

저작권양도서

(Copyright Transfer Form)

소속 : 인하대학교 정보통신공학과

성명 : 김소원, 서혜빈

학번 : 12181748, 12181785

논문제목 : BERT를 이용한 해양사고 보고서 분석

본인은 상기 논문을 2021학년도 2학기 정보통신프로젝트 최종 보고서 겸 결과 논문으로 제출하고자 합니다. 본 논문의 내용은 저자가 직접 연구한 결과인 것과 이전에 출판된 적이 없음을 확인합니다. 또한 공저자와 더불어 인하대학교 정보통신공학부에서 발간하는 논문집에 본 논문을 수록하는 것을 허락하며 제반 저작권을 정보통신공학부에 양도합니다.

2021년 12월 24일

저자 : 서혜빈



저자 : 김소원



정보통신공학과장 귀하

BERT를 이용한 해양사고 보고서 분석

Analysis of marine accident reports using BERT

김소원, 서혜빈
(Sowon Kim and Hyebeen Seo)

요약: 본 논문은 BERT 모델을 이용하여 두 가지 목적을 바탕으로 해양사고 보고서 내의 문장들을 자연어 처리 기술을 통해 분석하고 있다. 보고서 내의 문장 classification 및 clustering을 진행하고, 문장 내에서 중요 keyword를 추출하는 것이 목적에 해당한다. KoSentenceBERT를 이용하였으며, 문장 벡터를 얻는 방법 중 문장의 평균 벡터를 pooling 하는 방식을 채택했다. t-SNE, WordCloud와 같은 시각화뿐만 아니라, 실루엣 스코어, BLEU score 등의 evaluation을 통해 연구 결과를 제시하고 있다.

Abstract: In this paper, we use the BERT model to achieve two objectives and analyse the sentences in the marine accident report. First, we classify and cluster the sentences in the marine accident reports. To do this, we apply the mean vector of the sentence to obtain the sentence vector. Through visualization using t-SNE and WordCloud, we demonstrate the classification and clustering results of similar sentences. In addition, we extract important keywords from the sentences and evaluate our approach using the silhouette score and BLEU score.

Keywords: deep learning, marine accident, BERT, KoBERT, KoSentenceBERT, cosine similarity, natural language processing, data analysis

1. 서론

1. 개발동기

해양수산부 해양안전심판원에 따르면 해양사고 추이가 매년 증가 추세이다. 2016년에는 해양사고 발생 척수가 2,549건, 2020년에는 3,535건으로, 5년 사이에 사고 건수가 크게 증가한 것을 알 수 있다.[1] 따라서 해양사고의 원인을 파악하고 이를 예방하기 위한 노력의 필요성이 대두되고 있다.

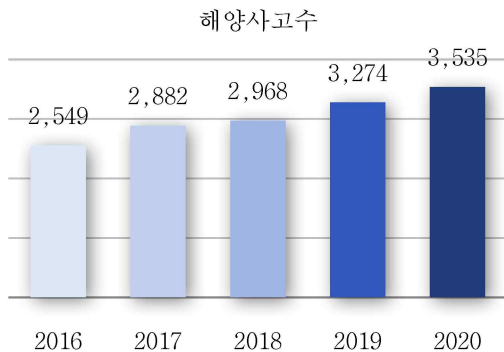


그림 1. 최근 5년 해양사고 통계

해양사고 보고서(이하 재결서)는 그림 2와 같이 발생한 사고에 대해 사고 원인, 사고 당일 날씨, 사고 원인 제공자, 집행 결과 등 자세하게 기술하고 있다.[2] 그렇기에 사고에 관해 한눈에 파악하기 어려우며, 읽고 이해하는 데에 많은 시간이 소요된다는 단점이 있다. 따라서 재결서를 자연어 처리 기술을 통해 분석하고자 한다.

해양사고관련자	B(****, ***)
장 구 취 지	「해양사고의 조사 및 심판에 관한 법률」 제38조 제1항의 규정에 의하여 심판 청구함 관여조사관 김 구 종
주 문	이 작업원부상자건은 프린세스호가 앞천항에서 하역작업 중, 1번 화물창 하(下)갑판 화물격재 완료 후 선원들이 중(中)갑판 맨줄 커버를 닫지 아니한 상태에서, 중갑판에 격재된 화물로 맨줄 주변이 어두워 이동하던 작업자들이 맨줄로 추락하여 발생한 것이나, 하역감독이 작업자들의 이동통로 안전을 미리 확인하지 아니한 것도 일부 원인으로 작용하였다. 해양사고관련자 A에게 시정할 것을 권고한다. 해양사고관련자 B에게 개선할 것을 명한다.
이 유	

그림 2. 재결서 일부 발췌

2. 목적

본 논문은 재결서 내에서 두 개의 목적을 실현하고자 하며, 다음과 같다.

