

# 한국어 문서요약 및 음성합성 통합 프레임워크 구축

김탁영 · 김지나 · 강형원 · 김수빈 · 강필성<sup>†</sup>

고려대학교 산업경영공학부

## Building an Integrated Framework of Korean Text Summarization and Text-to-Speech

Takyoun Kim · Jina Kim · Hyeongwon Kang · Subin Kim · Pilsung Kang

Department of Industrial & Management Engineering, Korea University

We propose an integrated text summarization and text-to-speech framework which summarizes Korean documents into a few sentences and reads them in a specific person's voice. In our framework, a pre-trained text summarization model (KoBART) is fine-tuned with an additional news-oriented text summarization dataset. Then, the fine-tuned model is compressed by knowledge distillation (DistilKoBART) to improve computational efficiency. For text-to-speech, Tacotron 2 and Waveglow models are used. To generate a natural speech sample, we design a task-specific transliteration module that converts numeric or English expressions into Korean. The experimental results show that the proposed framework effectively summarizes long documents and provides a human-like synthesized voice. The proposed framework can provide convenience such as fast information delivery to busy modern people or effectively deliver information to users in special situations such as drivers and people with low vision.

**Keywords:** Text Summarization, Text-to-Speech, KoBART, Distillation, Tacotron 2, Waveglow

### 1. 서론

문서요약(text summarization)은 긴 텍스트에서 핵심적인 내용을 추출 또는 생성하여 출력하는 것을 목적으로 하는 자연어 처리 분야의 과업이다. Radev *et al.*(2002)이 넓은 의미에서 정의하는 요약문은 원문 길이의 절반을 넘지 않는 텍스트이다. 따라서 자동문서 요약이 가능할 경우, 일상 혹은 인터넷 상에 등장하는 뉴스기사, 논문, 법률문서 등 많은 양의 텍스트 중 필요한 부분을 빠르게 파악해야 하는 상황에서 유용하게 사용될 수 있다.

문서 요약은 크게 원본 문서에서 중요한 몇 개의 문장을 선택하는 추출요약(extractive summarization)과 원문의 핵심 내용을

보존하면서 짧게 재구성하는 생성요약(abstractive summarization)으로 나눌 수 있다. 원문에서 문장 단위의 요약 과업을 수행할 경우, 추출요약은 원문에서 주요 문장을 선택하는 분류(classification) 혹은 순서화(ranking) 문제로 정의되며, 생성요약은 자연어 문장을 생성하는 언어 생성(natural language generation; NLG) 문제로 정의된다. 추출요약은 TextRank(Mihalcea and Tarau, 2004), LexRank(Erkan and Radev, 2004)와 같은 그래프 기반 방법론으로부터 연속적인 벡터 공간 내 인공신경망을 사용한 추출요약 방법론(Kågeback *et al.*, 2014)까지 다양한 연구가 진행되어 왔으나, 생성요약은 과업의 높은 난이도로 인해 비교적 최근에 인공신경망 기반의 NLG 모델의 발전과 함께 여러 모델들이 개발되었다(Rush *et al.*, 2015).

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.2021-0-00034, 과편화된 데이터의 적극 활용을 위한 시계열 기반 통합 플랫폼 기술 개발). 또한 이 연구는 한국산업기술진흥원의 산업인공지능인력양성사업의 지원을 받아 수행되었음(P0008691).

<sup>†</sup> 연락저자 : 강필성 교수, 02841, 서울시 성북구 안암로 145 고려대학교 산업경영공학부, Tel: 02-3290-3383, Fax: 02-929-5888,

E-mail: pilsung\_kang@korea.ac.kr

2021년 9월 14일 접수; 2021년 10월 18일 수정본 접수; 2021년 11월 30일 게재 확정.

대용량의 텍스트 데이터로 사전학습(pretraining)을 수행하고 특정 과업에 맞추어 미세조정(finetuning)을 수행하는 BERT(Devlin *et al.*, 2018) 모델의 등장 이후, 자연어처리의 많은 과업은 이러한 사전학습-미세조정 기반의 방법론을 기반으로 성능을 향상시켜 왔다(Liu and Lapata, 2019). 기계학습을 이용한 문서요약 과업 역시 지금까지도 사전학습-미세조정 기반의 모델이 가장 높은 성능을 보인다. 현재 공개되어 있는 방법론 중 가장 성능이 우수한 것으로 알려져 있는 방법론은 인코더-디코더(encoder-decoder) 구조의 사전학습 기반 자연어처리 모델인 BART(Lewis *et al.*, 2019)를 기반으로 하고 있다. 문서요약 과업에서 BART를 구축하기 위해서는 위키피디아(Wikipedia) 등의 대용량 문서로부터 얻은 텍스트로 먼저 사전학습을 진행한다. 그 후 원문과 요약문의 쌍으로 구성된 문서요약 데이터를 활용하여 미세조정을 수행한다. 사전학습을 수행하는 데이터는 다양한 경로를 통해 대량의 데이터를 수집하는 것이 가능한 반면, 미세조정을 위한 데이터의 경우 원문과 요약문에 대한 쌍이 반드시 필요하다는 점 때문에 제한된 데이터만이 가용하게 된다. 영어의 경우 CNN/Daily Mail(Nallapati *et al.*, 2016)과 같은 대량의 영문 요약 과업 전용 데이터셋이 존재하는데 반해, 2~3년 전까지만 하더라도 한글 요약 과업 전용 데이터셋의 규모는 매우 작았으며, 이는 동일한 모델 구조를 사용함에도 불구하고 영문 요약 모델에 비해 한글 요약 모델의 성능이 낮은 주요 원인으로 지목되어 왔다. 최근에는 정부의 디지털 뉴딜 정책의 일환으로 대량의 한국어 문서 요약 데이터셋이 구축되었으며,<sup>1)</sup> 민간 기업에서는 대용량의 한국어 텍스트로 사전학습을 진행한 KoBART를 공개하기도 하였다.<sup>2)</sup>

적절히 축약된 문서 요약문은 매우 다양한 분야에 활용할 수 있음에도 불구하고, 요약문의 품질을 향상시키는 요약 알고리즘 관점의 연구에 비해 요약문을 응용하는 관점의 연구는 상대적으로 미진한 실정이다. 특히 문서 요약이 신속하게 핵심 정보를 파악해야 하는 상황에서 유용한 점을 고려했을 때 바쁜 일정을 보내는 현대인들에게 편의를 제공할 수 있을 뿐만 아니라, 요약문 자체를 음성으로 변환하여 읽어줄 수 있다면 운전자나 저시력자와 같은 특수한 상황에 처한 사용자들에게 큰 효용을 제공해줄 수도 있다. 따라서, 본 연구에서는 문서 요약과 음성 합성을 결합하여 요약문의 품질을 향상시키고 동시에 이를 음성으로 변환하여 출력해줄 수 있는 통합 프레임워크를 제안한다.

본 연구에서 제안하는 문서 요약 및 음성 합성 프레임워크의 특징은 크게 세 가지이다. 첫째, 문서 요약의 품질을 향상시키면서도 동시에 빠른 추론 속도를 갖도록 모델 경량화를 추구하였다. 둘째, 단순히 요약문을 음성으로 합성해주는 것이 아닌, 사용자가 원하는 특정 인물의 음색을 모방할 수 있는 맞

춤형 음성 합성 모듈을 제공해줄 수 있도록 설계하였다. 마지막으로, 자연스러운 음성 합성을 위해 요약문에 포함된 숫자와 영어를 한글로 변환하는 음역(transliteration) 모듈을 구축하였다. 각 특징에 대한 보다 구체적인 설명은 다음과 같다.

