

이미지 캡션 및 재귀호출을 통한 스토리 생성 방법

전이슬¹, 조동하², 문미경^{2*}

¹동서대학교 컴퓨터공학과, ²동서대학교 소프트웨어학과

Automated Story Generation with Image Captions and Recursiva Calls

Isle Jeon¹, Dongha Jo², Mikyeong Moon^{2*}

¹Division of Computer Engineering, Dongseo University

²Dept. of Software, Dongseo University

요약 기술의 발전은 제작 기법, 편집 기술 등 미디어 산업 전반에 걸쳐 디지털 혁신을 이루어 왔고, OTT 서비스와 스트리밍 시대를 관통하며 소비자 관람 형태의 다양성을 가져왔다. 빅데이터와 딥러닝 네트워크의 융합으로 뉴스 기사, 소설, 대본 등 형식을 갖춘 글을 자동으로 생성하였으나 작가의 의도를 반영하고 문맥적으로 매끄러운 스토리를 생성한 연구는 부족하였다. 본 논문에서는 이미지 캡션 생성 기술로 스토리보드 속 사진의 흐름을 파악하고, 언어모델을 통해 이야기 흐름이 자연스러운 스토리를 자동 생성하는 것을 기술한다. 합성곱 신경망(CNN)과 주의 집중기법(Attention)을 활용한 이미지 캡션 생성 기술을 통해 스토리보드의 사진을 묘사하는 문장을 생성하고, 첫 번째 이미지 캡션을 KoGPT-2에 입력하여 생성된 새로운 글과 두 번째 이미지의 캡션을 다음 입력값으로 활용한 재귀적 접근 방안을 제안하여 전후 문맥이 자연스럽게 기획 의도에 맞는 스토리를 생성하는 연구를 진행한다. 본 논문으로 인공지능을 통해 작가의 의도를 반영한 스토리를 자동으로 대량 생성하여 콘텐츠 창작의 고통을 경감시키고, 인공지능이 디지털 콘텐츠 제작의 전반적인 과정에 참여하여 미디어 지능화를 활성화한다.

• 주제어 : 이미지 캡션, 인공지능, 자연어처리, KoGPT-2, 딥러닝, 미디어 지능화

Abstract The development of technology has achieved digital innovation throughout the media industry, including production techniques and editing technologies, and has brought diversity in the form of consumer viewing through the OTT service and streaming era. The convergence of big data and deep learning networks automatically generated text in format such as news articles, novels, and scripts, but there were insufficient studies that reflected the author's intention and generated story with contextually smooth. In this paper, we describe the flow of pictures in the storyboard with image caption generation techniques, and the automatic generation of story-tailored scenarios through language models. Image caption using CNN and Attention Mechanism, we generate sentences describing pictures on the storyboard, and input the generated sentences into the artificial intelligence natural language processing model KoGPT-2 in order to automatically generate scenarios that meet the planning intention. Through this paper, the author's intention and story customized scenarios are created in large quantities to alleviate the pain of content creation, and artificial intelligence participates in the overall process of digital content production to activate media intelligence.

• Key Words : Image Caption, Artificial intelligence, NLP, KoGPT-2, Deep Learning, Media Intelligence

Received 28 February 2023, Revised 13 March 2023, Accepted 15 March 2023

* Corresponding Author Mikyeong. Moon, College of Software Convergence, Dongseo University, 47, Jurye-ro, Sasang-gu, Busan, Korea.
E-mail: mkmooon@dongseo.ac.kr

I. 서론

기술의 발전은 문화를 바꾼다. 미디어 산업의 전통적인 제작 기법, 편집 기술 등은 디지털 전환을 통해 변화하였다. 콘텐츠 생산, 전달 및 관리 방식의 변화로 디지털 콘텐츠 생산이 증가하였고, 더불어 코로나 장기화와 IT 인프라 발달 및 실시간 스트리밍 서비스 등으로 디지털 콘텐츠 소비 또한 확대되었다.

국내에서는 디지털 기술에 초점을 두고 제작에서 유통까지의 과정에서 디지털 기술을 활용한 콘텐츠(음성, 이미지, 영상 등)를 디지털 콘텐츠라 정의하며, 최근 이에 빅데이터와 인공지능 기술 융합을 더 하여 디지털 콘텐츠를 자동으로 생성하는 연구가 활발히 진행되고 있다[1].

대표적으로 인공지능 기반 딥페이크 기술을 활용하여 아동용 콘텐츠 제작을 위한 연구를 진행한 바가 있다[2]. 또한 2016년 영국 런던 공상과학영화제에 출품한 '썬 스프링(Sunspring)'은 약 9분 정도의 러닝타임을 가지는 단편 SF 영화로 출품된 180여 개의 작품 중 10위 안에 선정된 바 있으며, 2020년에는 자연어처리 언어모델 GPT-3(Generative Pretrained Transformer-3)로 대본을 작성하여 제작된 약 3분 30초 길이의 단편 영화 '상품판매원(Solicitor)'이 인공지능으로 제작된 단편영상을 공개하는 유튜브 채널 'Calamity AI'에 공개되었다[3-4]. 그러나 인공지능을 통해 제작된 시나리오는 전체적인 흐름이 어색하여 짧은 단편 영화만 제작 가능하다는 한계가 분명히 존재하며 작가의 의도를 반영하고 스토리 흐름이 있는 글을 생성한 연구는 부족하다.