- 목적 1) 문장 classification & clustering
- 목적 2) 문장에서 중요 keyword 추출

목적 1을 통해 유사 문장끼리 분류하고, 군집화를 진행한다. 그리고 목적 2를 통해 문장에서 중요 키워드를 판단하여 해당 사고에 대해 보다 쉽게 파악하고자 한다.

II. 본론

1. 이론적 배경

1.1 BERT

BERT(Bidirectional Encoder Representations from Transformers)는 구글에서 개발한 NLP 기술로, 자연어 처리 분야에서 좋은 성능을 내는 범용 language model이다. BERT를 사용한 모델링 과정은 다음과 같다. 관련 대량 corpus를 encoder가 embedding 한 후, 이를 전이하여 fine-tuning을 진행한다.[3]

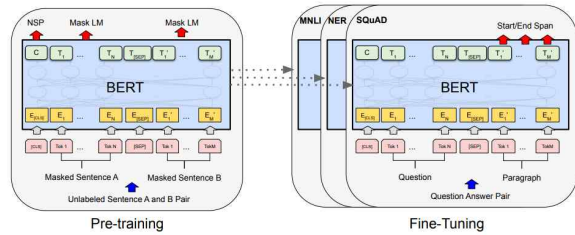


그림 3. BERT pre-training & fine-tuning 구조

1.2 KoBERT

KoBERT는 SKT open source로, SKT에 따르면 기존 BERT의 한국어 성능 한계를 극복하기 위해 개발되었다. 한국어 문장으로 이루어진 대규모 corpus를 학습하였으며, 한국어의 불규칙한 언어 변화의 특성을 반영하기 위해 데이터 기반 tokenization 기법을 적용하여 기존 대비 27%의 token으로 2.6% 이상의 성능 향상을 이끌었다. 파이토치(pytorch), 텐서플로우(tensorflow) 등을 포함한 다양한 딥러닝 API를 지원함으로써 많은 분야에서 언어 이해 서비스 확산에 기여하고 있다.[4]

1.3 KoSentenceBERT

KoSentenceBERT(이하 KoSBERT)는 BERT로부터 문장 embedding을 얻을 수 있는 SentenceBERT의 한국어 버전으로, KoBERT를 기반으로 하고 있다. 즉, BERT의 문장 임베딩 성능을 우수하게 개선한 모델이다.[5]

문장 벡터를 얻는 방법에는 세 가지가 있다. 첫째로, BERT의 [CLS] 토큰의 출력 벡터를 문장 벡터로 간주한다. 둘째로, BERT의 모든 단어의 출력 벡터에 대해서 평균 pooling을 수행한 벡터를 문장 벡터로 간주한다. 셋째로, BERT의 모든 단어의 출력 벡터에 대해서 max pooling을 수행한 벡터를 문장 벡터로 간주한다.[5] 본 논문에서는 두 번째 방법을 활용하였으며, 문장의 평균 벡터를 pooling 하는 방식을 채택했다.

1.4 Additive Margin Softmax

Additive Margin Softmax(이하 AdMSoftmax)는 L-softmax와 A-softmax처럼 m 부터 θ 까지 곱하는 대신, $\psi(\theta)$ 를 다음과 같이 변경함으로써 가산적인 방식으로 margin을 도입한 것이다.[6]

$$\psi(\theta) = \cos\theta - m$$

이는 L-softmax와 A-softmax의 $\psi(\theta)$ 보다 훨씬 간단하며 성능 또한 우수하다. 그리고 AdMSoftmax는 A-softmax와 유사한 무게와 편향을 정규화하지만 cosine 값을 확장하는 새로운 hyperparameter를 도입한다. 마지막으로 AdMSoftmax의 loss는 다음과 같이 정의할 수 있다.[6]

$$L_{AdMSoftmax} = -\frac{1}{n} \sum_{i=1}^n \log \frac{e^{s(\cos\theta_{y_i} - m)}}{e^{s(\cos\theta_{y_i} - m)} + \sum_{j=1, j \neq y_i}^c e^{s \times \cos\theta_j}}$$

2. 데이터셋

중앙해양안전심판원 사이트에서 수집한 재결서의 총개수는 816개이다. 각 재결서를 문장 별로 split 하면 약 73000개의 문장이 발생하고, 이후 전처리 과정을 거친다. 이때 train, test, validation set의 비율은 8:1:1이다. 최종 train set은 아래와 같다.

