

출발지 공간 연관성을 고려한 지역별 수단선택확률 추정 연구

Estimating Probability of Mode Choice at Regional Level by Considering Spatial Association of Departure Place

엄진기* · 박만식[†] · 여태영**

Jin Ki Eom · Man Sik Park · Tae-Young Heo

Abstract In general, the analysis of travelers' mode choice behavior is accomplished by developing the utility functions which reflect individual's preference of mode choice according to their demographic and travel characteristics. In this paper, we propose a methodology that takes the spatial effects of individuals' departure locations into account in the mode choice model. The statistical models considered here are spatial logistic regression model and conditional autoregressive model taking a spatial association parameter into account. We employed the Bayesian approach in order to obtain more reliable parameter estimates. The proposed methodology allows us to estimate mode shares by departure places even though the survey does not cover all areas.

Keywords : Mode choice, Spatial association, Utility function, Spatial logistic regression model, Conditional autoregressive model

요 지 일반적으로 교통수단선택 모형은 이용자의 인구 및 개인통행특성 등을 반영한 수단별 선호도를 효용함수로 구축하여 분석하고 있다. 본 연구에서는 이용자의 출발지에 대한 공간적 연관성을 수단선택모형에 고려한 방법을 제시하였다. 이를 위하여 공간적 연관성을 포함하는 공간로지스틱 회귀모형을 고려하였다. 신뢰성있는 추정값을 얻기 위해 베이저안 기법을 적용하였으며 이 연구에서 제시한 방법론은 수단선호도 조사가 이루어지지 않은 지역에 대해서도 수단분담률을 추정할 수 있을 것으로 기대된다.

주 요 어 : 수단선택모형, 공간적 연관성, 효용함수, 공간로지스틱 회귀모형, 조건부 자기회귀모형

1. 서 론

일반적으로 교통수요예측을 위해서는 4단계(통행발생, 통행분포, 수단선택, 통행배정) 과정을 통해서 장래수요예측을 수행한다. 통행발생단계에서는 교통분석존(traffic analysis zone; TAZ)을 설정하여 존단위 통행목적별 발생량을 추정하게 되며, 통행분포 단계에서는 통행발생량을 출발지와 도착지 존으로 분포시킨다. 분포단계에서 추정된 OD(Origin and Destination)는 세 번째 단계인 수단선택에서 추정된 수단별 분담률을 적용하여 각 수단별 OD를

구축된다. 마지막단계인 통행배정단계에서는 각 수단별 OD를 네트워크(도로망, 대중교통노선망)에 배정하여 교통량 및 수송량이 결정된다. 기존 4단계 수요예측 단계 중 수단선택에서는 수단별 수단선택 비율을 추정하기 위해 교통수단에 대한 이용자 선호도 조사를 수행하며 조사된 자료는 각 수단별 효용함수(utility function)를 추정하는데 이용된다. 추정된 효용함수는 일반적으로 수단선택에 영향을 미치는 통행시간, 통행비용변수 등을 포함하며 추정된 수단선택모형은 분석지역내의 수단별 통행시간, 비용 등을 고려한 평균적인 수단분담률을 제시한다. 따라서 통행자의 통행시간, 비용이 같으면 출발지 위치에 상관없이 유사한 수단분담률을 제시하게 되는 한계를 가지고 있다. 이는 선호도조사가 전수조사가 아닌 표본조사임에도 불구하고 조사응답자의 모든 출발지 또는 거주지에 대하여 대표성을 가지기 때문이다.

[†] 책임저자 : 교신저자, 정회원, 고려대학교, 의학통계학교실
E-mail : man.sik.park@gmail.com

TEL : (02)920-6439 FAX : (02)953-6439

* 정회원, 한국철도기술연구원, 철도교통물류연구실

** 정회원, 한국해양대학교, 데이터정보학과

통행자의 출발지에 의한 지역적인 공간효과(spatial effect)는 교통네트워크에 의한 접근성을 대변할 수 있는 접근통행시간 등에 의해 단순히 고려되므로 이용자들의 평균적인 통행비용에 의해 수단선호도가 결정되는 구조를 가지기 때문이다. 공간적 연관구조를 교통수단선택에 적용한 국내 연구 사례는 아직 없으며 외국의 경우 지역분석(Location Analysis)에 활용된 사례들이 있다. Varameth 외(2005)는 공간효과를 주저지 선택분석에 활용하기 위해 공간효과를 mixed logit 모형의 틀에서 적용하였으며 기존의 일반적인 다항로짓모형 등과 비교분석 하였다[1]. Hider 외(2005)는 공간효과를 포함한 다항로짓모형을 구축하여 거주자들의 주택선택에 대한 행태를 분석하여 거주자들의 주택선택은 공간적으로 이웃한 지역의 주거선택에 의해 영향을 받는 것으로 분석결과를 제시하였다[2].

