

원저

독립 추론 엔진을 활용한 딥러닝 기반의 범죄 분류 및 심각도 추론 기술

백명선¹, 이용태², 송형규^{3,4}

¹세종대학교 지능정보융합학과 교수

²한국전자통신연구원 책임연구원

³세종대학교 정보통신공학과

⁴세종대학교 지능형드론융합전공 교수

교신저자: 송형규, songhk@sejong.ac.kr

요약

범죄사건 접수 초기에 신속하고 효과적으로 해당범죄에 대응하기 위해 발생한 범죄의 유형과 심각수준에 대한 정보를 현장 파견 요원들에게 제공하는 것이 매우 효과적이다. 본 논문에서는 딥러닝 기술을 바탕으로 범죄의 유형을 분류하고 및 심각도를 추론하는 기술을 설계하고 성능을 검증한다. 설계된 기술은 범죄사건 접수 데이터인 텍스트 기반의 범죄사건 개요 정보를 바탕으로 해당 범죄의 유형분류 및 범죄 심각도 추론을 수행한다. 본 연구에서는 범죄의 심각 정도를 수치로 나타내는 범죄심각도 산출식을 선행 개발한다. 개발된 산출식은 범죄사건 별 심각도를 범죄의 위험정도 및 피해자의 피해정도를 고려하여 수치로 출력한다. 따라서 각 사건에 대해 범죄 유형 및 피해자의 피해정도가 정확히 파악된 경우 개별 사건에 따른 심각도를 효과적으로 산출하는 것이 가능하다. 본 논문에서 개발된 기술은 개별 사건이 접수되면 텍스트 기반의 사건 개요를 활용하여 해당 사건의 범죄 유형분류 및 범죄심각도를 추론하는 것이 가능하다. 또한 GUI가 포함된 소프트웨어 플랫폼 형태로 개발되어 현장 인원들이 직관적으로 사건 유형 및 심각도를 파악할 수 있다.

주제어

범죄 유형 분류, 심각도 산출, 범죄 추론, 과학치안, 스마트폴리싱

Open Access

Received: November 27, 2024

Revised: December 23, 2024

Accepted: December 23, 2024

Published: December 31, 2024

© 2024 Korean Data Forensic Society

This is an Open Access article distributed under the terms of the Creative Commons CC BY 4.0 license (<https://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Original Article

Deep learning-based crime classification and severity inference technology using an independent inference engine

Myung-Sun Baek¹, Yong-Tae Lee², Hyung-Kyu Song^{3,4}

¹Professor, Department of Intelligence and Information Convergence, Sejong University, Republic of Korea

²Principal Researcher, Electronics and Telecommunications Research Institute, Republic of Korea

³Professor, Department of Information and Communication Engineering, Sejong University, Republic of Korea

⁴Professor, Convergence Engineering for Intelligent Drone, Sejong University, Republic of Korea

Corresponding Author: Hyung-Kyu Song, msbaek@sejong.ac.kr

ABSTRACT

In order to respond to the crime quickly and effectively at the initial stage of receiving a crime case, it is very effective to provide information on the type and severity of the crime to field dispatched agents. In this paper, we design and verify the performance of a technology that classifies the type of crime and infers the severity based on deep learning technology. The designed technology performs crime type classification and crime severity inference based on text-based crime case summary information, which is crime case reception data. In this study, we develop a crime severity calculation formula that numerically represents the severity of the crime. The developed calculation formula outputs the severity of each crime case as a number by considering the degree of risk of the crime and the degree of damage to the victim. Therefore, if the crime type and the degree of damage to the victim are accurately identified for each case, it is possible to effectively calculate the severity for each case. The technology developed in this paper can classify the crime type and infer the crime severity of the case by utilizing the text-based case summary when an individual case is received. In addition, it is developed in the form of a software platform with a GUI, so that field personnel can intuitively identify the type and severity of the case.

KEYWORDS

Crime type classification, severity calculation, crime inference, scientific policing, smart policing

1. 서론

최근 인공지능/빅데이터를 발판으로 과학기술은 비약적으로 성장하고 있다. 과학기술 발전의 견인차 역할을 수행한 핵심 요소기술인 인공지능 기술은 스마트 의료기술분야, 음성인식/자연어처리 분야, 로봇/자율주행자동차 연구 분야, 인공지능기반의 무선신호 송수신 분야 등 공학분야 전반에 걸쳐 활발하게 연구되고 있다 [1-3]. 뿐만 아니라 주가 변동 예측 등의 경제분야, 및 다양한 예술 분야에서도 인공지능을 접목한 연구가 진행되고 있다 [4,5]. 인공지능 발달의 배경에는 빅데이터 기술이 존재한다. 빅데이터 기술은 인공지능이 원활하게 동작할 수 있도록 하는 또 다른 요소기술로 다양한 분야에 사용되고 있다 [6,7].

