

K-평균 군집화 기법을 이용한 공시지가 유사가격권의 설정

A Similar Price Zone Determination of Public Land Price Using K-means Clustering Technique

이성규*¹⁾, 홍성연²⁾, 박수홍³⁾

Yi Seong-Kyu, Hong Sung-Eon, Park Soo-Hong

인하대학교 지리정보공학과 석사과정(pla20@hanmail.net)¹⁾

인하대학교 지리정보공학과 박사과정(hongsu2005@hanmail.net)²⁾

인하대학교 지리정보공학과 조교수(shpark@inha.ac.kr)³⁾

요 약

비교표준지를 이용하여 개별공시지가를 산정하는 우리나라 제도 하에서 가장 중요한 문제는 개별필지 주변의 표준지 중에서 어떤 표준지를 선택·이용하여 지가를 산정해야 하는가이다. 그러나 지침상에서는 비교표준지 선정시 중요 요인으로 작용하고 있는 유사가격권에 대하여 수치적인 기준이 아닌 모호한 개념상으로 규정하고 있어 비교표준지 선정에 있어 부정확성을 초래하고 있다.

따라서, 본 연구에서는 객관적인 기준과 수치적인 기준의 부재로 많은 문제점을 발생시키고 있는 유사가격권 설정의 문제를 해결하고자 K-평균 군집화 기법을 활용하여 가격권을 설정하고 이에 대한 타당성을 제시하고자 한다.

산정하고 있다(이효상, 2001).

비교표준지를 이용하여 개별공시지가를 산정하는 우리나라 제도 하에서 가장 중요한 문제는 개별필지 주변의 표준지 중에서 어떤 표준지를 선택·이용하여 지가를 산정해야 하는가이다. 지침상에서는 비교표준지 선정시 용도지역, 유사가격권, 토지이용상황, 도로접면 등을 고려하도록 되어있다. 유사가격권을 제외한 요소들은 전산 코드가 되어 있어 용이하게 개별필지의 특성을 조사할 수 있다. 그러나 유사가격권의 경우는 비교표준지 선정시 지침상에서 상당히 중요한 요소로 취급되고 있으면서도 어느 정도의 범위를 유사가격권으로 설정할 수 있는가에 대해서는 개념상으로만 모

1. 서론

토지의 가격을 정확히 평가하고 공시하는 것은 토지정책의 성패를 좌우한다고 해도 과언이 아닐 정도로 중요한 일이라 하겠다. 그런데 공시지가제도 도입 초기 개별공시지가를 산정 하는데 있어 수작업에 의한 산정으로 지가에 대한 공신력과 정확성이 결여되었다. 이러한 문제를 해결하기 위해 건설교통부는 1994년부터 지가 자동화 산정 프로그램인 ALPA(Automatic Land Price Appraisal System)를 보급하였고, 현재는 이 프로그램에 의해 개별공시지가를

호하게 규정하고 있다(건설교통부, 2002).

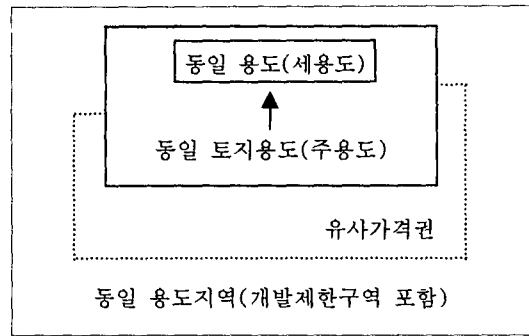
유사가격권의 설정에 관한 연구는 다양하지 못하며 다만, 비교표준지 선정 자동 선정 과정에서 유사가격 요소를 표준지와 개별지간의 지가차이로 유사가격권의 개념을 적용시키거나(박수홍·홍성언, 2003), 또는 유사가격을 도로접면 조건 등에 포함시켜 설정권을 해석한 선행연구가 있었다(박성규, 1999).

