

클러스터간 중첩성과 평균내부거리를 적용한 퍼지 클러스터링에 의한 위성영상 분류

전영준*, 김진일*

*동의대학교 컴퓨터공학과

e-mail:j4017@chol.com

A Study on the Satellite Image Classification Based on the Fuzzy Clustering using Overlap Measure and Average Intracluster Distance

Young-Joon Jeon*, Jin-Il Kim*

**Dept. of Computer Engineering, Dongeui University

요 약

본 논문에서는 평균내부거리를 적용한 퍼지 클러스터링 알고리즘과 클러스터들 사이의 중첩성을 이용한 위성영상의 분류 알고리즘을 제안하였다. 제안된 방법은 클러스터의 크기에 따라 큰 클러스터에는 많은 소속정도를 작은 클러스터에는 적은 소속정도를 부여함으로써 크기가 다른 클러스터가 존재하는 데이터 집합에 대해서도 분류의 효율성을 높였다. 클러스터들간의 중첩성을 이용한 평가를 통해 위성영상에 있어서 중첩되지 않은 화소는 각각의 분류항목에 포함시키고, 중첩된 화소들은 최대우도 분류를 수행한 후 각 화소에 대한 우도와 퍼지 클러스터링의 소속도를 비교 분석하여 최종 분류항목을 결정함으로써 분류를 효율적으로 할 수 있다.

1. 서론

원격탐사에 의한 정보는 토지의 체계적인 활용, 환경오염의 감시 및 통제, 지도제작 등에 요구되는 복잡 다양한 정보를 신속하고 정확하게 분류, 해석 하는데 있어서 커다란 잠재력을 가지고 있는 최첨단 과학기술로 각광받고 있으며, 그 활용 분야는 환경에서부터 토목, 농업, 해양, 지질, 임업, 수산업 등 각 분야에 걸쳐 급속도로 확산되고 있다. 최근 위성 탐측기의 급속한 발전에 의해 해상력이 향상되고 자료획득 시간이 단축됨에 따라 이용분야가 확대되고 있다. 이러한 정보의 추출을 위해 가장 자주 이용되는 방법은 패턴인식을 이용한 다중분광 분류이다. 영상의 분류는 영상을 구성하는 각각의 화소들 중 비슷한 분광 특성을 갖는 것끼리 집단화시켜 주는 방법으로서 일반적으로 이와 같이 구성된 각각의 집단을 분류항목이라고 한다. 영상의 각 화소는 분류를 수행한 후에 하나의 분류항목 또는 값을 갖게 된다. 이것은 영상으로부터 관심 대상인 정보를 추출

하기 위해 수행되는 것으로 주로 토지피복, 표면 온도, 대기 상태 등의 환경상태 등 광범위한 지역을 몇 종류 또는 몇 단계로 분류할 수 있는 정보를 추출하는 데 사용된다.

지금까지 연구되어 왔던 위성영상분류를 위한 기법에는 무감독 분류기법은 클러스터링(군집화) 과정이라고 하며, 분석자가 몇 개의 간단한 변수만을 지정하여 분류가 실시되며, 군집에는 군집의 중심, 밴드간의 공분산 등으로써 분광적 분포 특성이 표현된다. 영상에 대한 사전 지식을 가지지 못한 경우 적용이 가능한 분류기법이다. 감독분류 기법은 분석자가 영상 내에서 알고 있는 화소들을 지정하면, 각 군집의 중심, 밴드간 공분산 등을 이용하여 자료내의 모든 화소를 분석하여 가장 유사한 분포 특성을 가지는 군집에 할당시키는 과정이다[1]. 위성영상분류의 최근 연구 동향은 1980년대 후반에 들면서 인공지능, 퍼지, 신경망 이론이 본격적으로 등장하면서 기존의 통계적 이론에서 보다 개선된 제안들이 속속

소개되고 있다.

