The Journal of GIS Association of Korea, Vol. 2, No. 2, pp.175-188, October 1994

신경망을 이용한 원격탐사자료의 군집화 기법 연구

김광은* · 이태섭* · 채효석**

Study on Application of Neural Network for Unsupervised Training of Remote Sensing Data

Kwang-Eun Kim* · Tai-Sup Lee* · Hyo-Sok Chae**

요 약

본 연구에서는 최근 많은 분야에서 패턴인식을 위한 효과적인 기법으로 이용되고 있는 신경망 기법을 원격탐사자료의 군집화 기법으로서 적용하고자 하였다. 이를 위해 선택된 신경망 모델은 경쟁학습 신경망이며 이를 구성하는 각종 변수들을 재구성하여 원격탐사자료의 군집화를 위한 신경망 모델을 설정하였다.

본 신경망을 이용한 군집화 기법은 항공기를 이용하여 획득된 원격탐사자료를 이용하여 순차적(sequential) 군집화 기법 및 K 평균 군집화 기법과 비교되었다. 계산시간은 순차적 기법이나 K 평균기법에 비하여 더 많이 소요되나 정확도면에 있어서는 비교적 우수한 결과를 나타냈다.

ABSTRACT : A competitive learning network was proposed as unsupervised training method of remote sensing data. Its performance and computational requirements were compared with conventional clustering techniques such as Sequential and K-Means. An airborne remote sensing data set was used to study the performance of these classifiers. The proposed algorithm required a little more computational time than the conventional techniques. However, the performance of competitive learning network algorithm was found to be slightly more than those of Sequential and K-Means clustering techniques.

* 한국자원연구소(Korea Institute of Geology, Mining and Materials, 30, Kajung Dong, Yusong Gu, Taejon, Korea, Tel (042)868-3151)

** 전북대학교 자원공학과(Dept. of Mining and Mineral Resources Engineering, College of Engineering, Chonbuk Univ.)

서 론

원격탐사자료는 분류(classification)과정을 통하여 정량적 주제도(thematic map)로서 표현되어 진다. 분류를 통하여 작성된 주제도는 그 자체로서 정보를 표현하고 있을 뿐만 아니라, GIS 에서 원격탐사자료가 하나의 자료층으로서 이용될 수 있는 자료형태를 가지고 있다.

전통적으로 분류는 무감독(unsupervised) 기법과 감독(supervised)기법으로 나누어져 왔으나, 우선적으로 무감독분류를 통하여 분류항목의 수 및 각 분류항목의 군집중심, 밴드간 공분산 등을 구하고, 이들을 이용하여 감독분류를 행함으로서 최종 주제도를 작성하는 것이 일반적이고 효율적인 처리절차이다(Schowengerdt,1983). 최근의 상업화된 원격탐사자료 처리용 소프트웨어 시스템에서는 사용자가 영상으로부터 직접 교사자료(training data)를 선택하는 방법을 제공하고 있다. 이러한 방법에서는 선택된 교사자료들을 통계적으로 분석하여 각 분류항목의 군집중심, 밴드간 공분산 등을 계산하여 분류의 기준자료로 이용한다. 그러나 영상으로서 나타낼 수 있는 밴드의 수는 최대 3개에 불과할 뿐아니라 인간의 눈을 통하여 식별하기 어려운 미세한 분광특성차이가 존재할 수 있다. 따라서 최근에는 무감독분류를 통하여 자료를 일차 군집화한 뒤, 해석자가 다른 종류의 정보를 이용하여 군집화된 결과를 분석, 편집하여 분류항목별 이름, 군집중심, 밴드간 공분산 등을 새롭게 설정한 후 최종 분류를 수행하는 접합식(hybrid) 분류 방법이 보다 널리 이용되고 있다.

무감독분류는 자료값의 분포특성에 근거하여 분광특성이 서로 다른 여러개의 군집을 형성하는 군집화(clustering) 과정이다. 군집은 그 중심, 밴드간 공분산 등으로서 분광적 분포특성이 나타내어 진다. 형성된 각 군집의 중심, 밴드간 공분산 등을 가지고 자료내의 모든 화소를 가장 유사한 분광적 분포특성을 가지는 군집에 할당시키는 과정이 감독 분류이며, 그 유사성을 판별하는 기준에 따라 최소거리법(minimum distance), 평행사변형법(parallelepiped), 최대우도법(maximum likelihood) 등으로 나누어 진다. 따라서, 감독분류의 기준 자료가 되는 총 군집의 수, 각 군집의 중심, 밴드간 공분산 등을 계산하는 과정인 군집화는 사실상 분류 결과의 질을 결정짓는 매우 중요한 처리과정이라 할 수 있다. 특히, 최근에는 측정 밴드가 수십-수백개에 이르는 hyperspectral remote sensing 까지 등장하고 있어 효과적인 군집화 방법의 중요성이 강조되고 있다.

기존의 군집화 방법중, 순차적 군집화(sequential clustering)기법은 반복적 방법이 아니므로 계산시간이 적게 소요되는 장점이 있는 반면, 군집의 초기치가 좌상쪽 화소에서 시작되므로 좌상쪽 자료의 분광적 패턴에 크게 좌우되는 군집화 결과를 나타내는 경향이 있다. 뿐만 아니라 사용자가 군집의 최대반경, 군집간의 최소거리 등을 입력하여야 하므로 자료의 분포에 대한 사전지식이 있거나 시행착오적으로 여러번의 군집화를 수행하여야만 좋은 결과를 기대할 수 있다. K 평균 알고리즘이나 ISODATA 기법과 같은 반복적 방법의 경우, 군집중심의 초기값을 임의의 값으로 하므로 좌상쪽 자료에 의존되는 경향은 없다. 그러나 K 평균 알고리즘은 초기치에 의하여

그 군집화결과가 매우 다르게 나타날 수 있다는 문제점이 있다. ISODATA 기법은 군집화 도중에 각 군집의 중심간 거리가 가까운 두개의 군집을 하나의 군집으로 합치거나 분산이 너무 큰 군집을 서로 다른 두개의 군집으로 분리하는 기능이 있어 비교적 양질의 결과를 기대할 수 있으나 계산시간이 매우 많이 소요될 뿐 아니라 분산매개변수와 결합매개변수를 해석자가 결정하여야 하기 때문에 순차적 군집화기법에서와 마찬가지로 자료값의 분포에 대한 지식을 필요로 한다.

