

AI를 이용한 국방 기사의 효과적인 요약 모델 설계 및 성능 평가

남예린¹⁾, 최윤영¹⁾, 최종근²⁾, 권혁진^{2)*}

1) 서울과학기술대학교 인공지능응용학과, 2) 서울과학기술대학교 국방인공지능응용학과 교수

Designing Effective Summary Models for Defense Articles with AI and Evaluating Performance

Yerin Nam^{1)*}, YunYoung Choi¹⁾, JongGeun Choi²⁾, HyukJin Kwone^{2)*}

1) *SeoulTech University Research Institute*, 2) *SeoulTech University Professor*

Abstract : With the development of the Internet, the information in our lives has become fast and diverse. Especially in the field of defense, articles and information are pouring in from various sources every day, and fast information selection, understanding, and decision-making are required in the ever-changing situation. It is very cumbersome to go from platform to platform and read articles one by one to get the information you need. To solve this problem, this research aims to save time and provide quick access to the latest information by allowing you to quickly grasp key information from summarized content without having to read the entire article. This can improve efficiency by allowing defense professionals to focus more on important tasks rather than extensive information search and analysis.

Key Words : Defense, Information overload, Document summarization, News articles, Artificial intelligence, Efficiency, Military intelligence, Text summarization.

Received: May 31, 2024 / **Revised:** June 17, 2024 / **Accepted:** June 19, 2024

* corresponding author : HyukJin Kwone / SeoulTech University Professor /

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

1. 서론

최근 인공지능 연구 분야에서 다양한 기술들이 등장하면서, 국방 분야에 적용되는 인공지능 역시 많이 등장하고 있다. 과학기술정책연구원이 2020년 국방 전문가 40인(군인, 연구소, 방산기업, 대학)을 대상으로 실시한 설문 조사에 따르면 국방 인공지능 기술 도입에 대해 전체 응답자의 94%가 매우 중요하다고 보았다.[1] 이처럼 기술이 발전함에 따라 인공지능 기술을 군에 적용하고자 하는 노력이 증가하는 추세이다.

국방에서 활용될 인공지능 모델은 다음과 같은 특징이 있다. 첫째, 국방 인공지능 모델을 위한 학습 데이터가 부족하다. 국방 데이터는 보안규정 및 군 전용 네트워크 사용 등으로 접근이 제한되는 경우가 많으며, 데이터의 양이 적거나 불균형하여 학습하기에 불충분한 데이터일 가능성이 높다. 둘째, 성능에 신뢰성이 확보되어야 한다. 국방 인공지능 모델은 전투와 같이 어렵고 힘든 결정에 활용될 수 있다. 따라서 민간에서 사용하는 모델에 비해 그 요구 성능이 더욱 높아야 한다. 또한, 신속·정확하게 결과를 도출해야 하고, 그 결과에 대한 신뢰성 확보를 위해 설명력과 해석력이 따라야 한다. 셋째, 사이버 위협에 대한 보안대책이 강구되어야 한다. 국방 분야는 위협국에 의한 사이버 위협이 높다. 따라서 적대적 공격을 통해 인공지능 모델을 무력화시킬 가능성이 높으므로 이에 대응할 수 있는 강건함을 가져야 한다. 넷째, 인공지능프로그램의 간결성이 요구된다. 국방 인공지능 모델은 드론과 같은 소형 플랫폼으로 고성능 컴퓨팅 지원이 제한되는 환경에서 활용될 수 있다. 따라서 이러한 환경에서 사용될 수 있도록 간결하게 만들 필요가 있다.[2]

본 연구에서는 AIHub 데이터를 이용하여 딥러닝 모델인 T5와 BART 등을 학습시켜 국방 관련 뉴스 기사를 효과적으로 요약하는 연구를 수행한다. AIHub는 다양한 도메인에서의 데이터를 제공하는 플랫폼으로, 국방 분야의 특수성을 고려하여 이 데이터를 활용함으로써 모델의 성능을 향상시키는 것

에 중점을 두었다. T5와 BART는 최신의 딥러닝 모델 중 하나로, 특히 자연어 처리와 문서 요약에 강점을 지닌다. 이러한 모델을 AIHub 데이터에 적용함으로써 국방 분야의 뉴스 기사를 효과적으로 추출하고, 그 내용을 간결하게 요약하는 데 있어서 어떤 결과를 생성하는지를 분석하고자 한다.

본 논문은 국방 분야에서의 정보 수집과 이해를 돕기 위해, 뉴스 기사에 효과적인 요약을 통해 빠르고 정확한 의사결정을 지원하는 것을 목표로 한다. 특히 기존 연구에서는 국방 분야 문서 요약에 인공지능 적용에 있어서 모델의 신뢰성, 보안성 및 간결성 등 여러 가지 문제점을 다루지 않았으며, 본 연구에서는 이러한 문제를 보다 구체적으로 해결하고자 하였다. 모델의 신뢰성 확보를 위해 성능이 어느 정도 검증된 딥러닝 모델인 T5와 BART를 사용하였다. 이러한 모델은 자연어 처리와 문서 요약에 강점을 가지며, 특히 국방 분야의 특수한 요구 사항을 충족시킬 수 있는 성능을 보여준다. 또한, 모델의 요약 간결성을 확보하기 위해 제약 조건을 설정하고 모델 추론을 진행하였다. 이를 통해 신속함을 요구하는 국방 분야의 간결성과 신속성을 충족하는 것을 연구에서 입증한다.

또한, 본 연구는 기존 연구와는 달리 온디바이스(on-device) 환경에서 작동 가능한 모델을 제안하였다. 이는 하나의 기기 자체적으로 정보를 처리할 수 있어 클라우드 서버나 외부 연결이 필요 없다는 점에서 큰 차별점을 가진다. 이를 통해 군 내에서 다양한 기사 및 정보 파악 업무에 드는 인력을 대체할 수 있는 가능성을 제시하고자 한다.

