

텍스트마이닝 및 기계학습을 활용한 온나라시스템 문서분류 모델링

Document Classification Modeling Using Text Mining and Machine Learning in On-nara System

윤영빈¹⁾ · 문호석²⁾

Yeongbin Yun · Hoseok Moon

ABSTRACT

In the military, a substantial volume of official documents is processed daily in electronic format through the On-nara system. Documents dispatched from external sources are received by unit officers, and manually processing these documents is time-consuming. To solve this problem, the application of pre-trained language models in military administrative tasks could significantly enhance efficiency. However, introducing this technology into the military presents real-world challenges, such as cost and security issues. This study proposes a method for automatically classifying electronic official documents using text mining and machine learning. Specifically, a classifier based on XGBoost has achieved high accuracy, approximately 90%. The adoption of this artificial intelligence technology will facilitate the automation of routine administrative tasks, saving on manpower and time, and improve the overall work environment, allowing soldiers to concentrate more on their primary missions.

Keywords : Artificial Intelligence, Document Classification, Machine Learning,
On-nara System, Text Mining

논문접수일 : 2023년 11월 20일, 심사일 : 2023년 11월 27일, 게재확정일 : 2023년 11월 28일

1) 국방대학교 국방과학학과 석사과정

2) 국방대학교 국방과학학과 교수 / 교신저자(hsmoon0329@gmail.com)

1. 서론

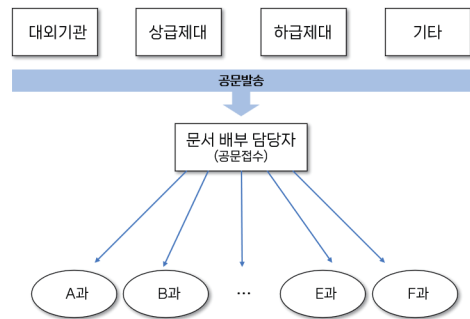
육·해·공군에 근무하는 모든 간부들은 ‘온나라 업무시스템(이하 온나라 시스템)’의 메모보고, 전자문서, 일정관리 등 기능을 이용하여 행정 업무를 수행한다. 온나라 시스템은 정부 기관, 관공서 등과도 연동되어있어 군내 공문처리 뿐만 아니라 대외업무도 효율적으로 처리할 수 있는 시스템이다. 이를 이용하여 각종 문서를 작성, 검토, 승인 및 공유하고 배포, 보관에 이르기까지 전체 문서 처리 프로세스를 관리할 수 있다. 또한, 부서별 업무 정보를 실시간으로 공유할 수 있기 때문에 협업에 용이한 체계이며, 강력한 보안 기능이 있어 인가되지 않은 사용자 접근을 제어하고, 필요시 문서를 비공개할 수 있다.

온나라 시스템 메뉴 중에서 ‘전자문서’ 메뉴는 상급부대, 타 기관·제대 등으로부터 접수한 공문, 부서 내에서 기안한 공문, 수기로 결재된 문서에 대한 근거를 남기기 위한 비전자문서 등록 등 다양한 기능을 제공한다.

특히 전자문서의 기능 중 가장 많이 활용되는 기능은 ‘공문 접수’ 기능이다. 공문은 공문서의 줄임말로 공공기관이나 단체에서 작성한 문서를 뜻하는데, 권한이 있는 군인(공무원 또는 군무원)이 작성하여 최종결재권자의 결재를 받으면 공문의 효력을 지닌다. 군에서 공문은 명령의 하달, 교육훈련 시행, 시설 협조, 비밀 수정 등 다양한 목적으로 사용되어 중요한 역할을 한다.

공문의 배부 절차는 <그림 1-1>과 같다. 대외기관, 상급부대 등 외부로부터 발송되어온 공문은 바로 담당 실무자에게 접수되는 것이 아니라 해당 부대의 문서 배부 담당자에게 종합되고, 문서수발 담당자는 각각의 공문의 제목과 기안지에 있는 정보를 확인하여 수신 부서로 배부한다. 많은 문서를 분류하는 담당자 입장에

서 공문 제목만으로 어느 부서로 배부되어야 하는지 명확히 구분되기도 하지만, 그렇지 않으면 공문의 전체 내용을 읽어보거나 부서에 일일이 연락하여 공문의 출처를 확인하여 올바르게 배부해야 한다. 또한 이러한 과정에서 공문이 잘못 배부되어 반송되는 사례도 종종 있으며, 이 경우 중요한 공문이 늦게 처리되어 실무자가 업무의 시기를 놓치는 문제가 발생할 수도 있다.



