This is the published version (version of record) of:


Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30009618

Copyright: ©2003 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.
THE APPLICATION OF ROUGH SET AND KOHONEN NETWORK TO FEATURE SELECTION FOR OBJECT EXTRACTION

LI PAN¹, HONG ZHENG², SAEID NAHAVANDI²

¹School of Remote Sensing and Information & Engineering, Wuhan University
129 Luoyu Road, Wuhan 430079, P.R.China
²School of Engineering and Technology, Deakin University, Geelong VIC3217, Australia
E-MAIL: panli@deakin.edu.au or zhenghong@21cn.com

Abstract:
Selecting a set of features which is optimal for a given task is a problem which plays an important role in a wide variety of contexts including pattern recognition, images understanding and machine learning. The paper describes an application of rough sets method to feature selection and reduction in texture images recognition. The proposed methods include continuous data discretization based on Kohonen neural network and maximum covariance, and rough set algorithms for feature selection and reduction. The experiments on trees extraction from aerial images show that the methods presented in this paper are practical and effective.

Keywords:
Rough set; Feature selection; Kohonen network

1 Introduction

Feature selection is the problem of choosing a small subset of features that is necessary and sufficient to describe target concept. It plays an important role in a wide variety of contexts including pattern recognition, image understanding, and machine learning. 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. 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 problem is that it is difficult to analyze, discover and generate the rules which are important for object classification. The second problem is that 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. 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. Some approaches of feature selection and classification using the rough set were published in some literatures. In these approaches, conventional algorithms and the rough set concept were combined in order to improve the accuracy of classification and reduce time costing. In this paper, a new improved algorithm is presented 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. Hence, original data have to be discretely normalized firstly. Focusing on the problem, this paper describes the discretization algorithm of values of features based on the Kohonen neural network and MAXCOV(maximum covariance). The approach is illustrated by its application to the problem of trees extraction from aerial images.

This paper is organized as follows: The continuous data discretization based on Kohonen neural network and MAXCOV is presented in Section 2. Section 3 is divided into two subsections. The first subsection overviews the rough set theory. The second subsection describes the algorithm of feature selection. In Section 4, experimental results are discussed, and the conclusion is given in Section 5.

2 The continuous data discretization based on Kohonen neural network

The process of converting 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. Discretization not only reduces the complexity and volume of dataset, but also serves as an attribute filtering mechanism. In recent years, many algorithms were presented, such as minmax entropy, the minmax error, minmax entropy and K-Means clustering [5]. The minmax entropy and the minmax error are supervised algorithms known as supervised K-Means clustering. The methods of equal interval and equal frequency are the simplest, but they are only used when the data are in uniform distribution. K-Means clustering is known as unsupervised learning and the K-Means clustering is unsupervised algorithm.

Let $S$ be a continuous data and let the domain of $S$ be the interval $[p, q]$. A partition $\varphi$ on $[p, q]$ is defined as the following set of $m$ subintervals

$$\varphi = \{ [p_0, p_1), [p_1, p_2), ..., [p_{m-1}, p_m) \}$$

where $p_0 = p$, $p_{i-1} > p_i$ for $i = 1, 2, \ldots, m$, and $p_m = q$. Thus, discretization is the process that produce a partition $\varphi$ on $[p, q]$.

In this paper, a partition is generated by the algorithm of Self-Organizing Feature-Mapping (SOM), in which interactions among competitive processing elements could be used to construct a network that could classify clusters of input vectors. When the interval is the partition, a suitable post-processing using maximum covariance is performed to select the best partition which has minimum interior variance and maximum internal variance. The following is the concrete algorithm:

**Algorithm1:** Continuous data discretization based on SOM

**Given:** $N$ texture features values of $M$ samples.

**Step1:** Given the number of category, each of the $N$ texture features is classified.

- Set the initial weight, which is a little random variable.
- Calculate the distance of the input and each output node.
- Adjust the join weight of the nodes with the minimum distance.

**Step2:** Continuous attribute values (texture feature values) are discretized to discrete interval.

- For each interval, calculate the within-groups and between-groups covariance.
- If the number of the classification is less than $m$ and the zero set appears, stop classifying.

- Select the best number of category in minimum within-groups and maximum between-groups covariance for each attribute.

**Step3:** Get symbol table.

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

### 3 Feature selection based on rough set

The continuous values of features have been 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.

#### 3.1 Rough set

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.

Some of the information system can be designed as a decision table

$$DT = \langle U, C \cup D, V, f \rangle$$

where $U$ is the universe, a finite set of $N$ objects $\{x_1, x_2, \ldots, x_N\}$, $C$ is a set of condition attributes, $D$ is a set of decision attributes, $V = \bigcup_{q \in C \cup D} V_q$, where $V_q$ is a domain (value) of the attribute $q \in Q$, 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 $q \in Q$, $x \in U$.

