

## 공간 연관규칙을 이용한 대형할인점의 입지 분석

이용익\* · 홍성언\*\* · 김정엽\*\*\* · 박수홍\*\*\*\*

### Analyzing the Location Decision of the Large-Scale Discount Store Using the Spatial Association Rules Mining

Yong-Ik Lee\*, Sung-Eon Hong\*\*, Jung-Yup Kim\*\*\*, and Soo-Hong Park\*\*\*\*

**요약 :** 본 연구의 목적은 1990년대부터 급속히 증가한 대형할인점에 대하여 입지영향인자를 추출하여 의사결정에 객관성을 확보하고 대량의 데이터베이스를 이용하여 숨겨진 유용한 정보를 입지 선정에 활용하는 것이다. 이를 위해 대형할인점이 입점하는데 미치는 인구학적 변수, 경제학적 변수 그리고 주변환경적 변수에 대한 다양한 인자를 통계자료를 수집하고 연구대상 지역의 공간 자료를 구축하여 공간 연관성 분석을 실시하여 공간 연관규칙을 추출하였다. 결과의 검증을 위해 추출된 규칙과 대형할인점의 매출액을 이용한 적용성 여부를 상호 비교하였다. 검증 결과 추출된 공간 연관규칙이 해당 대형할인점에 많이 부합할수록 매출액도 많은 것으로 나타났다. 본 연구를 통해 공간 연관규칙을 활용하여 객관적이고 매출에 이익을 주는 대형할인점의 최적입지 선정을 기대할 수 있다.

**주요어 :** 공간 데이터마이닝, 연관규칙, 입지 선정, 대형할인점

**Abstract :** The objective of this research is to achieve an objectivity of site decision after extracting site decision factors on a large-scale discount store(LSDS) and utilize any hidden information using the association rules mining through huge database. To catch this objective, we collect a census, economic , and environmental dataset related with locating of LSDS. And then, we construct a spatial data on the research area. These data is used for the extraction of a spatial association rules. To verify whether the extracted rules are suitability or not, we use the sales of some LSDS. As the result of test, the more sales, the more factors of the extracted rules relate with the sales it coincides. Consequently, the spatial association rules mining is efficient method which support the ideal site decision of LSDS.

**Key Words :** spatial data mining, association rules, site decision, large-scale discount store

\* 인하대 지리정보공학과 석사과정(Graduate Student, Dept. of Geoinformatics, Inha University), a78leekey@naver.com

\*\* 청주대 지적학과 전임강사(Full-time lecturer, Dept. of Welfare · land information, Cheongju University), hongsu2005 @hanmail.net

\*\*\* 인하대 지리정보공학과 박사과정(Ph.D. Candidate, Dept. of Geoinformatics, Inha University), jyfloo@empal.com

\*\*\*\* 인하대 지리정보공학과 부교수(Professor, Dept. of Geoinformatics, Inha University), shpark@inha.ac.kr

## 1. 서론

경영전략의 혁신을 통해 유통비용을 줄임으로서 기존의 소매업체와는 비교가 안 되는 저가판매 전략을 내세운 대형할인점<sup>1)</sup>이 1993년 11월 E-마트 창동점(서울시 도봉구 소재)을 시작으로 나타났고 소비자들은 이를 상당히 선호하였다. 이러한 연유로 국내 대기업 주도의 대형 유통업체와 외국의 유통업체 등 대차본을 중심으로 유통혁명을 주도하는 대형할인점은 1996년 유통 서비스 시장의 개방으로 까르푸를 필두로 국외 유통 업체들과 국내 유통업체들이 지방의 중소도시에 까지 진출하였다.

국내 대형할인점의 성장현황을 살펴보면, 1990년대 초 전통적인 유통경로인 제조 → 도매 → 소매 → 소비자의 체계가 점차 쇠퇴하면서 도매상과 독립소매점이 감소하고 새로운 형태의 매장, 즉 대형할인점, 편의점, 도·소매 체인점, 카테고리킬러 등의 신업태가 등장하면서 유통경로가 다변화되었다. 산업자원부의 2003년 「유통업계 전망 조사」 자료에 따르면, 국내 대형할인점은 1993년 개점한 이후, 성장세를 이어가고 있는데, 점포수의 기준으로 1999년에 이미 백화점과 추월하였고, 2001년에는 백화점 점포수의 2배를 초과하였다. 2003년에는 대형할인점의 전체 매출액이 백화점을 추월하여 유통업계 1위를 달성한 것을 확인할 수 있다. 그리고 2005년도 유통업계 전망 조사에서도 인터넷쇼핑 다음으로 7.2%의 고속 성장세를 이어 갈 것으로 예측하고 있다.<sup>2)</sup> 하지만 대형할인점에 대한 유통활동 특성과 입지적 특성을 고려하지 못한 채 과거의 소매업체에 대한 인식의 법적 근거와 기업의 근시안적인 의사결정에 의해 대형할인점의 입지가 선정되는 경우가 많다. 이로 인해 신도시 개발과 더불어 경쟁적으로 출점함으로서 특정지역에 점포의 과포화나 과잉 투자는 유통업체의 부담으로 작용하고 있다.

따라서 본 연구의 목적은 동(洞)별 통계자료와 공간 자료를 활용하여 공간 연관성 분석을 통해 대형할인점의 입지에 영향을 미치는 공간 연관규칙을 추출하는 것이다. 그리고 추출된 규칙과 대형할인점의 매출액을 비교하여 공간 연관성 분석을 통한 입지분석이 대형할인점의 최적의 입지를 선정하는데 있어서 의사결정지원

의 도구로 활용할 수 있는지에 대한 검증을 실시한다.

세부연구방법은 다음과 같다. 첫째, 국내외에서 선행된 대형할인점 입지분석 관련 연구를 통하여 입지선정 시 일반적으로 고려되는 인자가 어떤 것이 있는지를 확인하고, 동(洞)별 통계자료의 획득방법과 공간자료의 현황을 파악한다.

둘째, 대형할인점의 입지분석 실험을 위해 사례지역으로 서울시 강서구, 양천구, 영등포구를 선정하였다. 대형할인점 상권 내 입지영향인자 추출을 실험하기 위해 공간상의 점포별 상권을 결정하고, GIS를 이용하여 선택된 자료를 원시형태에서 실험에 사용될 통계자료와 공간자료를 구축하였다. 본 연구에서는 기존연구의 분석에서 제시한 점포를 중심으로 반경 2km의 존(zone)을 형성하여 자료들을 구축하였다.

