Land Cover Clustering of NDVI-drived Phenological Features

DongKeun Kim*, MyoungSeok Suh**, KyoungYoon Park***
*Department of Computer and Information Science, SeoNam University, Namwon ,Chonbuk, 590-170, Korea
Email: dgkim@tiger.seonam.ac.kr Tel: 0671-620-0276
**Department of Atmospheric Science, Seoul National University, Seoul, Korea
***ETRI, Taejon, Korea

Abstract

In this paper, we have considered the method for clustering land cover types over the East Asia from AVHRR data. The feature vectors such that maximum NDVI, amplitude of NDVI, mean NDVI, and NDVI threshold are extracted from the 10-day composite by maximum value composite(MVC) for reducing the effect of cloud contaninations.

To find the land cover clusters given by the feature vectors, we are adapted the self-organizing feature map(SOFM) clustering which is the mapping of an input vector space of n-dimensions into a one - or two-dimensional grid of output layer. The approach is to find first the clusters by the first layer SOFM and then merge several clusters of the first layer to a large cluster by the second layer SOFM.

In experiments, we were used the 8-km AVHRR data for two years(1992-1993) over the East Asia.

1. Introduction

Normalized difference vegetation index(NDVI) derived from AVHRR channels have been used to apply in large-scale area both for land-cover classification and for land-cover change detections[1,2,3,4]. NDVI depends on vegetation characteristics such as types , growing periods, and biomasses.

In this paper, we proposed the method for clustering land cover types using NDVI features derived from the AVHRR data. The NDVI-drived features are maximum NDVI, amplitude of NDVI, mean NDVI, and NDVI threshold, which are extracted from the 10-day composite NDVI by maximum value composite(MVC) for reducing the effect of cloud contaninations. To find the land cover clusters given by the 4-dimensional feature vectors, we are adapted self-organizing feature map(SOFM) clustering which is the mapping of an input vector space of n-dimensions into a one - or two-dimensional grid of output layer[7,8,9].

In experiments, we were used the 8-km AVHRR data for two years(1992-1993) over the East Asia.

2. Data and Methodology

Features Extraction

NDVI derived AVHRR has been use to estimate and monitor the characteristics such as biomass, productivity, leaf area, and percent vegetative ground cover. We used the transformed value by equation (1) to minimize the memory and the range of transformed values are from 0 to 200.

$$NDVI = \frac{CH_2 - CH_1}{CH_2 + CH_1} * 100 + 100 \qquad (1)$$

where CH_1 and CH_2 are reflectance value of channel 1 and channel 2, respectively.

We generated the daily NDVI by equation (1) from the 8-km AVHRR data. And then, to reduce the effect of cloud contaninations, the 10-day composite NDVI dataset were generated using maximum value composite(MVC) which is to obtain the maximum value at each pixel during the 10-day period.[1].

The four phenological features are extracted from the 10-day composite NDVI by MVC.

- Maximum NDVI : NDVI value at the time of peak greenness
- Mean NDVI : mean NDVI value of the 10-day composite data
- NDVI Amplitude: the difference between maximum and minimum NDVI
- NDVI threshold: (the number of composite data which is greater than NDVI threshold)*10.

Maximum and Mean NDVI are features to classify between the area with low NDVI value such as bare soil and water and the area with high NDVI value such as cultivation area and forest. NDVI Amplitude and threshold are used to classify between forest and agriculture, between a coniferous forest and a deciduous forest, etc. since they have seasonality about a growing season.

We produced the two phenological feature data dataset: (1) the one is generated from the annual 10-day composite NDVI(1992), (2) the another is from the composite for a two-year period(92-93).

Table 1. Statistics of NDVI-drived Phenological Features(1992)

	Minimum	Maximum	Mean	Std.dev
Maximum NDVI	98	181	154.31	19.19
Mean NDVI	95	170	133.11	14.02
NDVI Amplitude	101	181	141.27	19.29
NDVI threshold	0	150	136.79	38.59

Table 2. Statistics of NDVI-drived Phenological Features(1992-1993)

	Minimum	Maximum	Mean	Std.dev
Maximum NDVI	99	184	158.23	19.76
Mean NDVI	95	168	132.89	13.48
NDVI Amplitude	102	187	150.02	19.44
NDVI threshold	0	160	142.33	36.71

