

GIS와 공간데이터마이닝을 이용한 상업시설물의 입지패턴 분석

홍성언*, 이용익**

*청주대학교 지적학과

**지능형국토정보기술혁신사업단

*e-mail:hongsu2005@cju.ac.kr

Analysis of Commercial Facility Locational Pattern Using GIS and Spatial Data Mining

Sung-Eon Hong*, Yong-ik Lee**

*Cheongju University

**Korean Land Spatialization Group

요 약

입지분석은 공간 및 비공간적 특성이 중요하게 다루어져야 함에도 불구하고 공간데이터 타입(spatial data type), 공간관계(spatial relationship), 그리고 공간 자기상관성(spatial autocorrelation)의 복잡성에 기인한 처리의 어려움으로 인해 기하학적거리나 공간적 위치와 같은 단순 공간적 특성만 이용되었다.

본 연구에서는 서울시 대형할인점을 사례로하여 GIS에 의한 공간데이터와 비공간데이터(인구통계 등)를 통합 구축한 후, 공간데이터마이닝 기법을 이용하여 입지패턴(location pattern)을 분석·추출하여 보고자 한다.

1. 서론

일반적으로 상업시설물의 입지를 선정하기까지는 우선 매력적인 상권을 확인하고, 그 상권 내에서 가장 매력적인 입지들을 찾으며, 매력적인 입지들 가운데서 가장 최고의 입지를 선정하는 과정을 밟게 된다. 상권을 선정하기 위해서는 무수히 많은 요인에 중에서 입지에 의해 결정되는 공간적 특성은 매우 중요하게 다루어야 할 요소이다. 왜냐하면 공간적인 요소는 변경이 불가능하기 때문에 사전에 충분한 분석을 통해 결정되어야 한다.

현재까지 입지분석에 관한 연구의 경우, 통계자료와 현장조사를 통하여 입지이론 적용과 회귀분석이나 요인분석에 의한 통계 연구가 주로 이루어졌다. 그러나 최근 GIS 기술의 발달로 인해 공간데이터와 비공간데이터를 동시에 처리할 수 있게 됨에 따라 도시 상업 시설의 입지분석은 새로운 공간분석의 대상으로 많은 연구자에 의해 관심을 받고 있다.

하지만 입지분석은 공간적 특성이 중요하게 다루어져야 함에도 불구하고 공간데이터 타입, 공간관계(spatial relationship), 그리고 공간 자기상관성(spatial autocorrelation)의 복잡성에 기인한 처리의 어려움으로 인해 기하학적 거리(euclidean distance)나 공간적 위치와 같은 단순 공간적 특성만 이용되었다. 또한 의사결정을 지원함에 있어서 불확실성(uncertainty)을 배제하기 위해서 동종의 상업시설에 대하여 명확하고 강력한 규칙의 발견이 필요함에도 불

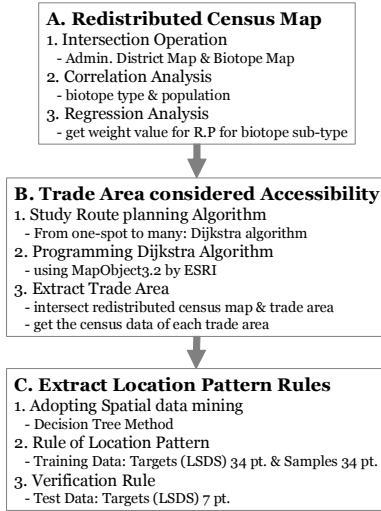
구하고, 지금까지 소개된 방법론에서는 공간데이터와 비공간데이터를 모두 의미 있게 동시에 적용한 사례가 없어, 공간적 요소와 비공간적 요소(인구통계 등)를 모두 적용한 효율적인 입지분석 체계가 요구된다.

본 연구에서는 서울시 대형할인점을 사례로하여 GIS에 의한 공간데이터와 비공간데이터(인구통계 등)를 통합 구축한 후, 공간데이터마이닝 기법을 이용하여 입지패턴(location pattern)을 분석·추출하여 보고자 한다.

