

基于粗糙集约简的神经网络集成及其 遥感图像分类应用

张东波^{1),2)} 王耀南²⁾

¹⁾(湖南大学电气与信息工程学院,长沙 410082) ²⁾(湘潭大学信息工程学院,湘潭 411105)

摘要 为降低集成特征选择方法的计算复杂性,提出了一种基于粗糙集约简的神经网络集成分类方法。该方法首先通过结合遗传算法求约简和重采样技术的动态约简技术,获得稳定的、泛化能力较强的属性约简集;然后,基于不同约简设计BP网络作为待集成的基分类器,并依据选择性集成思想,通过一定的搜索策略,找到具有最佳泛化性能的集成网络;最后通过多数投票法实现神经网络集成分类。该方法在某地区 Landsat 7 波段遥感图像的分类实验中得到了验证,由于通过粗糙集约简,过滤掉了大量分类性能欠佳的特征子集,和传统的集成特征选择方法相比,该方法时间开销少,计算复杂性低,具有满意的分类性能。

关键词 粗糙集 约简 神经网络集成 遥感图像分类

中图法分类号: TP391.41 TP753 文献标识码: A 文章编号: 1006-8961(2008)03-0480-08

Neural Network Ensemble Based on Rough Sets Reduction and its Application to Remote Sensing Image Classification

ZHANG Dong-bo^{1),2)}, WANG Yao-nan¹⁾

¹⁾(College of Electrical and Information Engineering, Hunan University, Changsha 410082)

²⁾(Institute of Information Engineering, Xiangtan University, Xiangtan 411105)

Abstract Neural network ensemble based on rough sets reduction is proposed to decrease the computing complexity of conventional feature ensemble selection algorithms. Firstly, a dynamic reduction technology, which integrates genetic algorithm and resample method, is used to get reduct sets that have stable and good generalization ability. Secondly, Multiple BP neural networks based on different reducts are built as base classifiers. According to the idea of selective ensemble, the best generalization ability neural network ensemble can be found by some search strategies. Finally, classification based on neural network ensemble can be completed by combination with vote rule. The method has been verified in the experiment of classifying Landsat 7 bands remote sensing image of chosen area. A number of feature sets of poor performance were discarded by reduction based on rough sets. Compared to conventional feature selection algorithms, the method takes less time, has lower computing complexity, and the performance is satisfying.

Keywords rough sets, reduction, neural network ensemble, remote sensing image classification

1 引言

1990年, Hansen等人提出了神经网络集成(neural network ensemble)方法^[1]。神经网络集成

通过训练多个神经网络并将其结论进行合成,可以显著地提高学习系统的泛化能力,是一种有广阔应用前景的工程化神经计算技术,已成为机器学习和神经计算领域的研究热点^[2]。

当前,对神经网络集成的研究主要集中在如何

基金项目:国家自然科学基金项目(60775047);湖南省自然科学基金项目(06JJ50112)

收稿日期:2006-05-09;改回日期:2006-11-06

第一作者简介:张东波(1973~),男,副教授,2007年获湖南大学控制科学与工程专业博士学位。主要研究方向为粗糙集、神经网络、遥感图像处理 and 模式识别等。E-mail: zhadonbo@yahoo.com.cn

生成集成中的个体网络 and 如何对个体网络的输出进行合成这两个方面。根据 Krogh 等人的理论分析^[3],为降低集成的泛化误差,在保证集成中的个体网络有足够预测精度的前提下,应尽可能使集成的个体网络误差不相关且具有较大的差异度。为生成具有较大差异度的个体网络,有各种各样的方法,如使用不同的目标函数^[4]、网络结构和隐层节点数^[5]、权初始值^[6]训练得到多个神经网络集成个体,或采用交叉验证技术^[3]、遗传算法^[7,8]、负相关学习^[9,10]、聚类技术^[11,12]来生成集成中的个体;其中,个体网络生成应用最广泛、最重要的技术是对样本进行扰动的 Boosting^[13]和 Bagging 算法^[14]。除此之外,基于特征选择的神经网络集成技术(即集成特征选择)也是一种很有效的个体网络生成技术,这在近年来很多文献中^[15-17]已得到验证,通过选择不同的特征子集,有利于提高个体网络的差异性和降低示例空间的误差相关性,从而有利于获得泛化能力强的神经网络集成。

以往文献中提出的集成特征选择,在给出综合预测精度和差异度的适应函数后,通过一定的搜索策略如爬山法^[18]、遗传算法^[16]、顺序前进法和顺序后退法^[19]等进行集成特征选择,由于搜索是在整个特征空间进行的,因此随着特征空间维数的增大,待搜索特征子空间急剧增加,对个体网络进行训练所需时间也激增。为解决这一问题,提出一种基于粗糙集约简的集成特征选择方法,来减少待集成的特征子集数目,通过粗糙集的知识约简能力,可以找到一定数量的约简特征属性作为生成个体网络的候选特征集。根据粗糙集中属性约简的特性,基于粗糙集约简的个体网络既有足够的分类预测精度,还可降低个体网络输入特征维数和结构的复杂性。由于不能保证基于不同约简设计的神经网络是误差不相关的,因此,对不利于集成泛化误差减小的个体网络予以剔除,进行选择集成是非常必要的。

