

키워드를 이용한 패러프레이즈 문장 생성

홍성태^o, 차정원
창원대학교

{ghdchlws123, jcha}@changwon.ac.kr

Paraphrase sentence generation using keywords

Seong-Tae Hong^o, Jeong-Won Cha
Changwon National University

요 약

패러프레이즈 연구를 위한 필수요소는 동일한 의미를 가지지만 표현이 다른 두 문장(입력 문장, 출력 문장) 쌍의 유무이다. 하지만 패러프레이즈 연구를 위한 공개된 학습 코퍼스의 수가 여전히 부족하고, 두 문장 쌍에 대한 표현들이 다양하지 않다는 문제점이 존재한다. 이러한 문제점을 해결하기 위해서 문장에서 키워드를 추출하여 문장을 생성하는 패러프레이즈 모델을 제안한다. 실질 형태소를 이용한 키워드 추출을 적용하였다. 키워드 입력을 두가지 방식(순서유지, 순서 변경)으로 적용한 실험을 통해 생성된 패러프레이즈 문장이 다양한 표현의 문장을 생성한 결과를 보였음을 확인하였다. 이러한 기법이 챗봇에서 사용자의 의도 탐지 및 MRC와 같은 자연어 처리의 여러 분야에 유용한 자원으로 사용될 것이다.

1. 서론

다양한 발화를 모델링하는 연구는 자연어 처리 분야에 꾸준히 지속되어 왔다. 최근 딥러닝 및 기계학습이 다양한 분야에 적용되고 있으며, 특히 챗봇이나 AI 고객 센터 등의 기술에 적용되고 있다. 챗봇 또는 AI 고객센터는 문장의 의미를 파악하고 고객이 문의한 항목에 정확한 답변을 제공함으로써 대기 시간을 단축할 수 있다는 장점을 가진다. 이러한 장점을 내세우기 위해서는 사용자들의 질문이나 의도 파악에 있어 동등한 의미 콘텐츠를 식별하고 생성하는 것은 중요한 기술이고, 다양한 발화를 모델링하는 요구는 자연어 처리 분야에서 꾸준히 있었으며 단어, 구 또는 문장과 동등한 의미 콘텐츠를 자동으로 식별하고 생성하는 것은 자연어 처리의 중요한 부분이다.

패러프레이즈의 사전적 정의는 더 명확하게 하기 위해 다른 단어 및 구를 사용하여 발표자 또는 글로 쓰거나 말한 것의 의미를 표현하는 것이다. 패러프레이즈 연구에 대한 공개된 학습 코퍼스의 수가 여전히 부족하며, 이를 보완 및 구축하기 위해서는 시간과 비용이 많이 소요된다는 문제점이 존재한다. 그리고 패러프레이즈를 유의어 및 동의어 등의 어휘로 대체하여 코퍼스를 구축하는 경우 단순히 단어 치환만 적용된 경우들이며, 특히 단어가 구로 치환되거나 구조가 치환되는 경우에 해당하는 패러프레이즈 표현들이 다양하지 않다는 문제점 또한 존재한다.

본 연구에서는 이러한 문제점을 해소하기 위해서 문장에서 키워드를 추출하여 문장을 생성하는 패러프레이즈 모델을 제안한다. 해당 모델은 시퀀스-투-시퀀스를 학습하기 위한 문장 쌍이 필수적이지 않아 한 문장만으로도 학습이 가능하다는 장점을 가지고 있다. 패러프레이즈 데이터셋을 이용하여 모델의 효용성에 대한 검증을 보인다.

2. 제안 방법

본 논문에서 제안한 패러프레이즈 생성 시스템은 그림 1과 같이 문장에서 키워드를 추출하여 문장을 생성하는 학습을 적용한다.

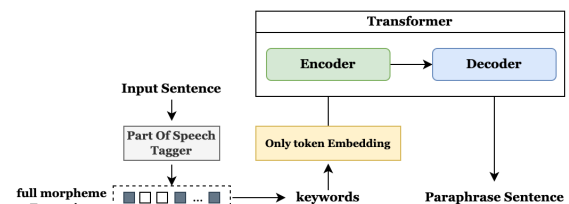


그림 1: 키워드를 이용한 패러프레이즈 문장 생성 모델 구조도

해당 모델은 KE-T5[1]를 사용하였으며, 패러프레이즈 데이터셋으로 미세조정(fine-tuning)을 진행하였다. 키워드를 입력으로 사용하여 패러프레이즈 문장을 생성하기 때문에 KE-T5에서 입력 토큰에 상대적 위치 인코딩(relative position encoding)을 제외하였다.

