

# 서울특별시 소지역 상권 매출에 영향을 미치는 요인분석을 위한 공간가중회귀모형 적용에 관한 연구

— 공간 빅데이터 활용을 중심으로 —

이명호\* · 홍상기\*\* · 신동빈\*\*\*

## A Study on Application of Spatial Weighted Regression for Analyzing the Influential Factors of Sales in Sub-region Commercial Area of Seoul

— Focusing on Geo-Spatial Big Data Utilization —

Lee, Myoung Ho\* · Hong, Sang Gi\*\* · Shin, Dong Bin\*\*\*

**요약:** 본 연구는 서울특별시 소지역 상권 매출에 영향을 미치는 요인분석을 위하여 공간 빅데이터 및 공간가중회귀모형의 적용방안을 제시한다. 이를 위해 먼저 선행연구를 검토하여 분석방법과 이론을 확립하였으며, 분석을 위한 변수와 데이터를 수집하여 소지역 단위로 구축하였다. 이후 공간가중회귀모형을 적용한 소지역 및 행정경계 단위의 회귀분석을 실시하였다. 이를 통해 국지적 R Square를 소지역 단위와 행정경계 단위에 시각화함으로써 분석결과를 비교하여 모형의 설명력을 평가하였다. 또한 독립변수별 회귀계수를 시각화하여 지역별 영향력을 측정하였다. 마지막으로 본 연구의 과정과 결과, 그리고 활용을 포함하는 결론을 제시하였다.

**주요어:** 공간 빅데이터, 소지역, 상권매출, 공간가중회귀분석

**Abstract:** This study suggests the application of the Spatial weighted regression model to analyze the factors affecting the sales of Sub-region commercial area in Seoul using Geo-spatial big data. For this purpose, First, the previous research was reviewed to establish the analytical methods and theories, and the variables and data for the analysis were collected and constructed in Sub-region units. After that, regression analysis was performed for the sub-region and Administrative boundary analysis unit applying the Geographical Weighted Regression Model. Through this, we visualized the local R squares in the Sub-region unit and the Administrative boundary analysis unit, and compared the analysis results to evaluate the explanatory power of the model. In addition, we visualized regression coefficients of major variables and measured regional influence. Finally, conclusions including the process, results, and application of this study are presented.

**Key Words:** Geo-Spatial Big Data, Sub-region, Commercial Sales, Geographical Weighted Regression Model

## I. 서론

오늘날 지역상권은 과거에 비해 그 정착과 확장 양

상이 보다 가속화, 다핵화하고 있다. 상권 형성은 빠르고 다양해지는 현대 소비자의 요구를 반영하고 있으며 이를 위해 소비자의 정보를 수집·활용하는 중요성이

\* 교신저자, 정회원 지오투정보기술, 주임(Corresponding Author, Member, Geotwo Technology Co., Ltd, Assistant Manager, mhlee@geotwo.com)

\*\* 공동저자, 정회원 안양대학교 도시정보공과 교수(Member, Professor, Department of Civil Engineering, Anyang University, skhong@aycc.anyang.ac.kr)

\*\*\* 공동저자, 정회원 안양대학교 도시정보공과 교수(Member, Professor Department of Civil Engineering, Anyang University, dbshin@anyang.ac.kr)

부각되고 있다. 근래에 들어 복잡한 소비자 행태와 상권변화요인을 도출하는 다양한 분석들이 시도되고 있으나, 이 또한 기존의 단순한 분석기법을 적용한 지역상권분석으로 공간현상과 보다 복잡한 사회현상이 고려되지 못할 뿐만 아니라 가치창출에도 한계가 있다. 특히 일반적인 통계자료를 근거로 단순 확산 모형을 적용함으로써 공간적인 상호작용에 대한 고려가 미흡한 상황이다. 따라서 과거 상권 분석 패러다임에서 벗어나 상권에 관한 다양한 외부요인을 적용하여 체계적으로 접근할 수 있는 새로운 분석 기법이 필요한 실정이다.

한편 정보통신기술의 발전에 따라 등장한 공간 빅데이터는 분야별 정보를 공간적으로 연계·활용함으로써 다양한 분야에 응용하는데 유용성이 클 것으로 예상하고 있다(안재성, 이양원, 2013). 공간 빅데이터는 미래 트렌드와 패턴에 대한 공간적인 맥락을 이해할 수 있을 뿐만 아니라 지역 상권 및 다양한 분야에서 적용 가능할 것으로 예상된다. 그러나 공간 빅데이터를 활용하여 지역 활성화 부문에 적용한 연구는 미흡한 실정이며 아직까지 공간 빅데이터의 파급력에 비해 관련 연구가 상당부분 미지의 영역으로 존재한다. 따라서 지역상권에 관한 다양한 요소와 공간적 요소들을 종합적으로 고려할 수 있는 공간 빅데이터 활용·분석기법이 필요하다.

이러한 맥락에서 본 연구는 공공, 민간에서 구축·생성되는 카드매출정보, 유동인구 등 공간 빅데이터를 활용하여 지역상권에 관한 인구, 고용, 소득 등의 매출요인을 도출하고 이를 공간적으로 연계하여 공간가중회귀모형을 소지역 상권매출에 관한 요인분석에 적용함으로써 분석결과를 통해 다양한 활용방안을 제시하는데 목적이 있다.

## 1. 연구방법

본 연구는 선행연구 및 이론적 고찰, 분석모형 및 자료의 구축, 지역상권의 매출영향력 분석 등 크게 3가지 부분으로 구분하여 연구를 수행하였다. 먼저 선행연구에서는 기존의 공간분석기법을 활용한 상권분석 연구와 빅데이터와 공간정보를 연계한 빅데이터 융합·활용 연구를 검토하였으며 이를 토대로 매출액 영향요인 분석방법을 선정하고 본 연구에서의 차별성을 제시하였다. 다음으로 이론적 고찰에서는 본 연구의 기초가 되

는 이론적 배경과 용어의 개념 등에 대해 고찰하였다. 분석모형 및 자료의 구축에서는 매출영향요인 및 요인 간 상호관계 등을 파악함으로써 분석변수를 구축하고 실증분석을 위한 분석 방법론과 기준, 절차 등을 설정하여 분석모형을 구축하였다. 이후 실증분석에 사용될 대상지 내 통계정보와 빅데이터 등을 검토, 수집하고 공간정보와 연계하여 데이터를 구축하였다. 실증분석에서는 서울특별시를 분석 대상으로 설정하고 지역 간 비교와 분석모형의 실효성을 파악할 수 있도록 행정구역 단위와 소지역 단위로 분석단위를 설정하였다. 앞서의 실증자료를 공간가중회귀, 공간통계분석 등을 연계하여 본 연구의 모형에 적용·분석함으로써 지역 간 매출 영향력을 도출하였다. 이를 토대로 다양한 부문의 활용방안을 제시하였다. 상기 연구방법을 토대로 본 연구의 절차와 흐름을 정리하면 다음 Figure 1과 같다.

