

베이지안 확률 기반 범죄위험지역 예측 모델 개발*

허선영¹ · 김주영² · 문태현^{2*}

Crime Incident Prediction Model based on Bayesian Probability*

Sun-Young HEO¹ · Ju-Young KIM² · Tae-Heon MOON^{2*}

요 약

범죄는 장소나 건축물 용도에 따라 발생빈도와 유형이 다르고, 그 장소를 이용하는 사람들의 특성 및 공간 구조 차이에 의해 다양하게 발생한다. 따라서 공간 및 지역특성을 포함한 공간 빅데이터를 활용하여 지역을 분석해 보면 범죄예방 전략을 마련할 수 있다. 아울러 빅데이터와 지능 정보화시대의 도래에 따라 예측적 경찰활동이 새로운 경찰활동의 패러다임으로 등장하고 있다. 이에 보편적인 지방도시 J시를 대상으로 3개년 동안의 7,420건의 실제 범죄사례를 바탕으로 도시공간의 물리·환경적인 특성을 분석하여 범죄발생공간을 규명하고, 위험지역을 예측해 보고자 하였다. 분석에는 다양한 빅데이터 중 범죄를 유발하는 도시 공간 내 물리·환경적 요소에 한하여 공간 빅데이터를 구축하여 공간회귀분석을 실시하였다. 다음으로 분석결과 도출된 가로폭, 평균 층수, 용적율, 1층 사용용도(제2종 근린생활시설, 상업시설, 유흥시설, 주거시설)을 변수로 베이지안확률 기반 범죄발생 위험성 예측 모형(CIPM: Crime Incident Prediction Model)을 개발하였다. 개발된 모델은 실제 범죄발생 지역과의 중첩분석 및 모델의 정확도를 판단하는 Roc curve 분석을 통해 AUC 값이 0.8로 모델이 적합한 것으로 나타났다. 개발된 모델을 토대로 사례지역의 범죄 위험도를 분석한 결과 범죄발생은 상업 및 유흥시설이 밀집된 지역과 건물층수가 높은 지역, 그리고 상업 및 유흥시설과 주거가 혼재해 있는 블록이 범죄발생 확률이 높은 것으로 나타났다. 본 연구는 단순히 범죄의 공간적 분포와 범죄발생 영향요인을 탐색하는 기존의 연구와 달리 범죄발생 예측모형을 확률론적 관점에서 개발하는 영역으로 한 단계 진전되었다는 점에 의의가 있다.

주요어 : 베이지안 확률, 공간분석, 범죄예방, 공간 빅데이터, 범죄발생위험지역, 예측모델

2017년 10월 31일 접수 Received on October 31, 2017 / 2017년 12월 14일 수정 Revised on December 14, 2017 / 2017년 12월 19일 심사완료 Accepted on December 19, 2017

* 이 연구는 한국연구재단의 기초연구사업(2017R1A2B4012254) 및 2017년도 경상대학교 연구년제연구교수 연구지원비에 의하여 수행되었으며 11th AURG International Symposium에서 발표된 논문을 수정·보완하였음.

1 경상대학교 공학연구원 Engineering Research Institute(ERI), Gyeongsang National University

2 경상대학교 도시공학과 Bk21+, ERI, Dept. of Urban Engineering, Gyeongsang National University

* Corresponding Author E-mail : thmoon@gnu.ac.kr

ABSTRACT

Crime occurs differently based on not only place locations and building uses but also the characteristics of the people who use the place and the spatial structures of the buildings and locations. Therefore, if spatial big data, which contain spatial and regional properties, can be utilized, proper crime prevention measures can be enacted. Recently, with the advent of big data and the revolutionary intelligent information era, predictive policing has emerged as a new paradigm for police activities. Based on 7420 actual crime incidents occurring over three years in a typical provincial city, “J city,” this study identified the areas in which crimes occurred and predicted risky areas. Spatial regression analysis was performed using spatial big data about only physical and environmental variables. Based on the results, using the street width, average number of building floors, building coverage ratio, the type of use of the first floor (Type II neighborhood living facility, commercial facility, pleasure use, or residential use), this study established a Crime Incident Prediction Model (CIPM) based on Bayesian probability theory. As a result, it was found that the model was suitable for crime prediction because the overlap analysis with the actual crime areas and the receiver operating characteristic curve (Roc curve), which evaluated the accuracy of the model, showed an area under the curve (AUC) value of 0.8. It was also found that a block where the commercial and entertainment facilities were concentrated, a block where the number of building floors is high, and a block where the commercial, entertainment, residential facilities are mixed are high-risk areas. This study provides a meaningful step forward to the development of a crime prediction model, unlike previous studies that explored the spatial distribution of crime and the factors influencing crime occurrence.

KEYWORDS : *Bayesian Probability, Spatial Analysis, Crime Prevention, Spatial Big Data, Crime Incident Risk Area, Prediction Model*

서 론

2017 노벨 경제학상에 세계적인 베스트셀러 ‘넛지(Nudge)’의 저자인 미국 시카고대 리처드 세일러 교수가 선정되었다. 넛지는 ‘팔꿈치로 슬쩍 옆구리 찌르기’라고 하면서 주의를 환기시켜 타인의 선택을 유도하는 부드러운 개입 방식을 정의하고 있다(Thaler and Sunstein, 2009). 넛지효과를 활용한 사례로 암스테르담 공항 화장실의 파리그림이 그려진 소변기를 통해 화장실을 깨끗하게 사용하라는 말이나 문구가 없어도 기존 대비 80% 이상 오염이 줄었을 정도로 효과가 있었던 사례가 대표적이다. 국내에

서도 횡단보도 대기구역에 바닥과 벽을 노랗게 칠한 공간에 어린이가 보행신호를 기다리게 하는 ‘옐로카펫’을 설치하여 자동차 운전자가 어린이를 발견할 확률을 높임으로써 교통사고를 예방하는데 효과가 있는 사례가 있다. 이러한 효과가 사회적으로 중요하게 인식되고 있으며, 범죄예방 분야에서도 찾아 볼 수 있다.

예를 들면, 도시에 깨진 유리창이 많을수록 범죄발생이 증가한다는 깨진 유리창 이론과 다른 시각에서 접근하는 범죄예방환경설계 (CPTED: Crime Prevention Through Environment and Design)가 대표적이라 할 수 있다. CPTED는 범죄가 발생한 후의 사후처리 보다 범죄에 취약한 지역의 환경을 디자인적으로 개선하여 공권

력 없이 범죄발생을 사전에 예방하는 이론이다. 즉 과거와 같이 형벌을 통해서 범죄를 억제하지 않고 범죄가 발생하는 공간 환경을 개선함으로써 범죄를 예방하려는 움직임으로 학계는 물론 실무에서 주목받고 있다. 하지만 한국의 현실상 범죄데이터가 기본적으로 비공개이기 때문에 학술적 연구가 제대로 이루어지지 못하고 있는 문제점이 지적되고 있다. 뿐만 아니라 범죄자료를 지리정보로 구축하지 못함으로써 고도의 범죄분석과 현장중심의 수사 및 방법정책 마련에 어려움이 있어 개선이 요구된다.