본 논문에서는 개연성 부분을 해결할 수 있도록 이미지 캡션 생성 기술로 스토리보드 속 이미지의 흐름을 파악하고, 딥러닝 모델을 통해 작가의 기획 의도를 반영한 스토리를 자동 생성하는 것을 기술한다. CNN(Convolutional Neural Network, CNN)과 주의 집중 기법(Attention Mechanism)을 활용한 이미지 캡션 생성 기술을 통해 스토리보드 속의 이미지를 묘사하는 문장을 생성하고, 첫 번째 이미지 캡션을 한국어 자연어 처리 모델 KoGPT-2(KoreanGPT-2)에 입력하여 생성된 여러 단편의 텍스트 중 분위기에 맞는 글을 선택하고 두 번째 이미지의 캡션을 다음 입력값으로 활용한 재귀적 접근 방안을 제안하여 전후 문맥이 자연스러운 스토리를 자동으로 생성한다. 또한, 인공지능 언어모델

이 생성한 글에 작가가 새로운 방향을 제시하고자 하는 경우, 작가가 직접 문장을 덧붙여 완성한 글을 다음 단계 학습에서 언어모델의 입력값으로 사용하여 기획 의도에 맞고 완성도 높은 작품을 생성한다.

본 논문의 기여도는 다음과 같다. 딥러닝 기술을 활용하여 작가의 의도를 반영하고 자연스러운 이야기 흐름을 가진 스토리를 대량으로 생성할 수 있어 창작의 고통을 경감시킬 수 있다. 또한, 드라마, 영화 등 스토리를 기반으로 한 콘텐츠 제작의 전반적인 과정에 인공지능이 참여하여 미디어 지능화를 현실화 시킬 것으로 기대한다.

본 논문의 구성은 다음과 같다. 2장에서는 이미지 캡션 생성 기술과 스토리 자동 생성 관련 연구에 관해 이야기하며, 3장에서는 이미지 캡션 생성 기술을 활용하여 인공지능을 통해 스토리를 자동으로 생성하는 방안을 제안한다. 4장에서는 제안 방안에 대한 실험 환경과 결과에 대해서 분석한다. 마지막으로 5장에서는 결론을 도출하고 향후 계획에 관해서 기술한다.

II. 관련연구

2.1 이미지 캡션 생성 기존 연구

스마트폰 보급 및 각종 기기 내 카메라 기술 발달로 이미지 데이터가 증가하여 이를 활용하는 연구가 발달하고 있다. 그중 이미지 캡션 생성 기술은 컴퓨터 비전(Computer vision)과 자연어 처리(Natural Language Processing)를 종합한 과제로 이미지를 인식하고 특징을 추출하는 이미지 처리 부분과 처리된 이미지 정보를 바탕으로 자연어를 생성하는 부분 그리고 두 부분을 연결하는 매핑 함수와 함께 구성된다.

이미지 캡션 생성 연구의 기본 프레임워크는 합성곱 신경망(CNN)을 통해 이미지 정보를 인코딩하고, 자연어를 처리하는 회귀신경망(Recurrent neural network, RNN)의 디코더 부분을 결합한 인코더-디코더 모델이다[5]. 이 기본 프레임워크를 활용하여 시력 저하로 어려움을 겪는 사람들에게 도움을 줄 수 있도록 이미지 내용을 인식하고, 이를 한국어로 번역해 음성으로 들려주는 인공지능 프로그램을 구현한 연구가 있다[6].

그러나 회귀신경망의 기술기 소설로 문장이 길어질 경우 앞 단어의 정보를 잃어버리는 장기 의존성(Long-Term Dependency) 문제가 있어 본 논문에서는

이를 개선한 LSTM(Long Short-Term Memory)의 구조의 디코더를 사용하고, 인코더에서 전체 이미지 픽셀 중 중요하다고 판단되는 픽셀에 가중치를 크게 부여하여 예측해야 할 단어와 연관이 있는 부분을 더 집중하는 주의 집중 기법(Attention)을 추가로 도입한 모델을 사용한다[7]. 또한, 연속된 이미지로부터 장면 사이의 전후 맥락(context) 및 인과 관계 이해하여 스토리 문장을 생성하기 위해 주의집중(Attention) 메커니즘을 사용한 딥러닝 모델을 제안한 연구가 있다[8].

