

감성분석을 이용한 재난문자의 긴급성 평가

주 상 임*, 강 현 주*, 오 승 희^o

Evaluating Urgency Levels of Emergency Alerts through Sentiment Analysis

Sang-Lim Ju*, Hyunjoo Kang*, Seung-Hee Oh^o

요 약

재난문자는 긴급한 상황에서 신속하고 정확하게 재난 정보를 전달할 수 있는 중요한 수단이지만, 재난문자 형태로 국민에게 전달하는 정보의 유형이나 건수가 증가함에 따라 국민의 피로감 증가와 신뢰도가 하락하고 있다. 따라서 본 논문에서는 재난문자 서비스가 얼마나 적절한 경보 정보를 전달하고 있는지를 평가하기 위해 감성분석 접근 방식을 활용하여 재난문자의 긴급성 평가 모델을 제안했다. 평가 모델은 대표적인 4가지 자연어 처리 알고리즘을 활용하여 개발되었다. 또한 긴급성 평가 기준과 방법론을 제시했다. 실험 결과에서는 4가지 알고리즘 중 BERT 알고리즘에 기반한 긴급성 평가 모델이 정확도 98%로 가장 뛰어난 학습 성능을 보였다. 그리고 학습된 4가지 알고리즘 기반 긴급성 평가 모델을 통해 2022년에 발령된 재난문자의 긴급성을 분류하여 평가했다.

키워드 : 감성분석, 긴급성 평가, 자연어 처리, 재난문자

Key Words : Emergency Alerts, Natural Language Processing, Sentiment Analysis, Urgency Level evaluation

ABSTRACT

Emergency alerts are crucial means for promptly and accurately conveying disaster information in urgent situations. However, as the type and frequency of emergency alerts delivered to the public increase, there has been a rise in public fatigue and a decline in trust. To evaluate the appropriateness of the emergency alert service, this study proposes models that evaluate urgency levels for emergency alerts based on sentiment analysis. The models are developed using four representative natural language processing algorithms. Furthermore, criteria and methodology for evaluating urgency levels are presented. In the experimental results, among the four algorithms, the urgency level evaluation model based on the Bidirectional Encoder Representations from Transformers algorithm showed the best learning performance with an accuracy of 98%. And through the four learned algorithm-based urgency level evaluation models, the urgency of the emergency alerts issued in 2022 was classified and evaluated.

※ 본 연구는 행정안전부 재난안전 부처협력 사업의 지원을 받아 수행되었습니다. (과제번호: 20008820)

♦ First Author : Electronics and Telecommunications Research Institute, imaward@etri.re.kr, 정희원

o Corresponding Author : Electronics and Telecommunications Research Institute, seunghee5@etri.re.kr, 정희원

* Electronics and Telecommunications Research Institute, hjkang@etri.re.kr

논문번호 : 202308-065-B-RN, Received August 29, 2023; Revised October 13, 2023; Accepted October 25, 2023

1. 서론

CBS (Cell Broadcast Service) 기술 기반의 재난문자는 긴급한 상황이나 재난 상황에서 중요한 경보 정보를 국민에게 신속하고 정확하게 전달할 수 있는 서비스이다. 특정 지역만을 타겟으로 재난 정보를 전달할 수 있으며, 해당 지역의 기지국은 자신에게 접속하고 있는 모든 단말에게 사용자 식별이나 통신 설정 없이 동시에 재난문자를 전달할 수 있는 장점이 있다. 이러한 장점으로 우리나라에서는 자연 재난, 사회 재난 등에 대한 경보 전달 수단으로 재난문자 서비스를 활발하게 사용하고 있다. 특히 중앙 정부에서 지자체로 송출 권한이 확대되면서 재난문자의 사용이 크게 증가하였다.

그림 1은 2012년부터 2022년까지 발령된 재난문자 건수를 보여준다^[1]. 그림에서 보이는 것처럼, 2017년 지자체로 송출 권한이 확대되면서 재난문자 발령 건수가 2배 이상 증가했다. 2020년에는 COVID-19에 대응하기 위해 재난문자를 적극적으로 활용함에 따라 발령 건수가 2019년에 비해 약 60배 이상 증가했다. 재난문자의 절대적인 건수가 증가함에 따라 국민이 받는 피로도가 증가했고, 신뢰도가 감소하는 문제가 제기되었다^[2,3]. 이로 인해 인터넷 상에서는 재난문자 수신을 거부하는 방법이 활발히 공유되고 있는 상황이다^[4]. 재난문자는 위급 상황에서 중요한 정보를 빠르게 전달, 수신할 수 있는 수단이다. 따라서 발령주체인 정부 기관 및 지자체는 불필요한 재난문자 발령이나 과잉 경보를 최대한 지양하고, 긴급한 상황에 대해서 명확하고 정확한 정보를 전달해야 한다. 이를 위해 재난문자의 내용 및 발령 기준을 명확하게 정립할 필요가 있다.

이러한 동기에 따라 본 논문에서는 기존에 발령되었던 재난문자가 국민에게 얼마나 적절한 경보 정보를 전달했는지 평가하는 것을 목적으로 한다. 이를 위해 자연

어 처리 기법 중 하나인 감성분석^[5]을 이용하여 재난문자의 긴급성을 평가한다. 감성분석은 텍스트 데이터에서 긍정적, 부정적, 또는 중립적 감성을 추출할 수 있다. 주로 서비스나 작품의 리뷰 평가, 사용자 피드백 관리 등에서 사용된다^[6-10]. 최근에는 금융이나 심리관련 분야에서도 활용되고 있다^[11,12].

이 논문의 목적은 경보 정보 전달에 관련하여 단어나 문장에 대해 긴급성을 평가할 수 있는 접근방법을 개발하는 것이지만, 우리가 아는 한 이러한 긴급성 평가에 대한 내용을 다루는 기존의 문헌은 없는 것으로 조사되었다. [13]에서는 자연어 처리 기술을 이용하여 단순히 COVID-19에 관련된 재난문자의 제공 목적을 분류한다. 따라서 이 논문에서는 긴급성 수준을 정의할 수 있는 기준을 정립하고, 그 기준에 따른 학습 모델을 설계하여 재난문자의 긴급성 레벨을 평가한다. 이를 위해 전통적인 모델부터 최근 가장 뛰어난 성능을 보이는 모델로, 서로 다른 접근 방식을 가진 4가지 자연어 처리 알고리즘 LR (Logistic Regression)^[14], LSTM (Long Short-Term Memory)^[15], CNN (Convolutional Neural Network)^[16], BERT (Bidirectional Encoder Representations from Transformers)^[17]을 활용한다. 이 알고리즘에 기반한 긴급성 평가 모델을 설계하고, 모델의 성능을 평가하여 제시한다.