하지만 미국, 영국, 호주, 일본 등에서는 이러한 한계를 극복하고 있으며, 최근에는 ICT 기술의 발전과 더불어 빅데이터 저장 및 분석이 가능한 환경이 조성되면서 다양한 범죄예측기법들이 학계와 실무에서 시도되고 있다. 특히 인터넷과 SNS 등과 같은 스마트 기기 사용이 일상화됨으로 인해, 실시간으로 생산되는 방대한 양의 빅데이터 정보를 처리하여 의미 있는 정보를 찾아내고 활용하고 있다. AI(Artificial Intelligence)가 이끄는 지능정보시대의 도래에 따라 이러한 발전 속도가 더 빨라질 전망이다. 빅데이터는 기업뿐만 아니라 정부와 개인, 연구 분야에서도 중요한 이슈로 다가왔으며, 홍수와 같은 자연재해, 질병발생, 교통 혼잡 등 매우 다양한 분야에서 활용하고 있다. 하지만 우리나라에서 범죄 안전 분야의 경우 범죄자료를 취득하기 어렵고, 예측에 대한 이론적 연구와 구현에 필요한 기술력 또한 아직 부족하여 새로운 도전분야로 남아있다.

범죄는 장소, 건축물 용도 등에 따라 범죄발생 특성이 상이하하며, 이는 그 장소를 이용하는 사람들의 특성과 공간 구조 차이에 의해 발생한다. 따라서 공간 및 지역특성을 포함한 공간 빅데이터를 활용하여 범죄예방 전략 대책을 마련한다면 효과적일 것이다. 실제로 미국이나 영국과 같이 몇몇 앞서가는 나라의 사례에서는 빅데이터 기반으로 범죄발생을 예측하여 예방하거나 범인 검거와 같은 사후대처 활동에 활용하여 도움을 주는 사례(Burch and Geraci, 2009; Smith, 2014)가 보고되고 있어 주목 받고 있다.

공간적 정밀도 측면에서 보면, 이들 사례는

공간속성을 포함한 상세한 자료를 사용함으로써 필지단위의 마이크로(micro)한 공간단위로 정밀한 범죄 분석이 가능하고, 공간적 범죄발생 장소를 예측하는 수준까지 도달해 있어 시사하는 바가 크다고 할 것이다. 그러나 우리나라는 아직 공간적으로 동이나 구 단위로 제공되는 집계자료를 이용함으로써 선진사례와 상당한 수준의 차가 있다고 할 수 있다.

이러한 배경에서 본 연구는 블록 단위의 마이크로한 공간범위를 대상으로 범죄발생과 공간의 물리적 특성간의 관계를 파악한 후, 범죄발생 가능성을 예측하는 모델을 개발해 보고자 한다. 본 연구에서 시도되는 모델은 사례도시의 실제 범죄 자료와 도시의 공간 빅데이터를 기반으로 하고 있다. 그러나 범죄발생에 미치는 영향 요소는 그 도시사회의 역사, 문화, 사회적 환경 등 매우 다양하지만 본 연구에서는 첫 단계로 도시의 물리적이며 환경적인 공간요소에 국한하여 검토해 보고자 한다. 나머지 요인을 반영한 모델은 본 연구 결과를 바탕으로 후속 연구로 진행할 예정이다.

선행연구 및 이론적 고찰

1. 선행연구 고찰

그동안의 범죄관련 연구는 다양한 분야에서 진행되고 있으나 크게 범죄발생에 대한 영향요인분석, 범죄지도화 및 통계적 공간분석, 범피 두려움, CCTV, 환경설계를 통한 범죄예방에 관한 연구 등으로 구분할 수 있다. 본 연구에서는 범죄발생에 대한 영향요인과 관련된 연구사례를 집중 검토하였다.

범죄발생에 대한 영향요인을 분석한 연구는 대부분 범죄발생 사회경제적 특성에 기반으로 한 연구가 대부분이다. 국외 연구 사례로서 Torruam and Abur(2014)은 나이지리아의 실업률, 인플레이션과 범죄 간의 인과 관계를 분석하였으며, 실업률이 높아지면 소득을 창출하는 기회가 줄어들 범죄를 유발한다고 주장하였다. Haider and Ali(2015)은 범죄에 영향을 주는 사회 경제적

요인을 분석하였으며, 인구밀도, 교육수준, 실업, 산업화, 송금액 중 인구밀도와 실업은 범죄에 긍정적이고 중요한 영향을 미치는 것을 밝혀내었다. 또한 교육수준과 송금액은 범죄발생과 반비례 관계가 있는 것을 확인하였다.

국내 연구는 주로 동 단위 또는 도시 단위로 범죄의 공간적 분포를 시각화하거나 공간분석을 시도하고 있다. 그러나 이들 연구도 전역적 차원에서 분석 및 해석은 가능하지만 국지적 차원에서 면밀한 분석 및 해석이 되지 않는 한계가 있다. 하지만 Heo and Moon(2013)는 상업지역을 대상으로 블록단위에서 도시환경적 영향요인 분석을 통하여 가로폭, 건물층수, 용적율, 제2종근린생활시설, 유흥시설, 상업시설 수가 많을수록 범죄가 많이 발생하는 것으로 분석하여 차별화 된다.

범죄관련 연구 분야에서 빅데이터를 활용한 연구는 많지 않다. 그 중에 Choi and Noh(2016)는 CCTV, 보안등 수, 인구수 등 공공에서 제공하는 빅데이터를 활용하여 범죄율과의 관계를 분석하였다. 또한 시간대, 요일별, 외국인 국적 등의 일부 데이터는 세분화하여 분석함으로써 사례지역의 범죄발생 패턴을 밝혀내고 각 패턴 별로 적합한 예방책을 제시하였다. 예측에 관한 연구 방법론에서 보면, 베이저안 모델(Bayesian model)을 이용한 사례들이 보고되고 있는데, GDP, 인플레이션, 재난, 재해 등에서 전통적 시계열 분석에 비해 높은 예측 정확도를 보여 주어 활용빈도가 늘어나고 있는 추세이다(Lee, 2011). 또한 베이저안 모델은 Deublein *et al.*(2013), Qin *et al.*(2005)의 교통사고예측, Wang *et al.*(2013)의 지진재난예측, Lee *et al.*(2016)의 홍수예측 분야 등에 적용한 사례가 있다. 국내의 경우, Park *et al.*(2015)는 범죄 관련 빅데이터에 베이저안 네트워크를 적용하여 범죄예측 알고리즘을 구현한 사례도 있다. 이 연구에서 베이저안 네트워크의 노드는 공간적, 인구적, 사회적 특성, 요일, 시간 등과 같은 기타 요인으로 구성하고 있다. 베이저안 네트워크를 통하여 예측한 값과 실제 데이터를 비교한 결과 약 71%의 예측 정확도를 나타낸 것이 보고되고 있어

범죄분야에서도 적용 가능성을 시사하고 있다.

