Pan, Li, Nahavandi, Saeid and Zheng, Hong 2003, The rough sets feature selection for trees recognition in color aerial images using genetic algorithms, *in Program & abstracts : the Second International Conference on Computational Intelligence, Robotics and Autonomous Systems : CIRAS 2003 : 15-18 December 2003, Pan Pacific Hotel, Singapore*, Center for Intelligent Control, National University of Singapore, Singapore.

The Rough Sets Feature Selection for Trees Recognition in Color Aerial Images Using Genetic Algorithms

Li Pan[1,2], Saeid Nahavandi[1], and Hong Zheng[1]
[1]School of Engineering and Technology, Deakin University
Geelong VIC3217, Australia
[2]School of Remote Sensing and Information & Engineering, Wuhan University
129 Luoyu Road, Wuhan 430079, P.R.China
panli@deakin.edu.au

Abstract

Selecting a set of features which is optimal for a given task is the problem which plays an important role in a wide variety of contexts including pattern recognition, images understanding and machine learning. The concept of reduction of the decision table based on the rough set is very useful for feature selection. In this paper, a genetic algorithm based approach is presented to search the relative reduct decision table of the rough set. This approach has the ability to accommodate multiple criteria such as accuracy and cost of classification into the feature selection process and finds the effective feature subset for texture classification . On the basis of the effective feature subset selected, this paper presents a method to extract the objects which are higher than their surroundings , such as trees or forest, in the color aerial images . The experiments results show that the feature subset selected and the method of the object extraction presented in this paper are practical and effective.

1. Introduction

The rough set that was presented by Z.Pawlak in 1982 is a new mathematical approach to imprecision, vagueness and uncertainty[1],The concept of reduction of the decision table based on the rough sets is very useful for feature selection. Because the decision table includes the condition attributes or features and the decision attributes or categories, the procedure of feature selection based the decision table is distinct and effective. The approaches of feature selection and classification using the rough set were studied in some papers [2,3,4,5]. In their approaches, there are two problems: (1) Because the problem reducing the decision tables based on rough sets is a difficult optimization problem, it is difficult for above methods to find out the global optimal solution or the effective feature subset. (2) The feature values or condition attributes were discretized firstly in above approaches, because continuous data or numeric attributes were proven to be rather unsuitable for the reduction of attributes using the rough sets. In fact, regardless of how effective are the methods of discretization, there is the difference or inconsistence between original data and discrete data. Besides, our application is in the domain of aerial images. It is difficult that symbol rulers are used in the texture classification in the aerial images. Focusing on the two problems, this paper uses two different form the decision tables, namely the decision tables based on continuous and discrete attributes and adopts genetic algorithm to find out the effective feature subset or the optimal reduct decision table of the rough sets. After the effective feature subset selected , we present the algorithm to extract the objects which are higher than their surrounding in color aerial images. The effective feature subset is used in the texture classification of high objects. The technique presented in this paper is illustrated by its application to the problem of trees extraction from aerial images.

This paper is organized as follows: Section 2 explains the method of feature selection based on the rough sets. Section 3 describes the procedure of reduct algorithm using genetic algorithm. Section 4 presents the method of texture classification. In Section 5, experimental results are discussed and the conclusion is given in Section 6.

2. Feature Selection Based on the Rough Sets

The feature selection is an important step in design of classification. Let us assume an image with size m×n is given (containing X categories), constituted with p-feature patterns y (classes). Let all p features of pattern generate a whole original feature set $T_{all} = \{t_1, t_2, \ldots t_p\}$. An optimal feature selection is a process of searching for a subset $T_{sub} = \{t_1, t_2, \ldots t_q\}$ ($T_{sub} \subseteq T_{all}, q < p$) under given a type of criterion, which guarantee better result of classification. Generally a image usually consists of many different classes. For example an image can be classified according to color, size, etc. Hence let assume that there is a family of indicernibility relation $I = \{I_1, I_2, \ldots I_n\}$ over the universe U, which is equal as a whole original feature set $T_{all} = \{t_1, t_2, \ldots t_p\}$. In rough set, if minimal subset $I^{'}$ I of I can determine knowledge about the universe, $\bigcap I = \bigcap I^{'}$ will be called a reduction of I, where $\bigcap I'$ is equal as a subset feature $T_{sub}=\{t_1,t_2,\dots t_q\}$. Hence, the process of features selection is that minimal subset attribute is fined. In this paper, we build firstly the decision table of the rough sets. The following content induce mainly the concept about the decision table and reduction.

