

국지적 조건부 자기회귀 모형을 활용한 교통사고 위험의 상대적 위험도 추정

임영빈*

*국민대학교 행정학과

e-mail: yblym0207@kookmin.ac.kr*

Estimating the Relative Risk of Road Traffic Injuries via a Localized Conditional Autoregressive Model

Youngbin Lym*

*Dept. of Public Administration, Kookmin University

요약

본 연구는 인명피해 교통사고의 상대적 위험도를 소지역 단위에서 정밀하게 추정하고, 공간적 불균형 및 국지적 고위험 지역을 식별하기 위해 계층적 베이지안 모형을 기반으로 한 국지적 조건부 자기회귀 모형을 적용하였다. 분석 대상은 미국 오하이오주 클리블랜드-엘리리아 광역권의 1,691개 센서스 블록그룹에서 2015년부터 2019년까지 발생한 46,013건의 인명피해 사고이며, 다양한 사회인구학적, 환경적, 물리적 특성을 설명변수로 활용하였다. Leroux CAR 모형과 국지적 LCAR 모형을 비교 적용하였으며, 베이지안 추론은 MCMC 기법을 통해 수행되었다.

Goodness of fit measure인 DIC 및 WAIC 결과값에 따라 클러스터 수가 4개인 국지적 CAR 모형이 가장 우수한 모형 적합도를 보였으며, 이질적인 공간 구조를 반영함으로써 기존 전역적 모형보다 높은 설명력을 보였다. 저임금 노동자 비율, 차량 미소유 가구, 활동 밀도, 토지이용 다양성, 통행 균형, 교차로 밀도 등은 인명피해 사고의 상대 위험을 유의하게 증가시키는 요인으로 나타난 반면, 인구밀도와 빈집 비율은 음의 영향을 미쳤다. 또한, 국지적 클러스터별로 통계적으로 유의한 수준의 절편 차이가 관측되었으며, 클러스터 4에서는 주변 지역 대비 상대 위험이 현저히 높은 이질적 공간 패턴이 확인되었다. 본 연구를 통해 기존 교통사고 위험 분석에서 고려하지 못했던 교통사고 상대 위험도의 국지적 공간 불균형을 정밀하게 파악할 수 있었다. 이를 더욱 발전시켜 지역 맞춤형 교통안전 정책 수립에 실증적 기반을 제공하고자 한다.

키워드: 국지적 조건부 자기회귀 모형, 공간 종속성, 공간 이질성, 인명피해사고, 소규모 지역, 상대 위험도

1. 서론

공간적 이질성(spatial heterogeneity)은 공중보건, 도시계획, 환경과학 등 다양한 분야에서 관측되는 주요한 특징으로, 특정 지역의 사회경제적 조건이나 환경적 요인이 공간적 패턴에 따라 변동성을 띠는 현상을 의미한다. 교통사고는 공간적 요인과 밀접하게 연결된 복합적 사회현상으로, 사고 발생의 패턴은 단순히 무작위적으로 분포하지 않고 지역의 물리적 환경, 교통 인프라, 사회경제적 요인, 정책적 개입 등 다양한 요소에 의해 공간적으로 구조화된 자기상관(spatial autocorrelation)을 형성한다. 이에 따라 교통사고의 발생 메커니즘을 이해하고 정책적 개입점을 도출하기 위해서는 공간통계학적 접근이 필수적이며, 특히 공간 단위별 위험도 추정은 실증 분석과 정책 설계 모두에 중요한 기반이 된다.