2. 관련연구

2.1 Transformer

Transformer는 어텐션 메커니즘(Attention Mechanism) 기반의 구조를 가지고 있다. 이는 Self-attention 메커니즘을 활용하여 입력 시퀀스의 각 위치에 대한 정보를 유연하게 고려할 수 있게

한다. 이러한 구조는 병렬 처리가 용이하며, 기존의 RNN 및 LSTM과는 달리 순차적인 계산을 수행할 필요가 없어 정보의 병목을 개선한다. 또한, 어텐션 메커니즘과 잔차 연결을 활용하여 높은 병렬 처리 성능을 보이고 이는 정보 처리의 신속성을 높인다. 또한, 각 서브층에 대한 층 정규화를 도입하여 학습을 안정화하고 모델의 성능을 향상시키는 효과를 얻는 모델이다.[3]

2.2 BART

BART는 Facebook에서 개발한 자연어 처리를 위한 딥러닝 모델 중 하나이다. Transformer의 양방향 인코더-디코더 구조를 기반으로 한다. 이 구조를 통해 BART는 문맥을 더 효과적으로 파악하고, 생성 과정에서 입력 정보를 유지할 수 있다. 특히, 문서 및 문장 요약 작업과 같이 입력과 출력 간의 상호 작용이 필요한 작업에 적합하다. 학습은 Masked Language Model (MLM)을 활용하여 이루어진다. 사전 훈련 단계에서 BART는 문장 내에서 일부 단어를 가리고 이를 복원하도록 학습함으로써 문맥을 고려한 언어 모델을 구축한다. 이는 다양한 자연어 처리 작업에 대한 사전 훈련을 강화하며, 입력 문장의 의미를 효과적으로 파악한다. 또한, 문서 및 문장 요약을 위해 문장의 순서를 중요하게 고려하여 학습한다. 이는 생성된 문장이 자연스럽게 흘러가고 문맥을 유지하도록 도와준다.[4]

2.3 T5

T5는 Text-To-Text Transfer Transformer의 약자로 Google에서 개발된 딥러닝 모델이다. Transformer 아키텍처를 기반으로 하며, 특히 어텐션 메커니즘을 통해 입력 토큰 간의 상호작용을 학습한다. 또한, 대량의 다양한 데이터를 사용하여 사전 훈련된 언어 모델을 만들고, 이를 특정 작업에 미세조정하여(Fine-Tuning) 성능을 높이는 전이 학습 방식을 채택한다. T5는 텍스트 형태의 입력을 받아 텍스트 형태의 출력을 생성하는 유연하고 강력

한 모델로 평가되고 있으며, 텍스트 기반 작업의 통일된 접근 방식은 자연어 처리 분야에서 주목을 받고 있다.[5]

2.4 Text Summarization

자동 텍스트 요약 기술은, 본문 내에서 추출한 단어를 조합한 추출 요약 (Extractive summarization) 과 본문에 쓰이지 않은 새로운 단어를 사용하여 문장을 만드는 생성 요약 (Abstractive summarization) 으로 나누어 볼 수 있다. 추출 요약은 생성 요약 대비 상대적으로 간단한 방법이며, 문서에서 중요하다고 판단되는 문장들을 하나의 요약본으로 생성하는 방법이다. 생성 요약은 원본을 재해석하여 본문과 다른 표현으로 요약을 생성한다. 추출 요약보다 방식이 복잡한 대신에 현실에 더 적합하여 성능 면에서는 생성 요약이 더 뛰어나다.[6]

추출 요약의 경우, 각 문장에 0 또는 1의 라벨을 부여하는 문제로 정의하여 요약본을 생성하고자 한 연구가 있었다. 해당 연구에서는 BERT의 변형인 BERT-SUM 모델을 활용하는데, 이는 BERT가 Masked-Language Model로 학습되어 출력 벡터가 문장이 아닌 토큰에 기반해 요약본 생성에 적합하지 않기 때문이다. 요약본 생성 문제에 적합하게 만든 BERT-SUM 모델로 입력된 문장들을 각각 벡터로 나타낸다. 그 후에 트랜스포머 층을 쌓아 해당 벡터로부터 문장 수준의 특징을 포착하고, 요약본 생성에 활용하도록 한다. 모델의 출력 레이어는 Sigmoid classifier로 구성하여 문장마다 0 또는 1의 라벨을 부여할 수 있도록 했고, 이를 통해 요약본에 포함할 문장과 포함하지 않을 문장을 구분한다.[7]

또 다른 연구에서는 추출 요약을 문장 순위 문제로 정의한다. 문서로부터 m 개의 문장을 선택해 요약본을 생성하는데, 각 문장의 연관성을 수치화해 문장마다 점수를 부여하고, 점수가 가장 높은 m 개의 문장을 선택해 요약본을 생성한다.[8]

생성 요약 분야에서는 Encoder-Decoder Framework를 사용해 생성 요약본을 구성하는 연

구가 있다. Encoder는 사전학습 된 BERT-SUM, Decoder는 무작위로 초기화된 6개의 층으로 이루어진 트랜스포머로 구성되었다. 해당 연구에서는 사전 학습된 Encoder와 무작위로 초기화된 Decoder 사이의 간극으로 인해 발생할 수 있는 학습 과정의 불안정함을 다루기 위해 새로운 최적화 기법을 고안했다. 이는 Encoder와 Decoder의 Optimizer Hyperparameter 값을 다르게 설정하여, 둘 사이의 간극을 줄이고자 하는 방식이었다.[9]

생성 요약 방식으로 긴 문서의 요약본을 생성하고자 하는 다양한 연구도 존재했다. 그중에는 Deep Communicating Agents를 제시해 긴 문서의 요약본을 생성하고자 한 연구가 있었다. 해당 연구에서는 긴 문서를 여러 개로 분할 후 하나의 Decode에 연결된 여러 개의 Encoder에 나누어 입력한다. 여기서 Encoder는 좋은 요약본을 생성하기 위해 강화 학습으로 훈련된다. 문서가 부호화되면 Agent들은 그 정보들을 모으고 서로 정보를 주고받는다. 이를 통해 중요한 사실을 놓치거나, 같은 내용을 반복하거나, 불필요한 세부 사항을 포함하는 등의 실수를 줄인다. Agent들은 그들의 정보를 Decoder에 전달하게 되는데, Decoder가 Agent들로부터 정보를 균형 있게 받을수록 더 좋은 성능을 나타냈다.[10]

