

PCM 알고리즘과 베이시안 분류의 통합기법

전영준, 김진일

동의대학교 컴퓨터공학과

j4017@chol.com, jikim@deu.ac.kr

Integrating Classification Method using PCM Algorithm and Bayesian Method

Young-Joon Jeon, Jin-Il Kim

Dept. of Computer Engineering, Dongeui University

요 약

본 논문은 PCM(Possibilistic C-Means) 알고리즘과 베이시안 분류 알고리즘을 통합한 고해상도 위성영상의 효과적인 분류방법을 제안하였다. 제안된 알고리즘은 학습데이터를 참고로 하여 PCM 알고리즘을 반복적인 과정없이 수행한다. 각 분류항목별로 분류된 데이터에서 평균내부거리 내부에 해당되는 데이터들을 선정하여 각 항목별 비율을 구한 후 베이시안 분류기법의 사전확률로 적용하여 분류를 수행한다. PCM 알고리즘은 각 데이터와 특정 클러스터와의 거리에 소속도를 부여하는 퍼지 C-Means 알고리즘과 달리 소속도를 각 데이터와 클러스터 중심간의 절대거리에 의존하는 방법으로 퍼지 C-Means 알고리즘이 가지는 상대성 문제를 해결하였다. 제안된 분류 기법을 고해상도 다중분광 데이터인 IKONOS 위성영상에 적용하여 분류를 수행한 후 최대우도 분류기법과 비교한다.

1. 서 론

인공위성에 의한 원격탐사에 대한 정보는 토지의 체계적인 활용, 환경오염의 감시 및 통제, 지도제작 등에 요구되는 복잡 다양한 정보를 신속하고 정확하게 분류, 해석하는데 있어서 커다란 잠재력을 가지고 있는 최첨단 과학기술로 각광 받고 있으며, 그 활용 분야는 환경에서부터 토목, 농업, 해양, 지질, 임업, 수산업 등 각 분야에 걸쳐 급속도로 확산되고 있다. 인공위성으로부터 얻어진 화상으로부터 지표면의 특징을 분류하여 피복도를 작성하는 것은 원격탐사 분야에서 중요하게 다루어지는 분야이다. 분류기법은 일반적으로 위성영상자료로부터 각종 주제도를 얻어내기 위한 효과적인 방법이다. 원격탐사위성영상의 분류는 영상을 구성하는 각각의 화소들 중 비슷한 분광특성을 갖는 것끼리 집단화시켜 주는 방법으로서 이것은 영상으로부터 관심 대상인 정보를 추출하기 위해 수행되는 것으로 주로 논, 밭, 나대지 등의 토지피복, 표면 온도, 대기 상태 등의 환경상태 등 광범위한 지역을 몇 종류 또는 몇 단계로 분류할 수 있는 정보를 추출하는 데 사용된다. 위성영상 분류의 최근 연구를 살펴보면 베이시안 분류기법[1]과 같은 기존의 통계적 이론을 확장하여 보다 개선된 내용을 제안한 연구, 원격탐사 영상의 분류에 있어 클래스의 경계를 확실히 정하기 어려운 경우에 효과적으로

이용되는 퍼지 이론을 적용한 연구[2], 신경망 이론들을 적용한 연구, 퍼지 이론과 신경망 등 여러 개의 분류기법을 혼합하여 분류 정확도를 향상시키는 연구 등 다양한 연구가 이루어지고 있다.

본 연구에서는 PCM(Possibilistic C-Means) 알고리즘과 베이시안 분류기법을 통합한 분류 방법을 제안하였다. PCM 알고리즘은 각 데이터와 특정클러스터와의 거리에 소속도를 부여하는 퍼지 C-Means 알고리즘과 달리 소속도를 각 데이터와 클러스터 중심간의 절대거리에 의존하는 방법으로 퍼지 C-Means 알고리즘이 가지는 대성 문제를 해결하였다. PCM 알고리즘에 의하여 각 분류항목별로 분류된 데이터에서 소속도가 0.5 이상인 데이터들을 선정하여 각 항목별 비율을 구한 후 베이시안 분류기법의 사전확률로 적용하여 분류를 수행한다. 제안된 분류기법을 IKONOS 위성영상에 적용하여 그 정확성을 비교하였다.

