

# CHARACTERISTICS OF THE DEGREE OF GRADE IN GRADE-ADDED ROUGH SET FOR LAND COVER CLASSIFICATION

Y. Ishii<sup>1</sup>, K. Iwao<sup>2</sup>, T. Kinoshita<sup>3,\*</sup>

<sup>1</sup> Graduate School of Agriculture, Ibaraki University, 300-0393, Ibaraki, Japan - 18am401h@vc.ibaraki.ac.jp

<sup>2</sup> Geological Institute of Geology and Geoinformation Geological Survey of Japan, Advanced Industrial Science and Technology, 305-8567, Ibaraki, Japan - iwao.koki@aist.go.jp

<sup>3</sup> College of Agriculture, Ibaraki University, 300-0393, Ibaraki, Japan - tsuguki.kinoshita.00@vc.ibaraki.ac.jp

Commission III, WG III/10

**KEY WORDS:** Grade-added rough set, Maximum likelihood classifier, Support vector machine, Membership function, Land cover classification, Multispec

## ABSTRACT:

This paper aims to clarify the meaning of the membership which is produced as by-products of land cover classification by Grade-added rough set (GRS). A new land cover classification method by using GRS was developed. The classification scheme of GRS which calculates membership (degree of grade) for each class is similar to those of MLC and SVM. But there are two things that are not clear. One is a meaning of the membership of GRS and the other is a reason why the larger membership in GRS employed works well. In this study, aerial images were used to visualize the relation of membership between GRS and existing classifiers, MLC and SVM. Furthermore, a model experiment in two-dimensional feature space was conducted. From these experiments, it was found that the meaning of degree of grade is a distance from a nearest training data of other class. That is, the meaning of membership of GRS is similar to that of SVM, because SVM also calculates a distance from boundary line which is determined by support vectors, while the meaning of membership of MLC is a distance from a centroid of own class. Also it was found that what the distance from the closest other class is given as the degree of grade implies that the higher the grade, the higher the certainty. In this research we could clarify some of the features of land cover classification using GRS.

## 1. INTRODUCTION

A new land cover classification method called Grade-added rough set (GRS) was developed (Ishii *et al.*, 2018). The rough set theory which is the basis of this approach was proposed by Pawlak (Pawlak, 1982).

In the process of land cover classification, the three methods, GRS, MLC and SVM have a similarity that it outputs a map of membership for each class and then classifies each pixel into an appropriate class referring to the membership. In GRS, the grade degree of each class is obtained for each pixel and classified into the class of the maximum grade degree. Pixels are allocated to their most likely class of membership in MLC (Foody *et al.*, 1992). In SVM, pixels are classified a class which the decision function takes maximum value when separating the class from the remaining class. In this way, these three classification methods once calculate the membership of each class and compare the membership with other classes. However, it is not clear why classifying to the class whose grade degree is maximum. Therefore, by examining the relation between GRS, MLC, and SVM membership value, we clarify what grade degree in GRS means in land cover classification.

In this research, we clarify the relation of membership between GRS and existing method, MLC and SVM using actual aerial photographs, then reveal the relation in two-dimensional modelled feature space, comparing the characteristics of membership values of GRS, MLC and SVM. This paper aims to clarify the meaning of the degree of grade of GRS in the land cover classification through these experiments.

## 2. MEMBERSHIP FUNCTIONS

We introduce the method of calculating the membership of the three classifiers used in this study.

### 2.1 GRS

GRS is a method developed based on the rough set theory. It is known that the classification will be successful to allocate a pixel into a class which is the highest degree of grade (Ishii *et al.*, 2018). The degree of grade is obtained as follows (Mori, *et al.*, 2004). A decision matrix is created for training data belonging to class  $k$  and not belonging to class  $k$ .  $M_{ij}^k$  is an element of decision matrix for class  $k$ .

