

Analysis of Disaster Safety Situation Classification Algorithm Based on Natural Language Processing Using 119 Calls Data

Su-Jeong Kwon[†] · Yun-Hee Kang^{††} · Yong-Hak Lee^{†††} · Min-Ho Lee^{††††} ·
Seung-Ho Park^{†††††} · Myung-Ju Kang^{††††††}

ABSTRACT

Due to the development of artificial intelligence, it is used as a disaster response support system in the field of disaster. Disasters can occur anywhere, anytime. In the event of a disaster, there are four types of reports: fire, rescue, emergency, and other call. Disaster response according to the 119 call also responds differently depending on the type and situation. In this paper, 1280 data set of 119 calls were tested with 3 classes of SVM, NB, k-NN, DT, SGD, and RF situation classification algorithms using a training data set. Classification performance showed the highest performance of 92% and minimum of 77%. In the future, it is necessary to secure an effective data set by disaster in various fields to study disaster response.

Keywords : Artificial Intelligence, Emergency Response, Natural Language Processing, Situation Classification, Machine Learning

119 신고 데이터를 이용한 자연어처리 기반 재난안전 상황 분류 알고리즘 분석

권수정[†] · 강윤희^{††} · 이용학^{†††} · 이민호^{††††} · 박성호^{†††††} · 강명주^{††††††}

요약

인공지능의 발달로 인하여 재난 분야에서는 재난대응 지원 시스템으로 이용되고 있다. 재난은 언제 어디서든 발생할 수 있으며, 재난 발생 시 소방청 119 신고접수대에 접수되는 신고는 크게 화재, 구조, 구급, 기타 신고 등 4가지로 구분된다. 119 신고에 따른 재난 대응도 그 종류 및 상황에 따라 다르게 대응된다. 본 논문에서는 119 신고 데이터 1280개 문서를 학습 데이터 셋을 이용하여 SVM, NB, k-NN, DT, SGD, RF 상황 분류 기계학습 알고리즘을 3 클래스로 테스트한 분류 성능은 최고 92%, 최소 77%의 성능을 보였다. 인공지능의 발달로 인하여 재난 분야에서는 재난 대응 지원 시스템으로 이용되고 있다. 재난은 언제 어디서든 발생할 수 있으며, 재난 발생 시 소방청 119 신고접수대에 접수되는 신고는 크게 화재, 구조, 구급, 기타 신고 등 4가지로 구분된다. 119 신고에 따른 재난대응도 그 종류 및 상황에 따라 다르게 대응된다. 본 논문에서는 119 신고 데이터 1280개 문서를 학습 데이터 셋을 이용하여 SVM, NB, k-NN, DT, SGD, RF 상황 분류 알고리즘을 3 클래스로 테스트한 분류 성능은 최고 92%, 최소 77%의 성능을 보였다. 앞으로 다양한 분야의 재난별 데이터 셋을 확보하여 효율적인 재난 대응 연구가 필요하다.

키워드 : 인공지능, 재난 대응, 자연어처리, 상황 분류, 기계학습

1. 서론

재난은 국민의 생명·신체 및 재산과 국가에 피해를 주거나 줄 수 있는 갑작스러운 사건으로 정의한다[1]. 태풍, 홍수, 호우와 같은 자연현상으로 인하여 발생하는 자연재난과 화재, 붕괴, 교통사고, 환경오염사고와 같은 인적 재난으로 구분할

수 있다. 최근 국가적으로 자연 재난과 인적 재난의 유형은 다양해지고 있으며, 복합재난 형태로 증가하고 있다.

재난 대응의 목표는 재난 상황을 빠르게 인지하고 정확한 행동 매뉴얼을 적용함으로써, 추가적인 인적 및 물적 피해를 최소화하는 것이다. 4차 산업혁명에는 정보통신기술을 기반으로 한 산업구조의 변화와 이에 따른 생산성 향상 특징으로 한다. 정보통신기술(ICT)의 발전은 전 산업 분야에서 기존의 체계를 혁신적으로 변화시키고 있다. 인프라 구축과 플랫폼의 형성으로 패러다임의 변화와 새로운 가치가 창출되고 있다. 텍스트, 이미지 등과 같은 비정형 데이터의 양이 점차 늘어나고 있으며, 이에 대한 정보화와 실시간 모니터링을 통한 유의미한 분석의 필요성이 재난 대응을 포함한 사회 분야에서 늘어나고 있다. 재난을 대응하기 위한 재난 관리 플랫폼은 재난 위험 저감 프로세스와 ICT 기술이 활용하여 위급 대응을 위

※ 이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2019-0-00103, 경험기반(빅데이터) 알고리즘의 재난 대응 AI Advisor 플랫폼 기술개발).