생성된 캡션을 살펴보면 연속된 이미지들의 전후 맥락이 잘 반영되었지만, 특정 이미지와 다소 다른 상황이 문장으로 표현되기도 하고 스토리의 전개 및 문장의 구조가 비교적 단순롭고 반복적이다.

본 논문에서는 스토리보드의 연속적인 이미지 각각의 캡션을 생성한 뒤, 자연어 처리 모델(Generative Pre-trained Transformer, GPT)에 입력하여 스토리를 생성한다. 스토리보드의 이미지 캡션 생성을 위해 CNN-LSTM, Attention 아키텍처를 활용하여 기획 의도에 맞으며 이야기 흐름이 자연스러운 스토리를 생성한다. 기존 연구에서 사용된 모델과의 성능 비교를 넘어 생성된 결과를 통해 미디어 콘텐츠 제작에 활용하여 그 영역을 확장한다.

2.2 기존 스토리 생성 연구

인공지능 자연어처리 기술을 활용하여 스토리를 생성한 기존 연구로 장단기 메모리(Long Short-Term Memory, LSTM) 모델을 사용하여 드라마 대본을 자동 생성한 연구가 있다[9].

드라마 대본의 텍스트 구조 패턴을 구현하였으나 점진적 소설 문제로 부자연스럽고 의미를 파악하기 어려운 문장을 생성하여 본 논문에서는 대용량의 데이터로 사전학습을 진행하고 한국어 문장 생성에 최적화되어 있는 KoGPT-2 활용하여 시나리오를 생성한다.

KoGPT-2는 영어 기반으로 학습한 GPT-2의 부족한 한국어 성능을 극복하기 위해 40GB 이상의 한국어 문장으로 학습된 한국어 디코더 언어모델이다[10].

연구[11]은 GPT-3 모델을 통해 여러 개의 스토리를 생성하고 사용자에게 선택지로 제공하여 작가의 의도를 반영한 시나리오를 생성할 수 있는 모델을 제안하였다. 또한, 기획 의도에 벗어난 선택지를 제공하거나 전후 맥락에 맞지 않는 문장을 생성할 경우, 작가가 직접 개입하여 이야기를 전개해 나아간다.

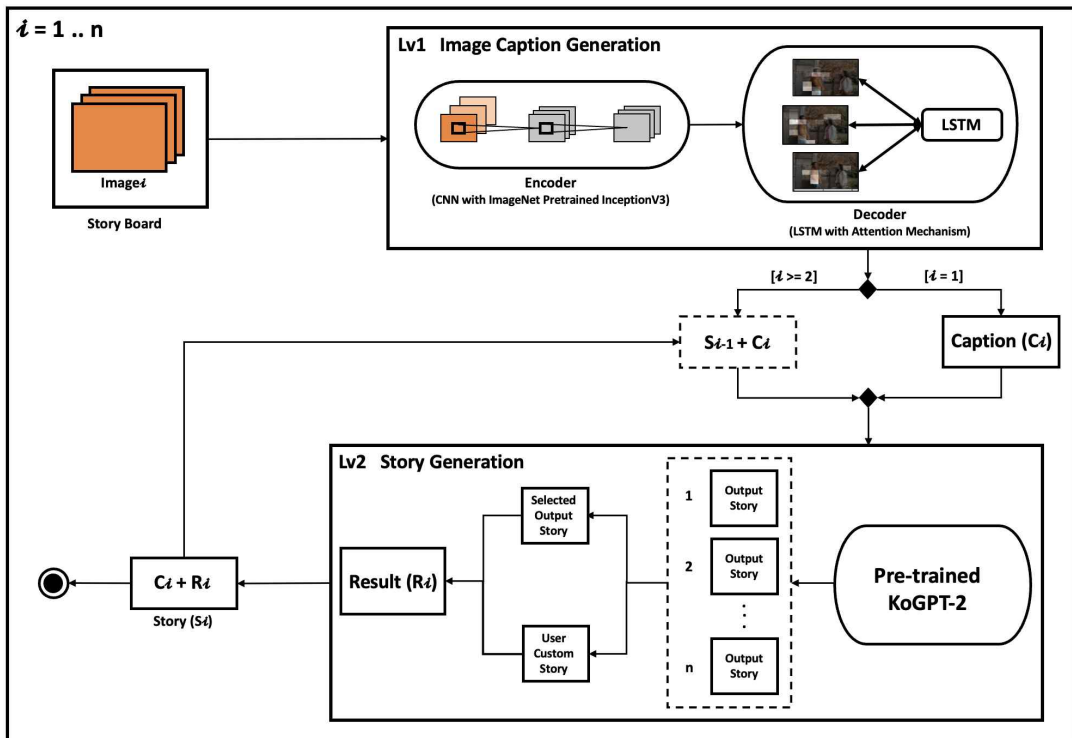


Fig. 1. System Architecture

본 논문에서 기획 의도 및 영화 스토리보드 속 이미지 내용 흐름에 맞는 스토리를 생성하기 위해 이미지 캡션 생성 기술과 연구[11]의 아키텍처를 함께 적용한 시스템을 제안한다.

