

유사구조 및 유사의미 문장 생성 방법

서혜인[○], 정상근^{*}, 정지수

충남대학교 컴퓨터융합학부

hyenee97@gmail.com, hugman@cnu.ac.kr, jisu.jung5@gmail.com

Semantic and Syntax Paraphrase Text Generation

Hyein Seo[○], Sangkeun Jung^{*}, Jeesu Jung

Department of Computer Science and Engineering, Chungnam National University

요 약

자연어 이해는 대화 인터페이스나 정보 추출 등에 활용되는 핵심 기술 중 하나이다. 최근 딥러닝을 활용한 데이터 기반 자연어 이해 연구가 많이 이루어지고 있으며, 이러한 연구에 있어서 데이터 확장은 매우 중요한 역할을 하게 된다. 본 연구는 자연어 이해영역에서의 말뭉치 혹은 데이터 확장에 있어서, 입력으로 주어진 문장과 문법구조 및 의미가 유사한 문장을 생성하는 새로운 방법을 제시한다. 이를 위해, 우리는 GPT를 이용하여 대량의 문장을 생성하고, 문장과 문장 사이의 문법구조 및 의미 거리 계산법을 제시하여, 이를 이용해 가장 유사하지만 새로운 문장을 생성하는 방법을 취한다. 한국어 말뭉치 Weather와 영어 말뭉치 Atis, Snips, M2M-Movie M2M-Reservation을 이용하여 제안방법이 효과적임을 확인하였다.

주제어: 자연어 이해, 문장 생성, 의미 유사도, 구조유사도

1. 서론

자연어 이해(Natural Language Understanding)란 인간이 발화하는 언어 현상을 기계적으로 분석해 컴퓨터가 이해할 수 있는 형태로 구조화하는 것을 말하며, 대화 인터페이스나 정보 추출 등에 활용되는 핵심 기술 중 하나이다. 자연어 이해는 자연어로부터 의미를 추출하고, 사용자의 의도를 유추하는 것을 목표로 한다.

본 연구에서는 자연어 이해영역의 데이터 확장(data augmentation) 문제에 있어서, 문장의 구조 및 의미 유사도를 반영하여 유사하지만 새로운 문장을 생성하는 방법을 소개한다.

최근 문장 생성 연구에 관한 다양한 연구가 진행되었다. 그중 Bidirectional Encoder Representations from Transformer(BERT)[1]와 Generative Pre-training(GPT)[2]는 텍스트 생성과 예측 분야에서 널리 사용되는 언어모델이다. 특히 GPT-2[2]는 유창하면서도 문법적인 문장을 생성해 높은 성능을 보인다.

또한, 자연어 문장의 의미를 벡터공간으로 프로젝션한 후, 자연어 처리를 수행하는 방식 연구가 진행되었는데, 이는 벡터공간 상에서 의미적으로 유사한 문장을 찾아낼 수 있다. 그중, [3,4]는 문장과 의미들은 서로 다른 표현 체계를 갖지만 두 표현 사이의 관계성인 ‘상호연관성(correspondence)’을 이용한 의미벡터 공간을 구성하는 연구를 제안하였고, 벡터공간 상에서 의미적으로 유사한 문장을 찾아낼 수 있음을 증명하였다.

본 논문에서는 주어진 자연어 이해 말뭉치로부터 문장의 구조 및 의미 유사도를 반영하여 유사하지만 새로운 문장을 생성하는 방법을 제안한다. 이를 위해 본 연구의 세부 목적은 다음과 같다.

1) 문장의 의미 혹은 구조 정보를 반영한 문장 생성 학습 방법을 제시한다.