문서요약 모델을 학습하기 위해서는 공개된 대용량 한국어 뉴스 문서요약 데이터셋을 이용하여 KoBART 모델을 학습하였다. Shleifer and Rush(2020)에 따르면 BART 기반 모델을 미세조정하는 과정에서 인코더와 디코더의 레이어 중 일부만을 추출하여 학습하는 Shrink-and-Finetune(SFT) 방법을 사용할 경우 더 가벼운 모델인 DistilBART를 구축할 수 있다고 알려져 있다. 따라서 본 연구에서도 한국어 데이터에 대해 SFT 방법을 사용하여 DistilKoBART를 구축하여 모델 경량화를 달성하였다. 이러한 방식으로 모델 경량화를 진행하게 되면 더 적은 수의 파라미터를 사용할 뿐만 아니라 요약 정량지표 관점에서 성능 또한 향상됨을 확인하였다.

이후 음성합성(text-to-speech) 모델 학습을 위해서는 한국어 발화로 구성된 KSS 데이터셋<sup>3)</sup>을 이용하여 음성합성 모델을 사전학습한 후, 합성하고자 하는 인물의 음성 발화를 사용하여 추가 학습을 수행하였다. 이는 합성하고자 하는 인물의 음성 발화를 구하기 어려운 현실적인 상황에서도 준수한 품질의 음성을 합성할 수 있도록 전이 학습(transfer learning) 기법을 이용한 것이다. 음성합성 모델은 Tacotron 2(Shen *et al.*, 2018)를 기반으로 하였고, 인코더의 결과물을 음성 발화로 변환하는 기존 WaveNet(Oord *et al.*, 2016) 모듈을 WaveGlow(Prenger *et al.*, 2019) 모듈로 교체하였다. 이는 음성 발화 생성 품질을 최대한 유지하면서 빠르게 결과물을 추론하여 실제 사용 상황에 잘 대응하기 위함이다.

추가적으로 문서요약과 음성합성 모델을 연결하는 추론 단계에서 요약문에 포함된 숫자나 영어를 한국어로 변환하는 음역(transliteration)모듈을 구축하였다. 음역모듈은 텍스트를 음성으로 변환하기 위해 학습해야 하는 음성 발화 및 텍스트의 쌍으로 구성된 데이터에 숫자 및 영어 표현이 충분하지 않은 현실적인 상황에 대응하는 역할을 수행한다. 음역모듈은 크게 규칙(rule) 기반 모듈과 신경망 기반 모듈을 사용하였다. 규칙 기반 음역모듈은 뉴스에서 자주 등장하는 날짜 및 개수를 표현하는 숫자를 빠르게 한글로 변환할 수 있다. 그리고 LSTM(Hochreiter and Schmidhuber, 1997) 구조를 갖는 신경망 기반 음역모듈은 다양한 형태로 등장하는 영어 표현을 효과적으로 한글로 변환할 수 있다. 문서요약과 음성합성, 그리고 음역모듈을 통합하여 실험을 수행한 결과 제안하는 프레임워크가 길이가 긴 원문의 내용을 효과적으로 요약한 후 음성 발화로 출력할 수 있음을 확인하였다.

본 논문의 구성은 다음과 같다. 먼저 제2장에서는 문서요약과 음성합성 분야의 세부 내용 및 과업에 따른 차이에 대해

1) <https://aihub.or.kr/aidata/8054>.

2) <https://github.com/SKT-AI/KoBART>.

3) <https://www.kaggle.com/bryanpark/korean-single-speaker-speech-dataset>.

설명한다. 이후 제3장에서는 제안하는 통합 프레임워크를 소개하면서 추가적으로 수행한 모델 경량화 및 음역 과정에 대해 설명한다. 제4장에서는 실험 설계 및 결과에 대해 소개하면서 제5장에서 본 연구의 결론과 추후 연구 방향에 대해 서술한다.

## 2. 관련 연구

### 2.1 문서요약

문서 요약은 결과물의 형태에 따라 크게 추출요약과 생성요약 두 가지로 구분된다. 추출요약은 원문의 내용 중 중요하다고 판단되는 문장의 인덱스를 추출하여 요약문으로 활용하는 방법론이다. 따라서 각 문장에 해당하는 인덱스를 분류(classification)하거나 순서(rank)를 매기는 방식으로 문제가 정의된다. 생성요약은 원문의 내용 중 핵심적인 내용을 간추려 새로운 문장으로 생성해내는 NLG 기반의 방법론이다. 새로운 문장을 구성해야 하기 때문에 추출요약 과업보다 난이도가 높지만 보다 자연스러운 요약문을 얻을 수 있다는 장점이 존재한다. 본 연구에서는 생성요약 과업이 현실세계에 적용하는 응용연구로서 더 적합하다고 판단하였기 때문에 생성요약 모델을 사용하여 실험을 진행하였다.

현재 생성요약 모델 중 가장 우수한 성능을 보이는 모델은 BART를 기반으로 하고 있다(Aghajanyan *et al.*, 2021). BART는 Facebook에서 공개한 인코더-디코더 형태로 구성된 모델로, 입력 텍스트에 다양한 노이즈(noise)를 추가하고 이를 복원하는 방식의 사전학습을 진행한다. 사전학습을 위해 byte-level Byte-Pair-Encoding(BPE)을 사용하여 토큰화 하였고, 사용된 노이즈는 총 다섯 가지로 각각 토큰 마스킹(token masking), 토큰 삭제(token deletion), 텍스트 채우기(text infilling), 문장 혼합(sentence permutation), 그리고 문서 회전(document rotation)으로 불린다.

토큰 마스킹 방법은 BERT 모델에서 제안한 방법과 같이 문장 내 임의의 토큰에 마스킹을 적용한 후 이를 복원하는 학습 방법이다. 토큰 삭제 방법은 입력 문장에서 임의로 선택된 토큰을 삭제한 후 삭제된 토큰의 위치를 찾는 학습 방법이다. 텍스트 채우기 방법은 SpanBERT(Joshi *et al.*, 2020)에서 영감을 받은 방법이며, 문장 내 토큰을 마스킹한 후 마스킹된 토큰이 총 몇 개인지를 예측하는 학습 방법이다. 이 때 한 개 이상의 연속된 토큰을 마스킹할 수도 있지만 어떠한 토큰도 마스킹하지 않은 경우에도 마스킹 토큰을 삽입하여 마스킹된 토큰의 개수를 예측해야 한다는 특징이 있다. 문장 혼합 방법은 문장 내 문장을 구분한 후 순서를 섞는 방식으로 노이즈를 추가하는 학습 방법이다. 마지막으로 문서 회전은 문서 내 임의의 토큰을 선택한 후 문서를 해당 토큰 위치부터 시작하도록 순서를 회전하여 실제 문서의 시작을 구분하도록 예측하는 학습 방법이다. 해당 논문의 실험 결과에 따르면 텍스트 채우기 방

법을 사용하여 사전학습을 진행했을 때 가장 우수한 성능을 보였다.

이후 BART 모델을 문서요약 과업에 맞게 미세조정하는 단계에서는 원문을 입력 데이터로, 대응하는 요약문을 레이블(label)로 활용한다. 디코더는 매 시점마다 미리 정의된 어휘 사전의 분포에서 가장 등장할 확률이 높은 단어를 선택하고 각 시점의 출력값을 취합하여 최종 요약문을 생성한다. 본 연구에서 사용한 KoBART의 어휘 사전은 30,000 차원으로 구성되어 있다.

### 2.2 음성합성

Tacotron(Wang *et al.*, 2017)은 Google에서 공개한 음성합성 모델이다. Tacotron에서는 음성합성을 위해 전통적인 방식으로 언어적인 특성과 음성 특성을 추출하는 대신, 텍스트와 음성의 쌍으로 구성된 많은 데이터로 학습한 단일 신경망으로 이를 대체하였다. Tacotron은 인코더-디코더와 어텐션(attention) 구조의 형태로 구성되어 있으며, 인코더에서는 텍스트를 캐릭터(character) 단위로 나눈 원-핫(one-hot) 벡터를 입력으로 받아 입력 텍스트에 대한 정보를 추출한다. 이 과정에서 문장 내의 정보를 추출하는 데 효과적이라 알려져 있는 CBHG(1-D convolutional bank + highway network + bidirectional GRU) 모듈을 사용한다. 디코더에서는 인코더에서 전달되는 정보와 어텐션 기법을 활용하여 음성 특징 벡터인 멜-스펙트로그램(mel-spectrogram)을 순차적으로 출력한다. 이후 CBHG 모듈을 사용하여 후처리를 진행한 후 Griffin-Lim(Griffin and Lim, 1984) 알고리즘을 사용하여 최종 음성 신호를 생성한다. <Figure 1>은 멜-스펙트로그램의 출력 예시를 나타낸다(Shen *et al.*, 2018).