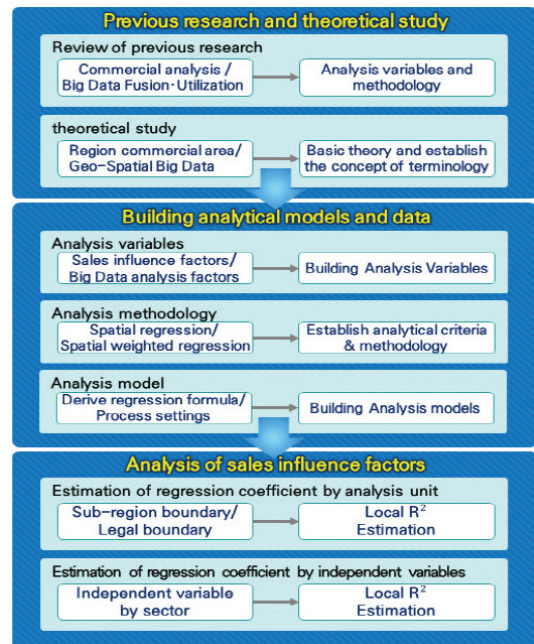


Figure 1. Research flow

## II. 선행연구 및 이론적 고찰

### 1. 선행연구 검토

상권분석은 전통적으로 많은 연구가 이루어진 분야

이다. 상권분석에 관한 연구분야에서는 GIS와 통계적 방법론에 대한 연구에 힘입어 더욱더 발전하여 보다 효과적이고 실용적인 방법론이 제시되고 있다. 먼저 GIS 공간분석기법을 적용한 상권분석 연구에서 정대석·김형보(2015)는 상권의 업종별 분포와 매출에 미치는 영향요인을 경기도 31개 시군을 대상으로 분석하였다. 이를 위해 다양한 업종에 대한 회귀분석 모형을 구축·적용하였으며 분석결과, 주요 요인별, 업종별, 매출에 대한 상관관계를 추정하였다. 이경주 등(2015)은 강원도 지역의 지역상권 현황진단과 상권활성화 및 발전방안 마련을 위해 상권 활성화에 큰 영향을 주는 매출액을 기반으로 공간회귀모형을 적용하여 지역상권의 현황을 진단하고 활성화를 위한 각종 요인들을 적용한 분석결과를 통해 정책적 시사점을 제시하였다. 정은애 등(2015)은 의류 소매업에서 유발되는 매출액에 대한 영향요인에 공간적인 요소를 고려하여 체계적인 접근을 시도하였다. 의류 소매업 매출에 영향을 주는 요인들을 인구특성, 토지이용특성, 입지특성으로 구분하여 파악하였다. 각 특성에 따라 최소제곱법모형, 일반최소제곱법모형, 공간계량모형 등을 구축하였으며 이들 모형 간의 비교·분석을 실시하였다. 이러한 분석결과는 상권에 대한 매출영향요인분석에서 기존에 통계요소만을 활용한 한계를 극복하고 공간요소를 고려하여 접근한 것에 큰 의의를 갖는다.

국내 빅데이터 융합·활용에 관한 연구는 아직 초기 단계로써 다양한 분야의 데이터를 융합, 분석하여 새로운 인사이트(Insights)를 발견할 수 있는 연구가 진행되고 있다. 이승주(2015)는 기존 통계적 분석에 한계를 극복하고 새로운 접근방법의 모색을 위해 주성분 분석에 기반한 빅데이터 분석 모형을 제시하였다. 기존 많은 분야에서 활용되는 다변량 통계분석기법인 주성분 분석을 응용하여 빅데이터 분석을 위한 새로운 표본추출 방법을 제안하였으며 통계적인 모의실험으로 그 성능을 평가하였다. 구자용(2015)은 공간 빅데이터의 공간적 분포를 통해 지역별로 발생하는 현상이나 특성 등을 파악하였다. 실증분석을 위해 서울시 SNS 데이터를 수집하였으며 공간정보와의 연계과정을 통해 지도 시각화를 수행하였다. 이후 지역별로 위치한 SNS 내용을 통해 분포특성을 파악하였으며 실제 장소에 대한 정보

량을 비교하여 분석모형에 객관성을 확보하였다. 허자연(2014)은 SNS(Social Network Services)상의 정보가 지역 방문자 수에 양(+)의 영향을 미치는 것을 파악함으로써 SNS 게시정보와 실제 공간현상의 상관관계를 밝혔다. 김관호 등(2013)은 지하철 이동패턴을 효과적으로 분석할 수 있는 기법개발을 위해 서울시 스마트카드 빅데이터를 활용하여 특정 관측지점에서 다른 지점으로 이동을 설명하는 클러스터링 기반 패턴 분석기법을 제안하였다. 제안된 분석기법을 이용하여 서울시 내 실제 지하철 데이터를 비교·분석하여 분석기법의 실효성을 검증하였다.

## 2. 선행연구 검토결과 및 차별성

선행연구를 검토한 결과 기존 전통적인 상권분석은 GIS 공간분석기법을 적용하여 그 정확도를 향상하는 연구들이 다양하게 진행되고 있음을 알 수 있었다. 또한 지역상권의 활성화 분석 및 매출분석 등 동태적 분석을 다루는 연구에서는 소비자 설문을 통하여 만족도 분석을 실시하거나 소비행태(재방문의사, 지출액 등)를 종속변수로 하는 회귀분석을 실시하여 매출의 영향요인을 밝히는 연구들이 진행되어 왔다. 그러나 기존 상권분석 및 활성화분석과 관련된 선행연구는 과학적 분석을 위해 공간분석과 통계분석을 적용하고 있으나 수학적 모델에 의존적이다. 또한 자료구득의 한계에 따라 대리변수를 사용하거나 시계열자료의 구축 미흡, 공간에 대한 고려부족 등의 한계를 갖는다. 반면 빅데이터 연구에서는 실시간 웹 정보, 민간의 빅데이터, SNS 비정형정보 등을 활용할 뿐만 아니라 공간참조 데이터와 GIS 공간분석기법과 연계하여 실제 공간상에 미치는 상호작용을 입증하였으며 빅데이터 기반의 새로운 방법론을 적용함으로써 새로운 가치를 발견할 수 있는 가능성을 시사했다. 따라서 과거 지역상권의 제한된 분석모형의 틀에서 벗어나 공간적인 요소와 함께 빅데이터 기반의 다양한 영향요인을 고려할 수 있는 공간 빅데이터 분석이 필요하다.

