Ensemble remote sensing classifier based on rough set feature partition

PAN Xin[1,2,3]**, ZHANG Shu-qing**[1]**, LI Xiao-feng**[1]**, NA Xiao-dong**[1,2]**, YU Huan**[1,2]

1. *Northeast Institute of Geography and Agricultural Ecology, Chinese Academy of Sciences*, *Jilin Changchun* 130012, *China*;
2. *Graduate University of Chinese Academy of Sciences*, *Beijing* 100039, *China*;
3. *School of Electrical & Information Technology, Changchun Institute of Technology*, *Jilin Changchun* 130012, *China*

Abstract: Supervised classification in remote sensing imagery is receiving increasing attention in current research. In order to improve the classification ability, a lot of spatial-features (e.g., texture information generated by GLCM) have been utilized. Unfortunately, too many features often cause classifier over-fit to a certain features' character and lead to lower classification accuracy. The traditional feature selection algorithms have an unstable classification performance which depends on the number of training samples. This study presents a rough set based ensemble remote sensing image classifier (briefly denoted as RSEC). It partitions feature set into a lot of reducts, and constructs training subset by utilizing these reducts. Each training subset trains an artificial neural network (ANN) classifier; the decisions from all the base classifiers are combined with a voting strategy. This approach can reduce input features to a single classifier, and it can avoid bias caused by feature selection. The RSEC classifier has been compared with the direct ANN method and the traditional feature selection method. It can be seen from the result that RSEC has better classification accuracy and more stable than the others.

Key words: ensemble classifier, rough sets, artificial neural network, feature selection

CLC number: TP751.1 **Document code:** A

1 INTRODUCTION

Land use/cover information has been identified as one of the crucial data components for many aspects of global change studies and environmental applications. The development of remote sensing technology has increasingly facilitated the acquisition of such information (Ouyang & Ma, 2006). How to extract accurate and timely knowledge about land use/cover from remote sensing imagery relies upon not only the data quality and resolution, but also the classification techniques used. Therefore, improvement of remote sensing classification accuracy is always a concern. Many data mining technologies e.g. Per-pixel based maximum likelihood, fuzzy classifications, object-oriented multi-resolution segmentation, artificial neural networks, decision tree-based classification and rule-based classification, have been used in supervised or unsupervised remote sensing classification (Leung *et al*., 2007).

Additional spatial-features (e.g., texture information generated by GLCM) derived from spectrum can provide more information for classifier to improve classification accuracy (Shaban & Dikshit, 2001). Unfortunately, not all the spatial features obtained by diverse parameter or method are helpful for the classification (Aguera, 2008). Huge amounts of irrelevant features may result in over-fitting and maybe lead to the classifier's poor performance (Lei *et al*., 2007).

In order to solve the problem, three methods are utilized at present: (1) Principal Component Analysis (PCA), which involves an orthogonal linear transformation procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables which have greater variance. The drawback of PCA is that many non-influenced features may be composed into dimensions and have higher proportion, which can decrease classification accuracy (Chou *et al*., 2006); meanwhile, with the lack of geographical meaning, the composed dimensions are difficult to explain. (2) Feature selection, which selects sub feature set by statistics, rough sets and genetic algorithm etc. Feature selection also has draw backs. First, many methods depend on the number of training samples, if training set is small then it would select small sub feature set, and useful feature may be filtered. Second, classifier's ability maybe unstable which produced by a small sub set selected from a large feature set, if feature number is large (especially its number larger than 100) many methods would be inefficient (Rokach, 2008).

Remote sensing data varied in character, we can't expect some sub feature set or some classifier have better performance

Received: 2008-09-22; **Accepted**: 2009-03-04

Foundation: National Natural Science Foundation of China (No. 40871188), Knowledge Innovation Program of Chinese Academy of Sciences (No.ZCX2-YW-Q10-1-3).

First author biography: PAN Xin (1978—), male, Ph.D. of Chinese Academy of Sciences. His research focuses on spatial data mining. E-mail: panxinpc@gmail.com

Corresponding author: ZHANG Shu-qing, E-mail: zhangshuqing@neigae.ac.cn

at all the circumstance. Combining outputs from multiple classifiers, known as ensemble classifier, is one of the standard and most important techniques for improving classification accuracy in machine learning (Bryll, 2003). For the ensemble method, the training set is randomly sampled into many subsets, classifiers trained by each subset and the output of these classifiers is combined to produce the output of the ensemble by voting strategy. Combining a set of diverse classifiers each has diverse classification space (misclassified samples not overlap) will lead to powerful classification ability.

Ensemble classifier training set split methods have two categories: resampling and feature partition. Bagging and boosting are the most popular methods of resampling, they obtained sub training set by random selection and add attendance rate of misclassified samples (Opitz & Maclin, 1999). Feature partition split feature set into several subsets, each subset has all the train object and a part of feature set (Tumer & Oza, 2003). Experiments show that feature partition method's performance is better than resampling (Rokach & Maimon, 2005). Well designed feature partition method can improve classification accuracy, reduce affect by too large dimension, increase training speed and avoid over-fitting. The key problem is how to obtained sub feature set which has good classification ability from feature set.

Rough set theory (RST) proposed by Pawlak (1982) is an extension of conventional set theory that supports approximations in decision making. The advantage of rough set theory is that it does not need any preliminary or additional information about data. It has also been conceived as a mathematical approach to analyze and conceptualize various types of data, especially to deal with vagueness or uncertainty (Pawlak, 1982, 1999). In rough set theory, reducts are particular subsets of features which provide the same information for classification purposes as the full set of features.

This study obtained several reducts from a feature set; each reducts trained an artificial neural network (ANN) classifier, and combined these classifiers into Rough Set Ensemble Remote Sensing Image Classifier (RSEC). The RSEC integrated with discretization algorithm, make rough set can be applied to continuous spatial feature data of remote sensing image. Different from popular discernable matrices method, on the basis of rough set this study proposed a novel algorithm derived from QuickReduct, it can obtain reducts which do not overlap except the core and more suited to large number of features. This study applied RSEC to remote sensing classification, and discuss its advantage by compared it with other classifiers.

2 PRELIMINARY KNOWLEDGE ON ROUGH SETS

According to Pawlak (1982, 1999), an information system (S) can be viewed as a table of data, consisting of objects (rows in the table) and features. It can be defined by a pair $S=(U, A)$, where:

(1) U is a nonempty finite set of objects called the universe of discourse;

(2) A is a nonempty finite set of features;

(3) For every $a \in A$, there is a mapping $a: U \rightarrow V_a$, where V_a is called the value set of a.

A decision table is an information system of the form $S=(U, A\cup\{d\})$, where $d \notin A$ is a distinguished feature called a decision feature.

With any $P \subseteq A \cup \{d\}$ there is an associated indistinguishable relation IND(P) given by:

$$\mathrm{IND}(P) = \{(x, y) \in U^2 \mid \forall a \in P, a(x) = a(y)\} \tag{1}$$

This corresponds to the indiscernible relation for which two objects are equivalent if and only if they have the same vectors of feature values for the features in P, i.e., if $(\boldsymbol{x}, \boldsymbol{y}) \in \mathrm{IND}(P)$, then $\boldsymbol{x}$ and $\boldsymbol{y}$ are indiscernible by features from P. The partition of U, determined by IND(P) is denoted by U/P or $U/\mathrm{IND}(P)$, which is simply denoted the set of equivalence classes generated by IND(P):

$$U / P = \otimes\{U / \mathrm{IND}(\{a\}) \mid a \in P\} \tag{2}$$

where $A \otimes B = \{X \cap Y \mid \forall X \in A, \forall Y \in B, X \cap Y \neq \phi\}$. The equivalence classes of the indistinguishable relation with respect to P are:

$$[x]_p = \{y \in U \mid (x, y) \in \mathrm{IND}(P)\} \tag{3}$$

Based on the indistinguishable relation, we can define lower and upper approximations. Let $X \in U$, X can be approximated using only the information contained within R:

R-lower approximation: $X_{\underline{R}} = \{x \mid [x]_R \subseteq X\}$ (4)

R-upper approximation: $X_{\overline{R}} = \{x \mid [x]_R \cap X \neq \phi\}$ (5)

If $X_{\underline{R}} \neq X_{\overline{R}}$ then the pair $(X_{\underline{R}}, X_{\overline{R}})$ is called a rough set. With the lower approximation and the upper approximation we can define the positive, negative and boundary regions for a set $X \in U$:

$$\mathrm{POS}_R(X) = X_{\underline{R}} \tag{6}$$

$$\mathrm{NEG}_R(X) = 1 - X_{\overline{R}} \tag{7}$$

$$\mathrm{BND}_R(X) = \mathrm{NEG}_R(X) - \mathrm{POS}_R(X) \tag{8}$$

An important notion in rough set is dependencies between features. Feature Q depending on R (feature dependency, denoted as γ) is defined by:

$$\gamma_R(Q) = \frac{\mathrm{Card}(\underset{X \subset \mathrm{IND}(Q)}{U} \mathrm{POS}_R(X))}{\mathrm{Card}(U)} \tag{9}$$

where Card(*) is the cardinality of a set. Q depends totally on R if $\gamma=1$, partially on R if $0<\gamma<1$ and not on R if $\gamma=0$.

Reducts are particular subsets of features which provide the same information for classification purposes as the full set of features. For a decision table $S=(U, A\cup\{d\})$, a reduct is formally defined as a subset R of the conditional feature set A such that $\gamma_R(d)=\gamma_A(d)$. A given data set may have many feature reduct sets, and the collection of all reducts is denoted by

$$R=\{X:X\subset A,\gamma_X(d)=\gamma_A(d)\} \qquad (10)$$

Core is the feature subset which cannot be deleted from any reduct, otherwise the discernibility of the system will decrease.

$$\text{Core}=\cap\{X:X\subset A,\gamma_X(d)=\gamma_A(d)\} \qquad (11)$$

Rough set theory discloses the fact that there exist multiple subsets of features which can keep the classification information of the original data. This would be helpful to construct the ensemble classifier.

3 CONSTRUCT ENSEMBLE CLASSIFIER BASED ON THE ROUGH SET

3.1 Feature partition algorithm

Feature partition algorithm can split decision table $S=(U, A\cup\{d\})$ feature set A into several subsets, and each subset keeps the classification information of the original data. Feature Partition can be written as follows:

FeaturePartition

Input: Decision table $S=(U, A\cup\{d\})$

Output: Reduct set R

Step 1: $R\leftarrow\phi$ core←FindCore(S) Initialize R to empty set and search core features by FindCore algorithm

Step 2: $A_{\text{candidate}}\leftarrow$A-{core} Set $A_{\text{candidate}}$ with candidate feature

Step 3: $X\leftarrow$FindReduct ($core$, $A_{\text{candidate}}$, S) Search a reduct from $A_{\text{candidate}}$ by FindReduct algorithm.

Step 4: WHILE $X\neq\phi$

$R=R\cup\{X\}$

$A_{\text{candidate}}\leftarrow A_{\text{candidate}}-\{X\}$

$X\leftarrow$FindReduct(core, $A_{\text{candidate}}$, S)

END WHILE

Step 5: return R

FeaturePartition algorithm calls FindCore and FindReduct algorithm. Core is the feature subset which cannot be deleted from any reduct. FindCore algorithm can search core based on Equ.(9) as follows:

FindCore

Input: Decision table $S=(U, A\cup\{d\})$

Output: Core

Step 1: Core←ϕ $C\leftarrow A$

Step 2: WHILE If exists $a\in C$ satisfy $\gamma_A<\gamma_{A-\{a\}}$

the feature a is core feature, add this feature into Core set

Core←Core∪{a}

$C\leftarrow C-\{a\}$

End While

Step 3: Return Core

Several methods to find reducts have been proposed in rough set research. Most of these methods are extended from the discernable matrices method (Skowron & Rauszer, 1992) or the QuickReduct algorithm (Chouchoulas & Shen, 2001). Discernable matrices method constructs matrix and obtains the prime implicants through representation of the Boolean function. QuickReduct algorithm adopts dependency degree (as shown in Equ.(9)) as reduct finding criterion. QuickReduct algorithm can't find all the reducts, but it's time complexity and space complexity relatively low when features number is large. FindReduct is derived from QuickReduct, it can improve QuickReduct by considering core features and candidate features:

FindReduct

Input: Core features set Core; Candidate feature set $A_{\text{candidate}}$; decision table $S=(U, A\cup\{d\})$;

Output: A reduct set reduct

Step 1: reduct←ϕ

Step 2: WHILE $\gamma_{\text{reduct}}<\gamma_A$

IF exists $a\in A_{\text{candidate}}$ make $\gamma_{\text{reduct}}<\gamma_{\text{reduct}\cup\{a\}}$ THEN

reduct←redcut ∪{a} $A_{\text{candidate}}\leftarrow A_{\text{candidate}}-\{a\}$

ELSE

cant find a redcut in $A_{\text{candidate}}$

return ϕ

END IF

END WHILE

Step 3: return reduct

3.2 Discretization

A decision table constructed by remote sensing data may contain integer-valued attributes (e.g., spectral bands) or real-valued attributes (e.g., texture information). These attributes usually have a large number of values (called continuous attributes). If the decision table with a large number of attribute values (relatively number of objects in U) is analyzed, then there is a very little chance that a new object will be properly recognized by matching its attribute value vector with the rows of this table, Therefore, discretization, which is a process that quantifies the numeric data into intervals, and assigns each interval a discrete value, is necessary to achieve a higher quality of classification.

A supervised classification task requires a training data set consisting of M examples, where each example belongs to only one of S classes. F indicates any of the continuous attributes from the mixed-mode data. Next, there exists a discretization scheme D on F, which discretizes the continuous domain of attribute F into n discrete intervals bounded by the pairs of numbers:

$$D:\{[d_0,d_1],(d_1,d_2],...,(d_{n-1},d_n]\},$$

To the rough set data in same interval can't be discerned and in different interval can be discerned. Lots of discretization algorithms have been proposed; in our study we adopt a simplest equal width algorithm. Interval number is the same as the class number adopted by lots of algorithms as default parameter (Kurgan & Cios, 2004).

3.3 Construct ensemble classifier

RSEC's construction have three steps: (1) Construct training set into decision table S=(U, $A\cup\{d\}$), and establish a discretized decision table using gray-level thresholds obtained by equal width algorithm. (2) Use FeaturePartition algorithm split feature sets into several reducts R_n. (3) Each R constructs a decision table S_n=(U, $R_n\cup\{d\}$), each decision table trains an artificial neural network (ANN). (4) Classifier decides an object's class by voting from each ANN's decision. As illustrated in Fig. 1:

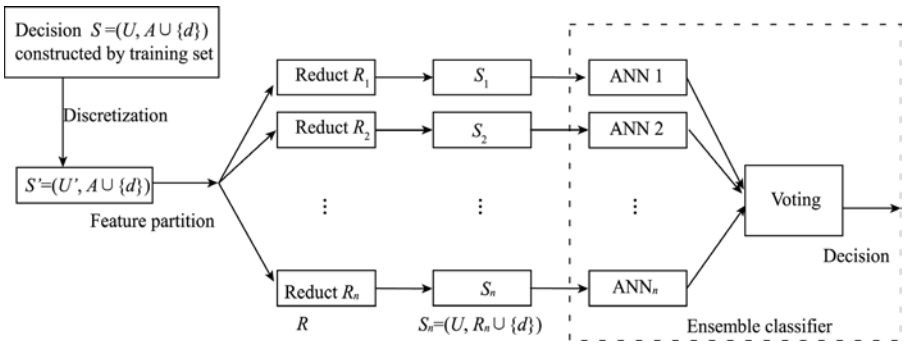

Fig. 1 The construct procedure of RSEC

Any classifiers (such as neural networks, decision tree and SVM, etc.) can act as sub-classifier, this article uses multi-layer perceptron artificial neural network as sub-classifier. ANN was designed by Matlab. Layer corresponding remote sensing data is input into the first lay, it uses tansig as transfer function. Output layer corresponds to the number of classes and raingd was used as training function.

