

<https://doi.org/10.7236/JIIBC.2022.22.4.23>  
JIIBC 2022-4-4

## 딥러닝 기반의 문서요약기법을 활용한 뉴스 추천

# News Recommendation Exploiting Document Summarization based on Deep Learning

허지욱\*

Jee-Uk Heu\*

**요약** 최근 스마트폰 또는 태블릿 PC와 같은 스마트기기가 정보의 창구 역할을 하게 되면서 다수의 사용자가 웹포털을 통해 웹 뉴스를 소비하는 것이 더욱 중요해졌다. 하지만 인터넷 상에 생성되는 뉴스의 양을 사용자들이 따라가기 힘들며 중복되고 반복되는 폭발하는 뉴스 기사에 오히려 혼란을 야기 시킬 수도 있다. 본 논문에서는 뉴스 포털에서 사용자의 질의로부터 검색된 뉴스후보들 중 KoBART 기반의 문서요약 기술을 활용한 뉴스 추천 시스템을 제안한다. 실험을 통해서 새롭게 수집된 뉴스 데이터를 기반으로 학습한 KoBART의 성능이 사전훈련보다 더욱 우수한 결과를 보여주었으며 KoBART로부터 생성된 요약문을 활용하여 사용자에게 효과적으로 뉴스를 추천하였다.

**Abstract** Recently smart device(such as smart phone and tablet PC) become a role as an information gateway, using of the web news by multiple users from the web portal has been more important things. However, the quantity of creating web news on the web makes hard to catch the information which the user wants and confuse the users cause of the similar and repeated contents. In this paper, we propose the news recommend system using the document summarization based on KoBART which gives the selected news to users from the candidate news on the news portal. As a result, our proposed system shows higher performance and recommending the news efficiently by pre-training and fine-tuning the KoBART using collected news data.

**Key Words** : BART, BERT, Document-Summarization, Recommendation, RNN, Seq2Seq

### 1. 서 론

인터넷 및 스마트 기기의 기술 발달과 사회 관계망 서비스(SNS)의 급격한 이용으로 인하여 대량의 정보들이 온라인상에서 생성, 소비되고 있다. 특히 다음(Daum)과 네이버(Naver)를 대표로하는 인터넷 포털 사이트의 중요도가 증가하면서 이러한 포털 사이트에서 뉴스를 보는

비율이 갈수록 증가되고 있다<sup>[1,2]</sup>. 한국언론진흥재단의 발표에 따르면 2021년 포털을 통한 사용자의 뉴스 이용률은 전년도인 2020년 보다 3.4% 증가한 79.2%를 차지하였으며 특히 스마트 폰이나 태블릿 PC등의 스마트 기기를 통한 이용은 20~40대가 90% 이상이라고 밝혔다<sup>[3]</sup>. 이러한 대량의 뉴스 이용으로 인하여 사용자들은 오히려 무방비하게 노출이 되면서 가짜뉴스와 같이 사용자가 원

\*정회원, 강남대학교 KNU참인재대학  
접수일자 2022년 6월 21일, 수정완료 2022년 7월 21일  
게재확정일자 2022년 8월 5일

Received: 21 June, 2022 / Revised: 21 July, 2022 /  
Accepted: 5 August, 2022

\*Corresponding Author: jeeukheu@kangnam.ac.kr  
College of KNU Charm-Injae, Kangnam University, Korea

하지 않는 정보를 접하게 되는 문제점이 발생되고 있다. 사용자는 자신이 원하는 뉴스를 접하기 위해서는 직접 뉴스 기사를 보면서 선택해야 하지만 실제로 짧은 시간 안에 기하급수적으로 생성되는 뉴스를 모두 확인하기는 불가능하다. 특히 올림픽 스포츠 같은 특정 이벤트가 발생되었을 때 각 신문사는 실시간으로 빠르게 뉴스를 생성하기 때문에 양질의 내용을 담지 못하고 내용이 비슷한 중복되는 뉴스를 사용자가 접하게 된다는 문제점이 발생된다<sup>4)</sup>.