2.1 문장 내 키워드 추출

입력으로 사용하는 키워드를 의미를 가지는 최소한의 단위인 ‘형태소’로 설정한다. 형태소 분석기는 창원대학교의 Espresso[2]를 사용하였으며, 형태소는 ‘실질 형태소’와 ‘형식 형태소’ 2가지로 분류할 수 있다. 실질 형태소는 어휘적 의미를 가지고 있는 형태소이고, 형식 형태소는 문법적 의미를 가지고 있는 형태소이다. 조사, 접사(접두사, 접미사), 어미(선어말어미, 어말 어미) 등과 같은 형식 형태소는 제외하고, 실질 형태소만 사용하도록 키워드를 추출하였다. 실질 형태소 중 키워드 추출에 사용되는 형태소는 표 1과 같다. 금액 관련 및

날짜 키워드 등의 숫자와 외국어는 문장 내 필요한 키워드라고 판단하여 실질 형태소로 추가하였다. 그리고 용언에 표시된 결합기호인 ‘-’는 제외하고 키워드로 사용하였다.

표 1: 키워드 추출 품사 태그 정보

대분류	소분류	세분류	예시
체언	명사	일반명사(NNG)	성인, 화재
		고유명사(NNP)	한국, 중국
	수사	수사(NR)	만, 억
용언	동사	동사(VV)	먹-, 되-
	형용사	형용사(VA)	같-, 없-
	보조용언	보조용언(VX)	싶-, 있-
	지정사	부정지정사(VCN)	아니
의존형태		어근(XR)	모호, 과도
기호		외국어(SL)	EBS, SNS
		숫자(SN)	1,2

3. 실험 및 성능

3.1 학습 코퍼스

패러프레이즈 데이터는 국립국어원에서 공개한 모두의 말뭉치의 유사 문장 말뭉치를 사용하였으며 학습데이터는 218,653개, 검증 데이터 27,332개, 평가 데이터 27,360개를 사용하였다.

3.2 실험 환경 설정

키워드를 이용한 패러프레이즈 문장 생성 시스템에서 사용한 모델은 KE-T5 Base이다. KE-T5 Base 모델의 멀티헤드 수는 12개이고, 레이어 수는 12개이며, 768차원의 벡터를 가지며, 사전 단어의 크기는 64,128개이다. 실험을 위한 모델의 하이퍼 파라미터는 배치 사이즈 128, 학습률 0.0001, 드롭 아웃 0.1, 입력 최대 길이 16, 출력 최대 길이 32, 조기 학습 종료 조건은 검증 데이터의 손실 함수를 3번 갱신하지 못하는 경우로 설정하였다.

3.3 실험 결과 및 분석

표 2는 키워드를 이용한 패러프레이즈 문장 생성 결과 성능을 보여주는 표이고, 표 3은 모델의 결과를 보여주는 예시이다.

표 2: 키워드를 이용한 패러프레이즈 문장 생성 결과 성능

평가방법	키워드 순서 유지	키워드 순서 변경
BLEU-1[3]	0.6515	0.4540
BLEU-2	0.5353	0.2907
BLEU-4	0.4566	0.2056
ROUGE-1[4]	0.7008	0.5566
ROUGE-2	0.5714	0.3514
ROUGE-L	0.6773	0.4503
CIDER[5]	0.5110	0.2695
chrF[6]	0.6870	0.5140

표 3의 ‘키워드 순서 유지’ 결과를 통해 실질 형태소가 추출된 순서를 따라하여 문장을 생성하는 경향을 보이는 것을 확인할 수 있었다. 본 연구에서 목표로 하고 있는 패러프레이즈 문장의 다양한 표현에 속하는 ‘문장 구조 치환’ 능력이 저하되는 것 또한 확인하였다.

이러한 문제점을 해소하기 위해서 입력에 나타나는 실질 형태소들의 순서를 섞어 더 다양한 문장 표현들이 나타나는 것을 목표로 추가 실험을 진행하였다.

표 2의 ‘키워드 순서 변경’은 키워드 순서를 변경한 패러프레이즈 문장 생성 결과 성능을 보여주는 표이고, 표 3에서 ‘키워드 순서 변경’ 모델의 결과를 확인할 수 있는 예시들이 있다. 표 3 내 (1)과 (2)의 원문장은 ‘키워드 순서 유지’의 원문장과 동일하다.

