

CHAID 技法에 의한 都市機能의 試論的 研究

梁 順 貞*

地理學에서는 地域의 特性을 규명하고자 수많은 計量的 分析手法을 사용하여 왔다. 본 고에서는 일종의 判別分析技法으로 최근에 도입된 CHAID 技法을 사용하여 都市와 都市機能에 관한 통계처리를 시도하였다. 2종류의 자료를 가지고 두 차례 처리를 실시하였는데, 하나는 인구 25만명 이상의 도시 20개를 예측변수로 하고, 行政, 市場, 金融機能 그리고 生産機能을 반응변수로 하여 도시의 기능을 분류해 내었다. 두번째 처리에서는 앞서 언급한 행정, 시장, 금융, 생산기능 이외에 交通, 教育, 醫療, 文化, 그리고 運送機能의 9가지를 예측변수로 선정하고, 首都圈, 釜山圈, 大邱圈, 光州圈, 忠淸圈의 5개 圈域을 반응변수로 하여 각 권역에서 탁월한 기능을 판별·분류해 내었다. 이상에서 CHAID 기법은 큰 양의 범주형 자료를 처리할 수 있고, 樹形圖로 결과를 산출하여 해석이 용이하므로 地域을 分類하거나 特性을 判別하는데 유용한 또 하나의 새로운 분석틀로 여겨진다.

主要語 : CHAID技法, 都市機能, 判別分析, 範疇形資料, 樹形圖

1. 연구목적 및 연구방법

地理學에서 地域研究에 사용되어 온 計量技法은 그 연구대상에 따라 다양하다. 계량기법은 지리학의 諸연구에 보다 이론적·법칙적인 근거를 마련하기 위해 연구목적과 대상에 맞게 응용될 수 있다. 본 연구에서도 이러한 취지로, 새로운 판별분석기법인 CHAID기법을 지리학에서 어떻게 이용할 수 있는가에 관해 살펴보기로 하겠다. 이를 위해 하나의 사례를 통해서 CHAID를 풀어 나가려 한다.

본 연구에서 다루려는 사례는 도시의 기능분류 및 권역별 기능판별이다. CHAID 판별·분류기법을 사용하여 우리나라 인구 25만 이상의 市級都市 20개를 대상으로 도시의 기능분류를 실시하였고, 또한 특별시와 직할시를 기준으로 한 5개의 권역인 수도권, 부산권, 대구권, 충청권, 광주권에 대해 어느 기능이 탁월한가를 살펴보았다.

우선 첫번째 연구대상도시인 20개의 도시를 예측변수로 선정하고 몇가지 도시기능을 반응변

수로 정하였다. 여기서 반응변수인 도시기능은 기존의 도시분류가 그러하듯이, 주로 경제기능에 근거하여 구체화되는 경우가 대부분이므로(남영우, 1992), 경제기능을 중심으로 하여 행정기능, 시장기능, 금융기능, 생산기능을 설명해 줄 수 있다고 생각되는 지표로 선정하였다. 즉, 해당 業體數(establishments)와 從業者數를 공통의 지표로 선정하였는데 이는 도시의 中心地機能을 설명할 때 항시 지표로 삼는 것이기 때문이다. 그리고, 그 이외에 해당기능을 설명해 줄 수 있다고 판단되는 몇 가지를 추가하였다.

두번째 사례의 경우에는 첫번째 사례와 달리 도시기능을 예측변수로 선정하고, 연구 대상지역인 5개 권역을 반응변수로 채택하였다. 그 기능으로는 앞서 서술한 4가지 기능에 교육기능, 문화기능, 의료기능, 교통기능, 운송기능이 더 포함된다. 또한 업체수와 종사자수를 공통적인 모든 기능의 지표로 삼았고 업체수와 종사자수를 파악하기 어려운 부분에 대해서는 나름대로 설명력을 가진 지표들을 임의로 선정하였다. 이 과정에서 교통기능은 주로 인구이동의 측면으로,

* 고려대학교 대학원 지리학과

운송기능은 주로 물자이동의 측면으로 구분하여 다루었다.