Tacotron2는 기존 Tacotron의 한계를 개선한 모델이다. 기존 Tacotron에서 최종 음성을 출력하기 위하여 사용하는 Griffin-Lim 알고리즘은 WaveNet과 같은 신경망 기반의 방법론보다 합성되는 음성의 품질이 낮다는 한계가 존재했다. 따라서 해당 연구에서는 기존 Tacotron 기반의 모델을 사용하지만 디코더를 통해 생성된 멜-스펙트로그램을 WaveNet 모델의 입력으로 사용하여 상대적으로 높은 품질의 음성을 합성하는 방법을 제안하였다.

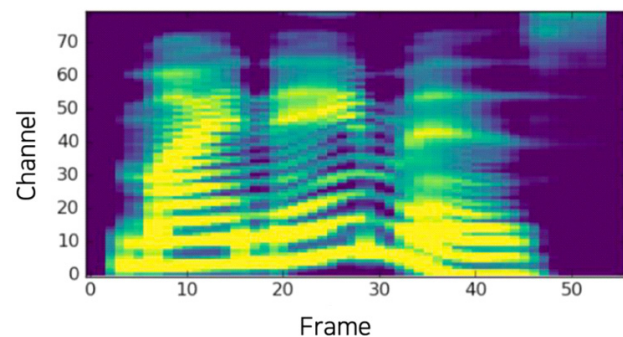


Figure 1. Example of Mel-Spectrogram

### 2.3 모델 경량화

모델 경량화(model compression)는 큰 모델이 좋은 성능을 보이지만 용량 및 추론 시간의 한계로 서비스 단계에서 활용할 수 없는 상황을 해결하기 위해 등장한 방법론이다. 그 중 지식 증류(knowledge distillation) 기법은 큰 모델의 지식(knowledge)을 정의하고, 상대적으로 작은 모델이 이를 전달받아 좋은 성능을 얻도록 유도하는 학습 방법이다(Hinton *et al.*, 2015). 지식 증류 기법은 자연어 처리, 이미지 처리, 음성 처리 등 다양한 도메인의 과업에서 범용적으로 활용 가능하다. 또한 모델의 구조에 의존하지 않고 경량화를 수행할 수 있다는 것이 장점이며 모델의 예측 확률 분포, 모델의 중간 레이어의 벡터 표현, 또는 모델 내 레이어 간 관계 등 다양한 지식을 정의하여 작은 모델에 전달할 수 있다(Gou, Yu, Maybank, & Tao, 2021).

## 3. 제안 방법론

### 3.1 통합 프레임워크

본 연구에서 제안하는 통합 프레임워크의 학습 및 추론 절차는 <Figure 2>에서 확인할 수 있다. 제안한 프레임워크에서는 관련 연구에서 소개한 딥러닝 기반의 KoBART(문서요약), Tacotron2(음성합성), Waveglow(음성합성) 세 가지가 사용되었으며, 문서요약과 음성합성이 별도의 학습 과정을 거친 후 추론 단계에서 통합된다.

문서요약 모델인 KoBART를 학습할 때는 위키피디아 등 일반적인 도메인의 텍스트로 사전학습된 모델을 기반으로 하여 수집한 뉴스 도메인의 데이터로 미세조정을 수행하였다. 음성합성 모델인 Tacotron2를 학습할 때는 단일 여성 화자의 발화로 구성된 KSS 데이터셋을 사용하여 모델을 사전학습한 후,

합성하고자 하는 인물의 음성 발화 데이터로 추가 학습을 수행하였다. 마지막 추론 단계에서는 문서요약 모델의 추론 결과인 요약문이 음성합성 모델의 입력 텍스트로 사용되어 요약문에 대한 합성된 발화를 출력한다.

본 연구에서 제안하는 통합 프레임워크의 활용도를 높이기 위해서는 요약문을 자연스러운 발화로 이어지게 하는 완성도 높은 음역 모듈이 필수적이며, 추론 단계에서의 연산 자원 효율화 및 시간 단축을 위한 모델 경량화도 함께 고려되어야 한다. 이를 위해 본 연구에서 사용된 음역모듈과 수행한 경량화 과정에 대해서는 각각 다음 두 절에서 설명한다.

### 3.2 음역모듈

음성합성 모델이 숫자나 영어 등 한글이 아닌 텍스트까지 발화로 합성하기 위해서는 숫자와 영어가 포함된 충분한 학습 데이터를 확보해야 한다. 그러나 현실적으로 합성하고자 하는 인물에 대한 많은 발화 데이터를 확보하는 것이 어려운 경우가 많다. 따라서 제안하는 프레임워크에서는 문서요약 모델의 추론 결과인 요약문 내 포함된 숫자와 영어 표현을 한글로 바꾸어 주는 음역모듈을 도입하였다.

모듈을 구축하기 위해 규칙(rule) 기반의 방법론과 신경망 기반의 방법론을 함께 활용하였다. 먼저 규칙 기반의 음역모듈은 숫자를 한글로 변환하는 상황에서 사용되었다. 예를 들어 ‘7시’를 ‘일곱 시’로, 그리고 ‘7월’을 ‘칠월’로 변환하는 등 학습 데이터가 속한 뉴스 도메인에서 충분히 등장할 수 있다고 판단되는 상황의 음역 표현에 대한 규칙을 미리 설정하였다. 이는 자주 등장하는 숫자 표현에 대한 빠른 추론이 가능하다는 장점이 존재한다. 뉴스 데이터의 특성 상 앞서 언급한 경우를 제외한 특수한 숫자 표현은 빈번하게 등장하지 않기 때문에 빠른 추론을 목적으로 할 경우 규칙 기반 음역모듈이 효

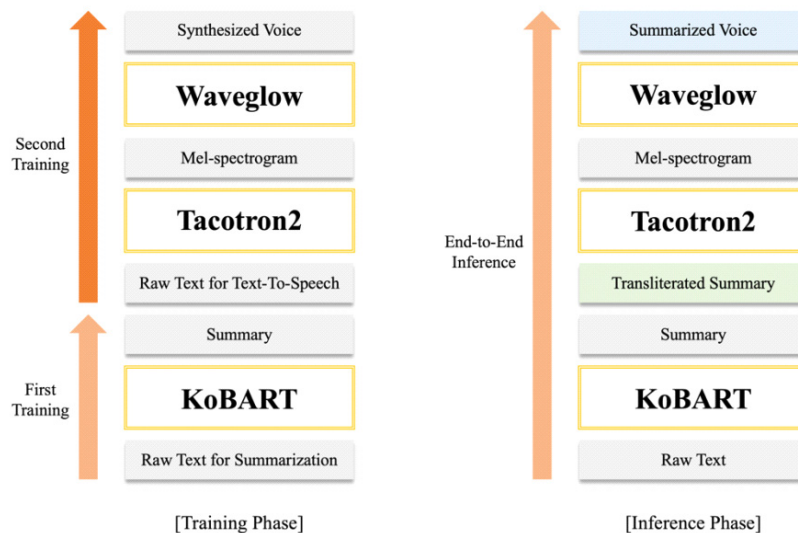


Figure 2. Overall Architecture

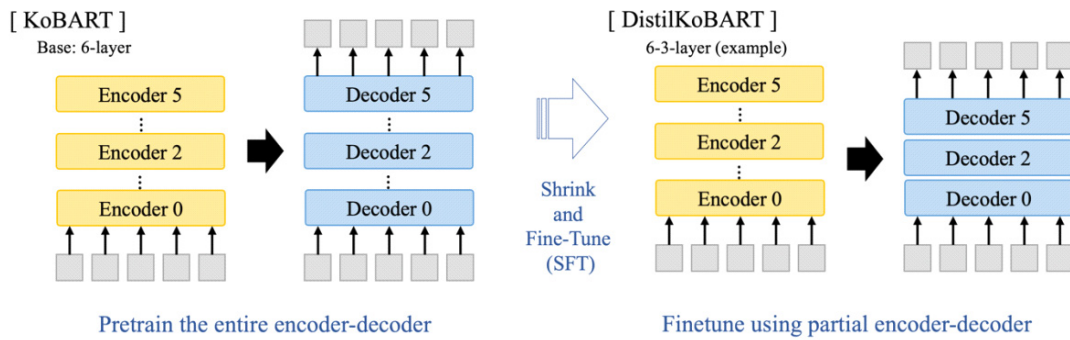


Figure 3. DistilKoBART Construction Process

과적이라 할 수 있다.

