

中图分类号: TP391 文献标志码: A 文章编号: 1006-8961(2011)08-1408-10

论文索引信息: 吴军, 王士同, 赵鑫. 正负模糊规则系统、极限学习机与图像分类 [J]. 中国图象图形学报, 2011, 16(8): 1408-1417

# 正负模糊规则系统、极限学习机与图像分类

吴军, 王士同, 赵鑫

(江南大学数字媒体学院, 无锡 214122)

**摘要:** 传统的图像分类一般只利用了图像的正规则, 忽略了负规则在图像分类中的作用。Nguyen 将负规则引入图像分类, 提出将正负模糊规则相结合形成正负模糊规则系统, 并将其用于遥感图像和自然图像的分类。实验证明, 其在图像分类过程中取得了很好的效果。他们提出的前馈神经网络模型在调整权值时利用了梯度下降法, 由于步长选择不合理或陷入局部最优从而使训练速度受到了限制。极限学习机 (ELM) 是一种单隐层前馈神经网络 (SLFN) 学习算法, 具有学习速度快, 泛化性能好的优点。本文证明了极限学习机与正负模糊规则系统的实质是等价的, 遂将其用于图像分类。实验结果说明了极限学习机能很好的利用正负模糊规则相结合的方法对图像进行分类, 实验结果较为理想。

**关键词:** 图像分类; 正负模糊规则系统; 极限学习机

## Positive and negative fuzzy rule system, extreme learning machine and image classification

Wu Jun, Wang Shitong, Zhao Xin

(School of Digital Media, Jiangnan University, Wuxi 214122 China)

**Abstract:** The positive fuzzy rules often were used only for image classification in the traditional image classification system, while the negative image classification rules were ignored in effect. Nguyen introduced the negative Fuzzy rules into the image classification, proposed a combination of positive and negative fuzzy rules to form the positive and negative fuzzy rule system, and then applied it to remote sensing image/natural image classification. Their experiments proved that their proposed method has achieved good results. However, since their method was realized using the feed forward neural network model which adjust the weights in the gradient descent, the training speed is very slow. Extreme learning machine (ELM) is a single hidden layer feed forward neural network (SLFN) learning algorithm, which has advantages such as quick learning, good generalization performance. In this paper, it proves that Extreme Learning Machine (ELM) and the positive and negative fuzzy rule system is essentially equivalent, so ELM can be naturally used for image classification. Our experimental results support this claim.

**Keywords:** image classification; positive and negative fuzzy rules; extreme learning machine

## 0 引言

随着数字图像处理技术的发展和实际应用的需

求, 图像分类识别技术应运而生, 并得到快速发展。图像分类技术是模式识别技术在图像处理领域中的具体应用, 其目的是研制出能代替人们完成图像分类和识别任务的并能够自动处理图像的计算机系

收稿日期: 2010-07-13; 修回日期: 2010-11-17

基金项目: 国家自然科学基金项目 (60704047); 国家自然科学基金项目 (9082002)。

第一作者简介: 吴军 (1986—), 男, 江南大学数字媒体学院计算机应用技术专业硕士研究生, 主要研究方向为图像分类。

E-mail: wujunqj@163.com。

统。图像分类的结果为实验和研究提供了很多重要信息。随着神经网络理论的发展,其在图像分类中的应用也不断增多。前馈神经网络具有很强的非线性逼近能力,但神经网络的学习速度往往不能满足人们的需求。Nguyen<sup>[1]</sup>提出利用正负模糊规则相结合的方法对图像进行分类,但传统的基于正负模糊规则相结合的前馈神经网络模型中所有的参数都需要调整,不同的参数之间存在相互依赖性且在误差反向传播过程中采用梯度下降法调整权值,容易进入局部最优,从而使学习速度受到限制,并对其广泛应用造成了很大影响。极限学习机(ELM)<sup>[2]</sup>与传统的神经网络有很大不同,它不需要对所有的权值进行调整,隐含层的权值和偏差是任意选取的,极大的提高了神经网络的学习速度。

在传统图像分类中,大多忽略了负规则在图像分类中的作用,只利用正规则对图像进行分类。在模糊规则系统中,一种典型的模糊分类规则 R<sup>[3-5]</sup>: IF  $x_{1k}$  is  $A_{r1}$  and  $x_{2k}$  is  $A_{r2}$  ... and  $x_{Nk}$  is  $A_{rN}$ , then  $y_{k1}$  is  $C_1$  with  $W_{r1}$  and  $y_{k2}$  is  $C_2$  with  $W_{r2}$  ... and  $y_{km}$  is  $C_m$  with  $W_{rm}$ 。其中,  $\mathbf{x}_k = [x_{1k}, x_{2k}, \dots, x_{Nk}]$  是  $N$  维输入,  $A_m, n = (1, 2, \dots, N)$  是相应的隶属度函数,  $W_{rm} \geq 0, m = (1, 2, \dots, M)$  是权值,  $\mathbf{y}_k = [y_{k1}, y_{k2}, \dots, y_{km}]$  是对应的输出,  $C_1, C_2, \dots, C_M$  是  $\mathbf{x}_k$  所属类。由此可以看出,在模糊规则库中,大部分的模糊分类规则只考虑了正规则(其权值  $W_{rm}$  为正数)。这对传统的数据挖掘来说是一种无形的限制。有些研究者虽然指出了负规则在图像分类中的重要性,但很少有人从事这一方面的研究,仅有极少数科研人员<sup>[6-8]</sup>提出一些算法挖掘这种类型的关联规则。在文献[1]中 Nguyen 将负规则引入图像分类,利用正负模糊规则相结合的方法对图像进行分类,并在图像分类过程中取得了很好的效果。利用极限学习机算法获得输出权值矩阵,并利用正负模糊规则相结合的方法对图像进行分类。

## 1 正负模糊规则系统

在模糊系统规则库中,假设 A 是规则的前提, B 是规则的结果。在正规则模糊系统规则库中,只有正规则,其权值为正值。典型的正规则的形式为:“IF A Then do B”,而负规则模糊系统规则库中,只包含负规则,其权值为负值,典型的负规则的形式为:“IF A Then do not do B”。在图像分类过程中,

这两种系统都只利用了图像的部分信息,从而造成了信息的浪费。

来看以下两个“IF...Then...”结构的模糊规则。

规则 1: IF 消费者年龄小于 18 岁

Then 他可以喝饮料但不可以喝酒

规则 2: IF 消费者年龄大于 18 岁

Then 他可以喝饮料也可以喝酒

从这个例子可以看出,负规则(规则 1)指导系统避免一些不良情况。当避开这些情况后,正模糊规则开始指导系统工作,系统有效地利用了负规则,基于这种系统,可以有效地控制青少年酗酒的不良习惯。将正负模糊分类规则相结合,将有效提高系统工作效率。正负模糊分类规则相结合的结构可以表示为:

规则 R: IF  $x_{1k}$  is  $A_{r1}$  and  $x_{2k}$  is  $A_{r2}$  ... and  $x_{Nk}$  is  $A_{rN}$ , Then  $y_{k1}$  is  $C_1$  with  $W_{r1}$  and  $y_{k2}$  is  $C_2$  with  $W_{r2}$  ... and  $y_{km}$  is  $C_m$  with  $W_{rm}$ 。