정보처리 분야에서는 사용자들에게 편의를 제공해주기 위해 문서요약기법(Document Summarization), 추천시스템(Recommend System), 질의응답(QA: Question Answering), 기계번역(Machine Translation), 개체명 인식(Named Entity Recognition) 등 다양한 기법들이 연구되어 왔다. 최근에는 기계학습 및 딥러닝 기술을 기반으로 한 자연어 처리 연구가 활발하게 진행되면서 성능도 빠르게 향상되고 있다. 문서요약기법은 대량의 문서의 문장과 단어를 분석하고 단문의 요약문을 생성하여 사용자에게 유용한 정보를 제공해주기 위한 기술이며 추천시스템은 대량의 정보중 사용자의 취향에 맞는 것만을 제공해주는 일종의 정보 필터링 기법이다.

현재 인터넷 포털 사이트에서는 사용자가 관심만한 뉴스를 요약하는 서비스가 제공되고 있다. 하지만 동일한 내용과 주제를 다루는 대량의 뉴스에서 추출된 요약문은 주어진 주제에 대한 핵심적인 내용만을 포함하고 있기 때문에 전체적인 문맥을 파악할 수 없다는 문제점이 생긴다. 결국 뉴스의 요약문을 제공받더라도 전체 맥락을 이해하기 위해서는 결국 뉴스 본문을 봐야 한다는 번거로움이 생기며 이 작업을 위해 내용이 중복되는 대량의 갖는 뉴스 기사들 중 하나를 선택해야한다.

본 논문에서는 대량의 뉴스 기사를 KoBART를 통해 생성된 요약문을 기반으로 동일한 주제의 문서들 중에서 사용자에게 제공할 대표 뉴스를 추천 해주는 뉴스 추천 시스템을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 딥러닝을 이용한 자연어 처리 연구들을 소개하고, 3장에서는 KoBART 기반의 문서요약기법을 활용한 뉴스 추천 기법을 제안한다. 4장에서는 실험 및 분석을 통하여 제안하는 우수성을 기술하고 마지막 5장에서는 향후 연구계획 결론을 맺는다.

## II. 관련 연구

### 1. 문서요약기법(Document Summarization)

문서요약기법은 크게 추출(Extractive)요약기법과 생성(Abstractive)요약기법으로 분류 된다. 추출요약기법은 주어진 문서를 분석하여 각 문장에 중요도를 점수로 계산하여 높은 점수를 획득한 문장 위주로 문서를 요약하는 반면 생성요약기법은 요약문 생성 시 자연어 처리(Natural Language Processing)를 적용하여 새로운 문장을 만든다. 생성요약기법은 생성되는 요약문의 문법 오류 그리고 전체적인 문맥에 따른 단어와 문장 선택의 어려움 등의 이유로 상대적으로 추출요약기법이 많이 연구되어 왔다. 하지만 딥러닝 기반의 순환 신경망(RNN: Recurrent Neural Network)연구가 활발하게 진행이 됨에 따라 양질의 문장을 생성하는 연구가 진행되면서 최근에는 생성요약기법 위주로 연구가 진행되고 있다.

순환신경망은 기본적으로 입력 층에 대한 출력 층을 다시 자신이 입력 층으로 넣는 순환 구조를 가진 신경망으로 주로 기계번역, 문서요약 등에 많이 활용된다. 초기의 순환신경망은 짧은 문장에 대해서는 효과적인 성능을 보였으나 단어의 수가 긴 문장이 입력 값으로 들어올 경우 성능이 저하되는 단점이 있었다. 이를 극복하기 위해 LSTM(Long-Short Term Memory)<sup>5)</sup>과 GRU(Gated Recurrent Units)<sup>6)</sup>가 연구되었으며 이후 LSTM 또는 GRU로 구성된 Seq2Seq (Sequence-to-Sequence)<sup>7)</sup> 모델이 발표되었다.

Seq2Seq 모델은 입력 값을 받는 인코더(Encoder)와 출력 값을 주는 디코더(Decoder)로 구성되어 있어 입력 값에 대한 출력 값을 생성하는 모델로 기계번역과 문서 요약에 효과적인 성능을 보였다. 하지만 Seq2Seq도 분석시 사전에 없는 단어(OOV: Out of Vocabulary)가 있을 경우와 입력 값이 길어지면 장기 의존성(Long-term dependency) 문제가 발생되어 성능이 저하된다는 단점이 있다.