SOFM Clustering

Self Organizing Feature Map(SOFM) proposed by Kohonen is based on unsupervised, competitive learning which provides a topology preserving mapping from the high dimensional space to a one - or two-dimensional lattice. The weight vectors update rule is the following:

$$W_i(t+1) = W_i(t) + h_{ci}(t)[X - W_i(t)] \text{ if } i \in N_c(t) \quad (2)$$

where h_{ci} is a gaussian weighting function,

$$h_{ci}(t) = \alpha_t \exp(-\| r_i - r_c \|^2 / \sigma_t^2).$$

where r_c is the coordinates of the best matching(winning) node which is the node with smallest Euclidean distance. $N_c(t)$ is a neighborhood of winning node c. $N_c(t)$, α_t and σ_t are decreasing functions of time.

To contain more than one output node in a cluster, we used the two layer SOFM which is merge several clusters of the first layer SOFM to a large cluster by the second layer. While a one-layer SOFM alwayes forms convex clusters given by the Voronoi cells, the clusters given by the two-layer SOFM are unions of convex volumes[9].

We removed a small cluster which the number of points in a cluster are less than the predefined threshold. That is accomplished by finding the next best matching node.

input vectors → first layer SOFM(n×n) → the second SOFM (k×k)

Fig. 1. the two-layer SOFM clustering

3. Experimental Results

In our experiments, we have considered land cover clustering over the East Asia from 8-km AVHRR data for two years(92-93). The water areas are excluded in input vectors. We used a two-dimensional SOFM, the output node of the first layer is 11×11(the second layer, 4×4). The weight vectors initialized randomly within $\pm 1\sigma$ of the centroid of feature space. The initial learning rate is $\eta = 0.9$ and the initial neighborhood function $N_c(t)$ is determined by the radius $R=5$ ($R=2$).

To train the weight vectors, we have sampled to a spatial resolution of interval 4×4 pixels from four input feature images(530×535). And then, in clustering stage, the whole resolution of images is assigned to the best matching node.

Fig. 2 shows the clustering result using feature data I(1992), and fig. 3 shows the result using feature data II(1992-1993). Table 3 and table 4 are statistics of clustersing results in Fig.2 and Fig.3, respectively.

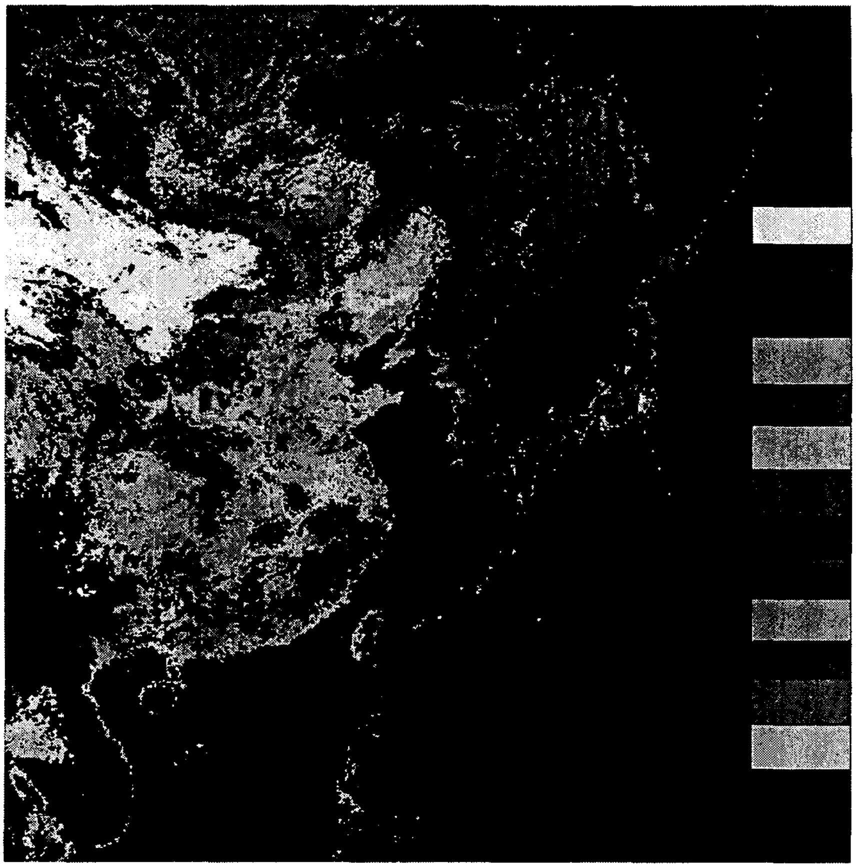

Fig.2. The clustering result using feature data I(1992)

