

## 단계구분도의 계급구간 설정을 위한 공간적 인접성의 최적화기법에 관해

손 일

### On the Optimization Methods of Spatial Proximity to Set the Class Intervals for the Choropleth Map

ILL SON

**요약** : 통계적 인접성을 증시하는 기존의 최적화 기법과는 달리 최대 BEI 지수를 추적하여 공간적 인접성을 최적화하는 새로운 최적화 기법을 개발하였다. 5가지 계급수에 대해 등간격, 시각구분법, 사분위법을 비롯해 최적화하여 얻은 최대 BEI, 최대 TEI를 나타내는 5가지 계급구분 방법으로 만들어진 25개의 단계구분도에 대해 계급구간 분류의 효율성을 분석하였다. 이때 사용된 평가기준은 TEI, BEI,  $C_F$ 이다. 최대 BEI와 최대 TEI를 나타내는 계급구간에서는 비교적 안정적이고 효율적인 분류 결과를 얻을 수 있었으나, 두 결과는 서로 상반된 결과를 보여주었다. 한편 시각구분법으로 얻은 분류 결과 역시 평가지수에서 높은 점수를 받았는데, 이는 새로운 지수인  $\alpha$ -index를 통해서도 확인할 수 있었다.

**주요어** : 계급구간, 최적화, BEI 지수,  $\alpha$ -index

**Abstract** : Differing from the existing method to optimize the statistical proximity, the new optimization method was developed to maximize the spatial proximity among the unit areas of choropleth map. 25 choropleth maps were made using 5 classification methods(maximum BEI, maximum TEI, equal interval, natural break, n-tile) for 5 classes(3-7) and were analyzed to elucidate the effectiveness of classification methods on the basis of 3 evaluation indices(TEI, BEI,  $C_F$ ). The results from the maximum TEI and maximum BEI were evaluated as relatively stable and effective compared to those from other classification methods. But they showed the opposite trends owing to the trade-off relationships. Meanwhile, the results from the natural break method got higher marks, which was also identified through the analysis using a new  $\alpha$ -index.

**Key Words** : class interval, optimization, BEI,  $\alpha$ -index

## 1. 서론

단계구분도란 면기호를 이용해 통계적 표면(statistical surface)을 나타내는 것이라고 간단히 정의할 수 있다(Robinson et al., 1995). 여기서 통계적 표면이란 지표의 기복과 같은 연속적인 3차원 자료모델이 아니라, 통계치가 집계되는 면단위, 다시 말해 행정구역의 경계선과 같은 불연속적인 경계로 나뉘어지는 3차원의 도수분포도와 같은 것이다. 따라서 단계구분도는 면단위로 집계된 자료를 여

러 계급으로 나누고 부여된 계급에 따라 해당되는 흑백 또는 컬러 면기호로 각 면단위를 표시해, 3차원의 자료모델을 2차원의 자료모델로 추상화한 것이다. 결국 이러한 2차원의 자료모델을 통해, 전체 지역의 공간적 패턴을 알아보려는 것이 단계구분도의 기본적인 목적이다.

단계구분도를 구축하는 데 있어서 가장 기본이 되는 요소는 단위지역의 크기, 모양 및 수, 계급의 수, 계급구간의 분류방법이다(이희연, 1995). 단위지역의 크기, 모양 및 수 그리고 계급의 수를 결정함

\* 경상대학교 사회교육학부 교수(Professor, School of Social Studies Education, Gyeongsang National University).

에 있어 여러 가지 논리적, 기술적, 시각적 제약이 가해지지만(정인철, 1998), 이 역시 지도제작자의 주관적 역할을 배제할 수 없다. 또한 이 두 가지 요소가 나름대로의 객관적인 기준을 가지고 결정되었다 하더라도, 세 번째 요소인 계급구간 분류방법에 따라 만들어지게 될 단계구분도의 수는 무수히 많다. 따라서 자료의 특성을 가장 잘 나타내는 지도를 선정하기란 매우 어려운 일이라 판단된다.

본 연구에서는 우선 계급구간 분류방법의 경우의 수를 계산할 수 있는 식을 개발하고, 이에 따른 단계구분도의 선택 및 객관성에 관해 논의하였다. 또한 계급구간 설정에 있어 통계적 인접성과 공간적 인접성이라는 두 가지 기준에 대해 간략히 살펴 본 후, 최적화 기법의 의미 및 기존의 최적화 방법에 대해 정리해 보았다. 마지막으로 계급내 자료 사이의 등질성과 계급간 이질성을 최대화하는 기존의 최적화 기법과는 달리, 인접한 면단위간의 양적 차이가 큰 경계가 계급간의 경계로 최대한 반영될 수 있는 새로운 최적화기법을 개발하였다. 특정 사례에 대해 기존의 분류방법과 새로이 개발된 분류방법을 이용하여 급간구분을 시도하고, 각각에 대해 여러 가지 평가기준을 이용해 단계구분도의 객관성을 확보할 수 있는 대안적 방법을 찾으려 했다.