4 EXPERIMENTAL RESULTS

This research chooses Landsat-5 TM image (October 30, 2006, orbit number 114/26, image size 684 × 844) covering the whole of Honghe National Nature Reserve (HNNR)'s. The land use/cover categories in the study area include Marsh Land, Forestland, Meadow, Dry Farmland and Paddy Field. Six spectral bands were used (including blue (Band 1), green (Band 2), red (Band 3), near-infrared (Band 4) and two mid-infrared (Band 5 and 7)). In order to study the relationship between large number of features and classification , this study utilizes eCognition to segment remote sensing image by object-oriented method and total 76 different object features were extracted (Table 1).

Table 1 Test data set's attribute list

Feature category	Object features	Number of features
Spectral	Mean	6
	Brightness	1
	Max-diff	1
	Standard deviation	6
GLCM	Contrast	6
	Correlation	6
	Dissimilarity	6
	Entropy	6
	Homogeneity	6
	Mean	6
	Standard deviation	6
GLDV	Contrast	6
	Entropy	6
	Mean	6
Shape	Length/width	1
	Shape index	1
SUM		76

Based on field experience aided, 1000 independent samples were extracted in the experiment, including 500 as training set and 500 as test set. In order to verify the relation among algorithms, features and sample numbers, 500 training set is further split into 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 numbers by random selection.

The construction of ensemble remote sensing image classifier based on Rough set can be divided into two steps: (1) Use Visual C# 2.0 implement discretization and feature partition based on rough set, each of *R* constructed a decision table and export to files. (2) Use Matlab construct Card(*R*) numbers of ANNs, each ANN was trained by the corresponding decision table. Classifier decides an object's class by voting from each ANN's decision. Because of the ANN's training may have equal gradients with several directions training function would randomly select the drop direction, this strategy may lead to the subtle difference of classification accuracy, even use the same set of training data. Therefore, our study gives the classification accuracy based on the mean value of 5 times training results.

In order to verify the classification ability, RSEC was compared with ANN classifier and feature selection + ANN (feature selection use QuickReduct algorithm). Three methods' classification accuracy can be seen from Fig. 2: RSEC's classification accuracy reaches minimum (85.96%) at 50 samples and maximum (92.36%) at 400, the classification accuracy is 91.8% at 500. ANN's classification accuracy is 63.84% at 50, it reaches the maximum (83.96%) at 250 samples, and 81.88% at 500. That of Feature selection + ANN reaches the maximum (83.92%) at 50 samples, and 77.96% at 500, reaching the minimum (74%) at 350 samples.

Only use ANN tends to converge to local minimum when sample number is small, its generalization ability is weak, therefore classification accuracy is low. When ANN use more samples this situation has improved, but still lower then RSEC's lowest classification accuracy at 50 samples. Feature selection + ANN's classification accuracy is unstable, it maybe decrease when samples number increase. Only RSEC's performance is stable (Table 2).

Table 2 Three methods classification accuracy in each number of training samples

Training set size	5 times training result's mean value		
	ANN	Feature selection + ANN	RSEC
50	<u>63.84</u>	**83.92**	<u>85.96</u>
100	74.80	79.88	88.00
150	80.92	78.8	91.84
200	82.88	78.44	89.44
250	**83.96**	77.76	91.52
300	80.12	77.24	91.00
350	79.44	<u>74.00</u>	90.28
400	81.92	83.04	**92.36**
450	83.08	76.76	92.04
500	81.88	77.96	91.80
Mean	79.284	78.78	90.424

Three methods' result comparison can be seen from Fig. 2. It used training set size as abscissa, classification accuracy comparison between different methods as ordinate. As can be seen from Fig. 2(a) the RSEC's classification accuracy compared with ANN is improved, especially at small samples, with the increase of training set size improved degree showed the downtrend. Fig. 2(b) show the comparison between RSEC and Feature selection + ANN, RSEC's classification accuracy is higher then feature selection + ANN at all the training set size. Fig. 2(b) is comparison between feature selection + ANN and ANN, Feature selection + ANN is unstable, classification accuracy is increase at 50, 100 and 400 samples, but decrease at 150, 200, 250, 300, 350, 450 and 500 samples.

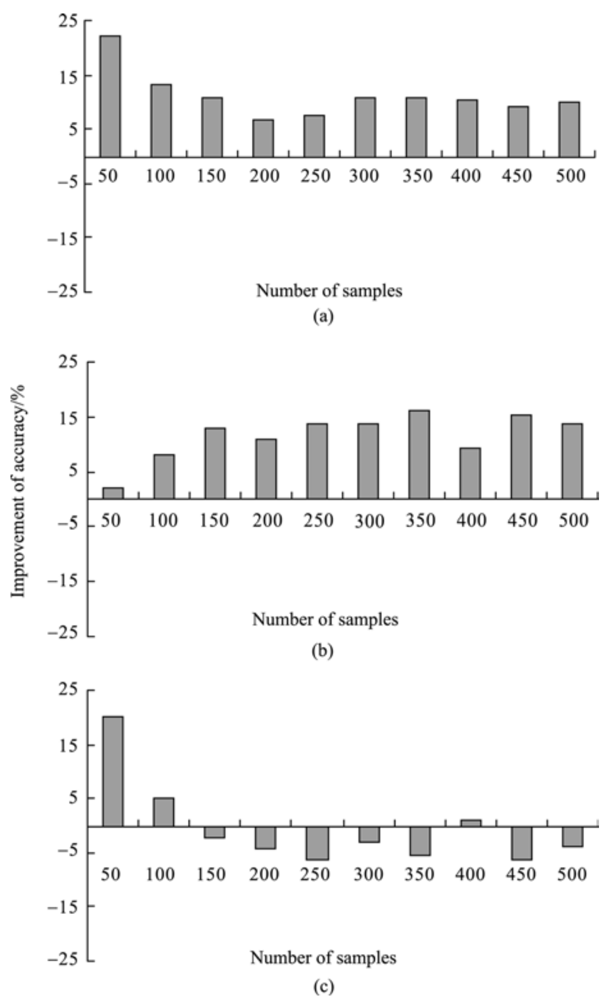

Fig. 2 The classification accuracy comparison of three methods
(a) The comparison of RSEC and ANN; (b) The comparison of RSEC and feature selection +ANN; (c) The comparison of feature selection +ANN and ANN

Three methods' running speed can be seen from Table 3 (the mean value of running 5 times, accurating to seconds). In the training stage ANN only need train a classifier model, so speed is fastest. Feature selection + ANN need feature selection before train classifier model, it's training speed is similar to ANN when samples is small, but significantly increased when samples are big. RSEC need traversal data several times and train several ANN, so training speed is slow, it reach 108 seconds at 500 samples. Remote sensing image classification can be performed after training stage; its speed has great relationship with classifier's complex degree. Feature selection + ANN is simplest so speed is fasted, ANN need all the features so its speed is slower. ANN need run several sub-classifier, running speed relate to sub-classifier's number, its speed is slowest, but classification efficiency is still high, whole image's classification only need 10 to 13 seconds. It will be seen that even RSEC need more training time, but classification efficiency is still high.

