

빅 데이터를 이용한 범죄패턴 분석 알고리즘의 구현

차경현*, 김경호*, 황유민*, 이동창**, 김상지**, 김진영*

Implementation of Crime Pattern Analysis Algorithm using Big Data

Gyeong Hyeon Cha*, Kyung Ho Kim*, Yu Min Hwang*, Dong Chang Lee**, Sang Ji Kim**, and Jin Young Kim*

요 약

본 논문에서는 빅 데이터를 이용하여 범죄 발생 패턴을 분석하는 알고리즘을 제안하고 구현했다. 제안된 알고리즘은 대검찰청에서 수집하여 공개한 범죄관련 빅 데이터를 사용하며, 표준편차 타원체 및 공간밀도 분석과 같은 공간통계분석을 통해 서울시의 2011-2013년 범죄발생 패턴을 분석했다. 범죄 발생 빈도수를 이용하여 범죄발생지역, 시간, 요일, 장소의 위험지수를 구했고, 범죄 패턴 분석 알고리즘을 통해 범죄 발생 확률을 구했다. 이를 통해 공간통계분석을 했다. 제안된 알고리즘의 구현 결과, 서울시의 각 구별로 범죄발생 패턴이 다르다는 것을 파악할 수 있었고, 다양한 범죄발생 패턴을 분석하고 범죄발생확률을 위험지수를 통해 수치화하여 위험도를 정량적으로 산출할 수 있었다.

Key Words : Big Data, Crime Pattern, Nearest Neighbor Index Analysis, Nearest neighbor distance Method, Density Analysis

ABSTRACT

In this paper, we proposed and implemented a crime pattern analysis algorithm using big data. The proposed algorithm uses crime-related big data collected and published in the supreme prosecutors' office. The algorithm analyzed crime patterns in Seoul city from 2011 to 2013 using the spatial statistics analysis like the standard deviational ellipse and spatial density analysis. Using crime frequency, We calculated the crime probability and danger factors of crime areas, time, date, and places. Through a result we analyzed spatial statistics. As the result of the proposed algorithm, we could grasp differences in crime patterns of Seoul city, and we calculated degree of risk through analysis of crime pattern and danger factor.

I. 서 론

최근 빅 데이터 기술의 발전으로 불과 몇 년까지만 해도 상상할 수 없었던 일들이 가능해지고 있다. 병원 내 네트워크에서 수집된 빅 데이터를 이용하여 환자 병명 진단 및 환자관리 등에 활용하고, 빅 데이터를 기반으로 하여 유권자 맞춤형 선거 전략으로 미국 대통령 선거에서 승리한 오바마 등이 빅 데이터 기술을 활용한 대표적인 사례이다.

이러한 빅 데이터를 이용한 사례 중 범죄를 예측하고 미리 예방하는 기술이 주목받고 있다. 미국 캘리포니아 주 Santa Clara 대학교의 수학과 교수인 G. O. Mohler는 지진·여진 예측 모델인 ETAS 모델을 이용하여 범죄예측 알고리즘을 연구하고 개발했다. ETAS 모델은 여진은 최초 지진의 진원 근처에서 동 시간대에 발생한다는 것에 근거하여 여진

발생을 예측하는 모델로 이는 범죄 예측에도 활용할 수 있다 [1]. 이 범죄예측 알고리즘을 통해 미국 LA 경찰청은 절도 발생률을 33% 줄였고, 폭력, 재산관련 범죄 역시 감소하는 성과를 올렸다. 또한 2013년 미국 보스턴 마라톤에서는 두 개의 폭탄이 결승지점에서 폭발하면서 3명의 사상자와 260 명의 부상자가 발생한 테러 사건이 일어났다. 당시 미국은 범인 검거를 위해 현장에 있는 CCTV와 대회 참가자들이 찍은 영상과 같은 비정형 데이터를 수집하고 10테라바이트나 되는 방대한 양의 빅 데이터를 분석하여 테러 용의자의 인상착의를 파악하고 검거 할 수 있었다.