셋째, 구축된 통계자료와 공간자료를 공간 연관성 분석을 위한 데이터 셋(set)으로 변환하고 이를 이용하여 대형할인점의 입지영향인자를 추출하기 위한 공간 연관성 분석을 하여 규칙을 추출한다.

넷째, 제시된 규칙들 중 원인과 결과가 서로 의미상 중복되는 것을 제거한 후 유용한 규칙을 선정한다.

마지막 검증 단계에서는 선별된 규칙을 바탕으로 대형할인점의 입지현황이 규칙에 얼마나 적절하게 입지하고 있는지 여부를 확인하고 해당 대형할인점의 매출액과 비교하여 실험결과를 검증한다.

## 2. 대형할인점의 연구동향과 공간 연관성 분석

### 1) 대형할인점의 연구동향

1990년대 중반부터 급속하게 증가하고 있는 대형할인점의 입지분석에 대해서 통계학, 경영학, 지리학, 도시공학 등 다양한 분야의 각 연구자들은 전문지식과 입지이론을 토대로 설명하고, 소매업 및 대형할인점의 최적입지선정에 대해서 연구를 하였다.

대형할인점의 분포패턴을 분석하고, 통계기법을 이용하여 입지요인을 설명한 연구(Zhou, 1999)는 대형

할인점의 입지 특성을 규명하기 위하여 공간적 상호관련성이 중요하다고 설명하고 있다. 공간적 확장이론과 일반 통계분석을 이용하여 입지특성을 규명한 최대식(1998)은 통계자료와 이론적 배경이 부족하여 단순 추측이지만 입점초기의 입지특성은 공간적 측면에서 중요하다는 점을 지적하였다. 이희연·김지영(2000)은 인자분석을 적용한 접근방법은 입지특성과 입지유형을 규명하였으나, 대형할인점의 입지특성을 반영하지 못한 일반적 특성만을 추출하고 GIS에 기반을 둔 공간적 상호작용 모델을 활용하여 각 상점의 매출액을 예측하였다. 응용된 인자분석을 통해 인구와 관련된 인자를 주요 인자로 밝힌 권용걸·강양석(2002)의 연구는 대형할인점과 연관되는 핵심적인 입지 인자를 제시하지는 못하였다. 다중회귀분석으로 입지요인을 분석한 이상규(2004)는 기존의 연구에 매출액이라는 대형 할인점의 성공적인 요소와 연관지어 입지 요인을 규명하려고 하였으나, 다중회귀분석의 이론적 취약점과 그 결과를 해석하는데 있어 인과관계를 혼돈할 가능성이 있다. 그리고 통계자료를 수집하는데 있어서 실제 상권의 범위(반경 2km)보다 큰 구(區)별 단위로 하여 면적별 가중치를 적용하는 방법을 사용함으로써 실제 지역에 거주하는 인구수를 고려하지 못하는 통계자료를 이용하였다.

기존 대형할인점의 입지분석과 관련한 연구의 경우, 여러 통계 분석 방법을 이용하여 입지 특성 분석에 관한 연구가 이루어졌으나 대체적으로 일반적인 특성분석에 국한되었다는 한계성이 있다. 또한, 일반 소매점에 비하여 대형할인점의 상권이 비교적 넓기 때문에 다양한 인문·사회적 통계자료를 구(區)별 단위로 구축하여 이용하였으나 이는 동일한 행정구역내에서도 동(洞)별 단위의 통계자료 수치가 면적 비율로 균일하게 적용되지 않기 때문에 올바른 통계자료의 사용이라고 보기에는 한계성이 있다. 그리고 기존의 연구에서는 공간자료를 이용한 분석에 있어서 소비의 주체가 되는 인구수를 가지고 거리 가중치를 고려하였기 때문에 소비자가 대형할인점을 이용하는 접근성(accessibility)에 대한 논의가 부족하였다. 따라서 본 연구에서는 동(洞)별 통계자료와 실제 상권의 자료를 GIS기반의 공간분석에 적용하여 만든 공간자료를 접근성을 고

려한 공간분석 데이터로 구축하여 공간 데이터 마이닝 기법인 공간 연관성 분석을 하여 대형할인점의 입지에 대한 공간 연관규칙을 도출하고자 한다.

## 2) 공간 연관성 분석

최근 전통적인 데이터마이닝<sup>3)</sup> 기술은 공간 데이터를 이용한 공간 데이터마이닝 기술로 확대되고 있다. 이는 대용량의 공간 데이터베이스로부터 어떠한 현상에 대하여 알려지지 않은 유용한 패턴들을 찾기 위한 프로세스이다(Shekhar, et al., 2001). 공간 데이터 탐색의 복잡성과 공간 객체들 사이의 다양한 관계성이 존재하기 때문에 공간 데이터마이닝은 전통적인 데이터마이닝 기술보다 패턴을 추출하는 과정이 어렵다(Shekhar, et al., 2001; Shekhar, et al., 2003). Koperski와 Han(1995)은 공간 데이터마이닝 기술들 중 공간 데이터베이스에서 공간·비공간 데이터들 간의 연관규칙(Agrawal, et al., 1993)을 추출하기 위한 공간 연관성 분석을 제시하였다. 이는 SQL(Structured Query Language)과 같이 공간 연관규칙에 사용되는 질의는 ‘교차한다(intersect)’, ‘이웃한다(adjacent)’ 등과 같은 관심 술어를 이용한다. 왜냐하면, 공간 정보는 기하학적 정보와 위상 정보를 가지고 있는데 공간 연관성 분석은 위상적 정보를 다루어 공간 연관규칙을 추출하기 때문이다. 기하학 정보로는 위치, 면적, 길이 등 정량적인 정보로 이루어져 있고, 위상적 정보에는 인접성(adjacency)(예를 들면 인도 A와 차도 B는 이웃한다), 계급성(hierarchy)(예를 들면 운동장 A는 학교 B의 내부에 있다) 등과 같은 객체간의 상대적 관계 정보를 다룬다.