이에 본 연구에서는 유동인구, 카드매출액 정보 등 공간 빅데이터를 활용하여 기존 연구의 한계를 극복할 수 있는 실제 매출액 정보를 활용함으로써 이에 대한 영향을 미치는 요인들을 분석한다는 측면에서 타 연구와

근본적인 차이를 갖는다. 또한 기존 행정경계 단위의 분석단위보다 규모가 작은 소지역 분석단위의 공간가중 회귀모형의 적용을 통해 보다 세분화된 지역을 분석함으로써 분석과정 및 결과의 일반화 및 활용도가 높다.

따라서 본 연구는 유동인구, 카드매출 등의 공간 빅데이터와 인구, 고용 등의 통계정보를 활용하여 지역상권 매출의 영향요인을 도출하고 공간 빅데이터 처리, 공간가중회귀모형을 적용하여 서울특별시 소지역 단위의 매출요인을 도출한다는 점에서 연구의 차별성이 있다.

### 3. 이론적 고찰

#### 1) 상권

상권(Commercial Area, Trading Area)이란 점포와 고객을 흡인하는 지리적 영역이며, 모든 소비자의 공간 선호(Space Preference)의 범위를 의미하기도 한다. 이로 보아 상권은 판매액의 비율을 고려하여 생각할 수 있는데, 대표적으로 상품 판매액의 약 70%를 차지하는 지역을 1차 상권, 다음 25%를 차지하는 지역을 2차 상권, 그 나머지를 3차 상권이라 말한다. 또한 상권은 의존형태에 따라 주거지형 상권, 도심지형 상권, 복합형 상권으로 정의되고 있다. 이러한 상권의 변화와 이동원인으로는 신규 지하철 노선개설, 버스터미널의 형성 및 이전, 버스정류장 이전 등 교통수단에서의 변화가 있으며 대형 할인점 및 백화점 개발 등 대형업체의 등장에 민감하게 반응한다. 또 공공 및 민간 기관의 이전, 대기업의 입지, 이전이 상권의 변화와 이동에 영향을 주고 배후지역의 소득수준 변화가 큰 요인으로 작용한다.

#### 2) 공간 빅데이터

공간 빅데이터의 정의를 위해서는 먼저 빅데이터에 관한 정의가 필요하다. 빅데이터는 아직까지 통일된 정의가 부재하나 여러 기관들에 따라 다양한 분야에서 여러 가지 관점으로 정의하고 있다. 먼저 Stefan, B. *et al.*(2013)은 전통적인 데이터베이스 소프트웨어로 수집, 저장, 관리, 분석이 어려울 정도의 큰 규모의 데이터로 정의하였으며, International Data Corporation(2011)은 다양한 데이터로 구성된 방대한 데이터로부터 고속 캡처, 데이터 탐색 및 분석을 통해 경제적으로 필요한

가치를 추출할 수 있도록 디자인된 차세대 기술 및 아키텍처로 정의하였다. 공간 빅데이터의 개념 또한 빅데이터와 마찬가지로 명확한 정의가 부재하며 일부 대학 및 학자에 의해 정의되고 있다. 먼저 Shashi Shekhar(2012)은 데이터의 양, 속도, 다양성 때문에 합리적인 노력으로 관리, 분석, 활용 등을 하기 위해 현재 컴퓨터의 능력을 초과하는 공간데이터세트로 정의하였다. 김미정 등(2013)은 공공·민간에서 생성되는 빅데이터와 공간정보를 융·복합한 데이터를 말하거나 세상에 존재하는 빅데이터를 공간 형태로 표현, 또는 가능한 모든 자료로 정의하였다.

앞서 살펴본 빅데이터 및 공간 빅데이터의 개념은 초기 단순 대용량 데이터 집합에서 점차 관리, 분석, 활용을 위한 새로운 기술 체계로 변화하고 있다. 즉, 공간 빅데이터의 개념은 다양한 형태의 모든 빅데이터와 공간정보를 융·복합한 데이터 집합이나 이러한 공간 빅데이터를 관리, 처리, 분석, 활용하는 동시에 미래 새로운 가치를 추출할 수 있는 기술체계의 확장된 의미로 해석할 수 있다. 본 연구에서는 공공에서 제공하는 인구, 고용 등의 통계정보와 민간에서 생성되는 유동인구, 스마트카드 데이터 등의 빅데이터를 기반으로 공간정보와 융합·활용하여 기존에 알 수 없었던 세분화된 분석을 수행하는데 초점을 둔다.

#### 3) 공간가중회귀분석

공간가중회귀모형(GWR: Geographically Weighted Regression)의 분석방법은 특정 지역을 중심으로 주변 지역의 독립변수와 종속변수 값을 공간적 가중치를 부여하는 방식으로 지역특성의 상관성을 분석하는 것이다. 일반회귀모형(Regression Analysis)은 회귀계수가 전지역에 적용되는 반면, 공간가중회귀모형은 각 지역별로 회귀계수가 다르다는 전제하에 지역별 회귀계수를 추정하는 모형이다. 공간가중회귀모형은 지역별로 상이한 설명변수와 종속변수들 간의 관계를 추정하는 방법이다. 특정 지역을 중심으로 주변 지역의 설명변수와 종속변수 값을 공간적 가중치를 부여하는 방식으로 재조정하여 각 지역의 회귀계수를 추정한다. 즉 가까이 위치하는 지역들 간에는 설명변수가 종속변수에 미치는 영향력 수준이 비슷함을 전제로 높은 가중치를 부여



하고, 거리가 멀어질수록 두 변수 간 관계성의 강도가 약해지는 것으로 보고 낮은 가중치를 부여한다. 다음 Eq. (1)은 GWR을 수학적으로 정의한 것이다.

$$y(i) = B(j)_0 + B(j)_1X_1 + \dots + B(j)_iX_i + \dots + B(j)_mX_m + \epsilon \quad (1)$$

여기서  $y_i$ 는 종속변수( $i = 1, 2, \dots$ )를 나타내며,  $x_m$ 는  $m$ 번째 독립변수를,  $B_m$ 는  $m$ 번째 회귀계수(매개변수)를,  $j$ 는 공간좌표,  $\epsilon$ 는 오차항을 나타낸다.

Eq. (1)이 일반적인 회귀모형과 다른 점은 각 항에  $j$ 가 붙는 것인데, 이는 모수의 추정치 주어진 위치( $j$ )에 대해 이루어지고 그 위치에 대해서만 의미를 가진다는 것을 나타낸다(이경주 등 2015). 지점에 따른 회귀계수의 추정식은 가중최소사승법에 따라 이뤄지며, 이에 대한 식은 다음과 같다.

$$\beta(j) = (X^TW(j)X)^{-1}X^TW(j)Y \quad (2)$$

Eq. (2)에서 공간가중치행렬  $W(j)$ 의 각 요소는 가중함수인 커널(kernel)에 따라 계산된다. 커널은 다양한 형태를 취할 수 있지만 일반적으로 가우스 형태를 사용하여 여러 회귀점을 할당하고 각 회귀점을 기준으로 일정한 거리 내의 사례에 대한 가중치를 준다(전해정, 2016).

$$w_i(j) = e^{-0.5(d_i(j)/h)^2} \quad (3)$$

여기서,  $w_i(j)$ 는 공간좌표에서  $i$ 의 관찰을 위한 가중치를 나타내며,  $d_i(j)$ 는 관측 값  $i$ 와 공간좌표 간의 거리를 나타내며  $j, h$ 는 대역폭을 나타낸다.