인공신경망 기법은 최근 수십년간 음성, 문자, 영상 등 각종 패턴의 인식을 위한 효과적인 기법으로서 이용되어 왔다. 본 연구에서는 다양한 신경망 모델중의 한 형태인 경쟁학습 신경망(competitive learning network)기법을 원격탐사자료의 군집화기법으로서 활용하는 데 있어 신경망의 구조 및 각종 변수들의 역할을 고찰하고자 한다. 또한 이러한 신경망 모델에 대한 고찰과 함께 기존의 군집화 기법들의 각종 문제점들을 고려하여 신경망을 이용한 효과적인 군집화 알고리즘을 구현하고, 이의 적용절차, 처리결과 및 문제점들을 K 평균 군집화 기법과 비교하고자 한다.

원격탐사자료 군집화 기법으로서의 경쟁학습 신경망

경쟁학습 신경망(Competitive learning network)

신경망 모델은 위상(topology), 활성화 함수(activation function) 및 학습규칙(learning rules) 등에 의하여 특징지워진다. 여기에서는 단순 경쟁학습 신경망, Kohonen(1982)의 자기조직화 특징지도(Self-Organizing Feature Map), 그리고 History Sensitive 경쟁학습 신경망에 대하여 그 특징을 고찰하고 이러한 고찰을 토대로 원격탐사자료의 효과적인 군집화를 위한 신경망 구조를 설정하고자 한다.

Fig. 1 에는 Kohonen의 자기조직화 특징지도 알고리즘의 신경망 구조를 도식적으로 나타냈다. 그림에서 보듯이 공통의 입력 유니트에서 2차원 격자로 배열된 N x N 개의 출력 유니트로의 연결강도를 조정함으로써 L 차원의 벡터인 입력 패턴을 양자화한다. 자기조직화 특징지도 알고리즘의 경우, 출력유니트는 위상학적으로 순서를 가지고 있어야 하나 단순 경쟁학습 신경망(simple competitive learning network)이나 History sensitive 경쟁학습 신경망의 경우 순서를 가질 필요가 없다. 이러한 Kohonen 의 자기조직화 지도 알고리즘의 구체적인 학습 순서는 Table 1 에 나타낸 바와 같다(Kung, 1993). 이때 만일 단계 3 에서 연결강도의 갱신을 출력 유니트

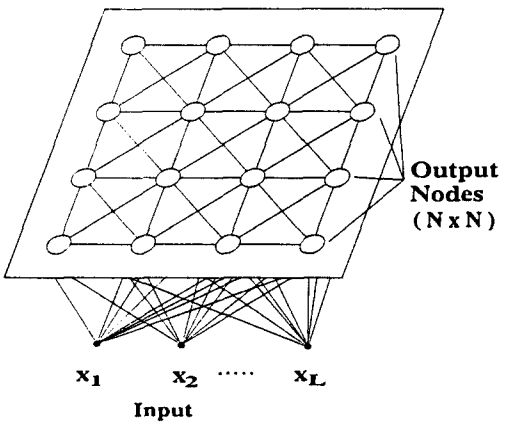


Fig. 1 A network for self-organizing feature map. The x_1, x_2,..., x_L are inputs, one for each component of the data (i. e., in the 6 band case, L is 6).

Table 1. Training phase in competitive learning network.

단계 1. 각 출력 유니트의 연결강도 벡터 $Wi(i=1, 2, 3,...N\times N)$를 임의의 작은 값으로 초기화 한다.

단계 2. 입력 벡터 X에 대하여 유클리디안 거리 $|X-Wi|$가 최소인 출력 유니트 i^*를 찾는다.

단계 3. 결정된 출력 유니트를 포함하는 일정한 영역내에 존재하는 출력 유니트의 연결강도를 다음과 같이 갱신한다.

$$\Delta Wj=\eta A(j, i^*)(X-Wj)$$

여기에서 $j\in j^*$이며, I^*는 i^* 주변의 일정영역을 나타낸다. η는 학습률(learning rate)이며 $A(j, i^*)$는 i^*를 중심으로 멀어질수록 그 값이 작아지는 함수(neighborhood function)이다.

새로운 입력 벡터에 대하여 단계 2부터 반복한다.

i^* 에 대하여만 수행한다면 단순 경쟁학습 신경망이라고 할 수 있으며 영역함수(neighborhood function) $A(j, i^*)$ 는 필요없게 된다. 또한 연결강도의 갱신을 출력 유니트 i^* 에 대하여만 수행하나 빈도함수 $E(i^*)$ 를 도입한다면 History sensitive 경쟁학습 방법이 된다(Kung, 1993). 여기에서, 빈도함수 $E(i^*)$는 그때까지의 출력 유니트 i^* 가 최소의 유클리디안 거리를 가지는 출력 유니트로 결정된 횟수에 의하여 결정되는 함수이다. 이제 이러한 신경망 학습법을 구성하는 여러가지 요인들을 고찰하고 원격탐사자료의 군집화를 위한 분류기로서의 신경망 구조 및 학습방법 알고리즘을 구현하고자 한다.