이 논문에서 제안하는 긴급성 레벨은 5가지로 구분된다. 레벨은 0~4까지 있으며, 4는 매우 높은 긴급성 레벨의 재난문자를 의미한다. 3은 상당한 위험 또는 긴급성을 가진 내용의 재난문자이고, 2는 중간 수준의 위험, 1은 낮은 위험 또는 관심을 요구하는 상황을 알리는 재난문자를 의미한다. 0은 사실상 경고나 긴급 상황을 의미하지 않는 것으로 정의했다. 여기에서, 긴급성 레벨에서 레이블 0에 해당하는 재난문자가 앞서 서술한 사용자가 재난문자를 차단하게 되는 원인 중 하나로 고려될 수 있다. 이 레이블에 해당하는 재난문자들은 긴급성이 없기 때문에, 불필요한 정보를 제공하거나, 과도한 알림(over-alerting)이거나, 메시지의 표현 방법을 수정할 필요가 있는 경우로 간주될 수 있다. 즉, 이러한 메시지들은 사용자의 피로감을 유발하며, 재난문자에 대한 전반적인 신뢰성을 감소시킬 수 있다.

본 논문의 구성은 다음을 따른다. 먼저 2장에서는 긴급성 평가 모델을 설계하기 위한 데이터 수집 방법과 데이터 셋에 대해 설명한다. 그리고 긴급성 평가를 위한 훈련 데이터의 레이블링 방법을 제안한다. 3장에서 긴급성 평가 모델을 설계하고, 4장에서 설계한 모델의 성능 평가와 2022년 재난문자에 대한 긴급성 평가 결과를 제시한다. 마지막으로 5장에서 결론을 맺는다.

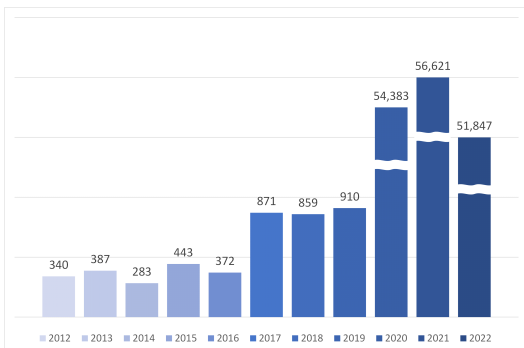


그림 1. 2012~2022 재난문자 발령 이력 [1]
Fig. 1. 2012~2022 Emergency alert history.

II. 데이터 수집과 레이블링

2.1 데이터 수집

재난문자의 긴급성 평가를 위해, 공공데이터포털에서 '행정안전부 재난문자방송 발령현황' 오픈 API를 통해 발령된 재난문자 데이터를 수집하였다. 본 연구에서 활용한 훈련 데이터는 2017년, 2018년, 2019년, 그리고 2021년에 발령된 재난문자들이다. COVID-19 관련 재난문자는 훈련 데이터에서 제외하였다. 앞서 그림 1에서 설명한 것처럼 우리나라는 COVID-19 대응을 위해 재난문자를 적극적으로 활용했는데, COVID-19 대응에 효과적인 측면도 있었지만 이로 인해 재난문자의 사용이 상당히 무분별해졌다. 특히 2020년에는 지역명과 확진자 수만을 나열하거나 확진자 동선을 제공하는 등 특수한 케이스들이 많다. 더불어, 기상과 관련된 재난문자에도 COVID-19 관련 행동 지침이 포함되는 등 일관성 없는 데이터 분포가 나타났다. 이러한 이유로, 이런 데이터들은 긴급성 평가 모델의 학습에 부정적인 영향을 미칠 가능성이 높다고 판단되었다. 2022년이나, 2023년 현재로서는 이러한 형태의 재난문자 발송이 거의 이루어지지 않아, 최근 재난문자의 긴급성을 평가하기 위한 훈련 데이터로서의 가치가 현저히 떨어진다. 따라서 본 논문에서는 이러한 데이터를 제외한다. 이러한 이유로 2020년 재난문자는 COVID-19와 관련된 내용이 대다수를 차지하기 때문에 전체적으로 제외하였다. 2021년에 발령된 재난문자 중에서도 COVID-19 관련 문구가 포함된 것들은 제외하였으며, 주로 자연재난, 사회재난, 실종정보와 관련된 데이터를 중점적으로 수집하였다.

2.2 훈련 데이터의 레이블링

우리의 문헌 조사에서 문장이나 단어의 긴급성을 정량적으로 평가할 수 있는 기존의 참고문헌을 찾을 수 없었다. 이에 따라, 본 논문에서는 단어의 사전적 의미와 저자의 주관적인 관점에서 문장의 긴급성 점수와 그에 따라 긴급성 레벨을 설정하는 기준과 방법을 제안한다. 제안하는 긴급성 레벨은 아래 식을 통해 도출된다.

$$s_n = \alpha_i + \max_{j=\{0, \dots, 11\}} (\beta_j), \quad (1)$$

$$l_n = \text{ceil}(s_n/2),$$

식에서 s_n 은 n 번째 훈련 데이터의 긴급성 점수를 의미하며, l_n 은 긴급성 레벨을 의미한다. 기본적으로 n 번째 훈련 데이터는 레벨 l_n 으로 레이블된다. 위 식에 의해 일관적으로 레이블링을 1차적으로 수행하지만, 수기로

표 1. α 에 대한 긴급성 점수
Table 1. Urgency score for α .

α_i	text	Score
α_0	Warning	4
α_1	Watch, Special Notice	3
α_2	Earthquake	4
α_3	Lifting or None of the above	0

입력된 문구로 이루어진 데이터 특성 상, 동일한 목적에 대해 다양한 표현으로 작성된 경우가 있기 때문에 연구자의 검토로 2차 레이블링 수정이 이루어진다.

식 (1)에서 α 는 주의보(Watch), 경보(Warning)와 같은 기상특보에 대한 점수를 나타낸다. 재난문자의 내용에 기상 재난에 관한 주의보나 경보의 발령 사항이 포함되어 있으면, 대부분의 사람들은 해당 단어만으로도 해당 재난의 위험도나 긴급성을 판단하는 데 큰 문제가 없다고 판단하였다. 따라서, 경보에 관한 재난문자의 경우 $\alpha_0 = 4$ 로 설정했고, 주의보에 대해서는 $\alpha_1 = 3$ 점을 매겼다. 지진(Earthquake)의 경우 경보나 주의보로 구분되지 않으므로, 지진에 관한 재난문자의 α 값 또한 $\alpha_2 = 4$ 로 결정하였다. 그리고, 경보나 주의보의 해제(Lifting)와 관련된 재난문자 또는 이러한 특보가 포함되지 않은 재난문자의 경우는 $\alpha_3 = 0$ 으로 설정하였다.

