Regular paper | Journal of the KIECS. pp. 1231-1236, vol. 19, no. 6, Dec. 31. 2024, t. 128, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2024.19.6.1231

웹소설 TTS를 위한 KoBERT 기반 문장 유형 판별 시스템

민태훈* · 박승민**

KoBERT-Based Sentence Type Classification System for Web Novel TTS

Tae-Hoon Min* · Seung-Min Park**

요 약

글을 읽는 데 장애가 있거나 직접 글을 읽는 시간이 부족하거나 모바일로 웹소설을 감상하는 것이 불편한 이용자들에게 TTS는 웹소설을 소비할 수 있는 유일한 통로이며 기업이 꼭 확보해야 하는 소비층이다. 그러나 현재 웹소설 TTS는 텍스트를 단일 목소리로 읽는 것이 전부이며 이는 이용자의 몰입을 어렵게 하고 진입장벽으로 작용한다. 따라서 생동감 있게 글을 읽어 이용자들을 몰입시키는 웹소설 TTS가 필요하다. 본 논문은 이전보다 생동감 있는 웹소설 TTS를 만들기 위해 글의 각 문장에 어떤 소리가 필요한지에 따라 그 유형을 해설문, 대사문, 의성문으로 정의한 뒤 이를 판별하기 위해 KoBERT 모델을 기반으로 한 문장 유형 판별 시스템을 제안한다. 해당 시스템은 결과적으로 높은 정확도를 보이며 문장을 분류했고 이런 식으로 나온 데이터는 분류된 문장 내에서도 조금 더 심층적인 판별을 진행하거나 개체명 인식이나 발화자 인식, 문장 사이의 관계 인식과 결합하여 더 높은 성능의 모델 연구에 도움이 된다.

ABSTRACT

For users who have difficulty reading text, don't have enough time to read directly, or find it inconvenient to enjoy web novels on mobile devices, TTS is the only way to consume web novels, making them an essential audience that companies must cater to. However, current web novel TTS systems merely read text in a monotonous single voice, which hinders user immersion and acts as a barrier to entry. Therefore, a more dynamic and immersive TTS system is needed. This paper proposes a sentence-type classification system based on the KoBERT model to create a more engaging web novel TTS. The system defines sentence types based on the commentary sentence, line sentence, and onomatopoeia sentence, based on the type of sound required. The system demonstrated high accuracy in classifying sentences, and the resulting data can be further utilized for more in-depth classification within the categorized sentences. Additionally, this data can be combined with techniques such as named entity recognition, speaker identification, and inter-sentence relationship detection to contribute to the development of more advanced models.

키워드

Text-to-Speech, Sentence Segmentation, KoBERT, NLP
Text-to-Speech, 문장 분류, KoBERT, 자연어 처리

* 동서대학교 연구원(20201634@dongseo.ac.kr)
** 교신저자 : 동서대학교 소프트웨어학과
· 접 수 일 : 2024. 10. 06
· 수정완료일 : 2024. 11. 08
· 게재확정일 : 2024. 12. 12
· Received : Oct. 06, 2024, Revised : Nov. 08, 2024, Accepted : Dec. 12, 2024
· Corresponding Author : Seung-Min Park
Dept. Software, Dongseo University
Email : sminpark@dongseo.ac.kr

Ⅰ. 서 론

여러 사용자가 쉽게 접근할 수 있도록 텍스트를 실제 사람 목소리로 변환하는 기술을 TTS라고 하며, 오늘날 이 기술은 음악, 서비스, 교육 등 다양한 분야의 텍스트 기반 콘텐츠를 활용하는 데 널리 사용되고 있다.

웹소설의 경우에도 네이버 시리즈, 카카오페이지 등의 대형 웹소설 플랫폼에서 TTS가 사용되는 데 글을 읽는 데 장애가 존재하거나 바쁜 시간 속에서 직접 글을 읽을 시간이 부족한 사람들을 위해 대신 글을 읽어주는 TTS는 손쉽게 웹소설 콘텐츠를 소비할 수 있는 통로가 되었다.