## 2. 계급구간 분류방법의 경우의 수와 객관성

다음 절에서 설명하겠지만 계급구간을 설정하는

표 1. 6개 면단위 통계치를 3개의 계급구간으로 나눌 경우의 분류방법

1 계급구간	2 계급구간	3 계급구간
X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub> X <sub>4</sub> X <sub>5</sub> X <sub>6</sub>
X <sub>1</sub>	X <sub>2</sub> X <sub>3</sub>	X <sub>4</sub> X <sub>5</sub> X <sub>6</sub>
X <sub>1</sub>	X <sub>2</sub> X <sub>3</sub> X <sub>4</sub>	X <sub>5</sub> X <sub>6</sub>
X <sub>1</sub>	X <sub>2</sub> X <sub>3</sub> X <sub>4</sub> X <sub>5</sub>	X <sub>6</sub>
X <sub>1</sub> X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub> X <sub>5</sub> X <sub>6</sub>
X <sub>1</sub> X <sub>2</sub>	X <sub>3</sub> X <sub>4</sub>	X <sub>5</sub> X <sub>6</sub>
X <sub>1</sub> X <sub>2</sub>	X <sub>3</sub> X <sub>4</sub> X <sub>5</sub>	X <sub>6</sub>
X <sub>1</sub> X <sub>2</sub> X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub> X <sub>6</sub>
X <sub>1</sub> X <sub>2</sub> X <sub>3</sub>	X <sub>4</sub> X <sub>5</sub>	X <sub>6</sub>
X <sub>1</sub> X <sub>2</sub> X <sub>3</sub> X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>

방법(분류방법)은 여러 가지이다. 하지만 어느 경우든지 선택된 계급구간은 가정할 수 있는 전체 모집단 중의 하나이며, 이렇게 선택된 분류방법의 수월성 혹은 객관성은 전체 모집단과의 비교를 통해 평가되어야 할 것이다. 실제로 모집단의 수는 어느 정도일까? 예를 들어 6개의 면단위에 해당되는 통계치를 올림차순 혹은 내림차순으로 정리하여 3개의 계급으로 나눌 경우, 아래와 같이 10가지의 계급구간 분류방법을 얻을 수 있다(표 1).

이상과 같은 모집단의 배열형태를 기본 알고리즘으로 이용하여 N개의 면단위와 R개의 계급수라는 조건하에서 만들어질 수 있는 단계구분도의 모든 경우의 수를 계산할 수 있는 아래와 같은 식을 개발하였다. 이 식에 따르면, 예를 들어 200개의 면단위를 6개 계급으로 나눌 경우 2,472,258,789개의 단계구분도가 만들어질 수 있다.

$$\sum_{I_1=1}^{N-R_1+1} \sum_{I_2=I_1+1}^{N-R_1+R_2} \sum_{I_3=I_2+1}^{N-R_1+R_2+R_3} \dots \sum_{I_{R-2}=I_{R-3}+1}^{N-1-I_{R-2}}$$

혹은

$$(N-1)C_R = \frac{(N-1)!}{(N-1-R)! \cdot R!}$$

따라서 어떤 특정 현상을 설명하기 위해 지리학 관련 문헌에 제시된 단계구분도는 이처럼 무수히 많은 단계구분도 중에서 선택된 하나의 지도이므로, 그것을 선택함에 있어 지도제작자의 주관을 결코 무시할 수 없다. 또한 동일한 자료로부터 전혀 다른 설명이 가능한 지도가 만들어질 수 있기 때문에, 단계구분도 제작 기법이 이용되기 시작한 19세기초부터 단계구분도의 객관성을 확보하기 위한 여러 가지 방법이 도입되었다. 그 결과 이에 대한 연구도 1960년대 이후 활발히 진행되었으며, 단계구분도 연구에 있어 가장 대표되는 학자로는 George F. Jenks(1961, 1971, 1977)를 들 수 있다. 앞에서 제시된 3가지 요소 이외에도 변수의 특성, 단계구분도의 범례 설정 등 고려해야 할 사항들이 무수히 많다. 따라서 단계구분도의 객관성을 확보하는 일은 대단히 중요하지만, 그렇다고 쉽게 이루어지는 문제는 결코 아니라고 판단된다.

### 3. 계급구간 설정 방법

계급구간 설정은 계급수의 결정과 함께 단계구분도 제작에 있어 핵심적인 요소인 동시에 지도제작자의 주관에 개입할 소지가 가장 많다. 계급구간을 설정함에 있어 일반적으로 두 가지 기준을 적용하는데, 통계적 인접성(statistical proximity)과 공간적 인접성(spatial proximity)이 그것이다.

#### 1) 통계적 인접성

통계적 인접성을 기준으로 계급구간을 설정할 경우 주로 면단위 자료 분포의 수치적, 통계적 특성에 따라 급간을 설정하는데, 일반적으로 제시되는 계급구간 설정 방법은 대부분 여기에 속한다. 등간격, 사분위, 누진법, 평균·분산법, 시각적 구분법, 최적화 등이 대표적인 방법이며, 통계적 인접성을 기준으로 하는 계급구간 설정 방법에 대한 분류는 Evans(1977)의 연구에 잘 정리되어 있다.

#### 2)공간적 인접성

이 방법에서는 비공간적 과정을 거쳐 통계적 인접성을 강조하는 기존의 최적화 기법과는 달리, 면단위의 인접성과 연계성을 강조한다. 다시 말해 특정 현상의 유무 혹은 수치적 유사성에 따라 인접한 면단위를 하나로 묶으려는(individual regionalization) 시도를 말한다. 대표적인 것으로는 Monmonier(1972)의 "continuity-biased" 알고리즘과 Johnston(1968)의 "linkage method"가 있다.

Monmonier는 자료의 크기에 따라 정리한 통계치를 해당 면단위 각각의 인접성을 기준으로 재배

열한 후(그림 1-가의 1단계 : 전환), 여러 가지 방법으로 계급구간을 설정하였다(그림 1-가의 2단계 : 집단화). 그림 1-나는 그림 1-가의 1단계 전환 과정을 보다 상세히 설명한 것으로 공간적 인접성의 강조 정도에 따라  $\omega$ 의 값이 달라지고, 그 결과 계급구간 설정에 이용될 continuity-biased된 자료가 달라진다. 결국 여기에도 지도제작자의 주관에 개입될 수밖에 없으며, 계급구간 설정 방법의 선택에서도 마찬가지이다.