본 연구에서는 교통수단선택에 있어 통행자의 출발지에 따른 공간적 연관구조가 영향을 미친다는 가정을 전제로 지역별로 수단분담률이 추정가능 하도록 하기 위해 이용자의 출발지를 중심으로 수단분담률 추정에 있어 공간적 연관구조를 모형에 포함시키는데 목적이 있다. 이를 위해 표본자료를 기반으로 출발지 위치를 GIS기능을 활용하여 획득한 후 개별 응답자의 지역 간 교통수단인 철도에 대한 선호도를 공간적 영향을 고려하여 추정하도록 한다. 따라서 조사되지 않은 지역의 수단선택 확률의 추정이 가능하고 공간적으로 수단전환 등 다양한 분석이 가능할 것으로 기대된다.

2. 공간모형의 고찰

2.1 공간자료

공간모형에서는 $s \in D \subset \mathbf{R}^d$ 를 d 차원 유클리드 공간(Euclidean space) 상의 연구자의 관심지역 D 에서의 일반적인 자료의 위치 지점이라 하고, 공간 위치지점 s 에서 관측 가능한 자료 $Y(s)$ 를 다변량 확률장 (multivariate random field)로 정의한다.

$$\{Y(s) : s \in D\}$$

집합 D 의 속성에 따라서 $\{Y(s) : s \in D\}$ 는 지리통계자료, 지역자료, 점 패턴 자료와 같이 세 가지의 형태로 나누어진다. 일반적으로 공간 자료는 특정 위치 또는 특정 지역에서의 관측 값이나 측정값으로 구성되며, 특히, 공간 자료는 지리통계자료(point-referenced data), 지역자료 (areal data), 공간 점 패턴자료 (spatial point patterns data)의 세 가지 유형으로 분류된다. 지리통계자료는 서울시 34개 측정기계가에서 관측되는 미세먼지 농도가 대표적인 예로써 측

정기계가 없는 지역의 예측을 목적으로 한다. 이와는 대조적으로 지역자료는 서울시 25개 각 구에서 집계한 특정 질환자의 수가 대표적인 예로써 모형의 적합을 목적으로 한다. 지금까지의 연구들을 살펴보면 지리통계 자료에 대한 분석 및 연구는 많은 반면에 상대적으로 지역자료에 대한 연구는 다소 미진하다. 이에 본 논문에서는 실제 자료가 지역자료에 적합하므로 지역자료의 모형화를 중점적으로 다루고자 한다.

2.2 공간로지스틱회귀모형

본 절에서는 교통수단(열차와 버스)에 대한 선호도를 분석하기 위한 이항분포(Binomial distribution) 기반의 공간로지스틱회귀모형(spatial logistic regression model)을 소개하며, 베이지안(Bayesian) 방법인 마코프 연쇄 몬테칼로(Markov chain Monte Carlo) 기법을 이용하여 공간로지스틱회귀모형 안에 포함되어 있는 모수를 추정하였다.

서울시의 25개 구별로 열차와 버스의 교통수단 선택에 대한 승객들의 선호도에 대한 공간적 예측을 위하여 이항분포모형을 제시하였으며, 공간적 연관성을 모형에 포함하기 위해 제시한 공간로지스틱회귀모형은 식 (1)과 같다.

$$Y(s_i) \sim B(n_i, \pi(s_i))$$

$$\log\left(\frac{\pi(s_i)}{1-\pi(s_i)}\right) = x(s_i)^T \beta + \nu_\rho(s_i) \quad (1)$$

여기서 B 는 성공의 확률(열차를 선택할 확률), $\pi(\cdot)$ 을 갖는 이항분포를 나타내고 $Y(s_i)$ 는 i 번째 지역(구)에서 발생한 관심 사건의 회수(열차를 이용한 경우)이며, 가산자료(count data)의 형태로 나타난다. 또한 $x(s_i)$ 는 i 번째 지역(구)에 대한 설명변수의 열벡터(column vector)이고 $\nu_\rho(s_i)$ 는 공간적 연관성을 나타내는 임의효과(random effect)를 의미하며 ρ 는 공간적 연관성의 강도(strength)를 나타내는 모수이다. 성공의 확률, $\pi(\cdot)$ 은 다음과 같은 가정을 한다. 모든 $i=1, \dots, n$ 에 대하여

$$0 \leq \pi(s_i) \leq 1.$$

일반화선형모형(generalized linear model)중 하나인 로지스틱회귀모형의 관측값은 앞에서 언급한 바와 같이 가산자료이다. 이를 연속형(continuous) 또는 범주형(categorical) 설명변수의 선형결합으로 표현하기 위해서는 종속변수, $Y(s_i)$ 또는 성공의 확률, $\pi(s_i)$ 의 적절한 변환이 필요하다. $Y(s_i)$ 의 분포가 이항분포이고 이는 지수분포족(Exponential distribution family)에 포함되므로 이 연구에서는 이러한 변

