
공공 빅데이터를 이용한 여성 대상 범죄 예방 시스템의 설계 및 구현

고성욱* · 오수빈* · 백세인* · 박혁주* · 박미화** · 이강우*

*동국대학교 컴퓨터공학과, **동국대학교 융합소프트웨어교육원

Design and Implementation of Crime Prevention System

Targeting Women by Using Public BigData

Sung-Wook Ko* · Su-Bin Oh* · Se-In Baek* · Hyeok-Ju Park* · Mee-Hwa Park** · Kang-Woo Lee*

*Department of Computer Science and Engineering, Dongguk University, Seoul, Korea

**Convergence Software Institute, Dongguk University, Seoul, Korea

E-mail : kso0611@naver.com

요 약

지역별로 여성 대상 강력범죄가 빈번하게 일어나는 구역을 표시해주는 범죄 지도를 이용한다면 경찰은 범죄가 빈번하게 발생하는 지역을 집중적으로 관리하여 범죄를 예방할 수 있고, 개인은 범죄 지역 정보를 참조하여 심각한 문제가 개인에게 발생하는 것을 피할 수 있다. 본 연구에서는 공공 데이터 포털에 공개된 여성 대상 범죄 데이터와 범죄자 데이터를 분석하여 지역별 위험도를 계산하고 개인의 연령대와 위치 등을 고려하여 맞춤형 위험 정보를 제공하는 범죄 예방 시스템을 제안한다. 성범죄가 일어난 지역에 대해서는 공공 데이터 포털에서 제공되는 범죄자 데이터를 크롤링하여 데이터를 수집한다. 성범죄가 일어난 지역에 대해서 범죄 통계 정보를 바탕으로 범죄 내용, 범죄자의 실제 거주지, 범죄 발생 지역, 형량, 범죄 횟수를 기반으로 범죄 위험도를 계산하고 해당 지역에 나타나는 위험 등급 정보를 기반으로 연령대별 범죄 발생 구역을 시각화한다. 시·군·구(동) 단위별로 계산된 지역별 위험도 점수는 GIS(Geographic Information System)와 연계하여 사용자의 위치와 연령대에 맞는 범죄 위험 정보로 제공된다.

ABSTRACT

If using crime map which represents criminal section that violent crimes targeting women frequently happened, the police could prevent additional crimes by positioning themselves intensively in expected crime zones and each individual could avoid being damaged by referring information of criminal zones. In this paper, by analyzing crimes targeting women and offender information which is provided in public-opened datum portal, we suppose a system which prevents crimes that calculates locational danger and, by considering location and age group of users, provides user-customized information of danger. By crawling the criminals datum which is provided in public-opened datum portal, It collects them. About the areas which happened sexual crimes, calculating danger of crime based on statistical crime information including criminal information, residence of offenders, areas which happened sexual crimes, sentences and the number of crime, this system is able to visualize the areas which sexual crimes happened based on information of danger grade representing on user's location. The score of danger calculated in location unit can provide criminal information according to location and ages of users by interacting GIS

키워드

빅데이터, 범죄예방시스템, 개인화 서비스, 지리정보시스템

※ 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 SW중심대학지원 사업의 연구결과로 수행되었음(R7116-16-1014)

I. 서론

여성을 대상으로 하는 성범죄는 2004년 11,105건에서 2013년 26,919건으로 10년 사이에 약 2.4배가 증가하였다[1]. 사회적으로 성범죄 문제가 심각하게 인식되어 성 범죄자에 대한 신상정보(성명, 주소, 신체정보, 사진 등)를 웹사이트를 통해 공개하는 성범죄자 신상공개제도를 시행하고 있다[4]. 신상공개는 사람들에게 성범죄의 경각심을 심어주고 성범죄자와 잠재적 범죄자의 범죄행위를 억제하여 범죄를 예방하기 위한 목적으로 활용될 수 있다.

그러나 현재 운영되는 성범죄자 알람 사이트는 모든 성범죄자의 정보를 나열하여 제공하는 형태로, 해당 사이트에서 직접 관심 지역을 텍스트로 검색해서 이용해야 하므로 성범죄에 대한 경각심을 심어주고 범죄행위를 예방한다는 성범죄자 신상공개의 목적과 비교하여 볼 때 효과가 떨어진다[3].

성범죄 관련 빅 데이터와 GIS(Geographic Information System)를 활용하여 성범죄 데이터를 분석하고 지도에 성범죄 발생 위험도를 표시한다면 사용자는 성범죄 위험지역에서 주의하며 행동하게 되어 범죄 예방에 더 큰 효과를 기대할 수 있을 것이다.

본 논문에서는 공공 데이터 포털에 공개된 성범죄의 정보를 분석하고 사용자의 위치와 연령대를 구분하여 여성 대상 범죄의 맞춤형 위험 정보를 서비스하는 시스템을 설계하고 구현한다.