其中,  $\mathbf{x}_k = [x_{1k}, x_{2k}, \dots, x_{Nk}]$ ,  $k = 1, 2, \dots, K$  是  $N$  维输入,  $C_1, C_2, \dots, C_M$  是  $\mathbf{x}_k$  所属类,  $A_m, n = (1, 2, \dots, N)$  是相应的隶属度函数,  $W_{rm} \in \mathbf{R}, r = (1, 2, \dots, R), m = (1, 2, \dots, M)$  是  $\mathbf{x}_k$  所属类的权值。当  $W_{rm}$  为正数时,表示  $\mathbf{x}_k$  属于  $C_m$  类的权重;当  $W_{rm}$  为负数时,就会减小  $\mathbf{x}_k$  所属类的选择范围。 $N, R, K, M$  分别表示输入的维数、规则的条数、输入集的个数,以及所需分类的类数。

这是一个多输入多输出的模糊神经网络,将其第  $m$  个输出定义为

$$y_{km} = \frac{\sum_{r=1}^R \beta_r(\mathbf{x}_k) W_{rm}}{\sum_{r=1}^R \beta_r(\mathbf{x}_k)} = \frac{\sum_{r=1}^R W_{rm} \beta_r(\mathbf{x}_k)}{\sum_{r=1}^R \beta_r(\mathbf{x}_k)} = \frac{\sum_{r=1}^R W_{rm} \exp\left[-\sum_{n=1}^N \frac{(x_{nk} - \mu_{rn})^2}{\sigma_{rn}^2}\right]}{\sum_{r=1}^R \exp\left[-\sum_{n=1}^N \frac{(x_{nk} - \mu_{rn})^2}{\sigma_{rn}^2}\right]} \quad (1)$$

式中,  $\beta_r(\mathbf{x}_k)$  分别表示为

$$\beta_r(\mathbf{x}_k) = \frac{\beta_r(\mathbf{x}_k)}{\sum_{r=1}^R \beta_r(\mathbf{x}_k)}$$

$$\beta_r(\mathbf{x}_k) = \exp\left[-\sum_{n=1}^N \frac{(x_{nk} - \mu_{rn})^2}{\sigma_{rn}^2}\right] \quad (2)$$

$\mu_{rn}, \sigma_{rn} r = (1, 2, \dots, R); n = (1, 2, \dots, N)$  分别为

相应高斯函数的中心和宽度,  $W_{rm}$ ,  $m = (1, 2, \dots, M)$  为输出权值。文中根据最大隶属度原则确定  $\mathbf{x}_k$  所属的类

$$\mathbf{y}_k = C_{m'}, m' = \max_{1 \leq m \leq M} (y_{km}) \quad (3)$$

图 1 是传统的基于正负模糊规则相结合的模糊神经网络模型<sup>[1]</sup>, 该模糊神经网络是一种前馈神经网络, 由输入层(第 1 层)、输出层(第 5 层)和隐含层组成(第 2 层、第 3 层、第 4 层)。

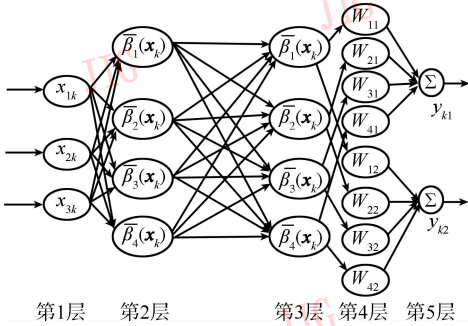


图 1 3 个输入 ( $N = 3$ ), 4 条规则 ( $R = 4$ ), 两个输出 ( $M = 2$ ) 对应的模糊神经网络

Fig. 1 Fuzzy neural network with 3 inputs ( $N = 3$ ) classes ( $M = 2$ ) and 4 fuzzy rules ( $R = 4$ )

第 1 层 该层中每个神经元将输入值  $x_{nk}$ ,  $n = (1, 2, \dots, N)$ ,  $k = (1, 2, \dots, K)$  直接传到下一层, 每个神经元的输出就是它的输入, 即  $O_{1n} = x_{nk}$ 。

第 2 层 本层中每个神经元代表一个模糊规则, 文中采用高斯函数作为隶属度函数, 第  $r$  个神经元的输出为

$$O_{2rk} = \beta_r(\mathbf{x}_k) = \exp\left[-\sum_{n=1}^N \frac{(x_{nk} - \mu_{rn})^2}{\sigma_{rn}^2}\right] \quad (4)$$

式中  $\mu_{rn}$ ,  $\sigma_{rn}$ ,  $r = (1, 2, \dots, R)$ ;  $n = (1, 2, \dots, N)$  是常数, 它们分别是高斯函数的中心和宽度。通过对网络训练得到最优值。

第 3 层 对各条规则的隶属度做归一化处理, 其输出为

$$O_{3r} = \bar{\beta}_r(\mathbf{x}_k) = \frac{\beta_r(\mathbf{x}_k)}{\sum_{r=1}^R \beta_r(\mathbf{x}_k)} \quad (5)$$

第 4 层 该层中每个节点都有一个输出权值  $W_{rm}$ ,  $W_{rm} = w_{rm0} + w_{rm1}x_{1k} + w_{rm2}x_{2k} + \dots + w_{rmN}x_{Nk}$ , 而参数  $w_{rml}$ ,  $r = (1, 2, \dots, R)$ ,  $m = (1, 2, \dots, M)$ ,  $l = (1, 2, \dots, N)$  是通过最小二乘法得到的。在文中,  $\mathbf{x}_k$  对应的输出  $\mathbf{y}_k$  由最大隶属度原则确定, 所以  $W_{rm}$  取负值时, 就会减少选择  $C_m$  类的可能性

( $W_{rm}$  所取的负值越多,  $y_{km}$  的值越小)。这一层有  $R \times M$  个神经元, 每个神经元的输出为

$$O_{4rm} = W_{rm} \frac{\beta_r(\mathbf{x}_k)}{\sum_{r=1}^R \beta_r(\mathbf{x}_k)} = W_{rm} \bar{\beta}_r(\mathbf{x}_k) \quad (6)$$

第 5 层 该层共有  $M$  个神经元, 每个神经元对应的输出为

$$y_{km} = \sum_{r=1}^R O_{4rm} = \sum_{r=1}^R \left[ W_{rm} \frac{\beta_r(\mathbf{x}_k)}{\sum_{r=1}^R \beta_r(\mathbf{x}_k)} \right] = \sum_{r=1}^R W_{rm} \bar{\beta}_r(\mathbf{x}_k) \quad (7)$$