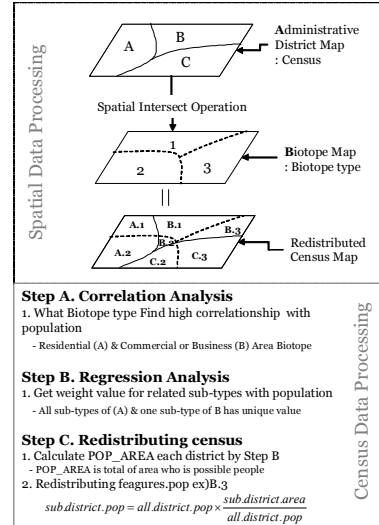
2. 입지분석 설계

본 연구에서는 대형할인점(Large-scale Discount Store : LSDS)에 대한 입지분석을 위해서 크게 3부분으로 나누어 입지분석 설계를 하였다. 첫째, 입지분석의 정확성을 높이기 위하여 행정구역 단위의 인구수를 정밀한 단위로 재분배한다. 둘째, 대형할인점의 상권을 결정함에 있어서 고객과의 접근성을 고려하기 위하여 도로의 연결성 정보를 활용한 경로탐색 알고리즘을 적용한다. 마지막으로 앞선 과정을 통하여 산정할 수 있는 점포별 데이터를 이용하여 공간데이터마이닝 기법을 적용한 후 의미있는 규칙 추출을 통하여 대형할인점의 입지패턴을 분석하여 보고자 한다[그림 1].

Locatoinal Analysis Methodology



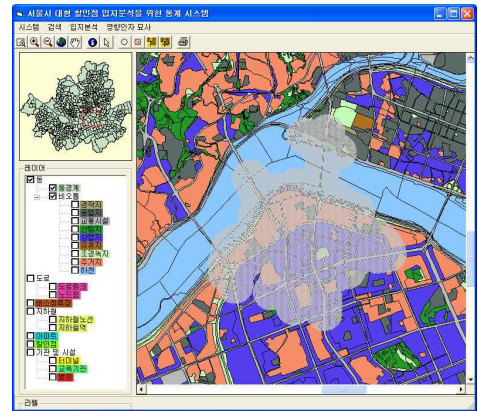
[그림 1] 입지분석 과정



[그림 2] 정밀한 단위로 인구재분배를 통한 인구수 산정 과정

구체적으로, 인구수의 산정과 관련하여서는 선행 연구와는 다르게 도시생태현황도와 행정구역도 등을 이용하여 정밀한 상권 인구 산정방법을 제시하고자 한다. 지금까지 입지분석을 위한 상권의 잠재고객을 추정할 하기 위해서는 상권의 크기에 비해 넓은 행정구역인 구나 동에 대하여 상권이 포함되는 면적을 이용한 단순 면적비례방법을 이용함에 따라 비합리적인 인구추정치가 되었다. 그러나 지리적으로 동질지역을 제외하고는 전 면적이라고 하는 지역 내부에는 서로 다른 여러 개의 이질지역으로 구성되어 있기 때문에 실제적인 값을 반영하지 못한다. 따라서 입지분석에 이용하는 상권의 인구추정치는 이러한 이질지역을 고려하여 인구를 추정하여야 한다. 본 연구에서는 인구지리학에서 말하는 경지인구밀도(physiological density of population)의 개념을 적용하기 위하여 도시생태현황도를 이용하여 유형별 생태현황과 인구와 상관관계를 분석하고, 상관성이 높은 비유흥 유형에 대하여 세부항목별 회귀분석을 이용하여 인구를 추정하였다. 과정은 [그림 2]와 같다.

상권분석 역시 기존의 원형상권 분석이 아니라, 상권내 잠재고객이 점포에 접근하기 위한 도로의 연결성을 반영하여 실질적인 상권을 결정할 수 있는 방법을 제시하였다. 즉, 경로탐색 알고리즘인 다익스트라 알고리즘(dijkstra algorithm)을 도로사상(road feature)에 적용하여 대상 대형할인점을 기준으로 1.7Km의 경로에 0.3Km의 버퍼를 생성한 후, 이 부분에 포함되는 범위에 대해서만 1차 상권으로 제한하였다. [그림 3]은 연구의 방법론을 적용하여 결정된 상권의 예이다.



[그림 3] 접근성을 고려한 상권 분석

3. 입지패턴 추출 및 검증

일반적인 데이터마이닝은 여러가지로 정의를 내리고 있는데, 간단히 말하면 대량의 데이터로부터 지식(knowledge)을 추출하는 일련의 과정을 말한다. 즉, 공간데이터 마이닝은 대용량의 공간 데이터베이스(spatial database)로부터 잠재되어 있는 유용한 패턴이지만 이전에 알려지지 않은 정보를 발견하기 위한 지식발견기법 이다. 공간데이터 타입, 공간관계, 그리고 공간 자시상관성(autocorrelation)의 복잡성 때문에 관심 있고 유용한 정보를 추출하는 것이 전통적인 데이터마이닝 기법보다 공간데이터 마이닝 기법이 더 어렵다.