출력 유니트의 위상(Topology)

Hung(1993) 은 경쟁학습 신경망을 원격탐사자료의 군집화에 적용하면서, Fig. 1 에 나타낸 것과 같이 2차원 격자로 배열된 출력 유니트를 설정했던 바 있다. 신경망 학습법에서의 출력 유니트는 원격탐사자료의 군집화시 생성가능한 군집의 최대 갯수가 된다. 학습은 매 입력벡터에 대하여 모든 출력 유니트와의 거리를 계산한 후, 최소 거리를 가지는 출력 유니트를 포함하는 일정 영역내에 존재하는 출력 유니트의 연결강도를 조정하는 방법으로 이루어지므로 출력 유니트의 수가 하나만 많아져도 전체 화소의 수만큼 거리계산 및 비교과정이 증가하여 총계산시간은 엄청나게 증가하게 된다. 2차원 격자로 배열된 출력 유니트를 설정하는 경우, 최대 군집의 개수를 자연수의 자승개로만 설정이 가능하다. 따라서 만일 해석자가 10개 정도의 군집이 형성되기를 원한다고 하여도 16개의 출력 유니트를 설정할 수 밖에 없어 계산시간의 손실이 매우 크다. 대개의 원격탐사자료의 양이 매우 방대함을

생각한다면 이는 엄청난 비효율이 될 수 있다. 따라서 본 연구에서는 1차원의 출력 유니트를 설정하기로 한다. 이에 따라 영역함수(neighborhood function) $A(j, i^*)$ 는 2차원 함수가 아닌 1차원 함수로서 바뀌어야 하며 이는 매우 간단히 이루어질 수 있다.

초기치의 설정 (Initial weight)

그 다음으로 고려해야 할 요소는 초기치의 설정 문제이다. 경쟁학습 신경망을 이용하는 학습 방법에서는 각 출력 유니트의 초기 연결강도를 임의의 값으로 설정한다. 이는 곧 임의의 군집중심으로부터 군집화를 시작한다는 의미이므로 K 평균 기법이나 ISODATA 기법과 마찬가지로 반복적으로 군집화가 이루어져야 한다. 이들 반복적인 군집화 기법에서는 반복횟수가 충분할 경우 초기치의 분포에 관계없는 군집화 결과를 보이는 것으로 알려져 있으나 초기치가 실제 자료값의 분포범위에서 매우 벗어나 있거나 반복횟수가 충분하지 않을 경우 초기치가 바뀜에 따라 군집화 결과가 다르게 나타나기도 한다. 특히 K 평균 기법의 경우에는 초기치가 달라짐에 따라 매우 다른 군집화 그 결과를 보이기도 한다. 이는 순차적 군집화 기법에서 좌상쪽 자료에 크게 의존하는 군집화 결과를 보이는 것과 같은 원리이다. 신경망을 이용하는 학습방법의 경우, 초기치를 아주 작은 임의의 값으로 하는 대신에 처음에는 연결강도를 조정하는 영역을 매우 넓게 하고(처음에는 거의 모든 출력유니트에 대하여 연결강도를 갱신한다) 영역함수도 비교적 모든 출력 유니트에 대하여 고른 값을 가지도록 설정한다. 따라서 사실상, 모든 출력 유니트들은 처음 입력되는 입력벡터를 중심으로 아주 작은 변화를 가지는 초기치들이 설정된다고 볼 수 있다. 일부 기존의 반복적 군집화 기법에서는 통계적으로 자료의 분포(평균, 분산)를 미리 계산하여 각 군집 중심의 초기치를 자료값의 범위내에 존재하는 임의의 값으로 결정하기도 하나 본 연구에서는 영역함수에 의하여 한두번의 반복에 의하여 초기치가 자연스럽게 자료값 분포 범위로 이동되므로 임의의 작은 값을 초기치로 설정하기로 하여 불필요한 계산과정을 피하도록 하였다.

학습률 (Learning rate)

다음으로 고려해야할 것은 학습률 η이다. Table 1에서 알 수 있듯이 η는 입력벡터와 연결강도와의 차에 곱하여 진다. 신경망 학습법에서는 대개 학습시간의 역수로 주어지며 Hung(1993)은 처음 얼마간의 반복시에는 아주 작은 값을 가지는 상수로 하며 반복이 진행됨에 따라 반복횟수에 반비례하는 값으로 선택하였던 바 있다. 이는 연결강도의 초기치가 임의의 작은 값으로서 실제 자료값의 분포와는 전혀 관계없는 값이므로 처음 얼마간은 반복횟수에 관계없이 연결강도가 충분히 갱신될 수 있도록 한 것이며, 어느 정도 반복과정후의 연결강도는 비교적 안정된 값으로 보고 반복횟수가 증가함에 따른 연결강도의 변화를 적게 주려는 의도이다. 그러나 본 연구에서는 연결강도의 초기치를 비교적 작은 값으로 설정하고 초기의 영역함수를 매우 넓은 영역에 대하여 작용하도록 함으로써 하나의 입력벡터에 의하여 모든 출력 유니트의 연결강도가 곧바로 실제 입력벡터값과 유사한 크기의 값으로 접근한다. 따라서 Hung(1993) 과 같이 처음 얼마간을 상수로 설정해야 할 이유가 없을 것

으로 판단된다. 본 연구에서는 경쟁학습 신경망 기법에서 가장 흔하게 설정되는 방법인 반복횟수의 역수를 기준으로 몇 가지 변형된 학습률을 적용 및 고찰하고자 한다.

영역함수 (Neighborhood function)