Eq. (3)에서  $h$ 인 대역폭(Bandwidth)값이 커질수록 회

귀점  $i$ 지역을 기준으로 더 넓은 범위의 데이터가 가중되며,  $h$ 값이 작아질수록 회귀점  $i$ 지역을 기준으로 더 좁은 범위의 데이터가 가중되는 구조를 갖는다. 관측치가 조사지역에 규칙적으로 분포할 경우 고정방식(Fixed Spatial Kernel)을 사용하고, 조사지역 주위에 관측치의 분포가 다양한 경우는 가변방식(Adaptive Spatial Kernel)을 사용한다. 그러나 이에 대한 판단이 확실치 않을 경우, 대개 가변방식을 채택한다. 적정 대역폭을 설정하기 위해 AIC(Akaike Information Criterion) 또는 CV(Cross Validation)가 사용되는데, 관찰 값과 추정 값의 차이 및 모형의 복잡성을 고려하는 AIC방식이 선호된다(정은애 등 2015).

#### 4) 매출 영향 요인

본 연구의 주요 관심대상인 매출액 영향요인은 상권 및 점포 방문에 따라 발생하는 매출이므로 매출액을 결정하는 필수지표로 구성되어 있는 상권결정요인을 검토하였다. 상권결정요인은 소비자와 관련된 수요요인과 판매자와 관련된 공급요인으로 나눌 수 있는데 수요요인은 인구, 소득수준, 인구통계적 특성을 세부요인으로 포함하며, 공급요인은 경쟁, 집적, 거리를 세부요인으로 포함한다. 상권의 결정에 있어 소비자와 관련된 수요요인으로는 크게 소득, 인구, 인구통계적 특성이라는 세 가지 요인이 있다(김성문, 2014). 지역의 배후인구는 지역상권을 결정하는 가장 중요한 요인이라는 것이 일반적인 통념이다(김도현 등 2014). 지역의 소득수준은 지역의 구매력을 나타내는 중요한 변수로 지역상권

Table 1. Commercial determinants

Classification		Explanation
Demand factor	Population	<ul style="list-style-type: none"> <li>• Basic and essential indicators for measuring the potential purchasing power of the commercial</li> <li>• Factors that have a great influence on store decisions and sales</li> </ul>
	Income level	<ul style="list-style-type: none"> <li>• Variables that indicate the economic level of the region</li> <li>• Indispensable indicator for store sales</li> </ul>
	Demographic characteristics	<ul style="list-style-type: none"> <li>• Demographic characteristics such as education level, occupation, and cultural propensity</li> <li>• Variables that describe consumer preferences</li> </ul>
Supply factor	competition	<ul style="list-style-type: none"> <li>• Whether or not a large store is within the scope of commercial</li> </ul>
	Agglomeration	<ul style="list-style-type: none"> <li>• Stores are geographically close</li> <li>• Agglomeration effects have a positive impact on determining commercial area</li> </ul>
	Distance	<ul style="list-style-type: none"> <li>• The physical distance between the store and the consumer affects the cost of moving the consumer</li> <li>• Sales are affected by the distance between the consumer and the store</li> </ul>

의 매출액에 영향을 미치는 요인으로 널리 알려져 있다(김성문, 2014). 연령구성, 성별구성, 교육수준, 문화적 특성 등의 인구 통계적 특성은 소비자의 기호에 영향을 미치고 이에 따라 지역상권의 결정에 영향을 미치는 중요한 요소로 여겨진다(정은애 등 2015). 상권 내에 유사한 점포가 입지하게 되면 경쟁관계가 발생하며, 유사한 점포가 동종업태와 이종업태인 경우 모두 경쟁관계가 발생한다. 경쟁관계와 매출액 사이의 관계는 경쟁이 선행하여 매출액에 영향을 미친다는 연구와 지역의 매출액 수준이 새로운 경쟁자의 출현을 유도한다는 연구관점이 서로 대립하고 있다(김성문, 2014). 그러나 상권분석에 관한 많은 연구에서는 대형점포와의 경쟁을 주로 고려하고 있다(신우진 등 2010). 집적은 다수의 업체들이 지역 내에 지리적으로 근접하여 입지하는 것으로 이를 통하여 소비자들에 대한 집객효과를 기대할 수 있다. 국내외 많은 연구들은 집객효과에 있어서 소매점의 집적이 긍정적인 영향을 미친다는 분석결과를 도출하고 있다. 지리적으로 인접한 지역에 집적되어 있는 상점군은 고객의 다양한 수요욕구를 충족시켜 줄 가능성이 높기 때문에 전체 상점군의 매출에 긍정적인 요인을 보이는 것으로 판단된다(김성문, 2014). 소비자의 거주지와 상점 간 지리적 거리는 소비자의 이동비용을 발생시키기 때문에 거리가 멀수록 소비자의 이용은 감소된다. 상점과 소비자 사이의 거리는 해당 상점의 선택하는데 주요한 결정요인이다. 이와 같은 가정은 Reilly의 소매 인력의 법칙의 근간을 이루고 있으며 이를 발전시킨 허프 모형에서 역시 중요한 핵심전제로 널리 알려져 있다.

### III. 분석모형 및 자료의 구축

#### 1. 분석변수 선정

본 연구의 상권매출 분석을 위해 먼저 주요 선행연구 및 이론적 배경 검토를 통해 수요요인, 공급요인 등 상권결정요인 도출하였다. 도출된 상권결정요인을 토대로 다음 표와 같이 대표성, 방향성, 이론적 근거, 자료구득가능성의 충족 여부를 근거하여 최종변수를 선정하였다.

Table 2. Criteria for selecting variables and data

Criteria element	Explanation
Representative	In order to determine the impact on sales of commercial products, a variable with a representative value that can be used to comprehensively judge whether specific variables are duplicated or not is selected
Directionality	Indicates whether the patterns in the sales factors by variables can show generalized results
Theoretical Basis	In previous researches related to existing sales factors, variables that are mentioned as having high ratios and existed theoretical basis are selected
The possibility of obtaining data	Spatial reference data must exist and be a spatial unit variable that can be linked and analyzed.