II. 관련연구

2.1 GIS(Geographic Information System)

GIS는 일반적으로 지리적 정보를 수집, 저장, 분석하여 새로운 형태의 유용한 정보로 표현할 수 있는 컴퓨터 응용프로그램을 말한다[5]. 다양한 데이터와 컴퓨터 지도를 결합하여 분석하여 지형 공간의 정보를 공간 안에 배열하고, 이를 바탕으로 통계와 자료를 산출한다. 지리 정보 시스템은 다양한 분야에서 활용되고 있으며, 범죄 분석을 위한 연구에도 활발하게 진행되고 있다.

범죄 관련 데이터 분석과 GIS 기술을 활용하여 사용자의 특성에 맞는 성범죄 발생 구역을 나타낸 지도를 제공한다.

2.2 범죄 예방 시스템

범죄 예방을 위한 공공 사이트가 제공되고 있으며 범죄 예방 시스템 연구들이 진행되고 있다.

그림 1은 치안, 재난, 교통안전을 위하여 위험 발생 지역을 사용자에게 제공해주고 있는 생활안전지도 홈페이지이다[2]. 생활안전지도는 유형

별 위험 지역을 한눈에 파악할 수 있지만, 여성의 안전을 위한 성범죄 발생 지역을 파악하는데 어려움이 있으며 연령대별 위험도 정보와 같은 맞춤형 정보를 제공하지 않는다.

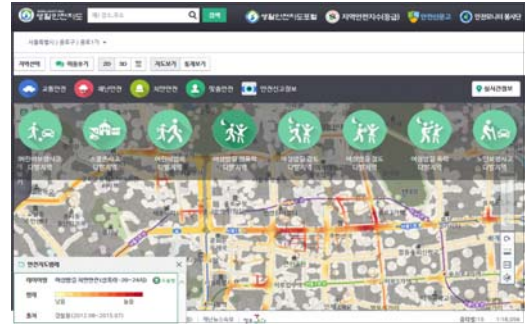


그림 1. 생활안전지도 홈페이지[2]

또한 범죄 예측을 위하여 마코프 체인을 적용한 위험도 확률 지도를 생성하는 방법을 제안한 연구가 진행되었다[6]. 이를 통하여 가로등의 설치, 치안센터 구축, 순찰차의 순찰 경로 등과 같이 범죄 예방에 도움이 될 수 있을 것이다. 공공 데이터 포털에서 제공하는 범죄자 데이터를 활용하여 여성의 안전을 위한 성범죄 발생 지도를 제공한다.

III. 범죄 발생 데이터 분석

3.1 시스템 구성도

본 논문에서 제안하는 여성범죄 예방시스템의 구성도는 그림 2와 같다. 공공 데이터 포털에서 제공되는 데이터를 크롤링(crawling)하여 자료를 수집한다. 수집된 공공 데이터는 속성 분류 및 불필요한 항목 제거(개인정보, 범죄발생시설물 등) 작업을 진행한 후, 재분류하는 전처리 과정을 거친다. 전처리를 거쳐 범죄 위험도 계산에 최종적으로 사용되는 데이터는 범죄발생년도, 범죄발생구역, 피해자 연령대, 성범죄 유형, 형벌, 형량 등이다. 범죄 지역 데이터와 범죄 유형과 연령대를 고려한 가중치 적용 및 위험도 데이터를 결합하여 지도로 표시하는 모듈은 R을 이용하여 시각화한다.

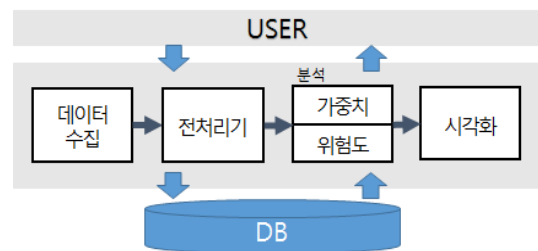


그림 2. 시스템 구성도

3.2 범죄 발생 데이터 수집

현재 2016년 09월을 기준으로 공공 데이터 포털 사이트에서 제공되고 있는 성범죄 발생 데이터의 일부는 표 1과 같다. 여성 대상 성범죄 데이터와 범죄자 데이터를 수집하였다. 수집된 성범죄자 데이터 중에서 불필요한 데이터는 제거하고, 데이터 분석을 위해 데이터의 범위를 재분류하여 데이터베이스에 저장한다. 범죄가 빈번하게 발생하는 장소(hot spot)를 연령대 별로 제공하기 위해 데이터 분석 및 재구성하여 주소 데이터를 좌표 값(위도, 경도)으로 변경한다.