Three methods' highest classification result can be seen from Fig. 3(b), (c) and (d) respectively. Fig. 3(a) is the original TM image. In Fig. 3(b) (ANN at 250 samples) the misclassification phenomenon of Forestland and Dry Farmland, Dry Farmland and Meadow are quite serious. Compared with Fig. 3(b), classification effect in Fig. 3(c) (Feature selection + ANN at 50 samples) was improved, but misclassification phenomenon are also significant at the border between different classes. Fig. 3(c) represents the classification result by RSEC, its misclassification and border is improved compared with the other two methods.

Table 3 The running speed comparison of three methods

Training set size	ANN		Feature selection + ANN		ANN	
	Training	Classification	Training	Classification	Training	Classification
50	6	8	6	6	47	13
100	6	8	7	5	49	12
150	7	7	9	6	41	13
200	8	6	14	6	61	10
250	8	7	15	6	67	10
300	9	8	17	6	67	10
350	10	8	22	5	82	10
400	11	7	26	5	97	11
450	11	8	32	6	119	10
500	12	8	32	6	108	10

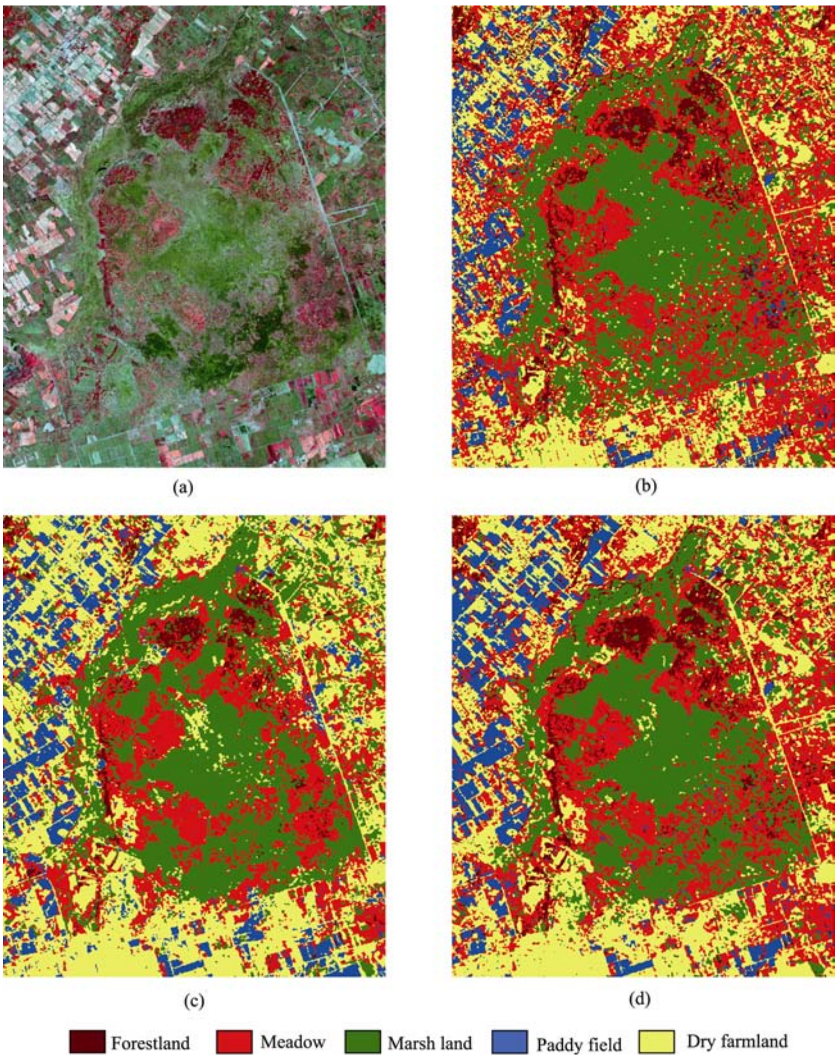

Fig. 3 The results of three classification methods' land use classification
(a) TM image (4, 3, 2); (b) ANN; (c) Feature selection +ANN; (d) RSEC

5 CONCLUSION AND DISCUSSION

Remote Sensing classification is a traditional and eternal topic. In order to improve classification accuracy when surface condition is very complex, additional spatial features on the basis of spectrum were adopted. Unfortunately, too many features would bring new challenge to remote Sensing classification. The effective combination of dimension reduction and spatial features will benefit to improve the classification accuracy.

Traditional feature selection method of rough set has draw backs of unstable and depends on samples' number excessively. RSEC can partition feature set into a lot of reducts, and constructs training subset by utilizing these reducts, the decisions from all the sub-classifiers are combined with a voting strategy, on the one hand it can reduce input features to a single classifier, on the other hand it can avoid bias caused by single classifier, and improve the classification accuracy. The RSEC classifier has been compared with the direct ANN method and the traditional feature selection method. It can be seen from the result that RSEC has better classification accuracy and it is also more stable than the others.

Because of the complex structure of RSEC, RSEC need traversal data several times when partition feature set, this would consume more time. Increase partition speed is the key problem for quickening the algorithm running speed. If granularity partition training set by rough set discern relation, construct decision feature and granularity index, would improve algorithm traversal data only one time. We will further study granularity partition and index construction, further improve RSEC's efficiency.

REFERENCES

Aguera F. 2008. Using texture analysis to improve per-pixel classification of very high resolution images for mapping plastic greenhouses. *ISPRS Journal of Photogrammetry & Remote Sensing*, **63**: 1—12

Bryll R, Gutierrez-Osuna R and Quek F. 2003. Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition*, 36: 1291—1302

Chou T Y, Lei T C and Shih I L. 2006. The study of knowledge expression on image classfication by using rought sets theory and PCA. Asian Conference on Remote Sensing, Mongolia: Curran Associates

Chouchoulas A and Shen Q. 2001. Rough Set-aided keyword reduction for text categorization. *Applied Artifcial Intelligence*, **15**(9): 843—873

Kurgan L A and Cios K J. 2004. CAIM discretization algorithm. IEEE *Transactions on Knowledges and Data Engineering*, **16**(2): 145—153

Lei T C, Wan S and Chou T Y. 2007. The comparison of PCA and discrete rough set for feature extraction of remote sensing image classification: A case study on rice classification. *Computational Geosciences*, **12** (1): 1—14

Leung Y, Fung T, Mi J S and Wu W Z. 2007. A rough set approach to the discovery of classification rules in spatial data. *International Journal of Geographical Information Science*, **21**(9): 1033—1058

Opitz D and Maclin R. 1999. Popular ensemble methods: An empirical study. *Journal of Artifcial Intelligence Research*, **11**: 169—198

Ouyang Y and Ma J W. 2006. Land cover classification based on tolerant rough set. *International Journal of Remote Sensing*, **24**: 3041—3047

Pawlak Z. 1982. Rough sets. *International Journal of Computer and Information Sciences*, **11**: 341—356

Pawlak Z. 1999. Rough classification. *International Journal of Human–Computer Studies*, **51**: 369—383

Rokach L and Maimon O. 2005. Feature set decomposition for decision trees. *Intelligent Data Analysis*, **9**: 131—158

Rokach L. 2008. Genetic algorithm-based feature set partitioning for classification problems. *Pattern Recognition*, **41**: 1676—1700

粗集属性划分的集成遥感分类

潘　欣[1,2,3]，张树清[1]，李晓峰[1]，那晓东[1,2]，于　欢[1,2]

1. 中国科学院 东北地理与农业生态研究所，吉林　长春　130012;
2. 中国科学院 研究生院，北京　100039;
3. 长春工程学院 电气与信息工程学院，吉林　长春　130012

摘　要：提出了一种基于粗集属性划分的遥感分类新方法，构造了基于粗集的集成遥感分类器。该分类器利用粗集理论将输入的属性集合划分为多个约减，利用这些约减构造多个训练子集。每个训练子集训练神经网分类器，在决策时将多个单个分类器的结果进行投票选举。这种方法即减少了单个分类器的输入属性个数，又避免了由于属性选取造成单一分类器在某些分类上的错误偏见。该分类器与神经网分类器方法，以及属性选取与神经网结合方法进行了比较。结果表明 RSEC 无论在分类精度上，还是在不同样本个数条件下的精度稳定程度上均有较好表现。

