ANALYSIS OF HYPERSPECTRAL VEGETATION INDICES BY ROUGH SET THEORY

Kuan-Tsung CHANG
Assistant professor
Dept. of Civil Eng., MUST, No. 1, Hsin-Hsing Rd., Hsin-Fong, Hsin-Chu 304, TAIWAN
ktchang@must.edu.tw

Edward WANG
Assistant professor
Dept. of Civil Eng., MUST, No. 1, Hsin-Hsing Rd., Hsin-Fong, Hsin-Chu 304, TAIWAN
ewang@must.edu.tw

Ge-Wen LEE
Assistant professor
Dept. of Civil Eng., CCIT, No. 190, Sanyuan 1st St., Tashi, Taoyuan 335, TAIWAN
bearlee@ccit.edu.tw

Jin-King LIU
Senior Researcher,
Energy and Resource Laboratories, Industrial Technology Research Institute
Hsin-Chu 311, TAIWAN
jkliu@itri.org.tw

Abstract: The analysis of vegetation types and detection of changes in vegetation patterns are keys to natural resource management and monitoring. Thus the detection and quantitative assessment of green vegetation is one of the major applications of remote sensing for environmental resource management and decision-making. The aim of this study is to analyze a series of slope-based vegetation indices (VIs) calculated from the hyperspectral data, received from the AVIRIS sensor, by using the Rough Set theory (RST). The RST, an analysis technology for the knowledge mining based on incomplete information, is used in this paper to extract the decision rules and significant VIs for land use/cover classification. The implicit knowledge and properties of slope-based VIs could be further understood by the result of the analyses. It is also shown in this study that RST is helpful for the land use/cover classification of hyperspectral data.
Keywords: Environmental resource management and decision-making, Land use/cover classification, Vegetation Index, Rough Set theory, Hyperspectral data.

1. Introduction

Taiwan located at a high-density developing, limited natural resources area. In many cases, over-expanded development and activities, such as slope cutting, Typhoon, earthquake, and deforestation, can sometimes increase the incidence of natural and human-made disasters. There is more demand for having real-time information, e.g. environmental change, natural resource management, and the condition of a disaster in hand. The vigorous development in the Remote Sensing technology can be use to satisfy the above mentioned demand.

A variety of information is available from sensors carried by both satellite and aircraft platforms. Those operating in the visible and near-infrared region of the electromagnetic spectrum provide measurements the correlate with an object's colour, which is often related to the chemical or mineralogical properties of that object. Data from thermal infrared sensors are related to the temperature and thermal properties of a target, while information about surface roughness and (over the land) moisture content can be derived from data collected in the microwave (radar) wavelengths. The spectral range of remotely-sensed data available in the late 1990s from orbiting satellites covers the full range from optical to microwave.

Satellite remote sensing platforms that are currently providing image data include Landsat-7, Terra, Aqua, Envisat, SPOT, NOAA, ERS, RADARSAT, IRS, the Japanese ADEOS-2, Meteosat and other geostationary meteorological satellite systems, and high-resolution commercial systems such as IKONOS and Quickbird, with 1 m or less spatial resolution. In addition, NASA's earth observing-1 carries an imaging spectrometer, Hyperion, capable of resolving 220

spectral bands (from 0.4 to 2.5 μm) with a 30 m spatial resolution. The instrument images a 7.5 km by 100 km land area per image. Experience gained with airborne imaging spectrometers helps researchers to understand the problems of handling such large data volumes. Potential technical problems include the development of methods of combining multi-source (multi-sensor) data, handling large volumes of high-resolution data, and selecting the optimum combination of bands to use for a particular application.

Healthy canopies of green vegetation have a very distinctive interaction with energy in the visible and near-infrared regions of the electromagnetic spectrum. In the visible regions, plant pigments (most notably chlorophyll) cause strong absorption of energy, primarily for the purpose of photosynthesis. This absorption peaks in the red and blue areas of the visible spectrum, thus leading to the characteristic green appearance of most leaves. In the near infrared, however, a very different interaction occurs. Energy in this region is not used in photosynthesis, and it is strongly scattered by the internal structure of most leaves, leading to a very high apparent reflectance in the near infrared. It is this strong contrast, then, most particularly between the amount of reflected energy in the red and near-infrared regions of the electromagnetic spectrum that has been the focus of a large variety of attempts to develop quantitative indices of vegetation condition using remotely sensed imagery [1-5].

Pawlak [6] introduced *rough set theory* (RST) in the early 1980s as a tool for representing and reasoning about imprecise or uncertain information. Based on the notion of indiscernibility and the inability to distinguish between objects, rough set theory deals with the approximation of sets or concepts by means of binary relations, typically constructed from empirical data. Such approximations can be said to form models of our target concepts, and hence in its typical use falls in under the bottom-up approach to model construction. As the methodology has matured, several interesting applications of the theory have surfaced, also in medicine. For example, in a medical setting, sets of interest to approximate could be the set of patients with a certain disease or outcome, or the set of patients responsive to a certain treatment. Until now, the RST has became an important tool for the knowledge discovery. There are many successful applications on medical diagnosis, decision analysis, machine learning, and information retrieval.

2. Definition of Vegetation analysis

Remote sensing is the science and art of obtaining information about an object, area, or phenomenon through the analysis of data acquired by a device that is not in contact with the object, area, or phenomenon under investigation. The remotely collected data can be of many forms, including variations in force, acoustic wave, or electromagnetic energy distributions [7]. The sunlight is a major resource providing electromagnetic energy of the Earth surface. When it propagates through the atmosphere, then interacts with objects or features on the earth surface. Different interactions will be happened because the material for objects is changed. Therefore, variable reflectance on different spectral bands can be used to analyze biomass amounts of vegetation.