$$M_{ij}^k = \left\{ \begin{array}{l} (a, \rho(\mathbf{x}_i, a), g_{ij}) | \rho(\mathbf{x}_i, a) \neq \rho(\mathbf{x}_j, a), \\ i \in K_k^+, j \in K_k^- \end{array} \right\}, \quad (1)$$

where,  $\mathbf{x}$  is training data,  $a$  is an attribute,  $\rho$  is an attribute value,  $g_{ij}$  is a grade degree which calculated by  $g_{ij} = \rho(\mathbf{x}_i, a) - \rho(\mathbf{x}_j, a)$ ,  $K_k^+ = \{i | \mathbf{x}_i \in C_*(D_k)\}$ , and  $K_k^- = \{i | \mathbf{x}_i \notin D_k\}$ . A threshold is defined as  $\alpha$ . If  $|g_{ij}| \geq \alpha$ , regard the sample as discernible from the other class. Then, calculate the disjunction about each attribute,

$$\mathcal{L}(M_{ij}^k) = (\rho(\mathbf{x}_i, a_1) \vee \rho(\mathbf{x}_i, a_2) \vee \dots \vee (\mathbf{x}_i, a_m), \arg \max(|g_{ij}|)) \quad (2)$$

\* Corresponding author

It is necessary to be discernible with all training data of other class, so calculate the conjunction. Furthermore, calculate the disjunction in order to need to be distinguished by at least one training data of class  $k$ .  $\mathbf{x}$  is redefined as each pixel of image.

$$D_k(\mathbf{x}) = \bigvee_{i \in K_k^+} \bigwedge_{j \in K_k^-} \mathcal{L}(M_{ij}^k) \quad (3)$$

where calculate  $\arg \min(|g_{ij}|)$  in conjunction and  $\arg \max(|g_{ij}|)$  in disjunction. For each class, calculate eq. (3), and each pixel allocates the class which maximize eq. (3). That is,

$$f(\mathbf{x}) = \arg \max_{i=1, \dots, n} D_i(\mathbf{x}) \quad (4)$$

If a pixel is not allocated any class or there are more two class of max membership, the pixel is allocated 0.

## 2.2 MLC

MLC is a method of finding likelihood and classifying pixels into a class that maximizes its likelihood, but actually it is a problem of classifying it into a class that minimizes the following formula derived on the basis of likelihood (Japan Remote Sensing Society, 2011).

$$D_k(\mathbf{x}) = \ln|\Sigma_k| + (\mathbf{x} - \mathbf{m}_k)^T \Sigma_k^{-1} (\mathbf{x} - \mathbf{m}_k) \quad (5)$$

where, each pixel is  $\mathbf{x}$ , a mean vector of class  $k$  is  $\mathbf{m}_k$ , variance-covariance matrix of class  $k$  is  $\Sigma_k$ .

$$f(\mathbf{x}) = \arg \min_{i=1, \dots, n} D_i(\mathbf{x}) \quad (6)$$

## 2.3 SVM

Let the  $i$  th decision function separating class  $i$  from the remaining class be given by

$$D_k(\mathbf{x}) = \mathbf{w}_k^T \phi(\mathbf{x}) + b_k \quad (7)$$

where,  $\mathbf{x}$  is each pixel,  $\phi(\mathbf{x})$  is a function of mapping input space to 1 dimensional feature space,  $\mathbf{w}_k$  is 1-dimensional coefficient vector and  $b_k$  is bias (Abe, 2011). For the 2-class problem, the sgn function of Eq. (7) becomes the decision function, but in the case of multi-class classification, the following function is the decision function.

$$f(\mathbf{x}) = \arg \max_{i=1, \dots, n} D_i(\mathbf{x}) \quad (8)$$

## 3. METHODS

### 3.1 Data sets

We used A Moderate Dimension Example published by A Freeware Multispectral Image Data Analysis System of Purdue University (Landgrebe and Biehl, 1994). This is a data set of 12 bands taken from the aircraft at a farmland in Indiana in June 1966. The size of the image is  $949 \times 220$  pixels, of which 70594 points are certain data whose classes are known.

Based on the experimental results using this data set by Purdue University (Landgrebe, 1997), we used 4 bands of 12 bands, Band 1, 6, 9 and 10. Based on the central limit theorem, about

70,000 points of data were divided into 31. 30 training data sets and a validation data set were created. Table. 1 shows class name, the number of pixels and the area proportion of 70594 points for each class. The difference between the minimum number of class and the maximum number of class is about 20 times. Since 31 divisions were conducted randomly, there are some differences in the area proportion of each class by the training data set, but it is almost the same as the data set.

