

범죄통계데이터를 활용한 카이제곱 검증 기반 범죄특징분석

임채은, 박원주*, 장광호**, 백명선*

충남대학교, *한국전자통신연구원, **경찰대학 치안정책연구소

collcr@o.cnu.ac.kr, *wjpark@etri.re.kr, **pathfinder@police.go.kr, *sabman@etri.re.kr

Analysis of Crime Characteristics Based on Chi-Square Using Crime Statistics Data

Chae-Eun Im, Wonjoo Park, Kwangho Jang*, Myung-Sun Baek**

ChungNam National Univ., *ETRI, **Police Science Institute

요약

본 논문은 경찰이 효과적으로 범죄예방과 범죄 대응을 수행할 수 있도록 범죄 발생에 영향을 미치는 요인들을 분석하고 유효성을 검증하였다. 범죄통계시스템 데이터와 범죄통계원표 데이터를 활용하여 범죄 발생에 영향을 미칠 수 있는 요인들을 추출하였고, 추출된 요인들이 지역(행정자치구)별 범죄발생량과 범죄 유형별 범죄발생량에 영향을 미치는 것을 카이제곱검증을 통해 수치적으로 증명하였다. 유효한 요인들 중 시간에 따른 행정자치구 별 범죄 발생량을 막대그래프로 시각화하여 연구 결과를 쉽게 파악할 수 있도록 하였다. 도출된 결과를 활용하면 경찰은 지역 및 범죄발생 요인에 따른 범죄특징을 참고하여 범죄예방과 사후범죄대응을 할 수 있다.

I. 서론

한국의 전체 범죄발생건수는 2008년 최고점을 기록한 이후 점차 감소하였고, 2011년 이후 증감을 반복하다가 2015년 이후부터 매년 감소하였다[1]. 이와 같이 범죄발생은 2015년을 기준으로 매년 감소하고 있지만 범죄 발생에 대한 시민들의 불안은 해소되지 않고 있다. 2018년 통계청이 발표한 '사회조사(가족, 교육, 보건, 안전, 환경)'에 따르면 국민들이 가장 많이 꼽은 사회의 가장 주된 불안요인은 '범죄발생(20.6%)'이다[2]. 범죄의 두려움은 사람들의 생활을 위축시키고 사회구성원들간에 불신감을 조장하며, 그로 인해 사회 전반에 악영향을 미치므로 범죄의 두려움을 감소시키기 위한 대책 마련이 필요하다[3]. 범죄의 두려움을 감소시키기 위해서는 경찰의 범죄예방·대응 활동이 활발하게 전개되어야 하는데[4], '빅데이터' 분석을 통해 순찰활동을 강화하고, 부족한 경찰인력 문제를 해결함으로써 범죄를 예방할 수 있다[5].

본 논문에서는 범죄통계원표 데이터와 범죄통계시스템 데이터를 활용하여 범죄발생에 영향을 미칠 것으로 예상되는 요인들을 추출하고, 추출된 요인들 중 지역별/범죄유형별 범죄발생량에 영향을 미치는 유효한 요인들을 카이제곱 검정을 통해 도출하였다. 상기 요인 중 시간 변화에 따른 지역별 범죄발생량을 그래프로 시각화하여 시간 대 별 범죄발생량을 쉽게 파악할 수 있도록 하였다. 도출된 결과에 따르면 범죄 발생에 영향을 미칠 수 있는 다양한 요인들을 쉽게 파악할 수 있으므로, 경찰이 상기 요인들을 고려하여 사전 범죄 예방 및 범죄 발생 시 효과적인 대응이 가능하도록 참고자료로 활용할 수 있다.

II. 본론

1. 데이터 전처리

본 연구에서 사용한 데이터는 경찰청에서 제공받은 2000년 0월 한달간의 범죄통계시스템과 범죄통계원표이다. 범죄통계시스템은 신고가 접수되었을 때의 범죄상황이 기록된 데이터이며 발생경찰서, 발생시간대, 발

생위치, 수사단서, 범죄수법 등 총 29개의 컬럼과 175,097개의 데이터를 포함한다. 범죄통계원표는 범죄통계시스템을 바탕으로 범죄상황을 정리한 데이터이며 발생데이터와 검거데이터로 나누어 진다. 범죄통계원표의 발생데이터는 검거수사기관, 죄명코드, 범행당시일기코드, 피해자성별, 피해자연령 등 총 71개의 컬럼과 145,196개의 데이터가 포함되고, 검거데이터는 검거지관서, 발생검거기간코드, 공범인원수, 범행도구종류 등 총 56개의 컬럼과 124,641개의 데이터를 포함한다.

데이터를 연구 분석 목적에 맞게 분석 가능한 상태로 만들기 위해 데이터 전처리를 한다. 연구 대상을 서울특별시로 한정하여, 범죄통계시스템과 범죄통계원표 데이터 중 서울에서 발생한 범죄만을 추출한다. 공통된 변수에 기반하여 추출된 범죄통계시스템과 범죄통계원표 데이터를 통합하였다. 통합한 데이터는 156개의 컬럼과 7,150개의 데이터를 포함한다. 통합한 데이터의 컬럼들 중에서 범죄특징을 분석할 수 있는 컬럼 7개를 추출한다. 추출한 컬럼들은 표 1과 같다.