The vegetation indices (VIs) can be classified into three groups: slope-based, distance-based, and orthogonal transformation VIs [8]. The slope-based VIs are simple arithmetic combinations that focus on the contrast between the spectral response patterns of vegetation in the red and near-infrared portions of the electromagnetic spectrum. They are so named because any particular value of the index can be produce by a set of red/infrared reflectance values that form a line emanating from the origin of a bi-spectral plot. Thus different levels of the index can be envisioned as producing a spectrum of such lines that differ in their slope shown as Fig. 1. This kind of VIs contains RATIO, NDVI, RVI, NRVI, TVI, CTVI, and TTVI can be used to present the amount of vegetation biomass.

Next, the distance-based group measures the degree of vegetation present by gauging the difference of any pixel's reflectance from the reflectance of bare soil. A key concept here is that a plot of the positions of bare soil pixels of varying moisture levels in a bi-spectral plot will tend to form a line (known as soil line). As vegetation canopy cover increases, this soil background will become progressively obscured, with vegetated pixels showing a tendency towards increasing perpendicular distance from this soil line. All of the members of this group (such as the Perpendicular Vegetation Index – PVI) thus require that the slope and intercept of the soil line be defined for the image being analyzed. This group including a serious of PVI, DVI, AVI, SAVI, TSAVI, MSAVI, and WDVI, can be used to eliminate the effect of soil brightness.

In contrast to the formal two groups, orthogonal indices undertake a transformation of the available spectral bands to form a new set of uncorrelated bands within which a green vegetation index band can be defined. This group has been approached through orthogonal transformation techniques such as the PCA, the GVI of the Kauth-Thomas Tasseled Cap transformation and the MGVI generated from the Wheeler-Misra orthogonal transformation.

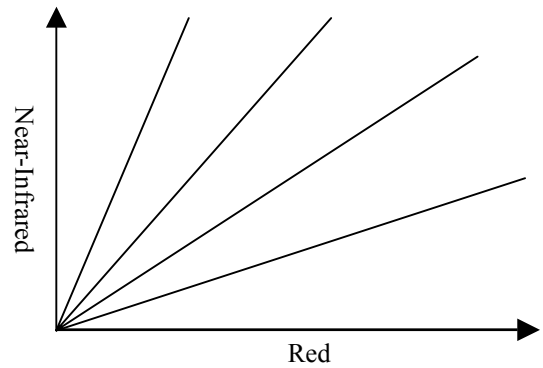


Fig. 1 A plot presents slope-based VIs

3. Rough set theory

Often, information on the surrounding world is imprecise, incomplete or uncertain. Still our way of thinking and concluding depends on information at our disposal. To conduct conclusion, we should be able to process uncertain and/or incomplete information. Formally, an information system, IS (or an approximation space), can be seen as a system

$$IS=(U, A), \qquad (1)$$

where U is the universe (a finite set of objects, U = $\{x_1, x_2, ..., x_m\}$) and A is the set of attributes (features, variables). Each attribute a$\in$ A (attribute a belonging to the considered set of attributes A) defines an information function $f_a : U \rightarrow V_a$, where V_a is the set of values of a, called the domain of attribute a. Moreover, a knowledge representation system containing the set of attributes A (now called condition attributes) and the set of decision attributes D is called a decision table.

In the ordinary set theory, crisp sets are used. A set is then defined uniquely by its elements, i.e, to define a set we have to point out its elements. The membership function, describing the belongingness of elements of the universe to the set, can attain one of the two values, 0 or1. It means that any element is either in or outside the set under consideration. This definition of the membership function does not take into the account the uncertainty of being an element of a given set of elements.

To deal with the uncertainty problems, the concept of fuzzy set was introduced, Fuzzy set is defined by the membership function which can attain values from the closed interval[0,1], allowing partial membership of the elements in the set. Fuzziness measures the degree to which an event occurs and not whether it occurs.

In the rough set theory (RST), membership is not the primary concept. Rough sets represent a different mathematical approach to vagueness and uncertainty. Definition of a set in the rough set theory is related to out information (knowledge) and perception about elements of the universe. In other words, we 'see' elements of the universe in the context of an available information about them. As a consequence, two different elements can be indiscernible in the context of the information about them and 'seen' as the same. Consider a simple example. Two acids with p*K*s of respectively p*K* 4.12 and 4.53 will, in many contexts, be perceived as so equally weak, that they are indiscenible with respect to this attribute. They are part of a rough set 'weak acids' as compared to 'strong' or 'medium' other category, relevant to the context of this classification.

For every set of attributes B$\subset$A, an indiscernibility relation ind(B) is defined in the following way: two objects, x_i and x_j, are indiscenible by the set of attributes B in A, if b(x_i) = b(x_j) for every b$\subset$B.

The equivalence class of Ind(B) is called elementary set in B because it represents the smallest discernible groups of objects. For any element x_i of U, the equivalence class of x_i in relation Ind(B) is represented as $[x_i]_{ind(B)}$. The construction of elementary sets is the first step in classification with rough sets.

The rough sets approach to data analysis hinges on two basic concepts, namely the lower and the upper approximations of a set (Fig.2) referring to:

-the elements that doubtlessly belong to the set, and

-the elements that possibly belong to the set.

