

러프 집합을 이용한 다중 분광 이미지 데이터의 분류

† 원성현*

‡ 이병성

‡ 정환목

† 지산전문대학 전산정보처리과
부산광역시 금정구 부곡3동 9번지

‡ 대구효성가톨릭대학교 전자·정보공학부
경북 경산군 하양읍 금락1리

Classification of Multi Spectral Image Data using Rough Sets

† Sunghyun Won*

‡ Byoungsung Lee

‡ Hwanmook Chung

† Dept. of Computer Info. Processing
Jisan Junior College

‡ Faculty of Electronic & Information Engineering,
Catholic Univ. of TaeguHyosung

Abstract

Traditionally, classification of remote sensed image data is one of the important works for image data analysis procedure. So, many researchers devote their endeavor to increasing accuracy of analysis, also, many classification algorithms have been proposed. In this paper, we propose new classification method for remote sensed image data that use rough set theory. Using indiscernibility relation of rough sets, we show that can classify image data very easily.

1. 서론

컴퓨터를 이용한 원격 탐사 이미지 데이터에 대한 처리 기술은 1972년 7월 23일 미 항공 우주국(NASA)의 ERTS-1이 발사된 후, 인공위성으로부터 수집된 데이터를 환경 감시, 국토 개발, 자원 탐사, 군사 동향 파악 및 농업 생산량 추정 등 여러 분야에 그 실용 가능성을 보여주었으며, 이러한 위성 이미지 분석 기술의 연구는 계속 발전되어 왔다[1,3].

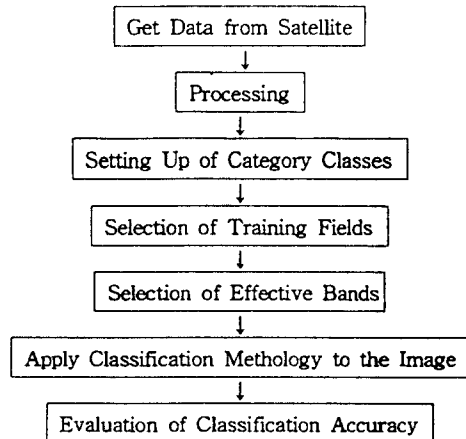
초창기 지표물 효과적으로 관리하기 위한 데이터로 주로 사용되었던 것은 항공 사진 등이었다. 그러나, 이것은 각국의 영공권 주장 등과 같은 국제법 문제에 부딪히게 되어 최근에는 인공위성에 의한 원격 탐사 데이터가 주로 이용되고 있는 실정이다[1].

원격 탐사 데이터에 대한 분류가 일반적인 화상 데이터의 분류와는 달리 더 어렵게 느껴지는 이유는 원격 탐사 데이터는 다중 분광 이미지 데이터이기 때문이다. 다중 분광 데이터란 탐사되는 대상을 하나의 파장대에서만 촬영하지 않고 여러 파장대에서 동시 병렬적으로 촬영된 데이터를 의미한다. 현재 인공 위성에서 주로 사용되는 Landsat TM 센서의 경우는 7개의 밴드로 대상을 탐사한다. 단일 밴드로 하지 않고 센서의 성능을 지속적으로 개선시켜 밴드의 수를 늘리는 이유는 탐사되는 데이터의 분석 정확도를 높이기 위함이다. 일반적으로 탐사 대상이 되는 객체들은 모두 빛에 의해 반사되는 분광 특성이 다르다. 즉, 숲(forest), 곡물(crop), 수역(water), 도시(urban) 등은 모두 빛에 대해 서로 다른 반사율을 보이기 때문에 단일 밴드 상에서 수집되었을 때 대상이 숲인 경우는 의미있는 데이터라 하더라도 도시 지역인 경우는 오차가 많이 발생할 수 있다. 그래서, 탐사 대상의 분광 특성들을 분석하고 밴드의 수를 가능하면 늘려서 실제 데이터의 분석에서는 분류 대상의 분광 특성을 잘 반영한 특정 밴드들을 선별하고 이들을 조합하여 분류를 수행하는 것이 분류 정확도를 높이는 한 방법이 될 수 있다. 그러므로, 복수개의 밴드를 조합하여 단일 범주로 귀속시키는 분류 작업은 매우 중요하고 이에 따라 위성 데이터의 신뢰성이 좌우된다.

