

공간 상호작용 모델에 대한 공간단위 수정가능성 문제(MAUP)의 영향

김감영*

Effects of the Modifiable Areal Unit Problem (MAUP) on a Spatial Interaction Model

Kamyoung Kim*

요약 : 공간 상호작용의 복잡성, 공간적 재현과 모델링의 필요성에 의해서 공간 상호작용 데이터의 합역이 불가피하다. 이러한 상황에서 본 연구의 목적은 공간 상호작용 데이터를 스케일을 달리하여 합역하거나 혹은 동일 스케일에서 합역 방식을 달리하여 합역하였을 때, 공간 상호작용 모델의 결과가 어떻게 달라지는지 평가하는 것이다. 공간 상호작용 데이터의 합역은 공간단위 수정가능성의 문제(Modifiable Areal Unit Problem: MAUP)를 야기한다. 공간 상호작용 데이터의 합역을 위하여 무작위로 구역 시드를 선정한 후 인접한 공간단위를 할당하는 방법, 구역 시드와 공간단위 사이의 인구 가중 거리를 최소화하는 방법, 구역 내 상호작용 비율을 최대화하는 방법, 구역 내 상호작용 비율을 최소화하는 방법을 사용하였다. MAUP의 영향을 평가하기 위한 공간 상호작용 모델로 기원지-목적지 제약 포아송 회귀 모델을 이용하였다. 분석 결과는 모델 잔차의 공간적 특성뿐만 아니라 파라미터 추정값, 적합도 등이 MAUP의 영향을 받는다는 것을 보여주었다. 모델은 합역 방식 보다는 합역 수준에 더 민감하게 반응하였고, 모델에 대한 스케일 효과는 구획 방식에 따라 상이하게 나타났다.

주요어 : 포아송 회귀 모델, 합역 수준 및 방식, 공간단위 수정가능성 문제, 공간 상호작용 데이터

Abstract : Due to the complexity of spatial interaction and the necessity of spatial representation and modeling, aggregation of spatial interaction data is indispensable. Given this, the purpose of this paper is to evaluate the effects of modifiable areal unit problem (MAUP) on a spatial interaction model. Four aggregation schemes are utilized at eight different scales: 1) randomly select seeds of district and then allocate basic spatial units to them, 2) minimize the sum of population weighted distance within a district, 3) maximize the proportion of flow within a district, and 4) minimize the proportion of flow within a district. A simple Poisson regression model with origin and destination constraints is utilized. Analysis results demonstrate that spatial characteristics of residuals, parameter values, and goodness-of-fit of the model were influenced by aggregation scale and schemes. Overall, the model responded more sensitively to aggregation scale than aggregation schemes and the scale effect on the model was varied according to aggregation schemes.

Key Words : Poisson regression model, aggregation scale and schemes, modifiable areal unit problem(MAUP), spatial interaction data

이 논문은 2009년도 정부재원(교육과학기술부 인문사회연구구역량강화사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2009-332-B00616)

* 경북대학교 사범대학 지리교육과 조교수(Assistant Professor, Department of Geography Education, Kyungpook National University), kamyoungkim@knu.ac.kr

1. 서론

인간, 자금, 물류, 정보 등 다양한 유형의 지역 간 흐름, 즉 공간 상호작용은 지역의 사회·경제적 변화의 동인으로, 다양한 연구 분야 및 의사결정, 예를 들어 지리학, 인구학, 도시계획, 교통, 경제학, 재난관리, 전염병학 등에서 중요한 요소이다(Guo, 2009). 왜냐하면 공간 상호작용은 어떻게 공간 과정이 작동하고 공간 환경에서 사람들이 어떻게 행동하는지를 보여주기 때문이다.

지역 변화 및 지역 간의 관계를 이해하는데 있어 공간 상호작용에 대한 이해의 중요성이 증대됨에도 불구하고 이러한 현상을 탐색, 분석, 시각화하는 것은 여전히 어려움으로 남아 있다. 다양한 분야에서 흐름 관련 데이터가 폭발적으로 늘어남에도 불구하고 이에 대한 탐색, 분석, 시각화가 부분적으로 이루어지고 있는 이유는 다음과 같다(Alvanides *et al.*, 2000). 첫째, 공간 상호작용 데이터를 손쉽게 다룰 수 있는 이용 가능한 틀(특히 지리정보시스템 환경에서)이 부족하다. 둘째, 공간 상호작용 데이터의 저장에 대한 요구와 처리 비용이 일반 공간 데이터보다 크다. 이는 공간 상호작용 데이터의 본질적인 특성에 기인하는데 공간단위가 n 개이고, 속성이 k 개인 경우 상호작용 데이터의 차원은 $n^2 \times k$ 이다. 또한 거의 모든 흐름은 방향성을 갖고 있고 공간적으로 집계될 때 비대칭적이다. 따라서 분석과 시각화에서 상호작용 행렬의 대각선 상부와 하부 모두를 고려해야 한다. 결과적으로 상호작용 데이터를 분석하고 모델링하는데 더 많은 시간과 노력이 요구된다. 셋째, 지역 간 상호작용은 사회의 다양한 측면에 의해서 유발되기 때문에 공간 상호작용 데이터의 본질을 이해하기 위하여 사회적, 정치적, 경제적 과정에 대한 이해를 필요로 하고, 이는 데이터의 분석을 더욱 복잡하게 만든다. 넷째, 공간 상호작용 데이터는 사용된 공간단위에 의존한다. 즉, 흐름을 포착하는 공간단위의 설정이 어떠냐에 따라 공간단위 사이 흐름의 양이 달라지고, 이를 바탕으로 수행한 분석 및 시각화의 결과가 달라진다. Barras *et al.*(1971)은 공간 상호작용의 공간단위 의존성을 다음과 같이 표현한다. “지역은 흐

름의 탐지기이다. 지역의 크기가 감소하면, 상호작용의 강도는 증가하는데, 여기에서 상호작용의 강도는 총 흐름 중 지역 경계를 가로지르는 흐름의 비율로 정의된다(p.140).” 즉 지역 설정 방식에 따라서 내부 흐름의 비율이 증가할 수도 감소할 수도 있다. 마지막으로 시간에 따른 흐름의 역동성을 다룰 수 있는 모델이 부족하다.

공간상에서 발생하는 인문과 관련된 상호작용 현상은 이산적인 이벤트이다. 이러한 이산적 이벤트를 재현하기 위하여 널리 이용된 방법은 지역을 중심으로 이벤트를 합역(aggregate)하는 것이다. 여기에서 지역은 기저의 이산적 흐름의 공간적 재현을 가능하게 하는 틀이 된다. 어떤 지역 체계는 기저의 흐름을 더 잘 반영하는 반면 어떤 체계는 그렇지 못하다. 즉 흐름의 재현 값은 지역 구분과 이에 따른 합역에 의존한다. 다시 말해서 지역 체계 없이 흐름 데이터는 없다. 지역 체계가 없는 상황에서 모든 흐름은 단일 공간단위 내부의 흐름이 된다. 이러한 공간 상호작용 데이터의 특성과 집계된 편익, 데이터의 기밀성 유지를 위하여 일반적으로 공표되는 공간 상호작용 데이터는 공간적으로 합역된 형태이다. 중력 모델을 발전시켜 만들어진 공간 상호작용 모델은 집합적 모델로 이렇게 합역된 데이터를 다룬다. 따라서 재현과 모델링을 위하여 공간 상호작용 현상의 합역은 불가피하다. 물론 공간 상호작용 연구에서 개별 흐름 행동을 모델링하는 개별 수준의 모델들, 예를 들어 효용이론(utility theory), 엔트로피 극대화 방법 등이 제시되어 있기도 하다(Webber, 1980).