대국민의 생활/안전을 보장하기 위한 치안 분야에서도 고도화된 치안기술 개발을 위한 인공지능/빅데이터 적용이 활발하게 연구되고 있다 [8]. 우선 경찰의 누적된 데이터를 디지털화 한 빅데이터 기반의 치안기술 연구 개발이 매우 활발히 진행 중이다 [9-11]. 문헌 [9]에서는 치안분야에서 ICT 기술이 적용될 수 있는 분야 및 활용사례 등을 소개하고, 데이터 확보, 통합DB 구축, 전문가집단의 양성, 유관기관 사이의 교류협력 등 빅데이터 기반의 경찰활동이 수행될 수 있는 다양한 빅데이터 환경 구축의 필요성 등이 논의되었다. 문헌 [10]에서는 치안분야에서 빅데이터가 활용될 수 있는 사례들과 더불어 이러한 빅데이터 기반의 치안기술 개발을 위한 다양한 공공연구 조직 설립 및 운영에 대한 방안을 논의하였다. 해당 문헌은 치안분야에서 빅데이터 활용서비스로 범죄 예측 서비스, 교통 관련 서비스, 영상 분석과 통합 관제 서비스, 신원 확인 (바이오 인식 기술) 서비스 등을 사례로 언급하고 있다. 또한 문헌 [11]은 보이스피싱 대응 수사를 위한 비정형 데이터 네트워크 기법의 수사에 대한 내용을 다루고 있다. 해당 문헌에서는 경찰이 보이스피싱 데이터를 처리하는 방식이 정교하지 않은 것에 대한 제언으로 비정형 수사데이터에서 전화번호, 계좌번호, 범죄 수법 등 핵심 용어를 추출하는 기술을 연구하여 설명하고 있다.

이와 더불어 인공지능 기술을 활용한 적극적인 스마트 치안에 대한 연구도 매우 활발히 진행되고 있다. 가장 활발하게 연구되고 있는 인공지능 기반의 치안 기술 분야는 범죄예측과 관련된 분야이다 [12-14]. 문헌 [12]는 빅데이터로 구축 가능한 범죄 정보와 도시지역 자료를 바탕으로 머신러닝 방식을 통해 범죄발생 위험지역 예측 모델을 개발하였다. 상기 모델을 통해 경찰력이 부족한 상황에서 효율성 있게 범죄에 대처할 수 있는 부가 정보를 제공할 수 있다. 문헌 [13]은 치안데이터, 공공데이터 등을 활용하여 기계학습 기반의 범죄 건수를 조기 예측하는 기술을 제안하였다. 상기 기술을 통해 효율적인 인력 배치를 위한 정보를 제공하는 등 실제 치안 인력을 지원하기 위한 방안을 마련하고 있다. 문헌 [14]는 범죄 접수단계에서 범죄의 유형과 범죄의 심각도를 예측하는 플랫폼을 설계하고 성능을 검증하였다. 해당 기술 기반으로 도출된 결과를 통해 경찰은 초동 대응방법 및 인력 배치 등에 참고할 수 있다.

본 논문은 문헌 [14]를 한국 치안상황에 적합한 시스템으로 확장 고도화한 연구내용을 다루고 있다. 본 논문에서는 위험 상황의 초기 인지를 위한 기계학습 기반의 범죄 유형 및 범죄 심각도를 추론하는 기술을 제안한다. 제안된 기술은 경찰에 접수된 사건 현황에 대해 수집된 텍스트 데이터를 분석하여 범죄 유형을 예측한다. 또한 이와 함께 범죄 심각도 추론을 수행한다. 상기 범죄 유형 예측 및 범죄 심각도 추론 기술을 개발하기에 앞서 범죄 심각도를 수치적으로 산출하는 산출식을 선행 개발한다. 범죄 심각도 산출식은 범죄 유형 및 피해자의 피해정도를 고려하여 범죄의 심각성을 산출하는 산출식이다. 따라서 제안된 기술은 접수된 사건이 어떤 유형인지를 예측한 후 얼마나 심각한 정도의 사건인지를 추론하여 해당 정보를 출력한다. 경찰은 접수된 사건에 대해 초동대응 및 현장 인력 배치 등에 상기 기술의 정보를 활용할 수 있다. 또한 개발된 기

술은 GUI기반의 SW 플랫폼 형태로 구현되어 현장 인력들이 쉽게 해당 결과물을 사용하는 것이 가능하다. 제안된 기술 및 플랫폼의 성능 검증을 위해 치안현장에서 실제 사용하는 사건 데이터인 형사사법정보시스템 (KICS: Korea Information System of Criminal Justice Services) 양식을 준용한 가상 KICS데이터를 생성하여 활용하였다. 가상 KICS데이터에서 사건 관련 텍스트데이터를 기반으로 범죄 유형 및 범죄 심각도를 예측하고 실제 결과와 비교하여 성능을 검증하였다.

본 논문은 다음과 같이 구성된다. 2절에서는 범죄의 심각성을 수치로 산출하는 범죄심각도 산출식을 설계한다. 3절에서는 사건관련 텍스트데이터를 사용하여 범죄유형을 예측하고 심각도를 추론하는 시스템을 구축한다. 4절에서는 설계된 시스템의 성능을 검증한다. 성능 검증을 위해 범죄의 유형을 예측하는 성능과 범죄의 심각도를 추론하는 성능을 각각 검증한다. 마지막으로 5절에서 논문의 결론은 논의한다.

2. 범죄 심각도 산출식

본 절에서는 사건의 내용에 기반하여 범죄의 심각도를 판단할 수 있는 범죄 심각도를 산출하는 식을 설계한다. 범죄 심각도는 범죄유형에 따른 기본 스코어와 피해자의 신체적, 금전적 피해 정도에 따라 산출되는 수치값이며 아래의 식과 같이 나타낼 수 있다.

$$CSS = BS \times (WS + WA) + WP + WM \quad (1)$$

위의 식에서 CSS는 범죄심각도 (Crime Severity Score)를 나타낸다. 산출식의 각 계수들은 다음과 같다.