GIS에 기반을 둔 공간분석을 통해 ‘객체 A와 B는 인접해 있다’라는 인접성을 밝히기 전에 ‘인접해 있다’라는 것을 공간적 술어를 사용하기 위해서는 인접한 정도를 프로세스화 하기 전 단계에서 조건으로 입력해야 한다. 그래서 공간적 제한성에 대하여 조사·연구한 후 공간적 위상 관계들에 대해서 정의할 수 있다. 하지만 공간 연관규칙의 경우 객체간의 거리의 정보를 이용하여 프로세스 내에서 신뢰도와 지지도를 바탕으로 규칙을 정의할 수 있기 때문에 수집된 공간 데이터

들 사이에서 일정한 수준의 인접성과 같은 위상 관계를 정의하여 이용할 수 있다(Koperski and Han, 1995).

공간 연관규칙은  $X \rightarrow Y$ 의 형태로 표현되며, X와 Y는 공간 솔어(adjacent\_to, near\_by, inside 등)를 포함한 조건식들의 집합이다. 예를 들어, “주유소는 고속도로에 인접해 있다 {is\_a(X, gas station)} \rightarrow near\\_by(Y, highway)”, (confidence:65%)”가 공간 연관규칙을 적용한 결과이다. 또 다른 예로, “작은 도시가 좋은 지형적 조건을 가진다면, 대부분 도로를 포함한다. (is\_a(X, small city has good terrain) \rightarrow near\_by(Y, road), (confidence:65%))”라는 규칙을 찾을 수 있다. 지지도(support)<sup>(4)</sup>와 신뢰도(confidence)<sup>(5)</sup>는 발견된 규칙의 성립률을 결정할 수 있는 인자이다.

이처럼 공간 연관성 분석은 공간상에서 발생한 사건에 대하여 비즈니스 데이터만을 이용하여 연관성을 분석하는 것뿐 만 아니라 공간 데이터를 이용하여 상호 연관성을 객관적인 측면에서 설명한다. 따라서 입지분석과 같은 다양한 요인에 의해 결정되는 의사결정지원에 있어서 숨겨진 연관성을 밝혀내어 의사결정자에게 유용한 정보를 제공할 수 있는 장점이 있다.

### 3. 대형할인점의 입지영향인자 추출

대형할인점의 입지 영향 인자를 추출하기 위한 연구 방법은 그림 1과 같다. 서울시 3개구에 위치하는 12개 대형할인점들로부터 반경 2km의 존(zone)을 형성한 후, 각 상권에 대해 행정구역도, 도로망도, 시설물도와 같은 공간자료와 인구관련, 경제관련 통계자료를 구축하였다.

구축된 자료들을 공간 연관성 분석에 적용하기 위해 변환 과정을 거쳐서 연관규칙을 발견하고 유용한 규칙을 선정하여 입지 영향 인자를 추출한다. 결과 검증을 위하여 주요 대형할인점의 2004년도 매출액 자료를 바탕으로 입지 인자와 매출액의 상관성을 비교 분석하였다.

### 1) 연구지역 선정

연구지역으로 선정된 강서 · 양천 · 영등포구는 서울 시로부터 뉴타운으로 지정되었다. 이를 기초자치단체는 인구증가율이 서울시 전체 평균 0.11%에 비해 강서구 0.94%, 양천구 1.71%, 영등포구 1.20%로 상당히 높은 지역이고, 서울시에서 많은 인구분포를 나타내는 지역이다(표 1).

또한 본 연구지역에는 12개의 대형할인점이 1994년 10월 영등포구 양평동에 코스트코 홀세일이 입점을 시작하여 2005년 9월 홈플러스 가양점이 입점하기 까지 대형할인점의 입점이 꾸준히 진행되고 있다.

본 연구대상지역에 대형할인점이 증가하는 이유는 서울시에서도 인구가 꾸준히 증가하고 인구도 많기 때문인데, 향후 이 지역에 대형할인점의 추가 입점이 예상되는 곳이다.

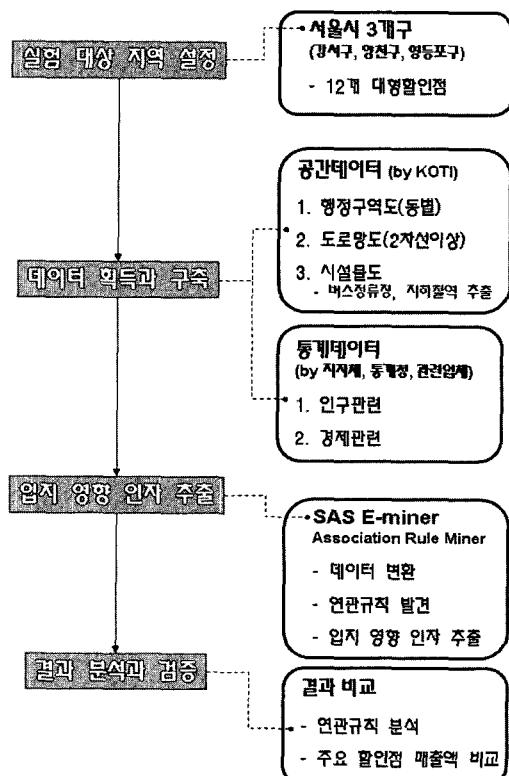


그림 1. 실험과정과 주요 내용

표 1. 연구지역 개요

	서울시	강서구	양천구	영등포구
인구(내국인)	10,287,847	541,671	495,098	406,261
면적(km <sup>2</sup> )	605.41	41.4	17.4	24.56
인구밀도 (1km <sup>2</sup> 당)	16,994	13,084	41,521	16,542
인구증가율(%)	0.11	0.94	1.71	1.20
가구수	3,780,305	193,973	169,326	154,167

자료 : 2004년, 서울시주민등록통계

## 2) 변수의 선정과 공간자료 구축

소매점의 성격을 바탕으로 입지에 영향을 미치는 다양한 인자를 공간자료와 통계자료로 구분한다. 공간자료는 입점한 대형할인점의 상권을 분석하기 위해 동별 행정구역도, 자동차 이용고객을 위한 2차선 이상의 도로망도, 대중교통수단이나 이용고객의 접근성을 확인 할 수 있는 버스 정류장이나 지하철역의 위치가 나타난 교통시설물도를 사용하였다.

### (1) 통계자료

대형할인점의 입지에 영향을 미치는 인자를 추출하기 위해서는 자료를 선별하고 주제별로 구분한다. 대형할인점의 규모는 일반 도매상점보다 크지만 판매의 성격이나 입점의 조건은 소매점의 성향이 강하다.