또한, 한국출판문화산업진흥원에 따르면 최근 웹소설을 사용하지 않는 이유 중 40%는 모바일이나 테블릿으로 웹소설을 감상하는 것이 불편하다고 나오는데 TTS를 활용하면 이런 소비층도 쉽게 웹소설에 접근하여 웹소설 시장의 성장을 촉진 시킬 수 있다.

이런 성장 가능성을 가진 TTS지만 현재 시장에 나온 기술은 단순히 텍스트를 읽는 수준에 그치고 있다. 사람처럼 자연스럽게 읽는 것은 가능하나 그 목소리는 단조로우며 문장 사이의 구분도 없이 단일 목소리로 읽는 것이 전부이다. 이는 TTS를 사용하는 이용자가 글에 몰입할 때 방해가 되고 신규 이용자에게 높은 진입장벽으로 작용한다.

사용자들이 글의 몰입하는 데 있어 소리가 얼마나 중요한지는 오디오북을 통해 알 수 있다. 오디오북 서비스 플랫폼 중 하나인 윌라 오디오북에 따르면, 특히 소설 오디오북의 경우 어떤 성우가 글을 읽고 연기하느냐에 따라 평가가 크게 달라질 정도로 오디오북에서 소리는 매우 중요한 요소이다.

즉 TTS 기술의 발전은 다양한 소비자층의 확보와 기존 콘텐츠의 새로운 소비 형태를 추가한다는 측면에서 웹소설 시장의 확장에 도움이 되는 것은 명백하다. 오디오북처럼 생동감 있게 웹소설을 읽을수록 이용자들은 콘텐츠에 깊게 몰입할 것이며 이는 이용자를 시장에 붙잡는 또 하나의 방법이 될 것이다.

그러나 웹소설과 같은 텍스트 매체를 생동감 있게 읽어주는 TTS 기술에 대한 연구는 상대적으로 부족한 편이며, 그중 국내 연구는 주로 대화체에서 발화자를 식별하는 방향으로 진행되었다. 초기 연구에서는 규칙 기반 방법론을 활용해 형태소 분석을 통해 문장에서 발화자의 후보 자질을 확인하고 이를 바탕으로 발화자를 찾아냈다[1, 2]. 이후 트랜스포머(Transformer) 개념의 도입으로, BERT 모델을 이용해 문맥을 분석하여 각 문장의 발화자를 인식하는 방식으로 발전했다[3, 4]. 그러나 지금까지의 연구는 주로 사람 간 대화체 문장에만 초점을 맞춰왔기 때문에, 다양한 문장 형식을 포함한 웹소설과 같은 텍스트에 적용하기에는 한계가 있다.

본 논문에서는 웹소설의 모든 문장에 TTS 기술을 적용하여 오디오북처럼 생동감 있는 웹소설 TTS를 구현하는 것을 목표로 한다. 이를 위해 각 문장이 출력하는 소리에 따라 문장의 형식을 새롭게 정의하고, 이를 기준으로 문장을 분류하는 방법을 제안한다.

Ⅱ. 소리 유형에 따른 문장 분석

웹소설 TTS를 위하여 먼저 웹소설의 문장마다 어떤 소리가 필요한지 알아내고 그 소리를 각 유형으로 묶을 필요가 있다. 이 경우 각 문장은 크게 해설문, 대사문, 의성문 총 3종류로 구분할 수 있다.

2.1 해설문 (Commentary Sentence)

해설문은 이야기를 설명하거나 묘사하는 문장으로 크게 두 가지 역할을 가진다. 첫 번째는 이야기의 배경, 사건의 진행 상황, 인물의 행동이나 감정, 심리를 독자에게 전달하는 정보 전달의 역할이며 두 번째는 대사와 대사 사이를 자연스럽게 연결하거나 톤의 높낮이를 통해 특정 분위기를 조성하여 이야기의 흐름을 유지하는 역할이다.