Let X denoted the subset of elements of the universe U (X$\subset$U). The lower approximation of X in B (B$\subseteq$A), denoted as $\underline{BX}$, is defined as the union of all these elementary sets which are contained in X.

More formally:

$$\underline{BX} = \{ x_i \in U \mid [x_i]_{ind(B)} \subset X \} \quad (2)$$

The above statement is to be read as: the lower approximation of the set X is a set of objects x_i, which belong to the elementary sets contained in X (in the space B).

The upper approximation of the set X, denoted as BX, is the union of these elementary sets, which have a non-empty intersection with X:

$$BX = \{Xi \in U \square [x_i]_{ind(B)} \cap X \neq 0\} \quad (3)$$

For any object x_i of the lower approximation of X (i.e. $x_i \in \underline{BX}$), it is certain that it belongs to X. For any object x_i of the upper approximation of X (i.e., $x_i \in BX$), we can only say that x_i may belong to X. The difference:

$$BNX = BX - BX, \quad (4)$$

is called a boundary of X in U.

If the lower and upper approximation are identical (i.e., BX=$\underline{BX}$), then set X is definable, otherwise, set X is undefinable in U. There are four types of undefinable sets in U:

1 if BX ≠ ∅ and BX ≠ U, X is called roughly definable in U;
2 if BX ≠ ∅ and BX = U, X is called externally undefinable in U;
3 if BX = ∅ and BX ≠ U, X is called interlly undefinable in U;
4 if BX = ∅ and BX = U, X is called totally undefinable in U;

where ∅ denotes an empty set. Additionally, the following notation can be introduced: POS_B (X) = $\underline{BX}$, called the B-positive region of X, is the set of these objects, which can, with certainty, be classified in the set X, $NEG_B(X)$ = U –$\underline{BX}$, called the B-negative region of X, is the set of objects, which without ambiguity, can be classified as belonging to the complement of X (or as not belonging to X), $BN_B(X)$, called the B-borderline region of X, is an undecidable area of the universe, i.e., none of the objects belonging to the boundary can, with certainty, be classified into X or – X, as far as the attributes B are considered.

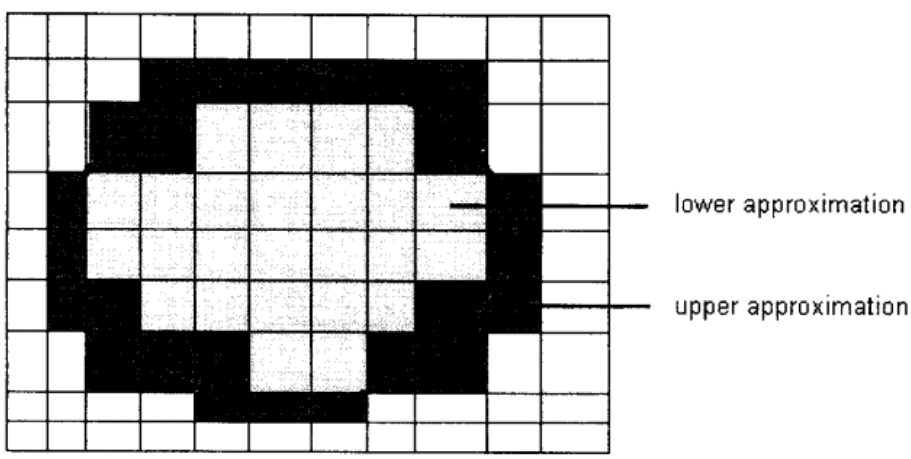


Fig. 2 The lower and the upper approximations of a set

Many different problems can be addressed by RST. During the last few years this formalism has been approached as a tool used in connection with many different areas of research. There have been investigations of the relations between RST and the Dempster-Shafer Theory and between rough sets and fuzzy sets. RST has also provided the necessary formalism and ideas for the development of some propositional machine learning systems. It has also been used for, among many others, knowledge representation; data mining; dealing with imperfect data; reducing knowledge representation and for analysing attribute dependencies. The notions of rough relations and rough functions are based on RST and can be applied as a theoretical basis for rough controllers, among others.

ROSETTA is a toolkit for analyzing tabular data within the framework of rough set theory. ROSETTA is designed to support the overall data mining and knowledge discovery process: From initial browsing and preprocessing of the data,

via computation of minimal attribute sets and generation of if-then rules or descriptive patterns, to validation and analysis of the induced rules or patterns [9]. The following commands are used in this paper :

1) Decision table import

A decision system can be read into a new ROSETTA project by selecting Open... from the main File menu, and will be placed immediately below the root of the Structures node in the project tree. It allows tabular data to be imported from a wide variety of data sources such as spreadsheets or relational databases, by means of ODBC. That data sources are available depends on which ODBC drivers that are installed on your system.

2) Discretization

Discretization amounts to searching for "cuts" that determine intervals. All values that lie within each interval are then mapped to the same value, in effect converting numerical attributes to attributes that can be treated as being categorical. The search for cuts is performed on the internal integer representation of the input decision table. Automatic "grouping" as a symbolic counterpart to intervals is not currently implemented in ROSETTA, but can be done manually.

Input to a discretization algorithm is a decision table, and a decision table is returned. Unless otherwise stated, the returned table is a discretized duplicate of the input table.