두 번째로 영어 표현에 대한 음역 작업은 LSTM 신경망 기반의 모듈을 사용하였다. 시퀀스 데이터를 효과적으로 처리할 수 있는 LSTM과 어텐션 기반의 모델을 사용함으로써 변환이 필요한 영어 텍스트가 입력으로 들어왔을 때 이에 대응하는 한글 표현을 출력할 수 있다. 신경망 기반의 음역모듈은 규칙 기반 음역모듈보다 상대적으로 추론 속도가 느리지만 다양한 형태의 입력에 대해서 상대적으로 안정적인 음역 품질을 보인다. 신경망 기반 음역모듈의 경우 웹 상에 공개된 기학습된 모델을 사용하였다.<sup>4)</sup>

### 3.3 모델 경량화

본 연구에서는 응용 서비스 측면에서 모델의 용량을 줄이고 각 단계에서 소요되는 추론 시간을 감소시키고자 경량화 작업을 수행하였다. 문서요약 모델에 대해서는 지식 종류 기법의 일종으로, 사전학습된 KoBART를 미세조정하는 과정에서 레이어의 일부를 추출한 후 학습을 수행하는 SFT 기법을 사용하여 DistilKoBART를 구축하였다(Shleifer and Rush, 2020). Figure 3은 DistilKoBART를 학습하는 과정을 나타낸 것이다. DistilKoBART는 추출하는 레이어의 개수에 따라 다양한 형태로 구축할 수 있는데, 본 연구에서는 인코더 3개-디코더 3개, 인코더 3개-디코더 6개, 그리고 인코더 3개-디코더 3개로 구성된 DistilKoBART를 구축하였다. 이후 실험 결과를 제시할 때 각 모델을 DistilKoBART-6-3, DistilKoBART-3-6, DistilKoBART-3-3으로 명명하였다.

음성합성 모델인 Tacotron2는 모델의 출력값인 멜-스펙트로그램을 음성으로 변환하기 위해 추가적으로 WaveNet 모듈을 사용한다. 그러나 자기-회귀(auto-regressive) 기반의 WaveNet 모듈은 음성 합성 속도가 느리다는 한계점이 존재한다. 따라서 본 연구에서는 해당 모듈 대신 Waveglow 모듈로 대체함으로써 추론 속도 측면에서의 경량화를 수행하였다. Waveglow

는 상대적으로 빠른 음성합성 속도를 보이는 flow 기반의 모델이며, 자기-회귀 형태가 아닌 역함수 변환이 가능한 함수를 사용하여 음성 데이터의 분포를 직접적으로 모델링하는 방법론이다(Prenger *et al.*, 2019). 모델 크기 측면에서는 WaveNet보다 크지만 추론 속도 측면에서는 매우 유리하며, 정성적으로 합성된 음성의 품질을 평가했을 때 두 모델의 결과물 간 유의미한 차이가 나지 않는다고 판단하였다.<sup>5)</sup>

## 4. 실험 및 결과

### 4.1 데이터셋 설명

문서요약 모델을 학습하기 위해서 2020년 AI 데이터셋 구축 사업의 일환으로 비플라이소프트 주관 하에 구축된 한국어 문서요약 텍스트 데이터 중 뉴스기사에 해당하는 생성요약 데이터를 사용하였다. 훈련, 검증, 평가 데이터의 수는 각각 30만 개, 1만 개, 1만 개로 구성되어 있으며 훈련 데이터는 정성적인 전처리를 통해 약 26만 개의 데이터를 최종적으로 사용하였다.

음성합성 모델은 전이 학습 기법을 사용하여 학습하였기 때문에 많은 양의 데이터로 사전학습을 진행하였다. 사전학습에 사용된 KSS 데이터셋은 단일 여성 화자가 약 13,000 건의 문장을 발화하여 공개된 데이터셋이며, 12시간 이상의 음성 발화로 구성되어 있다. 이후 합성하고자 하는 음성으로는 비교적 음성의 정체성이 뚜렷한 연예인인 장기하와 본 연구의 1저자를 선정하였다.

두 인물의 음성은 각각 다른 방법으로 수집하였다. 먼저 연예인 장기하의 음성은 오디오북 플랫폼인 ‘밀리의 서재’와 ‘네이버 오디오북’에서 오디오북을 구입하여 추출하였다. 추출한 음성 발화는 네이버의 ‘클로바 스피치’ 서비스를 이용하여 문장 단위로 분리하였고, 총 3,752 건의 문장에 대해 텍스트-음성 쌍으로 구성된 데이터를 확보하였다. 반면 본 연구저자의 음

4) [https://github.com/gritmind/engkor\\_transliterator](https://github.com/gritmind/engkor_transliterator).

5) <https://nv-adlr.github.io/WaveGlow>.

성 발화를 수집할 때에는 KSS 데이터셋에서 제공하는 스크립트를 직접 녹음하여 3,000건의 데이터를 수집하였다. 오디오 복과는 다르게 주변 소음이 포함되어 있기 때문에 별도의 전처리를 거쳐 소음 및 불필요한 공백을 제거하였다.

## 4.2 정량 평가 결과

### 4.2.1 Rouge Score

문서요약 모델을 정량적으로 평가하기 위해 Rouge 스코어를 사용하였다(Lin, 2004). Rouge 스코어는 사람이 생성한 정답 요약문과 모델이 생성한 요약문 간 질적 차이를 정량화하는 평가지표이다. 기본적으로 두 요약문 간 겹치는 토큰의 개수를 비교하는데, 한 번에 고려하는 겹치는 토큰의 개수에 따라 Rouge-1, Rouge-2 등으로 표기한다.

$$ROUGE-N = \frac{\sum_{S \in \{Reference Summaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{Reference Summaries\}} \sum_{gram_n \in S} Count(gram_n)} \quad (1)$$

Eq. (1)은 Rouge-N을 계산하는 식이며, 이는 정답 요약문 내 존재하는 N-gram에서 모델의 요약문과 겹치는 N-gram의 비율로 표현된다. 본 연구에서는 Rouge-1, Rouge-2, 그리고 Rouge-L을 사용하여 성능을 측정하였다. 여기서 Rouge-L은 Longest Common Subsequence를 의미하며, 정답 요약문과 모델 요약문 사이에서 공통적으로 등장하는 가장 긴 길이의 토큰 시퀀스를 계산하여 Rouge 스코어를 계산하는 방식이다.

Rouge 스코어는 Recall 기반의 평가지표이지만 문서요약을 평가하는 데 있어 변형된 계산 방법을 사용하는 경우도 존재한다. Eq. (1)의 분모는 정답 요약문에 해당하는 N-gram의 개수를 의미하지만, 모델 요약문에 해당하는 N-gram의 개수로 변형한다면 Precision 기반의 성능을 측정할 수 있다. 추가적으로 Recall 기반의 Rouge 스코어와 Precision 기반의 Rouge 스코어의 조화 평균을 통해 F1 기반의 Rouge 스코어를 유도할 수 있다. 본 연구에서는 F1 기반의 Rouge 스코어를 기준으로 모델 간 성능을 비교하였다.