† 비회원 : ㈜넥타르소프트 대리

†† 정회원 : 백석대학교 ICT학부 교수

††† 비회원 : ㈜넥타르소프트 기술이사

†††† 비회원 : (주)넥타르소프트 부사장

††††† 비회원 : ㈜넥타르소프트 대표이사

†††††† 종신회원 : ㈜넥타르소프트 연구소장

Manuscript Received : August 26, 2020

Accepted : September 17, 2020

* Corresponding Author : Myung-Ju Kang(kmjiziro@daum.net)

한 문맥 기반 위험 관리 시스템을 제공이 필요하다[2, 3]. 위험관리 시스템은 기술적, 관리 기능을 갖는 운영 솔루션을 통해 상당 부분의 위험 요소를 줄일 수 있다. 그러나 현행 재난 대응은 수많은 환경 변화의 연속성과 지난 경험의 장단점을 상호 연계하는 방대한 자료를 사람이 분석하여 최적의 위험 관리와 사고 대응 절차를 빠른 시간내에 도출이 용이하지 않은 문제점을 갖는다.

본 논문에서는 경험에 의존하는 재난 대응의 제약점을 해결하기 위해 재난 대응 분야의 119 신고를 처리하는 재난 대응 시스템의 음성 인식 엔진으로부터 출력된 텍스트를 기계학습을 통해 상황을 분류하기 위한 재난 분야 기계학습모델을 개발한다.

본 논문에서는 기계학습을 사용한 재난 상황 분류를 수행하며, 이를 통한 재난 대응 역량을 향상시킬 수 있는 방안을 제시한다. 본 논문의 구성은 다음과 같다. 2장에서 재난 대응 분야에 적용 가능한 딥러닝을 포함한 기계학습 기술을 설명한다. 3장에서 앞서 기술한 감독자 기반의 전통적인 기계학습 및 딥러닝을 재난 분류에 적용하는 방안에 관하여 설명한다. 4장에서는 기계학습을 기반으로 분류 실험환경 및 실험 결과를 기술한다. 5장에서는 결론을 맺는다.

2. 재난 대응 기계학습

2.1 기계학습 개요

기계 학습(Machine Learning) 시스템은 인간의 경험으로부터 학습하여 스스로 성능을 향상하는 시스템으로 특정 문제 해결에 필요한 경험적인 데이터를 바탕으로 지식, 즉 모델을 자동으로 구축하고 성능을 향상한다. 컴퓨터 프로그램이 학습한다는 것은 주어진 태스크에서 데이터를 제공할수록 성능이 좋아지는 과정을 말한다. 예를 들어, 음성인식 태스크에서 인식률로 성능을 측정하는 음성인식 시스템에서, 음성 데이터를 추가하여 인식률을 개선하도록 진행하여 학습을 수행한다[3].

문제 해결을 위해 경험(데이터)을 기반으로 시스템을 학습하기 위한 기법은 지도학습, 비지도학습 및 강화학습으로 구분한다. 지도학습은 훈련 데이터에 레이블을 사용하며, 획득한 정보를 사용할 수 있는 능력을 기반 작업을 반복적으로 수행하여 해당 능력을 개선할 수 있다. 지도학습은 입력과 출력 데이터가 모두 주어진 경우 입력으로 출력을 맞추는 문제로 이해할 수 있다. 주로 음성인식, 얼굴인식 등의 패턴인식이나 회귀분석(Regression)과 같은 경우에 사용된다.

문서분류는 미리 정의된 두 개 이상의 범주(Category) 또는 클래스(Class)에 대하여, 새로운 문서 집단이 입력되었을 때 미리 학습된 범주와 입력 문서 간의 유사도 비교를 통해 입력된 문서에 대한 범주 할당을 자동으로 해주는 기법이다[4-6]. 자동 문서 분류를 이용하여 대량의 문서를 수작업으로 분류하는데 소요되는 시간과 노력 등을 감소시킬 수 있으며, 효율적인 정보의 조직 및 검색을 가능하게 하는 이점이 있다. 본 논문에서는 위급 상황에 대한 분류를 위한 클래스 식별을 위해 사용한다.