환들 중 하나인 정준연결(canonical link)인 로짓변환(logit transformation)을 고려하였으며 아래의 로짓변환의 형태는 다음과 같은 구간을 갖게 된다. 모든 $i = 1, \dots, n$ 에 대하여

$$-\infty < \log\left(\frac{\pi(s_i)}{1-\pi(s_i)}\right) < \infty.$$

성공의 확률을 변환한 로짓은 식 (1)의 등호 오른쪽에 위치한 설명변수들의 선형결합과 동일한, 실수전체의 구간을 가정할 수 있게 된다. 아울러 공간적 연관성을 내포하는 임의효과인 $\nu_\rho(s_i)$ 에 대해서도 자연스럽게 정규분포(normal distribution) 가정을 할 수 있게 된다. 공간적 연관성의 모수를 임의효과인 $\nu_\rho(s_i)$ 에 포함시키기 위해 $\nu_\rho(s_i)$ 의 분산-공분산행렬(variance-covariance matrix)을 ρ 에 의해 표현가능한 모수적(parametric)인 모형들을 고려하게 된다. 이 연구에서는 조건부 자기회귀모형(conditional autoregressive model)을 고려하였다.

2.3 조건부자기회귀모형

조건부 자기회귀모형은 연구자의 관심지역 상에서 관측되어진 $\{Z(s_i); i = 1, \dots, n\}$ 가 정규 확률과정(Gaussian stochastic process)을 따르고 s_1, s_2, \dots, s_n 이 연구자의 관심지역인 집합 D 의 격자형태(lattice type)를 가진다고 가정한다.

$$\bigcup_{i=1}^n s_i = D, \quad s_i \cap s_j = \emptyset, \quad \forall i \neq j.$$

이러한 가정 하에서 $\{Z(s_i)\}$ 을 아래와 같은 조건부 자기회귀모형으로 표현할 수 있다.

$$Z(s_i) | \{Z(s_{-i})\} \sim N\left(\sum_{k=1}^n w_{ik} Z(s_k), \tau_i^2\right).$$

여기서, $\{Z(s_{-i})\} = \{Z(s_k); \forall k \neq i\}$ 이고, N 은 정규분포를 나타내며, $w_{ik}\tau_k^2 = w_{ki}\tau_i^2$ 이며, τ_i^2 는 i 번째 지역(구)에서의 조건부 분산(conditional variance)을 나타낸다. 공간적 연관성을 모형에 포함하기 위해 필요한 이웃행렬(neighborhood matrix), $\mathbf{W} = \{w_{ij}\}$ 은 0 또는 1의 값을 가지는 원소들로 구성된다. 즉, 서로 두 지역이 이웃하고 있으면 $w_{ij} = 1$, 그렇지 않으면 $w_{ij} = 0$ 을 가지는 이웃정보를 구성하고 조건부 분산은 $\tau_i^2 = \sigma_c^2 / w_{i+}$ 로 표현될 수 있다. 여기서, σ_c^2 은 공통 분산을 나타내고 $w_{i+} \equiv \sum_{j=1}^n w_{ij}$ 는 i 번째 지역(구)에 이웃한 지역들의 수를 의미한다.

이웃행렬, $\mathbf{W} = \{w_{ij}\}$ 와 조건부 분산으로 이루어진 행렬, $\mathbf{T} = \text{diag}\{\tau_1^2, \tau_2^2, \dots, \tau_n^2\}$ 을 이용하여 인수분해정리에 의한

$\mathbf{z} = (Z(s_1), Z(s_2), \dots, Z(s_n))^T$ 의 결합 확률분포(joint probability distribution)는 $\mathbf{z} \sim N_n(\mathbf{0}, (\mathbf{I} - \mathbf{W})^{-1}\mathbf{T})$ 로 표현될 수 있다(Besag, 1974)[3]. 여기서 N_n 은 n 차원 다변량 정규분포(n -dimensional multivariate normal distribution)을 의미하고 \mathbf{I} 은 단위행렬(identity matrix)을 나타내며 $\mathbf{0} = (0, 0, \dots, 0)^T$ 이다. $\mathbf{I} - \mathbf{W}$ 는 역행렬(inverse matrix)이 존재하고, $(\mathbf{I} - \mathbf{W})^{-1}\mathbf{T}$ 는 대칭(symmetric)인 양정치 행렬(positive-definite matrix)이다. 공간 상관성을 나타내는 모수 ρ 를 고려한 조건부자기회귀모형은 식 (2)와 같이 정의할 수 있다.

$$\mathbf{z} \sim N(\mathbf{0}, (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{T}) \tag{2}$$

이 때 $Z(s_i) \sim \text{CAR}(\rho)$ 라고 표현하며 이러한 조건부자기회귀모형이 식 (1)에서 공간적 연관성을 나타내는 임의효과인 $\nu_\rho(s_i)$ 에 대한 모형으로 적용할 수 있다.