본 연구에서는 클러스터에 있어서 평균내부거리를 적용한 퍼지 클러스터링 알고리즘과 클러스터들 사이의 중첩성을 고려한 최대우도 분류를 이용하여 위성영상의 분류를 수행하는 알고리즘을 제안하였다.

2. 관련연구

2.1 퍼지 C-Means 클러스터링

Bezdek에 의해 제안된 퍼지 C-Means 알고리즘은 목적함수기반(objective function based) 퍼지 규칙을 사용한다[2]. 퍼지 C-Means 알고리즘의 목적은 주어진 데이터 집합을 주어진 수의 퍼지 클러스터로 분할하는 것이다. 주어진 데이터 집합 $X = x_1, \dots, x_n$ 와 분할하고자 하는 군집 수 c 에 대해서 아래의 목적함수 J_m 을 최소화함으로써 퍼지분할 $F = F_1, F_2, \dots, F_c$ 를 구하는 것이다.

$$J_m(U, V; X) = \sum_{j=1}^n \sum_{i=1}^c u_{ij}^m \|x_j - v_i\|^2$$

여기서 u_{ij} 는 퍼지 클러스터 F_i 에 대한 x_j 의 소속정도를 나타내며, $(c \times n)$ 패턴행렬 $U = [u_{ij}]$ 의 원소가 된다. $V = v_1, \dots, v_n$ 는 퍼지 클러스터들의 중심 벡터의 집합이다. 퍼지 C-Means 군집분류 알고리즘은 각 화소들이 군집에 소속될 소속도와 군집중심을 생성한다. 따라서 위의 목적함수 J_m 의 값을 최소화시키는 u_{ij} 와 v_i 를 구하는 것이 목표이다. 여기에서 m 은 퍼지 식별자라고 하고 만약 $m=1$ 인 경우는 HCM이며, 퍼지 C-Means 알고리즘에서는 $1 < m < \infty$ 의 값을 사용하며 일반적으로 $m=2$ 값을 사용한다. 퍼지 C-Means 알고리즘에서는 목적함수 J_m 의 값을 최소화시키는 u_{ij} 와 v_i 를 다음과 같이 계산한다.

$$u_{ij} = \left[\sum_{k=1}^c \left(\frac{D_{ijA}}{D_{jkA}} \right)^{\frac{2}{m-1}} \right]^{-1}, \quad 1 \leq i \leq c; 1 \leq j \leq n,$$

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}, \quad i \leq i \leq c$$

$$D_{ij} = \|x_j - v_i\|_A > 0 \quad \forall i, j$$

3. 제안된 분류방법

본 연구에서는 위성영상에 대하여 분류항목에 따른 학습데이터를 선정하고, 선정된 학습데이터에 대하여 각각 평균값을 구하여 초기 중심값으로 설정하여 평균내부거리를 적용한 퍼지 클러스터링 알고리즘에 의해 클러스터링을 수행한 후에 구해진 클러스터에 대하여 클러스터들간의 중첩 정도를 계산한다. 중첩되지 않은 데이터는 각각의 분류 항목에 포함시키고, 중첩된 데이터들은 최대우도 분류를 수행한 후 각 데이터에 대한 우도와 퍼지 C-Means의 소속도를 비교 분석하여 최종 분류항목을 결정함으로써 최종분류결과를 얻는다. 그림 2는 제안된 분류 기법의 전체 수행 과정을 나타낸 것이다.