III. 제안 방법

이번 절에서는 이미지 캡션 생성 기술을 통해 스토리보드 속 이미지의 캡션을 생성하고, 이를 활용하여 스토리를 생성하는 방안에 대해 상세히 설명한다.

본 논문에서 제안하는 시스템 구조는 Fig. 1과 같이 크게 2단계로 구성된다. 첫 번째 단계는 Fig. 1의 Lv1. Image Caption generation 부분에 해당하며, Microsoft에서 제공하는 COCO(Common Objects in Context) 데이터셋으로 이미지 캡션 생성 모델을 사전학습 시킨 뒤, 스토리보드 속 이미지의 캡션을 순차적으로 생성한다. MS COCO는 영어 데이터로 학습되어 본 논문에서 생성된 스토리보드의 이미지 캡션을 일반적으로 일상에서 많이 사용되는 실시간 번역기를 통해 한국어 번역 과정을 거친 뒤 다음 단계에서 활용한다.

두 번째는 단계는 Fig. 1의 Lv2. Story Generation 단계로 사용자가 생성하고자 하는 장르의 데이터로 사전 학습된 KoGPT-2 모델에 이전 단계에서 생성된 이미지 캡션들을 입력하여 사용자의 기획 의도에 맞추어 전후 맥락이 자연스러운 스토리를 생성한다.

3.1 스토리 보드의 이미지 캡션 생성

장면(Scene)을 묘사하는 이미지, 인물의 대사, 촬영 방법 등으로 구성된 스토리보드는 영화, 방송, 뮤직비디오, 애니메이션, 게임 등 다양한 매체의 영상물을 기획하고 제작하는 데 중요한 역할을 한다. 그중 장면의 순서에 따라 나열된 이미지로 스토리보드 내용의 흐름을 이해할 수 있어 본 논문에서 딥러닝을 통해 스토리보드의 이미지를 묘사하는 캡션을 생성한 뒤, 기획 의도에 맞는 스토리를 자동으로 생성한다.

이미지 캡션 생성 모델의 사전학습을 위해 최근 Image Captioning 관련 연구에서 많이 사용되고 있는 대표적 데이터로 123,287장의 이미지와 886,284개의 문장(캡션)으로 구성되어 있으며 1개의 이미지에 약 5개의 캡션이 매핑된 MS COCO 데이터셋을 사용한다[12].

본 논문의 실험 환경에서 가용할 수 있는 범위로

제한하여 MS COCO의 이미지 캡션 데이터셋에서 무작위로 30,000개의 캡션을 추출하고, 그중 24,000개의 캡션은 훈련 데이터로 사용하고, 나머지 6,000개의 캡션은 검증 데이터로 사용한다. 이미지 캡션 생성을 위해 인코더에 합성곱 신경망(CNN)과 어텐션 메커니즘을 더한 LSTM 디코더를 활용한다. 인코더로는 각 이미지에서 특징을 추출하여 집중적으로 확인해야 할 픽셀의 정보를 전달하는 사전 학습된 구글의 Inception V3를 사용한다.

시스템의 중요한 부분이며 학습 모델의 정확도에 영향을 미치는 이미지 사전 처리 작업으로 이미지 디코딩 및 크기 조절 작업이 필요하다. 사전학습 모델마다 요구하는 이미지의 크기는 다르며 Inception V3 모델의 경우 299x299x3픽셀을 요구하기 때문에 해당으로 이미지 크기를 조절하고 픽셀값을 정규화하여 인코더를 학습시킨다.

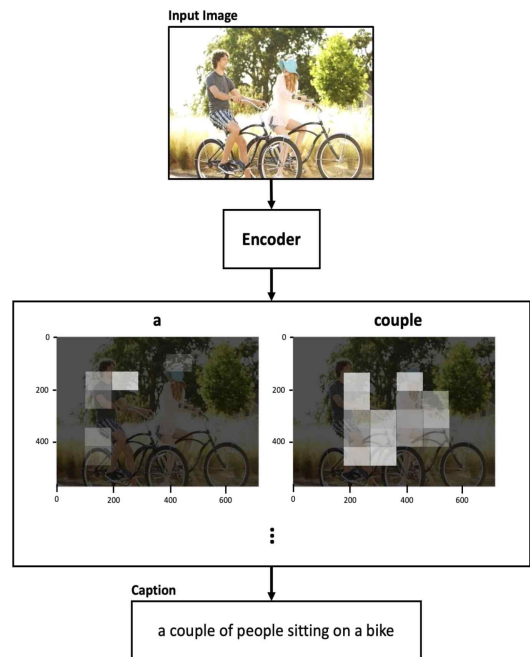


Fig. 2. Attention Mechanism

디코더를 학습시키기 위해 문장 데이터의 길이를 일정하게 만드는 패딩 작업과 문장을 단어 단위로 쪼개는 토큰화 과정을 통해 데이터 전처리 과정을 거친다. 디코더에는 LSTM 기반 주의 집중(attention) 모델을 사용하여 어떤 단어가 이미지의 어느 부분에 크게 반응하는지를 학습해 이미지를 묘사하는 최적의 문장을 생성한다(Fig 2).