- BS (Base Score): 범죄 유형 (죄종)에 따른 기본 가중치
- WS (Weight Sex): 피해자 성별 가중치
- WA (Weight Age): 피해자 연령 별 가중치
- WP (Weight Physical): 피해자의 신체 피해 정도에 따른 가중치
- WM (Weight Material): 피해자의 재산 피해 정도에 따른 가중치

상기 BS 기본 가중치는 죄종에 따른 가중치를 나타낸다. 본 논문에서는 경찰에서 일반적으로 고려하는 7종의 강력범죄를 고려하였으며 <Table 1>과 같이 각 범죄 유형 별로 점수를 할당하였다. 가중치는 단순히 범죄유형의 강력성만을 고려한 점수가 아닌 각 죄종별 부가되는 양형기준을 고려하여 할당한 값이다. 상기 죄종에 따른 기본 가중치에 곱셈연산으로 부가되는 가중치가 WS와 WA이다. 상기한 바와 같이 곱셈연산으로 부가되는 가중치는 사건에서 피해자의 신체와 직접적으로 연관된 정보를 바탕으로 설정된다. 상기 가중치는 피해자가 얼마나 범죄에 취약한가를 판별하는 직접적인 정보라고 판단하여 범죄의 심각도에 중요한 곱셈가중치에 할당된다. <Table 2>는 해당 가중치에 할당된 값을 나타낸다. 표에서와 같이 범죄피해에 상대적으로 취약한 여성에 다소 높은 값을 할당하였으며, 이와 동일한 관점에서 범죄에 취약한 어린아이와 고령층에 높은 값을 할당한다. 마지막 가중치는 범죄의 피해정도에 해당하는 가중치로 신체피해를 나타내는 WP와 재산 피해를 나타내는 WM이다. WP는 피해자가 신체에 얼마나 피해를 입었는지 정도에 따라 할당값이 결정되며 전치 단위로 평가한다. WM은 재산 피해 정도를 나타내

며 금액이 클수록 높은 값이 할당된다. WP와 WM에 할당된 값은 <Table 3>과 같다.

따라서 범죄 심각도는 식 (1)을 기반으로 각 사건 별 정보에 따라 결정되며 높은 값을 가질수록 심각도가 높은 것으로 판단할 수 있다.

<Table 1> BS 할당 값

	죄종	가중치
1	살인	100
2	강간	75
3	방화	70
4	강도	60
5	상해	60
6	폭행	55
7	절도	30

<Table 2> WS, WA 할당 값

가중치 대상 코드	가중방식	범위_최소값	범위_최대값	코드	가중치	비고
피해자 성 (WS)	Wx (곱하기)	1	1.2	남	1	
				여	1.2	
				불상	1.1	
피해자 연령 (WA)	Wx (곱하기)	1	2.5	6세이하	2.5	
				12세이하	2	
				15세이하	1.5	
				20세이하	1.2	
				30세이하	1	
				40세이하	1	
				50세이하	1	
				60세이하	1	
				60세초과	1.2	
				기타	1	
				미상	1	
	NULL	1	코드값이 NULL인 경우			

<Table 3> WP, WM 할당 값

가중치 대상 코드	가중 방식	코드	가중치	비고
신체피해정도 (WP)	W+ (더하기)	피해무	0	
		전치2주이하	0.8	
		전치1개월이하	1	
		전치2개월이하	1.2	
		전치4개월이하	1.4	
		전치6개월이하	1.6	
		전치6개월초과	2	
		사망	5	
재산피해정도 (WM)	W+(더하기)	피해무	0	
		1만원이하	0.1	
		10만원이하	0.2	
		100만원이하	0.5	
		1000만원이하	0.8	
		1억원이하	1	
		10억원이하	2	
		10억원초과	3	
		미상 (모름)	0.5	코드별 가중치 중간값 적용

3. 범죄 유형 및 심각도 예측 시스템

<Figure 1>은 제안된 범죄 유형 및 심각도 예측 시스템의 구조도이다. 제안된 시스템은 크게 아래와 같이 4가지 요소로 구성된다.

- build dictionary: 텍스트 기반의 데이터 소스에서 Feature Keywords 추출 알고리즘을 이용해서 키워드 사전(Dictionary)을 구축한다.
- build dataset: 데이터 소스와 키워드 사전을 이용해서 학습용 데이터셋(Feature Keywords + 심각도 + 범죄유형)을 생성한다.
- build model: 데이터 학습을 통해 딥러닝 기반의 예측 모델을 생성한다. 생성된 모델은 심각도 예측 모델과 범죄유형 예측 모델의 두 가지 이다.
- gui system: 범죄사실 텍스트 데이터를 기반으로 범죄유형과 범죄 심각도를 예측하는 GUI 시스템이다.

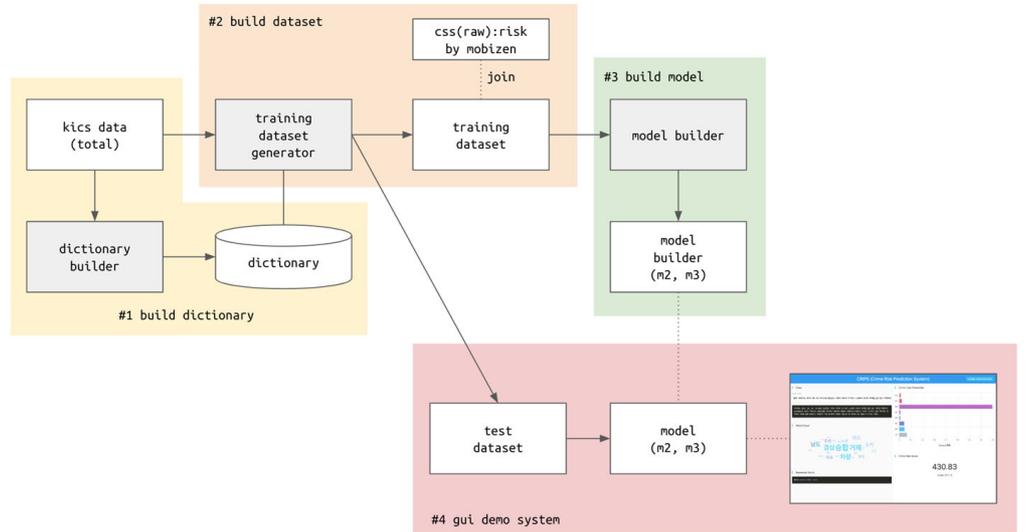
각 구성 요소에 대한 기능 및 동작 방식에 대한 자세한 설명은 아래 이어지는 항에서 수행한다.