따라서, 본 연구에서는 객관적인 기준과 수치적인 기준의 부재로 많은 문제점을 발생시키고 있는 유사가격권 설정의 문제를 해결하고자 K-평균 군집화 기법을 활용하여 가격권을 설정하고 이에 대한 타당성을 제시하고자 한다. 구체적으로, 연구에서는 군집화 인자로 당해 년도 개별공시지가와 필지간 공간상의 거리를 이용하여 유사가격권을 설정하였고, 이에 대한 타당성을 제시하고자 설정된 가격권내에 존재하는 개별필지에 대한 비교표준지 분포를 분석하였다.

2. 유사가격권과 K-평균 군집화

2.1 유사가격권의 개념과 문제점

비교표준지의 선정 기준의 경우, 조사대상 토지가 일반토지인 경우는 조사대상 토지와 동일한 용도지역(개발제한구역 포함) 안에 있는 유사가격권의 표준지 중에서 토지이용상황(주용도)이 같은 표준지를 선정하도록 되어 있다. 만일, 동일한 용도지역 내 토지이용상황이 같은 유사가격권의 표준지가 없는 경우에는 토지이용상황이 다르더라도 조사대상 필지 인근의 토지이용상황을 감안하여 유사가격권의 표준지를 선정하여야 한다고 규정하고 있다(건설교통부, 2003)[그림 1].



[그림 1] 비교표준지 선정 요령(개념도)

이렇게 현행 지침상에서는 비교표준지 선정시 중요 요인으로 유사가격권이라는 기준을 규정하고 있다. 그러나 유사가격권의 가격적 편차, 범위 등에 대해서는 수치적이고 구체적인 기준으로 제시하지 못하고 있다. 단지 땅값의 형성요인(도로조건, 건축규제, 주변여건 등)이 비슷하여 유사한 가격대를 형성하는 지역적 범위를 말한다고 개념적으로 정의하고 있다. 이렇게 모호한 유사가격권의 규정으로 비교표준지 선정에 있어 객관성과 합리성이 결여되고 있으며, 비교표준지 선정 자동화에 문제점으로 지적되고 있다.

선행 비교표준지 자동 선정에 관한 연구에서도 유사가격권의 모호성으로 인하여 비교표준지를 자동화된 방법론으로 선정할 경우 범위를 객관적이고 수치적인 기준으로 정립하기 곤란하여 자동화 구현에 한계성을 지적하였다. 그리고, 현행 수작업에 의한 비교표준지 선정과 자동 선정 결과가 정확하게 일치하지 않는 주요원인 중의 하나가 바로 유사가격권의 정의·설정 문제로 지적하고 있다(박수홍·홍성언, 2003).

이렇듯 유사가격권의 중요성에 비해 그에 대한 명확한 수치적인 기준을 제시하지 못하고 있어 지가 산정 관련 자동화에 있어 객관적인 적용의 어려움이 있다. 따라서, 현행 관련법률이나 지침에서 기술하고 있는 유사가격권에 관한 범위를 보다 명확

히 할 필요성이 있다.

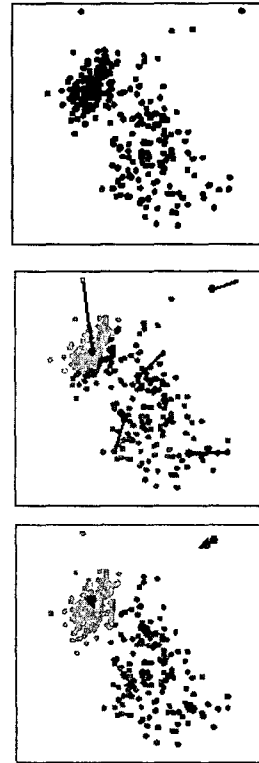
2.2 K-평균 군집화

객관적인 기준에 의해 구체적으로 유사 가격권을 설정하기 위해 연구에서 이용하고자 하는 K-평균 군집화 기법은 데이터 마이닝의 한 분야로서 군집화에 가장 널리 사용되고 있는 기법이다(이근수·김삼근, 1991). 이 기법은 MacQueen에 의하여 제안된 알고리즘으로서, [그림 2]와 같이 우선 패턴을 k개의 군집으로 나눈 후 군집에 포함되어 있는 패턴들의 평균을 클러스터의 중심값으로 계산한다. 그리고 이 중심값과 각 패턴과의 거리를 계산한 후 가장 거리가 가까운 클러스터에 패턴을 포함시키는 방법으로 조건은 다음의 식과 같다(조형기·민준영, 1996; 김윤식·모경주, 2000).

$$x_i \in C_j, \quad \|x_i - z_j\|^2 < \|x_i - z_k\|^2$$

여기서, $1 \leq i \leq N$, $1 \leq k \leq c$, $j \neq k$ 이며, N은 패턴 수, c는 군집 수, z는 군집의 중심값을 나타낸다.

