

텍스트 맥락 정보를 활용한 CSN기반 발화자 인식 시스템 개발

전승호*, 윤은**, 이하원***, 강민혁****, 김정원*****, 유길상*****

Development of a Candidate Scoring Network based Speaker Recognition System Utilizing Textual Context Information

Seungho Jun*, Eun Yun**, Hawon Lee***, Minhyeok Kang****, Jeungwon Kim*****,
and Gilsang Yoo*****

This research was supported by Basic Science Research Program through the
National Research Foundation of Korea(NRF) funded by the Ministry of Education(RS-2023-00246191)

요약

최근 다양한 디지털 플랫폼과 애플리케이션에서 발화자 인식 기술의 중요성이 부각되고 있다. 특히 OTT(Over The Top) 서비스, 오디오 북, 음성 인식 서비스 등에서 이 기술에 관한 관심이 높아지고 있다. 그러나 기존의 발화자 인식 기술은 음성 인식 영역에만 치중되어 있어서, 소설과 같이 음성 정보가 없는 분야에서 문맥을 기반으로 발화자를 특정하기 어렵다는 문제가 있다. 본 연구에서는 이러한 문제를 해결하기 위해 소설 속 인용문과 텍스트 맥락 정보를 활용하여 CSN(Candidate Scoring Network)기반의 발화자 인식 기법을 제안하였다. 시뮬레이션 결과, 테스트 데이터의 예측 정확도는 83.2%로 우수한 결과를 보였다. 본 연구에서 개발한 시스템은 소설 텍스트에서 발화자를 자동으로 식별하는 기능을 제공함으로써, 데이터 라벨링 작업이 필요했던 문제점을 해결하고 음성 기반 발화자 인식 서비스를 향상할 수 있을 것으로 기대한다.

Abstract

Recently, the importance of speaker recognition technology has been highlighted across various digital platforms and applications. This interest is especially growing in Over The Top(OTT) services, audiobooks, and voice recognition services. However, existing speaker recognition technologies are primarily focused on the area of voice recognition, posing a challenge in identifying speakers based solely on text, such as in novels, where no vocal information is available. To address this issue, this study proposes a speaker recognition technique based on the Candidate Scoring Network(CSN), utilizing quoted speech and contextual information from novels. Simulation results show that the system achieved an excellent prediction accuracy of 83.2% on test data. The system developed in this study is expected to resolve the issues requiring manual data labeling and enhance voice-based speaker recognition services by automatically identifying speakers in novel texts.

Keywords

multi-party conversation, speaker recognition, named entity recognition, candidate scoring network

- * 아주대학교 정치외교학과 학사과정
- ORCID: <https://orcid.org/0009-0008-7191-4770>
** 고려대학교 국어국문학과 학사과정
- ORCID: <https://orcid.org/0009-0003-9650-1888>
*** 고려대학교 환경생태공학부 학사과정
- ORCID: <https://orcid.org/0009-0009-1843-1451>
**** 고려대학교 산업경영공학부 학사과정
- ORCID: <https://orcid.org/0009-0009-0985-5850>
***** 숙명여자대학교 소프트웨어학부 학사과정
- ORCID: <https://orcid.org/0009-0009-1035-1369>

- ***** 고려대학교 정보대학 정보창의교육연구소 교수(교신저자)
- ORCID: <https://orcid.org/0009-0002-1085-5355>

- Received: Mar. 21, 2024, Revised: May 09, 2024, Accepted: May 12, 2024
· Corresponding Author: Gilsang Yoo
Creative Informatics and Computing Institute, Korea University, 145
Anam-ro, Seongbuk-gu, Seoul, Korea
Tel.: +82-2-3290-1674, Email: ksyoo@korea.ac.kr

1. 서 론

최근 음성 기반의 발화자 인식 서비스는 다양한 분야에서 높은 수요를 보이고 있다[1]. 넷플릭스, 티빙 등 여러 OTT(Over The Top) 서비스에서 제공하는 한국어 자막은 소비자들에게 호평받았으며, 이에 넷플릭스는 2023년 8월, 발화자와 대사를 표기하는 예능 전용 자막을 도입하였다[2]. 음성 인식 프로그램인 네이버 클로바노트(NAVER Clova Note)는 발화자를 구분한 음성 기록 기술을 선보였으며, 영상 편집 프로그램인 프리미어는 발화자를 구분하는 자동 자막 서비스를 출시한 바 있다. 이러한 기술들은 모두 기존에 발화자를 구분해 둔 대본이나 음성 인식에 기반을 두어 발화자를 구분하는 방식이다.

텍스트 내에서 발화자를 구분하려면 발화자를 미리 구분해 놓은 문서를 통한 사전 학습이 이루어져야 한다. 또한 이를 구축하기에는 상당한 데이터 라벨링 시간과 비용이 소요된다[3]. 이러한 어려움으로 인해 발화자 인식의 많은 연구가 음성 기반에 한정되어 연구되어왔다. 그러나 최근에는 오디오북 시장의 성장과 꾸준한 확대로 인해 텍스트 기반의 발화자 인식 기술에 대한 관심이 높아지고 있다. 미국 시장조사기관 '그랜드뷰리서치'에 따르면 오디오북의 전세계 시장 규모가 2021년 기준 42억 달러(약 6조원)에 달하였으며 2024년부터 2030년까지 연평균 성장률을 3.8%로 전망하고 있다[4].

2차 창작과 그 작품의 방영이 빠르게 일어나고 있는 웹소설 역시 발화자 인식 기술에 관한 연구가 필요한 실정이다. 2023년 문화체육관광부 실태조사 보고서에 따르면 대한민국에서 웹소설의 시장적 가치는 1조를 넘어 가파르게 성장하고 있는 것으로 보고되었다[5]. 국내에서 유명한 '재벌집 막내아들', '김비서가 왜 그럴까' 등의 드라마는 모두 웹소설을 원작으로 하고 있다. 이러한 사례들은 웹소설은 재창작이 활발하게 이루어지는 분야이며, 그 규모 역시 매우 높은 것으로 보고되고 있다. 드라마, 영화 등 2차 창작은 기존 웹소설을 재가공하는 것으로 시작된다. 이때 장면 구분을 위해 소설 속 시·공간적 배경을 구분하고, 인용문마다 발화자를 기재한다. 수동으로 이루어지는 이 작업은 많은 시간과 비용이 필요

하다. 웹소설에서 발화자 인식 과정을 AI로 자동화할 수 있다면 2차 창작 과정을 단순화할 수 있다.

오디오 북, 웹소설 등 산업 부문에서 텍스트 기반 발화자 인식 기술이 산업적으로 가치를 지님에도 불구하고 발화자 인식 연구의 대부분은 음성 인식 영역에 한정되어 왔다. 앞뒤 문맥을 기반으로 발화자를 인식하는 과정은 매우 복잡하며, 특히 3인 이상 인물이 번갈아 발화하는 대화문에서 발화자를 특정하기는 더욱 어렵다. 또한, 한국어는 주어의 생략이 잦아 발화문 앞뒤를 살피는 것만으로 발화자를 인식하기 어렵다. 이러한 이유로 한국어 텍스트 속 발화자 인식 연구는 더욱 부진하다.

음성 데이터에서의 발화자 인식 기술도 몇 가지 한계가 존재한다. 우선, 음성 기반 발화자 인식은 여러 제약 사항으로 정확도가 제한된다[6]. 가령 음성에 잡음이 포함될 경우, 상대적으로 먼 거리에 있는 발화자는 인식하지 못하는 경우가 있다. 또한, 발화가 겹치거나 여러 인물의 음역 주파수가 유사할 경우 각 발화자의 발화를 유의하게 구분해 내지 못한다. 특히 상담 기록, 소설 등 음성이 부재하며 텍스트 데이터만 있는 경우, 발화자를 전혀 인식할 수 없다. 이처럼 텍스트 기반의 발화자 인식은 음성 기반의 발화자 인식 기술을 보완하는 데도 중요한 역할을 할 수 있다.