<그림 1-1> 공문 배부 절차

이러한 공문 배부 상의 문제점이 어떤 것이 있는 지에 대해서 육군 2개 군단 및 8개 사단 문서 배부 담당자에게 업무 처리 시간을 확인한 결과, 군단에서는 문서 배부에 일일 평균 1시간 45분, 사단에서는 일일 평균 1시간이 소요되었다. 또한, 문서 배부 담당자에게 배부업무는 주요 업무가 아닌 수시로 처리해야 하는 업무로 일과 중에 접수된 공문을 수시로 확인하여 배부해야 하며, 당일 배부가 원칙이라 매일 빠짐없이 문서 배부 업무를 수행한다[1].

문서 배부업무는 보병 여단급 제대부터 군 최상위제대까지 모든 제대에서 실행되는 업무이며 우편번호나 주소를 보고 우편을 분류하는 것과 같이 단순한 업무에 속한다. 이러한 분류 작업은 인공지능을 통해서 구현할 수 있는 기능이며 다양한 사례로 실현되고 있다. 예를 들어 한국의 우정사업본부와 같은 미국의 USPS(United States Postal Service)는 인공

지능 기술을 활용하여 우편번호와 주소정보를 자동으로 수집하고 우편물을 분류한다. 이를 통해서 8~10명의 직원이 수일 동안 하던 일을 현재는 1~2명의 직원이 단 두 시간 만에 완료할 수 있게 되었다[2]. 또한 법률 자동화 회사인 로지스(LawGeex)에서는 인공지능과 20명의 변호사를 대상으로 11페이지 분량의 비밀 유지계약 5건을 처리하는 실험을 진행하였다. 실험 결과, 인공지능은 26초 만에 문서 검토를 완료했지만, 대형 로펌 등에서 근무하는 변호사들은 평균 92분이 걸렸다. 정확도 평가 결과에서도 인공지능은 94%의 정확도를 보인 반면 변호사들은 85%에 그쳤다[3].

본 연구는 온나라 시스템에 있는 공문접수 기능에서 접수된 공문을 자동으로 담당 실무자에게 분류하는 문서 배부 인공지능에 관한 연구이다. 이를 위해 2020년부터 2022년 전반기까지 접수된 공문 9,024개의 텍스트 정보를 수집하여 학습하였고, 2022년 후반기에 접수된 공문을 예측하여 정확도를 검증하였다. 유·무인 복합체계, 무인 로봇 등 전장에서 활약하는 인공지능을 개발하는 것도 중요하지만, 단순한 행정 업무를 인공지능으로 대체하여 전투 임무에 집중할 수 있는 환경을 조성하는 것도 중요한 분야이다.

2. 기존연구

텍스트 데이터는 대표적인 비정형데이터이며 이를 분석에 사용하기 위해서는 적절한 전처리 과정을 거쳐 정형화할 필요가 있다. 다양한 방법론에 따라 정형화한 데이터는 기계학습 기술 등에 훈련데이터로 활용해서 분류기를 개발하는 과정을 거친다.

텍스트 데이터의 전처리 과정에는 주로 BoW(Bag-of-Words), TF-IDF(Term Frequency-Inverse Document Frequency),

Word2Vec, 그리고 Doc2Vec 같은 기법이 활용된다. BoW와 TF-IDF는 단어의 빈도수를 분석하여 텍스트 내에서 중요한 단어를 식별하는데 사용되며, Word2Vec은 단어를 벡터화하여 문맥의 의미까지 고려하는 기법이고 Doc2Vec은 Word2Vec을 확장하여 문서 전체의 의미를 벡터로 변환하는 기법이다. 김성희(2008), 박찬정(2012)은 텍스트 전처리 과정을 거친 다음 전체 단어 벡터를 구성하고 대표 단어를 정한 다음 학습 문서에서 용어별 가중치를 TF-IDF를 이용하여 계산하여 활용하였다[4].