β 는 재난문자의 내용에서 긴급성을 표현하는 문구에 따른 점수를 나타낸다. 문구의 의도와 수신자에게 전달하고자 하는 메시지의 중요도에 따라 점수를 구분하였다. 재난에 대응하여 수신자에게 즉각적인 행동을 요구하는 문구는 긴급성이 가장 높다고 판단하여 가장 높은 점수를 설정했다. 예를 들자면, '대피하시기 바랍니다', '대피하십시오', '안전지대로 이동 바랍니다' 등은 재난 영향권에서 즉각적인 대피를 요청(Call for immediate

표 2. β 에 대한 긴급성 점수
Table 2. Urgency score for β .

β_i	text	Score
β_0	Call for immediate evacuation	7
β_1	Prohibition / Ban	3
β_2	Evacuation	3
β_3	Caution	2
β_4	Restraint / Refrain	2
β_5	Attention / Beware	1
β_6	Preparation / Anticipation	1
β_7	Control	1
β_8	Prepare / Readiness	1
β_9	Missing	1
β_{10}	Detour	1
β_{11}	None of the above	0

evacuation)하는 문구로써 보통 ‘즉시’나 ‘신속하게’와 같이 사용되며, β 값을 7점으로 설정했다. 다음으로, 수신자에게 어떠한 행동을 하도록 강하게 권고하는 의미를 내포하는 단어를 포함하는 문구일수록 높은 점수를 매겼다. 반면에 수신자에게 어떠한 행동을 하도록 권고하는 정도가 약하거나 단순히 정보를 전달하는 문구에 대해서는 낮은 점수를 설정했다. ‘금지(Prohibition/Ban)’는 법이나 규칙이나 명령 따위로 어떤 행위를 하지 못하도록 함을 뜻하는 말로써, 어떠한 행위를 금지하는 문구를 포함하는 재난문자는 강한 어조로 판단하여 β 값을 3점으로 설정했다. [18]에서 보듯이 ‘금지’를 강한 어조로 받아들이고 있음을 확인할 수 있다. 기상특보 발령 시 전달되는 일반적인 행동요령으로 선박 대피나 위험지역 대피와 같은 문구가 포함되는데, 이와 같은 문구에서 ‘대피’에 대해서 β 값으로 3점을 설정했다. 이런 설정을 통해, ‘주의보’와 함께 사용될 때 긴급성 레벨은 3이 되고, ‘경보’와 함께 사용될 때 긴급성 레벨은 4가 될 수 있다. 이외 ‘주의(Caution)’와 ‘자제(Restraint / Refrain)’를 권하는 문구에 대해서는 2점으로 설정했고, ‘유의(Attention / Beware)’, ‘대비(Preparation / Anticipation)’, ‘준비(Prepare / Readiness)’를 권하는 문구와 교통에 대한 ‘통제(Control)’, ‘우회(Detour)’를 알리는 문구, 노약자나 어린이의 실종(Missing)에 대한 재난문자에 대해서는 1점으로 설정했다. 그리고 모든 경우에 해당하지 않는 경우 0점으로 설정했다. 긴급성 점수에 대해서 표 1과 표 2에서 정리하여 보여준다.

식 (1)에서 명시된 것처럼 이 논문에서는 한 문구에 여러 행동에 대한 권고 내용들이 포함되더라도, 가장 높은 긴급성 점수만을 레이블링에 사용한다. 예를 들어 행정안전부예규 제14호 휴대폰 재난문자방송 표준문안에서 태풍주의보에 대한 문구는 다음과 같다:

[행정안전부] 오늘 00시 00지역 태풍주의보 태풍 관련 정보를 공유하고, 하천, 해안가 등 위험지역 접근금지, 논밭 관리행위 자제 등 안전에 주의 바랍니다.

이 문장에는 ‘주의보’가 포함되므로 $\alpha_1 = 3$ 이 고려된다. 그리고 접근금지, 관리행위 자제, 안전에 주의와 같은 문구 때문에 $\beta_1, \beta_3, \beta_4$ 가 고려된다. 여기서 $\beta_1 = 3, \beta_3 = \beta_4 = 2$ 이므로, 식 (1)에 따라 긴급성 점수 $s_n = 6$ 이고, 긴급성 레벨은 $I_n = 3$ 으로 결정된다. 실제 재난문자에 대한 긴급성 레벨을 계산한 결과의 예는 다음과 같다:

- i) 악양면 미점리 뒷산 산불발생으로 진화 중입니다. 미서, 미동, 대축, 소축, 하동읍 먹점마을 주민들은 즉시 안전지역으로 대피 바랍니다. $\Rightarrow I_n = 4$
- ii) 오늘 05시 50분 강원(강원 북부 산지) 호우경보, 산사태·상습 침수 등 위험지역 대피, 외출 자제 등 안전에 주의 바랍니다 $\Rightarrow I_n = 4$
- iii) 오늘 21시부터 한파주의보 발효로 급격한 기온 강하가 예상되오니 외출 자제, 수도관 동파예방, 농작물 냉해 피해가 없도록 조치하시기 바랍니다. $\Rightarrow I_n = 3$
- iv) 우리 지역에 강풍, 대설예비 특보가 발표 중으로 1월 28일 오전 발효 예정입니다. 주변 시설물 등을 점검하고 사전대비하시기 바랍니다. $\Rightarrow I_n = 2$
- v) 밤 사이 내린 눈으로 도로 결빙 등 블랙아이스로 인한 사고가 우려되오니 안전운전에 각별히 유의하시길 바랍니다. $\Rightarrow I_n = 1$
- vi) 겨울철 저체온증 예방법. 얇은 옷 여러 겹 끼입기, 손발 보온 유지하기, 핫팩 등 보온용품 챙기기, 면역력 관리하기, 최대한 휴식 많이 하기. 다 같이 실천합시다. $\Rightarrow I_n = 0$

표 3은 각 긴급성 레벨의 의미를 보여준다. I_n 이 4 이상이면 매우 높은 긴급성 레벨의 재난문자를 의미한다. 이 재난문자에는 즉시 대피하거나 반드시 지시사항을 준수할 것을 요청하는 내용이 포함된다. $I_n = 3$ 의 재난문자는 상당한 위험 또는 긴급성을 알리며, 수신자는 권고하는 조치를 즉시 준비하거나 수행해야 할 것이다. $I_n = 2$ 의 재난문자는 중간 수준의 위험 또는 긴급 상황을 나타낸다. 수신자는 지시사항을 확인하고 필요한 조치를 준비하거나 시작해야 할 수 있다. $I_n = 1$ 의 재난문자는 낮은 위험 또는 관심을 요구하는 상황을 알리는 내용이다. 이러한 메시지는 주의를 기울이도록 권고하지만, 즉각적인 대응은 요구되지 않을 수 있다. $I_n = 0$ 의 재난문자는 사실상 경고나 긴급 상황을 의미하지 않는다. 이는 과도한 알림(Over-alerting)인 경우에 해당될 수도 있고, 권고 표현 방식의 변경이 필요한 내용일 수도 있다. 주로 단순한 정보성 메시지가 $I_n = 0$ 으로 라벨

표 3. 긴급성 레벨: 0~4
Table 3. Urgency level: 0~4.