한편, Kohonen 의 자기조직화 특징지도 알고리즘에서 사용되는 영역함수(neighborhood function) 는 하나의 입력벡터가 입력되어 그 거리가 가장 가까운 출력 유니트가 결정되었을 때 연결강도를 갱신할 영역 및 정도를 결정하는 역할을 한다. 영역함수는 흔히 입력벡터로부터 거리가 가장 가까운 출력 유니트가 가장 큰 값을 가지며 이 출력 유니트로부터 멀어질 수록 작은 값을 가지게 설정된다. 원격탐사자료 군집화 기법의 관점에서는 이는 다음과 같은 의미를 가지는 함수로서 적용될 수 있다. 즉 초기 연결강도를 임의의 작은 값으로서 설정하였으므로 만일 이와 같이 어떤 영역에 대하여 연결강도를 갱신하지 않는다면 맨 처음의 입력벡터와 가장 거리가 가까운 출력 유니트의 연결강도만이 입력벡터의 수준으로 갱신되고 나머지 모든 출력 유니트들의 연결강도는 입력벡터의 크기와는 매우 거리가 먼 작은 값으로 남아있게 된다. 따라서 만일 이후의 입력벡터가 그전의 입력벡터와 크기 수준이 비슷하다면, 계속적으로 하나의 출력 유니트가 최소 거리를 가지는 출력 유니트로 결정될 가능성이 매우 커진다. 이렇게 되면 학습은 이루어지지 않는다. 따라서 원격탐사자료의 군집화를 위한 경쟁학습 신경망 알고리즘에 있어서 영역함수는 처음 얼마간의 입력벡터에 의하여 전체 초기치의 분포를 재조정하는 역할을 한다고 할 수 있다. 본 연구에서는 원격탐사자료의 특성을 고려하여 출력 유니트 j 에서의 영역함수 $A(j, i^*)$ 을 $\exp[-(R_{i^*j})^2 * T]$ 로서 설정하기로 한다. 여기에서 R_{i^*j} 는 입력벡터와 거리가 최소인 출력 유니트 i^* 로부터 출력 유니트 j 까지의 거리이며 T는 반복횟수이다. 이렇게 하면 출력 유니트 i^* 에 대하여는 항상 1의 값을 가지며 i^* 로부터 거리가 멀수록 작은 값을 가지게 된다. 또한 반복횟수가 증가될수록 그 영역이 감소되어 어느 정도의 반복후에는 거의 출력 유니트 i^*에 대하여만 1이 되며 다른 모든 출력 유니트에 대하여는 거의 0 에 가까운 값을 가지게 된다.

빈도 함수(History sensitive function)

빈도 함수 $E(i^*)$ 는 각 출력 유니트가 현재 가지고 있는 연결강도 값이 얼마나 많은 입력벡터에 의하여 학습되어진 결과인가를 나타내는 함수로 볼 수 있다. 즉, 많은 양의 입력벡터에 의하여 학습되어진 연결강도 값이, 새롭게 입력되어진 하나의 입력벡터에 의하여 크게 변화되어서는 안된다는 개념이라 할 수 있다. 이는 사실상 원격탐사자료의 군집화기법에 있어서 전혀 새로운 개념이 아니다. 즉 일반적인 원격탐사자료의 군집화 기법에서, 새로운 화소가 어떤 군집에 속한다고 결정되었을 때 그 군집중심은 그 군집에 이미 속해있는 화소의 개수를 고려한 가중평균으로서 결정됨과 같은 개념이다. 따라서 본 연구에서도 이와 같은 개념을 그대로 이용하여 $E(i^*)$ 를 한번의 반복과정중에 출력 유니트 i^* 가 최소 거리를 가지는 유니트로 결정되었던 횟수의 역수로서 설정하고자 한다. 지금까지의 고찰결과를 가지고 원격탐사자료의 군집화를 위한 신경망 구조 및 학습 알고리즘을 정리하면 Table 2와

Table 2. Training phase in modified competitive learning network.

단계 1. 각 출력 유니트의 연결강도 벡터 $\mathbf{W}i(i=1, 2, 3, \cdots N)$를 임의의 작은 값으로 초기화 한다.

단계 2. 입력 벡터 $\mathbf{X}$에 대하여 유클리디안 거리 $|\mathbf{X}-\mathbf{W}i|$가 최소인 출력 유니트 i^*를 찾는다.

단계 3. 모든 출력 유니트 j의 연결강도를 다음과 같이 갱신한다.

$$\Delta \mathbf{W}j=\eta \cdot \mathrm{A}(j, i^*) \cdot E(j) \cdot (\mathbf{X}-\mathbf{W}j)$$

여기에서 $\eta=1/\mathrm{T}$

$\mathrm{A}(j, i^*)=\exp[-\mathrm{R}_{i^*j}^2 \times \mathrm{T}]$

$E(j)=1/(\mathrm{K}_j+1)$

$\mathrm{T}=$반복횟수

$\mathrm{R}_{i^*j}=$출력 유니트 i^*와 출력유니트 j간의 거리

$\mathrm{K}_j=$그때까지 출력 유니트 j가 최소 거리 출력 유니트로 결정된 횟수

단계 4. 모든 입력 벡터에 대하여 단계 2, 단계 3을 적용한다.

단계 5. 모든 출력 유니트의 연결강도가 더이상 변화하지 않거나 원하는 크기이내의 변화를 보일 때까지 $\mathrm{K}_j=0$, $\mathrm{T}=\mathrm{T}+1$로 하고 단계, 2, 단계 3, 단계 4를 반복한다.

같이 표현될 수 있다.

적용성 고찰

본 기법의 적용성을 고찰하기 위하여 사용된 자료는 항공기를 이용하여 획득된 원격탐사자료로서 256×256 화소와 크기에 총 6개 밴드를 가지고 있다. 이의 False color composite 영상이 Fig. 3의 (a) 에 나타나 있다.

우선 이 자료에 본 연구에서 설정한 신경망 모델 및 변수들을 이용하여 군집화를 수행하여 보았다. 우선적으로 학습률을 반복횟수의 역수로 하여 적용하여 보았으며 출력 유니트의 수는 7개로 하였다. 출력 유니트의 수를 7개로 한 것은 자료의 대부분 지역이 논이며 그외에 물과 산이 약간 존재하는 비교적 단순한 토지피복을 보이기 때문이다. 모든 출력 유니트의 연결강도의 변화가 0.1% 이하로 되면 반복을 멈추도록 하였을 때 14회의 반복으로 군집화가 완료되었다. 학습률을 반복횟수의 역수의 제곱근으로서 설정하였을 때는 9회의 반복으로 군집화가 완료되었으며 반복횟수와 관계없는 상수(1.0) 하였을 때는 4회의 반복으로 군집화가 완료되었다. 직접적인 비교가 될 수는 없으나 Hung(1993) 이 128×128 화소 (3개밴드)의 자료를 군집화하면서 30회의 반복을 수행하였던 것과 비교하여 비교적 반복횟수가 적음을 알 수 있다.