2.4 모형 추정 방법

본 연구에서 제안된 공간로지스틱회귀모형의 모수 추정을 위하여 베이지안 기법을 이용하였으며 식 (2)에서 제시된 모형을 적용하기 위하여 모수들에 대한 사전분포(prior distribution)에 대한 정의가 선행되어야 한다. 본 연구에서는 주요 모수들에 대한 사전분포들은 다음과 같은 무정보 사전분포(non-informative prior distribution)를 갖는 접근방법을 이용하였다(Hughes-Oliver et al., 2009)[4].

$$\begin{aligned} \beta_p &\sim N(0, 10^3), p = 1, \dots, 4, & \nu_\rho(s_i) &\sim \text{CAR}(\rho), \\ \sigma_c^2 &\sim \text{gamma}(10^{-3}, 10^{-3}), \\ \rho &\sim \text{uni form}(\lambda_{\min}^{-1}, \lambda_{\max}^{-1}) \end{aligned}$$

여기서, λ_{\min} 과 λ_{\max} 은 $\mathbf{T}^{-1/2}\mathbf{W}\mathbf{T}^{-1/2}$ 의 최소, 최대 고유값(eigen value)을 의미하고, 공간상관모수인 ρ 는 이들 고유값의 역수를 범위로 가지게 된다.

3. 실증분석

3.1 조사자료 및 모형

본 연구에 이용된 자료는 경부고속철도 1단계 개통이전인 2001년에 경부고속철도와 연계교통체계 기본계획 수립을 위해 한국철도기술연구원에서 조사한 것으로 지역 간 교통수단에 대한 선호의식 조사 자료를 이용하였다[5]. 선호의식 조사는 승용차 이용자를 포함하여 주요 열차역, 버스터미널, 공항 등을 이용한 통행자를 대상으로 수행되었다. 본 연구에서는 서울시에서 출발하여 경상권, 전라권,

충청권을 통행하는 철도와 버스 통행자를 대상으로 출발지 주소를 기반으로 GIS프로그램을 활용하여 서울시의 구 단위로 표본을 추출하였다. Table 1은 각 구 단위로 철도 이용자에 대한 분포를 보여주고 있으며 모형구축을 위한 설명변수로 이용자들의 평균소득, 평균통행거리, 평균비용 등을 제시하고 있다. 출발지를 기준으로 수단선택의 공간적 연관성을 반영한 모형과 공간적 연관성을 고려하지 않은 일반적인 로지스틱모형을 비교하도록 한다.

Table 1. Summary of Data Sampled

서울시	평균소득 (만원)	평균거리 (km)	평균비용 (원)	철도이용자수	전체 응답자수
종로구	222	195.0	17691	13	23
중구	202	223.4	21684	14	22
용산구	215	247.1	24434	10	19
성동구	234	197.2	20006	12	16
광진구	241	241.2	22209	7	17
동대문구	207	191.5	18165	6	10
중랑구	217	231.1	22422	4	9
성북구	273	192.3	16888	8	20
강북구	238	118.3	11283	3	6
도봉구	252	181.1	19317	8	9
노원구	207	164.7	16582	13	18
은평구	207	192.7	18227	5	11
서대문구	295	198.1	18903	9	16
마포구	190	231.0	21866	13	20
양천구	278	239.2	23438	9	12
강서구	223	170.6	17411	6	9
구로구	213	203.9	23978	8	9
금천구	340	200.0	21256	6	8
영등포구	245	185.8	19768	34	39
동작구	210	238.8	20313	10	26
관악구	197	236.8	19976	8	25
서초구	239	246.9	20442	8	32
강남구	211	226.5	21477	17	55
송파구	219	202.3	19413	11	23
강동구	251	173.3	17000	4	6
계				246	460

모형 1. 일반 로지스틱회귀모형

$$Y(s_i) \sim B(n_i, \pi(s_i))$$

$$\log\left(\frac{\pi(s_i)}{1 - \pi(s_i)}\right) = x(s_i)^T \beta$$

모형 2. 공간 로지스틱회귀모형 (식 (1)과 식 (2) 참조)

2.2절에서 소개한 공간로지스틱회귀모형은 일반 로지스틱회귀모형에서의, 관찰개체(본 연구에서는 지역구 간의 독립성 조건을 가정하지 않는다. 즉, 관찰개체 간의 상호연관성을 가정한다. 특히, 특정 지역과 그 지역의 인접 지역들(이웃) 간에 선택수단은 서로 연관되어 있을 것으로 가정하고 지역들 간의 연관성 정도를 나타내는 구조를 일반 로지스틱모형에 포함한 형태가 바로 공간로지스틱회귀모형이다. 따라서 모형 2. 하에서 본 연구에서 고려한 설명변수들 뿐만 아니라 공간적 상호연관성의 효과도 특정 지역의 수단선택에 영향을 주는 지를 파악할 수 있다.