그는 천천히 손을 들어 문을 열었다, 며칠이 지나도 비는 그치지 않는다, 앞선 두 문장처럼 해설문은 작중에는 등장하지 않는 중립적인 존재가 내는 소리이기에 기본적으로 객관적이며 차분한 톤으로 글을 읽으나 상황에 따라 강약을 조절하는 방식으로 긴장감을 표현할 때 쓸 수 있다.

글을 해설하는 주체가 주인공인 1인칭 서술의 경우도 있으나 본 논문은 문장 분류를 원활히 진행하기 위해 고유의 문장 형식 특징이 뚜렷하고 웹소설에서

가장 많이 사용되는 형식은 3인칭 서술만을 고려하기로 한다.

2.2 대사문(line sentence)

대사문은 등장인물이 직접적으로 말하는 문장으로 크게 두 가지 역할을 가진다. 첫 번째는 인물 간의 대화를 통해 인물의 성격, 감정, 관계, 동기를 드러내 자연스럽게 독자에게 정보를 전달하여 이야기를 전개하는 역할이며 두 번째는 긴박하거나 차분한 감정적인 문장을 써 이야기에 생동감을 제공하여 독자의 몰입을 유도하는 역할이다.

'뭐지?', "내가 왜 그랬을까.", 앞의 두 문장처럼 의문이나 후회같이 해당 등장인물의 감정이 그대로 반영되는 소리가 쓰이며 작중에 등장하는 인물마다 고유한 목소리를 가지기 때문에 이를 구분하는 것도 중요하다.

작중에 등장하는 모든 인물의 소리를 따로 판별하는 것은 다른 기술과 결합할 필요가 있으므로 본 논문에서는 다루지 않으며 인물의 대사에는 대화와 독백으로 구분되기도 하나 결국 같은 등장인물이 말하는 것이므로 이 또한 구분하지 않는다.

2.3 의성문(Onomatopoeia Sentence)

의성문은 실제 사물의 소리를 모방해서 사용하는 문장으로 상황을 구체적으로 묘사하여 이야기에 생동감을 불어넣어 독자들의 몰입감을 높이는 역할을 한다.

파도 소리의 '철썩', 고양이 소리의 '야옹'처럼 의성문은 자연, 동물, 기계, 사람의 행동 소리 등 다양한 사물의 소리를 전달하기 때문에 그에 맞는 다양한 소리를 준비하고 구분할 필요가 있다.

수많은 의성문을 일일이 분류하기에는 데이터가 부족하므로 본 논문에서 의성어로 이루어진 문장은 전부 하나의 의성문으로 분류한다.

III. KoBERT 모델

본 논문에서는 동화나 소설의 문장을 여러 개의 카테고리 중 하나로 분류하는 다중 분류 모델을 사용한다. 따라서, 각 카테고리에 맞는 키워드를 찾기 위해 문장의 문맥을 정확하게 이해하는 모델을 사용할 필요가 있다.

이와 관련된 대표적인 모델로는 BERT와 ChatGPT가 있으나 ChatGPT의 경우 대화형 커뮤니케이션과 정보 교환을 위한 모델로 설계되어 다양한 기능을 수행하는 만큼 문장 분류와 같은 특정 작업에 맞춤형으로 훈련된 모델은 아닌 점과 다양한 국가에 서비스되는 다국어 모델이기에 한국어에 특화된 BERT보다는 한국어의 문법 구조나 어휘적 특성을 반영하지 못할 수도 있다고 판단되어 본 논문에서는 사용되지 않는다[5-6].

그렇게 선택된 BERT(: Bidirectional Encoder Representations from Transformers)는 딥러닝 기법을 활용한 기술로 2018년 구글에서 개발한 사전 학습(pre-trained) 언어 모델이다.