이러한 상황에서 공간 상호작용 데이터의 합역을 통한 구역 설계 문제는 다음과 같은 네 가지 목적을 가질 수 있다(Alvanides *et al.*, 2000). 첫째, 설계된 구역은 데이터 축소 및 기술을 위한 도구이다. 흐름의 양이 증가하면서 개별 흐름을 시각화하거나 분석하는 것이 어렵기 때문에 이산적 이벤트 혹은 상호작용 데이터의 양을 관리하고 기술하는 것을 용이하게 하기 위하여 이를 축소할 필요가 있다. 둘째, 왜곡과 편기를 최소화한 상태에서 구역 체계를 통하여 이산적 흐름 현상을 공간적으로 재현하기 위함이다. 구역을 설정하는 것은 결국 지역 사이의 상호작용을 포착하여 재현하기 위한

방법이다. 셋째, 구획 체계는 공간 상호작용의 분석 및 시각화를 위한 도구이다. 구역 체계가 형성되어야 집합적 모델인 공간 상호작용 모델을 적용하여 데이터를 분석하는 것이 가능하다. 또한 복잡한 흐름을 이해할 수 있는 형태의 지도로 시각화하기 위하여 반드시 공간단위의 구획화 및 이를 통한 흐름의 합역이 요구된다. 마지막으로 구획화는 모델링 문제를 용이하게 하고 모델 수행 능력을 강화시킬 수 있다. 일반적으로 공간단위 수가 많아질수록 공간 상호작용 모델을 수행하는 것이 어렵고, 모델의 수행 능력 또한 떨어진다. 또한 동일한 수의 구역을 갖더라도 어떻게 구역을 설계하느냐에 따라 모델의 수행 능력이 달라진다. 요약하면 공간 상호작용 데이터의 기술, 탐색, 분석, 시각화, 모델링을 위하여 체계적인 합역 방식이 요구된다.

공간 상호작용 데이터의 합역 문제는 '공간단위 수정가능성 문제(Modifiable Areal Unit Problem: MAUP)' (Openshaw and Taylor, 1981)와 연결된다. MAUP는 연구의 공간단위 선택은 작위적일 수 있으며, 그 연구의 결과는 선택된 공간단위에 따라 달라질 수 있음을 의미한다. 구체적으로 공간 분석 및 모델링의 결과는 데이터의 합역 수준(스케일 효과)과 합역 방식(구획 효과)에 따라 달라질 수 있다. Lee(1999)는 MAUP의 효과를 몇 가지로 일반화하여 정리하였다. 첫째, 분산이 큰 변수가 합역의 영향을 더 크게 받는다. 둘째, 합역 수준이 높아질수록 변수 간의 상관관계는 커진다. 셋째, 합역 수준이 높아질수록 회귀분석에서 결정계수 값은 커진다. 넷째, 스케일이나 구획방식에 따라 회귀계수는 상당한 편차를 보인다. 다섯째, 합역 수준이 높아질수록 공간적 자기상관의 값은 작아진다. 그러나 이러한 일반화는 주로 면(面, area) 단위 공간 데이터 합역에 초점을 둔 연구로부터 도출된 것으로, 지역 사이의 관련성을 재현하는 2차원적인 공간 상호작용 데이터의 합역이 재현과 모델링에 어떠한 영향을 미치는지에 대한 연구는 상대적으로 덜 되어 있다. 공간 상호작용의 복잡성, 공간적 재현 및 모델링의 필요성에 의해서 공간 상호작용 데이터를 합역하는 것이 필수불가결한 상황에서, 본 연구의 목적은 인구가동 데이터와 같은 공간 상호작용 데이터의 합역 수준을 달리하거나 동일 스케일에서 합역 방식을 달리 하였을

때 공간 상호작용 모델의 결과가 어떻게 달라지는지를 분석하는 것이다. 구체적으로 MAUP의 영향을 구획 체계의 밀집도, 모델 잔차의 공간적 특성, 파라미터 추정값, 적합도 등의 통계량을 이용하여 평가한다. 이를 통하여 기능지역 구분뿐만 아니라 모델링에 적합한 합역 방식이 무엇인지, 그리고 MAUP의 영향을 덜 받는 합역 방식이 무엇인지 살펴본다.

논문의 구성은 다음과 같다. 2장에서는 공간 상호작용 데이터의 합역 방법과 효과에 대한 기존 논의를 살펴본다. 3장에서는 공간 상호작용 데이터의 합역 개념과 합역에 필요한 목적함수를 살펴보고 목적함수별 합역 절차를 기술한다. 4장에서는 합역의 영향을 평가하기 위한 공간 상호작용 모델로 포아송 회귀 모델을 정의한다. 마지막으로 서울시 동간 인구이동을 사례로 스케일과 목적함수에 따라 합역하여 분석을 수행한 후 결과를 제시한다.

2. 문헌연구

공간 상호작용 데이터의 재현을 위하여 그리고 모델링을 위하여 오래 전부터 많은 연구들이 흐름 데이터의 합역 방법과 그 효과에 대하여 논의하였다. 합역의 궁극적 목적에 따라 기존 접근을 크게 데이터 중심 접근과 모델 중심 접근으로 구분할 수 있다.

먼저, 데이터 중심 접근은 공간 상호작용 데이터의 제반 특성을 고려하여 합역을 수행하는 것으로, 데이터 자체를 기반으로 노동시장, 집수구역, 상권, 판매권, 통근권 등의 기능지역 구분(functional regionalization)을 시도한 연구들이 이 범주에 속한다. 이러한 접근은 기저의 흐름 현상을 잘 반영하도록 구역 사이의 경계를 설정하고 설정된 경계를 바탕으로 구역의 특성을 파악하는데 초점을 둔다. Masser and Brown(1975)은 공간 합역 과정의 기저를 이루는 목적함수는 설계의 특성에 따라 다양할 수 있다고 전제한다. 하나는 합역된 기본 공간단위 사이에서 발생한 총 상호작용의 비율을 극대화하여 전체적으로 구역 간 이

동 비율을 최소화하는 Intramax이다. Intramax는 상호작용 행렬을 분할표(contingency table) 형태로 간주하고 주변합(marginal total)과 연관된 관측값과 기대값 사이의 차이를 이용하여 목적함수를 평가한다. 이들은 합역 방법으로 계층적 합역 절차를 제시하였다. 이 절차의 단계마다 관측값이 기대값보다 가장 큰, 즉 가장 높은 상호작용을 보이는 두 단위 지역이 결합된다. 다른 하나는 구역 간 흐름의 비율을 극대화하는, 그래서 구역 내 흐름을 최소화하는 Intramin이다. Intramin은 상호작용이 가장 작은 두 공간단위를 확인하여 결합한다. 이러한 계층적 합역 절차는 그리디(greedy) 알고리즘으로, 한번 결합된 두 공간단위는 이후 과정에서 분리되지 않기 때문에 국지적 최적해(local optimum)가 도출될 수 있는 제한점을 내포하고 있다.

Hirst(1977)는 Masser and Brown(1975)이 제시한 Intramax 목적함수는 결합 과정에서 상호작용 행렬의 행과 열의 합의 크기 변이를 고려하지 않았다고 지적하면서, 잔차에서 행과 열 합에서의 변이를 고려한 새로운 목적함수 계산식을 제시하였다. 한편 Masser and Scheurwater(1980)와 Slater(1981)는 기존에 제시된 여러 합역 절차를 비교 평가하였다. Masser and Scheurwater(1980)는 기존에 제시된 기능 거리(functional distance) 접근(Brown and Holmes, 1971) 및 절차 기반의 반복 비율 적합 방법(iterative proportional fitting)(Slater, 1976)의 문제점을 지적하면서 기능지역 구분 방안으로 Intramax가 적절함을 증명하였다. 한편 Slater(1981)는 계층적 합역 과정에 결합의 한계를 설정할 경우 좀 더 향상된 결과가 도출된다는 것을 보여주었다. Alvanides *et al.*(2000)은 이러한 계층적 합역 절차가 국지적 최적 해에 빠질 수 있음을 지적하면서 그 대안으로 자동구획절차(automated zoning procedure)를 이용한 합역 방법을 제시하였다. 이들은 표준화되지 않은 합역 전의 상호작용 행렬에서 구역 내 상호작용의 비율(전체 흐름에 대한 구역 내 흐름의 비)을 극대화하는 새로운 목적함수를 제시하고 Masser and Brown(1975)의 Intramax와 비교하였다. 그 결과 기능지역 구분의 중요한 요소라 할 수 있는 자족성(self-containment) 측면에서 전통적인 그

리고 지나치게 복잡한 Intramax보다 이들이 제시한 목적함수가 더 좋은 결과를 도출하였다.

