

GIS의 통계적 공간분석에 관한 연구

유 은 혜

서울대학교 지리학과 박사과정

1. 서론

1) 연구 배경 및 목적

위치 정보를 포함하고 있는 지리적 자료의 특성 상 전통적인 통계분석 기법의 적용은 통계적 모형의 유효성에 대한 그릇된 판별력을 갖게 함으로써 오인된 통계적 추론을 유도할 수 있다(Fischer, 1996). 그러나 최근 사회적 관심이 집중되고 있는 오염물질의 분포나 행정 구역 단위로 집계된 질병발생 자료에 대해 발표되고 있는 대부분의 연구들은 이와 같은 지리적 자료의 '공간 효과(spatial effects)'를 무시한 채 자료의 특성을 정확히 반영하지 못하고 있다.

이것은 주로 자료의 공간적 특성을 명확히 규명할 수 있는 용이하고 효과적인 방안이 부족하기 때문이다(Anselin, 1992). 물론, 최근 GIS 기술의 발전에 힘입어 이러한 문제에 대한 인식 및 그 해결을 위한 노력이 미국이나 영국 등에서 1990년대 초부터 본격적으로 이루어지고 있다¹⁾. 그러나 국내에서는 아직 이에 대한 선행 연구는 물론 그 필요성에 대한 인식마저도 부족한 실정이다.

이에 본 논문에서는 GIS를 활용하여 지리적 자료, 특히 사회 과학 분야에서 널리 활용되고 있는 센서스나 행정 구역별로 집계되는 면 단위(area unit) 자료의 '공간 효과'를 규명할 수 있는 기반을 마련하기 위해 1) GIS환경 내에 면 단위 자료에 대한 통계적 공간분석 도구(Statistical Analysis of Area Data)를 설계·구현하고, 2) 이를 이용하여 지리적 자료의 '공간 효과'를 고려한 공간분석의 사례 연구를 수행한다.

2) 연구 방법 및 자료

GIS 내에 통계적 공간분석 도구를 설계·구현하기 위해 선행적으로 지리적 자료의 '공간 효과'를 고려할 수 있는 '공간통계 분석기법'에 관한 정리·분류작업과 선별된 통계기법들을 GIS환경에서 사용하기 위한 GIS와 통계 모듈간의 연계방안 연구를 수행 한 후, 이를 기반으로 '면 단위 자료 분석을 위한 통계적 공간분석 도구(SAAD)'를 구현한다.

또한, SAAD를 이용해 실제 대기 오염과 호흡기 질환 자료의 상관성을 살펴본다.

한편, 논문에 사용된 자료로는 1995년 서울시의 구별 천식환자 수와 센서스와 같은 면 단위 집계 자료와 당시 구(區) 단위 행정구역 경계도, 서울시의 20개소 대기오염 자동 측정망으로부터 1995년 1년 동안 집계된 오존(O3), 분진(TSP), 일산화탄소(CO), 이산화 황(SO2), 이산화질소(NO2) 등의 일별 평균치를 사용하였다.

통계적 분석 도구(SAAD) 개발은 Java 1.1.6을 이용하였으며, 대기 오염원과 같은 연속 프로세스를 따르는 자료 분석을 위해 ESRI ArcViewGIS version 3.0을 사용하였다.

1) 'GIS환경 내에서 활용 가능하면서도 적절한 통계적 공간분석' 기법과 이를 위한 '분석 도구'의 개발을 위해 다음과 같은 연구가 진행 중이다. (미국 : NCGIS, 영국 : ESRC의 RRL(Regional Research Laboratories), 유럽 : GISDATA)

2. SAAD의 설계 및 구현

SAAD를 설계하기 위해서는 우선적으로 약 세 가지 요인을 고려할 필요가 있다. 면 단위 자료 분석에 적절한 분석적 프로세싱과 GIS와 통계 모듈간의 진정한 동적 상호작용을 지원해 줄 수 있는 통합 환경, 그리고 범용 분석도구로서의 사용 환경 등이 이에 해당한다.

1) 기능적 요구 분석

① GIS에서 활용 가능한 공간 통계 분석

GIS의 분석 도구로서 공간 통계 모듈을 추가하기 위해서는 GIS와 연계됨으로서 보다 세부적이고 현실적인 가정을 충족시킬 수 있는 수학적 정교성과 타학문 분야로의 응용성을 가지고 있을 뿐만 아니라, 공간적 이질성(spatial heterogeneity)과 같은 지리적 자료의 '공간 효과'를 그래픽 화 할 수 있는 공간 통계 분석 기법들의 선별이 필요하다.