3.2 모수추정결과

본 연구에서 제시한 공간로지스틱회귀모형의 모수를 추정하기 위하여 Tanner and Wong(1987)[6], Gelfand and Smith(1990)[7]에 의해서 소개된 마코프 연쇄 몬테칼로 방법을 적용하여 범용 베이저안 통계 패키지인 WinBUGS (Spiegelhalter et al., 2000)[8]를 사용하였으며, 추정된 모수값의 수렴상태를 알아보기 위하여 세 가지의 서로 다른 초기값을 이용하였으며, 사후분포(posterior distribution)를 통해 모수 추정값(posterior mean; 사후평균)을 얻을 수 있었다.

추정된 모수의 올바른 수렴 여부를 확인하기 위해 Gelman and Rubin(1992)[9]에 의해 제안된 방법을 이용하였다. 이 방법은 여러 개의 초기값을 가지고 수렴여부를 진단하는 것으로서 척도값이 1에 가까워지면 모든 초기값에 의해 생성된 모수들이 동일한 분포로 수렴한다는 것을 의미한다. 특정 지역의 이웃은 지역 구분 경계선을 공유하는 지역들로 정의한다.

본 연구에서는 제시하지 않았지만 공간로지스틱회귀모형에 대하여 세 개의 초기치를 적용한 결과 모수들의 사후분포 표본들에 대한 Gelman-Rubin 수렴진단 척도값이 1에 가깝게 나타나고 있으므로 생성된 모수들이 잘 수렴하여 올바른 추정 값을 확인하였다. Table 2는 공간적 연관성을 반영하지 않은 모형(모형 1)하에서의 추정값들을 나타낸다.

Table 2. Parameter estimates and confidence intervals from Model 1

	일반 로지스틱회귀모형 (모형 1.)	
	추정값	95%신뢰구간
상수항	0.340	-194.4 ~ 194.0
평균소득	0.013	-0.222 ~ 0.248
총거리	-1.291	-1.709 ~ -0.888
총비용	1.142	0.716 ~ 1.581

Table 2를 통해 알 수 있듯이 평균소득은 열차선호도에 영향을 미치지 않는 것으로 나타난 반면에 총거리와 총비

용은 열차 선호도에 영향을 미치는 것으로 나타났다. 총비용은 열차선호도에 정(+)의 영향을 미치는 반면에 총거리는 부(-)의 영향을 미치는 것으로 나타났다. 이는 상대적인 지역 간 교통수단인 버스에 비해 통행거리가 증가할수록 열차를 선호할 확률이 낮아지는 반면에 총비용이 증가할수록 상대적으로 저렴한 열차를 선호하는 경향이 반영된 것으로 분석된다.

Table 3은 공간적 연관성을 반영한 모형에 대한 추정값을 나타낸다. ρ 는 음의 추정값으로 나타났으나 95%신뢰구간 안에 0이 포함됨에 따라 통계적으로는 유의하지 않은 것으로 나타났다. 공간적 연관성을 반영하지 않은 모형의 추정값(Table 1)과 비교해 보면 역시 평균소득은 통계적으로 유의하지 않은 반면, 총거리와 총비용은 유의한 것으로 나타났다.

3.3 DIC와 MSPE 관점에서의 모형 비교

본 연구에서는 공간적 연관성을 반영하지 않은 모형과 반영한 모형을 비교하기 위한 기준들로 DIC(Deviance Information Criterion)와 MSPE(Mean Squared Prediction Error)를 고려하였다.

주어진 모형의 적합도를 평가하는 데 있어서 항상 문제가 되는 점은 모형이 복잡해질수록 항상 자료를 더 잘 적합할 수 있다는 것이다.

Table 3. Parameter estimates and confidence intervals from Model 2

	공간 로지스틱회귀모형 (모형 2.)	
	추정값	95%신뢰구간
상수항	0.389	0.147 ~ 0.658
평균소득	0.047	-0.239 ~ 0.330
총거리	-1.487	-1.995 ~ -1.011
총비용	1.289	0.789 ~ 1.828
공간적 연관성	-0.003	-0.046 ~ 0.031
공통분산	0.479	0.047 ~ 1.119

한편, 모형의 설명 능력은 우도함수의 값으로 측정될 수 있으나 복잡한 모형보다는 간단한 모형이 선호되어야 한다는 보수 절약의 원리에 의하여 추정되어질 모수의 수가 많은 것에 대해 벌칙(penalty)을 부가함으로써 모형의 절약성을 고려할 수 있다. 베이저안 기법에서 모형의 설명력과 절약성은 서로 상충되는 기준으로서 여러 가지 모형 선택의 판단기준을 함께 고려한 선택 기준이 필요하게 되는데 DIC 기준을 이용하였다.