### 4.2.2 문서요약

문서요약 모델의 정량적인 성능은 <Table 1>에서 확인할 수 있다. 우선 기본 요약 모델인 KoBART를 사용할 경우 Rouge-1,

2, L 스코어는 각각 0.507, 0.339, 0.403으로 나타났으며, 이 결과는 2020년 12월 종료된 한국어 문서 생성요약 AI 경진대회 리더보드 기준 5위에 해당하는 점수로서 상당히 우수한 요약 성능을 나타내고 있음을 알 수 있다.<sup>6)</sup> 서비스 활용 관점에서 모델 경량화의 효과를 살펴보면 우선 KoBART의 인코더와 디코더 레이어 수를 모두 3개로 축소시킨 경우 모델의 파라미터 수는 약 38% 감소한 반면, Rouge-1, 2, L 스코어가 근소하게 감소하는 것을 확인할 수 있다. 요약 성능 하락을 방지하는 효율적인 모델 구축을 위해 인코더와 디코더의 레이어 수만 각각 6개로 증가시킨 결과, 디코더의 수를 증가시킬 경우 인코더와 디코더 레이어 수가 3개인 모델에 비해 모델의 파라미터 수가 약 38% 증가한 반면, Rouge-1, 2, L 스코어가 더욱 하락하여 효과적이지 않은 것으로 나타났다. 반면에, 인코더의 레이어 수를 6개로 증가시킬 경우 인코더와 디코더 레이어 수가 3개인 모델에 비해서는 파라미터의 수가 약 28% 증가하였으나, 기본 요약 모델에 비해서는 여전히 21% 감소한 파라미터의 수를 가지면서도 Rouge-1, 2, L 스코어가 모두 근소하게 향상되어 요약 모델의 성능을 유지하면서도 효율성을 향상시키는 것을 확인하였다.

## 4.3 정성 평가 결과

### 4.3.1 문서요약

정답 요약문과 모델 요약문을 비교한 결과는 <Table 2>에서 확인할 수 있다. 정량 평가 시 가장 높은 성능을 보였던 DistilKoBART-6-3을 사용하여 모델의 요약문을 출력하였다. 첫 번째 예시에서는 정답 요약문과 모델의 요약문이 거의 일치하는 것을 확인할 수 있다. 이는 모델이 주어진 문서 내에서 핵심을 잘 포착한다는 것을 의미한다. 두 번째 예시의 모델 요약문에서는 ‘맞춤형 금융서비스’를 제공하는 주체가 되는 핵심 키워드인 ‘서민금융지역협의체’를 먼저 제시하고 세부 내용을 후술하였다. 이는 가독성 측면에서 정답 요약문보다 유리한 요약 방법이다. 특히 세 번째 예시의 경우, 사람이 작성한 정답 요약문에 ‘여행체임’이라는 오타가 존재하는 반면 모델의 요약문은 오타 없이 간결한 문장을 생성하는 것을 확인할 수 있다. <Table 2>의 요약문에 대응하는 원문은 <Appendix>의 <Table A1>에서 확인할 수 있다.

Table 1. Rouge-F1 Score of Summarization Models

	# of parameters	Rouge-1	Rouge-2	Rouge-L
KoBART	120M	0.507	0.339	0.403
DistilKoBART-3-3	74M	0.501	0.331	0.399
DistilKoBART-3-6	102M	0.496	0.327	0.392
DistilKoBART-6-3	95M	<b>0.510</b>	0.340	<b>0.408</b>

6) <https://dacon.io/competitions/official/235673/leaderboard>.

**Table 2.** Difference between Reference and Model Summary

Reference Summary	Model Summary
달콤커피는 가수이자 화가로 전천후 활동 중인 ‘진한 음색’ 솔비가 오는 15일과 22일 오후 8시부터 각각 달콤커피 논현점과 산본 퍼트인점에서 배란다라이브 공연을 개최한다고 5일 밝혔다.	달콤커피는 가수이자 화가로 전천후 활동 중인 ‘진한 음색’ 솔비를 ‘이 달의 아티스트’로 선정하고, 오는 15일과 22일 오후 8시부터 각각 달콤커피 논현점과 산본 퍼트인점에서 배란다라이브 공연을 실시한다고 5일 밝혔다.
서민금융통합지원센터를 주축으로 익산시는 다양한 맞춤형 금융 서비스를 안내하고 제공받을 수 있도록 주민들에게 서민금융 지원 제도를 알리고 맞춤형 서비스를 연계해 줄 ‘서민금융 지역협의체’를 구축한다고 밝혔다.	익산시는 4일 근로복지공단 익산지사, 지역자활센터, 소상공인시장진흥공단 익산센터, 지역 소재 금융회사 등 유관기관과 간담회를 갖고 ‘서민금융 지역협의체’를 구축했다고 밝혔으며 협약에 따라 각 기관은 앞으로 시민 개개인의 상황과 목적에 맞는 다양한 맞춤형 금융서비스를 안내하고 제공받을 수 있도록 지원한다.
‘여행체험 1번지 전북’을 추진 중인 전북도가 최근 한국으로의 관광이 급증하고 있는 베트남을 공략하기 위해 하노이에서 전북관광 설명회를 개최하고 세일즈콜 등 홍보활동을 전개하며 지난 27일부터 30일까지 하노이에서 진행되는 ‘2019 베트남 하노이 국제관광전에 참가해 전북의 관광자원을 집중 홍보한다고 밝혔다.	‘여행체험 1번지 전북’을 추진 중인 전북도가 최근 한국으로의 관광이 급증하고 있는 신흥 시장 베트남을 공략하기 위해 하노이에서 전북관광 설명회 및 현지 여행사 방문 세일즈콜 등 홍보활동을 적극 전개한다고 28일 밝혔다.

**Table 3.** Transliteration Results

Original Sentence	Transliterated Sentence
범정부 코로나19 백신도입 TF는 이날 이렇게 밝히고, 이스라엘과 백신 교환으로 공급되는 화이자 백신은 오는 7일 오전 7시 15분에 인천공항에 들어온다고 설명했다.	범정부 코로나일구 백신도입 티에프는 이날 이렇게 밝히고, 이스라엘과 백신 교환으로 공급되는 화이자 백신은 오는 칠일 오전 일곱시 십오분에 인천공항에 들어온다고 설명했다.
AI전문기업 마인즈랩 이종미 전무는 “25종 이상의 지능형 CCTV 서비스가 개발된 일본과 달리 우리나라는 아직 초기 수준”이라며 “정부 주도의 데이터 확보와 실증 연구 사업으로 우리 역시 개발속도를 높일 수 있게 됐다”고 말했다.	에이아이전문기업 마인즈랩 이종미 전무는 “이십오종 이상의 지능형 시시티브이 서비스가 개발된 일본과 달리 우리나라는 아직 초기 수준”이라며 “정부 주도의 데이터 확보와 실증 연구 사업으로 우리 역시 개발속도를 높일 수 있게 됐다”고 말했다.
기상청은 “총 강수량이 100mm가 넘는 곳이 있고 순간적으로 초속 20m 안팎의 강한 돌풍이 불거나 우박이 떨어질 가능성이 있다”며 주의를 당부했다.	기상청은 “총 강수량이 백밀리미터가 넘는 곳이 있고 순간적으로 초속 이십미터 안팎의 강한 돌풍이 불거나 우박이 떨어질 가능성이 있다”며 주의를 당부했다.

4.3.2 음역모델

숫자 및 영어 표현이 포함된 뉴스 기사 문장에 대해 음역처리를 수행한 결과는 <Table 3>에서 확인할 수 있다. 시간과 날짜를 언급하는 숫자 표현과 약어를 지칭하는 영어 표현이 많이 등장하는 첫 번째 사례와 두 번째 사례에서 본 연구에서 사용한 음역모델이 한글 표현을 상황에 맞게 잘 변환하였다. 또한 ‘mm’, ‘m’ 등 단위를 나타내는 특수 표현이 등장하는 세 번째 사례에서도 정확하게 한글 발화 형태로 변환해주는 것을 확인할 수 있다.

