

Geodemographics의 연구기법을 활용한 서울시 지역유형 분석 연구

우현지* · 김영훈**

Analysis of Area Type Classification of Seoul Using Geodemographics Methods

Woo, Hyun Jee* · Kim, Young Hoon**

요약 : Geodemographics(GD)는 지리적 패턴 분석을 위해 사회경제적, 행태적 자료의 이용에 관한 연구 기법으로써 동일한 근린 혹은 이웃한 지역에 거주하는 사람들은 유사한 인구적, 행태적 특성을 보인다는 전제를 바탕으로 유사한 지역을 도출하고 그 지역들의 공간적 특징을 유형화하는 연구 방법이다. 따라서 본 연구는 이러한 GD 개념의 활용성과 지역 유형화의 유용성을 확인하기 위하여 서울시 행정동 및 관련 인구 센서스 자료를 바탕으로 서울시 지역 유형화에 대한 연구를 진행하였다. 본 연구는 Ward와 K-means 및 관련 통계 기법에 의한 센서스 변수들의 표준화와 군집화를 통해 총 13개의 사회경제적 특성을 공유하는 지역을 도출하고 이를 바탕으로 서울시 13개 각각의 지역에 대한 지리적 성격을 제시하였다. 마지막으로 본 연구에서는 GD의 분석기법이 유사한 공간적 성격을 갖는 새로운 공간 단위 및 군집을 탐색하고 유형화하는데 효과적인 공간기법이 될 수 있음을 제시하고자 하였다.

주요어 : Geodemographics, 지역유형분류, Ward 방법, K-means 군집분석, Penetration Index

Abstract : Geodemographics(GD) can be defined as an analytical approach of socio-economic and behavioral data about people to investigate geographical patterns. GD is based on the assumptions that demographical and behavioral characteristics of people who live in the same neighborhood are similar and then the neighborhoods can be categorized with spatial classifications with the geographical classifications. Thus, this paper, in order to identify the applicability of the geographical classification of the GD, explores the concepts of the geodemographics into Seoul city areas with Korea census data sets that contain key characteristics of demographic profiles in the area. Then, this paper attempt to explain each area classification profile by using clustering techniques with Ward's and k-means statistical methods. For this as as as, this paper employs 2005 Census dataset released by Korea National Statistics Office and the neighborhood unit is based on Dong level, the smallest administrative boundary unit in Korea. After selecting and standardizing variables, several areas are categorized by the cluster techniques into 13, this paps as distinctive cluster profiles. These cluster profiles are used to cthite a short description and expand on the cluster names. Finally, the results of the classification propose a reasonable judgement for target area types which benefits for the people who make a spatial decision for their spatial problem-solving.

Key Words : Geodemographics, area type classification, Ward method, K-means clustering technique, Penetration Index

1. 서론

Geodemographics(GD)는 사람들이 살고 있는 장소와 그 장소에서 거주하고 있는 사람들간의 관계를 분석하는 연구방법으로, 특정 계층의 장소적 성격을 파악하기 위해서는 우선 그 지역의 거주하는 사람들의 인구통계학적인 정보의 분석이 필요하다는 접근 방법을 기초로 발전하였다(Harris *et al.*, 2005). Birkin and Clarke(1998)는 인구학(demography)이 사람들의 유형과 그들의 역학관계에 대해 초점을

두는 학문적 접근이라면 GD는 지역과 공간에 따라 유형화되는 사람들과 이들간의 역학관계를 연구하는 분야로 구분하였다.

이러한 분석을 위해 GD는 통계 데이터가 제공되는 최소 공간단위를 기반으로 데이터베이스를 구축하고, 다양한 공간 데이터 분석기법을 통해 유사한 인구통계학적, 행태적 특성을 공유하는 지역 유형을 도출하고 그 지역의 의미와 특성을 탐구하는데 도입될 수 있다(Harris, *et al.* 2005). 또한 GIS의 공간분석 기법과 결합하여 마케팅 분야를

* 한국교원대학교 지리교육과 석사, 충남발전연구원 환경생태연구부 초빙연구원

** 한국교원대학교 지리교육과 조교수

중심으로 상용화되어 다양한 사회경제적인 영역에서 활용되고 있다(Brown, 1991; Longley and Clarke, 1995; Clarke, 1999; Harris *et al.*, 2005)

이에 따라 지역 연구와 분석을 위한 객관적이고 과학적인 방법에 기초한 각종 지역 및 근린 유형에 대한 해석이 요구되며(Clarke, 1999; 홍일영, 2000) 각 지역 유형이 복잡한 도시 공간 체계 내에서 어떻게 기능하는지에 대한 통계적 프로파일의 제작과 합리적 공간의사결정지원을 위한 GD 기반의 지역 유형 분류의 연구가 시도되고 있다(Feng and Flowerdew, 1999; Webber, 2007).

따라서 본 연구의 목적은 GD 연구 기법에 기초하여 서울시를 사례 지역으로 인구통계학적 특성에 기초하여 지역적으로 서로 유사한 사회경제적 특성을 설명하고자 하였다. 서울시의 경우, 우리나라의 인구와 자원, 중요한 사회경제적 기능 및 역할이 집중되어 있는 공간으로서 우리나라의 도시 체계에서 가장 상위 기능을 차지하고 있는 공간이다. 또한 서울은 우리나라의 다른 어느 도시보다 도시 기능과 역할이 복잡하며 사회분화 정도가 높은 공간 구조로 구성 되어 이러한 구조속에 내재된 유사한 지리적 패턴을 규명하는데 적합한 지역이라고 할 수 있다. 그러므로 본 연구에서 제시하는 GD의 기법과 분석내용들은 향후 서울시를 대상으로 공간의사결정이 요구되는 각종 공공 서비스나 공간 마케팅 전략 수립의 방향설정을 위한 근거를 제공하는데 유용하게 활용될 수 있다고 할 수 있다.

이러한 연구 목적을 달성하기 위해 본 논문은 우선 연구 시점 당시 센서스 공표의 최소 공간단위인 행정동을 GD 분석의 기본 공간단위로 설정하고, 518개 행정동(2007년 기준)을 대상으로 서울시와 서울시 각 구청을 대상으로 '2005년 '인구·주택 총조사' 자료와 재산세 납부 현황, SGIS(Statistical Geographic Information Service)에서 제공하는 데이터를 인구통계 자료로 이용하였다(통계청, 2005)¹⁾. GD 분석의 유용성과 객관적인 분석 결과의 도출을 위해서는 공간단위가 작을수록 다양한 지역 분류 및 유형화가 가능하기 때문에 전수 조사를 기반으로 한 사회경제적 변수의 제공이 가능한 소지역인 '행정동'기반의 자료를 이용하여 GD 분석을

수행하였다. 이러한 자료를 바탕으로 관련되는 세부 변수의 선택과 표준화 과정을 거쳐 계층적, 비계층적 방법(또는 클러스터링 방법)을 통한 군집 분석 후에 최종적으로 각 군집유형별 해석을 통해 서울시의 사회경제적 지역유형에 대한 특징을 도출하였다.