본 논문에서는 다중 밴드로 수집된 위성 데이터에 대한 분류를 위해 러프 집합 이론을 도입한다. 이를 통해 러프 집합의 식별 불능 관계(indiscernibility relation)가 위성 데이터의 분류에도 이용될 수 있음을 보인다.

2. 기존의 이미지 분류 기법

일반적으로 위성을 이용한 원격 탐사의 연구는 2가지 분야가 있는데 첫째는 위성의 장치에 관한 연구이며 둘째는 위성으로부터 취득한 데이터의 분석을 통한 실용적 응용 기술에 관한 연구이다. 후자인 위성 데이터의 처리 개요는 다음과 같다[1,3].



(그림 2-1) 위성 이미지 분류를 위한 처리 과정

전통적으로 사용되어오던 원격 탐사 다중 분광 이미지 데이터에 대한 분류는 크게 교사 분류와 비교사 분류 기법으로 나뉘어 진행되어왔다.

2.1 교사 분류[3]

교사 분류란 분류 항목에 해당하는 학습 데이터를 그 항목의 대표적 분광 특징을 갖는다는 가정 하에 분류 과정의 지표로 삼는 기법을 말한다.

최초, 학습 단계에서는 분류 항목에 따른 학습 데이터를 선정하여 그의 다중 분광 밴드에 대한 화소값들을 입력하여 학습시킨 후, 분류 단계에서는 분류하고자 하는 후보 화소들을 앞서 학습된 분류 항목과 가장 유사한 항목으로 분류한다. 마지막으로 출력 단계에서는 분류 결과를 지도 형태나 데이터 파일 또는 컬러 이미지 등으로 출력하게 되는데 중요한 점은 분류에 앞서 학습 단계를 거치는 것이다.

이러한 교사 분류 기법은 학습 데이터의 각 밴드에 대한 상한값과 하한값을 구하여 그 범위 내에서 분류를 하는 방법인 평행육면체 분류(Parallelepiped Classification), 화소가 속하는 각 분류 항목의 평균 백터와 그로부터의 거리를 계산하여 가장 가까운 항목에 귀속시키는 최소 거리 분류(Minimum Distance Classification), 분류 항목의 분포가 정규 분포를 이룬다는 가정 하에 공분산을 이용하여 최소 거리 분류를 도입한 Mahalanobis 거리 분류, 임의의 화소가 각 범주에 속할 경우에 발생하는 확률 밀도 함수를 모든 분류 항목에 대해 구하고 그 중 확률이 가장 커지는 경우의 범주로 결정하는 최대 공산 분류(Maximum Likelihood Classification) 등이 있다. 이 중, 마지막의 최대 공산 분류가 가장 일반적으로 사용되고 있다.

2.2 비교사 분류[3]

비교사 분류는 군집화라고도 하는데 이는 분석자의 분류 항목에 대한 정의에 따라 결합 혹은 분리시켜 같은 속성을 갖는 데이터들끼리 무리지어주는 방법이다.

이러한 비교사 분류 기법은 분류하고자하는 이미지의 한쪽에서 순서대로 분석되 화소 간의 거리가 임계값보다 크면 병합하고, 작으면 분리하는 방법으로 군집화하는 순차적 군집, 초기에 지정된 군집 중심을 기준으로 각 화소를 가장 가까운 군집에 소속됨을 가정하고 그로부터 각 군집의 평균을 새로 계산하여 중심값을 조금씩 변화시키면서 군집화하는 ISODATA 군집 등이 있다.

3. 인공지능을 이용한 분류 기법

2절에서 소개된 대부분의 분류 기법들은 공분산, 유클리드 거리, 확률 밀도 함수, 정규 분포 등의 통계 이론을 기본 원리로 하고 있다. 그러나, 최근 신경망, 퍼지 이론 등의 새로운 인공지능 기법들이 응용 면에서 만족할 만한 결과를 보이면서 이들을 위성 데이터의 분류에도 널리 활용되고 있다. 1993년 Sahu 등에 의해 제안된 이진 다이아몬드 신경망에 의한 분류는 신경망을 이용한 학습에 의한 분류라는 특징이 있긴 하나 교사 분류의 한 종류인 최대 공산 분류보다 좋은 결과를 보이지는 못했다[2]. 그러나, 기존의 통계적 기법들이 데이터가 정규 분포를 이루는 특징이 있을 때 효과적이라는 한계가 있지만 Sahu 등에 의한 방법은 불규칙 데이터 분포 상태에서도 분류가 가능하다는 진일보한 면을 보인데 큰 의미가 있다. 역전파 신경망을 이용한 분류도 여러 연구자들에 의해 제안되어 비교적 용이한 연산 과정에 의해 분류가 수행될 수 있음을 보였다[4,5]. 또한, 분류를 위해 분류 공간을 퍼지 부공간으로 나누어 퍼지 규칙을 자동 생성한 후 이를 통해 분류 정확도를 높이는 연구도 진행되었다[1].