关键词：集成分类器，粗集，神经网，属性选取

中国分类号：TP751.1　　文献标识码：A

1 引　言

土地利用/覆盖及其变化是全球环境变化中的重要因子，遥感技术的提高增强了获取这些信息的能力(Ouyang & Ma, 2006)。如何及时准确地从遥感影像自动地获取土地利用/覆盖信息，不但依赖数据源的质量和分辨率，而且依赖分类技术的选取。目前出现了很多遥感分类方法如：最大似然分类、模糊集、神经网、基于分类树和基于分类规则的分类方法等(Leung 等, 2007)。

在光谱基础上进一步考虑其空间特征(如灰度共生矩阵生成的纹理特征)可以为分类器提供更多信息，从而提高分类精度(Shaban & Dikshit, 2001)。但是并非所有的参数和方法获得的空间特征对分类有帮助(Aguera，2008)，过多的分类属性可能导致分类器的过度拟合并降低分类精度(Lei 等, 2007)。

为了解决属性过多这一问题，目前主要使用的方法是：(1)主成分分析(PCA)，它通过属性的线性组合构成一组正交的矢量替代原有的属性，并选择方差贡献率高的前几个矢量达到减少分类属性的目的。PCA 主要缺陷是：它可能将与分类不相关的属性加入正交矢量，并占较高的比重，从而影响分类精度(Chou 等, 2006)；由于缺乏实际的地理意义，组合后的矢量难以解释。(2)属性选取(feature selection)，它通过统计、粗集和遗传算法等方法选取对于分类有利的属性子集。属性选取方法也存在不足，很多属性选取算法依赖于训练集样本的多少，如果训练集较少则可能获得较小的属性子集，很多和分类相关的属性有可能被过滤掉；在大量属性中获取一个小的属性子集所产生的分类器，其分类能力和精度一般不稳定；在属性较多的情况下(尤其是个数超过100)，许多属性选取算法的效率很低(Rokach, 2008)。

遥感影像数据变化多样，不能期望一组属性子集或者一个分类器在所有情况下有较好的表现。将多个分类器的决策进行集成构成集成分类器(ensemble classifier)，是在机器学习领域提高分类精度的一个重要的手段(Bryll, 2003)。集成分类器将一个训练集划分为多个子集，每个子集训练一个单独的分类器，多个分类器对需要分类的对象进行判断，以投票选举的形式得出最终的结果。将分类空间特征不同(即在错误分类方面不交叠)的多个分类集成，将取得较好的分类效果。

集成分类器训练集的划分，可以分为重复采样(resampling) 和属性划分(feature partion)两大类。重

收稿日期：2008-09-22；修订日期：2009-03-04

基金项目：国家自然科学基金项目(编号：40871188)和中国科学院知识创新工程重要方向项目(编号：ZCX2-YW-Q10-1-3)。

第一作者简介：潘欣(1978—　)，男，中国科学院东北地理与农业生态研究所博士，长春工程学院讲师。主要从事空间数据挖掘方面的研究。E-mail: panxinpc@gmail.com。

通讯作者：张树清，E-mail: Zhangshuqing@neigae.ac.cn。

复采样主要包括 bagging 和 boosting 两类方法，它们通过随机选取和加大错误分类样本出现的几率构成训练子集(Opitz & Maclin, 1999); 属性划分将训练集全部属性划分为多个子集，每个子集包含全部训练样本和部分属性(Tumer & Oza, 2003)。实验表明属性划分方法由于分类器之间差异较为明显，所以可以获得比重复采样方法更好的分类效果(Rokach & Maimon, 2005)。设计较好的属性划分算法，可以提高分类精度，减小属性维度过高所造成的影响，提高训练速度并防止过度拟合。其关键问题是如何从一个属性集合中获得多个具有良好分类能力的属性子集。

由 Pawlak 在 1982 年提出的粗集理论(rough set theory, RST)是对传统集合论在近似决策方面的扩展。相对于模糊集，粗集的一个最大优势是对于要分析的数据它不需要任何先验或附加的信息；尤其是在处理不确定、不一致的数据方面，粗集是一个有力的数学分析工具(Pawlak, 1982, 1999)。在粗集理论中，属性的约减(reducts)是保持原有属性分类能力的最小属性子集。

通过一个属性集获得多组属性约减，每个属性约减训练一个神经网分类器，然后将多个分类器组合构成“基于粗集的集成遥感影像分类器”(rough set ensemble remote sensing image classifier, RSEC)。本分类器将离散化算法集成到分类器之中，使得粗集可以适用于连续的遥感影像空间属性数据。与其他粗集集成神经网使用区分矩阵作为约减获取手段不同，本分类器使用改进的 QuickReduct 算法来找出除了核外不相交的多个属性子集，使其更适合应用于大量属性数据。本研究将其应用于实际遥感影像分类，通过与其他方法的比较，探讨了新建的分类器优势。

2 粗集基本理论

根据 Pawlak 的粗集理论，一个信息系统 S 可以被看作是一个数据表。该表由对象(行)和属性(列)组成。它可以由对 $S=(U, A)$ 表示，其中：论域 U 是非空有限集合；A 是非空有限的属性集合；对于 A 中的每一个元素 $a\in A$，存在一个映射 a: $U\rightarrow V_a$，其中 V_a 是 a 取值的集合。一个决策表就是形如 $S=(U, A\cup\{d\})$ 的信息系统，其中 $d\notin A$ 是决策属性。对于任意的属性集合 $P\subseteq A\cup\{d\}$ 存在一个不可区分关系 IND(P):

$$\mathrm{IND}(P)=\{(x,y)\in U^2 \mid \forall a\in P, a(x)=a(y)\} \qquad (1)$$

也就是说如果 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 是 P 上的不可区分关系，那么 $\boldsymbol{x}$, $\boldsymbol{y}$ 矢量在 P 的所有属性上都相等。根据不可区分关系可以确定 U 的划分表示为 U/P 或 $U/\mathrm{IND}(P)$:

$$U/P=\otimes\{U/\mathrm{IND}(\{a\}) \mid a\in P\} \qquad (2)$$

其中 $A\otimes B=\{X\cap Y \mid \forall X\in A, \forall Y\in B, X\cap Y\neq\phi\}$。一个基于 P 不可区分关系的等价类可以定义为:

$$[x]_p=\{y\in U \mid (x,y)\in \mathrm{IND}(P)\} \qquad (3)$$

根据不可区分关系可以定义上下近似集，令集合 $X\in U$, X 可以由下面两个集合近似的表示:

$$R\text{-下近似集:}\quad X_{\underline{R}}=\{x \mid [x]_R\subseteq X\} \qquad (4)$$

$$R\text{-上近似集:}\quad X_{\overline{R}}=\{x \mid [x]_R\cap X\neq\phi\} \qquad (5)$$

如果 $X_{\underline{R}}\neq X_{\overline{R}}$ 那么对 $(X_{\underline{R}}, X_{\overline{R}})$ 就称之为粗糙集。根据上下近似集对于一个集合 $X\in U$ 可以进一步定义正域、负域和边域:

$$\mathrm{POS}_R(X)=X_{\underline{R}} \qquad (6)$$

$$\mathrm{NEG}_R(X)=1-X_{\overline{R}} \qquad (7)$$

$$\mathrm{BND}_R(X)=\mathrm{NEG}_R(X)-\mathrm{POS}_R(X) \qquad (8)$$

粗集的一个重要概念是属性之间的依赖度。属性 Q 对于属性 R 的依赖程度可以定义为(属性依赖度，由 γ 表示):

$$\gamma_R(Q)=\frac{\mathrm{Card}(\bigcup_{X\subset \mathrm{IND}(Q)}\mathrm{POS}_R(X))}{\mathrm{Card}(U)} \qquad (9)$$