2.5 News Summarization

이 논문에서는 딥러닝 텍스트 요약 모델의 데이터 편향 문제를 해결하는 기법을 제시하고 있다. 하나의 Task로 학습한 모델은 비슷한 계열의 다른 Task에서 요약을 제대로 하지 못한다는 것을 문제 삼아 이를 데이터 편향 문제로 정의한다. 이 편향 문제를 해결하기 위해 고유명사 마스킹 기법과 길이를 임의로 줄이거나 늘리는 길이 변화 기법을 제시한다.[11]

또 다른 연구로는 RNN과 LSTM을 이용하여 기사를 요약하는 시스템이 있다. 이는 뉴스 기사를 입력하면 문장 단위로 분리를 한 후 그래프를 생성한 다음에 텍스트 랭크를 적용하여 요약 기사를 출력하는 형식의 시스템이다.[12]

상술한 선행 연구는 대부분 LSTM 혹은 RNN 기반의 요약 모델이 목표였으며 기사 제목을 위주로 요약하려는 시도였다. 따라서 본 연구는 기사 제목이 주어지지 않고 본문만으로 텍스트를 요약하기 위한 학습 기법을 제안하고 실제 실험을 진행하여 그 실효성을 검토한다. 첫 번째는 Transformer 기반의 모델들을 사용하는 것이다. 두 번째는 영어로 사전 학습된 모델들을 한국어로 미세조정 시킨다. 이에 한 번 더 신문 기사 데이터셋으로 모델들을 추가 학습한 후 평가하는 것이다.[13]

2.6 Text Summarization Evaluation

텍스트 요약 모델의 성능 측정에 사용될 수 있는 지표로는 BLEU(Bilingual Evaluation Understudy)가 있는데, 이 지표는 기계 번역의 성능을 측정하기 위해 고안되었으나 요약 모델 평가에도 사용이 가능하다. BLEU 스코어는 여러 개의 정답 문장과, 기계가 번역한 여러 개의 문장들이 있으면, n-gram word에 대한 정밀도(Precision)에 기반해서 성능을 측정한다.[14]

ROUGE(Recall-Oriented Understudy for Gisting Evaluation)는 요약 모델 평가를 위해 고안되었다. ROUGE는 요약 평가를 위해 인간이 만든 요약본과 컴퓨터가 생성한 요약본 사이의 재현율(Recall)에 기반하여 모델의 성능을 평가한다. ROUGE에는 4가지 종류가 있는데, 그중에 ROUGE-N의 경우, N-gram 동시 발생 통계로 점수를 계산한다. 여기서 N은 n-gram의 길이를 나타내고, 기준 요약본과 모델 생성 요약본에 동시 발생하는 n-gram의 최대 개수를 의미한다. 즉, 기준 요약본과 모델 생성 요약본에 동시에 존재하는 단어의 개수가 많을수록 모델은 높은 점수를 받는다.

ROUGE-L(Longest Common Subsequence)은 기계에 의해 생성된 요약본과 예시 요약본의 단어 배열의 유사성을 측정하는데, 여기에 LCS를 이용한다. ROUGE-L은 단순히 단어 수를 세는 것이 아닌 문자열의 짝을 맞추는 식으로 평가를 진행한다.

ROUGE-Lsum은 ROUGE-L의 변형이다.

ROUGE-Lsum은 요약본을 문장 단위로 나누고, 각 문장 사이의 유사성을 측정하는 방식으로 평가를 진행한다.[15] METEOR(Metric for Evaluation of Translation with Explicit Ordering)는 원래 기계 번역의 성능을 평가하기 위해 개발되었다. 그러나, 기계 요약의 성능을 평가하는 데에도 사용될 수 있다. METEOR는 문법과 의미를 고려하여 기계가 생성한 요약본과 예시 요약본 사이의 유사성을 평가한다. METEOR는 다양한 유형의 오류를 다룰 수 있어 평가에 유용하게 사용되는데, 그 예로는 단어 순서 오류, 동의어와 모호성 문제, 그리고 문서의 유창성도 고려할 수 있다.[16]

METEOR는 다양한 요소를 평가에 반영할 수 있기에 유용하지만, 단독으로 사용하는 것은 피해야 한다. METEOR는 본래 기계 번역 평가를 위해 고안되었으므로 기계 생성 문서의 유창성은 평가할 수 있으나, 기계 요약에서 중요하게 여겨지는 일관성을 평가하기에 한계가 있을 생성하는 등의 요약 방식을 충분히 고려하지 못할 가능성이 있기 때문이다. 따라서, METEOR는 다른 지표와 함께 사용하여 다양한 측면의 평가를 진행하는 것이 효과적이다.

2.7 Text Summarization Related Algorithms

문장 분석 대신, Attention 알고리즘을 활용한 딥러닝 모델을 활용하여 영문 리뷰 데이터와 국문 신문 기사 데이터에 대한 추상적 요약문을 생성한다. 실험 결과는 제안된 모델이 단어의 의미를 중점적으로 고려해 성공적으로 영문 리뷰 데이터를 요약하는데 성공했음을 보여준다. 국문 텍스트의 경우에도 전처리 어려움에도 불구하고 유의미한 예측 요약문을 생성하는 결과를 도출했다. 수기 확인 및 설문조사 결과는 생성된 요약 콘텐츠가 주요 단어 및 추상적 개념을 효과적으로 담아내어 문장을 효과적으로 요약한다는 것을 확인하였다.[17]

TextRank 알고리즘은 텍스트 처리를 위한 그래프 기반 랭킹 모델로, 정점 정보와 정점 간의 Global, Local 관계를 모두 고려한다. 이 알고리즘은 가중 그래프에서 동작하며 추출 방식에 중점을

두고 있다. 또한 텍스트를 정점으로 식별하고, 텍스트 간의 관계를 간선으로 나타낸다. 이후 그래프 기반 랭킹 알고리즘을 반복하여 수렴할 때까지 실행하고, 최종적으로 정점을 스코어에 따라 정렬한다. 이 알고리즘은 키워드 추출과 문장 추출에 활용되며, 특정 키워드 앞뒤의 단어 Co-Occurrence를 분석하여 그래프를 생성하고 활용한다.[18]