传统的正负模糊规则系统在遥感图像和自然图像的分类过程中取得了较好的图像分类效果, 但也存在明显的不足: 1) 传统的前反馈神经网络中所有参数都需要调整, 不同参数间往往存在相互依赖性。2) 通过梯度下降法来调整权值, 而梯度下降法具有局部最优值, 从而无法达到全局最优点。3) 学习率的选取具有不确定性, 学习率过小, 则学习速度慢; 学习率过大, 算法会不稳定。4) 由于在误差反向传播过程中采用梯度下降法调整权值, 在样本数量较大时, 神经网络的训练速度不够理想。5) 采用误差的反向传播方法调整权值很容易通过训练, 影响神经网络的推广性。基于以上几点不足, 引入极限学习机理论<sup>[2]</sup>。

## 2 极限学习机

### 2.1 单隐层前馈神经网络

对任意  $N$  个独立的样本  $(\mathbf{x}_k, \mathbf{y}_k)$ , 其中  $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kN}]^T$ ,  $\mathbf{y}_k = [y_{k1}, y_{k2}, \dots, y_{kM}]^T$ , 典型的含  $R$  个隐含层节点, 激活函数为  $f(\mathbf{a}_r \cdot \mathbf{x}_k + b_r)$  的单隐层神经网络的数学模型为

$$O_k = \sum_{r=1}^R \mathbf{w}_r f_r(\mathbf{x}_k) = \sum_{r=1}^R \mathbf{w}_r f(\mathbf{a}_r \cdot \mathbf{x}_k + b_r) \quad (8)$$

$k = (1, 2, \dots, K)$

式中,  $\mathbf{a}_r = [a_{r1}, a_{r2}, \dots, a_{rN}]^T$  是输入层到隐含层第  $r$  个神经元的域值向量,  $b_r$  为第  $r$  个神经元偏差,  $\mathbf{w}_r = [w_{r1}, w_{r2}, \dots, w_{rM}]^T$  是隐含层第  $r$  个神经元到输出层的权值。

含  $R$  个隐含层神经元, 激活函数为  $f(x)$  的标准单隐层前馈神经网络可以没有任何误差的逼近  $N$  个不同样本。即存在  $\mathbf{a}_r$ ,  $b_r$ ,  $\mathbf{w}_r$  满足

$$y_k = \sum_{r=1}^R w_r f(a_r \cdot x_k + b_r) \quad (9)$$

$k = (1, 2, \dots, K)$

上式可以写成如下形式

$$HW = Y \quad (10)$$

式中

$$H(a_1, a_2, \dots, a_R, b_1, b_2, \dots, b_R, x_1, x_2, \dots, x_K) = \begin{bmatrix} f(a_1 \cdot x_1 + b_1) & f(a_2 \cdot x_1 + b_2) & \dots & f(a_R \cdot x_1 + b_R) \\ f(a_1 \cdot x_2 + b_1) & f(a_2 \cdot x_2 + b_2) & \dots & f(a_R \cdot x_2 + b_R) \\ \vdots & \vdots & \vdots & \vdots \\ f(a_1 \cdot x_K + b_1) & f(a_2 \cdot x_K + b_2) & \dots & f(a_R \cdot x_K + b_R) \end{bmatrix}_{K \times R} \quad (11)$$

$$W = \begin{bmatrix} w_1^T \\ w_2^T \\ \vdots \\ w_R^T \end{bmatrix}_{R \times M} \quad Y = \begin{bmatrix} y_1^T \\ y_2^T \\ \vdots \\ y_K^T \end{bmatrix}_{K \times M} \quad (12)$$

$a_r \cdot x_k$  是  $a_r$  和  $x_k$  的内积,  $H$  是隐含层对应的输出矩阵。

如果激活函数在任何区间上都是无限可微的, 现有的资料表明(文献[1]中已有详细的证明): 对于  $N$  个独立的训练样本  $(x_k, y_k)$ , 其中  $x_k \in \mathbf{R}^N$ ,  $y_k \in \mathbf{R}^M$ , 权值  $a_r, b_r$  是随机选取的, 若隐含层神经元个数等于训练样本的个数, 即  $R = K$ , 则隐含层对应的输出矩阵为方阵且以概率 1 可逆, 并且  $\|HW - Y\| = 0$ 。但是在很多情况下, 由于训练样本数量庞大, 所取的隐层神经元个数往往小于训练样本的个数, 即  $R < K$ , 对于任意给定的正数  $\varepsilon$ , 存在  $R^*$  ( $R^* < K$ ) 个神经元的单隐层前馈神经网络使得  $\|H_{K \times R} W_{R \times M} - Y_{K \times M}\| \leq \varepsilon$ 。

对于传统的单隐层神经网络, 往往希望通过训练神经网络得到特定的  $\hat{a}_r, \hat{b}_r, \hat{W}$  ( $r = 1, 2, \dots, R$ ), 使得

$$\|H(\hat{a}_1, \dots, \hat{a}_R, \hat{b}_1, \dots, \hat{b}_R) \hat{W} - Y\| = \min_{\hat{a}_r, \hat{b}_r, \hat{W}} \|H(a_1, \dots, a_R, b_1, \dots, b_R) \hat{W} - Y\| \quad (13)$$

上式等价于代价函数

$$E(\Theta) = \sum_{k=1}^K \left( \sum_{r=1}^R w_r f(a_r \cdot x_k + b_r) - y_k \right)^2 \quad (14)$$

式中, 参数  $\Theta = [a_r, b_r, w_r]$ 。当  $H$  未知时, 常用梯度下降法来调整参数  $\Theta = [a_r, b_r, w_r]$  使  $\|HW - Y\|$  达到最小, 而参数  $\Theta$  的学习规则为

$$\Theta^{\text{new}} = \Theta^{\text{old}} - \eta \frac{\partial E(\Theta)}{\partial \Theta} \quad (15)$$

式中,  $\eta$  是学习率。

较为流行的反传学习算法, 梯度在误差的方向传播过程中能被有效地计算出来, 但是反传算法也有以下较为明显的缺陷:

- 1) 学习率较难确定, 若学习率太小, 则网络的学习速度缓慢; 若学习率过大, 会导致学习过程不稳定。
- 2) 存在局部最优解, 当局部最优值与全局最优值相差较远时很难达到最优值。
- 3) 在样本数量较大时, 对神经网络的训练速度不够理想。
- 4) 当神经网络过度学习时往往会影响网络的泛化能力, 在训练过程中要有合适的验证和停机方法。

为解决上述梯度下降法带来的问题, Huang<sup>[2]</sup> 提出更有效的单隐层前馈神经网络学习算法。单隐层神经网络的激活函数在无限可微的情况下, 输入权值和偏差可以任意选取而不需要调整。在网络开始训练时, 一旦将随机值赋给输入权值和隐层偏差, 隐含层对应的输出矩阵  $H$  将保持不变。在固定输入权值和隐层偏差的情况下, 训练单隐层前馈神经网络就等价于寻找线性系统  $HW = Y$  的一个最小范数最小二乘解  $W$ , 式(13)就可以写成如下形式

$$\|H(\hat{a}_1, \dots, \hat{a}_R, \hat{b}_1, \dots, \hat{b}_R) \hat{W} - Y\| = \min_{\hat{W}} \|H(a_1, \dots, a_R, b_1, \dots, b_R) \hat{W} - Y\| \quad (16)$$