2 基本理论

2.1 粗糙集基本概念

一个信息系统可以用四元组来表示 $S = (U, A, V, f)$, 其中, U 为论域, 表示非空有限对象集; A 为特征属性集合; $V = \bigcup_{a \in A} V_a$, V_a 是属性 a 的值域; 信息函数 $f: U \times A \rightarrow V$ 为每一个对象的属性赋予属性值。 S 可以简记为 $S = (U, A)$ 。

任意的子集 $G \subseteq A$ 定义论域 U 上的一个二元关系 $ind(G)$, 称为不可分辨关系, 根据不可分辨关系 $ind(G)$, 可以对论域 U 进行等价类划分, 记作 U/G , 满足: $U/G = \{X_1, X_2, \dots, X_n\}; X_i \subseteq U, X_i \neq \emptyset, X_i \cap X_j = \emptyset; \bigcup_{i=1}^n X_i = U$; 其中 $i \neq j, i, j = 1, 2, \dots, n$ 。

任意的 $X_i \in U/G$ 称之为基本等价类, 其中包含 x 的等价类记作 $G(x)$ 。若 $a(x)$ 表示对象 x 关于属性 a 的取值, 如果 $x, y \in G(x)$, 则 $\forall a \in G$, 有 $a(x) \equiv a(y)$ 。

集合 X 如果能表达成 U/G 中某些等价类的并时, 则称集合 X 是 G 可定义集或 G 精确集, 否则集合 X 是 G 不可定义集或 G 粗糙集。给定集合 X , $\overline{G}(X)$ 和 $\underline{G}(X)$ 分别称为 X 的上、下近似集:

$$\begin{cases} \overline{G}(X) = \bigcup \{M(x) \mid M(x) \subseteq X, x \in U\} \\ \underline{G}(X) = \bigcup \{M(x) \mid M(x) \cap X \neq \emptyset, x \in U\} \end{cases} \quad (1)$$

$BN_G(X) = \overline{G}(X) - \underline{G}(X)$ 称为 X 的 G 边界域, $POS_G(X) = \underline{G}(X)$ 称为 X 的 G 正域, $NEG_G(X) = U - \overline{G}(X)$ 称为 X 的 G 负域。

当一个信息系统的属性集 A 由条件属性 C 和决策属性 D 组成时, 此时的信息系统也称之为决策系统, 记作 $S = (U, C \cup D)$ 。条件属性 C 和决策属性 D 之间往往存在着某种程度的依赖与关联, 这种依赖性可以用依赖度来定义:

$$\gamma_C(D) = |POS_C(D)| / |U| \quad (2)$$

$|*|$ 表示求集合的基数, $POS_C(D) = \bigcup_{x \in U/D} C(X)$, 称为 D 的 C 正域, $\gamma_C(D)$ 衡量了根据属性 C 描述的知识, 论域对象能够被正确划分到相应决策类的比例。

由于条件属性之间的依赖与关联, 对决策属性来说, 并非所有的条件属性都是必要的, 从而引出属性约简问题, 约简可以理解为在保证分类能力不变的前提下, 最简单地表示决策系统的决策属性对条件属性集的关联和依赖。

如果条件属性 C 的非空子集 C' , 满足条件:

$$(a) \gamma_C(D) = \gamma_{C'}(D);$$

$$(b) \text{不存在 } C'' \subset C', \text{使得 } \gamma_{C''}(D) = \gamma_{C'}(D)。$$

则称 C' 为 C 的一个关于 D 的约简。通常约简不是唯一的, 一个信息系统的多个约简在保证系统分类能力不变的基础上, 从不同输入特征子空间对信息系统进行描述, 从而提供了大量的互补冗余信息。

2.2 神经网络集成泛化误差分析

假设通过神经网络集成学习来模拟函数 $f: \mathbf{R}^m \rightarrow \mathbf{R}^n$, N 个神经网络 $f_i (i = 1, \dots, N)$ 构成集成网

络,各网络分别赋予权值 $\omega_i (i = 1, \cdots, N), \omega_i \geq 0$ 且 $\sum_{i=1}^N \omega_i = 1$ 。神经网络集成的输出为参与集成的各神经网络的加权平均,为简便起见,假设每一个网络只有一个输出分量(即 $n = 1$),但相关的结论很容易推广到多变量输出情况。

设 $x \in \mathbf{R}^m$ 满足概率分布 $p(x)$,在输入 x 时,理想输出为 $d(x)$,第 i 个成员网络 f_i 的实际输出为 $f_i(x)$,则