PLSA(Probabilistic Latent Semantic Analysis)는 문서와 단어 간의 확률적 관계를 주제를 활용하여 모델링하는 알고리즘이다. 주어진 문서에서 주제를 뽑은 뒤 해당 주제에 속하는 문서와 단어를 추출한다. 이를 통해 문서와 단어의 잠재적인 의미 구조를 파악할 수 있다. PLSA는 정보 검색과 주제 모델링에 활용된다.[19]

2.8 국방 분야 기사 요약

국방 기사 데이터를 대상으로 맞춤형 정보 분석을 수행하는 가상의 시스템을 제안하고, 정보 분석을 위한 기반 기술들을 소개한다. 데이터 마이닝과 관계 추출을 통한 맞춤형 정보를 분석하는 게 주 관점이다.[20]

국방 관련 기사를 자동으로 수집하여 LDA(Latent Dirichlet Allocation)로 주제를 분류하고 본문을 요약하며, 기본적인 감성 분석까지 가능한 국방 기사 자동 분석 시스템 구축 방안에 대해 소개한다. 텍스트 마이닝의 도메인으로 국방 관련 기사를 사용한 것이 주 관점이다.[21]

이러한 기존 연구들은 문서 요약이라는 분야에 있고, 중요한 문장을 데이터 마이닝 기법으로 추출하거나 기존 문장의 단어들을 조합하여 생성하는 방식에 집중해왔다. 하지만 본 연구는 새로운 문장 혹은 기존 단어들을 참고하여 생성함과 동시에 국방 분야에 걸맞는 간결성, 온디바이스 가능성 등에 집중하고자 한다.

3. 연구 방법

3.1 연구절차



[Figure 1] Research Development diagram

본 연구에서는 현재 공개된 모델 중에 한국어로 작성된 뉴스에 대해 충분한 성능을 낼 수 있는지 시험하여, 한국어 국방 기사 요약에 가장 적합한 모델을 여러 지표를 통해 검증하고자 한다. 관련 연구를 통해 국방 분야 성능을 검증한 사례가 없는 관계로 본 연구에서는 아래 제시된 활용할 요약 모델들을 선정하여, 오픈 소스 데이터셋과 웹 크롤링을 통해 얻은 기사 전문을 모델 학습 및 성능 확인에 사용했다. 학습된 모델들은 ROUGE, METEOR, TOXICITY 지표를 통해 성능 검증을 하였다.

3.2 데이터 수집과 전처리

본 연구에서는 AIHub의 문서 요약 텍스트 데이터셋 중 신문기사 데이터셋을 이용하여 문어체 데이터에 대한 추상적 요약을 수행하는 모델의 성능을 평가한다. 문서 요약 데이터셋 중에서도 신문 기사 Json 파일을 사용하였다. 실험을 위해 주어진 문서 요약 데이터는 최소한의 전처리 과정을 거쳤다 Documents에서 Media_name, Id, text, Abstractive, Extractive를 추출하여 Pandas Data Frame으로 변환하였다. 데이터는 195,186개의 Train 데이터, 48,797 Validation 데이터, 30,122 개의 Test 데이터로 구성되어 총 27만 건이 있고, 요약문은 54만 건으로 구성되었다.

또한, 원문으로부터 변형 없이 그대로 선택된 3개의 추출 요약 문장과 원문의 내용을 바탕으로 재작성된 생성 요약 문장도 데이터셋에 포함되어 있다.[22]

추가적으로 국방 관련 키워드 5개를 이용해 웹 크롤링한 기사를 시험 데이터셋으로 활용하여 모델

이 생성한 요약본을 직접 확인하고 평가하였다.

3.3 모델 및 하이퍼파라미터 설정

본 연구에서 사용된 모델은 BART, T5, Ko-T5, Ko BART이다. 모든 모델의 Train 하이퍼파라미터는 Eval Steps 500, Batch Size 2, Epochs 10, Learning Rate $2e-5$, Weight Decay 0.01로 동일하게 설정하였다. Inferenece 하이퍼파라미터는 length_penalty 2.0, max_length 100, min_length 10, num_beams 8로 전부 동일하게 설정했다.

3.4 환경 설정

웹 크롤링에 관한 환경은 VS(Visual Studio)에서 진행하였다. 모델에 관한 연구는 구글 클라우드 플랫폼의 Colab GPU (엔비디아 V100 (35GB in Pro))에서 수행하였다.

3.5 웹 크롤링

웹 크롤링은 네이버 뉴스에 한정하여 적용하였다. 전개도는 그림 2와 같다. 네이버 뉴스를 크롤링하여 제목, 링크, 본문을 추출하고 이를 데이터프레임으로 정리하여 CSV 파일로 저장하는 과정을 수행한다. 먼저 필요한 라이브러리들을 불러오고 웹 드라이버를 설정한 후, 페이지 URL을 생성하는 함수와 페이지 번호를 조정하는 함수를 정의한다. 그 후, 사용자에게 검색어, 크롤링 시작 페이지, 종료 페이지를 입력받아 이를 바탕으로 크롤링할 URL을 생성하는 방식이다. 다음으로 Selenium 모듈을 사용하여 네이버 뉴스 검색 페이지를 불러오고, 각 검색 결과의 URL을 수집한다. 그리고 각 뉴스의 링크를 열어서 본문 URL을 얻어오고 해당 URL이 네이버 뉴스인 경우에만 리스트에 추가한다. 수집된 뉴스의 링크를 활용하여 각 뉴스의 제목과 본문을 크롤링하고, 이를 데이터프레임으로 정리하여 최종적으로 CSV 파일로 저장한다.

