

# 공간적 자기상관성의 정도에 따른 MAUP에서의 스케일 효과 연구 - LBSNS 데이터를 중심으로 - A Study on Scale Effects of the MAUP According to the Degree of Spatial Autocorrelation - Focused on LBSNS Data -

이영민\* · 권 필\*\* · 유기윤\*\*\* · 허 옹\*\*\*\*  
Lee, Young Min · Kwon, Pil · Yu, Ki Yun · Huh, Yong

## 要 旨

포인트 속성의 위치 기반 소셜 네트워크 서비스(Location-Based Social Network Services, LBSNS) 데이터를 멀티스케일의 타일맵상에 효과적으로 시각화하기 위해서는 격자 기반으로 군집화하여 표현해야 할 필요성이 있다. 이때 격자의 크기 및 개수를 결정해야 하는데, 이에 대한 기준은 정해진 것이 없으며 데이터의 종류와 분석 목적에 따라 달라지므로 연구자의 주관에 개입될 수밖에 없다. 이때 연구 결과에 영향을 끼치는 공간단위 임의성의 문제(Modifiable Areal Unit Problem, MAUP)가 발생한다. 본 연구에서는 LBSNS 중 지오태깅(geotagging)된 트위터(Twitter) 데이터를 대상으로 하여 이러한 MAUP의 영향을 스케일 효과(scale effect)의 측면에서 탐색해 보고자 하였다. 이를 위해 공간오차모델(spatial error model)을 이용하여 데이터의 공간적 자기상관성(spatial autocorrelation)의 정도를 조절하였으며, 이에 대해 격자의 크기를 달리함에 따른 공간적 자기상관성의 변화를 Moran's  $I$ 를 통해 분석하였다. 실험 결과, 원 데이터에는 양의 공간적 자기상관성이 존재하는 것을 확인하였으며, 이러한 경우에는 공간오차모델의 공간자기회귀계수(spatial autoregressive coefficient)의 값이 증가할수록 공간적 자기상관성이 감소하는 것을 알 수 있었다. 이러한 특성을 이용하여 트위터 데이터의 공간적 자기상관성의 강도를 5단계로 조절하였으며, 각 단계에 대하여 격자의 크기를 9단계로 나누어 각각에서의 Moran's  $I$ 를 계산하였다. 그 결과, 함역 수준이 높아질수록 공간적 자기상관성이 증가하다가 격자의 크기가 600m에서 1,000m 사이일 때 감소하는 것을 알 수 있었으며, 공간적 자기상관성이 강할수록 MAUP에서의 스케일 효과는 감소하는 경향이 있는 것을 확인하였다.

핵심용어 : LBSNS, MAUP, Moran's  $I$ , 공간적 자기상관성, 공간오차모델

## Abstract

In order to visualize point based Location-Based Social Network Services(LBSNS) data on multi-scaled tile map effectively, it is necessary to apply tile-based clustering method. Then determining reasonable numbers and size of tiles is required. However, there is no such criteria and the numbers and size of tiles are modified based on data type and the purpose of analysis. In other words, researchers' subjectivity is always involved in this type of study. This is when Modifiable Areal Unit Problem(MAUP) occurs, that affects the results of analysis. Among LBSNS, geotagged Twitter data were chosen to find the influence of MAUP in scale effects perspective. For this purpose, the degree of spatial autocorrelation using spatial error model was altered, and change of distributions was analyzed using Moran's  $I$ . As a result, positive spatial autocorrelation showed in the original data and the spatial autocorrelation was decreased as the value of spatial autoregressive coefficient was increasing. Therefore, the

Received: 2016.01.20, accepted: 2016.03.01

\* 정회원 · 서울대학교 공과대학 건설환경공학부 박사과정(Member, Doctoral Student, Department of Civil and Environmental Engineering, Seoul National University, [daldanka@snu.ac.kr](mailto:daldanka@snu.ac.kr))

\*\* 정회원 · 서울대학교 공과대학 건설환경공학부 박사과정(Doctoral Student, Department of Civil and Environmental Engineering, Seoul National University, [pil0706@snu.ac.kr](mailto:pil0706@snu.ac.kr))

\*\*\* 정회원 · 서울대학교 공과대학 건설환경공학부 정교수(Member, Professor, Department of Civil and Environmental Engineering, Seoul National University, [kiyun@snu.ac.kr](mailto:kiyun@snu.ac.kr))

\*\*\*\* 교신저자 · 대한지적공사 공간정보연구원 선임연구원(Corresponding Author, Senior Researcher, LX Spatial Information Research Institute, [huhyong78@lx.or.kr](mailto:huhyong78@lx.or.kr))

intensity of the spatial autocorrelation of Twitter data was adjusted to five levels, and for each level, nine different size of grid was created. For each level and different grid sizes, Moran's  $I$  was calculated. It was found that the spatial autocorrelation was increased when the aggregation level was being increased and decreased in a certain point. Another tendency was found that the scale effect of MAUP was decreased when the spatial autocorrelation was high.

**Keywords :** LBSNS, MAUP, Moran's  $I$ , spatial autocorrelation, spatial error model

## 1. 서론

GPS 기술의 발달과 스마트 기기의 보급 확산으로 인해 트위터(Twitter), 페이스북(Facebook), 인스타그램(Instagram) 등 위치 기반 소셜 네트워크 서비스(Location-Based Social Network Services, 이하 LBSNS)에 대한 사용이 활발히 이루어지고 있다. LBSNS 사용자들은 지오태깅(geotagging)<sup>1)</sup> 기능을 활용함으로써 자신이 작성하는 콘텐츠와 관련된 위치정보(경위도좌표)를 함께 업로드할 수 있다. 그렇기 때문에 LBSNS를 통해 생성된 데이터들은 포인트의 속성을 지니는 공간 객체라고 볼 수 있다.