Table 1. Multispec test data set

No.	Class	Pixels	Area proportion (%)
1	Alfalfa	3375	4.8
2	Bare Soil	1230	1.7
3	Corn	10625	15.1
4	Oats	5781	8.2
5	Red Clover	12147	17.2
6	Rye	2385	3.4
7	Soybeans	25133	35.6
8	Wheat 1	7827	11.1
9	Wheat 2	2091	3.0

## 3.2 Experimental design

In this study, land cover classifications using Multispec data set were conducted by three methods, GRS, MLC, and SVM. Then, the maps of membership are outputted for each class. From the membership map, 70549 points whose correct classes were known were extracted. When classification was conducted, the data set of the original 8-bit Multispec was normalized a mean to 0 and variance to 1, and then it multiplied the number by 1000. Assume a threshold  $\alpha = 1$  in GRS. Radial basis function (RBF) was used for the SVM kernel. In this paper, hyperparameters of SVM were set to  $\sigma = 1.0$  and regularization parameter  $C = 100.0$ . This classification was repeated for 30 data sets. Overall accuracy of maps was evaluated, because it is better to know how accurate each classifier is for examining the correlation of membership.

Furthermore, experiments were conducted to investigate the differences of three classifiers using a two-dimensional feature space. We assumed two classes and examined how each classifier draws a class boundary. Training data of each class was set to five (1st row in Fig. 3). It is known that MLC is a method considering variance of each class, but it is not clear GRS and SVM can consider the effect of variance of each class. So, experiments were carried out with three variance cases (1st row in Fig. 3) to investigate whether GRS and SVM have such effect.

## 4. RESULTS AND DISSCUSION

Land cover classifications were conducted by GRS, MLC and SVM using Multispec data set. Fig. 1 shows overall accuracy with 30 training data sets. SVM has the highest average accuracy, but the variance is large. On the other hand, MLC has the smallest variance.