기존 자연어 처리 모델은 문장을 왼쪽에서 오른쪽으로 읽는 단방향 방식으로 사용되었으나 BERT는 양방향 Transformer 인코더를 사용하여 문장의 앞뒤 문맥을 모두 고려하여 단어들을 분석하고 글의 문맥을 이해한다. 이를 통해 다양한 자연어 처리 태스크에 뛰어난 성능을 보이는 데 특히 어순이 자유롭고 문맥에 따라 단어의 의미가 크게 변하는 한국어 문장 분류에 매우 적합한 특징이라 할 수 있다.

이 점은 본 논문 같은 웹소설 속 문장을 분류하는 점에서 이점이 큰데 입력 문장의 문맥을 고려하여 문장을 이해하기에 소설에서 자주 나타나는 긴 문장에 받는 영향을 최소로 한 상태에서 문장의 어떤 부분이 중요한지를 단어 간의 상관관계를 학습한 뒤 키워드를 중심으로 카테고리를 분류하는 데 높은 성능을 보인다[7].

그리고 BERT는 Self-Attention 메커니즘을 활용하여 각 문장의 단어를 벡터로 변환한 뒤 그 의미를 표현하기에 문장 전체의 의미를 보존함과 더불어 각 문장의 유사성을 수치화하여 이러한 벡터 표현을 바탕으로 다양한 분류 모델을 통해 문장을 분류할 수 있다.

또한, 대규모 코퍼스로부터 사전 학습을 통해 전이학습된 모델이기에 입력데이터와 분야가 겹치는 데이터를 미리 학습한 모델로 Fine-tuning을 진행할 경우 비교적 적은 수의 데이터라도 특정 범주에 문장을 분류하는 데 우수한 성능과 빠른 속도를 보여주는 장점이 존재하며 이는 직접 데이터셋을 만들 때 시간과

비용을 크게 줄여주기에 매우 큰 이점이라 할 수 있다[8].

본 논문에서 사용하는 데이터셋은 한국어만을 사용하기 때문에 이에 적합한 모델인 SKT Brain에서 한국어 텍스트 데이터를 기반으로 사전 학습하여 개발한 KoBERT 모델을 사용한다. KoBERT 모델은 대량의 한국어 소설을 학습하여 문학 텍스트의 감정도 분류할 정도로 문맥에 대한 이해가 탁월하기에 타 모델 대비 높은 성능을 기대할 수 있다[9].

Ⅳ. KoBERT 기반 문장 분류 방법

4.1 데이터셋 준비

인어 공주, 성냥팔이 소녀 등 저작권이 만료된 소설 혹은 동화에서 1,112개의 문장을 추출하여 문장 분석에서 해설문, 대사문, 의성문으로 정의된 3가지 문장 유형을 클래스로 레이블링하였고 표 1과 같은 데이터셋을 자체 제작했다.

표 1. 모델링에 필요한 데이터 셋
Table 1. Data set required for modeling

Form	Sentence
Commentary Sentence	said the boy.
line Sentence	Are you okay?
Onomatopoeia Sentence	splash!

4.2 모델 학습 및 성능 평가 방법

자체 제작된 데이터셋을 라벨링하여 전처리를 진행하고 데이터를 학습시키기 위한 Train set과 성능평가를 위한 Test set으로 나누었다.

전처리된 데이터를 사용하여 KoBERT 모델에서 요구하는 입력 형태로 변환시키고 파라미터를 설정하여 해설문, 대사문, 의성문의 3가지 클래스에 따라 문장을 분류하는 모델을 구축한 뒤 학습을 진행한다. 전체 모델 학습 과정은 그림 1과 같다.

