C,D) = \{O_1, O_3, O_4, O_6\}$ 。而其品质指标 $\gamma^\beta(C,D) = 4/7 \approx 0.571$; 而当选约简 $\{C_4\}$, β 取为 (0.667, 0.75] 时, β 正域 $POS^\beta(C_4,D) = \{O_2, O_4, O_5, O_7\}$, $\gamma^\beta(C_4,D) = 4/7 \approx 0.571$ 。虽然约简前后 γ^β 不变, 即 $\gamma^\beta(C,D) = \gamma^\beta(C_4,D) = 0.571$, 但正域中的元素发生了变化, 即 $POS^\beta(C,D) \neq POS^\beta(C_4,D)$, 说明约简改变了原决策系统的分类, 将导致决策系统产生不同的分类决策规则。

2.3 β 近似约简的选取原则

由于 β 近似约简的非唯一性, 同时和品质指标 γ^β 及 β 取值都有较大的关系, 因此, β 近似约简的选取是一个值得研究的问题^[11,12], 有很多的因素需要考虑。还是以表 1 的信息表为例, 对 β 近似约简有关的一些问题进行的实例分析如下:

若 $\beta = 1$ (对应传统粗糙集模型), 则:

$$\begin{aligned} \gamma^1(C,D) &= \text{card}(POS^1(C,D)) / \text{card}(U) \\ &= \text{card}(\{O_1, O_3, O_4, O_6\}) / \text{card}(U) \approx 0.571 \end{aligned}$$

若 $\beta = 0.55$, 则:

$$\begin{aligned} \gamma^{0.55}(C,D) &= \text{card}(POS^{0.55}(C,D)) / \text{card}(U) \\ &= \text{card}(\{O_1, O_3, O_2, O_4, O_5, O_6, O_7\}) / \text{card}(U) \\ &= 7/7 = 1 \end{aligned}$$

在传统粗糙集模型中, 分类品质 $\gamma^1 = 0.571$, 其中 $\{O_1, O_3, O_4, O_6\}$ 被正确分类, 不能决策的样本为 $\{O_2, O_5, O_7\}$; 当 β 降为 0.55 时, 所有对象都能归入到一定的决策类, 分类品质提高到 $\gamma^{0.55} = 1$, 其中 $\{O_1, O_3, O_4, O_5, O_6, O_7\}$ 可以被正确分类, 被错分类的样本只有 $\{O_2\}$ 。尽管 β 减小后, 出现错分类, 但总体的正确分类率从 57.14% 提高到 85.71%, 明显改善了分类性能。

表 2 给出了不同 β 区间对应的 β 约简。随 β 的减小, β 近似约简长度趋于缩短, 同时从图 1 中可知 γ^β 反而提高。 $\gamma^\beta(C,D) = 1$ 时, β 允许取值区间为 (0.5, 0.667], 可能的 β 近似约简有 $\{C_1, C_3\}, \{C_3, C_6\}, \{C_4\}, \{C_2, C_5\}$, 但选约简 $\{C_1, C_3\}, \{C_3, C_6\}$ 时, β 区间仅为 (0.5, 0.6], 出现下限收缩, 因此 $\{C_1, C_3\}, \{C_3, C_6\}$ 实际上并不能提供和全部属性 C 完全一致的信息。而对于约简 $\{C_4\}, \{C_2, C_5\}$, 如果仅从长度上考虑, 应取 $\{C_4\}$, 但选约简 $\{C_4\}$ 时, 被错分类样本有 O_2 和 O_6 ; 而选约简 $\{C_2, C_5\}$ 时, 只有 O_2 被错分。所以, 从尽可能减小错分, 提高分类正确率来看, 约简取为 $\{C_2, C_5\}$ 时分类性能更好。

表 2 不同 β 区间域对应的 β 近似约简

β 区间域	β 近似约简
(0.5 0.6]	$\{C_1, C_3\}, \{C_3, C_6\}, \{C_4\}, \{C_2, C_5\}$
(0.6 0.667]	$\{C_4\}, \{C_2, C_5\}$
(0.667 0.75]	$\{C_4\}, \{C_1, C_2, C_5\}, \{C_2, C_3, C_5\}, \{C_2, C_5, C_6\}$ $\{C_1, C_4, C_5\}, \{C_2, C_3, C_4\}, \{C_2, C_4, C_5\},$
(0.75 1.00]	$\{C_3, C_4, C_5\}, \{C_4, C_5, C_6\}$

通过前文的异常分析和实例分析, 在综合考虑分类品质, 约简长度, β 允许取值区间域, 避免和降低约简异常影响等因素基础上, 本文提出如下 β 近似约简选取原则:

- (1) 根据设定的分类品质指标的下限 γ_{\min}^β , 计算所有分类品质指标 $\gamma^\beta \geq \gamma_{\min}^\beta$ 的 β 近似约简;
- (2) 在原则(1)得到的 β 近似约简中, 选取错分类样本数不大于某一阈值 Δ (可参考基于全部属性 C 分类时的错分类样本数来确定一适当的值)的 β 近似约简;
- (3) 在原则(2)选取的 β 近似约简中, 寻找和全体属性 C 具有最相似 β 取值区间域的 β 近似约简, 同时剔除掉出现分类异常的约简;
- (4) 在上述处理后, 寻找具有最短长度的 β 近似约简。

3 变精度粗糙集神经网络的结构

基于“强耦合”的粗糙集神经网络结构设计^[5,13], 本文构建了图 2 的粗糙集神经网络。

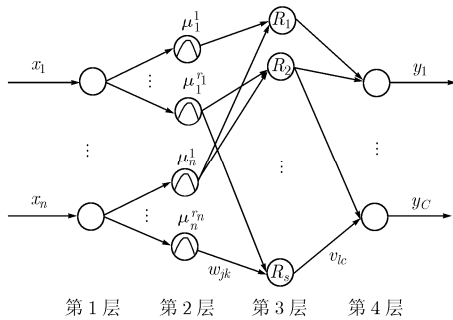


图 2 粗糙集神经网络结构

第 1 层为输入层, 输入 $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, $x_i (i = 1, 2, \dots, n)$ 是输入特征变量。

第 2 层为离散量化层, 分别将输入变量 x_i 离散化为 r_i 个不同值, 激励函数为高斯函数:

$$\mu_i^j = \exp\left[-\frac{x_i - c_{ij}}{\delta_{ij}}\right]^2 \quad (3)$$

c_{ij} , δ_{ij} 是变量 x_i 离散化到第 $j (j = 1, 2, \dots, r_i)$ 档的平均值和方差, 可由训练样本进行估算。

第 3 层为规则层, 规则的获取和描述依据的是变精度粗

糙集模型。该层神经元和第 2、4 层神经元的连接关系, 由规则的前件和后件来决定, 该层第 k 个神经元 R_k 的输出:

$$O_k^3 = f_3\left(\sum_{j=1}^q w_{jk} \mu_i^j + \beta_k\right) \quad (4)$$

$k = 1, 2, \dots, s$, $j = 1, 2, \dots, q$, s 是规则数, q 是和第 k 条规则相关的条件属性数, f_3 是规则层的激励函数, 取为正切 S 函数 tansig。

第 4 层为输出决策层, 每一个神经元代表一个决策类, 第 c 个神经元的输出:

$$O_c^4 = f_4\left(\sum_{l=1}^p v_{lc} O_l^3 + \theta_c\right) \quad (5)$$

$c = 1, 2, \dots, C$, $l = 1, 2, \dots, p$, C 是决策类数, p 是和第 c 类相关的规则数, f_4 是输出层的激励函数, 取为对数 S 函数 logsig。

网络权值参数 w_{jk} , v_{lc} , 及神经元阈值 β_k , θ_c , 可由反向传播 BP 算法来进行训练。

4 纹理图像分类实验

本文在文献[10]的纹理图像集中, 选择了 10 类纹理图像 (D1, D3, D20, D47, D49, D53, D55, D78, D82, D85) 进行分类。采用基于奇异值分解 SVD(Singular Value Decomposition)和主成分分析 PCA(Principal Components Analysis)方法进行特征提取^[14]。每一个纹理图像均被分割为 400 个不重叠的 32×32 的子图像, 总共获得 4000 个样本图像。选其中 1000 个样本进行训练, 500 个作为校验集, 其余 2500 个样本作为测试集。

本文最终在 6 维 PCA 特征空间设计分类器。样本条件属性为 $\{C_1, C_2, C_3, C_4, C_5, C_6\}$, 由等频离散法将每一个属性划分为 3 个等级, 在 Pawlak 粗糙集模型下, 约简为本身, 即 $\{C_1, C_2, C_3, C_4, C_5, C_6\}$, 从决策表中提取到 335 条规则。而在变精度粗糙集模型下 (取 $\gamma(C, D) \geq 0.9$), 此时 β 取值区间为 (0.5 0.67], 计算出所有的 β 近似约简后, 根据约简选取原则, 最后选择 β 近似约简为 $\{C_1, C_2, C_3, C_4, C_5\}$, 相应提取到 158 条规则。根据提取的规则, 分别构建普通粗糙集神经网络 RNN 和变精度粗糙集神经网络 VPRNN。