神经网络集成在输入 x 下的输出

$$f(x) = \sum_{i=1}^N \omega_i f_i(x) \quad (3)$$

神经网络集成的泛化误差:

$$E = \int p(x) (\bar{f}(x) - d(x))^2 dx \quad (4)$$

第 i 个成员网络 f_i 的泛化误差:

$$E_i = \int p(x) (f_i(x) - d(x))^2 dx \quad (5)$$

神经网络 f_i 与神经网络集成 $\bar{f}$ 的差异度:

$$A_i = \int p(x) (f_i(x) - \bar{f}(x))^2 dx \quad (6)$$

根据 Krough 等人^[3]的理论分析,有

$$E = \bar{E} - \bar{A} \quad (7)$$

$\bar{E} = \sum_{i=1}^N \omega_i E_i$ 是各神经网络泛化误差的加权平均值,
 $\bar{A} = \sum_{i=1}^N \omega_i A_i$ 是各神经网络与神经网络集成的差异度的加权平均值。

从式(7)可知,要降低神经网络集成的泛化误差,一方面要尽可能减少各成员网络的泛化误差,另一方面要提高集成中各成员网络的差异度。差异度 $\bar{A}$ 度量了神经网络集成中各网络的相关程度,若集成是高度偏向的,即对于相同的输入,集成中所有网络都给出相同或相近的输出,此时,集成的差异度接近于零,其泛化误差接近于各网络泛化误差的加权平均,反之,若集成中各网络是相互独立的,则集成的差异度较大,其泛化误差将远小于各网络泛化误差的加权平均值,因此,要增强神经网络集成的泛化能力,就应尽可能使集成中各成员网络的误差互不相关。

3 基于粗糙集约简神经网络集成原理

3.1 基于粗糙集约简的神经网络集成模型

本文在基于神经网络集成理论基础上,利用粗

糙集约简构造用于集成的个体网络,充分利用约简所提供的互补冗余信息来提高多神经网络集成分类系统的泛化性能。根据粗糙集约简的特性,一方面,集成中的个体网络的预测精度有足够的保证;另一方面,由于不同的约简,对应不同的特征子空间,因而有利于提高个体网络的差异性。图 1 是基于粗糙集约简的神经网络集成模型。图中 NN1、NN2...NNn 分别指代第 1 个,第 2 个,第 n 个个体网络。

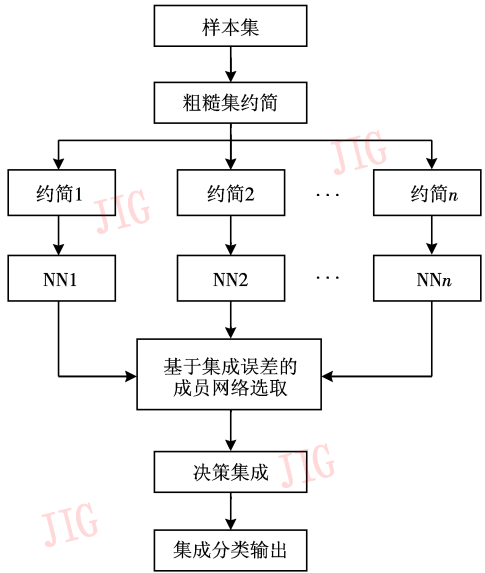


图 1 基于粗糙集约简的神经网络集成模型

Fig. 1 Model of neural network ensemble based on rough set reduction

3.2 粗糙集约简的生成

粗糙集约简的求取是一个 NP(nondeterministic polynomial)完全问题,在特征维数较高及样本数较多时,计算量非常大,本文采用遗传算法来求取约简,对其原理做了简要的介绍,更详细的阐述可参看文献[20]。

3.2.1 遗传算法求约简

给定信息系统 $S = (U, A)$,为表示对象 x 和论域中其他对象的可区分性,可以定义可区分布尔逻辑函数:

$$f_A(x) = \prod_{y \in U} \left\{ \sum \hat{a} \mid a(x) \neq a(y) \right\} \quad (8)$$

对整个信息系统来说,论域中所有对象的可区分性函数可以定义为

$$g_A(U) = \prod_{x \in U} f_A(x) \quad (9)$$

在定义了可区分性函数后,信息系统的属性约简问题可以转换为可区分函数的质蕴涵求取问题,

而质蕴涵很容易转化为多重集形式。

举例:

$$M((\hat{a} + \hat{b}) \cdot (\hat{a} + \hat{b}) \cdot (\hat{c})) = \{(a, b), (a, b), (c)\}$$

M 表示多重集构造算子。最后,约简问题即可转化为多重集的最小命中集问题,即

$$B \in \text{RED}(A) \Leftrightarrow B \in \text{MHS}(M(g_A(U))) \quad (10)$$

MHS 表示求最小命中集。

用遗传算法求属性约简的原理如下:

(1) 染色体编码 每一个属性按照在 U 中的原始位置编成 0 或 1 的染色体串。如果染色体中基因的编码为 0, 表示该位置的属性被约简, 反之, 表示对应位置上的属性没有被约简。

