

적대적 생성 신경망을 이용한 문법 오류 교정

권순철^o, 유환조, 이근배

포항공과대학교 컴퓨터공학과
{theincluder, hwanjoyu, gblee}@postech.ac.kr

Grammatical Error Correction Using Generative Adversarial Network

Soonchoul Kwon, Hwanjo Yu, Gary Geunbae Lee

요약

문법 오류 교정은 문법적으로 오류가 있는 문장을 입력 받아 오류를 교정하는 시스템이다. 문법 오류 교정을 위해서는 문법 오류를 제거하는 것과 더불어 자연스러운 문장을 생성하는 것이 중요하다. 이 연구는 적대적 생성 신경망(GAN)을 이용하여 정답 문장과 구분이 되지 않을 만큼 자연스러운 문장을 생성하는 것을 목적으로 한다. 실험 결과 GAN을 이용한 문법 오류 교정은 MaxMatch F_{0.5} score 기준으로 0.4942을 달성하여 Baseline의 0.4462보다 높은 성능을 기록했다.

주제어: 문법 오류 교정, 적대적 생성 신경망, CoNLL shared task

1. 서론

문법 오류 교정(Grammatical error correction; GEC)이란 문법적으로 틀린 문장을 입력으로 받아서, 문장에 포함된 문법 오류들을 교정한 문장을 출력하는 과제이다. 좋은 GEC 모델의 결과는 1. 문장에 포함된 문법 오류를 제거해야 하고, 2. 입력 문장의 의미를 유지해야 하며, 3. 사람이 보았을 때 자연스러운 문장을 생성해야 한다.

GEC을 위해 과거에는 언어 모델(Language model), 컨퓨전 행렬(Confusion matrix) 등을 이용하는 통계적 방법이 사용되었으나, 현재는 GEC 과제를 ‘문법적으로 틀린 언어’에서 ‘문법적으로 올바른 언어’로 번역하는 것으로 보고 앞의 언어를 뒤의 언어로 번역하는 기계 번역(machine translation) 기술이 사용되고 있다. 현재 가장 높은 GEC 성능을 보이는 모델들은 기계 번역 방법을 사용하고 있다.

특히 인공지능을 이용한 기계 번역(Neural machine translation; NMT) 기술에 의해 GEC는 더욱 높은 성능을 얻게 되었다[1, 2]. NMT 방법을 이용한 GEC는 통계적 방법을 이용한 GEC를 위해 필요했던 오류 유형 레이블이나 오류 유형에 종속적인 방법론, 철자법 교정기 등이 필요하지 않아 현재까지 가장 높은 성능을 보여주고 있다.

이 연구에서는 적대적 생성 신경망(Generative adversarial network; GAN)을 이용하여 실제 문장과 유사한 문장을 생성함으로써 문법 오류를 교정하는 방법을 제안한다. 제안하는 방법은 오류 교정을 기계 번역과 더불어 스타일 전송(Style transfer) 문제로도 봄으로써 기계 번역의 손실 함수와 GAN의 손실 함수를 함께 사용한다. GAN을 이용해서 오류가 포함된 입력 문장이 정답 문장과 가까워지도록 훈련함으로써 1. 정답 문장처럼 문법 오류를 제거하고, 2. 입력 문장의 의미를 유지하고, 3. 정답 문장과 구분할 수 없을 정도로 자연스러운 문장을 생성할 수 있다.

2. 관련 연구

현재 GEC 시스템에 대한 연구는 NMT를 이용한 것이 주류가 되어 있다. 대부분의 NMT는 인코더-디코더 구조를 갖고 있다. 인코더(Encoder)는 입력 문장을 분석해 문장 전체의 임베딩을 얻는 역할을 하고, 디코더(Decoder)는 문장 임베딩과 인코더의 중간결과를 이용해 문법에 맞는 문장의 각 단어를 생성한다.

[1]은 multilayer convolution neural network(CNN)와 복사 방법론(Copying mechanism)을 이용했다. [2]은 트랜스포머(Transformer) 구조에 BERT[3]와 유사한 사전학습과 복사 방법론을 적용한 네트워크와 훈련 방법을 제안했다.

[4]은 GEC 시스템을 스타일 전송 문제로 보고 해결을 시도했으나, 여러 언어의 병렬 코퍼스를 필요로 하는 반면 문법 오류를 위한 데이터로 훈련하지 않았다. 그 결과 언어 스타일 간의 전환에 비하여 문법 오류 스타일의 전송은 성공적이지 않았다.

