

Echelon Analysis

문승호¹⁾ · 신재경²⁾

요 약

인공위성을 이용한 원격탐사가 일반화되어 있는 선진국의 경우 인공위성의 관측자료를 이용한 여러 가지 분야의 연구가 활발하게 이루어지고 있다. 이러한 국제적 경향은 결과적으로 원격탐사정보와 지리정보시스템 기술의 가져오게 되는 계기가 되었다. 이로 인해 환경조사 및 자원탐사, 홍수, 가뭄, 쓰레기 처리, 농업 식량예측, 사막화 해양오염 등 여러 가지 응용 분야에 유용하게 사용되고 있다.

에쉬론 해석법(Myers et. al., 1997)은 이러한 원격탐사에 의해 얻은 각종 공간 데이터 해석을 위해 개발된 것으로, 공간데이터의 위상적 구조를 계통적이고 객관적으로 발견하기 위한 해석법이다. 본 연구에서는 한국의 시·도별 인구 데이터에 본 해석법을 적용해 본다.

주제어 : echelon analysis, spatial data, echelon dendrogram

1. 공간 데이터의 echelon 해석

최근 원격탐사(remote sensing)에 사용되는 관측기(censor)의 성능이나 기술의 향상과 함께, 자연환경에 관한 다양한 데이터 관측이 쉬워졌다. 이 원격탐사에 의해 물체로부터 반사 또는 방사되는 전자파의 차이를 이용해서 지구상의 물체·물질의 상태를 조사할 수 있다. 이 때, 전자파란 사람이 볼 수 있는 빛(청, 녹, 적외 가시광), 자외선, 적외선 등을 총칭한 것이다. 반사특성은 물체에 따라 서로 달라서 식물, 물, 토양 등은 전혀 다른 패턴이 된다.

따라서, 각 물체로부터 반사되는 전자파의 강약 데이터를 몇 개의 밴드(band, 波長帶)로 나눠서 측정함으로써 그 물체에 대한 식별(/분광특성)과 함께 다양한 정보를 얻을 수 있다.

원격탐사에서 반사율 데이터는 각 화소(pixel)를 요소로 하는 배열데이터이며, 이들 배열데이터는 각 밴드마다 얻어지는데, 공간적 위치구조를 지닌 다변량 데이터로 취급할 수 있다. 또한 각 밴드의 이러한 영상 자료는 동일 지역을 나타내는 독립적인 영상자료로서 저장된다.

이처럼 remote sensing data는 공간적 위치구조를 자료로서 제공하는데, 이와 같은 공간 데이터 구조를 해석하기 위해서는 데이터를 가시화한 데이터의 2차원, 3차원적 그래픽 표현이 유효하다. 지리정보시스템(GIS: geographic information systems)은 이와 같은 공간 데이터 해석에 있어서 강력한 도구로 사용되고 있다. 그렇지만 이러한 도구를 이용한 공간 데이터 구조의 객관적인 취급은 곤란함을 동반하는 것이 일반적인 현상이다.

이러한 상황에서 echelon 해석은 배열상에 분류된 지도상의 1변량 데이터에 대해 공간적 위치를 표면상의 데이터 高低(濃淡)를 바탕으로 분할, 공간 데이터의 위상적 구조를 계통적이고 객관적으로 발견하기 위해서 개발된 해석법이다. echelon 해석에서 사용되는 echelon 지도나 구조를 계층적으로 나타낸 echelon dendrogram은 공간 데이터의 구조를 꼭 맞게 표현한 그래프이다. 또한, echelon 해석은 1변량 데이터에 축약된 remote sensing data 뿐만 아니라 위치정

1 608-738 부산시 남구 우암동 부산외국어대학교 정보통계학과, 부교수
2 641-773 경남 창원시 사림동 창원대학교 통계학과, 부교수

Echelon Analysis

보를 제시해 줌으로써 각종 지리 데이터에도 적용이 가능하다.

2. 데이터 고저가 1차원적인 수평 위치인 경우

데이터 고저가 1차원적인 수평위치로 주어졌을 경우, 공간(수평) 위치 x 에서의 데이터 값(고저를 나타냄)을 h 라 하면 데이터는 (x, h) 로 표시된다. 그림 2-1에 나타낸 것과 같은 데이터가 있을 경우, 위상적(位相的)으로 동일한 9개 계급으로 나눌 수가 있다.