모델 중심 접근은 모델의 수행 능력 관점에서 공간 상호작용 데이터의 합역을 시도하고 평가한다. Openshaw(1977b)는 합역 과정에서 명시적으로 중력 모델을 위하여 설계된 목적함수를 이용하였다. 즉 합역된 데이터에 대한 공간 상호작용 모델의 적합도(goodness-of-fit)가 최적화되도록 구획 체계를 다시 설계하였다. 이를 통하여 고정 파라미터(fixed parameter)를 갖는 공간 상호작용 모델에 가장 적합하도록 공간 상호작용 데이터를 재구획화하는 것이 가능하다는 것을 보여주었다. 또한 공간 상호작용 모델의 성능과 파라미터 값의 구획 체계의 선택에 따라 달라짐을 보여줌으로써 공간 상호작용 모델이 MAUP의 영향을 받는다는 것을 보여주었다. Batty and Sikdar(1982a; 1982b; 1982c; 1982d)는 공간 상호작용 모델들은 공간 상호작용 현상과 독립적으로 존재하는 것이 아니라는 가정 하에서 합역에 따라 모델의 파라미터가 얼마만큼 달라지는지를 평가하였다. 이를 통하여 모델 구조의 선택은 직면한 문제의 원인이 될 수 있음을 밝혔다. 이는 합역뿐만 아니라 모델 구조 선택이 모델 파라미터 값의 변화에 영향을 줄 수 있음을 의미한다. Putman and Chung(1989)은 무작위, 기본 공간단위 수, 면적, 인구, 저소득 가구 수 등 공간 체계 정의와 관련된 기준을 바탕으로 다섯 가지 상이한 합역 방법을 제시하였고, 이에 따라 다변량 다중 파라미터 상호작용 모델의 결정계수(R^2)와 최상-최하 우도비(likelihood ratio)의 비율을 평가하였다. 그 결과 체계적인 합역 방법이 적합도와 파라미터 신뢰성 관점에서 더 좋은 결과를 제시하였다. 이는 공간 체계 합역에 대한 상이한 알고리즘은 모델의 적합도에 대한 상이한 분포를 산출할 수 있기 때문에 체계적인 합역 절차가 필요함을 의미한다.

Amrhein and Flowerdew(1992) 또한 세 가지 상이한 합역 알고리즘을 이용하여 포아송 회귀 모델에 대한 영향을 평가하였지만 이전 연구들과는 다른 결론에 도달하였다. Openshaw(1977b), Batty and Sikdar(1982a; 1982b; 1982c; 1982d), Putman and Chung(1989) 등은 MAUP의 스케일 효과와 구획 효과 모두

공간 상호작용 모델의 파라미터 추정값과 모델의 적합도에 주요한 영향을 준다는 결론에 도달하였지만, Amrhein and Flowerdew(1992)의 경우 구획 효과에 대한 증거를 찾을 수 없다고 결론을 내렸다. 이들은 합역의 특성 및 정도를 평가하기 위하여 어떠한 통계량을 사용하는지가 중요하다는 점을 지적하였다. 이전 연구에서 주로 사용된 통계량은 모델의 적합도와 파라미터 값이다. 다른 연구와 달리 이 연구는 세 가지 상이한 구획화 방법을 무작위로 반복 수행하여 측정된 결과 값을 평균하여 비교하였다. 해상도가 가장 높은 공간단위에서 얻은 결정계수 값과 합역 모델의 결정계수 값의 차이를 비교한 결과 스케일에 따른 차이는 그렇게 크지 않았으며, 동일 스케일에서 모델 적합도의 차이 또한 크지 않았다.

요약하면, 데이터 중심 접근은 흐름 데이터의 특성을 반영하도록 합역을 시도하였고, 모델 중심 접근은 사용된 모델의 성능을 극대화하거나 성능을 평가하기 위하여 합역을 시도하였다. 전자는 합역 자체가 목적이고 후자는 모델링이 목적이다. 이러한 이유로 전자는 공간 상호작용 데이터의 합역 결과를 형상과 같은 구역의 지리적 특성을 이용하여 평가하였고, 후자는 합역 과정에서 공간 상호작용 현상 자체를 크게 고려하지 않은 상태에서 모델의 결과에 초점을 맞추었다. 본 연구에서는 데이터 중심 접근과 모델 중심 접근을 결합하는 종합적 접근을 시도한다. 즉 데이터 관점에서 합역을 위한 목적함수를 정의하고, 합역된 데이터에 공간 상호작용 모델을 적용하여 그 영향을 살펴본다. 구획의 밀집도, 모델 잔차의 공간적 특성, 파라미터 및 적합도 등의 측면에서 MAUP의 영향을 평가한다.

3. 공간 상호작용 데이터의 합역

1) 공간 상호작용 데이터의 합역

인구이동과 같은 지역 간 상호작용은 이산적인 공간 현상임에도 불구하고 흐름을 재현하고 모델링하기 위하여 집합적으로 평가될 필요가 있다. 집합적으로 재

현된 이러한 공간 상호작용은 지도화 기법을 이용하여 시각화되고 공간 상호작용 모델을 이용하여 분석된다. 그러나 이러한 시각화 및 분석은 지역 간의 관계를 다룬다는 공간 상호작용의 본질적 특성 때문에 기본 공간단위가 많아질수록 어렵게 된다. 이러한 이유로 정보의 손실을 최소화하면서 공간 상호작용 데이터를 그룹화 하는 합역 및 군집 절차가 요구된다(Masser and Brown, 1978). 공간 상호작용 데이터의 합역은 n 개의 공간단위로 구성된 $n \times n$ 공간 상호작용 행렬을 m 개의 공간단위로 구성된 $m \times m$ 행렬로 축소하는 과정이다($n > m$). $n \times n$ 행렬을 $m \times m$ 행렬로 축소하는 것은 개념적으로는 간단해 보이지만 실재는 그렇지 않다. 왜냐하면 공간 상호작용 행렬은 지역 사이에 발생한 공간 현상을 추상화한 것이기 때문에 행렬의 축소 과정에서 이러한 현상이 발생한 공간적 상황(공간단위들 사이의 관계)을 고려하지 않고 행렬 정보만으로 축소하는 것은 불가능하기 때문이다. 즉 인접성과 같은 공간단위 사이의 공간 관계를 참조해야 하며, 축소를 위한 일정한 목적이 있어야 한다. 상호작용 행렬을 축소하기 위해서는 먼저 행렬이 표현하고 있는 기본 공간단위 사이의 흐름 정보뿐만 아니라 기본 공간단위 사이의 인접성 정보가 필요하다. 합역은 이러한 정보를 바탕으로 일정한 목적(예를 들어 구역 내 이동을 극대화)이 최적화되도록 인접한 기본 공간단위를 결합한 후 합역된 구역 사이의 흐름을 행렬 형태로 재집계하는 과정이다.

2) 공간 상호작용 데이터 합역을 위한 목적함수

공간 상호작용 데이터의 합역에서 중요한 것은 목적함수를 정의하는 것이다. 지금까지 다양한 목적함수가 제시되었고, 목적함수의 선택은 구획화의 목표에 의존하게 된다. 기능지역 구분을 위한 데이터 중심 접근의 경우 구역 내 상호작용을 극대화하는 목적함수를 선택하고, 모델 중심 접근에서는 구역 간 상호작용을 극대화하는 형태로 목적함수를 정의하거나 모델의 성능을 극대화하는 목적함수와 선택한다. 본 연구에서는 Masser and Brown(1975)과 Alvanides *et al.*(2000)의 논의를 바탕으로 다음과 같은 세 개의 목적함수를 정

의하여 공간 상호작용 데이터의 합역에 이용하였다:

$$\text{Dis-Min: } \sum_m \sum_{i \in N_m} p_i d_{im} \quad (1)$$

$$\text{Intra-Min: } \frac{\sum_m \sum_{i \in N_m} \sum_{j \in N_m} f_{ij}}{S} \quad (2)$$

$$\text{Intra-Max: } \frac{\sum_m \sum_{i \in N_m} \sum_{j \in N_m} f_{ij}}{S} \quad (3)$$

여기에서 m 은 구역 시드(seed) 혹은 구역 인덱스이고, p_i 는 기본 공간단위 i 의 인구, d_{im} 은 기본 공간단위 i 와 구역 시드 m 사이의 거리(본 연구에서는 기본 공간단위의 중심점과 구역 시드의 중심점 사이의 직선거리), N_m 는 구역 m 에 할당된 기본 공간단위의 집합($N_m \subseteq N$), f_{ij} 는 기본 공간단위 i 와 j 사이의 흐름의 양, S 는 합역 전 상호작용 행렬에서 전체 흐름의 양을 나타낸다. 식 (1)은 구역 시드와 기본 공간단위 사이의 인구 가중 거리의 합을 최소화한다. 식 (1)을 Dis-Min으로 명명한다. 식 (2)는 구역 내 상호작용의 비율을 최소화하는 목적함수로, 개별 구역에 할당된 기본 공간단위 사이에 발생한 상호작용의 합을 전체 상호작용으로 나누어 계산한다. 이 목적함수를 Intra-Min으로 명명한다. 식 (3)은 Alvanides *et al.*(2000)이 제시한 목적함수와 동일한 것으로 구역에 할당된 기본 공간단위 사이에 발생한 상호작용의 비율을 극대화, 즉 구역 내 상호작용의 비율을 극대화한다. 이 목적함수를 Intra-Max¹⁾로 명명한다. Intra-Max는 Masser and Brown(1975)이 제시한 것과 달리 표준화된 값이 아닌 상호작용의 절대량을 이용하여 계산된다.