2. 이론적 배경 및 선행 연구

Geodemographics의 역사적 선구자는 런던에서 'Life and Labour of People of London'이라는 저서를 남긴 Charles Booth이다(Harris *et al.*, 2005). 그는 1886년 런던의 빈곤 상태를 조사하여 사회경제적 특성을 토대로 지역을 분류하였는데 이러한 시도는 영국의 사회 구조 연구에서의 최초의 사회적 분류 연구성과였다. 그의 고전적 연구는 이론적, 방법론적 기여와 더불어 20세기 초 영국 사회지리학의 밑그림을 제공하였을 뿐만 아니라(Vickers *et al.*, 2005) 20세기 미국의 경험적인 사회학의 연구 방법에도 영향을 주어 시카고학파의 도시생태학 연구에 많은 기여를 하였다(Harris *et al.*, 2005). 이러한 시카고학파의 도시생태학은 사회지역분석(Social Area Analysis)과 요인생태학(Factorial Ecology)으로 발전하게 되고(최병두, 2002) 이후 컴퓨터 분석 기법이 발달함에 따라 다양한 사회경제적 데이터와 관련 통계학적 기법을 바탕으로 거주구조의 생태적 특성을 밝히는 연구로 진전되었다(차종천 등, 2004). 이러한 흐름에서 지난 20년간 북미와 유럽에서 정보 산업의 빠른 성장에 따라, 지리 데이터의 급증과 데이터 처리기술은 GD에 관한 기본적인 접근들로부터 지리적 방법론을 더욱 발전시키는 새로운 기회를 제공하고 있다.

Geodemographics의 이론적 기반과 방법론, 그리고 그것의 활용에 대한 연구는 국내보다는 국외에서 주로 이루어지고 있다. 국내 관련연구의 경우, 김태호(2004)는 국·공립 유치원의 입지선정을 위해 GD의 개념에 기반하여 우리나라 센서스 데이터를 활용한 GD의 적용 가능성을 모색하였다. GD 시스템의 공공적 활용에 주목하여, 국공립 유치원의 입지를 결정하는 문제에서 GD 분석방법을 이용하여 수요자의 선호를 조사함으로써 보다 효율적이고 실효성 있는 정책결정을 유도할 수 있음을

제시하였다.

또한 GD의 방법론적 측면에서 구축된 공간 데이터베이스로부터 어떤 규칙적인 패턴이나 관계를 찾아내기 위해서 공간 데이터마이닝 작업이 요구되며, 지역유형내 동질성과 지역유형간 이질성을 잘 드러낼 수 있는 대표적인 기법이 클러스터링이다. 이성규 등(2006)은 클러스터링 기법을 이용하여 개별공시지가 산정과정에서 객관적인 기준 부재로 문제점이 발생되고 있는 유사가격권역을 평균연결법과 K-means 기법을 혼합하여 구획하였다. 이 연구에서는 군집화의 객관적 정확도를 평가하기 위해 Rand Index를 통해 도출된 최적 군집수일 때, 군집 내의 동질성과 군집 간의 이질성이 높음을 검증하였다. 또한 군집 내 개별필지와 비교표준지 일치정도를 분석한 결과 높은 부합률을 보이는 것을 밝혀냈다. 나성호(2002)는 고객세분화를 위한 클러스터링 기법 중 K-means 기법과 코호넨 네트워크의 분류 성능에 관해 연구한 결과 CCC 지수와 PST2 통계량을 동시에 고려할 경우 군집의 수가 증가할수록 K-means 기법이 코호넨 네트워크보다 정확한 분류를 하는 것을 밝혀내었다.

클러스터링의 결과를 분석하기 위해 어떤 지표를 사용하여 군집을 구성하는 대상들의 속성을 파악하고, 지역유형 간의 상대적인 차이를 밝혀낼 것인지의 연구는 다음과 같다. Harris *et al.*(2005)은 지역 유형을 프로파일링하고 다른 지역 유형과의 특성을 비교하는 해석적 기준으로 Penetration Rate와 Penetration Index를 언급하였다. 이 지수를 기준으로 크거나 작은 변수들의 값을 통해 지역 유형을 Profiling하는데 사용될 수 있다. 이상일(2007)은 이질적인 두 집단이 어느 만큼 공간적으로 분리되어 있는지를 측정하는 문제를 다루기 위한 '표준화 집중도' 방법을 제안하였는데, Penetration Index를 사용할 때 거주지 분리의 불균등성은 측정할 수 있지만 공간적 집중도는 측정하지 못하는 단점을 극복하는 방법으로 활용될 수 있다.

그 밖의 GD와 유사한 연구의 흐름을 보이는 CRM과 G-CRM에 관한 경영학 분야의 연구가 진행되고 있다. CRM과 G-CRM은 주로 기업이 보유한 고객정보를 기반으로 목표 고객층이 분포하고 있는 공간을 탐색하고 기업의 목적에 적합한 마케팅 전략을 수립하는 것을 목적으로 특정 지역의

고객 성향을 밝혀내는데 클러스터링 등과 같은 공간데이터마이닝 방법을 주로 활용하고 있다(황정래·이기준, 2002).

3. Geodemographics의 방법론적 접근

1) 데이터베이스의 구축

GD를 위한 분석의 기초 작업으로 연구 목적에 적합한 데이터베이스의 구축이 필요하다. 데이터들은 대부분 센서스 자료에서 얻어지는데, 이유는 센서스가 국가기관에 의해 행해지기 때문에 가장 완전하고 신뢰성이 높은 사회·경제적 데이터이고(Vickers *et al.*, 2005), 기본적으로 전수 조사로서 데이터 편중에 대한 우려가 적고 다양한 공간적 스케일과 인구통계학적인 변수들을 폭넓게 포함하고 있기 때문이다(김태호, 2004). 국내 센서스 공표의 최소단위인 읍/면/동은 전통·문화 지역 정체성 등에 따라 역사적 배경을 갖고 성장해 온 단위이고(구자문, 2000), 주요 공공자원의 질적 수준 및 도심과 부심 접근성 측면에서 거의 동질적인 공간이기 때문에(송명규, 1992) 이를 GD 분석의 기본 단위로 설정하고 데이터를 구축할 수 있다. 중요한 것은 기 구축된 데이터베이스 내에서 연구의 목적과 자료 상황에 따라 특징을 적절하게 나타낼 수 있는 최소한의 변수를 선별하는 것이다. 이와 더불어 연령, 성별, 혼인여부, 소득, 직업, 교육 등에 근거한 변수들은 개인의 사회·경제적 특성을 나타내고 주택에 관련된 변수들은 지역간 이질성을 대변해 줄 수 있는 지표가 될 수 있다.

