

마스크 언어 모델 기반 비병렬 한국어 텍스트 스타일 변환

배장성⁰¹ 이창기¹ 황정인² 노형종²

¹강원대학교 컴퓨터학과

²엔씨소프트 NLP Center Language AI Lab

jseffort88@gmail.com, leeck@kangown.ac.kr, {jihwang, nohhj0209}@ncsoft.com

Unpaired Korean Text Style Transfer with Masked Language Model

Jangseong Bae⁰¹ Changki Lee¹ Hyungjong Noh² Jeongin Hwang²

¹Department of computer Science, Kangwon National University

²NLP Center Language AI Lab, NCSOFT

요약

텍스트 스타일 변환은 입력 스타일(source style)로 쓰여진 텍스트의 내용(content)을 유지하며 목적 스타일(target style)의 텍스트로 변환하는 문제이다. 텍스트 스타일 변환을 시퀀스 간 변환 문제(sequence-to-sequence)로 보고 기존 기계학습 모델을 이용해 해결할 수 있지만, 모델 학습에 필요한 각 스타일에 대응되는 병렬 말뭉치를 구하기 어려운 문제점이 있다. 따라서 최근에는 비병렬 말뭉치를 이용해 텍스트 스타일 변환을 수행하는 방법들이 연구되고 있다. 이 연구들은 주로 인코더-디코더 구조의 생성 모델을 사용하기 때문에 입력 문장이 가지고 있는 내용이 누락되거나 다른 내용의 문장이 생성될 수 있는 문제점이 있다. 본 논문에서는 마스크 언어 모델(masked language model)을 이용해 입력 텍스트의 내용을 유지하면서 원하는 스타일로 변경할 수 있는 텍스트 스타일 변환 방법을 제안하고 한국어 긍정-부정, 채팅체-문어체 변환에 적용한다.

주제어: 텍스트 스타일 변환, 마스크 언어 모델, 비병렬 데이터

1. 서론

텍스트 스타일 변환(text style transfer)은 입력 스타일(source style)로 쓰여진 텍스트의 내용(content)을 유지하며 목적 스타일(target style)의 텍스트로 변환하는 문제이다. 텍스트 스타일 변환에는 긍정적인 문장을 부정적인 문장으로, 구어체를 문어체로, 고전 문장을 현대 문장으로 변환하는 것을 예로 들 수 있다.

텍스트 스타일 변환은 시퀀스 간 변환 문제(sequence-to-sequence)로 기존 기계 학습 모델[1,2,3]을 이용해 해결할 수 있는데, 이 방법으로 스타일 변환 문제를 해결하기 위해서는 대량의 병렬 말뭉치가 필요하다. 하지만 각 스타일에 대응되는 병렬 말뭉치를 구하기 어려운 문제가 있고, 말뭉치를 직접 구축하는 경우에도 많은 비용이 필요하다. 따라서 최근에는 병렬 말뭉치 없이 비병렬 말뭉치를 이용해 텍스트 스타일 변환을 수행하는 방법들이 연구되고 있다[4-9].

비병렬 말뭉치를 이용한 텍스트 스타일 변환 연구는 적대적 학습을 통해 입력된 텍스트에서 내용과 스타일을 분리한 후 디코더 등을 이용해 변환된 문장을 생성하는 방법, 그리고 입력된 문장에서 스타일에 해당하는 토큰(attribute marker)을 찾은 뒤, 해당 토큰을 삭제한 후, 이를 목적 스타일 단어로 대체하거나 생성하는 방법 등이 있다.

스타일 토큰을 삭제한 문장과 스타일 속성(attribute)을 표시하는 값을 생성 모델의 입력으로 사용하는 기존 연구는 스타일 변환을 생성 모델을 통해 해결하기 때문에, 원래의 문장이 가지고 있는 내용이 누락되거나 다른

내용의 문장이 생성될 수 있는 문제점이 있다. 또한 스타일과 내용을 각각 학습해야 하기 때문에 학습이 더 어려워지는 문제점이 있다.