3) 공간 상호작용 데이터의 합역 방식

공간 상호작용 데이터의 합역 방식은 정의된 목적함수의 형태에 따라 달라진다. Dis-Min은 선정된 구역 시드와 기본 공간단위 사이의 거리 관계를 평가하여 구역 시드에 배정하기 때문에 입지-할당(location-allocation) 문제에 해당하고, Intra-Max와 Intra-Min은 구역 시드 지정 없이 기본 공간단위를 분할하기 때문에 할당 문제에 해당한다. 이러한 차이에 의해서 합역 알고리즘이 달라진다. 먼저 Dis-Min 목적함수는 p -

median 문제의 목적함수와 동일한 형태이기 때문에, 이 문제에 대한 휴리스틱 기법의 하나인 Teitz and Bart(1968)의 교환 알고리즘(Interchange algorithm)을 이용하여 합역을 수행하였다. 이 알고리즘은 먼저 n 개의 기본 공간단위 중에서 임의로 m 개를 구역 시드로 선택한 후 나머지 기본 공간단위를 인구 가중 거리가 최소가 되도록 할당한다. 다음 선택된 구역 시드(m 개)와 선택되지 않은 기본 공간단위($n-m$ 개)를 바꿀 경우 가장 크게 목적함수를 향상시키는 것을 교환하고 이러한 과정을 몇 차례 반복하여 해를 찾는다(Kim, 2004).

Intra-Max와 Inter-Min 목적함수를 이용하여 공간 상호작용 데이터를 합역하기 위하여 Openshaw (1977a)가 제시한 자동구획절차를 이용하였다. 기존 연구에서 공간 상호작용 데이터를 합역하기 위하여 널리 사용되었던 방법은 Masser and Brown(1975)이 제시한 계층적 군집 절차이다. 그러나 이 방법은 그리디 알고리즘이기 때문에 도출된 해가 전역적 해가 아닐 수 있다. 왜냐하면 합역 과정에서 한번 결합된 공간단위는 다시 분리되지 않기 때문이다. 이러한 문제점을 극복하고 보다 향상된 구획 체계를 도출하기 위하여 몇몇 연구에서 자동구획절차를 이용하여 공간 상호작용 데이터를 합역하는 시도가 이루어졌다(Alvanides *et al.*, 2000; Kim *et al.*, 2009; Kim *et al.*, 2010). 공간 상호작용 데이터 합역을 위한 자동구획절차 알고리즘은 다음과 같다. 먼저, n 개의 기본 공간단위를 m 개 구역에 무작위로 할당하여 m 개 구역으로 구성된 구획 체계를 생성하고, 이 구획 체계에 대하여 Intra-Max 혹은 Intra-Min 목적함수 값을 평가한다. 다음 m 개의 구역 중 임의로 하나를 선택한 후 다른 구역과 경계를 공유하는 기본 공간단위 목록을 생성한 후 임의로 하나의 기본 공간단위를 선택한다. 연속성(contiguity) 제약조건 하에서 선택된 기본 공간단위를 연결한 다른 구역에 할당할 경우 목적함수가 향상되면 할당을 수용한다. 다음 선택된 구역의 경계를 구성하는 단위 지역 목록을 갱신하고 목록의 모든 기본 공간단위를 대상으로 이전 과정을 반복한다. 선택된 구역에 대한 평가가 완료되면 다른 구역을 선택하여 이전 과정을 목적함수의 향상이 없을 때까지 반복한다. 이상에서 논의한 세 개

의 목적함수를 이용하는 합역 방법 외에 무작위로 m 개의 구역 시드를 선택한 후 나머지 기본 공간단위를 최근접 구역 시드에 할당하는 무작위 방법을 이용한다 (Random). 따라서 본 연구에서는 Dis-Min, Intra-Max, Intra-Min, Random 등 네 개의 서로 다른 합역 방식을 적용하여 공간 상호작용 데이터를 합역한다.

4. 공간 상호작용 모델

공간 상호작용 데이터의 합역이 공간 상호작용 모델

에 미치는 영향을 평가하기 위하여 포아송 회귀 모델 (Poisson regression model)을 사용한다. 잔차가 동일 정규분포를 따른다고 가정할 수 없고, 사례로 사용되는 인구이동은 이산적인 현상이기 때문에 이산적 분포를 가정하는 것이 바람직하다(Flowerdew and Aitkin, 1982). 따라서 공간 상호작용을 위한 모델로서 포아송 회귀 모델이 적합하다. 공간 상호작용을 위한 모델로서 포아송 분포를 사용하기 위한 핵심 가정은 목적지 i 로부터 기원지 j 까지의 개별 흐름은 서로 독립적이고, 지역 쌍 사이의 상호작용은 다른 지역 쌍 사이의 상호작용과 독립적이라는 것이다(Tiefelsdorf, 2003). 포아송 회귀 분석을 이용하여 미지의 파라미터를 추정하기



Figure 1. Examples of districts according to aggregation scale and schemes.
합역 수준과 방식에 따른 구획체계의 예

위한 로그-선형 식은 다음과 같다:

$$\ln(\mu_{ij}) = \lambda + \lambda_o \ln(p_i) + \lambda_D \ln(p_j) + \beta \ln(d_{ij}) \quad (4)$$

이 식에서 μ_{ij} 는 기원지 i 와 목적지 j 사이에서 추정된 흐름의 양을, p_i 와 p_j 는 각각 기원지 i 와 목적지 j 의 인구를, d_{ij} 는 기원지 i 와 목적지 j 사이의 거리를 나타낸다. λ , λ_o , λ_D , β 는 추정될 파라미터이다.

이러한 기본 포아송 회귀 모델은 사전 정보의 사용 유무에 따라서 기원지, 목적지 혹은 기원지와 목적지 모두에 제약을 부여하는 형태로 확장될 수 있다 (Fotheringham and O'Kelly, 1989). 식 (4)는 기원지로부터의 전출량과 목적지로의 전입량이 모두 알려지지 않은 비제약 모델로, 사전 정보 없이 기원지와 목적지 사이의 상호작용량을 추정한다. 기원지 제약은 기원지로부터의 전출량이 알려진 경우 기원지 i 로부터 나간 전출량에 대한 추정치와 이 관측치가 동일하도록 하는 것이고, 목적지 제약은 목적지로의 전입량이 알려진 경우로 목적지로 들어온 전입량에 대한 추정치와 이 관측값이 동일하도록 하는 것이며, 기원지-목적지 제약은 기원지로부터의 전출량과 목적지로의 전입량이 모두 알려진 경우로 위 두 조건을 동시에 만족시키도록 하는 것이다(Tiefelsdorf and Boots, 1995). 본 연

구에서 분석하는 인구이동 자료의 경우 기원지와 목적지에 대한 전출 및 전입량 정보가 있기 때문에 기원지-목적지 제약 모델을 사용한다. 이 모델에 대한 포아송 회귀 식은 다음과 같다:

$$\ln(\mu_{ij}) = \lambda + \lambda_i^o + \lambda_j^D + \beta \ln(d_{ij}) \quad (5)$$

이 식에서 λ_i^o 는 기원지 제약을 만족시키도록 하는 기원지 특수 요인(origin specific factor)이고, λ_j^D 는 목적지 제약을 만족시키도록 하는 목적지 특수 요인(destination specific factor)이다.