GAN은 실제 데이터 x 와 유사한 데이터를 생성하는 방법이다. GAN은 생성기(Generator; G)와 판별기(Discriminator; D)의 두 네트워크로 이루어진다[5]. 생성기는 입력된 무작위 노이즈 z 로부터 실제 데이터와 유사한 데이터 $G(z)$ 를 생성한다. 판별기는 실제 데이터 x 또는 생성기에 의해 생성된 데이터 $G(z)$ 를 입력으로 받아 이것이 생성된 데이터일 확률을 구한다. 이에 대한 GAN 손실 함수 L 은

$$L = E_x[\log(D(x))] + E_z[\log(1 - D(G(z)))]$$

이 된다. 판별기는 이 손실 함수를 최소화함으로써 두 데이터를 구분하려 하고, 생성기는 생성된 데이터 $G(z)$ 를 x 와 유사하게 만듦으로써 손실 함수를 최대화한다. 잘 훈련된 생성기는 위의 손실 함수가 최대화되는, 즉 판별기가 구분할 수 없을 정도로 실제 데이터 분포와 유사한 데이터를 생성할 수 있게 된다.

이러한 GAN은 이미지 분야의 스타일 전송 분야에서 성

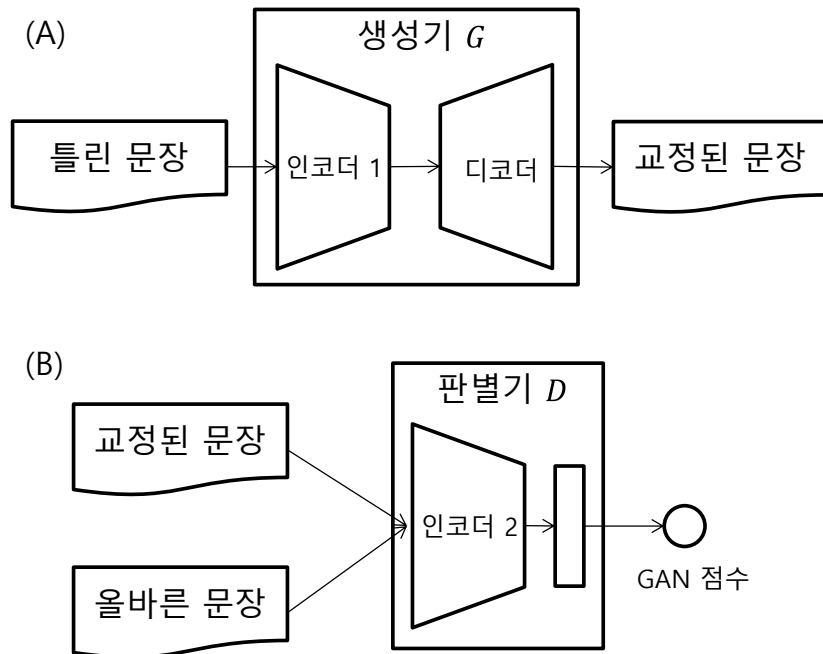


그림 1. GAN을 이용한 문법 오류 교정 훈련 모델

공적인 결과를 가져왔다. [6]은 GAN을 이용하여 인물사진을 에모지(Emoji) 그림으로 전송한다. [7]은 cycle-consistency loss를 이용해 두 종류의 스타일을 양방향으로 전송할 수 있다. 위의 스타일 전송 연구들은 한 스타일에 대응되는 다른 스타일의 데이터가 없는 환경, 즉 unpaired dataset에 대해 진행이 가능한 모델이다.

3. GAN을 이용한 문법 오류 교정

NMT는 GEC라는 과제를 ‘문법적으로 틀린 언어’를 문법적으로 올바른 언어’로 번역하는 문제로 본다. 그러나 다른 관점에서 이 과제를 ‘문법적으로 틀린 문장의 스타일’을 ‘문법적으로 올바른 문장의 스타일’로 바꾸는 것으로 보았을 때, 이 과제는 스타일 전송 문제로 볼 수 있다. GAN은 이미지의 스타일 전송 분야에서 좋은 결과를 얻을 수 있다는 것이 알려져 있다. 따라서 GAN의 원리를 자연어 처리 분야인 GEC에도 활용할 수 있다.