통계자료의 경우는 그 내용이 매우 광범위하고 구축 방법, 시기, 목적, 그리고 조사기관에 따라 다양한 자료를 획득할 수 있다. 본 연구에서는 소매점 형태의 대형할인점의 상권을 분석하기 위해서 동별로 구축된 자료를 통계청과 지방자치단체의 통계 목록을 확인하여

표 2. 동별 통계자료 변수명

그룹	변수명	개별 변수명
인구학적 변수	총인구, 총가구수, 면적, 아파트수(가구수)	
경제학적 변수	지방세 징수총액, 사업체수, 노동자수, 자가용 자동차 보유대수	
주변환경적 변수	재래시장수, 재래시장 면적, 병원수, 은행수	

은행, 시장, 병원, 인구, 가구 등의 인자를 선택하였다. 또한 한국체인스토어협회에서 발행하는 유통업체연감 등 문헌자료와 웹을 통해 표 2와 같이 인구학적 변수, 경제학적 변수, 주변환경적 변수로 구분하여 구축하였다.

인구학적 변수는 매출액과 직접 관련이 있는 대형할인점의 이용자 규모를 측정할 수 있고 소비활동의 근간이 되는 것들이다. 경제학적 변수는 실제 거주민들의 소득을 간접적으로 확인하여 경제 수준을 판단할 수 있는 근거이다. 마지막으로 주변환경적 변수는 대형할인점이 위치한 상권 내에 존재하는 상업활동을 지원할 수 있는 인자를 선택하였다.

### (2) 공간자료

본 연구에서 이용된 공간자료는 건설교통부와 교통개발연구원이 '국가교통DB구축사업'의 일환으로 제작한 것으로, 국토지리정보원의 항공사진을 기반으로 작성한 NGIS(국가기본 수치지도)를 수정·보완하여 교통관리·투자분석 등 교통목적으로 재구축한 축척 1/5,000의 교통주제도이다.<sup>6)</sup>

대형할인점의 상권분석도를 구축하기 위하여 연구지역의 행정구역도(a\_admin\_line), 도로망도(a\_road), 그리고 시설물도에서 추출한 버스정류장(a\_bus), 지하철역(a\_sub\_station)을 추출하였다. 행정구역도의 경우는 동별 행정구역도를 기준으로 구축되어 있어 다른 통계자료와 쉽게 관련지울 수 있다. 그리고 도로망도

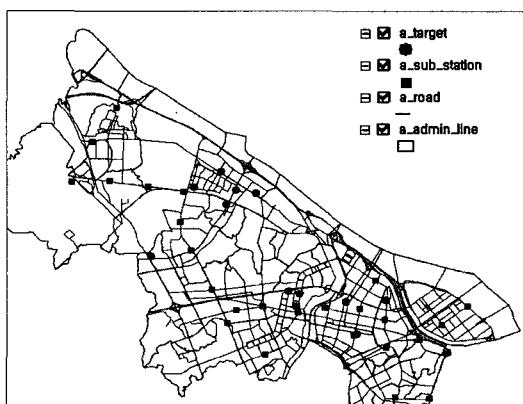


그림 2. 연구지역내 대형할인점 분포

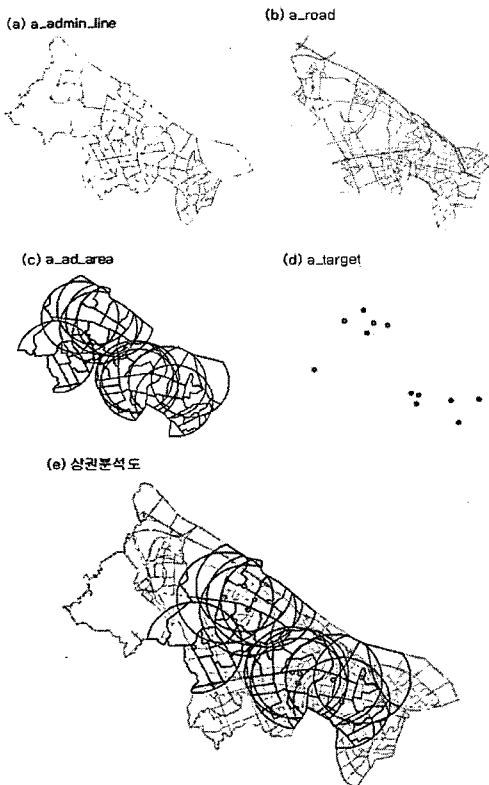


그림 3. 상권분석도 레이어

의 경우는 2차선 이상의 접근성과 통행량이 많은 주요 도로만을 선택하여 이용하였다. 전국 단위의 교통 주제도에서 연구지역을 추출하기 위하여 속성 중 디스트릭트(District)\_ID를 이용하여 행정구역도, 도로망도, 그리고 시설물도를 대상으로 새로운 레이어에 각각의 주제에 맞게 추가하여 4개의 공간 자료를 구축하였다. 연구의 중심인 대형할인점의 위치는 직접 지도화를 실시하였다(그림 2).

대형할인점의 상권분석을 위하여 상권의 범위를 결정하는데 있어 점포를 중심으로 반경 2km의 구역을 1차 상권으로 정하였다(이상규, 2004; 이희연 · 김지영, 2000). 그래서 각 대형할인점의 상권을 추출하여 행정 구역도의 동별 경계와 인터섹트(intersect)를 실시하여 77개의 세부 존을 설정하고, 각 대형할인점의 상권과 행정구역도(그림 3a)를 중첩 분석하여 상권도(그림 3c)를 구축하였다. 이렇게 구축된 레이어들을 중첩하여 상권분석도(그림 3e)를 제작하였다. 본 연구에서는

a\_ad\_area와 a\_admin\_line의 면적과 행정구역명 등의 속성 정보를 이용하여 a\_target에 각 대형할인점의 상권면적 비율을 적용하여 동(洞)별 통계자료를 구축하였다.

### 3) 대형할인점 입지관련 공간 연관규칙 추출

#### (1) 공간 연관성 분석

공간 연관성 분석을 위해 상용 어플리케이션인 SAS E-Miner를 사용하였다. 공간 연관성 분석을 수행하기 위해서는 a\_target 테이블에 구축된 통계자료의 연속형 변수를 직접 이용할 수 없기 때문에 범주형 변수 또

표 3. 분석자료 설명

변수명	설명	역할
ID	대형할인점 ID	ID
ITEM	대형할인점 특성	TARGET

는 문자형 변수로 변환해야 한다. 또한 대형할인점을 ID 변수로 하고 대형할인점의 특성을 품목(ITEM) 변수로 지정하였다. 연관성 분석에서는 목표변수를 여러 개 지정할 수 없기 때문에 품목변수를 목표변수(TARGET)로 지정하였다(표 3).