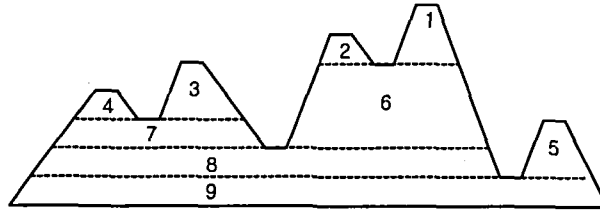


그림 2-1 echelon 해석에 있어서 동일 위상 영역으로의 분할

이들 계급이 각각 echelon이며, 그림 2-1에 주어진 번호가 echelon 번호(echelon number)이다. echelon 1부터 5는 피크(peak)이며 1차 echelon이라 부른다. echelon 6은 2개 이상의 피크(1차 echelon 1과 2)의 파운데이션(foundation)이며, 2차 echelon이라고 한다. 마찬가지로 echelon 7은 1차 echelon 3과 4의 파운데이션이며 2차 echelon이다. echelon 8은 2개 이상의 2차 echelon 6과 7의 파운데이션이며 3차 echelon이라 부른다. echelon 9는 echelon 5와 8의 파운데이션이지만, echelon 8은 3차 echelon 5는 1차이므로 echelon 차수는 3차이다. 또한, echelon 9는 루트(root)라고도 한다.

그림 2-1에서 보여주는 데이터 구조는 그림 2-2와 같은 echelon 덴드로그램(dendrogram)으로 주어진다.

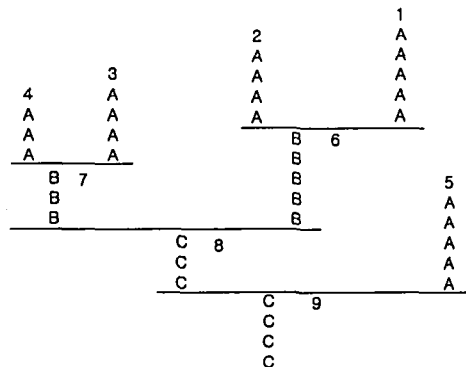


그림 2-2 echelon dendrogram

3. 데이터 고저가 간단한 구조인 2차원 공간인 경우

데이터 고저가 2차원 공간에 주어진 경우, 공간의 위치 (x, y) 에 있어서의 데이터 값(고저를 나타냄)을 h 라 하면, 데이터는 (x, y, h) 로 표현된다. 일반적으로 지형도(地形圖)와 같이 데이

터 고저가 비교적 간단한 구조로 주어진 경우, 이들 세 변수간에는 연속함수 $h=f(x, y)$ 의 관계가 있다.

이제 그림 3-1과 같은 2차원 공간상의 지형도를 고려해 보자.

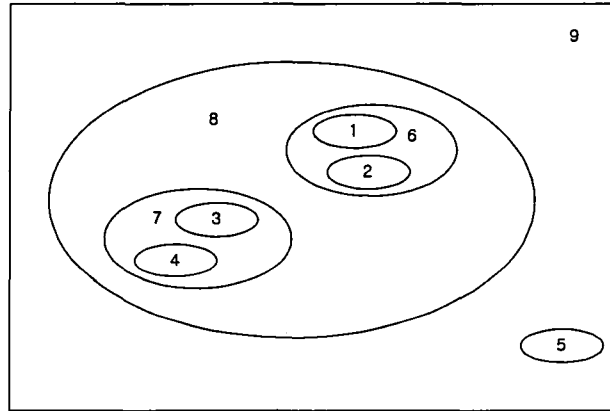


그림 3-1 2차원 공간상의 지형도와 echelon

이 지형도에서는, base 9 위에 작은 피크 5와 큰 파운데이션 8이 있으며, 파운데이션 8 위에는 두 개의 파운데이션 6과 7이 있다. 또한, 파운데이션 6 위에는 2개의 피크 1과 2, 파운데이션 7 위에는 2개의 피크 3과 4가 있다. 피크 1부터 5는 번호순으로 값이 작으며, 파운데이션 6이 7보다 값이 클 경우, 그림 3-1에서의 데이터 구조도 그림 2-2와 같은 echelon dendrogram으로 주어진다.