In ROSETTA, algorithms for automatic discretization generally fall into one of three categories:

1. Each attribute is considered in isolation, and no knowledge of any outcome or decision attribute is employed in the process. These algorithms are said to be univariate and unsupervised.
2. Only one condition attribute is considered at a time, but is done so in conjunction with the decision attribute. These algorithms are said to be univariate and supervised.
3. All condition attributes are considered simultaneously, and are done so in conjunction with the decision attribute. These algorithms are said to be multivariate and supervised.

An entropy/MDL algorithm is implemented in the software, based on recursively partitioning the value set of each attribute so that a local measure of entropy is optimized [10]. The minimum description length principle defines a stopping criterion for the partitioning process. Missing values are ignored in the search for cuts. If no cuts are found for an attribute, the attribute is left unprocessed.

3) Reducer

The "Reducer" algorithms are used for computing reducts or reduct approximations. Note that any attribute subset is in this context considered to be an approximation to a reduct. Input to a Reducer algorithm is a decision table, and a set of reducts is returned. The returned reduct set may possibly have a set of rules attached to it as a child. A reduct is a collection of attribute indices into the table the reduct belongs to.

Two main types of discernibility are currently supported by ROSETTA. In addition, reducts of both these types can be computed modulo the decision attribute or not.

1. Full: Computes reducts relative to the system as a whole, i.e., minimal attribute subsets that preserve our ability to discern all relevant objects from each other.
2. Object: Computes reducts relative to a fixed object, i.e., minimal attribute subsets that preserve our ability to discern that object from the other relevant objects. Generally, instead of fixing a single object x, we select a subset X of U, and process each x 2 X sequentially. That is, we first compute the minimal attribute subsets that discern the first object in X from all other relevant objects in U, before proceeding to compute the minimal attribute subsets that discern the second object in X from all other relevant objects in U, etc.

Table 1. Options for selecting subsets of U.

Option	Subset
All	X = U
Index	X = {x}
Value	X = {x 2 U | a(x) = v}
File	X = {x 2 U | x is listed in a file}

In the table, an object is specified by a 0-based index into U. For files, one 0-based object index should appear on each line in the file. Alternatively, a line can consist of a 0-based index set contained in curly braces. Lines in other formats

than the ones described are ignored. For reducts relative to an object, the set X can be selected in different ways, as shown in Table 1.

Invokes a variation of a simple greedy algorithm to compute a single reduct only, as described by Johnson [9]. The algorithm has a natural bias towards finding a single prime implicant of minimal length.

The reduct B is found by executing the algorithm outlined below, where S denotes the set of sets corresponding to the discernibility function, and w(S) denotes a weight for set S in S that automagically gets computed from the data.

1. Let B = ∅;.
2. Let a denote the attribute that maximizes □w(S), where the sum is taken over all sets S in S that contain a. Currently, ties are resolved arbitrarily.
3. Add a to B.
4. Remove all sets S from S that contain a.
5. If S =∅ ; return B. Otherwise, go to step 2.

Suppose for computing approximate solutions is provided by aborting the loop when "enough" sets have been removed from S, instead of requiring that S has to be fully emptied.

The support count associated with the computed reduct equals the reduct's hitting fraction multiplied by 100, i.e., the percentage of sets in S that B has a non-empty intersection with.

4. Case study

1) Test dataset

The test dataset (see Fig. 3) of the experiment is an AVIRIS image acquired from the website of School of Electrical and Computer Engineering at Purdue University [11]. The image covers an agriculture field on northwest Indiana in 1992. The original dataset has 224 spectral bands from 450 to 2500nm with 10 nm spectral resolution. The image size of the test area is 145×145. The number of bands is 220 after removing 4 noisy bands. The ground truth data including 16 classes is shown in the right part of the Fig 3. The ID number of each class and its pixel number is shown as Table 2.