### 2. 딥러닝 기반 자연어 처리(Natural Language Processing based on Deep Learning)

자연어 처리 분야는 인공지능의 한 분야로 사람의 언어를 기계가 이해하고 사용할 수 있도록 처리하는 분야라고 할 수 있으며 기계번역, 챗봇(Chatbot), 문법 수정, 문장생성 등 다양하게 활용된다. 과거에는 주어진 텍스트의 형태소 분석, 품사, 구문구조 분석 등 문법의 규칙과 패턴을 기반으로 연구가 진행되었으며 추후 통계기반으로 한 확률적인 접근으로 연구가 진행되었다.

최근에는 신경망(Neural Network)기술을 기반으로 주어진 문서의 단어나 문장을 임베딩(Embedding)하여 벡터로 표현하여 자연어를 분석하는 연구가 활발하게 진행되고 있다. 단어 수준 임베딩은 Word2Vec<sup>[8]</sup>, GloVe<sup>[9]</sup>, FastText<sup>[10]</sup>와 같은 기법들이 있으며 문장 수준의 임베딩은 ELMo<sup>[11]</sup>, BERT<sup>[12]</sup>, GPT<sup>[13]</sup> 기법들이 있다. 특히 BERT는 2018년 구글에서 공개한 모델로 자연어 처리 분야에서 매우 우수한 성능을 보이고 있으며 현재 기계번역, 문서요약, 질의응답 등 다양한 자연어 처리 분야에 활용되고 있다.

BERT는 Seq2Seq의 단점인 장기 의존성을 보완하기 위하여 연구된 어텐션(Attention)기법을 기반으로 한 트랜스포머(Transformer)를 적용한 모델이다. 어텐션 기법은 인코더와 디코더의 구조에서 어텐션 모듈을 넣음으로써 임베딩된 각 벡터의 집중도를 높이는 방법으로 정보 손실을 줄여 기계번역의 성능을 높였다. BERT는 사전 학습(Pre-Training)과 파인 튜닝(Fine\_Tuning)을 통하여 다양한 자연어 처리 분야에 활용되며 주어진 시퀀스에 대한 분석의 방향이 양방향(Bidirectional)으로 작동하기 때문에 성능이 더욱 향상된 결과를 보여주고 있다. 이후 Seq2Seq 모델을 기반으로 하여 BERT모델과 GPT모델을 하나의 형태로 접목시킨 BART를 페이스북에서 공개했다. BART<sup>[14]</sup>모델은 BERT보다 더 좋은 성능을 보이고 있으며 특히 문서요약에 우수한 성능을 보여주고 있다.

### 3. 뉴스 추천(News Recommendation)

뉴스 추천 시스템은 특정 주제에 대한 수많은 뉴스를 분석하여 사용자가 관심을 가질만한 양질의 뉴스를 선택해서 추천해주는 기술이다. 스마트 기기의 보급과 인터넷의 발달로 인하여 뉴스 포털의 이용이 증가함에 그 중요성은 더욱 부각되고 있다. 뉴스 추천을 위해서는 사용자의 패턴, 성향, 대중의 관심 등 다양하게 고려해야 할 요소들이 있으며 최근에는 딥러닝을 기반으로 한 뉴스 추천시스템이 많이 연구되고 있다.

Okura<sup>[15]</sup>의 연구에는 뉴스를 임베딩하여 벡터로 표현하고 사용자의 정보를 GRU 모델에 적용하여 후보 뉴스와의 벡터 계산을 통한 뉴스 추천 시스템을 제안하였으며, An et al<sup>[16]</sup>의 연구에서는 신경망을 기반으로 사용자의 뉴스클릭을 통한 장/단기 흥미도를 분석하여 GRU 모델에 적용한 추천시스템을 제안하였다.