4.3.3 음성합성

일반적으로 음성합성 모델을 평가하기 위해서는 임의의 평가자를 선별하여 주관적인 판단을 통해 점수를 측정하는 Mean Opinion Score 방법을 사용한다(Shen et al., 2018). 본 연구에서

는 음성 합성 모델의 출력 결과를 누구나 확인하여 주관적으로 평가할 수 있도록 데모 페이지(7)에 공개하였다. <Figure 4>와 <Figure 5>는 합성된 음성을 실제 음성과 비교하는 데모 페이지의 일부이다. 먼저 <Figure 4>에서는 뉴스 기사 원문을 요약하고 이를 음성으로 합성하는 일련의 절차에 대한 예시를 보여주고 있다. 먼저 뉴스 원문에 대해서 사람이 직접 요약한 Gold Summary와 본 연구에서 사용한 DistilKoBART(6-3)의 요약문을 보여주고 있으며, 본 연구에서 사용한 요약 모델이 실제 Gold Summary와 완전히 일치하지 않더라도 원문의 핵심 내용을 모두 포함하여 요약을 하는 것을 확인할 수 있다. 이후 Transliterated Summary에는 음성 합성을 위해 숫자 및 특수기호를 음역으로 변환한 결과를 보여주고 있으며, 마지막으로 Text-to-Speech에서는 Transliterated Summary를 특정인(연예인 장기하)의 목소리로 읽어주는 결과를 청취할 수 있게 구성

7) <https://youngerous.github.io/kobart-voice-summarization/>.

하였다. <Figure 5>는 음성 합성의 정성적인 비교를 위하여 두 명의 화자(연예인 장기하 및 본 논문의 제1저자)의 실제 음성과 합성된 음성을 청취할 수 있는 예시이다. 합성된 장기하의 음성과 실제 음성을 비교하기 위해 음성 합성 모델 학습에 사용하지 않은 외부 인터뷰 음성 발화를 사용하였고, 마찬가지로 같은 문장에 대한 본 연구 저자의 음성 발화도 함께 비교하였다.

#### 4.4 실험 결과 요약

본 장에서는 한국어 문서요약 모델과 음성합성 모델에 대한 정량적, 정성적 평가를 수행하였다. 제안하는 경량화된 문서

요약 모델은 기존 KoBART 모델보다 더 적은 파라미터를 가지면서도 좋은 성능을 보이고, 정성적으로 판단하였을 때 문서 내의 핵심 내용을 잘 포착함을 확인하였다. 특히 사람이 작성한 정답 요약문에서 오다가 존재할 수 있는 반면, 잘 학습된 모델은 올바른 단어를 사용하여 요약할 수 있기 때문에 구축 비용과 정확도 측면에서 효과적이라는 사실을 보여주었다. 음성합성 모델에 대해서도 학습한 목소리와 유사한 음성을 생성한 결과를 데모 페이지를 통해 공개하였다. 또한 문서요약 모델과 음성합성 모델을 연결하는 음역모델은 다양한 상황에서 등장하는 숫자 및 영어 표현을 효과적으로 한글로 변환하였다. 이는 주어진 데이터를 잘 학습한 모델이 추가적인 비용을 지불하지 않고도 문서를 잘 요약하여 음성으로 결과를 제공해

		Content
1	Original Document	[서울경제] 정상적인 학사운영이 불가능한 사립대학과 학교법인에 폐쇄·해산 명령을 내릴 수 있도록 한 고등교육법은 헌법에 어긋나지 않는다는 헌법재판소의 결정이 나왔다. 헌재는 지난 2011년 폐쇄 조치된 세림학원 이사장 정모씨 등이 '학교폐쇄·법인해산 명령'을 규정한 고등교육법이 위헌이라며 낸 헌법소원 사건에서 재판관 전원일치 의견으로 합헌 결정했다고 2일 밝혔다. 전남 강진에서 성화대학을 운영하는 세림학원은 지난 2011년 6~7월 교육부 특별감사에서 설립자가 교비 등 65억원을 횡령하고 수업일수 미달학생 2만3,848명에게 부당하게 학점을 부여한 사실이 적발됐다. 교육부가 현장조사한 결과 아예 진행하지 않는 강의로 많았다. 성화대학은 시정 요구를 받고도 이행하지 않다가 같은해 11월 학교를 폐쇄하고 법인을 해산하려는 정부의 명령이 내려지자 행정소송을 냈다. 헌법소원은 상고심 재판 중 제기했다. 성화대학은 2016년 4월 대법원에서 패소를 확정 받고 최종 폐쇄했다. 헌재는 "정상적인 학사 운영이 불가능한 학교를 그대로 방치하는 것은 사회적으로 많은 혼란을 야기할 수 있다"며 "우리나라는 사립학교도 공교육 체계에 편입시켜 국가 등의 지도·감독을 받도록 함과 동시에 많은 재정적 지원과 각종 혜택을 부여하는데 목적달성이 불가능한 학교법인을 존치 시킨다면 학교법인에 특혜만을 제공하게 된다"고 지적했다.
1	Gold Summary	헌재는 지난 2011년 폐쇄 조치된 세림학원 이사장 정모씨 등이 '학교폐쇄·법인해산 명령'을 규정한 고등교육법이 위헌이라 주장한 사건에 정상적인 학사 운영이 불가능하다 판단, 각종 혜택을 부여하는데 목적달성이 불가능하다 판단하여 2016년 4월 대법원에서 패소 확정 받고 폐쇄했다.
1	DistilKoBART (6-3) Summary	지난 2011년 폐쇄 조치된 세림학원 이사장 정모씨 등이 '학교폐쇄·법인해산 명령'을 규정한 고등교육법이 위헌이라며 낸 헌법소원 사건에서 정상적인 학사운영이 불가능한 사립대학과 학교법인에 폐쇄·해산 명령을 내릴 수 있도록 한 고등교육법은 헌법에 어긋나지 않는다는 헌법재판소의 결정이 나왔다.
1	Transliterated Summary	지난 이천십일년 폐쇄 조치된 세림학원 이사장 정모씨 등이 '학교폐쇄·법인해산 명령'을 규정한 고등교육법이 위헌이라며 낸 헌법소원 사건에서 정상적인 학사운영이 불가능한 사립대학과 학교법인에 폐쇄·해산 명령을 내릴 수 있도록 한 고등교육법은 헌법에 어긋나지 않는다는 헌법재판소의 결정이 나왔다.
1	Text-to-Speech	▶ 0:00 / 0:21 ————— 🔊 ⋮

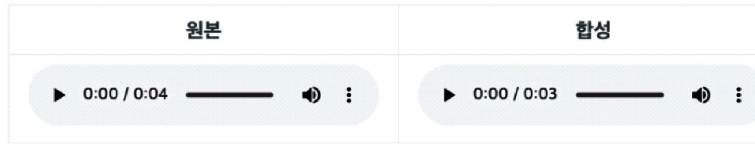
Figure 4. An Example of Text-to-speech with Document Summarization



## 장기하 및 저자

“딱 두 번 해봤을 뿐이긴 하지만, 나는 서핑을 좋아한다.”

- 장기하



- 저자

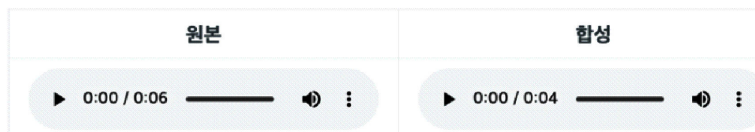


Figure 5. Text-to-Speech Comparison

야 하는 다양한 상황에서 활용될 수 있음을 시사한다.