실제 Monmonier의 연구 결과에 의하면  $\omega$ 값을 증가시켜 공간적 인접성을 강조하더라도 그가 선택한 지도 효율성의 기준인 구획된 지도지역(map region)의 수는 그다지 감소하지 않았다. 또한 자료가 일정한 경향을 가진 채 분포하고 있을 경우 F-검정치가 조금 감소했으나, 자료가 불규칙하게 분포한 경우에는 F-검정치가 크게 줄어드는 결과를 초래하였다. 다시 말해 패턴의 단순성(pattern simplicity)과 집단내 등질성(within-group homogeneity) 사이에 trade-off 관계가 나타나기 때문에 이 역시 지도제작자의 주관성이 개입될 여지가 있다.

통계적 인접성을 기준으로 계급구간을 설정할 경우, 지리학에서 중요시하는 면단위간의 공간적 인접성이나 연계와 같은 변수는 무시되고 단지 개별 면단위가 정해진 계급구간에 편입될 뿐이다(typological regionalization). 실제로 자료의 분포 특성에 따라 적절한 설정 방법을 이용한다지만(예를 들어, 정규분포를 하고 있는 자료의 경우 평균·분산법을 이용), 결국 각각의 설정 방법에 대한 객관적 평가는 "계급구간내 동질성을 최대화하고, 계급

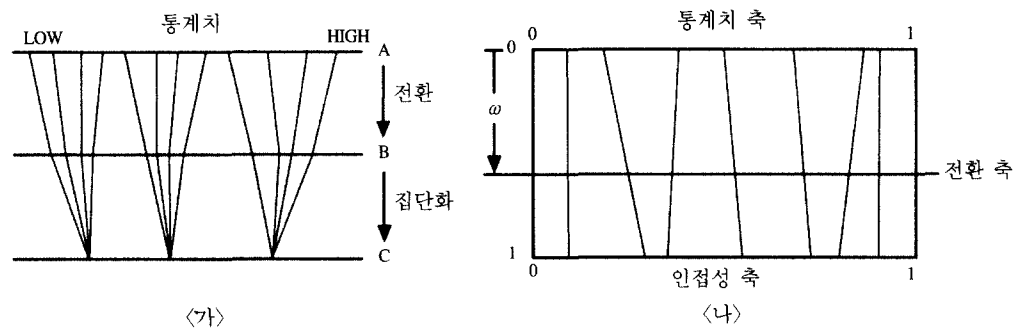


그림 1. 공간적 인접성을 고려한 계급구간 분류 모형(Monmonier, 1968, 212에서 전재)

구간 간에는 이질성을 최대화하는” 기준, 다시 말해 F-검정치 혹은 Jenks의 GVF 값과 같은 통계치, 다시 말해 통계적 인접성에 의존할 수밖에 없다.

한편, 공간적 인접성을 기준으로 계급구간을 설정할 경우, 자료 자체의 수치적 분포를 인정하면서도 가능한 한 이웃한 면단위가 유사한 통계치를 갖고 있다면 하나의 지도지역(map region)으로 통합시키는 방법이다. 이 경우 분할된 지도지역의 수, Jenks & Caspall(1971)의 BEI 지수, MacEachren (1982)의 map complexity 지수 등이 평가 기준이 될 수 있다. 하지만 너무 공간적 인접성만을 강조할 경우 F-검정치가 감소하거나 계급구간이 겹칠 수 있으며, 고려해야 할 변수가 너무 많아 최적화 프로그램을 만들기가 어렵다는 단점도 있다.

### 3. 직적화 기법

#### 1) 기존의 최적화 기법

기존의 최적화 기법은 면단위의 공간적 인접성에 대한 고려 없이 단지 자료의 통계적 인접성만을 강조한 것으로, 계급내 동질성과 계급간 이질성이 최대화되는 계급구간을 설정하는 방법이다. 앞에서 지적한 대로 200개의 면단위 자료를 6개 계급으로 분류할 경우 24억 여개의 단계구분도가 만들어질 수 있으며, 이 모든 경우의 F-검정치 혹은 GVF값을 계산·비교해 최대값을 갖는 한 가지 분류방법을 찾아내는 작업이다.

따라서 Jenks & Caspall(1971)의 연구를 필두로 최적화 기법에 관한 연구가 진행되었다. 초기의 최적화 기법에서는 고속 계산능력 및 알고리즘의 부재로 만족할만한 결과를 얻지 못했다. 즉, Jenks & Caspall(1971)의 최적화 기법에서는 임의의 초기 계급구간을 설정해 주고 iteration의 수를 한정시켰기 때문에, global optimum 보다는 local optimum에 빠질 위험이 있었다. 그후에 개발된 Jenks(1977)와 Lindberg(1990)의 기법에서는 Fisher(1958)의 알고리즘을 이용해 최적화를 시도하였다.