이와 같이 범죄발생 위험지역을 예측하기 위해서 다른 분야에서 활용되는 사례를 바탕으로 방법론적 탐구와 연구개발이 필요하다. 하지만 국내에서는 개인정보보호 등의 이유로 범죄발생 데이터 공개 및 활용이 어렵기 때문에 수행된 연구가 거의 전무하며, 범죄발생 데이터 확보 없이 예방적 차원의 위험지역 예측이 필요한 실정이다. 이에 본 연구에서는 기초연구로서 범죄 데이터가 없어도 도시의 물리적이며 환경적인 공간요소를 기반으로 범죄발생 위험지역 예측 모델을 개발하여 실용화할 수 있는 방안을 모색하고자 한다.

2. 이론적 고찰

베이저안 모델은 수만 건의 표본 데이터를 생성한 후, 그 중 명확한 데이터로 확률에 의해 추론하는 방식으로 객관적인 결과를 제공하므로 단순한 영역의 문제해결에는 매우 유용하다. 이에 범죄발생 위험 지역 예측을 위하여 베이저안 확률기반 추론을 시행하였다. 베이저안 확률은 새로운 사건이 발생했을 때 다음 사건이 발생할 가능성을 확률로 계산하는 것이다. 수식으로 표현하면, 새로운 사건을 A, 다음 사건을 B라고 했을 때, 확률 A인 $P(A)$ 는 사전확률(Prior probability)이라고 하고, $P(A|B)$ 는 사후 확률(Posterior probability)이라고 한다. $P(B|A)$ 는 우도(Likelihood)라고 한다. 베이저확률을 통한 확률변수 간의 관계식은 식 1과 같다.

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(A) \cdot P(B|A) + P(\bar{A}) \cdot P(B|\bar{A})} \quad (1)$$

단, $P(B) > 0$

사전확률은 관측자가 알고 있던 사건으로부터 나온 확률이며, 우도는 알고 있던 사건이 발생했다는 전제하에서 다른 사건이 발생할 확률을 의미한다. 사후확률을 직접적으로 계산할 수 없는 상황일 때 사전확률과 우도를 이용하면 사후확률을 구할 수 있다. 위 식에 기반한 베이저안

추론은 연구자가 관심을 갖는 모든 것(모수, 예측치 등)에 대한 불확실성을 확률분포로 나타낸다는 가정에서 출발한다. 본 연구에서는 베이지안 기법 중 복잡한 모형에서 빈도론적 방법보다 모수 추론이 쉬워 반복측정 및 군집 자료 분석에 유용하며, 머신러닝의 여러 분야에서 효과적인 MCMC(Markov Chain Monte Carlo) 시뮬레이션을 활용하여 분석하였다.

자료구축 및 연구방법

미국 캘리포니아 LA에서는 범죄예측 서비스인 프레드 폴(PredPol)을 적용해 범죄를 예측하고 있다. 프레드 폴은 UCLA의 몰러 산타클라라 대학교수가 개발했는데, 영화 ‘마이내리티 리포트’ 처럼 범죄자를 예측하지는 못한다. 다만 과거 범죄가 일어난 지역의 빅데이터를 이용하여 6개월 단위로 기계학습을 통해 특정 장소에서 언제 어떤 유형의 범죄가 일어날지 예측해 준다. 현재 LA뿐만 아니라 애틀랜타 등 여러 도시에 프레드 폴을 활용하고 있다. 이들 지역의 범죄 예방률은 기존 대비 15~20% 정도 향상된 것으로 나타났다. UAE의 두바이도, 범죄 예측 시스템을 도입할 전망이다. 2016년 12월 스페이스 이미징 미들 이스트(Space Imaging Middle East)는 두바이 경찰청에 범죄를 예측할 수 있는 시스템을 구축할 계획이라고 밝히고 있다(Yoo, 2017). 이처럼 사례는 적지만 빅데이터를 이용한 범죄예측 시스템 개발이 조금씩 진행되고 있다.

데이터 측면에서 보면, 범죄는 발생 지역의 사회경제적, 인구학적, 물리적 환경특성에 영향을 받기 때문에, 무작위로 발생하는 것이 아니라, 환경특성에 따라 집중과 반복되는 발생 패턴이 있다(Newburn and Sparks, 2004). 이에 범죄발생에 영향을 미치는 환경적 특성을 분석하여 특정 지역의 범죄발생 예측 정확성을 높이기 위해서는 빅데이터의 일반적인 특성을 공유하면서 위치적 속성과 장소적 특성을 가지는 다양한 공간데이터를 통한 환경특성과 공간적 분석이 선행되어야 될 것으로 판단된다.

한편 공간 빅데이터에 대한 개념 정의가 적용 분야에 따라 다소 차이가 있어 우선 본 연구에서도 연구의 범위와 성격에 맞도록 정의해 두자 한다. 사례를 보면, Kim *et al.*(2013)는 공공부문에서 축적해온 행정정보와 트위터, 인터넷 정보 등 민간부문의 다양한 종류의 빅데이터들을 공간정보와 융합한 데이터, 또는 공간정보화가 가능한 모든 데이터라고 정의하였다. Yu *et al.*(2014)는 정형·반정형·비정형 공간 빅데이터를 효율적으로 수집·저장·관리하는 동시에 공간정보와 융합된 다양한 속성정보에 대해 실시간·통합 분석을 수행하여 의미 있는 정보를 추출함으로써 미래에 대응할 수 있는 기술이라 정의하였다. Kim(2015)은 행위 주체에 대한 정보와 함께 행위가 발생한 시간과 장소에 대한 시공간적 속성을 보유하며, 단일한 개체로서 거대한 것이 아니라 미시적인 기록들의 방대한 집합체라고 정의하였다. Kim *et al.*(2016)는 위치정보와 해당 위치의 시계열 정보를 가지고 공간화 할 수 있는 모든 데이터를 빅데이터라고 하였다.