上式的最小范数最小二乘解为

$$W = H^\dagger Y \quad (17)$$

式中,  $H^\dagger$  为隐层输出矩阵  $H$  的 Moore-Penrose 伪逆。该特殊解具有以下特性。

- 1) 最小的训练误差。特殊解  $W = H^\dagger Y$  是线性系统的一个最小二乘解, 因此, 可以达到最小训练误差

$$\|H \hat{W} - Y\| = \|H H^\dagger Y - Y\| = \min_{\hat{W}} \|H W - Y\| \quad (18)$$

然而, 对于大部分学习算法而言, 由于局部最小值或是训练次数等原因而无法达到最小的训练误差。

- 2) 最小权值范数。特殊解  $W = H^\dagger Y$  在线性系统  $HW = Y$  所有最小二乘解中具有最小范数

$$\|\hat{W}\| = \|H^\dagger Y\| \leq \|W\|$$

$$\forall W \in \{W: \|HW - Y\| \leq \|H z - Y\|, \forall z \in \mathbf{R}^{R \times M}\} \quad (19)$$

Bartlett<sup>[9]</sup> 指出, 对前馈神经网络而言, 当同时具有很小的训练方差和很多较小的连接权值时, 网络推广性和网络连接的权值有关, 权值越小, 其所具有

的推广性也就越好。该方法不仅有最小的训练误差,而且有最小的权值范数,因此,该方法有较好的推广性。

### 2.2 极限学习机学习算法

基于上述,文献[2]把一种学习算法简单的单隐层前馈神经网络叫做极限学习机。给定一个训练数据集  $S = \{(\mathbf{x}_k, \mathbf{y}_k) \mid \mathbf{x}_k \in \mathbf{R}^N, \mathbf{y}_k \in \mathbf{R}^M, k = 1, 2, \dots, K\}$ , 激活函数  $f(x)$ , 隐含层神经元个数为  $R$ 。

- 1) 将输入权值  $\mathbf{a}_r$  和隐含层偏差  $b_r$  赋随机值,  $r = 1, 2, \dots, R$ 。
- 2) 计算隐含层对应的输出矩阵  $\mathbf{H}$ 。
- 3) 计算输出权值

$$\mathbf{W} = \mathbf{H}^\dagger \mathbf{Y} \quad (20)$$

式中,  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_K]^T$ 。

极限学习机具有很强的逼近和分类能力,能有效地处理多元问题。极限学习机算法可以对输入权值  $\mathbf{a}_r$  和隐含层偏差  $b_r$  赋随机值,涉及的计算量很小,算法简洁高效,不需要设置学习率,不存在局部极值,得到的输出权值具有全局最优性,具有训练速度快、泛化性能好的优点。极限学习机算法中计算单元(极测机神经网络的隐含层节点数)在一定范围内的选取,计算结果不是很敏感,可通过交叉检验优化选取的方法获得最佳的计算单元数。极限学习机不仅能最小化经验风险(最小训练误差),而且能最小化推广误差(最小权值范数),具有较好的推广性。

### 3 基于极限学习机的正负模糊规则系统的图像分类

极限学习机与正负模糊规则系统的实质是等价的。二者都是利用神经网络来实现基于正负模糊系统的图像分类,二者的激活函数的类型也是一致的。由 ELM 算法得到的输出权值  $W_{rm}$  是实数,  $r = (1, 2, \dots, R)$ ,  $m = (1, 2, \dots, M)$ ; 正负模糊系统规则库中规则的典型形式为

规则 R: IF  $x_{1k}$  is  $A_{r1}$  and  $x_{2k}$  is  $A_{r2}$  ...and  $x_{Nk}$  is  $A_{rN}$ , Then  $y_{k1}$  is  $C_1$  with  $W_{r1}$  and  $y_{k2}$  is  $C_2$  with  $W_{r2}$  ...and  $y_{kM}$  is  $C_M$  with  $W_{rM}$ , 式中的权值  $W_{rm}$  也是实数;而正规则模糊系统中的模糊规则的典型形式

规则 R<sup>[3-5]</sup>: IF  $x_{1k}$  is  $A_{r1}$  and  $x_{2k}$  is  $A_{r2}$  ...and  $x_{Nk}$  is  $A_{rN}$ , then  $y_{k1}$  is  $C_1$  with  $W_{r1}$  and  $y_{k2}$  is  $C_2$  with  $W_{r2}$  ...and  $y_{kM}$  is  $C_M$  with  $W_{rM}$ , 式中的权值  $W_{rm}$  是正数;同理,负模

糊规则系统中权值  $W_{rm}$  是负数。由此可见,极限学习机与正负模糊规则系统是等价的,而不可能与正规则模糊系统或负规则模糊系统等价。

相对传统的前馈神经网络的 5 层网络结构,利用单隐层前馈神经网络来实现,输出权值的学习采用极限学习机算法。二者的不同主要体现在网络的结构和参数的学习上:传统的前反馈神经网络结构,网络中所有参数都需要调整,不同参数间往往存在相互依赖性,并且通过梯度下降法来调整权值,而梯度下降法具有局部最优性,当局部最优值与全局最优值相差较远时很难达到最优值,并且在样本数量较大时,对神经网络的训练速度不够理想;极限学习机则不存在上述问题,其激活函数中的高斯函数的中心和宽度矩阵可以随机赋值,通过求 Moore-Penrose 伪逆的方法求得权值矩阵  $\mathbf{W}$ 。该方法不仅有最小的训练误差和最小的权值范数,而且训练速度很快,得到的权值是全局最优的,因此该方法在图像分类中有较好的推广性。由此将极限学习机算法引入图像分类,提出基于极限学习机的正负模糊规则系统。

由式(1)(2)定义的前馈神经网络的输出可知,所提出的模糊系统的激活函数为

$$f(x_k) = \bar{\beta}_r(x_k) = \frac{\beta_r(x_k)}{\sum_{r=1}^R \beta_r(x_k)} = \frac{\exp\left[-\sum_{n=1}^N \frac{(x_{nk} - \mu_{rn})^2}{\sigma_{rn}^2}\right]}{\sum_{r=1}^R \exp\left[-\sum_{n=1}^N \frac{(x_{nk} - \mu_{rn})^2}{\sigma_{rn}^2}\right]} \quad (21)$$

由式(10)知,文中的模糊系统的隐含层对应的输出矩阵为

$$\mathbf{H}(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_R, \sigma_1, \sigma_2, \dots, \sigma_R) = \begin{bmatrix} \bar{\beta}_1(\mathbf{x}_1) & \bar{\beta}_2(\mathbf{x}_1) & \dots & \bar{\beta}_R(\mathbf{x}_1) \\ \bar{\beta}_1(\mathbf{x}_2) & \bar{\beta}_2(\mathbf{x}_2) & \dots & \bar{\beta}_R(\mathbf{x}_2) \\ \vdots & \vdots & & \vdots \\ \bar{\beta}_1(\mathbf{x}_K) & \bar{\beta}_2(\mathbf{x}_K) & \dots & \bar{\beta}_R(\mathbf{x}_K) \end{bmatrix}_{K \times R} \quad (22)$$