이러한 공간 자기상관 및 공간적 이질성을 고려한 대표적인 통계모형 중 하나가 베이지안 통계 기반의 조건부 자기회귀 모형(Conditional Autoregressive model, CAR)이다(Besag,

1974; Besag et al., 1991). CAR 모형은 인접 지역 간의 상호작용을 수학적으로 구조화함으로써 공간 단위별 위험도의 정확한 추정을 가능하게 하며, 교통사고, 범죄, 질병 역학 등 다양한 분야에서 활발히 적용되고 있다. 그러나 전통적인 CAR 모형은 모든 지역 간 공간 의존성이 동일하게 작용한다는 전제 하에 정적인 인접 행렬을 사용한다는 점에서 중요한 한계를 지닌다. 즉, 전역적(global) 평활화를 전제로 하여 공간적으로 급격한 위험도 변화나 국지적 클러스터(local clusters)가 존재할 경우, 이러한 미시적 구조를 과도하게 평활화(over-smoothing)하여 실제 공간적 불균형을 왜곡할 수 있다.

교통사고와 같은 복합적이고 국지적 맥락에 따라 달라지는 사회현상에서는 특정 지역이 주변과는 통계적으로 유의미하게 다른 위험도를 보이는 이질적인 공간 구조를 형성할 수 있다. 예컨대, 고령 운전자 밀집 지역, 특정 교차로 구조적 위험, 또는 정책 사각지대 등은 경계를 따라 분절(discontinuities)되거나 소규모 고위험 집합(cluster)을 형성할 수 있으며, 이러한 국지적 특성을 정확히 탐지하지 못하면 위험도 추정의 정확도는 물론, 정책적 개입의 정밀성도 저하된다.

이러한 문제의식에 기반하여 최근의 공간통계학 연구에서는 전통적인 CAR 모형의 한계를 극복하고 국지적 이질성을 반영하는 확장된 공간모형, 즉, 국지적 조건부 자기회귀 모형(Localized CAR)이 제안되고 있다. 국지적 조건부 자기회귀 모형은 공간적 자기상관의 강도를 지역별로 조정함으로써 동질적인 공간 영역 내에서는 평활화를 유지하면서도, 경계 영역에서는 비평활화를 허용하는 방식으로 작동한다. 특히 Lee & Mitchell (2012) 및 Rushworth et al. (2017)의 연구에서는 인접 지역 간의 연결성 자체를 확률적으로 모델링함으로써 자동으로 공간적 경계를 탐지하고, 클러스터의 존재를 반영할 수 있는 베이저안 프레임워크를 제안하여 이론적·실증적 유효성을 입증하였다.

교통사고 분석에 국지적 CAR 모형을 도입하는 것은 단지 통계적 적합도를 개선하는 것을 넘어, 고위험 지역을 보다 정밀하게 식별하고, 국지적 특성에 기반한 공간 맞춤형 정책 제언으로 이어질 수 있다는 점에서 그 장점이 크다 하겠다. 본 연구는 이러한 문제의식을 바탕으로, 인명피해 교통사고의 상대 위험도 분석에 Lee & Sarran (2015)의 국지적 CAR 모형을 적용함으로써 기존의 전역적 접근이 간과할 수 있는 지역 간 이질성과 공간적 불연속성을 정밀하게 반영하는 분석틀을 제시하고자 한다.

2. 연구방법

2.1 분석데이터

본 연구는 임영빈(2024)의 연구에서 활용한 미국 오하이오주 북동부에 위치한 광역 클리블랜드 지역 내 1,691개 소지역인 센서스 블록그룹(census block group)을 대상으로 수행하였다. 연구에서 활용한 교통사고 데이터는 2015-2019년 사이에 발생한 46,013건의 인명피해 사고로, 각 블록그룹의 5년 동안의 누적사고 건수를 집계하여, 소규모 지역의 인명피해 교통사고의 상대 위험도를 추정하고자 하였다.

분석에서 활용한 변인은 사회인구학적, 환경적, 물리적 요인을 포함한 다양한 변수로 구성되어 있다. 주요 변수로는 인구밀도(PopDen), 활동밀도(ActDen), 저임금 노동자 비율(LowWage), 차량 미소유 가구 비율(NoCar), 빈집 비율(Vacancy) 등의 사회인구학적 지표와, 토지이용 다양성(Landuse), 고용·주거 혼합도(WorkEmpMix), 통행 균형지수(TripEq), 도로 밀도(RoadNetwork), 보행자 중심 및 자동차 중심 교차로 밀도(Street, Intersection), 차량주행거리(VMT)가 포함되었다. 아울러, 지역의 도시화 수준을 반영하기 위해 농촌(Rural), 교외(Suburban), 도시(Urban) 구분의 범주형 변수도 도입하였다. VMT는 블록그룹 별 교통량의 지역적 차이를 반영하기 위한 것으로 변수의 추정이 이루어지지 않고 오프셋(offset)으로 적용하여 사고의 누적발생빈도가 아닌 상대 위험도 추정에 활용하였다.