其中 Card(*)一个集合的势。如果 Q 完全依赖于 R, 那么$\gamma=1$；部分依赖时，$0<\gamma<1$ 如果完全不依赖，那么$\gamma=0$。

约减(Reducts)是属性集的特殊子集，它能提供和属性的全集相同分类信息。对于一个决策表 $S=(U, A\cup\{d\})$，如果一个子集 R 可以和 A 具有关系 $\gamma_R(d)=\gamma_A(d)$，那么可以称之为一个约减。一个数据集可以提供多个约减，约减集是约减的集合可以表示为:

$$R=\{X: X\subset A, \gamma_X(d)=\gamma_A(d)\} \qquad (10)$$

核(Core)是属性集合中的关键属性，如果删除核属性那么整个属性集的分类能力下降，所有的约减的必须包含核属性。核可以表示为所有约减的交集:

$$\mathrm{Core}=\cap\{X: X\subset A, \gamma_X(d)=\gamma_A(d)\} \qquad (11)$$

任何核属性不能从约减中删除，否则将引起依赖度γ的降低。通过在一个属性集当中找到多个属性约减，就可以构成多个用于训练分类器的属性子集，每个子集保持了和属性全集相同的分类能力，可为构造集成分类器提供帮助。

3 基于粗集属性划分的集成分类器构建

3.1 属性划分算法

属性划分算法可以将决策表 $S=(U, A\cup\{d\})$中的属性集合 A 划分为多个属性子集，每个子集是一个

属性的约减并具有和整个属性集 A 相同的分类能力。属性划分算法 AttributePartition 如下:

AttributePartition

```
Input:    决策表 S=(U, A∪{d})
Output:   划分后的约减集合 R
Step 1:   R ← φ        core←FindCore(S)      初始化 R 为空集，并通过 FindCore 算法查找核。
Step 2:   A_candidate←A-{core}               候选的属性放在 A_candidate 中。
Step 3:   X←FindReduct(core, A_candidate, S)  在 A_candidate 属性集合中查找一个约减放入 X 中。
Step 4:   WHILE  X ≠ φ
              R = R ∪ {X}                    将 X 添加到 R 中
              A_candidate←A_candidate- {X}   将已经添加到 R 中的属性从 A_candidate 中移出。
              X←FindReduct(core, A_candidate, S)
          END WHILE
Step 5:   return R                           算法结束返回划分后的结果 R。
```

AttributePartition 算法调用了核查找算法 FindCore 和约减查找算法 FindReduct 算法。核是对于分类重要属性，所有的约减中都包含所有核属性。FindCore 算法根据属性依赖度公式(9)作为核查找的标准(FindCore)，其内容如下:

FindCore

```
Input:    决策表 S=(U, A∪{d})
Output:   属性的核 Core
Step 1:   Core ← φ      C ← A
Step 2:   WHILE  存在 a∈C 使得 γ_A < γ_{A−{a}}
属性 a 就是核属性，将其加入 Core 中
Core ← Core ∪ {a}
C ← C − {a}
          End While
Step 3:   Return Core
```

在一个属性集合当中获得约减集算法主要包括 Discernable matrices 算法(Skowron & Rauszer, 1992)和 QuickReduct 算法(Chouchoulas & Shen, 2001)。Discernable matrices 算法构造区分矩阵并将矩阵中的析取范式转换为合取范式，每一个合取范式就是一个约减。算法 QuickReduct 采用属性依赖度(公式(9))作为查找属性约减的标准。QuickReduct 算法虽然不能找到所有的约减，但是当属性数较多的情况下它的时间复杂度和空间复杂度相对较低。约减查找算法 FindReduct 是在 QuickReduct 基础上考虑到候选属性范围和核属性的改进算法:

FindReduct

```
Input:    核 Core;  候选属性 A_candidate;  决策表 S=(U, A∪{d});
Output:   一个约减  reduct
Step 1:   reduct ← φ
Step 2:   WHILE   γ_reduct < γ_A
              IF  存在  a ∈ A_candidate  使得  γ_reduct < γ_{reduct∪{a}}  THEN
                  reduct ← redcut ∪ {a}            A_candidate ← A_candidate − {a}
              ELSE
                 无法在 A_candidate 中找到约减集，返回空集合表示查找失败
                     return φ
              END IF
          END WHILE
Step 3:   return reduct
```

3.2 属性数据的离散化

由遥感影像数据构成的决策表包含整型属性(如:光谱信息)或实数属性(如: 纹理信息)。这些属性值的个数很多，称之为连续属性(Continuous attributes)。如果粗集直接分析这些属性很难得应有的分析效果，所以需要将连续属性离散化(Discretization)。

离散化就是将大量个数的数据按照一定的阈值分成段。对于一个属性 $a\in A$，由不同阈值构成的分段可以表示为:

$$D:\{[d_0,d_1],(d_1,d_2],\ldots,(d_{n-1},d_n]\}$$

处于不同的分段数据是可区分的，而处于同一分段的数据是不可区分的。离散化算法的主要任务就是决定一个属性数据分段的个数和阈值。离散化算法方法有很多，在本文中采用最简单的等距离的离散化方法(Equal width)。离散化分段的个数采用很多相关算法默认值与分类个数相同(Kurgan & Cios, 2004)。

3.3 集成分类器的结构

构造集成分类器的步骤如下: (1)将训练数据构成决策表 $S=(U, A\cup\{d\})$，对该决策表进行离散化，获得离散化的决策表 $S'=(U', A\cup\{d\})$。(2) 使用属性划分算法 AttributePartition 将属性集 A 划分为多个约减的集合 R。(3) 使用集合 R 中的每个约减 R_n 构成多个决策表 $S_n=(U, R_n\cup\{d\})$，每个决策表训练一个多层神经网感知器。(4) 将多个神经网集成，对于一个对象的分类以投票选举的形式产生结果。构造过程如图 1。

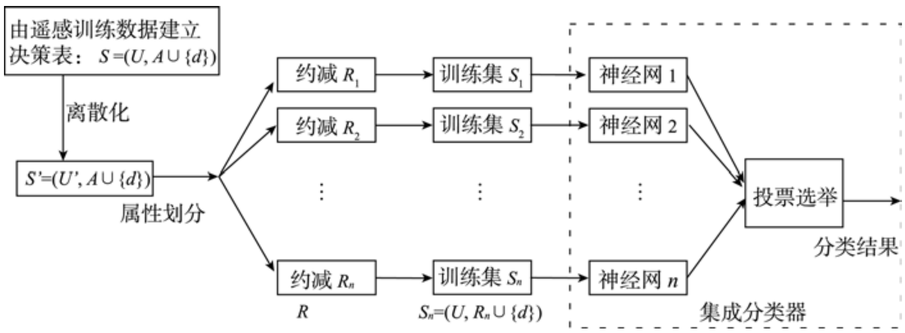

图 1 基于粗集的集成遥感分类器的构建过程

对于子分类器的选择可以任意选定(如神经网、决策树和 SVM 等)，本文中 RSEC 分类器采用了多层神经网感知器作为子分类器。

子分类器所采用的神经网为多层感知器(multi-layer perceptron, MLP)BP 网，利用 Matlab 设计实现。第一层输入层为对应的遥感属性数据，使用传递函数 tansig。输出层个数直接对应分类个数，采用 purelin 作为传递函数。 训练采用 traingd 作为训练函数。

4 影像分类实验

选用位于三江平原的洪河自然保护区的 Landsat-5 TM 影像(2006-10-30，轨道号 114/26，截取的试验区影像大小 684×844)，土地利用/覆盖类型包括: 沼泽、林地、草地、旱地和水田共 5 种。选取的波段为 1—5 和 7 共 6 个波段。为了研究多个属性和分类的关系，首先使用 eCognition 软件对遥感影像进行面向对象分割。根据实地调查采样，共获取了 1000 个已知分类数据。其中 500 个作为训练样本, 500 个作为验证数据。为了验证算法、属性、样本个数之间的关系，进一步将 500 个训练样本使用随机函数分别选取 50, 100, 150, 200, 250, 300, 350, 400, 450 和 500 个样本作为训练数据。