title	link	content
		고팍스 코인원 이적한지 두 달 만에 퇴사... AI 소프트웨어 기업 팔린티에서 새 출발
[[단독]이중준 코인원 CSO, 가상자산 떠나	https://n.news.naver.com/mnews/article/417/00009	이중준 전 코인원 CSO, /사친=뉴스1 이중준 코인원 최고전략책임자(CSO)가 부임 3달 만에 인공지능(AI) 소프트웨어 기업 '팔린티'에 테크직했다. 지난해 6월 가상자산사업자(VASP) 변경 신고에 어려움을 겪던 고팍스의 대표로 부임했지만 그해 9월 퇴사한 그는 코인원에서도 23일 가상자산 업계에 따르면 지난해 12월 말 코인원을 퇴사한 이중준 전 CSO는 현재 팔린티에 'deployment strategist'(실명전락 담당)에는 인공지능(AI) 시스템으로 방위산업시장에서 두각을 나타내고 있는 곳으로 최근 우크라이나와 러시아 전쟁에서 데이터 분석 플랫폼 '고팍스'를 인수한 "이중준 전 CSO가 작년 말 퇴사를 그만둔 것이 맞다"고 확인했다. 1981년생인 이 전 대표는 KAIST 전산학과를 졸업하고 마이크로 엔지니어로 일했다. 이후 미국 하버드대 로스쿨에서 학위를 취득한 뒤 홍콩 골드만삭스를 거쳐 메리츠증권 피생상품 본부장을 역임한 피생본부장 시절 증권업계 최연소 임원으로 주목받았고 메리츠증권에 국내 최고 피생상품 기업으로 성장시켰다는 평가가 나온다. 그는 2소 고팍스를 세운 이준행 창업주와의 인연으로 고팍스 CSO 겸 부대표가 됐다. 고팍스는 당시 가상자산 산업에 대한 높은 이해와 전문성뿐 (IT)업계 및 금융업계의 이론과 실무를 두루 겸비한 최고의 전문가라며 영입 이유를 밝혔다. 하지만 시장 상황이 악화되면서 "글로벌 최고 금융 기업으로 발돋움하는 데 주도적인 역할을 하고 싶다"는 그의 바람도 무위에 그쳤다. 고팍스는 현재 세계 2위 가상자산 거래소 'FTX'에 흔들리면서 위기를 맞았다. 바이낸스가 구원투수로 등장했지만 금융정보분석원(FIU)이 고팍스의 VASP 변경 신고에 미온적인 태도를 보였다. 이중준 당시 부대표가 지난해 6월 신입 대표로 영전하면서 해당 직업을 이겼으나 선임 46일 만에 대표직을 내려놓았다. 그해 9월 초 CSO로 자리를 옮겼다. 시장 점유율이 한자리수에 머물러 있는 코인원은 올해 7월 가상자산법 시행에 맞춰 사업전략을 구성하기 위해 지난해 이중준 전 CSO는 "코인원과 안 좋게 끝낸 것은 아니다"라며 "가상자산 대신 다른 분야에 도전하고 싶었다"고 말했다.

[Figure 4] Web crawl CSV

3.6 카카오톡 API



[Figure 3] KakaoTalk service

본 연구에서 카카오톡 API는 연구자의 토큰으로 진행하였다. 전개도는 그림 3과 같다. 먼저, 웹 크롤링을 통해 추출한 상위 5개의 뉴스 기사에 대해 텍스트 템플릿을 생성한다. 해당 템플릿을 이용하여 카카오톡 메시지를 보내는 작업을 수행한다. 이때, 각 메시지에는 기사의 제목, 요약된 내용, 그리고 기사의 링크가 포함되어 있다. 마지막으로, 메시지 전송에 대한 응답 결과를 확인하고, 성공 여부 및 오류 메시지를 출력한다. 이를 통해 뉴스 기사를 카카오톡으로 전송하는 자동화된 프로세스를 구현하였다.

3.7 모델 성능 추가 검증

뉴스를 요약하는 경우, 특히나 사실 검증과 차별에 대한 경계가 중요하다. 따라서, 요약 모델은 문서의 중요한 내용을 잘 담는 것뿐만 아니라 차별, 허구 등의 유해한 내용을 거르는 능력 또한 필요하다. 그러나, ROUGE 지표는 단순히 Gold Summary와 모델의 요약본 사이에 중복되는 단어의 개수만을 확인하므로 요약본의 내용을 검증하기 어렵다. 이를 보완하기 위해, 모델 평가에 Toxicity와 Meteor 지표를 추가 도입했다. Toxicity를 통해 기계 생성 요약본의 유해성을 검증하고자 했다. 추가로 Meteor

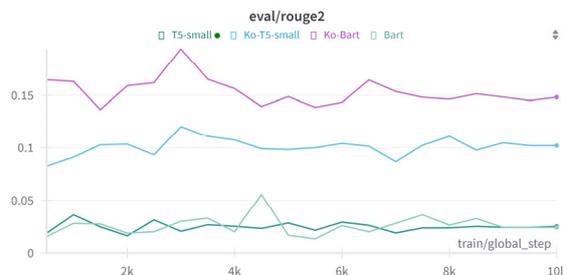
지표를 도입하여 기계 생성 요약본의 단어 순서 문제 혹은 동의어 문제뿐만 아니라 유창성도 모델 평가에 고려할 수 있도록 했다. Toxicity는 사전 학습된 Hate Speech Classification Model을 이용하여 0에서 1 사이의 숫자로 표현되었다. Meteor는 nltk 패키지의 meteor_score 라이브러리를 이용해서 측정되었고, 마찬가지로 0에서 1 사이의 숫자로 점수가 매겨졌다.

4. 연구 결과

4.1 웹 크롤링

그림 4는 웹 크롤링 후 나온 결과 파일이다. 단순히 크롤링만 진행하였기에 Content에 문장 부호가 그대로 담겨 있는 것을 확인할 수 있다. 또한, 사용자에게 원 기사를 제공하기 위해 뉴스 링크도 포함하고 있다.

4.2 요약 모델 학습 성능



[Figure 5] ROUGE-2 in Train Set

실험에 대한 평가는 ROUGE-1과 ROUGE-2를 기준으로 진행하였다. ROUGE는 요약 모델 평가를 위해 고안된 지표로, 인간이 만든 요약본과 기계가 생성한 요약본 사이의 재현율(Recall)에 기반하여 모델의 성능을 평가한다. ROUGE에는 여러 종류가 있는데, 그중 ROUGE-1과 ROUGE-2는 각각 단 일 단어(n-gram)와 두 단어 쌍(n-gram) 사이의 일치율을 측정한다. 또한, ROUGE score를 이용해 모델의 추론으로 생성된 자연어를 사람이 미리 작성 한 정답 요약문(Ground truth)과 비교, 대조하여 성능을 검증한다.[23]

그림 5는 Validation Set에서의 ROUGE-2 로 0.1 이상의 성능을, 한국어로 파인튜닝되지 않은 T5-small과 Bart는 그 이하의 성능을 보였다.