## III. 제안하는 시스템

### 1. 시스템의 구성과 흐름

그림 1은 본 논문에서 제안하고 있는 딥러닝 기반의 문서요약기법을 활용한 뉴스 추천 시스템을 보여주고 있다. 우선 사용자가 뉴스들을 포털을 통하여 찾고자 하는 뉴스에 대한 질의어를 입력하여 관련된 후보 뉴스들을 검색하고, 검색된 후보 뉴스들은 전처리 과정을 통해서 사전 학습된 KoBART Summarization 모듈의 입력 값으로 들어간다. 전처리 과정에서는 학습에 적용하기 부적절한 불용어(Stopword)와 특수 문자들을 제거되며 정보가 부족한 단문으로 이루어진 뉴스들을 필터링 한다. 그 후 생성된 요약문을 기반으로 후보 뉴스들중 요약문의 내용을 가장 잘 반영하고 있는 뉴스를 선택하여 사용자에게 추천해준다.

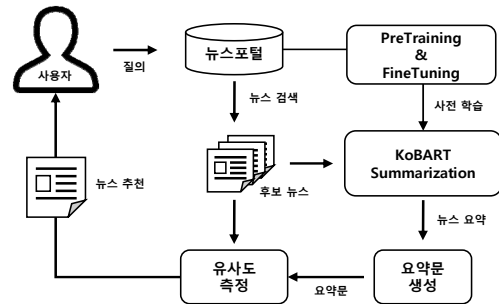


그림 1. 제안하는 뉴스 추천 시스템 구성도  
 Fig. 1. Proposed News Recommend System Architecture

### 2. 뉴스 요약 단계

본 논문에서는 입력된 뉴스의 요약을 위하여 STK-AI에서 공개한 한국어 KoBART 모델(<https://github.com/seujung/KoBART-summarization>)을 이용하여 요약문을 생성하였다. KoBART는 한국어 텍스트를 학습한 한국어 인코더-디코더 언어모델로 학습에 사용된 데이터 셋은 Dacon 한국어 문서 생성요약 AI 경진대회의 학습데이터를 이용하였다.

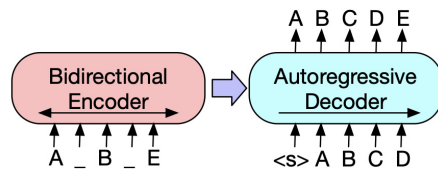


그림 2. BART의 기본 구조  
 Fig. 2. A Schematic of BART

BART의 손상된 문서를 입력 값으로 받아 기존 문서로 복구 시키는 되돌리는 것으로 학습하는 Denoising auto-encoder 기반의 사전학습 방법 모델로 그림2와 같이 Seq2Seq 구조로 구성된다. 그림2에서 보이는 바와 같이 BART는 양방향 인코더(Bidirectional Encoder) 구조를 갖는 BERT의 형태이며 마스크 토큰(Mask Token)이 포함된 손상된 텍스트를 입력 값으로 받는다. 반면 BART의 디코더는 GPT의 Autoregressive Decoder 구조를 가지고 있으며 인코더를 거쳐 출력된 값을 받아 마스크 토큰 추측하여 복원하여 학습하는 구조를 갖는다.

BART는 사용자가 질의를 입력한 뉴스포털로부터 검색된 후보뉴스들을 입력 값으로 받아 요약문을 생성한다. 문서요약은 파인 튜닝과정에서 원본 뉴스의 중요 문장 또는 키워드 등과 같은 정보를 찾아내고 재구성하여 생성요약기법으로 요약문을 생성한다.

### 3. 뉴스 추천 단계

뉴스 요약 단계를 통해서 생성된 후보뉴스들의 요약문은 사용자에게 뉴스를 추천하기 위하여 후보뉴스들과의 유사성을 비교한다. KoBART를 통해서 임베딩된 요약문 기본적으로 후보 뉴스들의 본문에 대한 중요한 내용일 압축되어 있으므로 후보 뉴스 중 요약문의 내용이 충분히 반영된 뉴스를 유사도 측정을 통해서 선택한다. 문서와 문장 간의 유사도를 측정하기 위해서는 자카드 유사도 (Jaccard Similarity), 유클리디언 유사도 (Euclidean Similarity) 그리고 코사인유사도 (Cosine Similarity) 등 다양한 측정법이 있다. 본 논문에서는 후보뉴스와의 유사성을 분석하기 위해서 코사인 유사도를 적용하여 계산한다. 벡터로 표현된 문서 A와 문서 B의 유사도를 구하는 방법은 (1)과 같다.