2) 표준화 방법

각각의 변수들은 수집 당시부터 측정 단위가 다르고, 자료의 구조에 있어서 상이한 분포형태를 보이고 있기 때문에 각 변수들의 단위를 동등하게 나타내도록 자료에 대한 표준화가 선행되어야 한다²⁾. 표준화에는 Z-score 표준화, Range 표준화, Inter-decile range 표준화 방법 등이 사용될 수 있다. Z-score 표준화는 정규분포를 가정하는데, 어떤 관찰치 X의 값이 그 분포의 평균으로부터 표준편차의 몇 배 정도나 떨어져 있는가를 표준화

Range standardization $Z_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$ <공식 1>

Inter-decile range standardization $Z_i = \frac{x_i - x_{\text{med}}}{x_{90th} - x_{10th}}$ <공식 2>

된 확률변수 Z로 나타낸다(홍두승, 2005). 그러나 Z-score 표준화 방법은 비대칭적 자료에서 나타나는 데이터 이상치의 영향을 간과할 수 있기 때문에(Vickers et al. 2005), 상대적으로 비대칭적 자료에는 Range 표준화 방법과 Inter-decile range 표준화 방법을 사용한다. 이들 표준화의 공식은 다음과 같다(공식 1, 공식 2).

이때 x_{\max} 는 변수 x의 최대값이고 x_{\min} 은 변수 x의 최소값을 의미한다. 1에 가까운 값이 상대적으로 큰 값이다. x_{med} 는 변수 x의 중앙값이고, x_{90th} 는 변수 x의 90분위수이고, x_{10th} 는 변수 x의 10분위수이다(Vickers et al., 2005).

3) 군집분석 방법

변수들을 동등한 단위로 표준화했다면, 각 케이스(행정동)들을 유사한 특성을 갖는 집단으로 분류하는 작업이 필요하다. 클러스터링 기법은 케이스들을 분류할 때 각 케이스들의 값을 거리(Distance)로 환산하여 가까운 거리에 있는 케이스들을 하나의 집단으로 묶음으로써 군집 내에서 각 케이스들이 서로 유사하도록 만든다. 이에 따라 분류된 케이스들로 구성된 군집들은 군집내 동질성이 높고 군집 간 이질성이 높아지는 결과를 나타낸다(정훈진, 2003). 거리를 측정하는 방법에는 유클리드(Euclidean) 거리, 유클리드 제곱(Squared Euclidean) 거리, 민코우스키(Minkowski) 거리 등이 있는데, 유클리드 거리는 가장 일반적으로 사용되는 거리

측정방법이고<공식 3, 공식 4> 유클리드 제곱거리는 계산이 편리하고 Ward 군집방법을 위한 척도로 사용된다(송문섭·조신섭, 2008).

이러한 거리 척도에 기초해 집단 내의 변량에 대한 집단 간의 변량을 최대화시키는 방법인 계층적 군집방법과 비계층적 군집방법이 있다(이성규 등, 2006). 비계층적 방법인 K-means 방법은 우선 케이스를 K개의 군집으로 나눈 후 군집에 포함되어 있는 케이스들의 평균을 군집의 중심값으로 계산한다. 그리고 군집의 중심값과 각 케이스와의 거리를 계산한 후 가장 거리가 가까운 클러스터에 케이스를 포함시키는 방법이다.

K-means 클러스터링 기법은 임의로 군집의 개수를 정해야하므로, 최적의 군집수를 정하기 위해 다양한 기준 통계량을 제공해 주는 계층적 분석방법을 통해 군집의 개수를 정함으로써 이를 보완할 수 있다. 일반적으로는 Dendrogram의 검토를 통해 군집이 병합되는 단계마다 군집을 형성하는 기준이 되는 근접도(proximity) 계수가 상대적으로 큰 변화를 보일 경우 군집 개수를 결정한다(이경미, 1995). 또한 군집수에 대응되는 기준 통계량을 검토하여 급격한 변화를 보이는 곳에서 군집의 개수를 정하는 방법을 많이 사용하고 있다(김기명·전명식, 1998).

따라서 본 연구에서는 계층적 군집방법을 통해 군집수와 초기 군집들의 중심값을 도출하고 K-means 방법에 이 값들을 대입하여 각 케이스들을 가장 가까운 중심값을 갖는 군집으로 재할당하는 과정

유클리드 거리 $d = \sqrt{\sum_{i=1}^p (X_{1i} - X_{2i})^2}$ <공식 3>

유클리드 제곱거리 $d = \sum_{i=1}^p (X_{1i} - X_{2i})^2$ <공식 4>

- X_{1i} : 케이스 1의 변수 i의 좌표
- X_{2i} : 케이스 2의 변수 i의 좌표
- p: 변수의 수.

을 거쳐 최종적인 군집을 결정하는 방법을 적용하였다.

4) 군집의 기술과 해석

군집분석의 결과로 지역유형이 분류되면, 각 유형에 대한 해석이 요구된다. 각 군집유형에 속한 변수들의 대표성을 다른 군집유형에 속한 변수들의 대표성과 비교함으로써, 군집을 구성하고 있는 대상들의 속성을 파악하고 지역 유형의 특성을 가장 잘 설명할 수 있는 명칭을 부여할 수 있다(정훈진, 2003). 지역유형별 상대적 차이를 파악하여 각 지역유형의 특성을 밝혀내기 위하여 Penetration Index(P-Index)를 산출하는 방법을 적용하였다(공식 5). 예를 들어 이 지수는 각 지역유형(군집)의 전체 인구에 대한 각 지역유형에 속한 특정 인구 집단의 비율을 모든 지역유형의 전체 인구의 총합에 대한 각 지역유형에 속한 특정 인구집단 인구 총합으로 나눈 백분율에 해당한다(Harris *et al.* 2005). 예를 들면 100을 기준으로 하여 100보다 상대적으로 크거나 작은 값을 가지는 변수가 그 지역유형의 특성을 결정하는 변수가 될 수 있다.

즉, 어떤 군집 1의 변수 1에 대한 P-Index의 의미는 군집 1의 전체 인구에 대한 변수 1의 0-9세 인구의 비율을 서울시 전체 인구에 대한 0-9세 인구 비율로 나눈 백분율이 될 수 있다. 따라서 서울시 전체 인구에 대한 0-9세 인구 비율보다 군집 1 전체 인구에 대한 0-9세 인구 비율이 높다면 군집 1의 변수 1은 대표성을 띠고 있다고 할 수 있다.