앞서 언급했듯이 인공지능 기법을 이용한 대부분의 분류 기법들이 갖는 공통적인 특징은 불규칙 분포 상태의 데이터들에 대해서도 효과적으로 분류할 수 있다는 것이

다.

4. 러프 집합을 이용한 분류

4.1 러프 집합에서의 정의[6,7]

러프 집합에서 속성 값을 기술하는 함수 P_X 는 다음과 같이 정의한다.

[정의 1]

$$P_X: O \rightarrow V; P_X(q) = P(x, q)$$

또한, 대상 x_i 와 x_j 가 어떤 속성 집합 O 에 관하여 식별할 수는 대상들임을 나타내는 식별 불능 관계 $\text{ind}(O)$ 를 다음과 같이 정의한다.

[정의 2]

$$P(x_i, O) = P(x_j, O)$$

$$(x_i, x_j) \in \text{ind}(O)$$

$$R = \text{ind}(O)$$

따라서, x_i, x_j 가 $P \subset O$ 인 P 에 대하여 식별할 수 없는 경우 다음과 같이 나타낸다.

$$(x_i, x_j) \in \text{ind}(P)$$

$$\text{ind}(P) = \bigcap_{R \in P} \text{ind}(q)$$

특히, $P=O$ 인 경우 $(x_i, x_j) \in \text{ind}(O)$ 라면 x_i 와 x_j 는 정보 시스템에서 식별할 수 없는 것으로 한다.

$\text{ind}(q)$ 는 동치 관계이기 때문에 이것에 의한 X 의 분할은 상집합이 된다.

$$X/\text{ind}(q) = \{[x_i] \mid x_i \in X\}$$

4.2 러프 집합을 이용한 분류 절차

러프 집합을 이용한 다중 분광 이미지 데이터의 분류를 위해 다음과 같은 참조표 [표 4-1]이 있다고 가정하자. 이 참조표에 대한 해석은 다음과 같이 수행한다. 즉, v_{11} 은 밴드 1(B_1)에서 범주 1(C_1)로 분류되기 위한 방사량 값이다. 그러나, 하나의 범주가 항상 하나의 값으로 수집되는 것만은 아니기 때문에 각 밴드별로 같은 범주로 귀속되는 방사량 값들의 표준편차(σ)를 그 오차의 한계로 두기로 한다.

다음 [표 4-2]는 센서에 의해 수집된 각 화소의 밴드별 방사량 값이다. 이 표에 대한 해석은 다음과 같이 한다. p_{v11} 은 화소 P_1 을 밴드 1(B_1)에서 수집한 실제 방사량 값이다.

[표 4-1] 범주와 밴드 간의 참조표

범주(C) \ 밴드(B)	C_1	C_2	...	C_n
B_1	$v_{11} \pm \sigma_1$	$v_{12} \pm \sigma_2$...	$v_{1n} \pm \sigma_n$
B_2	$v_{21} \pm \sigma_1$	$v_{22} \pm \sigma_2$...	$v_{2n} \pm \sigma_n$
⋮	⋮	⋮	⋮	⋮
B_m	$v_{m1} \pm \sigma_1$	$v_{m2} \pm \sigma_2$...	$v_{mn} \pm \sigma_n$

(단, n 은 범주의 수, m 은 밴드의 수, v_{ij} 는 i 번째 밴드의 범주 j 에 대한 방사량 값, σ_j 는 범주 j 로 분류되는 화소의 각 밴드별 방사량 값의 표준편차)

[표 4-3]은 [표 4-1]과 [표 4-2]를 조합하여 탐사된 화소가 어떤 범주에 귀속되는지를 나타낸다. 이 때 가장 적절한 범주를 선택하는 원칙은 첫째, 탐사된 방사량이 [표 4-1]에서 제시된 범위 내에 속하면 그 범주로 귀속시키고, 둘째, 속하는 범위가 없다면 가장 가까운 범주로 귀속시킨다. 둘째의 경우가 발생하는 이유는 센서의 공간 해상 능력에 따라 하나의 화소 내에 두 개 이상의 범주가 혼재한 경우가 실제 존재하기 때문이다. 이 표에 대한 해석은 C_i의 경우 탐사된 화소 P₁이 [표 4-1]의 참조표에 의해 범주 i일 가능성이 가장 높음을 의미한다.