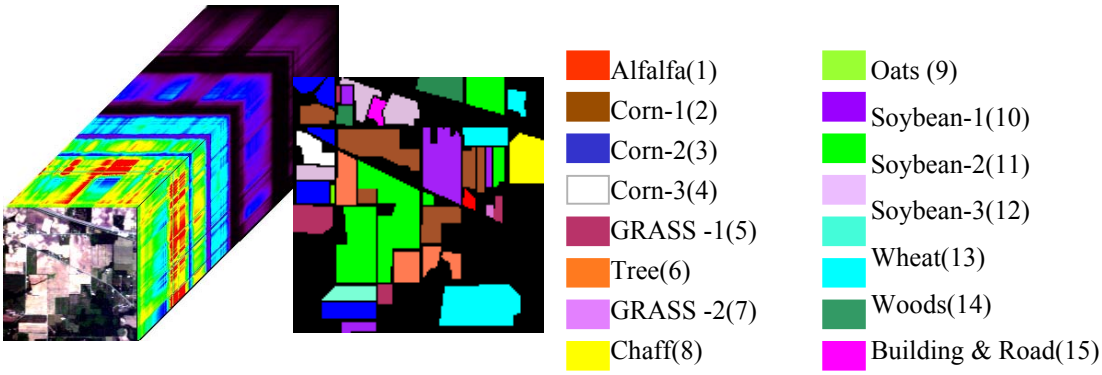


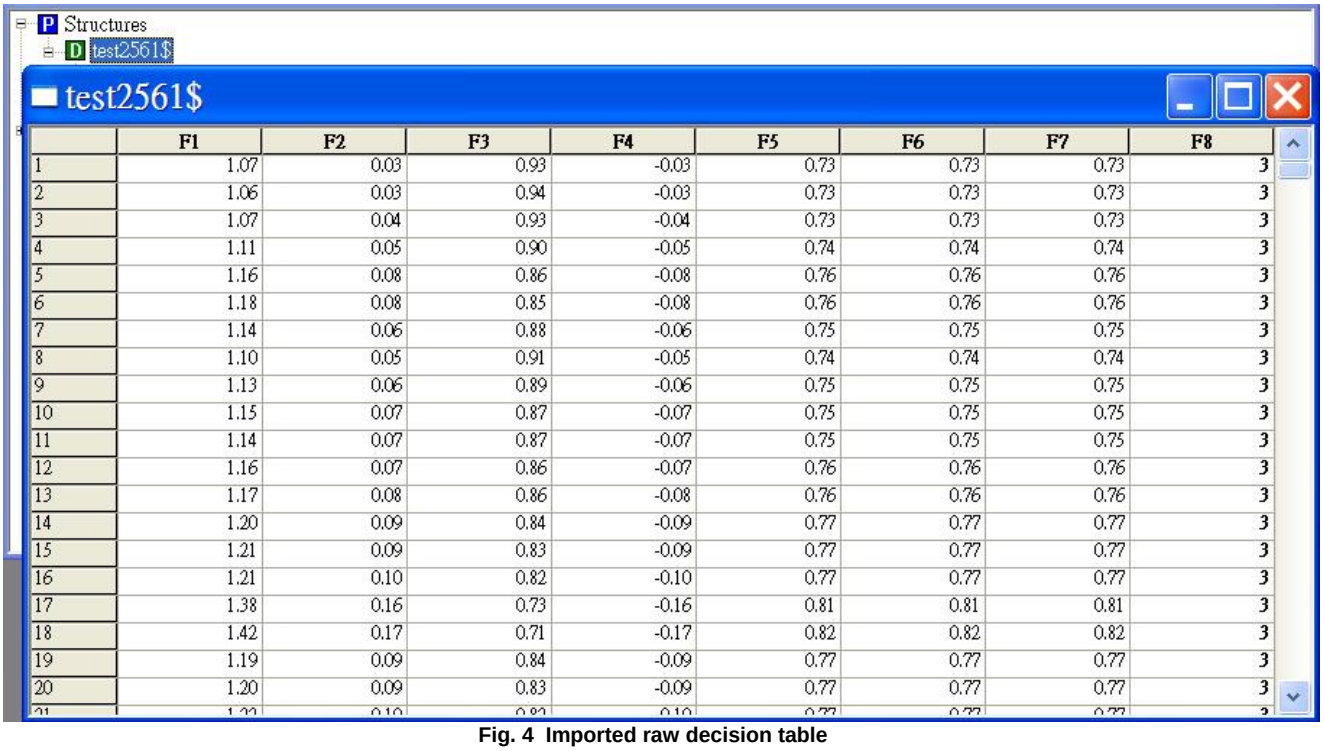
Fig. 3 The used AVIRIS image and its corresponding ground truth

Table 2. The ID number and its pixel number for each Land-use category

Land-use category	Pixel number	Land-use category	Pixel number
Alfalfa(1)	54	Oats(9)	20
Corn – 1(2)	1434	Soybean – 1 (10)	968
Corn –2(3)	834	Soybean – 2 (11)	2468
Corn - 3(4)	234	Soybean - 3(12)	614
Grass-1(5)	497	Wheat(13)	212
Tree(6)	747	Woods(14)	1294
Grass-2(7)	26	Builds and Roads (15)	380
Chaff (8)	489	Steel (16)	95