2. PCM 알고리즘

Bezdek이 제안한 FCM(Fuzzy C-Means) 알고리즘은 각 데이터와 특정 클러스터 중심과의 거리에 소속 정도를 부여한다[2]. 그러나 FCM 알고리즘은 데이터의 각 클러스터에 관한 소속함수 값이 1이 되는 확률적 제약 조건을 이용하므로 소속함수 값이 소속성(belonging)이나 적합성(compatibility) 정도의 직관적인 개념과 항상 일치하지 않는다.

FCM의 경우에는 “equal evidence”와 “ignorance”를 구분할 수 없기 때문에 특히, 잡음이 있는 상황에서 심각한 문제점을 야기시킨다. 이는 FCM의 경우 소속함수 값이 다른 클러스터와 관계가 있는 상대적인 값이기 때문이다. 이러한 문제점을 개선하기 위해 Raghunath는 PCM(Possibilistic C-Means) 알고리즘을 제안하였다. PCM의 경우는 소속함수 값이 다른 클러스터와 관계가 없고, 속성 데이터와 해당 클러스터 중심간의 거리에만 의존하기 때문에 이러한 문제점을 개선할 수 있다. 그러나 PCM 방법을 적용하기 위해서는 각 클러스터의 평균 내부거리를 알아야 한다. 본 논문에서는 공분산 행렬을 적용한 FCM 알고리즘인 퍼지 Gustafson-Kessel 알고리즘을 반복없이 수행하여 평균내부거리를 얻고 이를 바탕으로 PCM 알고리즘을 적용하였다. PCM 알고리즘의 목적함수와 알고리즘은 다음과 같다[3].

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

여기서 η_j 는 각 그룹의 평균내부거리이다. Krishnapuram은 η_j 를 다음과 같이 정의하였다.

$$\eta_j = K \frac{\sum_{i=1}^n \mu_{ij}^m d_{ij}^2}{\sum_{i=1}^n \mu_{ij}^m} \quad (2)$$

N 은 속성 데이터의 개수이고, $K=1$ 이다. PCM 알고리즘의 소속도 μ_{ij} 는 다음과 같이 구할 수 있다.

$$\mu_{ij} = \frac{1}{1 + \left(\frac{d_{ik}}{\eta_j}\right)^{\frac{1}{m-1}}} \quad (3)$$

3. 베이시안 분류기법

베이시안 분류기법은 각 분류항목에 대한 각 화소의 우도를 구하고 최대우도 분류항목에 화소를 분류하는 방법으로서 베이즈의 이론에 기초를 두고 있다. 어떤 화소에 대해 분류 결정을 하기 위해서는 화소가 특성치 X 를 가질 때 w_i 에 속할 확률인 사후확률을 알아야 하며 이 확률 $P(w_i | x)$ 을 우도라고 하고 베이즈 정리에서 유도된다.

$$P(w_i | x) = \frac{P(x | w_i)P(w_i)}{P(x)} \quad (4)$$

$P(w_i)$: 영상에서 분류항목 w_i 가 발생할 사전확률
 $P(x | w_i)$: 학습 데이터로부터 계산된 확률
 (분류항목 w_i 로부터 X 가 관측되는 조건부 확률)
 $P(x)$: 위치 X 에서 어떤 분류항목으로부터 픽셀을 찾을 확률 => $P(x) = \sum_i P(x | w_i)P(w_i)$
 $P(x | w_i)P(w_i)$: 어떤 화소가 특성치 x 를 가지며 동시에 w_i 에 속할 확률

위성영상과 같이 다차원 다중분광 데이터를 분류할 경우 다차원 정규확률밀도 함수를 이용한다. 이 때 어떤 화소가 특성치 x 를 가지며 동시에 w_i 에 속할 확률 $P(x | w_i)P(w_i)$ 를 구하면 다음과 같다.

$$P(x | w_i)P(w_i) = \frac{P(w_i)}{(2\pi)^{N/2} \left| \sum_{i=1}^c \mu_{ij} \right|} \exp\left[-\frac{1}{2}(X - U_i)^T \sum_{i=1}^c \mu_{ij}^{-1} (X - U_i)\right] \quad (5)$$

계산식의 간편화를 위해 양변에 로그를 취하여 다음과 같은 판별함수를 얻을 수 있다.