3.1. 키워드 사전 구축 (Build Dictionary)

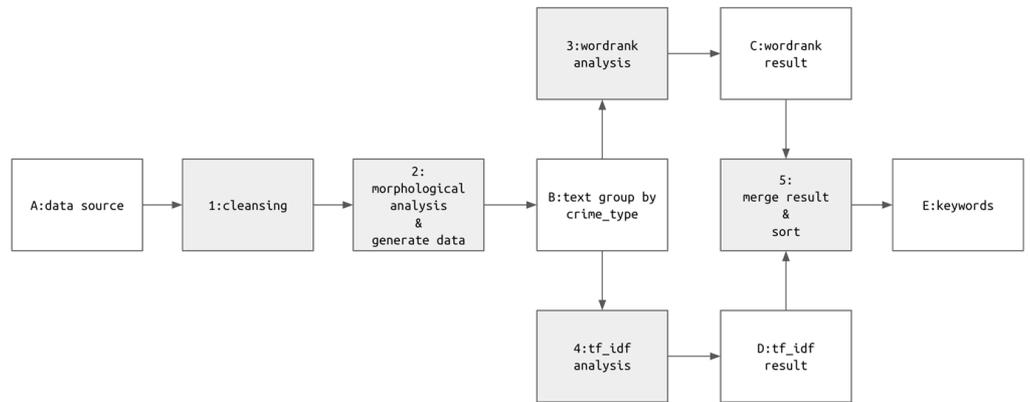
<Figure 2>는 Feature Keyword를 추출하는 프로세스를 나타내는 블록도이다. 본 기술의 특징은 wordrank 분석 기법[15,16]과 tf_idf 분석 기법[17,18]을 모두 사용하여 키워드를 추출하는 것이다.

그림에서 입력데이터 소스는 텍스트 기반의 수사 데이터로 범죄 건 별로 입력된다. 이렇게 입력된 데이터는 정제(cleaning)과정을 거치게 된다. 정제 과정에서는 데이터 소스에서 키워드 추출에 필요 없거나 방해가 될 수 있는 노이즈가 제거된다. 이후 형태소 분석 (morphological analysis)을 수행한다. 형태소 분석 절차에서는 일반명사(NNG)와 동사(VV)만 추출한다. 이 과정에서 추출된 데이터는 wordrank analysis [15,16]에 적용하기 위해 검거죄종(강력 7대)별로 Grouping과정을 거치게 된다. 상기 과정을 통해 B: text group by crime_type 데이터가 출력된다.

이후 wordrank 분석을 통해 특징 키워드를 추출하고 그 결과를 C: wordrank list에 기록한다. 또한 이 과정에서는 사용자에게 의해 정의된 불용어(stopwords)를 제거한다. 또한 이와 동시에 tf-idf 분석을 통해 키워드를 추출하고 그 결과를 D:tf_idf list에 기록한다. 마지막으로 wordrank분석과 tf_idf 분석 결과를 합병하고, 높은 순으로 정렬한다. 합병 시에는 동일한 키워드를 찾아서 둘의 평가 점수를 합하는 방식으로 합병을 수행한다. 위의 절차를 기반으로 각 범죄유형별로 높은 점수를 받은 상위 10개의 키워드를 추출하여 feature keywords를 추출한다.



<Figure1> 시스템 구조도



<Figure 2> 키워드 추출 프로세스

3.2. 데이터셋 생성 (Build Dataset)

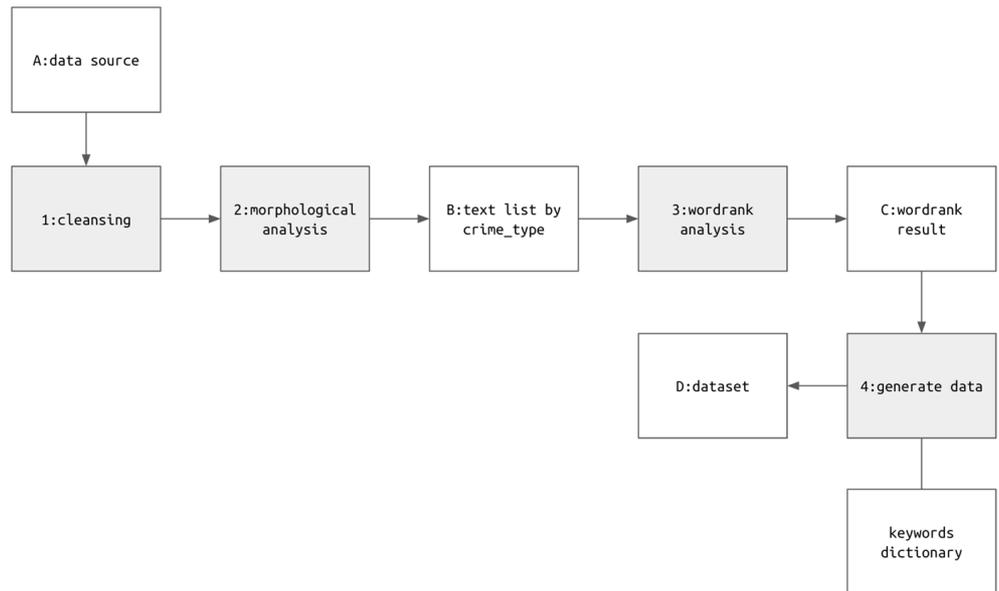
<Figure 3>은 데이터셋을 생성하는 과정을 나타내는 블록도이다. 데이터셋 구축 절차에서도 키워드사전 구축 절차와 동일한 데이터 정제, 형태소 분석 및 Grouping 절차가 수행된다. 상기 절차를 거친 출력 데이터 B: text group by crime_type가 wordrank 분석기로 입력된다. wordrank 분석을 기반으로 데이터의 각 사건 별 키워드를 추출하고 C:wordrank result에 기록한다.