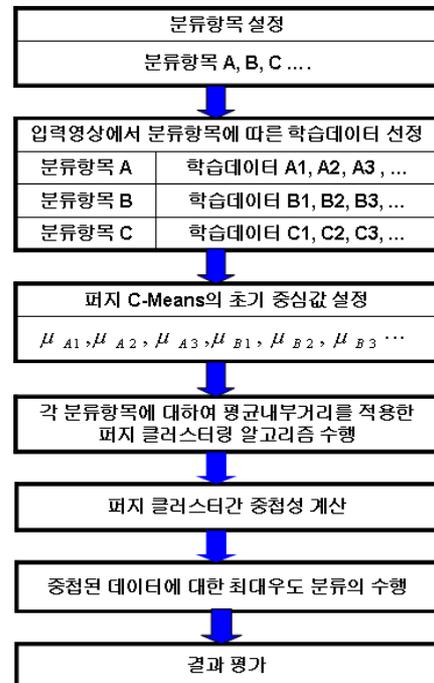


그림 1. 제안된 분류 기법의 전체 수행 과정

퍼지 C-Means 방법은 각 화소들이 클러스터의 중심에 속하는 소속정도를 부여함으로써 화소로부터 각 클러스터 중심까지의 거리가 같으면 같은 소속정도를 가진다. 그러므로 중심탐색에 있어 클러스터의 크기가 다른 경우 문제점이 발생한다. 이러한 문제점을 해결하기 위해 평균내부거리를 이용한 클러스터링 방법을 이용하였다. 각 화소가 내부클러스터에 속하는 정도를 소속정도로 줌으로 클러스터의 크기에 상관없이 중심값 탐색을 잘 할 수 있도록 하였다. 내부클러스터는 평균내부거리 안쪽에 속한 화소들의 집합을 의미한다. 그리고 내부클러스터의 크기는 평균내부거리에 의해 결정된다.

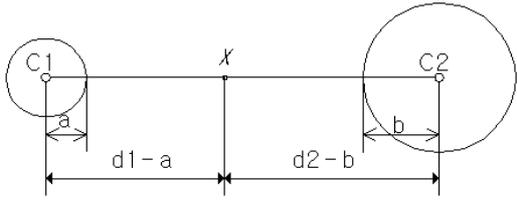


그림 2. 평균내부거리를 이용한 클러스터링 방법

제안된 알고리즘은 각 클러스터의 응집성을 고려한 퍼지 C-Means 알고리즘의 목적함수를 사용하였다. 그리고 이 목적함수를 최소화하는 소속정도를 다음처럼 정의한다.

$$u_{ij} = \left\{ \sum_{k=1}^c \left(\frac{D_{ik}}{D_{kj}} \right)^{\frac{2}{m-1}} \right\}^{-1}$$

$$D_{ij} = d_{ij} - \omega_i$$

여기서 D_{ij} 는 j 번째 화소로부터 i 번째 내부클러스터까지의 거리를 의미한다. ω_i 는 각 클러스터의 평균내부거리를 의미하며 클러스터의 전체 모양과 크기에 관계되며 ω_i 를 클러스터의 평균내부거리에 비례하도록 정의하였다. K 는 상수이다.[3][4][5]

$$\omega_i = K \frac{\sum_{j=1}^n u_{ij}^m d_{ij}^2}{\sum_{j=1}^n u_{ij}^m}$$

평균내부거리를 이용한 클러스터링 알고리즘의 과정은 다음과 같다.

단계1: $c(2 \leq c \leq n)$ 값과 $m(1 \leq m \leq \infty)$ 값을 결정한다.

단계2: 다음 조건을 만족하는 퍼지 c 분할 $U^{(0)}$ 을 초기화한다.

단계3: 각 클러스터에 대한 중심값 $v_i^{(l)}$ 을 구한다.

$$(l = 0, 1, 2, \dots)$$

단계4: $D_{ij} = d_{ij} - \omega_i$ 에서 각 화소로부터 각 내부클러스터까지의 거리 D_{ij} 를 구한다.

단계5: $u_{ij} = \left\{ \sum_{k=1}^c \left(\frac{D_{ik}}{D_{kj}} \right)^{\frac{2}{m-1}} \right\}^{-1}$ 에서 새로운 소속정도 $U^{(l+1)}$ 을 계산한다. 단, $\eta_i \neq d_{ij}$ 일

때

단계6: 만약 $\|U^{(l+1)} - U^{(l)}\| \leq \epsilon$ 이면 알고리즘을 종료하고 그렇지 않으면 단계 3으로 이동하여 반복 수행한다.