표 1. 범죄 발생 데이터의 일부분

실제거주지	범죄발생 구역	피해자 연령대	형별	형량 (개월)	범죄 횟수
서대문구 연세로 54길 14	서대문구	10	미성년자에 대한 간음, 추행	36	2
성동구 성수일로 52	성동구	20	주거침입(강간, 강제추행)	30	1
양천구 자양로 2길 25	강동구	30	강제추행	12	1
성북구 삼안로 16길 28	경기중부	40	미성년자에 대한 간음, 추행	36	4
중랑구 사기정로 31길 8-3	중랑구	50	강제추행	18	3

3.3 위험도 평가

범죄 위험 지도를 생성하기 위해 각 위험 요소별 가중치를 계산하고 범죄 발생 위험도를 계산한다. 공공 데이터 포털에서 공개된 성범죄 데이터를 분석하여 지역별 범죄 가중치, 성범죄 유형별 가중치, 피해자 연령대별 가중치, 형량에 따른 가중치로 총 4가지의 가중치를 계산한다. 서울과 서울 인근 범죄 발생 데이터를 대상으로 가중치와 위험도를 계산하였다.

가. 위험 요소별 가중치 계산

1) 지역별 범죄 가중치 (LCW, Locational Crimes Weights)

지역별 범죄 가중치는 성범죄자의 실제 거주지와 범죄 발생 구역이 일치할 경우 계산하는 가중치를 의미하고, 식 (1)과 같이 계산한다.

서울 지역을 대상으로 한 범죄 발생 전체 건수에서 실제 거주지의 지역과 범죄 발생 지역이 일치 하는 건수의 비율은 34%이다. 범죄 발생 지역에 해당하는 건에 1.34의 가중치를 부여한다.

$$LCW = \frac{L}{C} + 1 \quad (1)$$

- L : 성범죄자의 실제 거주지와 범죄 발생 지역이 일치하는 건수
- C : 전체 범죄 발생 건수

2) 성범죄 유형별 가중치 (SCW, Sexual Crimes Type Weights)

성범죄 유형별 가중치는 해당 지역에서 발생한 성범죄 유형별 전체 건수에서 성범죄 유형에 따른 건수의 비율을 계산하여 해당 성범죄 유형 별로 가중치를 부여한다. 성범죄 유형별 가중치는 식 (2)와 같이 계산한다.

$$SCW = \frac{T}{B} \quad (2)$$

- T : 성범죄 유형별 건수

3) 피해자 연령대별 가중치 (AW, Ages Weights)

피해자 연령대별 가중치는 전체 발생된 범죄의 횟수에서 해당 피해자의 연령대별 범죄 발생 건수의 비율을 계산하고 연령대에 따른 가중치를 부여한다. 피해자 연령대별 가중치는 식 (3)과 같이 계산한다. 피해자 연령대는 10대, 20대, 30대, 40대, 50대로 구분한다. 10대의 범위는 10세 이하의 아동을 포함하고, 50대의 범위는 60대 이상까지 포함한다.

$$AW = \frac{A}{C} + 1 \quad (3)$$

- A : 피해자 연령대별 범죄 발생 건수

4) 형량 가중치 (SW, Sentence Weights)

형량에 따른 가중치는 수집 데이터에서 최소 형량과 최대 형량의 값을 계산하여 0~1 사이 값으로 정규화하고 식 (4)와 같이 가중치를 부여한다. 서울 지역을 대상으로 한 데이터를 분석한 결과 최대 형량은 10년(120개월), 최소 형량은 4개월로 형량 가중치는 0.03 ~ 1의 값으로 계산할 수 있다.

$$SW = \frac{S}{M} \quad (4)$$

- S : 해당 범죄 건의 형량(개월)
- M : 최대 형량(개월)

나. 범죄 발생 위험도 계산

1) 범죄 발생 위험도 (CDR, Crimes Danger Risk)

범죄 발생 위험도는 앞서 계산한 위험 요소별 가중치를 이용하여 식 (5)와 같이 계산한다. 가중치와 위험도는 상황에 따라 요소 값을 변경할 수 있다.

$$CDR = LCW \times SCW \times AW \times SW \quad (5)$$

- LCW : 지역별 범죄 가중치 - 식(1)
- SCW : 성범죄 유형별 가중치 - 식(2)
- AW : 피해자 연령대별 가중치 - 식(3)
- SW : 형량 가중치 - 식(4)

IV. 구현

4.1 개발 환경

본 논문에서 제안한 시스템을 구현하기 위한 개발 환경은 표 2와 같다.

표 2. 시스템 구축 환경

H/W	Processor	Intel(R) Core(TM) i5-2500 @ 3.30GHz (4CPUs), ~3.3GHz
	Memory	8GB RAM
S/W	OS	Windows 10 Pro 64bit
	R	R-3.3.1 64bit
	R library	ggplot2, ggmap

4.2 구현 결과

성범죄가 발생한 지역에 따른 위험 요소별 가중치와 위험도를 계산하여 사용자의 위치와 연령대에 맞는 범죄 발생 지도를 구현하였다.