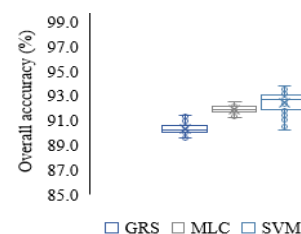


Figure 1. Overall accuracy

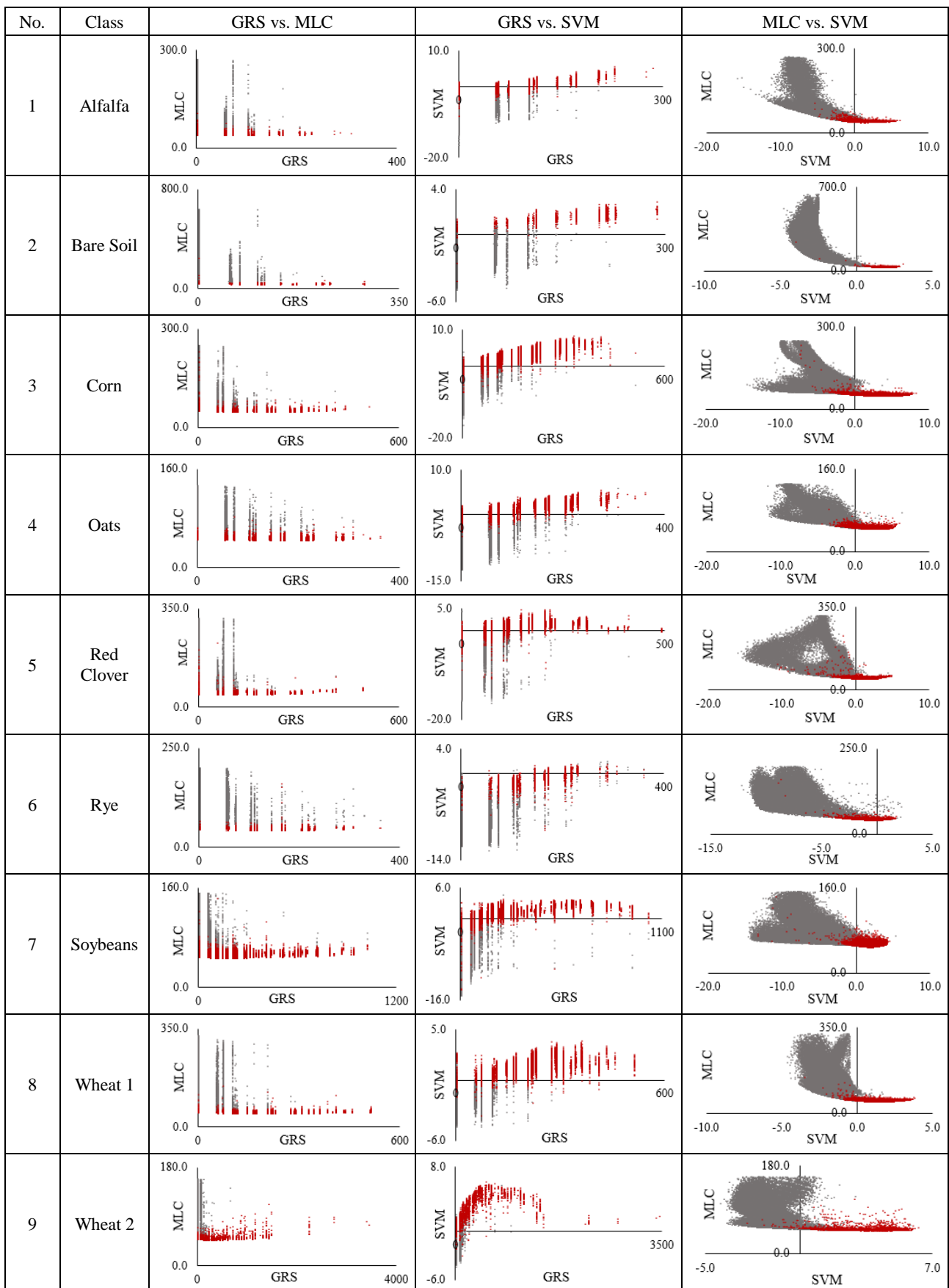


Figure 2. Scatter plots of membership between two classifiers

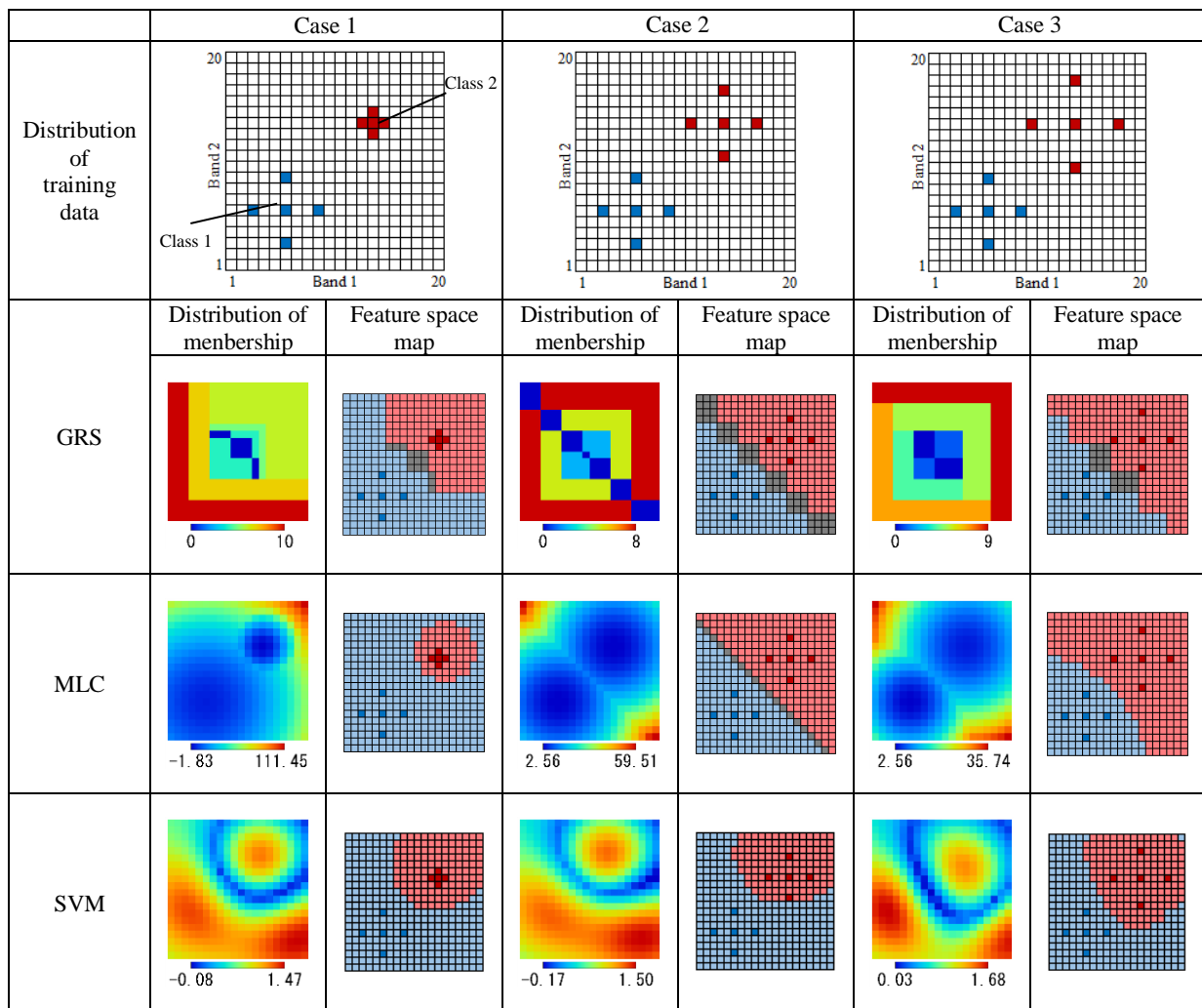


Figure 3. Distribution of training data, membership and classification map in feature space in three cases

Then, using Multispec image, a trend of membership was examined. Each classifier outputted nine class membership maps for each 30 training data sets. In addition, the correlation of membership of GRS and MLC, GRS and SVM, MLC and SVM was examined for each class. The number of points used for examining the correlation were 70594 points where the correct class was known. As a result, similar results were shown from different training data sets, so only a result of one training set is represented in Fig. 2. The graphs in Fig. 2 are scatter plots of membership of two classifiers. The horizontal axis indicates a membership of one classifier, and the vertical axis indicates that of the other. Although the membership of GRS is integer value, that of MLC and SVM is continuous value. The red points in each graph are the membership which is actually correct answer, and grey points are the membership not chosen that class. There are about 70,000 points in each graph, but some points overlap.

Focusing on the distribution of correct classes (red points) in GRS and MLC, the membership of MLC is mostly low, but the membership of GRS takes various values. On the other hand, paying attention to the distribution of grey points, the membership of MLC exponentially increases as the membership of GRS decrease.

Many of GRS and SVM have slightly positive correlation. However, as the membership of GRS gets larger, the membership

of SVM become larger, but there are also cases where membership of SVM is slightly lower when GRS membership is larger.

In MLC and SVM, they have a common point that there is a distribution of the correct class (red points) in the lower right to a distribution of all points. In other words, when the membership of SVM is small, MLC takes wide range of membership, but when the membership of SVM is large corresponds to only when the membership of MLC is small. This means that the membership of MLC and SVM almost agrees with reasoning for correct class, but the membership of other classes shows different prediction depending on the deference of classification scheme.

Next, Fig. 3 shows the results of experiments using a two-dimensional feature space. The horizontal axis represents band 1, and the vertical axis represents band 2. A range of value is from 1 to 20. As shown in the first row of Fig. 3, three training data sets including different variance class are considered. It was assumed that blue training data is class 1 and red training data is class 2. Case 1 has smaller variance of class 2 than class 1. Case 2 has the same variance of class 1 and class 2. Case 3 has larger variance of class 2 than class 1. In all cases, the centroids of the classes are fixed. The rows from second to fifth in Fig. 3 represent the distribution of membership (left side) and the map of the classification result in feature space (right side).