클러스터링 과정이 끝나면 분류항목별로 새롭게 생성된 클러스터들 사이에 중첩정도를 확인한다. 중첩성은 퍼지 C-Means 알고리즘에 의해 계산된 퍼지 클러스터들에 대한 평가 인덱스로 사용된다. 두 클러스터간 중첩정도가 낮을수록 공간에 잘 분포하게 되어 있다.

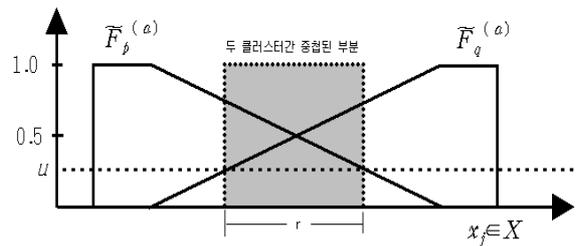


그림 3. 소속정도 μ 에서 두 클러스터간의 중첩성 계산

본 연구에서는 클러스터간 중첩된 화소를 판단하는데 이용하였다. 두 클러스터간 중첩된 화소를 판단하여 중첩정도가 아주 높으면 하나의 학습데이터로 병합한다. 따라서 중첩정도가 낮고 분리성이 높은 클러스터를 학습데이터로 재설정하여서 감독분류를 수행한다. 그리고 분류결과를 분류항목별로 재설정한다. 클러스터간 중첩성을 판단하기 위해 주어진 소속정도(u)에 대한 클러스터간 중첩함수를 이용한다. 두 퍼지 클러스터 F_p, F_q 와 주어진 소속정도 u 에 대해서 중첩함수 $f(u)$ 는 다음과 같다[6].

$$f(u) = \sum_{j=1}^n \delta(x_j, u; \bar{F}_p, \bar{F}_q) w(x)$$

$$\delta(x_j, u; \bar{F}_p, \bar{F}_q) = \begin{cases} 1.0 & \text{if } u \leq u_{F_p}(x_j) \text{ and } u \leq u_{F_q}(x_j) \\ 0.0 & \text{otherwise} \end{cases}$$

$\delta(x_j, u; \bar{F}_p, \bar{F}_q)$ 는 데이터 x_j 에 대해서 소속정도 μ 에서 두 클러스터가 중첩하는지를 결정한다. 두 클러스터 모두 μ 값 이상이면 중첩성은 1.0의 값을 가진다. 그렇지 않은 경우는 0.0의 값을 가진다. 그림 4는 소속정도 μ 에서 두 클러스터간의 중첩성 계산을 도식화 한 것이다. 전체 데이터 $x_j \in X$ 에 대해서

$x_j \in r$ 만이 중첩된 것으로 판단한다. 퍼지 클러스터간 중첩 여부를 확인하여 두 클러스터간 중첩되지 않은 데이터는 각각의 분류항목으로 할당한다. 중첩된 데이터 부분들은 학습데이터를 이용하여 최대우도 분류를 수행한다. 그래서 각 분류항목에 대한 우도를 구한다. 중첩 데이터에 대한 우도와 퍼지 소속도를 비교하여 우도가 가장 높고 소속도가 큰 항목으로 분류를 한다. 우도와 소속도가 각기 다른 분류항목으로 나타나는 중첩 화소는 상, 하, 좌, 우, 대각 방향의 이웃한 화소를 검사하여 분류항목을 결정한다.