4. 데이터 고저가 2차원 공간의 배열상에 주어진 경우

remote sensing data와 같이 데이터 고저(高低)가 $n \times m$ 배열상의

$$D_{ij} = \{(x, y) \mid x_{i-1} < x < x_i, y_{j-1} < y < y_j\}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m$$

digital 값 h 로 주어진 공간 데이터인 경우, 데이터는 (i, j, h) 로 나타낸다. 이와 같은 공간 데이터의 경우 이들 세 변수의 관계를 나타내는 함수 $h=f(i, j)$ 는 離散的이며 복잡한 함수가 되어 구조도 복잡하게 된다. 그림 4-1과 같은 5×5 배열상의 데이터가 주어질 때, 아래와 같은 방법으로 echelon 해석이 진행된다. 단, 인접한 데이터란 대상으로 하고 있는 echelon에 포함되는 데이터에 인접하는 데이터 집합을 뜻한다.

2	24	8	15	3
10	1	14	22	5
4	13	19	23	25
20	21	12	11	17
16	6	9	18	7

2	24	8	15	3
10	1	14	22	5
4	13	19	23	25
20	21	12	11	17
16	6	9	18	7

그림 4-1 5×5 배열상의 데이터

Echelon Analysis

(1). peak의 발견

peak에 속하는 데이터 값은 동일한 peak에 속하는 데이터 이외의 인접하는 데이터 값보다 크다. 그림 4-1의 데이터에 있어서는 4개의 peak가 있다.

① 제1피크

최대값은 25이다. 우선, 25는 제1피크에 포함된다. 25에 인접하는 데이터의 최대값은 23으로, 23은 (25, 23)에 인접하는 데이터보다 크므로 제1피크에 포함된다. (25, 23)에 인접하는 데이터의 최대값은 22로, 22는 (25, 23, 22)에 인접하는 데이터보다 크므로 22도 제1피크에 포함된다. (25, 23, 22)에 인접하는 데이터 중에서 최대값은 19이다. 하지만 19는 (25, 23, 22, 19)에 인접하는 21보다 작으므로 제1피크에 속하지 않는다. 따라서 제1피크는 데이터 25, 23, 22에 의해 구성되며, echelon 번호는 1(En 1)이다. 이들 데이터는 동일한 피크 이외의 인접하는 데이터보다 크다.

② 제2피크 이후

제1피크를 제외한 데이터의 최대값은 24이다. 우선, 24는 제2피크에 포함된다. 24에 인접하는 데이터의 최대값은 14이지만 인접하는 23보다 작기 때문에 제2피크에 속하지 않는다. 그러므로 제2피크 (En 2)는 24에 의해서만 구성된다. 동일한 순서에 의해 제3피크 (En 3)는 21, 20, 제4피크 (En 4)는 18에 의해 구성된다.

(2). foundation의 발견

① 제1파운데이션

4개 피크에 포함되는 데이터를 제외한 최대값은 19이다. 19는 제1피크 (En 1)와 제3피크 (En 3)의 파운데이션이며 echelon번호는 5가 된다 (En 5). (En 1, 3, 5)에 인접하는 데이터의 최대값은 17이다. 그러나 17은 제4피크의 18보다 작으므로 17은 En 5에 속하지 않는다. En 5는 En 1과 En 3의 parent이며 이 관계는 echelon 번호를 이용해서 5(1 3)로 표시된다. 이후 파운데이션을 찾게 되면 En 1과 En 3은 사용되지 않고 대표로 En 5를 이용한다.

② 제2파운데이션

En 1부터 5에 포함되는 데이터를 제외한 최대값은 17이다. 17은 En 5와 En 4의 파운데이션이며 En 6이다. En 6은 En 5와 En 4의 parent가 되며 6(5(1 3)4)이다. 이후, En 1, En 3, En 4, En 5는 대표로 En 6을 이용한다. En 1부터 6에 포함되는 데이터를 제외한 최대값은 16이다. 이 16은 (En 1, 3, 4, 5, 6)에 인접하는 데이터보다 크므로 En 6에 속한다. En 1부터 6 (단, 16은 En 6에 포함된다)에 포함되는 데이터를 제외한 최대값은 15이다. 이 15도 En 6에 속한다. En 1부터 6(단, 16, 15는 En 6에 포함된다)에 포함되는 데이터를 제외한 최대값은 14이다. 그렇지만 14는 (En 1, 3, 4, 5, 6)에 인접하는 데이터의 제2피크의 24보다 작으므로 14는 En 6에는 속하지 않는다.