따라서 본 논문은 텍스트 기반의 발화자 인식 기술을 제안하고 이를 통해 음성에만 국한되었던 기존 발화자 인식 서비스를 개선하고자 포털 웹소설 데이터를 활용하여 발화자 학습 모델을 구축하였다. 웹소설은 2차 창작이 활발한 분야로, 발화자 인식이 필수적이다. 발화자 인식 과정은 많은 시간과 노력이 들어가는 과정이며, 2차 창작 과정에서 데이터 라벨링 근로자의 수작업 비중을 많이 차지한다. 따라서 본 논문의 모델을 활용하여 발화자를 빠르고 효율적으로 찾을 수 있도록 하였다. 발화자는 RoBERTa(Robustly Optimized BERT Pretraining Approach) 기반의 개체명 인식 기법으로 추출되며, 인용문마다 발화자 후보군을 찾고, 주변 맥락을 고려하여 인용문에 대해 가장 높은 확률을 가진 발화자를 선정하였다. 본 연구에서 활용된 데이터는 네이버 웹소설 데이터로 선정하고 수집 및 구축하였다.

구축된 사전 데이터는 작품별로 다양한 특성이 있다. 구어체가 많은 작품이 있고, 이와 반대로 문어체가 대다수인 작품이 있다. 작가에 따라서 시점이 달라지기도 하고, 작품의 배경에 따라서 다양한 단어들 나온다. 이와 같은 특성을 고려할 때, 본 논문이 제시하는 모델은 웹소설 뿐 아니라 다양한 종류의 텍스트에서도 적용될 수 있을 것으로 기대한다.

본 논문의 구성은 다음과 같다. 제2장은 관련 연구로, BERT(Bidirectional Encoder Representations from Transformers) 언어 모델과 개체명 인식(NER, Named Entity Recognition), 다자 간 대화에서의 발화자 분석 기술에 대해 기술하였다. 제3장에서는 본 논문에서 제안하는 CSN(Candidate Scoring Network) 기반 발화자 예측 모델의 전체 흐름도를 제시하고, 각 과정에 대해 기술하였다. 제4장에서는 본 모델의 성능을 분석하고 제5장에서는 기대효과 및 향후 연구에 관하여 기술하였다.

II. 관련 연구

2.1 다자간 대화 연구 동향

다자간 대화(MPC, Multi-Party Conversation)는 대화 시스템 분야 중 하나로, 주어진 다자간 대화에서 '누가' '누구에게' '무엇을' 말하는지를 분석하는 분야이다[7]-[9]. MPC의 선행 연구를 분석한 Jia-Chen Gu의 연구에서는 MPC의 영역을 크게 대화 맥락 모델링과 대화 요소 모델링의 영역으로 구분하였다[10]. 대화 맥락은 담화 구분과 정보의 흐름 등을 나타낼 방식을 연구하는 영역이며, 대화 요소 모델링은 발화자(Speaker)와 발화 대상자(Addressee), 그리고 발화(Utterance)를 모델링하는 영역이다. 선형 모델로 나타낼 수 없는 다자 간 대화 속 복잡성은 딥러닝의 발전과 함께 본격적인 연구가 시작되었다.

B. Lee and Y. S. Choi[11]은 이후로 다자 간 대화를 시각적 흐름(Graphical flow)으로 나타내려는 시도가 있었으며, Z. Shi and M. Huang[12]은 예측 기반 신경망 모델을 이용해 담화를 나타냈다. Wang은 별도의 예측 구조가 없는 엣지(Edge) 중심의 그래프 신경망 모델로 각 발화 쌍 간 정보를 단계별로 업데이트하는 모델을 제시해 2021년 SOTA(State

Of The Art)를 차지하였다[13].

발화자 예측 모델은 크게 발화자를 알고 있을 때와 모르고 있을 때의 두 경우로 나뉘어 모델이 개발됐으며, 전자의 SOTA 모델은 2021년 출판된 논문의 BERT[14] 기반 MPC-BERT[15] 모델이며, 후자의 SOTA 모델은 2019년 개발된 모델로 word2vec과 SVM(Support Vector Machine)을 사용하여 단어 수준의 임베딩을 만든 후 다수의 LSTM(Long Short-Term Memory) 층을 거쳐 문맥을 이해하는 AC-SVM 모델[16]이다. 이 중 MPC-BERT는 인풋에 대해 토큰 임베딩, 세그먼트 임베딩, 위치 임베딩 외 'speaker embedding'을 추가해 훈련한다. 그 결과물로 얻은 임베딩에 대해 비선형 함수 및 정규화를 거친 뒤, 'Reply-to Utterance Recognition', 'Identical Speaker Searching'과 'Pointer Consistency Distinction'에 대한 예측을 수행한다. 성능 평가는 발화문마다 발화자와 발화 대상자의 주석이 달린 두 개의 IRC 벤치마크 데이터셋에 대해 발화 대상자 인식(Addressee recognition), 발화자 인식(Speaker identification), 그리고 특정 발화에 대한 대답 선택(Response selection)을 얼마나 잘 예측하였는지에 대한 평가로 이루어졌다. MPC-BERT의 제안 모델은 텍스트에서 발화자를 인식하는 우리의 목적에 부합하나, 대상 문장의 길이가 길어질수록 그 정확도가 급격히 하락한다는 점에서 부적격하다고 판단하였다.

또 다른 발화자 구분 모델은 중국어 소설 World of Plainness을 데이터로 하여 발화자를 예측한 CSN-SAPR(Candidate Scoring Network-Speaker Alternation Pattern Based Revision)[17] 모델이다. CSN(Candidate Scoring Network)은 특정 발화문의 발화자 이름은 발화문 주변에서 1회 이상 언급될 것이라고 가정하며, 발화문마다 주변을 스캔하여 발화자 후보군을 구하고 후보별로 점수를 계산해 최종 발화자를 특정하는 방법이다. 이러한 시도는 78.5%의 정확성을 보였다. 그러나 하나의 소설로 학습/검증/테스트 데이터셋을 모두 구성하여 평가하였기 때문에, 과적합 문제와 범용성을 확신할 수 없다. 따라서 본 연구에서는 베이스라인 모델로 이 CSN 모델을 선택하였고, 입력 형태와 학습 방법을 변경하여 한글 소설에 적용하는 동시에 60개의 소설을 학습 데이터로 활용하여 다양한 뉘앙스를 학습할 수 있게 한다.

2.2 언어 모델

자연어 처리(Natural language processing) 분야는 문맥을 고려하는 언어 모델의 개발을 목적으로 연구되었다. 단어 수준 임베딩에서 나아가 Bi-LSTM(Bidirectional Long Short-Term Memory)으로 양방향 문맥을 고려한 ELMo(Embeddings from Language Models) 모델 개발 이후 주목받은 문장 수준 임베딩 방식은 Transformer의 개발로 새로운 패러다임에 접어들었다. Transformer 모델은 RNN(Recurrent Neural Network) 방식이나 CNN(Convolutional Neural Network) 방식과 달리, 순환적 코드를 기반으로 하지 않으면서 Attention을 통해 텍스트 내 맥락을 매 순간 고려한다. BERT는 Transformer의 인코더 구조를 이용한 모델로, Auto Encoding 방식으로 mask 토큰을 예측하며 학습한다. BERT는 레이블이 존재하지 않는 텍스트 데이터에 MLM(Masked Language Model), NSP(Next Sentence Prediction) 기법을 활용한 비지도 학습을 통해 강력한 언어 표현을 생성한다[14].

klue-roberta-large는 RoBERTa 모델을 한국어 KLUE 데이터셋으로 학습시킨 BERT 모델의 일종이다. RoBERTa란 BERT의 학습 과정에서 NSP 학습을 제거하고, 동적 마스크를 사용하는 등 학습 방식에 변화를 주어 성능을 향상한 모델이다[18]. KLUE(Korean Language Understanding Evaluation) 벤치마크는 한국어 자연어 이해 벤치마크로 토픽 분류(TC, Topic Classification), 의미 유사도(STS, Semantic Textual Similarity), 자연어 추론(NLI, Natural Language Inference), 개체명 인식(NER, Named Entity Recognition), 관계 추출(RE, Relation Extraction), 의존 구문 분석(DP, Dependency Parsing), 기계 독해(MRC, Machine Reading Comprehension), 대화 상태 추적(DST, Dialogue State Tracking)의 총 8개의 태스크에 대해 평가할 수 있는 데이터셋으로 구성되어 있다[19]. 본 논문에서는 RoBERTa(klue-base_large)에 포털 웹소셜 데이터를 추가로 학습시켜 미세 조정(Fine-tuning)을 진행하였다.