3.2 스토리 생성을 위한 재귀적 접근 방안

본 논문에서 한국어 기반 스토리 생성을 위해 텍스트의 다음 단어를 잘 예측하고 문장 생성에 최적화된 GPT 언어모델을 활용했으며, 영어 문장으로 학습된 GPT 대신 40GB 이상의 한국어 문장을 학습한 KoGPT-2를 사용한다.

학습 데이터 수집을 위해 한국저작권위원회에서 저작권 보호기간이 지난 저작물을 자유롭게 이용할 수 있도록 개설한 웹사이트 '공유마당'에서 저작권이 만료된 한국 소설 492권을 수집하였다.

Table 2. Learning Data (Genre: Romance)

Id	Title	Author	Chapter	Content
1	햄릿의 연인	김지혜	1	밤공기 가..
2	햄릿의 연인	김지혜	2	계절의..
...
1293	애인	김내성	10	황교장 은..

Table 3. Data Detail

Data Detail
[romance] 죄(罪)와 벌(罰) 이무영 1 경관이 쓴 피스톨에 범인인 교회지기가 쓰러지자 관중석에서는 벌써 의자 젖혀지는 소리가 요란하였다. 그러나 화면은 아직도 계속되고 있다.

로맨스 장르의 스토리 생성을 목표로 하여 수집한 학습 데이터 중 로맨스 장르만 추출하여 총 153권의 소설을 수집하였다.

대부분 소설에서 시간이나 배경이 변동되거나 주요 인물들이 등장할 때, 이야기 전개에 변동이 생기는 경우 챗터로 구분 지어 이야기를 전개한다. 이러한 특성을 고려하여 언어모델이 이야기 전개의 일관성을 유지하며 스토리를 생성시킬 수 있도록 로맨스 소설을 챗터 단위로 구분하는 데이터 정제과정을 거쳐 table 2와 같이 id(번호), title(제목), author(작가), chapter(챗터), content(본문)로 학습 데이터를 구성하였다.

별도의 챗터없이 전개되는 소설의 경우 소설 전체를 10등분으로 나누고 임의의 챗터를 생성하여 총 1,293개 챗터로 구분된 이야기를 학습하였다.

로맨스 소설을 사전 학습한 KoGPT-2 모델에 Lvl Image Caption Generation 단계에서 생성된 첫번째 이미지의 캡션을 입력하여 N개의 스토리를 생성한다. 그중 이미지와 어울리거나 사용자의 기획 의도에 맞는 스토리를 선택하면 첫 번째 이미지 캡션과 사용자가 선택한 스토리, 두 번째 이미지 캡션을 더한 입력값을 KoGPT-2 모델에 넣어 이전 이야기 흐름과 이어지는 새로운 N개의 스토리를 생성한다.

위 과정은 사용자가 스토리 생성 작업을 종료하기 전까지 반복하여 진행되는데, 만약 사용자가 언어모델이 생성한 글에 새로운 이야기 흐름을 제시하고자 한다면 사용자가 직접 입력하여 이전에 생성된 스토리들과 함께 다음 언어모델 학습 때 입력값으로 전달한다.

본 논문에서 제안한 재귀적 접근 방안은 언어모델에 하나의 입력값을 이용해 스토리를 생성한 연구와 이미지 캡션을 단편적으로 활용하여 스토리를 생성한 연구보다 전후 문맥을 고려하고 사용자의 기획 의도에 맞는 스토리를 생성하는 것을 확인할 수 있다.

IV. 검증

본 논문에서는 이미지 캡션 생성 기술과 자연어처리 언어 모델을 활용하여 스토리보드의 이야기 흐름에 맞는 로맨스 장르의 스토리를 자동으로 생성하는 것을 목표로 한다.

이미지 캡션 생성 모델 학습을 위한 실험 데이터는 본 논문의 실험 환경에서 가용할 수 있는 범위로 제한하여 MS COCO의 이미지 캡션 데이터셋에서 무작위로 30,000개의 캡션을 추출하여, 그중 24,000개의 캡션은 훈련 데이터로 사용하고, 나머지 6,000개의 캡션은 검증 데이터로 사용하였다.

인공지능으로 스토리를 자동 생성하기 위해 저작권이 만료된 153권의 로맨스 소설을 챗터로 구분하여 총 1,293개의 이야기로 KoGPT-2 언어모델을 학습시켰다. 본 연구에서 활용한 모델 학습을 위해 T4(GPU), 32GB(RAM)의 자원을 지원하는 Google Colab Pro 환경에서 진행하였다.