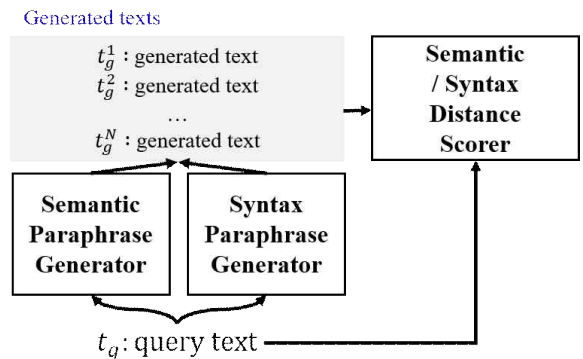


그림 1. 유사 문장 생성 프레임워크

2) 질의 문장과 생성된 문장과의 의미 유사도 및 구조 유사도 계산법을 소개한다.

3) 이를 순위화하여 상위 K개 문장과 질의 문장이 실제로 같은 의미 및 구조를 보이는지 확인한다.

그림 1은 본 연구의 핵심 아이디어를 그림으로 나타낸 것이다.

본 고의 구성은 다음과 같다. 2장에서는 기존 관련 연구를 소개하고, 3장에서는 유사 문장 생성 프레임워크에 대해 소개한다. 4장에서는 본 연구의 실험 구성과 결과를 소개하며, 마지막으로 5장에서는 결론을 소개한다.

2. 관련 연구

최근 문장 생성 연구에 관한 다양한 연구가 진행되었다. 그중 Bidirectional Encoder Representations from Transformer(BERT)[1]와 Generative Pre-training(GPT)[2]는 텍스트 생성과 예측을 위해 널리 사용되는 언어 모델이다. GPT를 활용한 다양한 문장 생성 연구가 진행되었는데, [5]는 GPT-2의 생성 기능을 활용하여 라벨링된 데이터를 어떠한 지도(supervision) 없이 패러프레이즈 데이

* 교신 저자(Corresponding Author)

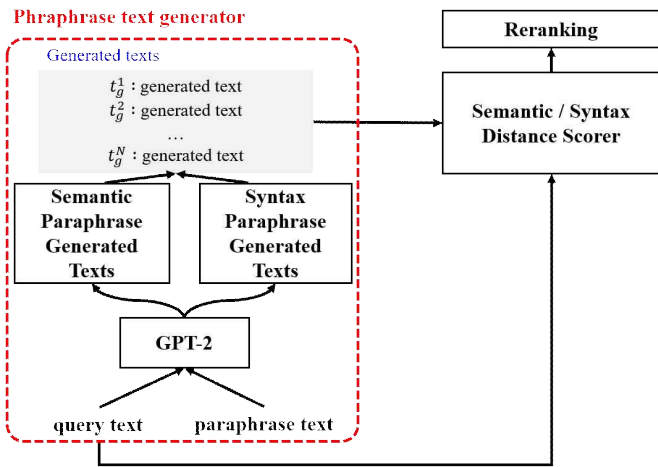


그림 2. 유사 문장 생성기, 벡터 거리 기반 유사도 측정, 재순위화 네트워크 구조

터를 생성하고, 이는 데이터 확장에 사용될 수 있음을 증명하였다.

또한, 데이터 확장(data augmentation) 기법 연구도 진행되었다. 데이터 확장은 컴퓨터 비전 등의 분야에서 활발하게 사용되고 있는 기법이다. 그러나 이미지를 뒤집거나, 방향을 살짝 트는 등의 변형을 가해도 라벨 값에 큰 영향을 미치지 않는 비전 분야와 달리 텍스트에서는 단어 하나만 바뀌어도 문장의 의미가 달라지기 때문에 데이터 확장을 적용하기 쉽지 않았다. EDA[6]는 자연어 처리에서 쉽게 떠올릴 수 있는 데이터 확장 기술이 실제로 성능 향상에 유의미한 영향을 미칠 수 있으며, 데이터 확장 과정을 통해 생성된 문장들이 기존 라벨의 성질을 잘 따른다는 연구를 소개하였다. 또한, 문장으로부터 약간의 구문 변환을 통해 새로운 훈련 데이터 확장 기법 연구도 있었다[7]. 이러한 연구들은 보유한 데이터에 적절한 변경을 가해 새로운 데이터를 얻어 학습 데이터의 양을 증가시키는 방식을 취했다.