I_n	Urgency Level
above 4	Very High or Critical
3	High
2	Moderate
1	Low
0	None or Negligible

링된다.

표 1과 표 2, 표 3, 식 (1)은 이 논문에서 제안하는 긴급성 점수 및 그 평가 규칙이다. 독자는 본 논문의 방법론에 근거하여 다른 긴급성 점수나 긴급성 레벨 결정 규칙을 제안할 수 있을 것이다.

III. 긴급성 평가 모델 설계

긴급성 평가 모델은 II장에서 설명한 것처럼 수집 및 레이블링된 데이터에 대해 텍스트 전처리를 수행하고, 4가지의 서로 다른 접근 방식을 가진 학습 모델에 기반하여 설계된다. 긴급성 평가 모델은 Windows 11환경에서 python 3.8과 백엔드 엔진으로 Tensorflow 2.10.0을 사용하여 설계된다.

3.1 텍스트 전처리

재난문자는 90자로 제한된 용량에서 담당자가 직접 입력하여 발령하는 방식으로, 띄어쓰기가 무시된 경우가 종종 발생한다. 예를 들면, ‘[울릉군청] 금일(7일) 강풍주의보, 풍랑경보 발효 강풍에 의한 낙석,지붕탈락 등 낙하물사고의 위험이 있으니, 해안가 접근금지 및 외출자제하여 주시기 바랍니다.’는 90자 내에서 최대한 많은 정보를 담기 위해 띄어쓰기가 무시되어 있다. 일관적이고 올바른 학습을 위해서라면, 이러한 오류를 바로 잡아주어야 한다. 이를 위해 이 논문에서는 네이버 맞춤법 검사기를 이용한 파이썬용 한글 맞춤법 검사 라이브러리 ‘py-hanspell’을 활용한다.

또한 모든 재난문자의 앞에는 발령기관 명칭(예: [울릉군청], [행정안전부] 등)이 작성되어 있다. 이러한 부분도 모델이 학습하는 데 있어서 불필요한 부분이기 때문에 제거할 필요가 있다.

표 4. 알고리즘 장단점 비교
Table 4. Strengths and weaknesses by algorithm.

Algorithm	Strengths	Weaknesses
LR	Simple and fast for small datasets	Not suitable for non-linear data
LSTM	Effective for sequential data	High risk of overfitting
CNN	Fast training and inference times	Assumes spatial hierarchies, not optimal for sequences
BERT	Captures context in both directions	Requires substantial computational resources

3.2 긴급성 평가 모델 설계

긴급성 평가 모델 설계를 위해 LR, LSTM, CNN, BERT 알고리즘을 고려한다. LR은 전통적인 기계 학습 알고리즘으로, 텍스트 분류 작업에서 널리 사용된다. 감성분석에 있어서 기본적으로 사용되는 모델이다. LSTM은 RNN의 변형된 모델로, 시퀀스 데이터의 패턴을 학습하는 데 특화되어 있다. 텍스트 데이터와 같은 시퀀스 데이터에서 적합하다. CNN은 컴퓨터 비전 분야에서 주로 사용되지만, 지역적인 패턴을 학습하는 특성으로 NLP 분야에서도 활용될 수 있다. 마지막으로 BERT는 유명한 사전 훈련된 언어 모델 중 하나로, Fine-tuning 방식을 사용하여 다양한 NLP 작업에 적용 가능하다. 표 4는 4가지 알고리즘의 장단점을 비교하여 보여준다.

본 논문은 위 4가지 모델을 통해 긴급성 평가에 대한 다양한 접근 방식을 제공하고, 각각의 알고리즘이 긴급성 평가에 어떤 영향을 미치는지 비교 분석하여 제시한다. 그림 2는 본 논문에서 고려하는 4가지 모델에 기반한 긴급성 평가 모델의 흐름도를 보여준다.

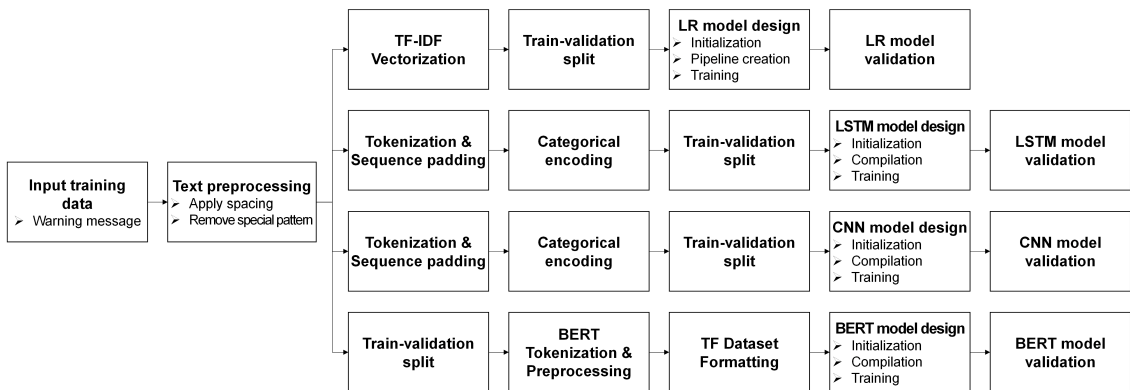


그림 2. 긴급성 평가 모델
Fig. 2. Model for urgency level evaluation.