이를 위해 품목의 빈도수가 비슷하도록 하고 각 품목의 의미와 분석의 편이성을 고려하여 SAS 에디터에서 연속형 변수를 서울시 평균값과 연구 대상지역의 평균값을 기준으로 하여 범주형 변수로 변환하였다. 그 결과, 대형할인점 관련 특성을 10개의 품목으로 구분하고 각 품목에 대하여 2~3개의 범주로 구분하였다(표 4). 따라서 연구지역 내에는 12개의 대형할인점이 있지만, 여기서 사용할 자료의 개수가 대형할인점 당 10개의 품목이 생기므로 총 120개의 품목이 생긴다.

E-Miner에서는 연관성 분석에서 일반적 연관성 분석과 시차 연관성 분석을 수행하는데, 본 연구에서는 시간에 대한 변수가 정의되지 않았으므로 일반적인 연관성 분석을 수행하였다. 규칙을 추출하기 위한 환경 설정에 있어서 최소 지지도, 연관규칙에 있는 품목수의 상한 등을 지정할 수 있다. 본 연구에서는 비교적

표 4. 공간 연관성 분석 자료

품목값	원자료내용	기준 값	특성
BUS0, BUS1	500M이내 버스정류장(개소)	15	공간적 접근 성 변수
SUB0, SUB1	1000M이내 지하철역	유/무	*연구지역의 평균만 적용
RO0, RO1	주변도로접근(개)	2	
POP0, POP1, POP2	인구(명)	200000, 245360	
H_A0, H_A1, H_A2	가구당 아파트 비 (1가구 당 아파트 거주 비)	0.39, 0.49	인구학적 변수
P_A0, P_A1, P_A2	인구밀도 (<명/km <sup>2</sup> )	1699.4, 2125.9	
TXP0, TXP1, TXP2	개인당 지방세 부담(원)	811688, 823300	
PW0 ,PW1, PW2	인구당 노동자 비 (1인당 노동자 비)	0.36, 0.469	경제학적 변수
CAR0, CAR1 CAR2	가구당 자가용 자동차 보유대수 비 (1가구당 자동차 보유비)	0.65, 0.75	
BK0, BK1 BK22	상권내 은행수(개소)	30, 47	주변환경적 변수

적은 수의 트랜잭션(품목 수 : 120개)을 이용하기 때문에 지지도가 높은 규칙을 찾기 위해서 규칙 발생 빈도를 60%로 제한함으로써 총 12개의 대형할인점에서 5개 이하의 대형할인점에서 발생하는 규칙을 제거하는 조건을 명시하였기 때문에 최소 지지도는 41.67%로 높게 설정 되었다. 또한 최대 세 가지 품목이 연관성을 가지는 규칙을 추출되도록 환경설정을 하였다.

#### 4. 결과분석과 검증

##### 1) 결과분석

E-Miner를 이용한 공간 연관성 분석을 실시한 결과 84개의 유용한 규칙이 도출되었다. 본 연구에서는 대형할인점의 입지에 영향을 미치는 다양한 인자값을 판단하기 위함이기 때문에 인자값을 결정하는 기준을 단

계별로 적용하여 분석하였다.

예로, ‘X → Y, support(66.67%), Confidence (88.89%), Lift(1.19)’라는 규칙은 ‘대형할인점이 입점한 위치에서 반경 2km이내의 상권 영역에 품목 X가 발생한다면, 품목 Y도 발생한다. 이는 지지도 66.67%를 가지고 신뢰도 88.89%로써 향상도가 1이상이므로 양의 상관관계를 가지는 규칙이다’라고 해석할 수 있다. 여기서 향상도가 의미하는 양의 상관관계라는 것은 조건X가 있는 경우가 조건X가 없는 경우보다 발생할 확률이 커진다는 것을 의미한다. 또한 결과를 이해하는데 있어서 유의해야 할 것은 결과 규칙을 해석함에 있어서 조건과 결과를 직접적인 인과관계로 보기보다는 상호관련성으로 보아야 한다는 것이다.

이에 따라, 품목별로 적정한 조건에 따라 결과들을 확인하면 실제 84개의 규칙이 존재하지만 규칙에서 제시한 조건과 결과의 역의 관계나 포함관계를 고려한다면 비교적 간단한 입지 영향 인자를 추출할 수 있다. 따라서 본 연구에서 향상도를 바탕으로 1과 가까운 1.07의 향상도를 가지는 연관규칙을 배제하고 조건과 결과를 바탕으로 의미를 부여하였다(표 5). 그 결과로 21개 규칙 조건에 대하여 35개의 결과를 얻을 수 있다. 이는 대형할인점 입지 영향 인자로 이용할 수 있다.

다음은 표 5에서 제시한 유용한 규칙들 중에서 중복되는 의미를 제외시킨 대표적인 규칙을 읽기 쉽게 정리하였다.

▲ 규칙1: [번호 1] 주변 은행수가 30보다 작으면, 연지방세는 811,688원 이하를 내고 상권내 노동인구비는 0.36이다. (지지도 : 41.67%, 신뢰도 : 100%, 향상도 1.71).

▲ 규칙2: [번호 2-2] 상권내 은행수가 47개소 보다 많으면, 지하철역은 1개 이상 존재하고 도로접근성이 우수하다. (지지도: 41.67%, 신뢰도: 83.33%, 향상도 1.67)

▲ 규칙3: [번호 3] 상권내 버스정류장이 15개소 이하이면, 상권인구가 20만명~24만 5천명이 거주한다. (지지도: 41.67%, 신뢰도: 62.5%, 향상도 1.5)