이렇게 기술된 지역유형의 특성은 특정 목적을 가진 기업이나 기관들에게 지리인구통계학적인 Profile을 제공함으로써 고객에 대한 객관적 판단의 근거를 제시할 수 있다. 또한 기업의 목표시장(Target Market)으로 어느 곳을 마케팅 영향권으로 설정해

야 할 것인지와 그 지역에 어떠한 상품이 적합할 것인지에 대한 의사결정에도 활용될 수 있다.

3. GD를 통한 서울시 군집유형 분석

1) 서울시 군집유형 분석

본 연구에서는 2005년 '인구주택 총조사'에서 동 단위로 공표되는 변수들과 통계청 웹사이트 SGIS에서 획득한 변수들, 그리고 각 구청 통계연보에서 구축한 데이터베이스에서 인구 구성, 가구 구성, 교육 수준, 주택 유형, 주택 점유형태, 주택 나이, 주택 규모, 재산 등 8개 범주에 대하여 하위 37개 변수를 선정하였다(표 1). 이들 변수들은 상대적으로 비대칭적 자료에 사용할 수 있는 Range 표준화 방법으로 표준화하여 0과 1사이의 값을 갖도록 하였다.

표준화 후에 계층적 군집분석 방법 중 지역 구분의 경우 가장 많이 이용되는 Ward 방법과 Ward 방법의 척도로 사용되는 유클리드 제곱거리 계산 방법을 적용하여 유의미한 군집수와 군집의 중심값을 도출하였다. 유의미한 군집수를 선정하기 위해서 군집의 병합과정을 보여주는 덴드로그램을 통해 군집이 병합되는 단계마다 군집을 형성하는 기준이 되는 근접도(proximity) 계수가 현저하게 차이가 나는 단계에서 군집수를 결정하는 방법을 사용하였다. 군집 수에 대응되는 통계량을 확인하여 군집수를 선정하는 방법은 각각의 통계량을 통해 군집수를 해석하는 조건을 만족시키지 않았기 때문에 전자의 방법으로 13개의 초기 군집을 선정하였고, K-means 군집분석을 통해 최종적으로 13개의 지역유형에 할당된 중심값(각 케이스의 평균값)과 동들의 정보를 도출하였다. 최종적으로 분류된 13개 지역유형은 <그림 1>에 나타나 있다. <표 2>는 13개 군집별로 분포된 서울시 행정동 비율을

$$\text{Penetration Index} = \frac{n / \sum_{n=1}^k n}{N / \sum_{n=1}^k N} \times 100 \dots\dots\dots \text{<공식 5>}$$

- n: 각 지역유형의 특정 변수에 해당하는 수
- N: 전체 케이스의 특정 변수에 해당하는 수
- k: 케이스의 수

표 1. 8개 범주의 하위 37개 변수

주요 영역	ID	변수 이름	주요 영역	ID	변수 이름
인구구성	v1	Age 0-9	주택 유형	v19	단독주택
	v2	Age 10-19		v20	아파트
	v3	Age 20-29		v21	연립주택+다세대주택
	v4	Age 30-39		v22	비거주용 건물주택
	v5	Age 40-49		v23	자가
	v6	Age 50-64	주택 점유형태	v24	전세
	v7	Age 65+		v25	월사글세
	v8	인구밀도		주택나이	v26
	v9	미혼	v27		Age 1990-1999
	v10	배우자있음	v28		Age 1980-1989
	v11	사별+이혼	v29		Age 1980-
교육수준	v12	미취학+초교+중교	주택규모	v30	1-9평
	v13	고교+전문대		v31	9-19평
	v14	대학교+대학원		v32	19-29평
가구구성	v15	1세대가구		v33	29-39평
	v16	2세대가구		v34	39-49평
	v17	3세대가구		v35	49-69평
	v18	비혈연가구		v36	69평 이상
		재산		v37	재산세

나타낸 수치이다.

2) 서울시 GD Profiling

(1) 서울시 군집 유형 해석

지역유형을 결정하는 핵심적 변수들을 토대로 지역유형의 특성을 해석하기 위해 각 유형에서 어떤 변수가 대표성을 보이는지 <표 1>의 37개 사회경제적 변수들에 대한 P-Index 값을 도출하여 지역 유형의 특성을 인구구성, 가구구성, 교육수준, 주택유형 및 소유형태, 주택상대 등 총 5개 특성으로 분류하였다. <그림 2>는 이러한 지역 유형에 대한 행정동의 분포패턴이며 각 군집별로 집중과 분산 패턴을 가지적으로 쉽게 파악할 수 있다.

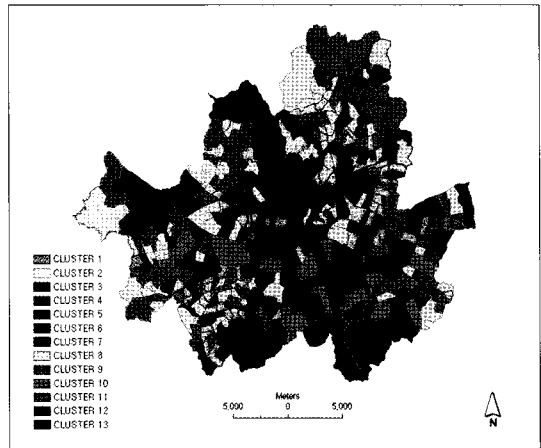


그림 1. 서울시 13개 군집유형의 지도화

표 2. 군집별 케이스(동) 수

군집	군집별 동수	군집	군집별 동수	군집	군집별 동수
1	22(4.3%)	2	69(13.3%)	3	57(11.0%)
4	38(7.3%)	5	24(4.6%)	6	53(10.2%)
7	77(14.7%)	8	69(13.3%)	9	49(9.5%)
10	23(4.4%)	11	22(4.3%)	12	5(1.0%)
13	10(1.9%)	유효 518 / 결측 0			