LBSNS에서 추출한 포인트 데이터를 타일맵 형태로 제공되는 멀티스케일의 지도상에 나타내고자 할 때, 정보의 가독성을 위하여 군집화(clustering)하여 표현해야 한다. 여러 군집화 기법 중 격자(grid) 기반의 군집화 기법은 대상 공간을 일정한 격자 구조로 구성하고, 모든 군집화 과정을 이 격자 구조 내에서 실행하는 방식으로, 데이터 객체 수에 대해 독립적이고 미리 계산된 격자의 수에 의존하기 때문에 계산량이 적고 처리 속도가 빠르다는 장점이 있다. LBSNS 데이터에 대하여 격자 기반 군집화 작업을 수행하기 위해서는 지도의 각 줌 레벨(zoom-level)별로 적절한 격자의 크기 및 개수를 결정해야 한다. 그러나 이러한 격자 크기에 대한 기준은 정해진 것이 없고 데이터의 성격과 활용 목적에 따라 달라지므로 연구자의 주관에 개입될 수밖에 없다. 이때 분석 결과에 영향을 미칠 수 있는 ‘공간단위 임의성의 문제(Modifiable Areal Unit Problem, 이하 MAUP)’가 발생한다. 본 연구에서는 이러한 문제를 최소화하기 위한 과정으로써 LBSNS 중 트위터 포인트 데이터에 대하여 MAUP의 영향을 스케일 효과(scale effect)의 측면에서 알아보고자 하였다.

이를 위해 공간오차모델(spatial error model)을 이용하여 트위터 포인트 데이터에 대해 공간적 자기상관성(spatial autocorrelation)의 정도에 따른 분포를 생성하고, 각각에 대하여 격자의 크기를 달리함에 따른 공간적 자기상관성의 변화를 Moran's  $I$ 를 통해 알아보고자

하였다.

본 논문의 구성은 다음과 같다. 2장에서는 MAUP와 관련된 이론적 배경 및 선행연구를 고찰하였으며, 3장에서는 공간적 자기상관성의 정도에 따른 MAUP에서의 스케일 효과 탐색을 위한 실험 과정 및 결과를 작성하였다. 4장에서는 본 연구에 대한 결론과 향후 연구 과제를 도출하였다.

## 2. 이론적 배경 및 선행연구 고찰

### 2.1 MAUP의 개념

트위터, 페이스북, 인스타그램 등 위치정보를 포함하는 LBSNS 데이터들은 해당 데이터가 참조하는 장소에 대한 좌표 정보를 포함하고 있기 때문에 이를 공간상에 표현할 때 포인트의 형태를 취하게 된다. 어떠한 공간적 범위, 예를 들면 서울시를 대상으로 수집된 포인트 데이터를 특정 공간 단위, 예를 들면 행정동 또는 격자를 기준으로 분석하기 위해서는 포인트 속성의 원데이터를 폴리곤 속성의 데이터에 결합하여 사용해야 한다. 이때 연구자는 분석에 사용할 공간 단위를 선택하거나 기존의 공간 단위를 조작하여 새로운 공간 단위를 만들어내야 하는데(Lee, 1999), 이에 대한 명확한 기준은 정해진 것이 없기 때문에 연구자의 판단에 따를 수밖에 없다.

새로운 공간 단위를 구축하는 문제는 지역구분(regionalization)과 밀접한 관련성을 가지는데, 이는 실질적으로 작은 스케일의 지역을 큰 스케일의 지역으로 합역(spatial aggregation)하는 것을 의미한다. 그런데 이때 합역양식(aggregation scheme)에 따라서 동일한 지역을 다양한 방식으로 나눌 수 있다. 즉 합역양식에 의해 다양한 스케일의 지역을 생성할 수도 있고, 같은 스케일에서 서로 다른 구획을 가지는 지역을 생성할 수도 있는 것이다(Lee, 1999).

이러한 합역 과정에서 발생하는 문제점을 MAUP라고 하는데, Openshaw(1984)에 의하면 “연구에 사용되는 공간 단위의 선정은 기본적으로 임의적이며, 연구의 결과는 선정된 공간 단위와 체계적인 관련성을 가진다.” 이렇게 분석에 사용되는 공간 단위를 어떻게 설정하느냐에 따라 분석 결과에 영향을 끼치는 것을 MAUP라고 한다.

1) 글, 사진, 동영상 등 디지털 매체 내에 해당 데이터와 관련된 위치정보를 삽입시키는 것

MAUP는 대부분 공간 단위가 자의적으로 한정된 상태에서 가변적이라는 사실에서 기인하며, 그렇기 때문에 공간 단위는 상이한 스케일이나 구획을 만들기 위해서 합여되거나 형태가 변형될 수 있다(Jelinski and Wu, 1996). 즉 MAUP는 스케일 효과(scale effect)와 구획 효과(zoning effect)의 두 가지 관점을 갖는다. 스케일 효과는 서로 다른 스케일의 공간 단위를 대상으로 분석을 실시했을 때 스케일에 따라 결과가 달라지는 것을 의미하며, 구획 효과는 동일한 스케일에서 공간 단위의 재그룹화를 통해 발생하는 서로 다른 결과를 말한다(Kwan and Weber, 2008).

MAUP가 발생하는 가장 큰 이유는 공간 단위들의 합여이 이루어질 때 스케일 효과와 구획 효과로 인하여 변수들의 변량과 공변량이 변화하기 때문이다. 즉, 합여는 일반적으로 변수의 변량을 감소시키게 되는데, 이는 공간 단위들이 합쳐지면서 균질효과(smoothing effect)가 발생되기 때문이다. 예를 들어, 합여이 이루어지기 전의 수치적 이상치들(outliers)이 다른 공간 단위의 값들과 합쳐지면서 평균을 향해 수렴되는 경향이 있기 때문이다(Lee, 1999).