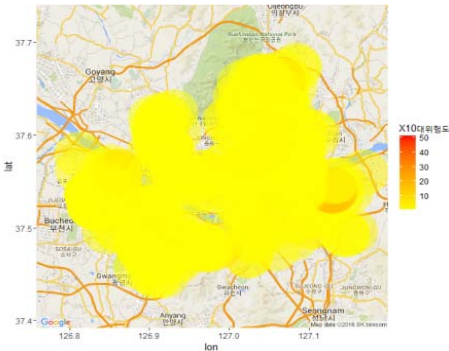


그림 3. 10대 맞춤 성범죄 위험 지도

서울과 서울인근 지역을 대상으로 10대와 50대의 연령대에 맞춰진 성범죄 위험 지도를 시각화 하였다. 그림 3은 10대 맞춤 성범죄 위험 지도의 결과이며, 그림 4는 50대 맞춤 성범죄 위험 지도의 결과이다. 각 연령대에 해당되는 성범죄 유형, 발생 위치, 건수는 다르기 때문에 연령대에 따라 성범죄 위험 지도가 다르게 보이는 것으로 확인할 수 있다.

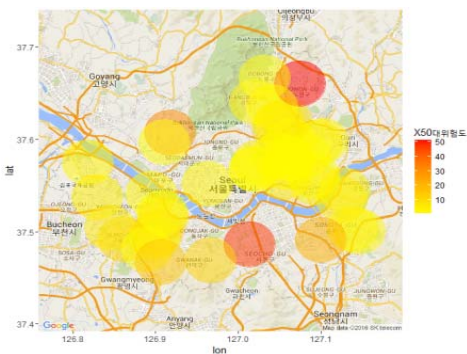


그림 4. 50대 맞춤 성범죄 위험 지도

서울과 서울인근이 지역이 아닌 특정 지역에 대한 전체 연령대에 맞춰진 성범죄 위험 지도를 시각화 하였다. 그림 5는 아동, 10대, 20대, 30대, 40대, 50대 맞춤 성범죄 위험 지도의 결과 이다.

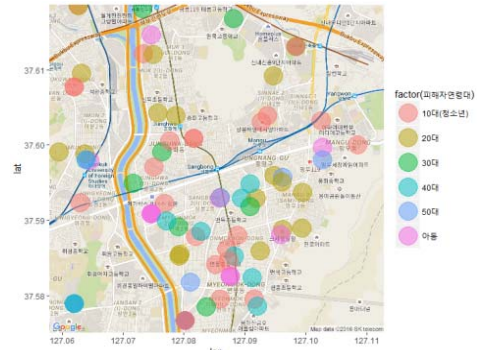


그림 5. 전체 연령대의 중량구 성범죄 위험 지도

V. 결론

본 연구에서는 공공 데이터 포털에 공개된 여성 대상 성범죄 데이터와 범죄자 데이터를 분석하여 위험 요소별 위험도를 계산하고 개인의 연령대와 위치를 고려하여 맞춤형 위험 정보를 제공하는 범죄 예방 시스템을 제안하였다.

성범죄가 일어난 지역에 대해서는 공공 데이터 포털에서 제공되는 범죄자 데이터를 크롤링하여 데이터를 수집하여 분석하였다. 성범죄가 일어난 지역에 대해서 지역별 범죄 가중치, 성범죄 유형별 가중치, 피해자 연령대별 가중치, 형량에 따른 가중치와 범죄 발생 위험도를 계산하였다. 이를 기반으로 연령대별 성범죄 발생 위험 지도를 시각화 하였다.

제안한 방법을 통하여 연령대별 사용자에게 맞춰진 성범죄 위험 지도를 제공할 수 있으며, 성범죄 예방에 도움이 될 것으로 기대한다. 또한 더 많은 데이터를 활용한다면 보다 정확한 시스템을 구축하는데 도움이 될 것으로 보인다.

참고문헌

- [1] 국가통계포털, www.kosis.kr
- [2] 생활안전지도, www.safemap.go.kr
- [3] 성범죄자 알림e, www.sexoffender.go.kr
- [4] 심희기, “신상공개와 정당화근거와 적절한 공개대상과 공개기 준의 탐색”, 저스티스, 통권 65호, p.24-38, 2002.
- [5] 이현희, “GIS(지리정보시스템)를 이용한 범죄지역 분석”, 한국사회학회 사회학대회 논문집, pp.327-331, 2000.
- [6] 노찬숙, 김동현, “마코프 체인 기반의 범죄 발생 위험도 확률지도 생성 모델”, 한국정보기술학회논문지, 제10권, 제10호, pp.89-98, 2012.