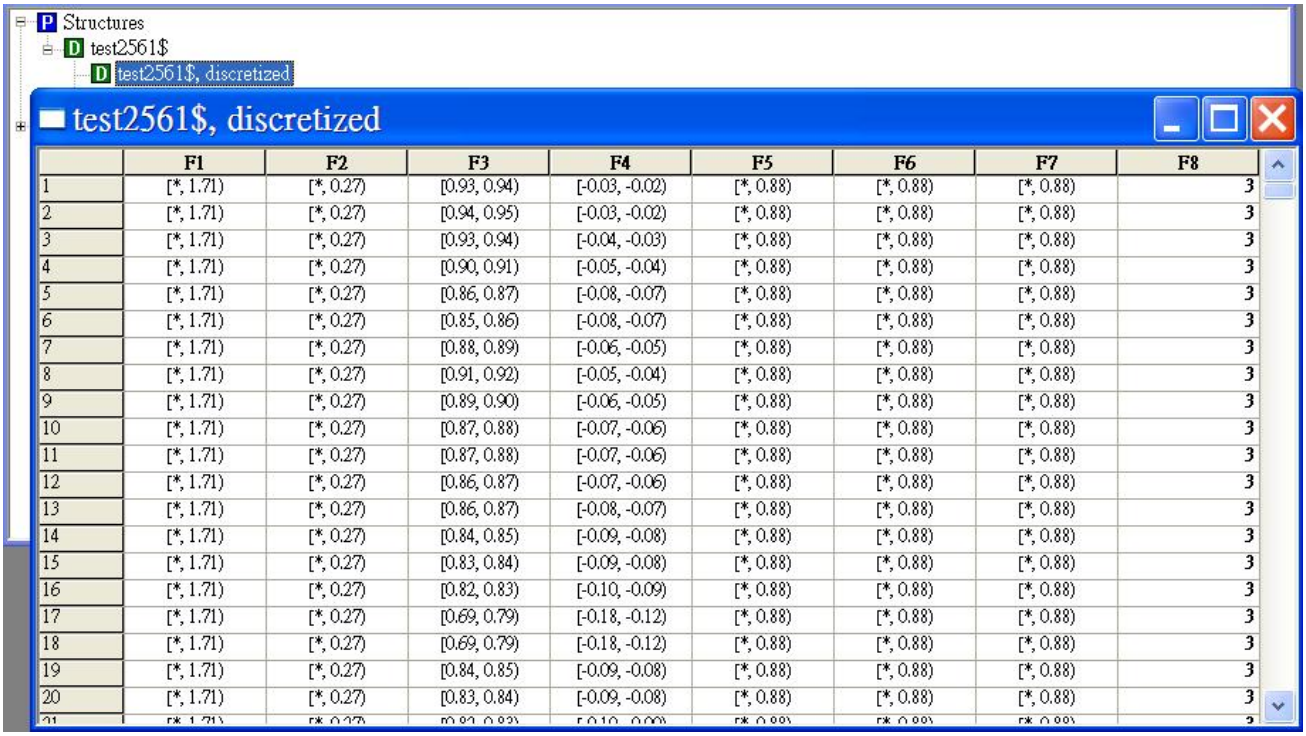
2) Results and analysis

It is obvious from figure 4 that there is no bare soil in the test database. None of the distance-based VIs can be obtained. To understand the vegetation indices generated from the narrow-band spectrum and its effect on the ratio of land use and cover information ranging from red (Band No. 23 to 29) to near-infrared (Band No. 39 to 71) were selected to determine the slope-based VIs. Statistical analyses were performed by the author in the past to study the variation of the average values and standard deviations of seven slope-based VIs[12]. The research result obtained after the seven slope-based VIs and land categories were input into ROSETTA can be shown in figure 4. It is then discretized through Entropy/MDL method because attributes are real numbers. Attributes can be divided into several intervals using this technique.



	F1	F2	F3	F4	F5	F6	F7	F8
1	1.07	0.03	0.93	-0.03	0.73	0.73	0.73	3
2	1.06	0.03	0.94	-0.03	0.73	0.73	0.73	3
3	1.07	0.04	0.93	-0.04	0.73	0.73	0.73	3
4	1.11	0.05	0.90	-0.05	0.74	0.74	0.74	3
5	1.16	0.08	0.86	-0.08	0.76	0.76	0.76	3
6	1.18	0.08	0.85	-0.08	0.76	0.76	0.76	3
7	1.14	0.06	0.88	-0.06	0.75	0.75	0.75	3
8	1.10	0.05	0.91	-0.05	0.74	0.74	0.74	3
9	1.13	0.06	0.89	-0.06	0.75	0.75	0.75	3
10	1.15	0.07	0.87	-0.07	0.75	0.75	0.75	3
11	1.14	0.07	0.87	-0.07	0.75	0.75	0.75	3
12	1.16	0.07	0.86	-0.07	0.76	0.76	0.76	3
13	1.17	0.08	0.86	-0.08	0.76	0.76	0.76	3
14	1.20	0.09	0.84	-0.09	0.77	0.77	0.77	3
15	1.21	0.09	0.83	-0.09	0.77	0.77	0.77	3
16	1.21	0.10	0.82	-0.10	0.77	0.77	0.77	3
17	1.38	0.16	0.73	-0.16	0.81	0.81	0.81	3
18	1.42	0.17	0.71	-0.17	0.82	0.82	0.82	3
19	1.19	0.09	0.84	-0.09	0.77	0.77	0.77	3
20	1.20	0.09	0.83	-0.09	0.77	0.77	0.77	3

Fig. 4 Imported raw decision table



Structures
test2561$
test2561$, discretized

test2561$, discretized

	F1	F2	F3	F4	F5	F6	F7	F8
1	[*, 1.71)	[*, 0.27)	[0.93, 0.94)	[-0.03, -0.02)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
2	[*, 1.71)	[*, 0.27)	[0.94, 0.95)	[-0.03, -0.02)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
3	[*, 1.71)	[*, 0.27)	[0.93, 0.94)	[-0.04, -0.03)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
4	[*, 1.71)	[*, 0.27)	[0.90, 0.91)	[-0.05, -0.04)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
5	[*, 1.71)	[*, 0.27)	[0.86, 0.87)	[-0.08, -0.07)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
6	[*, 1.71)	[*, 0.27)	[0.85, 0.86)	[-0.08, -0.07)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
7	[*, 1.71)	[*, 0.27)	[0.88, 0.89)	[-0.06, -0.05)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
8	[*, 1.71)	[*, 0.27)	[0.91, 0.92)	[-0.05, -0.04)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
9	[*, 1.71)	[*, 0.27)	[0.89, 0.90)	[-0.06, -0.05)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
10	[*, 1.71)	[*, 0.27)	[0.87, 0.88)	[-0.07, -0.06)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
11	[*, 1.71)	[*, 0.27)	[0.87, 0.88)	[-0.07, -0.06)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
12	[*, 1.71)	[*, 0.27)	[0.86, 0.87)	[-0.07, -0.06)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
13	[*, 1.71)	[*, 0.27)	[0.86, 0.87)	[-0.08, -0.07)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
14	[*, 1.71)	[*, 0.27)	[0.84, 0.85)	[-0.09, -0.08)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
15	[*, 1.71)	[*, 0.27)	[0.83, 0.84)	[-0.09, -0.08)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
16	[*, 1.71)	[*, 0.27)	[0.82, 0.83)	[-0.10, -0.09)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
17	[*, 1.71)	[*, 0.27)	[0.69, 0.79)	[-0.18, -0.12)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
18	[*, 1.71)	[*, 0.27)	[0.69, 0.79)	[-0.18, -0.12)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
19	[*, 1.71)	[*, 0.27)	[0.84, 0.85)	[-0.09, -0.08)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3
20	[*, 1.71)	[*, 0.27)	[0.83, 0.84)	[-0.09, -0.08)	[*, 0.88)	[*, 0.88)	[*, 0.88)	3