## 2.2 선행연구 고찰

MAUP에 대한 연구는 국외뿐 아니라 국내에서도 지리학과 공간정보공학 분야를 중심으로 활성화되어 있으며(Cheong and Park, 2015), 스케일 효과와 구획 효과의 관점에서 다양한 연구가 진행되고 있다. 예컨대, Lee(1999)는 미국의 기능지역 설정 방식에 초점을 두고, 공간 단위를 합여하는 과정에서 MAUP가 발생한다고 보았다. 또한 MAUP는 스케일 효과와 구획 효과로 구분되어 나타난다고 보고, 같은 변수에 대해 각각 다른 기능지역을 설정하여 결과를 비교하였다. 결과적으로는 통계치가 합여양식에 따라 변화하는 것을 지적하기 보다는 다양한 통계치 중에서 유의미한 것은 무엇이며, 그 통계치를 도출한 공간 단위의 합여양식이 무엇인지를 평가할 수 있는 기준을 제시하고자 하였다.

보다 실증적인 연구로써 Oh(2002)는 시공간적 분석에서 MAUP를 보완할 수 있는 방법에 대해 연구하였다. 이를 위해 최소 크기의 격자 단위로 지가 변동에 대한 시공간 데이터베이스와 지도를 구현하였으며, 분석 기준 시점에서의 속성 데이터와 공간 데이터를 일치시킴으로써 MAUP를 보완할 수 있음을 밝혔다. 또한 Lee and Kim(2013)은 개별공시지가와 주택실거래가 사이의 공간적 불일치에 주목하였으며, 이러한 차이에 대한 공간적 변이를 시각적으로 제시하였다. 또한 개별공시지가나 개별주택가격 산정에 고려될 수 있는 지역의 특

성 변수들의 MAUP 효과를 실증적으로 보여주었다.

국의 선행연구 중 본 연구에서와 같이 스케일 효과에 중점을 둔 연구로, He et al.(2007)은 MAUP가 생태학 분야의 연구에서 차지하는 중요성을 인지하고, 식물 군집의 실제 공간 분포 패턴을 가장 잘 나타내는 최적의 스케일을 탐색하기 위하여 MAUP의 스케일 효과를 이용하였다. 또한 Viegas et al.(2007)은 교통분석권역을 설정할 때 MAUP의 영향을 최소화 하고자 관련 변수에 대해 MAUP의 스케일 효과를 분석하였다. Swift et al.(2008)은 마시는 물의 수질과 위장병 간의 상관관계에 대하여 공간 합여의 영향과 MAUP의 효과를 알아보고자 하였으며, 이를 위해 스케일 효과와 구획 효과를 모두 분석할 수 있는 9개의 서로 다른 공간 단위를 구성하고 그 결과를 비교하였다.

이렇듯 공간 분석 및 통계적 분석 결과에 영향을 미치는 MAUP의 일반적인 작동 기제에 대해서는 다양한 분야에서 오랫동안 연구되어 왔지만 이를 LBSNS 데이터에 적용한 사례는 찾아보기 힘들다. MAUP와 관련된 연구는 1980년대부터 꾸준히 진행되어 왔지만 그에 반해 LBSNS라는 것은 2000년대 초반, 스마트폰의 등장 이후에 생겨난 것이기 때문에 두 분야의 접점이 없었던 것이 그 이유이다. 따라서 본 연구는 새로운 형태의 데이터라고 할 수 있는 LBSNS 데이터를 대상으로 공간적 자기상관성에 따른 MAUP의 효과를 탐색하는데 의의를 갖는다.

## 3. 공간적 자기상관성의 정도에 따른 MAUP에서의 스케일 효과 탐색

### 3.1 사용 데이터 및 데이터 전처리

실험에 사용한 LBSNS 데이터는 246,984개의 지오태깅된 트위터 포인트 데이터이며, 이에 대한 분포는 Fig. 1과 같다.

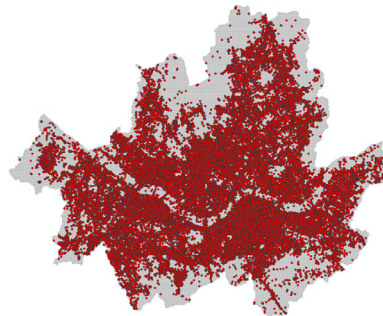


Figure 1. The spatial distribution of original Twitter point data

이러한 포인트 속성의 데이터를 MAUP의 스케일 효과 분석에 이용하기 위해서는 이를 폴리곤 속성의 데이터에 병합하는 과정이 필요하다. 이때 변수는 각 폴리곤 내에 포함된 포인트의 개수이므로 카운트 데이터의 속성을 지니게 된다. 특정 공간 단위로 합산된 카운트 속성의 데이터는 LBSNS 데이터의 가장 일반적인 형태라고 할 수 있다. 이러한 카운트 데이터를 대상으로 공간적 군집성 여부를 파악할 때는 해당하는 카운트 데이터가 서로 비교 가능한 형태로 제시되어야 할 필요성이 있다. 그렇기 때문에 변수는 기본적으로 빈도나 비율 데이터의 형태를 취하게 되는데, 이때 비율 데이터의 유형은 비율, 밀도, 행비중, 열비중으로 더욱 세분화할 수 있다(Cho, 2013). 행비중은 관심 집단이 전체 공간단위 내에서 차지하는 비중을, 열비중은 관심 집단이 공간단위 간에서 갖는 비중을 의미하며(Lee, 2007), 행비중은 특화도(specialization), 열비중은 집중도(concentration)에 초점이 맞춰져 있다고 볼 수 있다(Cho, 2013).