5. 분석 결과

공간 상호작용 모델에 대한 MAUP의 영향을 평가하기 위하여 2008년 서울시 동간 인구이동 데이터를 이용하였다(통계청, 2008, 인구이동통계). 이 데이터는 438개 동 사이의 인구이동을 표현하고 있다. 이 인구이동 데이터를 $m=200, 175, 150, 125, 100, 75, 50, 25$ 등 여덟 개 스케일에서 네 가지 합역 방식인 Random, Dis-Min, Intra-Max, Intra-Min을 이용하여 합역하였

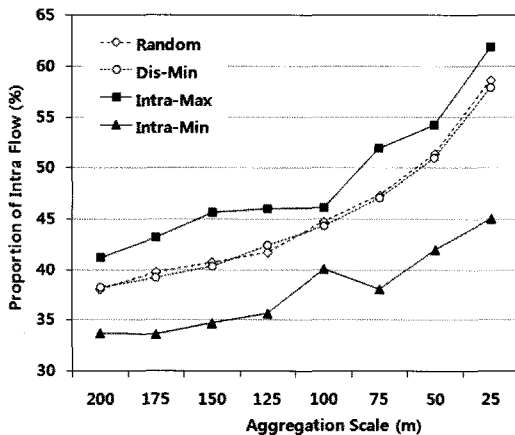


Figure 2. Change of proportion of intra flow according to aggregation scale and schemes.
합역 수준과 방식에 따른 구역 내 이동 비율 변화

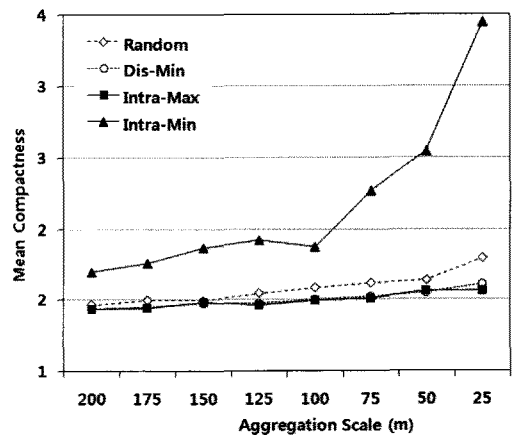


Figure 3. Change of compactness according to aggregation scale and schemes.
합역 수준과 방식에 따른 밀집도 변화

다. Dis-Min을 위한 교환 알고리즘, Intra-Max와 Intra-Min을 위한 자동구획절차 등 제반 분석 과정은 ESRI ArcGIS(버전 9.3)의 VBA(Visual Basic for Applications)를 이용하여 구현되었다. Figure 1은 합역 수준과 방식에 따라 달리 산출된 구획 체계의 예를 보여주고 있다. 기원지-목적지 제약 포아송 회귀 모델을 실행하기 위하여 SPSS(버전 18.0)의 GENLOG를 이용하였다. GENLOG는 주요 구성 요소로서 범주형 변수를 갖는 모델에 대한 적합도, 가설 검정, 파라미터 추정을 위한 일반적 절차이다. GENLOG는 파라미터 추정을 위하여 최대우도 추정 절차를 활용한다. 최대우도 추정 방법은 분포 가정 하에서 우도함수의 최대화를 통하여 모수를 추정하는 방법으로, GENLOG는 Newton-Raphson 방법에 토대한 반복재가중최소제곱(iteratively reweighted least squares)을 이용하여 최대우도 추정치를 찾는다. GENLOG는 포아송과 다항 분포를 다룰 수 있으며, 관측값과 기대값의 차이인 잔차를 단순, 표준화 및 조정된 형태로 제시할 수 있다.

포아송 회귀 모델의 분석 결과를 살펴보기에 앞서 합역 스케일과 방식을 달리하여 도출된 구역 체계 사이의 상호작용 형태와 구역의 형상적 특성을 먼저 살펴보자. Figure 2와 3은 각각 합역 수준과 방식의 변화에 따른 구역 내 인구이동 비율과 밀집도²⁾의 변화를 보여주고 있다. 구역 내 인구이동 비율의 경우, 구역 내 인구이동 비율을 극대화한 Intra-max가 가장 높은 수치를 보인 반면, 이를 최소화한 Intra-Min이 가장 낮은 수치를 보였다. Random과 Dis-Min은 이 두 극단의 중간에 놓이면서 값에서 큰 차이를 보이지 않았다. 한편 합역 방식과 상관없이 합역 수준이 높아지면서 구역 내 이동 비율은 증가하는 경향을 보였다. 이는 Ravenstein(1885)의 이동 법칙(law of migration) 중 하나인 단거리 이동 특성에 기인한다. 합역 수준이 높아지면서 상대적으로 상호작용량이 많은 연결한 공간단위들이 결합되면서 구역 내 이동 비율이 높아진 것이다.

Figure 3의 밀집도 추정치의 경우 1에 가까우면 가까울수록 밀집도가 높고, 값이 커지면 커질수록 밀집도가 낮아진다. Intra-Min은 다른 합역 방식에 비하여 큰 값을 보여, 밀집도가 가장 낮음을 알 수 있다. 또한

합역 수준이 높아지면서 Intra-Min에 의한 구획체계의 밀집도는 급격하게 약화됨으로써, 스케일 변화에 가장 민감하게 반응함을 알 수 있다. Intra-Min의 경우, 구역 내 이동을 최소화하는 과정에서, 연결하지만 상대적으로 상호작용이 적고 중심점 사이의 거리가 떨어져 있는 공간단위들을 결합하면서 길쭉한 모양을 갖는 구역이 형성되었고 이로 인하여 밀집도가 떨어졌다(Figure 1). 이와는 달리 Random, Dis-Min, Intra-Max는 상대적으로 근접한 공간단위들을 결합하면서 밀집도가 높게 나타났고 합역 수준이 높아지더라도 그 변화는 크지 않았다. 전체적으로 Intra-Max가 밀집도 측면에서 가장 좋은 결과를 제시하였다. Figure 2와 3을 통하여 Intra-Max가 구획된 구역의 형상적 측면에서 MAUP의 영향을 가장 적게 받으며 기능지역 구분 측면에서 다른 방식보다 더 적합하다는 것을 알 수 있다.

합역 스케일 및 방식에 따른 과대 및 과소 추정된 잔차의 공간적 특성을 파악하기 위하여 평균 거리³⁾를 산출하였다. Figure 4에서 보는 것처럼, 합역 수준이 높아지면서, 즉 구획할 구역의 수가 감소하면서 과대 추정된 잔차와 과소 추정된 잔차의 평균 거리는 상이한 경향을 보인다. 과대 추정된 잔차의 평균 거리는 합역 수준이 높아지면서 증가하는 반면, 과소 추정된 잔차의 평균 거리는 감소하는 경향을 보인다. 또한 동일 스케일 내에서 과대 추정된 잔차의 평균 거리는 과소 추정된 잔차의 평균 거리보다 크게 나타나 포아송 회귀 모델이 먼 구역 사이의 이동을 과대 추정하였고, 근거리 구역 사이의 이동을 과소 추정하였음을 알 수 있다. 한편 동일 스케일에서 합역 방식에 따른 평균 거리의 차이는 크지 않다. 종합하면 포아송 회귀 모델에서 공간적으로 분리된 구역 사이의 흐름은 과대 추정되는 경향이, 공간적으로 근접한 구역 사이의 흐름은 과소 추정되는 경향이 있으며, 합역 방식 보다는 합역 수준에 더 민감하게 반응하고 그 정도는 과대 추정된 잔차의 평균 거리가 더 심하다.