3. 유사 문장 생성 프레임워크

본 연구에서는 주어진 자연어 이해 말뭉치로부터 문장을 생성하고, 생성된 문장 중 의미와 구조가 모두 유사한 문장을 추려내는 방법에 대해 제안한다. 본 논문에서 제안하는 프레임워크는 유사 문장 생성기(Paraphrase text generator), 벡터 거리 기반 유사도 측정, 재순위화(Reranking) 방법론으로 구성된다. 그림 2는 전체적인 학습 네트워크 구조이다.

1. 유사 문장 생성기
2. 벡터 거리 기반 유사도 측정
3. 재순위화 방법론

각 방법론에 대한 자세한 설명은 3.1과 3.2, 3.3에 기재하도록 한다.

3.1 유사 문장 생성기(Paraphrase text generator)

유사 문장 생성기(Paraphrase text generator)는 입력된 질의 문장(query text)과 유사 문장(paraphrase text)을 통해 의미(semantic) 또는 구조(syntax)가 같은

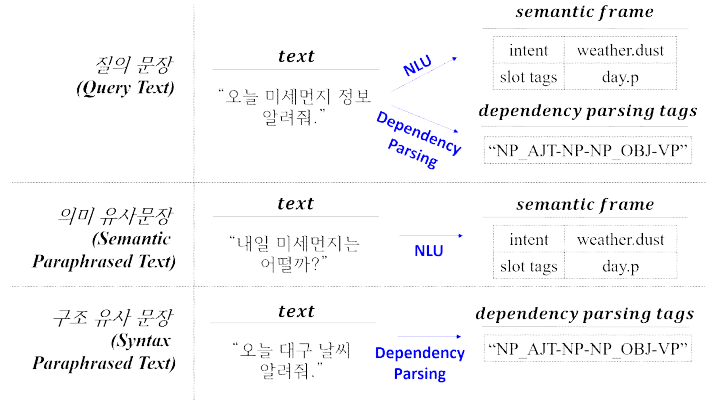


그림 3. 질의 문장에 따른 의미 및 구조 유사 문장

문장들을 생성한다. 문장 생성기의 입력은 질의 문장과 유사 문장으로 구성되며, 이 둘을 매핑한 $P(\text{paraphrase text} | \text{query text})$ 값이 최대화할 수 있는 모델을 학습한다. 문장 생성기의 입력인 유사 문장은 아래와 같이 2가지 유형에 대해 모델링 시도하였다:

1. 의미 유사 문장(Semantic paraphrased text)
2. 구조 유사 문장(Syntax paraphrased text)

일반적으로 패러프레이징(paraphrasing)이란 같은 의미지만 다르게 표현된 문장을 뜻하지만, 본 논문에서는 질의 문장으로부터 의미가 유사한 문장 또는 구조가 유사한 문장들을 패러프레이징된 문장이라 표현하였고, 표현의 자연스러움을 위해 ‘유사’로 의역하였다. 또한, 유사성의 범위를 문장의 의미 및 문법구조로 한정 지었는데, 의미 유사 문장의 경우, 훈련 데이터에서 문장의 의미정보를 나타내는 intent와 slot tags가 같은 문장끼리 짝지어 (query text, semantic paraphrased text)의 형태로 구성된다. 구조 유사 문장의 경우, 문장의 문법구조가 같은 문장끼리 짝지어 같은 형태로 구성되는데, 본 연구에서는 문장의 문법구조를 나타낼 수 있는 형태로 ‘의존 구문분석(Dependency parsing)’을 활용하였다. 의존 구문분석을 통해 문장의 문법적 구조를 나타내고, 단어들 사이의 관계를 정의하여 문장의 의존 관계를 추출하였다.