본 연구에서는 Lindberg의 최적화 프로그램의 결과가 모든 경우를 계산한 결과와 일치하는 지를 판단하기 위해, 아래와 같이 포트란 언어를 이용한

전수계산 알고리즘을 개발하고 이를 이용해 그 결과를 비교해 보았다. 그 결과는 정확히 일치하였고 면단위의 수가 50개를 넘지 않을 경우 계산 시간이 짧아 어느 알고리즘을 이용하더라도 큰 문제가 없었다. 그러나 면단위 수가 이보다 훨씬 많을 경우 Lindberg의 프로그램은 전수를 계산하지 않고 최대값만을 추적하기 때문에 계산 속도에 있어 본 연구의 전수계산 알고리즘 결과를 훨씬 능가하였다.

```
DO 100 I1=1, N-5
T1 = I1 + 1
DO 100 I2=T1, N-4
T2 = I2 + 1
DO 100 I3=T2, N-3
T3 = I3 + 1
DO 100 I4=T3, N-2
T4 = I4 + 1
DO 100 I5=T4, N-1
100 CONTINUE
```

\* 이 전수계산 알고리즘은 계급 수가 6개이고 여기서 N는 면단위의 수를 의미한다

#### 2) 인접한 면단위 간의 차이를 중시하는 최적화 기법

본 연구에서 개발하고자 하는 최적화 기법이란 기존의 최적화 기법과는 달리 통계적 인접성과 공간적 인접성을 동시에 고려하는 Jenks & Caspall(1971)의 BEI(Boundary Error Index) 지수를 최적화하는 과정을 말한다. BEI 지수란, 우선 이웃한 단위지역의 경계선(arc)를 경계로 두 단위 지역의 통계치 차이를 모두 계산한 후, 차이가 큰 arc가 실제로 단계구분도의 경계선으로 이용되는 정도를 나타내는 것이다. 예를 들어 i 개의 arc가 단계구분도의 경계선으로 이용되었다면, 차이가 큰 순서로 상위 i 개의 arc 값의 합에 대해 실제로 단계구분도 경계선으로 이용된 i개 arc 값의 합과의 비율을 계산한 것이다. 따라서 BEI 지수의 최대값은 1.0이며, 이 경우 가장 이상적으로 단위지역간의 중요한 경계들이 단계구분도의 경계로 모두 반영된 경우를 말한다.

그림을 통해 설명하면 다음과 같다. 즉, 그림 2-가는 6개 단위지역으로 나누어진 가상적인 자료모델을 나타내고 있고, 인접한 단위지역 간의 경계선

손 일

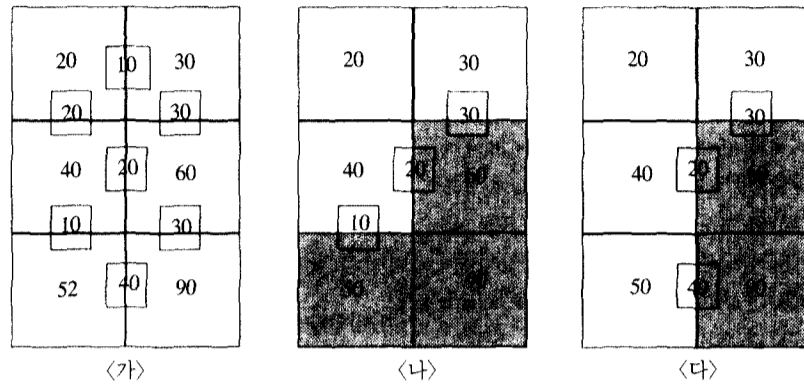


그림 2. BEI 지수를 설명하기 위해 6개의 단위지역으로 이루어진 가상 자료모델(가)로부터 작성된 2종류의 단계구분도(나와 다).

에는 두 단위지역의 통계치 차이를 보여주고 있다. 그림 2-나와 그림 2-다는 각기 2개 지도지역으로 나뉘어진 단계구분도이다. 그림 2-가에서 단위지역 간의 차이를 내림차순으로 정리해보면 40, 30, 30, 20, 20, 10이다. 그림 2-나의 경우 지도지역의 경계로 사용된 arc의 수는 3개이고, arc를 경계로 단위지역 간의 통계치 차이의 합은  $60(30+20+10)$ 이다. 따라서 단위지역 통계치 간의 차이 중 상위 3번째까지의 합이  $100(40+30+30)$ 이므로 BEI 지수는 0.6이 된다. 한편 그림 2-다의 경우 지도지역의 경계로 사용된 arc의 수는 3개이고 arc를 경계로 단위지역 간의 통계치 차이의 합은  $90(30+20+40)$ 이기 때문에 BEI 지수는 0.9가 된다. 따라서 그림 2-다의 경우 그림 2-나에 비해 주요한 경계들(단위지역간 차이가 큰 경계들)이 지도지역의 경계로 반영된 비율이 높아, 공간적 인접성이 보다 제대로 반영된 단계구분도로 인정될 수 있다.

본 연구에서 개발된 최적화 프로그램은, 주어진 면단위의 수와 계급 수에서 만들어질 수 있는 모든 단계구분도에 대해 BEI 지수의 계산을 계속 반복하면서(iteration) 최대 BEI 지수를 찾아내고, 이때 최대 BEI 지수를 생성하는 계급구간을 선정하는 것이다. Lindberg(1990)의 경우와 같이 TEI 지수의 최적화는 Fisher(1958)의 알고리즘을 이용해 프로그램화가 가능하지만, BEI 지수의 경우 면단위 간의 공간적 관계를 고려해야 하기 때문에 Fisher의 알고리즘을 일반적으로 적용할 수 없어 앞에서 지적인 전수계산 알고리즘을 적용하였다.

## 4. 실험

### 1) 실험 방법 및 자료

실험에 이용되는 지도는 수도권 53개 시·군에 대한 1995년도 인구밀도를 나타낸 단계구분도이다. 계급수는 3개에서 7개까지 모두 5가지의 계급수를 설정하였는데, 이에 대한 논의는 정인철(1998)의 연구를 참조하길 바란다. 다음으로 계급구간 설정 방법은 우선 일반적으로 많이 사용하고 있는 natural break, equal interval, n-tile을 이용했으며, 통계적 인접성을 기준으로 TEI 지수가 최대가 되는 계급구간을 설정하였다. 마지막으로 공간적 인접성을 기준으로 BEI 지수가 최대가 되는 계급구간을 설정하였다. 이러한 방식으로 설정된 총 5가지 계급구간 설정 방법을 3~7개 계급 수에 적용시켜 단계구분도를 작성하였다. 결국 이러한 과정에서 만들어진 총 25개의 단계구분도를 실험 자료로 이용하였다.