3.2.1 Logistic regression기반 긴급성 평가 모델

LR은 통계적 분류 모델 중 하나로, 주어진 입력 특성을 바탕으로 각 label에 속할 확률을 예측한다. 본 논문에서는 재난문자의 긴급성을 다중 클래스 분류 문제로 접근하여, LR을 적용하였다. 재난문자 내용을 모델이 학습할 수 있는 숫자 형식으로 변환하기 위해 TF-IDF (Term Frequency-Inverse Document Frequency) 방식을 사용하여 텍스트 전처리된 문장을 벡터화하였다. TF-IDF는 문장 내에서 단어의 중요성을 평가하는 기중치 방식으로, 단순한 단어 빈도보다는 단어의 중요성을 반영하여 특징을 추출할 수 있다. 이렇게 TF-IDF 방식으로 벡터화된 데이터를 LR 모델에 입력하여, 긴급성 레벨을 예측하도록 모델을 훈련시켰다. 훈련 데이터와 검증 데이터의 비율은 각각 전체 데이터의 75%, 25%이다.

3.2.2 LSTM기반 긴급성 평가 모델

텍스트 데이터를 모델에 입력하기 위해 Tokenizer를 이용하여 텍스트를 토큰으로 변환하였다. 다음으로, LSTM은 모든 입력 데이터가 동일한 길이를 가져야 한다. 재난문자는 최대 길이가 90글자로, 전달하는 정보의 양에 따라 길이가 서로 다르다. 따라서 전체 문장의 길이를 최대 문장 길이와 동일하게 맞추기 위해 패딩 처리하였다. 토큰화된 단어를 벡터 공간에 임베딩하고, LSTM 층에서 텍스트 시퀀스의 패턴을 학습하였다. 모델을 학습하는 과정에서 과적합(Overshooting)이 발생했으며, 이를 해결하기 위해 SpatialDropout1D를 통해 임베딩 레이어의 출력에 대해 공간 드롭아웃(30%)을 적용했다. 다음으로 LSTM 레이어 내에서 드롭아웃(50%)과 순환 드롭아웃(50%)을 적용했고, LSTM 레이어의 기중치에 L2 정규화를 적용하여 과적합을 방지하였다.

3.2.3 CNN기반 긴급성 평가 모델

CNN은 LSTM과는 대조적으로 긴 시퀀스의 패턴을 학습하는 데는 덜 효과적이지만, 짧은 범위의 지역적 패턴 인식에는 뛰어난 것으로 알려져 있다. 텍스트 데이터에서 CNN은 주어진 윈도우 크기 내에서의 연속된 단어나 문자의 패턴을 인식한다. 본 논문에서는 1차원의 컨볼루션 연산을 수행했다. 윈도우 크기는 5로 설정했고, GlobalMaxPooling1D를 통해 고차원의 피쳐맵을 저차원으로 압축하였다.

3.2.4 BERT기반 긴급성 평가 모델

BERT에 기반한 긴급성 평가 모델을 설계하기 위해

Hugging Face의 Transformers 라이브러리를 사용했다. 이 라이브러리는 다양한 사전 훈련된 언어 모델을 제공하는데, 이 논문에서는 한국어로 작성된 재난문자를 다루기 때문에 BERT Tokenizer로 ‘bert-base-multilingual-cased’ 모델을 사용하였다. ‘bert-base-multilingual-cased’ 모델은 104개의 다양한 언어에 대해 훈련된 BERT 모델로, 한국어 데이터에 대해 적합하다. Tokenizer를 사용하여 데이터를 BERT 입력 형식에 맞게 전처리하고, 이 데이터를 Tensorflow의 Dataset 객체로 변환하여 학습에 사용했다. 모델 컴파일 단계에서 Adam optimizer를 사용하여 학습 속도와 안정성을 향상시키고, SparseCategoricalCrossentropy 손실 함수로 다중 클래스 분류 문제를 처리하였다. 이때 BERT 모델의 학습률은 5e-5를 적용했다.

IV. 실험 및 평가 결과

앞서 서술했던 것처럼, 본 논문에서는 긴급성 평가 모델 설계를 위해 2017년, 2018년, 2019년, 2021년에 발령된 재난문자 데이터를 활용했다. 연도별 데이터 건수는 다음과 같다: 2017년 871 건, 2018년 859 건, 2019년 910 건, 그리고 2021년 4,780 건. 훈련 및 검증 데이터의 레이블별 비율은 4: 14.6%, 3: 31.1%, 2: 6.4%, 1: 38%, 0: 9.9%로, 2와 0 레이블의 데이터 수가 상대적으로 부족하다. 이 장에서는 위와 같은 비율을 가진 훈련 및 검증 데이터로 학습된 모델들의 성능을 평가한다. 또한 학습된 모델에 대해 2022년 발령된 재난문자 중 COVID-19와 관련된 재난문자를 제외한 6,489건을 테스트 데이터로 입력하여, 2022년에 발령된 재난문자는 어느 정도의 긴급성을 갖고 있는지 분석하고 개선 방안을 논의한다.

4.1 모델 성능 평가

모델의 성능 평가를 위해 Scikit-learn(sklearn)의 metrics 모듈에서 제공하는 classification_report를 사용한다. classification_report를 통해 분류 모델의 평가 지표인 정밀도(Precision), 재현율(Recall), f1-score, 정확도(Accuracy)를 분석한다.

정밀도는 모델이 양성(True positives)으로 예측한 것 중 실제로 양성에 속하는 비율이다. 즉, 예측한 결과 중에서 실제 레이블과 동일한 것의 비율을 의미한다. 잘못된 양성(False positives) 예측이 적을수록 값이 높다. 그림 3은 4가지 모델에 대해서 레이블별 정밀도를 보여준다. 모든 모델이 0.85 이상 정밀도를 가지는 것으로 분석되었다. 특히 BERT 기반 긴급성 평가 모델은

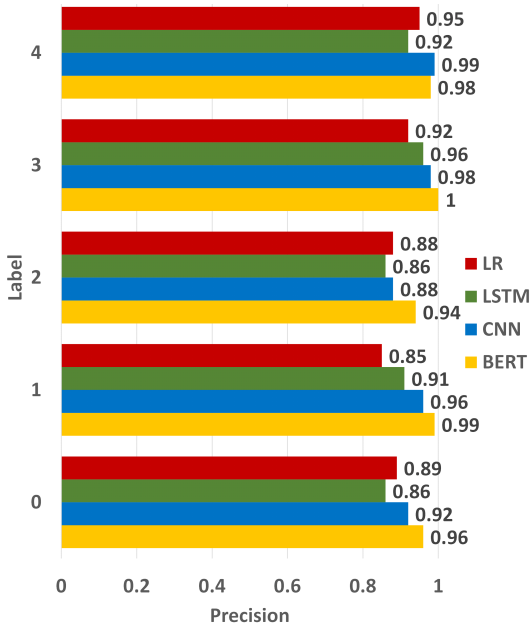


그림 3. 알고리즘 및 레이블에 따른 정밀도
Fig. 3. Precision by algorithm and label.