2.2 연구방법

본 연구는 계층적 베이저안 기반의 일반화 선형 혼합 모형을 채택하고 있는데, 설명변수의 고정효과 이외 공간이 가진 확률효과를 설명하기 위해 조건부 자기회귀 모형을 활용하였다. 전역적 공간의 영향은 Leroux et al. (2000)이 제안한 모형(Leroux CAR, LCAR)을 적용하였고, 국지적 클러스터 파악을 위해 Lee & Sarran (2015)이 제안한 Localized CAR 모형을 적용하였다. Localized CAR 모형에서는 공간의 영향이 국지적으로 발생하며, 국지적 클러스터별 절편을 통해 공간적으로 급격한 변화 또는 비유사성을 설명하고자 하였다. 이러한 이해를 바탕으로 본 연구에서는 설명변수만의 고정효과 모형(GLM)과 전역적 조건부 자기회귀 모형인 LCAR 모형, 그리고 국지적 클러스터 기반의 Localized CAR 모형을 활용하여, 미국 오하이오주 북동부 광역 클리블랜드 내 소규모 지역의 인명피해 사고의 상대적 위험도에 대하여 고찰하고자 하였다.

3. 분석결과

본 연구에서 제안한 모형의 분석은 베이저안 모형의 추정에서 일반적으로 활용되는 마르코프 체인 몬테카를로(Markov Chain Monte Carlo, MCMC)를 통해 이루어졌다. MCMC 체인 1개에 초기 번인(burn-in) 샘플 50,000개, MCMC 샘플 550,000개 및 MCMC 체인 샘플의 자기상관을 줄이기 위해 매 100번째의 샘플을 가져올 수 있도록 thin = 100으로 지정해 주어 총 5,000개(500,000/100)의 MCMC 샘플을 도출하였다.

[표 1]은 본 연구에서 고려한 모형들에 대해 MCMC를 통해 도출된 결과를 정리한 것으로, 고정효과 모형(GLM)과 전역적 조건부 자기회귀 모형인 LCAR 및 국지적 CAR 모형의 분석 결과를 담고 있다. 국지적 CAR 모형의 경우 클러스터의 숫자를 2에서 6까지 적용하여 분석을 수행하였는데, G = 4의 경우가 최적의 모형으로 산출되어 이를 [표 1]에 정리하였다. 일반적으로 베이저안 모형에서는 DIC와 WAIC가 낮을수록 모형의 성능이 뛰어난 것으로 알려져 있다. 따라서 본 연구에서의 최적 모형은 클러스터가 4개인 국지적 CAR 모형이며, 이 모형을 바탕으로 통계적 추론이 이루어졌다.

[표 1] 모형의 성능 비교

Fit Measure	GLM	LCAR	Localized CAR (G=4)
DIC	32,289.24	11,006.25	9,935.07
WAIC	32,680.93	10,667.06	10,822.49
Loglikelihood	-16,130.86	-4,049.31	-4,656.52

[표 2]는 국지적 CAR 모형의 분석결과를 정리한 것으로, 우선 고정효과를 살펴보면, 저임금 노동자의 비율, 차량을 소유하지 않은 가구의 비율, 활동 밀도, 토지이용의 다양성, 통행 발생 및 통행 유발 균형, 및 고용 및 주거의 다양성 지수, 도로 밀도 및 자동차 중심 교차로 밀도가 높아질수록 인명피해 사고의 상대적 위험이 유의미하게 높아지는 것으로 나타났다. 한편, 인구밀도와 빈집 비율 변수의 경우 유의미한 음의 영향을 미치는 것으로 나타났다. 그리고 농촌지역에 비해 교외 및 도시화 지역의 인명피해 위험이 높은 것으로 추정되었다.