表 3 是 RNN 和 VPRNN 的比较, 其中, RNN 规则数, 需调节的权值数, 训练时间都要远大于相应的 VPRNN。显然 VPRNN 比 RNN 结构更为精简, 训练时间更短。而从 MSE(样本平均均方差)的比较来看, 虽然, RNN 的训练误差更小, 但是校验误差反而高于 VPRNN, 这也表明, VPRNN 对未知样本的分类能力强于 RNN, 即 VPRNN 具有更强的泛化推广能力。

表 4 是训练样本的分类结果, 除比较了 RNN 和 VPRNN 的分类性能外, 也给出了直接采用粗集规则进行分类的结果。在直接采用粗集规则进行分类时, 所有的训练样本均可

表3 RNN网络和VPRNN网络的比较

	规则数	需调节的权值数	训练时间(s)	训练集MSE	校验集MSE
RNN	335	2407	1578	0.003731	0.006977
VPRNN	158	1029	861	0.003835	0.005668

表4 各类纹理的训练样本分类精度(%)

	粗集规则分类				粗集神经网络分类	
	经典粗集规则		变精度粗集规则		RNN	VPRNN
	正确分类样本	不能决策样本	正确分类样本	不能决策样本		
D1	89.00	0.00	86.00	0.00	96.00	94.00
D3	97.00	0.00	91.00	0.00	97.00	97.00
D20	95.00	0.00	83.00	0.00	98.00	96.00
D47	89.00	0.00	83.00	0.00	93.00	94.00
D49	84.00	0.00	75.00	0.00	95.00	95.00
D53	93.00	0.00	93.00	0.00	99.00	98.00
D55	100.00	0.00	99.00	0.00	99.00	97.00
D78	97.00	0.00	96.00	0.00	100.00	100.00
D82	80.00	0.00	77.00	0.00	97.00	97.00
D85	81.00	0.00	71.00	0.00	94.00	91.00
全部	90.50	0.00	85.40	0.00	96.80	95.90

决策分类。经典粗集规则(条件属性6个)和变精度粗集规则(条件属性5个)中,经典粗集规则对训练样本中信息的描述更为精细,因此,其分类性能优于变精度粗集规则,两者的分类精度分别为90.50%和85.40%。表5是测试样本的分类结果,对测试样本来说,由于训练样本集规模较小,而且由训练集分析获得的规则仅适用于训练样本,因此,在对大量的未知样本进行分类时,不可避免会遇到很多不能决策的样本,正确分类率急剧下降,变为71.64%。而变精度粗糙集模型由于允许一定的错分类,使得能被决策分类的样本数大大增加,不能决策的样本比例从12.64%降为4.56%,而正确分类率从71.64%上升为75.60%。

经典粗集规则和变精度粗集规则对测试样本的正确分类率分别为71.64%和75.60%,但同时分别存在12.64%和4.56%的不能决策样本。而集成了神经网络技术的RNN和VPRNN推广泛化能力大大增强,不仅避免了不能决策的问题,而且对测试样本的正确分类率分别大幅提高到89.08%和90.42%。

VPRNN对训练样本的正确分类率(95.90%)低于RNN的正确分类率(96.80%),但大部分测试样本在VPRNN中的分类性能都有一定程度的提高,从而使得整体正确分类率达到90.42%高于RNN的89.08%,说明VPRNN具有更强的推广泛化能力。VPRNN对训练样本的分类精度低于RNN,是因为变精度粗糙集约简属性少,因而构造的信息粒度较

表5 各类纹理的测试样本分类精度(%)

	粗集规则分类				粗集神经网络分类	
	经典粗集规则		变精度粗集规则		RNN	VPRNN
	正确分类样本	不能决策样本	正确分类样本	不能决策样本		
D1	72.00	13.20	70.40	5.20	91.60	94.00
D3	65.60	9.20	74.40	5.20	88.00	91.60
D20	70.80	17.60	72.80	8.80	91.20	94.40
D47	65.20	16.00	72.00	5.20	83.20	86.00
D49	69.60	6.80	64.80	4.00	88.00	84.00
D53	83.20	4.40	84.40	2.00	96.80	95.60
D55	85.20	14.00	99.60	0.40	94.00	95.80
D78	93.20	6.40	96.40	1.20	98.00	98.80
D82	67.60	10.40	71.20	3.60	90.00	92.00
D85	44.00	28.40	50.00	10.00	70.00	72.00
全部	71.64	12.64	75.60	4.56	89.08	90.42