#### 2. 자료의 구축

소지역 상권의 매출액에 미치는 영향 요인 분석을 위한 데이터는 앞서 도출된 변수에 따라 데이터 수집, 데이터 가공 등을 통해 분석 데이터를 구축하였다. 앞서 선정된 분석변수를 기준으로 각각 수요요인, 공급

Table 3. Variables affecting sales

Classification		Variable		Data
Independent variable	Demand factor	population	Residential population	100m grid center point
			Floating population (Big Data)	Points of interest (POI) in Sub-region
		Employment	Working population	100m grid center point
		Income level	Average income in the region	100m grid center point
	Supply factor	Agglomeration	Number of stores in the region	Building POI
		Distance	Included in model	100m grid center point
Dependent variable	Sales		Card sales (Big Data)	Sub-regional aggregate sales



Figure 2. Comparison of Census Output Areas (Left) with sub-region boundaries (Right)

요인 등 공간 빅데이터, 통계정보 등을 수집하였다. 데이터 수집은 공간 빅데이터에 해당하는 카드매출, 유동인구, 상점의 수와 기존 통계정보에 해당하는 주거인구, 직장인구, 평균소득(소득분위 및 소득분위별 평균소득) 등을 수집하였다. 다음으로 좌표정의 및 속성 매칭에서는 기존의 시군구, 법정동 행정경계 보다 세분화된 분석단위인 소지역 경계를 기준으로 각각의 데이터를 집계하기 위해 공간 데이터 타입을 모두 점(Point) 형태로 구축하였다. 주거인구, 직장인구, 지역의 평균소득의 경우 100M 격자 중심점 데이터를 활용하며 유동인구, 지역 상점의 수는 POI(Point of Interest)를 활용한다. 카드매출의 경우 본래의(Original) 데이터가 소지역 기준 거래내역으로 이루어지는 텍스트이며 공간 참조 식별자는 소지역 ID로 구성되어 있다. 이러한 소지역 경계를 기준으로 거래내역 데이터를 집계하게 되며 이를 위한 방법은 ArcGIS를 활용하여 테이블 조인을 수행하여 최종적인 소지역 기준 매출 집계 데이터를 구축하였다. 본 연구는 시군구, 법정동 등 한정된 분석단위에서 벗어나 소지역(Micro Zone) 상권에서의 매출액 영향요인을 분석하는데 의의가 있다. 소지역단위는 다음 그림과 같이 법정동, 집계구, 기초단위구 경계 보다 더욱 세분화된 공간단위이다.

소지역 경계의 가장 큰 특징은 각종 마이크로 데이터를 취합하여 행정구역 단위 보다 정확한 실제 상점, 배후지 등에 위치정보를 획득할 수 있다는 점이다. 소지역 경계는 통계청의 통계구역과 다르게 명확한 정의나 통일되는 작업지침이 부재하나 일반적으로 집계구 경계를 기준으로 실제 건축물, 지번, 필지 등을 참조하여 도로, 하천, 철도, 산능성 등과 같은 준항구적인 명확한 지형지물을 기준으로 폴리곤 편집 프로그램의 기능인 분할 폴리곤으로 세분화하여 구성할 수 있다. 그리

나 본 연구에서는 직접조사에 소요되는 비용 문제 등으로 인하여 공간 빅데이터 연구단<sup>1)</sup>의 내부 자료인 소지역 경계 자료를 참조하여 소지역 자료를 구축하였다. 공간단위 세분화 방법에 따라 서울시 기준 19,940개의 소지역 경계가 구축되었다. 이러한 소지역 경계는 본 연구에서 선정한 분석변수를 통합하여 각종 분석에 활용할 수 있다. 다음 그림은 행정동, 집계구, 소지역 등 공간 데이터 분석 단위의 차이를 나타낸다. 본 연구에서 활용하는 공간 빅데이터는 카드매출액과 유동인구자료를 활용한다. 여기서 종속변수는 지역의 매출액이다. 이 자료는 민간통신사의 자료를 취합하여 구축하고 있는 공간 빅데이터 연구단의 내부 자료를 활용하였다. 카드매출데이터는 2014년 1월~2015년 12월 간의 월별 소지역에서 발생한 데이터를 Text형태로 구축하였다. 분석에 필요한 데이터구축을 위해 대용량 데이터에 읽기, 쓰기 및 정렬 등이 가능한 IBM 사의 SPSS Statistics 24를 활용하였다. SPSS를 통해 공간참조가 가능한 소지역 ID를 기준으로 특정 케이스(case) 선택을 위한 데이터베이스 질의어(Query, like function)를 입력하여 이에 만족하는 서울시 소지역에 대한 결과를 반환 받았다. 유동인구 정보의 경우에도 카드매출정보와 마찬가지로 SPSS와 ArcGIS를 활용하여 X, Y 좌표를 소지역 속성에 매칭하였다. 다음 집적은 공급요인에 해당하는 변수로써 본 연구의 대상인 서울특별시 전체의 상점의 수를 집적으로 판단한다. 분석·연계를 위해서 소지역별로 전체 상점의 수의 총합을 집계하여 변수로 활용하였다. 공간 빅데이터에 대한 구축 정보를 정리하면 다음 Table 4와 같다.

Table 4. Data construction

Classification	Data	RawData	RawData
Geo- Spatial Big Data	Card Sales	14,566,241	7,525,969
	Floating population	3,674,082	165,361
	Store data	2,505,817	530,623
demand supply factors data	Residential population	959,660	29,638
	Working population	421,514	24,388
	Average income	959,660	29,638

### 3. 분석모형 구축

본 연구는 공공, 민간에서 생성되는 공간 빅데이터를 활용하여 서울시 내 지역상권의 매출액에 영향을 주는 변수를 선정하고 그 영향력을 파악·검증하여 다양한 분야에 활용방안을 제시하는데 목적이 있다. 이를 위해 본 연구에서는 카드매출, 유동인구, 상가업소 수 등 공간 빅데이터와 매출액 영향을 주는 주거인구, 평균소득, 직장인구 데이터를 구축하였다. 매출액 영향력 추정을 위해 활용하는 방법론을 그 간 사회과학분야에 활용되어온 회귀분석 이론을 지리학법칙과 응용한 GWR을 활용한다. 또한 GIS 공간분석 수행에 있어 종종 발생하는 가변적공간단위문제(MAUP, Modifiable Areal Unit Problem)를 자유롭게 하고 보다 세분화된 분석이 가능한 소지역 분석단위를 활용하였다.

이를 위해 모든 독립변수는 점 데이터 형태로 구축하였으며 각각의 독립변수는 소지역 단위에 합계되어 분석된다. GWR 모형에 적용하기 위하여 설정한 종속변수는 공간 빅데이터의 카드매출 데이터이다. GWR의 독립변수로는 주거인구, 유동인구, 직장인구, 평균소득, 상가업소 수 등 5개 변수를 설정하였다.