또한, 사회 과학분야에서 흔히 접할 수 있는 면 단위 자료에 적합한 분석 기법으로서 자료의 공간적 의존성과 이질성 등을 다양한 스케일에서 분석할 수 있는 기반을 제공해 줄 수 있는 분석 기법이여야 한다.

② 안정적이면서도 동적인 상호작용을 제공하는 통합환경

GIS환경에서 진정한 공간분석을 수행하기 위해서는 통계 모듈과 GIS의 안정적이면서도 동적인 상호작용 지원이 매우 중요하다. 극단적으로 말해, 성공적인 대화식 자료 분석은 GIS모듈과 통계 모듈간의 지리적 자료와 분석 결과의 효율적인 전송에 달려있다고 볼 수 있다(Bailey, 1995; Anselin, 1997).

실질적인 구현을 위해 현재까지 제시된 개념적 구조는 크게 느슨한 결합 방식(loose-coupling), 긴밀한 결합 방식(close-coupling), 총체적 결합 방식(fully integrated) 등으로 구분할 수 있다.

그러나 최근 가장 일반적으로 사용되고 있는 연계 방식인 두 모듈간의 '긴밀한 결합방식'이나 비교적 기술적으로 구현이 용이한 '느슨한 결합 방식' 등은 근본적인 설계방식과 개발 목적의 차이로 인해 진정한 동적 상호작용을 제공할 수 있는 분석환경 조성에 한계가 있다.

③ 범용의 분석 도구

지리적 전문가 뿐만 아니라 비전문가를 대상으로 하는 공간분석 도구인 SAAD는 그 기능적 구성 요소와 더불어 사용환경 또한 이질적인 환경의 불특정 다수의 사용자를 대상으로 설계되어야 한다. 이러한 맥락에서 지리적·GIS적 전문가 또는 비전문가에 관계없이 이용될 수 있는 ESDA 중심의 분석 기법 구현과 사용 환경의 선택은 SAAD 개발에 매우 중요한 고려 요인이 된다.

2) SAAD의 개념적 설계 및 구현

본 시스템 설계 및 구현에서 가장 중점적인 사항은 면 단위 자료에 적합한 통계적 분석 도구의 설계 및 구현으로서 면 단위로 집계된 자료의 공간적 패턴과 변이(variation)의 탐색과 같은 분포 메커니즘을 이해하기 위한 가장 기본적이면서도 핵심적인 탐색적 분석 기법에 해당하는 공간적 자기상관(spatial autocorrelation)이라는 공간 분석기법을 다양한 유형의 공간구조와 국지적 스케일에서 적용할 수 있는 분석도구를 마련하였다.

특히, Getis와 Ord가 제안한 G, G*와 Anselin의 Local Moran'I, Local Geary-type'c 등과 같은 LISA(Local Indicator of Spatial Association) 유형의 분석기법들은 단위지역의 규모가 매우 크거나

공간적 구조가 안정적이지 않은 지역에 공간적 자기상관(spatial autocorrelation)기법을 적용함으로써 발생하는 통계적 분석기법의 추론적 오류를 최소화하고, 국지적 수준에서의 공간적 상관 구조를 규명할 수 있다. 이와 같은 통계량들은 군집성 분석, 군집의 스케일, 주변 지역과의 공간적 상관관계('양의 상관성' 또는 '음의 상관성') 파악, '국지적 구조의 불안정성(local instability)' 규명 등과 더불어 '지도화'를 통한 공간 자료의 탐색적 분석이 가능하다.

또한 GIS와 통계 모듈간의 보다 안정적인 자료 교환을 위해 단일 자료모형으로 설계·구현하였으며, 통계 모듈과 일반 GIS환경을 위한 공용의 인터페이스를 제공하였다. 즉, 전반적인 시스템의 구성방식을 일종의 'fully integrated system'의 통합 방식으로 구현하고자 노력하였다.

이외에도 범용의 분석 도구로서의 활용성을 감안해 개발 방식으로서 플러그인이나 별도의 Helper 프로그램을 설치할 필요 없이 WWW 브라우저안의 임베딩을 통해 모든 지역의 클라이언트를 대상으로 실시간 전송이 가능하며, 기계 중립언어로 설계되어 특정 하드웨어(HW) 및 운영 체제(OS) 플랫폼에 관계없이 구동할 수 있고, 객체 지향언어로 구성되어 코드의 재 사용성·이식성 등이 뛰어난 자바 애플릿 형태를 채택하였으며, 사용자 수준과 관계없이 복잡한 통계적 공간분석 과정을 처리할 수 있도록 위젯(widget) 중심의 사용자 위주(user-friendly)의 인터페이스로 설계하였다.