예를 들어, 그림 3에서 질의 문장인 “오늘 미세먼지 정보 알려줘.”의 intent는 weather.dust, slot tags는 day.p임을 알 수 있다. 여기서 day.p는 ‘오늘’, ‘내일’과 같은 일반적인 day에 대해 지칭하는 대명사를 뜻한다. 또한, 이 문장에 대해 의존 구문분석 결과로 나오는 태그 셋을 문장 형태로 표현하면 “NP_AJT-NP-NP_OBJ-VP”가 된다. 따라서, 의미 유사 문장은 intent와 slot tags가 같은 “내일 미세먼지는 어떨까”가 될 수 있고, 구조 유사 문장은 “오늘 대구 날씨 알려줘”가 될 수 있음을 확인할 수 있다.

BERT[1]와 GPT-2[2] 모두 텍스트 생성 및 예측 분야에서 뛰어난 성능을 보이지만, 본 논문에서는 GPT-2[2]를 사용하여 문장을 생성하였으며, 한국어 말뭉치의 경우 KoGPT-2[1], 영어 말뭉치는 GPT-2 사전학습 언어모델을 활용하여 구현하였다. 또한, 의미 유사 문장과 구조 유사 문장으로 각각 100문장씩 생성하였다.

1) <https://github.com/SKT-AI/KoGPT2>

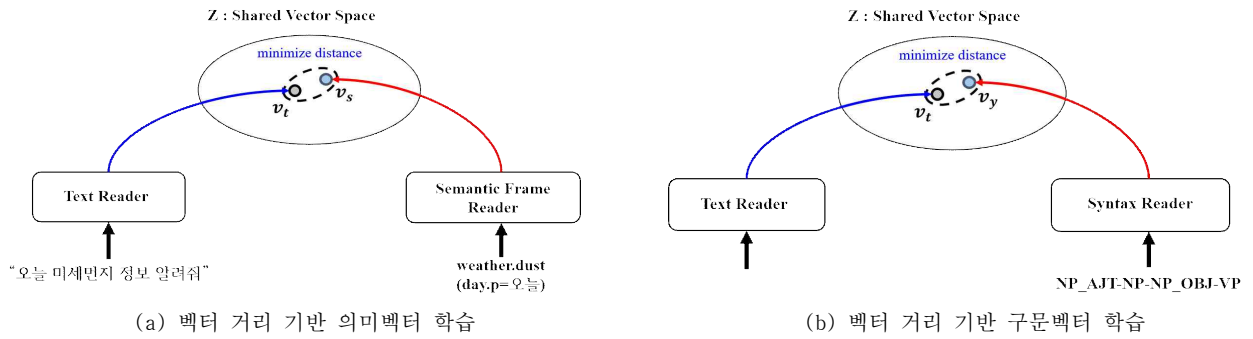


그림 4. 벡터 거리 기반 학습

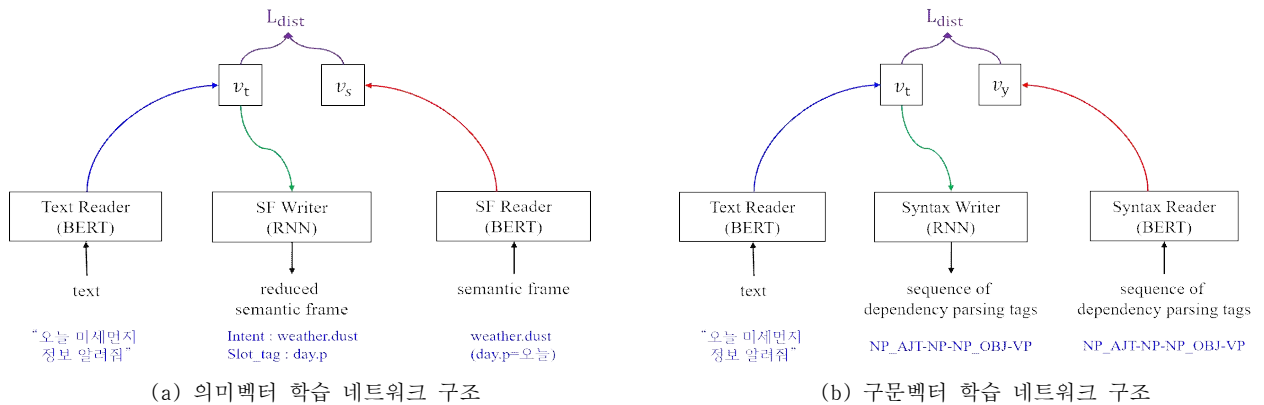


그림 5. 의미벡터 및 구문벡터 학습 프레임워크

3.2 벡터 거리 기반 유사도 측정

본 논문에서는 질의 문장과 생성된 문장 사이의 유사도를 측정하기 위해 문장을 벡터공간으로 프로젝션한 후, 두 문장 벡터 사이의 거리를 유사도로 표현하였다. 이는 문장의 의미 및 구조적 특징에 기반한 평가 방식에 해당된다.

의미유사도 측정은 [3,4]의 의미벡터 학습 방법 (Semantic Vector Learning)을 활용하였다. 또한, 이를 문장의 문법구조에 대해 학습할 수 있도록 확장하여 구문벡터 학습 프레임워크(Syntax Vector Learning Framework)를 구현하였고, 이를 통해 구조유사도를 측정하였다.