본 논문에서는 입력 텍스트의 내용을 유지하면서 삭제된 스타일 토큰을 원하는 스타일을 토큰으로 변경하기 위해 삭제된 토큰을 마스크 언어 모델을 이용해 생성하는 방법을 통해 스타일 변환 문제를 해결하고자 한다. 이를 위해 BERT(Bidirectional Encoder Representations from Transformers)[10] 모델을 이용한 한국어 텍스트 스타일 변환 방법을 제안하고 긍정-부정, 채팅체-문어체 변환에 적용한다.

2. 관련 연구

텍스트 스타일 변환 연구는 기계학습 모델에 필요한 병렬 말뭉치의 유무에 따라 구분할 수 있다. 일반적으로 병렬 말뭉치를 이용할 수 있는 경우 더 좋은 품질의 모델을 얻을 수 있으나 학습에 사용하기 위한 대량의 병렬 말뭉치를 확보하기 어려운 문제점이 있고 말뭉치를 새로 구축하더라도 많은 시간과 비용이 요구된다.

[3]의 연구는 병렬 말뭉치를 사용한 연구로 transformer 모델 기반 생성모델과 대량의 비병렬 말뭉치를 이용해 사전학습하고, 사전학습된 모델과 소량의 병렬 말뭉치를 이용해 미세조정을 수행하는 방법을 통해 스타일 변환 모델을 학습시켰다.

비병렬 말뭉치를 이용한 연구들은 같은 도메인을 구성하고 있는 각각의 스타일에 해당하는 비병렬 말뭉치를 수집하여 모델 학습에 사용한다. 예를 들어, 영화 리뷰

도메인에서 긍정적인 리뷰와 부정적인 리뷰는 점수를 기반으로 쉽게 수집할 수 있다. 이렇게 수집된 데이터들은 각 문장들이 일대일로 대응되지는 않지만, 각각의 스타일 전체의 묶음으로는 대응된다고 할 수 있고, 각각의 스타일 또한 잘 표현하고 있다고 할 수 있다.

[4,5,6]의 연구는 생성 모델과 적대적 학습 방법을 통해 해석 가능한 잠재 표현을 생성하는 방법이다. 이 방법은 입력된 텍스트의 내용과 스타일을 분리하고, 분리된 내용과 스타일을 표현하는 속성 값을 생성 모델의 입력으로 사용해 변환된 문장을 생성하는 연구를 진행했다.

[7]의 연구는 스타일과 내용을 분리하지 않는 방식의 연구를 진행하였다. Transformer 모델과 판별기를 이용해 적대적 학습 방법과 유사한 방법으로 텍스트 스타일 변환을 수행하는 모델을 제안했다.

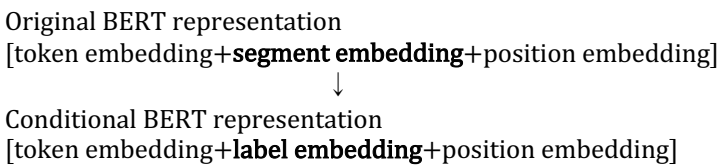
[8,9]의 연구에서는 먼저 비병렬 말뭉치에 대해 각각의 스타일을 구분할 수 있는 분류기를 학습하고, 학습된 분류기를 이용해 입력된 텍스트에서 스타일에 해당하는 토큰을 찾고 삭제한다. [8]에서는 N-gram 빈도수를, [9]에서는 transformer 모델을 이용해 스타일 토큰을 찾는다. 그리고 해당 토큰을 삭제하여 내용만 남은 텍스트와 목적 스타일을 나타내는 속성 값을 함께 생성 모델의 입력으로 사용해 변환된 문장을 생성하는 연구를 진행했다.