The rough sets theory deals with information represented by a table called an information system or decision table. This table consists of objects and attributes. The entries in the table are the categorical values of the features and possible categories.

The decision table of feature selection can be designed as

$$DT=<U,C\cup D,V,f> \tag{2-1}$$

where U is a finite set of N objects $\{x_i,x_i,\dots x_N\}$,Q is a finite set of attributes, C is a set of condition attributes or feature attributes, D is a set of decision attributes or categories attributes, $Q\subseteq C\cup D$. $v=\bigcup_{q\in C\cup D}V_q$,where V_q is a domain (value)of the attribute $q\in C\cup D$, and $f:U\times(C\cup D)\rightarrow V$ is a total decision function(information function, decision rule in DT) such that $f(x,q)\in V_q$ for every $q\in C\cup D$ $x\in V$.

For a given decision table DT , a given subset of attributes $A\subseteq Q$ determines the approximation space $AS=(U,IND(A))$ in DT. For a given $A\subseteq Q$ and $X\subseteq U$,the A-lower approximation $\underline{A}X$ of set X in AS and A-upper approximation $\overline{A}X$ of set X in AS are defined as follows:

$$\underline{A}X=\{x\in U:[x]_A\subseteq X\}=\bigcup\{Y\in A^*:Y\subseteq X\} \tag{2-2}$$

$$\overline{A}X=\{x\in U:[x]_A\cap X\neq\emptyset\}=\bigcup\{Y\in A^*:Y\cap X\neq\emptyset\} \tag{2-3}$$

where A^* denotes the set of all equivalence classes of IND(A). The process of finding a smaller set of attributes than original one with same classificatory power as original set is called attribute reduction. A reduction is the essential part of an information system which can discern all objects discernible by the original information system. A core is a common part of all reduces. Given an decision table DT condition and decision attributes $Q=C\cup D$,for a given set of condition attributes $A\subset C$ we can define a positive region $POS_A(D)$ in the relation IND(D),as

$$POS_A(D)=\bigcup\{\underline{A}X\mid X\in IND(D)\} \tag{2-4}$$

The positive region $POS_A(D)$ contains all objects in U which can be classified without error into distinct classes defined by IND(D) based only on information in the relation IND(A).

For a decision table DT and a subset of attributes $A\subseteq C$ an attribute $a\in A$ is called dispensable in the

NO	CON$_1$	CON$_2$	...	CON$_p$	DES
SAM$_1$	Value$_{11}$	Value$_{12}$	...	Value$_{1p}$	Class$_1$
SAM$_2$	Value$_{21}$	Value$_{22}$	...	Value$_{2p}$	Class$_2$
SAM$_3$	Value$_{31}$	Value$_{32}$	...	Value$_{3p}$	Class$_3$
SAM$_4$	Value$_{41}$	Value$_{42}$	...	Value$_{4p}$	Class$_4$
SAM$_5$	Value$_{51}$	Value$_{52}$	...	Value$_{5p}$	Class$_5$

(a)

NO	1	0	...	0	DES
SAM$_1$	Value$_{11}$	Value$_{12}$	...	Value$_{1p}$	Class$_1$
SAM$_2$	Value$_{21}$	Value$_{22}$	...	Value$_{2p}$	Class$_2$
SAM$_3$	Value$_{31}$	Value$_{32}$	...	Value$_{3p}$	Class$_3$
SAM$_4$	Value$_{41}$	Value$_{42}$	...	Value$_{4p}$	Class$_4$
SAM$_5$	Value$_{51}$	Value$_{52}$	...	Value$_{5p}$	Class$_5$

- Genetic algorithm
- Fuzzy C-Mean
- Continuous data
- Discrete attribute

(b)

NO	CON$_1$	...	CON$_q$	DES
SAM$_1$	Value$_{11}$	...	Value$_{1q}$	Class$_1$
SAM$_2$	Value$_{21}$	...	Value$_{2q}$	Class$_2$
SAM$_3$	Value$_{31}$	...	Value$_{3q}$	Class$_3$
SAM$_4$	Value$_{41}$	...	Value$_{4q}$	Class$_4$
SAM$_5$	Value$_{51}$	...	Value$_{5q}$	Class$_5$

Texture classification

(c)

Figure 1. An example of relative reduction of the decision table. (a)original decision table. (b) the middle result. (0 and 1 denotes the irrelevant and relevant features respectively) (c) a relative reduction of the decision table.

set A if IND(A)=IND(A-{a}). The set of all indispensable attributes in the set $A \subseteq C$ is called a core of A in DT and it is denoted by CORE(A).