## IV. 매출액 영향 요인 실증분석

### 1. 분석의 개요

상권 매출에 미치는 영향요인 분석을 위해 앞서 도출된 분석변수와 데이터, 분석모형을 토대로 공간분석 소프트웨어 중 하나인 ArcGIS 10.1 활용하여 서울특별시 전체 지역상권의 매출액에 미치는 영향요인을 실증 분석한다. 종속변수와 독립변수에 대한 실증자료는 소지역 단위로 구축하였으며 최종적으로 선정된 종속변수는 공간 빅데이터의 카드매출데이터이며, 독립변수는 종속변수에 영향을 주는 주거인구 수, 유동인구 수, 직장인구 수, 지역상권 내 상점의 수로 선정하였다. 분석결과는 각각의 독립변수별로 구체적인 영향력 수준을 해석하였다. 본 연구에서의 분석모형을 검증하기 위해 Local R Square와 각 독립변수의 회귀계수를 지역별로 시각화하여 비교·분석하였다. 회귀계수의 추정치는 분석 단위의 개수만큼(19,440) 도출되므로 많은 양의 계수를 해석하기 보다는 군집 및 분포패턴 중심으로 그룹화하여 지역 간 차이를 보았다. 또한 특정 주제(Theme)의 데이터를 표현하고 지역 간 차이를 확인할 수 있는 대표적인 지도 시각화 방법으로 단계구분도(Choropleth Map)로 활용하며 급간구분방법은 최소 값,

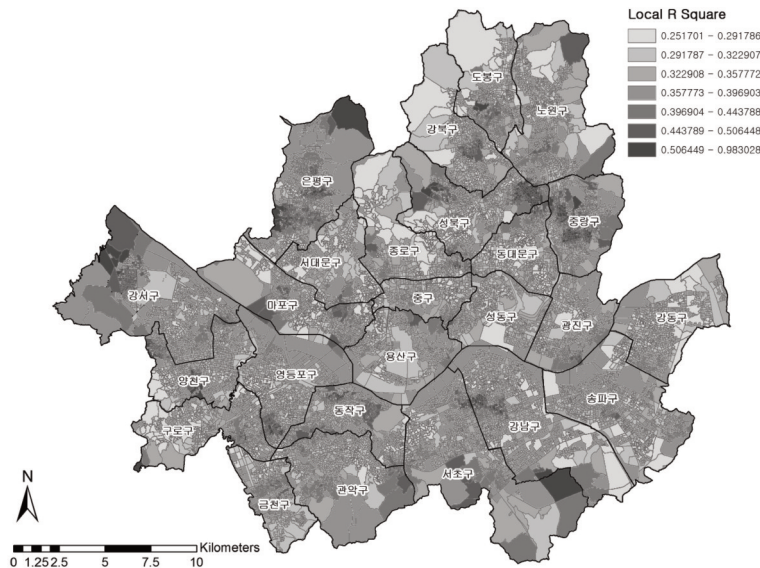


Figure 3. 소지역 단위 Local R square 분석결과



중위 값, 최대 값의 영향을 적절히 배분하여 표현할 수 있는 Natural Breaks 급간구분법을 적용하였다.

## 2. Local R square

본 연구는 종속변수 매출액에 대한 영향력 분석을 위해 매출영향요인에 해당되는 독립변수를 소지역 단위로 집계하여 공간가중회귀분석을 실시하였다. 또한 분석결과와 직관적인 확인이 가능하도록 읍면동 단위 집계 분석을 추가하여 비교하였다. 공간가중회귀분석 모형은 독립변수마다의 회귀계수를 분석하며 그 결과를 통해 종속변수인 매출액에 미치는 영향력을 지역별로 파악하는데 활용 할 수 있다. 다음 식은 앞서 도출한 회귀식에 본 연구의 변수를 대입하여 서울시 내 상권에서 매출액  $y_j$ 의 회귀계수를 추정하기 위한 GWR 모형을 나타낸다.

$$sales(j) = B(j)_0 + B(j)_1 X_{Residential\ Population} + B(j)_2 X_{Floating\ Population} + B(j)_3 X_{Working\ Population} + B(j)_4 X_{Average\ Income} + B(j)_5 X_{Number\ of\ Store} \quad (4)$$

서울특별시를 소지역으로 구분한 결과 19,440개의 분석단위가 생성되었으며, 19,440개의 회귀식을 통해 국지적 R Square를 분석하였다. 즉 본 연구의 회귀식이 모든 소지역 단위 19,440개에 적용되어 소지역별 회귀

계수와 국지적 R Square가 산출되는 것이다. Figure 3의 소지역 분석결과를 보면 용산구, 마포구 일대, 영등포구 일대, 동대문구 일대 등에서 R Square값이 상대적으로 낮은 것을 알 수 있다. 반면 붉은 색으로 표시된 지역은 양천구 일대, 은평구, 강남구 일대, 강북구 일대, 동작구 등에 속해 있는 상권으로 나타났다. Figure 4의 서울시 동 단위 분석결과를 살펴보면 마찬가지로 용산구, 마포구 일대, 영등포구 등이 약 0.40 이상의 R Square값으로 나타내어 본 연구의 독립변수가 큰 영향을 미치는 것으로 나타났다. 하지만 소지역 분석결과와 대조적으로 읍면동 분석은 서대문구, 강동구, 동대문구 등에서는 전반적으로 낮은 R Square 값을 나타내어 본 연구의 독립변수 이외에도 다양한 변수에 의한 영향이 높음을 알 수 있었다.

### 3. 독립변수의 영향력 추정

소지역 상권의 매출에 미치는 영향력 분석에서는 주거인구, 유동인구, 직장인구, 지역의 평균소득, 상점의 수 등 개별 독립변수의 영향력을 분석하였으며 이중 명확한 패턴탐색과 주변지역 간 차이가 분명한 주거인구, 유동인구를 채택하여 제시하고자 한다. 분석방법은 전체적인 맥락의 R Square 추정과 동일하게 공간가

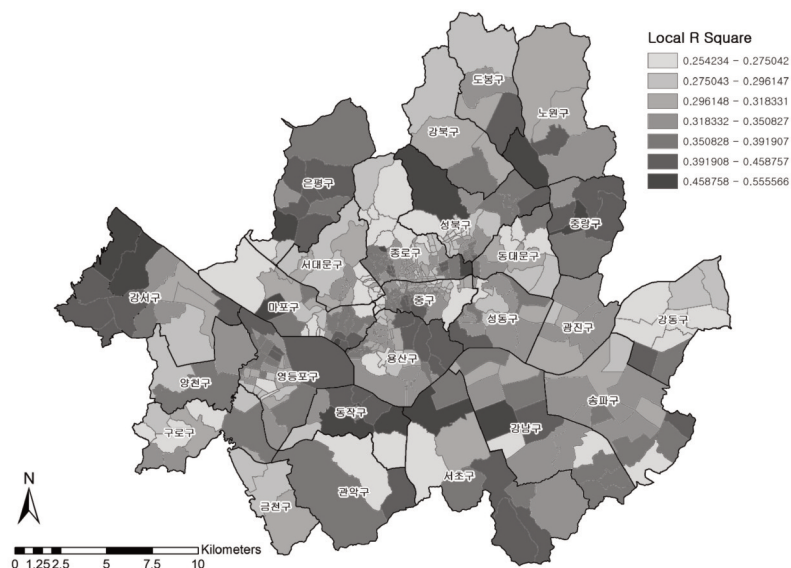


Figure 4. 읍면동 단위 Local R square 분석결과

중회귀모형을 적용한 독립변수 마다 회귀계수를 추정하였다.