Fig. 5 Discreted decision table

For simplicity only the resulting extraction of reduct sets of red (band No. 25) and near-infrared (band No. 61, 63) through the Johnson Method are shown in Table 3-4. F1 through F7 represents seven VIs such as ratio, ndvi, rvi, nrvi, tvi, ctvi and ttvi, repectively. The following conclusions are drawn from the results:

1. F6 and F7 (i.e. ctvi and ttvi) can be treated as superfluous attributes for zero reduct sets are discovered
2. F3 and F4 (i.e. rvi and nrvi) out of F1 to F5 demonstrate the highest probability of showing the reduct sets. It implies that these two are relatively important on the analysis of 16 kinds of terrain category.
3. From the information collected, VIs values have no particular impact on Type 4, 5, 6, 11 and 15 (i.e. corn field, grass, tree, soybean-2, buildings and roads). It is concluded that the slope-based VIs are independent of soil background effect for VIs are unable to rectify them.

Table 3. Reduct sets for the combination of red (band no. 25) and NIR (band no. 61)

CLASS 1	CLASS 2	CLASS 3	CLASS 4	CLASS 5	CLASS 6	CLASS 7	CLASS 8
{F4} {F1, F3} {F3}	{F3} {F3, F4} {F4} {F1, F3} {F2, F5} {F1, F5}	{F3} {F3, F4} {F4}	{F3} {F3, F4} {F4} {F1, F3}	{F4} {F3, F4} {F1, F3} {F3} {F1} {F1, F2} {F1, F5}	{F4} {F1, F3} {F1} {F1, F5} {F2, F5} {F1, F2} {F3} {F3, F4}	{F4} {F1, F3}	{F4} {F1, F3} {F1} {F3, F4}
CLASS 9	CLASS 10	CLASS 11	CLASS 12	CLASS 13	CLASS 14	CLASS 15	CLASS 16
{F1, F3} {F3, F4}	{F3} {F3, F4} {F4} {F1, F3}	{F3, F4} {F3} {F4} {F1} {F1, F3} {F1, F2}	{F3, F4} {F3} {F4} {F1, F3}	{F1, F3} {F4} {F1} {F1, F2}	{F1, F3} {F1} {F1, F2} {F1, F5}`	{F1, F3} {F1} {F1, F2} {F1, F5} {F3, F4} {F4} {F3} {F2, F5}	{F3} {F3, F4}

Table 4. Reduct sets for the combination of red (band no. 25) and NIR (band no. 63)

CLASS 1	CLASS 2	CLASS 3	CLASS 4	CLASS 5	CLASS 6	CLASS 7	CLASS 8
{F4} {F2, F3} {F3}	{F3, F4} {F3} {F4} {F2, F3} {F1}	{F3} {F4} {F3, F4}	{F3, F4} {F3} {F4} {F2, F3}	{F4} {F3} {F3, F4} {F2, F3} {F1} {F1, F2}	{F4} {F2, F3} {F3, F4} {F1} {F1, F2} {F1, F5} {F3}	{F4} {F2, F3}	{F4} {F2, F3} {F1} {F3, F4}
CLASS 9	CLASS 10	CLASS 11	CLASS 12	CLASS 13	CLASS 14	CLASS 15	CLASS 16
{F2, F3} {F4}	{F3} {F3, F4} {F4} {F2, F3}	{F3} {F3, F4} {F4} {F1} {F2, F3}	{F3, F4} {F3} {F4}	{F2, F3} {F4} {F3, F4} {F1} {F3}	{F2, F3} {F1} {F1, F2} {F1, F5}	{F2, F3} {F1} {F1, F2} {F1, F5} {F3, F4} {F3} {F4}	{F3} {F3, F4}

Except for finding the reduct sets of each land category, the interpretation rules for land use/cover classification can be acquired in this study. The generated rules results shown in figure 6 indicates that the prerequisite are listed in the left and the results are listed on the right.