다음으로 합역 스케일과 방식에 따라 포아송 회귀 모델의 파라미터가 어떻게 변하는지 살펴보았다. 분석에서 기원지-목적지 제약 모델이 사용되었기 때문에 거리 파라미터(β)와 상수(λ) 값의 변화만을 살펴볼 수 있다. 상수는 특징적인 패턴을 보이지 않았다. Figure

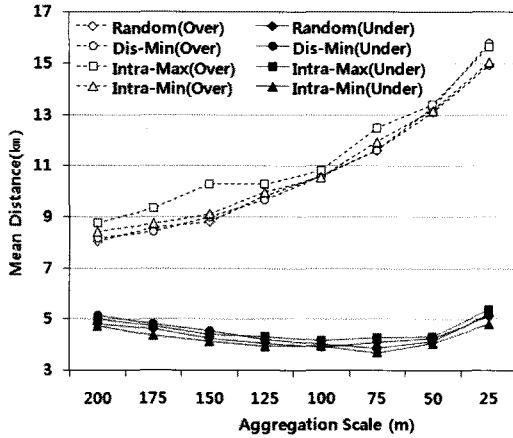


Figure 4. Change of mean distance of overestimated and underestimated residuals. 과대 및 과소 추정 잔차의 평균 거리 변화

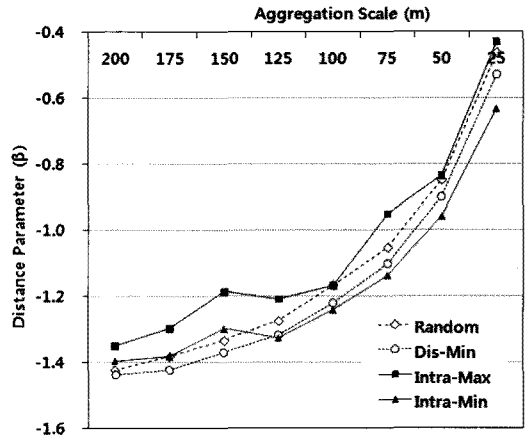


Figure 5. Change of distance parameter according to aggregation scale and schemes. 합역 수준과 방식에 따른 거리 파라미터 값의 변화

5는 합역 수준과 방식에 따른 거리 파라미터 값의 변화를 보여주고 있다. 모든 경우에 거리 파라미터가 음의 값을 보여 거리는 인구이동과 역의 관계에 있음을 알 수 있고 이러한 거리 마찰 효과는 합역 수준이 높아지면서 줄어드는 것을 알 수 있다. 이는 합역 수준이 높아지면서 구역 사이의 상호작용량은 증가하는 반면 평균 거리는 크게 변하지 않음으로써 거리 마찰 효과가 줄어든 것으로 해석할 수 있다. 예를 들어, Intra-Max

의 경우 $m=25$ 에서 평균 구역 간 거리는 13.18km, 평균 인구이동량은 1,296명인 반면, $m=200$ 에서 평균 구역 간 거리는 13.01km, 평균 인구이동량은 30명이다. 동일 합역 수준에서 대체적으로 Intra-Max의 거리 파라미터 값이 가장 작은 반면, Intra-Min의 값이 상대적으로 크게 나타났다. Intra-Max의 경우 상대적으로 밀집도가 높은 구역이 형성되면서 구역 간 이동량에서 편차가 상대적으로 적은 반면, Intra-Min의 경우 밀집

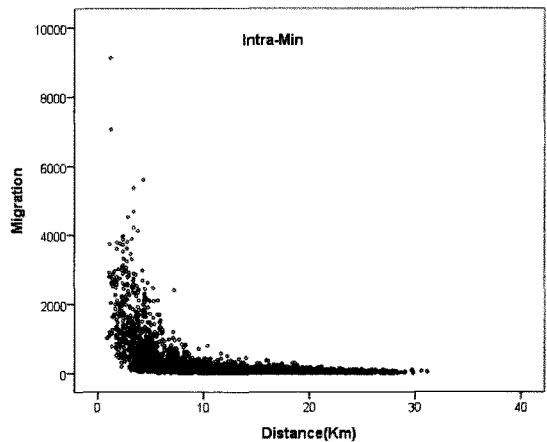
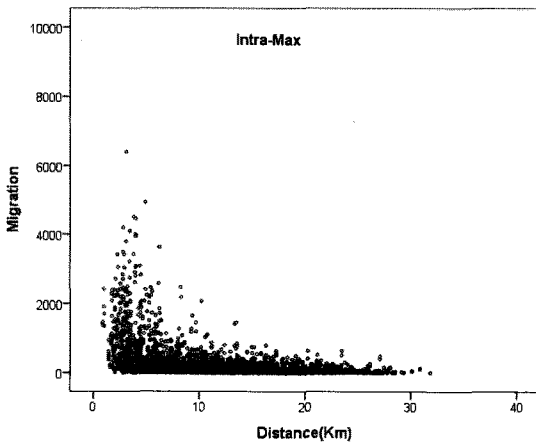


Figure 6. Relationship between distance and migration($m=75$). 거리와 인구이동량의 관계($m=75$)

도가 떨어지는 신장형의 넓은 구역이 형성되면서 특정 구역 사이에는 극단적으로 이동량이 많은 반면 나머지 지역 사이에는 이동량이 적어지면서 값의 편차가 커지게 되었고 이러한 값의 분포 차이가 거리 파라미터 값에 영향을 주었다. Figure 6의 거리와 상호작용량과의 관계를 통하여 이러한 합역 방식의 차이에 의한 거리 마찰 효과의 차이를 확인할 수 있다. Intra-Max의 경우 Intra-Min과 비교하여 거리가 증가하면서 상호작용량이 상대적으로 느리게 감소하는, 그래서 거리의 영향이 상대적으로 작게 나타나고 있다. 이러한 차이에 의해서 Intra-Max의 거리 파라미터 값이 Intra-Min의 거리 파라미터 값보다 작게 추정되었다.

Figure 7과 8은 각각 합역 스케일과 방식에 따른 포아송 회귀 모델의 적합도 측정치인 우도비와 이탈도(deviance)의 변화를 보여주고 있다. 우도비는 관측값과 예측값의 차이를 이용하여 산출된다. 관측값과 예측값의 차이가 크면 클수록 우도비 통계량은 커지는데, 이는 모델의 적합도가 낮아짐을 의미한다. 반대로 차이가 작으면 작을수록 우도비 통계량은 작아지고 모델의 적합도는 높아진다. 합역 수준에 따른 우도비의 변화를 살펴보면(Figure 7), Intra-Max의 경우 합역 수준이 증가하면서 적합도가 점차 높아지는 경향을 보인 반면, Intra-Min의 경우 합역 수준이 증가하면서 적합

도가 빠르게 낮아지는 대조적인 경향을 보인다. Random과 Dis-Min은 합역 수준이 증가하여도 적합도에서 거의 변화를 보이지 않다가 $m=25$ 에서 높아지는 패턴을 보이고 있다. 이를 통하여 합역 수준에 따른 적합도의 변화는 합역 방식에 따라 달라짐을 알 수 있다. 합역 방식별 적합도를 비교하면 Intra-Max가 다른 방식에 비하여 대체적으로 높은 적합도를 보여, 이 방식이 다른 방식에 비하여 모델링의 관점에서 좀 더 부합한다고 할 수 있다. 합역 방식에 따른 적합도의 차이는 합역 수준에 따라 다르게 나타났다. 합역 수준이 낮은 상태에서는 합역 방식별 차이가 크지 않은 반면, 합역 수준이 증가하면서 합역 방식별 차이가 커졌다.

한편 포아송 회귀 모델의 적합도를 평가하기 위하여 또 다른 통계량인 이탈도를 측정하여 그래프화 하였다(Figure 8). 이탈도는 모델의 부적합 정도를 측정하기 위한 것으로, 일반화 선형 모델(generalized linear models)에 대한 적합도 측정을 위하여 Nelder and Wedderburn(1972)이 제시한 통계량이다. 특정 모델의 부적합 정도는 $-2 \times \text{최대 우도값}$ 으로 측정되는데, 이탈도는 관심 모델(concerned model)과 완전 모델(complete model) 사이의 부적합 정도의 차이를 의미하고, 포아송 분포의 경우 다음과 같은 식에 의해서 계산된다:

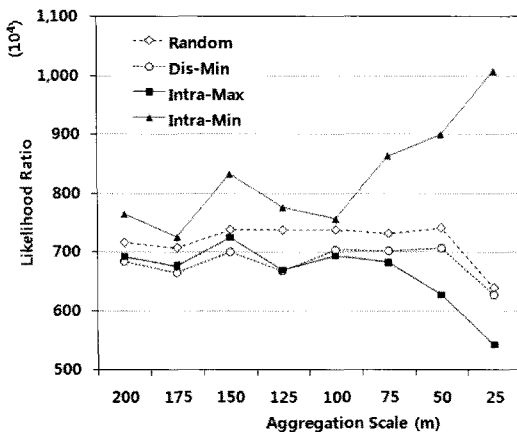


Figure 7. Change of likelihood ratio according to aggregation scale and schemes.
합역 수준과 방식에 따른 우도비의 변화

