

심각범죄의 효과적 대응을 위한 기계학습 기반의 범죄 유형 및 범죄위험스코어 예측 기술 연구

백명선¹⁾, 이용태¹⁾, 장광호²⁾, 이희조³⁾, 박원주¹⁾

1) 한국전자통신연구원, 2) 경찰대학 치안정책연구소, 3) (주)모비젠

{sabman, ytle, wjpark}@etri.re.kr, pathfinder@police.go.kr, heejo@mobigen.com

A Study on the Estimation Technique of Crime Types and Crime Risk Score based on Machine Learning for Effective Response of Severe Crimes

Myung-Sun Baek¹⁾, Yong-Tae Lee¹⁾, Kwangho Jang²⁾, Heejo Lee³⁾, Wonjoo Park¹⁾

1) ETRI, 2) Police Science Institute, 3) Mobigen Co., Ltd.

요 약

본 논문에서는 지능형 범죄 및 심각 범죄에 대한 효과적인 대응을 위해 범죄 신고 접수 단계에서 범죄의 유형과 범죄의 심각정도를 예측하는 플랫폼을 설계하고 성능을 검증한다. 범죄 심각정도를 결정하기 위해 범죄의 심각정도를 수치로 나타낼 수 있는 범죄위험스코어 생성을 선행 개발한다. 범죄위험스코어는 범죄의 행위 및 피해정도에 따른 심각도를 수치화 한 것이다. 상기 기본 정보를 바탕으로 신규 범죄신고가 접수되었을 경우, 신고내용을 기계학습을 통해 분석하고 이를 바탕으로 범죄의 유형과 범죄위험스코어를 예측한다. 이렇게 신고 단계에서 예측된 범죄 유형 및 심각도를 참고하여 경찰은 초동 대응방법 및 인력 배치를 효과적으로 수행할 수 있다.

I. 서 론

다양한 통신기기/첨단 장비들을 활용한 지능형 범죄와 더불어 피해자의 신체 및 재산을 크게 손상시키는 심각범죄의 발생 빈도가 점차 증가하고 있다. 이와 같은 지능형 심각 범죄에 효과적으로 대응하기 위해 빅데이터/인공지능 기반의 ICT 기술을 활용한 지능형 치안 관련 기술들의 개발이 활발히 진행되고 있다 [1]-[2]. 지능형 범죄의 효과적인 초동 대응을 위해서는 범죄 신고 접수단계에서 범죄의 종류 및 심각성을 빠르게 예측할 수 있는 기술이 필요하다.

본 논문에서는 범죄 신고 접수단계에서 범죄의 유형을 예측하고 심각도를 수치로 분석하여 표출하는 기술을 설계하고 성능을 검증한다. 또한 범죄 심각도를 수치화 하기 위해 범죄 수법 및 피해 정도에 따른 범죄위험스코어를 계산하는 방법을 제시한다. 개발된 기술은 신고 접수 단계에서 범죄 유형 및 심각도를 예측하여 제공하므로 신고 접수 경찰은 즉시 출동 여부/전문인력 배치 등 효과적인 대응에 해당 정보를 참고하는 것이 가능하다.

II. 범죄 위험스코어 설계

본 논문에서 개발한 예측 시스템을 위한 범죄 위험스코어는 범죄 신고 접수 단계에서 획득되는 범죄자의 범죄 유형 및 피해자의 피해정도 등의 정보를 기반으로 다음과 같이 설정하였다.

위의 식에서 각각의 변수에 대한 설명 및 설정 값은 표 1 과 같다. 본 논문에서는 표 1 의 값을 사용하였지만 설계된 시스템의 운용상의 용이성과 실제 환경에 적용성 향상을 위해 변수 값 조정 및 변수의 추가 삭제 또한 가능하다.