(2) 适应度函数 在求属性约简时, 一方面希望约简的长度短, 另一方面希望该约简有很高的命中系数, 因此, 定义适应度函数:

$$f(B) = (1 - \alpha) \frac{|U| - |A|}{|U|} + \alpha \frac{|\{MS_i \in MS \mid MS_i \cap B \neq \emptyset\}|}{|MS|} \quad (11)$$

该式前一项用来衡量约简的长度, 后一项用来衡量约简的命中系数, α 是综合约简长度和命中系数的参数。 $MS = \{MS_i\}$ 是式 (9) 的质蕴涵对应的多重集。

(3) 选择操作 采用基于精英策略的适应度比例选择, 对于给定规模为 n 的群体 $P = \{b_1, b_2, \dots, b_n\}$, 个体 $b_j \in P$ 的适应值为 $f(b_j)$, 其选择概率为

$$p(b_j) = \frac{f(b_j)}{\sum_{k=1}^n f(b_k)}, j = 1, 2, \dots, n$$

若当前代最佳个体适应值小于上一代群体最大适应值, 则保留上一代的最佳个体, 并将其替代当前代中的最差个体。

(4) 交叉操作 将选择后的新一代种群中的个体放入交配池, 随机配对后形成父代, 依据给定的交叉概率 p_c , 将一对父代个体的基因链随机地在同一位置切断, 部分交换重组后产生新一代个体。

(5) 变异操作 变异是将选中的基因取反从而生成新个体的过程。根据事先给定的变异概率 p_m , 选择要产生变异的染色体, 随机选择待变异的基因进行取反, 得到新的个体。

在遗传算法进化的过程中, 如果连续几代种群的平均适应度的差值都不再有所改善, 即认为算法已收敛, 已经得到了一定程度的属性约简, 结束遗传算法。

在本文实验中, 各参数的取值情况: $\alpha = 0.4$, 种群规模 = 70, 交叉概率 $p_c = 0.3$, 变异概率 $p_m = 0.05$, 连续 10 代种群的平均适应度的差值都不再有所改善, 则算法终止。

3.2.2 结合遗传算法求约简和重采样技术的动态约简算法

由于粗糙集约简对样本比较敏感, 鲁棒性不强, 在实际模式分类器设计过程中, 通常将样本划分为训练样本和测试样本, 并根据训练本来进行分类器设计, 在随机抽取训练样本集时, 如果重复多次, 则随着每次随机抽取训练样本集的不同, 计算得到的粗糙集约简也有一定的差异, 为找到比较稳定的, 泛化能力较强的约简集合, 采用了动态约简技术^[21], 原理如下:

(1) 对样本集按一定比例, n 次随机采样抽取样本构成子系统集 $S = \{S_1, S_2, \dots, S_n\}$, 每一个子系统 $S_i = (U_i, A)$, $U_i \subseteq U$, A 是待约简的特征属性集;

(2) 对每一个子系统 $S_i \in S, i = 1, 2, \dots, n$, 根据遗传算法计算其属性约简 $\text{RED}_{S_i}(A)$;

(3) 统计所有在第 (2) 步得到的约简出现频率。

上述过程中, 某个约简出现的频率越高, 则该约简越稳定, 泛化能力也越强, 所有属于 $\text{RED}_{S_i}(A), i = 1, 2, \dots, n$ 的约简 B , 如果出现的频率超过给定阈值 T_f , 则作为最终的约简输出:

$$\text{DRED}(A, T_f, S) = \left\{ B \subseteq A \mid \frac{|\{S_i \in S \mid B \in \text{RED}_{S_i}(A), i = 1, \dots, n\}|}{|S|} \geq T_f \right\} \quad (12)$$

3.3 成员网络选取

尽管通过上述约简生成方法, 能够得到大量的约简, 但并不能保证基于各约简设计的神经网络是误差不相关的, 彼此相似的个体网络不一定能促使集成泛化误差降低, 甚至可能起相反的作用^[22], 而造成基于全部约简的神经网络集成不一定具有最好的集成泛化能力, 关于这一点, Zhou^[23]、吴建鑫^[24]等人的文献中有具体的分析和论述, 并由此引出了选择性神经网络集成的思想。

由于本文的遥感图像分类是一个模式分类问题, 在模式分类问题中, 神经网络集成的泛化误差可做如下分析, 同样为简便起见, 相关阐述都是基于二类别分类, 但相关结论可以推广到多类别问题。

假设通过 N 个神经网络 $f_i (i = 1, \dots, N)$ 集成来模拟函数 $f: \mathbf{R}^m \rightarrow \mathcal{L}, \mathcal{L} = \{-1, +1\}$ 是类别集, 神经网