지도학습은 정답과 정답에 대한 레이블로 모델을 학습하는 방법이다. 비지도학습은 입력 상태에서만 입력 특성을 찾는

다. 학습문서 집합은 문서분류학습모델 구성을 위해 사용되며, 테스트문서는 문서 분류에 사용된다. 재난 분류 시스템은 전체 문서 집합으로부터 일정한 개수의 문서를 선택한 후 이에 대한 카테고리를 재난 대응 후 전문가에 의해 부여하여 학습 문서 집합을 구성한다. 본 시스템은 문서 분류 알고리즘을 선택하여 문서 분류 학습모델을 생성한다. 생성된 문서학습 모델은 문서분류기를 이용하여 실질적인 문서분류 예측을 수행한다. Fig. 1은 학습과 예측 과정을 보인 것이다.

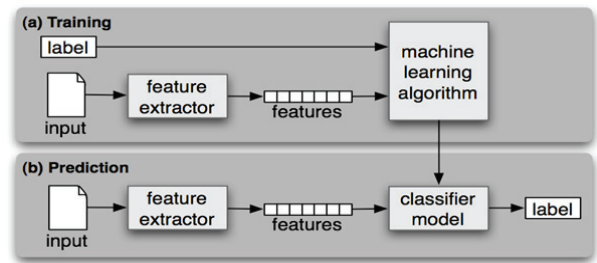


Fig. 1. The Process of Training and Prediction

2.2 기계학습 알고리즘

기계학습 기법을 이용한 문서분류의 문제를 다룬다. 이 기법에서는 학습 방법이 문서분류의 정확도를 크게 좌우하기 때문에 기계학습 연구는 문서 분류의 문제에 있어 중요하다. 문서분류시스템의 문서분류 정확도를 결정하는 요인을 두 가지 측면으로 요약할 수 있다. 첫 번째는 학습 문서 집합의 구성 방법이다. 문서분류를 위해서는 임의의 문서 공간 상에서 일정량의 문서집합을 택하여 학습 문서 집합을 구성한 후, 이를 이용하여 문서분류함수를 만든다. 따라서 어떤 문서를 학습문서로 선택할 것인가는 문서 분류 시스템에서 중요하다. 학습 문서집합의 구성 방법과 관련하여 학습문서 선택의 문제 외에도 선택한 학습 문서집합의 재구성 문제도 문서분류의 성능에 큰 영향을 미친다.

주어진 학습 문서집합을 있는 그대로 이용하여 문서분류함수를 만들 수도 있으나, 이를 지능적으로 재구성하여 이용하면 보다 정확한 문서분류함수를 얻을 수 있다. 학습기는 특성이 결정한 학습자료와 목표 변수(또는 레이블)의 쌍을 이용하여 각 클래스의 학습데이터로 사용하여 적절한 학습 알고리즘을 적용하여 학습한 후 새로운 데이터를 분류하기 위한 모델을 생성한다.

문서분류 시스템의 정확도를 결정하는 두 번째 요인으로 문서분류함수를 추정하는 알고리즘을 들 수 있다. 학습 문서집합 구성 방법에 의해 채택된 학습 문서집합을 이용하여 최종적으로 문서분류함수를 만들어 내는 것이 이 알고리즘의 역할이다. 의사결정트리(Decision Tree), k-NN (k-Nearest Neighbor, 최단인접 기법), SVM(Support Vector Machine), 신경망(Neural Networks), Naïve Bayes 알고리즘 등이 이에 속한다[7-12].

의사결정트리(Decision Tree)는 어떤 항목에 대한 관측값과 목표값을 연결시켜 주는 예측 모델이다. 입력 변수를 바탕으로 목표 변수의 값을 예측하는 모델을 생성하는 것을 목표로 한다[13]. 결정 트리의 '학습'은 학습에 사용되는 자료 집

합을 적절한 분할 기준 또는 분할 테스트에 따라 부분 집합들로 나누는 과정이다. 의사결정트리는 다음의 장점을 갖는다.

- a) 결과를 해석하고 이해하기 쉽다.
- b) 자료를 가공할 필요가 거의 없다.
- c) 수치 자료와 범주 자료 모두에 적용할 수 있다.
- d) 안정적이다.
- e) 대규모의 데이터 셋에서도 잘 동작한다.

Fig. 2와 같이 SVM(Support Vector Machine) 학습모델은 학습 데이터를 확보하고 데이터 커널 기반의 학습을 수행한 후 데이터를 클래스별로 분류한다. SVM은 2개의 범주를 분류하는 이진 분류기의 한 종류로서, 주어진 많은 벡터들을 두 개의 그룹으로 나눌 때, 각 그룹에 속한 벡터중 Hyperplane까지의 거리가 가장 가까운 벡터들이 먼 거리를 가지도록 두 개의 그룹으로 분류하는 것이다. 녹색과 파란색 두 개의 범주로 분류하고자 할 때, 두 그룹을 나누는 실선이 Hyperplane이 되는 것이고, 점선에서 이 Hyperplane까지의 거리가 최대가 되도록 Hyperplane을 결정하는 것이다.