먼저 잠재 수요변수 중 하나인 주거인구 변수를 소지역 기준으로 시각화하였다. 서울특별시 매출액에 미치는 영향력의 지역 차이를 확인해 보면 주거인구의 영향력이 가장 큰 지역은 마포구, 양천구 등이다. 이는 해당 지역의 인구가 매출액에 크게 영향을 주는 것을 의미하며, 외부요인의 영향이 다른 지역에 비해 비교적 적다는 것을 의미한다. 반면 영향력이 낮은 지역은 구로구 일대, 용산구 일대로 나타났다. 이는 지역 내 배후인구보다 제3의 요인으로 인한 결과일 수 있으며 또한 지역 내 인구의 수요가 적은 지역으로 해석할 수 있다.

읍면동단위 분석결과를 대체적으로 소지역 단위분석과 비슷한 결과를 나타냈으나 동대문구, 성동구, 송파구 등이 매출액에 상당한 영향 미치는 것으로 나타났으나 이는 가변적공간단위문제로 나타난 결과임을 알 수 있다. 즉 일부분의 수치로 같은 공간단위의 전체를 대표되어 표시되는 문제이다.

유동인구의 영향력이 가장 크게 나타나는 지역은 성

북구로 나타났다. 이는 해당 지역의 매출액이 특정 지역의 유동인구에 크게 영향을 받는다는 것을 의미하며 직·간접적인 잠재수요가 높은 지역임을 뜻한다. 반면 양천구, 은평구 일대에서는 상대적으로 낮은 영향력을 보이는데 이는 해당지역의 유동인구가 타 상점으로 흡수되었을 가능성이 높거나 실제 유동인구가 많더라도 상점 매출액에 미치는 영향은 상대적으로 낮은 지역으로 해석할 수 있다. 은평구 지역은 음(-)의 영향력이 가장 크게 나타나며 이는 인접지역인 서대문구의 일부에서 수요를 흡수하여 나타난 결과로 해석할 수 있다. 또한 강남의 경우 양(+) 영향력과 음(-) 영향력의 지역이 상존하며 양, 음의 영향력이 서로 대립하는 특징이 있다. 일반적으로 강남구 유동인구는 타 지역보다 상대적으로 많으나, 다양한 음(-)에 요인으로 인해 대체적으로 영향력이 낮은 지역으로 도출되었다.

읍면동 단위 분석결과, 매출액에 대한 유동인구의 영향력이 양(+)을 나타내는 지역으로 영등포구, 동작구, 강서구 일대, 용산구 일대 등으로 도출되었다.

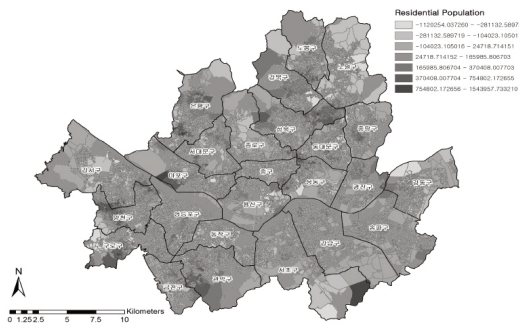


Figure 5. 주거인구에 대한 소지역 단위 분석결과

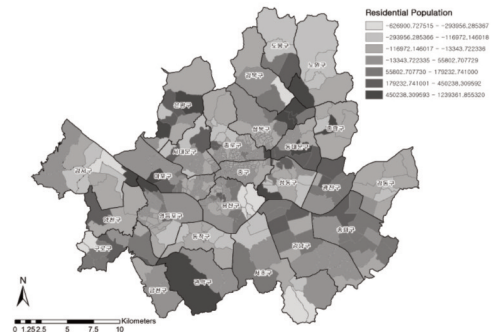


Figure 6. 주거인구에 대한 읍면동 단위 분석결과

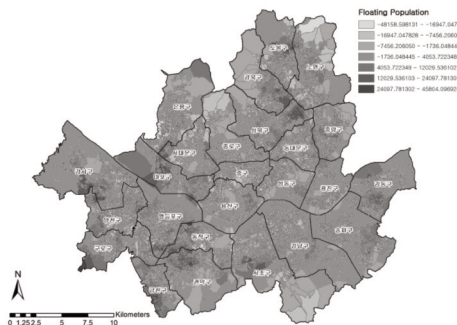


Figure 7. 유동인구에 대한 소지역 단위 분석결과

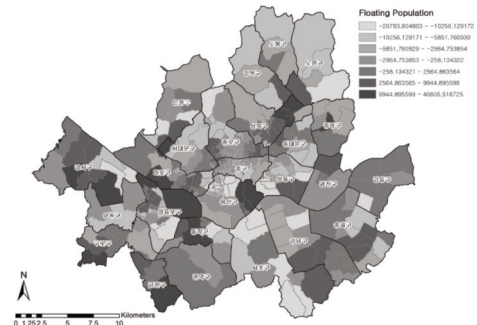


Figure 8. 유동인구에 대한 읍면동 단위 분석결과

### 3. 분석결과와 활용방안

본 연구의 분석과정과 결과 그리고 분석모형을 통해 학술적, 정책적, 기술적 측면으로 구분하여 다음과 같은 활용방안을 기대할 수 있다.

먼저 학술적 활용방안으로는 기존 지역상권 분석모형에서 다양한 요인을 고려하지 못한 점, 공간적 이질성과 가변적 공간단위 문제 등 그동안 지속적으로 지적되어온 자료구축과 활용에 대한 문제점을 최소화할 수 있는 방법과 가능성을 제공한다는 측면에서 큰 의의가 있다. 또한 상권 활성화 등과 관련된 학술연구를 보다 더 확장시켜 공간 빅데이터 처리 및 분석 모형을 적용하거나 응용함으로써 지역개발과 기타 관련 학술 분야에서 일종의 가이드라인 형태로 활용될 수 있을 것으로 기대된다.