본 연구는 특정 관심 집단이 전체 공간에서 차지하는 공간적 특화도를 파악하는 것이라기 보다는 공간단위 간의 비교를 통해 LBSNS 데이터가 보다 활발한 공간단위, 즉 집중도를 알고자 하는 것에 가깝기 때문에 열비중 비율을 사용하였다. 따라서 모든 공간 단위에 대하여 해당 공간 단위 내에 포함된 트위터 데이터의 개수를 전체 트위터 데이터의 개수(246,984)로 나누었으며, 그 후 이에 대해 표준점수(z-score)를 이용한 표준화를 실시하였다.

### 3.2 공간적 자기상관성의 정도에 따른 MAUP에서의 스케일 효과 분석

#### 3.2.1 공간오차모형을 이용한 공간적 자기상관성의 정도 조절

식 (1)과 같이 벡터  $y$ 를 이용하면 격자의 공간적 자기상관성의 정도를 조절할 수 있다(Anselin, 1995). 이는 공간오차모형에서 잔차를 분리해낸 것으로, 공간적 자기상관성이 있는 잔차와 없는 잔차를 구분하는 데 쓰이기도 한다.

$$y = (I - \rho W)^{-1} \epsilon \quad (1)$$

여기서  $I$ 는 전체 공간 단위 개수만큼의 배열을 가지는 단위행렬이며,  $\rho$ 는 공간자기회귀계수(spatial autoregressive coefficient)로, 0.0과 0.9 사이의 값을 가진다. 본 연구에서는  $\rho$ 값으로 0.2, 0.3, 0.4, 0.5, 0.6을 사용하였으며, 이때  $\rho$ 의 값이 클수록 공간적 자기상관성이 강한 분포를 만들어낼 수 있다. 그러나 벡터  $\epsilon$

에 공간적 자기상관성이 존재하는 경우에는 그 반대임을 확인할 수 있었는데, 이는 뒤에서 보다 자세히 다루도록 한다.  $W$ 는 행표준화된 공간가중행렬,  $\epsilon$ 는 해당 변수값을 의미한다.

#### 3.2.2 Moran's $I$ 를 이용한 공간적 자기상관성 분석

공간 데이터는 순수한 자신만의 정보를 가지고 있을 뿐 아니라 지리적인 정보를 함께 포함하고 있다는 점에서 다른 데이터들과 구별된다. Lee(2001)에 따르면 이렇게 위치 정보가 포함된 데이터들은 공간적 의존성(spatial dependence)을 가지고 있기 때문에 서로 독립적으로 존재할 수 없다. 만약 전통적인 선형분석 방법들로 공간 데이터를 분석하는 경우, 사회경제현상이나 인구현상 또는 자연현상이 공간상에서 나타내는 특이인 공간적 의존성 및 상호작용을 통제하지 못한다(Doreian, 1981). 즉 Tobler(1970)의 지리의 제1법칙에서와 같이 공간상의 객체들은 공간상에 무작위(random)하게 있지 않고, 서로 간에 영향을 주고받으며 존재한다고 할 수 있다. 이와 같이 지리적 공간상에서의 공간 객체 간 상호 의존성을 공간적 자기상관성이라고 한다. 일반적으로 통계 분야에서 사용하는 상관계수는 두 변수 간에 존재하는 관계의 특성과 정도에 대한 정보를 제공하는 반면에 공간적 자기상관성은 두 변수 사이의 관계가 아닌 하나의 변수가 여러 관측치들 사이에서 갖는 관계를 의미한다(Jung and Son, 2009).

Moran's  $I$ 는 이러한 공간적 자기상관성을 전역적(global)으로 측정하는 도구로, 거의 -1에서 1 사이의 값을 가지며, -1에 가까울수록 음의 상관관계, 1에 가까울수록 양의 상관관계, 그리고 0은 랜덤하다는 것을 의미한다(Moran, 1950).

공간적 자기상관성에 따른 MAUP에서의 스케일 효과를 알아보기 위하여 식 (1)을 통해  $\rho$ 값 별로 계산된  $y$ 에 대하여 격자의 크기를 달리함에 따른 Moran's  $I$ 값을 계산하였다. Moran(1950)에 의하면 Moran's  $I$ 는 식 (2)와 같이 계산할 수 있다.

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (2)$$

여기서  $w_{ij}$ 는 공간가중행렬(spatial weights matrix),  $n$ 은  $i$ 와  $j$ 로 이루어진 모든 공간 단위의 개수,  $x_i$ 는  $i$ 번째 변수,  $\bar{x}$ 는  $x$ 의 평균을 의미한다.

### 3.3 실험 적용 및 결과

실험을 위하여 서울시를 공간적 범위로 하는 격자 데이터를 200m부터 1,000m까지 격자의 크기(한 변의 길이,  $l$ )별로 생성하였다. 격자의 크기가 커질수록 합역의 수준이 높아지므로 전체 격자의 개수는 적어지게 된다. 격자의 크기에 따른 격자 개수( $n$ )는 Table 1과 같다.

공간결합연산(spatial join)을 통해 격자 데이터에 대하여 격자 크기별로 각각 트위터 포인트 데이터를 결합하였다. 그리고 앞에서 언급한 카운트 데이터의 표준화 방식을 이용하여 격자의 크기별로 변수를 표준화하였다. 그 후에 Moran's  $I$ 를 이용하여 공간결합된 본래 데이터의 공간적 자기상관성을 계산하였다. 이때 공간인접관계(spatial contiguity relationship)는 상하좌우로만 이동하는 룩(look)의 움직임 방식을 이용하였으며, 공간가중행렬은 경계를 접하고 있는 공간 단위 쌍에 대해서는 1의 값을, 그렇지 않은 경우에는 0의 값을 부여하는 단순이항가중(simple binary weighting) 방식을 사용하였고, 각 값은 행표준화(row-standardization)되었다. 실험은 ArcMap 10.3의 ArcToolbox, GeoDa, 그리고 MATLAB R2013a를 이용하여 진행되었다.