4. SAAD를 이용한 사례 연구

사례연구는 대기 오염이 호흡기 질환에 미치는 영향을 평가하기 위한 기초 연구로서, 서울시의 25개 행정 구(區) 단위로 집계된 천식 환자와 대기오염의 분포 패턴을 탐색하고, 그 상관성을 추적하였다.

대기오염원의 분석은 ArcViewGIS 3.0을 이용했으며, Kriging이라는 보간 방법을 통해 대기오염원의 분포 패턴을 살펴보고, 이 과정을 통해 서울시 전역에 걸친 오염 표면을 추정할 수 있었다.

한편, 각 구 단위로 집계된 천식환자의 분포 패턴의 탐색은 기본적인 자료 탐색 기법(e.g> histogram, box plot ...)에서부터 정량적인 자기 상관 분석에 이르기까지 다양한 분석 절차를 통해 이루어 졌다. 특히, 천식 환자의 발병은 전체 인구수에 비해 상대적으로 드물게 발생하는 사건이기 때문에, 모집단의 분포적 특성을 고려한 상대적 위험도, 포아송 확률지도, 베이지언 추정도와 같은 위험도를 통해 천식 발생의 공간적 패턴을 살펴보았다.

이와 같은 확률론적 접근 이외에도 지도에 나타난 공간적 패턴을 일반화하고 정량화하기 위해서는 공간적 자기상관과 같은 분석이 필요하다. 공간적 자기상관은 공간 구조의 정의 유형과 가중치 함수에 의해 많은 영향을 받기 때문에 본 연구에서는 SAAD의 이웃 정의 기준인 경계선 공유 여부와 인접한 이웃의 차수, 면 객체 중심점간의 거리에 대해 각각의 공간 이웃 구조를 정의하였고, 공간적 자기상관(spatial autocorrelation) 분석과 LISA 기법을 적용하여 다양한 스케일의 공간적 상관성을 탐색하였다.

패턴 탐색 결과 몇몇 지역이 이례적인 군집 현상을 띄고 있었으며, 14세 이하의 천식 환자 발병에 대해 공간적 자기상관을 발견할 수 있었다.

이외에도 대기오염과 천식 발병간의 상관성 파악을 위해 회귀 모형을 사용하였다. 회귀 모형에는 두 변인간의 직접적인 상관성을 파악할 수 없기 때문에 소득이라는 일차 설명 변인을 추가하였다.

5. 결론

이상과 같은 연구 결과로 도출한 '면 단위 자료의 통계적 공간분석 도구(SAAD)'는 지리적 자료의 특성을 고려한 이론적인 공간통계 분석기법들을 GIS환경 내에 접목시킴으로써 보다 다양한 방

식의 공간 구조 분석이 가능함을 보여주는 시범적인 GIS의 공간분석 도구로서, 앞으로 공간적 회귀 모형과 같은 보다 정교한 공간 분석 이론들을 추가·보완함으로써 공간정보과학으로서 GIS의 입지를 더욱 강화시킬 수 있을 것이다. 뿐만 아니라, 차후에 데이터베이스와 공간 데이터베이스 관리 시스템과의 연결구조로 확장될 경우, 현 시스템의 가장 큰 문제점이라고 할 수 있는 공간 연산 처리 속도 등의 문제가 해결될 것으로 기대된다.

또한 SAAD 적용 사례 연구로 살펴본 대기오염과 호흡기질환의 상관성 연구는 보다 마이크로스케일에서 대기오염의 호흡기 질환에 관한 평가를 위한 기초자료로서 상당히 설득력 있는 근거가 될 수 있었다. 또한 이와 같은 접근 방식의 연장선상에서 보다 다양한 스케일의 자료 수집과 설득력 있는 설명변수의 유입, 시간차원 등을 고려한 시공간 예측 모형의 산출은 대기오염원의 변화에 따른 호흡기 질환 발생과 관련된 연구에 많은 도움이 될 것으로 예상된다.

참고문헌

- Anselin, L and Getis, A (1992) Spatial statistical analysis and geographical information systems. The Annals of Regional Science 26, 19-33.
- Bailey, T.C. and Gartell (1995) Interactive spatial data analysis Harlow, Longman Scientific and Technical.
- Fischer, M.M Scholten, J. and Unwin, (1996) Geographic information systems, spatial data analysis and spatial modelling: and introduction, Spatial Analytical Perspectives on GIS. (Eds) Fischer, M.M. Scholten. H.J and Unwin, Taylor & Francis.