앞서 밝힌 전수계산 알고리즘을 이용해 만들어진 프로그램에서는 maximum TEI, maximum BEI, maximum OEI를 추적할 수 있을 뿐만 아니라, 특정 계급구간을 입력할 경우 해당되는 TEI, BEI, OEI 지수를 생성할 수 있는 option을 부여하였다. 또한 전수계산 알고리즘을 이용하여 최소 지도지역이 나타나는 계급구간을 추적하기 위한 최적화 프로그램을 작성하였다. 하지만 최소 지도지역이 나타나는 계급구간을 추적하는 최적화 프로그램은 통계적 인접성을 전적으로 무시하기 때문에 본 연

구의 계급구간 분류방법에서는 제외하였다. 한편 최소 지도계역을 추적하는 최적화 프로그램에도 특정 계급구간을 입력할 경우 해당되는  $C_F$ ,  $C_V$ ,  $C_L$ 를 생성할 수 있는 option을 부여하였다.

## 2) 평가기준의 설정

단계구분도의 오차를 측정하는 기준으로는 Jenks & Caspall(1971)의 OEI, TEI, BEI와 MacEachren(1982)의  $C_F$ ,  $C_V$ ,  $C_L$  이외에도 여러 다른 기준들이 있으나, 나머지는 이상의 6개 기준과 대동소이하므로 평가 기준을 6개로 한정하였다. 그러나 OEI와 TEI 둘 모두는 집단내 등질성과 집단간의 이질성을 추구하는 통계적 기준이지만, OEI는 단위지역의 면적을 가중치로 산정한다는 점을 제외하고는 TEI 수와 개념적으로 동일한 것이다. 따라서 본 연구에서 OEI 수치는 평가기준에서 제외하였다. BEI는 이웃한 면단위간의 통계치 차이가 큰 경계가 지도지역을 구획하는 경계로 사용되는 정도를 나타내는 지수이며,  $C_F$ ,  $C_V$ ,  $C_L$ 는 지도 패턴의 단순성을 측정하는 지수이다. 따라서 5가지 계급수 각각에 대한 5가지 계급구간, 다시 말해 총 25개의 단계구분도에 대해 앞서 이야기한 5가지의 평가기준을 적용하였다. 본 연구에서 설정한 5가지 평가기준을 보다 자세히 설명하면 다음과 같다.

- TEI(Tabular Error Index) :

$$\frac{\text{각 계급내 분산의 합}}{\text{전체 분산}} = \frac{\sum_{j=1}^m \sum_{k=1}^{n_j} (Z_{kj} - \bar{Z}_j)^2}{\sum_{i=1}^n (Z_i - \bar{Z})^2}$$

- BEI(Boundary Error Index) =  $\frac{\sum_{j=1}^n ARC_j}{\sum_{i=1}^n ARC_i}$

- $n$  : map region 구분에 사용된 inner arc의 수
- $ARC_j$  : map region 구분에 사용된 j번째 arc를 경계로 이웃한 면단위의 통계치 간의 차이.
- $ARC_i$  : 인접한 면단위 통계치 간의 차이 중에서 상위 n번째까지의 합

- $C_F = \frac{\text{분류 후 지도지역의 수}}{\text{분류 전 지도지역의 수}}$

- $C_E = \frac{\text{분류 후 변의 수}}{\text{분류 전 변의 수}}$

- 이 경우 외곽의 변까지 포함

- $C_V = \frac{\text{분류 후 꼭지점의 수}}{\text{분류 전 꼭지점의 수}}$

- 이 경우 외곽 변과의 꼭지점도 포함

## 3) 실험 결과 및 평가

단계구분도의 평가기준으로 설정된 5가지 평가 지수중 우선 TEI 지수는 계급구간내 등질성과 계급간의 이질성을 반영하는 지수, 다시 말해 통계적 인접성을 나타내는 지수이다. 한편 BEI는 주요 면 단위 경계의 반영정도를 나타내는 변수로 공간적 인접성을 반영하는 변수이며, 나머지  $C_F$ ,  $C_E$ ,  $C_V$ 는 지도 패턴의 단순성을 나타내는 평가지수이다. 각기 성격이 다른 평가지수에 대한 결과를 종합적으로 판단하기가 어려워 평가의 용이성을 위해 각각의 평가지수에 대한 순위를 부여하였다.

우선 표 2의 이해를 돕기 위해 예를 들어 설명하면, 자료 첫 행의 0.8785, 0.7856, 0.5136, 0.8906, 0.6459는 계급수를 3으로 하고 5가지 분류방법(natural break, equal interval, n-tile, maximum TEI, maximum BEI)에 따라 계급구간을 구분할 경우 각각의 TEI 지수를 나타낸 것이다. 앞 절에서 설명한 바와 같이 TEI 지수와 BEI 지수는 그 값이 클수록, 반대로  $C_F$ ,  $C_E$ ,  $C_V$ 는 그 값이 작을수록 계급구간이 효과적으로 설정되었음을 의미한다. 따라서 첫 행의 괄호안 수치는 TEI 지수를 기준으로 각각의 계급구간 설정 방법의 결과를 순위로 나타낸 것이다.