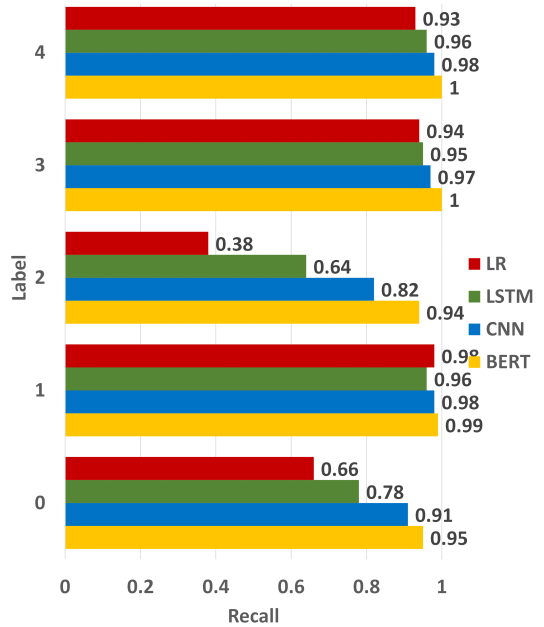


그림 4. 알고리즘 및 레이블에 따른 재현율
Fig. 4. Recall by algorithm and label.

다른 모델의 비해 더욱 높은 정밀도를 보였다. LSTM과 CNN 역시 0.9 이상의 높은 정밀도를 보이며 효과적인 성능을 나타냈다. LR 모델은 다른 딥러닝 기반 모델에 비해 상대적으로 낮은 정밀도를 보여주었다.

그림 4는 4가지 모델에 대해서 레이블별 재현율을 보여준다. 재현율은 실제 양성 클래스에 속한 데이터에 대해 모델이 양성으로 예측한 데이터 비율을 의미한다. 이를 통해 모델이 얼마나 많은 양성을 제대로 예측했는지를 확인할 수 있다. 재현율에 있어서도 BERT는 가장 우수한 성능을 보였다. 레이블 4와 3에 대해서는 1에 달하는 성능을 보였고, 레이블 2에서 가장 낮은 0.94 재현율을 달성했다. LSTM과 CNN 모델 또한 대체로 높은 재현율을 보였으나, BERT에 비해 일부 레이블에서는 약간의 성능 차이를 보였다. 특히, 레이블 0과 2에서는 BERT가 다른 모델들에 비해 상당한 성능 향상을 보였다. LR 모델은 특히 레이블 2에서 낮은 재현율을 보이며, 이는 해당 모델이 복잡한 패턴이나 불균형한 데이터 분포를 잘 파악하지 못하는 것으로 고려된다.

그림 5는 4가지 모델에 대해서 레이블별 f1-score를 보여준다. f1-score는 정밀도와 재현율의 조화 평균으로, 두 지표 모두를 고려하여 전체적인 모델의 성능을 평가하는 데 사용된다. 특히 이 논문에서 다루는 재난문자 데이터처럼 불균형한 데이터셋에서 모델 성능을 평가할 때 유용하다. 그림에서 보이는 것처럼, BERT 모

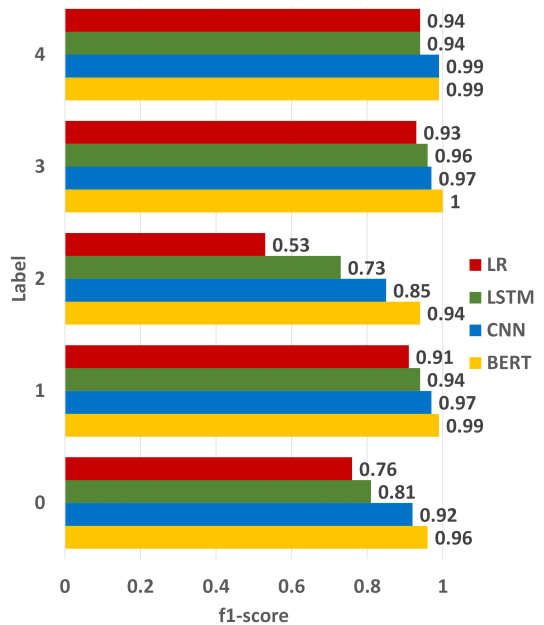


그림 5. 알고리즘 및 레이블에 따른 f1-score
Fig. 5. f1-score by algorithm and label.

델은 모든 레이블에서 가장 높은 f1-score를 보였다. 특히 레이블 2에서 BERT는 0.94의 f1-score로, 다른 모델들과 비교했을 때 가장 높은 성능을 보였다. BERT가 복잡한 패턴의 데이터나 불균형적인 데이터에 대해서

도 뛰어난 성능을 보이는 것으로 분석된다. LSTM과 CNN 모델도 전반적으로 높은 f1-score를 보였지만, BERT에는 미치지 못하는 성능을 나타냈다. LR 모델은 특히 레이블 2에서 0.53의 상대적으로 낮은 f1-score를 기록하였다. 데이터 숫자가 많은 레이블 1, 3, 4에 대해서는 대부분의 모델들이 높은 f1-score를 보였다.

표 5는 4가지 모델의 정확도를 보여준다. 이 정확도는 전체 예측 중에서 올바르게 예측된 비율을 의미한다. 결과를 보면, BERT와 CNN 같은 심층 학습 모델들이 재난문자 데이터 패턴을 학습하는 데 더 효과적이라는 것을 보여준다. 특히 그림 3~5와 표 5를 종합적으로 분석해보면, BERT가 가장 우수한 성능을 보여주는데, 이 이유는 주로 그 특징적인 구조와 작동 원리 때문이다. BERT는 양방향 인코더와 전이 학습(Transfer Learning)의 특성으로 문장의 전체 문맥을 이해할 수 있으며, 이는 특히 긴급성이 요구되는 상황에서 문맥적 뉘앙스를 파악하는 데 매우 효과적이다. 또한 'attention mechanism'을 사용하여 재난 상황의 긴급성을 나타내는 키워드와 같은 중요한 정보에 집중하여, 분류 정확도를 높이는 데 결정적인 역할을 한다. 반면, 전통적인 모델인 LR는 상대적으로 단순한 구조로 인해 복잡한 패턴, 불균형한 데이터에 대한 학습 능력이 제한적인 것으로 분석되었다. 전체적으로 모델 성능 평가 결과를 종합해보면, 재난문자 데이터에 대해 BERT 기반 긴급성 평가 모델이 다른 모델들에 비해 더욱 정확하게 재난문자의 긴급성을 분류할 수 있는 것으로 분석되었다.