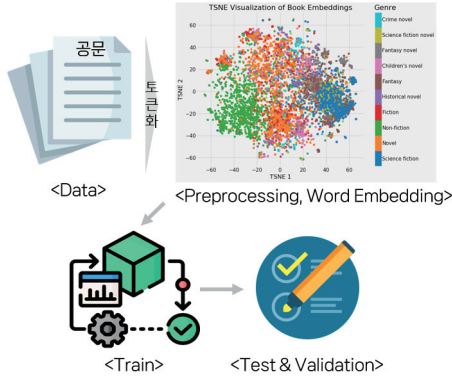
텍스트 데이터를 이용하여 지도학습에 활용한 기존 연구도 다양하다. 김성희(2008)는 신경망(Neural Network), 의사결정나무(C5.0, CHAID), KNN(K-Nearest Neighbor)을 활용하여 미국 국립의학도서관 MeSH의 8개의 주제별 범주를 분류하였으며[5], 김정미(2017)는 Word2Vec을 활용한 RNN(Recurrent Neural Network)기반 모델을 새롭게 제안하였다[6].

김도우(2017)는 Word2Vec과 Doc2Vec을 CNN(Convolution Neural Network)에 적용하는 방법을 제안하였다. 이 논문에서는 문서 분류에 Doc2Vec을 활용하는 것이 훨씬 효과적이고 한국어 신문 기사 분류 간 높은 성능을 보여주었다[7]. 정지수(2019), 송진주(2019)는 Doc2Vec을 이용하여 문서의 특징을 추출하고, 이를 바탕으로 SVM(Support Vector Machine)과 로지스틱 회귀 모델을 구축하였다[8][9].

BoW 및 TF-IDF방식은 핵심 단어 추출에는 유용하나, 문서의 전반적인 문맥을 이해하고 파악하는 데는 한계가 있다. 이에 대한 해결책으로 Word2Vec과 Doc2Vec을 함께 활용하여 단어 간의 의미적 관계를 보존하고 전체 문서를 벡터화하는 방식이 최근에는 유용하다고 평가받고 있다. 본 연구에서는 다양한 방식들을 고려하여 문서 분류 모델을 성공적으로 구현할 수 있는 벡터화 방식을 사용하였다.

3. 연구 방법

3.1 연구 절차



<그림 3-1> 세부 연구 절차

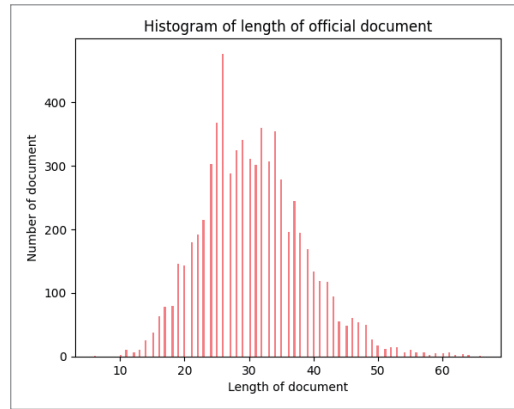
본 논문의 연구 절차를 간략화하면 <그림 3-1>과 같다. 먼저, 2020년 1월 1일부터 2022년 12월 31일까지 온나라 시스템 전자문서 공문의 제목과 수신처에 해당하는 텍스트 데이터를 수집하였다. 이후 탐색적 자료분석(Exploratory Data Analysis)을 통해 텍스트 데이터가 가지고 있는 내재된 특성을 확인하기 위해 시각적으로 표현하여 확인하였다. 그리고 텍스트에 대한 정보를 단위별로 나누는 자연어 토큰화(Tokenizing) 과정을 거쳐 불용어 처리, 어간 추출 및 일치화 등 전처리를 과정을 실시하였다.

텍스트 데이터는 대표적인 비정형데이터로 컴퓨터로 분석하기 위해선 숫자 형태인 벡터로 변환하는 벡터화(Word Embedding) 과정이 필요한데, Python의 사이킷런 라이브러리를 활용하면 텍스트 데이터에서 빈도수를 기준으로 특징을 추출하는 ‘CountVectorizer’, 단어의 빈도수와 문서 빈도수를 활용한 ‘TfidfVectorizer’ 등을 이용할 수 있으며, 사전학습 언어 모델을 이용하여 벡터화하는 방법도 사용할 수 있다. 이번 연구에서는 단어 빈도수 기반의 벡터화

방식인 CountVectorizer를 이용하였다. 이후에 학습단계에서는 데이터의 특징에 따라 단순한 회귀모형부터 랜덤포레스트, XGBoost와 같은 앙상블 모형까지 복잡한 모형에 적합시키며 결과를 비교하였다.