표 1. 최종 train set

index	text	label
0	당직을 맡은 항해사 A는...	0
1	이 선박은 사고 발생 전...	0
2	해양사고 관련자 A는...	1
3	사랑호는...	0
4	선박사이에는...	4
...

최종 데이터셋의 열 수는 다음과 같다. train set은 20438개, test set은 3591개, validation set 3590개이다.

3. 모델 설계

3.1 설계 블록도

3.1.1 KoSBERT 알고리즘

목적 1의 경우 문장의 평균 벡터를 pooling 한다. 목적 2의 경우 단어를 학습하기 때문에 각 단어의 벡터를 pooling 한다. 그리고 각 벡터값을 통해 코사인 유사도를 구하여 연구를 진행하였다.

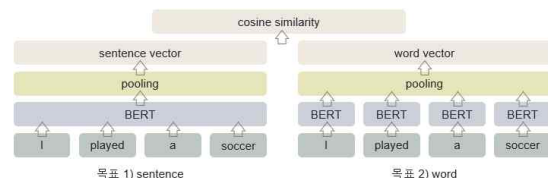


그림 4. 각 목적에 따른 KoSBERT 구조

3.1.2 설계 블록도

본 연구의 전반적인 flow는 그림 5와 같다.

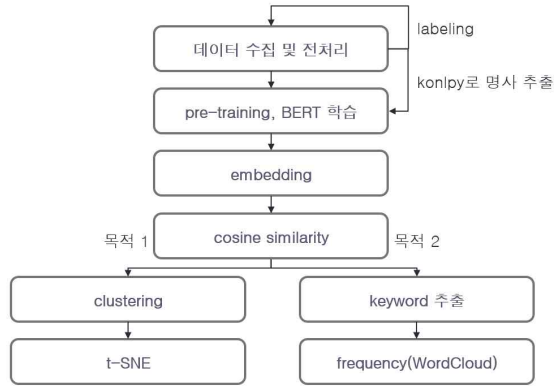


그림 5. 전체 설계 flow

3.2 데이터 수집 및 전처리

3.2.1 크롤링(crawling) 및 전처리

사용 데이터는 재결서로, 중앙해양안전심판원 사이트에서 수집하였다. selenium라는 프레임워크를 이용하였으며, webdriver라는 API를 통해 chrome 브라우저를 제어하였다.

모든 재결서를 하나의 dataframe으로 만든 후, 'Wn'을 기준으로 split을 진행하였다. 이는 한 행에 하나의 문장이 있도록 하기 위함이다. 그리고 정규표현식을 이용해 특수문자를 제거하여 전처리를 진행하였다.

3.2.3 라벨링(labeling)

BERT는 self-supervised 모델이다. 따라서 label 값을 바탕으로 문장을 학습한다. 그러나 본 팀이 수집한 데이터는 지정된 label 값이 존재하지 않으므로, 일일이 label 값을 지정해준다. multi-class labeling 방식을 통해, 예로 문장 내에 화재, 불 등의 단어가 포함될 경우 1로 준다. 즉 유사한 사고원인별로 동일한 숫자를 부여한다. 지정 label은 다음과 같다.

표 2. 사고원인별 label 값

label	단어
0	해당 사항 없음(사고와 관련 없는 문장)
1	화재, 폭발, 불, 발열, 열기, 화염, 열감
2	좌초, 부유물감김, 운항저해, 접촉, 충돌, 침몰, 전복, 해양오염, 오염, 장애물, 기관손상, 조타장치손상, 추진축계손상, 시설물손상, 손상
3	부상, 실종, 사망, 행방불명, 상처
4	경계소홀, 졸음운항, 위법, 위반, 미준수, 부적절, 불이행, 음주, 알코올, 술

3.2 BERT pre-training

기존 KoSBERT 모델을 이용하면 재결서 데이터에 대한 정보가 없어 결과를 도출해도 성능이 떨어질 것이라 예상했다. 따라서 pre-training을 진행하며 모델에 재결서 데이터를 넣어 한 번 더 학습을 해 성능을 높이고자 했다.