한편 이때 물부분이 군집화되어가는 과정을 Table 3에 나타냈다. Table 3에서 알 수 있

듯이 초기치는 30 이하의 작은 임의의 값으로 설정되어 있음에도 3 회의 반복으로 군집중심에 상당히 접근하고 있음을 알 수 있다. 학습률을 1.0 으로 하였을 때는 4 번째의 반복단계에서 군집화가 종료되었으나 학습률을 반복횟수의 역수나 이의 제곱근으로 하였을 때는 4 회 이후의 반복단계부터 매우 느린 속도로 군집중심이 변화되고 있음을 알 수 있다. Table 3 으로부터 알 수 있듯이 학습률은 군집화를 종료하는 반복횟수를 결정하며, 이 때 군집화 결과도 다르게 나타난다.

한편 초기치와 학습률에 관계없이 첫번째 반복단계에서 모두 같은 값을 보이고 있어 본 신경망 기법이 초기치에 관계없이 항상 같은 결과를 보임을 유추할 수 있다. 이를 확인하기 위하여 군집화 종료기준을 연결강도의 변화가 1% 이하일 때로 하고 5 번의 군집화를 수행하였다. 이때는 6 회의 반복으로 군집화가 종료되었으며 물부분의 군집화과정을 Fig. 2 에 파란색에 해당하는 파장대인 밴드 1과 근적외 파장대인 밴드 4를 기준으로 나타냈다. 초기치는 임의의 작은 값으로 하였으므로 좌하쪽에 임의의 위치에 분포하고 있으나 첫번째 반복을 통하여 하나의 지점으로 모이고 있다. 이후 두번째 반복과정을 통하여 실제 군집중심에 상당히 접근하며 세번째 반복과정이후로는 군집중심이 거의 변화하지 않는다. 이는 본 신경망 기법이 초기치의 분포에 관계없이 하나의 군집중심으로 수렴함을 나타내고 있다.

군집화 결과를 평가하는 일반적인 방법으로서는 군집중심간의 거리를 비교하는 방법과 각 군집의 밴드별 분산을 고찰하는 방법이 있다. Table 4 에는 신경망 기법과 K 평균 기법의 적용 결과를 군집중심간의 거리로서 비교하여 나타냈다. 여기에서 K 평균 기법에서는 대개 반복횟수 4 회 이내에서 군집화가 종료되었다. 그러나 K 평균 기법은 초기치의 분포에 크게 의존하는 결과를 보임을 적용과정을 통하여 확인할 수 있었다. 사실상 임의의 초기치를 설정해서는 원하는 결과를 얻기가 매우 어려웠으며 자료값의 분포에 대한 통계적 자료를 이용하여 초기치를 실제 자료값이 분포하는 범위내의 값으로 하여야 원하는 군집화 결과를 얻을 수 있었다. 본 연구에서는 미리 각 밴드별 평균 및 표준편차를 계산하여 초기치를 평균 ±2.0x표준편차 내에 분포하도록 설정하고 총 군집의 개수를 7 개로 군집화를 수행한 결과를 이용하였다. 결과의 비교를 쉽게 하기 위하여 Table 4 에서 군집중심간 거리가 가장 크게 나타난 결과는 역상으로 나타냈다. 전체적으로 K 평균 기법의 적용 결과에서 비교적 신경망 기법에서보다 더 큰 군집중심간 거리를 보이고 있다. 한편 Table 5 에는 각 군집의 밴드별 분산을 나타냈으며 분산이 가장 작게 나타나는 결과를 역상으로 표현하였다. Table 5 에서 Cluster 5 를 제외하고는 전체적으로 신경망 기법이 K 평균 기법보다 우수한 결과를 보이고 있으며 학습률을 반복횟수의 역수로 하였을 때 가장 작은 값의 분산을 나타내는 경우가 많았다. 따라서, 군집중심간 거리에 있어서는 K 평균 기법이, 군집의 분산에 있어서는 신경망 기법이 더 양호한 결과를 보이고 있어 이러한 비교만을 통해서 두 기법의 우열을 판단하기는 어렵다.

한편, 학습률을 반복횟수의 역수로서 하고 신경망 기법을 이용하여 형성된 군집을 가지고 전체 자료를 최소거리법으로서 분류한 결과를 Fig. 3의 (b) 에 나타냈다. 우하쪽에 있

Table 3. Variation of weights with iterations.

Iter.	Band / η	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6
Initial Value	1/T	25.00	8.00	16.00	18.00	3.00	13.00
	$\sqrt{(1/T)}$	13.00	12.00	2.00	12.00	20.00	9.00
	1	17.00	23.00	1.00	10.00	25.00	4.00
1	1/T	155.37	150.31	117.73	99.93	72.76	95.27
	$\sqrt{(1/T)}$	155.37	150.31	117.73	99.93	72.76	95.27
	1	155.37	150.31	117.73	99.93	72.76	95.27
2	1/T	110.71	106.31	67.62	35.15	58.29	88.02
	$\sqrt{(1/T)}$	110.15	106.06	66.11	32.36	57.53	87.67
	1	110.76	106.98	66.42	31.78	57.33	87.55
3	1/T	108.15	103.48	63.90	31.26	57.35	87.25
	$\sqrt{(1/T)}$	109.13	104.93	64.78	31.15	57.22	87.41
	1	110.70	106.93	66.29	31.59	57.28	87.52
4	1/T	107.77	102.88	63.53	31.27	57.44	87.17
	$\sqrt{(1/T)}$	108.79	104.47	64.45	31.08	57.22	87.41
	1	110.70	106.93	66.29	31.59	57.28	87.52
5	1/T	107.61	102.55	63.40	31.40	57.54	87.14
	$\sqrt{(1/T)}$	108.55	104.14	64.23	31.06	57.24	87.32
6	1/T	107.52	102.34	63.33	31.53	57.62	87.13
	$\sqrt{(1/T)}$	108.38	103.89	64.07	31.07	57.26	87.29
7	1/T	107.47	102.19	63.29	31.65	57.70	87.12
	$\sqrt{(1/T)}$	108.25	103.69	63.95	31.08	57.28	87.26
8	1/T	107.43	102.08	63.27	31.83	57.76	87.12
	$\sqrt{(1/T)}$	108.15	103.53	63.86	31.10	57.30	87.24
9	1/T	107.41	101.99	63.26	31.83	57.76	87.13
	$\sqrt{(1/T)}$	108.06	103.40	63.78	31.12	57.32	87.23
10	1/T	107.40	101.93	63.26	31.90	57.81	87.13
11	1/T	107.39	101.88	63.26	31.96	57.89	87.14
12	1/T	107.38	101.83	63.26	32.01	57.92	87.14
13	1/T	107.38	101.80	63.26	32.06	57.95	87.14
14	1/T	107.37	101.77	63.26	32.10	57.97	87.15