3.2 탐색적 데이터 분석

전체 데이터의 개수는 6,874개로 제목을 의미하는 ‘document’ 변수와 수신처를 의미하는 ‘label’로 구성되어 있으며, 반응변수에 해당하는 ‘label’ 변수는 작전, 인사, 정보, 지통, 군수 등 12개 부서명으로 구분되어 있다. ‘document’ 변수는 전처리 과정 이전의 텍스트 데이터로 문서 제목이 포함되어있다. 각각의 데이터들에 대해 문서 제목의 길이를 확인해보면 <그림 3-2>와 같다.



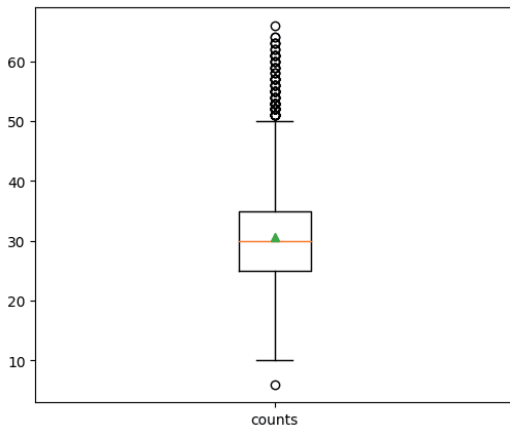
<그림 3-2> 온나라 전자문서 공문 제목 길이 분포

문서 제목의 길이는 10자 정도의 짧은 길이부터 60자까지 분포돼 있다. 많은 공문의 제목은 25자에서 35자 사이에 많이 몰려있는 정규 분포의 형태를 보이는데, 이러한 특성은 공문을 작성하는 실무자가 제목을 작성할 때 적절한 길이로 표현하려는 습관이나 노력이 반영된 결과라고 볼 수 있다. 문서 제목에 대한 기술통계량은 <표 3-1>과 같다.

<표 3-1> 공문 제목 길이 기술통계량

최대값	최소값	평균	표준 편차	중앙값	1사분 위수	3사분 위수
66	6	31	8	30	25	36

길이의 최대값은 66이다. 최소값은 6이고 평균의 경우 약 31 정도 된다. 중간값은 30자로 평균보다 조금 작은 값을 가진다. 이 값에 대해 상자 그림(Boxplot)을 그려보면 <그림 3-3>과 같다. 그림을 보면 일부 길이가 긴 데이터가 꽤 있다는 것을 확인할 수 있다. 중앙값과 평균값은 전체 데이터로 보면 중앙 쪽에 위치하고 있다.



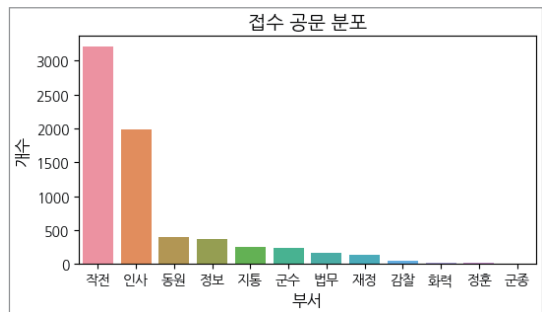
<그림 3-3> 공문 제목 길이 상자 그림

워드클라우드(Word Cloud)는 텍스트 데이터에서 단어의 빈도를 시각적으로 표현하는 방법인데 텍스트 내에서 자주 등장하는 단어는 더 큰 글꼴로, 드물게 등장하는 단어는 더 작은 글꼴로 표현하여 데이터 내의 주요 개념과 키워드를 한눈에 파악할 수 있는 방법이다. 이를 활용하여 시각화한 결과는 <그림 3-4>와 같다. 워드클라우드를 보면 ‘결과 보고’, ‘일일 명령’, ‘점검 결과’ 등의 어휘가 가장 많이 사용된 것을 확인할 수 있다.