위와 같은 사례들을 통해 빅 데이터 기술을 공공 안전을 위해 범죄 예측이나 범인 검거 등에 활용할 수 있다는 것을 알 수 있다. 현재 해외뿐만 아니라 국내에서도 사회안전망 구축을 위해 빅 데이터 기술을 활용하는 방안에 대해 끊임없

* 이 논문은 2014년 미래창조과학부의 재원으로 SW융합기술고도화 사업의 지원을 받아 수행된 연구임(S1004-14-1010).

*광운대학교 전자융합공학과 유비쿼터스 통신 연구실 (chagyeonghyeon@kw.ac.kr, gentle@kw.ac.kr, yumin@kw.ac.kr, jinyoung@kw.ac.kr)

** (주)위니텍 (goldie64@naver.com, chocoajong@gmail.com)

접수일자 : 2014년 10월 21일, 수정완료일자 : 2014년 11월 04일, 최종게재확정일자 : 2014년 11월 13일

이 논의되고 연구 중에 있다.

빅 데이터의 기본적인 목적은 다양한 데이터의 수집과 처리를 통해 그 목적에 맞는 데이터를 분석하고, 저장된 데이터를 통해 쉽게 예측할 수 없는 정보의 흐름을 읽는데 그 목적이 있다. 빅 데이터 분석을 위한 기법에는 자연어 처리 기술 기반 Text Mining, 평판 분석 Opinion Mining, 네트워크 상에서 허브 역할 사용자를 찾는 Social Network Analytics, 유사 특성의 군집을 찾는 Cluster Analysis 기법 등이 있다.

이에 따라 본 논문에서는 관련된 선행연구를 바탕으로 대검찰청에서 수집하여 공공데이터 포털 사이트에 공개한 범죄 관련 빅 데이터를 이용하여 데이터 간의 유사 특성을 찾기 위해 범죄발생 패턴을 분석하는 알고리즘을 구현했다. 2장에서는 본 논문에서 제안하는 알고리즘에서 사용하기 위해 수집된 범죄 관련 빅 데이터에 대해 설명하고, 3장에서는 위험지수와 수집된 범죄 관련 빅 데이터를 이용한 범죄패턴 분석 알고리즘을 제안하였으며, 4장에서는 공간통계분석과 제안한 알고리즘의 구현 결과, 5장에서는 결론으로 끝을 맺는다.

II. 범죄 관련 빅 데이터의 정의

본 논문에서는 대검찰청에서 2011년부터 2013년까지 3년간 수집한 범죄관련 빅 데이터를 이용하여 범죄패턴을 분석했다. 범죄 관련 빅 데이터는 지역, 시간, 요일, 장소 별로 범죄발생 빈도수가 포함되어 있으며, 본 논문에서는 서울특별시를 분석 지역으로 선정해 범죄패턴을 분석했다. 전체 데이터 구성은 표 1과 같다.

범죄발생지역 빅 데이터는 서울특별시 각 구별로 절도, 살인, 강도, 방화, 강간, 폭행, 상해, 공갈, 약취와 유인, 체포와 감금, 폭력행위 등 처벌에 관한 법률 위반, 도박과 복표, 과실치사상, 업무상 과실치사상 등 위 상기 범죄 종류에 따라 범죄 빈도수로 분류 되어 있다. 범죄발생지역에서는 빈도수가 2012년이 171,796건으로 가장 높았다. 표 2는 5대 범죄(살인, 강도, 강간, 절도, 폭행) 발생 빈도수가 높은 5개의 구의 데이터를 나타낸다.

표 1. 범죄 관련 빅 데이터의 전체 구성.

데이터 제공	대검찰청
기간	2011 ~ 2013
수집지역	서울특별시

(단위 : 건)

	2011	2012	2013
범죄발생지역	161,097	171,796	169,473
범죄발생시간	811,116	851,722	842,504
범죄발생요일	811,116	851,722	842,504
범죄발생장소	811,116	851,722	842,504

표 2. 강남구, 5대 범죄 발생 빈도수.