이 기법은 초기 군집수를 결정하는 것이 중요한데 초기 k개의 클러스터의 중심값을 주는 방법에는 주어진 패턴에서 처음 k개의 패턴을 추출하여 중심값을 하는 방법과 임의로 k개를 추출하여 중심값으로 하는 방법이 있다. 본 연구에서는 유사가격권의 설정과 관련한 선행 연구가 다양하지 못하기 때문에 대상지역에 대하여 군집수를 다양하게 설정하여 군집화를 시행하였다.

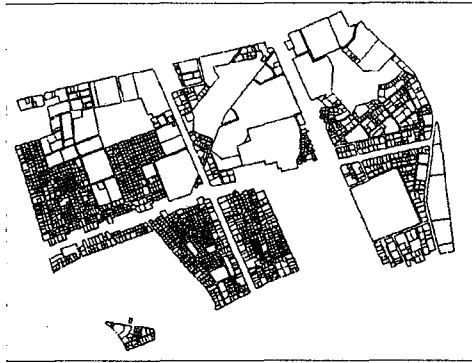


[그림2] K-평균 군집화 기법의 군집화 과정

3. 실험 및 분석

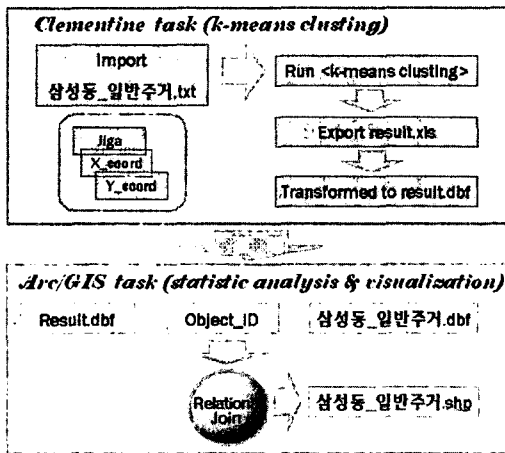
3.1 실험

K-평균 군집화 기법을 이용하여 실제 유사가격권을 설정하고자 서울시 강남구 삼성동일부 지역(일반주거지역)을 연구지역으로 선정하였다. 데이터는 강남구에서 구축하여 놓은 토지특성 도면을 이용하였다. 데이터 구축시 특수토지나 공공용지의 경우는 일반토지와 비교표준지 선정하는 기준이 상이하기 때문에 이를 포함할 경우 군집화에 오류가 발생할 수 있기 때문에 이는 배제하였다[그림 3]. 군집화는 마이닝 툴로 잘 알려진 SPSS社의 Clementine 6.5 소프트웨어를 이용하였으며, 군집화된 결과는 Arc/GIS를 이용하여 분석을 하였다.



[그림 3] 연구대상지역
(강남구 삼성동 일반주거지)

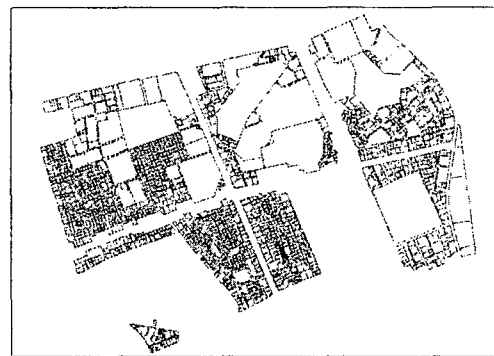
K-평균 군집화 기법은 군집화 이전에 미리 군집수를 결정하여야 한다. 기법의 특성상 특정 군집수가 연구 결과를 잘 반영한다고 결론 지을 수 있는 근거가 없으므로, 본 연구에서는 다양하게 군집수를 적용하였다. 즉, 군집수를 10, 20, 30, 40, 50, 으로 나누어 적용하였다. 이렇게 군집수를 미리 정의하고, 군집화를 수행하게 되는데 개략적인 과정은 [그림 4]와 같이 군집화에 필요한 속성(지가, X·Y좌표)을 먼저 추출하고 Clementine 소프트웨어를 이용하여 군집화를 수행하였다. 그리고, Arc/GIS를 이용하여 시각화 및 통계치 분석을 하였다.