## 5. 결론

본 연구에서는 현실 상황에서 길이가 긴 문서에 담긴 정보를 받아들이기 여유가 없는 경우가 많음을 인지하고 문서 내의 핵심 내용을 요약하여 음성으로 출력하는 프레임워크를 제안하였다. 특히 모델이 현실 상황에서 사용되어야 한다는 점을 고려하여 문서요약 모델에는 사전 학습 및 미세 조정 기반의 KoBART 모델과 지식 증류 기법을 결합한 DistilKoBART 모델을 활용하였고, 음성합성 모델에서는 많이 확보할 수 있는 음성 데이터로 사전 학습 후 타겟이 되는 목소리를 사용하여 목표하는 음성을 안정적으로 생성하는 전이 학습 기반의 방법론과 빠르게 음성을 합성하는 Waveglow 모델을 사용하는 방식으로 효율성을 추구하였다. 또한 학습 데이터가 충분하지 않은 상황에서 별도의 음역모듈을 구축하여 요약문에 등장하는 다양한 숫자 및 영어 표현을 적절한 한글 표현으로 변환하여 한글이 아닌 숫자 및 영어 표현에 대한 음성 표현을 효과적으로 구축하였다. 제안하는 프레임워크는 특히 많은 텍스트를 읽기 어려운 운전자나 텍스트 자체를 읽기 힘들어 하는 저시력자 등 특수한 상황에 처한 사용자에게 편의를 제공할 것으로 기대된다.

더 나은 품질의 요약된 음성 결과물을 얻기 위해 몇 가지 개선 방향 역시 존재한다. 우선 문서 요약 모델의 결과물이 사실에 기반하였는지를 판단하는 것이다. 본 연구에서 수행한 실험에서는 전반적으로 사실에 기반한 요약문이 도출되었으나, 최근의 연구까지도 문서요약 성능을 정량적으로 평가할 때에는 정답 요약문과 모델의 요약문이 겹치는 정도를 비교하는

단순한 방식을 사용하기 때문에 사실과 다른 요약문이 등장하지 않는다고 보장하지 못하는 상황이다. 따라서 요약문이 실제 상황을 반영하는지 평가하는 모델과 지표에 대한 연구가 필요하다. 두 번째로는 영어 및 숫자가 등장하는 보다 많은 상황에서 안정적으로 한글 음성 발화를 출력하는 것이다. 본 연구에서는 숫자 및 영어를 표현하는 음성 발화가 부족한 상황을 보완하기 위해 음역모듈을 구축하였지만, 숫자 및 영어를 포함하는 다양한 음성 발화를 구축할 수 있다면 별도의 음역모듈이 없어도 안정적으로 요약된 음성을 출력할 수 있을 것이다. 마지막으로 음성합성 모델이 합성할 수 있는 문장의 최대 길이를 개선하는 것이다. 모델의 크기가 한정되어 있기 때문에 현재는 한 번에 생성할 수 있는 발화의 양이 제한되어 있는 상황이다. 원 문서의 길이가 길어질수록 요약문 역시 길어질 가능성이 크기 때문에 하나의 요약문에 대한 음성을 합성할 때 여러 번의 추론을 거쳐야 하는 경우도 발생한다. 따라서 길이가 긴 입력 데이터를 잘 처리하는 모델의 사용을 고려하거나 모델의 추론 속도를 개선하여 결과를 빠르게 도출하는 연구가 필요할 것이다.

## 참고문헌

- Aghajanyan, A., Gupta, A., Shrivastava, A., Chen, X., Zettlemoyer, L., and Gupta, S. (2021), Muppet: Massive Multi-task Representations with Pre-Finetuning, arXiv preprint arXiv:2101.11038.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018), Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv preprint arXiv:1810.04805.
- Erkan, G., and Radev, D. R. (2004), Lexrank: Graph-based Lexical Centrality as Salience in Text Summarization, *Journal of Artificial*

- Intelligence Research*, **22**, 457-479.
- Gou, J., Yu, B., Maybank, S. J., and Tao, D. (2021), Knowledge Distillation: A Survey, *International Journal of Computer Vision*, **129**(6), 1789-1819.
- Griffin, D., and Lim, J. (1984), Signal Estimation from Modified Short-time Fourier Transform, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **32**(2), 236-243.
- Hinton, G., Vinyals, O., and Dean, J. (2015), Distilling the Knowledge in a Neural Network, arXiv preprint arXiv:1503.02531.
- Hochreiter, S., and Schmidhuber, J. (1997), Long short-term Memory, *Neural Computation*, **9**(8), 1735-1780.
- Joshi, M., Chen, D., Liu, Y., Weld, D. S., Zettlemoyer, L., and Levy, O. (2020), Spanbert: Improving pre-training by Representing and Predicting Spans, *Transactions of the Association for Computational Linguistics*, **8**, 64-77.
- Kågebäck, M., Mogren, O., Tahmasebi, N., and Dubhashi, D. (2014), Extractive Summarization Using Continuous Vector Space Models, *Paper presented at the Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC)*.
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. (2019), Bart: Denoising Sequence-to-sequence pre-training for Natural Language Generation, Translation, and Comprehension, arXiv preprint arXiv:1910.13461.
- Lin, C.-Y. (2004), Rouge: A Package for Automatic Evaluation of Summaries, *Paper Presented at the Text Summarization Branches out*.
- Liu, Y., and Lapata, M. (2019), Text Summarization with Pretrained Encoders, arXiv preprint arXiv:1908.08345.
- Mihalcea, R., and Tarau, P. (2004), Textrank: Bringing order into text, *Paper Presented at the Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*.
- Nallapati, R., Zhou, B., Gulcehre, C., and Xiang, B. (2016), Abstractive Text Summarization Using Sequence-to-sequence rnns and Beyond, arXiv preprint arXiv:1602.06023.
- Oord, A. V. D., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. (2016), Wavenet: A Generative Model for Raw Audio, arXiv preprint arXiv:1609.03499.
- Prenger, R., Valle, R., and Catanzaro, B. (2019), Waveglow: A flow-based Generative Network for Speech Synthesis, *Paper presented at the ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
- Radev, D. R., Hovy, E., and McKeown, K. (2002), Introduction to the Special Issue on Summarization, *Computational Linguistics*, **28**(4), 399-408.
- Rush, A. M., Chopra, S., and Weston, J. (2015), A Neural Attention Model for Abstractive Sentence Summarization, arXiv preprint arXiv:1509.00685.
- Shen, J., Pang, R., Weiss, R. J., Schuster, M., Jaitly, N., Yang, Z., ... Skerrv-Ryan, R. (2018), Natural tts Synthesis by Conditioning Wavenet on Mel Spectrogram Predictions, *Paper presented at the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
- Shleifer, S., and Rush, A. M. (2020), Pre-trained Summarization Distillation. arXiv preprint arXiv:2010.13002.
- Wang, Y., Skerry-Ryan, R., Stanton, D., Wu, Y., Weiss, R. J., Jaitly, N., ... Bengio, S. (2017), Tacotron: Towards end-to-end Speech Synthesis, arXiv preprint arXiv:1703.10135.

## 저자소개

**김탁영:** 성균관대학교 미디어커뮤니케이션학과에서 2020년 학사 학위를 취득하였다. 현재는 고려대학교 산업경영공학부에서 석사과정으로 재학 중이다. 연구 분야는 모델 경량화와 대화시스템이다.

**김지나:** 고려대학교 산업경영공학부에서 2020년 학사학위를 취득하였다. 현재는 고려대학교 산업경영공학부에서 석박통합과정으로 재학 중이다. 연구 분야는 이상치탐지와 자연어처리이다.

**강형원:** 인천대학교 산업경영공학과에서 2020년 학사학위를 취득하였다. 현재는 고려대학교 산업경영공학부에서 석사과정으로 재학 중이다. 연구 분야는 제조/IT 이상치탐지와 강화학습이다.

**김수빈:** 고려대학교 통계학과에서 2021년 학사학위를 취득하였다. 현재는 고려대학교 산업경영공학부에서 석사과정으로 재학 중이다. 연구 분야는 자연어처리이다.

**강필성:** 서울대학교 산업공학과에서 2003년 학사, 2010년 박사 학위를 취득하였다. 이후 현대카드 과장, 서울과학기술대학교 조교수로 근무하였으며, 현재는 고려대학교 산업경영공학부 부교수로 재직 중이다. 연구 분야는 정형 및 비정형 데이터를 활용한 데이터마이닝 및 기계학습 알고리즘 개발 및 제조/IT/공공분야 응용이다.