아래의 표는 재결서를 이용한 pre-training 전과 후의 평균 코사인 유사도를 구한 것이다. 평균 코사인 유사도를 바탕으로 성능을 판단하였다. 실제 pre-training 결과를 표 3을 통해 확인해 보면 후의 결과가 더 좋음을 확인할 수 있다. 따라서 재결서로 pre-training을 진행한 결과를 본 연구에 사용하였다.

표 3. pre-training 결과

전후	평균 코사인 유사도
재결서 학습 전	0.34
재결서 학습 후	0.8230

3.3 BERT fine-tuning 및 training

KoSBERT를 기반으로 전처리 된 데이터셋을 통해 fine-tuning 및 training을 진행하였다. 이제 fine-tuning에 해당하는 epoch 값을 중점적으로 한 hyperparameters 선정 과정뿐만 아니라, training 과정에서의 classification을 중점적으로 살펴보자.

3.3.1 hyperparameters

KoSBERT의 hyperparameters는 다음과 같이 설정하였다.

$L = 15$, $H = 768$, $A = 12$

max length = 128

optimizer = Adam

learning rate = $2e-5$

hidden size = 768

표 4는 적절한 epoch 횟수를 찾기 위해 epoch를 1부터 3까지 증가시켜 학습을 진행한 결과를 표로 나타낸 것이다. 이때, 다른 hyperparameters는 앞의 값을 그대로 사용하였다.

표 4. epoch 별 결과 비교

epoch	loss	코사인 유사도
epoch 1	0.23	0.8230
epoch 2	0.30	0.61
epoch 3	0.27	0.71

위 결과를 통해 epoch 횟수를 1로 설정했을 때의 loss 값이 0.23으로 가장 적을 뿐만 아니라 코사인 유사도도 가장 높게 도출되었음을 확인했다. 따라서 최종 모델의 epoch 값을 1로 지정하였다.

epoch 횟수가 2일 때와 3일 때의 결과도 우수하

였지만 epoch를 1로 설정하고 학습을 진행했을 때의 성능이 가장 높아 최종 모델의 epoch 값을 1로 지정하였다. 주어진 환경의 한계로 epoch 횟수를 더 증가시켜 연구를 진행하지는 못했지만, 만약 epoch 횟수를 크게 증가시킬 수 있었다면 더욱 최적의 epoch 횟수를 선정할 수 있었을 것이라 예상된다.

3.3.2 classification

앞서 지정했던 라벨의 수는 0~4로, 총 5개이다. 그러나 현재 사용 중인 BERT 모델은 hidden size가 768로, 길이가 768인 벡터로 embedding 된다. 따라서 768차원을 5차원으로 분류하는 classification 과정을 진행해야 한다. 이때 5차원으로 분류하는 이유는 라벨의 수가 5개이기 때문이다. 본 논문에서는 classification 과정에서 AdMSoftmaxLoss를 이용하였다.

3.3.3 training

데이터셋의 text와 label을 모델에 넣어 함께 학습을 진행했다. back propagation 할 때 기울기를 계산하고, gradient가 특정 단어에 너무 fitting 되지 않도록 상한선을 제한했다. 이를 통해 layer가 점점 깊어질수록 back propagation 때 gradient를 계산하면 값이 너무 커지거나 작아지는 현상이 발생하는 것을 막는다. 이후 학습 과정에서 앞서 언급했던 AdMSoftmaxLoss를 이용하여 분류, 즉 classification을 수행한다. 마지막으로 문장의 평균 벡터를 output으로 pooling 하는 과정을 거친다.

3.4 clustering 및 t-SNE

3.4.1 clustering

서론에서 언급했던 목적 1은 유사 문장끼리 군집화하는 것이다. 본 논문에서는 clustering 기법 중 k-Means를 사용하였다. 최적의 hyperparameters를 찾기 위해 여러 parameter를 설정하여 결과를 도출한 결과는 아래와 같다. 표 5에서의 score는 실루엣 스코어를 의미한다.