이상과 같이 각 기능을 설명해 줄 수 있다고 생각되는 지표를 채택하기는 하였으나, 그 지표의 설명력이 실제로 어느 정도인지는 밝혀내지 못하였다. 그리고 각 기능을 매우 잘 나타내 주는 지표가 있다고 하더라도, 그 지표에 해당하는 통계자료의 입수 불가능이라는 문제를 가지고 있기 때문에, 지표 선정에 어느 정도 한계가 있었다.

다음으로, 선정된 변수들은 범주화되었다. 변수를 범주화한 것은 CHAID 기법이 범주형 자료를 처리하는 기법이므로 이에 적절하게 그 입력 형태를 조작하기 위한 것이다. 우선 첫번째 사례의 예측변수인 도시에 있어서, 각각의 지표에 대한 특정한 기준이 있는 것이 아니므로, 일반적 범주화(grouping) 방법인 四分位(quartile)를 이용하였다. 즉, 각각의 지표에 대해 25%, 50%, 75%에 해당하는 값을 산출한 후에 4개의 범주로 구분하였다. 그리고 반응변수인 도시기능의 범주화는, 각 지표마다 단위와 기준이 다르므로 각각의 상이한 지표를 범주화하였는데, 이로 인해 각 기능별 범주의 수는 다르게 나타난다.

두번째 사례의 예측변수인 도시기능의 경우에는 각 지표의 평균(X)과 표준편차(SD)를 구한 후, ① $X \sim X + SD$ ② $X + SD \sim X + 2SD$ ③ $X + 2SD \sim X + 3SD$ ④ $X + 3SD \sim$ 의 4개 범주로 나누고, 종속변수인 연구지역은 각 권역에 해당하는 市級都市를 1로 邑級都市를 0으로 각각 범주화하였는데, 이는 시급도시에 도시기능이 강할 것으로 예상되기 때문이다. 이상의 자료를 가지고 SPSS PC+CHAID를 통해 통계처리를 실행하였다.

2. CHAID 기법의 특징

CHAID(Chi-Squared Automatic Interaction Detection)는 AID(Automatic Interaction Detection)라고 하는 판별분석기법을 G.V. Kass가 개발해 낸 알고리즘이다(Kass, 1980). AID는 반

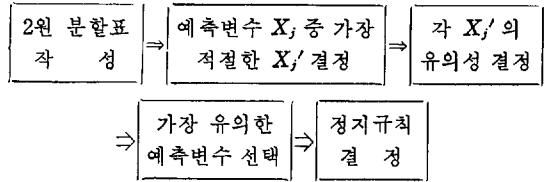


그림 1. CHAID 알고리즘의 주요단계

응변수와 예측변수간의 상관관계를 알아보기 위해 사용되는 기법으로써 원 자료집합의 연속적인 이진분리(binary split)로 보다 동질적인 그룹으로 분리하는 것이다(서혜선, 1991). 이러한 기법을 Kass가 범주형 자료에 적합하도록 발전시킨 것이 CHAID이다. 즉, 반응변수에 관련된 예측변수들을 순서대로 찾아내고, 이들 변수들이 결과에 얼마만큼의 영향을 주는가를 알아내는데 유용하다(허명희 등, 1993). CHAID 기법은 반응변수를 잘 설명하도록 상호 독립적인 부분집단(Sub-groups)으로 자료를 분리하는 단계적 절차이다(서혜선, 1991). Kass의 CHAID 알고리즘의 주요단계를 도식화하면 그림 1과 같다.