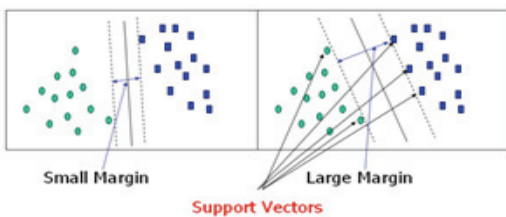


Fig. 2. SVM Training Model

나이브베이지스 분류(Naïve Bayes Classification)는 특성들 사이의 독립을 가정하는 베이즈 정리를 적용한 확률 분류기의 일종으로 1950년대 이후 광범위하게 연구되고 있으며, 1960년대 초에 텍스트 검색 커뮤니티에 단순 베이즈, 독립 베이즈 등의 이름으로 소개되었다. 나이브 베이즈 분류는 텍스트 분류에 사용됨으로써 문서를 여러 범주(예: 스팸, 스포츠, 정치)중 하나로 판단하는 문제에 대한 해결 방법으로 제시되며, 독립성 가정으로 나이브베이지스 분류를 매우 편리하고 유용하게 사용할 수 있다. 나이브베이지스 분류기는 지도학습 알고리즘으로 분류기(Classifier)가 실행되기 이전에 학습벡터를 통한 학습기(Trainer)에 의해 학습이 이루어져야 한다.

딥러닝(Deep Learning)은 머신러닝(Machine Learning)의 한 분야이다. 머신러닝은 컴퓨터에게 문제를 해결할 수 있는 정확한 알고리즘을 제공하지 않고, 문제 해결능력을 배우게 하는 방법이다[13, 14]. 딥러닝에서는 컴퓨터가 인공 신경망(Artificial Neural Network)을 사용하여 문제 해결 능력을 배운다. 인공신경망이란 인간의 뇌에 있는 신경망(Neural Network)을 모사하여 만든 모델이다. 이러한 인간의 특징에 주목하여, 컴퓨터에게 인간과 비슷한 문제 해결 과정을 부여하는 방법이 바로 딥러닝이다. 2012년에 영상에 포함된 다양한 객체를 인식하는 CNN(Alex Net) 모델이 개발된 이래, 그 성능을 개선한 객체 인식을 위한 모델들(GoogLeNet, VGGnet, ResNet, DenseNet 등)이 개발되었으며 기본 CNN 구조를

바탕으로 하여 영상을 복원하는 auto-encoder 모델이나, 의미 정보를 바탕으로 영상의 영역을 나누는 U-Net과 같은 모델, 그리고 영상에 포함된 특정 대상을 감지하는 faster RCNN 모델 등 다양한 변형들이 개발되고 있다. Fig. 3은 Alex Net의 학습과정을 보인 것이다. 딥러닝은 뛰어난 성능과 높은 모델의 확장성(Scalability)으로 인해 많은 주목을 받았고, 요즘 산업계에서도 활발하게 적용되고 있다.

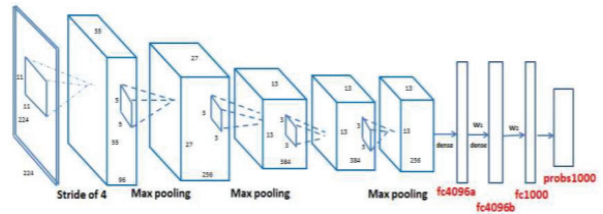


Fig. 3. AlexNet for 2D Image Data Analysis

CNN은 텍스트 자료의 효과적 학습과 분석을 위해 개발된 딥러닝 모델로, 1차원 벡터모델 자료에 대하여 Convolution과 Pooling 연산 과정을 반복적으로 거치면서 텍스트로 부터 의미 있는 특징 정보를 추출하는 과정을 학습하고, 이후 마지막 단계에서 기존의 MLP와 같은 Fully Connected Layer 계산을 통하여 분류를 수행하도록 설계된다.

3. 재난 대응 환경 기계학습 모델

재난 대응 기술은 재난상황에 따른 적절한 재난 대응 시나리오를 운영할 수 있어야 한다. 재난이 발생했을 때 재난유형 및 상황에 맞는 대응 시나리오를 관리자에 제공할 수 있어야 한다.