(단위 : 건)

	살인	강도	강간	절도	폭행
강남구	44	158	1,738	13,323	6,736
송파구	23	91	931	11,933	4,993
관악구	37	137	1,224	9,727	6,307
서대문구	36	83	1,134	10,019	4,661
영등포구	51	102	934	8,950	5,214

범죄발생시간 빅 데이터는 범죄별로 범죄 발생 빈도수를 새벽(04:00~06:59), 아침(07:00~08:59), 오전(09:00~11:59), 오후(12:00~17:59), 저녁(18:00~19:59), 밤(20:00~03:59) 시간 대 별로 분류 해놓은 빅 데이터이다.

범죄관련 빅 데이터는 절도 범죄가 가장 빈도수가 많았고, 토요일 밤 노상에서 범죄 발생 빈도가 가장 높게 나타났다.

III. 범죄패턴 분석 알고리즘

3.1 Cluster Analysis(CA)

본 논문에서 범죄관련 빅 데이터 분석을 위해 사용 된 기법은 군집분석(Cluster Analysis, CA)이다. 군집분석은 개인 또는 여러 개체를 유사한 속성을 지닌 대상들끼리 그룹화 하는 기법이다. 군집분석은 크게 계층적 군집분석과 비 계층적 군집분석으로 분류할 수 있다. 계층적 군집분석은 대상 간의 거리에 의하여 가장 가까이 있는 대상들로부터 시작하여 결합해 가는 방법으로써 나무모양의 계층적 구조를 형성해 나가는 방법이다. 계층적 군집분석(Hierarchical Clustering)은 군집의 형성되는 과정을 정확하게 파악할 수 있다는 장점이 있지만 자료의 크기가 크면 분석하기 어렵다는 단점이 있다. 비 계층적 군집분석은 수가 정해진 군집의 중심에서 가장 가까운 개체를 선택해 포함시키는 방법으로 큰 자료도 빠르고 쉽게 분류할 수 있지만 군집의 수를 미리 정해 주어야 하고, 군집을 형성하는 중심에 따라 군집의 결과가 달라진다는 어려움이 있다. 따라서 보통 계층적 군집분석을 통해 대략적인 군집의 수를 파악하고, 이를 이용하여 비 계층적 군집분석을 시행한다.

3.2 위험지수

본 논문에서는 데이터들의 패턴을 분석하기 위하여 우선 수집된 범죄 데이터들의 위험지수를 계산한다. 위험지수는 범죄발생횟수의 상대적 발생 빈도를 0~1 사이 값으로 처리한 것으로 발생 빈도를 단계별로 도식화 할 수 있다[2][3]. 위험지수를 구하는 방법은 다음 식 1과 같이 정의하였다.

$$\text{위험지수} = \frac{\text{범죄발생빈도수}}{(\text{최대값} - \text{최소값})/0.6} \quad (1)$$

범죄데이터는 각 년도 별로 분류되어 있으므로 범죄 종류에 따라 위험지수를 구한 후 년도 구분 없이 위험지수 값을 구해주면 된다. 아래 표 3~6에는 지역, 장소, 시간, 요일 별로 5대 범죄에 따른 위험지수를 나타낸다.

3.3 공간통계분석

본 논문에서는 수집된 범죄데이터의 전체적인 패턴을 알기 위하여 공간통계분석을 실시했다. 공간통계분석(Spatial statistics analysis)에는 일반적으로 활용되는 표준편차 타원체(Standard deviational ellipse)와 밀도 분석(Density analysis)을 이용하였다[4]. 표준편차 타원체는 점 사상의 산술적 평균과 표준거리를 이용하여 타원체를 도식화하는 것으로 사건의 분포 경향, 점사상의 방향성을 직관적으로 파악할 수 있게 해준다. 밀도 분석은 연구지역을 동일한 크기의 격자(Grid)로 나눈 후 격자에 집계되는 점 빈도수를 2차원으로 표현하는 방법이다[5]. 이를 통해 연구지역 내 사건들의 공간적 밀도를 분석하고 알고리즘의 모델링이 가능하며 공간통계분석 중 최근린지수(Nearest neighbor index, NNI) 분석법을 이용해 범죄패턴을 분석했다. NNI 분석은 점들의 공간적 분포패턴이 상호작용에 의해 형성된 것인지 혹은 임의적으로 형성된 것인지를 밝히는 방법이다. 임의의 한 점으로부터 다른 점들까지의 거리를 측정한 후, 각 지점에서 가장 가까운 다른 점까지의 거리를 평균하여 관측된 평균 최근린거리(Nearest neighbor distance, NND)를 정한다. 점 분포패턴으로부터 기대되는 평균 NND를 구한 후, 기대되는 NND에 대한 관측된 평균 NND의 비율로 NNI를 도출하게 되는데, 이 지수를 통해 점 분포 패턴에서 각 점들의 간격이 임의적인 점 분포 패턴에서의 간격에 비해 어느 정도 차이가 발생하는지를 알 수 있다[6]. NNI가 1인 경우는 완전히 임의적인 분포유형을 의미하며 NNI가 1보다 큰 경우에는 분산 분포를 나타내고, NNI가 1보다 작은 경우는 점들이 모여 있는 군집적 분포를 의미한다[7].