[3,4]는 문장과 의미틀(semantic frame)은 서로 의미적으로 상응해야 한다는 ‘상호연관성’을 벡터공간 상에서의 거리로 측정하고, 학습 방법을 제시하였다. 여기서 의미틀이란 사람이 읽고 의미를 부착할 수 있는 형태의 의미기호를 담고 있는 틀을 말한다. 따라서 그림 4-a와 같이 문장으로부터 투사된 벡터 v_t 와 의미틀로부터 투사된 v_s 는 벡터 공간상에서 가까워져야 함을 의미한다. 마찬가지로 그림 4-b는 문장으로부터 투사된 벡터 v_t 와 문장의 문법구조로부터 투사된 v_y 가 벡터 공간상에서 서로 가까워져야 함을 의미한다.

그림 5-a는 [4]의 의미벡터 학습 프레임워크이며, 본 연구에서 문장 읽기장치(Text Reader), 의미틀 읽기장치(Syntax Frame Reader)는 모두 BERT-base-cased 사전학습 언어 모델을 활용하여 구현하였다. 구문벡터 학습 프레임워크의 경우 의미벡터 학습 프레임워크를 확장하여 문장의 ‘문법 구조(syntax structure)’를 학습할 수 있도록 의미를 읽기장치를 문법구조 읽기장치(Syntax Reader)로 확장하였으며, 문법구조 읽기장치의 입력은 의존 구문분석 결과의 태그열로 사용하였다. 구문벡터 학습 프레임워크의 문장 읽기장치와 문법구조 읽기장치 또한 BERT-base-cased 사전학습 언어 모델을 활용하여 구현하였다.

각 프레임워크에 대한 학습 방법 및 손실함수는 [4]에서 제안한 방법을 활용하였다. 그림 5-a와 같이 문장 읽기장치로부터 투사된 벡터 v_t 와 의미를 읽기장치로부터 투사된 벡터 v_s 는 벡터 공간상에서 서로 가까워져야 한다. 이를 거리비용이라 하고, 아래와 같이 정의한다.

$$L_{dist} = dist(v_t, v_y)$$

이때 다양한 벡터 간 거리함수를 사용할 수 있는데, 본 연구에서는 cosine 거리함수를 사용하였다. 구문벡터 학습 프레임워크의 학습 방법 및 손실함수 또한 의미벡터 학습 프레임워크와 같은 방식으로 구현하였다.