이처럼 기존 연구에서는 다소 포괄적인 개념에서 공간 빅데이터를 정의하였다. 하지만 범죄안전 분야를 다루고 있는 본 연구에서는 연구특성을 반영하여 범죄안전 분야의 공간 빅데이터를 ‘공간, 상황, 시간 등 위치정보를 가지는 방대한 데이터’라고 정의하고자 한다. 국내의 공공 및 민간에서 유·무료로 제공되는 빅데이터는 개별·집계구·동 등 다양한 공간 범위로 데이터를 구성하고 있다. 이러한 공간 빅데이터를 사례지역에서 수집하여 범죄 데이터와 연계한 예측모형을 개발하고자 한다.

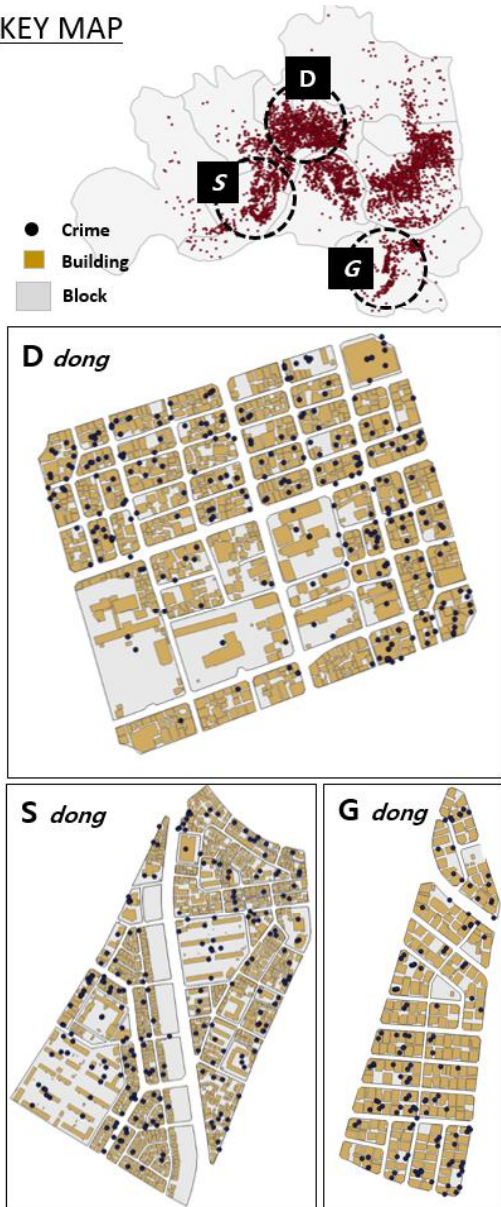
사례지역은 국내의 보편적인 지방도시 J시를 대상으로 하였으며, 범죄발생 자료는 해당 지역 경찰서의 범죄관리대장을 분석에 활용 하였다. 여기에는 범죄발생일자, 행정동, 발생장소, 범죄수법, 5대 범죄유형별 등으로 기록되어 있다. 본 연구에서는 5대 범죄(강간, 절도, 폭력, 살인, 강도)로 한정해서 공간분석기법을 적용하기 용이하게 QGIS(Quantum Geographic Information System)를 활용하여 구축했으며, 필지별 건물

속성데이터 등을 연계하여 분석하였다. 범죄발생 자료는 2008년 2,347건, 2011년 2,561건, 2012년 2,512건으로 총 7,420건이며 3개년도 자료를 대상으로 구축하였다. 지역적으로 보면 범죄로부터 가장 취약하고, 군집적으로 발생하는 지역 중 그림 1과 같이 범죄발생 건수가 많

고 다양하여 분석에 적합한 세 지역¹⁾을 선정하였다. 그 결과 D지역은 356건, S지역은 248건, G지역은 151건의 범죄가 발생한 것으로 파악되었다. 공간적으로 보면, D지역은 60개 블록과 701개 건축물, G지역은 66개 블록과 944개 건축물, S지역은 22개 블록과 225개 건축물이 존재하며, 총 148개의 블록과 1,870개의 건물로 구성되어 있다.

분석에 사용된 공간 빅데이터는 범죄발생에 영향을 미치는 물리·환경적 영향요소와 관련된 선행연구를 기반으로 빅데이터로 구축 가능한 데이터를 위주로 4개 요인(건물주용도, 1층용도, 가로폭, 건물층수)과 선행연구에서 사용되지 않은 2개요인(용적율, 건폐율)을 추가하여 사용하였다. 그 중 가로폭, 건물 층수, 용적율, 건폐율의 경우 각 건물 및 블록별 평균값을 사용하였다. 각 블록 건물들의 1층 주용도(제1종근린생활시설·제2종근린생활시설²⁾, 유흥시설, 상업시설, 업무시설, 주거시설)는 현장을 방문하여 조사하였다.

KEY MAP



**베이지안 확률기반 범죄위험지역
예측모델**

1. 베이지안 확률 기반 범죄위험지역 예측모델 개발

범죄발생 위험지역을 예측하기 위해서는 범죄발생 요인에 근거해야 되는데 범죄가 발생한 공간(장소)을 분석하면 공통적인 특징들을 확인할 수 있다. 그 중에서 중요한 특징 중 하나는 범죄에 취약한 물리적 환경이 조성되어 있다는 점이다(Choi and Kang, 2012). 따라서 베이지안 확률기반 범죄위험지역 예측모델의 정확도를 높이기 위해 범죄발생을 유발하는 영향요인을 찾아내고, 요인을 중심으로 베이지안 모델을 개발해야 한다.

본 연구에서 범죄발생에 영향을 미치는 요인은 공간회귀분석으로 탐색하였으며, GeoDa 프로그램을 활용하였다. 공간회귀분석에는 일반회귀분석모형(Ordinary Least Square, OLS), 공간회귀분석모형인 공간지체모형(Spatial Lagged Model,

FIGURE 1. Crime status in study area

SLM), 공간오차모형(Spatial Error Model, SEM)을 적용하여 분석하였다. OLS는 일반선형회귀 분석에서 많이 사용하지만 공간분석의 경우는 오차항의 상관성이 존재하지 않는 것으로 전제하고 분석하는 것이고, SLM은 공간자기상관을 가지고 있는 종속변수를 독립변수로 회귀시켜 분석한다. SEM은 공간분석에서 많이 활용되는 기법으로 OLS에 공간적 관계성을 고려하여 분석하는 특징이 있다. 여기에서 공간회귀모형의 적합도는 R^2 , 로그우도(log likelihood), AIC, SC로 검토하는데, SLM, SEM은 OLS보다 일반적으로 로그우도가 증가하고, AIC (Akaike Info Criterion)와 SC(Schwarz Criterion)는 감소하는 경향이 있다(Anselin, 1995).