As a result, the degree of grade becomes smaller towards the center of the feature space in GRS, and when the degree of grade is the same, the point is unclassified. Unclassified points are stored 0. The degree of grade of both classes in the vicinity of the boundary line is a little different, and a class having a higher degree of grade is selected. Observing the degree of grade of each cell inferred from the training data revealed that the degree of grade stored in each cell indicates the distance from the closest training of the opponent class in the feature space. Therefore, it was revealed that to calculate the degree of grade is to find the closest training data of the opponent class, and the degree of grade shows how far from the closest training data of the opponent class. Extracting rules by disjunction of training data of the same class means to adopt the point with the larger degree of grade. This indicates that the larger grade of degree, the larger distance to the closest training data of opponent class. Furthermore, it can be interpreted that more reliable information is adopted because this is more reliable training data as the distance to the opponent is greater. That is, the greater the degree of grade, the higher the certainty of being their own class. In classification by GRS, a map of membership for each class is created, and among them each point classifies into a class with the highest grade of degree. It was confirmed that the reason for this is that the degree of grade can be regarded as equivalent to the degree of certainty. Also, the logic that unclassification occurs at the boundary because the certainty of both classes is of equal degree is reasonable. In addition, it can be said that GRS can take into account the problem that whether variance can be considered. In fact, the concept of variance does not appear in the GRS. However, comparing when the variance of class 2 is large and when it is small, the degree of grade stored in class 1 is different. For example, when the centroids are fixed, the larger the variance of class 2 is, the closer the distance is to the side closer to class 1 than class 2's centroid, and the further away from class 1 than class 2's centroid, the further the distance becomes. This indicates what an operation of taking a disjunction for rules extracted from each training data of the same class is easier to adopt larger grade of degree by a relatively large variance class and smaller grade of degree by a relatively small variance class. As a result, it can be seen from the second row in Fig. 3 that the class with large variance has a scheme that dominates the relatively large region in the feature space as compared with the small class.

Next, in the case of MLC, it was confirmed that the membership increases as the distance from the centroid obtained from the training data, and the class boundary is drawn where the magnitude of the membership changes between classes. MLC is a method that takes variance and inclination into consideration from the theory. From the results in Fig. 3, it is also found that the change with the distance from the centroid is larger for classes with smaller variance. And grey cells indicates unclassified which occurs by same membership in Case 2. In addition, MLC is fundamentally similar to the nearest neighbour method, and it is a different property from GRS in that it calculates how close the distance from the centroid of opponent class is.

Fig. 4 shows a graph of the distribution of membership of GRS and MLC on class 1 of Case 1 on a line AB in the feature space. Fig. 4 (a) is a position of line AB in feature space, and Fig.4 (b) is membership of GRS and SVM on a line AB. Normalization that the average is 0 and variance is 1 for each membership, was performed, and then, took minimum value from all membership to make them positive numbers. From Fig. 4, we can explain the relation between GRS and MLC in Fig. 2. When the membership of the MLC takes a relatively low value, it takes a variety of membership in the GRS in Fig. 2. That coincides with the graph

to the left side in Fig. 4 (b). In addition, when the membership of GRS takes a relatively small value, MLC has a wide range of values. That coincides with the graph to the right side Fig. 4 (b).

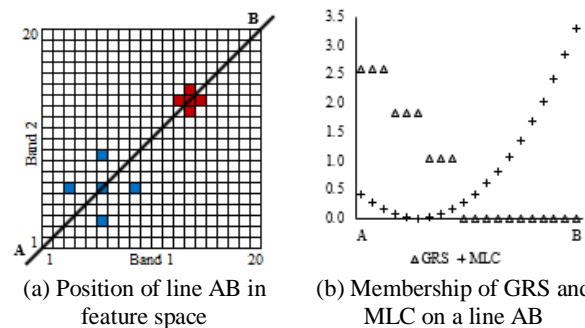


Figure 4. Comparison of class 1 membership between GRS and MLC in Case 1

In the case of SVM, it was confirmed that the boundary line is first obtained from the support vector, then the distance from it is given as membership, and as the result of adopting a class with a large distance, the boundary is finally drawn by change points of the magnitude of membership between classes. In SVM, it was not clear from this experimental result whether the effect of class variance was taken into consideration.

Fig. 5 (b) is a graph of comparing the distribution of Class 1 membership on a line AB in the feature space between GRS and SVM in Case 1. The trend is similar between GRS and SVM. Fig. 5 (b) explains to the tendency of distribution between GRS and SVM in Fig. 2.

