The Reduction Subset Based on Rough Sets Applied to Texture Classification

Li Pan[1,2], Saeid Nahavandi[1], and Jinhui Lan[1,3]
[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
[3]Department of Precision Instruments, Tsinghua University
Beijing 100084, P. R. China
panli@deakin.edu.au

Abstract

The rough set is a new mathematical approach to imprecision, vagueness and uncertainty. The concept of reduction of the decision table based on the rough sets is very useful for feature selection. The paper describes an application of rough sets method to feature selection and reduction in texture images recognition. The methods applied include continuous data discretization based on Fuzzy c-means and , and rough set method for feature selection and reduction. The trees extractions in the aerial images were applied. The experiments show that the methods presented in this paper are practical and effective.

1. Introduction

Feature selection is the problem of choosing a small subset of features that is necessary and sufficient to describe target concept. The importance of feature selection is due to the potential for speeding up the processes of both concept learning and classifying objects, reducing the cost of classification, and improving the quality of classification. Feature selection has long been the focus of researchers of many fields – pattern recognition, image understanding and machine learning. In general, the methods that have existed can be classified into two categories: (1) the filter or open-loop approach [1] do not consider the effect of selected features on a whole processing algorithm performance. (2) the wrapper or closed-loop methods [2] are based on feature selection as a wrapper around a classifying algorithm relying on which the relevant attributes are determined. Although the wrapper approach has certain advantages, it is not as general as the filter approach. The approach to feature subset selection proposed in this paper is an instance of filter approach. It utilizes a rough sets for feature subset selection.

The work presented here was motivated by our experience in using conventional feature selection algorithms for difficult image understanding problems involving texture classification. In the case there can be a lot of features and complex interactions among the features. On such problems conventional features selection algorithms can be used ,but there are two problems. The first, it is difficult to analyze, discover and generate the rules which are important for object classification. The second, some methods need special conditions. In order to solve the problems, this paper employs rough sets to implement feature selection. The rough set is a new mathematical approach to imprecision, vagueness and uncertainty [3,4]. The concept of reduction of the decision table 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 process of feature selection and classification using the rough set was studied in some papers [6,7,8]. In their approaches, conventional algorithms and the rough set concept were combined in order to improve the accuracy of classification and reduce time costing, Then, on the basis of the attributes reduced in the decision table, this paper presents a new algorithm to select the optimal image feature, which is more direct and simpler than above methods. In the real world, many feature values are continuous data, which are proven to be rather unsuitable for the extraction of concise symbolic rules. Also, rule conditions that comprise of singular continuous values have poor predictions [7]. Hence, original data have to be discretely normalized firstly. Focusing on the problem, this paper describes a discretization algorithm of values of features based on Fuzzy C-Means cluster. The technique is illustrated by its application to the problem of trees extraction from aerial images.

2. Overview of Rough Sets

The rough sets theory has been developed for knowledge discovery in databases and experimental data sets. The rough sets theory based on the concept of an upper and a lower approximation of a set, the approximation space and probabilistic and deterministic models of sets. The rough sets theory deals with information represented by a table called an information system. 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

$$DT = < U, C \cup D, V, f > \qquad (2\text{-}1)$$

where U is a finite set of N objects $\{x_i, x_i, \ldots x_N\}$,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 $q \in C \cup D \quad x \in V$.

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:

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

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

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

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

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 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. The following section describes two algorithms : continuous data discretization and feature selection based on rough sets.

3. 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 $T_{sub} = \{t_1, t_2, \ldots t_q\}$. Hence, the process of features selection is that minimal subset attribute is fined. In this paper, we discretized firstly the decision table based on continuous attributes. On the basis of the decision table based on discrete attributes , the effective feature subset is selected using rough sets.

3.1 The continuous data discretization based on Fuzzy C-means cluster

The process of convening data sets with continuous attributes into input data sets with discrete attributes is called discretization. Generally, the data of texture features of aerial images are continuous data. Henceforth, the necessity to discretize continuous data to discrete intervals, where each interval can be represented by a label. Disretization not only reduces the complexity and volume of dataset, but also serves as an attribute filtering mechanism. In our work, we employed discretization method based on Fuzzy C-Means which is a sort of supervised algorithm. The following is the concrete algorithm:

Algorithm1: Continuous data discretization based on Fuzzy C-Means clustering

Given: N texture features values of M samples.

Step 1: Given the number (H) of category the each set of texture features are classified.

- Grouping feature values according to their fuzzy membership by region yields.

$$U=\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ u_{c1} & u_{c2} & \cdots & u_{cn} \end{bmatrix} \qquad (3\text{-}1)$$

- To determine the feature value membership as follows:

$$J_m(U,v,X)=\sum_{i=1}^{c}\sum_{k=1}^{n}(m_{ik})^m D_{ik} \qquad (3\text{-}2)$$

Step 2: Continuous attribute values (texture feature values) are discretized to discrete interval.

- The each discrete interval can be denoted by a label (1,2,...N).
- Merge the same rows in the symbol table.

Step 3: Get symbol table.

The continuous values of features have discretized into symbol table. When category attributes or decision attribute is added to the symbol table, we obtain the decision table which is necessary for our algorithm.