2.3 개체명 인식(NER) 모델

개체명 인식은 비정형 텍스트에서 인명, 단체, 장소, 시간 등 미리 정의된 분류에 속하는 정보를 추출하는 기술로 자연어 처리의 영역 중 하나이다. 분류 방식은 NER 모델마다 다른 기준을 적용한다. 일반적으로 NER 모델이 추출 대상으로 하는 개체명은 크게 이름, 시간, 수량 표현으로 구성되며, 각 개체명은 실세계에서 유일한 개체를 표현하는 이름 혹은 이름을 대신할 수 있는 표현이다. 최근에는 딥러닝 모델 기반 개체명 인식을 주로 사용한다[20].

본 연구는 한국해양대학교에서 제공하는 NER 모델을[21] 사용하였다. 이 모델은 맥락을 고려해야 하기에 문맥 기반 방식인 BERT를 기반으로 개체명을 인식한다[22]. 사용한 모델은 총 23,964개의 문장으로부터 최장 일치율 원칙으로 하여 10 종류의 주석으로 개체명을 구분하였다. 개체의 범주는 크게 개체이름, 시간 표현, 수량 표현으로 구분하였으며, 세부 범주는 인물(PER), 기관(ORG), 지명(LOC), 기타(POH), 날짜(DAT), 시간(TIM), 기간(DUR), 통화(MNY), 비율(PNT), 기타 수량 표현(NOH)이 있다. 이 중 인명, 지명, 기관명과 날짜, 시간을 사용할 정보로 추출하였고 실제로 본 논문에서 활용된 것은 인명, 지명, 날짜, 시간이다. 인명은 텍스트 내의 발화자 후보군을 구성하는 데 사용하였다. 지명과 날짜 그리고 시간은 최종 결과물에 장면별로 장소와 시간을 나타내기 위해 사용되었다.

III. 제안한 발화자 인식 모델

제안한 텍스트 기반 발화자 인식을 위한 전체 시각화 모델링 과정은 그림 1과 같다. 먼저 웹소셜 데이터를 수집하고, NER 기반으로 장소, 시간, 그리고 네임리스트(등장인물 목록)를 추출한다. 그리고 전처리를 거쳐서 학습 단위인 인스턴스를 구성한다. 인스턴스를 모델에 넣어서 발화자, 인용문, 주변 맥락 정보를 기반으로 하여 발화자를 인식한다. 최종적으로 인식한 발화자와 앞에서 추출한 장소, 시간을 넣어서 최종 결과를 출력한다.

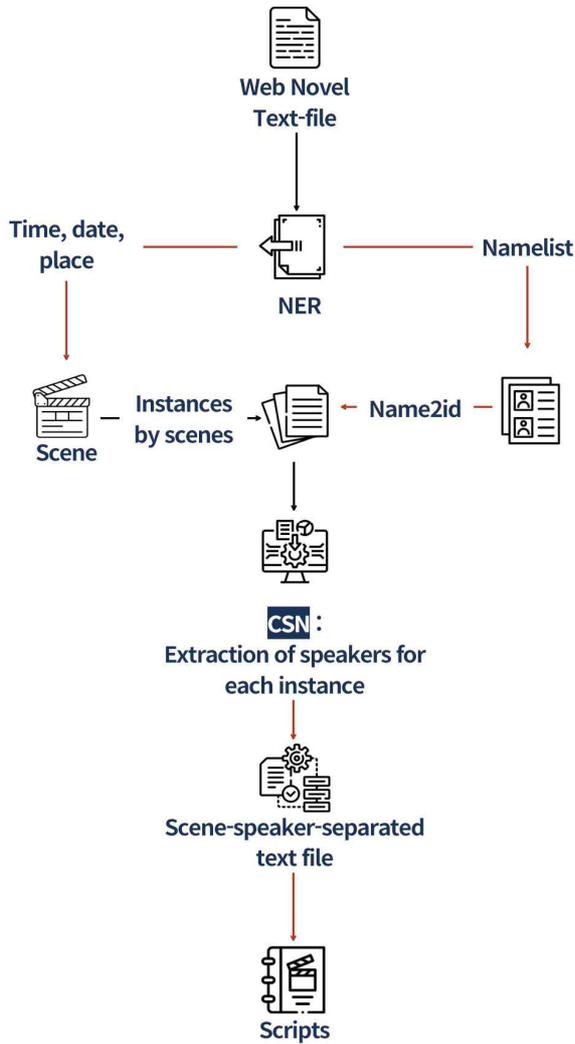


그림 1. 제안한 모델의 전체 흐름도
Fig. 1. Overall flowchart of the proposed model

3.1 데이터 수집

학습용 데이터는 다양한 장르로 구성된 포털 웹소설 총 66개의 작품을 수집하였다. 이 중, 등장인물의 수가 100명 이상이거나, 작품 특성상 발화자의 이름 언급이 적어 발화자 구분이 어려운 경우를 제외하고 63편을 선정하였다. 포털에 서비스 중인 웹소설은 크게 두 개의 장르(로맨스/로맨스 판타지)로 나눌 수 있다. 수집한 작품 중 60편의 작품은 학습 및 검증용으로 활용하였다. 또한 정확성 평가를 위한 테스트 데이터셋으로, 랜덤으로 추출한 세 작품(로맨스 판타지 두 작품과 로맨스 한 작품으로 구성)을 활용하였다. 정답 유무를 확인하기 위해 등장인물이 사진으로 라벨링 된 작품만을 활용하였다.

웹소설에는 발화자가 레이블된 작품과 그렇지 않은 작품이 있는데, 발화자 정보가 표기된 작품은 그림 2와 같이 발화문의 왼쪽에 발화자의 인물 이미지 또는 이름이 삽입되어 있다.

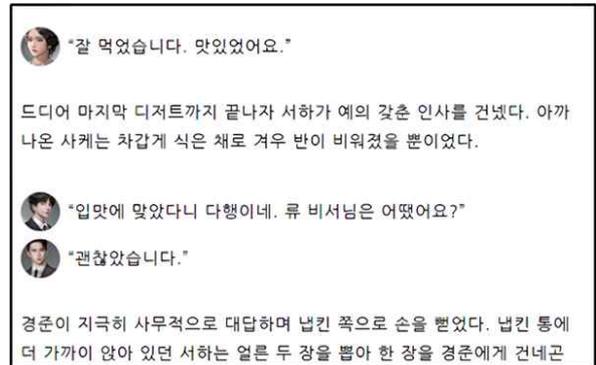


그림 2. 발화자 사진이 삽입된 소설 문장의 예
Fig. 2. Example of a novel sentence with an inserted speaker's photo

웹소설은 작품별로 다양한 특징을 가진다. 줄 바꿈, 인용문의 비율은 작품별로 다르다. 수집된 작품은 각 회차가 200~500줄로 구성되어 있다. 각 작품의 전 회차 중 인용문의 비율은 그림 3과 4와 같이 중앙값은 36%로 나타났다. 그림 3의 x축은 발화문을 전체 문장으로 나눈 값을, y축은 특정 비율에 해당하는 작품의 수를 나타낸다. 수집된 데이터셋에 포함된 소설 작품 중 절반은 발화문이 전체 문장의 36% 이하를 차지하고, 나머지 절반은 36% 이상으로 대화문의 비중이 유용하게 구성된 것을 확인할 수 있다.