▲ 규칙4: [번호 13] 도로접근성이 우수하고 자동차비가 0.75이상이면, 연간 지방세 납세액이 811,688원

표 5. 공간 연관규칙 결과 분석

번호	조건	결과	대형할인점 입지 영향 인자 분석
1	BK0	TXP0 & PWO	• 은행수 30↓ : 지방세 811,688원↓, 노동인구비 0.36↓
2	BK2	TXP2 & SUB1 SUB1 & RO1	• 은행수 47↑ : 지방세 823,300원↑, 지하철 유 • 은행수 47↑ : 지하철 유, 도로접근 우수
3	BUS0	POP1	• 버스정류장 15↓ : 인구 20만~24만 5천
4	CAR2	P_A2 & PW0 PW0 & H_A2 RO1 & PW0	• 자동차비 0.75↑ : 인구밀도 2125명↑, 노동인구비 0.36↓ • 자동차비 0.75↑ : 아파트비 0.49↑, 노동인구비 0.36↓ • 자동차비 0.75↑ : 도로접근 우수, 노동인구비 0.36↓
5	H_A2 & CAR2	PW0	• 자동차비 0.75↑, 아파트비 0.49↑ : 노동인구비 0.36↓
6	H_A2	CAR2	• 아파트비 0.49↑ : 자동차비 0.75↑
7	P_A2 & CAR2	PW0	• 인구밀도 2125명↑, 자동차비 0.75↑ : 노동인구비 0.36↓
8	PW0 & BK0	TXP0	• 노동인구비 0.36↓, 은행수 30↓ : 지방세 811,688원↓
9	PW0 & CAR2	TXP0	• 노동인구비 0.36↓, 자동차비 0.75↑ : 지방세 811,688원↓
10	PW0 & H_A2	CAR2	• 노동인구비 0.36↓, 아파트비 0.49↑ : 자동차비 0.75↑
11	PW0	H_A2 & CAR2 P_A2 & CAR2 TXP0 & BK0 TXP0 & CAR2 TXP0 & RO1 TXP0 & SUB0	• 노동인구비 0.36↓ : 아파트비 0.49↑, 자동차비 0.75↑ • 노동인구비 0.36↓ : 인구밀도 2125명↑, 자동차비 0.75↑ • 노동인구비 0.36↓ : 지방세 811,688원↓, 은행수 30↓ • 노동인구비 0.36↓ : 지방세 811,688원↓, 자동차비 0.75↑ • 노동인구비 0.36↓ : 지방세 811,688원↓, 도로접근 우수 • 노동인구비 0.36↓ : 지방세 811,688원↓, 지하철 무
12	RO1 & BK2	SUB1	• 도로접근성 우수, 아파트비 0.49↑ : 지하철 유
13	RO1 & CAR2	PW0	• 도로접근성 우수, 자동차비 0.75↑ : 지방세 811,688원↓
14	RO1 & PW0	CAR2 TXP0	• 도로접근성 우수, 노동인구비 0.36↓ : 자동차비 0.75↑ • 도로접근성 우수, 노동인구비 0.36↓ : 지방세 811,688원↓
15	RO1	SUB1 & BK2	• 도로접근성 우수 : 지하철 유, 은행 47↑
16	SUB0	TXP0 & PW0	• 지하철 무 : 지방세 811,688원↓, 노동인구비 0.36↓
17	SUB1 & BK2	RO1 TXP2	• 지하철 유, 은행 47↑ : 도로접근 우수 • 지하철 유, 은행 47↑ : 지방세 823,300원↑
18	SUB1	RO1 & BK2 TXP2 & BK2	• 지하철 유 : 도로접근 우수, 은행 47↑ • 지하철 유 : 지방세 823,300원↑, 은행 47↑
19	TXP0	PW0 & BK0 PW0 & CAR2 RO1 & PW0 PW0 & SUB0	• 지방세 811,688원↓ : 노동인구비 0.36↓, 은행수 30↓ • 지방세 811,688원↓ : 노동인구비 0.36↓, 자동차비 0.75↑ • 지방세 811,688원↓ : 노동인구비 0.36↓, 도로접근 우수 • 지방세 811,688원↓ : 노동인구비 0.36↓, 지하철 무
20	TXP2 & BUS1	BK2	• 지방세 823,300원↑, 버스정거장 15↑ : 은행 47↑
21	TXP2	SUB1 & BK2	• 지방세 823,300원↑ : 지하철 유

이하이다. (지지도: 50%, 신뢰도: 100%, 향상도 1.33)

▲ 규칙.....(이하 생략)

이와 같이 공간 연관규칙에서는 조건과 결과로 입지 영향인자를 설명할 수 있다. 문장으로 만들면 분석 결과를 이해하는데 비전문가에게도 도움을 줄 수 있다.

표 6. 대형할인점 매출액을 이용한 결과 검증

번호	조건	결과	이마트 가양점	홈플러스 영등포점	롯데마트 영등포점	까르푸 목동점	까르푸 가양점	이마트 신월점
		524억원	506억원	202억원	311억원	174억원	71억원	
		3500평	17000평	3500평	3200평	3700평	900평	
1	BK0	TXP0 & PWO	○				○	○
2	BK2	TXP2 & SUB1 SUB1 & RO1		○ ○	○ ○	X ○		
3	BUS0	POP1	○	○	X			○
4	CAR2	P_A2 & PW0 PW0 & H_A2 RO1 & PW0	○ ○ X			○ X ○	X ○	
5	H_A2 & CAR2	PW0	○					
6	H_A2	CAR2	○		X			
7	P_A2 & CAR2	PW0	○	X		○		
8	PW0 & BK0	TXP0	○				○	○
9	PW0 & CAR2	TXP0	○			○	○	
10	PW0 & H_A2	CAR2	○					
11	PW0	H_A2 & CAR2 P_A2 & CAR2 TXP0 & BK0 TXP0 & CAR2 TXP0 & RO1 TXP0 & SUB0	○ ○ ○ ○ X X			X ○ X ○ ○ X	○ X ○ ○ ○ ○	X X ○ ○
12	RO1 & BK2	SUB1		X	○			
13	RO1 & CAR2	PW0		X		○	○	
14	RO1 & PW0	CAR2 TXP0		X X		○ ○	○ ○	X ○
15	RO1	SUB1 & BK2		○	○	○	X	X
16	SUB0	TXP0 & PW0	○				○	
17	SUB1 & BK2	RO1 TXP2		○ ○	○ ○	○		
18	SUB1	RO1 & BK2 TXP2 & BK2		○ ○	○ ○	○ X		
19	TXP0	PW0 & BK0 PW0 & CAR2 RO1 & PW0 PW0 & SUB0	○ ○ X ○			X ○ ○ X	○ ○ ○ ○	○ X ○ X
20	TXP2 & BUS1	BK2			○			
21	TXP2	SUB1 & BK2			○			
계			○:18 X:4	○:9 X:5	○:10 X:2	○:16 X:9	○:18 X:3	○:9 X:9

매출액 자료 : 2004년, 유통업제연감 보고서

이때 중요한 것은 지지도와 신뢰도를 바탕으로 여러 연관규칙 중에서 유용한 것을 찾는 것이고, 각 조건과 결과에 포함된 인자에 대한 값이 최종 대형할인점 입지 영향 인자이다.