CLUSTER 9
seoul

5,000 Meters 0 5,000



군집 7

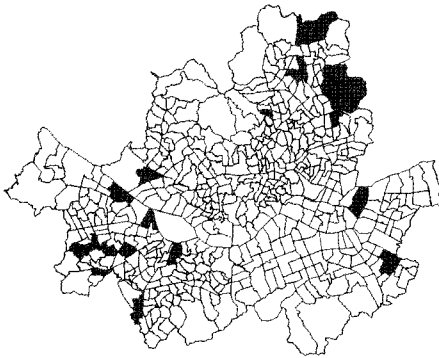


CLUSTER 10
seoul

5,000 Meters 0 5,000



군집 8



CLUSTER 11
seoul

5,000 Meters 0 5,000



군집 9

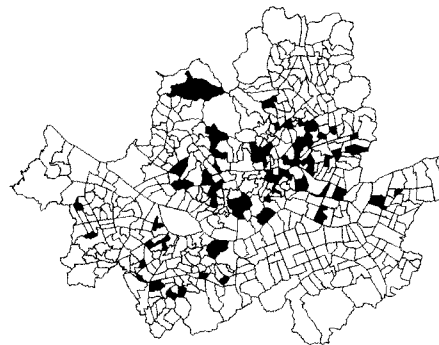


CLUSTER 12
seoul

5,000 Meters 0 5,000



군집 10

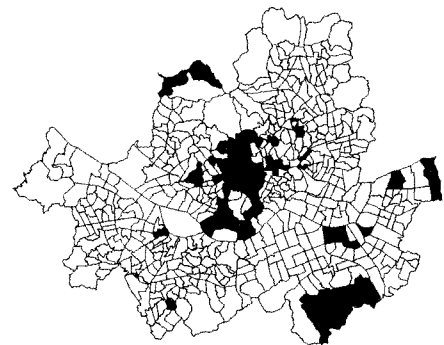


CLUSTER 7
seoul

5,000 Meters 0 5,000



군집 11

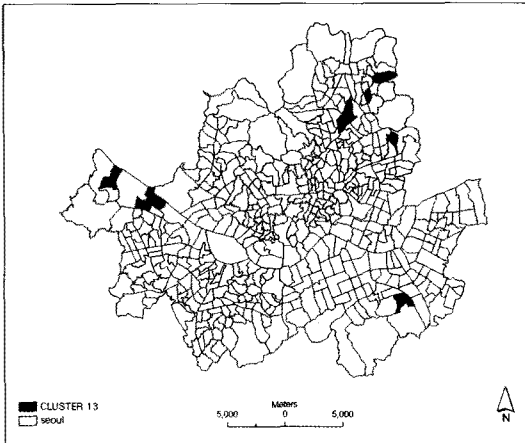


CLUSTER 6
seoul

5,000 Meters 0 5,000



군집 12



군집 13
그림 2. 군집1-13의 공간적 분포패턴

다음은 GD분석에 의한 군집의 공간적 분포 패턴을 바탕으로 각 군집의 유사한 지리통계적 패턴에 대한 공간적 해석이다. 군집 1의 경우, 0-19세, 40-49세 인구 비율과 기혼인구 비율, 고학력 인구 비율이 높고, 2세대가구 비율이 매우 높다. 아파트에 거주 가구가 대부분이고 자가 소유 비율이 비교적 높으며, 1980년대 지어진 10평대 주택 비율이 매우 높은 특징을 보인다.

군집 2는 0-9세, 30-39세, 50세 이상 인구 비율과 사별, 이혼인구 비율이 평균 이상이다. 3세대가구와 비혈연가구의 비율이 비교적 높다. 대졸 이상의 인구비율이 상당히 낮고, 평균학력 수준 이하의 인구비율이 높다. 연립, 다세대 주택에 거주하는 가구 비율이 매우 높고, 자가소유 비율과, 2000년대 건축된 10평대의 주택 비율이 비교적 높다.

군집 3은 인구면에서 연령대별 고른 구성을 보인다. 2,3세대가구 비율이 비교적 높고, 교육수준 또한 높다. 아파트 거주 가구와 자가소유 가구 비율이 매우 높고, 2000년대 지어진 19-39평대 주택 비율이 비교적 높다.

군집 4는 0-19세, 40-49세의 인구 비율과 기혼인구 비율이 비교적 높다. 2,3세대가구의 비율이 상당히 높고, 대졸 이상의 인구 비율이 비교적 높다. 아파트 거주 가구와 자가소유 가구의 비율, 1990년대 이후에 지어진 19-39평대 주택의 비율이 상당히 높다.

군집 5는 20-29세 인구 비율과 미혼인구 비율이

상당히 높다. 1세대가구와 비혈연가구의 비율이 높은 특징을 보인다. 오피스텔 거주 가구와 월사글세 가구의 비율, 2000년대에 지어진 주택의 비율이 상당히 높다. 교육수준 또한 상당히 높은 지역 유형이다.

군집 6은 20-29세, 50세 이상 인구 비율과 미혼, 사별, 이혼인구 비율이 비교적 높다. 1세대와 비혈연가구의 비율과, 저학력의 인구 비율이 상당히 높다. 단독주택과 오피스텔 거주 가구와 월사글세 가구 비율이 매우 높고, 1980년대 이전에 건축된 1-9평대, 40평대 이상의 주택 비율이 상당히 높은 특징을 보인다.

군집 7은 20-29세, 65세 이상 인구 비율과 미혼, 사별, 이혼인구 비율 그리고 1세대와 비혈연가구의 비율이 상당히 높다. 교육수준에 있어서는 저학력의 인구 비율이 상당히 높고, 평균 학력의 인구 비율이 비교적 높다. 단독주택과 오피스텔 거주 가구 비율과 임대 가구의 비율이 높다. 1980년대 이전에 건축된 낮은 주택이 많고 40평대 이상의 주택 비율이 상당히 높다.

군집 8은 20-29세, 65세 이상의 인구 비율과 사별, 이혼인구 비율이 비교적 높다. 1세대와 비혈연가구의 비율과, 평균 교육수준 이하의 인구 비율이 비교적 높다. 단독주택 거주 가구의 비율이 매우 높고, 월사글세의 비율이 특히 높다. 1980년 이전에 건축된 낮은 주택이 많고, 40-49평의 주택 비율이 상당히 높다.

군집 9는 연령대별 고른 인구구성을 보이고, 사별, 이혼인구 비율이 평균 이상이다. 교육면에서 평균수준 이하의 인구 비율이 비교적 높다. 단독주택과 연립·다세대주택 가구 비율이 상당히 높고, 전세가구 비율이 평균 이상이다. 2000년대 건축된 50평대 이상의 주택비율이 비교적 높다.

군집 10은 10-19세, 40-49세 인구 비율이 비교적 높고, 기혼인구 비율과 2세대가구 비율이 상당히 높다. 4년제 대학 이상의 교육을 받은 인구 비율이 매우 높다. 아파트 거주 가구와 자가소유 가구 비율과, 1980년대와 그 이전에 건축된 30-60평대의 주택 비율이 상당히 높다.