络集成的输出采用相对多数投票法。现有 m 个示例样本,期望输出为 $[d_1, d_2, \dots, d_m]^T$, $d_j \in \{-1, +1\}$ ($j=1, 2, \dots, m$) 是第 j 个样本的期望输出,第 i 个个体网络在这些样本上的实际输出为 $[f_{i,1}, f_{i,2}, \dots, f_{i,m}]^T$, $f_{i,j} \in \{-1, +1\}$ ($i=1, 2, \dots, N$; $j=1, 2, \dots, m$) 是第 j 个样本在第 i 个个体网络的实际输出。显然,如果第 i 个个体网络输出和期望一致,则 $f_{i,j}d_j = 1$, 否则 $f_{i,j}d_j = -1$ 。个体网络 i 在 m 个样本上的泛化误差为

$$E_i = \frac{1}{m} \sum_{j=1}^m \text{error}(f_{i,j}d_j) \quad (13)$$

函数 $\text{error}(x)$ 定义如下:

$$\text{error}(x) = \begin{cases} 1 & x = -1 \\ 0.5 & x = 0 \\ 0 & x = 1 \end{cases} \quad (14)$$

根据投票法, N 个网络的集成在第 j 个样本的投票统计为

$$\text{sum}_j = \sum_{i=1}^N f_{i,j} \quad (15)$$

则第 j 个样本的集成输出为

$$\bar{f}_j = \text{sgn}(\text{sum}_j) \quad (16)$$

函数 $\text{sgn}(x)$ 定义如下:

$$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases} \quad (17)$$

显然, $\bar{f}_j \in \{-1, 0, +1\}$, ($j=1, 2, \dots, m$), 对于第 j 个样本, 如果实际集成输出和期望输出一致, 则 $\bar{f}_j d_j = 1$, 如果实际集成输出和期望输出不一致, 则有 $\bar{f}_j d_j = -1$, 如果在计算 $\bar{f}_j$ 时, 个体网络输出 $+1$ 和 -1 的次数相同, 则 $\bar{f}_j d_j = 0$ 。集成的泛化误差为

$$E = \frac{1}{m} \sum_{j=1}^m \text{error}(\bar{f}_j d_j) \quad (18)$$

如果从集成中去掉个体网络 k , 此时其余的网络在第 j 个样本的集成输出为

$$\bar{f}_j^{-k} = \text{sgn}(\text{sum}_j - f_{k,j}) \quad (19)$$

去掉个体网络 k 后新的集成泛化误差为

$$E^{-k} = \frac{1}{m} \sum_{j=1}^m \text{error}(\bar{f}_j^{-k} d_j) \quad (20)$$

从式(18)和式(19)可知, 只要去掉满足式(21)的个体网络 k , 此时 $E^{-k} \leq E$, 集成泛化误差不会增大, 反而可能减小。

$$\sum_{j=1}^m \{ \text{error}(\text{sgn}(\text{sum}_j) d_j) - \text{error}(\text{sgn}(\text{sum}_j - f_{k,j}) d_j) \} \geq 0 \quad (21)$$

上述分析表明, 在独立训练若干个体网络后, 可先进行选择操作, 剔除部分不利于集成泛化误差减小的个体网络, 剩余网络的集成具有比全体网络集成更好的泛化能力。

尽管从理论上讲, 两类别问题, 可以根据式(21)实现神经网络集成中成员的选择, 但对于多类别问题, 则情况要复杂的多, 在实际应用中, 可以通过独立的校验集根据式(18)来估算集成的泛化误差, 并根据一定的搜索策略找到具有最小泛化误差的最佳集成网络。

假设基于粗糙集约简, 已经设计了 N 个神经网络分类器, 则从中找到合适的成员网络集成的搜索策略很多, 如穷举法, 可以计算 N 个分类器所有可能组合(除空集外, 总共有 $2^N - 1$ 种可能)的集成泛化误差, 选择泛化误差最小的组合。当 N 不大时, 该方法很容易找到具有最好泛化性能的神经网络集成, 但是如果 N 较大(如 $N > 30$)则穷举法的计算量太大, 此时, 具有全局最优搜索性能的遗传算法是一种较好的选择。

3.4 基于多数投票法的集成策略

假设神经网络集成要解决的是一个 n 类的模式分类问题, 集成的规模为 N , 各模式类分别记作 $c_1, c_2, \dots, c_n$, 类别 i 的输出编码为 $[0, \dots, 0, 1, 0, \dots, 0]$, 其中, 除第 i 个元素为 1 外, 其余元素全为 0。通过期望的输入输出编码映射关系对 N 个成员网络进行训练, 训练后, 对于测试模式 P , 每一个成员网络都会给出一个输出向量, 其中, 第 k 个成员网络的输出为列向量 $[a_{1,k}, a_{2,k}, \dots, a_{n,k}]^T$, $k=1, \dots, N$, 如果 $a_{i,k} = \max_j(a_{j,k})$, 则 $a_{i,k} = 1$ 且 $a_{j,k} = 0, j \neq i$ 。根据投票法, N 个网络的集成输出为

$$P_{\text{ensm_out}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_{1,1} + a_{1,2} + \dots + a_{1,N} \\ a_{2,1} + a_{2,2} + \dots + a_{2,N} \\ \vdots \\ a_{n,1} + a_{n,2} + \dots + a_{n,N} \end{bmatrix} \quad (22)$$