베이저안 적합도 기준인 DIC는 일반적인 적합도 기준인 AIC(Akaike Information Criterion), BIC(Bayesian Information Criterion)와 같이 모형의 복잡성을 고려한 적합도

기준이다.

모형 적합도는 $D(\theta) = 2\log[f(\mathbf{y}|\theta)] + 2\log[h(\mathbf{y})]$ 의 사후분포 평균인 $\bar{D} = E_{\theta, \mathbf{y}}[D(\theta)]$ 이며 $h(\mathbf{y})$ 는 모형에 의하여 영향을 받지 않는다. $\theta = (\beta_1, \beta_2, \beta_3, \beta_4, \rho, \sigma_c^2)$ 이고 모형의 복잡성은 모형에 영향을 주는 모수의 갯수인 $p_D = \bar{D} - D(\hat{\theta})$ 로 측정되어질 수 있다. 여기서 $D(\hat{\theta})$ 는 모수의 사후평균, $\hat{\theta}$ 에서 평가된 데비언스의 값이다. 결과적으로 $DIC = \bar{D} + p_D$ 로 정의되며 작은 값일수록 모형 적합이 잘 되었다고 할 수 있다.

교통수단 선호도 모형을 위하여 제시된 두 모형에 대하여 가장 적합한 모형을 선택하기 위하여 앞서 설명한 DIC와 아래와 같이 정의되는 MSPE를 통하여 최적 모형을 선택하였다. 물론, MSPE는 DIC와는 달리 모형에 포함된 모수의 수에 대한 벌칙은 고려되지 않지만 DIC와 마찬가지로 모형과 실제자료와의 괴리 정도를 측정하는 적합기준으로 많이 사용된다. DIC와 MSPE 척도는 값이 작을수록 모형이 주어진 자료에 잘 적합되었다고 할 수 있다.

$$MSPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{Y(s_i)}{n_i} - \hat{\pi}(s_i) \right)^2$$

여기서, $\hat{\pi}(s_i)$ 는 각 모형 하에서 구한 i 번째 지역(구)에서 열차를 선택할 사후확률(posterior probability 혹은 posterior proportion)값이다. 표 4는 모형 비교를 위한 DIC와 MSPE 척도의 값을 제시하였다. 먼저, MSPE 관점에서는 공간 영향을 고려하지 않은 모형 1(일반 로지스틱 회귀모형)보다 모형 2(공간 로지스틱 회귀모형)가 훨씬 더 좋은 적합 정도를 보이고 있다. 하지만 이는 모형 2에 포함된 모수가 모형 1에 비해 많음으로 인해 우연히 발생된 결과일 수 있다. 따라서 모수의 개수에 대한 벌칙을 고려하는 적합기준인 DIC를 각 모형에 적용한 결과 모형 1이 공간적 상관구조를 고려한 모형 2에 비해 주어진 자료를 설명하는 데 다소 부적합하다는 판단을 내리게 된다.

Table 4. Diagnostic measures from Model 1 and Model 2

	일반 로지스틱		공간 로지스틱	
	추정값	95%신뢰구간	추정값	95%신뢰구간
DIC	116.9		106.5	
MSPE	0.017	0.015 ~ 0.022	0.010	0.006 ~ 0.016

3.4 열차 선호도에 대한 사후확률 지도

Fig. 1은 서울시 25개구의 지역 간 통행자의 열차 선호도에 대한 실제 확률(=열차를 이용한 통행자 수/전체 통행자 수)을 나타낸 그림이다. 관측된 실제 비율을 모형 1과 모형 2를 적용한 결과 Fig. 2와 Fig. 3과 같은 사후확률에 대한 지

도를 구축하였다. Fig. 2는 공간적 연관성을 반영하지 않은, 획일화된 평균적인 수단 선호도에 바탕을 둔 일반 로지스틱모형(모형 1)에 의해 얻어진 열차선호도에 대한 사후확률지도이고 Fig. 3은 공간적 연관성을 반영한, 즉 모형 1과는 다소 차별화된 공간 로지스틱회귀모형(모형 2)로부터 얻어진 열차선호도에 대한 사후확률의 결과를 나타내고 있다.

서울시에서 출발한 지역 간 통행자중 열차를 이용한 사람들에 대한 선호도 조사결과를 서울시 구별단위로 확률을 제시한 결과 Fig. 1에서 보는 바와 같이 강북지역이 강남지역에 비해 열차 선호도가 높은 것으로 나타났으며 강남지역의 경우 동작구, 관악구, 서초구, 강남구의 열차 선호도는 상대적으로 낮은 것으로 분석되었다.