표 5. 알고리즘별 정확도
Table 5. Accuracy by algorithm.

	LR	LSTM	CNN	BERT
Acc.	0.89	0.92	0.96	0.98

4.2 2022년 재난문자 긴급성 평가

학습된 모델을 통해 2022년 발령된 재난문자 중 COVID-19와 관련된 재난문자를 제외한 6,489건에 대하여 긴급성을 평가한다. 참고로, 2022년 재난문자 데이터에는 실종경보 문자 2,373건이 포함되어 있다.

그림 6에서 보이는 것처럼, 레이블 1은 2022년 재난문자 중 가장 빈도가 높은 레이블로 확인되었고, LR 모델이 4509건으로 가장 많은 메시지를 분류하였다. BERT는 3911건으로 가장 적게 분류하였다. 반면에 레이블 2에 대해서는 BERT가 386건으로 가장 많이 분류하였으며, LR, LSTM이 각각 182건, 202건이 레이블 2에 해당할 것이라고 분류했다. 앞서의 학습 평가 결과에서 BERT는 다른 모델들에 비해 전반적으로 높은 정

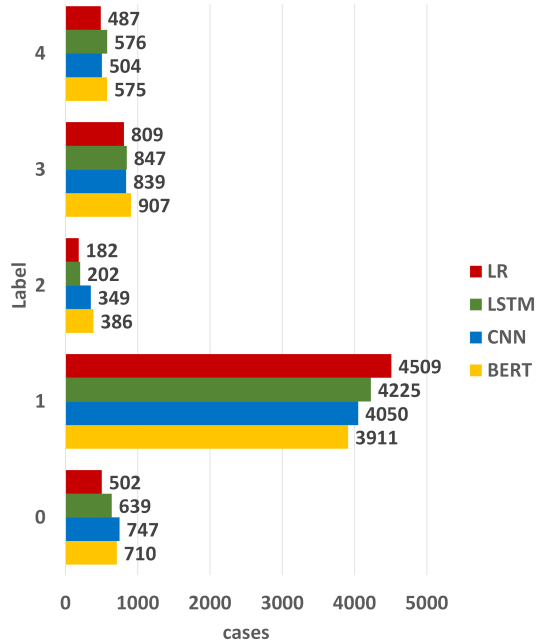


그림 6. 알고리즘 및 레이블에 따른 분류 건수
Fig. 6. Number of classifications by algorithm and label.

밀도와 재현율을 보였는데, 이러한 성능이 2022년에 발령된 재난문자 분류 결과에도 반영되어서 이러한 결과가 도출된 것으로 고려된다. 특히, 레이블 1에 대한 분류 결과를 보면, LR과 LSTM은 다른 레이블로 분류되어야 할 재난문자들을 1로 잘못 분류했을 가능성이 높다. 반면 BERT는 높은 정밀도를 바탕으로 다른 레이블로 분류되어야 할 재난문자를 1로 잘못 분류하지 않았을 것으로 고려할 수 있다. 레이블 2에 대해서는 LR과 LSTM이 각각 182건, 202건으로 비교적 낮은 건수를 기록했는데, 이는 두 모델이 레이블 2에 속하는 재난문자를 제대로 분류하지 못한 것으로 고려할 수 있다. BERT는 386건을 레이블 2로 분류했는데, 이 결과는 앞서의 학습 평가에서 BERT의 높은 재현율이 실제 데이터에도 잘 반영되었음을 의미한다고 볼 수 있다.

앞서 서론에서 서술했던 것처럼, 레이블 0에 해당하는 재난문자들은 긴급성이 없거나, 재난문자로 전달이 불필요한 과도한 알림(over-alerting)이거나, 재난정보 전달의 표현 방법을 수정할 필요가 있는 경우로 고려될 수 있다. BERT 모델을 이용하여 2022년에 분류한 결과, 레이블 0으로 판단된 재난문자는 총 710건이었다. 이것은 본 논문에서 설정한 긴급성 기준을 충족하지 못한 테스트 재난문자가 11% 임을 의미한다. 긴급성 판단 기준을 달리 설정한다면, 이 값은 변경될 수 있다. 다만 본 논문에서 설정한 긴급성 점수 및 레벨을 기준으로

이 결과에 대해 논의하자면, 11%에 해당하는 재난문자는 긴급하지 않거나, 표현 방법에 개선의 여지가 있는 것으로 해석될 수 있다. 앞서의 학습 평가 결과에서 BERT가 높은 정밀도와 재현율을 보였던 것을 감안하면, 이러한 분류 결과는 신뢰성 있게 간주될 수 있다. 따라서, 이러한 결과는 현재 재난문자 발령 방식 및 내용에 대한 개선이 필요함을 의미한다.

4.3 긴급성 평가 활용방안

재난문자 발령권한을 가진 기관 및 담당자의 수가 증가하고 활용도가 넓어지면서, 재난문자 발령에 있어서 긴급도에 대한 발령 기준이 모호해지고 있다. 과도한 알람이나 중복된 내용, 불필요한 정보를 최소화함으로써, 사람들이 실제로 긴급한 상황에서 적절하게 대응할 수 있도록 하는 것이 중요하다.

본 연구를 통해 개발된 긴급성 평가 모델을 재난문자 발령 담당자들에게 의사결정지원으로 제공되면 지자체 별로 통일된 긴급도 기반의 재난문자 발령을 지원할 수 있을 것이다. 특히, 실제 재난문자 발령 환경에서 BERT와 같은 알고리즘에 기반한 긴급성 평가 모델을 사용하여 작성한 재난문자의 긴급성을 평가하고, 평가 결과를 반영하여 재난문자를 발령하는 절차를 수립한다면 재난문자를 수신하는 국민들이 느끼는 감성적 중요도 측면에 효과적인 것으로 사료된다.

V. 결 론

본 논문은 자연어 처리 기법 중 하나인 감성분석 접근방식에 기반하여 재난문자에 대한 긴급성을 평가하는 모델을 개발하고 평가하였다. 네 가지의 알고리즘을 적용하여 설계한 각 모델은 90% 이상의 높은 정확도를 보였으며, BERT 기반의 모델은 특히 98%의 뛰어난 정확도를 기록하였다. 이 모델을 통해 2022년의 6,489건 재난문자를 분석한 결과 약 11%에 해당하는 재난문자를 발령하는데 있어서 내용 작성이나 발령 절차에 개선이 필요한 것으로 고려되었다. 물론 11%는 본 논문에서 제시하는 긴급성 평가 기준을 따랐을 때 수치로, 기준 정의에 따라 변경될 수 있다.