이 연구에서는 GAN을 이용해 문법적으로 틀린 문장을 교정하여 문법적으로 올바른 문장과 구분할 수 없는 문장을 생성하는 것을 목표로 한다(그림 1). 이 때 GAN의 생성기가 새 문장을 생성하는 역할을, 판별기는 두 문장을 구분할 수 없게 만드는 역할을 한다. 생성기는 널리 사용되는 NMT 기반의 GEC 시스템을 그대로 사용할 수 있다(그림 1.A). 판별기는 교정된 문장 또는 올바른 문장을 입력으로 받아 이 문장이 올바른 문장일 확률(GAN 점수)을 출력한다(그림 1.B). 이 때 사용하는 손실 함수 L 은 GEC loss와 GAN loss를 합쳐서 사용한다:

$$L = L_{GEC} + \alpha L_{GAN}$$

$$L_{GEC} = -\mathbb{E}_{x \in x_c} [\log(P_G(x))]$$

$$L_{GAN} = -(\mathbb{E}_{x \in x_c} [\log(D(x))] +$$

$$\mathbb{E}_{x \in x_w} [\log(1 - D(G(x)))])$$

x_c 와 x_w 는 각각 올바른 문장과 틀린 문장을 의미한다. α 는 두 손실 함수 사이의 비율을 의미한다. 위 손실 함수를 최소화함으로써 올바른 문장과 구분할 수 없는 문장을 생성할 수 있다.

4. 실험 및 결과

4.1. 실험 설정

생성기로 사용하는 GEC 네트워크 구조는 [1]의 multilayer CNN 구조를 사용했다. Multilayer의 CNN의 인코더(그림 1의 인코더1)와 디코더는 각 7 층으로 되어 있고, 각 층의 은닉층은 500 차원, 출력층은 512 차원으로 되어있다. 인코더의 각 층은 fully connected layer, CNN layer, gated linear unit, residual connection으로 이루어져 있다. 디코더의 각 층은 gated linear unit, CNN layer, attention layer, residual connection으로 이루어져 있다. 입출력 단위로 단어를 사용할 경우 신규 단어나 오타를 입/출력할 수 없기 때문에 이 문제를 해결하기 위해 subword 단위로 입력 및 출력한다.

판별기는 인코더(그림 1의 인코더2)와 fully connected layer으로 이루어져 있다. 판별기의 인코더는 생성기의 인코더와 같은 구조로 이루어져 있지만 weight를 공유하지는 않는다. Fully connected layer는 판별기의 512차원 출력을 받아 1차원의 GAN 점수를 출력한다.

표 1. GAN을 이용한 GEC 성능

	Precision	Recall	F _{0.5}
Baseline [1]	0.5988	0.2211	0.4462
GAN	0.6594	0.2478	0.4948
Grundkiewicz [11]	0.6677	0.3449	0.5625

GEC loss와 GAN loss의 비율을 나타내는 α 는 1을 사용했다. Dropout은 0.2, momentum은 0.99를 사용했다.

훈련 데이터로는 NUS Corpus of Learner English (NUCLE)[8]¹와 NAIST Lang-8 Learner Corpora[9]²을 사용했다. 테스트 데이터로는 CoNLL-2014 Shared Task의 테스트 데이터를 사용했다[10]³. 성능 지표로는 GEC 과제에서 사용되는 MaxMatch (M2) F_{0.5} score를 사용했다[11]. M2 score는 오류 교정의 단위를 구 (phrase)로 취급하고 GEC 결과가 정답과 비교해서 얼마나 교정되었는지 평가한다.

4.2. 성능 평가

표 1은 GAN을 사용한 실험과 GAN을 사용하지 않은 baseline 실험의 성능 비교 결과이다. 실험 결과 GAN을 이용한 GEC 방법이 0.4948 M2 F_{0.5} score를 달성해서, 0.4462를 보인 baseline 방법보다 0.05 가량 높은 성능을 거두었다. 또한 이 논문에서는 출력층이 512차원으로, [1]에서 1024차원을 사용한 것과 비교해서 parameter의 수가 절반 정도인데도 [1]에서 보고된 F_{0.5} score인 0.4638보다 훨씬 높은 M2 F_{0.5} score을 달성했다. 현재 문법 오류 교정 분야에서 최고 성능을 보인 Grundkiewicz[11]의 시스템보다는 낮은 성능을 보였으나, 이 시스템은 통계 기반과 인공신경망 기반 기계 번역 모델을 함께 사용하고, 이 외에 언어 모델(language model)과 스펠링 오류 검사기(spell-check component) 등 추가 모듈을 사용하였다.