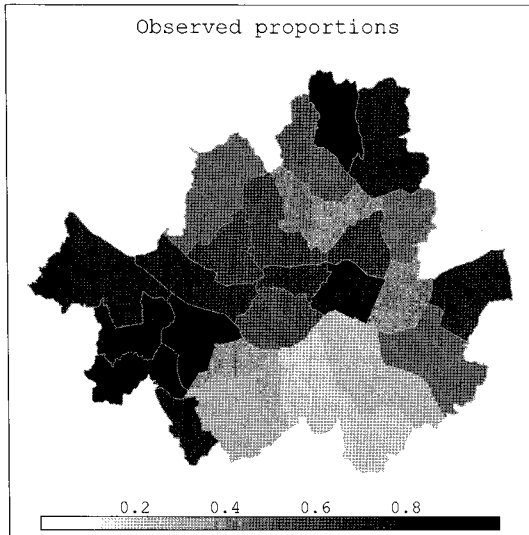


Fig. 1. Observed probability map for train preference

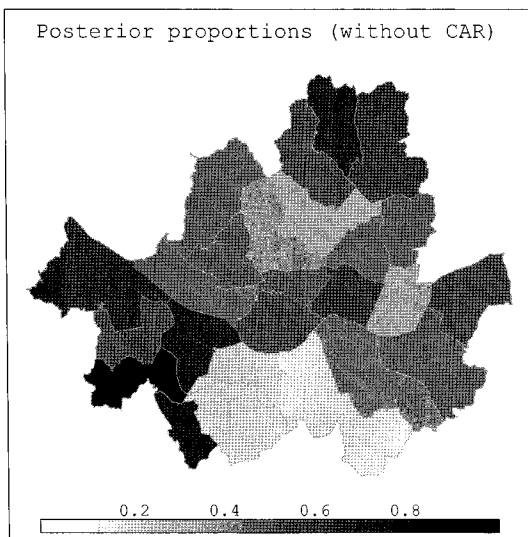


Fig. 2. Posterior probability map for train preference from Model 1

모형의 비교를 위해서 열차수단 선호도의 공간적 연관성을 반영하지 않은 모형(Fig. 2)과 반영한 모형(Fig. 3)을 Fig. 1과 비교해 보면 열차 선호도가 공간적 연관성을 반영하지 않은 로지스틱모형보다 공간적 연관성을 반영한 로지스틱모형에서 얻은 결과와 유사함을 알 수 있다.

이는 Fig. 4를 통해서 보다 분명해 진다. 즉, 특정 수단의 선택에 있어서 설명변수의 효과와 더불어 인접지역 간의 유사성의 효과를 모형에 포함시킴으로써 보다 신뢰성 있는 예측 및 적합이 이루어졌음을 각 모형에서 얻은 사후확률과 실제 확률을 비교해 봄으로써 알 수 있다.

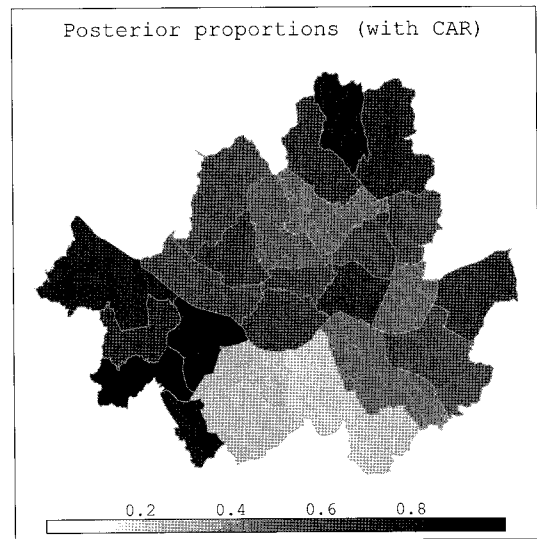


Fig. 3. Posterior probability map for train preference from Model 2

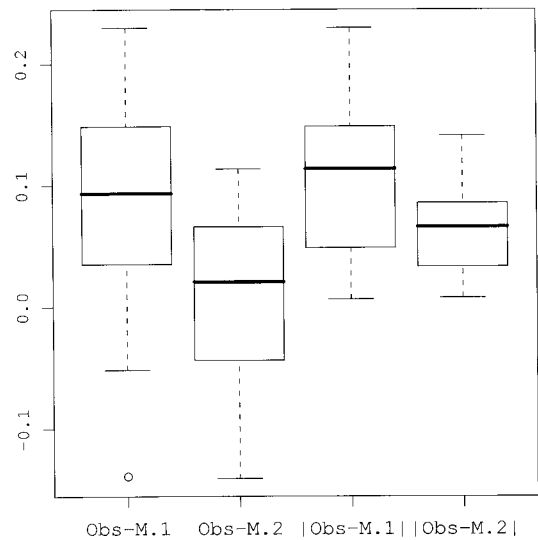


Fig. 4. Box plot for comparison between observed probability and posterior probability from Model 1 and 2

실제 확률(Obs)에서 각 모형(M.1, M.2)에서 얻은 사후확률을 뺀 차이값(Obs-M.1, Obs-M.2)을 이용하여 구성된 상자그림을 보면 일반 로지스틱회귀모형(M.1) 하에서의 사후확률은 전체적으로 과소추정된 경향을 보이고 있으나 공간 로지스틱회귀모형(M.2)은 차이값의 분포가 0을 중심으로 대칭적인 형태를 취하고 있으므로 안정적인 적합이 이루어졌음을 알 수 있다.