표 1 범죄위험스코어 도출 변수

스코어변수	기본값	범위_최소값	범위_최대값	코드	변수 값
피해자 성별 (WS)	2	1	1.2	남	1
				여	1.2
				불상	1.1
피해자 연령 (WA)	2	1	2.5	6 세 이하	2.5
				20 세 이하	2~1.5
				30~60 세 이하	1
				60 세 초과	1.2
				미상	1
신체 피해정도 (WP)	100	0	5	피해무	0
				전치 1 개월 이하	0.8~1
				전치 6 개월 이하	1.2~1.6
				전치 6 개월 초과	2
				사망	5
재산 피해정도 (WM)	30	0	3	피해무	0
				100 만원 이하	0.1~0.5
				1 억이하	0.8~1
				10 억원 초과	2~3
범죄수법 (WR)	50	0	2	미상	0.5
				강도 (수법에 따라)	2~1.5
				절도 (수법에 따라)	1.2~1
				미분류	0

$$CRS = (WS + WA) + WP + WM + WR \quad (1)$$

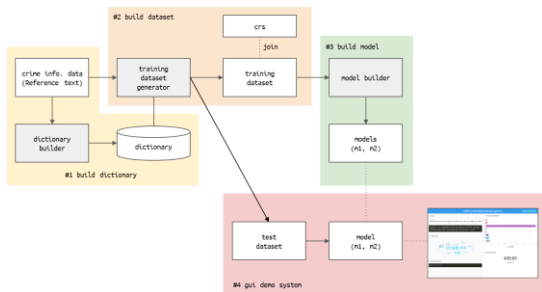


그림 1. 범죄 유형 및 범죄위험스코어 예측 시스템 구조도

III. 범죄 유형 및 범죄위험스코어 예측 시스템

그림 1 은 기계학습 기반의 범죄 유형 및 범죄위험스코어 예측 시스템의 구조를 나타내는 구조도이다. 그림에서와 같이 범죄 유형 및 범죄위험스코어 예측 시스템은 크게 네 가지 기능블록으로 구성되어 있다. 첫번째 블록인 build dictionary 블록은 텍스트로 구성된 범죄 정보를 포함하는 범죄 reference 소스로부터 feature keywords 를 추출하여 키워드 사전을 구축한다. 상기 feature keywords 추출을 위해 word rank [3]와 tf-idf [4]가 사용되었다. 범죄 reference 소스는 실제 경찰에서 작성 및 사용하는 양식에 따른 가상수사데이터가 사용되었다. build dataset 블록은 데이터 소스와 키워드 사전을 이용하여 학습용 데이터셋을 생성한다. 생성된 학습용 데이터셋에는 feature keywords, 범죄위험스코어, 범죄유형이 포함된다. build model 은 상기 생성된 학습용 데이터셋을 기반으로 딥러닝구조를 사용하여 아래와 같은 두가지 예측모델을 생성한다.

- m1: 범죄위험스코어 예측모델
- m2: 범죄 유형 예측 모델

마지막 gui demo system 블록은 범죄 신고 접수 정보를 입력받아 예측값을 표출하는 GUI 블록이다.

구축된 예측 시스템에는 테스트 동작을 수행하는 동안 실제 정답 값 (범죄 유형과 범죄위험스코어)을 기반으로 예측의 정확도를 분석/검증하는 기능이 포함되어 있다. 따라서 개발된 예측 시스템의 성능을 분석하고 지속적으로 성능 향상을 수행하는 것이 가능하다.

IV. 모의실험 결과 및 결론

설계된 시스템의 성능을 검증하기 위해 7 대 강력범죄에 해당하는 범죄사실 데이터를 사용하였다. 그림 2 는 범죄 유형 및 범죄위험스코어 예측 시스템 성능을 보여준다. 그림 2-(a)는 실제 시스템에 입력된 범죄 사실 데이터 및 해당 입력값에 따라 추출된 Keyword vector 를 보여준다. 입력된 범죄 사실 데이터의 자세한 내용은 표 2 와 같다. 그림 2-(b)는 입력값을 바탕으로 예측한 범죄 유형 및 범죄위험스코어 결과 이다. 표 2 에서와 같이 입력된 범죄 데이터는 절도에 해당하며, 실제 위험도는 351.4 이다. 개발된 예측 시스템에서 추정한 범죄 유형은 절도가 92%로 가장 높았으며, 예측된 범죄위험스코어는 344.77 이다. 따라서 실제 입력된 값과 매우 유사한 결과를 예측하는 것을 확인할 수 있다. 또한 그림 2-(c)에서와 같이 범죄위험스코어 예측 및 범죄 유형예측의 전체 성능은 모두 0.8 이상으로 우수한 결과를 제공하는 것을 확인할 수 있다. 개발된 예측 시스템은 접수된 범죄 내용을 기반으로 범죄 유형 및 범죄위험스코어를 예측하여 표출하므로, 경찰의 효과적인 초동대응을 위한 참고자료로 활용될 수 있다.