표 5. k-Means hyperparameters 비교 결과

init	n_init	max_iter	tol	score
k-means++	10	300	1e-04	0.265
k-means++	10	500	1e-04	0.264
k-means++	10	800	1e-04	0.264
k-means++	30	300	1e-04	0.264
k-means++	30	300	3e-04	0.264
k-means++	10	300	3e-04	0.262
random	10	300	1e-04	0.264
random	10	300	3e-04	0.262

표 5를 통해 2번째 행이 실루엣 스코어가 가장 높게 도출되는 것을 확인할 수 있다. 따라서 init은 k-means++로, n_init은 10으로, max_iter는

500, tol은 1e-04로 선정한다.

k-means의 경우 군집의 수 k를 지정해야 한다. 따라서 최적의 군집 수를 선정하기 위해 시각화한 결과는 그림 6과 같다.

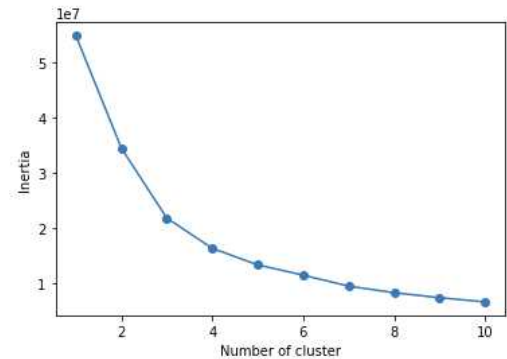


그림 6. k-means의 k 수 선정

위 결과를 통해 k의 수는 5로 선정하였다.

3.4.2 t-SNE

t-SNE(t-distributed stochastic neighbor embedding)는 기존 라이브러리 sklearn에서 제공하는 TSNE를 이용하였다. 문장 평균 벡터인 786차원의 embedding 값을 바탕으로 n_components의 값을 3으로 주어 3차원으로 축소한다. 이후 classes의 값에 앞서 clustering으로 구한 군집 결과 label을 줌으로써 각 군집을 표현해주었다.

3.5 keyword 추출

서론에서 언급했던 목적 2는 문장에서 중요 keyword를 추출하는 것이다. BERT 모델은 self-supervised 학습을 기반으로 하고 있으나 현재 데이터셋에 존재하는 수많은 단어들을 일일이 label 값을 부여하는 것은 어려운 일이라고 판단했다. 따라서 unsupervised 학습으로 진행하였다.

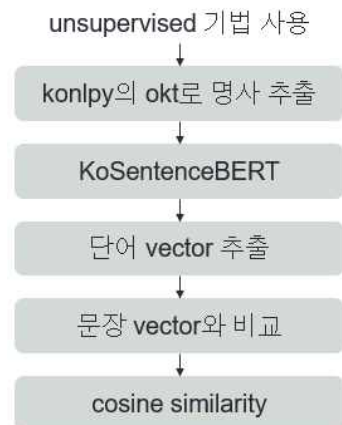


그림 7. 중요 keyword 추출 과정

가장 먼저 전처리된 데이터셋에서 konlpy의 okt를 이용해 'text' 열에서 명사를 추출하였다.

표 6. 명사 추출 결과

text	nouns
당직을 맡은 항해사 A는...	['당직', '항해사', ...]
이 선박은 사고 발생 전...	['선박', '사고' ...]
해양사고 관련자 A는...	['해양', '사고', ...]
...	...

각 문장에서 추출된 명사는 해당 문장에서 추출되었음을 구분해주기 위해 2차원 리스트로 정리하였다.

이후 KoSBERT를 이용하여 query에는 한 문장을, corpus에는 해당 문장에서 추출된 단어를 반복문으로 하나씩 넣어주며 한 문장과 한 단어 사이의 코사인 유사도를 구하였다. 그리고 문장 별로 코사인 유사도가 높은 순대로 sorting을 진행하였다. 즉, A 문장과 A 문장 내에서 추출된 단어들을 하나씩 반복 비교하였고, 코사인 유사도가 높으면 중요 keyword로 판단하였다.