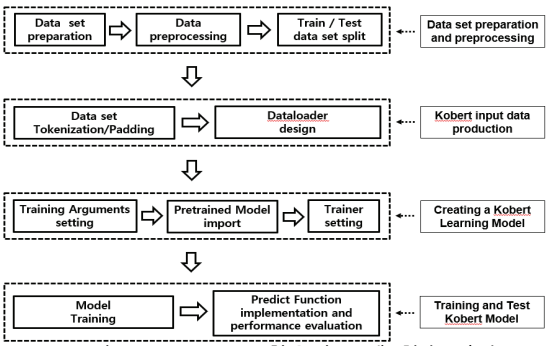

그림 1. KoBERT를 활용한 모델 학습 과정
Fig. 1 Model training process using KoBERT

학습이 완료된 모델은 앞서 만들어 놓은 Test set을 활용하여 성능평가를 진행한다. 성능평가는 다중분류 모델에서 주로 사용되는 정확도 검사(Accuracy, ACU)를 사용한다[10].

$$Accuracy: \frac{TP+TN}{TP+TN+FP+FN} \quad \cdots (1)$$

True Positive (TP), True Negative (TN),
Fasle Positive (FP), False Negative (FN)

4.3 결과

계산식 (1)의 정확도 검사 결과 아래 그림과 같은 1.0 즉, 100%에 근접하는 굉장히 높은 정확도를 보였으며. 시뮬레이션에서도 그림 2와 같이 임의로 넣은 문장을 잘 분류하였다.

```
1. Please enter a sentence: "Who are you?"
>>> The input is a line Sentence.

2. Please enter a sentence: Bang!!!
>>> The input is an Onomatopoeia Sentence.

3. Please enter a sentence: He was surprised.
>>> The input is a Commentary Sentence.

4. Please enter a sentence: 'What is this?'
>>> The input is a line Sentence.

5. Please enter a sentence: She then left her beloved and headed to another place.
>>> The input is a Commentary Sentence.
```

그림 2. 정확도 검사 결과와 모델 시뮬레이션
Fig. 2 Model simulation

Ⅴ. 결 론

본 논문에서는 오디오북처럼 생동감 있게 글을 읽는 웹소설 TTS를 목표로 하였다. 해당 TTS를 위하여 먼저 글의 문장이 출력할 소리에 따라 문장 유형을 해설문, 대사문, 의성문으로 정의한 다음 이를 기준으로 한국어 문장 분류에 적합한 KoBERT 모델을 기반으로 문장 유형을 판별하는 방법을 실행하였고 100%에 근접하는 높은 정확도를 달성하였다.

5.1 시사점

해당 결과는 기존의 연구가 대화체 위주로 분류된 것과는 달리 소설 부류의 텍스트 매체 글 전문을 대상으로 한 문장 분류의 가능성을 보여주며 이는 이전보다 적극적인 소설 텍스트 데이터 수집을 기대할 수 있다.

5.2 한계점 및 향후 연구 과제

현존하는 대부분의 웹소설은 저작권으로 인해 데이터 학습을 시킬 수 없어 비슷한 매체인 고전 소설이나 동화로 대체했다는 점에서 웹소설 TTS라는 목표에서는 조금 거리가 있다는 한계점이 존재한다.

향후 연구로는 해설문의 1인칭과 3인칭, 대사문의 대화와 독백 혹은 여러 인물로 나누어지는 경우, 그리고 다양한 의성어의 구분 등 본 논문에서 정의한 유형보다 조금 더 심층적으로 들어간 판별에 도전하거나 문장 사이의 관계 인식, 발화자 인식, 개체명 인식 연구와 같은 다른 기술과 결합해 기존보다 높은 성능의 모델을 목표로 한다.[11, 12, 13].

감사의 글

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 지원을 받아 수행되었음 (2019-0-01817)

References

[1] H. Min, S. Kim, and J. Park, “Automatic Speaker Identification in Fairytales towards Robot Storytalling.” *The Korean Society for Cognitive Science, Busan, South Korea*, Oct. 2012, pp. 77-83.

[2] J. Son and S. Kim, “Patent document classification automation system based on association rules,” *Korean Institute Of Industrial Engineers, Yeosu, South Korea*, May 2013, pp. 575-586.