군집 11은 0-19세, 40-49세의 인구 비율과 기혼인구 비율이 비교적 높다. 2세대, 3세대가구 비율과, 4년제 대학 이상의 교육을 받은 인구 비율이

표 3. 각 군집별 Profile

군집	각 군집별 Profile
1	학령자녀를 둔 고학력의 2세대가구가 거주하는 오래된 소형 아파트 지역
2	낮은 학력의 3세대와 비혈연가구가 거주하는 신형의 연립 다세대지역
3	고학력의 2, 3세대 가구로 구성된 20~30평대 새 아파트 밀집지역
4	학령자녀를 둔 2, 3세대 가구가 거주하는 높은 학력의 현대식 아파트 지역
5	고학력의 미혼 인구가 세 들어 거주하는 오피스텔 지역
6	무주택의 낮은 학력의 미혼과 노년인구로 구성된 단독주택과 오피스텔이 혼재된 도심지역
7	낮은 학력의 미혼과 노년인구가 거주하는 넓고 조밀한 혼합 주택지역
8	평균학력 이하의 빈곤한 미혼과 노년 인구가 세 들어 거주하는 넓고 조밀한 단독주택지역
9	평균학력 이하의 고령의 인구로 구성된 단독 주택과 연립 다세대 주택지역
10	청소년기 자녀를 둔 고학력의 장년부부가 거주하는 부유한 중형 아파트지역
11	어린 자녀를 둔 평균 학력 수준 이상의 가구가 거주하는 현대식 소형 아파트 지역
12	고학력의 20~30대 미혼 인구가 세 들어 사는 단독주택과 오피스텔 지역
13	무주택의 홀로된 노년과 장년이 포함된 2~3세대 가구가 거주하는 소형 아파트 지역

상당히 높다. 아파트 거주 가구와 자가소유 가구의 비율과, 1990년대 건축된 10평대, 20평대의 주택 비율이 상당히 높다

군집 12는 20-29세, 30-39세의 인구 비율이 높고, 특히 미혼인구의 비율이 매우 높으며, 1세대와 비혈연가구 비율이 상당히 높다. 교육수준에 있어서는 고학력 인구 비율이 매우 높다. 단독주택과 오피스텔에 거주하는 월사글세 가구의 비율이 매우 높고, 1990년대 건축된 큰 평형대의 주택비율이 높은 특징을 보인다.

군집 13은 65세 이상 인구 비율과 사별, 이혼인구 비율이 상당히 높다. 2세대가구의 비율 또한 상당히 높고, 3세대가구의 비율도 비교적 높다. 아파트 거주 가구의 비율과, 월사글세 가구의 비율이 상당히 높고, 1990년대 건축된 1-9평대 주택 비율이 매우 높다. 반면 교육수준이 상당히 낮다. <표 3>은 이러한 13개의 군집별 특징을 요약하여 각 군집별 profile을 제시한 자료이다.

(2) 군집유형간 특성 비교

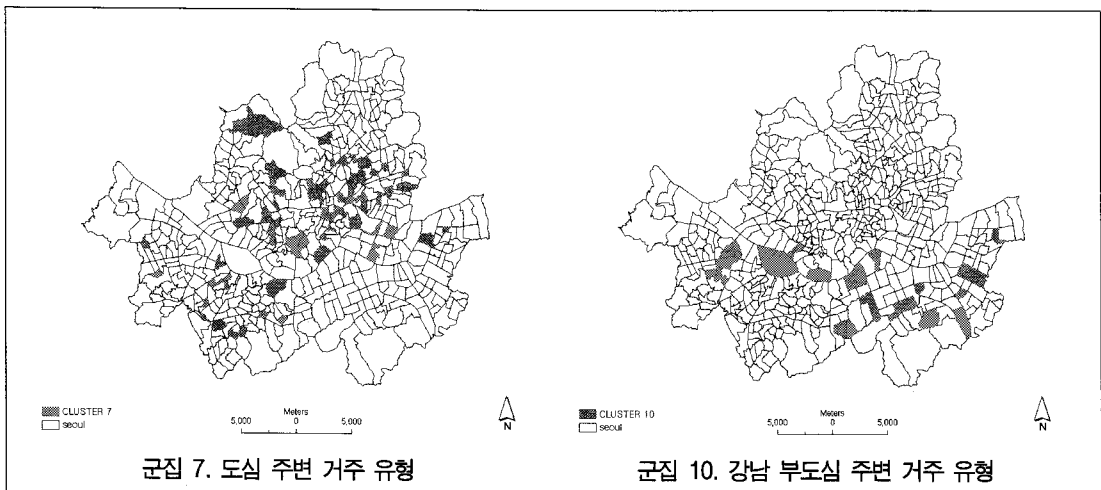


그림 3. 다른 특성을 보이는 도심 주변과 강남의 군집

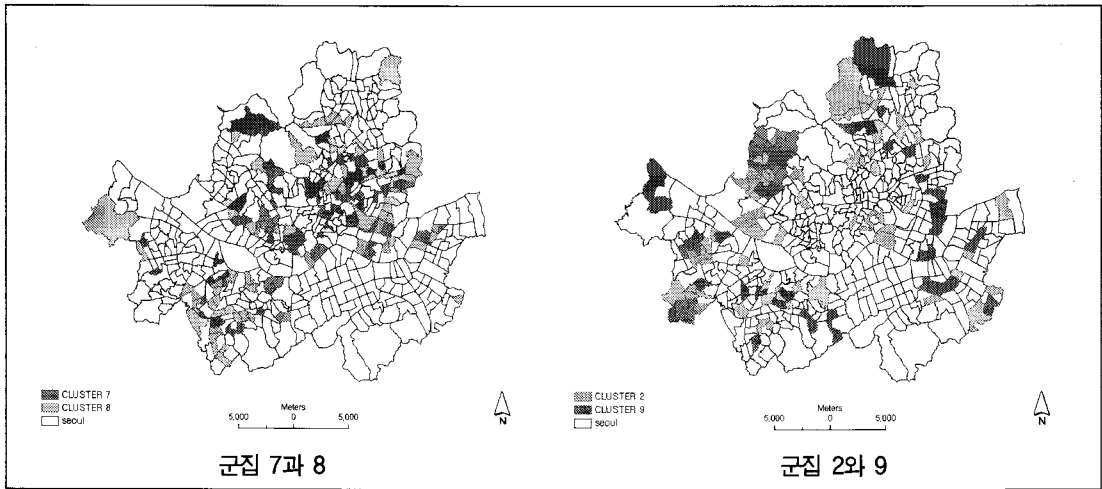


그림 4. 유사한 특성을 공유하는 군집간의 지리적 근접성

이를 바탕으로 각 군집의 특성을 해석하고, 유사한 인구통계학적 특성을 공유하는 지역들이 서울시의 도시공간에서 어떤 패턴으로 분포하고 있는지 지리적 맥락에서 해석할 수 있다.

<그림 3>은 인구통계학적으로 대조적인 두 지역 유형의 사례를 보여 주는 것으로 도심적 성격의 거주지역과 부도심적 성격의 거주 지역의 특징을 보여 주는 지역 유형의 공간적 분포이다.