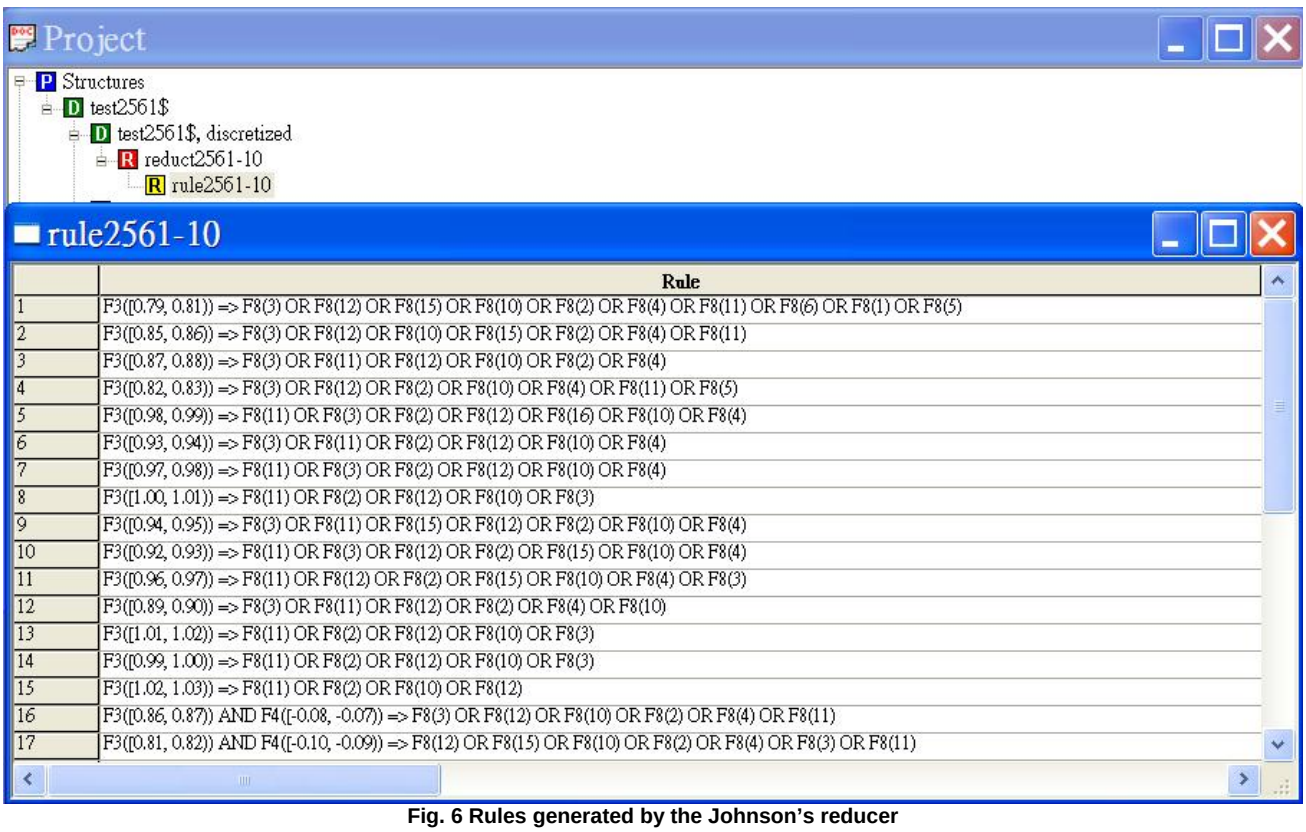


Fig. 6 Rules generated by the Johnson's reducer

5. Conclusions

In this study, the RST is proved to extract effectively the decision rules and significant slope-based VIs for the land use/cover classification. Other comments are as follows:

1. Reduct sets are not found in ctvi and ttvi after reduction. They can be treated as superfluous attributes.
2. For the rest of 5 VIs, the Reduct sets are frequently found in rvi and nrvi. Therefore they are considered critical in

categorizing the land type.

3. For the information collected, all VIs are irrelevant in corn field, grass, tree, soybean-2, buildings and roads. It is suspected that the slope-based VIs are unable to rectify the soil background effect.

References

[1] Maselli, Fabio and Felix Rembold, 2001. Analysis of GAC NDVI Data for Cropland Identification and Yield Forecasting in Mediterranean African Countries, Photogrammetric Engineering & Remote Sensing, Vol. 67, 5:593-602

[2] Moran, M. S., Y. Inoue & E. M. Barnes, 1997. Opportunities and Limitations for Image based Remote Sensing in Precision Crop Management, Remote Sensing Environment, 61:319-346.

[3] Tucker, C. J., 1979. Red and photographic infrared linear combinations for monitoring vegetation, Remote Sensing of Environment, 8:127-150.

[4] Wiegand, C. L., Mass, S. J., Aase, J. K., Hatfield, J. L., Pinter, P. J., Jackson, R. D., Kanemasu, E. T., and Lapitan, R. L., 1992. Multisite analysis of spectral-biophysical data for wheat, Remote Sensing of Environment, 42:1–21.

[5] Yang, C. and James Everitt, 2000. Relationships Between Yield Monitor Data and Airborne Multispectral Digital Imagery, 5th International Conference on Precision Agriculture.

[6] Pawlak, Z., 1991. Rough Sets Theoretical Aspects of Reasoning about Data, Kluwer Academic Publisher, Dordrecht, Netherlands.

[7] Lillesand, T. M. and R. W. Kiefer, 1994. Remote Sensing and Image Interpretation, 3rd ed., John Wiley & Sons, Inc.

[8] Clark Labs, 2001. IDRISI 32 Release 2 User Manual, University of Clark, U. S. A..

[9] Øhrn, A., 2000. ROSETTA Technical Reference Manual, Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. 66 pages.

[10] Dougherty, J., R. Kohavi, and M. Sahami, 1995. Supervised and unsupervised discretization of continuous features. In A. Prieditis and S. Russell, editors, Proc. Twelfth International Conference on Machine Learning, pages 194–202. Morgan Kaufmann.

[11] **URL:** http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/

[12] Chang, K. T. and J. K. Liu, 2004. Hyperspectral Vegetation Indices for the Land-use/Cover Classification, SE2004, NCHU, Taichung, pp 185-192. (in Chinese)