4.1 스토리보드의 이미지 캡션 생성

본 실험에서 생성하고자 하는 스토리의 줄거리는 다음과 같다. “남녀 주인공이 도서관에서 만나 식사하러 간다.” 이에 스토리보드에는 줄거리를 표현할 수 있는 연속된 3장의 이미지를 선정하였고, CNN-LSTM, Attention 알고리즘을 통해 각각의 이미지 캡션을 Fig. 3과 같이 생성하였다. 생성된 캡션은 영문이므로 실시간 번역기를 활용해 한국어로 번역하였다.

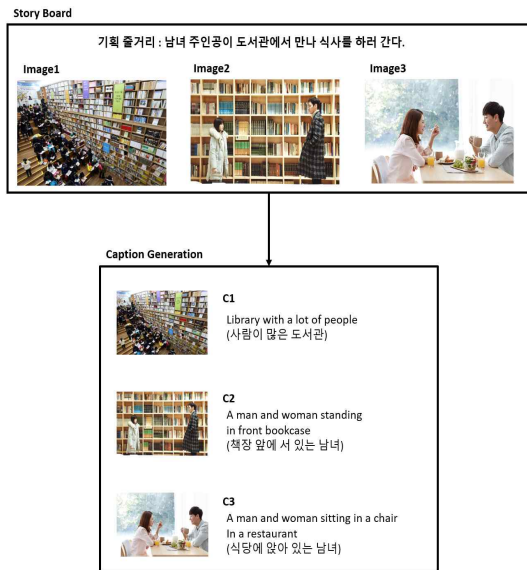


Fig. 3. Image Caption Generation

4.2 스토리 생성

자연어처리 언어 모델이 자동으로 스토리를 생성하기 위해 이전 단계에서 생성된 3개의 한국어 캡션을 입력값으로 활용한다. 기획 의도에 맞고 이야기 흐름이 연결되는 스토리를 생성하기 위해 본 논문에서 제안한 재귀적 접근 방안의 성능을 검증하고자 4가지 케이스로 나누어 실험하였다.

첫 번째 케이스는 로맨스 소설로 사전학습된 KoGPT-2 모델에 하나의 이미지 캡션을 넣어 스토리를 생성하는 일반적인 방법으로 Table 4와 같은 결과를 얻었다. 생성된 스토리를 살펴보면 4번째 문장부터 이야기 흐름이 맞지 않고 어색한 문장을 생성하는 것을 확인할 수 있다.

Table 4. The result of conventional system

Input 1	C1
Result	
<p>[C1] 사람이 많은 도서관 사서실 자료실 등을 돌아다니면서 자료들을 살펴봤다. 예휴 하여튼 난 내일 드디어 개강이야. 한숨을 푹푹 쉬어보았다. 자료실에 도착하니 내 옆에 조용히 서있는 이 젊은 남자가 보였다. 눈이 휘둥그레진 서영이 뭔가 말하려는데 갑자기 그 남자가 서영을 쳐다보았다. 그 남자가 있는데 갑자기 방문이 확 열렸다 닫혔다 문이 닫히고 서영의 방 안으로 들어갔다 서영을 발견한 순간 그 남자로 인해 그 남자는 몸을 돌려 그 사물이 날아들고 말았나 봤나</p>	

Table 5. The result of story generation using captions

Input 1	C1
Input 2	C2
Input 3	C3
Result	
<p>[C1] 사람이 많은 도서관 사서실 자료실등을 돌아다니면서 자료들을 살펴봤다. 예휴 하여튼 난 내일 드디어 개강이야 한숨을 푹푹 쉬어보았다 자료실에 도착하니 내 옆에 조용히 서있는 이 젊은 남자가 보였다 눈이 휘둥그레진 서영이 뭔가 말하려는데 갑자기 그 남자가 서영을 쳐다보았다.</p> <p>[C2] 책장 앞에서 있는 남녀 그들이 문을 열고 들어갔을때는 이미 낮이었다. 그는 무표정하게 무언가를 생각하고 있는듯 했고 그대로 서 있던 모습도 서글프게 보였다. 하지만 기억하고 싶지 않았던 것은 푸름이 기억하지 못했던 부분이었다 당시 그는 수진의 뺨에 난 상처에 잠시 멍이 들고 깨어났을 뿐이다.</p> <p>[C3] 식당에 앉아 있는 남녀 연인사동반 가족들로 모였다 차사장이 어르신들의 인사를 받고 식사도 하는 동안 아이들은 시원한 밤바람 속에 웅알거렸다. 큰언니와 아이들이 눈부시게 반짝거리고 있었지만 말이. 나름대로 이쁘다는 말도 들었다</p>	

두 번째 케이스는 스토리보드의 이미지 캡션 3개를 KoGPT-2에 순차적으로 입력하여 3가지 스토리를 생성하는 방법으로 그 결과는 Table 5와 같다. 생성된 캡션을 순차적으로 모델에 입력하였지만, 첫 번째 이야기에서는 ‘서영’이라는 인물이 등장하지만, 두 번째 이야기에서는 ‘푸름’과 ‘수진’이라는 인물이 등장하고 세 번째 이야기에서는 ‘차사장’과 관련된 인물

들이 등장한다. 이전 내용과 흐름이 달라 3가지 결과 가 같은 스토리를 이야기하는지 알아차리기 어렵다.