표 1. 범죄 특징 분석을 위해 추출된 항목

no.	column
1	범죄대분류
2	발생위치-시군구
3	발생시간대
4	발생요일
5	피해자 연령
6	발생장소
7	범행당시일기 (날씨)

2. 카이제곱 검정기반 범죄발생 유효 요인 도출

카이제곱 검정이란 두 범주형 변수가 서로 상관이 있는지 독립인지를 판단하는 통계적 검정방법이다[6]. 카이제곱 검정통계량(X^2)은 표 2와 같이 분할표를 만들어 계산할 수 있다. 카이제곱 검정통계량을 구하는 방법은 식 1과 같다.

표 2 범죄발생지역과 날씨의 독립성 검정 분할표

	만월	맑음	비	암흑	폭풍우	흐림
강남구	3	279	45	13	1	42
강동구	0	147	34	14	0	21
...
중랑구	0	180	16	14	0	24

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

O_{ij} = i 번째 행과 j 번째 열에 있는 분할표의 관측값
 E_{ij} = i 번째 행과 j 번째 열에 있는 분할표의 기대값

카이제곱 검정을 활용하여 2가지 분석을 진행한다. 먼저 서울 행정자치구별 총 범죄 발생량과 날씨, 장소, 시간, 요일, 피해자연령 사이의 상관성 검증한다. 그리고 범죄 유형별 발생건수와 날씨, 장소, 시간, 요일, 피해자연령 사이의 상관성을 확인한다. 상관성은 카이제곱 검정을 통해 도출된 p-value 값으로 확인할 수 있다. 상기 추출된 p-value값이 0.05보다 같거나 작으면 두 범주형 변수가 서로 상관이 있다는 것을 의미한다. 아래의 표 2는 범죄통계시스템과 범죄통계원표를 사용하여 범죄 발생에 영향을 미칠 수 있는 요인들과 지역별/범죄유형별 범죄발생량 사이의 카이제곱검정결과를 보여준다. 행정자치구별 범죄 발생량의 경우 요일을 제외한 날씨, 장소, 시간, 피해자연령이 범죄발생에 영향을 줄 수 있는 요인으로 분석되었고, 범죄 유형별 발생건수의 경우 날씨, 장소, 시간, 요일, 피해자연령이 유효한 요인으로 분석되었다.

표 2 행정자치구별 범죄·범죄유형과 범죄요인의 p-value값

	행정자치구별 총 범죄 발생량	범죄 유형별 발생량
날씨	0.00000000000000631	0.00003304
장소	0.00000000000000631	0.00000000000000022
시간	0.0000000137	0.00000000000000022
요일	0.1066	0.000003365
피해자연령	0.0000005184	0.00000000000000022

그림 1은 범죄 발생시간에 따른 지역(행정자치구)별 총 범죄 발생량을 정규화 하여 시각화한 막대그래프이다. 그림에서와 같이 행정자치구에서 시간의 변화에 따라 범죄발생량이 영향을 받은 것을 확인할 수 있다. 구로구는 주로 새벽에 범죄가 많이 발생하고 중구는 오후에 범죄가 많이 발생

하는 등, 시간변화에 따른 지역별 범죄발생 특징을 파악하는 것이 가능하다.



그림 1 범죄 발생시간대별 행정자치구별 총 범죄 발생량

III. 결론

본 논문은 범죄통계시스템 데이터와 범죄통계원표를 기반으로 지역별/범죄유형별 범죄발생량에 영향을 미치는 유효한 요소들을 추출하고 카이제곱 검정을 활용하여 상관성을 검증하였다. 상관성을 가지는 요인들을 바탕으로 행정자치구별 범죄 발생건수와 범죄유형별 발생건수를 막대그래프로 시각화 하고 범죄 특징을 파악하였다. 파악한 범죄 특징은 경찰의 효과적인 출동, 효율적인 순찰 배치 등과 같은 사전범죄예방과 사후범죄대응을 위한 참고자료로 활용할 수 있다.

ACKNOWLEDGMENT

이 논문은 2020년 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임[No.2018-0-00440, 위험 상황 초기 인지를 위한 ICT 기반의 범죄 위험도 예측 및 대응 기술 개발].

참 고 문 헌

- [1] 경찰청, 경찰범죄통계, 2018, (<http://www.police.go.kr>).
- [2] 통계청, 2018년 사회조사(가족, 교육, 보건, 안전, 환경), 2018, (<http://kostat.go.kr>)
- [3] 박정선, 이성식, “범죄두려움에 관한 다수준적 접근”, 한국콘텐츠학회논문지 15권 11호 pp.74-82, 2015.
- [4] 경찰대학교, 치안정책연구소, ‘치안전망 2013 보고서’
- [5] 김형자, “빅데이터가 범죄 예방하는 세상”, TTA저널/정보통신표준화소식 173호 pp58-59., 09.2017.
- [6] 홍선경, 이상훈, 문양세, “카이제곱 검정 통계량과 유관계수의 안전한 다자간 계산”, 한국정보과학회 학술발표논문집, 12.2014.