模式 P 被划分到第 j 类(c_j), 如果 $y_j = \max_i(y_i)$ 。

4 实验结果及分析

实验中, 选用某地区 1 幅 1024×1026 的 Landsat TM 7 波段的遥感图像进行实验, 通过目视解译, 该地区地类主要分为 5 类, 植被(d_1), 河流(d_2), 高密度建成区(d_3), 低密度建成区(d_4), 空地(d_5)。总

共抽取了7217个样本,其中 30% 的样本作为训练集(2 165个),20% 的样本作为校验集(1 443个),其余 50% 的样本作为测试集(3 609个)。各类样本的期望输出编码为 $d_1(1,0,0,0,0)$, $d_2(0,1,0,0,0)$, $d_3(0,0,1,0,0)$, $d_4(0,0,0,1,0)$, $d_5(0,0,0,0,1)$ 。

由于各波段的灰度值已经是 0~255 的离散值,因此,不再作离散化处理,可以直接用来构成决策表。为获得比较稳定的和泛化能力较强的粗糙集约简,在采用动态约简技术求取粗糙集约简时,每次从训练样本集中抽取 30% 的样本,共重复 40 次,其中总共出现的约简数 56 个,而出现 15 次以上的只有 12 个,设频率阈值 $T_f = 15/40 = 0.375$,则选择出现频率 15 次以上的 12 个约简作为最终的粗糙约简集,如表 1 所示。

表 1 出现频率 15 次以上的粗糙集约简
Tab.1 Rough set reductions which appears
morn than fifteen

编号	约简波段	编号	约简波段
1	B4,B6,B7	7	B3,B5,B6,B7
2	B3,B4,B6	8	B2,B4,B5,B6
3	B4,B5,B6	9	B1,B3,B4,B5
4	B1,B3,B5,B6	10	B1,B2,B5,B7
5	B1,B5,B6,B7	11	B1,B2,B4,B5
6	B2,B5,B6,B7	12	B1,B2,B3,B6,B7

基于上述候选的 12 个约简,可以构造 12 个神经网络作为基分类器,神经网络选用单隐层的 BP 神经网络,隐层节点数均取为 15,输出神经元为 5,对应 5 类地物,输入神经元由约简中所含的波段数

决定,BP 网络由反向传播算法训练后作为待集成的基分类器。

根据选择性集成的思想,还需要从这 12 个个体网络中选出合适的成员网络,使得集成的泛化误差最小。由于总共的组合数为 $2^{12} - 1 = 4095$,搜索的空间不是很大,可采用简单的穷举法来实现。集成的泛化误差由抽取出来的校验集进行估算。经过搜索,具有最小泛化误差的网络集成由表 1 中编号为 2、4、9、11、12 的 5 个约简对应的个体网络组成,得到一个最佳集成网络。

表 2 给出了各个体网络在测试集上的分类精度,其中,个体网络 8 具有最佳分类精度 97.41%,个体网络 12 分类精度最差为 95.05%,平均分类精度为 96.62%。整体上来说,各个体网络都具有较高的分类精度,正是因为各个体网络是基于粗糙集约简基础上设计的,所以分类性能得到了保证。表 3 对神经网络集成的性能进行了比较,由本文方法得到的最佳集成网络分类精度达到 97.82%,比最佳个体网络、全部个体网络集成及不进行属性约简,而由全部波段信息设计的神经网络的分类精度都要高。网络集成的分类精度(全部个体网络集成、最佳集成网络)都要高于平均分类精度及任何单个网络的分类精度,最佳集成网络分类精度比平均精度提高 1.2 个百分点。表 4 给出了分别对前 1~12 个个体网络进行集成时的分类精度,从表中可知,集成的分类精度并不是简单的随集成规模的增加而提高,前 5 个神经网络的集成精度 97.74% 甚至高于全部 12 个神经网络集成精度 97.54%,实验结果表明,进行神经网络选择性集成是非常必要的。

表 2 各个体网络对测试样本的分类精度
Tab.2 Classification precision of individual networks for test samples

	编 号											
	1	2	3	4	5	6	7	8	9	10	11	12
分类精度(%)	97.27	97.07	96.87	96.69	96.69	96.85	96.97	97.41	96.64	95.43	96.52	95.05
平均精度(%)	96.62											

注:加粗的第 8、12 个体网络分别具有最高和最低分类精度

表 3 神经网络集成的分类精度比较
Tab.3 The classification performance comparation of neural network ensemble

