



빅 데이터를 이용한 범죄패턴 분석 알고리즘의 구현

범죄 발생 빈도수를 이용하여 범죄발생지역, 시간, 요일, 장소의 위험지수를 구했고, 범죄 패턴 분석 알고리즘을 통해 범죄 발생 확률을 구했다.

범죄 관련 빅 데이터(지역, 시간, 요일, 장소 별 범죄발생 빈도수)

범죄 패턴 분석 알고리즘

Cluster Analysis(군집 분석)

계층적 군집분석을 통해 대략적인 군집의 수를 파악하고, 이를 이용하여 비 계층적 군집분석을 시행

위험지수

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

0에서 1사이의 데이터 별 위험지수를 구할 수 있다. 후에 범죄 발생 지수 구할 때 사용!

ex) 지역, 장소, 시간, 요일 별로 5대 범죄에 따른 위험지수

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

공간통계 분석

수집한 범죄데이터의 패턴을 알기 위해 표준편차 타원체(Standard deviational ellipse)와 밀도 분석(Densityanalysis) 을 이용

- 표준편차 타원 - 점 사상들의 산술적 평균과 표준거리를 이용하여 타원체를 도식화
- 밀도 분석 - 연구지역을 동일한 크기의 격자(Grid)로 나눈 후 격자에 집계되는 점 빈도수

⇒ 공간적 밀도 분석/ 알고리즘 모델링 가능!

- 최근린지수(Nearest neighbor index, NNI)분석법을 이용해 범죄패턴 분석

▼ 최근린지수(NNI)

임의의 한 점으로부터 다른 점들까지의 거리를 측정한 후, 각 지점에서 가장 가까운 다른 지점까지의 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 도출함.

NNI == 1 : 완전 임의적인 분포유형

NNI > 1 : 분산분포

NNI < 1 : 점들이 모여있는 군집적 분포

범죄패턴 분석 알고리즘

트리 형태로 패턴을 형성하는 것이 목표!

범죄발생지수(PRTDS)는 범죄발생빈도를 비교할 수 있는 척도

$$P_{RTDS} = R_n + T_n + D_n + S_n.$$



PRTDS : 범죄발생지수

ex) 강남구(R23)에서 토요일(D7) 밤(T6) 노상(S4)에서 절도 범죄로 패턴이 형성

$$P_{RTDS} = 0.88 + 0.21 + 0.6 + 0.6 = 2.29.$$

$P \geq 1.8$: 범죄발생빈도가 높은 것으로서 범죄가 일어날 가능성이 크다는 것을 의미함.

$P < 1.5$: 범죄 발생 가능성이 적다고 판단, 패턴에서 제외

⇒ 구한 범죄 발생 지수는 단계별로 지도에 원으로 맵핑(Mapping)이 되고, 공간통계 분석에서 밀도 분석에 이용

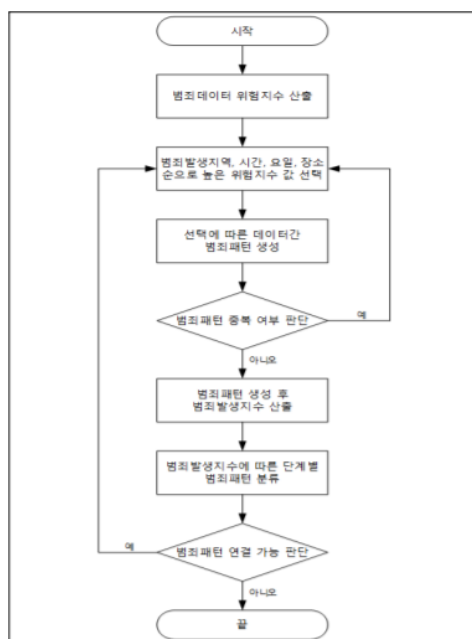


그림 1. 범죄패턴 생성 알고리즘.

범죄 패턴 생성하기

- R,T,D,S 순으로 각 데이터 별로 높은 위험지수 값을 선택 → 패턴 형성
- 패턴이 생성된후에는 각 데이터별 위험지수를 합산해 범죄발생지수를 산출
- 산출된 범죄발생지수를 이용해 그 값에 따라 단계별로 위험도를 분류

결과

범죄관련 빅 데이터의 공간적 분포 패턴 결과

▼ 5대 범죄를 범죄 밀도 분석한 결과

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 2. 범죄발생지 및 타원체 중심점.

→ 이걸 범죄 밀도 분석, 이를 다시 NNI 분석법으로 분석해 공간적 분포 확인

NNI 분석 수식

NNI로 분석한 결과

$$d_m = \frac{1}{n} \sum_{i=1}^n d_i.$$

$$d_k = 0.5 \sqrt{\frac{A}{n}}.$$

$$NNI = \frac{d_m}{d_k}.$$

d_i : i 지점에서 가장 가까운 거리

d_m : 한 지점에서 가장 가까운 다른 지점까지의 평균

A : 대상지역의 면적

표 8. 5대 범죄유형별 최근린지수.

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

▼ NNI 분석 결과

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.

범죄관련 빅 데이터 패턴 알고리즘 분석 결과

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

▼ 분석 결과

분석 결과를 통해 중랑구는 토요일 노상에서 절도 범죄와 폭행 범죄가 발생할 확률이 높고, 살인 범죄는 토요일 단독 주택에서 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 보아 금요일 사무실에서의 도박 범죄에 대한 단속과 조치가 필요할 것이다