$$\text{Cosine Similarity}(A, B) = \text{Cos}(A, B) \frac{AXB}{|A||B|} \quad (1)$$

$$\begin{aligned} \text{News}_{\text{recommend}} \\ = \text{argmax}_{d \in D} \text{Cos}(d_i, \text{KoBART}_{\text{sum}}) \end{aligned} \quad (2)$$

(2)는 후보 뉴스  $D = \{d_1, d_2, \dots, d_n\}$  중 KoBART로 생성된 요약문  $\text{KoBART}_{\text{sum}}$  과 유사도가 높은 뉴스 문서를 사용자에게 추천하기 위한 뉴스  $\text{News}_{\text{recommend}}$  의 계산식을 보여주고 있다. (2)에 의해 계산된  $\text{News}_{\text{recommend}}$  는 뉴스 포털을 통해 사용자가 입력한 질의어에 대한 적절한 뉴스로써 추천된다.

## IV. 실험 및 결과

### 1. 실험 데이터

본 논문에서는 실험 데이터로 네이버 포털 뉴스 (<https://news.naver.com/>)를 크롤링하여 수집하였다. 네이버 포털 뉴스는 “사회”, “정치”, “경제”, “IT과학”, “오피니언”, “생활문화”, “세계”의 7 가지 항목으로 나뉘어져 있으며 본 논문에서는 “사회”와 “정치” 항목의 뉴스들을 수집하여 실험 데이터로 사용했다. 수집은 2022년 1월부터 2022년 3월 까지 3개월 간 약 90일 동안 생성된 네이버 인터넷 뉴스 기사 33만 건을 일별로 수집하였다. 수집된 뉴스 기사들 중 300자 이하의 단문으로 문장으로 이루어진 2~3문장의 짧은 단문으로 이루어지거나 이미 지 위주로 작성된 제목만 있는 약 4만 건의 뉴스 기사들은 실험에서 제외하였다. 수집된 네이버 뉴스의 1, 2월 데이터를 KoBART의 학습데이터로 사용하였으며, 3월 데이터를 테스트 데이터로 사용하였다.

표 1. 구글 트렌드 일별 인기 급상승 검색어  
Table 1. Google daily search trend

요일	검색어
3월 9일(수)	윤석열, 출구조사, 투표율, 대선, 이재명, 대통령, 대선결과, 김건희, 윤석열 공약
3월10일(목)	보배드림, 의료민영화, 허경영, 영탁, 여가부 폐지, 한동훈, 박영미, 윤석열 당선
3월11일(금)	이근, 김은혜, 멜론, 이찬원, 기안84
3월12(토)	임영웅, 이광수, 임희숙, 이상민 의원, 박지현.

사용자가 검색하고자 하는 질의어 선정을 위해서 구글 트렌드(<https://trends.google.com>)를 통해 인기 급상승 검색어를 선정하여 실험에 사용했다. 수집된 질의어는 “사회”, “정치”, “연예”, “스포츠” 등의 항목들이 주로 이루어져있다. 표 1은 구글 트렌드 일별 인기 급상승 검색어의 예시를 보여주고 있다. 검색에 사용된 검색어는 주중에 “사회”, “정치”의 질의어가 주로 사용되었으며 주말에는 “스포츠”, “연예”의 질의어사 사용되어 요일에 따라 사용자들의 관심이 집중되는 질의어의 성격을 확인할 수 있었다.

### 2. 평가 방법

KoBART를 통해 입력된 후보 뉴스들에 대한 뉴스요약 평가 방법으로는 ROUGE(Recall-Oriented Understudy for Gisting Evaluation)<sup>[17]</sup>를 사용하였다. ROUGE는

문서요약을 평가하기 위한 지표이며, 요약 모델이 생성한 요약본을 정답으로 주어진 참조본과 대조해서 성능 점수를 계산하는 방법으로 사용된다. ROUGE-N의 평가 방법은 아래의 수식(3)과 같다.

$$ROUGE-N = \frac{\sum_{S \in S_H} \sum_{g_n \in S} C_m(g_n)}{\sum_{S \in S_H} \sum_{g_n \in S} \alpha(g_n)} \quad (3)$$