基于粗集的集成遥感影像分类器的构建主要分为两步: (1)利用 Visual C# 2.0 程序实现离散化和基于粗集的属性划分算法，并划分 R 中的每一个约减构造决策表导出; (2)利用 Matlab 建立 Card(R)个多层神经网感知器，每个感知器使用对应决策表进行训练。分类时集成感知器的输出进行投票选举。由于使用神经网训练时会有多个方向的梯度相等，此时训练函数采用随机选取下降方向的策略，这导致相对同一组训练数据，在每次训练之后分类精度会有细微的不同。为此，本研究给出的分类精度是 5 次训练并分类取均值结果(表 1)。

为了验证分类能力, RSEC 同直接采用神经网和属性选取+神经网(属性选择是使用 QuickReduct 算法输出的约减)这两种算法进行比较，它们也使用同 RSEC 子分类器一样的 BP 网，只是输入属性有所不同。3 种算法的分类精度如表 2。使用 RSEC 遥感影

表 **1** 实验数据的属性列表

属性分类	计算方法	属性个数
对象光谱	均值	6
	亮度	1
	最大差分	1
	标准差	6
灰度共生矩阵(GLCM)	对比度	6
	相关性	6
	相异性	6
	熵	6
	均质性	6
	均值	6
	标准差	6
灰度差分矢量(GLDV)	对比度	6
	熵	6
	均值	6
对象形状指数	长宽比	1
	形状指数	1
总计		76

表 **2** 每种训练样本个数的三种方法 **5** 次分类的精度均值

训练样本个数	5 次分类的精度均值		
	直接采用神经网	属性选取+神经网	RSEC
50	63.84	**83.92**	85.96
100	74.80	79.88	88.00
150	80.92	78.8	91.84
200	82.88	78.44	89.44
250	**83.96**	77.76	91.52
300	80.12	77.24	91.00
350	79.44	74.00	90.28
400	81.92	83.04	**92.36**
450	83.08	76.76	92.04
500	81.88	77.96	91.80
平均	79.284	78.78	90.424

像分类器分类的精度，在 50 个训练样本的时候为分类精度最低(85.96%)，而 400 个样本时达到最高(92.36%)，在 500 个样本的时候分类精度为 91.8%。神经网分类器在使用 50 个分类样本时精度为 63.84%，在 250 个样本的时候分类精度达到最高(83.96%)，在 500 个样本的时候分类精度为 81.88%。先属性选取然后使用神经网(属性选取+神经网)的方法在 50 个样本时达到最高(83.92%)，在 500 个样本的时候达到 77.96%,在 350 个样本的时候达到最低(74%)。

仅采用神经网进行分类，在样本较少的情况下很容易收敛于局部最小点，泛化能力较低，所以分类精度较低。在使用较多的样本时这种情况有所改善，但仍然不及 RSEC 分类器在 50 个样本时分类精度。属性选取+神经网方法的分类精度不稳定，训练样本升高反而有可能降低分类精度。只有 RSEC 分类器的分类精度一直表现比较稳定。

3 种方法比对之后的结果如图 2。横坐标为训练样本的个数，纵坐标为不同算法分类精度提高的程度。该图表示对应一定数量的样本，一个算法和另外一个算法分类精度的差别，如果大于 0 表示精度提高，反之精度降低。图 2(a)为 RSEC 相对于神经网分类精度有所提高，特别是在样本较少的情况下表现尤为明显，随着样本数的增加，其分类精度的提高程度呈下降趋势。图 2(b)为 RSEC 和属性选取+神经网分类精度比较结果，与属性选取+神经网分类器相比, RSEC 分类精度有所提高。图 2(c)为属性选取+神经网方法与仅使用神经网分类精度对比情况，属性选取+神经网方法分类精度不稳定，在样本数为 50，100 和 400 时提高了分类精度；但在 150,

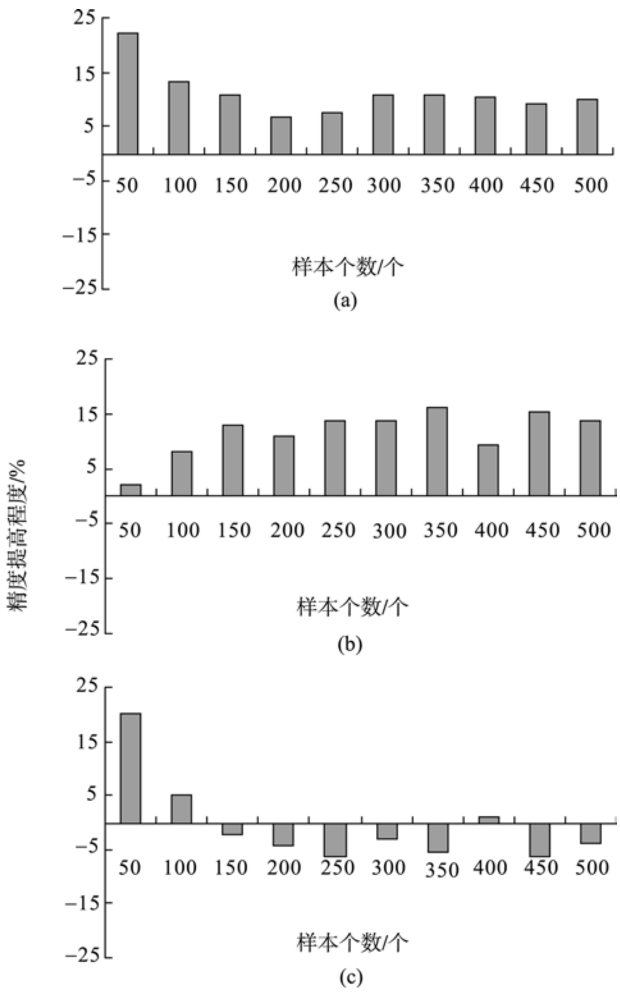

图 2 3 种方法分类精度的比较

(a) RSEC 和直接使用神经网进行比较; (b) RSEC 和属性选取+神经网比较; (c)属性选取+神经网和直接使用神经网进行比较

200, 250, 300, 350, 450和500个样本时，分类精度反而下降。对于直接使用神经网进行分类方法，由于属性较多所以训练算法很快收敛于一个局部极小值(尤其是样本数较少的时候)，但该训练结果泛化能力较低，达不到很高的分类精度。对于属性选取+神经网方法，仅从众多属性中选取一组属性，在某些样本分类精度有所提高，但是某些样本个数情况下反而不如使用全部属性进行分类的精度，这说明它受样本个数的影响比较大。RSEC 分类器在所有样本情况下其分类精度要超过其他 2 种分类器，这说明使用多个子分类器共同进行决策，可以获得稳定的分类效果提高，并受样本个数影响较小。

3 种方法的运行时间对比如表 3(该表数据是经 5 次运行的均值，精确到秒)。在训练阶段直接采用神经网方法由于仅需要训练一个模型，所以速度最快。属性选取+神经网方法需先遍历一次数据进行属性选取，然后再训练神经网，在样本数较少的情况下和直接采用神经网方法比较接近，但是在样本数较多的情况下训练时间明显增加。RSEC 方法需对训练数据先进行多次遍历，然后同时训练多个神经网，训练时间较长，在 500 个样本时达到最大 108s。在模型训练完毕后对整个遥感影像进行分类，和模型的复杂程度有很大关系。由于属性选取+神经网方法仅使用部分属性，模型最为简单，分类速度最快。直接采用神经网方法的需使用全部属性，模型较为复杂，分类速度稍慢。RSEC 需要同时运行多个子分类器，分类速度和子分类器的个数有关，速度较其他两种方法慢，但是分类效率很高，分类整个图像需要 10—13s。可见 RSEC 分类器虽然需要耗费较多的训练时间，但是分类较好。