이러한 ROUGE score는 자동 기사 요약 Train Set에 대해 Ko-BART와 Ko-T5 모델이 높은 성 능을 낸다는 것을 보인다. Ko-T5는 T5보다 Rouge-1에서 3.2배, Rouge-2에서는 3.6배 더 높 은 성능을 보였고, Ko-Bart는 Bart보다 Rouge-1 에서 3.2배, Rouge-2에서는 5.2배 더 높은 성능을 냈다. 이는 한국어로 파인튜닝된 모델이 기사 내용 이해 및 주요 키워드 선택하는 방식을 더 잘 학습함 을 보인다.

또한, 학습 진행 과정에서 ROUGE score가 지속 적인 향상을 보이지 않으므로, 성능이 학습 횟수에 단순 비례하지는 않는다는 점을 알 수 있다.

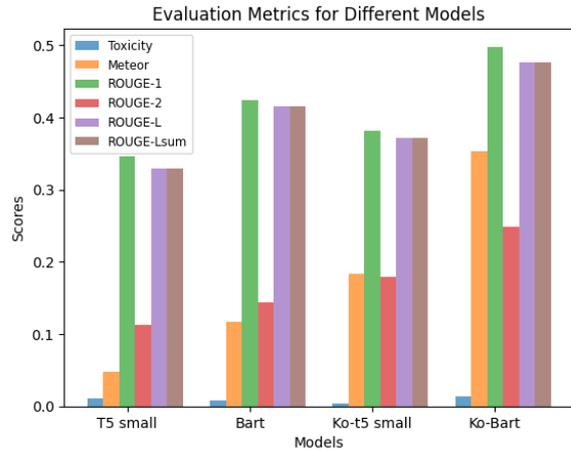
<Table 1> ROUGE-1, ROUGE-2 in Train Set

Model	Rouge1	Rouge2
Ko-T5 small	0.3127	0.097
Ko-bart	0.4174	0.1416
T5 small	0.0974	0.0265
Bart	0.1303	0.0269

4.4 모델 성능 추가 검증 결과

Toxicity는 기계 학습 모델이 생성한 텍스트에서 유해하거나 불쾌감을 주는 내용을 평가하는 지표로, 모델이 안전하고 윤리적인 출력을 생성하는지 확인

하는 데 사용된다. 이 지표는 특히 민감한 주제를 다루는 뉴스나 소셜 미디어 콘텐츠와 같은 응용 분 야에서 중요하다.



[Figure 6] Performance by Model for Test Set

METEOR는 기계 번역의 성능을 평가하기 위해 개발된 지표로, 기계가 생성한 문장과 인간이 작성 한 문장 간의 의미적 유사성을 평가한다. METEOR 는 단어 형태, 동의어, 어순 등을 고려하여 보다 세 밀하게 평가하며, BLEU보다 인간 평가와의 상관성 이 높다는 장점이 있다.

앞서 설명한 Toxicity, Meteor 지표를 추가하여 각 모델이 Test set에 대해 내놓은 요약본의 성능 을 검증하였다.

Toxicity는 4개의 모델 모두 근소한 차이를 보였 다. 유해성 부분에서는 비슷한 성능을 낸다고 생각 했으나 데이터셋에 유해한 정보를 담고 있는 뉴스가 포함되지 않았을 가능성도 존재한다.

Meteor는 모델에 따라 유의미한 성능의 차이를 보였다. 한국어로 학습이 한 번 더 되었던 모델들이 강세를 보였다. Meteor는 번역 품질을 측정하는 매 트릭이므로 해당 모델들은 문장 구조, 어휘 등의 부 문에서 강점을 보여 높은 점수를 받았다. 이를 통해 한국어로 학습이 한 번 더 진행된 모델들이 문법적 으로 더 자연스러운 요약본을 생성한다는 것을 확인 하였다.

ROUGE1/L/Lsum은 조금 다른 양상을 보였다.

요약 전 문장

'대한민국 5G 홍보대사'를 자처한 문재인 대통령은 "넓고, 체중 없는 \\'통신 고속도로\'가 5G"라며 "대한민국의 대전환이 이제 막 시작됐다"고 기대감을 높였다. '문 대통령은 8일 서울 올림픽공원에서 열린 5G플러스 전략발표에 참석해 "5G 시대는 우리가 생각하고, 만들면 그것이 세계 표준이 되는 시대"라며 "5G는 대한민국 혁신성장의 인프라"라고 강조했다. "산업과 시대에 고속도로가 우리 경제의 '대동맥' 역할을 했듯, 5G가 4차 산업혁명 시대의 고속도로가 돼 새로운 기회를 열어 줄 것이라 설명했다.", '문 대통령은 "5G가 각 분야에 융합되면, 정보통신산업을 넘어 자동차, 드론(무인항공기), 로봇, 지능형 폐쇄회로TV(CCTV)를 비롯한 제조업과 벤처에 이르기까지 우리 산업 전체의 혁신을 통한 동반성장이 가능하다"고 밝혔다. "세계 최초 상용화에 성공한 5G가 반도체를 이을 우리 경제의 새 먹거리가 될 것이라 관측이다.", '정부는 2026년 세계 5G 시장 규모가 1161조원에 달할 것으로 보고 있다.', '작년 반도체 시장 규모가 529조원인 점을 고려하면 2배 이상 큰 미래 시장이 창출되는 셈이다.', '문 대통령은 아직은 국민에게 다소 낯선 5G 시대의 미래상을 친절히 설명해 눈길을 끌기도 했다.', '문 대통령은 "지금 스마트폰으로 충분인데, 5G가 왜 필요하지?"라고 생각할 수 있다"며 "4세대 이동통신은 \\'아직은\' 빠르지만 가까운 미래에는 결코 빠르지 않다"고 했다.', '그러면서 "자동차가 많아질수록 더 넓은 길이 필요한 것처럼 사람과 사물을 연결하고, 데이터를 주고받는 이동통신망도 더 넓고 빠른 길이 필요하다"고 덧붙였다.', '문 대통령은 세계 최초 상용화에 성공한 우리 5G 기술을 널리 알리는 홍보대사를 자처하기도 했다.', '5G 시장을 선점하기 위한 각국의 경쟁이 뜨겁게 돌아오던 만큼 정부 차원에서 적극 지원하겠다는 방침이다.', '문 대통령은 "평창동계올림픽 360도 중계, 작년 4·27 남북 정상회담 때 프레스센터에서 사용된 스마트워치러기 기회가 생기면 대통령부터 나서서 우리의 앞선 기술을 홍보하겠다"고 말했다.