분석한 각 모형들의 추정 결과값을 비교한 결과 우선, R^2 는 OLS, SLM, SEM 순서대로 커

지는 것을 확인 할 수 있었다. 다만 모형의 적합성 지수에서는 D, S, G 지역의 경우 SEM의 AIC, SC, Log likelihood 값이 작아 세 모형 가운데 현상을 가장 잘 설명하는 것으로 분석되었다(표 1 참조).

SEM 분석 결과 표 2와 같이 D지역에서 정(+)의 관계를 가지는 것은 가로폭, 건물층수, 용적율, 제2종 근린생활시설, 유흥시설, 상업시설, 주거시설로 나타났다. 이를 통해 제2종 근린생활시설, 유흥시설, 상업시설이 많을수록 범죄발생건수가 많은 것으로 확인 할 수 있고, 건물층수가 높을수록 범죄발생건수도 많은 것으로 나타났는데, 이는 바닥 면적이 넓어지면서 인구와 토지이용이 늘어나게 되는 점이 반영된 것으로 판단되며, Heo and Moon(2013) 연구의 결과와 같이 용도의 복합비율이 높아지는 점도 작

TABLE 1. Comparison of Model Fit Results

| Model | OLS | | | SLM | | | SEM | | |
|----------------|--------|--------|--------|--------|--------|---------|--------|--------|--------|
| | D | S | G | D | S | G | D | S | G |
| R2 | 0.34 | 0.51 | 0.61 | 0.53 | 0.65 | 0.70*** | 0.56 | 0.73 | 0.81 |
| Log likelihood | -66.63 | -81.63 | -61.12 | -64.17 | -78.57 | -58.82 | -63.02 | -75.54 | -54.47 |
| AIC | 146.29 | 177.28 | 136.20 | 144.38 | 173.12 | 133.68 | 140.01 | 165.11 | 122.98 |
| SC | 156.08 | 186.11 | 145.03 | 154.46 | 183.17 | 143.74 | 148.84 | 173.92 | 131.77 |

* < 0.1 ** < 0.05 *** < 0.01

TABLE 2. Spatial Error Model(SEM) Result

| Model | D | S | G |
|--------------------------------------|----------|-----------|-----------|
| Constant | -8.73* | -23.86** | -13.32*** |
| Road width | 0.02*** | 0.04*** | 0.04*** |
| Number of building floors | -0.01** | -0.01*** | -0.01*** |
| Floor area ratio | 3.72** | 7.21*** | 3.13*** |
| Building coverage | -62.72 | -129.77 | -0.01** |
| Type I neighborhood living facility | -3.14*** | -166.11 | -128.92** |
| Type II neighborhood living facility | 95.62*** | 239.69*** | 131.26*** |
| Accommodations | -2.27** | -0.01** | -0.01 |
| Entertainment facilities | 0.95* | 8.63** | 4.47* |
| Commercial facilities | 7.03** | -29.46 | -22.97 |
| Office facilities | 2.41 | -2.06 | 59.72 |
| Residential facilities | 76.88** | 132.81 | 48.13 |
| Breusch-Pagan | 10.67** | 21.14** | 13.24*** |
| Likelihood ratio | 7.25*** | 12.18*** | 17.98*** |
| LM-Lag | 0.84 | 4.55** | 7.24 |

* < 0.1 ** < 0.05 *** < 0.01

용하는 것으로 판단된다. 주택의 경우는 주거용도가 많아짐에 따라 범죄에 노출될 가능성이 상대적으로 높아진 결과라고 분석된다(Lee, 2004). 다음으로 약한 부(-)의 관계가 있는 것으로 나타난 것은 제1종 근린생활시설, 숙박시설로 분석되었다. 범죄발생이 적은 블록을 현장 조사한 결과, 부(-)의 관계인 제1종 근린생활시설에 해당되는 편의점, 경찰서, 동사무소, 소방서 등이 다수 입지한 블록이었으며, 이들 시설은 출입구 자동 통제 및 보안시스템, CCTV 등이 설치되어 있어 상대적으로 감시가 어우어져 발생 위험성이 적은 것으로 판단된다. 즉 다양한 용도들이 혼합되어 조성되어있는 블록이 사람들의 활동을 증가시켜 범죄발생 위험성이 증가하지만 오히려 서로 감시하는 기회가 늘어남에 따라 범죄발생이 적은 것으로 분석된다.

지역별로 보면, S지역은 가로폭, 용적률, 유흥시설이 범죄율과 정(+)의 관계를 나타내는데 비해 건물층수, 숙박시설은 부(-)의 관계를 가지는 것으로 나타났다. G지역은 가로폭, 용적율, 제2종근린생활시설, 유흥시설이 정(+)의 관계를 나타내고 있으며, 건물층수, 건폐율, 제1종근린생활시설이 부(-)의 관계를 가지는 것으로 나타났다. 도시공간의 물리·환경요인을 중심으로 범죄다발지역의 범죄발생 영향요인 분석한 결과 가로폭, 건물층수, 용적율, 제2종근린생활시설, 유흥시설, 상업시설, 주거시설 수가 범죄발생에 영향을 주는 요인으로 분석되었다.

이상에서 추출된 변수로 베이지안 모델을 개발하고자 한다. 이때 AMOS 프로그램을 활용하였다. 모델은 범죄발생여부를 종속변수로 하고, 앞서 공간회귀분석에서 영향 요인으로 추출된 가로폭, 평균 층수, 용적율, 제2종 근린생활시설(1층 용도), 상업시설(1층 용도), 유흥시설(1층 용도), 주거시설(1층 용도)을 독립변수로 모델을 구성하였다. 또한 관측변수들이 설명하지 못하는 부분이 있으므로 오차항을 생성하였다. 이때 각 변수들은 측정단위가 다르므로 z-score를 이용하여 투입하였다. 범죄발생 위험지역을 예측하기 위하여 작성된 연구모델(CIPM: Crime Incident Prediction Model)은 그림 2와 같다.