본 연구에서 제시한 결과는 통계자료를 바탕으로 대형할인점의 입지 영향 인자를 설명할 수 있을 뿐 아니라 공간적 객체인 버스정류장이나 지하철역과 대형할인점간의 공간적 연관성을 고려한 입지 영향 인자도 추출할 수 있었다. 이러한 인자는 연관규칙을 통해 얻어진 조건과 결과로 귀결될 수 있다. 즉, 추출된 규칙에서 얻어진 조건과 결과는 대형할인점이 입지하는데 있어서 고려되어야 할 인자로 2개 이상의 인자가 상호 연관성을 가지고 있다고 설명될 수 있다. 또한 하나의 규칙에서 조건과 결과에 나열된 인자가 직접적 인과 관계로 보는 것 보다는 앞서 지적한 연관성을 바탕으로 설명될 수 있다.

## 2) 검증

대형할인점의 최적입지를 가장 잘 판단할 수 있는 것이 매출액 자료이다. 그래서 본 연구에서는 2004년도 대형할인점의 매출액 자료를 이용하여 본 분석의 결과가 타당한지 판단하고자 한다. 이러한 방법은 현재 입지한 위치가 적정한 위치에 입점하였는가에 대하여 대형할인점의 관리자에게 매우 유용한 정보를 제공할 수 있다. 표 6과 같이, 매출액과 매장면적에 대비하여 몇 개의 규칙이 각 대형할인점의 입지조건에 부합하는지를 확인하였다.

84개의 공간연관규칙 중 추출된 35개의 항목을 대상으로 실제 대표적인 6개의 주요 대형할인점에 대해 적합성 여부를 확인하였다. 그 결과 매출액이 가장 높은 이마트 가양점이 가장 많은 항목에서 부합하는 것을 확인하였고, 가장 매출액이 저조한 이마트 신월점은 매장면적이 가장 작고 부적합과 부합이 동일한 수로 나오는 것을 확인할 수 있다. 이는 신월점에 대한 입지 여건이 다른 매장에 비해 좋지 않기 때문에 매장의 규모를 작게 한 것으로 해석 할 수 있다. 홈플러스 영등포점의 경우는 부합과 부적합의 비율이 까르푸 목동점에 비해 낮게 나왔지만, 실제 매장의 규모를 비교해 본

다면 매출액은 저조한편이다. 그래서 홈플러스에서는 대규모 매장을 설립하여 입지여건의 불리함을 극복한 경우이다. 전체적으로, 6개의 주요 대형할인점은 본 연구에서 추출한 35개의 공간 연관규칙에 대하여 최소 12개에서 25개에 대하여 규칙의 조건에 부합하였고 결과의 과반수까지 만족하는 것을 확인하였다. 이는 추출된 규칙이 각 대형할인점의 입지에 영향을 주는 인자를 고려한 것으로 볼 수 있다.

특이할 사항은 까르푸 가양점으로 규칙에 가장 많은 수가 적합한 것을 확인하였는데, 경쟁업체인 이마트 가양점과 불과 800m정도 떨어져 있어 경쟁업체에 의한 영향과 규칙에 부적합한 것이 인구밀도 항목이므로 소비의 규모를 결정하는 인구관련 인자가 도로의 접근성이 보다 주요한 입지영향인자인 것을 확인하였다.

마지막으로 경쟁점의 영향과 매장규모의 영향을 배제하기 위하여 이마트 가양점, 롯데마트 영등포점 그리고 까르푸 목동점에 대하여 매출액과 규칙에 대한 적합성을 분석해 보면, 부적합 비율을 따져보면 많은 규칙에서 적합성 유무를 판단하고 부적합이 적으면 매출액이 많다는 것이 확인되었다.

이와 같이, 대형할인점의 입점에 관한 공간연관규칙을 검증한 결과 추출된 규칙은 각 대형할인점에 대하여 의미 있는 상호연관성이 존재한다는 것을 확인하였다.

## 5. 결론

본 연구는 대형할인점 입지 영향 인자를 추출하기 위하여 공간 데이터마이닝 기법인 공간 연관규칙을 적용하였다. ArcGIS ver 9.0을 사용하여 수치지도를 가지고 상권을 분석하고 통계청과 지방자치단체를 통해 획득한 통계자료를 해당 상권에 취합하여 연관성분석을 위한 원시 자료를 구축하였다. 구축한 원시자료를 공간 연관성 분석을 통하여 공간 연관규칙을 발견하고 이를 분석하여 입지 영향 인자에 대하여 '조건→결과'의 규칙으로 추출하였다. 추출된 규칙들 중 충복되는 것과 향상도가 1에 가까운 것을 제거하여 최종 공간 연관규칙으로 선정하였다. 그 결과로 일정한 형식의 입

지영향인자를 선별할 수 있었다.

즉, 공간 연관성 분석을 실시한 결과 84개의 유용한 규칙을 도출할 수 있었다. 이 규칙에 대하여 향상도를 바탕으로 1과 가까운 1.07의 향상도를 가지는 연관규칙을 배제하고 조건과 결과를 바탕으로 의미를 부여한 결과 21개 규칙 조건에 대하여 35개의 대형할인점의 입지 영향인자를 도출할 수 있었다.

연구 결과를 검증하기 위하여, 본 연구에서는 대형 할인점의 매출액과 비교하여 그 타당성을 분석하였고, 표 4에서 제시한 입지 영향 인자들의 부합과 부적합의 비율에 따라 매출액과 상관관계가 있는 것을 확인할 수 있었다.

본 연구에서는 입지분석 기법의 새로운 형태로 기존의 일반 통계분석과 공간분석을 이용한 입지분석 등의 기법을 벗어나 새롭게 떠오르는 공간 데이터마이닝 중 공간 연관규칙을 적용하였다는 점에서 그 의의를 들 수 있다. 그러나 공간 연관성 분석을 적용함에 있어 연구 지역에 대하여 대형할인점의 수가 적어서 충분한 이론적 근거를 바탕으로 규칙을 제시하지 못하였다.

향후 서울시나 전국을 대상으로 하여 상당히 많은 수의 대형할인점을 이용하여 보다 정밀한 공간연관규칙을 발견하여 입지 영향 인자를 추출해야 할 것이다. 또한, 본 연구를 통하여 공간연관규칙의 활용 기대성을 높일 수 있었다.