3. 한국어 텍스트 스타일 변환

본 논문에서는 두개의 한국어 텍스트 스타일 변환 방법을 제안한다. 하나는 입력 문장에서 분류기를 이용해 스타일 토큰을 삭제하고, 삭제된 토큰을 Conditional BERT[11] 모델을 이용해 생성하는 방법이다. 다른 하나는 사전학습된 두개의 언어 모델을 이용해 토큰을 삭제하고, 삭제된 토큰을 Padded Masked Language Model(Padded MLM) 통해 생성하는 방법이다[12].

3.1 Conditional BERT 기반 텍스트 스타일 변환

Conditional BERT는 기존 BERT 모델에서 마스크된 단어를 생성할 때, 생성되는 단어가 입력된 레이블 정보에 영향을 받도록 설계된 모델이다. [그림 1]은 Conditional BERT 모델이다. 모델은 기존 BERT 모델과 동일한 구조를 지니고 있으나 모델의 입력으로 사용되는 표현을 아래와 같이 변경한다.



Conditional BERT 모델의 학습은 사전학습된 BERT 모델을 변경된 입력 정보(레이블 정보)와, 스타일 변환에 사용할 도메인 데이터로 미세조정하는 방법을 통해 학습한다. 학습이 끝난 Conditional BERT 모델을 이용해 스타일 변환을 하기 위한 순서는 다음과 같다. 먼저 입력 텍

트에서 스타일을 지닌 토큰을 찾기 위해 소스 스타일과 목적 스타일을 구분할 수 있는 분류기를 학습한다. 분류기는 각 도메인(감성분석:긍정-부정, 문체변환:채팅체-문어체)에 대해 따로 학습하며, 사전학습된 BERT 모델을 미세조정하여 사용한다. 미세조정이 완료된 BERT 모델에 스타일을 변환할 입력 문장을 넣고, 문장에서 [CLS] 토큰과 각 단어 사이의 어텐션 값이 높은 상위 n개 만큼의 제거할 토큰을 [MASK] 토큰으로 변경한다.

다음으로 스타일 토큰 삭제 작업이 끝난 마스킹된 문장을 학습이 완료된 Conditional BERT 모델의 입력으로 사용한다. 이 때, 입력 문장과 함께 목적 레이블을 같이 넣어주면 스타일이 변경된 문장을 출력하게 된다. 이 모델에서는 마스킹된 토큰을 생성할 때 문맥(context)에만 의존하는 기존 BERT 모델과 달리 레이블 정보에 영향을 받게 된다.

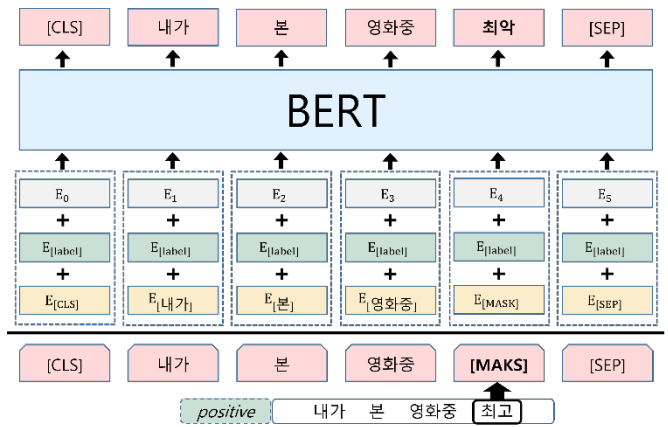


그림 1. Conditional BERT 모델 구조

3.2 Padded Masked Language Model 기반 텍스트 스타일 변환

Padded MLM은 n개의 마스크된 토큰을 n개의 토큰으로 생성하는 기존 BERT 모델과 달리 가변 길이의 토큰을 생성할 수 있는 모델이다. 가변길이의 토큰을 생성하기 위한 Padded MLM의 수식은 아래와 같다.