<Appendix>

Table A1. Raw Text for Summarization

Index	Raw Text
1	<p>가수이자 화가로 전천후 활동 중인 ‘진한 음식’ 솔비가 달콤커피 배란다라이브 무대를 통해 오랜만에 보컬리스트로서의 깊이 있는 매력을 선보인다. 다남의 프랜차이즈 커피 전문 브랜드 달콤커피는 가수 솔비를 ‘이달의 아티스트’로 선정하고, 달콤커피 매장에서 배란다라이브 공연을 실시한다고 5일 밝혔다. 배란다라이브는 오는 15일과 22일 오후 8시부터 각각 달콤커피 논현점과 산본피트인점에서 열린다. 신청 방법은 18일까지 달콤커피 공식 앱을 통해 응모 가능하며, 논현점 또는 산본피트인점에서 만원 이상 구매한 영수증 응모를 통해서도 참여할 수 있다. 당첨자는 오는 11일(논현점), 19일(산본피트인점) 달콤커피 공식 페이스북을 통해 발표되며, 당첨자에게는 1인 당 2매의 티켓이 지급된다. 11월의 아티스트로 선정된 솔비는 지난 2006년 혼성그룹 ‘타이퐁’의 리드보컬로 데뷔, 그룹과 솔로 활동을 병행하며 진한 음식과 파워풀한 가창력으로 꾸준히 사랑받고 있는 실력과 보컬리스트다. 최근 ‘음악하는 솔비’와 ‘미술하는 권지안(솔비의 본명)’의 협업, ‘셀프 콜라보레이션’이라는 독창적인 작업을 통해 음악과 미술의 경계를 넘어서신 자신만의 장르를 개척하는 등 다재다능한 매력으로 폭넓은 지지를 받고 있다. 솔비는 달콤커피 배란다라이브를 통해 특유의 파워풀한 가창력과 호소력 짙은 음식으로 오랜만에 ‘명품 보컬리스트’로서의 존재감을 유감없이 발휘할 예정이다. 또 현장 이벤트를 통해 팬들과 소통하며, 특유의 상큼발랄하고 사랑스러운 매력을 선보이는 등 달콤한 시간을 함께할 계획이다. 한편, ‘달콤커피 배란다라이브’는 매일 정동하, 김태우, 알리, 허각, 한동근, 플라이트 더스카이, 홍대광, 하이어뮤직, 라비, 권진아 등 역량있는 뮤지션들을 ‘이달의 아티스트’로 선정하고, 전국의 달콤커피 매장에서 라이브 공연을 실시하는 달콤커피만의 특별한 문화 이벤트다.</p>
2	<p>주민들에게 서민금융 지원 제도를 알리고 맞춤형 서비스를 연계해 줄 협업체계가 구축된다. 익산시는 4일 서민금융통합지원센터(센터장 이재갑)를 주축으로 근로복지공단 익산지사, 익산지역자활센터, 소상공인시장진흥공단 익산센터, 지역 소재 금융회사 등 유관기관과 간담회를 갖고 ‘서민금융 지역협의체’를 구축했다고 밝혔다. 협약에 따라 각 기관은 앞으로 시민 개개인의 상황과 목적에 맞는 다양한 맞춤형 금융서비스를 안내하고 제공받을 수 있도록 지원한다. 특히 소득기준 초과로 근로복지공단의 생계 자금지원이 어려운 근로자가 통합지원센터를 통해 지역 금융회사에서 대출을 받거나 긴급자금이 필요한 기초수급자가 통합지원센터와 연계해 미소금융 대출과 재산형성 지원을 받을 수 있도록 연계할 예정이다. 아울러 이 날 간담회에서는 서민금융지원 강화를 위해 향후 운영방안과 참여기관별 지원제도를 공유해 유기적인 안내가 가능하도록 종합홍보물 제작 등에 대해서도 논의됐다. 익산시는 “서민금융 지역협의체의 활성화에 주력하고 금융·고용·복지 기관들과 유기적으로 협업해 시민들이 몰라서 이용하지 못하는 일이 없도록 다양한 매체를 통해 지속적으로 시민 홍보에 힘쓰겠다”고 밝혔다. 한편 지난 2017년 개소한 익산 서민금융통합지원센터(☎1397)에서는 창업·사업운영·생활안정 등 필요자금을 담보나 보증 없이 저금리 지원, 생계자금 대출 보증 지원, 저금리 전환대출 지원, 채무조정 등 수요자가 한 자리에서 종합상담, 심사 및 지원까지 제공받을 수 있도록 서민금융진흥원, 신용회복위원회, 한국자산관리공사, 미소금융이 참여하고 있다. 또한 전산시스템도 구축해 서민을 위한 맞춤형 원스톱 금융·복지 상담서비스를 운영하고 있으며 6월 말 현재 총 1,780건(미소금융 1,240건, 신용회복위원회채무조정 240건, 자산관리공사 바뀔드림론 등 300건)의 서민 금융 서비스가 지원됐다.</p>
3	<p>‘여행체험 1번지 전북’을 추진 중인 전북도가 최근 한국으로의 관광이 급증하고 있는 신흥 시장 베트남을 공략하기 위해 모든 행정력을 집중하고 있다. 도는 베트남 관광객 유치에 하노이에서 전북관광 설명회 및 현지 여행사 방문 세일즈콜 등 홍보활동을 적극 전개한다고 28일 밝혔다. 이에 지난 27일부터 오는 30일까지 베트남 하노이에서 진행되는 ‘2019 베트남 하노이 국제관광전(2019 VITM)’에 참가해 전북의 다양한 관광자원을 집중 홍보한다. 특히, 도는 베트남 관광객들의 기호를 면밀히 분석해 ‘맞춤형 관광자원’ 홍보에 나서고 있다. 베트남 관광객들은 ‘봄 꽃’과 ‘가을 단풍’ 등을 선호하는 경향이 있기 때문에 전북관광 홍보관에서는 봄·가을 중심의 계절여행 및 미식여행 상품을 중점 소개한다는 전략이다. 또한, 베트남에서 전북 관광 인지도를 높이기 위해 페이스북 등 소셜 미디어를 활용한 포토 이벤트도 실시하고 있다. 아울러 28일에는 베트남 현지 30여개 아웃바운드 여행사를 초청해 ‘전북관광 설명회’도 개최했다. 29일에는 하노이 소재 주요 여행사를 방문, 1:1 맞춤형 전북여행 상담 등 세일즈콜을 실시하면서 베트남인들이 호응할 수 있는 전북관광을 홍보할 계획이다. 베트남 현지에서 전북관광 상품을 기획·운영 중인 사랑 트래블 에이전시(sarang travel agency) 대표 ‘당 티 투이 번(Dang Thi Thuy Van)’은 “봄과 가을 중심으로 전북관광 상품을 판매하고 있다”면서 “올해는 명상체험 코스도 추가해 홍보하고 있다”고 말했다. 김희욱도 관광총괄과장은 “베트남 관광객의 주요 관광지로 한국이 부상함에 따라 한국방문객 수가 매년 큰 폭으로 증가하고 있다”면서 “전북도에서는 해외관광시장 다변화를 위한 베트남 대상 관광마케팅에 많은 노력을 들일 계획이다”고 밝혔다. 한편, 최근 3년 동안 한국을 찾은 베트남 관광객은 100만 명을 넘는 상태며, 연도별로는 지난 2016년 25만 1,402명, 2017년 32만 4,740명, 지난해 45만 7,818명으로 큰 폭의 증가추세를 보이고 있다.</p>