표 4: 정성 평가 결과

분류 기준	키워드 순서 유지	키워드 순서 변경
입력과 동일한 문장	18(18%)	7(7%)
에러문장	19(19%)	24(24%)
일부(어미, 조사 등)만 변경된 문장	61(61%)	43(43%)
문장구조 변경 문장	2(2%)	26(26%)

표 4는 각 항목별 무작위로 추출한 100개의 문장을 2명의 평가자에게 동일하게 평가 받아 평균을 낸 결과이며 각 무작위 추출된 항목별 원문장은 동일하다. 표 4에서 키워드 순서 변경 실험의 결과 입력 키워드 순서의 변경으로 인해 키워드 사이에 적절한 문장 형식을 만들어내지 못하는 경우가 5%p 증가하였으나 동일한 문장을 생성하는 경우는 11%p 감소하였고 정답 문장과 다른 문장 구조를 가지는 형태적 변경이 24%p 증가하였다.

그에 비해 키워드 순서 유지 실험은 동일한 문장이 조금 더

표 3: 생성된 패러프레이즈 문장 예시

키워드 순서	구분	예시
유지	입력 예측	사회심리학 학문 문화 제도 규범 인간 심리 미치 영향 대하 연구 사회심리학이라는 학문은 문화, 제도, 규범 등이 인간 심리에 미치는 영향에 대해 연구한다.
변환	입력 예측	대하 미치 사회심리학 심리 연구 규범 문화 제도 영향 학문 인간 사회심리학은 문화, 제도, 규범 등이 인간 심리에 미치는 영향에 대해 연구하는 학문이다.
원문장		사회심리학이라는 학문은 문화, 제도, 규범 등이 인간 심리에 미치는 영향에 대해서 연구한다.
유지	입력 예측	녹차 세포 손상 유해산소 제거 주 심혈관 질환 예방 효과 있 또한 녹차는 세포를 손상시키는 유해산소를 제거해 주기 때문에 심혈관 질환을 예방하는 효과도 있다.
변환	입력 예측	주 손상 유해산소 효과 질환 녹차 심혈관 예방 제거 세포 있 또 심혈관 질환을 예방하는 효과도 있는데 녹차가 세포를 손상시키는 유해산소를 제거해 주기 때문이다.
원문장		또한 녹차가 세포를 손상시키는 유해산소를 제거해 주기 때문에 심혈관 질환을 예방하는 효과도 있다.

많았고 주로 문장구조 변경보다는 일부만 변경된 문장을 생성하는 결과를 보였다. 정량평가 지표 결과에서는 키워드 순서 변경이 실제 정답 문장과 비교하기 때문에 키워드 순서 유지에 비해 낮은 성능을 보였으나 정성평가 결과 문장 구조가 변경된 문장이 많이 생성되었고, 정답 문장과 다른 문장을 생성하는 결과를 보였다. 이는 정량평가 성능이 낮은 이유에 대한 근거가 된다.

4. 결론 및 정리

본 논문에서는 문장에서 키워드를 추출하여 문장을 생성하는 패러프레이즈 모델을 제안하였고, 결과 단일 문장만으로도 패러프레이즈 문장을 생성할 수 있다는 장점을 가진다. 제안한 모델에 대한 효용성을 검증하기 위해서 공개된 패러프레이즈 데이터셋을 이용하였다.

키워드 순서 유지 실험 결과 정량적 평가는 BLEU-4 기준 0.4566, ROUGE-L 0.6773의 성능을 보였으나 정성적 평가 결과 입력과 동일한 문장 혹은 일부만 변경된 문장을 생성함을 확인하였다. 문장 구조 치환 능력이 현저히 떨어지는 결과를 보였으며, 키워드 순서 변경 실험은 키워드 순서 유지 실험에 비해 구조가 변경된 문장을 출력하는 경우가 24%p 증가하였다. 진행한 실험들의 분석 결과, 키워드 유지 실험에 비해 다양한 표현 및 구조를 가진 문장들을 출력하는 것을 확인할 수 있었다.

Acknowledgement

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00354, 비정형 텍스트를 학습하여 쟁점별 사실과 논리적 근거 추론이 가능한 인공지능 원천기술)

참고 문헌

- [1] K. AIRC, “Ke-t5: Korean english t5,” Mar. 2021.
- [2] A. L. changwon national university, “Espresso 2: United korean language understanding engine.”
- [3] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: A method for automatic evaluation of machine translation,” in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02*, (USA), p. 311–318, Association for Computational Linguistics, 2002.
- [4] C.-Y. Lin, “ROUGE: A package for automatic evaluation of summaries,” in *Text Summarization Branches Out*, (Barcelona, Spain), pp. 74–81, Association for Computational Linguistics, July 2004.
- [5] R. Vedantam, C. Lawrence Zitnick, and D. Parikh, “Cider: Consensus-based image description evaluation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [6] M. Popović, “chrF: character n-gram f-score for automatic mt evaluation,” in *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pp. 392–395, 2015.