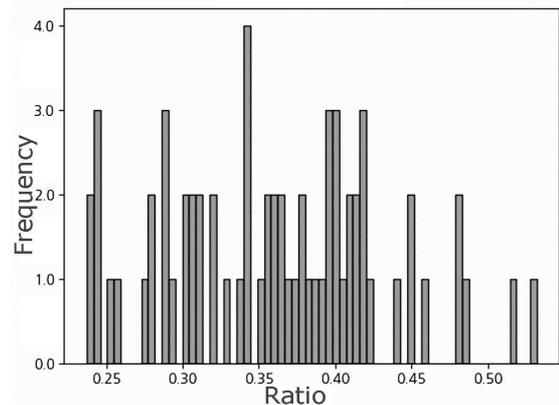


그림 3. 전체 문장에서 발화문이 차지하는 비율에 따른 작품 수 분포

Fig. 3. Plot of the distribution of the number of works as a function of the percentage of utterances in the total sentence

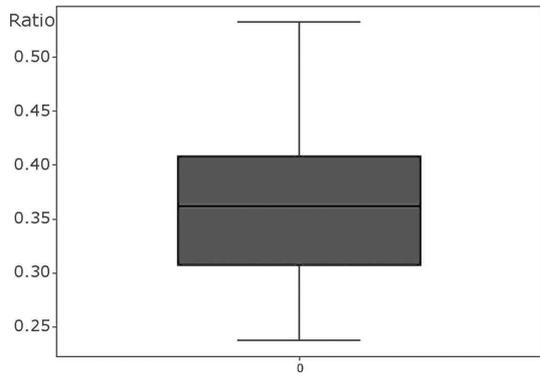


그림 4. 전체 문장에서 발화문이 차지하는 비율을 나타낸 boxplot

Fig. 4. Boxplot showing the percentage of utterances in the total sentence

```

김 대리가 폭탄 발언을 한 직후, 각자 다른 반응이 터져 나왔다.

https://novel-phinf.pstatic.net/20210827_73/novel_1630055233943tEtso_PND
https://novel-phinf.pstatic.net/20210902_237/novel_1630568469780Gnusk_PN
https://novel-phinf.pstatic.net/20210827_73/novel_1630055233943tEtso_PND
니, 김 대리 저 아저씨가 딱 그럴네.
https://novel-phinf.pstatic.net/20210902_237/novel_1630568469780Gnusk_PN
https://novel-phinf.pstatic.net/20210827_73/novel_1630055233943tEtso_PND
세대다. 으흐흐, 예진이랄 민규라니, 오하하하하!
마찬가지로 심각하게 상황을 지켜보고 있던 최한석은 배꼽을 잡고 웃어 버렸고, 유미는 이

은 우주의 공허함을 가득 담은 것만 같은 눈동자로 말하니 김 대리를 응시했다.

그런 그렇고 하루 또한 김 대리가 이런 사람인지는 몰랐다.

이게 말로만 들던 연애고자인가?

점심을 먹으며 팀원들과 나누는 말들로 그가 지금 연애를 하지 않고 있는 줄로리는 것은
    
```

그림 5. Raw 파일
Fig. 5. Raw file

```

{
  "id": "1002534",
  "epi_title": "떡볶이 좀 드셔 보셨나 (2)",
  "title": "신과 함께 요리천재",
  "epi": "100",
  "contents": [
    "...",
    "...",
    "김 대리가 폭탄 발언을 한 직후, 각자 다른 반응이 터져 나왔다.",
    "20210827_73 : -푸흐흐흐흐!",
    "20210902_237 : \"세상에나...\"",
    "20210827_73 : -이야, 요새 젊은이들이 삼포 세대다 사포 세대다 그러더니,",
    "20210902_237 : \"연애, 결혼, 출산을 포기하세요?\"",
    "20210827_73 : -아니. 연애, 결혼, 출산에 연애 세포까지 포기해서 사포 세",
    "마찬가지로 심각하게 상황을 지켜보고 있던 최한석은 배꼽을 잡고 웃어 버렸고",
    "은 우주의 공허함을 가득 담은 것만 같은 눈동자로 말하니 김 대리를 응시했",
    "그런 그렇고 하루 또한 김 대리가 이런 사람인지는 몰랐다."
  ]
}
    
```

그림 6. 소설 별 콘텐츠를 누적한 json 파일
Fig. 6. Json file accumulating content by novel

3.2 데이터 전처리

크롤링을 통해 수집된 raw file은 회차마다 별도

의 파일로 저장되어 있다. 그림 5에서와 같이 Labeled 데이터는 <링크: 대사> 형태로 저장되어 있으며 줄 바꿈이 많다. 대사 옆의 링크는 각 이미지 파일이며, 해당 링크를 활용하여 등장인물별로 고유 id를 부여한다. 그리고 각 소설의 제목과 회차의 제목, 내용을 분류하여 그림 6과 같이 json 형태의 파일로 저장한다.

3.3 등장인물 추출과 네임리스트 생성

학습 데이터에 없는 새로운 소설의 발화자도 인식할 수 있도록, 등장인물 추출과 네임리스트 생성을 자동화하기 위해 한국 해양대학교 NER 모델을 활용하였다. 소설마다 등장한 인명을 추출하고, 동일인의 이름과 별칭을 같은 id로 관리하도록 후처리하였다. 후처리 대상은 다음과 같다.

1) 성, 이름, 성 + 이름이 별도의 이름으로 인식된 경우: 인명에서 두 번째 이후 음절의 태그인 'I-PER' 이후 개체명으로 인식되지 않은 음절의 태그인 '_' 토큰이 있고 다시 'I-PER'이 있을 때 이들을 하나의 이름으로 인식하도록 처리한다. 예를 들어 '산토스 로레인'에서 '로'의 태그가 'I-PER'일 경우 '산토스 로레인'을 하나의 이름으로 인식한다. 또는 하나의 단어에 기존 디셔너리 내 이름과 완전히 동일한 이름이 있을 경우, 이들을 하나의 인물로 처리한다. 예를 들어 '김선우'가 디셔너리에 있고 '선우'가 나올 경우 혹은 그 반대의 경우에 이들을 하나의 인물로 인식하고 동일한 인물의 리스트로 관리된다.

2) 명사 뒤 조사까지 함께 저장하거나 마지막 음절을 두 번 반복해 저장하는 등의 오류: NER를 활용하여 동일 인물이 30회 이상 언급된 경우에 한해 주요 등장인물로 간주하고 후처리를 진행한다. 이때 30회라는 기준은 소설마다 다소 차이가 있을 수 있으므로 유동적인 임계값(Threshold)으로 설정한다. 이 기준에 따라 66개의 소설에서 등장인물의 언급 횟수를 추출한 후 이를 디셔너리로 정리하고, 포털 사이트에서 사용된 인물 이름과 연관된 링크를 매칭시켜 라벨링 된다. 다양한 작품을 분석한 결과, 예를 들어 '김선우우', '선우의', '선우에게'와 같은 이름이 10회 이상 언급된 것을 확인할 수 있다.

임계값을 5회에서 30회 사이로 조절하여 실험해 본 결과, 30회 이상 언급되었을 때 인물이 유의미하게 등장한다는 것을 귀납적으로 확인하였다.

3) 이름이 한 글자일 경우: 실제 인물명인 경우보다 오류인 경우가 많아 후처리 대상으로 하였다.

3.4 NER기반 발화자 인스턴스(Instance) 생성

NER에서 Namelist를 받은 후 이를 활용하여 모델의 입력 단위인 인스턴스로 데이터를 가공하였다. 인스턴스는 모델 학습 과정에서 활용된다. 각 인스턴스의 구성은 다음과 같다.