표 2의 분석 결과를 토대로 순위를 종합해 보면 표 2 최하단의 결과와 같이 maximum TEI, maximum BEI, natural break, equal interval, n-tile 순으로 높은 점수를 얻었다. 물론 이러한 결과는 maximum TEI와 maximum BEI의 경우 2가지 평가지수(TEI, BEI) 가운데 하나는 항상 1위를 한다는 점을 반영하고는 있으나, 나머지 평가지수에서도 비교적 높은 점수를 얻고 있다. 따라서 지도제작자의 목적에 따라 통계적 인접성을 최적화하거나 공간적 인접성을 최적화하여도 다른 분류방법에 비해 비교적 안정되고 합리적인 계급구간 분류

손 일

결과를 얻을 수 있다고 판단된다.

한편 표 2의 분석 결과에 따르면 패턴의 단순성을 나타내는 3가지 지수( $C_F$ ,  $C_E$ ,  $C_V$ )에 대한 순위가 모든 계급수에서 각각의 급간설정 방법에 대해 그 결과가 동일하였다. 따라서  $C_E$ 와  $C_V$  제외한  $C_F$ 만을 패턴의 단순성을 나타내는 평가지수로 채택하여도 큰 문제가 없는 것으로 판단된다. 이상의 결과를 바탕으로 비교적 순위에 따른 평가 점수가 낮은 equal interval과 n-tile을 제외한 나머지 3가지 분류방법(maximum TEI, maximum BEI, natural break)에 대해 3가지 평가기준(TEI, BEI,  $C_F$ )을 적용한 순위 결과를 도표로 나타냈다(그림 3). 이 그림에서 확인할 수 있는 가장 특징적인 사실은 통계적 인접성(maximum TEI)에 근거한 분류와 공간적 인접성(maximum BEI)에 근거한 분류가 상

반된 순위 결과를 보여준다는 사실이다. 이는 Monmonier(1968)의 지적처럼 두 가지 기준에 따른 최적화의 분류결과가 서로 trade-off의 관계, 즉 통계적 인접성을 강조할 경우 공간적 인접성이, 공간적 인접성을 강조할 경우 통계적 인접성이 훼손될 수 있음을 나타내고 있다.

그림 3에서 또 다른 특징은 natural break를 기준으로 한 분류의 경우 비교적 안정된 순위 결과를 얻고 있다는 점이다. 이를 확인하기 위한 한 가지 방법으로 그림 4와 같은  $\alpha$ -index라는 새로운 지수를 상정하였는데, 이 지수는 3가지 평가지수를 벡터적으로 합성하여 얻은 지수이다. 물론  $\alpha$ -index를 얻기 위해 이용된 3가지 평가지수 이외에도 여러 가지 평가지수가 있을 수 있으며 이들 지수간의 상관관계 역시 인정할 수 밖에 없다. 그러나 선정

표 2. 5가지 평가기준에 대한 25개의 단계구분도(5가지 계급, 5가지 계급구간)의 평가 결과

class index	grouping	natural break	equal interval	n-tile	maximum TEI	maximum BEI
3	TEI	0.8785(2)	0.7856(3)	0.5136(5)	0.8906(1)	0.6459(4)
	BEI	0.9539(2)	0.9206(4)	0.7492(5)	0.9391(3)	1.0000(1)
	$C_F$	0.1754(3)	0.1053(2)	0.1930(4)	0.1930(4)	0.0526(1)
	$C_E$	0.3416(3)	0.2360(2)	0.5776(5)	0.3789(4)	0.1988(1)
	$C_V$	0.1238(3)	0.0762(2)	0.1524(5)	0.1429(4)	0.0381(1)
4	TEI	0.8992(3)	0.9030(2)	0.5907(5)	0.9406(1)	0.6585(4)
	BEI	0.9647(2)	0.9416(4)	0.8105(5)	0.9421(3)	0.9790(1)
	$C_F$	0.1930(2)	0.2105(3)	0.3860(5)	0.2982(4)	0.0720(1)
	$C_E$	0.3478(2)	0.3665(3)	0.7205(5)	0.6087(4)	0.2112(1)
	$C_V$	0.1429(2)	0.1619(3)	0.3619(5)	0.2571(4)	0.0571(1)
5	TEI	0.9505(3)	0.9277(4)	0.6647(5)	0.9646(1)	0.9574(2)
	BEI	0.9546(3)	0.9217(4)	0.8364(5)	0.9597(2)	0.9810(1)
	$C_F$	0.2632(2)	0.2105(1)	0.4211(5)	0.3509(3)	0.3509(3)
	$C_E$	0.4410(2)	0.3975(1)	0.7329(5)	0.6398(3)	0.6708(4)
	$C_V$	0.2381(2)	0.1905(1)	0.4000(5)	0.3143(3)	0.3238(4)
6	TEI	0.9748(2)	0.9335(4)	0.6997(5)	0.9811(1)	0.9659(3)
	BEI	0.9630(3)	0.9569(4)	0.8753(5)	0.9852(2)	0.9931(1)
	$C_F$	0.3684(3)	0.1930(1)	0.4561(5)	0.3509(2)	0.3860(4)
	$C_E$	0.6522(3)	0.3975(1)	0.7516(5)	0.6460(2)	0.6894(4)
	$C_V$	0.3333(3)	0.1714(1)	0.4381(5)	0.3048(2)	0.3619(4)
7	TEI	0.9853(2)	0.9653(4)	0.7181(5)	0.9872(1)	0.9744(3)
	BEI	0.9817(3)	0.9482(4)	0.8719(5)	0.9852(2)	0.9955(1)
	$C_F$	0.4561(3)	0.2632(1)	0.5439(5)	0.3509(2)	0.5088(4)
	$C_E$	0.7516(3)	0.4907(1)	0.8075(5)	0.6460(2)	0.7826(4)
	$C_V$	0.4381(3)	0.0952(1)	0.5238(5)	0.3048(2)	0.4857(4)
순위 합계 (TEI+BEI+ $C_F$ )		38(3)	45(4)	74(6)	32(1)	34(2)