유사한 의미를 갖는 문장들은 v_t 를 기준으로 서로 가깝게 뭉쳐지기 때문에, 거리 기반으로 유사한 의미를 갖는 문장을 검색할 수 있다. 본 논문에서는 질의 문장과 생성된 문장들 사이의 벡터 거리를 측정하여 가장 가까운 문장을 선정하였고, 이 두 문장 사이의 거리를 유사도로 사용하였다. 구문벡터 학습 프레임워크에 대해서도 같은 방식으로 계산하였다. v_t 를 기준으로 문법구조가 유사한 문장들이 벡터공간 상에 가까이 위치하게 된다.

표 1. 실험 말뭉치들의 특성

Dataset	Weather	Atis	Snips	M2M-M	M2M-R
# of training	6,993	3,621	12,658	3,321	10,653
# of testing	2,998	753	685	2,291	5,707
# of intent tags	13	22	7	21	26
# of slot-tags	14	79	39	5	9
# of dependency parsing tags	29	43	43	47	47

3.3 재순위화(reranking) 방법론

질의 문장과 생성된 문장과의 벡터 거리 기반 유사도를 측정하면, 문장별 의미 유사도와 구조유사도가 도출된다. 본 논문에서는 두 유사도를 모두 고려해 재순위화(reranking) 하였고, 그 결과로 상위 K개의 문장을 추출하여 질의 문장과 생성된 문장과의 유사도를 측정하였다. 본 연구에서는 생성된 문장 중 상위 10개의 문장만을 활용하였다.

4. 실험

본 연구의 목적은 GPT로부터 대량의 문장을 생성하고, 생성된 문장 중 의미와 구조가 모두 유사한 문장을 추려내는 것이다. 본 논문에서 제안한 유사 문장 생성 프레임워크의 타당성을 검증하기 위해 생성된 문장이 실제로 의미 및 구조적으로 유사한지 확인해보았고, 생성된 문장 중 비문의 비율과 유형에 대해 살펴보았다.

4.1 실험 구성

먼저 GPT-2를 통해 의미 유사 문장과 구조 유사 문장에 대해 각각 100문장씩 생성하였고, 이를 벡터공간 상의 문법구조 및 의미 거리를 이용해 질의 문장과 유사하지만 새로운 문장을 분석하였다. 새로운 문장을 분석하기 위해 생성된 200문장 중 이미 훈련 데이터에 있던 문장은 제외하고 실험을 진행하였다.

본 논문에서 제안한 방법론인 유사구조 및 유사의미 문장 생성 기법의 타당성을 한국어 기반 Weather 말뭉치와 영어 기반 Atis¹⁾ [8], Snips²⁾ [9], Machines Talking To Machines(M2M-Movie³⁾ [10], M2M-Restaurant [10] 말뭉치를 이용하여 검증하였다. 각 말뭉치의 특성은 표 1에서 살펴볼 수 있다.

한국어 말뭉치의 경우 KoGPT-2 사전학습 언어모델 활용하였고, 영어 말뭉치의 경우 GPT-2 사전학습 언어모델을 활용하여 구현하였다. 의미 유사 문장 생성기(semantic paraphrase text generator)의 입력은 질의 문장과 유사 의미 문장이 쌍을 지어 (질의 문장, 유사 의미 문장)의 형태로 구성되고, 구조 유사 문장 생성기(syntax paraphrase text generator)의 경우 (질의 문장, 유사 구조 문장)의 형태로 GPT-2의 입력으로 들어가게 된다. GPT-2 모델의 레이어는 따로 조정하지 않고, 그대로 사용하였다.

1) <https://github.com/yvchen/JointSLU/tree/master/data>

2) <https://github.com/snipsco/snips-nlu>

3) <https://github.com/google-research-datasets/simulated-dialogue>

표 2. Weather 도메인에서의 유사 문장 생성 결과 예시

질의 문장	Top	생성 문장
내일 온도 알려줄래?	1	내일모레에 덥니?
	2	내일모레에 더워?
	3	오늘 기온 알려줘.
	4	오늘 온도 알려줄래?
	5	내일모레에 따듯하니?