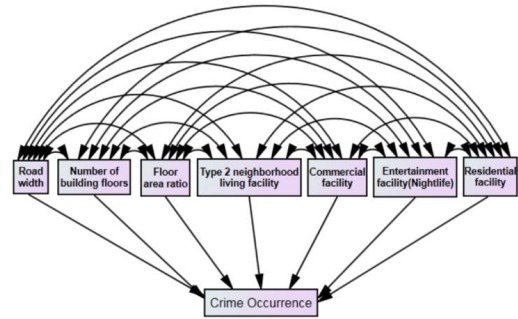


FIGURE 2. Route figure of the crime risk

베이지안 추론 시 범죄발생은 확률적 접근이므로 대상지의 2008, 2011, 2012년도 평균 범죄건수를 토대로 평균 범죄건수보다 높은 블록과 낮은 블록으로 구분하였다. 왜냐하면 계산 결과는 발생건수가 아닌 평균이하 또는 평균이상으로 어느 정도의 확률로 범죄가 발생할 것인가를 계산하게 되기 때문이다. 건수에 상관없이 범죄가 발생한 모든 블록을 범죄발생 블록으로 설정한다면 예측값이 실제 범죄 취약지점을 반영할 수 없기 때문이기도 하다. 베이지안 모형 개발의 일반적 연구흐름에 따라 사례지역 내의 일부지역(예를 들면 임의의 '가' 지역) 데이터로 모형을 구축하고, 모형의 정밀도를 검증하기 위해 사례지역 나머지 지역(예를 들면 임의의 '나' 지역)을 대상으로 모형 예측값과 실측값을 비교하여 검증하게 된다. 이를 위해 우선 '가' 지역의 데이터를 다양하게 조합하여 사전 분포정보로서 2만개의 샘플 데이터를 생성하였다. 다음으로 모형의 정확도를 높이기 위하여 10회 머신러닝을 수행한 후, 그 평균값을 활용하였다. 베이지안 모델에서 산출된 결과값을 Y라고 한다면 예측 값은 원인들에 대한 조건부 기댓값으로 계산할 수 있으며, 베이지안 모형 이론에 따라 식 2와 같이 표현된다. 모형의 적합성은 사후예측 p값(posterior predictive p value)을 활용하여 검증하는데, 본 사례의 경우 p값은 0.5로 모형이 적합한 것으로 나타났다.

$$Y_i = 0.025X_1 + 0.08X_2 + 0.052X_3 + 0.045X_4$$

$$4+0.079X_5-0.1X_6+0.089X_7+e \quad (2)$$

- Y_i : 범죄발생 위험성
- X₁ : 평균 가로폭
- X₂ : 평균 건물 층 수
- X₃ : 평균 용적률
- X₄ : 상업시설 수(1층 주용도)
- X₅ : 제2종 근린생활시설 수(1층 주용도)
- X₆ : 주거시설 수(1층 주용도)
- X₇ : 유흥시설 수(1층 주용도)
- e : 오차항

베이지안 모델을 활용하여 추론한 결과, 평균 가로폭, 평균 건물 층수, 평균 용적률, 1층 주용도가 상업시설, 제2종 근린생활시설, 유흥시설 수는 범죄발생 위험성에 양의 영향을 미쳤으며, 1층 주용도가 주거인 건물의 수는 음의 영향을 미치는 것으로 나타났다. 또한 변수 중 주거시설(1층 주용도)이 -0.1로 부의 영향을 미치므로 1층을 주거시설로 이용하는 건축물이 많을수록 범죄발생 위험성이 낮아진다. 양의 영향을 미치는 변수 중 유흥시설(1층 주용도)은 0.089로 범죄발생 위험성에 가장 큰 영향을 미치므로, 1층을 유흥시설로 사용하는 건축물이 많을수록 범죄발생 확률이 높아지는 것으로 나타났다. 평균 가로폭은 0.025로 범죄발생 위험성에 가장 적은 영향을 미치는 것으로 나타났다.

2. 범죄위험지역 예측 모델 검증

구축된 범죄위험지역 예측 모델의 유효성 및 정확도를 검증하기 위해 ROC(Receiver Operating Characteristic) 곡선을 활용하여 분석하였다. ROC 곡선은 모형의 민감도(Sensitivity) 및 특이도(Specificity)로 설명된다. 민감도란 모형을 통해 예측된 값이 긍정일 때 실제로 긍정일 확률을 의미하며, 특이도는 예측된 값이 부정일 때 실제로 부정일 확률을 의미한다.

ROC 곡선을 활용한 모델의 검증결과는 ROC 곡선(파란색 계단모양선)의 아래 면적(Area Under Curve; AUC)의 값이 지표로 사용된다. AUC는

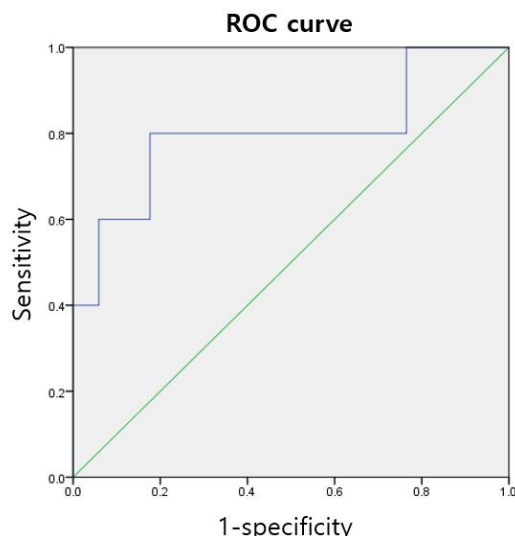


FIGURE 3. CIPM accuracy analysis results

ROC 곡선의 아래 면적을 의미하며 AUC 값이 1에 가까울수록 모형의 정확도가 높다고 할 수 있다. 보통 AUC수치에 따라 비정보적(AUC=0.5), 덜 정확한 (0.5<AUC≤0.7), 중등도의 정확한 (0.7<AUC≤0.9), 매우 정확한(0.9<AUC<1)와 완벽한 검사(AUC=1)로 분류할 수 있다 (Greiner *et al.*, 2000).

본 연구에서 구축한 모형의 AUC 값은 0.8로 나타나 베이지안 추론을 활용한 모델의 범죄발생 위험성 예측의 정확도가 80%로 높은 편이며, 베이지안 확률기반 범죄위험예측 모델의 신뢰도가 높음을 의미한다(그림 3 참조).