③ 제3파운데이션(root)

En 1부터 6에 포함되는 데이터를 제외한 최대값은 14이다. 14는 En 2와 En 6의 파운데이션이며 En 7이다. En 7은 En 2와 En 6의 parent가 되며 7(2 6(5(1 3)4))이다.

④ root

En 1부터 6에 포함되지 않는 13이하의 데이터는 루트가 된다.

이상에 의해서 이 5×5 배열 데이터의 구조는 그림 4-2와 같은 echelon dendrogram에 의해 주어진다.

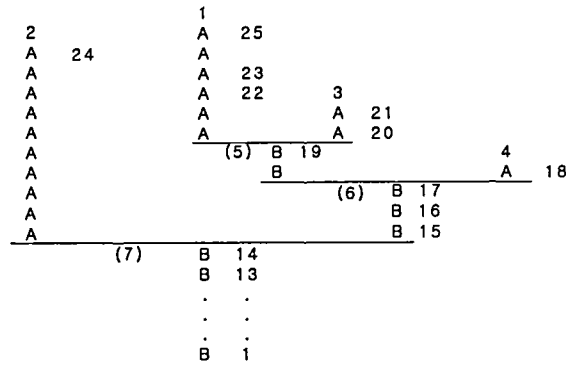


그림 4-2 5×5 배열 데이터의 echelon dendrogram

5. 지리 데이터의 경우

지리 데이터의 경우는 지역의 인접정보를 제공해 줌으로써 echelon 해석이 가능해 진다. 여기서는 아래 한국의 16개 시도별 인구데이터를 예로써 들어보자.

시도명	총인구	15세 미만 비율	15세~64세 비율	65세 이상 비율
전국	48,021,543	0.206	0.720	0.074
서울특별시	10,263,336	0.182	0.760	0.057
부산광역시	3,770,536	0.181	0.754	0.064
대구광역시	2,525,109	0.206	0.732	0.062
인천광역시	2,564,598	0.225	0.718	0.057
광주광역시	1,383,765	0.230	0.711	0.058
대전광역시	1,403,164	0.222	0.721	0.057
울산광역시	1,055,618	0.243	0.715	0.042
경기도	9,544,496	0.236	0.705	0.059
강원도	1,552,407	0.194	0.708	0.098
충청북도	1,496,520	0.205	0.699	0.095
충청남도	1,918,137	0.197	0.686	0.118
전라북도	2,006,454	0.201	0.692	0.107
전라남도	2,099,308	0.190	0.684	0.126
경상북도	2,784,704	0.191	0.697	0.112
경상남도	3,106,502	0.211	0.701	0.088
제주도	546,889	0.221	0.696	0.083

(출처: 2001년도 주민등록인구, 통계청)

이 자료에 대한 에쉬론 분석 결과는 발표 당일 제시하기로 함.

참고문헌

栗原考次(2001), データの科學, 放送大學振興會, 日本.

Echelon Analysis

- F. Ishioka and K. Kurihara(2002), The Evaluations of Spatial Structure based on Echelon Measures, Proceedings of the 2nd Joint Conference on Recent Development in Statistics, 109-118.
- K. Kurihara and H. Hong(2002), Detection of hotspots for Geospatial Data with Echelon Analysis based on Spatial Scan Statistics, Proceedings of the 4th Conference of the IASC, 189-192.
- K. Kurihara, W.L. Myers and G.P. Patil (2000), Echelon analysis of the relationship between population and land cover pattern based on remote sensing data, COMMUNITY ECOLOGY, 1(1), 103-122.
- W.L. Myers and G.P. Patil (2002), Echelon analysis, Encyclopedia of Environmetrics, Vol. 2, 583-586.
- W.L. Myers, G.P. Patil and K. Joly (1997), Echelon approach to areas of concern in synoptic regional monitoring, Environmental and Ecological Statistics, 4, 131-152.