generate data 블록에서는 wordrank 분석 결과에서 키워드 사전에 있는 단어들만 추출하고, 데이터셋 형식에 맞도록 벡터 데이터를 만들어서 전체 데이터셋에 추가한다.

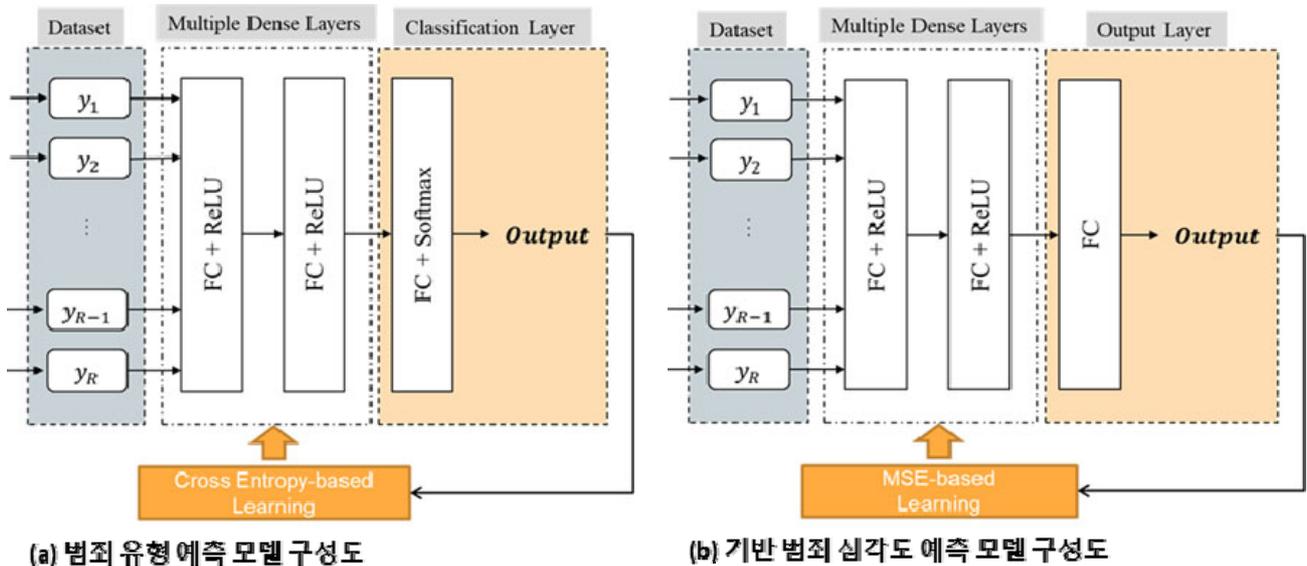
3.3. 예측 모델 생성 (Build Model)

<Figure 4>는 범죄유형을 예측하는 딥러닝 기반 예측모델 구성도와 범죄 심각도를 예측하는 딥러닝 기반 예측모델 구성도이다. 그림에서와 같이 두 예측모델 모두 2개의 ReLU기반의 fully

Connected Layer와 1개의 출력 layer로 구성되어 있다. 범죄유형 예측모델의 경우 출력레이어는 softmax기반의 Classification Layer로 구성되어 있다. 이때 최종 출력은 7대 강력범죄 유형과 기타유형을 나타낼 수 있도록 길이 8의 벡터이다. 모델의 학습을 위해서 Categorical Cross Entropy 기반의 Loss값을 이용한다. 반면 범죄 심각도 예측 모델에서는 수치 값 기반의 범죄심각도를 예측하기 위해 Output Layer는 1개의 값을 출력하며 Activation Function은 적용되지 않는다. 또한 모델의 학습을 위해 MSE (Mean Square Error)값이 Loss로 사용된다. <Table 4>는 예측 모델의 Hyper-Parameter를 나타낸다.