- 1) 몇 번째 인스턴스인지 나타내는 인덱스
- 2) 발화문을 중심으로 앞/뒤 각 10문장(발화문 포함 21문장)을 묶은 본문
- 3) 본문 중심에 위치한 발화문의 발화자의 식별 인자 (아이디)
- 4) 그 외 부수 정보 (소설 제목 및 장면 정보)

그림 7은 전처리를 마친 인스턴스의 예시이다. 11번째 문장인 발화문을 대상으로 인스턴스를 생성하였다. 인스턴스의 구분자(Instance index) 이후, 인용문을 중심으로 앞뒤 10줄을 하나의 인스턴스로 생성한다. 훈련 및 추론 과정에서 모델은 앞뒤의 맥락문을 활용하여 인용문의 발화자를 인식한다. 장면의 맨 앞 10줄, 맨 뒤 10줄 내의 발화문은 부족한 줄을 공백으로 패딩(Padding)하여 인스턴스를 생성했다.

Instance index: 5114 (8줄 전략) 그러는 사이 &C00&이 나타났고, 지긋한 나이의 &C01&를 보고 흘린 듯 겹두 사람은 갠스레 웃으며 대화는 나눴다. "할아버님, 돌려 깎기가 잘 안 돼요. 너무 두껍게 잘려요." - 발화자를 예측하 "괜찮다. 나는 원래 두꺼운 오이를 좋아해." &C01&은 몇 안남은 치아로 오이를 즐겁게 씹어 먹었다. (8줄 후략) Speaker: &C01& Speaker_index: 01 Name: 비서 실세 Scene: 106 Place : None Time : None

그림 7. 발화자 인스턴스 예시
Fig. 7. Speaker instance examples

NER 모델을 거쳐 발화자가 특정되면 'Speaker'로 발화자의 코드 이름이 추가된다. 만약 NER 모델에서 해당 발화자가 특정되지 않으면 '다인지정'으로

추가된다. 해당 부분은 NER로 추가하지 못한 등장인물 목록을 가리키며, 이는 사용자가 별도로 지정하여 발화자를 인식할 수도 있다. 'Scene'은 드라마/영화 등 2차 창작에서 사용할 수 있는 장면 번호로, 작품 속 장면 구분자(***, '...' 등)를 기준으로 인덱싱한 숫자이다. 장면마다 인식된 장소와 시간을 각각 Place, Time으로 기록하였다. 이렇게 추출한 시공간적 배경은 장면 단위의 최종 결과를 출력할 때 활용된다. 그림 8은 NER 모델을 활용해 추출한 장면, 시간 예시로 DAT은 시간, LOC는 지명을 나타낸다. 예를 들어, "어제 오후 3시에 런던에서 만난다"라는 문장에서 '어제 오후 3시'와 '런던'은 각각 시간과 장소를 나타내는 개체로 추출된다.

이때 인스턴스 내 인용문의 발화자는 &C00& 형태의 코드 이름으로 변환하였다. C 뒤의 숫자 정렬 방식은 전체 소설에서 해당 등장인물의 언급 빈도를 내림차순으로 정렬하였다. 이러한 방식을 통해 새로운 소설마다 별도의 네임리스트를 생성하거나 소설을 학습시키지 않아도 발화자를 추론할 수 있으며, 네임리스트를 비교적 간결하게 유지할 수 있다. 또한 모델에서 발화 빈도가 높은 등장인물에 높은 가중치를 부여해 정확도를 높일 수 있다.

TOKEN	TAG
[CLS]	O
7	B-DAT
##월	I-DAT
##의	O
-	O
스	B-LOC
##때	I-LOC
##인	I-LOC
##.	O
[UNK]	O
[SEP]	O

그림 8. NER 모델 출력 결과
Fig. 8. NER model output results

3.5 CSN기반 발 화자 분류

3.4절에서 생성한 인스턴스는 아래의 과정에 따라 모델 학습에 쓰인다. 모델은 발화자, 인용문, 인용문 주변 맥락 정보를 활용하여 각 인용문의 발화자를 인식한다.

발화자는 인스턴스 내에서 언급이 된 발화자 해당 인용문의 발화자일 확률이 가장 높은 등장인물을 찾는 방식으로 찾는다. 따라서 발화자 후보를 설정하고, 후보자별로 인용문의 발화자일 확률을 찾는다. 발화자 후보는 Namelist에 해당하는 발화자들을 인스턴스 내에서 찾는 방식으로 진행된다. 인용문 정보는 인스턴스에서 추출한다. 그리고 인용문 주변의 앞뒤 맥락 정보는 발화자와 인용문 사이의 정보 (CSS, Candidate Specific Segments)를 활용한다. CSS는 발화자 후보의 이름이 언급된 문장과 인스턴스의 중심에 존재하는 발화문을 포함하여, 둘 사이에 존재하는 문장들로 구성된다. 예를 들어 한 인스턴스에 두 명의 등장인물 이름이 언급될 경우, 이로부터 두 개의 CSS를 얻는다. 만약 특정 인물의 이름이 인스턴스에서 2회 이상 등장한다면, 즉 한 후보에 대하여 2개 이상의 CSS가 만들어진다면 발화문으로부터 가까운 거리(글자 수 기준)에 위치한 후보가 선택된다. 한 인스턴스 내에서 CSS는 언급되는 등장인물의 수만큼 만들어진다. 이후 각 CSS는 BERT에 입력되어 맥락을 표현하고 인공지능망에 입력되어 발화자를 찾아낸다.

[&C00&의 CSS]
 그러는 사이 &C00&이 나타났고, 지긋한 나이의 &C01&를 보고 흘린 듯 걸음 두 사람은 캔스에 웃으며 대화를 나눴다.
 "할아버님. 돌려 깎기가 잘 안 돼요. 너무 두껍게 잘려요."
 [&C01&의 CSS]
 "할아버님. 돌려 깎기가 잘 안 돼요. 너무 두껍게 잘려요."
 "괜찮다. 나는 원래 두꺼운 오이를 좋아해."
 &C01&는 몇 안남은 치아로 오이를 즐겁게 씹어 먹었다.

그림 9. 발화자 인스턴스로부터 만들어지는 CSS 예
 Fig. 9. Example CSS created from the speaker instance

그림 9는 "할아버님. 돌려 깎기가 잘 안 돼요. 너무 두껍게 잘려요."의 발화자를 찾는 작업을 수행하는 CSS의 예시이다. 앞뒤 10문장으로부터 발화자 후보 ‘&C00&’과 ‘&C01&’ 두 명을 찾고, 해당 발화자의 이름이 언급되는 문장부터 인용문 문장까지를 인스턴스화 한다.

RoBERTa 모델에는 전체 웹소설 데이터를 띄어쓰기를 포함한 입력으로 넣는다. 이후, 그 출력값에서 인용문과 값을 가져온다. 그리고 발화자, 인용문 그리고 CSS를 합친 후 중요한 특징을 추출하기 위해서 max-pooling을 진행한다. 최종적으로 인용문별

로 발화자 후보별 점수가 나오게 되고, 인용문의 발화자일 확률이 가장 높은 후보가 최종 선정된다.

학습을 위한 하이퍼파라미터는 다회의 테스트를 거쳐 다음과 같이 최적화하였다. Loss function으로 Margin Ranking Loss를, optimizer로는 Adam을 사용하였다. Learning rate는 1e-5, 0.5의 Dropout 값과 Batch Size를 256으로 설정하였다. 총 10 Epoch를 돌았으며, Window Size (ws)는 10으로 하고 모델 학습을 진행하였다.