通过上述第 3 部分可知文中的模糊系统的输出权值  $\mathbf{W}$  通过求 Moore-Penrose 伪逆得到,用式(17)计算得到,形式如下

$$\mathbf{W} = \mathbf{H}^\dagger \mathbf{Y}_q \quad (23)$$

式中,  $\mathbf{Y}_q = [\mathbf{y}_{q1} \ \mathbf{y}_{q2} \ \dots \ \mathbf{y}_{qK}]^T_{M \times K}$ ,  $\mathbf{y}_{qk}$  是模糊系统的期望输出,其形式如下:

$$\mathbf{y}_{qk} = (y_{qk1}, y_{qk2}, \dots, y_{qkM})^T = \begin{cases} (1, 0, \dots, 0)^T & \mathbf{x}_k \in \text{class } C_1 \\ (0, 1, \dots, 0)^T & \mathbf{x}_k \in \text{class } C_2 \\ \vdots \\ (0, 0, \dots, 1)^T & \mathbf{x}_k \in \text{class } C_M \end{cases} \quad (24)$$

文中的模糊系统由输入层、输出层和隐含层(且隐含层只有一层)组成。

第 1 层(输入层) 在这一层中,每个神经元的输入就是它的输入,  $O_{1n} = x_{nk}, n = (1, 2, \dots, N), k = (1, 2, \dots, K)$ 。

第 2 层(隐含层) 本层中每个神经元代表一条模糊规则,文中采用高斯函数作为隶属度函数,第  $r$  个神经元的输出为

$$O_{2rk} = \bar{\beta}_r(\mathbf{x}_k) = \frac{\beta_r(\mathbf{x}_k)}{\sum_{r=1}^R \beta_r(\mathbf{x}_k)} = \frac{\exp\left[-\sum_{n=1}^N \frac{(x_{nk} - \mu_{rn})^2}{\sigma_{rn}^2}\right]}{\sum_{r=1}^R \exp\left[-\sum_{n=1}^N \frac{(x_{nk} - \mu_{rn})^2}{\sigma_{rn}^2}\right]} \quad (25)$$

第 3 层(输出层) 在该层中,每个神经元决定着神经网络的输出,输出权值  $W_{rm}$  通过求 Moore-Penrose 伪逆的方法得到,每个神经元的输出为

$$y_{km} = \sum_{r=1}^R O_{2rk} W_{rm} = \sum_{r=1}^R \bar{\beta}_r(\mathbf{x}_k) W_{rm} \quad (26)$$

所提出的正负模糊规则系统充分考虑了负规则在图像分类中的作用,相对于传统的基于正规则模糊系统或负规则模糊系统有明显的优越性。不仅如此,文中还将正负模糊规则系统与 ELM 联系起来,并证明了二者实质上是等价的。从而使我们的模糊系统又明显的具有 ELM 的特征,即参数可以随机赋值,学习速度快,泛化性能好。

所提出的模糊系统在图像分类过程中的具体步骤:

1) 参考文献[1]中的样本提取方法,在原图像上选取特征明显、最具有代表性的区域作为训练样本,并对样本进行预处理(包括对训练样本做归一化处理以及输入格式的转化)。

2) 对所提出的模糊系统的输入神经元到隐含层神经元的连接权值——高斯函数的中心矩阵  $\mu$  和宽度矩阵  $\sigma$  赋随机值。

3) 计算出文中模糊系统的隐含层输出矩阵  $\mathbf{H}$ , 并计算出输出权矢量  $\mathbf{W} = \mathbf{H}^+ \mathbf{Y}_q$ 。

4) 测试我们的模糊系统,对原图像进行预处理,对图像的输入格式进行转化,并做归一化处理,由训练好的模糊系统根据式(26)计算出结果,并将分类后的得到的点还原成图像。

## 4 实验结果及分析

利用文中所提方法分别对遥感图像和自然图像分类,并分别与模糊 C 均值聚类(FCM)算法、BP 神经网络算法,以及 SVM 算法的分类结果进行比较。文中的 SVM 算法是利用 C 语言编译的 SVM 工具包 1: LIBSVM<sup>[10]</sup>实现的,其核函数是径向基函数。实验的硬件平台为: AMD Athlon × 2 CPU,其主频为 2.0 GHz,内存为 1GB。在 MATLAB R2009a 环境下编写了图像分类程序。实验结果表明,文中的方法对样本的训练速度很快,并且有较好的图像分类效果,在自然图像和遥感图像的分类上都表现出明显的优越性。

### 4.1 遥感图像分类

基于以上分析,将文中所提方法应用到遥感图像分类中<sup>[11]</sup>,图 2(a)为原始图像,图像大小  $200 \times 200$ ,图(b)中红色部分是训练样本,共计选择 3 800 个样本点,用这些特征对模糊系统进行训练,使其输出误差平方和达到最小。在这个实验中,输入神经元个数为 3 ( $N = 3$ ),规则条数为 40 ( $R = 40$ ),输出神经元个数为 3 ( $M = 3$ )。对于遥感图像中的每个点,将其降维形成 R、G、B 3 个灰度值,并将同一个点的 R、G、B 值作为 ELM 神经网络的 3 个输入。期望输出城市、植被、水体这 3 类,分别表示为  $[1 \ 0 \ 0]$ 、 $[0 \ 1 \ 0]$ 、 $[0 \ 0 \ 1]$ 。在训练结束之后,用训练好的神经网络对图像分类。在该实验中,BP 神经网络采用的是 3-40-3 的结构,包含输入层、输出层和隐含层,输入层的节点数为 3,隐含层的节点数为 40,输出层的节点数为 3,隐含层和输出层分别采用双曲正切 S 型激活函数和线性激活函数。BP 神经网络算法以及 FCM 算法都是调用 MATLAB 中自带的函数,并且辅助参数都是默认值,SVM 算法的辅助参数亦设置为默认值。

为了得到更加合理的实验结果,实验中所有算法都将运行 10 次,实验结果取平均值。文中方法用 0.225 0 s 的 CPU 时间获得了训练误差均方根为 0.002 6,而 BP 神经网络算法在训练样本时