<그림 3-4> 공문 제목 워드클라우드

마지막으로 기계학습 모델의 훈련 및 평가 과정에서 중요한 요소인 라벨값의 비율을 <그림 3-5>와 같이 시각화하였다. ‘작전’은 3,204개, ‘인사’는 1,985개 등 불균형한 분포를 보이고 있다. 이러한 불균형은 다양한 부서로부터의 데이터 수집이 충분하지 않아 발생한 클래스 불균형 문제로 본 연구에서는 ‘작전’, ‘인사’ 라벨에 해당되는 데이터만을 분석대상으로 선정하였으며, 클래스 균형을 맞추고 실험을 하였다.



<그림 3-5> 데이터 클래스 분포

4. 연구 결과

4.1 데이터 전처리 결과

앞서 데이터를 분석한 결과를 바탕으로 데이터 전처리 과정을 진행하였다. 전처리 과정에는 Python 언어의 numpy, pandas, konlpy, tensorflow 등의 라이브러리를 사용하였다. 공문 제목에 특수문자 혹은 숫자 등이 종종 보이

는데 분석과정에서 중요하지 않다고 판단하여 제거하고 한글과 영어 문자만을 남겼다. 이후 konlpy 라이브러리의 형태소 분석기 중 okt 형태소 분석기를 사용하여 형태소를 분석함과 동시에 어간 추출을 하였다.

다음으로 불용어 제거는 공문서 제목이 한글로 작성되었기에, 기존에 영문 위주의 불용어 사전을 활용할 수 없었다. 본 연구에서는 연구자가 직무지식과 탐색적자료분석 결과를 바탕으로 불용어 사전을 정의하고 그것을 기초로 불용어 제거를 하였다. 전처리가 완료된 결과의 일부가 <표 4-1>에 정리되어 있다. 모든 문서 제목에 대해서 <표 4-1>과 같이 정리 결과가 만들어졌다. 전처리 결과를 요약해보면 공문 제목을 구성하는 단어는 약 2,300개이며, 각각의 제목은 최소 2개에서 최대 14개 단어로 구성되어 있고, 6개 단어로 이루어진 제목이 가장 많았다.

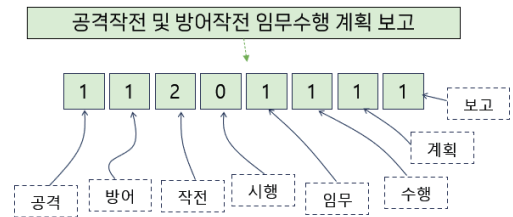
<표 4-1> 문서의 전처리 결과 예

문서	전처리 결과
1	'CCPT', '모의', '운용', '요원', 'BSC', '선발', '협조'
2	'군단', '경계', '작전', '지침', '수정', '가문', '하달'
3	'주둔지', '경계', 'CCTV', '사업', '물품', '증감', '현황', '보고'
4	'분기', '군단', '상황', '관계', '회의', '계획', '보고'
...	...

4.2 텍스트 데이터 벡터화 및 모델링

텍스트 데이터 전처리가 완료된 후에는, 컴퓨터가 이해할 수 있는 형태로 데이터를 변환해야 한다. 이러한 변환 과정을 '벡터화'라고 하

는데, 본 연구에서는 다양한 벡터화 방법 중 문서를 단어의 빈도로 변환하여 BOW를 생성하는 단어 빈도수 기반 벡터화 방법을 사용하였다. BoW 기법을 간단히 표현하면 <그림 4-1>과 같다. 각 문서는 고정된 길이의 벡터로 변환되며, 각 요소는 해당 문서에서 특정 단어가 출현한 횟수를 나타낸다. BoW 기법은 텍스트 데이터 벡터화 기법 중 가장 단순한 방법으로 텍스트 데이터를 직관적으로 정형화시키는 방법이다.