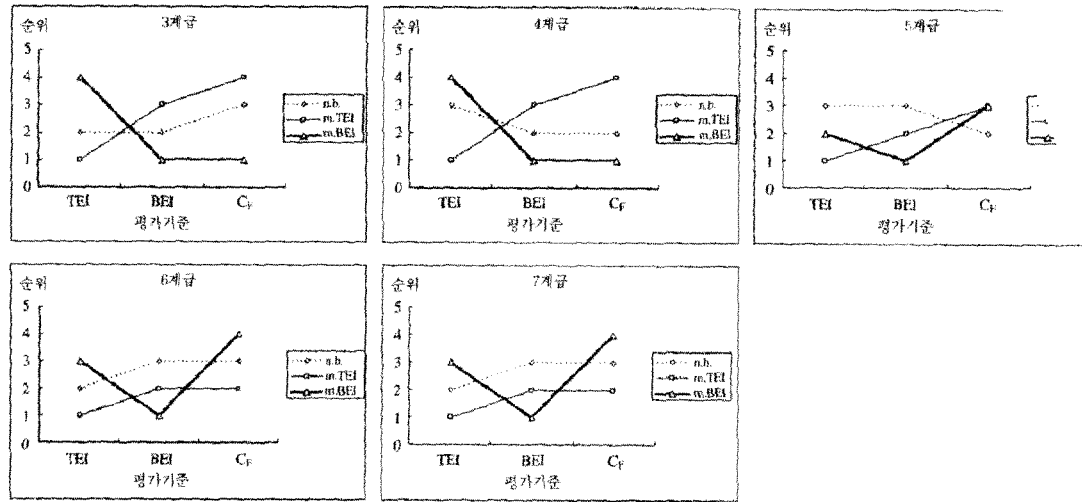
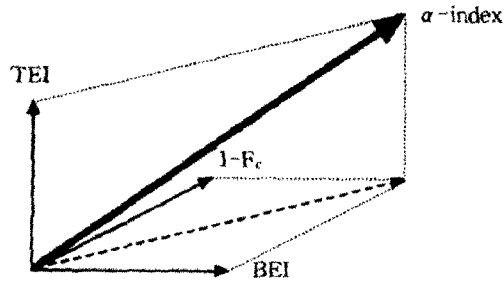


그림 3. 계급수별 3가지 분류방법에 대한 3가지 평가기준을 적용한 순위 결과



$$\alpha = \sqrt{TEI^2 + BEI^2 + (1 - F_c)^2}$$

그림 4. 3가지 평가지수를 벡터적으로 합성한  $\alpha$ -index에 대한 개념 모형

된 3가지 평가지수는 공간 패턴의 확인이라는 단계구분도의 작성 목적에 가장 잘 부합하는 평가지수라고 판단되며, 기존의 연구(Jerks and Caspell, 1971; MacEachren, 1982; Slocum, 1999)에서도 이들 지수의 효용성에 대해 인정한 바 있다. 하지만 TEI와 BEI 지수는 수치가 높을수록 효율적인 분류 결과를 나타내나  $C_F$ 는 그 반대여서, 지수의 특성에 따라  $C_F$  대신  $1-C_F$ 를  $\alpha$ -index 계산에 이용하였다. natural break, maximum TEI, maximum BEI 3가지 분류방법에 대해 각 계급수별로 구한  $\alpha$ -index의 값은 표 3과 같다.

앞에서 지적한 바와 같이 통계적 인접성(m. TEI)과 공간적 인접성(m. BEI)과의 trade-off 관계

표 3. 계급수별 3가지 분류방법에 대한  $\alpha$ -index

계급수	구분방법	n. b.	m. TEI	m. BEI
3		1.5368	1.5252	1.5214
4		1.5461	1.5049	1.5011
5		1.5354	1.5076	1.5167
6		1.5088	1.5344	1.5153
7		1.4934	1.5383	1.4771

로 인해, natural break를 기준으로 분류한 경우 비교적 효율적인 분류 결과를 얻을 수 있었다. 한편 계급수가 적은 경우 natural break를 이용한 분류의  $\alpha$ -index이 높았으며, 계급수가 많아짐에 따라 m. TEI를 기준으로 한 분류의  $\alpha$ -index가 높았다. 이는 계급수 증가에 따라 m. TEI에 비해 natural break와 m. BEI의  $C_F$  지수가 급격하게 상승한 데 그 원인이 있는 것으로 판단되나 이들 변수간의 상관관계에 대해서는 보다 자세한 연구가 필요하리라 여겨진다.

### 5. 결론 및 제언

단계구분도를 구축하는 데 있어서 가장 기본이 되는 요소는 단위지역의 크기, 모양 및 수, 계급의 수, 계급구간의 분류방법이다. 본 연구에서는 주어진 단위지역의 수와 계급수로부터 만들어질 수 있



는 모든 경우의 수를 계산할 수 있는 수식을 개발해 보았다. 실제로 만들어질 경우의 수는 무수히 많아 자료의 특성을 가장 잘 나타내는 지도를 선정하기란 어쩌면 불가능한 일이며, 이런 과정에서 지도제작자의 주관이 개입될 여지는 아주 많다.