$S_H$ 는 정답 요약 데이터  $S$ 의 집합을 의미하며  $C_m(g_n)$ 은 생성된 뉴스의 요약문이 정답 요약문  $g_n$  과의 N-gram 만큼의 얼마나 일치하는지를 나타낸다. 본 논문에서는 ROUGE 측정 시 사용자의 질의에 의해 검색된 후보 뉴스기사들 중 사용자에게 추천하기 위해 선택된 뉴스의 제목을 정답 요약문으로 설정하였다. KoBART의 성능의 비교를 위하여 사전학습된 Baseline 모델과 본 논문에서 실험을 위해 수집된 2022년 1, 2월 네이버 뉴스를 학습 데이터를 파인 튜닝(Fine Turning)하여 학습한 결과를 비교하였다.

### 3. 실험 결과

표 2.는 본 논문에서 수집한 네이버 뉴스를 학습한 KoBART의 생성된 요약문과 비교를 위하여 Seq2Seq, Transformer, 그리고 사전 학습된 KoBART 기반의 생성 요약 모델로 생성한 요약문의 비교 결과이다. 표 2에서 확인 가능하듯이 KoBART의 요약문이 Seq2Seq와 Transformer 기반의 기존 생성 요약 모델보다 우수한 성능이 나왔다. 또한 사전 학습된 KoBART의 성능보다 새롭게 뉴스 데이터를 수집하여 학습하고, 파인 튜닝을 했을 때의 결과가 더 우수하게 나온 것을 확인할 수가 있다. 이것은 기존 KoBART의 사전학습 시 사용된 학습 데이터에서 부족한 요약에 필요한 정보들이 새롭게 수집된 뉴스 데이터 정치와 사회분야에 관련된 용어나 데이터를 추가적으로 학습되고 미세조정 되어 더욱 풍부한 정보를 기반으로 학습한 결과라고 할 수 있다.

표 2. 뉴스 요약 결과

Table 2. The results of the news summarization

Model	ROUGE-1	ROUGE-2	ROUGE-L
Seq2Seq	0.275	0.122	0.252
Transformer	0.301	0.136	0.315
KoBART_Base	0.354	0.177	0.335
KoBART_NewsData	<b>0.377</b>	<b>0.199</b>	<b>0.338</b>

표 3. 뉴스 추천 예시

Table 3. The example of the news recommendation

질의어	여가부 폐지
후보 뉴스 (제목)	<ul style="list-style-type: none"> <li>● 하태경"여가부 폐지 조정 안 한다...민주당과 협치 원칙 지킬 것"</li> <li>● 하태경"이준석 공 압도적...여가부 폐지 공약 조정 안해"</li> <li>● 하태경"민주당과 협치...여가부 폐지 조정 안 한다"</li> <li>● 하태경"여가부 폐지 조정 안 할 것..2030 지지 얻는데 이준석 공약이"</li> </ul>
생성된 요약문	국민의힘 선거대책본부에서 게임특별위원장으로 활동한 하태경 의원은 윤석열 당선인이 내건 여가가족부 폐지 공약 조정은 안 할 것이라고 밝혔다.
추천 뉴스 (제목)	하태경 "민주당과 협치..여가부 폐지 조정 안 한다"

표3은 본 논문에서 제안하는 뉴스 추천 시스템의 결과의 예시를 보여주고 있다. 예시에 사용된 검색어는 '여가부 폐지'이며 검색어를 통해 뉴스 포털에서 검색된 후보 뉴스들의 목록을 보여주고 있다. KoBART는 해당 후보 뉴스의 기사를 입력 값으로 받아 학습된 모델에 의해 요약문을 생성한다. 추천 시스템은 생성된 요약문을 기반으로 해당 요약문의 내용이 가장 잘 반영된 뉴스를 사용자에게 추천해준다.

## V. 결 론

최근 스마트폰의 급격한 보급과 이용으로 다수의 사용자의 뉴스포털 이용에 대한 중요도가 커지면서 뉴스의 소비 중요성도 커지게 되었다. 하지만 인터넷 상에 생성되는 뉴스의 양을 사용자들이 따라가기 힘들며 폭발하는 뉴스 기사에 오히려 혼란을 야기 시킬 수도 있다.