Above are some important concepts of the rough sets. In this paper, we utilize the decision table (DT) and reduction based on DT. Because the problem reducing the decision tables is a difficult optimization problem, this paper employs genetic algorithm to search the relative reduct of decision table or the effective feature subset. Figure1 shows the procedure of the decision table reduced. The following section describes the algorithm of reducing decision table or feature selection.

3. Algorithm

Feature selection techniques generally involve both a search algorithm and a criterion function. This paper employs genetic algorithm to search the effective feature subset and the generalization accuracy of classification is taken as the criterion function, namely the fitness function. Some recent attempts in applying GAs for feature selection are reported [6,7]. The traditional feature selection algorithm was improved using genetic algorithm [7]. Our algorithm was inspired by the method [6],which offered an approach to feature subset selection for neural network pattern classifier. In this paper, we use genetic algorithm to find out the effective reduct decision table of the rough set based on continuous or discrete attributes .

The main issues in applying GAs to any problem are selecting an appropriate representation, an adequate evaluation function, and defining a suitable genetic operation. These issues will be described as follows.

3.1 Encoding scheme

Now a set of individuals are defined in a population generated during t generation cycles. $P(t)=\{I_k| k=1,2,\ldots m\}$, where m is the number of individuals or the population size. The size affects both the ultimate performance and the efficiency of GAs. Each individual is generated by some encoded form known as a chromosome. Here a chromosome is the vector consisting of a single bit for each feature, with a 1 indicating that the feature participates in classification, and a 0 indicating that it is omitted.

3.2 Crossover and mutation

The operations, namely, crossover and mutation, are then following the encoding scheme. A crossover position is randomly chosen in the encoded string. The crossover operation applied to two parents produce offsprings. Mutation is carried out by selection a value 0 or 1 on a randomly selected positions of a parent string.

3.3 Fitness function

The fitness function is used to evaluate the goodness of a chromosome (solution). In this study, the fitness function has to combine two different criteria – the accuracy of the classification function realized by the Fuzzy C-Mean method and the cost of performing classification. The accuracy of the classification function can be estimated by calculating the correct rate of each reduct decision table using the Fuzzy C-Mean method. The measures of the cost of classification suggest the number of condition attributes in the decision table needed for classification. Here, we choose a relatively simple form of a 2-criteria fitness function defined as follows:

$$Fitness(i) = \frac{correct\_rate(i)}{1+\lambda * num(i)} \qquad (3\text{-}1)$$

Where $Fitness(i)$ is the fitness of the feature subset represented by individual i, $correct\_rate(i)$ is the test accuracy of the fuzzy c—mean classifier using the feature subset represented by i, Because the cost of classification is in proportion to the number of the selected features, $num(i)$ is set to the number of the condition attributes or features of subset represented by individual i. In additional, the parameter the λ is the weight of the number of features and discussed in detail in the experiment. Obviously, the greater the value of fitness, the better the performance of the selected feature subset. The following contents describe that the effective feature subset is used in texture classification.

4. Texture Classification Based the Feature Subset Selected

We have applied feature selection for the purpose of texture classification in color aerial images. In this paper, Trees or forests are interested objects. Trees or forests are a kind of natural scenes which are not structured and cannot be represented easily by regular rules. Texture and color features are important cues for trees extraction in color aerial images. In practice, if only texture and color features are used ,the results of trees extraction are inaccurate. Our method combines texture features, color and height information to overcome these disadvantages. A technique similar to ours was discussed in [8], which dealt with realistic and thus more complex scenes. But it needed high resolution aerial images and DTM. In this paper, low resolution aerial images and digital elevation model (DEM) are used in trees extraction. At first ,according to the DEMs, the original color aerial images are segmented into the high and low objects. High objects include trees or forest, houses , bridges and so on. In order to refine trees or forest, the high objects are classified by Fuzzy C-Mean based on the effective feature subset. The procedure is described as following.

We start with DEM data an automatically generated by the digital photogrammetry system --Virtuozo. The resolution of the images by which DEM data are obtained may be lower than original images. According to following algorithm , high and low objects are obtained.