표 3. Atis 도메인에서의 유사 문장 생성 결과 예시

질의 문장	Top	생성 문장
i need a flight from charlotte to miami	1	i need a flight tomorrow from philadelphia to
	2	i want a flight from ontario to chicago on cont
	3	what is the distance between pittsburgh
	4	what is the twapacity of airport
	5	what 's the capacity of airport

벡터공간 상의 문법구조 및 의미 거리를 측정하기 위한 의미벡터 학습 프레임워크(그림 5-a)와 구문벡터 학습 프레임워크(그림 5-b)의 구현은 BERT 사전학습 언어 모델을 활용하여 구현하였다. 한국어 말뭉치의 경우 BERT-base-multilingual-cased를, 영어 말뭉치의 경우 BERT-base-cased를 활용하여 구현하였으며, 별도의 어휘 처리는 하지 않고 일반 자연어 말뭉치를 통해 학습된 어휘를 그대로 사용하였다.

4.2 실험 결과

문장 생성 결과 표 2와 3은 Weather와 Atis 영역에서 벡터 공간상의 문법구조 및 의미 거리를 이용한 유사 문장 생성 결과 예시를 나타낸다. 질의 문장은 테스트 데이터 예제이며, Top은 질의 문장으로부터 문법구조 및 의미 거리가 가까운 문장 5개를 나타낸 것이다. 표 2에서 질의 문장인 “내일 온도 알려줄래”의 intent는 weather.temperature로 날씨 중 온도를 나타내는 문장이다. 이 문장으로부터 생성된 의미와 구조가 모두 유사한 문장 5개를 확인해보면 모두 기온 또는 온도에 관한 문장임을 알 수 있고, 문장의 문법구조 또한 의문문으로 끝나거나, “내일 온도 알려줄래”와 같이 “명사-명사-동사”의 구조를 가짐을 확인할 수 있다. 표 3을 보면, “i need a flight from charlotte to miami”의 intent는 “flight”로 기본적인 항공권 예약에 대한 문장임을 알 수 있다. 이 문장으로부터 생성된 문장을 확인해보면, 항공권 예약에 대한 문장이거나, 지역 간 거리 또는 공항의 크기에 대한 문장임을 알 수 있다. 또한, 문장의 문법구조를 살펴보니 Top 1과 Top 2의 경우 질의 문장과 상당히 유사한 문법구조를 가짐을 확인할 수 있다.

본 논문에서 제안한 유사 문장 생성 기법을 통해 의미와 문법구조가 유사하지만, 훈련 데이터에 없는 새로운 문장 생성 결과를 보였다. 그러나 생성된 문장이 완전한 문장이라고 보장할 수 없다. 특히 표 3에서 확인할 수 있듯이, 문장이 불완전하거나 단어의 철자에 오류가 있는 경우도 발견되어 이에 대해 분석해보았다.

표 4 말뭉치별 문장 분석 통계(600문장 샘플링 후 분석)

Dataset	Good Sentence	Incorrect Sentence				
		오생성 비율	UNK	반복 생성	비문	부분 출력
Weather	87.1%	12.9%	0.5%	1.8%	9.8%	0.8%
Atis	75.8%	24.2%	-	2.2%	3%	19%
Snips	60.5%	39.5%	1.3%	3.7%	4.6%	29.8%
M2M-M	69.2%	30.8%	-	2.6%	6.2%	22%
M2M-R	76%	24%	-	1.6%	3.2%	19.2%

표 5. Weather 도메인에서의 오생성 문장 타입 별 예제

오생성 유형	Incorrect Sentence
UNK	오늘 해가 <unk>까? 내일도 햇빛 <unk>하지? 내일모레 철원군의 적설량 얼마나 되 되는지 좀 찾아줘
반복 생성	내일모레 충남대의 적설량이 얼마나 될지 좀 될지 좀 말해줄 수 있어? 자외선량이 내일은을까?
비문	내일모레 춘천의 기상정보 좀 찾아와지 알고있니?
부분 출력	내일모레 장안원의 적설량 좀 알아봐줄 수 다음 주 목요일에 구미시 비 얼마나.