[그림 4] 유사가격권 설정을 위한 군집화 과정

3.2 결과 분석

결과의 분석은 각 구간별 군집화의 시행 결과가 현재 유사가격권과 어느 정도 반영하고 있는가를 분석하기 위해 군집내 개별 필지의 비교표준지 수를 분석하였다. 그리고, 가장 잘 부합한다고 판단되는 군집 구간을 선별하여 구간별 각 군집이 포함하고 있는 필지 수, 지가 평균, 표준편차 등을 분석하였다. [그림 5]는 군집수를 20개로 하였을 경우에 군집화가 이루어진 예이다.



[그림 5] 군집화 결과(k=20)

<표 1>은 구간별 각 군집내 표준지수와 개별필지의 비교표준지 수를 나타낸 것이다. 결과를 분석해 보면, 군집구간을 10으로 정의하였을 경우, 개별필지의 비교표준지가 가장 많이 분포하고 있고, 군집수를 증가시켜 정의할 경우, 상대적으로 비교표준지가 감소하는 것으로 나타났다. 이것은, 군집수에 따라 상대적으로 포함할 수 있는 필지 수가 감소하기 때문으로 분석된다.

<표 2>는 일치율이 높다고 판단되는 군집구간(10, 20)을 선별하여 포함 필지수, 평균 지가, 표준편차를 산정한 것이다. 결과를 분석해 보면, 20구간의 군집들은 대체적으로 10구간에 비해 가격편차가 적은 것으로 나타났다. 이것은 군집이 늘어날 수록 필지를 포함할 수 있는 범위는 감소하게 되기 때문으로 분석된다. 그리고, 주변 지

가와 지가 불균형이 발생하는 필지들은 대체적으로 군집형성이 되지 않거나 아주 적은 필지들로 군집화가 이루어지는 것으로 나타났다.

<표 1> 구간별 각 군집내 개별필지의 비교표준지 분포 현황

군집구간(수)	필지 수	표준지 수	비교표준지 수	일치율 (%)
10	1,973	80	1,637	86.5
20	1,973	80	1,435	75.81
30	1,973	80	1,182	62.44
40	1,973	80	929	49.08
50	1,973	80	836	44.16

4. 결론

본 연구에서는 현행 모호한 규정으로 많은 문제점을 발생시키고 있는 유사가격권 설정의 문제를 해결하고자 K-평균 군집화 기법을 활용하여 가격권을 설정해본 결과

다음과 같은 결론을 얻을 수 있었다. 먼저, 유사가격권에 대하여 가격 수준별로 정의가 가능하였다. 그리고, 군집내 개별필지와 일치하는 비교표준지를 조사한 결과 군집수를 대략 10으로 설정할 경우 86.5%가 일치하여 군집화의 타당성을 제시할 수 있었다. 그리고, 주변지에 비해 가격이 매우 고지가인 필지들은 군집화가 이루어지지 않거나 아주 작은 수로 군집화가 이루어지는 것을 분석할 수 있어 지역적으로 문제가 되고 있는 지가불균형 지역을 관리하거나 해결하는데 효율적으로 이용될 수 있을 것이다. 또한 개별필지의 유사가격권내 표준지들을 정확하고 효율적으로 확인이 가능하기 때문에 정확한 비교표준지 선정이 이루어질 수 있을 것이다.