$$D_i(x) = \ln P(w_i) - \frac{N}{2} \ln 2\pi - \frac{1}{2} \ln \left| \sum_{i=1}^c \mu_{ij} \right| - \frac{1}{2} (X - U_i)^T \sum_{i=1}^c \mu_{ij}^{-1} (X - U_i) \quad (6)$$

3. 제안된 분류 알고리즘

본 연구에서는 PCM 알고리즘과 베이시안 분류기법을 통합한 분류 방법을 제안하였다. PCM 알고리즘에 의하여 각 분류항목별로 분류된 데이터에서 소속도가 0.5이상인 데이터들을 선정하여 각 항목별 비율을 구한 후 베이시안 분류기법의 사전확률로 적용하여 분류를 수행한다. 위성영상은 다차원 데이터이다. 제안된 알고리즘은 훈련 데이터를 이용하여 초기 중심값을 설정한 후 퍼지 Gustafson-Kessel 알고리즘을 반복 없이 수행한다. 그리고 퍼지 Gustafson-Kessel 알고리즘의 소속도를 이용하여 평균내부거리를 구한다. 평균거리 내부에 소속된 데이터들의 항목별 비율을 베이시안 분류 기법의 사전확률로 이용한다. 제안된 분류기법의 수행 과정은 다음과 같다.

입력 영상에서 분석자가 먼저 분류를 위한 분류항목을 설정하고, 분류항목별로 입력영상에 대하여 여러 영역을 훈련 데이터로 선정한다. 분류항목별 훈련데이터에서 평균값을 구하여 퍼지 알고리즘의 초기 중심값으로 설정한다. 공분산 행렬 F_i 를 구하고, 그것을 이용하여 중심과의 마하라노비스 거리를 계산한다. 초기에는 중심에 대한 소속도가 구해지지 않은 상태이므로 모든 화소들의 소속도 $\mu_{ij} = 1$ 로 설정한다. 위성영상이미지 데이터를 $I = \{X_j | j=1, \dots, M\}$ 로 나타낸다. b 는 위성영상의 밴드(채널)을 나타낸다. V 는 각 항목별 중심벡터를 나타낸다.

$$F_i = \frac{\sum_{j=1}^M \mu_{ij}^m (X_{jb} - V_{ib})(X_{jb} - V_{ib})^T}{\sum_{j=1}^M \mu_{ij}^m} = \frac{\sum_{j=1}^M \mu_{ij}^m \begin{bmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ M \\ x_{jN} \end{bmatrix} - \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ M \\ v_{iN} \end{bmatrix} \cdot \left[\begin{bmatrix} x_{j1} \\ x_{j2} \\ \vdots \\ M \\ x_{jN} \end{bmatrix} - \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ M \\ v_{iN} \end{bmatrix} \right]^T}{\sum_{j=1}^M \mu_{ij}^m} \quad (7)$$

마하라노비스 거리 d_{ij}^2 는 다음과 같이 계산된다.

$$d_{ij}^2 = \left(\begin{pmatrix} x_{ij1} \\ x_{ij2} \\ \vdots \\ x_{ijN} \end{pmatrix} - \begin{pmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{iN} \end{pmatrix} \right)^T \cdot \left[|F_i|^{-\frac{1}{b}} \cdot F^{-1} \right] \cdot \left(\begin{pmatrix} x_{ij1} \\ x_{ij2} \\ \vdots \\ x_{ijN} \end{pmatrix} - \begin{pmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{iN} \end{pmatrix} \right) \quad (8)$$

각 분류항목에 속할 소속도를 계산한다. 이 소속도는 훈련데이터의 평균값과 각 화소와의 거리에 대한 소속도이다.

$$u_{ij} = \left[\sum_{k=1}^c \left(\frac{d_{ij}^2}{d_{jk}^2} \right)^{\frac{2}{m-1}} \right]^{-1}, \quad 1 \leq k \leq c; \quad 1 \leq j \leq N \quad (9)$$

소속도에 의하여 각 항목별로 입력된 영상의 데이터를 분류한다. 각 항목별로 분류된 결과를 이용하여 중심값을 새롭게 설정한다. 중심값의 갱신은 다음 식에 의하여 이루어진다.