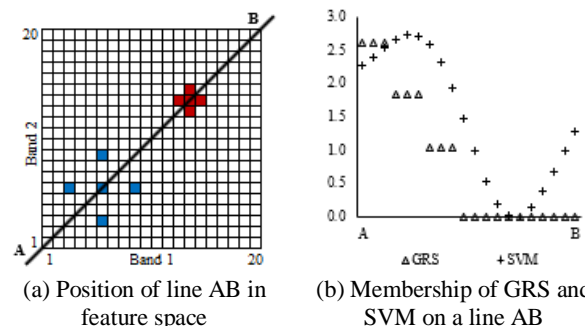


Figure 5. Comparison of class 1 membership between GRS and SVM in Case 1

The degree of grade in GRS is the distance from the closest training data of the other class, and the membership is the distance from the boundary obtained by the support vector which is the closest distance to the other class. That is, similar in that the distance is obtained from the relation with other class. Meanwhile, the membership of MLC is the distance from the centroid of its own class. From this point it differs greatly from GRS and SVM. (Foody and Mathur, 2006) and (Brown, Lewis and Gunn, 2000) have already used the membership value of SVM to solve the mixed pixel problem. Therefore, what the way of finding membership in GRS is similar to that of SVM means that GRS has the possibility of applying to mixed pixel estimation. GRS has advantages which does not need the adjustment of hyperparameter like SVM, so we plan to consider GRS application to mixed pixel estimation in the future.



## 5. CONCLUSIONS

GRS, MLC, SVM have a common point in classifying based on the membership of each class for land cover classification. In this paper, we discussed the significance of membership of each class generated by land cover classification using GRS, MLC or SVM. First, using the Multispec data set published by Purdue University, land cover classification is performed with these three classifiers. Then, membership maps of each class were output, and the distributions of membership of GRS vs. MLC, GRS vs. SVM and MLC vs. SVM are examined. Although these results can not be said to be completely correlated, they showed that they are any relations to each other. In order to investigate about these relations, a model experiment was conducted in a two-dimensional feature space. As a result of classification of 2 classes with 5 training data per class, the characteristics of drawing boundary in the feature space of each classifier was clarified. In GRS, the distance between its own class and the closest point of other class is given as the degree of grade. In addition, comparing the degree of grade of the two classes and adopting the larger class means to adopt a class with a higher certainty class. In MLC, it was confirmed that a boundary was drawn where the distance as the magnitude of the value adding the effect of variance and inclination from the centroid of its own class changes between two classes. In SVM, it was confirmed that the distance from the boundary obtained from the support vector is given as a membership value, and the place where the distance from the line changes is the boundary in the feature space. It is concluded that the way of classification of GRS is close to SVM in that it focuses on the distance to the other class. SVM is already used for estimating mixed pixels. So there is a possibility to use GRS as a more practical method for estimating mixed pixels.

Pawlak, Z., 1982. Rough Sets. In: *International Journal of Computer & Information Sciences*, 11(5), pp. 341–356. doi: 10.1007/978-3-7908-1776-8.

## REFERENCES

- Abe, S., 2011. Introduction to support vector machine for pattern recognition. *Morikita publishing*, 224pp. (in Japanese).
- Brown, M., Lewis, H. G. and Gunn, S. R., 2000. Linear Spectral Mixture Models and Support Vector Machines for Remote Sensing. *Bayrischer Rundfunk*, 38(5), pp. 2346–2360.
- Foody, G. M., Campbell, N. A., Trodd, N. M. and Wood, T. F., 1992. Derivation and applications of probabilistic measures of class membership from maximum-likelihood classification. In: *Photogrammetric Engineering & Remote Sensing*, 58(9), pp. 1335–1341. doi: 10.3851/IMP1936.
- Foody, G. M. and Mathur, A., 2006. The use of small training sets containing mixed pixels for accurate hard image classification: Training on mixed spectral responses for classification by a SVM. In: *Remote Sensing of Environment*, 103(2), pp. 179–189. doi: 10.1016/j.rse.2006.04.001.
- Ishii, Y., Bagan, H., Iwao, K. and Kinoshita, T., 2018. Comparison of land cover classification using rough set theory. In: *39th ACRS 2018 PROCEEDING*.
- Japan Remote Sensing Society (Ed.), 2011. Remote Sensing Understanding from Basics. *Rikoh Tosho*, pp.242-245 (in Japanese).
- Landgrebe, D., 1997. Multispectral Data Analysis: A Moderate Dimension Example©. In: *Electrical Engineering*, pp. 1–10.
- Landgrebe, D. and Biehl, L., 1994. MultiSpec © A Freeware Multispectral Image Data Analysis System Documentation for MultiSpec. Purdue Research Foundation <https://engineering.purdue.edu/~biehl/MultiSpec/documentation.html> (9 January 2019).
- Mori, N., Tanaka, H. and Inoue, K. (Eds), 2004. Rough sets and kansei. *Kainundo*, pp.131-181 (in Japanese).