<Figure 3> 데이터셋 생성 과정



(a) 범죄 유형 예측 모델 구성도

(b) 기반 범죄 심각도 예측 모델 구성도

<Figure 4> 딥러닝 기반 범죄 유형 및 심각도 예측 모델 구성도

<Table 4> 범죄 유형 및 심각도 예측 모델 Hyper-Parameter

Layer	Size	Activation
Crime Type Prediction Model		
Input layer	70	-
FC 1	128	ReLU
FC 2	64	ReLU
FC 3 (Output)	8	Softmax
Crime Risk Score Prediction Model		
Input layer	70	-
FC 1	128	ReLU
FC 2	64	ReLU
FC 3 (Output)	1	None

4. 개발 결과 및 성능 분석

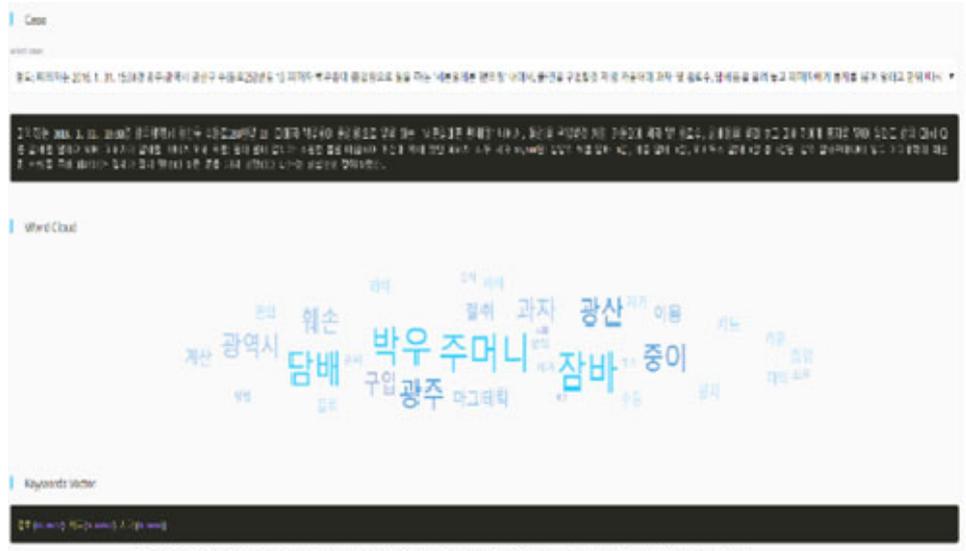
<Figure 5>는 본 논문에서 설계한 예측 기술이 적용된 GUI System이다. 그림에서 입력된 범죄 내용은 <Table 4>와 같다. 그림에서와 같이 개발된 시스템 결과물은 범죄사실을 입력하면 이에 해당하는 Word Cloud, Keyword Vector, Crimt Type Probability, Crime Risk Score를 출력한다. 각각의 의미는 아래와 같다.

- Word Cloud: 범죄사실 텍스트 데이터에서 형태소를 추출하고, 이에 대한 WordRank 가중치로 워드 클라우드를 표현한다.
- Keyword Vector: 범죄 사실 텍스트 데이터에서 키워드 벡터의 가중치를 표현한다.
- Crime Type Probability: 범죄 사실 텍스트 데이터를 기반으로 범죄 유형을 예측한 결과를 확률로 표현한다.
- Crime Risk Score: 범죄 사실 텍스트 데이터를 기반으로 범죄 심각도를 예측한 결과를 출력한다.

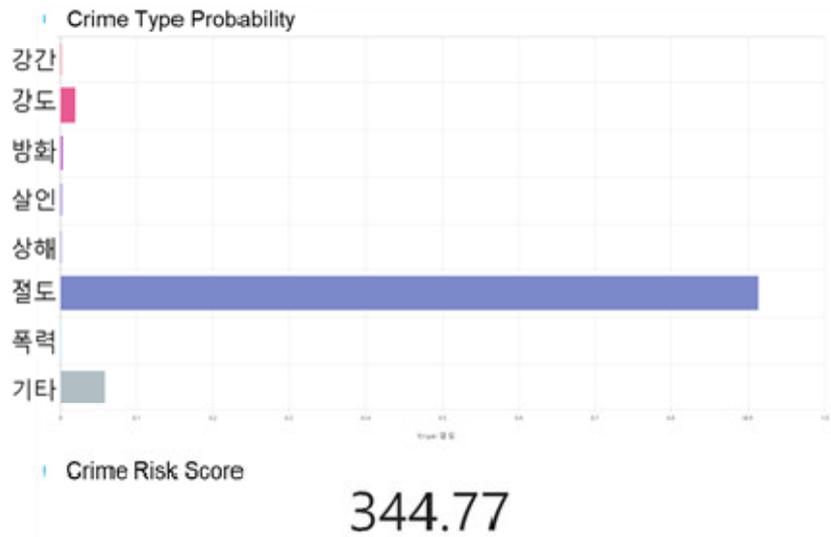
상기와 같은 다양한 기능을 통해 개발된 시스템의 예측 결과 뿐만 아니라 예측을 위한 동작결과 또한 점검하는 것이 가능하다. <Table 5>는 상기 <Figure 5>의 결과를 도출하기 위해 사용된 입력데이터를 보여준다. 표에서와 같이 입력된 내용은 범죄 유형중 절도에 해당한다. 또한 정확한 정보에 따른 범죄 위험도는 351점이다. 개발된 시스템은 범죄 유형을 92%의 확률로 절로도 예측하였으며 범죄 위험도 또한 344.77로 실제 위험도와 유사한 결과를 도출하였다.

<Figure 6>은 다른 범죄 사실 데이터 기반의 결과를 보여준다. 해당 결과의 입력값은 <Table 6>과 같으며 방화에 해당하는 내용을 담고 있다. <Figure 6>에서와 같이 개발된 시스템은 약 79%의 확률로 범죄 유형을 방화로 예측하였으며 430.83의 범죄 위험도를 예측하였다. 범죄 위험도에는 다소 차이가 있지만 범죄 유형은 실제 방화에 적합하게 예측하였다. 범죄 위험도의 차이가 큰 것은 텍스트 내에 피해사실 등의 정보가 많이 포함되어 있지 않기 때문이다.