이 기법을 적용한 논문으로는 徐惠善(1991)의 것이 있고, 韓國分類學會 主題發表(1993)가 있을 뿐이다. 전자는 당뇨병 임상자료를 사용, 당뇨병의 합병증 유발요인에 대한 판별분석을 시도한 사례연구로, 당뇨병의 합병증 중에서 6가지를 반응변수로 선정하였고, 이에 영향을 줄 수 있는 11개 변수를 예측변수로 선택하였다. 이에 의하여 11개 예측변수 중에서 어느 변수가 반응변수인 합병증에 영향을 주는가를 판별해 내었다. 후자는 자동차 충돌사고 자료를 이용하여, 그 부상정도에 관련된 변수들을 판별 및 분류한 것이다. 충돌사고에서의 부상정도를 반응변수로 하고, 차체중량, 운행속도, 운전자의 나이, 자동차의 모델년도 등 4개를 예측변수로 선정하여, 이들 예측변수가 반응변수인 부상정도에 얼마만큼의 영향을 주는가를 살펴보고, 또한 이에 관련된 변수들을 순서대로 판별하였다.

이상의 적용된 사례에서 보듯이 CHAID는 명목형 자료를 범주화한 것으로 충분히 결과를 도출해 낼 수 있는 기법이므로 유용하다고 할 수 있다. 그런데 범주의 형태를 Kass는 3가지로 분류하였다. 하나는, 명목형 변수로, 모든 범주들

의 자유로운 병합이 허용되는 'free'형, 둘째는 범주수준이 순서형인 변수로 인접 범주들의 병합(merge)만 허용되는 'monotonic'형, 그리고 셋째는 위 두가지 유형의 중간형태로 변수의 속성이 기본적으로는 순서형이지만 마지막 범주가 결측치이기 때문에 어느 범주와도 병합이 허용되는 'floating'형이다(허명희 등, 1993). 명목형 변수를 사용하는 경우에는 정보의 손실이 없으나 순서형 변수를 사용하는 경우에는 범주화하는 과정에서 정보의 손실이 발생한다. 따라서, 순서형 변수를 택할 경우, 보다 많은 양의 자료를 사용한다면 정보의 손실을 어느 정도 극복해 낼 수 있다.

3. CHAID 기법의 지리학적 응용

도시지리학에서 주로 도시기능을 분류해 온 계량수법은 고전적인 방법인 H.J. Nelson의 도시기능 분류방법이었다(Nelson, 1955). 판별분석을 적용한 사례도 매우 적어, Q. Ahmad의 인도의 도시분류(Ahmad, 1965), E. Casetti의 이탈리아 지역구분(Casetti, 1964), 그리고 K. Suzuki의 일본 농업지역과 공업지역 구분(Suzuki, 1968) 등이 있다. 국내에서 간행되고 있는 주요 학술잡지에 나타난 계량적 기법을 조사한 경우에도 판별분석은 그다지 널리 사용되고 있지 않음을 알 수 있다(남영우, 1992).

이런 시점에서 CHAID 기법은 기존 판별·분류기법과 또 다른 특성을 가지고 있어 유용하게 사용될 수 있으리라 본다. 특히 인문지리 분야에서 숫자로 명시할 수 없는 명목형자료를 일정기준에 의해 범주화시킴으로써 원하는 결과를 얻어낼 수 있는 장점이 있다. 또한 변수만 적절히 선정된다면 CHAID는 SPSS PC+에서 손쉽게 결과를 산출해 낼 수 있으므로 기존의 방식보다는 비교적 간단하게 처리할 수 있다.

4. 수형구조의 분류결과

1) 개별도시의 기능분류

앞서 언급한 자료를 가지고 CHAID를 이용하

여 분석을 실시하였다. 변수는 예측변수인 도시 20개와 반응변수인 도시기능 4개로 구성되었으며 반응변수인 행정기능, 시장기능, 금융기능, 생산기능에서 각각의 범주들을 병합한 판정기준은 $\alpha=5\%$ 유의수준으로 선택하였다. 표 1은 행정기능에서 모든 예측변수에 대해 X_j' 의 유의성을 결정하는 검정통계량의 값이 유의수준을 상회하여 형성된 병합 그룹들을 보여준다. '여기서 다루어진 변수의 유형은 전체가 모두 'monotonic'형에 해당한다. 이 표에서 서울은 범주화한 4그룹중에서 범주 1과 범주 2가 하나의 그룹으로 병합되고 범주 3은 단일그룹으로 구성되어, 결국 이 예측변수는 2개의 부분집단(sub-groups)으로 형성되는 것이다. 이 표에서는 전체가 2개의 부분집단으로 병합되었음을 볼 수 있다.