大,对已知样本描述较为粗糙。而VPRNN对测试样本的分类精度高于RNN,是由于变精度粗糙集允许一定程度的错误分类率存在,能够描述统计意义上的数据模式或者概率上的不确定分类问题,从而具有较强的近似决策能力。原来在RNN中难以正确决策的样本,VPRNN在统计和概率的意义上进行了近似决策,因而提高了分类性能。

5 结束语

(1)VPRNN和RNN一样,都具有明确的语义,两者均类似于局部逼近神经网络,因此有较快的收敛速度,两者在训练迭代次数上差别不是很大,但VPRNN由于具有更为精简的结构,需调整的权值数少于RNN,所以训练时间大大缩短;

(2)VPRNN是一种具有近似决策能力的神经网络,尤其适用于噪声数据的处理及小样本情况下的分类系统设计,与RNN相比,明显具有更强的推广泛化能力。

如何根据实际要求,兼顾分类品质指标 γ^β 和 β 取值仍是一个有待解决的问题。如何在保证分类性能的条件下,选取恰当的离散化精度,也值得进一步研究。

参考文献

- [1] 张东波,王耀南,易灵芝. 粗集神经网络及其在智能信息处理领域的应用. 控制与决策, 2005, 20(2): 121-126.
Zhang D B, Wang Y N, and Yi L Z. Rough neural network and its application to intelligent information processing. *Control and Decision*, 2005, 20(2):121-126.
- [2] Banerjee M, Mitra S, and Pal S K. Rough fuzzy MLP: knowledge encoding and classification. *IEEE Trans. on Neural Networks*, 1998, 9(6): 1203-1216.
- [3] Yu C Y, Wu M H, and Wu M. Combining rough set theory

- with neural network theory for pattern recognition. International Conference on Robotics, Intelligent Systems and Signal Processing, Changsha, 2003: 880-885.
- [4] Chen S Y and Yi J K. A fuzzy neural network based on rough sets and its applications to chemical production process. Proceedings of Info-tech and Info-net, Beijing Oct 29- Nov 1, 2001, 4: 405-410.
- [5] Wu Z C. Research on remote sensing image classification using neural network based on rough sets. Proceedings of Info-tech and Info-net, Beijing Oct 29-Nov 1, 2001, 4: 279-284.
- [6] Ziarko W. Variable precision rough sets model. *Journal of Computer and System Sciences*, 1993, 46(1): 39-59.
- [7] Malcolm J B and Michael J P. Variable precision rough set theory and data discretisation: an application to corporate failure prediction. *The International Journal of Management Science*, 2001, 29(6): 561-576.
- [8] Zhao Y Q, Zhang H C, and Pan Q. Classification using the variable precision rough set. RSFDGrC 2003, Chongqin, 2003, LNAI 2639: 350-353.
- [9] Malcolm J B. The identification of low-paying workplaces: An analysis using the variable precision rough sets model. RSCTC 2002, Malvern, PA, USA, 2002, LNAI 2475: 530-537.
- [10] Brodatz Textures [Online]. Available: <http://sipi.usc.edu/database/database.cgi? volume=textures&image=24>, 2005. 10.18.
- [11] Malcolm J B. Reducts within the variable precision rough sets model: A further investigation. *European Journal of Operational Research*, 2001, 134(3): 592-605.
- [12] Malcolm J B. An investigation of β -reduct selection within the variable precision rough sets model. RSCTC'2000, Banff, Canada, October 16-19, 2000, LNAI 2005: 114-122.
- [13] 李永敏, 朱善君, 陈湘辉, 等. 根据粗糙集理论进行 BP 网络设计的研究. *系统工程理论与实践*, 1999, 19(4): 62-69.
- Li Y M, Zhu S J, and Chen X H, *et al.* A rough set approach to BP neural network designing. *Systems Engineering-Theory & Practice*, 1999, 19(4): 62-69.
- [14] Swiniarski R W and Hargis L. Rough sets as a front end of neural-network texture classifiers. *Neurocomputing*, 2001, 36 (1-4): 85-102.
- 张东波: 男, 1973 年生, 副教授, 博士生, 从事粗糙集、神经网络、图像处理和模式识别等的研究。
- 王耀南: 男, 1957 年生, 教授, 博士生导师, 从事人工智能、智能控制、图像处理和模式识别等的研究。
- 黄辉先: 男, 1957 年生, 教授, 从事智能交通控制、优化算法等的研究。