는 산의 그림자 부분중 일부가 물로 분류되었으나 이는 (a) 의 False color composite 영상을 통해서도 알 수 있듯이 산 그림자 지역의 분광적 특성이 물부분과 실제로 매우 유사했기 때문으로 판단된다.

Fig. 3의 (c) 에는 K 평균 군집화 기법을 이용하여 군집화를 수행한 후 최소거리법을 이용하여 분류한 결과를 나타냈다. 앞에서 언급하였듯이 이 분류 결과는 미리 각 밴드별 통계적 처리를 거친 후에 군집화를 수행한 결과이다. 앞에서 검토하였던 군집중심간 거리 및 각 군집의 밴드별 분산만을 통하여는 확인할

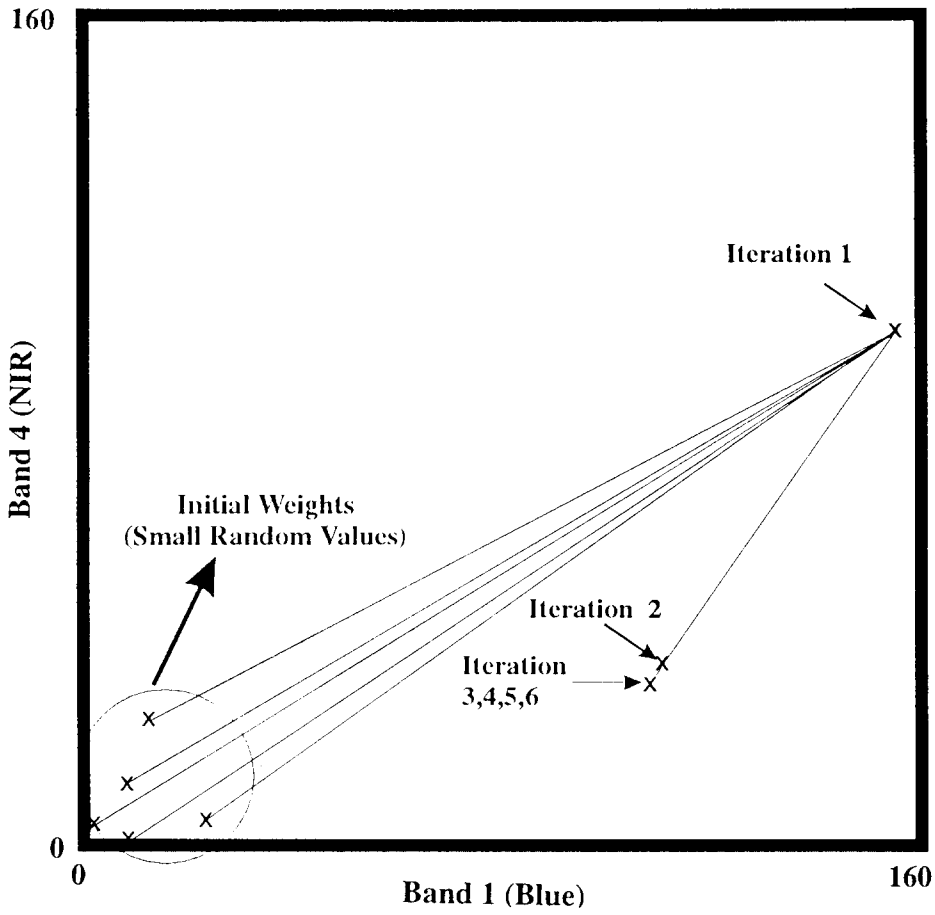


Fig. 2 Convergency of weights with iterations.

수 없었으나 우하쪽의 산그림자 지역이외에도 좌하쪽의 일부 논들이 물로 분류되고 있어 정확도면에서 신경망기법에 뒤지는 결과를 보이고 있음을 알 수 있다. 반면에 K 평균 군집화 기법은 군집화 완료 기준을 0.1% 로 하여도 4회 정도의 반복으로 군집화를 완료하여 계산 시간면에서는 신경망 기법보다 상당히 효율적이었다.

Fig. 3의 (d) 에는 순차적 군집화 적용결과를 가지고 최소거리법을 이용하여 감독분류를 수행한 결과를 나타냈다. 순차적 군집화에서는 입력변수인 최대 군집반경 및 군집간 최소거리에 따라 군집화 결과가 매우 다르게 나타난다. 따라서 최적의 군집화를 위해서는 수차례의 시행착오를 거쳐서 적절한 최대 군집반경 및 군집간 최소거리를 결정해야 한다. 그림에 나타낸 결과는 최대 군집반경을 5, 군집간 최소거리를 10으로 한 경우의 군집화 결과이다. 그림에서 알 수 있듯이 전체적으로 많은 논들이 좌상쪽 논들과 같은 분류항목으로 분류되

Table 4. Distance between cluster means.