중첩 화소에 대한 우도는 최대우도(Maximum likelihood) 분류기법에 의해 수행하였다. 최대우도분류기법은 하나의 분류 항목에 대한 확률함수가 정규 밀도 함수에 근사 한다고 가정하고서, 트레이닝 데이터로부터 취득된 통계 값들을 이용하여 영상의 화소들을 특정한 항목들로 분류하는 알고리즘이다[7]. 즉, 각 클래스에 대한 화소 자료의 우도를 구하고 최대 우도 클래스에 그 화소를 분류하는 방법이다. 화소 자료 X 가 분류항목(클래스) i 에 속하는 우도는 다음과 같다.

$$P(X | w_i) = \frac{1}{(2\pi)^{N/2} |\Sigma_i|^{1/2}} \cdot A$$

$$A = \exp \left[-\frac{1}{2} (X - U_i)^T \Sigma_i^{-1} (X - U_i) \right]$$

X = 데이터 벡터

U_i = 분류항목 i 에 대한 N 개 밴드의 평균 벡터

Σ_i = 분류항목 i 의 공분산 행렬

C_{kl} = 두 개의 밴드들 사이의 공분산(k 와 l 밴드)

$|\Sigma_i|$ = 공분산 행렬 Σ_i 의 determinant

Σ_i^{-1} = Σ_i 의 역행렬

$(X - U_i)^T$ = 벡터 $(X - U_i)$ 의 전치행렬

3. 결론

본 연구에서는 미지 영상 화소를 무조건 단일 분류항목에 할당하는 것이 아니라 평균내부거리를 이용한 퍼지 알고리즘의 소속도를 고려하여 우도가 큰 분류항목에 할당하게 함으로서 정확한 판단을 할 수 있도록 하여 분류의 성능을 개선 할 수 있다. 각각의 화소에서부터 내부클러스터까지의 거리로 소속정도를 부여하므로 경계 영역에 있는 애매한 데이터에 대해서 클러스터의 크기에 상관없이 소속정도를 잘 부여하게 됨을 알 수 있다. 따라서 위성영상의 분류에 있어서 분류가 애매한 지역의 분류 정확도를 개선할 수 있다. 본 논문에서는 클러스터의 크기에 따

라 큰 클러스터에는 많은 소속정도를 작은 클러스터에는 적은 소속정도를 부여함으로 크기가 다른 클러스터가 존재하는 데이터 집합에 대해서도 분류의 효율성을 높였으며, 클러스터들간의 중첩성을 이용한 평가를 통해 중첩되지 않은 데이터는 각각의 분류항목에 포함시키고, 중첩된 데이터들은 최대우도 분류를 수행한 후 각 데이터에 대한 우도와 퍼지 클러스터링의 소속도를 비교 분석하여 최종 분류항목을 결정함으로써 분류를 효율적으로 할 수 있다.

향후 다양한 위성영상을 이용하여 알고리즘의 적용성에 대한 실험 및 분석이 뒤따라야 하겠다.

참고문헌

- [1] John A. Richards, 'Remote Sensing Digital Image Analysis : An Introduction', Second, Revised and Enlarged Edition, pp. 229-262, Springer-Verlag, 1994.
- [2] N.R. Pal and J.C. Bezdek, "On cluster validity for the fuzzy c-means model", IEEE Transactions on Fuzzy Systems, Vol. 3, No. 3, pp.370~379, 1995.
- [3] R. Krishnapuram, H. Frigui and O. Nasraoui, " Fuzzy and Possibilistic shell Clustering algorithms and their application to boundary detection and surface approximation," IEEE Trans. on Fuzzy Systems, Vol. 3, No. 1, pp. 29-60. 1995.
- [4] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," IEEE Trans. on Fuzzy Systems, Vol. 1, No. 2, pp. 98-110, 1993.
- [5] 유현제, 조석제, "평균내부거리를 적용한 퍼지 클러스터링" 한국정보처리학회 추계 학술발표논문집 제6권 제2호, 1999.
- [6] 김대원, 이광형, "클러스터간 중첩성과 분리성을 이용한 퍼지 분할의 평가 기법", 한국 퍼지 시스템 학회, Vol.13, No.4, pp.455~460, 2003.
- [7] B.Gorte and A. Stein. "Bayesian classification and class area estimation of satellite images using stratification," IEEE Trans. On Geoscience and Remote Sensing, 36(3):303, 1998.