요약 후 문장

T5-small	8 '5G', 5G 4., 0, 0 TV(CCTV) "
Bart	문재인 대통령은 8일 서울 올림픽공원에서 열린 5G플러스 전략발표에서 "5G는
Ko-T5-small	문 대통령은 5G플러스 전략발표에 참석해 대한민국 혁신성장의 인프라라고 강조하며 4차 산업혁명 시대의 고속도로가 돼 새로운 기회를 열어 줄 것이라고 말했다.
Ko-Bart	대통령은 8일 서울 올림픽공원에서 열린 5G플러스 전략발표에 참석해 넓고, 체중 없는 '통신 고속도로'가 되어 혁신성장의 기회를 열어 줄 것이라 설명하며 우리 산업 전체의 혁신을 통한 동반성장을 이루겠다는 기대감을 드러냈다.

[Figure 7] Comparison between pre-summarized and post-summarized sentence models

앞선 지표와는 다르게 한국어로 학습된 Ko-t5보다 Bart가 더 높은 성능을 보였다. ROUGE 메트릭은 주로 추출 요약의 성능을 측정하는 데 중점을 둔다. 따라서, Bart 계통의 모델이 중요한 단어, 문장 추출에서 우수한 성과를 낸 것으로 보인다.

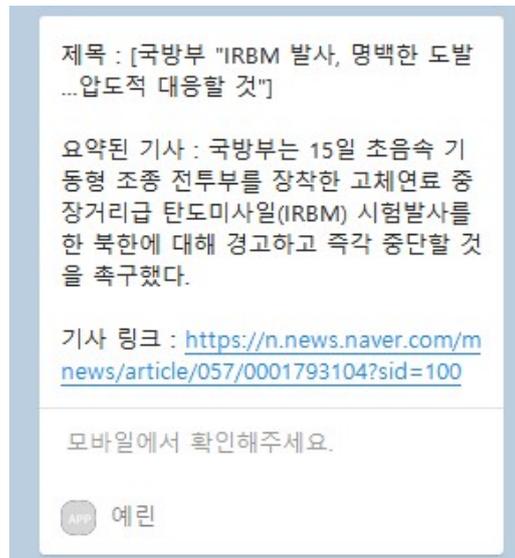
전체적으로 비교해보면, 한국어로 추가학습을 진행한 것이 성능 향상에 기여하는 바가 있는 것을 확인했다. 또한, Ko-Bart가 뉴스 요약에 가장 적합한 모델인 것으로 나타난다. Bart는 인코더와 디코더 모두 자기 회귀 구조를 가지는데, 이 부분에서 시퀀스 데이터를 처리하는 데 이점을 가진다. 이전 단계의 출력을 현재 단계의 입력으로 사용하여 문맥을 고려할 수 있고, 이를 통해 문서의 전체적인 의미를 이해하고 재구성하는 데에 강점을 보이는 것으로 추정된다.

4.5 네이버 뉴스에서 적용

그림7은 추가로 웹 크롤링을 통해 얻은 기사 중 하나의 전문을 요약 모델에 넣었을 때 각 모델이 내놓은 요약 결과이다. 한국어로 사전학습 되지 않은 T5-small과 Bart는 단어 몇 개만 생성하거나 단어를 잘못 표기하는 등 문장의 완성도가 현저히 떨어지는 모습을 보였다. 반면, Ko-T5-small과 Ko-Bart의 경우 완성된 문장을 내놓았다. 즉,

ROUGE score와 요약 문장의 완성도가 비례한 모습을 보였다. 그러나, Ko-T5-small이 생성한 요약본에는 동일한 단어가 중복적으로 생성된다는 점을 그림 7을 통해 확인할 수 있다. 이는 주어진 요약본과 같은 단어를 사용한 횟수를 이용해 성능을 평가하는 ROUGE 지표의 영향인 것으로 판단하였다.

4.5 카카오톡 서비스



[Figure 8] KakaoTalk Demo Service

그림 8은 실제 카카오톡 서비스 데모 버전이다. 데모 버전은 국방 키워드로 웹 크롤링을 한 후 모델

로 요약을 진행한 결과물이다. 요약된 기사를 몇 개 보낼지는 사용자가 선택할 수 있으며, 요약 기사의 원 기사 링크를 첨부할 수 있다.

4.6 실제 서비스와 비교

그림 9는 실제 Gpt-4o에 뉴스 기사를 넣고 요약해달라고 한 후 나온 결과와 Ko-Bart의 결과이다. Gpt-4o에게는 따로 문장을 간결하게 하라고 지시하거나 추가 조건을 주지 않는 이상 간결하게 내지 않는다는 점이 존재한다. 하지만 본 연구에서 진행한 모델은 문장 길이, 단어 반복 등 일정 제약 조건을 걸었기 때문에 보다 간결한 문장을 생성한다는 것을 확인할 수 있다.

5. 결론

본 연구에서는 4개의 모델 중 Ko-Bart가 한국어 국방 기사 요약에 가장 적합한 모델이라는 것을 확인하였다. Ko-Bart가 가장 좋은 성능을 낸 요인은 한국어로 추가학습을 진행한 점, 자기 회귀 방식을 활용해 시퀀스 처리에 유리한 점인 것으로 추측한다.

또한, Ko-Bart는 온디바이스로 학습 및 추론할 수 있는 가능성을 시사하여 국방 분야와 같은 특수

한 상황에서도 활용될 수 있을 것으로 보인다. 이는 데이터 보안과 실시간 처리의 필요성이 높은 국방 분야에서 특히 중요하다. 온디바이스 학습 가능성으로 인해 실무 적용에도 적합하다고 판단된다.

문장의 간결성 역시 문장의 길이 혹은 단어 반복 등을 조절하여 적절한 문장 생성을 함과 동시에 간결성을 띄게끔 하였다. 이로 인해 국방 분야에서의 문서 요약이라는 타이틀에 적합하다고 판단된다.