$$W_{i:j}$$

$$= (w_1, \dots, w_{i-1}, [MASK], [MASK], [MASK], [MASK], w_{i+3}, \dots, w_{|w|})$$

$$\mathcal{L}(w_{i:j}|w_{i:j}; \theta) = \prod_{t=i}^j P_{MLM}(w_t|W_{i:j}; \theta)$$

$$\times \prod_{t=j+1}^{i+n_p-1} P_{MLM}([PAD]_t|W_{i:j}; \theta)$$

Padded MLM을 학습하기 위해 입력 문장에 고정된 n_p 개의 마스크를 씌우고 $i:j$ 개의 토큰을 정답으로, $j+1:i+n_p-1$ 개의 토큰을 [PAD] 토큰으로 학습하게 된다. 실제 모델 학습은 사전학습된 BERT 모델을 위의 수식으로 미세조정하는 방식으로 진행되었다.

표 1. 긍정-부정 변환 결과 예제

	원문	Conditional BERT	Padded MLM
1	볼 거 없을때 봐도 재미 없는 영화!	볼 거 없을때 봐도 재미 있는 영화 ㅋ	볼 거 없을때 봐도 재미 있는 영화!
2	람보는 전쟁씬 말고도 볼게 많음	람보는 전쟁씬 말고도 볼게 없음	람보는 전쟁씬 말고도 볼게 없더라
3	내가 기억하는 가장 따뜻한 영화중 하나	내가 기억하는 가장 쓰레기 영화중 하나	내가 기억하는 가장 최악의 영화중 하나
4	중간에 화면이 좀 끊기는것 빼곤 넘 좋았어요~	중간에 화면이 좀 끊기는것 빼곤 넘 재미없어요~	중간에 화면이 좀 끊기는것 빼곤 볼 거리 하나도 없음.
5	외쳐! 갓 토르! 그냥 노잼이어도 10점준다	외쳐! 갓 토르! 그냥 노잼이라 2점준다	외쳐! 갓 더스툼! 다같이 총10점 준다
6	피는 물보다 진하다. 이 한마디로 모든게 표현되는 영화.	피는 물보다 진하다; 이 한마디로 모든게 표현되는 영화ㅋ	피는 물보다 진하다. 이 한마디로 모든게 잘못된 영화.
7	이거 어렸을때 보고 오늘 사서 엄마하고 봤습니다.역시 명화더군요	이거 어렸을때 보고 오늘 커서 엄마하고 봤습니다.역시 만화더군요	이거 어렸을때 보고 오늘 사서 엄마하고 봤습니다.역시 미국영화였어~

표 2. 채팅체-문어체 변환 결과 예제

	원문	Conditional BERT	Padded MLM
1	작성 성공 했습니다. ^^	작성 성공 했습니다...	작성 성공 했습니다..
2	통제 현황 좀 댓글로 부탁드립니다~~	통제 상황 좀 댓글로 알려주십시오.	통제 현황 좀 댓글로 부탁드립니다.
3	그래도 아무것도 못하고 죽는분들을 위해...	그래도 아무것도 못하고 죽는분들을 위해 갑니다.	그래도 아무것도 못하고 죽는분들을 생각해야 한다.
4	피씨로 하는데 오늘 아침부터 총 2번 튕겼다 엔씨야 빨리 안정화좀	피씨로 하는데 오늘 아침부터 총 2번 튕겼다 이젠 캐릭 안정화좀	피씨로 하는데 오늘 아침부터 총 2번 통화했다"고 말했다고 말했다.
5	악용은 안되었으면 하지만...	악용은 안되었으면 합니다	악용은 당연 하지만...
6	함께라는 말을 지~대로 느낄수있는 현맹.	함께라는 말을 지~대로 느낄수있는 부족.	함께라는 말을 느낄수있는 것 입니다.
7	답변 부탁드립니다	제안 부탁드립니다	. 잘 부탁드립니다