범죄발생 위험성 예측 결과치를 검증하기 위해 분석에 사용된 지역과 인접해 있으며 유사한 환경을 가지고 있는 S지역의 다른 블록을 대상으로 비교 분석해 보았다. 그 결과 그림 4 같이 범죄발생 위험성이 높은 지역에서 대다수의 범죄가 발생한 것을 확인할 수 있어 모형이 타당함을 알 수 있다. 중첩결과에서 범죄발생 위험성이 가장 높은 지점은 소규모 상업시설 등이 단지 내에 위치한 대규모 아파트 단지로 나타났다. 이는 1층 주용도가 주거시설이긴 하지만, 평균 건물 층수가 주변 건물들에 비해 매우 높기 때문에 주택 수와 인구가 많고, 활동량도 많

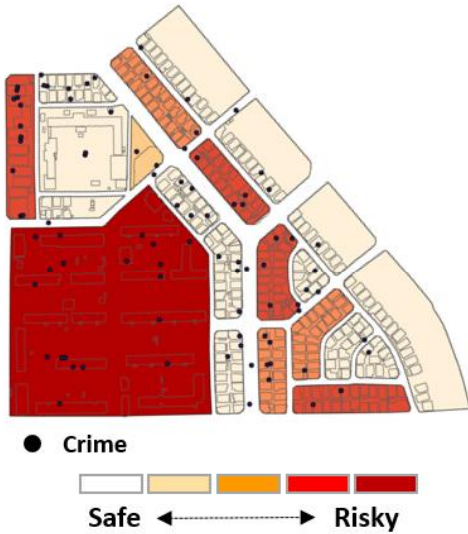


FIGURE 4. CIPM verification results

이 범죄발생위험성이 높게 나타난 것으로 분석된다. 실제로도 범죄가 가장 많이 발생한 것으로 나타났다. 또한, 상업시설이 밀집된 블록에서 범죄 위험도가 높았으며, 단독주택만 위치한 블록보다는 상업시설, 유흥시설 등과 주거시설이 혼재해 있는 블록에서 범죄발생 확률이 높은 것으로 나타났다. 이와 같은 결과는 상식적으로 보더라도 큰 차이는 없으나 구체적인 확률로 범죄발생 위험성을 계산 가능하다는 점에서 차이가 있다고 할 것이다.

베이지안 추론을 통해 대상지의 범죄발생 위험지역을 예측하여 위험도를 5단계로 구분하여 지도화한 결과는 그림 5와 같다. 주거시설이 대부분인 G지역보다 1층 주용도가 상업시설, 제2 근린시설, 유흥시설 등과 주거시설이 섞여 있는 S지역에서 범죄발생 확률이 더 높았으며, 대부분이 상업시설, 제2중 근린생활시설, 유흥시설인 D지역에서 가장 높은 것으로 나타났다. 또한 골목길보다는 대로에 접하여 있는 블록에서 범죄발생 위험성이 높았다. 실제 각 블록별 범죄발생 건수와 범죄발생 위험성을 매칭해본 결과 예측치가 낮은 대부분의 지역에서 평균보다 적은 확률로 범죄가 발생한 것을 확인할 수 있다.



FIGURE 5. Crime Incident risk prediction area mapping

결론

본 연구는 범죄발생에 영향을 미치는 도시 내 물리·환경적 공간 빅데이터를 구축하여 범죄발생 위험지역을 확률적으로 찾아내어 보다 범죄로부터 안전한 도시를 만드는 데 기여하고자 하였다. 이를 위해 먼저 공간 빅데이터의 개념을 정의하고, 공간통계분석을 이용하여 사례지역에서 범죄를 유발하는 물리·환경적 영향요인을 추출하였다. 그 결과 사례지역에서는 도시의 물

리·환경적 특성을 나타내는 요소가 범죄발생에 영향을 주고 있는 것으로 나타났다. 다음으로 이를 활용하여 베이지안 모형으로 범죄발생 위험성 예측 확률모형을 개발하여 범죄 위험도를 분석하였다. 그 결과, 범죄발생은 상업 및 유흥시설이 밀집된 지역과 건물 층수가 높은 지역, 그리고 상업 및 유흥시설과 주거가 혼재해 있는 블록에서 범죄가 높은 확률로 발생하는 것으로 나타났다. 마지막으로 본 연구에서 개발한 모델의 검증에 위해 실시한 ROC곡선 검토에서 AUC 값이 80%로 나타나 모델이 적절한 것으로 확인되었다. 본 연구의 결과는 우리가 상식적으로 생각하는 바와 크게 차이가 나지 않으나, 범죄발생 위험성을 정량적인 확률로 계산할 수 있었던 점에서 특징이 있다고 할 수 있다.

본 연구 결과는 범죄발생 위험지역에 물리적 환경요소가 영향을 주고 있으므로 이에 대한 개선이나 도시계획적 대응을 통한 범죄발생 감소와 예방이 가능함을 시사하고 있다. 현실적으로는 도시계획 관점에서 범죄발생 확률의 차이에 따라 지역의 건축물이나 공간구조를 개선하는 방법으로 범죄발생을 줄일 수 있을 것이다. 구체적으로 보면 상업시설 및 근린생활시설 등의 입지 및 규모를 선별적으로 허용하는 토지이용계획이나 지구단위계획으로 조절이 가능할 것이다. 또한 경찰 측에서는 범죄발생 위험 확률이 높은 곳을 중심으로 CCTV와 같은 감시체계를 구축하거나 경찰력을 증대하는 등의 조치로 범죄로부터 안전한 도시를 앞당길 수 있을 것이다.

본 연구는 단순히 범죄사건의 공간적 분포를 시각화하고, 범죄발생 영향요인을 탐색하는 기존의 연구들과 달리 범죄발생 예측모형을 개발하는 연구로 한 단계 발전하는 계기가 되었다는 점에서 중요한 의의가 있다. 이는 전 세계적으로 확산되고 있는 빅데이터 혁명과 같은 급변하는 시대에 학술연구가 현상을 관찰하고 설명하는 수준을 뛰어 넘어, 현실적 문제를 적극적으로 해결하는데 기여해야 된다는 공감대와 부합하는 것이라 할 수 있다. 이와 더불어 기존의 연구가 범죄의 원인을 밝혀내는데 주력했다면 본 연구는 범죄발생을 예측하여 실질적인 사회문제 해

결에 기여할 수 있다는 점에서 의의가 있다.

하지만 본 연구는 많은 한계가 있다. 우선 범죄 공간 빅데이터를 충분히 구축하기 어려워 분석의 공간범위가 다소 작다는 점이다. 이외에도 범죄에 미치는 영향도 물리적 요인 뿐 아니라 도시사회 현상에 관한 많은 요인이 분명히 있을 것이나, 이를 공간 빅데이터로 구축하지 못하여 적용하기 어려웠다는 점이다. 이러한 한계점은 개발된 모형의 정밀성과 예측 결과의 신뢰성에 대한 심각한 오류가 발생할 수 있다. 그러나 이는 국내 현실과 본 연구가 시작하는 단계에서 극복해야 할 과제이며, 추후 충분한 공간 빅데이터 구축과 정교한 예측모델의 개발로 개선이 가능할 것이다. **KAGIS**

주

- 1) 범죄발생 Data의 경우 개인재산 보호 및 사생활 보호 측면에서 정확한 지명이나 구체적인 위치, 스케일바, 방위표시 등을 표시하지 않았다.
- 2) 제1종 근린생활시설은 슈퍼, 소매점, 이·미용원, 일반 목욕탕, 주민센터, 경찰서, 소방서 등이며, 제2종 근린생활시설은 일반음식점, 학원, 부동산중개소, 극장 등이다.