실험 결과, Table 2와 Fig. 2를 보면, 본래 데이터에 99%의 유의확률로 공간적 자기상관성이 존재하며 이는 격자의 크기가 증가할수록 함께 증가하는 경향을 보이는 것을 알 수 있다. 이를 좀 더 자세히 살펴보면, 격자의 크기가 증가할수록 Moran's  $I$ 의 값이 계속적으로 증가하다가 800m에서 감소하였으며, 1,000m에서 다시  $I$ 의 값이 증가하는 것을 알 수 있다. 이와 같은 결과는

Table 1. The number of grids( $n$ ) according to sizes of grids( $l$ )

$l$ (m)	200	300	400	500	600	700	800	900	1,000
$n$	15,708	7,099	4,063	2,625	1,855	1,380	1,074	858	697

Table 2. Result for Moran's  $I$  calculation of original data

$l$ (m)	Moran's $I$	z-score	p-value
200	0.418060	65.550872	0.000000
300	0.436099	51.335283	0.000000
400	0.479180	42.600342	0.000000
500	0.506554	36.078012	0.000000
600	0.521796	31.255140	0.000000
700	0.559468	28.802971	0.000000
800	0.525105	23.873066	0.000000
900	0.487763	19.880516	0.000000
1,000	0.542454	19.916419	0.000000

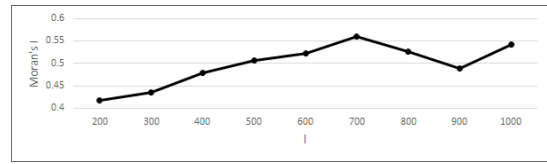


Figure 2. Changes of Moran's  $I$  of original data according to sizes of grids( $l$ )

공간 단위가 커질수록 공간적 의존성은 감소하는 경향이 있다는 일반적인 사실(Can 1996; Lee, 1999; Getis, 2010)과는 반대되는 것으로, 이는 본래 데이터가 가지고 있는 공간적 의존성이 매우 강하기 때문인 것으로 보인다.

원래는 식 (1)에서  $\rho$ 의 값을 증가시킬수록 공간적 자기상관성의 정도가 증가하나, 본래 데이터에 공간적 자기상관성이 존재하는 경우에는 반대의 결과를 얻을 수 있음을 추가 실험을 통해 확인할 수 있었다. 즉 식 (1)에서 본래 변수인 벡터  $\epsilon$ 에 공간적 자기상관성이 존재하는 경우,  $\rho$ 의 값이 증가할수록 공간적 자기상관성은 감소하게 된다. 본 연구의 목적은 공간적 자기상관성의 정도에 따라 구성된 분포에 대하여 각각에서 격자의 크기 변화가 공간적 자기상관성에 미치는 영향을 파악하기 위한 것이므로 원 데이터의 공간적 자기상관성을 제거하기보다는 공간적 자기상관성의 정도를 조절하는 것에 의미를 두었다.

$\rho$ 값을 0.2부터 0.6까지 차례로 적용한 결과, 본래 값, 즉  $\rho$ 가 0.0일 때의 결과(Table 2, Fig. 2)보다 격자의 크기에 따른 Moran's  $I$ 값의 변동이 작아진 것을 확인할 수 있었다. 이는 본래 값의 분포는 왼쪽으로 치우친 포아송 분포를 따르는 반면  $\rho$ 값에 의해 재조정된 변수 값은 정규분포를 따르기 때문이다.

Table 3의 Moran's  $I$ 값을 보면, 모든  $\rho$ 값에 대하여 격자의 크기별 값의 분포는 모두 양의 공간적 자기상관성을 보이는 것을 알 수 있다. 이를 보다 상세히 살펴보면, 먼저  $\rho$ 값이 0.2일 때는 격자의 크기가 증가할수록 Moran's  $I$ 의 값도 계속적으로 함께 증가하다가  $l$ 의 값이 1,000m일 때  $I$ 의 값이 감소하였다.  $\rho$ 값이 0.3일 때 역시 마찬가지로 Moran's  $I$ 의 값이 격자의 크기와 함께 증가하다가 900m부터 값이 감소하는 것을 알 수 있다.  $\rho$ 값이 0.4일 때를 보면, 격자의 크기가 증가함에 따라 Moran's  $I$ 값이 감소와 증가를 반복하고 있으나 전체적으로는 증가하는 경향성을 보이다가 800m에서 감소하는 모습을 보인다.  $\rho$ 값이 0.5일 때 역시 Moran's  $I$ 의 값이 감소와 증가를 반복하다가 600m에서 값이 비교적 큰 폭으로 감소하였으며,  $\rho$ 값이 0.6일 때는 격자

Table 3. Result for Moran's  $I$  according to spatial autoregressive coefficient( $\rho$ ) and the sizes of grids( $l$ )

$\rho$	$l(m)$	Moran's $I$	z-score	p-value
0.2	200	0.754630	117.789053	0.000000
	300	0.755241	88.489484	0.000000
	400	0.756396	66.770230	0.000000
	500	0.757361	53.583998	0.000000
	600	0.767341	45.439262	0.000000
	700	0.772549	39.268824	0.000000
	800	0.779344	34.823289	0.000000
	900	0.787925	31.333441	0.000000
	1,000	0.787806*	28.245810	0.000000
	0.3	200	0.581416	90.753944
300		0.596325	69.871589	0.000000
400		0.600067	52.976201	0.000000
500		0.603825	42.725119	0.000000
600		0.615643	36.471461	0.000000
700		0.616982	31.362866	0.000000
800		0.630710	28.182195	0.000000
900		0.618753	24.609169	0.000000
1,000		0.611096	21.925774	0.000000
0.4		200	0.513682	80.182714
	300	0.506261	59.321032	0.000000
	400	0.513879	45.373947	0.000000
	500	0.513378	36.330464	0.000000
	600	0.516336	30.592129	0.000000
	700	0.537165	27.318495	0.000000
	800	0.527922	23.597647	0.000000