	最佳个体网络	全部个体网络集成	全部波段构成网络	最佳集成网络	传统集成特征选择方法
分类精度(%)	97.41	97.54	97.35	97.82	97.91

表 4 前 1 ~ 12 个体网络集成的分类精度

Tab.4 The classification precision of anterior 1 ~ 12th individual neural network ensemble

	集成规模											
	1	2	3	4	5	6	7	8	9	10	11	12
分类精度 (%)	97.27	97.21	97.47	97.47	97.74	97.57	97.35	97.49	97.59	97.49	97.66	97.54

显然,基于粗糙集约简的神经网络集成和以往的集成特征选择方法相比,搜索空间变小、花费的时间减少、复杂性得到降低。若采用以往基于整个特征空间的集成特征选择,由 7 个波段构成的待搜索的特征子空间可能的组合数为 $2^7 - 1 = 127$ 。相应需设计和训练 127 个分立的神经网络基分类器,在进行选择性集成时,可能的集成组合有 $2^{127} - 1$,由于搜索空间太大,穷举法不再适用,可采用遗传算法

找到的最好集成分类精度为 97.91% (表 3),略好于本文方法,但从时间开销和计算复杂性来说,并不合算,而粗糙集约简过滤了大量分类性能欠佳的特征子集,在更少的时间开销和计算复杂性的基础上得到了令人满意的效果。是一种有效的进行神经网络集成的集成特征选择方法。

图 2 是本文方法得到的分类结果图,其他情况下的分类图限于篇幅,不再赘述。

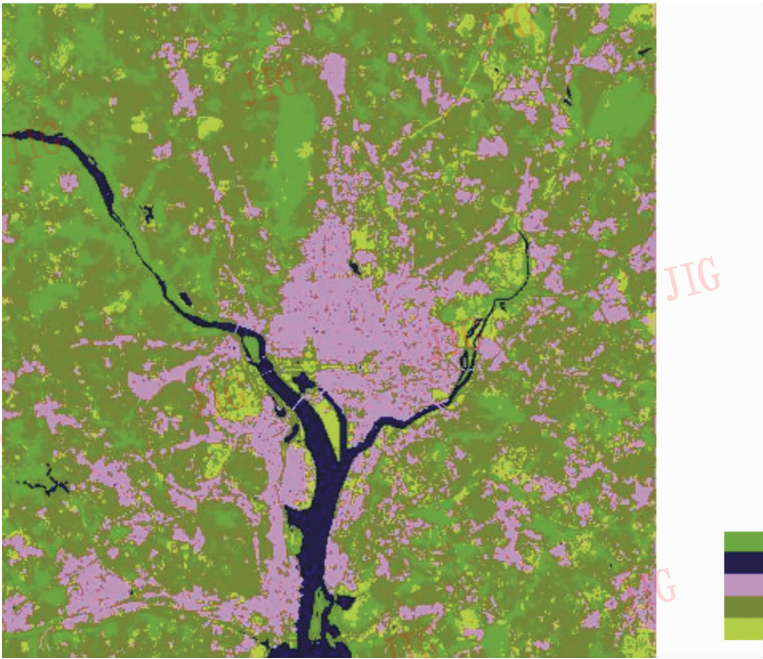


图 2 本文方法的分类图

Fig.2 Classification image of method in this paper

5 结 论

集成特征选择方法是近年来研究较多的一种有效的神经网络集成的个体网络生成方法,但传统方法存在搜索空间大、时间开销大、计算复杂性高的缺陷,本文通过粗糙集约简,在保证个体网络的分类性能基础上,过滤掉大量分类性能欠佳的特征子集,从而极大的降低了时间开销和复杂性。尽管如此,本方法也有不足和值得进一步研究改进的地方:

- (1) 如果系统特征属性过少,则能获得的差异性较大的约简数会很少,此时采用本文方法难以发挥作用;
- (2) 可以考虑将扰动特征属性的粗糙集约简技术和比较成熟的对样本进行扰动的 Boosting 和 Bagging 方法结合起来,以生成精确且差异度大的个体网络,以便进一步提高集成的泛化能力。

参考文献 (References)