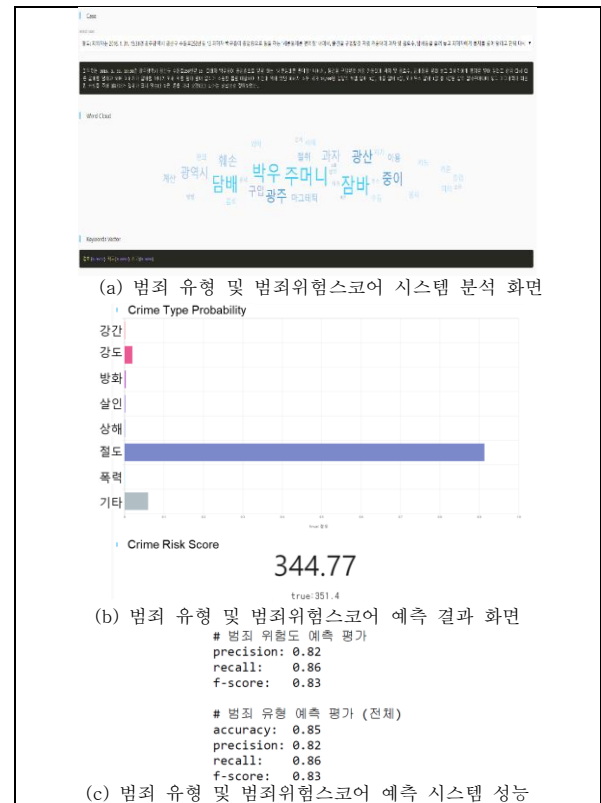


그림 2. 개발된 범죄 유형 및 범죄위험스코어 예측 시스템 GUI 및 성능

표 2. 입력된 범죄 내용

피의자는 2016. 1. 31. 15:30 경 광주광역시 광산구 수동로 xxx 길 피해자 xxx 이 종업원으로 일하는 편의점 내에서, 물건을 구입할것 처럼 카운터에 과자 및 음료수, 담배등을 올려 놓고 피해자에게 봉지를 넣어 달라고 한뒤 다시 다른담배를 달라고 하여 피해자가 담배를 꺼내기 위해 뒤를 돌아 볼 때 갑자기 소홀한 틈을 이용하여 계산대 위에 있던 피해자 소유 시가 13,500 원 상당의 보행 담배 1 갑, 레종 담배 1 갑, 피비우스 담배 1 갑 총 3 갑을 상의 장바구니에 넣고 마그네틱이 훼손된 카드를 주어 피해자가 결제가 되지 않는다 하던 돈을 가져오겠다고 나가는 방법으로 절취하였다.

ACKNOWLEDGMENT

이 논문은 2020 년도 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] 방준성, 박원주, 윤상연, 신지호, 이용태, “지능형 치안 서비스 기술 동향”, [ETRI] 전자통신동향분석, vol. 34, no. 1, pp. 111-122, 2019. 2.
- [2] 이상운, 엄경록, 황재원, 박원주, 장광호, 이용태, “치안정보 분석을 위한 데이터뱅크 설계에 관한 연구”, 대한전자공학회 학술대회, pp. 697-698, 2019. 11.
- [3] 김현중, 조성준, 강필성, “KR-WordRank: WordRank 를 개선한 비지도학습 기반 한국어 단어 추출 방법”, 대한산업공학회지, vol. 40, no. 1, pp. 18-33, 2014. 2.
- [4] 박대서, 김화중, “TF-IDF 기반 키워드 추출에서의 의미적 요소 반영을 위한 결합벡터 제안,” 한국정보기술학회논문지, vol. 16, no. 2, pp. 1-16, 2018. 2.