0.4	900	0.515875	20.530277	0.000000
	1,000	0.525191	18.846444	0.000000
0.5	200	0.448295	69.977094	0.000000
	300	0.442765	51.884657	0.000000
	400	0.450537	39.780953	0.000000
	500	0.461732	32.676939	0.000000
	600	0.454729	26.941664	0.000000
	700	0.466822	23.748966	0.000000
	800	0.465328	20.805845	0.000000
	900	0.470371	18.730099	0.000000
	1,000	0.474174	17.018611	0.000000
	0.6	200	0.404244	63.102140
300		0.405854	47.559697	0.000000
400		0.413898	36.547982	0.000000
500		0.431003	30.508900	0.000000
600		0.408442	24.203580	0.000000
700		0.415998	21.163569	0.000000
800		0.411032	18.381443	0.000000
900		0.418718	16.671687	0.000000
1,000		0.406748	14.609953	0.000000

\* The values which are reduced significantly are represented by bold type.

크기가 500m까지 Moran's  $I$ 의 값이 증가하다가 600m에서 큰 폭으로 감소한 후에 감소와 증가를 반복하고 있다(Table 3, Fig. 3).

$\rho$ 값과  $l$ 값에 따른 Moran's  $I$ 의 전체적인 경향성을 보면, 합역의 수준이 높아질수록, 즉 격자의 개수는 줄어들고 크기는 증가할수록 Moran's  $I$ 의 값이 커지다가

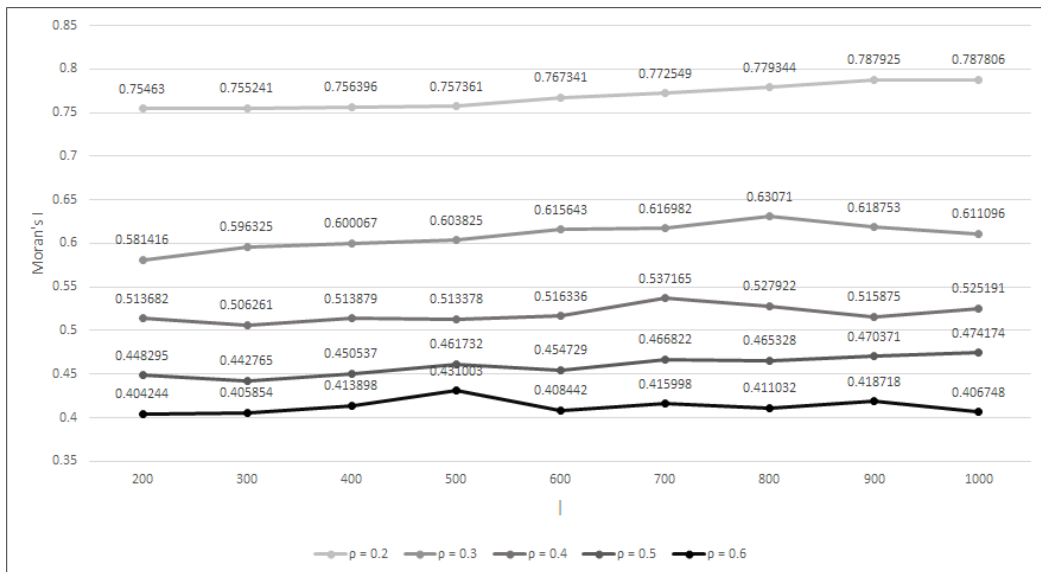


Figure 3. Changes of Moran's  $I$  according to spatial autoregressive coefficient( $\rho$ ) and the sizes of grids( $l$ )