한편 확률효과의 경우 공간효과의 전역적 분산인 τ^2 의 분포 이외 클러스터별 절편의 분포가 각각 $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ 로 나타나고 있으며, 그 값들은 통계적으로 유의미하게 다르게 분포되어 있음을 알 수 있다.

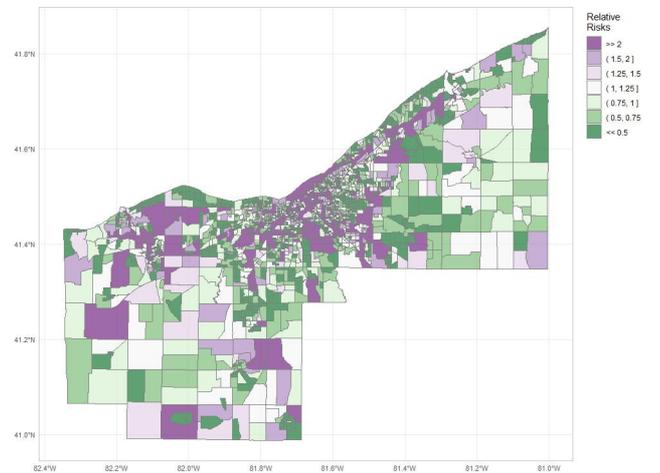
[표 2] :Localized Cluster-based CAR model의 분석결과

Variables	Median	0.025Q	0.975Q
Fixed effects			
LowWage	0.168	0.105	0.229
LogPopDen	-0.58	-0.658	-0.499
NoCar	0.103	0.052	0.152
ActDen	0.108	0.072	0.148
Landuse	0.25	0.208	0.292
TripEq	0.089	0.051	0.127
WorkEmpMix	0.139	0.1	0.179
RoadNetwork	0.098	0.007	0.192
Street	-0.035	-0.102	0.036
Intersection	0.143	0.103	0.18
Vacancy	-0.105	-0.154	-0.054
Suburban	0.281	0.063	0.489
Urban	0.448	0.205	0.677
Random effects			
λ_1	-1.348	-1.669	-0.875
λ_2	-0.694	-0.89	-0.486
λ_3	-0.487	-0.733	-0.213
λ_4	0.764	0.52	1.02
τ^2	1.445	1.102	1.873
δ	1.008	1	1.03