III. 결과 및 분석

1. 목적 1)의 결과

1.1 2차원 t-SNE

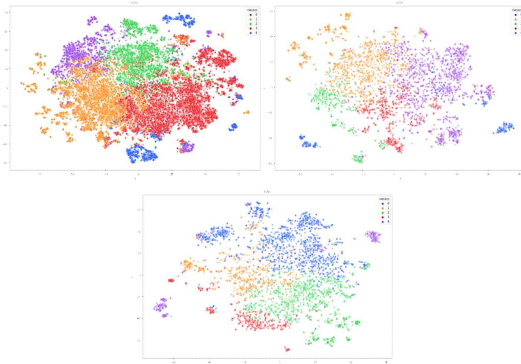


그림 8. 각 train, test, valid t-SNE 결과

1.2 3차원 t-SNE

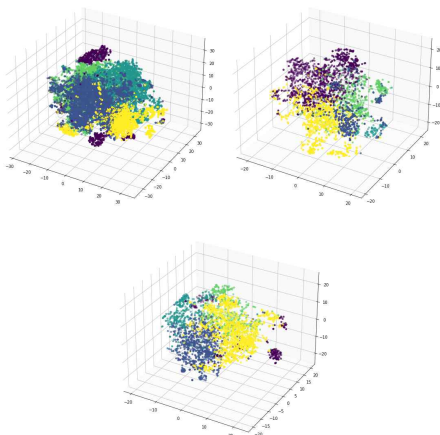


그림 9. 각 train, test, valid t-SNE 결과

t-SNE 2차원, 3차원 시각화 결과 육안으로는 군집화가 비교적 잘 되었다고 판단된다. 그러나 정확

한 evaluation을 위해 실루엣 스코어를 통해 판단하고자 한다.

1.3 Silhouette score

$$s(i) = \frac{(b(i) - a(i))}{(\max(a(i), b(i)))}$$

Silhouette score를 구하는 공식은 위와 같다.[7]

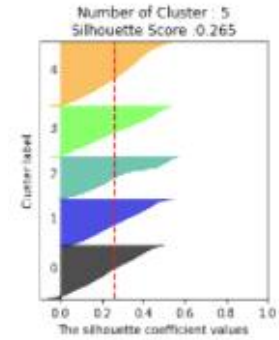


그림 10. Silhouette score 시각화

그림 10에서 확인할 수 있듯이 개별 군집의 평균 실루엣 계수 값이 비교적 균일하게 위치하고 있다. 또한 평균 실루엣 스코어는 0.265가 도출된 것을 확인할 수 있다.

2. 목적 2)의 결과

2.1 중요 keyword 추출 결과

sentence가 다음과 같을 때 아래 keyword가 중요 keyword로 추출된 것을 확인할 수 있다.

sentence : 이 화재사건의 최초 목격자 C는 부두에 정박 중인 대왕호 기관실 쪽에서 불꽃과 연기이는 것을 보고 신고하였다 진술하였다

keyword : ‘화재’, ‘부두’, ‘목격자’, ‘최초’, ‘대왕호’, ‘신고’, ‘연기’

2.2 BLEU score

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N W_n \log p_n\right)$$

BLEU(Bilingual Evaluation Understudy) score의 공식은 위와 같다.[8] 앞선 과정을 통해 추출된 중요 keyword와 사람이 판단한 중요 키워드 간의 유사성을 수치화한 것으로, 본 논문에서는 표본 약 100여개를 랜덤 추출하여 일일이 BLEU score 값을 구하였다. 그 후, 평균 BLEU score를 구하여 중요 키워드가 잘 추출되었는지에 대해 판단했다.

표 7. 평균 BLEU score 결과

평균 BLEU score
0.864

평균 BLEU score가 0.864가 도출되었고, 해당

결과를 통해 중요 keyword가 잘 추출되었음을 판단할 수 있다.

2.3 WordCloud

그림 10의 첫 번째는 기존 전처리된 데이터셋에서의 단어 빈도를, 두 번째는 중요 keyword 추출 결과에서의 단어 빈도를 WordCloud로 시각화한 결과이다.



그림 11. 중요 키워드 추출 전후 WordCloud

두 결과가 다르게 나온 것을 확인할 수 있다. 두 번째 결과를 보면 첫 번째 결과와 달리 해양사고, 충돌 등의 단어의 빈도가 커진 것을 알 수 있다. 이는 앞서 분석을 통해 문장에서 중요 키워드, 즉 사고와 관련된 단어가 추출되었기 때문임을 예상할 수 있다.

그림 12는 앞서 clustering 결과, 군집별로 중요 키워드를 추출한 후 빈도를 WordCloud로 시각화한 결과이다. 각 순서대로 0 군집부터 4 군집까지이다.



그림 12. 군집 별 중요 키워드 빈도 WordCloud

하나의 군집을 예로 결과를 살펴보자. 0 군집에 해당하는 첫 번째 WordCloud를 보면 맑은, 기상, 파고 등 날씨와 관련된 단어의 빈도가 큰 것을 확인할 수 있다.