표 3. 지역구별 5대 범죄 위험지수.

	살인	강도	강간	절도	폭행
용산구	0.33	0.53	0.25	0.31	0.45
도봉구	0.3	0.26	0.12	0.31	0.4
노원구	0.24	0.28	0.27	0.42	0.6
강서구	0.67	0.52	0.31	0.54	0.7
영등포구	0.73	0.52	0.39	0.59	0.74

표 4. 장소별 5대 범죄 위험지수.

	살인	강도	강간	절도	폭행
아파트	0.55	0.16	0.39	0.23	0.1
단독주택	0.6	0.22	0.44	0.3	0.1
노상	0.51	0.6	0.6	0.6	0.6
사무실	0.09	0.05	0.09	0.07	0.04
공장	0.02	0	0.01	0.03	0.01

표 5. 시간대별 5대 범죄 위험지수.

	살인	강도	강간	절도	폭행
새벽	0.13	0.25	0.19	0.22	0.1
아침	0.07	0.05	0.12	0.18	0.04
오전	0.17	0.1	0.12	0.2	0.07
오후	0.35	0.22	0.31	0.2	0.18
저녁	0.16	0.08	0.14	0.2	0.09
밤	0.67	0.65	0.72	0.21	0.64
미상	0.26	0.32	0.47	0.23	0.14

표 6. 요일별 5대 범죄 위험지수.

	살인	강도	강간	절도	폭행
일	0.29	0.43	0.31	0.52	0.56
월	0.22	0.24	0.24	0.55	0.46
화	0.74	0.51	0.58	0.55	0.52
수	0.35	0.31	0.26	0.56	0.51
목	0.52	0.4	0.45	0.56	0.51
금	0.31	0.29	0.26	0.59	0.54
토	0.82	0.84	0.84	0.6	0.6

3.4 범죄패턴 분석 알고리즘

본 논문에서는 주어진 범죄데이터들의 분석을 통해 데이터를 트리 형태로 패턴을 형성 하는데 그 목적이 있다. 데이터들끼리 패턴을 형성하여 데이터 간의 유사성을 비교하고, 패턴이 형성된 데이터들의 의미를 파악한다. 표 7은 범죄패턴 분석 알고리즘에 사용된 매개변수에 대해 나타낸다.

범죄 데이터 간 패턴 생성을 위해 각 범죄 종류에 따라 범죄발생지역 위험지수 R_n 에서부터 범죄발생시간 위험지수 T_n 중 높은 T_n 값을 선택하고, 그 다음 범죄발생요일 위험지수 D_n 에서 높은 D_n 값을 선택한다. 마지막으로 범죄발생장소 위험지수 S_n 에서 높은 S_n 값을 선택해 패턴을 생성한다. $R_n - T_n - D_n - S_n$ 와 같이 생성된 패턴은 선택된 위험지수 값들은 서로 합산해 식 2와 같이 범죄발생지수 P_{RTDS} 를 산출한다.

$$P_{RTDS} = R_n + T_n + D_n + S_n. \quad (2)$$

표 7. 매개변수 정의.