REFERECNES

Anselin, L. 1995. Local indicators of spatial association-LISA. *Geographical Analysis* 27(2):286-306.

Burch, J.H. and Geraci, M.N. 2009. Data-Driven Approaches to Crime and Traffic Safety(DDACTS): operational guidelines. *Police chief*. 76(7):18-23.

Choi, M.J. and K.S. Noh. 2016. Exploratory study on crime prevention based on bigdata convergence -through case studies of Seongnam City-. *Journal of Digital Convergence* 14(11):125-133 (최민제, 노

- 규성. 2016. 빅데이터 융합 기반 범죄예방에 관한 탐색적 연구 - 성남시 사례 분석을 통해 -. 디지털융복합연구 14(11):125-133).
- Choi, Y.S. and S.J. Kang. 2012. A study on the foreign crimes in ethnic places in South Korea. Korean Institute of Criminology. p.300 (최영신, 강석진. 2012. 외국인 밀집지역의 범죄와 치안실태 연구. 형사정책연구원 연구총서. 300쪽).
- Deublein, M., M. Schubert, B.T. Adey, J. Köhler, and M.H. Faber. 2013. Prediction of road accidents: a Bayesian hierarchical approach. Accident Analysis & Prevention 51: 274-291.
- Greiner, M., D. Pfeiffer, and R.D. Smith. 2000. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Preventive veterinary medicine 45(1):23-41.
- Haider, A. and A. Ali. 2015. Socio-economic determinants of crimes: a cross-sectional study of punjab districts. International Journal of Economics and Empirical Research 3(11):550-560.
- Heo, S.Y. and T.H. Moon. 2013. Analysis of urban environmental impact factors in crime hotspot. Journal of Korea Planning Association 48(6):223-234 (허선영, 문태현. 2013. 범죄다발지역의 도시 환경적 영향요인 분석. 국토계획 48(6):223-234).
- Kim, D.H. 2015. The use of spatial big data for planning support : case of building-energy data for Seoul Metropolitan Area. The Korea Spatial Planning Review 87:163-178 (김동한. 2015. 계획지원을 위한 공간 빅데이터의 분석과 활용. 국토연구 87:163-178).
- Kim, G.H., C.M. Jun, H.C. Jung, and J.H. Yoon. 2016. Providing service model based on concept and requirements of spatial big data. Journal of the Korean Society for Geospatial Information Science 24(4):89-96 (김근한, 전철민, 정휘철, 윤정호. 2016. 공간 빅데이터의 개념 및 요구사항을 반영한 서비스 제공 방안. 한국지형공간정보학회지 24(4):89-96).
- Kim, M.J., D.J. Kim, and Y.J. Lee. 2013. Spatial big data utilization for the national land policy. Korea Research Institute for Human Settlements. pp.21-69 (김미정, 김대중, 이영주. 2013. 과학적 국토정책을 위한 공간 빅데이터 활용방안. 국토연구원. 21-69쪽).
- Lee, C.Y. 2011. Long-term crude oil price forecast using the Bayesian model. POSRI Business & Economics Research, 11(2): 58-86 (이철용. 2011. 베이지안 모형을 이용한 중장기 국제유가 전망 연구. POSRI 경영경제연구 11(2):58-86).
- Lee, S.H., J.E. Kang, and C.S. Park. 2016. Urban flood risk assessment considering climate change using Bayesian probability statistics and GIS: a case study from Seocho-Gu, Seoul. Journal of the Korean Association of Geographic Information Studies 19(4):36-51 (이상혁, 강정은, 박창석. 2016. 베이지안 확률통계와 GIS를 연계한 기후변화 도시홍수 리스크 평가: 서울시 서초구를 대상으로. 한국지리정보학회지 19(4):36-51).
- Lee, S.W. 2004. Implication of the urban plan on crime occurrence of Seoul. Seoul Development institute. p.36 (이성우. 2004. 서울시 범죄발생의 도시계획적 함의. 서울시 정책개발연구원. 36쪽).
- Newburn, T. and R. Sparks. 2004. Criminal justice and political cultures: national and international dimensions of crime control. Willan Publishing, UK. p.224.
- Park, J.H., G.H. Cha, K.H. Kim, D.C. Lee,

- K.J. Son, and J.Y. Kim. 2015. Implementation of crime prediction algorithm based on crime influential factors. *Journal of Satellite, Information and Communications* 10(2):40-45 (박지호, 차경현, 김경호, 이동창, 손기준, 김진영. 2015. 범죄발생 요인 분석 기반 범죄예측 알고리즘 구현. *한국위성정보통신학회논문지* 10(2):40-45).
- Thaler, R.T. and C.R. Sunstein. 2009. *Nudge improving decisions about health, wealth, and happiness*. Penguin Books, U.S. p.320.
- Smith, A.M. 2014. Police.uk and data.police.uk: developing open crime and justice data for the UK. *JeDEM-eJournal of eDemocracy and Open Government* 6(1): 87-96.
- Torruam, J.T. and C. Abur. 2014. The relationship between unemployment, inflation and crime: an application of cointegration and causality analysis in Nigeria. *Journal of Economics and Sustainable Development* 5(4):131-137.
- Qin, X.H., L. Liu, and Y. Zhang. 2005. A traffic accident prediction method based on Bayesian network model. *Computer Simulation* 22(11):230-232.
- Wang, J., X. Gu, and T. Huang. 2013. Using Bayesian networks in analyzing powerful earthquake disaster chains. *Natural hazards* 68(2):509-527.
- Yoo S.M. 2017. Big data crimenable possible. *The Science Times*. Available at: <http://www.sciencetimes.co.kr/?news=%EB%B9%85%EB%8D%B0%EC%9D%B4%ED%84%B0%EB%A1%9C-%EB%B2%94%EC%A3%84-%EC%98%88%EC%B8%A1-%EA%B0%80%EB%8A%A5&s=%EB%B9%85%EB%8D%B0%EC%9D%B4%ED%84%B0>(Accessed June 22, 2017) (유성민. 2017. 빅데이터로 범죄예측가능. *사이언스타임즈*).
- Yu, S.C., W.W. Choi, D.B. Shin, and J.W. Ahn. 2014. A study on concept and services framework of geo-spatial big data. *Journal of Korea Spatial Information Society* 22(6):13-21 (유선철, 최원욱, 신동빈, 안종욱. 2014. 공간 빅데이터의 개념 및 서비스 프레임워크 구상에 관한 연구. *한국공간정보학회지* 22(6):13-21). **KAGIS**