이 논문의 접근 방법은 다음과 같은 한계점을 가진다. i) 현재 단어나 문장에서 긴급한 정도를 평가할 수 있는 신뢰성 있는 기준이 없는 환경에서 저자의 주관에 따라 긴급성 점수가 정립되었다. 향후 연구에서는 평가 모델의 신뢰성을 강화시키는 작업이 이루어질 수 있을 것이다. ii) 훈련 데이터의 레이블링에 있어서 최대한 객관적인 기준을 설정하고 그 기준을 따르려고 노력했

지만, 언어의 복잡성으로 인해 모든 데이터가 설정한 기준에 부합하지 않을 수 있다. 이러한 부분에 대해서는 레이블링 하는 연구자의 주관이 개입하게 된다. 이러한 주관성이 결과에 영향을 줄 수 있다. iii) 사용자의 특정 상황에 따라 개인이 느끼는 긴급성이나 중요도는 다를 수 있다. 따라서 Geo-fencing과 같이 방송 영역을 특정할 수 있는 기술을 융합하여 사용자 맞춤형 정보를 제공하면서, 이 정보에 대한 긴급성을 평가하는 시스템이 연구될 수 있을 것이다.

향후 본 논문에서 제안하는 방법론을 기반으로 긴급성 평가 모델을 발전시킴으로써, 재난문자 문안 작성 기준과 재난문자 발령 절차의 효율적인 개선에 기여할 수 있을 것으로 기대한다.

References

- [1] S.-H. Oh, H. Kang, S.-L. Ju, W.-S. Jung, and M. Kim, *Technical report of emergency alert (version 2.0)*, 2023. (<https://doi.org/10.22648/ETRI.2023.B.000005>)
- [2] K. Park, K. Pyo, J. Jeong, W. Whang, and H. Kim, "An empirical analysis of the importance and satisfaction of disaster text broadcasting," *J. Korean Urban Manag. Assoc.*, vol. 35, no. 4, pp. 179-191, 2022. (<https://doi.org/10.36700/KRUMA.2022.12.35.4.179>)
- [3] S.-H. Han, "Effective use of cell broadcast service (CBS) by disaster type," *Crisisonomy*, vol. 16, no. 8, pp. 1-14, 2020.
- [4] The Financial News, <https://www.fnnews.com/news/202303211832433108>
- [5] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," *Artif. Intell. Rev.*, vol. 55, pp. 5731-5780, 2022. (<https://doi.org/10.1007/s10462-022-10144-1>)
- [6] A. A. Francisca, W. Chen, and N.-M. Henry, "Text-based emotion detection: Advances, challenges, and opportunities," *Eng. Reports*, vol. 2, no. 7, Jul. 2020. (<https://doi.org/10.1002/eng2.12189>)
- [7] N Pavitha, et. al., "Movie recommendation and sentiment analysis using machine learning," *Global Transitions Proc.*, vol. 3, no.

1, 2022.
(<https://doi.org/10.1016/j.gltp.2022.03.012>)

[8] J.-S. So and P.-S. Shin, "Rating prediction by evaluation item through sentiment analysis of restaurant review," *J. The Korea Soc. Comput. and Inf.*, vol. 25, no. 6, pp. 81-89, Jun. 2020.
(<https://doi.org/10.9708/jksoci.2020.25.06.081>)

[9] Y. Kim and S. R. Jeong, "Competitive intelligence in Korean ramen market using text mining and sentiment analysis," *J. Internet Comput. and Serv.*, vol. 19, no. 1, pp. 155-166, Feb. 2018.
(<https://doi.org/10.7472/jksii.2018.19.1.155>)

[10] M. Irfan Marwat, et al., "Sentiment analysis of product reviews to identify deceptive rating information in social media: A sentideceptive approach," *KSII Trans. Internet and Inf. Syst.*, vol. 16, no. 3, Mar. 2022.
(<https://doi.org/10.3837/tiis.2022.03.005>)

[11] C. Qian, et al., "Understanding public opinions on social media for financial sentiment analysis using AI-based techniques," *Inf. Process. & Manag.*, vol. 59, no. 6, 2022.
(<https://doi.org/10.1016/j.ipm.2022.103098>)

[12] N. V. Babu and E. G. M. Kanaga, "Sentiment analysis in social media data for depression detection using artificial intelligence: A review," *SN COMPUT. SCI.*, vol. 3, no. 74, 2022.
(<https://doi.org/10.1007/s42979-021-00958-1>)

[13] S. Park, Y. Kim, and M. Kim, "A design for intent classification models with covid-19 disaster alerts data," *Korea Computer Congress*, pp. 1810-1812, 2021.

[14] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, vol. 398, John Wiley & Sons, 2013.

[15] D. Wang and E. Nyberg, "A long short-term memory model for answer sentence selection in question answering," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, vol. 2, pp. 707-712, 2015.
(<https://doi.org/10.3115/v1/P15-2116>)

[16] A. Krizhevsky, I. Sutskever, and E Hinton

Geoffrey, "Imagenet classification with deep convolutional neural networks," *Advances in NIPS*, pp. 1097-1105, 2012.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, pp. 4171-4186, 2019.
(<https://doi.org/10.48550/arXiv.1810.04805>)

[18] The indigo, <https://theindigo.co.kr/archives/50296>

주 상 임 (Sang-Lim Ju)



2016년 : 충북대학교 전파통신공학과 공학석사

2021년 : 충북대학교 전파통신공학과 공학박사

2021년~현재: 한국전자통신연구원 재난안전지능화융합센터 박사후연수연구원

<관심분야> 이동통신, MIMO, 재난정보시스템, 디지털라디오

[ORCID:0000-0002-7584-8941]

강 현 주 (Hyunjoo Kang)



1999년 : 안동대학교 컴퓨터공학과 학사

2002년 : KAIST(구, 한국정보통신대학원대학교) 정보통신공학 석사

2002년~현재 : 한국전자통신연구원 재난안전지능화융합센터 책임연구원

<관심분야> 5G 긴급재난문자 고도화, 재난정보전달 플랫폼 기술, 유무선 통신 네트워크 프로토콜 기술

오 승 희 (Seung-Hee Oh)



2001년 : 이화여자대학교 컴퓨터
학과 석사

2019년~현재 : 충북대학교 전과
통신공학전공 박사과정

2001년~현재 : 한국전자통신연
구원 재난안전지능화융합센터
책임연구원/기술총괄

<관심분야> 긴급재난문자, 재난정보전달, 예경보시스
템, 복합재난 모델링, 네트워크 보안

[ORCID:0000-0001-5185-8435]