Table 3. Statistics of clustersing results in Fig.2

Clusters	mean	std.	# of points
0	107.0, 104.7, 104.1, 0.0	2.01, 1.87, 2.04, 0.00	9345
1	156.5, 124.4, 156.1, 133.1	4.08, 3.52, 4.42, 5.17	7757
2	169.3, 135.2, 165.4, 141.1	4.19, 5.15, 4.64, 7.32	28591
3	112.5, 106.8, 108.9, 106.1	3.13, 1.73, 5.34, 2.07	1417
4	140.0, 116.1, 139.6, 130.0	5.67, 3.10, 6.86, 5.34	2895
5	156.9, 130.3, 149.5, 149.1	4.69, 3.78, 3.75, 4.45	6912
6	167.8, 143.1, 152.0, 154.3	4.78, 4.31, 4.38, 1.67	18987
7	119.8, 111.1, 114.9, 127.0	5.42, 2.44, 6.05, 8.82	7274
8	141.9, 121.4, 137.9, 143.2	4.12, 2.70, 4.21, 3.75	3269
9	155.8, 134.6, 140.6, 154.4	3.75, 4.34, 3.08, 1.56	12705
10	164.6, 146.8, 140.1, 155.0	4.37, 3.87, 3.57, 0.00	8238
11	130.4, 117.6, 122.9, 147.3	4.95, 2.71, 5.86, 5.05	3017
12	145.6, 129.0, 129.2, 154.7	3.99, 4.41, 4.83, 1.16	6079
13	153.7, 140.7, 126.6, 155.0	3.15, 3.52, 6.28, 0.00	7482
14	165.4, 154.6, 124.0, 155.0	4.58, 4.41, 6.58, 0.00	12305

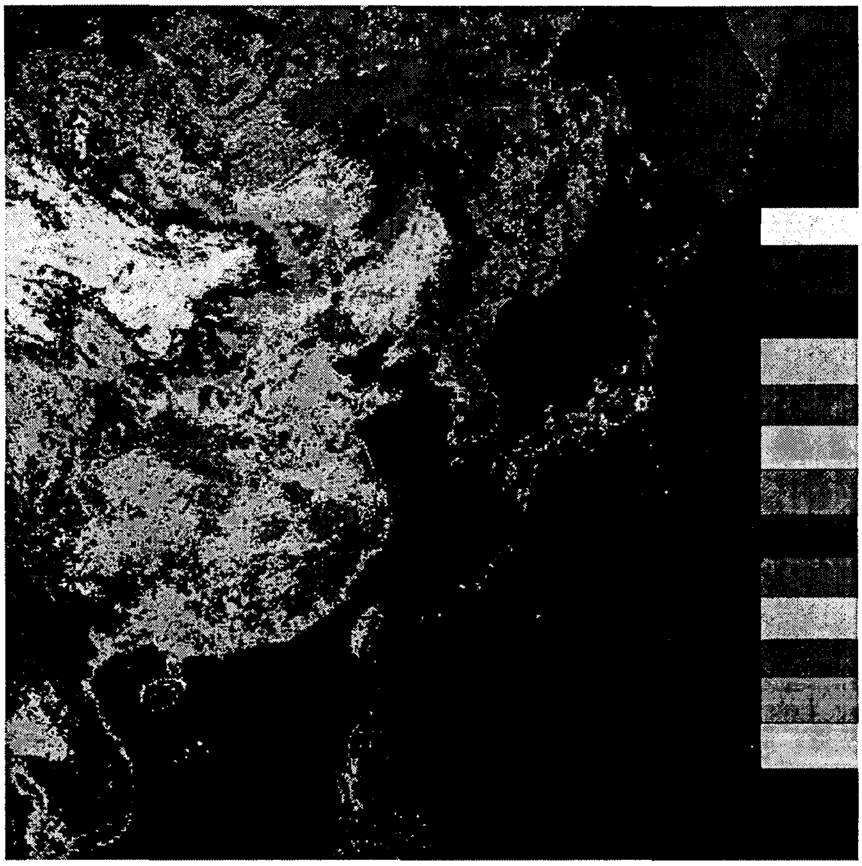

Fig.3. the result using feature data II(1992-1993)