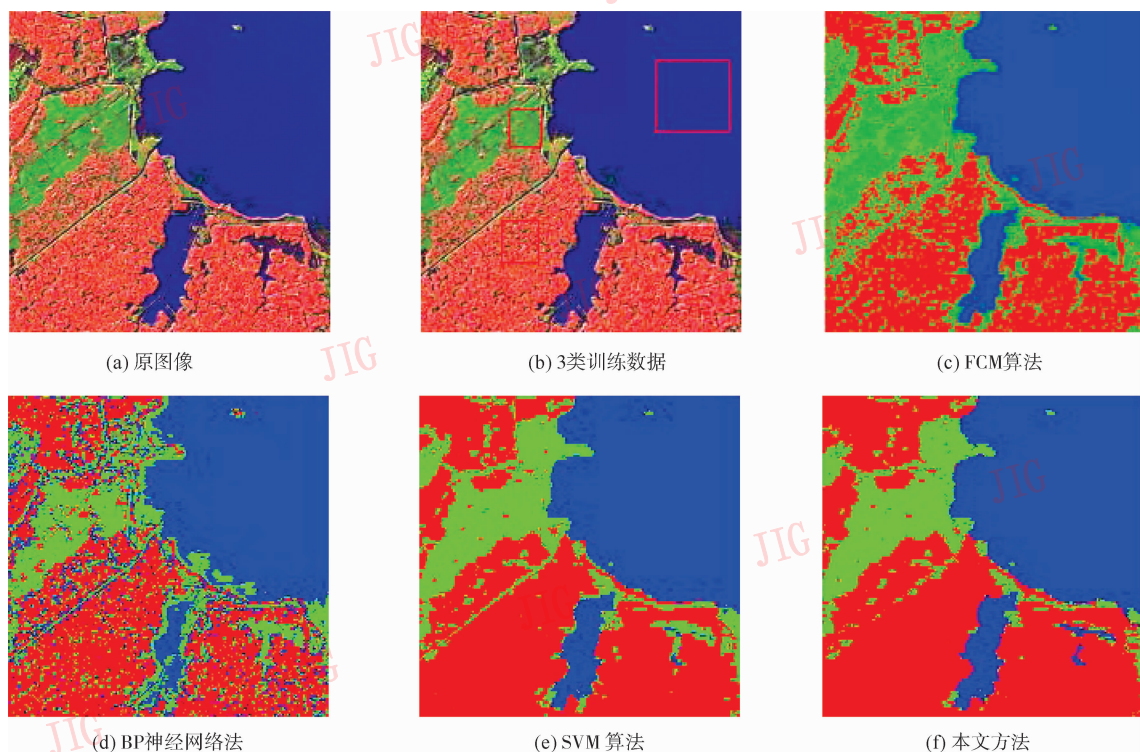


图2 遥感图像及其分类结果

Fig. 2 Classification results for remote sensing image

用了 148.720 3 s 的 CPU 时间达到了训练误差均方根为 0.047 3, SVM 算法训练样本时用了 0.422 0 s 的 CPU 时间达到了训练误差均方根为 1.420 8, 文中方法的学习速度是 SVM 算法的 1.875 6 倍, 是传统 BP 神经网络算法的 660 倍; 文中方法在测试时用时 0.304 7 s CPU 时间, 而 BP 神经网络算法的测试用时为 0.404 7 s CPU 时间, SVM 算法的测试时间为 0.503 1 s CPU 时间。文中方法的测试时间为 BP 神经网络算法的 75.29%, 是 SVM 算法的 61.1%。相对于 FCM 算法而言, FCM 算法的聚类时间为 1.950 0 s CPU 时间, 文中方法的测试时间仅为 FCM 算法聚类时间的 15.63%, 而 FCM 算法的聚类时间是文中方法的训练时间与测试时间之和的 3.681 3 倍。由此可见, 文中方法不论是在训练阶段还是在测试阶段都有很好的表现。

从图 2 中可以看出, 文中所提方法比 FCM 算法的分类结果要好得多, 相对 BP 神经网络的分类图像而言, 对图像两类分界处的处理要好很多, 而对 SVM 算法而言, 文中所提方法受噪声的影响也非常小, 在蓝色区域(水体)的处理效果更出色, 准确率更高。

## 4.2 自然图像分类

在这组实验中, 对自然图像分类<sup>[12]</sup>, 并将分类结果与 BP 神经网络算法、FCM 算法, 以及 SVM 算法比较。实验中所有算法都将运行 10 次, 实验结果取平均值。取测试图片——图 3(a)、图 4(a) 来自伯克利数据库, 图片 5(a) 来自文献[13]。在该部分做 3 个实验, 原始图像如图 3(a)、图 4(a)、图 5(a) 所示, 图像大小分别为  $200 \times 165$ 、 $200 \times 135$ 、 $150 \times 150$ ; 在 3 幅图像的原图上分别加入均值为 0、方差为 0.05 的高斯白噪声。将噪声图像的输入划分成  $5 \times 5$  的滑动邻域窗口, 图像的边缘用边界复制的方法填充。为训练文中的神经网络, 分别将图 3(b)、图 4(b)、图 5(b) 中的红色区域作为训练样本。对图 4、图 5 这两个实验采用 25 个输入 ( $N = 25$ )、100 条规则 ( $R = 100$ ) 和 3 类输出 ( $M = 3$ ) 的文中所述的模糊系统, 而图 5 对应的实验采用 25 个输入 ( $N = 25$ )、60 条规则 ( $R = 60$ ) 和 3 类输出 ( $M = 3$ ) 的模糊系统, 并对模糊系统进行训练。将  $5 \times 5$  的窗口对应的 25 个点输入到已训练好的模糊系统, 判断滑动邻域窗口的中心属于哪一类。

图 3(d)、图 4(d)、图 5(d) 是对应的 BP 神经网的分类结果, 在该实验中, BP 神经网络包含

输入层、输出层和一个隐含层,输入层的节点数为 25,图 3、图 4 对应的实验的隐含层的节点数为 100,图 5 对应的隐含层的节点数为 60,输出的节点数为 3,隐含层和输出层分别采用双曲正

切 S 形激活函数和线性激活函数。FCM 算法以及 BP 神经网络算法是调用 MATLAB 中自带的函数,且辅助参数是默认值;SVM 算法的辅助参数亦设置为默认值。