Instance index: 22
 회진이 끝나고 몹시 지친 기색이 역력한 &C01&가 목덜미를 주무르며 교수실로 돌아 뒤편에서 동료 의사인 &C05&가 &C01&를 부르며 걸음으로 다가왔다.
 "윤 교수님, 요즘 물결이 왜 그래?"
 "죽을 때라도 됐나 보지."
 "그러니까 조금 쉬라고. 젊은 거 믿고 그러다 진짜 혹 간다?"
 너스레를 떨다 그의 손에 &C05&의 손이 스치자, 순간 &C01&의 눈빛이 차갑게 돌변 &C01& 연립병원 흉부외과 최연소 교수.
 그의 외모도 한뼉했지만, 선한 얼굴과는 달리 독설가에 도가 지나친 결백증으로 병원 "미안 안 달았다, 안 달았어."
 그의 날 선 눈빛에 움찔한 &C05&가 손을 위로 뻗으며 결백을 주장했다.
 "괜찮아, 장갑 껴있으니까."
 다소 날카롭게 말을 건넌 &C01&가 주머니에 손을 꽂아 넣었다.
 저로서 무슨 의사를 해먹겠단 거냐고. 그래서 아직도 네가 모태솔로인 거라고. 혀를 끝끝 차면서도 그에게로 향하는 모든 이들의 시선은 365일 뜨거웠다.
 웬만한 성인 남성은 올라다본 적 없는 장신에, 창백해 보일 만큼 하얀 피부와 붉은 입 인정하지 않지만, &C01&는 남녀노소 선망하는 피조물, 그런 완벽에 가까운 남자였다. 그런 남자의 얼굴이 요즘 들어 여자가 많이 아파 보인다. 언제나 혈색이 돌던 그의 입 "건강관리는 20대부터 하는 거라니까?"
 걸음을 멈춘 &C01&가 미간을 잔뜩 좁혔다.
 "의도가 뭐야?"
 "20대랑 조금이라도 가까운 때 하란 소리지, 한 살이라도 젊을 때!"
 Speaker: &C01&
 Speaker_index: 01
 Category: latent
 Name: 너에게 닿으면
 Scene: 2
 Place: 교수실, 연립병원
 Time: None

그림 10. CSN 모델 입력값
 Fig. 10. CSN model input value

회진이 끝나고 몹시 지친 기색이 역력한 &C01&가 목덜미를 주무르며 교수실로 돌아 뒤편에서 동료 의사인 &C05&가 &C01&를 부르며 걸음으로 다가왔다.
 &C05& : "윤 교수님, 요즘 물결이 왜 그래?"
 &C01& : "죽을 때라도 됐나 보지."
 &C05& : "그러니까 조금 쉬라고. 젊은 거 믿고 그러다 진짜 혹 간다?"
 너스레를 떨다 그의 손에 &C05&의 손이 스치자, 순간 &C01&의 눈빛이 차갑게 돌변 &C01& 연립병원 흉부외과 최연소 교수.
 그의 외모도 한뼉했지만, 선한 얼굴과는 달리 독설가에 도가 지나친 결백증으로 병원 &C05& : "미안 안 달았다, 안 달았어."
 그의 날 선 눈빛에 움찔한 &C05&가 손을 위로 뻗으며 결백을 주장했다.
 &C01& : "괜찮아, 장갑 껴있으니까."
 다소 날카롭게 말을 건넌 &C01&가 주머니에 손을 꽂아 넣었다.
 저로서 무슨 의사를 해먹겠단 거냐고. 그래서 아직도 네가 모태솔로인 거라고. 혀를 끝끝 차면서도 그에게로 향하는 모든 이들의 시선은 365일 뜨거웠다.
 웬만한 성인 남성은 올라다본 적 없는 장신에, 창백해 보일 만큼 하얀 피부와 붉은 입 인정하지 않지만, &C01&는 남녀노소 선망하는 피조물, 그런 완벽에 가까운 남자였다. 그런 남자의 얼굴이 요즘 들어 여자가 많이 아파 보인다. 언제나 혈색이 돌던 그의 입 &C05& : "건강관리는 20대부터 하는 거라니까?"
 걸음을 멈춘 &C01&가 미간을 잔뜩 좁혔다.
 &C01& : "의도가 뭐야?"
 &C05& : "20대랑 조금이라도 가까운 때 하란 소리지, 한 살이라도 젊을 때!"

그림 11. CSN 모델의 출력 결과
 Fig. 11. CSN model output results

발화자 후보 중 점수가 가장 높은 것이 발화자로 선정되어 그림 11과 같이 발화자 인식 결과를 출력한다. 이때 발화자는 앞서 부여한 코드 이름의 형태로 출력되는데, 이를 다시 등장인물의 이름으로 바꾸고 미리 추출해 두었던 시공간적 배경 정보에 기반하여 장면별 시나리오로 변환하는 후처리 과정을 거친다. CSN 모델의 입력값과 출력 결과는 각각 그림 10, 11과 같다.

IV. 실험 결과

4.1 모델 학습 및 평가

모델의 성능을 평가하기 위해서 학습용 데이터셋과 검증용 데이터셋은 같은 작품을 사용하였으며, 8:2 비율로 인스턴스를 분리하여 구성하였다.

모델의 학습과 평가를 위해 활용한 작품 수를 목적과 장르에 따라 분류하면 표 1과 같다. 웹소설은 장르별로 다른 특성을 가진다. 따라서 테스트 데이터셋은 랜덤으로 추출한 세 작품(로맨스 판타지 두 작품, 로맨스 한 작품)으로 구성하였다.

표 1. 목적별, 장르별로 활용한 소설의 수
Table 1. Number of novels utilized by purpose and genre

Category	Romance fantasy	Romance	Fantasy	Total
Training set	40	15	5	60
Validation set	40	15	5	60
Test set	2	1	0	3

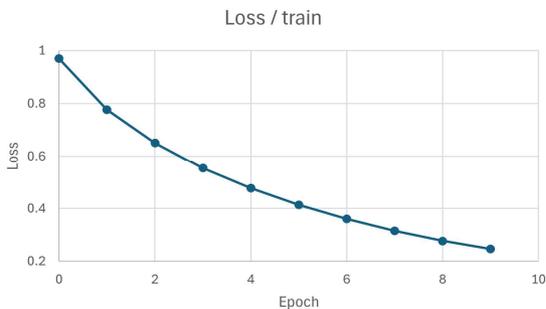


그림 12. 학습 데이터셋의 loss
Fig. 12. Loss of train dataset

학습용 데이터의 loss와 성능은 그림 12와 그림 13과 같다. 학습이 진행됨에 따라 train 데이터의

loss는 계속해서 감소하여 0.20을 기록하였고 정확도는 93.8%를 기록하였다. 검증용 데이터의 성능을 그림 14와 그림 15와 같다. validation 데이터의 loss는 0.52를 기록하였고 정확도는 83.2%를 기록하였다.

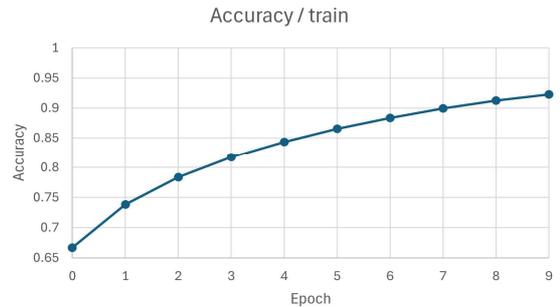


그림 13. 학습 데이터셋의 정확도
Fig. 13. Accuracy of train dataset

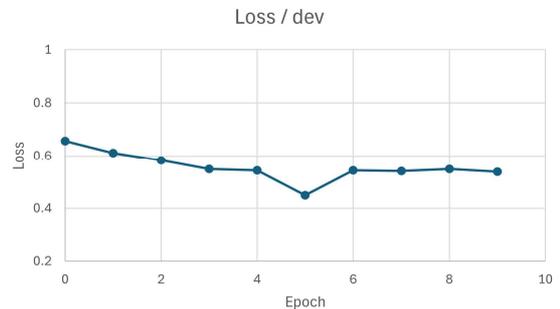


그림 14. 검증용 데이터셋의 loss
Fig. 14. Loss of validation dataset

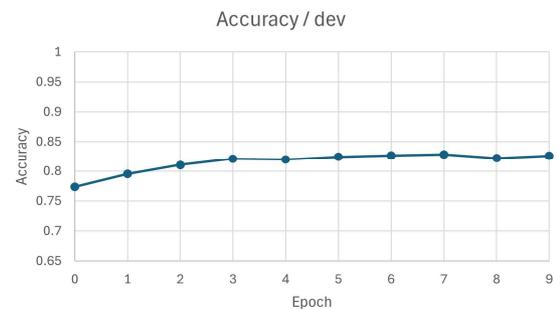


그림 15. 검증용 데이터셋의 정확도
Fig. 15. Accuracy of validation dataset

4.2 소설 데이터 적용 결과

학습된 모델의 성능을 평가하기 위해, 미리 분리해둔 두 개의 로맨스 판타지 소설(“괴물 여기사는 육아휴직을 원한다”, “패륜 폭군이 짐승 같아서 죄송합니다”)과 한 개의 로맨스 소설(“내 남편의 연인”)을 활용하였다.