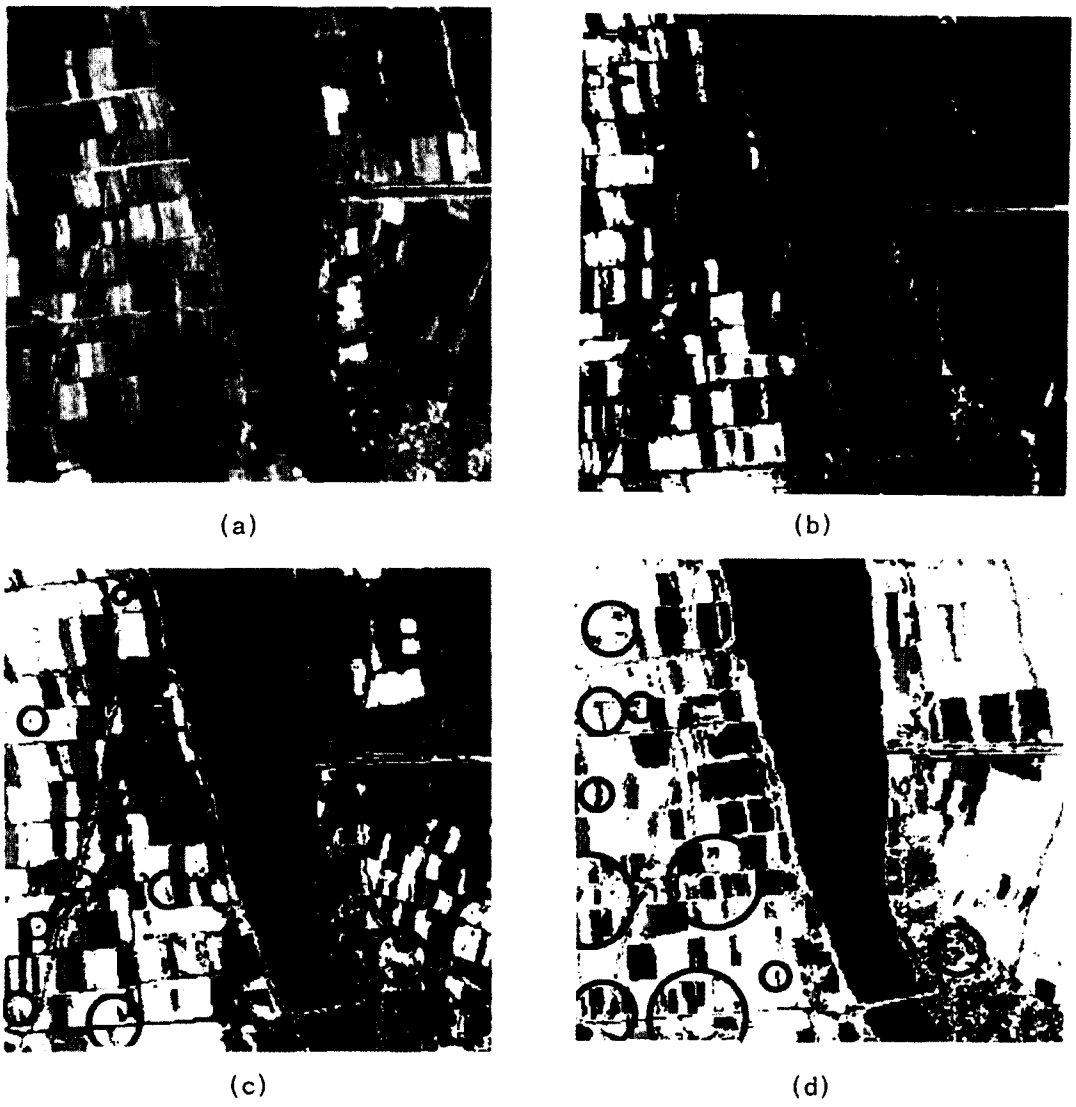
CLUSTER / CLUSTER		1	2	3	4	5	6	7
1	$\eta = 1/T$	-	71.05	39.48	134.67	136.30	72.74	57.98
	$\eta = \sqrt{(1/T)}$	-	76.62	36.08	138.60	134.56	73.46	55.20
	$\eta = 1$	-	79.78	39.44	140.72	134.15	79.83	59.65
	K-means	-	79.78	39.44	140.71	134.15	79.84	59.65
2	$\eta = 1/T$		-	55.99	125.49	184.75	66.18	110.50
	$\eta = \sqrt{(1/T)}$		-	64.23	134.56	186.17	72.70	111.08
	$\eta = 1$		-	70.67	142.57	188.37	83.18	114.94
	K-means		-	70.67	142.57	188.37	83.18	114.94
3	$\eta = 1/T$			-	97.56	175.26	33.38	96.56
	$\eta = \sqrt{(1/T)}$			-	104.54	170.36	37.71	90.43
	$\eta = 1$			-	103.46	173.45	41.38	98.20
	K-means			-	103.44	173.47	41.37	98.21
4	$\eta = 1/T$				-	267.48	66.68	190.50
	$\eta = \sqrt{(1/T)}$				-	269.39	70.28	191.15
	$\eta = 1$				-	270.66	66.80	197.02
	K-means				-	270.66	66.79	197.02
5	$\eta = 1/T$					-	208.33	78.86
	$\eta = \sqrt{(1/T)}$					-	207.23	80.49
	$\eta = 1$					-	213.06	78.05
	K-means					-	213.07	78.05
6	$\eta = 1/T$						-	129.70
	$\eta = \sqrt{(1/T)}$						-	127.17
	$\eta = 1$						-	137.38
	K-means						-	137.39

η = Learning rate, T = Iteration

Table 5. Variances of each cluster.