또한 표 2는 행정기능에 있어 다중나무 분류방법에 의해 병합된 이후의 카이제곱과 P-값을 보여주고 있다. 이것이 최종적으로 분류되어 나타난 결과가 그림 2에 해당한다. 이 그림을 보면 행정기능의 경우 20개 예측변수 각각에 대해 가장 적절한 판정기준으로부터 선택된 $X_j'(j=1, 2, \dots, 20)$ 중에 유의수준 5%에서 가장 유의한 변수로 대구를 선택하고 있다. 대구는 범주가 2개인 그룹으로 나뉘어지는데, 범주 1에는 원자료의 범주 1($0\% \leq A \leq 25\%$)이 해당되고, 범주 2에는 원자료의 범주 2($26\% \leq A \leq 50\%$)가 해당된다. 가장 유의한 변수 대구의 첫번째 범주에서, 26.87%는 대구의 첫번째 그룹이면서 행정기능 전체(범주 1, 2, 3, 4)에 걸친 총 100% 중 행정기능 범주 1에 대한 행의 백분비를 나타내 준다. 또한, 대구의 두번째 범주에서, 33.33%는 대구의 두번째 그룹이면서 행정기능 범주 1에 대한 행의 백분비이다. 행정기능 첫번째 그룹이 선택된 부분집합에서 독립적으로 재분석을 실시하였으나, 더 이상 유의한 변수가 존재하지 않아 하나의 분리를 마치게 된다. 이 작업은 5% 유의수준에서 더 이상 유의한 변수가 없을 때까지 계속 진행되어 분류작업을 끝낸다.

이상에서 보듯이 행정기능은 가장 유의한 변수로 대구만을 판별해 낼 수 있다. 나머지의 도시기능인 시장기능, 금융기능, 생산기능

표 3. 예측변수의 범주별 그룹화(수도권)

예 측 변 수	범주의 그룹화	부분집단의 수
공무원수(A1)	1, 2	4
관공서 및 주요기관수(A2)	1	4
시장의수(B1)	1, 2, 3, 4	2
시장 건물 연면적(B2)	1, 2, 3, 4	2
금융기관의 수(C1)	1, 2	4
예금액(C2)	1	4
대출액(C3)	1	4
제조업체수(D1)	1, 2	4
제조업 종사자수(D2)	1, 2	4
도로율(E1)	1, 2, 3, 4	2
인구 100인당 차량대수(E2)	1	4
중·고등학교수(F1)	1, 3	4
중·고등학생수(F2)	1, 3	4
대학(교)의 수(F3)	1, 2	4
대학(교) 학생수(F4)	1, 2, 4	1
의료 종사자수(G1)	1, 2	4
도서관수(H1)	1, 4	1
문화시설수(H2)	1, 2, 3	4
운송업체수(L1)	1, 3	4
운송업 종사자수(L2)	1, 2	4