Fig. 4는 재난사고 현장에서의 신고접수에서 출동까지의 과정을 보인 것으로 음성인식(Voice Recognition) 후 변환된 텍스트는 한국어 처리시스템(Korean Language Processing System)에서 언어처리를 통해 AI 조력자의 재난 분류 시스템의 의사결정(Decision Support)의 입력으로 제공되며, AI Advisor는 재난 분류 결과를 의사결정에 활용한다[2].

재난 대응 기술은 재난 상황에 따른 적절한 재난 대응 시나리오를 운영할 수 있어야 한다. 재난이 발생했을 때 재난 유형 및 상황에 맞는 대응 시나리오를 관리자에 제공할 수 있어야 한다.

재난 분류를 위한 과정은 다음과 같다. 본 과정을 통해 재난분류 결과는 재난 대응 추천시스템에 의해 학습 문서집합에 추가될 수 있다.

- 1) 재난 대응 전문가는 후보 학습문서를 입력으로 받아 클래스를 부여함
- 2) 문서분류 학습모델 생성을 위해 문서집합을 입력으로 전달받음
- 3) 전달받은 문서에서 주요한 특징을 선택하기 위해 클래스별 주요 키워드 선별 후 학습문서 구성을 수행함
- 4) 문서 분류 알고리즘을 사용하여 문서 분류 학습모델을 생성함

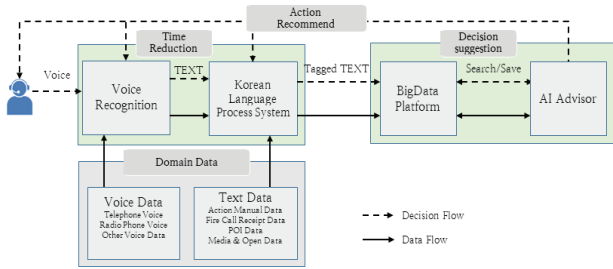


Fig. 4. Concept of 119 Disaster Calls using AI Advisor

- 5) 문서분류기는 재난 대응 분류 시스템으로 부터 테스트 문서를 전달받아 분류를 요청받아 문서 분류를 수행함
- 6) 재난 대응 추천시스템은 문서분류 결과를 사용하여 학습문서 추천을 수행하며, 학습 문서집합에 추가함

4. 상황인지 기계학습 분류 실험결과

기계학습 모델을 개발하기 위한 학습 데이터는 소방본부 119상황센터에서 개인정보 없는 데이터를 전달된 음성정보를 음성인식 STT(Speech to Text)를 이용하여 텍스트로 변환한 후 획득된 데이터로부터 재난분류 및 사건정보의 추출 등 특징을 추출한다. 이 분석 결과를 기반으로 사건 신고정보와 연결된 상황을 인지하여 재난유형을 분류한다. 축적된 자료를 바탕으로 한 다양한 상황의 인과관계 예측을 통해 여러 상황에 신속하게 대응할 수 있다.

4.1 데이터 셋

벡터공간 모델에서 질의와 각 문서는 용어 공간 내의 벡터로서 표현한다. 유사도의 측정은 두 벡터를 계산하여 얻는다. 색인어 k_i 와 문헌 d_j 가 주어졌을 때, (k_i, d_j) 쌍의 가중치라고 하며, 이 가중치는 문헌의 의미적 내용을 설명하기 위한 색인어의 중요도를 정량화한다. 시스템 내의 색인어 수를 t 라 하고 k_i 를 색인어라 하면, 모든 색인어 집합 $K=\{k_1, \dots, k_t\}$ 가 된다. 문헌 d_j 에서의 색인어 k_i 의 가중치는 (d_j, k_i) 이고 따라서, 문헌 내에 한번도 출현하지 않은 색인어의 가중치는 0 이 된다. 문헌은 색인어 벡터 = $\langle w_{1j}, w_{2j}, \dots, w_{tj} \rangle$ 로 표현된다. 벡터 공간 모델에서 색인어 가중치는 상호독립적이라고 가정하며, 색인어-문헌 쌍 (k_i, d_j) 의 가중치 w_{ij} 와 (k_{i+1}, d_j) 의 가중치 $w_{i+1, j}$ 는 서로 무관하다.