3.2 Algorithm

Although there are many algorithms of rough set theory used to reduce, they are the same in that by the dependent properties of attributes, find a reduced set of the attributes, providing by removing superfluous attributes, without losses in classification power or the reduced information system. The algorithms can be divided into two kinds. One is attribute reductionism based on of algebraic set . It has some advantages such as clear steps and distinct meaning, but is not very efficient under cases. The other is reasoning method being based upon logical operation, which is easy to carry out and give a fast algorithm, but is abstract. Hence, on the basis of the above two methods, we describe the algorithm of image feature selection based on the rough set , in which the decision table is made up of a group of formulas that can be processed by logical adjusting. By judging formulas or rules whether they are contrary to each other or not , we decide the compatibility of the decision table. On the other hand , if there are conflict of formulas or rules when an attribute is to be omitted, the decision table is not compatible and the attribute is not eliminated . Generally speaking ,there are various noises in our data, so it is unrealistic, in most case, to expect that the decision rules obtained from the data. Therefore, if less than 15% rules are contrary, the decision table is still compatible. The following is concrete algorithm.

Algorithm 2: Feature selection using rough set

Given: $DT=<U,C\cup D,V,f>$
where U is the universe, C is a set of condition attributes, D a set of decision attributes, $v=\bigcup_{q\in C\cup D}V_q$,$V_q$ is is a domain (value)of the attribute, $f:U\times(C\cup D)\rightarrow V$ is total decision function.

Step 1. $DX\in U\times C\rightarrow V$ is a rule of a decision table (DT) based condition attribute C . $DX\in U\times D\rightarrow V$ is a rule of a decision table (DT) based decision attribute D.

- For more than 85% rules: if i≠j DX(i)=DY(j) and DY(i)=DX(j) the decision table is compatible, then go to step2.

- For more than 85% rules: if i≠j and DX(i)=DX(j) but DY(i) ≠DY(j), then the decision table: $DT=<U,C\cup D,V,f>$ is not compatible, calculate $POS_A(D)=\bigcup\{\underline{A}X|\ X\in IND(D)\}$, $A\subseteq C$ go to step3.

Step 2: When r (r⊆C) is eliminated for more than 85% rules: if i≠j, DX(i)=DY(j) and DY(i)=DX(j), then r(r⊆C) (is redundant attribute (IND(C)=IND(C-{r})), otherwise, it is reserved go to step2 again. Do this until all attributes are finished.

Step 3: if $POS_{A-\{r\}}(D)=POS_A(D)$,r (r⊆A)can be omitted, otherwise, it can not be omitted , repeat step3 until all attributes are completed.

Step 4: Obtain $DT_{sub}=<U,C'\cup D,V',f'>$ $C'\subseteq C$ $V'\subseteq V$ $f'\subseteq f$.

4. Experiments

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. In our experiment , 12 color aerial images are used. The photography scale of from different aerial images. 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

Table 1. The set of 16 texture and color features are used in this paper

NO	Feature	Model
1	Mean	Local statistics
2	Standard deviation	Local statistics
3	Skewness	Local statistics
4	Kurtosis	Local statistics
5	Contrast	CLCM
6	Entropy	CLCM
7	Inertia	CLCM
8	Energy(E5L5)	LTT
9	Energy(E5R5)	LTT
10	Energy(E5S5)	LTT
11	Energy(L5S5)	LTT
12	I_1	RGBSCM
13	I_2	RGBSCM
14	I_3	RGBSCM
15	hue	HIS
16	saturation	HIS

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) are computed. In addition, 5 color features are used, which include hue , saturation, I_1, I_2 and I_3. Table 1 lists all features used in our experiments.

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. The continuous data are discretized firstly by **algorithm1**, 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 **algorithm2** .

The one of original color aerial images used in the experiment of trees extraction is shown in Figure 1 (A). On the basis of the effective feature subset selected, trees are extracted by Fuzzy c-means with the effective feature set in color aerial image. The results are shown in Figure 1 (B). We did experiments on 12 color aerial images using the reduced feature subset. Fuzzy c-means cluster method is employed to extract trees from aerial images. The average recognition rate can reach 80%, while the recognition rate based on original feature set is 71%, and the time costing is large (details see[9]).

5. Conclusions

The results presented in this paper indicate that the reduction algorithm based on the rough set offer an attractive approach to solving the feature subset selection problem in texture classification of aerial image. Because texture feature values are continuous data, we have proposed the discretization algorithm based on Fuzzy C-Means and get the decision table. On the basis of the decision table based on discrete attributes. we have selected the subset reduced using feature selection using rough set. The experiment results prove that the features subset selected by our algorithm is effective. Our further work is extensible to remote sensing images.

References:

[1] H.Liu, R. Setiono, A Probabilistic Approach to Feature Selection—A Filter Solution. Machine Learning : LCML'96, Bari, Italy, pp.319-327, 1996.

[2] Jihoon Yang, Vasant Honavar, Feature Subset Selection Using a Genetic Algorithm. Journal of Intelligent Information Systems, Vol. 16, pp.215-228, 2001.

[3] H. Vafaie, K. D. Jong, Robust Feature Selection Algorithms. In: Proc. of International Conference on Tools with AI, Boston, Massachusetts, pp.356-363, 2001.

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

[5] Z. Pawlak, Rough Set Approach to Knowledge-Based Decision Support, European Journal of Operational Research , pp.48-57,1991.

[6] R. W. Swiniarski, Rough Sets as A Front End of Neural Networks Texture Classifiers, Neurocomputing , pp.85-102, 2001.

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

[8] M. R. Chmielewski, Global Discretization of Continuous Attributes as Preprocessing for Machine Learning, Int. Journal of Approximate Reasoning, pp.319-331, 1996.

[9] 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.

(A)

(B)

Figure 1. An example of trees recognition from aerial image based on the reduced subset. (A) original color aerial image. (B) The result of trees extraction (white area denote trees and black area denote non-trees);