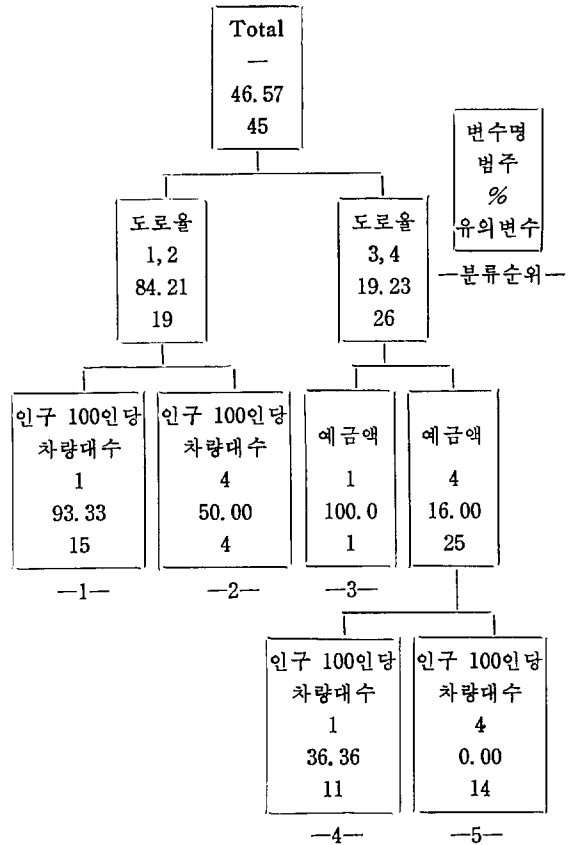


표 4. 수도권 수형구조 분류방법에 의한 결과표

예 측 변 수	병합 후의 χ^2	P-값
E1	18.6 (df=1)	4.8E-5
H2	15.3 (df=1)	0.0003
E2	12.6 (df=1)	0.0004
G1	12.9 (df=1)	0.0007
A1	12.9 (df=1)	0.0007
F1	12.9 (df=1)	0.0007
A2	11.1 (df=1)	0.0009
C2	11.1 (df=1)	0.0009
L2	11.1 (df=1)	0.002
C1	11.1 (df=1)	0.002
F2	11.1 (df=1)	0.002
C3	7.9 (df=1)	0.005
D1	8.8 (df=1)	0.006
B2	9.5 (df=1)	0.006
L1	6.4 (df=1)	0.02
B1	6.5 (df=1)	0.03
D2	5.0 (df=1)	0.05
F3	4.4 (df=1)	0.07
F4	0.0 (df=0)	1.0
H1	0.0 (df=0)	1.0

그림 3. 수도권의 樹形構造 분류결과

수도권은 가장 유의한 변수로 도로율을 선택해 내고 있으며 차례로 예금액과 인구 100인당 차량 대수를 판별해 내고 있다($\alpha=0.05$).

그룹 1에는 원자료의 범주 1($14\% \leq E1 \leq 26\%$)과 범주 2($27\% \leq E1 \leq 37\%$)가 포함되고, 그룹 2에는 원자료의 범주 3($38\% \leq E1 \leq 49\%$)과 범주 4($50\% \leq E1$)가 포함된다. 'monotonic' 변수는 인접하고 있는 범주와만 병합이 가능하기 때문에 이런 결과가 나타나게 되었다. 그리고 첫번째 사례에서와 마찬가지로 가장 유의한 변수인 도로율의 첫번째 그룹에서 84.21%는 도로율 첫번째 그룹이면서 수도권 전체(범주 1, 2, 3, 4)에 걸친 총 100% 중 수도권 그룹 1에 대한 행의 퍼센트를 나타내 준다. 또한 도로율의 두번째 그룹에서, 19.23%는 도로율 두번째 그룹이면서 수도권 그룹 1에 대한 행의 백분비이다. 도로율 첫번째 그룹이 선택된 부분집합에서 독립적으로 재분석을 실시하여 그 다음 유의한 변수로 인구

표 5. CHAID 분석을 통해 산출된 개별도시의 유의한 변수들

예측변수	행정기능	시장기능	금융기능	생산기능
서울		*		*
부산			*	*
대구	*			
인천				
광주				
대전				
수원				
성남				*
안양				
부천			*	
광명				
안산				
고양				
청주				
전주				
포항				
창원				
울산				
마산				
진주				

* 유의수준 $\alpha=0.05$ 를 기준으로 분류한 결과 유의한 것으로 선정된 변수들

100인당 차량대수(E2)를 선택해 내고 있다. 인구 100인당 차량대수는 2개의 그룹으로 분리되며, 도로율의 첫번째 그룹과 인구 100인당 차량대수의 첫번째 그룹으로 이루어지는 부분집합에서 독립적으로 재분석을 실시하게 된다. 유의한 변수가 더 이상 존재하지 않으면 하나의 분류작업을 마치게 된다.