Table 6. The result of story generation through recursive input

Input 1	C1
Input 2	S1 + C2
Input 3	S2 + C3
Result	
<p>C1] 사람이 많은 도서관 사서실 자료실등을 돌아다니면서 자료들을 살펴 봤다. 에휴 하여튼 난 내일 드디어 개강이야 한숨을 푹푹 쉬어보았다 자료실에 도착하니 내 옆에 조용히 서있는 이 젊은 남자가 보였다 눈이 휘둥그레진 서영이 뭔가 말하려는데 갑자기 그 남자가 서영을 쳐다보았다.</p> <p>[C2] 책장 앞에 서있는 남녀 민우는 서영에서 눈을 떼지 못했다. “서영” 놀란 빛이 역력한 서영의 눈을 보며 민우가 인상을 썼다 그의 말에 서영의 눈썹이 꿈틀 움직였다 서영은 순간 망설이는 듯 눈을 깜빡였다.</p> <p>[C3] 식당에 앉아 있는 남녀 민우가 서영에게 말을 건넨다. “여기” 서영은 인상을 쓰며 고개를 끄덕였다. 뭐라고 대답하는 민우의 얼굴에 췌려왔다. 아무래도 도서관부터의 갑작스러운 만남이 달갑지 않은 서영이다.</p>	

세 번째 케이스는 본 논문에서 제안하는 스토리보드의 첫 번째 이미지 캡션으로 생성된 스토리와 두 번째 이미지 캡션을 함께 입력값으로 넣는 재귀적 접근 방안이다.

이전 결과들과 달리 table 6을 보면 3가지 스토리 속 주인공들의 이름이 동일하며 전후 문맥이 자연스러운 스토리를 생성하는 것을 확인할 수 있다. 그리고 “남녀 주인공이 도서관에서 만나 식사하러 간다.” 라는 실험 전 기획 의도와도 부합한다.

마지막으로 인공지능이 생성한 스토리에 작가가 새로운 이야기 흐름을 제시하고자 한다면, 세 번째 케이스를 기반으로 작가가 직접 개입하여 스토리를 완성하고 다음 학습 시, 입력값으로 반영하여 새로운 스토리를 생성한다(Table 7).

작가의 개입 유무의 연구 결과를 비교하였을 때, 작가의 개입을 통해 생성된 스토리가 전후 문맥이 더 자연스럽고, 로맨스 장르에 더 가까운 느낌이 드는 스토리였다.

본 논문에서는 길이가 긴 스토리를 생성함에도 기

획 의도에 부합하고 전후 문맥을 고려하여 자연스러운 흐름을 가진 스토리를 생성하였다.

Table 7. The result of story generation through recursive input including the author's involvement

Input 1	C1
Input 2	S1 + C2
Input 3	S2(the author intervened) + C3
Result	
<p>C1] 사람이 많은 도서관 사서실 자료실등을 돌아다니면서 자료들을 살펴 봤다. 에휴 하여튼 난 내일 드디어 개강이야 한숨을 푹푹 쉬어보았다 자료실에 도착하니 내 옆에 조용히 서있는 이 젊은 남자가 보였다 눈이 휘둥그레진 서영이 뭔가 말하려는데 갑자기 그 남자가 서영을 쳐다보았다.</p> <p>[C2] 책장 앞에 서있는 남녀 (KoGPT-2 모델이 생성한 부분) 민우는 서영에서 눈을 떼지 못했다. “서영” 놀란 빛이 역력한 서영의 눈을 보며 민우가 인상을 썼다 그의 말에 서영의 눈썹이 꿈틀 움직였다 서영은 순간 망설이는 듯 눈을 깜빡였다.</p> <p>(추가 - 작가가 개입한 부분) 잠깐의 정적이 흐르고 민우가 다시 입을 열었다. “같이 밥 먹으러 갈래?”</p> <p>[C3] 식당에 앉아 있는 남녀 아침부터 민우와 서영은 함께 식사를 한다. “저기” 민우의 말에 서영의 얼굴에는 웃음이 피어났다. 서영의 웃음을 보는 민우는 새삼 더 호기심이 생겼다. 그는 사람을 응시하며 고개를 끄덕였다.</p>	

V. 결 론

빅데이터와 딥러닝 네트워크의 융합으로 소설, 대본 등의 형식을 갖춘 문장을 생성하는 연구가 지속적으로 이어져 오고 있다. 기존 인공지능을 통해 문장을 생성하는 연구들은 얼마나 문장다운 문장을 생성할 수 있는지에 초점이 맞추어져 있었다면, 본 논문에서는 이미지 캡션 생성 기술을 활용하여 자연스러운 흐름을 가진 스토리를 생성하였다.