<Figure 7>은 개발된 시스템의 전체 예측 성능을 검증한 결과를 보여준다. 그림에서와 같이 범죄 유형 예측 및 범죄 위험도 예측 모두 0.8이상의 값을 달성하며 우수한 성능을 도출하는 것을 확인할 수 있다. 또한 개발된 GUI는 직관적으로 내용들을 표출하고 있으므로 치안 현장에서 일선 경찰의 의사결정에 효과적으로 도움을 줄 것으로 예상된다.



(a) 범죄 유형 및 범죄 심각도 시스템 분석 화면

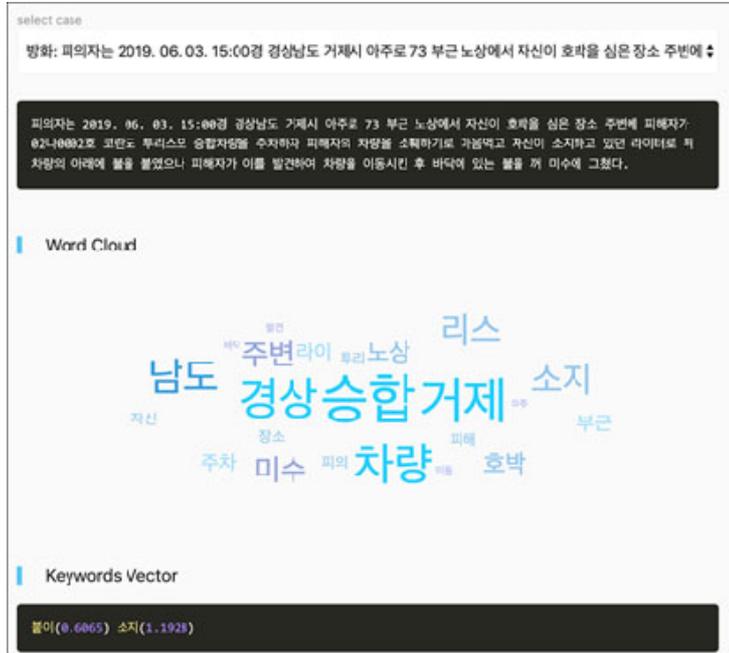


(b) 범죄 유형 및 범죄 심각도 예측 결과 화면

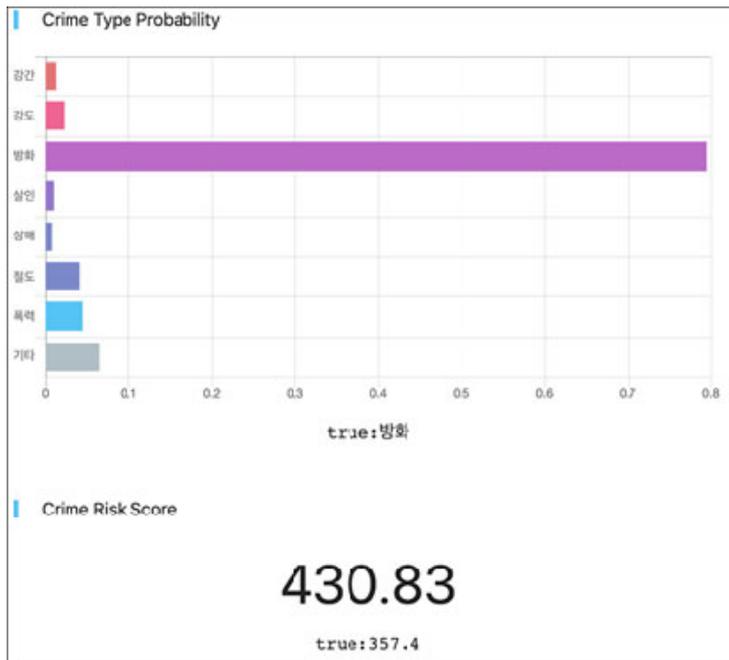
<Figure 5> 절도 관련 범죄사실이 입력된 GUI 기반 개발 시스템

<Table 5> 텍스트 기반 절도 범죄 사실 입력 데이터

피의자는 2016. 1. 31. 15:30경 광주광역시 광산구 수동로xxx길 피해자 xxx이 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 감시가 소홀한 틈을 이용하여 계산대 위에 있던 피해자 소유 시가 13,500원 상당의 보합 담배 1갑, 레종 담배 1갑, 뽕비우스 담배 1갑 총 3갑을 상의 잠바주머니에 넣고 마그네틱이 훼손된 카드를 주어 피해자가 결제가 되지 않는다 하면 돈을 가져오겠다고 나가는 방법으로 절취하였다.



(a) 범죄 유형 및 범죄심각도 시스템 분석 화면



(b) 범죄 유형 및 범죄심각도 예측 결과 화면

<Figure 6> 방화 관련 범죄사실이 입력된 GUI 기반 개발 시스템

<Table 6> 텍스트 기반 방화 범죄 사실 입력 데이터

피의자는 2019. 06. 03. 15:00경 경상남도 거제시 아주로 73 부근 노상에서 자신이 호박을 심은 장소 주변에 XXX가 02나0002호 코란도 투리스모 승합차량을 주차하자 피해자의 차량을 소취하기로 마음먹고 자신이 소지하고 있던 라이터로 위 차량의 아래에 불을 붙였으나 피해자가 이를 발견하여 차량을 이동시킨 후 바닥에 있는 불을 꺼 미수에 그쳤다.

```

# 범죄 위험도 예측 평가
precision: 0.82
recall:    0.86
f-score:   0.83

# 범죄 유형 예측 평가 (전체)
accuracy:  0.85
precision: 0.82
recall:    0.86
f-score:   0.83