Step 1. DEM are mapped to range (0—255 gray level) in order to form the image of DEM. As a result, different gray levels denote different elevation.

Step 2. With the DEM image, original image is divided into many regions with same size.

Step 3. The edges in the DEM image are extracted by Sobel algorithm. These edges reflect the local changes of elevations of objects. According to the edge image, we compute the segment threshold for each region.

Step 4. According to the threshold of region, the original image is segmented into a binary image, in which 1 represents high objects and 0 represents low objects. The high objects include trees, houses ,bridges and so on.

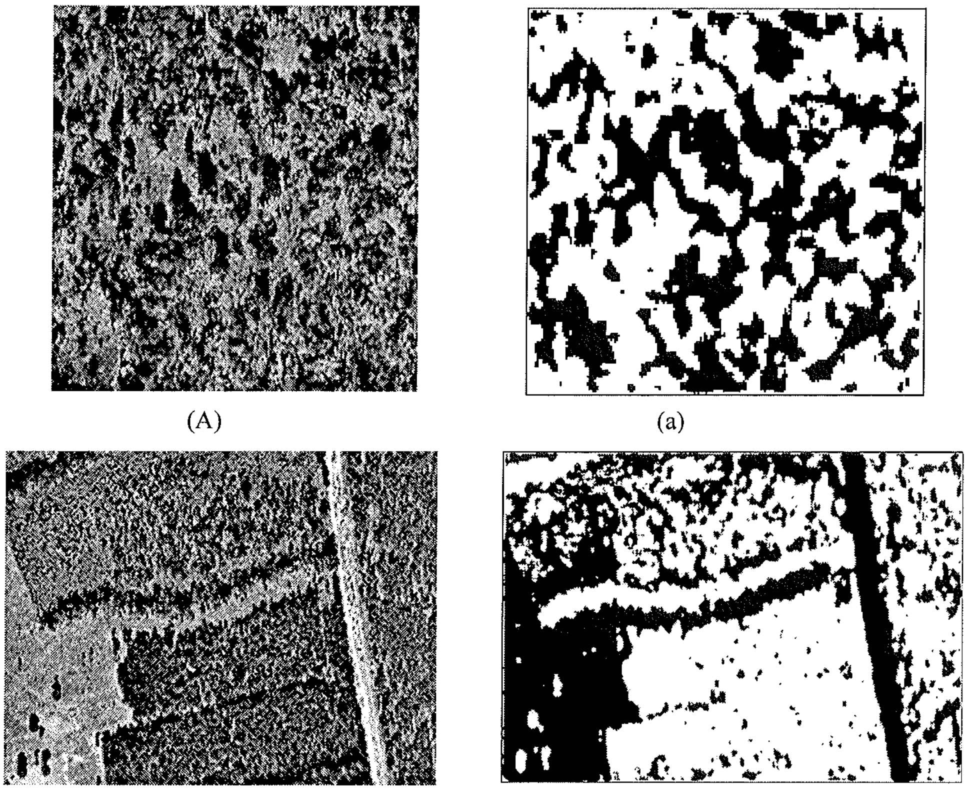
The next step is to refine trees from high objects by Fuzzy C-Mean clustering based the effective feature subset selected. Fuzzy C-Mean clustering algorithm (FCM) was introduced by J. C. Bezdek [9]. In this paper , FCM is used in three different processing procedures :

- The continuous data are discretezed by FCM.
- The samples are classified by FCM in the reduction of decision table.
- Trees are extracted by FCM in the high objects.

In above three procedures, the classified objects are different. But basic algorithm is the same. The principle of Fuzzy C--Mean will be discussed in detail in [9].

5. Experimental Results

In our experiment , 12 color aerial images are used. The photography scale of color aerial images is 1:8000. The principle focal is 152.987mm. The scanning resolution is 96 μm. The photo size is 23cm ×23cm. A total of 200 samples are selected, which include density trees, sparse trees, houses, roads, grass ,river and ground. At first, the color aerial images samples in the RGB space are converted into in the HIS (hue, saturation, and intensity) space. In the intensity of the color aerial images samples , a total of 11 texture features per pattern (pixel)[10] are computed in our experiments. In addition, five color features are



(A) (a) (B) (b)

Figure 2. (A) and (B) are original color aerial images; (a) and (b) are the results of classification, where white area denote trees and black area non-trees.