테스트 실험 결과는 그림 16, 17과 같다. 실험 결과, 테스트 데이터에서의 평가 결과에 근접한 평가 결과를 보여주고 있으며, loss는 최저 0.63, 정확도는 최대 80.2%를 기록하였다. 전체 데이터에서 인용문별로 제시되는 평균 발화자 후보의 수는 3.64명으로 나타났다. 즉, 무작위로 예측하는 경우($100/3.64 = 27.4\%$ 의 정확도)와 비교했을 때, 제안 모델은 80.2%의 좋은 예측력을 보였다.



그림 16. 실제 소설 데이터셋의 loss
Fig. 16. Loss of an actual novel dataset

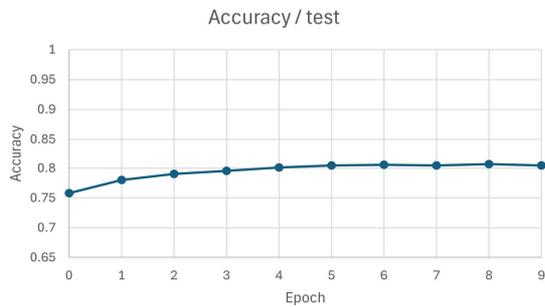


그림 17. 실제 소설 데이터셋의 정확도
Fig. 17. Accuracy of an actual novel dataset

4.3 구현 결과

마지막 단계로 인용문별로 발화자를 붙이고, id로 구분된 인물들을 이름으로 바꿔주는 과정을 거쳐 최종 결과물이 생성된다. 장면별 구분 및 시·공간 정보를 반영해 생성된 결과는 그림 18과 같다. 최종 결과물을 통해 장면별로 인식한 Place, Time이 가장 위에 기술되어 있고 인용문별로 자동 인식된 발화자의 이름이 포함된 것을 확인 할 수 있다.

Chen[17]의 방법과 제안한 방법의 차이점은 표 2와 같다. 우선 Chen은 훈련, 검증, 테스트에 사용하는 데이터를 모두 하나의 소설에서 추출한 데 비해,

제안한 방법은 총 63개의 데이터를 활용하였다. 특히 테스트 데이터는 학습과 검증 단계에서 사용하지 않은 소설을 추가로 활용하여 실험함으로써 학습하지 않은 다른 소설에 적용한 결과를 보여주지 않은 Chen의 모델과 달리, 새로운 소설에 대해서도 발화자 추론이 가능하다.

둘째, 제안 방법은 등장인물을 자동으로 인식해 수동 라벨링 단계를 제거하였다. Chen의 방식은 모델 학습 전, 작품 내 모든 등장인물과 그의 별칭을 기록해야 하므로 실질적으로 사용하기에 어려움이 있었다. 이에 비해 제안한 방식은 NER 모델로 인물 인식을 자동화하고 새로운 소설의 학습 및 추론을 가능하게 하였다.

#2 (장소: 교수실, 연립병원)
회진이 끝나고 몹시 지친 기색이 역력한 선우가 목덜미를 주무르며 교수실로 돌아오던 뒤편에서 동료 의사인 헤지가 선우를 부르며 걸음로 다가왔다.

헤지: "윤 교수님, 요즘 물골이 왜 그래?"
선우: "죽을 때라도 됐나 보지."
헤지: "그러니까 조금 쉬라고, 젊은 거 믿고 그러다 진짜 혹 간다?"

너스레를 떨다 그의 손에 헤지의 손이 스치자, 순간 선우의 눈빛이 차갑게 돌변했다. 윤선우. 연립병원 흉부외과 최연소 교수. 그의 외모도 한몫했지만, 선한 얼굴과는 달리 독설가애 도가 지나친 결벽증으로 병원

헤지: "미안. 안 닿았다, 안 닿았어."

그의 날 선 눈빛에 움찔한 헤지가 손을 위로 뻗으며 결백을 주장했다.

선우: "괜찮아, 장갑 껴있니까."

다소 날카롭게 말을 건넨 선우가 주머니에 손을 꽂아 넣었다. 저러서 무슨 의사를 해먹겠다 거냐고. 그래서 아직도 네가 모태솔로인 거라고. 혀를 끌끌 차면서도 그에게로 향하는 모든 이들의 시선은 365일 뜨거웠다. 웬만한 성인 남성은 올려다본 적 없는 장신에, 장백해 보일 만큼 하얀 피부와 붉은 입술 인정하기 싫지만, 선우는 남녀노소 선망하는 피조물, 그런 완벽에 가까운 남자였다. 그런 남자의 얼굴이 요즘 들어 어딘가 많이 아파 보인다. 언제나 혈색이 돌던 그의 입술

헤지: "건강관리는 20대부터 하는 거라니까?"

걸음을 멈춘 선우가 미간을 잔뜩 좁혔다.

선우: "의도가 뭐야?"
헤지: "20대량 조금이라도 가까울 때 하란 소리지, 한 살이라도 젊을 때!"

그림 18. 최종 결과물

Fig. 18. Final result

표 2. 기존 방법과 제안 방법의 차별점

Table 2. Differences between traditional and proposed methods

Category	Chen et al[17]	Proposed method
Number of works for training/testing	1	63
Automatic labeling feature	X	O
Performance (Accuracy)	78.5%	83.2%

마지막으로 발화자 인식 정확도를 높였다. 기존 방법은 한 가지 작품으로 학습 및 테스트를 진행하여 과적합 문제가 해결되지 않았으며, 발화자 인식 성능은 78.5%로 확인되었다. 제안한 방법은 과적합 없이 더욱 일반화된 모델임에도 불구하고, 83.2%의 정확도를 보이며 기존 모델보다 4.7% 더 높은 정확도를 보였다.

V. 결론 및 향후 과제

본 논문에서 제안된 CSN기반 시스템은 웹소설과 같이 텍스트만 존재하는 미디어에서 발화자를 자동으로 식별하고 태깅하는 데 중점을 두고 있다. 기존의 음성 인식 기반 발화자 인식 기술과 달리, 본 시스템은 발화자 후보, 인용문, 그리고 그 인용문이 나타나는 맥락을 포괄적으로 분석하여, 각 인용문에 대해 가장 가능성이 높은 발화자를 예측하였다. 이 과정에서, 다양한 자연어처리 기법과 기계 학습 알고리즘이 활용되었으며, 이를 통해 각 인용문의 발화자를 효과적으로 식별할 수 있는 모델을 구현하였다. 실험 결과, 본 시스템은 웹소설 데이터에 대해 평균 80.2%의 인식 정확도를 보였으며, 이는 학습 단계에서 보인 83.2%의 테스트 정확도와 유사한 성능을 보였다. 이를 통해 복잡한 다자간 대화가 포함된 텍스트에서도 본 모델이 높은 신뢰도로 발화자를 인식할 수 있음을 증명하였다. 이는 특히 음성 기반 발화자 인식이 여러 발화자의 목소리가 겹치는 상황에서 정확도가 떨어질 수 있는 문제를 텍스트 기반 접근 방식으로 효과적으로 해결할 수 있음을 확인하였다. 또한, 본 시스템의 적용 가능성은 한국 웹소설을 기반으로 한 K-드라마 시장에서 더욱 두드러지게 활용될 수 있다. 가령, 웹소설을 오디오북으로 전환하거나, 드라마의 대본을 준비하는 과정에서 대사와 발화자 정보를 자동으로 태깅함으로써 제작 과정을 간소화하고 창작자가 콘텐츠 제작에 더 집중할 수 있다.