IV. 결론

본 논문에서는 BERT를 이용해 재결서를 분석하여 문장 분류 및 중요 키워드 추출 방법에 대해 제

안하였다. 2개의 목적을 달성하기 위해 supervised와 unsupervised 학습을 모두 이용하여 KoSBERT 모델에 적용하였다. 목적 1인 문장 classification 및 clustering의 경우, supervised 학습으로 진행했으며 이를 위해 문장마다 label을 부여하였다. 목적 2인 중요 keyword 추출의 경우 각 단어마다 label을 부여하는 것은 어렵다고 판단하여 unsupervised 학습으로 진행하였다.

pre-training과 fine-tuning, training 과정은 목적 구분 없이 공통적으로 진행하였으며, pre-training 과정에서 재결서를 한 번 더 학습해 줌으로써 전보다 훨씬 높은 코사인 유사도를 유도했다. 이는 기존 모델과는 달리 pre-training을 통해 해양 사고에 관한 학습을 진행했기 때문이다. fine-tuning 과정에서는 hyperparameters를 선정했다. epoch 횟수를 1부터 3까지 지정하여 모델을 학습시키면서 loss와 학습 결과 코사인 유사도를 구하였다. 결과적으로 epoch 1일 때 loss가 가장 적고, 코사인 유사도는 높게 도출되어 epoch 횟수는 1로 선정하였다. 마지막으로 training 과정에서는 AdMSoftmaxLoss를 이용하여 classification을 진행하였다.

이후 목적 1을 위해 k-Means를 이용하여 clustering을 진행하고, clustering label 값을 저장했다. 이후 t-SNE를 이용하여 768차원의 corpus embedding을 2, 3차원으로 축소하고, classes에 cluster label을 주어 군집을 표시해 주었다. 이 결과의 evaluation은 실루엣 스코어로 진행하였으며, 각 군집의 실루엣 계수 값이 비교적 균일하고 0.265의 값을 도출하였다.

목적 2의 경우 konlpy를 이용해 문장별로 명사들을 추출한 후, 한 문장과 한 단어 사이의 코사인 유사도를 구하였다. 그리고 코사인 유사도가 높은 순대로 정렬을 진행해 중요 키워드를 판단해 주었다. evaluation은 BLEU score를 이용했으며, 0.864라는 비교적 높은 결과를 도출하였다.

본 논문의 활용 가능성을 통한 기대효과는 다음과 같다. 더 세부적인 분석을 통해 해양 사고에 대한 인사이트를 얻을 수 있다는 것이다. 어떤 사고가 발생했는지 이미 알고 있을 때, 해당 사고별로 재결서를 취합한다. 이후 유사 문장끼리 분류 및 군집화를 진행하고, 사고 관련 단어가 들어간 문장만 추출한다. 그리고 중요 키워드를 추출하여 사고 원인 등에 대한 잠재 변수 또는 또 다른 인사이트를 알아낼 수 있다는 점 등을 기대효과의 예로 들 수 있다.

참고문헌

- [1] 중앙해양안전심판원 해양사고 통계(승인번호 : 제123020호)
- [2] 중앙해양안전심판원 해양안전심판재결 재결서
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: 언어 이해를 위한 양

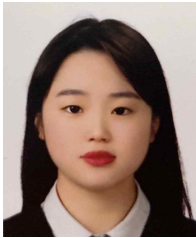
방향 트랜스포머 사전 학습(BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding), Dec. 2018.

- [4] SKT KoBERT Open Source, <https://sktelecom.github.io/project/kobert/>
- [5] 딥러닝을 이용한 자연어 처리 입문, sentence Bert
- [6] Additive Margin Softmax Loss, <https://arxiv.org/pdf/1801.05599.pdf>
- [7] Choi, Hyunwong, Comparative analysis of electricity consumption at home through a silhouette-score prospective, Apr. 2019.
- [8] 김고은, 어텐션 기반 시퀀스 투 시퀀스 모델을 이용한 감정기반 한국어 대호 응답 생성, Jul. 2020.



서 혜 빈

2018년 ~ 현재 인하대학교 정보통신공학과 학사 과정 재학 중.
2022년 8월 졸업 예정.
관심분야는 데이터 분석, 인공지능



김 소 원

2018년 ~ 현재 인하대학교 정보통신공학과 학사 과정 재학 중.
2022년 2월 졸업 예정.
관심분야는 인공지능, 백엔드 개발