또한 차이의 절대값(|Obs-M.1|, |Obs-M.2|)을 비교한 결과 모형 1(M.1)의 변동이 모형 2(M.2)에 비해 훨씬 심한 것을 쉽게 파악할 수 있다. 따라서 공간 상관성을 반영한 모형을 통해 통행자의 수단 선호도를 예측하는 것이 보다 정확하고 신뢰성있는 결과를 도출하는 것으로 판단된다.

4. 결론

본 연구에서는 지역 간 교통수단인 철도에 대하여 출발지(구단위)를 중심으로 공간적 연관성을 반영한 새로운 교통수단선택 모형을 구축하였다. 이용자의 출발지 공간적 연관성을 반영한 모형은 그렇지 않은 모형에 비해 수단선택 확률을 보다 정확하게 추정하는 것으로 분석되었으며 특히 지역적 특성을 반영할 수 있는 장점을 가지는 것으로 분석되었다. 그러나 본 연구에서 적용한 데이터는 서울시 구단위의 수단선택자료로서 충분하지 못한 한정된 표본으로 인해 공간적 특성을 반영하는 파라미터의 추정치가 통계적으로 유의하지 않은 것으로 분석되어 향후 통계적으로 유의한 모형을 구축하기 위해서는 보다 많은 데이터의 수집이 필요하다.

공간적 연관성을 반영한 수단선택 모형은 향후 철도시설과 같이 특정 지역에 신규서비스가 이루어질 경우 이에 따른 공간적 영향을 반영한 수단분담의 추정이 가능하므로 향후 활용도가 높을 것으로 분석된다. 또한 대중교통 우선 정책에 따라 여러 가지 정책적인 변수에 의해 타 수단으로부터 열차 수단으로 전환이 될 경우(modal shift) 수단 전환에 따른 환경개선효과(배기가스 절감효과) 등을 지역적으로 분석하는데 유용하게 이용될 것으로 판단된다.

본 연구는 서울시 25개 구 단위로 모형을 구축하여 분석

하였으나 보다 많은 자료를 이용할 경우 서울시의 540여개 동단위로 수단선호도의 분석이 가능할 것으로 판단된다. 그러나 현재 모형에서는 승용차와 항공 수단이 제외되어 향후 연구과제는 다양한 수단의 영향에 대한 공간적 분석이 필요할 것으로 판단된다.

참고문헌

1. Varameth, V., Kazuaki, M., Yoshiyuki, T.(2005), "Mixed Logit Model Framework with Structuralized Spatial Effects: A Test of Applicability with Area Unit Systems in Location Analysis," *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 6, pp. 3789-3802.
2. Haider, M. and Miller, E.J.(2000), "Effects of Transportation Infrastructure and Location on Residential Real Estate Values: Application of Spatial Autoregressive Techniques," *Transportation Research Record 1722*, TRB, National Research Council, Washington, D.C., 2000, pp. 1-8.
3. Besag, J.(1974). "Spatial interaction and the statistical analysis of lattice systems (with Discussions)," *Journal of the Royal Statistical Society, Series B*, Vol. 36, 192-236.
4. Hughes-Oliver, J. M., Heo, T-Y, Ghosh, S.(2009), "An Autoregressive point source model for spatial processes," accepted in *Environmetrics*.
5. 경한국철도기술연구원(2003), "부고속철도 연계교통체계 구축 기본계획 수립 연구."
6. Tanner, M. A. and Wong, W. H.(1987), "The calculation of posterior distributions by data augmentation (with discussion)," *Journal of the American Statistical Association*, Vol. 82, pp. 528-550.
7. Gelfand, A. E. and Smith, A. F. M.(1990), "Sampling based approaches to calculating marginal densities," *Journal of the American Statistical Association*, Vol. 85, pp. 398-409.
8. Spiegelhalter, D. J., Thomas, A., and Best, N. G.(2000), "WinBUGS Version 1.3 User Manual." Cambridge: Medical Research Council Biostatistics Unit.
9. Gelman, A. and Rubin, D. B.(1992), "Inference from iterative simulation using multiple sequence (with Discussion)," *Statistical Science*, Vol. 7, 457-511.

접수일(2009년 2월 13일), 수정일(2009년 6월 3일),
게재확정일(2009년 9월 3일)