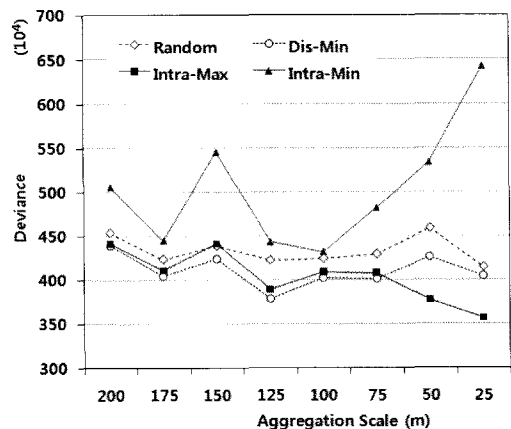


Figure 8. Change of deviance according to aggregation scale and schemes.
합역 수준과 방식에 따른 이탈도의 변화

$$2 \left\{ \sum_{i=1}^n \sum_{j=1}^n f_{ij} \ln \left(\frac{f_{ij}}{\hat{u}_{ij}} \right) - \sum_{i=1}^n \sum_{j=1}^n (f_{ij} - \hat{u}_{ij}) \right\} \quad (6)$$

여기에서 f_{ij} 는 i 와 j 지역 사이에서 관측된 상호작용량이고, \hat{u}_{ij} 는 모델에 의해서 추정된 상호작용량이다. 결국 이탈도는 모델이 설명하지 못한 데이터의 변이를 측정하기 때문에(Flowerdew and Aitkin, 1982), 이탈도가 크면 클수록 모델의 설명력이 낮다는 것을 의미하고, 반대로 이 값이 작으면 작을수록 모델의 설명력이 높다는 것을 의미한다. Figure 8에서 보는 것처럼, 합역 수준과 방식에 따른 이탈도의 변화는 우도비의 변화와 유사하다. Intra-Max의 경우 합역 수준이 증가하면서 이탈도가 감소하는, 그래서 모델의 설명력이 높아지는 경향을 보인 반면, Intra-Min의 경우 합역 수준이 증가하면서 이탈도가 증가하는 경향을 보여 모델의 설명력이 떨어지고 있다. 한편 Random과 Dis-Min의 경우 합역 수준에 따른 이탈도의 증감 경향이 뚜렷하지 않다. 합역 방식별로 비교를 하면 대체로 Dis-Min과 Intra-Max가 낮은 이탈도를 보이고 있다. 합역 방식별 편차는 합역 수준에 따라 달라짐을 알 수 있다. 우도비와 이탈도를 통한 적합도 평가를 종합하면, 모델의 수행 관점에서 Intra-Max가 가장 좋은 결과를 제시하였다. 또한 Intra-Max와 Intra-Min이 합역 수준에 따라 적합도의 변이가 커, MAUP에 민감하게 반응한 반면, Random과 Dis-Min은 합역 수준에 따른 적합도의 차이가 크게 나타나지 않아 MAUP에 상대적으로 덜 민감하게 반응하였다.

6. 요약 및 결론

본 연구의 목적은 인구이동 데이터와 같은 공간 상호작용 데이터를 스케일을 달리하여 합역하거나 혹은 동일 스케일에서 합역 방식을 달리하여 합역하였을 때 공간 상호작용 모델의 결과가 어떻게 달라지는지를 평가하는데 있다. 공간 상호작용의 복잡성, 공간적 재현 및 모델링의 필요성에 의해서 공간 상호작용 데이터를 합역하는 것은 불가피한 일이다. 공간 상호작용 데이

터의 합역은 공간단위 수정가능성의 문제(MAUP)를 야기한다. 즉 합역의 스케일이나 합역 방식에 따라 분석 및 모델링의 결과가 달라질 수 있다. 본 연구에서는 합역의 방법으로 1) 무작위로 구역 시드를 선정하고 근접한 공간단위를 할당하는 무작위 방법(Random), 2) 구역 시드와 단위 지역 사이의 인구 가중거리를 최소화하는 방법(Dis-Min), 3) 구역 내 상호작용 비율을 최대화하는 방법(Intra-Max), 4) 구역 내 상호작용 비율을 최소화하는 방법(Intra-Min) 등 네 가지를 사용하였다. Dis-Min은 교환 알고리즘의 목적함수로 정의하여 구획화에 적용하였고, Intra-Max와 Intra-Min은 자동 구획절차의 목적함수로 정의하여 적용하였다.

합역 수준 및 방식을 달리하여 산출된 구획 사이의 상호작용 데이터를 포아송 회귀 모델을 이용하여 MAUP의 영향을 분석한 결과는 다음과 같다. 첫째, 합역 수준이 높아지면서 밀집도가 떨어지는 경향을 보였고 그 정도가 가장 심한 것은 Intra-Min이었다. 동일 스케일에서 Intra-Max, Dis-Min, Random에 의해서 형성된 구획체계는 높은 밀집도를 보인 반면, Intra-Min에 의해서 형성된 구획체계는 낮은 밀집도를 보였다. 전체적으로 Intra-Max가 거의 모든 합역 수준에서 높은 밀집도를 보였고 합역 수준에 따른 밀집도의 변이도 크지 않았다. 이는 Intra-Max를 이용하여 설정된 구획 체계가 MAUP의 영향을 가장 적게 받으며 형상적 측면에서 가장 향상된 기능지역 구분을 가능하게 함을 의미한다. 둘째, 과대 및 과소 추정된 잔차의 공간적 특성을 살펴보기 위하여 평균 거리를 살펴본 결과, 합역 수준이 높아지면서 과대 추정된 잔차의 평균 거리는 증가하는 반면, 과소 추정된 잔차의 평균 거리는 감소하는 경향을 보였다. 동일 스케일에서 과대 추정된 잔차의 평균 거리는 과소 추정된 잔차의 평균 거리보다 크게 나타나 포아송 회귀 모델이 원거리 이동을 과대 추정하였고 근거리 이동을 과소 추정하였음을 알 수 있다. 그러나 동일 스케일에서 합역 방식별 평균 거리 차이는 크지 않았다. 잔차의 평균 거리의 경우 합역 방식보다는 합역 수준에 더 민감하게 반응하였고 그 정도는 과대 추정된 잔차의 평균 거리가 더 심하였다. 셋째, 거리 파라미터의 경우 합역 수준이 높아지면서 거리 마찰의 영향이 감소하는 경향을 보였는데, 이는

합역에 의해서 구역간 평균 이동량은 증가한 반면, 평균 거리는 거의 변하지 않았기 때문이다. 동일 스케일에서 Intra-max의 거리 파라미터 값이 가장 작은 반면, Intra-Min의 값은 상대적으로 크게 나타났다. 이는 Intra-Min을 이용하여 합역할 경우 극단적으로 큰 혹은 작은 구역이 형성되면서 구역간 이동량의 편차가 커졌기 때문이다. 마지막으로 포아송 회귀 모델의 우도비와 이탈도와 같은 적합도 측정치가 합역 수준과 방식에 따라 어떻게 달라지는지를 분석한 결과, Intra-Max의 경우 합역 수준이 높아지면서 모델의 적합도가 향상된 반면, Intra-Min의 경우 적합도가 악화되었다. Intra-Max와 Intra-Min이 합역 수준에 상대적으로 민감하게 반응한 반면, Random과 Dis-Min은 스케일의 영향을 상대적으로 덜 받았다. 동일 스케일에서 Dis-Min과 Intra-Max가 상대적으로 높은 적합도를 보였다.

이상의 분석 결과는 포아송 회귀 모델은 잔차의 공간적 특성뿐만 아니라 파라미터 추정값, 적합도 등에서 MAUP의 영향을 받는다는 것을 증명한다. 전체적으로 포아송 회귀 모델의 결과는 합역 방식 보다는 합역 수준에 더 크게 영향을 받으며, 동일 수준에서의 합역 방식의 차이는 스케일에 따라 달라진다. Intra-Min의 경우 밀집도, 우도비, 이탈도 등 구획의 공간적 특성과 회귀 모델의 적합도 측정치들이 합역 수준이 높아지면서 악화된 반면, Intra-Max는 향상된 결과를 보였는데, 이는 지역 구분에 초점을 둔 데이터 관점뿐만 아니라 모델링 관점에서 Intra-Max를 이용한 구획화가 보다 적합하다는 것을 의미한다. 또한 모델의 적합도 측면에서 Intra-Max와 Intra-Min이 스케일 효과에 상대적으로 민감하게 반응한 반면 Random과 Dis-Min은 덜 민감하게 반응하였는데, 이는 스케일 효과는 구획 방식에 따라 상이하게 나타난다는 것을 의미한다.