매개변수	정의
R_n	범죄발생지역 위험지수
T_n	범죄발생시간 위험지수
D_n	범죄발생요일 위험지수
S_n	범죄발생장소 위험지수
P_{RTDS}	범죄발생지수

한번 생성된 패턴은 다시 생성되지 않고, 다른 경우의 수로 패턴을 생성한다. 범죄발생지수는 소수점 둘째자리까지

단계별로 분류된다(범위:0-2.5). 예를 들어 강남구(R_{23})에서 토요일(D_7) 밤(T_6) 노상(S_4)에서 절도 범죄로 패턴이 형성된다면 $R_{23}-T_6-D_7-S_4$ 와 같은 패턴이 형성되고 그때 범죄 발생수치는 다음 식 3과 같다.

$$P_{RTDS} = 0.88 + 0.21 + 0.6 + 0.6 = 2.29. \quad (3)$$

또 만약 강남구(R_{23})에서 화요일(D_3) 오전(T_3) 지하철(S_{14})에서 절도 범죄로 패턴이 형성된다면 그때 패턴은 $R_{23}-T_3-D_3-S_{14}$ 이며, 그때 범죄발생지수는 다음 식 4와 같다.

$$P_{RTDS} = 0.88 + 0.2 + 0.55 + 0.01 = 1.64. \quad (4)$$

범죄발생지수는 범죄발생빈도를 비교할 수 있는 척도로서 그 값이 1.8 이상이 되면 범죄발생빈도가 높은 것으로서 범죄가 일어날 가능성이 크다는 것을 의미한다. 데이터 간 패턴 생성 시 범죄발생지수가 1.5 이하인 값들은 범죄 발생 가능성이 적다고 간주해 패턴에서 제외한다. 구해진 범죄발생지수는 단계별로 지도에 원으로 맵핑(Mapping)이 되고, 공간통계 분석에서 밀도 분석에 이용된다. 범죄패턴 생성 알고리즘은 그림 1과 같다.

그림 1에서 보듯이 각 범죄데이터의 위험지수를 위에서 언급한 방법을 통해 각 데이터별로 위험지수를 구한다. 위험지수 산출 후에 데이터 간 패턴 생성을 위해 범죄발생지역 위험지수 R_n 에서부터 T_n, D_n, S_n 순으로 각 데이터 별로 높은 위험지수 값을 선택한다. 선택에 따라 범죄데이터들 간의 패턴이 형성된다. 패턴 생성 시 기존에 생성된 패턴 여부를 확인해, 중복된 패턴이 생성되지 않도록 하고, 패턴이 생성된 후에는 각 데이터별 위험지수를 합산해 범죄발생지수를 산출한다. 산출된 범죄발생지수를 이용해 그 값에 따라 단계별로 위험도를 분류한다.

IV. 알고리즘 구현 결과

4.1 범죄관련 빅 데이터의 공간적 분포 패턴 결과

범죄데이터 중 표본수가 비교적 많은 5대 범죄(절도, 살인, 강간, 폭행, 강도)를 공간통계분석을 통해 분석한 결과 서울 특별시 각 구별의 경우 그림 2와 같이 나타났다. 범죄 발생 중심점은 강남구, 송파구, 종로구 중심으로 3개가 나타났으며, 각 타원체마다 방향성을 띄고 있었다. 총 범죄의 경우에는 강남구, 송파구, 서초구 일대가 가장 범죄 밀도가 높았고, 그 다음으로 영등포구, 동작구, 양천구 일대가 높게 나타났고, 종로구, 성북구 순으로 범죄 밀도가 높았다. 그리고 폭행, 절도 범죄는 노상에서 많은 패턴 분포가 일어났고, 살인 범죄는 단독주택, 도박 범죄는 사무실에 특히 강하게 분포되는 것을 확인할 수 있었다. 그림 2의 범죄 밀도 분석을 보다 수