Padded MLM을 이용한 스타일 변환에서는 3.1의 방법과 달리 스타일 토큰을 삭제하기 위해 분류기를 사용하지 않았다. 분류기를 사용하는 대신 입력 스타일과 목적 스타일 도메인에 대해 각각 Padded MLM을 학습한 후, 실제 입력 문장에 [MASK] 토큰을 씌운다. 그 다음 입력 및 목적 Padded MLM에서 단어 등장 확률의 차이를 구하고, 그 차이가 가장 큰 단어를 새로운 단어로 생성하는 방법을 사용한다. 이를 수행하기 위한 수식은 아래와 같다. $\hat{W}_{i,j}^{target}$ 은 목적 Padded MLM으로 생성한 단어, $W_{i,j}$ 는 실제 입력된 단어를 나타낸다.

$$TargetScore(i, j) = \mathcal{L}(\hat{W}_{i,j}^{target} | W_{i,j}; \Theta_{target}) - \mathcal{L}(W_{i,j} | W_{i,j}; \Theta_{target})$$

$$SourceScore(i, j) = -\max[0, \mathcal{L}(\hat{W}_{i,j}^{target} | W_{i,j}; \Theta_{source}) - \mathcal{L}(W_{i,j} | W_{i,j}; \Theta_{source})]$$

$$Score(i, j) = TargetScore(i, j) + SourceScore(i, j)$$

4. 실험 및 결과

본 논문에서는 한국어 텍스트 스타일 변환에 두개의 비병렬 실험 데이터를 사용한다. 먼저 한국어 긍정-부정 변환에는 네이버 영화 리뷰 데이터¹를 사용한다. 다음으로 채팅체-문어체 비병렬 변환에는 직접 수집한 데이터를 사용하였으며 문어체는 뉴스 기사로 가정하고 채팅체는 게임 게시판에 게시된 글로 가정하였다. 문어체는 네이버 뉴스 기사를 크롤링 한 약 50만 사용하였고, 채팅체는 엔씨소프트 playNC² 리니지M, 리니지2M 게시판에서

¹ <https://github.com/e9t/nsmc>

² <https://www.plaync.com>

크롤링한 약 50만 문장을 사용하였다.

실험에 사용한 BERT 사전학습 모델은 ETRI에서 학습한 KorBERT³를 사용하였다. Conditional BERT 모델의 미세조정 사용된 파라미터는 입력 길이 64, 배치 크기 32, 에폭 50이고, 긍정 문장과 부정 문장 분류기의 학습에는 입력 길이 64, 배치 크기 128, 에폭 100을 사용하였고 문체 변환 및 영화리뷰 도메인에 대해 동일한 파라미터 값을 사용하였다. Padded MLM 모델의 학습에는 사용한 파라미터는 다음과 같다. 영화리뷰 도메인에는 입력 길이 128, 배치 크기 64, 에폭 100을 사용하였고, 문체 변환 데이터에도 동일한 값의 파라미터를 사용하였다.

[표 1]은 각 방법을 통해 긍정-부정 변환을 수행한 결과이다. 전체 예제에서 ‘따뜻한’, ‘쓰레기’, ‘최악의’와 같이 주로 감성분석 도메인에서 극성을 나타내는 단어를 변경하는 것을 확인할 수 있다. 1-4번 예제는 스타일 변환이 잘 수행된 결과, 5번 예제는 Padded MLM이, 6번 예제는 Conditional BERT 모델이, 7번 예제는 두 모델 모두 스타일 변환에 실패한 예제이다. 4번 예제를 통해 Padded MLM 모델이 가변길이 생성을 통해 Conditional BERT 모델 보다 더 풍부하게 단어를 생성할 수 있음을 알 수 있다.