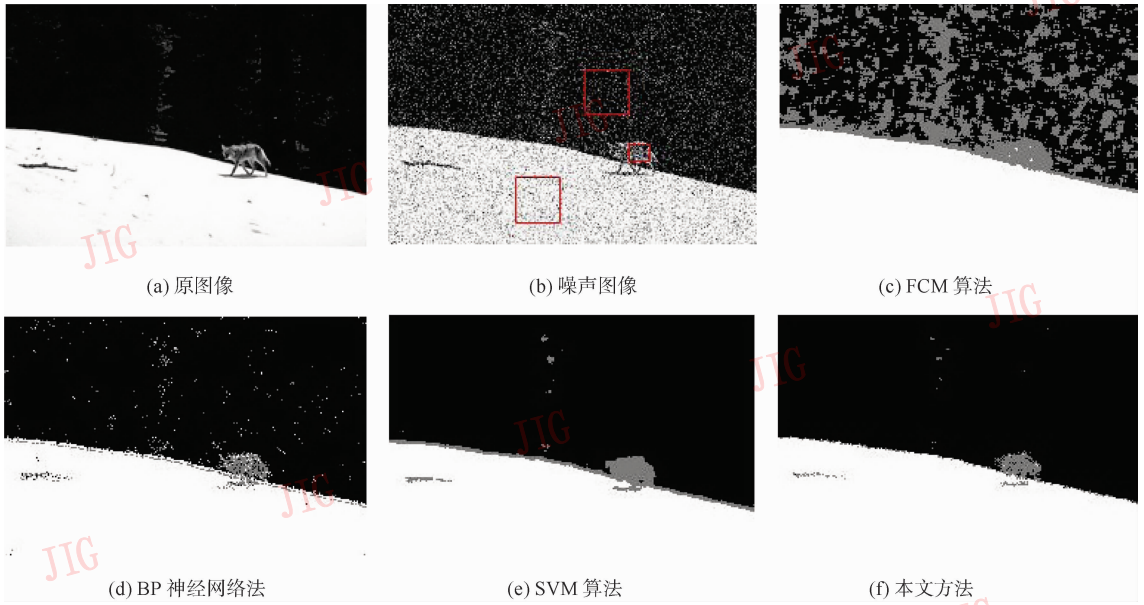


图 3 雪地图像及其分类结果  
Fig. 3 Classification results for snow image

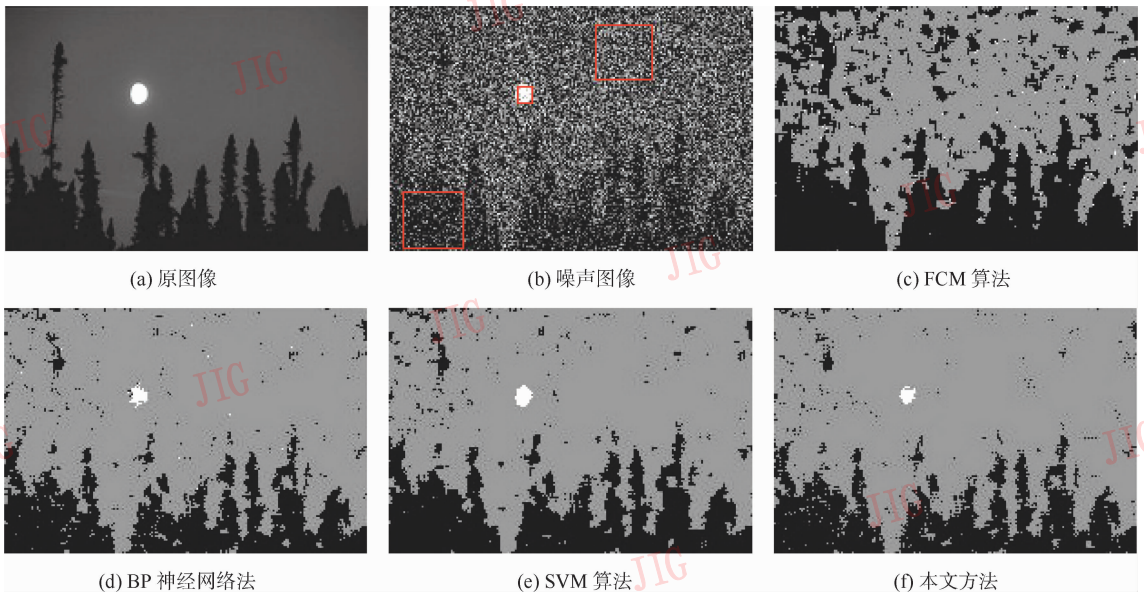


图 4 月色图像及其分类结果  
Fig. 4 Classification results for moonlight image

在图 3 对应的实验中,共计选择 3 300 个点作为训练样本点。期望输出结果为雪、狼和树 3 类,分别表示为  $[1 \ 0 \ 0]$ 、 $[0 \ 1 \ 0]$ 、 $[0 \ 0 \ 1]$ 。在实验中,文中方法的训练用时为

0.614 1 s 的 CPU 时间,得到的最小误差均方根为 0.077 9;BP 神经网络算法的训练用时为 218.386 0 s 的 CPU 时间,获得的最小训练误差为 0.147 1;SVM 算法训练用时为 0.762 5 s 的 CPU

时间,获得的最小训练误差为 2.455 6。在这个实验中,文中方法训练样本的速度是 SVM 算法的 1.241 7 倍,是传统 BP 神经网络模型的 355 倍;文中方法在测试时用时 0.964 1 s CPU 时间,SVM 算法的测试用时为 2.012 5 s CPU 时间,BP 神经网络算法的测试用时为 1.081 3 s CPU 时间。可见,文中方法的测试速度是 SVM 算法的 2.087 4 倍,是 BP 算法的 1.121 6 倍。相对于 FCM 算法而言,FCM 算法的聚类时间为 6.212 5 s CPU 时间。FCM 算法的聚类时间是文中方法的训练时间与测试时间之和的 3.936 4 倍。

在图 4 对应的实验中,选择了 2 700 个训练样本点。期望输出结果为月亮、天空和树 3 类,分别表示为  $[1 \ 0 \ 0]$ 、 $[0 \ 1 \ 0]$ 、 $[0 \ 0 \ 1]$ 。在实验中,文中方法用 0.456 3 s CPU 时间获得了训练误差均方根为 0.177 7;BP 神经网络算法的训练用时为 185.942 2 s 的 CPU 时间,获得的最小训练误差为 0.204 6;SVM 算法训练用时为 1.393 8 s 的 CPU 时间,获得的最小训练误差为 2.111 5。在这个实验中,文中方法训练样本的速度是 SVM 算法的 3.054 6 倍,是传统 BP 神经网络算法的 407 倍;文中方法在测试时用时 0.623 4 s CPU 时间,

SVM 算法的测试用时为 5.546 9 s CPU 时间,而 BP 神经网络算法的测试用时为 0.731 2 s CPU 时间。相对于 FCM 算法而言,FCM 算法的聚类时间为 2.164 1 s CPU 时间,FCM 算法的聚类时间是文中方法的训练时间与测试时间之和的 2.004 4 倍。

在图 5 对应的实验中,一共选择了 1 200 训练样本,期望输出结果为白、灰和黑 3 类,分别表示为  $[1 \ 0 \ 0]$ 、 $[0 \ 1 \ 0]$ 、 $[0 \ 0 \ 1]$ 。在实验中,文中方法的训练用时为 0.262 5 s 的 CPU 时间,得到的最小误差均方根为 0.086 0;SVM 算法的训练用时为 0.131 3 s 的 CPU 时间,获得的最小训练误差为 1.994 6;BP 神经网络算法的训练用时为 96.900 0 s 的 CPU 时间,获得的最小训练误差为 0.1477。在这个实验中,文中方法训练样本的速度是传统 BP 神经网络模型的 369 倍。文中方法在测试时用时 0.551 6 s CPU 时间,而 SVM 算法的测试用时为 0.790 7 s CPU 时间,BP 神经网络算法的测试用时为 0.614 1 s CPU 时间。相对于 FCM 算法而言,FCM 算法的聚类时间为 3.840 6 s CPU 时间,FCM 算法的聚类时间是文中方法的训练时间与测试时间之和的 4.717 6 倍。