4.3. 문법 오류 교정 결과 분석

표 2는 오류가 포함된 문장에 대한 GAN으로 훈련하지 않은 시스템과 GAN으로 훈련한 문장을 비교한 것이다. GAN으로 훈련한 모델은 그렇지 않은 모델이 교정하지 못한 오류(mock of → mock)나 잘못 교정한 오류(your conditions → your condition)의 경우가 적었다. 그러나 오류를 교정하지 못한 경우(and vice versa → or not)도 여러 번 나타났다.

표 2. 문법 오류 교정 결과 분석

입력	This knowledge <u>maybe</u> relavant to them , and vice versa .
정답	This knowledge <u>may be</u> relevant to them or not .
Baseline	This knowledge <u>may</u> relavant to them , and vice versa .
GAN	This knowledge <u>may be</u> relavant to them , and vice versa .

입력	People and friends often <u>mock of</u> your conditions .
정답	People and friends often <u>mock</u> their conditions .
Baseline	People and friends often <u>mock of</u> your condition .
GAN	People and friends often <u>mock</u> your conditions .

5. 결론

이 연구를 통해 같은 네트워크 구조를 갖더라도 GAN을 이용한 훈련 방법을 통해 GEC의 성능을 높일 것이라는 것을 보였다. 추후 연구로 conditional GAN, data augmentation 등의 새 방법론을 적용한 연구를 시도할 예정이다.

6. Acknowledgement

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2013-0-00179, (엑소브레인-3세부) 컨텍스트 인지형 Deep-Symbolic 하이브리드 지능 원천 기술 개발 및 언어 지식 자원 구축)

참고문헌

- [1] Chollampatt, Shamil, Hwee Tou Ng. "A multilayer convolutional encoder-decoder neural network for grammatical error correction." Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- [2] Zhao, Wei, Liang Wang, Kewei Shen, Rouyu Jia, Jingming Liu. "Improving grammatical error correction via pre-training a copy-augmented architecture with unlabeled data." arXiv preprint arXiv:1903.00138. 2019.
- [3] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. "Bert: Pre-training of deep

¹ <https://www.comp.nus.edu.sg/~nlp/corpora.html>

² <https://sites.google.com/site/naistlang8corpora/>

³ <https://www.comp.nus.edu.sg/~nlp/conll14st.html>

- bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805. 2018.
- [4] Korotkova, Elizaveta, Agnes Luhtaru, Maksym Del, Krista Liin, Daiga Deksnė, Mark Fishel. "Grammatical Error Correction and Style Transfer via Zero-shot Monolingual Translation." arXiv preprint arXiv:1903.11283. 2019.
- [5] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio. "Generative adversarial nets." In Advances in neural information processing systems (pp. 2672-2680). 2014.
- [6] Taigman, Yaniv, Adam Polyak, Lior Wolf. "Unsupervised cross-domain image generation." arXiv preprint arXiv:1611.02200. 2016.
- [7] Zhu, Jun-Yan, Taesung Park, Phillip Isola, Alexei A. Efros. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.
- [8] Dahlmeier, Daniel, Hwee Tou Ng, Siew Mei Wu. Building a large annotated corpus of learner English: The NUS corpus of learner English." In Proceedings of the eighth workshop on innovative use of NLP for building educational applications (pp. 22-31). 2013.
- [9] Mizumoto, Tomoya, Mamoru Komachi, Masaaki Nagata and Yuji Matsumoto. "Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners." In Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP), pp.147-155. Chiang Mai, Thailand, November 2011.
- [10] Ng, Hwee Tou, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, Christopher Bryant. "The CoNLL-2014 shared task on grammatical error correction." Proceedings of the Eighteenth Conference on Computational Natural Language Learning: Shared Task. 2014.
- [11] Dahlmeier, Daniel, Hwee Tou Ng. "Better evaluation for grammatical error correction." Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2012.
- [12] Grundkiewicz, Roman, and Marcin Junczys-Dowmunt. "Near human-level performance in grammatical error correction with hybrid machine translation." arXiv preprint arXiv:1804.05945

2018.