BAND / CLUSTER		1	2	3	4	5	6
1	$\eta = 1/T$	**97.52**	**76.79**	**93.08**	**195.15**	**151.37**	**255.01**
	$\eta = \sqrt{(1/T)}$	104.58	77.95	99.73	198.13	160.22	276.01
	$\eta = 1$	131.54	96.198	126.48	212.21	168.98	290.33
	K-means	131.52	96.18	126.46	212.17	168.95	290.28
2	$\eta = 1/T$	**213.95**	203.65	**198.59**	734.01	**132.37**	**273.28**
	$\eta = \sqrt{(1/T)}$	226.77	**202.01**	204.79	593.97	136.63	286.59
	$\eta = 1$	264.11	223.88	226.61	**534.95**	154.17	324.76
	K-means	264.11	223.88	226.61	**534.95**	154.17	324.76
3	$\eta = 1/T$	47.35	**41.65**	**46.43**	**101.44**	**91.96**	**170.22**
	$\eta = \sqrt{(1/T)}$	**46.92**	47.67	59.13	123.34	111.79	186.67
	$\eta = 1$	58.35	58.91	74.24	142.02	126.01	203.95
	K-means	58.33	58.89	74.22	141.99	125.97	203.88
4	$\eta = 1/T$	29.75	55.06	26.84	**12.80**	5.45	**9.26**
	$\eta = \sqrt{(1/T)}$	26.22	41.09	24.41	14.17	**5.14**	9.38
	$\eta = 1$	19.80	**28.66**	**18.49**	14.74	5.93	10.46
	K-means	**19.17**	**28.66**	**18.49**	14.73	5.93	10.46
5	$\eta = 1/T$	152.12	201.24	820.14	654.78	456.22	772.72
	$\eta = \sqrt{(1/T)}$	145.03	185.29	780.77	627.83	451.47	760.26
	$\eta = 1$	116.13	133.81	689.45	582.25	**443.04**	747.54
	K-means	**115.99**	**133.67**	**689.40**	**582.06**	**443.04**	**747.47**
6	$\eta = 1/T$	**39.08**	**47.88**	**60.34**	**114.58**	62.73	115.08
	$\eta = \sqrt{(1/T)}$	46.92	58.78	67.61	141.36	47.14	95.67
	$\eta = 1$	48.40	60.43	68.04	149.56	**35.60**	**75.20**
	K-means	48.40	60.44	68.05	149.63	35.61	75.22
7	$\eta = 1/T$	300.11	157.31	229.68	708.97	240.53	477.76
	$\eta = \sqrt{(1/T)}$	284.35	154.05	207.85	**698.18**	**238.06**	**477.34**
	$\eta = 1$	**243.49**	**151.88**	**202.07**	698.83	249.13	491.57
	K-means	243.51	151.92	202.22	698.86	249.15	491.67

η = Learning rate, T = Iteration



(a) (b)

(c) (d)

Fig. 3 (a) False color composite image of sample data(band 4, 3, 2).
(b) Classified result by Minimum distance classifier (trained by competitive learning network).
(c) Classified result by Minimum distance classifier (trained by K-Means clustering technique). Misclassified pixels are shown in the black circles.
(d) Classified result by Minimum distance classifier (trained by Sequential clustering technique). Misclassified pixels are shown in the black circles.

었다. 이는 순차적 군집화 기법의 특징적인 단점으로서 앞에서 언급했던 바 있다. 또한 좌측 일부분과 좌하쪽의 많은 논들이 물로 분류되어 분류의 정확도면에서 매우 떨어지는 결과를 보이고 있다. 단, 더 많은 시행착오를 통하여 보다 적절한 최대 군집반경 및 군집중심간 최소거리를 찾는다면 다소 향상된 군집화 결과를 기대할 수 있을 것이다. 순차적 군집화는 반복과정이 필요없어 계산시간면에서 효율적이라고 하나 이와같이 적절한 입력변수를 찾기 위한 노력 및 시간을 필요로 하므로 양질의 결과를 얻기위한 총 투자시간을 고려할 때 신경망기법이나 K 평균 및 ISODATA 등의 반복적 기법보다 효율적이지 않을 수도 있다.

결 론

본 연구에서는 최근 많은 분야에서 패턴인식을 위한 효과적인 기법으로 이용되고 있는 신경망 기법을 원격탐사자료의 군집화 기법으로서 적용하고자 하였다. 이를 위해 선택된 신경망 모델은 경쟁학습 신경망이며 이를 구성하는 각종 변수들을 재구성하여 원격탐사자료의 군집화를 위한 신경망 모델을 설정하였다.

본 기법은 항공기를 이용하여 획득된 원격탐사자료를 이용하여 순차적 군집화 기법 및 K 평균 군집화 기법과 비교되었다. 비교결과, 각 기법의 우열을 논하기는 어려우나 신경망 군집화 기법의 장,단점은 다음과 같이 요약될 수 있다.

1. 자료값의 분포에 대한 사전지식을 필요로 하지 않는다.
2. 초기치에 의존하는 경향이 없어 임의의 초기치를 설정할 수 있으므로 사전 통계적 처리가 불필요하며 총 군집의 수만 같다면 항상 동일한 출력 결과를 얻을 수 있다.
3. K 평균 기법에 비하여 비교적 많은 반복 횟수가 요구되어 다소 많은 계산시간이 필요하나 정확도면에 있어서는 순차적 군집화나 K 평균 군집화 기법에 비하여 비교적 우수하다.

자료값의 분포에 대한 사전지식을 필요로 하지 않으며 초기치에 관계없이 항상 동일한 군집화 결과를 나타내는 것은 매우 큰 장점으로 생각되며 향후 영역함수 및 학습률에 대한 보다 많은 고찰을 통하여 계산시간측면에서 향상될 수 있다면 더욱 효과적인 군집화 기법으로서 활용이 기대된다.

참 고 문 헌

C. C. Hung, 1993, Competitive learning networks for unsupervised training, Int. J. Remote Sensing, Vol.14, No.12, pp.2411－2415.

R. A. Schowengerdt, 1983, Techniques for image processing and classification in remote sensing, Academic Press, pp. 147－148.

S. Y. Kung, 1993, Digital neural networks, Prentice Hall, pp.85－90.

T. Kohonen, 1982, Analysis of a simple self－organizing process, Biological Cybernetics, Vol.44, pp.135－140.