表 3　3 种方法运行时间对比

样本个数	直接采用神经网		属性选取+神经网		RSEC	
	训练	分类	训练	分类	训练	分类
50	6	8	6	6	47	13
100	6	8	7	5	49	12
150	7	7	9	6	41	13
200	8	6	14	6	61	10
250	8	7	15	6	67	10
300	9	8	17	6	67	10
350	10	8	22	5	82	10
400	11	7	26	5	97	11
450	11	8	32	6	119	10
500	12	8	32	6	108	10

图 3(b)、(c)、(d)分别为神经网、属性选取+神经网、RSEC 3 种方法中精度最高情况下的分类效果图，图 3a 为原始 TM 图像。图 3(b)(分类器：神经网，样本数 250)，该方法对于林地与草地、及旱地和草地误分现象比较严重。与图 3(b)相比，图 3(c)(分类器：属性选取+神经网，样本数 50)的分类情况有所改善，但是在不同地类边界均有误分情况发生，在水田和旱地的边界有误分情况尤为明显。图 3(d)为使用 RSEC 分类器的分类效果图，被误分和分类边界较前两种方法有较大改善。

5　结　论

遥感分类是遥感研究领域的一个古老而又永恒的课题，当地表情况复杂时为了提高分类精度，往往在原有的光谱特征基础上添加新的空间属性信息。但是信息的增加会给遥感分类方法提出新的挑战，过多的属性会使分类混淆降低分类精度，研究数据降维方式，实现光谱、纹理等空间属性的有效组合，将有利于遥感分类精度的提高。

传统的粗集属性选取算法存在分类能力不稳定、且依赖于样本个数的缺点。本文首次提出基于粗集的集成遥感分类器(RSEC)算法。RSEC 方法是将一个大的训练集划分为多个训练子集，用多个分类的投票选取的方法构造分类器，该方法一方面解决了对于一个分类器输入的属性过多的问题，将属性集划分为多个约减，每组属性约减均由较少的属性构成，防止单个分类器过度拟合；另一方面由于分类结果由多个分类器投票产生，解决了一个分类器在某些分类方面的偏见，提高了分类精度。将该算法应用于洪河自然保护区的遥感影像分类，取得了较好的分类效果。并与常用神经网分类器和属性选取+神经网两种方法相比，RSEC 有较好表现。

由于 RSEC 分类器结构较为复杂，查找属性集的划分需要多次遍历训练数据，这将耗费较多时间。提高查找属性集划分的速度是进一步提高算法运行效率的关键。如通过粗集的不可区分关系对训练数据进行粒度划分，建立决策属性和粒度间的索引关系，就可以使算法仅遍历一次训练数据便可获得属性集的划分。我们将进一步研究粗集粒度划分与建立索引技术，提高 RSEC 分类器的效率。

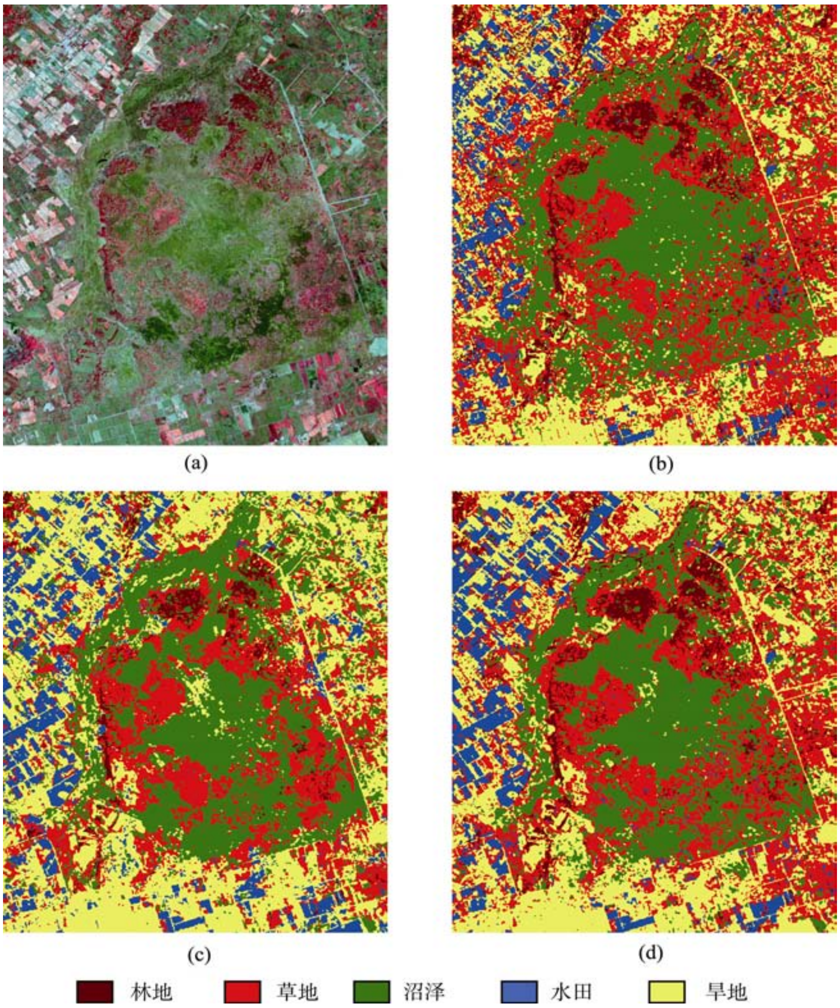

图 3　3 种分类方法的土地利用分类结果

(a)TM 影像(4, 3, 2); (b)神经网分类; (c)属性选取+神经网; (d)RSEC 分类

REFERENCES

Aguera F. 2008. Using texture analysis to improve per-pixel classification of very high resolution images for mapping plastic greenhouses. *ISPRS Journal of Photogrammetry & Remote Sensing*, **63**: 1—12

Bryll R, Gutierrez-Osuna R and Quek F. 2003. Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition*, **36**: 1291—1302

Chou T Y, Lei T C and Shih I L. 2006. The study of knowledge expression on image classfication by using rought sets theory and PCA. Asian Conference on Remote Sensing, Mongolia: Curran Associates

Chouchoulas A and Shen Q. 2001. Rough Set-aided keyword reduction for text categorization. *Applied Artifcial Intelligence*, **15**(9): 843—873

Kurgan L A and Cios K J. 2004. CAIM discretization algorithm. IEEE *Transactions on Knowledges and Data Engineering*, **16**(2): 145—153

Lei T C, Wan S and Chou T Y. 2007. The comparison of PCA and discrete rough set for feature extraction of remote sensing image classification: A case study on rice classification. *Computational Geosciences*, **12** (1): 1—14

Leung Y, Fung T, Mi J S and Wu W Z. 2007. A rough set approach to the discovery of classification rules in spatial data. *International Journal of Geographical Information Science*, **21**(9): 1033—1058

Opitz D and Maclin R. 1999. Popular ensemble methods: An empirical study. *Journal of Artifcial Intelligence Research*, **11**: 169—198

Ouyang Y and Ma J W. 2006. Land cover classification based on tolerant rough set. *International Journal of Remote Sensing*, **24**: 3041—3047

Pawlak Z. 1982. Rough sets. *International Journal of Computer and Information Sciences*, **11**: 341—356

Pawlak Z. 1999. Rough classification. *International Journal of Human–Computer Studies*, **51**: 369—383

Rokach L and Maimon O. 2005. Feature set decomposition for decision trees. *Intelligent Data Analysis*, **9**: 131—158

Rokach L. 2008. Genetic algorithm-based feature set partitioning for classification problems. *Pattern Recognition*, **41**: 1676—1700