본 연구에서는 입지분석에 있어서 기존의 통계학적 접근이 아닌 최근 대두되는 공간데이터마이닝 기법 중 의사결정 트리분석을 이용하여 서울시 대형할인점을 사례로 하여 입지패턴을 추출하였다. 입지패턴을 추출함에 있어 CART(Classification and Regression Tree), CHAID(Chi-squared Automatic Interaction Detection), C4.5와 같은 다양한 의사결정트리 알고리즘을 이용하였다.

[표 1] 입지분석 변수

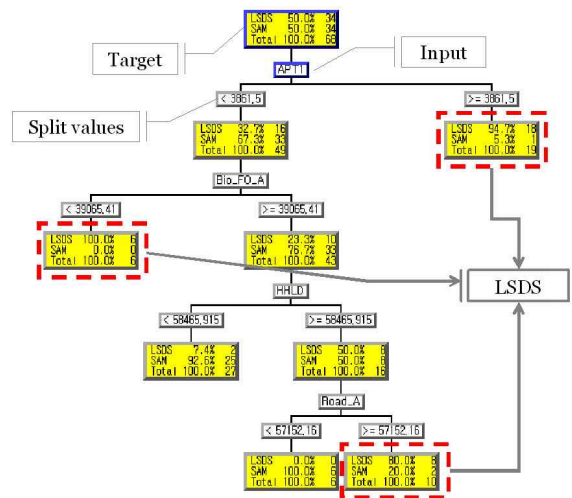
Name	Measurement	Note or Value
TARGET	Binary	LSDS/SAMPLE
BIO_RE_A	Interval	Total area of Biotope type A in trade area
BIO_BU_A	Interval	Total area of Biotope type B in trade area
BIO_IN_A	Interval	Total area of Biotope type C in trade area
BIO_TR_A	Interval	Total area of Biotope type D in trade area
BIO_GR_A	Interval	Total area of Biotope type E in trade area
BIO_FA_A	Interval	Total area of Biotope type F in trade area
BIO_ST_A	Interval	Total area of Biotope type G in trade area
BIO_FO_A	Interval	Total area of Biotope type H in trade area
BIO_UN_A	Interval	Total area of Biotope type I in trade area
POP	Interval	Total population in trade area
HHL D	Interval	Total household in trade area
APT_H_C	Interval	Total APT's household in trade area
CAR	Interval	Total car in trade area
SUM_IN	Interval	Total income in trade area
SUM_FOOD	Interval	Total food-expenditure
SUBWAY	Binary	Exist subway station or not in the 500m of radius
BUS	Interval	Total bus stop in the 500m of radius
EDU	Interval	Total education facility in the 500m of radius
HS	Interval	Total hospital facility in the 500m of radius
ROAD_A	Interval	Total Road area in the 500m of radius
APT_1	Interval	Total APT's household in the 500m of radius

그리고 위에서 구축된 정밀단위로 재분배된 인구통계 데이터와 접근성을 고려한 상권분석 데이터를 기초로 서울시 대형할인점 총 34곳의 데이터 셋을 이용하여 추출하였다. [표 1]은 입지분석 변수를 나타낸 것이다.

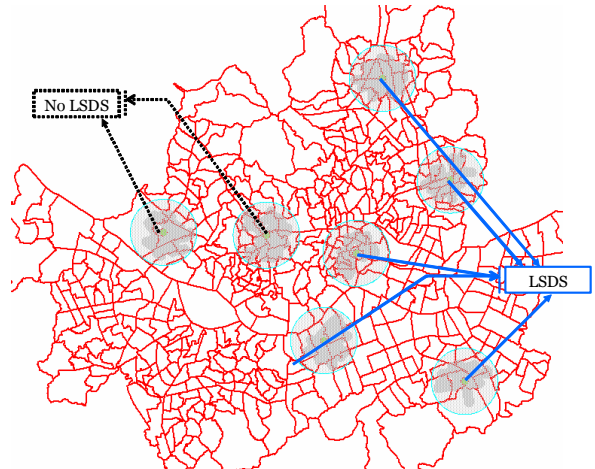
[그림 4]는 의사결정트리 분석을 이용한 대형할인점의 입지패턴을 분류하여 본 것으로 최종 노드의 결과(5개)가 의사결정트리 분석을 이용한 규칙이 된다. 이 중에서 점선으로 박스처리가 된 부분이 대형할인점의 입지패턴을 보여주는 규칙이다. 점선으로 표현된 규칙 중 첫 번째 규칙은 APT1이 3861.5 이상일 경우 대형할인점으로 분류가 됨을 의미한다. 두 번째 규칙은 APT1이 3861.5 보다 작으면서 Bio_FO_A가 39065.41 보다 작을 경우 대형할인점으로 분류되는 규칙을 보여준다. 마지막 규칙은 APT1이 3861.5 보다 작으면서 Bio_FO_A는 39065.41 이상이고, HHL D는 58465.915 이상, Road_A는 57152.16 이상일 경우 대형할인점으로 분류됨을 보여준다. 이상의 규칙을 통해 대형할인점을 선정하는 입지패턴을 확인할 수 있다.