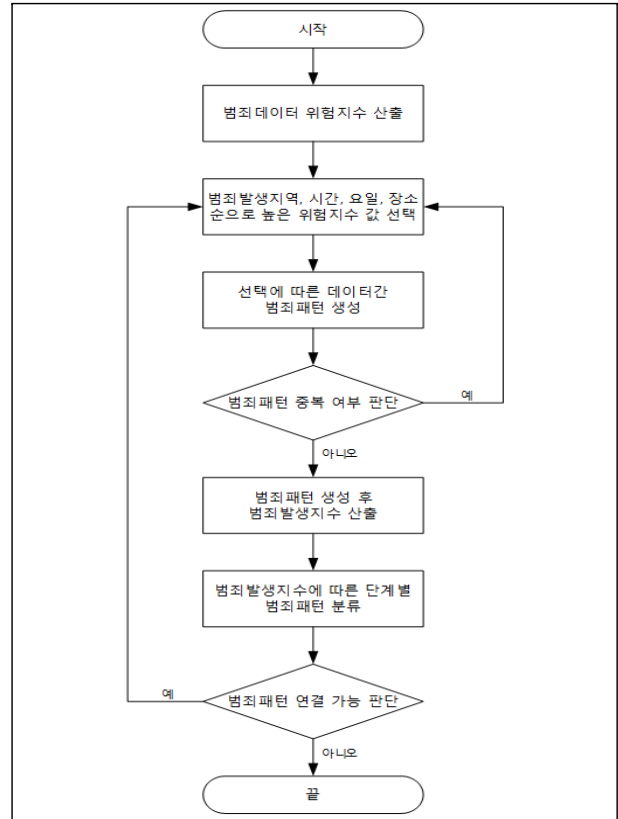


그림 1. 범죄패턴 생성 알고리즘.

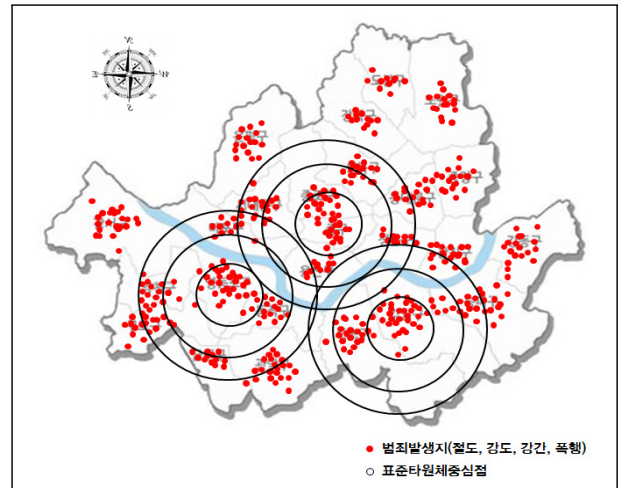


그림 2. 범죄발생지 및 타원체 중심점.

치화된 정보로 공간적 분포 패턴을 확인하기 위해 NNI 분석법을 이용해 분석하였다. 앞서 분석한 5대 범죄를 NNI 분석한 결과는 표 8에 제시하였다.

표 8. 5대 범죄유형별 최근린지수.

범죄 종류	최근린지수
살인	1.248
절도	0.296
강간	0.403
폭행	0.301
강도	0.315

NNI는 기대되는 NND d_k 에 대해 구해진 평균 NND d_m 의 비율로 구하고, 이 값은 분포패턴에서 각 점들의 간격이 임의적인 점분포 패턴에서의 간격에 비해 어느 정도 차이가 있는지를 나타낸다. 여기서 d_i 는 i 지점에서 가장 가까운 거리를 의미하며, d_m 은 한 지점에서 가장 가까운 다른 지점까지의 평균을 의미한다. 그리고 A 는 대상지역의 면적이다. 이를 정리하면 아래 식 5, 6, 7로 나타낼 수 있다.

$$d_m = \frac{1}{n} \sum_{i=1}^n d_i. \quad (5)$$

$$d_k = 0.5 \sqrt{\frac{A}{n}}. \quad (6)$$

$$NNI = \frac{d_m}{d_k}. \quad (7)$$

NNI 분석 결과 살인을 제외한 범죄들은 서로 인접하여 발생하는 군집하는 성향을 보였으며, 절도와 폭행은 낮은 NNI 값으로 보아 강한 군집 패턴을 보이는 것을 확인할 수 있었다. 그리고 살인범죄는 다른 범죄들에 비해 빈도수가 적고, 분산된 형태를 보이는 것을 확인했다. NNI가 1이 넘는 것으로 보아 임의적인 범죄 분포 패턴을 보이는 것을 확인할 수 있었다.