1. Hansen L K, Salamon P. Neural network ensembles [J]. IEEE

- Transactions on Pattern Analysis and Machine Intelligence, 1990, **12**(10): 993 ~ 1001.
- 2 Zhou Zhi-hua, Chen Shi-fu. Neural network ensemble[J]. Chinese Journal of Computers, 2002, **25**(1): 1 ~ 8. [周志华, 陈世福. 神经网络集成[J]. 计算机学报, 2002, **25**(1): 1 ~ 8.]
 - 3 Krogh A, Vedelsby J. Neural network ensembles, cross validation and active learning[A]. In: Tesauro G, Touretzky D, Leen T eds, Advances in Neural Information Processing Systems 7 [M], Cambridge, MA, USA: MIT Press, 1995: 231 ~ 238.
 - 4 Hampshire J, Waibel A. A novel objective function for improved phoneme recognition using time delay neural networks[J]. IEEE Transactions on Neural Networks, 1990, **1**(2): 216 ~ 228.
 - 5 Cherkauer K J. Human expert level performance on a scientific image analysis task by a system using combined artificial neural networks [A]. In: Proceedings of the 13th AAAI Work shop on Integrating Multiple Learned Models for Improving and Scaling Machine Learning Algorithms[C], Portland, OR, USA, 1996: 15 ~ 21.
 - 6 Maclin R, Shavlik J W. Combining the predictions of multiple classifiers: using competitive learning to initialize neural networks [A]. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence[C], Montreal, Canada, 1995: 524 ~ 530.
 - 7 Yao X, Liu Y. Making use of population information in evolutionary artificial neural networks[J]. IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics, 1998, **28**(3): 417 ~ 425.
 - 8 Zhou Z H, Wu J X, Jiang Y, *et al.* Genetic algorithm based selective neural network ensemble [A]. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence[C], Seattle, WA, USA, 2001, **2**: 797 ~ 802.
 - 9 Zeke S, Chan H, Kasabov N. Fast neural network ensemble learning via negative-correlation data correction [J]. IEEE Transactions on Neural Networks, 2005, **16**(6): 1707 ~ 1710.
 - 10 Fu Xiang-hua, Feng Bo-qin, Ma Zhao-feng, *et al.* Method of incremental construction of heterogenous neural network ensemble with negative correlation [J]. Journal of Xi'an Jiaotong University, 2004, **38**(8): 796 ~ 799. [傅向华, 冯博琴, 马兆丰等. 增量构造负相关异构神经网络集成的方法[J]. 西安交通大学学报, 2004, **38**(8): 796 ~ 799.]
 - 11 Li Kai, Huang Hou-kuan. A selective approach to neural network ensemble based on clustering technology[J]. Journal of Computer Research & Development, 2005, **42**(4): 594 ~ 598. [李凯, 黄厚宽. 一种基于聚类技术的选择性神经网络集成方法[J]. 计算机研究与发展, 2005, **42**(4): 594 ~ 598.]
 - 12 Fu Qiang, Hu Shang-xu, Zhao Sheng-ying. Clustering-based selective neural network ensemble [J]. Journal of Zhejiang University, 2005, **6A**(5): 387 ~ 392.
 - 13 Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences, 1997, **55**(1): 119 ~ 139.
 - 14 Breiman L. Bagging predictors [J]. Machine Learning, 1996, **24**(2): 123 ~ 140.
 - 15 Ho T K. The random subspace method for constructing decision forests[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, **20**(8): 832 ~ 844.
 - 16 Opitz D. Feature selection for ensembles [A]. In: Proceedings of 16th National Conference on Artificial Intelligence[C], AAAI Press, 1999: 379 ~ 384.
 - 17 Ling Jin-jiang, Chen Zhao-qian, Zhou Zhi-hua. Feature selection based neural network ensemble method[J]. Journal of Fudan University(Natural Science), 2004, **43**(5): 685 ~ 688. [凌锦江, 陈兆乾, 周志华. 基于特征选择的神经网络集成方法[J]. 复旦学报(自然科学版), 2004, **43**(5): 685 ~ 688.]
 - 18 Cunningham P, Carney J. Diversity versus quality in classification ensembles based on feature selection[A]. In: DeMántaras R L, Plaza E. (eds.), Proceedings of 11th European Conference on Machine Learning[C], Barcelona, Spain, Springer, 2000: 109 ~ 116.
 - 19 Tsymbal A, Pechenizkiy M, Cunningham P. Diversity in ensemble feature selection [R]. Technical Report, TCD-CS-2003-44 Trinity College Dublin, 2003: 1 ~ 38.
 - 20 Vinterbo S, Ohrn A. Minimal approximate hitting sets and rule templates[J]. International Journal of Approximate Reasoning, 2000, **25**(2): 123 ~ 143.
 - 21 Bazan J, Skowron A, Synak P. Dynamic reducts as a tool for extracting laws from decision tables[A]. In: Proceedings of Symposium on Methodologies for Intelligent Systems[C], Charlotte, NC, Springer-Verlag, 1994: 346 ~ 355.
 - 22 Liu Y, Yao X. Simultaneous training of negatively correlated neural networks in an ensemble[J]. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 1999, **29**(6): 716 ~ 725.
 - 23 Zhou Z H, Wu J X, Tang W. Ensembling neural networks: many could be better than all[J]. Artificial Intelligence, 2002, **137**(1, 2): 239 ~ 263.
 - 24 Wu Jian-xin, Zhou Zhi-hua, Shen Xue-hua, *et al.* A selective constructing approach to neural network ensemble[J]. Journal of Computer Research & Development, 2000, **37**(9): 1039 ~ 1044. [吴建鑫, 周志华, 沈学华等. 一种选择性神经网络集成构造方法[J]. 计算机研究与发展, 2000, **37**(9): 1039 ~ 1044.]