[표 4-2] 화소들의 밴드별 방사량 값

화소(P) \ 밴드(B)	P ₁	P ₂	...	P _k
B ₁	pv ₁₁	pv ₁₂	...	pv _{1k}
B ₂	pv ₂₁	pv ₂₂	...	pv _{2k}
⋮	⋮	⋮	⋮	⋮
B _m	pv _{m1}	pv _{m2}	...	pv _{mk}

(단, m은 밴드의 수, k는 탐사한 화소의 수, pv_{ij}는 i번째 밴드에서 수집한 화소 j의 방사량 값)

[표 4-3] 탐사된 화소들이 귀속될 범주

화소(P) \ 밴드(B)	P ₁	P ₂	...	P _k
B ₁	C ₁	C ₁	...	C _k
B ₂	C ₁	C ₁	...	C _k
⋮	⋮	⋮	⋮	⋮
B _m	C ₁	C ₁	...	C _k

마지막으로 러프 집합의 식별 불능 관계 (indiscernibility relation) ind(B)를 참고문헌 [8]의 방법으로 분류한다.

5 적용 예

5.1 다중 밴드 상의 문제 설정

다중 분광 밴드 상에서의 다음과 같은 참조표[표 5-1]이 오랜 실험에 의해 만들어졌다고 가정하자.

[표 5-1] 범주와 밴드간의 참조표

범주 \ 밴드	1	2	3	4	5	6	7	8
1	10	21	29	40	52	61	72	80
2	9	20	30	44	51	58	70	80
3	10	22	30	38	52	57	73	80
4	9	20	31	41	52	60	70	82
5	11	21	31	41	52	59	71	83
6	10	21	31	40	53	55	72	83
7	9	22	32	39	50	60	70	82
표준편차	0.7	0.8	0.9	1.8	0.9	1.9	1.1	1.3

그리고, 탐사에 의해 [표 5-2]가 수집되었다고 가정하자.

[표 5-2] 화소들의 밴드별 방사량 값

화소 \ 밴드	1	2	3	4
1	31	67	55	40
2	29	70	54	65
3	37	69	55	38
4	29	66	57	41
5	28	72	59	41
6	27	70	55	40
7	32	68	50	39

5.2 러프 집합을 이용한 분류와 해석

5.1의 [표 5-1]과 [표 5-2]를 통해 탐사된 화소들이 귀속될 가능성이 가장 높은 범주를 구하면 다음 [표 5-3]과 같다. 셀 내의 수치는 범주 번호를 나타낸다.

[표 5-3] 탐사된 화소들이 귀속될 범주

화소 \ 밴드	1	2	3	4
1	3	7	5	4
2	3	7	5	7
3	4	7	5	4
4	3	7	6	4
5	3	7	6	4
6	3	7	6	4
7	3	7	5	4

[표 5-3]을 화소 1, 2, 3, 4에 대한 러프 집합의 식별 불능 관계 개념을 이용하여 다음과 같이 분류한다.

$$P/\text{ind}(P_1) = \{[1, 2, 4, 5, 6, 7], [3]\} \text{ ---- (5-1)}$$

$$P/\text{ind}(P_2) = \{[1, 2, 3, 4, 5, 6, 7]\} \text{ ---- (5-2)}$$

$$P/\text{ind}(P_3) = \{[1, 2, 3, 7], [4, 5, 6]\} \text{ ---- (5-3)}$$

$$P/\text{ind}(P_4) = \{[1, 3, 4, 5, 6, 7], [2]\} \text{ ---- (5-4)}$$

$$P/\text{ind}(P_1, P_2) = \{[1, 2, 4, 5, 6, 7], [3]\} \text{ --- (5-5)}$$

$$P/\text{ind}(P_1, P_3) = \{[1, 2, 7], [3], [4, 5, 6]\} \text{ -- (5-6)}$$

$$P/\text{ind}(P_1, P_4) = \{[1, 4, 5, 6, 7], [2], [3]\} \text{ -- (5-7)}$$

$$P/\text{ind}(P_1, P_2, P_3, P_4) = \{[1, 7], [2], [3], [4, 5, 6]\} \text{ ---- (5-8)}$$