4.2 범죄관련 빅 데이터 패턴 알고리즘 분석 결과

두 번째 실험에서는 범죄관련 빅 데이터 패턴 알고리즘의 결과 값을 분석하였다. 표 9는 토요일 밤 노상에서 절도가 일어날 범죄발생지수를 5개의 지역구별로 나타낸 것이다.

분석 결과 강남구가 2.3으로 범죄발생지수가 가장 높았고, 다음 송파구, 은평구 순이었다. 범죄발생지수가 1.8 이상이면 범죄발생빈도가 높은 것으로서 범죄에 대한 대비와 예방이 필요하다는 것을 의미한다. 반면에 도봉구와 금천구는 발생지수가 1.8 이하이므로 그에 비해 범죄발생빈도가 다른 구에 비해 토요일 밤 노상에서 범죄 빈도가 낮은 것을 확인할 수 있었다.

다음 분석 결과는 중랑구에서 절도, 살인, 상해, 도박 범죄가 각각 일어날 최대 범죄발생확률을 분석하였다. 각 범죄 종류마다 범죄 발생확률이 높은 시간과 요일, 장소별로 구분하여 범죄 패턴을 형성하였다. 표 10은 중랑구의 범죄발생지수를 나타낸다.

분석 결과를 통해 중랑구는 토요일 노상에서 절도 범죄와 폭행 범죄가 발생할 확률이 높고, 살인 범죄는 토요일 단독주택에서 범죄 분포가 높게 나타났다. 그리고 도박 범죄는 금요일 밤 사무실에서 범죄발생지수가 높게 나타났다. 절도 범죄와 폭행 범죄는 모두 범죄발생지수가 1.8 이상이므로 두 범죄가 모두 일어날 가능성이 크나 폭행 범죄가 수치가 더 높은 것으로 보아 폭행 범죄에 대한 예방과 조치가 필요할 것이다. 도박 범죄는 금요일에 발생지수가 높은 것으로 보

표 9. 토요일 밤 노상에서 절도 범죄발생지수.

	P_{RTDS}
강남구	2.3
송파구	2.21
은평구	1.87
도봉구	1.72
금천구	1.7

표 10. 중랑구 절도, 살인, 상해, 도박 범죄발생지수.

	발생요일	발생장소	P_{RTDS}
절도	토요일	노상	1.82
살인	토요일	단독주택	2.46
폭행	토요일	노상	2.21
도박	금요일	사무실	2.18

아 금요일 사무실에서 도박 범죄에 대한 단속과 조치가 필요할 것이다.

V. 결론

본 논문에서는 서울특별시 각 구별을 연도별 범죄(절도, 살인, 강도, 방화, 강간, 폭행, 상해, 공갈, 악취와 유인 등) 빅 데이터를 이용하였다. 그 데이터의 위험지수를 계산하여, 범죄 패턴 알고리즘을 이용하여 범죄발생지수를 산출하였다. 그 범죄발생수치를 이용하여 지도에 맵핑함으로써 공간통계 분석을 통해 범죄 밀도 분석과 공간적 분포를 확인하였다. 공간통계분석을 보다 수치화된 정보로 파악하기 위해 NNI 분석법을 이용해 5대 범죄에 대한 군집 성향을 파악하였다.

특히 제안된 알고리즘을 통한 분석으로 범죄 발생은 각 구별 인구 및 도시생활 등에 따라 다른 범죄 분포 패턴을 보이는 것을 알 수 있었고, 범죄 발생확률을 위험지수를 통해 수치화함으로써 그 값을 단계별로 나눌 수 있는 지표를 제시하였다. 그리고 범죄 패턴 생성 알고리즘을 통해 범죄 데이터간의 여러 가지 패턴을 생성할 수 있는데, 다양한 패턴을 분석함으로써 여러 범죄 데이터 분석이 가능하다.

추후 연구에서는 정확한 범죄 예측으로 활용되기 위해 더욱더 상세하고 다양한 범죄관련 빅 데이터를 사용해 GIS를 통한 핫스팟 분석으로 범죄 예측이 필요할 것으로 판단된다. 나아가서 알고리즘을 통해 현재 범죄 예측 및 치안 유지에 필수적인 경찰서, 파출소 등에 활용될 수 있을 것으로 기대된다.