도로율의 첫번째 부분집합에서의 작업이 끝나면 두번째 부분집합에서의 분류작업이 시작된다. 도로율의 두번째 그룹이 선택된 부분집합에서 독립적으로 재분석을 실시하여 다음의 유의한 변수로 예금액(C2)을 판별해 내었다. 예금액 또한 2개의 범주로 분리되며, 첫번째 범주는 5% 유의수준하에 더 이상 유의한 변수가 없으므로 분류를 끝마친다. 두번째 범주는 도로율의 두번째 그룹과 예금액과 두번째 그룹으로 이루어진 부분집합에서 독립적으로 재분석을 실행하여 다음으로 좋은 변수인 인구 100인당 차량대수를

표 6. CHAID 분석을 통해 산출된 권역별 유의한 변수들

예측변수	수도권	부산권	대구권	광주권	충청권
공무원수				*	
관공서 및 주요기관수		*			*
시장의 수					
건물연면적					
금융기관의 수					
예금액	*	*			
대출액					
제조업체수					
제조업 종사자수				*	
도로율	*				*
인구 100인당 차량대수	*		*		
중·고등학교의 수		*	*		*
중·고등학생수					
대학(교)의 수				*	
대학(교) 학생수					
의료 종사자수			*		
도서관수					
문화시설수					
운송업체수					
운송업 종사자수					

* 유의수준 $\alpha=0.05$ 를 기준으로 분류한 결과 유의한 것으로 선정된 변수들

선정한다. 이 작업은 5% 유의수준에서 더 이상 유의한 변수가 없을 때까지 계속 진행된다. 예금액은 두개 그룹으로 나누어지고 예금액의 두번째 그룹의 부분집합에서 첫번째 범주로 인구 100인당 차량대수 $1/2(11\% \leq E2 \leq 14\%)$ 과 $2/2(E2 \leq 22\%)$ 를 분리함으로써 모든 분류를 끝마친다.

결과적으로 수도권에서는 가장 유의한 변수로 도로율 그리고 인구 100인당 차량대수와 예금액을 판별해 낼 수 있으며, 이것은 교통기능과 금융기능으로 설명할 수 있다. 나머지 부산권, 대구권, 광주권, 충청권의 경우에도 이와 같은 절차로 유의한 변수를 판별해 내게 된다. 분류결과를 보면, 부산권은 관공서 및 주요기관수로 나타나는 행정기능, 중·고교수로 나타나는 교육기능, 그리고 예금액으로 나타나는 금융기능으로 분류되었다. 대구권의 경우 중·고교

수로 나타나는 교육기능과 의료 종사자수로 나타나는 의료기능 그리고 인구 100인당 차량대수로 나타나는 교통기능으로 분류되었다. 광주권의 경우는 공무원수, 제조업 종사자수 그리고 대학 및 대학교 수가 유의한 변수로 판정되어 행정기능, 생산기능 그리고 교육기능이 분류되었다. 끝으로 충청권에 있어서는 도로율, 관공서 및 주요기관수 그리고 중·고교수 등의 변수가 유의한 것으로 나타나, 교통기능, 행정기능, 교육기능의 3가지로 판별되었다.

이제까지 행한 CHAID 분석의 모든 수행결과 중에서 유의한 변수들만을 표로 작성해 보면 다음과 같다(표 5와 6을 참조).