### 註

- 1) 자연녹지지역의대형할인점등설치·운영에관한고시 제2조 (정의) 이 고시에서 사용하는 용어의 정의는 다음 각호와 같다 1. “대형할인점”이라 함은 상법상 회사가 개설한 판매시설로서 유통산업발전법 제2조에서 규정한 대규모점포의 업태 중 대형점과 저가지향형 대규모점포(상품을 통상의 판매 가격보다 저렴한 가격으로 계속 판매하는 대규모점포를 말한다.)로 선정된 자연녹지지역 안의 판매시설을 말한다.
- 2) 2005년 산업자원부 유통업계 전망 조사에 따르면, 인터넷 쇼핑의 성장률은 9.1%, 대형할인점 7.2%, TV홈쇼핑 6.0%, 재래시장 5.0%, 편의점 3.9%, 슈퍼마켓 1.1%의 순이고 백화점이 0.5%로 성장률이 가장 낮을 전망이다.
- 3) Han · Kamber(2001)은 데이터마이닝을 ‘데이터베이스, 데이터 웨어하우스, 또는 다른 정보저장소에 저장된 대용량

의 데이터로부터 유용한 지식을 발견하는 과정’으로 정의 하였음.

- 4) 데이터 집합 X에 대하여 전체 트랜잭션에서 X가 차지하는 비율로 sup(X)로 표시함. sup(X) = (X를 포함하는 트랜잭션 수)/(전체트랜잭션 수)
- 5) 규칙에 대한 강도를 나타내는 척도로써 규칙  $X \rightarrow Y$ 로의 규칙이 존재한다면 X를 포함하는 트랜잭션 중 Y를 동시에 포함하는 트랜잭션의 비율로 conf( $X \rightarrow Y$ )로 표시함. conf( $X \rightarrow Y$ ) = ( $X$ 와  $Y$ 를 동시에 포함하는 트랜잭션의 수)/(X를 포함하는 전체 트랜잭션 수)
- 6) 교통주제도는 베셀 타원체를 기준타원체로 삼고, 단일원점 체계(128E, 38N)를 적용하였으며, 원점좌표는 X(N)=600,000과 Y(E)=400,000으로 정의되었음.

### 文獻

- 권용걸 · 강양석, 2002, “대형할인점 입지 결정 요인에 관한 연구,” 국토연구, 37(1), 207-217.
- 김근형 · 황병웅 · 김민철, 2004, “중요지지도를 고려한 연관규칙 탐사 알고리즘,” 정보처리학회논문지, 11-D(3), 545-552.
- 산업연구원, 1999, 유통 신업체 경쟁력 연구.
- 산업자원부, 2002, 자연녹지지역의 대형할인점 등 설치·운영에 관한 고시.
- 삼성경제연구소, 2001, 국내 할인점시장 현황과 성장전략.
- 서울특별시, 2004, 2004년 서울시 주민등록통계.
- 서울특별시, 2005, 2005년 9월 서울시 산업활동동향 보고서.
- 심규석, 2001, “특집 1부, 데이터 마이닝–걷히는 안개를 바라보면서–,” 마이크로소프트 5월호, 210-217.
- 이병길, 2003, “비지니스 GIS에서 공간 데이터마이닝 기법 을 이용한 상권추출,” 한국GIS학회, 11(2), 171-184.
- 이상규, 2004, “대형할인점의 매출액 결정에 있어서 입지 요인의 영향에 관한 연구,” 국토연구, 40, 35-52.
- 이희연 · 김지영, 2000, “대형할인점의 입지적 특성과 상권 분석에 관한 연구,” 국토연구, 35(6), 61-80.
- 임영문 · 최영두, 2000, “연관규칙을 이용한 데이터 분석에 관한 연구,” 산업경영시스템학회지, 23(61), 115-126.
- 조성호 · 박순호, 1996, “GIS기법을 이용한 도시공공서비스 시설의 입지분석,” 한국지역지리학회지, 2, 69-85.
- 조영아, 1999, Arc/View를 이용한 광주·전남지역 공간 연

- 관 규칙 탐사, 전남대학교 대학원 석사학위 논문.  
최대식, 1998, 대형 할인점의 입지특성에 관한 연구, 서울대학교 대학원 석사학위 논문.
- 한국체인스토어협회, 2004, 2004년 유통업체연감 보고서.
- Agrawal, R., Imielinski, T., and Swami, A., 1993, Mining association rules between sets of items in large databases, *ACM-SIGMOD*, 22(2) 207-216.
- Han, J. and Kamber, M., 2001, *Data mining: concepts and techniques*, Morgan Kafmann Publisher, San Francisco.
- Koperski, K. and Han, J., 1995, Discovery of spatial association rules in geographic information databases, *4th Int. Symp. SSD'95*, 47-66.
- Ladner, R., Petry, F.E., and Cobb, M.A., 2003, Fuzzy set approaches to spatial data mining of association rules, *Transactions in GIS*, 7(1), 123-138.
- Shekhar, S., Huang, Y., Wu, W., and Lu, C., 2001, What's Spatial about Spatial Data Mining: Three Case Studies, *Data Mining for Scientific and Engineering Applications*. in kumar, V., R. Grossman, C. Kamath, and R. Namburu(eds.) Kluwer Academic Publishers, Norwell, MA, USA.

- Shekhar, S., Zhang, P., Huang, Y., and Vatsavai, R., 2003, Trends in Spatial Data Mining, in *Data Mining: Next Generation Challenges and Future Directions*, H. Kargupta, in kargupta, H., A. Joshi, K. Sivakumar, and Y. Yesha (eds.), AAAI/MIT Press, Cambridge, MA, USA.
- Zhou, T., 1999, Show me the location: A GIS approach on discount store location study, *ESRI Proceedings* 98, 1-22.

교신 : 이용익, 402-751 인천 남구 용현4동 인하대학교 지리정보공학과 GIP 연구실(이메일 : a78leekey@naver.com, 전화 : 032-867-7605)

Correspondence : Yongik Lee, GIP Lab., Department of Geoinformatics, Inha University, 253, Yonghyun-dong, Nam-gu, Incheon, 402-751, Korea(e-mail : a78leekey@naver.com, Tel : 032-867-7605)

최초투고일 06. 4. 17.

최종접수일 06. 6. 12.