[3] S. Hwang and D. Kim, “BERT-based Classification Model for Korean Documents,” *The J. of Society for e-Business Studies*, vol. 25, no. 1, Feb. 2020, pp. 203-214. https://doi.org/10.7838/jsebs.2020.25.1.203

[4] G. Gang and O. Kwon, “Automatic Speech Style Recognition Through Sentence Sequencing for Speaker Recognition in Bilateral Dialogue Situations,” *J. of Intelligence and Information Systems*, vol. 27, no. 2, June 2021, pp. 17-32. https://doi.org/10.13088/jiis.2021.27.2.017

[5] Y. Oh, “Generative AI Jeonse Fraud Prevention System,” *J. of the Korea Institute of Electronic Communication Sciences*, vol. 19, no. 1, Feb. 2024, pp. 173-180. https://doi.org/10.13067/JKIECS.2024.19.1.173

[6] C. Lee and H. Beak, “A study on the need for AI Literacy according to the development of Artificial Intelligence chatbot,” *J. of the Korea Institute of Electronic Communication Sciences*, vol. 18, no. 03, June 2023, pp. 421-426. https://doi.org/10.13067/JKIECS.2023.18.3.421

[7] J. Min, S. Na, and J. Shin, and Y. Kim, “BERT for Transition-based Korean morphological analysis and POS tagging.” *Korean Institute of ingormation Scientists and Engineers,* Pyeongchang, South Korea, Dec. 2019, pp. 401-403.

[8] J. Lee, “Comparison of Sentiment Classification Performance of for RNN and Transformer-Based Models on Korean Reviews,” *J. of the Korea Institute of Electronic Communication Sciences*, vol. 18, no. 4, Aug. 2023, pp. 693-700. https://doi.org/10.13067/JKIECS.2023.18.4.693

[9] G. Park and Y. Jeong, “Korean Daily Conversation Topics Classification Using

KoBERT." *Korean Institute of ingormation Scientists and Engineers*, Jeju, South Korea, June 2021, pp. 1735-1737.
[10] E. You, G. Choi, and S. Kim, "Study on Extraction of Keywords Using TF-IDF and Text Structure of Novels," *J. of the Korean Society of Computer and Information*, vol. 20, no. 2, Feb. 2015, pp. 121-129. G704-001619.2015.20.2.005
[11] C. Nam and K. Jang "Predicate Recognition Method using BiLSTM Model and Morpheme Features", *J. of the Korea institute of Information and Communication Engineering*, vol. 26, no. 1, Jan. 2024, pp. 24-29. https://doi.org/10.6109/jkiice.2022.26.1.24
[12] S. Jun, E. Yun, H. Lee, M. Kang, J. Kim, and G. Yoo, "Development of a Candidate Scoring Network based Speaker Recognition System Utilizing Textual Context Information," *J. of Korean Institute of Information Technology*, vol. 22, no. 5, May 2024, pp. 151-163. http://dx.doi.org/10.14801/jkiit.2024.22.5.151
[13] J. Lee, "Performance Comparison and Error Analysis of Korean Bio-medical Named Entity Recognition," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 19, no. 4, Aug. 2024, pp. 701-708. https://doi.org/10.13067/JKIECS.2024.19.4.701

저자 소개

민태훈(Tae-Hoon Min)

2024년 동서대학교 소프트웨어학과 졸업(공학사)

2020년 ~현재 동서대 소프트웨어학과 학부생
※ 관심분야 : 기계학습, 딥러닝, 패턴인식, 자연어처리

박승민(Seung-Min Park)

2010년 중앙대학교 전자전기공학부 졸업(공학사)
2019년 중앙대학교 대학원 전자전기공학과 석박사통합과정 졸업(공학박사)
2019년 ~현재 동서대학교 소프트웨어학과 조교수
2022년 ~현재 동서대학교 AI+X융합연구센터장
2021년 ~현재 산업인공지능 표준화포럼 운영위원
※ 관심분야 : 인공생명, 패턴인식, 뇌-컴퓨터 인터페이스, 기계학습