5. 분류의 문제점 및 한계

우선 예측변수로 사용한 도시기능에서 그 지표의 설명력 내지는 기여도를 측정하지 않은 것에 문제가 있다. 즉 2~3개로 선정한 지표들이 그 기능을 어느 정도 설명해 줄 수 있는가 하는 것이다. 또한 선택한 변수가 실제로 더 있었으나 관측치의 수가 적어서 유의수준 5%에서는 유의한 변수를 판별해 내지 못하였다. 따라서 관측치의 자료를 충분히 입수하는 것이 필요하다. 그리고 ‘圈域’이라고 하는 지역구분 자체에도 다소의 문제점을 내포하고 있다. 즉 실제의 통계자료 자체는 행정구역 단위로 분류되어 있어 기능지역으로 구분되는 용어의 의미를 모호하게 만들고 있다.

CHAID 기법은 범주형 자료에 적절한 기법이므로 유용하기는 하나 지리학에서 적용하는 경우 변수선택에 있어 몇가지 점들을 고려해야 하며, 나름대로 지역의 의미에 적절하게 가공하여 사용할 필요가 있다.

앞서 수행한 분석에 있어서는 자료의 양이 그다지 크지 않으므로 반응변수의 양을 늘릴 경우, 그 결과는 앞에 산출한 결과와 달라질 수도 있을 것이다. 따라서 본 논문의 결과는 제시된 자료에 한하여 산출된 것이므로 확대해석 하기에 다소 무리가 있을 것으로 보인다.

6. 요약 및 결론

본 논문에서는 우리나라 인구 25만명 이상의 도시 20개를 선정하여 도시의 기능을 판별분석하고, 5개 지역에 대한 도시기능을 판별하기 위하여 Kass의 CHAID(Chi-Squared Automatic Interaction Detection) 기법을 사용하였다. 전자의 경우 예측변수로 우리나라 인구 25만 이상의 市級都市를 선정하였고, 반응변수로는 行政機能, 市場機能, 金融機能과 生産機能의 4가지로 정하여 각 기능에 따른 지표를 선정하였다. 후자의 경우 예측변수로 行政機能, 市場機能, 金融機能, 生産機能, 交通機能, 教育機能, 醫療機能, 文化機能 그리고 運送機能의 9가지로 정하고 각 기능에 따른 지표를 20개 선정하였다. 반응변수로는 수도권, 부산권, 대구권, 광주권, 충청권의 5개 권역으로 정하여 각 권역에 포함되는 市級都市와 邑級都市를 대상으로 분석을 실시하였다.

분석의 결과 행정기능은 대구만을 유의한 변수로 판별해 내었고, 시장기능 역시 서울만을 분류해 내었다. 금융기능은 부천과 부산을 판별·분류하였으며 생산기능은 서울과 부산 그리고 성남을 분류해 내었다. 한편, 수도권에서는 도로율·인구 100인당 차량대수로 나타나는 교통기능과 예금액으로 나타나는 금융기능으로 분류되었다. 부산권의 경우는 관공서 및 주요기관수로 나타나는 행정기능, 중·고등학교수로 나타나는 교육기능과 예금액으로 나타나는 금융기능으로 분류되었으며, 대구권의 경우 중·고등학교수로 나타나는 교육기능과 의료 종사자수로 나타나는 의료기능, 그리고 인구 100인당 차량대수로 나타나는 교통기능으로 분류되었다. 광주권의 경우는 공무원수, 제조업 종사자수 그리고 대학 및 대학교 수가 유의한 변수로 판정되어, 행정기능, 생산기능 그리고 교육기능이 분류되었다. 끝으로 충청권에 있어서는 도로율, 관공서 및 주요기관수 그리고 중·고등학교수 등의 변수가 유의한 것으로 나타나, 교통기능, 행정기능, 교육기능의 3가지로 판별해 내었다.

이상의 결과는 본 연구에 사용된 자료에 한정되며, 반응변수 자료의 양이 많아지게 될 경우 위의 결과와 달라질 가능성을 배제할 수 없다. 따라서 본 연구결과의 확대해석에는 다소의 부담이 따르며, CHAID 기법을 이용한 지역분류의 연구가 시도되어야 할 것이다.