<그림 4-1> 텍스트 데이터 벡터화 과정

하지만 분리된 단어는 약 2,300여 개로 모든 단어를 학습하기에는 차원이 너무 커 단어 빈도수 누적 비율을 고려하여 각 단어의 중요성을 정량적으로 평가하는 방법(이창용, 2016)을 활용하여, 누적 비율 60%를 차지하는 상위 100개의 단어를 핵심단어로 선정하였다. 이는 선정된 100개의 단어가 전체 문서의 60%를 설명한다는 의미를 갖는다. 또한 데이터의 설명력은 40% 정도 감소하지만 약 95.6%의 차원 축소 효과가 있다고 해석할 수 있다[10].

이렇게 변환된 벡터는 학습 데이터와 평가 데이터에 모두 적용되며, 각 벡터의 길이는 100으로 동일한 값을 갖는다. 마지막으로 벡터화가 완료된 정수형 벡터를 모델에 적용하고 라벨값을 예측하는 모델을 만들었다. 이 과정에서 알고리즘은 단순한 모델인 의사결정나무(Decision Tree)부터 각종 앙상블 모델까지 여러 모델을 사용하여 정확도를 비교하였다.

4.3 기계학습을 통한 분류 결과

전처리와 벡터화를 통해서 구축된 문서 제목에 대한 학습 데이터에 대해서 기계학습 방법을 적용하여 문서분류 인공지능 모델링을 실시하였다. 문서 분류 모델링을 위해서 2020년, 2021년 데이터를 학습 데이터로, 2022년 데이터를 평가 데이터로 구분하여 실시하였고 의사결정나무, SVM, 랜덤포레스트, XGBoost 알고리즘을 사용하였다. 평가지표는 정확도(Accuracy)를 사용하였고 결과는 <표 4-2>와 같다.

<표 4-2> 기계학습을 통한 문서 분류 결과

알고리즘	의사결정나무	SVM	랜덤포레스트	XGBoost
정확도	87.3%	89.0%	90.3%	90.9%

<표 4-2>의 정확도는 학습 데이터로 모델을 만든 후에 평가 데이터에 모델을 적용하여 평가된 값이다. 실험 결과 앙상블 모형인 XGBoost를 활용한 모델 결과가 가장 높은 정확도가 나왔다. 또한, 비교적 간단한 모형이면서 해석력이 좋은 의사결정 나무도 87.3%의 정확도를 보여주고 있다. 특히 문서를 분류하는데 걸리는 시간이, 평가 데이터로 사용하는 1,200개의 공문서 분류에 1초 정도였다. 이는 사단 사령부에 근무하는 실무자가 하루에 1시간씩 소요되는 업무를 자동화 시킬 수 있다는 것을 보여주는 것이다. 이를 육군 전체로 확장해보면 육군의 사단 사령부에만 적용한다고 가정하였을 때 하루에 40시간을 절약하는 효과가 있고, 문서 배부를 하는 담당자가 있는 여단부터 군단, 지상작전사령부, 육군본부 등 모든 계대로 확대한다면 시간 절약 효과뿐만 아니라 이로 인한 다른 중요한 업무에 집중할 수 있는 기회비용이 크게 향상 될 수 있을 것이다.

5. 결론 및 향후 연구

2019년 구글이 발표한 BERT(Bidirectional Encoder Representations from Transformers)와 ChatGPT(Chat Generative Pre-trained Transformer)의 기반이 되는 GPT-4 같은 대규모 사전학습 언어 모델은 인공지능 분야에서 혁신적인 변화를 가져왔다. 이러한 모델들은 감정 분석, 텍스트 분류 등 다양한 작업에서 높은 성능을 보여주며, 방대한 데이터를 사전학습 단계에서 학습했기 때문에 새로운 학습을 할 때 적은 양의 데이터로도 높은 성능을 달성할 수 있다. 군 인트라넷 환경에서 이러한 사전학습 모델을 활용하면 행정 업무의 효율성을 크게 향상시킬 수 있을 것이다. 그러나 사전학습 모델을 군에 도입하는 데는 소요 비용이나 보안 문제 등 다양한 현실적인 어려움이 존재한다. 이번 연구의 실험에서 사전학습 모델을 사용하지 않은 이유도 인트라넷 환경에서 활용하기 어려운 방법이기 때문이다.