Goodness of fit	Value
DIC	9,935.07
WAIC	10,822.49
Log-likelihood	-4,656.52

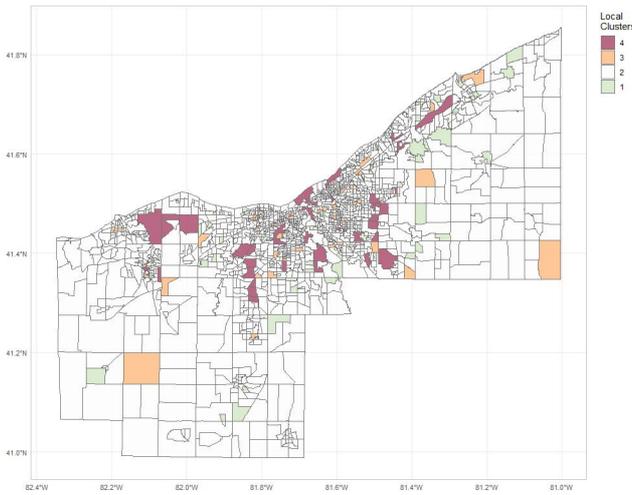
4. 논의 및 결론

본 연구에서는 설명변수만의 고정효과 모형과 전역적 공간 모형인 LCAR 모형, 국지적 조건부 자기회귀 모형을 통해 인명피해 교통사고의 상대 위험도에 있어서의 공간의 영향을 파악하고자 하였다. LCAR 및 국지적 CAR 모형 모두 GLM에 비해 매우 높은 모형 설명력을 보여주었으며, 광역 클리블랜드 내 소지역 인명피해 교통사고를 이해함에 있어 관측불가능한 공간적 불확실성 또는 공간적 요인을 고려하는 것이 타당함을 시사한다. [그림 1]은 전역적 조건부 자기상관 회귀 모형인 LCAR 모형을 기반으로 한 교통사고의 상대 위험도의 공간적 분포를 나타낸다. 보라색으로 제시된 곳들이 주변 센서블록그룹의 평균에 비해 2배 이상의 위험이 있는 곳이며, 녹색으로 나타나는 곳들은 상대적으로 위험이 낮은 곳들을 의미한다.



[그림 1] 전역적 LCAR 모형 기반의 상대 위험도

[그림 1]에 따르면, 교통사고 위험이 높은(또는 낮은) 곳들의 경우 인접한 이웃 역시 사고의 상대위험도가 높은(낮은) 공간적 자기상관을 보이고 있음을 알 수 있다. 즉, CAR 모형에서는 평활화를 통해 사고의 상대 위험도가 인접한 이웃의 정보를 반영하여 추정되기 때문에, 주변 지역과 유사한 값을 지니게 된다. 하지만, [그림 1]을 구체적으로 살펴보면, 특정 지역의 경우 인접한 주변 지역과 유의미하게 다른 상대 위험도를 보이는 이질적인 공간 구조를 형성하고 있음을 알 수 있다. 이러한 국지적 특성을 반영하여 국지적 CAR 모형을 추정하였으며, [그림 2]와 같이 제시하였다.



[그림 2] 국지적 CAR 모형 기반의 국지적 클러스터 분포

[표 2]에 따르면 클러스터 별 절편의 사후 중간값(posterior median)은 -1.348(클러스터 1), -0.694(클러스터 2), -0.487(클러스터 3), 그리고 0.764(클러스터 4)로 추정되며, 해당 클러스터에 속하는 센서스블록들의 인명피해 사고의 상대위험도가 인접한 지역과 유사하지 않고, 차이가 발생하는 현상을 설명하고 있다. [그림 2]의 클러스터 4에 해당하는 블록그룹의 경우 주변지역의 상대위험도가 낮음에도 불구하고 높은 것으로 나타나는 이 질적인 공간 구조를 가지고 있음을 파악할 수 있으며, 이러한 국지적 특성을 정확히 탐지하고자 하는 본 연구의 목적이 이루어졌다고 판단된다.

참고문헌

[1] 임영빈, “데이터마이닝과 베이지안 기법을 활용한 소규모 지역의 부주의 운전으로 인한 교통사고 위험 모형 구축 -미국 오하이오주 북부 지역을 대상으로-”, *Crisisonomy*, 제 20권 3호, pp. 106-116, 3월, 2024년.

[2] J. Besag, “Spatial interaction and the statistical analysis of lattice systems,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 192-236, 1974

[3] J. Besag, Y. Jereny, and M. Annie, “Bayesian Image Restoration, with Two Applications in Spatial Statistics.” *Annals of the Institute of Statistical Mathematics*, vol. 42, no. 1, pp. 1-20, 1991.

[4] D. Lee and R. Mitchell, “Boundary detection in

disease mapping studies,” *Biostatistics*, vol. 13, no. 3, pp. 415-426, Jul. 2012

[5] D. Lee and C. Sarran, “Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies,” *Environmetrics*, vol. 26, no. 7, pp. 477-487, Nov. 2015

[6] B. G. Leroux, X. Lei, and N. Breslow, “Estimation of disease rates in small areas: A new mixed model for spatial dependence,” in *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, M. E. Halloran and D. Berry, Eds. New York, NY: Springer, 2000, vol. 116, pp. 179-191.

[7] A. Rushworth, D. Lee, and C. Sarran, “An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 66, no. 1, pp. 141-157, Jan. 2017