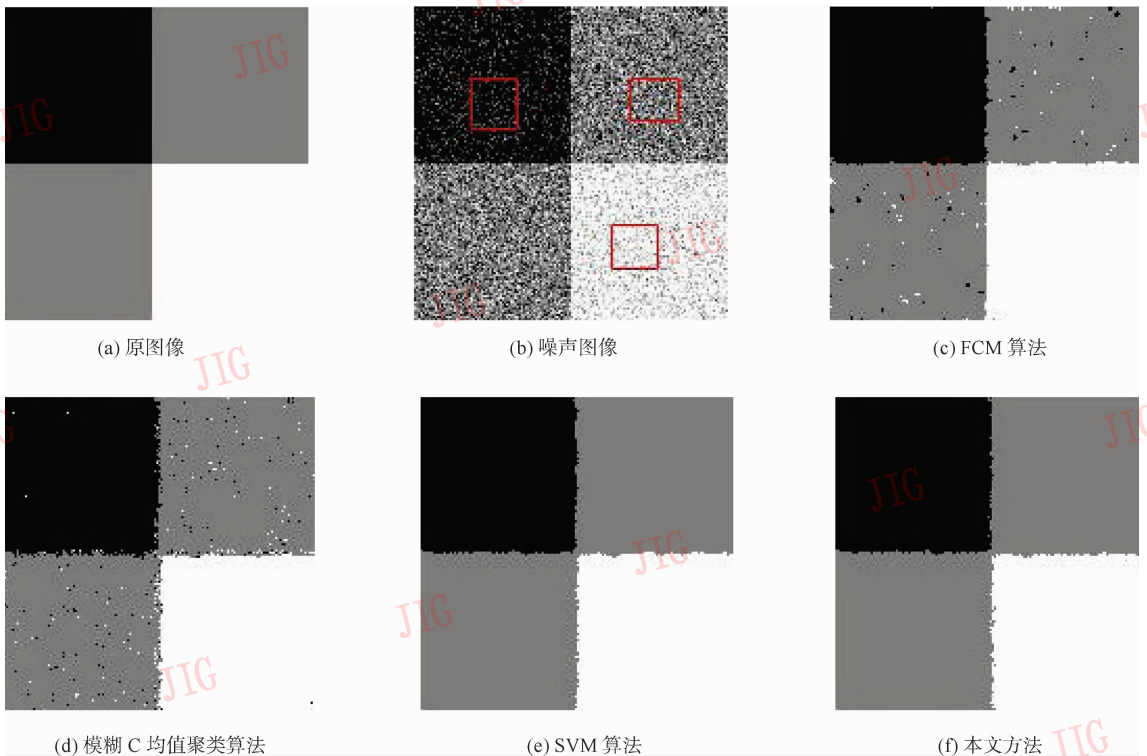


图 5 3 色图像及其分类结果

Fig. 5 Classification results for three-color image

由以上分析可以看出,文中所提的模糊系统训练样本的速度比传统的BP神经网络算法要快很多,很好地解决了神经网络训练和测试速度不理想的缺陷。与FCM算法、BP神经网络算法,以及SVM算法相比较而言,我们提出的方法对图像分界处和图像边缘的处理有明显的优势,同时有较好的抗噪性,不仅能较好的将图像分类,而且可以起到降噪的作用。从以上几个实验可以看出,文中提出的方法是有效的和理想的。

## 5 结论

所述的极限学习机算法,能有效地利用图像的负信息,实验说明该方法在图像分类过程中取得了较好的效果。正负模糊规则相结合形成的正负规则模糊系统的实质等价于极限学习机。这种方法不仅能对遥感图像进行有效的分类,而且对自然图像也有很好的分类效果。极限学习机不仅有最小的训练误差和最小的权值范数,而且训练速度很快。极限学习机算法可以对输入层权值和隐含层偏差赋随机值,涉及的计算量很小,算法简洁高效,得到的输出权值是全局最优的,因此,该方法在图像分类中有较好的推广性。但该方法也存在不足,对于大型的图像,极限学习机方法仍然不是很有效,这主要由于求隐含层对应的输出矩阵的Moore-Penrose伪逆时有很大的计算复杂性。如何克服以上的缺点仍是今后研究工作的重点。

## 参考文献 (References)

- [ 1 ] Nguyen T M, Wu J Q M. A combination of positive and negative fuzzy rules for image classification problem [ C ] // Proceedings of the 2008 Seventh International Conference on Machine Learning and Applications, Washington DC, USA: IEEE Computer Society, 2008: 741-746.
- [ 2 ] Huang Guangbin, Zhu Qinyu, Siew Chee-Kheong. Extreme learning machine: theory and applications [ J ]. Neurocomputing, 2006, 70 ( 1-3 ): 489-501.
- [ 3 ] Mandai D P, Murthy C A, Pal S K. Formulation of a multivalued recognition system [ J ]. IEEE Transactions on Systems, Man and Cybernetics, 1992, 22 ( 4 ): 607-620.
- [ 4 ] Pal S, Manda D P. Linguistic recognition system based on approximate reasoning [ J ]. Information Sciences, 1992, 61 ( 1-2 ): 135-161.
- [ 5 ] Ishibuchi H, Yamamoto T. Rule weight specification in fuzzy rule-based classification systems [ J ]. IEEE Transactions on Fuzzy Systems, 2005, 13 ( 4 ): 428-435.
- [ 6 ] Wu X, Zhang C, Zhang S. Efficient mining both positive and negative association rules [ J ]. ACM Transactions on Information Systems, 2004, 22 ( 3 ): 381-405.
- [ 7 ] Teng W, Hsieh M, Chen M. On the mining of substitution rules for statistically dependent items [ C ] // Proceedings 2002 IEEE International Conference on Data Mining, Washington DC, USA: IEEE Computer Society Press, 2002: 442-449.
- [ 8 ] Savasere A, Omiecinski E, Navathe S. Mining for strong negative associations in a large database of customer transactions [ C ] // Proceedings of the Fourteenth International Conference on Data Engineering. Washington DC, USA: IEEE Computer Society Press, 1998: 494-502.
- [ 9 ] Bartlett P L. For valid generalization, the size of weights is more important than the size of networks [ J ]. Advances in Neural Information Processing System, 1996, 9 ( 1 ): 134-140.
- [ 10 ] Ferrari S, Stengel R F. Smooth function approximation using neural networks [ J ]. IEEE Transactions on Neural Networks, 2005, 16 ( 1 ): 24-38.
- [ 11 ] Zhang Dongbo, Wang Yaonan. Neural network ensemble based on rough sets reduction and its application to remote sensing image classification [ J ]. Journal of Image and Graphics, 2008, 13 ( 3 ): 480-487. [ 张东波, 王耀南. 基于粗糙集约简的神经网络集成及其遥感图像分类应用 [ J ]. 中国图象图形学报, 2008, 13 ( 3 ): 480-487. ]
- [ 12 ] Fu Yan, Wang Yaowei, Wang Weiqiang, et al. Content-based natural image classification and retrieval using SVM [ J ]. Chinese Journal of Computers, 2003, 26 ( 10 ): 1261-1265. [ 付岩, 王耀威, 王伟强, 等. SVM用于基于内容的自然图像分类和检索 [ J ]. 计算机学报, 2003, 26 ( 10 ): 1261-1265. ]
- [ 13 ] Guo Yanhui, Cheng H D. New neutrosophic approach to image segmentation [ J ]. Pattern Recognition, 2009, 42 ( 5 ): 587-595.