본 연구의 실험 결과 분류 정확도는 약 90% 정도로 확인되었다. 이에 더해 파라미터 조정이나 새로운 변수 추가 등 다양한 시도로 본 연구보다는 더 높은 정확도로 향상시킬 수 있을 것이다. 일례로 구글이 데이터 과학자 커뮤니티인 캐글(Kaggle)에서 실시한 대회 중 ‘Bag of Words Meets Bags of Popcorn’이라는 유명한 대회가 있다. 이 대회는 인터넷 영화 데이터 베이스(IMDB)에서 나온 영화 평점 데이터를 활용한 문제로, 영화 리뷰 텍스트와 평점을 이용하여 해당 리뷰의 감정(긍정 혹은 부정)을 예측하는 이항 분류 문제였다. 대회 초반에는 정확도가 80% 후반을 기록하다가 최종 우승팀의 정확도는 99.3%였다. 이처럼 이번 연구 결과를 바탕으로 정확도 개선을 위한 향후 연구가 이어진다면 문서 배부를 인공지능으로 완벽히 대체할 수 있는 전자문서 분류기 개발이

가능할 것이라 기대된다.

군 내에서 생성되고 있는 인트라넷 자료, 수기 문서 등 국방 데이터를 활용해 인공지능 기술을 적용할 수 있는 분야는 무궁무진하며 특정 문제에 대한 인식과 해결에 관한 아이디어가 있다면 충분히 해결할 수 있다. 본 연구도 ‘단순하고 반복적인 이 업무를 자동화 시킬 수 없을까?’ 하는 생각으로 출발하여 데이터 수집 방법과 해결 방안에 대한 계획으로 시작하였다. 이처럼 인공지능을 잘 활용하면 간단한 업무를 자동화하고, 계획문서를 작성할 때 많은 도움을 받을 수 있다. 이러한 노력이 모여 머지않은 미래에 행정 업무 패러다임을 혁신적으로 변화시켜 인간의 창의적인 에너지를 보다 고도의 문제 해결과 전략적 사고에 집중할 수 있도록 하여 전반적인 업무 환경의 발전에 기여할 것이다.

참 고 문 헌

- [1] 육군본부. “육규 151 사무관리 및 일상명령 발령 규정 제35조”. 2023.
- [2] NVIDIA. 『Sharpening Its Edge: U.S. Postal Service Opens AI Apps on Edge Network』. 2021.
- [3] LawGeex. 『20 Top Lawyers Beaten by Legal AI』. 2018.
- [4] 박찬정, 성동수, 이건배. “기계 학습을 이용한 특허 문서의 자동 IPC 분류”. 한국정보기술학회논문지, 10(4). 2012.
- [5] 김성희, 엄재은. “기계학습을 이용한 문서 자동분류에 관한 연구”. 정보관리연구, 39(4), 2008.
- [6] 김정미, 이주홍. “Word2vec을 활용한 RNN기반의 문서 분류에 관한 연구”. 한국지능시스템학회, 27(6), 2017.
- [7] 김도우, 구명완. “Doc2Vec과 Word2Vec을 활용한 Convolutional Neural Network 기반 한국어 신문 기사 분류”. 정보과학회, 44(7), 2017.
- [8] 송진주, 강승식. “Doc2Vec을 이용한 특허 문서 자동 분류”. 한국정보처리학회, 2019.
- [9] 정지수, 지민규, 고명현, 김학동, 임현영, 이유림, 김원일. “문서 유사도를 통한 관련 문서 분류 시스템 연구”. 방송공학회논문지, 24(1), 2019.
- [10] 이창용, 문호석. “텍스트마이닝을 이용한 북한 보도동향 및 북한 도발과의 연관성 분석”. 국방연구, 2016.

저 자 소 개



윤영빈 (E-mail: yb6470@gmail.com)

2014 육군사관학교 토목환경공학 학사

현재 국방대학교 군사운영분석전공 석사과정

관심분야 : 텍스트 마이닝, 데이터과학, 국방 AI



문호석 (E-mail: hsmoon0329@gmail.com)

1994 육군사관학교 화학과 학사

2003 고려대학교 전자공학 석사

2006 고려대학교 산업공학 박사

2010 고려대학교 통계학 박사

현재 국방대학교 국방과학학과 교수

국가안전보장문제연구소

군사과학연구센터장

관심분야 : KCTC 빅데이터분석, 데이터과학,
국방 AI, M&S