다음으로 정책적 측면에서는 공간 빅데이터 처리를 통해 세분화된 지역의 모니터링체제로 활용가능하다. 공간 빅데이터의 세밀한 자료는 쇠퇴지역 및 정책 사각지역을 관리할 수 있도록 지원하거나 시계열 자료의 구축을 통해 새로운 패턴을 탐색할 수 있다. 이를 통해 정책지원의 세분화, 관리방안 등에 실질적으로 활용될 수 있다. 또한 공간정보 및 빅데이터 기반의 데이터 구축을 통해 다양한 분석이 가능한 기초DB로 활용가능할 것으로 예상된다. 분석 측면에서는 분석의 파급성과 효과성, 시급성 등을 고려하여 포괄적인 분야의 데이터를 축적하고 시스템 및 플랫폼으로 개발하여 고도화된 서비스가 가능할 것으로 기대된다. 기술적 측면에서는 공간 빅데이터 기반 처리, 분석기법과 기존 사회과학분야에서 활용되는 통계분석 원리 등을 응용하여 범용적이고 확장성이 뛰어난 기술로 활용될 것으로 기대된다. 또한 대용량 처리, 분석뿐만 아니라 공간분석의 원리를 기반으로 관련 분야의 응용기술 발전에도 기여될 것으로 기대된다.

를 공간적으로 연계하여 공간가중회귀모형의 적용을 통해 매출 영향요인을 분석하고 그 결과물을 통해 다양한 활용방안을 제시하는 목적으로 수행하였다.

본 연구에서는 서울특별시를 중심으로 공간 빅데이터 기반 공간가중회귀모형(GWR)을 적용하여 기존 상권분석에서 도출하기 어려운 소지역 단위별 회귀계수와 국지적 R Square를 추정하였다. 소지역 단위 분석은 지역상권의 매출액 영향요인의 공간적 이질성을 효과적으로 파악하고 소지역별 차이와 상권분석과 관련된 논의의 폭을 확장하고자 하는 시도이다. 또한 기존 지역상권 분석연구가 개별 대상지나 시군구, 읍면동과 같은 특정 지역을 중심으로 이루어진 문제와 대비반수, 집계 데이터의 활용에 따른 한계를 벗어나 세분화된 지역을 도출하고자 하였다. 아울러 다양한 데이터의 처리 및 분석을 통해 도출한 결과를 지도상에 시각화함으로써 서울특별시의 매출 요인의 영향력 수준을 파악하고 실제 상권의 세분화된 분포특성 등을 실증적으로 제시하였다.

실증분석 결과 매출액에 미치는 변수는 지역에 따라 상이하게 나타났으며 본 연구에서 선정된 변수 이외에도 제 3의 변수가 존재함을 알 수 있었다. 특히 읍면동 단위의 비교를 통해 기존 공간 데이터의 분석단위에 따라 발생하는 오류를 최소화할 수 있는 결과를 도출하였다.

그간 정통적인 상권분석과 GIS 분석이 가지고 있는 한계를 극복할 수 있는 소지역 단위는 보다 세분화된 지역에서의 지역상권 매출 영향력을 체계적으로 분석할 수 있다. 이러한 분석모형은 대용량 데이터에 내포하고 있는 패턴을 발견하여 새로운 가치를 생성하는 가능성을 제공한다는 점에서 다양한 분야에서 활용될 것으로 기대된다.

### 謝 辭

이 논문은 2015년 정부(국토교통부)의 재원으로 공간정보 융복합핵심인재 양성사업의 지원을 받아 수행된 연구임(관리번호: 2015-06-03)

### V. 결 론

본 연구는 공공, 민간에서 구축·생성되는 카드매출 정보, 유동인구 등 공간 빅데이터를 활용하여 지역상권에 관한 인구, 고용, 소득 등의 매출요인을 도출하고 이

## 註

- 1) 공간 빅데이터 연구단은 국토교통부 국토공간정보연구사업 ‘국토공간정보의 빅데이터 관리, 분석 및 서비스 플랫폼 기술 개발’(16NSIP-B081011-03)을 말한다.

## 文獻

- 구자용, 2015, “공간정보 빅데이터의 지도화와 공간적 분포 특성에 관한 연구,” 국토지리학회지 49(3): 349-360.
- 김관호·오규협·이영규·정재운, 2013, “스마트카드 빅데이터를 이용한 서울시 지하철 이동패턴 분석,” 한국전자거래학회지 18(3): 211-222.
- 김도현·한혜근, 2014, 집합 대형할인점의 매출에 미치는 입지 결정요인, 집합건물법학, 13:123-152.
- 김미정·김대중·이영주, 2013, 과학적 국토정책을 위한 공간 빅데이터 활용방안, 경기:국토연구원.
- 김성문, 2014, 소매업태의 매출액 결정요인에 관한 연구, 건국대학교, 박사학위논문.
- 신우진·문소연, 2010, “기업형 슈퍼마켓의 상권 잠식 강도에 관한 연구,” 국토계획 45(7): 207-219.
- 안재성·이양원, 2015, “웹 지도 이용자의 관심집중지역 추출을 위한 빅데이터 플랫폼 기반의 공간 패턴 분석,” 국토지리학회지 47(4): 443-451.
- 이경주·홍성조·고석관, 2015, “공간가중회귀모형을 이용한 강원도 지역상권 현황진단 및 발전방안 연구,”

한국은행 지역본부 공동연구보고서 모음집 5(2): 241-300.

- 이승주, 2015, “주성분 분석을 이용한 빅데이터 분석,” 한국지능시스템학회지 25(6): 592-599.
- 정대석·김형보, 2014, “상권 업종별 분포 및 매출 영향요인 분석 경기도 31개 시군을 대상으로,” GRI 연구논총 16(2): 101-122.
- 정은애·성현곤·노정현, 2015, “공간자기상관성을 고려한 의류 소매업 매출액 영향요인분석,” 대한국토도시계획학회지 50(5): 215-231.
- 전해정, 2016, “GWR모형과 GIS를 이용한 주택가격 추정에 관한 연구. 주거환경학회,” 주거환경 통권 14(1): 1-11.
- 허자연·정연주·정창무, 2014, “소셜네트워크서비스와 상권 활성화에 관한 연구,” 관광경영연구 18(4): 80-82.
- IDC, 2011, Extracting Value from Chaos.
- Stefan, B., Davide, C., and Paul, W., 2013, Big Data: What's your plan?, McKinsey & Company, Accessed, Nov 15.
- Shashi Shekhar, 2012, Spatial Big Data, AAG-NIH Symp on Enabling a National Geospatial cyberinfrastructure for Health Research.

접 수 2019년 05월 21일  
최종수정 2019년 06월 10일  
게재확정 2019년 06월 14일