전후 문맥이 자연스러운 스토리를 생성하기 위해 순서대로 나열된 스토리보드 이미지를 활용하였고, 원하는 분위기의 글을 생성하기 위해 사용자가 원하는 장르의 글을 사전 학습한 KoGPT-2 언어모델을 사용하였다.

스토리보드의 첫 번째 이미지 캡션으로 생성된 글과 두 번째 이미지 캡션을 함께 입력값으로 넣는 재귀적 접근 방안을 제안하여 사용자가 기획한 줄거리와 스토리보드의 이미지 내용 흐름에 맞는 스토리를 생성하는 것을 확인할 수 있었다.

본 연구를 통해 인간의 전유물로 여겨지는 창조의 영역에서 인공지능을 활용하여 인간은 영감을 얻고, 기술은 인간의 도움을 얻어 공존할 수 있음을 확인하였다. 또한, 드라마, 영화 등 스토리를 기반으로 한 콘텐츠 제작의 전반적인 과정에 인공지능이 참여하여 미디어 지능화를 현실화 시킬 것으로 기대한다.

향후에는 KoGPT-2보다 방대한 한국어 데이터를 학습한 카카오브레인의 GPT-3 기반 KoGPT를 활용하여 조금 더 자연스러운 스토리를 생성하는 연구를 진행할 예정이며, 본 연구에서 제안한 모델을 기반으로 국내 환경에 적합한 고품질 이미지와 한국어 캡션 데이터를 구축하여 성능을 개선하고자 한다.

ACKNOWLEDGMENTS

본 논문은 2022년도 동서대학교 “Dongseo Cluster Project” 지원을 받아 수행된 것임 (DSU-20220002)

REFERENCES

[1] Kim, D. E, “Management System of Invasive Alien Species Threating Biodiversity in Korea and Suggestions for the Improvement”, *Journal of Environmental Impact Assessment*, 27(1), 33-55(2018). <https://doi.org/10.14249/EIA.2018.27.1.33>

[2] Ministry of Environment. 2014. 1st Management Plan of Alien Species (2014-2018).

[3] Da Bin Kim · Kyung Ah Koo. “A Study on the Current Status and Improvement of Ecosystem Disturbance Species”, 29(4), 59-81(2021).

[4] Ikeda T, Asano M, Matoba Y, Abe G, “Present status of invasive alien raccoon and its impact in Japan”, *Global environmental research*, 8(2), pp. 125-131, 2004.

[5] National Institute of Ecology, “Ecological studies of alien species (VI)”, Seocheon National Institute of Ecology

[6] Jihae Kim et al, “Comparison of Fine Grained Classification of Pet Images Using Image Processing and CNN”, *Journal of Broadcast Engineering (JBE)* Vol. 26, No. 2, pp. 175-183, 2021.

[7] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, Quoc V. Le; “Learning Transferable Architectures for Scalable Image Recognition”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8697-8710

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun; “Deep Residual Learning for Image Recognition”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778

[9] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, Zbigniew Wojna; “Rethinking the Inception Architecture for Computer Vision”, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818-2826

[10] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alexander A Alemi; “Inception-v4, inception-resnet and the impact of residual connections on learning”, *Thirty-first AAAI conference on artificial intelligence*. 2017.

[11] Simonyan, Karen, and Andrew Zisserman; “Very deep convolutional networks for large-scale image recognition”, *arXiv preprint arXiv:1409.1556* (2014).

[12] Kaiming He Ross Girshick Piotr Dollár, “Rethinking ImageNet Pre-training” *arXiv:1811.08833v1* (2018).

[13] Shorten, C., Khoshgoftaar, T.M., “A survey on Image Data Augmentation for Deep Learning”, *J Big Data* 6, 60 (2019). <https://doi.org/10.1186/s40537-019-0197-0>

[14] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database”, 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

저자소개

전 이 슬 (Isle Jeon)



2012년 2월 : 대구한의대학교
화장품약리학과 졸업(학사)
2021년~현재 : 동서대학교
컴퓨터공학과 석사 과정
관심분야 : 컴퓨터공학, 인공지능,
자연어처리 등

조 동 하 (Dongha Jo)



2017년 3월~현재 : 동서대학교
소프트웨어학과
관심분야 : 컴퓨터공학, 딥러닝,
인공지능 등

문 미 경 (Mikyeong Moon)



1992년 2월 : 이화여자대학교
전자계산학과 졸업(이학석사)
2005년 2월 : 부산대학교
컴퓨터공학과 졸업(공학박사)
2006~2008년 : 부산대학교
정보컴퓨터공학부 연구교수
2008년~현재 : 동서대학교
소프트웨어학과 교수

관심분야 : IoT 응용개발 방법, 소프트웨어 품질평가,
소프트웨어 프로덕트라인 개발방법 등