전체 학습 데이터 셋은 1280 문서이며, 각 문서는 2638의 색인 특징으로 이루어졌다. 전체 데이터 셋은 구급 891, 화재 176, 구조 111, 미분류 55, 기타 47으로 전체 데이터 셋의 69.60%가 구급으로 구성되었으며, 상대적으로 화재는 13.75%, 구조는 8.67를 차지한다. 각 특징은 발생 값에 따라 0과 1의 값을 특징값으로 지정하여 계산한다.

실험에서의 기계학습 알고리즘 적용시 데이터셋은 구급, 화재, 구조, 미분류 및 기타의 5개 클래스에 대한 레이블을 구성하여 사용하였으며, 이 중 구급, 화재 및 구조만을 대상으로 3개 클래스에 대한 학습을 수행하였다

4.2 실험 환경 구성

전통적인 기계학습 실험을 위해서는 Python 3.6 환경에서 구현된 기계학습 오픈소스 라이브러리인 scikit-learn 0.19.1를 사용한다. 라이브러리의 구성은 크게 지도 학습, 비지도 학습, 모델 선택 및 평가, 데이터 변환으로 구성되어 있다. Python 3.6을 사용하여 데이터 전처리를 위한 프로그램을 작성하며, 데이터셋의 전처리를 위해서는 numpy 1.6.1를 사용하여 벡터공간을 구성한다. CNN 모델 구성을 위해서는 keras 2.1.1을 백그라운드 엔진으로는 Tensorflow 1.14.0을 사용한다. 데이터 학습을 위한 GPU의 cuda 라이브러리는 10.0을 사용한다. 경량의 학습모델 운영을 가정하여 Jetson Nano에 Ubuntu 18.04인 Linux jetson-4-3을 사용한다.

4.3 기계학습 분류 결과 및 분석

6가지 기계학습을 수행한 결과는 Table 1과 같다. 분류 문서는 5개의 레이블을 갖는 자료를 대상으로 학습데이터로 80%, 테스트 데이터로 20%를 사용하였다. 전체적인 분류 성능은 최고 87%, 최소 71%이며, SVM과 SGD, RF가 80% 이상으로 상대적으로 높은 분류 성능을 보였으며, NB는 73%와 k-NN는 71%의 성능을 보였다.

Table 1. Comparison of Performance with regard to Classification Algorithms

Classification Algorithm	5 Classes	3 Classes	Performance Improvement
SVM	87	92	5.7
NB	73	77	5.47
k-NN	71	81	14.08
DT	78	89	14.1
SGD	82	92	12.19
RF	81	87	7.40
Average	78.67	88.33	9.92

Table 1은 성능 결과에 대한 분석, 1,280개의 적은 데이터 셋과 편향이 큰 학습 데이터를 기반으로 기계학습 수행한 결과를 보인 것으로 SVM와 SGD은 적은 데이터 셋에서 분류가 우수함을 보였다. 5개의 클래스에서 3개의 클래스로 변경되어 KNN과 DT의 성능개선은 다른 학습 알고리즘 보다 우수함을 보였다.

Epoch 값으로 10을 설정한 후 CNN 학습수행 후 Epoch 값 = 10, 전체자료 학습 데이터 90%, 테스트 데이터 10%, 5개의 클래스 학습과정에서의 결과 학습손실과 정확률을 보인 것으로 정확률은 84.20%를 손실 함수인 RMS 0.6737을 보였다. Fig. 5와 6은 5개 클래스 학습과정에서의 손실율과 정확률을 보인 것이다.

Epoch 값으로 10을 설정한 후 CNN 학습수행 후 Epoch 값 = 10, 전체자료 학습 데이터 90% 테스트 데이터 10%, 미분류와 기타를 제외한 후 3개의 클래스 학습과정에서의 결과 학습손실과 정확률을 보인 것으로 정확률은 92.31%를 손실 함수인 RMS 0.6737을 보였다. Fig. 7은 3개 클래스 학습과정에서의 정확률을 보인 것이다.