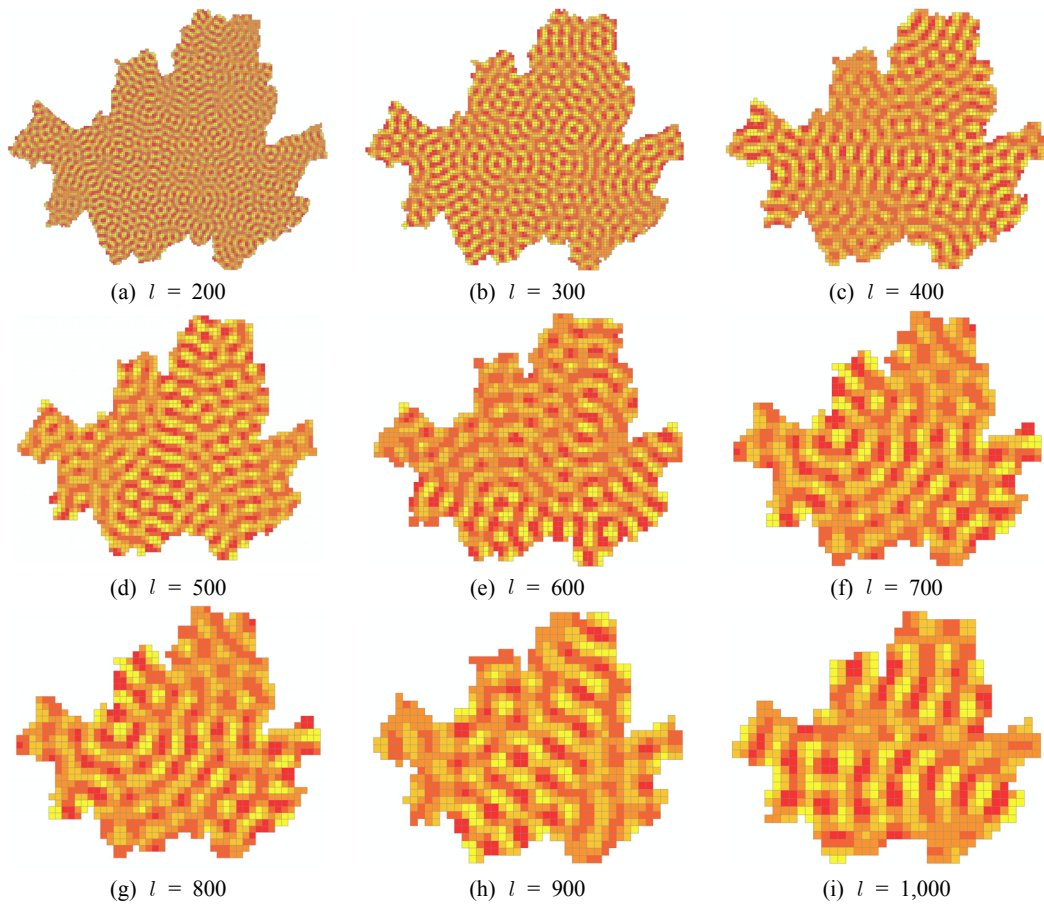


Figure 4. Distribution maps according to the size of the grids( $l$ ) when  $\rho = 0.4$

$l$ 의 값이 600m에서 1,000m 사이에서 감소하는 것을 알 수 있다. Moran's  $I$ 의 값이 증가한다는 것은 공간적 자기상관성이 증가한다는 것인데, 이는 곧 합역이 진행될수록 인접한 공간 단위들 간의 수치적 유사성이 증가한다는 것을 의미한다. 이때 공간적 자기상관성이 강할수록 값이 변화하는 지점인 변곡점이 우측으로 이동하는 것을 확인할 수 있었다(Fig. 3). 이러한 변곡점은 MAUP의 영향력을 최소화할 수 있는 최적의 공간 단위를 추측하는 데 도움을 줄 수 있다. 공간적 자기상관성이 최소가 되는 지점이 MAUP의 영향을 가장 적게 받는 지점이라고 볼 수 있기 때문이다. 그러나 세부적인 경향성을 보면 격자의 크기가 변화할 때마다 공간적 자기상관성의 값은 증가와 감소를 반복하였다.

Fig. 4(a)에서부터 Fig. 4(i)는  $\rho$ 값이 0.4일 때 격자의 크기에 따라 변화하는 분포도를 나타내고 있는데, 격자의 크기가 변화함에 따라 전체적인 분포 패턴 또한 함께 변화하고 있는 것을 알 수 있다.

위의 결과에 대하여 공간적 자기상관성에 따른 MAUP에서의 스케일 효과를 알아보기 위하여 각  $\rho$ 값 별로 격자의 크기에 의한 Moran's  $I$ 값에 대한 편차의 평균을 계산해 보았다. 그 결과, 공간적 자기상관성이 강할수록, 즉  $\rho$ 의 값이 작을수록 MAUP에서의 스케일 효과는 감소하는 경향이 있는 것을 확인할 수 있었다 (Table 4).

Table 4. Result for the average of deviation of Moran's  $I$

$\rho$	Average of deviation of Moran's $I$
0.2	0.00417675
0.3	0.00861350
0.4	0.00874163
0.5	0.00674162
0.6	0.01018725

#### 4. 결론

본 연구는 포인트 속성을 지니는 LBSNS 데이터를 멀티스케일의 지도상에 군집화하여 표현하고자 할 때 발생할 수 있는 MAUP를 최소화 할 수 있는 방안을 알아보고자 하는 목적을 가지고 있다. 이에 따라 LBSNS 중 트위터 포인트 데이터에 대하여 MAUP에서의 스케일 효과를 공간적 자기상관성의 측면에서 알아보았다. 이를 위해 공간오차모형을 이용하여 원 데이터의 공간적 자기상관성의 정도를 조절하였으며, Moran's  $I$ 를 이용하여 격자의 크기를 달리함에 따른 공간적 자기상관성의 변화를 파악하였다.

실험 결과, 원 데이터에 공간적 자기상관성이 강하게 존재할 때에는 공간오차모형의 공간자기회귀계수( $\rho$ )값이 증가함에 따라 공간적 자기상관성이 감소하는 결과를 보이는 것을 확인하였다. 이를 이용하여  $\rho$ 값을 0.2에서부터 0.6까지 차례로 적용함으로써 트위터 데이터에 대해 공간적 자기상관성의 강도를 조절하였고, 각각에 대하여 격자의 크기별로 Moran's  $I$ 의 값의 변화를 살펴보았다. 그 결과, 격자의 합역 수준이 높아질수록, 즉 격자의 크기가 증가할수록 공간적 자기상관성이 함께 증가하다가  $I$ 값이 600m에서 1,000m 사이에서 감소하는 전체적인 경향성을 확인할 수 있었다. 또한 공간적 자기상관성이 강할수록 MAUP에서의 스케일 효과는 감소하는 경향이 있는 것으로 확인되었다. 그러나 이러한 결과를 모든 데이터에 대하여 일반화하기는 어려우며, 데이터의 종류와 특성에 따라 달라질 수 있음을 밝혀둔다.

LBSNS 데이터를 격자 기반으로 클러스터링하여 표현하고자 할 때 MAUP의 효과를 최소화할 수 있는 격자의 크기를 결정하는 데 이와 같은 방법론이 하나의 척도가 될 수 있을 것으로 보인다. 또한 본 연구를 통해 통계적 결과는 공간 단위의 선택에 의존적이라는 것을 알 수 있었다.