For a given information system $S$, a given subset of attributes $A \subseteq Q$ determines the approximation space $AS = (U, IND(A))$ in $S$. For a given $A \subseteq Q$ and $X \subseteq U$, the $A$-lower approximation $AX$ of set $X$ in $AS$ and $A$-upper approximation $AX$ of set $X$ in $AS$ are defined as follows:

$$AX = \{ x \in U : [x]_A \subseteq X \} = \bigcup \{ Y \in A^* : Y \subseteq X \}$$

$$AX = \{ x \in U : [x]_A \cap X \neq \emptyset \} = \bigcup \{ Y \in A^* : Y \cap X \neq \emptyset \}$$

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 information system S
condition and decision attributes \( Q = C \cup D \), for a given set
of condition attributes \( A \subseteq C \) we can define a positive
region \( POS_A(D) \) in the relation \( IND(D) \), as
\[
POS_A(D) = \bigcup \{ DX | X \in IND(D) \}
\]
(5)

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 an information system \( S \) 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 \( S \), and it
is denoted by \( CORE(A) \).

### 3.2 Algorithm

The feature selection is an important step in design of
classification. Let us assume an image \( m \times n \) 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.
In this paper, the criteria of feature selection is described
using concept of reduction defined by theory of rough sets.

Generally there are many classification patterns of
images. For example an image can be classified according to
color, size, etc. Hence let assume that there is a family of
indiscernibility 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 \( I' \) 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.

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
categories. 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 present the algorithm for image feature
selection based on 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 based on rough set**

**Given:** \( DT = \langle U, C \cup D, V, f > \)

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

**Step1:** \( DX \cup U \times C \rightarrow V \) 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 \( D \).

1. For more than 85% rules: if \( i \neq j \) and \( DX(i)=DY(j) \) and
   \( DY(i)=DX(j) \), the decision table is compatible, then go to step 2.
2. For more than 85% rules: if \( i \neq j \) and \( DX(i)=DX(j) \),
   but \( DY(i) \neq DY(j) \), then the decision table:
   \( DT = \langle U, C \cup D, V, f > \) is not compatible, calculate
   \( POS_A(D) = \bigcup \{ DX | X \in IND(D) \} \), \( A \subseteq C \) go to step 3.

**Step2:** When \( r (r \subseteq C) \) is eliminated, for more than
85% rules: if \( i \neq j \), \( DX(i)=DY(j) \) and \( DY(i)=DX(j) \), then
\( r \subseteq C \) (is redundant attribute ( \( IND(C) = IND(C-\{r\}) \) ),
otherwise, it is reserved. go to step 2 again. Do this until
all attributes are finished.

**Step3:** if \( POS_A-\{r\}(D) = POS_A(D) \) \( \forall (r \subseteq A ) \) can be
omitted, otherwise, it can not be omitted , repeat step 3
until all attributes are completed.

**Step4:** Obtain \( DT_{sub} = \langle U, C' \cup D, V', f' > \) \( C' \subseteq C \)
\( V' \subseteq V \) \( f' \subseteq f \).
4 Experiments

In our experiment, a total of 200 samples were used. These samples include density trees, sparse trees, houses, roads, grasses, rivers and lands which come from different aerial images. Being limited of space, only 20 samples are listed in Table 1. A total of 11 texture features per pattern (pixel) were computed. In the first row of Table 1, a, b, ..., k represents the mean of gray, variance of gray, skewness of gray, extrusion degree of gray, energy, absolute value of gray, Law's energy template1, Law's energy template2, Law's energy template3, Law's energy template4, adaptive energy template[6], respectively. Because the values of texture features or attributes are continuous data, it is necessary to discretize all continuous data by the Algorithm 1 and generate symbol decision table—Table 2. Because the same rows are merged, there are different row numbers in Table 2, in which category information or decision attribute listed in twelfth column (D) that has four classes (dense trees, sparse trees, house, other).

Depending on the decision attributes or categories of samples, the decision Table 2 is reduced by Algorithm 2. At last, we obtain the reductionism decision table, where a and k is and D is decision attribute. In other word, the texture feature set \{a, k\} is minimal set of the original feature set.

We have applied feature selection for the purpose of trees object extraction in aerial images. An example of trees extraction from aerial images is shown in Figure 1, where Fig 1(a) shows a original aerial image and the results of trees extraction are shown in Fig 1(b). The effective texture feature set \{a, k\} is used to classify each pixel of aerial images into one of two classes (trees and non-trees). Figure1(b) shows extraction results, in which black and white area denotes non-trees and tress respectively.

We did experiments on 100 aerial images using the reduced feature subset. The Fuzzy 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[7]).

5 Conclusion

The results presented in this paper indicate that the reduction algorithm based on the rough set offer an attractive approach to solve the feature subset selection problem in object extraction from aerial image. Because texture feature values are continuous data, we have proposed the discretization algorithm based on Kohonen network and got the decision table (Table 2). On the basis of Table 2, we have selected the reduced subset using our Algorithm 2. The experiment results prove that the feature subset selected by our algorithm is effective. Our further work is to combine rough sets and genetic algorithms to solve the problem of mass feature selection.

Figure 1. An example of trees recognition from aerial image based on the reduced subset. (a) An original aerial image (23×23cm). (b) The result of trees recognition (white colour denotes trees).

Acknowledgements

The research was supported by National Nature Science Funding of China under grant No.40271094. This support is gratefully acknowledged.

References:


Table 2. Decision table (D—decision attribute)

<table>
<thead>
<tr>
<th>No</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>g</th>
<th>h</th>
<th>i</th>
<th>j</th>
<th>k</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3. Reduced decision table

<table>
<thead>
<tr>
<th>No</th>
<th>a</th>
<th>k</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>