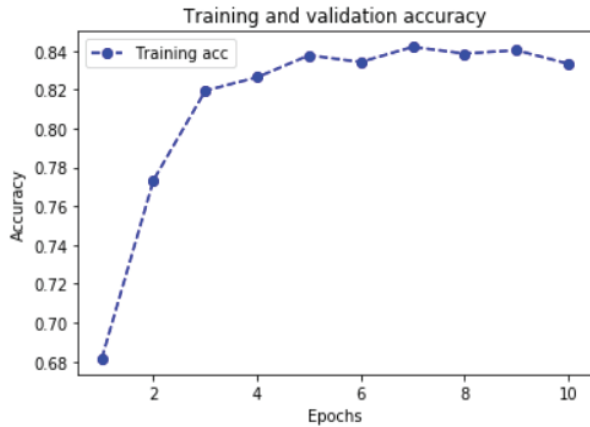


Fig. 5. Precision of 5 Classes CNN Classification

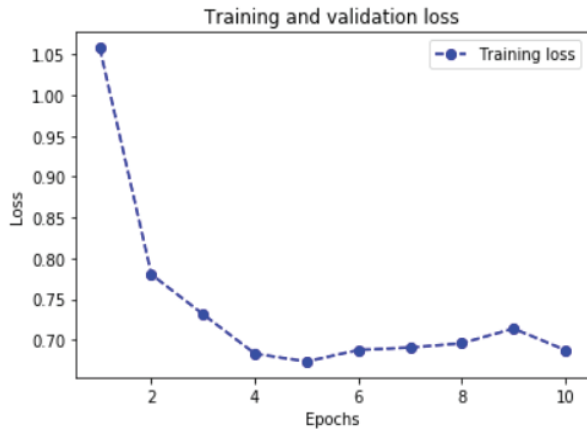


Fig. 6. Loss rate of 5 Classes CNN Classification

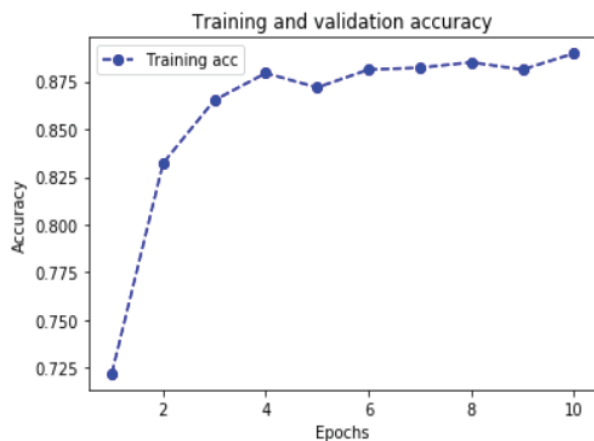


Fig. 7. Precision of 3 Classes CNN Classification

5. 결 론

본 논문에서는 119 신고 접수대에서 처리하는 재난 대응 시스템의 음성인식 엔진으로부터 출력된 텍스트를 기계학습을 이용하여 상황을 분류하기 위한 재난 분야 기계학습 모델을 개발하였고, 119 신고 데이터 셋과 머신러닝을 이용하여 상황인지 기계학습 분류 실험한 결과를 바탕으로 상황 분류

알고리즘을 분석하였다. 119 신고체계에서 신고자의 상태에 따른 적절한 대응책들을 제시하여 효율적인 대응을 위한 AI Advisor 기능을 이용하여 신고자의 주변 환경과 상황 등을 파악을 위해 감독자 기반 기계학습 알고리즘을 사용하여 재난 상황인지를 분류한 알고리즘들을 분석하였다. 제안된 기계학습 분류 알고리즘과 함께 추가적으로 재난 안전 SNS, 기상정보, 센서 데이터 등 기존의 재난 관련 정형 데이터 및 비정형 데이터(텍스트, 영상)를 연계 통합하여 분석하면 보다 효율적인 재난 대응 체계 연구가 더 필요하다.

References

- [1] D. Sarwar, M. Ramachandran, and A. Hosseinian-Far, 2017. "Disaster Management System as an Element of Risk Management for Natural disaster Systems using the PESTLE framework," In Communications in Computer and Information Science (pp. 191-204). (Communications in Computer and Information Science; Vol. 630). Springer Verlag. https://doi.org/10.1007/978-3-319-51064-4_16
- [2] Myung-Ju Kang, Sang-Cheon Lee, Yong-Hak Lee, and Yun-Hee Lee, "Emergency Response Artificial Intelligence Advisor Technology based Big Data," *Korea Information Processing Society Review*, Vol.26, No.6, pp.18-26. Jan. 2019.
- [3] Thomas M. Mitchell. 1997. "Machine Learning," (1 ed.) McGraw-Hill, Inc., New York, NY, USA.
- [4] G. Salton, "Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer," Addison-Wesley, 1989
- [5] Y. Yang and J. O. Pedersen, "A Comparative Study on Feature Selection in Text Categorization," *Proc. Of the 14th International Conference on Machine Learning ICML-97*, pp.412-429, 1997.
- [6] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *Proc. European Conference on Machine Learning (ECML)*, pp. 137-142, 1998.
- [7] Alpaydin, E. (2014). "Introduction to Machine Learning. Cambridge," MA: MIT Press. ISBN: 978-0-262-02818-9
- [8] S. J. Russell and P. Norvig, "Artificial Intelligence: A Modern Approach, Third Edition, Prentice Hall, 2010.
- [9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, Vol.20, No.3, pp.273-297, 1995.
- [10] A. Ben-Hur, D. Horn, H. Siegelmann, and V. Vapnik, "Support vector clustering," *Journal of Machine Learning Research*, Vol.1, pp.125-137, 2001.
- [11] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?" *International Conference on Database Theory*, pp.217-235, 1999.