참 고 문 헌

[1] W. B. Liu and L. Ma, "A recent application of the ETAS model and a proposed method for prediction of strong

aftershocks," *Journal of pure and applied geophysics*, vol.163, no.11, pp. 2513-2528, Dec. 2006.

- [2] W. Ma, J. Chen, and P. Chen, "Illegal activities hotspot analysis based on GIS methods," in *proceedings of 2011 2nd IEEE International Conference on Emergency Management and Management Sciences (ICEMMS)*, pp. 270 - 273, Beijing, China, Aug. 2011.
- [3] S. Rahim and T. Sun, "ICTs based crime control model: An application based study of Gilgit-Baltistan, Pakistan," in *Proceedings of 2011 10th International Conference on Electronic Measurement & Instruments (ICEMI)*, pp. 1-6, Chengdu, China, Aug. 2011.
- [4] A.C. Alegria, H. Sahli, and E. Zimanyi "Application of density analysis for landmine risk mapping," in *Proceedings of 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)*, pp. 223-228, Fuzhou, China, June 2011.
- [5] L. Fajian, J. Zhang, and H. Zhang, D. Chen "Spatial structure of excellent tourism resources in Suzhou, China: A GIS-based approach," in *Proceedings of 2011 19th International Conference on Geoinformatics*, pp. 1-4, Shanghai, China, June 2011.
- [6] C. Peng, C. Tao, and Y. Hongyong, "GIS based crime risk analysis and management in cities," in *Proceedings of 2010 2nd International Conference on Information Science and Engineering (ICISE)*, pp. 3721-3724, Hangzhou, China, Dec. 2010.
- [7] N. Ishii, I. Toriin, and H. Tanaka, "Mapping of nearest neighbor for classification," in *Proceedings of 2013 IEEE/ACIS 12th International Conference on Computer and Information Science (ICIS)*, pp. 121-126, Niigata, Japan, Dec. 2010.

저자

차 경 현(Gyeong Hyeon Cha)



- 2014년 7월 : 광운대학교 전자융합공학과 졸업
- 2014년 8월 ~ 현재 : 광운대학교 전과공학과 석박사통합과정

<관심분야> : 데이터마이닝, 디지털통신, 5G 이동통신

김 경 호(Kyung Ho Kim)

학생회원



- 2013년 2월 : 광운대학교 전과공학과 졸업
- 2013년 3월 ~ 현재 : 광운대학교 전과공학과 석박사통합과정

<관심분야> : 디지털통신, 스마트그리드, 데이터마이닝, 5G 이동통신

황 유 민(Yu Min Hwang)

학생회원



- 2012년 2월 : 광운대학교 전과공학과 졸업
- 2012년 3월 ~ 현재 : 광운대학교 전과공학과 석박사통합과정

<관심분야> : 5G 이동통신, 재난 통신, WPAN, 협력통신, Big data 분석

이 동 창(Dong Chang Lee)



- 1990년 2월 : 경북대학교 전자공학과 학사
- 2006년11월 : 파마닉스/아트시스템 상무
- 2013년 2월 ~ 현재 : 위니텍해외사업본부 부장

<관심분야> : USN, Machine Vision, 영상처리, Multi-Vision System, 패턴 인식

김 상 지(Sang Ji Kim)



- 1990년 2월 : 동서대학교 소프트웨어공학과 석사
- 2006년 8월 : 쓰리비시스템 대리
- 2009년 6월 ~ 현재 : 위니텍 해외사업부 과장

<관심분야> : 신호처리, 영상처리, 패턴인식, 암호화, Water Mark

김 진 영(Jin Young Kim)

종신회원



- 1998년 2월 : 서울대학교 전자공학과 공학박사
- 2001년 2월 : SK텔레콤 네트워크연구소 책임연구원
- 2001년 3월 ~ 현재 : 광운대학교 전자융합공학과 교수

<관심분야> : 디지털통신, 가시광통신, UWB, 부호화, 인지무선통신, 4G 이동통신