(5-1)의 경우 밴드 1, 2, 4, 5, 6, 7은 화소 1에 대해 모두 같은 범주로 귀속되기 때문에 이중 어떤 밴드들의 조합을 활용해도 무방하다. 그러나, 밴드 3은 이들과 서로 다른 범주로 귀속되기 때문에 조합해서는 안된다. (5-6)의 경우, 화소 1과 화소 3을 동시에 고려해야 하는 경우 밴드 1, 2, 7 내에서의 밴드 4, 5, 6 내에서는 조합 가능하지만 이들 간에는 조합 불가능하다. (5-8)의 경우 화소 1, 2, 3, 4가 모두 동시에 고려되어야 하는 경우, 밴드 1과 7은 조합 가능, 밴드 4, 5, 6은 조합 가능 그러나 이들과 밴드 2, 밴드 3은 상호 조합 불가능하게 된다.

[표 5-3]을 밴드 1, 2, 3, 4, 5, 6, 7에 대하여 동일한 방법으로 분류하면 밴드 2에 대해서만 화소 2와 4가 식별 불능 관계가 되고 나머지 모든 화소들은 모든 밴드에 대해 식별 가능 즉, 서로 다른 범주로 분류될 수 있음을 보인다.

6. 결론

러프 집합은 주로 근사성, 식별 불능 관계 등을 이용하여 속성 집합들을 분류하는 도구이다. 본 논문에서는 이 특징을 이용하여 각 밴드별로 서로 다르게 나타나는 방

사람을 통해 탐사 대상이 되는 화소들의 밴드를 손쉽게 분류해낼 수 있음을 보였다. 그러나, 본 논문에서 사용한 데이터는 Landsat-TM이나 SPOT-HRV 등의 실제 위성 센서에 의해 수집된 데이터는 아니고 가상으로 만든 임의의 데이터이다. 그럴 수밖에 없었던 이유는 실제 데이터를 사용하여 실험을 하게 되면 그 작업이 너무 방대하기 때문이다. 그러므로, 실제 데이터를 사용했을 때, 동일한 결과가 생성될 지에 대해서는 의문의 여지가 있다.

향후로는 실제 위성 데이터와 지형도를 이용해서 본 논문에서 제안하는 방법을 적용하여 만족할 만한 결과를 얻어내는 실험이 수반되어야 하고, 하나의 화소가 실제로 두 개 이상의 범주 특성으로 분석되었을 때, 과연 이것을 어떻게 분류할 것인지에 대한 문제(통상 이 문제를 혼합 화소(MIXEL)의 분류 문제라고 칭함)를 해결하기 위해 러프 집합을 이용할 수 있는지에 대한 연구를 진행해야 할 것이다.

7. 참고문헌

- [1] 김진일, "위성 탐사 이미지 분류를 위한 다중 분광 패턴 공간에서의 퍼지 규칙 생성에 관한 연구", 서강대학교 대학원, 박사학위 논문, 1994.
- [2] Y.Salu and J. Tilton, "Classification of Multispectral Image Data by th Binary Diamond Neural Network and by Nonparametric, Pixel - by - Pixel Methods", IEEE Trans. on Geoscience and Remote Sensing, Vol.31, No.3, May, 1993.
- [3] ERDAS Field Guide, ERDAS Inc., 1991.
- [4] H.Bischof, W.Schneider and A.J.Pinz, "Multispectral Classification of Landsat-Image Using Neural Networks", IEEE Trans. on Geoscience and Remote Sensing, Vol.30, No.3, May, 1992.
- [5] P.D.Heermann and N.Khazenie, "Classification of Multispectral Remote Sensing Data Using a Back Propagation Neural Network", IEEE Trans. on Geoscience and Remote Sensing, Vol.30, No.1, Jan., 1992.
- [6] Z.Pawlak, Rough Sets, Theoretical Aspects of Reasoning about Data, Kluwer Academic Publisher, 1991.
- [7] Z.Pawlak, "Rough Sets", International Journal of Computer and Information Science, 1982.
- [8] 정구범, 정관득, "러프 집합을 이용한 관계 데이터 베이스 모델의 구성 및 해석", 한국퍼지및지능시스템학회 추계 학술 대회 논문집, 6권 2호, 1996.