또한 본 연구에서는 양의 공간적 자기상관성이 존재하는 경우에 대해서만 그 강도를 조정하여 실험을 진행하였는데, 추후에는 공간적 자기상관성이 존재하지 않는 경우와, 음의 공간적 자기상관성이 존재하는 경우에 대한 연구도 추가적으로 진행되어야 할 것으로 보인다. 비록 공간 데이터의 경우, 음의 공간적 자기상관성이 존재하는 경우가 드물기는 하지만 서로 다른 공간적 자기상관성이 존재할 때 MAUP의 효과가 어떻게 변화하는지를 살펴보는 것은 의미가 있을 것이다. 또한 MAUP에서의 스케일 효과뿐 아니라 구획 효과에 대한 연구도 함께 이루어져야 할 것이다. 뿐만 아니라

LBSNS 데이터가 MAUP에서 가지는 특성을 알기 위해서는 다른 인문사회적 데이터와의 비교분석도 이루어져야 할 것으로 생각된다.

공간 데이터를 다룰 때 필연적으로 발생하는 MAUP 효과는 다루기 어려운 문제임에 틀림이 없다. 그럼에도 불구하고 MAUP는 분석 결과의 신뢰도를 치명적으로 감소시킬 수 있는 중요한 문제이기 때문에 연구자들의 지속적인 관심이 필요한 분야이다. 따라서 공간 단위의 합역으로 인한 군집효과와 공간적 자기상관성을 감소시킬 수 있는 합역양식을 적용함으로써 MAUP에 대한 영향력을 최소화하는 방법론에 대한 연구가 계속적으로 진행되어야 할 것이다.

#### 감사의 글

본 연구는 국토교통부 국토공간정보연구사업의 연구비지원(15CHUD-C061156-05)에 의해 수행되었습니다.

#### References

1. Anselin, L., 1995, Local indicators of spatial association-LISA, *Geographical Analysis*, Vol. 27, No. 2, pp. 93-115.
2. Can, A., 1996, Weight matrices and spatial autocorrelation statistics using a topological vector data model, *International Journal of Geographical Information Systems*, Vol. 10, No. 8, pp. 1009-1017.
3. Cheong, J. S. and Park, J. H., 2015, Test of the scale effect of MAUP in crime study: analyses of sex crime using nation-wide data of eup-myon-dong and si-gun-gu, *The Journal of the Korea Contents Association*, Vol. 15, No. 10, pp. 150-159.
4. Cho, D. H., 2013, Trends and methodological issues in spatial cluster analysis for count data, *Journal of the Korean Geographical Society*, Vol. 48, No. 5, pp. 768-785.
5. Doreian, P., 1981, Estimating linear models with spatially distributed data, *Sociological Methodology* 12, pp. 359-388.
6. Getis, A., 2010, Spatial autocorrelation, In *Handbook of applied spatial analysis*, Springer Berlin Heidelberg, pp. 255-278.
7. He, Z., Zhao, W. and Chang, X., 2007, The modifiable areal unit problem of spatial heterogeneity of plant community in the transitional zone between



- oasis and desert using semivariance analysis, *Landscape Ecology*, Vol. 22, No. 1, pp. 95-104.
8. Jelinski, D. E. and Wu, J., 1996, The modifiable areal unit problem and implications for landscape ecology, *Landscape Ecology*, Vol. 11, No. 3, pp. 129-140.
  9. Jung, D. Y. and Son, Y. G., 2009, A analysis on the spatial features of the neighborhood trade area using positive spatial autocorrelation method, *Journal of the Korean Society for Geospatial Information System*, Vol. 17, No. 1, pp.141-147.
  10. Kwan, M. P. and Weber, J., 2008, Scale and accessibility: Implications for the analysis of land use-travel interaction, *Applied Geography*, Vol. 28, No. 2, pp. 110-123.
  11. Lee, S. I., 1999, The delineation of function regions and Modifiable Areal Unit Problem(MAUP), *Journal of Geographic and Environmental Education*, Vol. 7, No. 2, pp. 757-783.
  12. Lee, S. I., 2001, Developing a bivariate spatial association measure: an integration of Pearson's  $r$  and Moran's  $I$ , *Journal of Geographical Systems*, Vol. 3, No. 4, pp. 369-385.
  13. Lee, S. I., 2007, Spatial statistical approach to residential differentiation (1): Developing a spatial separation measure, *Journal of the Korean Geographical Society*, Vol. 42, No. 4, pp. 616-631.
  14. Lee, G. H. and Kim, K. Y., 2013, A study on the spatial mismatch between the assessed land value and housing market price: Exploring the scale effect of the MAUP, *Journal of the Korean Geographical Society*, Vol. 48, No. 6, pp. 879-896.
  15. Moran, P. A., 1950, Notes on continuous stochastic phenomena, *Biometrika*, Vol. 37, No. 1/2, pp. 17-23.
  16. Oh, C. W., 2002, Spatio-temporal analysis of land price variation considering modifiable area unit problem, *The Journal of Geographic Information System Association of Korea*, Vol 10, No. 2, pp. 185-199.
  17. Openshaw, S., 1984, The modifiable areal unit problem, *Geo Abstracts University of East Anglia*.
  18. Swift, A., Liu, L. and Uber, J., 2008, Reducing MAUP bias of correlation statistics between water quality and GI illness, *Computers, Environment and Urban Systems*, Vol. 32, No. 2, pp. 134-148.
  19. Tobler, W. R., 1970, A computer movie simulating urban growth in the Detroit region, *Economic Geography*, Vol. 46, pp. 234-240.
  20. Viegas, J. M. and Martinez, L. M., 2007, Effects of the modifiable areal unit problem on the delineation of traffic analysis zones, *Environment and Planning, B: Planning and Design*, advance online publication, pp. 1-20.