6. 향후 계획

선행 연구들을 분석한 결과, 기존에는 한국어 문서를 적용하여 텍스트 요약 성능검증한 결과를 제시한 연구가 이루어지지 않아 한국어 문서에서 가장 높은 성능을 내는 요약 모델이 무엇인지 알 수 없었다. 따라서, 후속 연구를 진행하게 된다면 T5와 Bart 외에도 SOTA¹⁾에 등극한 모델들을 한국어로 추가학습 시키고자 한다. 이를 통해 국방 분야에 특화된 보안, 간결성, 정확성 등의 부분에서 좋은 성능을 내는지도 연구할 계획이다. 또한, 온디바이스 외에도 국방 분야 도메인에 특화된 기법들을 적용하여 연구할 예정이다.

정량적인 평가 방법만을 이용해 모델을 평가 및 학습하게 되면 기계에 의해 생성되는 요약 문서가

한국 대통령 문재인은 5G 기술을 홍보하는 대사로 나서 "5G는 대한민국의 대전환이 시작됐다"고 강조했습니다. 5G는 혁신성장을 위한 기반이며, 다양한 산업 분야에서 혁신과 동반성장이 가능하다고 설명했습니다. 또한 5G가 정보통신산업뿐만 아니라 자동차, 드론, 로봇, CCTV 등 다양한 분야에 융합되면서 미래 시장의 큰 부분을 차지할 것으로 기대되고 있습니다. 5G의 성공은 우리 경제에 새로운 기회를 제공할 것으로 예측되며, 정부는 2026년에는 세계 5G 시장이 1161조원에 달할 것으로 보고 있습니다. 문재인 대통령은 미래의 5G 시대를 국민에게 설명하면서 "4세대 이동통신은 가까운 미래에는 빠르지 않다"며 5G의 필요성을 강조했습니다. 또한 대통령은 세계 최초 상용화에 성공한 한국의 5G 기술을 널리 알리기 위해 홍보대사로서 적극 활동할 것을 약속했습니다.

문 대통령은 5G플러스 전략발표에 참석해 대한민국 혁신성장의 인프라라고 강조하며 4차 산업혁명 시대의 고속도로가 돼 새로운 기회를 열어 줄 것이라고 말했다.

< GPT-4 >

< Our Model >

[Figure 9] Chat Gpt-4o와 우리 모델 요약 결과 비교

1) State-Of-The-Art의 줄임말이다.

실제 인간이 원하는 방향과는 달라질 가능성이 존재한다. 따라서, 정량적인 평가인 ROUGE, Meteor 외에도 인간이 직접 평가에 개입할 수 있는 정성적인 성능 측정 방식을 도입할 예정이다.

References

1. 국방대학교 산학협력단, “국방 인공지능 추진을 위한 전문조직 구성 및 인력관리 연구”, No. 21, No. 25, 2021.
2. 국방대학교 국가안전보장문제연구소, 『국방 인공지능 모델 기술과제 분석과 발전방안 연구』, No. 2, 2022.
3. Vaswani, A., Proceedings of the 31st Conference on Neural Information Processing Systems, “Attention is all you need”, No. 1, No. 3, 2017.
4. Mike Lewis, Computation and Language, “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension”, No. 1, No. 5, 2019.
5. Colin Raffel, “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer”, Google Research, No. 1, No. 6, 2019.
6. 남원경, 이지수, 장백철, 『한국컴퓨터정보학회』 제27권 제8호, “사전 학습된 BERT 기반 모델을 이용한 구어체 텍스트 요약”, No. 2쪽, 2022.
7. Liu, Yang, Mirella Lapata, “Text summarization with pretrained encoders.” No. 50, 2019.
8. Narayan, Shashi, Shay B. Cohen, Mirella Lapata, “Ranking sentences for extractive summarization with reinforcement learning.” No. 4, No. 6, 2018.
9. Liu, Yang, Mirella Lapata, “Text summarization with pretrained encoders.”, No. 5, 2019.
10. elikyilmaz, Deep communicating agents for abstractive summarization, No. 4 2018.
11. 조준희, 오하영, 『한국정보통신학회논문지』 Vol 26, 『딥러닝 텍스트 요약 모델의 데이터 편향 문제 해결을 위한 학습 기법』, No. 1, 2022.
12. 설경호, 『딥러닝 모델을 이용한 한국어 기사 자동요약 시스템 설계 및 구현』, 한양대학교, No. 45, 2018.
13. Kaplan, “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Processing”, Duke University, 2nd Edition Tom Henighan & Tom B. Brown, No. 1, No. 4, 2019.
14. Papineni & Kishore, “Bleu: a method for automatic evaluation of machine translation.” In: Proceedings of the 40th annual meeting of the Association for Computational Linguistics, No. 311, No. 318, 2002.
15. LIN & Chin-Yew, “Rouge: A package for automatic evaluation of summaries.” In Proceedings of the Workshop on Text Summarization Branches Ou, No. 74, No. 81, 2004.
16. Banerjee & Satanjeev, “METEOR: An automatic metric for MT evaluation with improved correlation with human judgments.” In: Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization, No. 65 No. 72, 2005.
17. 이소연, 최지은, 유선용, “Attention 알고리즘 기반 요약 콘텐츠 생성 방안 연구”, 한국디지털콘텐츠학회, No. 1, No. 2, 2021.
18. Mihalce & Rada, “Textrank: Bringing order into text.”, In: Proceedings of the 2004 conference on empirical methods in natural language processing, No. 404, No. 411,

2004.

19. Thomas Hofmann, "Probabilistic Latent Semantic Analysis", Appears in Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence, No. 15, 2013.
20. 최중환, 임채오, "국방 기사 데이터를 이용한 맞춤형 정보 분석 시스템, 국방기술품질원, No. 7, No.9, 2010.
21. 김현중, 김우주, 국방 기사 자동 분석 시스템 구축 방안 연구, 한국군사과학기술학회지 제21권 제1호, No. 10, 2018.
22. 비플라이소프트, 문서 요약 텍스트, AIHub, 2020.
23. 남원경, 이지수, 장백철, "사전 학습된 BERT 기반 모델을 이용한 구어체 텍스트 요약", 『한국컴퓨터정보학회』 제27권 제8호, No. 5, 2022.