[표 2]는 각 방법을 통해 채팅체-문어체 변환을 수행한 결과이다. 1-3번은 문체 변환을 잘 수행한 예제이다. 1번 예제는 채팅체에 사용되는 이모티콘 표현을 두 모델 모두 마침표로 변환하였다. 2번 예제의 경우 Padded MLM은 ‘부탁드려요~~’를 ‘부탁드립니다.’로 완벽하게 변환하였고, Conditional BERT는 ‘알려주십시오.’로 변환했지만, 입력 문장의 내용을 유지하고 있다. 3번 예제에서 Padded MLM이 생성한 문장을 보면 문어체로 가정한 뉴스 도메인에서 사용하는 문장을 그대로 가져온 것을 볼 수 있다. 6번 예제에서 Conditional BERT 모델이 ‘혈맹.’을 ‘부족’으로 변환하였는데 단어 수준에서는 도메인의 내용을 잘 변경한 것으로 보이지만, 원래의 내용을 보존하기 위해 마스킹에 제한을 두어 문어체로의 변환은 이루어지지 않은 것으로 볼 수 있다.

실험 결과 긍부정 변환에서는 주로 극성을 띤 단어를 변경하고, 문체 변환에서는 문장 뒤쪽의 단어를 변환하는 것을 확인할 수 있었다. 한국어의 경우 대부분 종결어미에 의하여 문체가 결정되기 때문에 모델이 이를 학습한 것으로 보인다. 따라서 문체 변환을 수행할 때 마스킹하는 토큰의 제약을 통해 7번 예제와 같은 실패 결과를 줄일 수 있을 것으로 보인다.

5. 결론

본 논문에서는 마스크 언어 모델을 이용한 비병렬 한국어 텍스트 스타일 변환 모델을 제안하였다. 실험 결과, 병렬 데이터 없이도 긍부정과 문체 변환이 가능함을 보였고, 변환 과정에서 각 도메인이 가지는 특성을 실험 결과를 통해 확인할 수 있었다. 추후 연구에서는 스타일

변환에 사용되는 비병렬 데이터의 노이즈를 줄이는 방법과, 기존 스타일 변환 모델의 성능 향상 방법을 연구할 예정이다.

감사의 글

이 논문은 엔씨소프트의 “인터넷 채팅체에 적응하는 말뭉치 자동 확장 연구” 과제의 지원을 받아 수행된 연구임

이 논문은 한국연구재단의 “딥러닝 기반의 한국어 텍스트 스타일 변환 기술 연구” 과제의 지원을 받아 수행된 연구임 (No. NRF2021R1F1A106440311)

참고문헌

- [1] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." *Advances in neural information processing systems*. 2014.
- [2] Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems*. 2017.
- [3] 박다솔, 장두성, 차정원. "TGST: 트랜스포머 기반의 한국어 생성적 스타일 변환." *한국컴퓨터종합학술대회*, 2020.
- [4] John, Vineet, et al. "Disentangled representation learning for non-parallel text style transfer." *arXiv preprint arXiv:1808.04339*. 2018.
- [5] Hu, Zhiting, et al. "Toward controlled generation of text." *International Conference on Machine Learning*. 2017.
- [6] Fu, Zhixin, et al. "Style transfer in text: Exploration and evaluation." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. No. 1. 2018.
- [7] Dai, Ning, et al. "Style transformer: Unpaired text style transfer with out disentangled latent representation." *arXiv preprint arXiv:1905.05621* (2019).
- [8] Li, Juncen, et al. "Delete, retrieve, generate: A simple approach to sentiment and style transfer." *arXiv preprint arXiv:1804.06437* (2018).
- [9] Sudhakar, Akhilesh, Bhargav Upadhyay, and Arjun Maheswaran. "Transforming delete, retrieve, generate approach for controlled text style transfer." *arXiv preprint arXiv:1908.09368* (2019).
- [10] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
- [11] Wu, Xing, et al. "Conditional bert contextual

³ <https://www.plaync.com>

- augmentation." International Conference on Computational Science. Springer, Cham, 2019.
- [12] Malmi, Eric, Aliaksei Severyn, and Sascha Rothe. "Unsupervised Text Style Transfer with Masked Language Models." Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020.