[12] T. K. Ho, "Random decision forests," *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Montreal, QC, pp.278-282, Aug. 1995.

[13] I. Ben-Gal, A. Dana, N. Shkolnik, and G. Singer, "Efficient Construction of Decision Trees by the Dual Information Distance Method," *Quality Technology & Quantitative Management*, Vol.11, No.1, pp.133-147, 2014.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, Vol.521, No.7553, pp.436-444, 2015. <https://doi.org/10.1038/nature.14539>.



이 민 호

<https://orcid.org/0000-0002-5444-0306>
 e-mail : dpmhl@nectarsoft.co.kr
 1991년 동국대학교 컴퓨터공학과
 1994년 현대상선 IT기획팀장
 2000년 현대상선 대만 IT 주재원
 2007년 현대유엔아이 사업부장
 2017년 타임게이트 영업본부장
 2018년 ~ 현 재 (주)벡타르소프트 부사장
 관심분야 : 음성인식, 빅데이터, 머신러닝, 인공지능



권 수 정

<https://orcid.org/0000-0002-8014-552X>
 e-mail : sjkwon@nectarsoft.co.kr
 2017년 동국대학교 컴퓨터공학과(공학석사)
 2018년 ~ 2019년 트리니티디아아이비(주)
 2019년 ~ 현 재 (주)벡타르소프트 대리
 관심분야 : 자연어처리, 머신러닝, 인공지능



박 성 호

<https://orcid.org/0000-0001-6939-1095>
 e-mail : sh_park@nectarsoft.co.kr
 1980년 건국대학교 전자과(공학사)
 1988년 메킨토시 OS 한글 Localize
 1994년 커팅플로터 2.5D CAD 프로그램
 대한민국 최초 개발
 1994년 ~ 현 재 (주)벡타르소프트 대표이사
 2005년 국내 최초 디지털 무전기 디스패치 시스템 개발(VRS)
 2018년 Nectar Wizer Buddy 음성인식 엔진 개발
 관심분야 : 음성인식, 머신러닝, 무선통신시스템



강 윤 희

<https://orcid.org/0000-0002-6977-3779>
 e-mail : yhkang@bu.ac.kr
 1989년 동국대학교 컴퓨터공학과(공학사)
 1991년 동국대학교 컴퓨터공학과(공학석사)
 2002년 고려대학교 컴퓨터학과(이학박사)
 2000년 ~ 현 재 백석대학교 ICT학부 교수
 관심분야 : 클라우드컴퓨팅, 그리드 컴퓨팅, 머신러닝, 인공지능



강 명 주

<https://orcid.org/0000-0002-1139-7184>
 e-mail : kmjziro@daum.net
 1988년 동국대학교 컴퓨터공학과(공학사)
 1991년 동국대학교 컴퓨터공학과(공학석사)
 1993년 ~ 1997년 제주국제대학교
 정보보호공학과 교수
 2000년 ~ 2008년 (주)클릭큐 연구소장
 2009년 ~ 2012년 자바정보기술(주) SI사업본부장
 2013년 ~ 2017년 트리니티(주) 빅데이터사업본부장
 2018년 ~ 현 재 (주)벡타르소프트 연구소장
 관심분야 : 빅데이터, 자연어처리, 기계번역, 인공지능



이 용 학

<https://orcid.org/0000-0002-5680-8901>
 e-mail : leeyh@nectarsoft.co.kr
 1990년 동국대학교 컴퓨터공학과(공학사)
 1992년 동국대학교 컴퓨터공학과(공학석사)
 1998년 동국대학교 컴퓨터공학과(공학박사)
 1998년 ~ 2003년 김포대학 컴퓨터계열
 조교수

2004년 ~ 2013년 (주)백상정보통신 연구소장
 2017년 ~ 2019년 (주)빌리언21 기술이사
 2019년 ~ 현 재 (주)벡타르소프트 기술이사
 관심분야 : 빅데이터, 자연어처리, 인공지능