Table 1. Comparison between the results of two different algorithm for feature subset selection.

Data	Parameter λ	Feature selected	Correct rate
Continuous attributes	0.005	7	97.2%
	0.01	6	98.3%
	0.05	5	92.3%
	0.1	4	88.3%
Discrete attributes	0.005	7	96.1%
	0.01	6	97.2%
	0.05	5	93.4%
	0.1	4	88.3%

used, which include hue , saturation, I_1, I_2 and I_3.

The original data is represented by the 200×17 decision table based on continuous attributes, where 200 is the number of training samples and 16 condition attributes or features are selected. Besides, there is one decision attribute or class information in the table. We use two different algorithms to find out the effective feature subset. The first algorithm uses directly the decision table based on the continuous data. All the values of features are normalized to the range from 0 to 1. In the second algorithm , the continuous data are discretized firstly by Fuzzy c-means, where the classifying number of each feature is the two . On the basis of the decision table based on the discrete data, the effective feature subset is searched by the genetic algorithm. The results of two methods are listed in Table 1. There are different four weights: λ=0.1,0.05,0.01,0.005 and different correct rates are obtained. Because the correct rates are more important than the classifying cost ,the parameter λ is 0.01 in this paper. The comparison results show that the continuous attributes are the same effective as the discrete attributes.

In addition, the parameters used in the genetic algorithm are listed as follows:

- Population size: 50 ;
- Number of generation: 500;
- Probability of crossover: 0.9;
- Probability of mutation: 0.5;

6. Conclusions

This paper describes an approach to find out the effective feature subset based on the decision table of the rough set using genetic algorithm. Generally, before the decision tables are reduced, the continuous attributes of the decision table are discretezed firstly. In this paper, the decision tables based on the continuous data and discrete attributes are reduced using genetic algorithm. The results of comparison between the two algorithms show that continuous attributes are same useful as discrete data. Our application is in the domain of aerial images. The values of texture features are continuous data. The symbol rules are not effective in the classification of aerial images. Besides, on the basic of the effective feature subset selected , this paper presents an approach to extract the objects which are higher than their surroundings ,such as trees or forests. The experiment results show that the feature subset selected and the method of classification are effective and practical.

In practice, due to remote sensing images have more information than the aerial images, trees or forest are extracted in remote sensing images. Although we have dealt with color aerial images, the technique is extensible to multispectral images. But ,the geometric resolution of remote sensing images is lower than aerial images, the rate of trees recognition will decrease. In addition, it is difficult for remote sensing images to obtain quickly the height information. Hence, the algorithm of extracting trees should be improved further.

References

[1] Z. Pawlak, Rough Set Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht, Boston, London, 1991.

[2] R. W. Swiniarski, L.Hargis, Rough sets as a front end of neure-networks texture classifiers, Neurocomputing 36, pp.85-102, 2001.

[3]Hoe Kok Meng, A Multi-Mechanism Pipeline for Generating Symbolic Rules From Un-Annotated Data, Proc. NSF Workshop, Kuala Lumpure,pp.1-6, 2001.

[4]Suresh K. Choubey, Jitender S. Deogun ,Vijay V. Raghavan, On Feature Selection and Effective Classifiers, Journal of Asis 49, 5 ,pp.423-434,1998.

[5]P. Lingras, Unsupervised Rough Set Classification Using Genetic Algorithm, Journal of Intelligent Information Systems, 16, pp.215-228, 2001.

[6] Jihoon, Yang, Vasant, Honavar, Feature subset selection using a genetic algorithm. Journal of Intelligent Information Systems, 16, pp.215-228,2001.

[7]H. Vafaie, K. D. Jong, Robust Feature Selection Algorithms. International Conference on Tools with AI, Boston, Massachusetts, pp.356-363,1993.

[8] W. Eckstein, C. Steger, Fusion of digital terrain model and texture for object extraction. Proceeding of 2nd Airborne Remote Sensing Conference. San Francisco, pp.1-10,1996.

[9]J. C.Bezdek, Convergence Theory for Fuzzy c-Means: Counterexamples and Repairs, IEEE Trans. Syst., September/October, 1987.

[10] L. Pan, The Study on Forest Area Recognition

from Color Aerial Image and Its Application for Automatic Aerial Triangulation, Ph.D. thesis, Wuhan University,2001.