34개 데이터셋으로부터 추출된 규칙은 검증 과정이 필요하다. 그래서 본 연구에서는 서울시의 대형할인점 47개 중에서 서울시 경계 지역에 위치한 6개의 할인점을 제외하고 7개를 검증 데이터 셋을 선정하였다.

검증은 E-Miner에서 제공하는 Assessment, score, insight node를 이용하여 실시하였다. 그 결과 7개의 검증 데이터 셋 중에서 5개가 대형할인점으로 분류되는 것을 확인할 수 있었다[그림 5].



[그림 4] 의사결정트리 분석을 이용한 대형할인점의 입지패턴 분류



[그림 5] 검증 데이터 셋의 위치(대형할인점 7곳)

4. 결론

본 연구에서는 서울시 대형할인점을 사례로하여 GIS에 의한 공간데이터와 비공간데이터(인구통계 등)를 통합 구축한 후, 공간데이터마이닝 기법을 이용하여 입지패턴(location pattern)을 분석·추출하여 보고자 하였다.

이를 위해 기존 입지분석과는 다르게 인구수 산정이나 상권분석에 있어 정확도와 신뢰성을 향상시키기 위하여 새로운 방법론을 제시하였다. 즉, 정밀 인구수 산정방법, 접근성에 기초한 상권결정방법 등을 제시하였다.

그리고 상기와 같이 구축된 데이터를 기반으로 서울시 대형할인점의 입지패턴을 분석하여 보고자 대형할인점 총 34곳의 데이터 셋을 구축한 후, 공간데이터마이닝을 이용하여 입지패턴을 추출·분석하여보았다.

결과, 입지분석 변수 중 ATP1가 3861.5이상일 경우, APT1가 3861.5 보다 작으면서 Bio_FO_A가 39065.41 보다 작을 경우, APT1가 3861.5 보다 작으면서 Bio_FO_A는 39065.41 이상이고, HHLD는 58465.915 이상, Road_A는 57152.16 이상일 경우 대형할인점으로 분류됨을 알 수 있었다.

연구에서 추출한 입지패턴의 타당성을 검증하고자 서울시 대형할인점 중 7곳을 선정하여 실제 대형할인점으로 분류가 되는지를 확인한 결과, 7곳의 대형할인점 중 5곳이 대형할인점으로 분류되는 것을 확인할 수 있었다.

연구에서는 서울시의 대형할인점을 사례로 입지분석을 시행하여 보았으나 연구의 방법론은 지역 또는 대형할인점 외에 다른 상업시설물에도 효율적으로 적용이 가능할 것으로 판단된다.

참고문헌

- [1] 권용걸, 강양석, “대형할인점 입지 결정 요인에 대한 연구”, 국토계획, 제37권 제1호, pp. 207-217, 2002.
- [2] 백영기, “소매중심지 입지를 위한 GIS기반의 공간적 의사결정지원시스템”, 대한지리학회지, 제36권 제3호, pp. 278-291, 2001.
- [3] 이상규, “대형할인점의 매출액 결정에 있어서 입지요인의 영향에 관한 연구”, 석사학위논문, 인하대학교 대학원, 2004.
- [4] 이희연, 김지영, “대형할인점의 입지적 특성과 상권 분석에 관한 연구”, 국토계획, 제35권 제6호, pp. 61-80, 2000.
- [5] 조혜중, 새인구론: 인구의 공간적·사회적 접근, 푸른길, 2006.
- [6] 최대식, “대형할인점의 입지특성에 관한 연구”, 석사학위논문, 서울대학교 대학원, 1998.
- [7] Dunne, P.M., Lusch, R.F., Retailing (5th edition), South-Western College Publisher, 2004.
- [8] Hartshon T.A, Interpreting the City: An Urban

Geography, John Willey & Sons, Inc., New York, p. 387, 1992.

- [9] Han, J., Kamber, M., Data Mining: Concepts and Techniques, The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor Morgan Kaufmann Publishers, USA, 2000.
- [10] Shekhar S, Zhang P, Huang Y, Vatsavai R, Trends in Spatial Data Mining, as a Chapter of Book: Data Mining: Next Generation Challenges and Future Directions by H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha(eds.), AAAI Press, 2004.
- [11] Shekhar, S., Chawla, S., Spatial Databases: A Tour, Prentice Hall, 2003.
- [12] Zhou, T., “Show Me the Location: A GIS Approach on Discount Store Location Study”, ESRI Proceeding 98, pp. 1-22, 1999.