본 논문에서는 뉴스 포털에서 사용자의 질의로부터 검색된 뉴스후보들 중 KoBART 기반의 문서요약 기술을 활용한 뉴스 추천 시스템을 제안하였다. 실험을 통해서 KoBART의 성능이 사전훈련보다 더욱 우수한 결과를 보여주었으며 KoBART로부터 생성된 요약문을 환용하여 사용자에게 효과적으로 뉴스를 추천하였다.

향후 연구로 추천 시스템 구축에서 있어서 중요한 요소인 인터넷 뉴스 언론사를 비롯하여 사용자의 정보, 성향, 검색과 소비된 뉴스의 로그 기록들을 수집하여 데이터를 확보하고 더욱 다양한 연구를 진행할 예정이다.

## References

- [1] PARK, Raegun, et al, "Application of Advertisement Filtering Model and Method for its Performance Improvement", Journal of the Korea Academia-Industrial cooperation Society, Vol 20, No. 11, pp. 1-8. 2020.  
DOI: <https://doi.org/10.5762/KAIS.2020.21.11.1>
- [2] Jeong, Eun-Hee, "Regional Image Change Analysis using Text Mining and Network Analysis", The Journal of Korea Institute of Information, Electronics, and Communication Technology, Vol 15, No 2, pp. 79-88. 2022.
- [3] Korea Press Foundation, "Audience Survey, 2021", 2021.
- [4] Pritchard, Duncan, "Good news, bad news, fake news", The Epistemology of Fake News, pp. 46-67, 2021.  
DOI: <https://doi.org/10.1093/oso/9780198863977.003.0003>
- [5] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory", Neural computation, Vol. 9, No. 8, pp. 1735-1780, 1997.
- [6] Chung, Junyoung, et al, "Empirical evaluation of gated recurrent neural networks on sequence modeling", arXiv preprint arXiv:1412.3555. 2014.
- [7] Sutskever, I., Vinyals, O., and Le, Q. "Sequence to sequence learning with neural networks", Advances in neural information processing systems, 2014.
- [8] Mikolov, T., Corrado, G., Chen, K. & Dean, J, "Efficient estimation of word representations in vector space", Proc. of the International Conference on Learning Representations, ICLR 2013, 1-12.
- [9] Pennington, J., Socher, R., & Manning, C., "Glove: Global vectors for word representation", In Proceedings of the 2014 conference on empirical methods in natural language processing, pp. 1532-1543, Oct, 2014.
- [10] Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T., "Enriching word vectors with subword information". Transactions of the Association for Computational Linguistics, pp. 135-146, 2017.
- [11] Matthew E. Peters, et al, "Deep Contextualized Word Representations", Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol, 1, pp. 2227-2237, 2018.  
DOI: <http://dx.doi.org/10.18653/v1/N18-1202>
- [12] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K., "Bert: Pre-training of deep bidirectional transformers for language understanding", arXiv preprint arXiv:1810.04805. 2018.
- [13] Radford, Alec, et al. "Improving language understanding by generative pre-training", 2018.
- [14] Lewis, Mike, et al, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension", arXiv preprint arXiv:1910.13461, 2019.
- [15] Okura, Shumpel, et al. "Embedding-based news recommendation for millions of users". In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1933-1942., 2017.
- [16] A. Mingxiao, et al. "Neural news recommendation with long-and short-term user representations", In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 336-345. 2019.
- [17] Lin, C. Y., "Rouge: A package for automatic evaluation of summaries". In Text summarization branches out, pp. 74-81, 2004.

## 저 자 소 개

### 허 지 욱(정회원)



- 2007년 : 한림대학교 컴퓨터공학과 학사 졸업,
  - 2009년 : 한양대학교 컴퓨터공학과 석사 졸업,
  - 2016년 : 한양대학교 컴퓨터공학과 박사 졸업,
  - 현재 : 강남대학교 KNU참인재대학 교양교수부 조교수
- 관심분야 : 멀티미디어 점보검색, SNS 분석, 빅데이터 분석, 다중문서요약, 자연어처리, 기계학습, 딥러닝 등

※ 본 연구는 2021학년도 강남대학교 교내연구비 지원에 의해 수행되었음.