본 연구는 하나의 사례 분석을 통하여 공간 상호작용 모델에 대한 MAUP의 영향을 평가한 것으로, 이 연구의 결과만으로 MAUP의 영향을 일반화하는 것은 어렵다. 공간 상호작용 모델에 대한 MAUP의 영향을 일반화하기 위하여 본 연구의 관찰 사항을 뒷받침할 수 있는 다양한 사례연구가 요구된다. 또한 본 연구의 경우 단순 포아송 회귀 모델을 사용하였는데 향후 연구에서는 모델의 설명력을 높일 수 있도록 모델을 정교

화하고 이에 대한 MAUP의 영향을 평가하는 과정이 요구된다. 나아가 MAUP가 공간 데이터의 본질적 특성인 공간적 의존성에서 기인한다고 할 때(Lee, 1999), 공간적 의존성 측면에서 포아송 회귀 모델에 대한 MAUP의 영향을 평가하는 것은 의미 있는 연구가 될 것이다. 마지막으로 공간적 의존성 요소를 모델링하는 고유함수 공간 필터링(eigenfunction spatial filtering) 접근(Chun, 2008)이나 자기회귀 기반 공간 계량경제학적 접근(LeSage and Pace, 2008)에 대한 MAUP의 영향을 분석하는 것 또한 흥미로운 연구 주제가 될 것이다.

주

- 1) Masser and Brown(1975)의 Intramax와 구분하기 위하여 중간에 하이픈을 삽입하여 사용하였다.
- 2) 가장 보편적인 밀집도 측도인 '둘레길이/(2√π√면적)'를 이용하여 구획 별 밀집도를 계산한 후 평균하였다.
- 3) 평균 거리는 단위 지역 사이의 중심점 사이의 직선거리와 잔차를 곱하여 산출한 가중 평균의 형태를 띤다. 양 혹은 음의 잔차에 거리를 곱한 후 양 혹은 음의 잔차의 합으로 나누어 산출하였다.

참고문헌

- Alvanides, S., Openshaw, S., and Duke-Williams, O., 2000, Designing zoning systems for flow data, in Atkinson, P. and Martin, D., (eds.), *GIS and Geocomputation. Innovation in GIS 7*, Taylor and Francis, London, 115-134.
- Amrhein C. G. and Flowerdew R., 1992, The effect of data aggregation on a Poisson regression-model of Canadian migration, *Environment and Planning A*, 24(10), 1381-1391.
- Barras, R., Broadbent, T. A., Cordy-Haynes, M., Massey, D. B., Robinson, K., and Willis, J., 1971, An operational urban development model for Cheshire, *Environment and Planning A*, 3(2),

- 115-234.
- Batty M. and Sikdar, P. K., 1982a, Spatial aggregation in gravity models. 1. An information-theoretic framework, *Environment and Planning A*, 14(3), 377-405.
- Batty M. and Sikdar, P. K., 1982b, Spatial aggregation in gravity models: 2. One-dimensional population density models, *Environment and Planning A*, 14(4), 525-553.
- Batty M. and Sikdar, P. K., 1982c, Spatial aggregation in gravity models: 3. Two-dimensional trip distribution and location models, *Environment and Planning A*, 14(5), 629-658.
- Batty M. and Sikdar, P.K., 1982d, Spatial aggregation in gravity models: 4. Generalisations and large-scale applications, *Environment and Planning A*, 14(6), 795-822.
- Brown, L. A. and Holmes, J. H., 1971, The delimitation of functional regions, nodal regions and hierarchies by functional distance approaches, *Journal of Regional Science*, 11(1), 57-72.
- Chun, Y., 2008, Modeling network autocorrelation within migration flows by eigenvector spatial filtering, *Journal of Geographical Systems*, 10(4), 317-344.
- Flowerdew, R. and Aitkin, M., 1982, A Method of Fitting the Gravity Model Based on the Poisson Distribution, *Journal of Regional Science*, 22(2), 191-202.
- Fotheringham, A. S. and O'Kelly, M. E., 1989, *Spatial Interaction Models: Formulations and Applications*, Kluwer Academic Publishers, Dordrecht, Netherlands.
- Guo, D., 2009, Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data, *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 1041-1048.
- Hirst, M. A., 1977, Hierarchical aggregation procedures for interaction data: a comment, *Environment and Planning A*, 9(1), 99-103.
- Kim, K., 2004, Heuristic approaches for p-median location-allocation problem, *Journal of Geography Education*, 48, 14-30.
- Kim, K., Lee, G., and Shin, J., 2010, A study on reconstructing of local administrative districts using spatial analysis and modeling, *Journal of the Korean Association of Regional Geographers*, 16(6), 673-688 (in Korean).
- Kim, K., Shin, J., Lee, G., and Cho, D., 2009, A location model and algorithm for visiting health-care districting for the rural elderly, *Journal of the Korean Geographical Society*, 44(4), 813-831 (in Korean).
- Lee, S., 1999, The delineation of function regions and modifiable areal unit problem, *Journal of Geographical and Environmental Education*, 7(2), 757-783 (in Korean).
- LeSage, J. P. and Pace, R. K., 2008, Spatial econometric modeling of origin-destination flows, *Journal of Regional Science*, 48(5), 941-967.
- Masser, I. and Brown, P. J. B. (eds.), 1978, *Spatial Representation and Spatial Interaction*, Martinus Nijhoff Social Sciences Division, Boston.
- Masser, I. and Brown, P. J. B., 1975, Hierarchical aggregation procedures for interaction data, *Environment and Planning A*, 7(5), 509-523.
- Masser, I. and Scheurwater, J., 1980, Functional regionalisation of spatial interaction data: an evaluation of some suggested strategies, *Environment and Planning A*, 12(12), 1357-1382.
- Nelder, J. A. and Wedderburn, R. W. M., 1972, Generalized Linear Models, *Journal of the Royal Statistical Society A*, 135(3), 370-384.
- Openshaw, S. and Taylor, P. J., 1981, The modifiable areal unit problem, in Wrigley, N. and Bennett, R. J., (eds.), *Quantitative geography: a British view*, Routledge, London, 60-69.
- Openshaw, S., 1977a, A geographical solution to scale and aggregation problems in region-building, partitioning, and spatial modelling, *Transactions of the Institute of British Geographers*, NewSeries2, 459-472.
- Openshaw, S., 1977b, Optimal zoning systems for

- spatial interaction models, *Environment and Planning A*, 9(2), 169-184.
- Putman, S. H. and Chung, S-H., 1989, Effects of spatial system design on spatial interaction models. 1: The spatial system definition problem, *Environment and Planning A*, 21(1), 27-46.
- Ravenstein, E. G., 1885, The laws of migration, *Journal of the Statistical Society of London*, 48(2), 167-235.
- Slater, P. B., 1976, A hierarchical regionalization of Japanese prefectures using 1972 interprefectural migration flows, *Regional Studies*, 10(1), 123-132.
- Slater, P. B., 1981, Comparisons of aggregation procedures for interaction data: An illustration using a college student international flow table, *Socio-Economic Planning Sciences*, 15(1), 1-8.
- Teitz, M. B. and Bart, P., 1968, Heuristic Methods for Estimating the Generalized Vertex Median of a Weighted Graph, *Operations Research*, 16(5), 955-961.
- Tiefelsdorf, M. and Boots, B., 1995, The specification of constrained interaction models using the SPSS loglinear procedure, *Geographical Systems*, 2, 21-38.
- Tiefelsdorf, M., 2003, Misspecifications in interaction model distance decay relations: a spatial structure effect, *Journal of Geographical Systems*, 5(1), 25-50.
- Webber, M. J., 1980, A theoretical analysis of aggregation in spatial interaction models, *Geographical Analysis*, 12(2), 129-141.
- 교신: 김감영, 702-701, 대구광역시 북구 대학로 80 경북대학교 사범대학 지리교육과(이메일: kamyoungkim@knu.ac.kr, 전화: 053-950-5861)
- Correspondence: Kamyoung Kim, Department of Geography Education, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu, 702-701, Korea (e-mail: kamyoungkim@knu.ac.kr, phone: +82-53-950-5861)

최초투고일 2011. 2. 9
수정일 2011. 4. 15
최종접수일 2011. 4. 19