다중나무 판별분석기법은 큰 양의 범주형 자료를 처리하는데 유용하며 의학계에서 사용되고 있는 신뢰성 있는 기법이다. 단지 지역분류의 경우 많은 양의 자료가 필요하다는 부담 때문에 너무 지나친 지역분류로 자료 입수가능성의 문제를 야기시켜, 적정 수준의 지역선정 및 통합을 필요한 과제로 남기고 있다. 이러한 부분만 보완된다면 지리학에서 지역을 분류하거나 특성을 판별해 내는데 유용한 기법이라고 생각된다.

본 연구는 필자의 고려대학교 대학원 석사학위논문 수정·요약한 것이다. 지도해 주신 남영우 교수님과 통계처리에 협조해 주신 고려대학교 박사과정의 서혜선씨께 감사드립니다.

(投稿 1994년 7월 30일)

文 獻

- 남영우, 1985, 都市構造論, 法文社.
 _____, 1992, 計量地理學, 法文社.

- _____, 1993, “分類技法의 地理學的 應用”, 師大論集, 18, 157-166.
 內務部, 1993, 韓國都市年鑑.
 徐惠善, 1991, CHAID를 利用한 判別分析—糖尿病 臨床資料의 事例分析—, 高麗大學校 大學院 碩士學位論文.
 崔在憲, 1987, 韓國의 都市體系에 관한 研究, —金融의 空間構造分析을 通하여—, 서울 大學校 教育大學院 碩士學位論文.
 韓國都市行政研究院, 1993, 全國統計年鑑 (上) (中) (下).
 허명희·서혜선·정성원, 1993, CHAID 判別/分類技法의 소개, 韓國分類學會.
 洪慶姬, 1982, “韓國都市의 機能分類: 1960~1978年の 變化를 中心으로,” 地理學, 26, 大韓地理學會, 1-14.
 Beaujeu-Garnier, J., G. Chabot, 1971, *Urban Geography*, Longman Group Ltd, London.
 Kass, G., 1980, An Exploratory Technique for Investigating Large Quantities of Categorical Data, *Applied Statistics*, 29(2), 119-127.
 Nelson, H.J. 1955, A Service classification of America cities, *Economic Geography*, 31, 189-210.

An Introductory Study on the Urban Functions Using CHAID technique

Soon-Jeong Yang*

Summary

To this day, a number of quantitative analytical methods have been employed in clarifying regional characteristics in the discipline of geography. This paper attempted, as a part of application of those quantitative analyses, to

make clear the urban functions and consequently the urban characteristics statistically by adopting newly-introduced CHAID, a sort of discriminant analysis technique. The processing of data was conducted in two phases. To begin with, the urban functions were classified after designating twenty cities—the population of each city counting 250,000 or more—as predictor variable,

* Korea University Graduate

and at the same time four major urban functions like administration, marketing, finance and production as response variable.

And then, preeminent functions of individual region were discriminated and concurrently classified by treating the remaining traffic, education, medicare, culture and transportation functions as predictor variable, and the following five regions as response variable: Metropolitan Seoul Area, Pusan region, Taegu region, Kwangju region and Chungcheong region.

According to the result of this analysis, marketing and administration are emerged as meaningful functions in Seoul and Taegu respectively. As for the finance function only Pusan and Pucheon can be discriminated. Seoul, Pusan and Seongnam reveal their dominancy in production function.

To take a look at the result of the latter analysis, the Metropolitan Seoul Area shows, among other functions, strong traffic and finance

functions. When it comes to Pusan region, administration, education and finance functions are recorded as a leading ones, and Taegu region is preferable in education, medicare and transportation functions. In case of Kwangju region administration, production and education functions are discriminated from any other functions. Chungcheong region shows similar aspect with only traffic function replacing the production function of Kwangju region.

Based on aforementioned analysis, it can be said that the CHAID technique, which is capable of processing large amount of categorical data and, by presenting its outcome in the form of dendrogram, facilitates the interpretation work, is an effective, meaningful means to classify and discriminate certain geographical regions and their characteristics.

Key Words: CHAID Technique, urban function, discriminant analysis, categorical data, dendrogram.