표 4는 말뭉치별 문장 분석 통계를 나타낸다. 600문장만 샘플링 후 분석해보았는데, 여기서 Good Sentence는 정상적으로 생성된 문장의 비율을 나타낸다. 오생성 문장의 경우 4종류(UNK, 반복생성, 비문, 부분 출력)로 분류하였고, 각각의 통계는 표 4의 Incorrect Sentence에서 확인할 수 있다. 영어 말뭉치의 경우 한국어 말뭉치보다 문장 길이가 길어짐에 따라 반복생성 및 부분 출력 현상이 높게 나타남을 확인할 수 있다. 오생성 문장 타입 별 예제는 표 5에서 확인할 수 있다.

본 연구에서는 새로운 데이터 생성에 집중하였지만, 향후 연구로 오생성 문장 비율을 줄이기 위한 기제가 필요할 것으로 확인된다.

5. 결론

본 연구에서는 주어진 자연어 말뭉치를 통해 문장을 생성하고, 생성된 문장 중 의미와 구조가 모두 유사한 문장을 추려내는 방법에 대해 제안하였다. 문장 생성을 위해 GPT-2 사전학습 언어모델을 활용하였으며, 질의 문장과 생성된 문장 사이의 유사도를 측정하기 위해 연관성 학습을 통한 의미벡터 학습 모델을 사용하였다. 벡터 공간상에서 문장간 문법구조 및 의미 거리 측정을 통해 질의 문장과 문법구조와 의미가 모두 유사하지만, 훈련 데이터에는 없는 새로운 문장을 생성할 수 있음을 보였다.

제안한 연구 결과에 기반하여 향후 문장의 의미 및 구조적 특징을 활용한 시각화 방법과 이에 기반하여 생성 문장 선별 보조 장치로 활용할 수 있는 연구들이 진행될 수 있을 것으로 기대한다.

감사의 글

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임

(No.2020-0-01441)

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2019-0-00004, 준지도학습형 언어지능 원천기술 및 이에 기반한 외국인 지원용 한국어 튜터링 서비스 개발)

이 논문은 2019년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2019R1F1A1060601)

참고문헌

- [1] DEVLIN, Jacob, et al, Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
- [2] RADFORD, Alec, et al, Language models are unsupervised multitask learners, OpenAI Blog, 1.8: 9, 2019.
- [3] JUNG, Sangkeun; LEE, Jinsik; KIM, Jiwon. Learning to embed semantic correspondence for natural language understanding, In: Proceedings of the 22nd Conference on Computational Natural Language Learning, p. 131-140, 2018.
- [4] JUNG, Sangkeun. Semantic vector learning for natural language understanding. Computer Speech & Language, 56: 130-145, 2019.
- [5] HEGDE, Chaitra; PATIL, Shrikumar. Unsupervised Paraphrase Generation using Pre-trained Language Models. arXiv preprint arXiv:2006.05477, 2020.
- [6] WEI, Jason; ZOU, Kai. Eda: Easy data augmentation techniques for boosting performance on text classification tasks, arXiv preprint arXiv:1901.11196, 2019.
- [7] MIN, Junghyun, et al. Syntactic data augmentation increases robustness to inference heuristics, arXiv preprint arXiv:2004.11999, 2020.
- [8] Patti J Price, "Evaluation of spoken language systems: The atis domain," In Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990.
- [9] Alice Coucke, Alaa Saade, Adrien Ball, Th´eodore Bluche, Alexandre Caulier, David Leroy, Cl´ement Doumouro, Thibault Gisselbrecht, Francesco Calta girone, Thibaut Lavril, et al, Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. arXiv preprint arXiv:1805.10190, 2018.
- [10] Pararth Shah, Dilek Hakkani-Tür, Gokhan Tür, Abhinav Rastogi, Ankur Bapna, Neha Nayak, Larry Heck, Building a Conversational Agent Overnight with Dialogue Self-Play, arXiv preprint arXiv:1801.04871, 2018.