다만, 군집화된 결과를 비교하는 과정에서 각 필지별 비교표준지들이 해당 소관청의 지적공무원을 통해 나온 결과이므로 본 연구의 결과를 이와 비교한다는 데 있어 어느 정도 적절한지에 대한 기준이 모호하다.

<표 2> 구간별 군집내 필지 수, 평균, 표준편차 산출 결과

구분	k=10			구분	k=20		
	필지 수	평균	표준편차		필지 수	평균	표준편차
1군집	107	4000560	194474	1군집	66	4149848	96475
2군집	1	9500000	0	2군집	0	9500000	0
3군집	4	6587500	146704	3군집	3	5793330	58891
4군집	5	7200000	413521	4군집	1	7050000	0
5군집	395	2181443	218309	5군집	236	2147711	98410
6군집	24	4960833	156149	6군집	2	5390000	110000
7군집	3	8426666	278248	7군집	1	8400000	0
8군집	171	3328011	217698	8군집	121	3319752	118436
9군집	1222	1522242	172860	9군집	587	1537819	55057
10군집	41	4520487	103545	10군집	22	4921818	85047
				11군집	1	6450000	0
				12군집	360	1797944	82648
				13군집	76	3706184	102955
				14군집	3	7500000	0
				15군집	96	2597500	145021
				16군집	1	8970000	0
				17군집	1	8780000	0
				18군집	39	4531538	93576
				19군집	1	8100000	0
				20군집	355	1307549	63741

따라서, 연구에서 설정한 군집수(10개)가 개별필지의 비교표준지를 상당수 포함하고는 있지만 군집결과가 절대적인 기준이 될 수 없을 것이다. 군집수 결정에 관한 기준 정립은 향후 다양하고 광범위한 연구지역에 대한 실험을 통하여 결정되어야 할 것이다. 아울러 유사가격권 설정 연구에 지가와 거리라는 두 가지 인자를 사용했는데 이외에 영향을 줄 수 있는 다른 인자에 대한 연구도 함께 병행되어야 할 것이다.

참고문헌

- 건설교통부, 2003, 2003년도 적용 개별공시지가 조사·산정 지침.
- 국토개발연구원, 1997, “공시지가의 균형성 제고 방안”, 국토연 97-23
- 구자훈·김성희, 1999, “GIS를 활용한 개별공시지가 산정 및 도로개설에 따른 토지보상비 산정 방법론”, 한국 GIS 학회지, 제7권 제1호.
- 김성희·정병호·김재경, 2002, 의사결정 분석 및 응용, 영지문화사
- 김윤식·모경주·윤인섭, 2000, “클러스터링 기법을 이용한 공정 데이터의 압축 저장 기법에 관한 연구”, 한국가스학회지, 제4권 제4호, pp.58-64.
- 박수홍·홍성언·김현석·김정엽, 2003, “공간 다기준 의사결정 방법을 이용한 개별공시지가 비교표준지 선정”, 한국GIS학회지, 제11권 제1호, pp.1-11.
- 박성규, 1999, “토지 평가의 자동화를 위한 GIS의 적용에 관한 연구”, 조선대학교 대학원 박사학위 논문.
- 박정호, 1999, “공시지가제도에 관한 연구”, 동의대학교 대학원 석사학위 논문
- 안종욱, 2000, “개별공시지가의 효율적인 산정을 위한 GIS 적용에 관한 연구”, 건국대학교 행정대학원 석사학위, pp.40~43.
- 이근수·김삼근, 1991, “군집화 알고리즘에 관한 고찰”, 안성대학교 산업대학원 논문집, 제23권, pp.175-185.
- 이효상, 2001, “개별공시지가산정을 위한 토지특성테이블 구축에 관한 연구”, 명지대학교 산업대학원 석사학위 논문.
- 조형기·민준영·최종욱, 1996, “클러스터링 방법을 이용한 차종인식 모형”, 한국정보처리학회 논문지, 제3권 제2호, pp.369-380.
- 홍길순, 1998, “개별공시지가 제도의 발전 방향에 관한 연구”, 중앙대학교 대학원 석사학위 논문
- Rand, W. M., 1971, “Objective Criteria for the Evaluation of Clustering Methods”, Journal of American Statistical Association, Vol.66, pp.846-850.