군집 7과 10의 경우, 명동을 중심으로 하는 도심 주변의 거주지역인 군집 7과 강남·서초·송파구 중심의 군집 10의 인구통계학적 특성은 매우 대조적임을 알 수 있다. 즉 서울시 공간구조에서 도심과 강남 부도심의 두 축을 중심으로 유사한 특성들이 뚜렷하게 분류되는 특징으로 해석될 수 있다. 군집 7의 경우, 20대와 65세 이상 인구, 높은 인구 밀도, 낮은 학력, 단독주택과 오피스텔, 월사글세, 1세대 또는 비혈연가구, 낮은 주택의 변수에서 대표성을 보인다. 그러나 군집 10의 경우, 10대와 40대, 2세대가구, 높은 학력, 자기 소유의 중형 아파트, 재산세의 변수에서 대표성을 보인다. 이러한 결과는 오래되고 상대적으로 낙후한 도심을 중심으로 한 강북과 1970년대 이후의 강남지역의 개발을 통한 생활권의 분리가 이들 지역과 지리적으로 가까운 지역들의 인구통계학적 특성들을 유사하게 변화시킨 결과라고 해석할 수 있다(차충천 등, 2004).

이와 대조적으로 그림 4는 지리적으로 근접성을 보이면서 매우 유사한 사회경제적 특성을 보이는

지역 유형을 보이는 군집들 간의 사례이다.

군집 7과 지리적 근접성을 보이는 군집 8의 경우에는 군집 7과 상당히 유사한 특성을 보이지만 20대와 65세 이상 인구를 제외한 연령대의 비율이 높아지고 65세 이상 인구의 대표성이 상당히 완화되는 지역적 특징이 있다. 또한 이 군집 유형에서는 단독주택과 연립, 다세대 주택의 비율이 조금씩 높아지고 오피스텔의 비율은 약간 감소한다. 주택 건설 시기면에서 1990년대 지어진 주택비율이 상승하고, 주택 규모면에서 큰 평형의 비율이 증가한다. 군집 2와 군집 9도 인구통계학적 특성이 상당히 유사한 군집으로 지리적으로도 상당히 인접해 있으면서 이 군집들은 주택과 관련한 변수에서 상대적으로 군집 9의 주택들이 더욱 최근에 건축되었고 주택 소유 측면에서 전세의 비율이 높게 나타나는 것으로 해석된다.

5. 결론

본 연구는 GD의 개념을 도입하여 국내 상황에 적합한 지역유형 분류를 시도하고, 그에 따른 공간적 분포패턴을 설명하고자 하였다. 이를 위해 서울시의 2007년 기준 518개 행정동을 대상으로 데이터베이스를 구축하고 변수들의 선택과 표준화 과정을 거쳐, 군집분석 기법을 이용하여 유사한 인구통계학적 특성을 공유하는 지역들이 서울시의 도시공간에서 어떤 패턴으로 분포하고 있는지를 지

리적 맥락에서 해석을 시도하였다.

서울시를 대상으로 GD의 방법론을 적용한 사례에서 유사한 인구통계학적 특성을 갖는 13개의 지역 유형을 도출하였고, 그 지역의 특성을 공간적 분포패턴과 연관시켜 해석하였다. 한 예로, 도심과 강남 부도심의 두 축을 중심으로 유사한 특성들이 뚜렷하게 분류되는 특징이 나타났다. 이것은 오래되고 상대적으로 낙후한 도심을 중심으로 한 강북과 1970년대 이후의 강남지역의 개발을 통한 생활권의 분리가 이들 지역과 지리적으로 가까운 지역들의 인구통계학적 특성들을 유사하게 변화시킨 결과라고 할 수 있다.

이러한 연구 성과 및 의의를 바탕으로 향후 연구 내용으로는 우선 개별 변수들간의 오차에 관한 통계학적 검증에 대한 심도있는 분석이 필요하다. 본 논문은 GD 활용성에 초점을 두고 연구를 진행했음에도 불구하고 GD 분석을 위해서는 선정된 변수들간의 독립성(independence of error)이나 동일한 분산 정도의 측정(homoscedasticity), 다중공선성(multicollinearity) 등과 같은 통계학적 분석 시도 및 분석 알고리즘에 대한 이론적 연구도 필요하다. 또한 지역 유형의 의미와 특징의 해석에 있어 공간 단위에 대한 인구 규모를 고려한 P-Index 해석이 필요하다. 즉 동일한 P-Index 값이라고 하더라도 각 공간단위의 인구 규모가 크다면 그 비중값이 갖는 통계적 유의성은 더 커질 수 있기 때문에(Rogerson, 1999; 이상일, 2007), 대표성의 확인과 공간 단위의 상이한 규모 수준을 고려할 수 있는 지표의 개발이 필요하다. 그리고 본 연구는 인구 센서스 및 재산세 자료에 국한하여 분석하였지만, 다양한 지방자치단체의 인구통계관련 속성 정보 및 자료를 활용한다면 보다 현실성있는 GD 분석을 시도할 수 있다. GD에서 이용할 수 있는 속성 자료의 범주에서 우선적으로 고려할 수 있는 자료가 최소 구역의 각종 인구 관련 센서스 자료임에도 불구하고 센서스에서 제공하는 자료만으로는 다양한 목표 유형의 지역 탐색에 필요한 최적의 변수의 정보를 획득하는데 한계가 있다. 그렇기 때문에 여가, 선호, 가치 등의 정보 등을 제공하는 민간 기업의 고객 데이터베이스, 각 정부 기관이나

연구기관에서 자체적으로 구축한 다양한 인구통계 자료의 확보와 이용을 통해 GD 분석의 정확성을 높이는데 중요하다. 또한 Fuzzy 분석과 같은 알고리즘 기법을 고려하여, 도출된 결과와 실제 패턴과의 일치성을 토대로 최적의 군집분석 기법을 선택하는 과정이 고려될 수 있고 군집의 특성을 분석하면서 개인의 특성을 추론할 때 발생하는 생태적 오류(Ecological Fallacy)의 위험이 발생할 수 있기 때문에 이를 감소시킬 수 있는 방안들에 대한 모색이 필요하다. 이러한 오류를 해결하기 위한 방안으로 집계구라는 소지역 기반의 지리통계정보가 구축된다면 향후 GD의 연구에 있어 생태적 오류를 감소시키고 정확한 결과를 도출하는데 도움이 될 것이다.