```

<Figure 7> 개발 시스템의 전체 예측 성능

5. 개발 결과 및 성능 분석

본 논문에서는 경찰의 사건 접수 초기에 범죄의 심각성을 인지하고 범죄의 유형을 특정하여 효과적으로 초동대응할 수 있도록 정보를 제공해 주는 범죄 유형 예측 및 위험도 추론 기술을 개발하여 성능을 검증하였다. 개발된 시스템은 범죄 사실관련 텍스트 데이터를 바탕으로 범죄의 유형과 심각성을 나타내는 위험도를 예측한다. 범죄 위험도 정의를 위해 범죄 죄종 및 피해자의 범죄 피해정도에 따른 범죄 심각도 산출식을 선행 개발하였다. 개발한 산출식은 범죄 유형의 강력정도와 피해자의 범죄 취약정도 및 범죄 취약정도에 따라 심각도를 수치적으로 산출한다. 산출된 심각도는 높은 값을 가질수록 범죄의 위험도가 높은 것을 의미한다. 제안된 기술은 상기 심각도 산출식을 통해 산출된 위험도를 텍스트기반으로 학습하여 추론하는 것이 가능하다. 또한 범죄 텍스트 기반 사건 내용을 바탕으로 범죄 유형을 예측 분류할 수 있다. 예측 검증된 예측 성능은 범죄 유형 및 범죄 위험도 모두 0.8이상의 우수한 값을 달성하는 것이 가능하다. 제안된 기술은 다양한 결과를 확인할 수 있도록 GUI기반의 시스템으로 구현되어 치안 현장인력이 손쉽게 시스템을 사용할 수 있도록 개발하였다. 따라서 경찰의 사건 접수 초기 해당 시스템을 이용하여 범죄의 유형 및 위험도관련 정보를 획득할 수 있으며, 이를 통해 범죄 유형 특정, 현장 인력 배치 등 초동대응을 보다 효과적으로 수행하는 것이 가능하다.

참고문헌(References)

- [1] Baek MS, Kwak SW, Jung JY, et al. 2019. Implementation Methodologies of Deep Learning-Based Signal Detection for Conventional MIMO Transmitters. *IEEE Transactions on Broadcasting*, 65(3), 636-642.
- [2] Jeon JH, Lee KC. 2020. Standardization Trends on Artificial Intelligence in Medicine. *ETRI Trend*, 35(2), 113-126.
- [3] Kim Y, Hwang H, Shin J. 2020. Optimization of object detection and inference time for autonomous driving. *The Journal of Korean Institute of Communication and Information Sciences*, 45(4), 722-726.
- [4] Kim HJ, Sung JY. 2019. A stock price prediction based on recurrent convolution neural network with weighted loss function. *KTSDE*, 8(3), 123-128.
- [5] Choi HS, Shon YM. 2017. Case analysis and characteristics of the convergence between artificial intelligence and art creation activities. *Korea Science & Art Forum*, 28, 289-299.
- [6] Park JK, Park EY. 2019. Development of scalable big data storage system using network computing technology. *The Journal of Korean Institute of Communication and Information Sciences*, 23(11), 1330-1336.
- [7] Jeong JH, Jo KI, Kang HJ, et al. 2019. The case study on big data for the development of the footwear industry. *KIICE*, 23(2), 104-106.
- [8] Jang J. 2020. *Smart Policing*. Parkyoungsa Press.
- [9] Kim YS. 2016. Information and communications technology in the field of public security: crime prevention and response system. *Journal of convergence security*, 16(6), 23-32.
- [10] Kwon H, Lee J. 2017. Cases of using Big Data in the field of security and designing a desirable public research organization. *KOTIS*, 1303-1320.
- [11] Jang J, Kim H. 2020. Research of investigation technology to create a network using unstructured data: Focusing on voice phishing response support. *The Journal of Police Science*, 20(2), 93-117.
- [12] Heo JY, Kim JY, Moo TH. 2018. Predicting crime risky area using machine learning. *Journal of Korean Association of Geographic Information Studies*, 21(4), 64-80.
- [13] Lee M, et al. 2019. Early prediction of crime cases based on machine learning models. *Journal of the Korean Industrial Engineering Society Fall Conference*. Korean Institute of Industrial Engineers, 515-532.
- [14] Baek MS, Park W, Park J, et al. 2021. Smart Policing Technique With Crime Type and Risk Score Prediction Based on Machine Learning for Early Awareness of Risk Situation. *IEEE Access*, 9, 131906-131915, doi: 10.1109/ACCESS.2021.3112682
- [15] Kim HJ, Cho S, Kang P. 2014. KR-WordRank : An Unsupervised Korean Word Extraction Method Based on WordRank. *Journal of the Korean Institute of Industrial Engineers*, 40(1), 18-33.
- [16] Kritikopoulos A, Sideri M, Varlamis I. 2007. Wrodrandk: a method for ranking web pages based on content similarity. *BNCOD07*, 1-9, Glasgow UK.
- [17] You ES, Choi GH, Kim SH. 2015. Study on extraction of keywords using TF-IDF and Text structure of novels. *Journal of the Korea Society of Computer and Information*, 20(2), 121-129.
- [18] Park DS, Kim HJ. 2018. A proposal of joint vector for semantic factor reflection in TF-IDF based keyword extraction. *Journal of KIIT*, 16(2), 1-16.