한편 본 연구에서는 기존의 통계적 인접성을 최적화하는 기법과는 달리 전수계산 알고리즘을 이용하여 공간적 인접성을 최적화하는 프로그램을 작성하였다. 여기서 최적화의 대상 평가지수는 BEI로, 이웃한 면단위간의 통계치 차이가 큰 경계가 지도지역을 구획하는 경계로 사용되는 정도를 나타내는 지수이다. 5가지 계급수(3~7개)별로 5가지 분류방법(maximum TEI, maximum BEI, equal interval, n-tile, natural break)으로 만들어진 25개의 단계구분도에 대해, 5가지 평가기준(TEI, BEI, C<sub>F</sub>, C<sub>E</sub>, C<sub>V</sub>)을 이용해 계급구간 분류 결과를 분석하였으며 그 결과는 다음과 같다.

첫째, maximum TEI와 maximum BEI의 경우 비교적 높은 점수를 얻고 있어, 지도제작자의 목적에 따라 통계적 인접성을 최적화하거나 공간적 인접성을 최적화하여도 다른 분류방법에 비해 비교적 안정되고 합리적인 계급구간 분류결과를 얻을 수 있었다. 둘째, 패턴의 단순성을 나타내는 3가지 지수(C<sub>F</sub>, C<sub>E</sub>, C<sub>V</sub>)에 대한 순위가 모든 계급수에서 각각의 계급설정 방법에 대해 그 결과가 동일하여, C<sub>F</sub>만을 패턴의 단순성을 나타내는 평가지수로 채택하여도 큰 문제가 없는 것으로 판단된다. 셋째, 통계적 인접성(maximum TEI)에 근거한 분류와 공간적 인접성(maximum BEI)에 근거한 분류가 상반된 순위 결과를 보여준다는 사실이다. 이는 두 가지 기준에 따른 최적화의 분류결과가 서로 trade-off의 관계, 즉 통계적 인접성을 강조할 경우 공간적 인접성이, 공간적 인접성을 강조할 경우 통계적 인접성이 훼손될 수 있음을 나타내고 있다. 넷째, natural break를 이용한 계급구간 분류의 경우, 순위 결과에서도 확인될 수 있었지만  $\alpha$ -index를 이용한 분석 결과 역시 비교적 안정되고 효율적인 분류 결과를 얻을 수 있었다.

계급구간을 분류함에 있어 지도제작자의 의도에 따라 통계적 인접성 및 공간적 인접성을 최적화하는 기법 모두를 사용할 수 있다. 하지만 이 경우 우수리가 남는 계급경계를 택할 가능성이 많아 체

계적이며 판독에 용이한 계급구간을 설정하는 데 어려움이 따른다. 이는 natural break의 경우도 마찬가지이다. 따라서 시간과 비용이라는 제한적 요소를 감안한 현실적인 계급구간 설정방법을 제시하면 다음과 같다. 즉, natural break 혹은 최적화 결과를 기준으로 우수리가 없는 몇 가지 대안적 계급구간 분류 안을 마련하고, 이들에 대해 BEI, TEI, C<sub>F</sub> 지수들을 검토하여 최종적인 계급구간을 설정하는 것이다. 이렇게 얻어진 결과 역시 수없이 많은 단계구분도 중의 하나이므로, 지도제작자는 자료의 의미를 가장 잘 표현할 수 있는 지도의 제작에 최선을 다해야 할 것이다.

## 文獻

- 손 일, 1998, "커뮤니케이션 이론에 대한 대안과 지리적 시각화," 한국지역지리학회, 4(1), 27-41.
- 손 일, 1999, "단계구분도의 계급구간 설정에 관해," 대한지리학회 춘계학술대회 요약집, 7-10.
- 이희연, 1995, 지도학: 주제도 제작의 원리와 기법, 범문사.
- 정인철, 1998, "단계구분도 제작을 위한 최적 분류 방법에 대한 소고," 부산지리, 7, 1-12.
- Evans, I. S., 1977, The selection of class interval, *Transactions, Institute of British Geographers (new series)*, 2(1), 98-124.
- Fisher, W. D., 1958, On grouping for maximum homogeneity, *Jour Am. Stat. Assoc.*, 53(284), 789-798.
- Jenks, G. F. and Caspall, F. C., 1971, Error on choroplethic maps: Definition, measurement, reduction, *AAAG*, 61(2), 217-244.
- Jenks, G. F. and Knos, 1961, The use of shaded pattern in graded series, *AAAG*, 51(3), 316-34.
- Jenks, G. F., 1977, Optimal data classification for choropleth maps, *Occasional Paper*, No.2, Dept. of Geography, University of Kansas, 24.
- Johnston, R. J., 1968, Choice in classification: The subjectivity of objective methods, *AAAG*,

- 58(3), 575-589.
- Lindberg, M. B., 1990, Fisher: A turbo pacal unit optimal partitions, *Computer & Geosciences*, 16(5), 717-732.
- MacEachren, A. M., 1982, Map complexity: comparison and measurement, *The American Cartographer*, 9(1), 31-46.
- Monmonier, M. S., 1972, Continuity-biased class-interval selection: A method for simplifying pattern on statistical maps, *Geographical Review*, 62(3), 203-228.
- Monmonier, M. S., 1980, The hopeless pursuit of purification in cartographic communication: A comparison of graphic-arts and perceptual distortions of graytone symbols, *Cartographica*, 17(1), 24-39.
- Robinson, A. H., 1982, *Early Thematic Mapping in the History of Cartography*, The University of Chicago Press.
- Robinson, A. H. et al., 1995, *Elements of Cartography*, John Wiley & Sons.
- Slocum, T. A., 1999, *Thematic Cartography and Visualization*, 293, Prentice Hall, New Jersey.

(2001년 9월 4일 접수)