향후 연구 과제로는 소설 내에서 비주류 인물이나 엑스트라 같은 낮은 빈도로 등장하는 인물의 인식을 향상하기 위한 연구가 진행될 예정이다. 또한, 여러 발화자가 동시에 같은 인용문을 말할 때 발생할 수 있는 인식 오류를 줄이기 위해 알고리즘

의 정확도를 높이는 연구를 진행할 계획이다.

References

- [1] E. Lee, "Speech Recognition Services", *Asti Market Insight*, Vol. 2022, No. 67, pp. 10-13, Jun. 2022.
- [2] Netflix introduces closed captioning for entertainment that identifies speakers and lines of dialogue, https://mobile.newsis.com/view.html?ar_id=NISX20230808_0002406282 [accessed: Apr. 18, 2024]
- [3] X. Anguera, S. Bozonnet, N. Evans, C. Fredouille, G. Friedland, and O. Vinyals, "Speaker diarization: A review of recent research", *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, Vol. 20, No. 2, pp. 356-370, Feb. 2012.
- [4] Books Market Size, Share & Trend Report, <https://www.grandviewresearch.com/industry-analysis/books-market> [accessed: Apr. 18, 2024]
- [5] K. Kim, "Press Release on the Status of the Web Novel Industry in 2022", *Ministry of Culture, Sports and Tourism*, pp. 1-6, Sep. 2023.
- [6] S. Sreedharan and C. Eswaran, "A Review On Speaker Verification: Challenges And Issues", *International Journal of Scientific & Technology Research*, Vol. 8, No. 3, Aug. 2019.
- [7] F. Xu and G. Li, "Multiple Information Extraction and Interaction for Emotion Recognition in Multi-Party Conversation", *2023 IEEE Symposium on Computers and Communications (ISCC)*, Gammarth, Tunisia, pp. 977-982, Jul. 2023. <https://doi.org/10.1109/ISCC58397.2023.10218135>.
- [8] S. Mizuno, N. Hojo, S. Kobashikawa, and R. Masumura, "Next-Speaker Prediction Based on Non-Verbal Information in Multi-Party Video Conversation", *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, pp. 1-5, May 2023. <https://doi.org/10.1109/ICASSP49357.2023.10094679>.

- [9] X. Quan, S. Wu, J. Chen, W. Shen, and J. Yu, "Multi-Party Conversation Modeling for Emotion Recognition", in *IEEE Transactions on Affective Computing*, pp. 1-17 May 2023. <https://doi.org/10.1109/TAFFC.2023.3273589>.
- [10] J.-C. Gu, C. Tao, and Z.-H. Ling, "Who Says What to Whom: A Survey of Multi-Party Conversations", *Proc. of the Thirty-First International Joint Conference on Artificial Intelligence Survey Track*, pp. 29-34, Nov. 2023. <https://doi.org/10.24963/ijcai.2022/768>.
- [11] B. Lee and Y. S. Choi, "Graph Based Network with Contextualized Representations of Turns in Dialogue", *arXiv:2109.04008*, Sep. 2021. <https://doi.org/10.48550/arXiv.2109.04008>.
- [12] Z. Shi and M. Huang, "A deep sequential model for discourse parsing on multi-party dialogues", *Proc. of the AAAI Conference on Artificial Intelligence*, Vol. 33, No. 1, pp. 7007-7014, Dec. 2018. <https://doi.org/10.48550/arXiv.1812.00176>.
- [13] A. Wang, et al., "A Structure Self-Aware Model for Discourse Parsing on Multi-Party Dialogues", *Proc. of the Thirtieth International Joint Conference on Artificial Intelligence Main Track*, pp. 3943-3949, Aug. 2021. <https://doi.org/10.24963/ijcai.2021/543>.
- [14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding", *arXiv:1810.04805*, Oct. 2018. <https://doi.org/10.48550/arXiv.1810.04805>.
- [15] J.-C. Gu, C. Tao, Z.-H. Ling, C. Xu, X. Geng, and D. Jiang, "MPC-BERT: A Pre-Trained Language Model for Multi-Party Conversation Understanding", *arXiv:2106.01541*, Jun. 2021. <https://doi.org/10.48550/arXiv.2106.01541>
- [16] M. G. Baysar, P. Cavalin, C. Pinhanez, and B. Zadrozny, "Learning Multi-Party Turn-Taking Models from Dialogue Logs", *arXiv:1907.02090*, Jul. 2019. <https://doi.org/10.48550/arXiv.1907.02090>.
- [17] Y. Chen, Z.-H. Ling, and Q.-F. Liu, "A Neural-Network-Based Approach to Identifying Speakers in Novels", *Proc. Interspeech 2021*, pp. 4114-4118, Aug. 2021. <https://doi.org/10.21437/Interspeech.2021-609>.
- [18] Y. Liu, et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach", *arXiv:1907.11692*, Jul. 2019. <https://doi.org/10.48550/arXiv.1907.11692>.
- [19] S. Park, J. Jeong, J. Lee, W. Lim, S. Kim, J. Choi, and J. Yoon, "KLUE: Korean Language Understanding Evaluation", *arXiv:2105.09680*, May 2021. <https://doi.org/10.48550/arXiv.2105.09680>.
- [20] H. Son, Y. Han, K. Nam, S. Han, and G. Yoo, "Development of a News Trend Visualization System based on KPF-BERT for Event Changes and Entity Sentiment Analysis", *The Journal of Korean Institute of Information Technology*, Vol. 22, No. 1, pp. 203-213, Jan. 2024. <https://doi.org/10.14801/jkiit.2024.22.1.203>.
- [21] NER model, <https://github.com/kmounlp/NER/> [accessed: Apr. 18, 2024]
- [22] S. Han, D. Yu, B.-W. On, and I. Lee, "Empirical Study on the Loss Functions of Contrastive Learning-based Multi-scale BERT model for Automated EssayScoring", *The Journal of Korean Institute of Information Technology*, Vol. 21, No. 9, pp. 51-63, Sep. 2023. <https://doi.org/10.14801/jkiit.2023.21.9.51>.

저자소개

전 승 호 (Seungho Jun)



2023년 12월 : 고려대학교 지능정보
SW아카데미 3기 수료(640H)
2018년 3월 ~ 현재 : 아주대학교
정치외교학과 학사과정
관심분야 : 자연어 처리, 딥러닝,
데이터사이언스, 사회과학,
국제통상

윤 은 (Eun Yun)



2023년 12월 : 고려대학교 지능정보
SW아카데미 3기 수료(640H)
2019년 3월 ~ 현재 : 고려대학교
국어국문학과 학사과정
관심분야 : 자연어 처리, 딥러닝,
머신러닝

이 하 원 (Hawon Lee)



2023년 12월 : 고려대학교 지능정보
SW아카데미 3기 수료(640H)
2017년 3월 ~ 현재 : 고려대학교
환경생태공학부 학사과정
관심분야 : 딥러닝, 머신러닝,
생물정보학, 자연어처리

강 민 혁 (Minhyeok Kang)



2023년 12월 : 고려대학교 지능정보
SW아카데미 3기 수료(640H)
2020년 3월 ~ 현재 : 고려대학교
산업경영공학부 학사과정
관심분야 : 자연어 처리, 딥러닝,
머신러닝, 데이터사이언스,
빅데이터 분석

김 정 원 (Jeongwon Kim)



2023년 12월 : 고려대학교 지능정보
SW아카데미 3기 수료(640H)
2019년 3월 ~ 현재 :
숙명여자대학교 컴퓨터과학전공
학사과정
관심 분야 : 딥러닝, 머신러닝,
데이터사이언스, 자연어 처리

유 길 상 (Gilsang Yoo)



2010년 3월 ~ 현재 :
(사)한국컴퓨터게임학회 이사
2021년 3월 ~ 현재 : 고려대학교
정보대학 정보창의교육연구소/
지능정보 SW아카데미 교수
2023년 3월 ~ 현재 : (사)한국미디어
아트산업협회 수석부회장
관심분야 : 데이터사이언스, 데이터 시각화, 빅데이터
분석, 3D영상 콘텐츠, 머신러닝, 딥러닝, 컴퓨터교육