$$v_i = \frac{\sum_j \mu_{ij}^m \cdot x_j}{\sum_{j=1}^N \mu_{ij}^m} \quad (10)$$

분류결과를 이용하여 각 항목별로 평균내부거리 η_i 를 구한다. 평균내부거리는 각 데이터로부터 해당 항목의 중심값까지의 거리를 평균한 값으로 클러스터의 크기와 밀도에 비례한다. 평균내부거리는 식(2)을 이용하여 구한다. 평균내부거리를 구한 후 식(3)을 통하여 소속도를 갱신한다. 평균내부거리 안쪽에 속하는 데이터들의 집합을 내부클러스터라 정의한다. 내부클러스터는 식(3)에 의해 소속도가 $\mu_{ij} \geq 0.5$ 에 속하는 화소들이다. 내부클러스터에 속하는 화소들에 대하여 각 항목별 비율을 구한다. 이를 베이지안 분류의 사전확률 $P(w_i)$ 로 이용한다. 그리고 내부클러스터에 해당되는 데이터에 대하여 식(7)을 이용하여 공분산 행렬을 구한다. 내부클러스터의 데이터들은 정규분포를 보여주고 있으므로 이를 이용하여 공분산 행렬과 마하라노비스 거리를 구하여 베이지안 알고리즘에 적용함으로써 정확도를 높여줄 수 있다. 베이지안 분류는 다음과 같이 수행한다.

$$D_i(x) = \ln P(w_i) - \frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |F_i| - A \quad (11)$$

$$A = \frac{1}{2} (x_{ijb} - v_{ib})^T \left[|F_i|^{-\frac{1}{b}} \cdot F^{-1} \right] (x_{ijb} - v_{ib})$$

$D_i(x)$ 의 값이 최대가 되는 분류항목으로 화소를 분류한다.

4. 실험

본 논문에서는 제안된 알고리즘을 테스트하기 위해서 입력 영상으로 IKONOS 위성영상을 이용하였다. IKONOS 위성영상은 크기가 1000×1000 화소를 가진 영상이다. 이 영상은 4개의 밴드로 구성된 다중분광 위성영상이다.

4개의 분류항목 Water, Forest, Urban, Soil 을 선정한 후 입력 영상에서 3~5개 지역을 훈련데이터로 선정하였다.

표 1. 훈련데이터에 대한 분류 정확도 비교

		전체정확도
훈련데이터의 수		52500
분류방법	최대우도분류	94%
	제안된 분류기법	97%

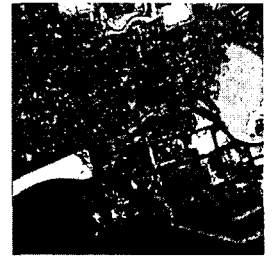


그림 1. IKONOS 위성영상에 대한 제안된 알고리즘의 분류결과 영상

본 연구에서는 IKONOS 다중분광 위성영상에서 선정한 훈련 데이터들이 최종 분류 후 처음 선정된 항목으로 잘 분류되었는가를 확인하였다. 그 결과 일반적인 최대우도 분류보다 좋은 결과를 얻을 수 있었다.

5. 결론

본 논문은 PCM 알고리즘과 베이지안 분류알고리즘을 통합한 고해상도 위성영상의 효과적인 분류방법을 제안하였다. PCM 알고리즘은 각 클러스터에 속하는 정도가 명확한 속성 데이터는 가능한 높은 소속도를 부여하고, 그렇지 못한 속성데이터는 낮은 소속도를 부여한다. 따라서 PCM 알고리즘의 평균내부거리 안쪽에 속하는 데이터는 명확한 속성데이터의 집합으로서 이를 베이지안 분류의 사전확률로 적용하여 분류를 수행한 결과 정확도를 향상시켜 주었다. IKONOS 고해상도 위성영상을 이용하여 일반적으로 많이 사용되는 베이지안 분류 기법과 비교한 결과 분류 정확도가 향상되었음을 확인할 수 있었다. 향후 연구로서는 고해상도 위성영상의 도시지역의 세분화된 분류를 할 수 있는 연구가 뒤따라야 하겠다.

참고문헌

- [1] B.Gorte and A. Stein, "Bayesian classification and class area estimation of satellite images using stratification," IEEE Trans. On Geoscience and Remote Sensing, Vol.36, No.3, pp.803-812, 1998.
- [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 and J. M. Keller, "A possibilistic approach to clustering," IEEE Trans. on Fuzzy Systems, Vol. 1, No. 2, pp.98-110, 1993.