Table 4. Statistics of clustering results in Fig.3

Clusters	mean	std.	# of points
0	107.3, 104.3, 105.3, 0.0	1.83, 1.85, 1.64, 0.00	7659
1	168.7, 130.7, 168.5, 140.8	3.81, 3.64, 4.98, 6.93	15680
2	175.3, 140.9, 171.3, 151.7	3.37, 5.0511, 3.97, 5.66	22439
3	112.2, 106.4, 109.3, 103.4	1.77, 1.53, 3.53, 1.60	1858
4	156.0, 120.0, 158.1, 132.4	5.98, 4.27, 6.39, 5.21	4193
5	169.5, 144.2, 154.6, 159.6	4.83, 5.32, 3.34, 0.97	11092
6	169.8, 142.6, 162.9, 157.7	5.18, 4.04, 2.83, 2.26	12761
7	123.3, 110.9, 119.5, 127.6	8.50, 2.96, 10.22, 11.35	8737
8	147.3, 122.3, 145.6, 148.7	4.36, 3.23, 4.76, 4.11	3179
9	159.6, 133.5, 149.7, 158.5	3.90, 3.90, 3.18, 1.81	12016
10	159.3, 130.9, 157.9, 152.3	3.93, 4.80, 2.86, 5.19	8561
11	134.1, 118.0, 126.8, 152.7	5.81, 2.89, 6.36, 5.83	4052
12	150.3, 127.9, 138.8, 158.7	4.30, 3.53, 4.90, 3.80	7114
13	158.7, 139.9, 138.7, 159.9	3.03, 3.46, 5.19, 0.09	5955
14	169.7, 152.2, 142.1, 159.9	4.52, 4.30, 6.95, 0.37	10977

4. Conclusions

We have considered the method for clustering land cover types by two-layer SOFM over the East Asia from the 8-km AVHRR data. The feature vectors such that maximum NDVI, amplitude of NDVI, mean NDVI, and NDVI threshold is extracted from the 10-day composite by maximum value composite for reducing the effect of cloud contaninations. We first found out 121 clusters by the first layer SOFM and then merge clusters into a larger cluster by the second layer SOFM. Our method is simple and efficient. The clustered data analysis using ground truth data and land-cover change detections will be a feature work.

References

[1] Gutman, G. G., "Vegetation indices from AVHRR, An update and feature prospects," Remote Sens. Environ. , 35, pp.121-136, 1991.

[2] Holben, B.N., "Characteristics of maximum-value composite images from temporal AVHRR data," Int. J. Remote Sens. , 7, pp.1417-1434, 1986.

[3] Lambin, E.F. and D.Ehrlich, " The surface temperature-vegetation index space for land cover and land-cover change analysis," Int. J. Remote Sens., 17, pp.463-487, 1996.

[4] MyoungSeok Suh, Kwangmi Jang, and DongKeun Kim, "Land Cover Classification over Korean Peninsular Using Phenological Data Derived from NOAA/AVHRR," Proceedings of international Symposium on Remote Sensing, 6th Annual Workshop of EMSEA, pp.39-44, Nov.5-7, 1997.

[5] MyoungSeok Suh, Kwangmi Jang and KyoungYoon park, "Examination of Compositing Techniques Using NOAA/AVHRR Data," Proceedings of international Symposium on Remote Sensing, 6th Annual Workshop of EMSEA, pp.58-65, Nov.5-7, 1997.

[6] A.K. Jain and R.C.Dubes, Algorithms for Clustering Data, Prentice Hall, 1988.

[7] A.S.Pandya and R.B.Macy, Pattern Recognition with Neural Networks in C++, CRC Press, 1996.

[8] Jari A. Kangas, Teuvo K. Kohonen, and Jorma T.Laaksonen, "Varients of Self-Organizing Maps," IEEE Trans. on Neural Networks, vol.1, no.1, pp.93-99, 1990.

[9] Jouko Lampinen, and Erkki Oja, " Distortion Tolerant Pattern Recognition Based on Self-Organizing Feature Extraction," IEEE Trans. on Neural Networks, vol.6, no.3, pp.539-547,1995.