주

- 1) 본 연구의 속성 데이터는 동 단위로 수집되어 있기 때문에 공간 DB인 서울시의 동 단위 행정구역 경계 지도를 사용하여 데이터 정보의 공간 단위별 표현이 가능하도록 하였다. 서울시의 동 단위 행정구역 경계 지도는 2007년의 행정경계를 기준으로 25개구 518개동으로 구분된 데이터를 사용하였다. 그러나 '인구·주택 총조사' 데이터는 2005년의 522개동의 행정경계를 기준으로 수집되었고 통계청 통계지리정보 서비스(SGIS)에서 제공하는 데이터는 2007년의 행정경계를 기준으로 제공되었다. 그러므로 공간 단위의 일치를 위해서 2007년 행정구역 경계를 기준으로 데이터를 통합하여 공간 단위를 재조정하였다.
- 2) 각각의 변수들은 데이터 수집 당시부터 측정 단위가 다르고, 자료의 구조에 있어서 상이한 분포형태를 보이고 있기 때문에 반드시 자료에 대한 표준화가 선행되어야 한다. 예를 들어 관측값을 대상으로 정규분포 Q-Q도표(quantile-quantile plot)를 이용하여 변수들의 정규성을 검증할 수 있다. 산점도의 모양이 직선에 가까우면 정규성을 가정하며, 직선과 거리가 먼 곡선의 형태를 보이면 정규성을 가정할 수 없는 자료로 해석할 수 있다(송문섭·조진섭, 2008). 예를 들어 2세대기구의 산점도가 직선에 가까운 모양을 보인다면 정규성을 가정할 수 있는 반면, 월사글세의 산점도의 경우, 곡선의 형태를 보인다면 정규성을 가정할 수 없다고 해석한다. Z-score 표준화는 데이터의 정규성을 가정하고 있기 때문에 비정규성을 보이는 자료를 사용했을 경우 자료를 왜곡할 수 있다. 따라서 본 연구에서는 상대적으로 비대칭적 자료에 사용할 수 있는 범위 표준화 방법으로 변수들을 표준화하였다.

문헌

- 구자문, 2000, 센서스를 위한 수치지도의 개발과 센서스트랙의 지정에 관한 연구, 한국인구학, 23, 181-196.
- 김기명·전명식, 1998, SAS 군집분석, 자유아카데미, 서울.
- 김태호, 2004, Geodemographics를 활용한 공공 시설물의 입지에 관한 연구, 서울대학교 석사학위논문.
- 나성호, 2002, 고객세분화를 위한 군집분석 중 K-평균 군집분석과 코호넨 네트워크 분류 성능에 관한 연구, 서울대학교 석사학위논문.
- 송명규, 1992, 지방공공재가 소득계층별 주거지 분화에 미치는 영향에 관한 연구, 서울대학교 박사학위논문.
- 송문섭·조신섭, 2008, SAS를 이용한 통계자료분석, 자유아카데미, 서울.
- 이경미, 1995, 덴드로그램의 동적 접근에 의한 계층적 집락분석의 소프트웨어 구현, 통계연구, 3, 43-53.
- 이상일, 2007, 거주지 분화에 대한 공간통계학적 접근(I): 공간 분리성 측도의 개발, 대한지리학회지, 42, 616-631.
- 이성규·홍성언·박수홍, 2006, 평균연결법과 K-means 혼합클러스터링 기법을 이용한 공시지가 유사가격권역의 설정, 대한지리학회지, 41(1), 121-135.
- 정훈진, 2003, K-평균군집모형에서 군집수 결정 방법에 관한 연구, 중앙대학교 석사학위논문.
- 차종천·유홍준·이정환, 2004, 서울시 계층별 주거지역 분포의 역사적 변천, 백산서당, 서울.
- 최병두, 2002, 현대사회지리학, 한울아카데미, 서울.
- 홍두승, 2005, 한국의 중산층, 서울대학교출판부, 서울.
- 홍일영, 2000, GIS의 마케팅 응용에 관한 연구: 소매점 상권분석을 중심으로, 지리교육논집, 44, 26-39.
- 황정래·이기준, 2002, gCRM과 공간데이터마이닝, 한국공간정보시스템학회 춘계학술대회 논문집, 38-44.
- Birkin, M. and Clarke, G., 1998. GIS Geodemographics, and Spatial Modeling in the U.K. Financial Service Industry, *Journal of Housing Research*, 9, 87-111.
- Brown, P., 1991, Exploring geodemographics, in Masser, I. and Blackmore, M.(eds.), *Handling Geographical Information: Methodology and Potental Applications*, Longman Scientific and Technical, Essex, 221-258.
- Clarke, G., 1999, Geodemographics, marketing and retail location, in Pacione, M. *Applied Geography: Principles and Practice*, Routledge. London, 577-592.
- Feng, Z. and Flowerdew, R., 1999, The use of fuzzy classification to improve geodemographic targeting in Gittings, B.(ed.), *Innovations in GIS 6*, Taylor & Francis, London, 133-144.
- Harris, R. Sleight, P. and Webber, R., 2005, *Geodemographics, GIS and Neighbourhood Targeting*, John Wiley and Sons, London.
- Martin, D., 1998, Automatic neighbourhood identification from population surfaces, *Computers, Environment and Urban Systems*, 22, 107-120.
- Longley, P., and Clarke, G., 1995, GIS for Business and Service Planning, John Wiley and Sons.
- Openshaw, S., 1995, Marketing spatial analysis: a review of prospects and technologies relevant to marketing, in Longley, P. and Clarke, G. (eds.), *GIS for Business and Service Planning*, GeoInformation International, Cambridge, 150-166.
- Rogerson, P. A., 1999, The detection of clusters using a spatial version of the Chi-square goodness-of-fit statistics, *Geographical Analysis*, 31, 130-147.
- Vickers, D. Rees, P. and Birkin, M., 2005, *Creating the National Classification of Census Output Areas: Data, Methods and Results*, Working Paper No.2, School of Geography, University of Leeds, U.K.
- Webber, R., 2007, Geodemographics, in Wise, S. and Craglia, M.(eds.), *GIS and Evidence-Based Policy Making*, CRC Press, London,

43-68.

통계청, 2005, 인구주택총조사보고서

통계청 통계지리정보서비스(SGIS) <http://gis.nso.go.kr>

go.kr

- 교신 : 김영훈, 363-791, 충북 청원군 강내면 다락리 산 7
한국교원대학교 지리교육과 gis@knue.ac.kr, 전화: 043-

230-3641, 팩스: 043-231-4948

Corresponding author, Kim, Young Hoon, 363-791,
Department of Geography Education, Korea National
University of Education, San 7, Darak-Ri, Cheongwon
-Gun, Chungbuk, gis@knue.ac.kr, Tel: 043-230-3641,
Fax: 043-231-4948

(접수: 2009.6.13, 수정: 2009.7.11, 채택: 2009.8.8)