

한국어 문법 오류 교정 모델을 위한 문장 단위 디노이징 학습법

김훈래¹, 김윤수^{1,2}, 이근배^{1,2}

포항공과대학교 인공지능대학원¹, 포항공과대학교 컴퓨터공학과²

{hoonrae, yunsu.kim, gblee}@postech.ac.kr

Sentence Unit De-noising Training Method for Korean Grammar Error Correction Model

Hoonrae Kim¹, Yunsu Kim^{1,2}, Gary Geunbae Lee^{1,2}

Graduate School of Artificial Intelligence, Pohang University of Science and Technology¹

Computer Science and Engineering, Pohang University of Science and Technology²

요약

문법 교정 모델은 입력된 텍스트에 존재하는 문법 오류를 탐지하여 이를 문법적으로 옳게 고치는 작업을 수행하며, 학습자에게 더 나은 학습 경험을 제공하기 위해 높은 정확도와 재현율을 필요로 한다. 이를 위해 최근 연구에서는 문단 단위 사전 학습을 완료한 모델을 맞춤법 교정 데이터셋으로 미세 조정하여 사용한다. 하지만 본 연구에서는 기존 사전 학습 방법이 문법 교정에 적합하지 않다고 판단하여 문단 단위 데이터셋을 문장 단위로 나눈 뒤 각 문장에 G2P 노이즈와 편집거리 기반 노이즈를 추가한 데이터셋을 제작하였다. 그리고 문단 단위 사전 학습한 모델에 해당 데이터셋으로 문장 단위 디노이징 사전 학습을 추가했고, 그 결과 성능이 향상되었다. 노이즈 없이 문장 단위로 분할된 데이터셋을 사용하여 디노이징 사전 학습한 모델을 통해 문장 단위 분할의 효과를 검증하고자 했고, 디노이징 사전 학습하지 않은 기존 모델보다 성능이 향상되는 것을 확인하였다. 또한 둘 중 하나의 노이즈만을 사용하여 디노이징 사전 학습한 두 모델의 성능이 큰 차이를 보이지 않는 것을 통해 인공적인 무작위 편집거리 노이즈만을 사용한 모델이 언어학적 지식이 필요한 G2P 노이즈만을 사용한 모델에 필적하는 성능을 보일 수 있다는 것을 확인할 수 있었다.

주제어: 한국어 문법 교정, Grammatical Error Correction

1. 서론

문법 오류 교정(Grammatical Error Correction; GEC)은 텍스트 입력에 존재하는 문법적인 오류들을 탐지하여 이를 맞춤법에 맞도록 고치는 작업을 의미한다. 기존 GEC는 규칙 기반으로 작동하여 정확한 작동을 위해서는 전문 지식을 바탕으로 시스템을 구성할 수밖에 없었기 때문에 전문가의 참여가 필수적이었다. 하지만 최근의 GEC는 대규모 데이터를 활용한 인공 신경망을 바탕으로 작동하여 전문 지식 없이 작동하는 시스템을 구축할 수 있게 되었다.

GEC 모델은 한국어 학습자의 학습 도구에 사용되거나 다른 자연어 처리 신경망의 입력 데이터를 교정하기도 하며, 자연어 생성 혹은 음성 인식 모델의 출력을 교정하여 시스템의 성능을 향상하는 데 사용되는 등 여러 분야에 사용되어 그 중요도가 높다고 볼 수 있다.

본 연구에서는 한국어 문단들로 이루어진 데이터셋을 사용하여 사전 학습된 T5[1] 모델과 한국어 맞춤법 교정 데이터셋을 사용하여 GEC 모델을 훈련시켰으며, 그 결과 F0.5 점수 기준 0.79의 성능을 보였다. 하지만 대부분의 GEC 작업은 문장 단위로 이루어지며, 맞춤법 교정 데이터셋 또한 문장 단위로 이루어져 있어 해당 모델의 사전 학습 단계만으로는 GEC 작업을 수행하기에 부족하다. 또한, 영어 데이터셋을 사용하여 사전 학습된 모델에 문장 단위 디노이징 사전 학습 단계를 추

가로 진행했을 때 성능이 향상되었다는 학습법[2]이 제안된 바 있다. 따라서 본 연구에서는 한국어 문장 단위 디노이징 사전 학습 단계를 추가하여 한국어 GEC 모델의 성능을 향상하고자 하였다. 우선, 노이즈가 가장 적은 신문 기사 데이터셋을 문장 단위로 분할한 데이터셋을 구축했고, 해당 데이터셋의 각 문장에 편집거리 기반 자모 단위 노이즈와 Grapheme to Phoneme(G2P) 노이즈를 추가하여 디노이징 사전 학습을 위한 데이터셋을 구축하였다. 먼저 문장 단위 분할의 효과를 검증하기 위해 노이즈를 추가하지 않고 문장 단위 분할만 수행한 데이터셋을 사용하여 모델에 사전 학습을 추가로 진행하였다. 모델의 입력 문장과 목적 문장이 같은 오토인코더와 같은 학습을 진행했고, 그 결과 기존 F0.5 점수 0.79보다 높은 0.84의 성능을 보였다. 그리고 디노이징 사전 학습의 효과를 검증하기 위해 문장 단위로 분할된 후 노이즈를 추가한 데이터셋을 사용하여 모델에 사전 학습을 추가로 진행하였다. 모델의 입력은 노이즈가 추가된 문장이고 목적 문장은 노이즈가 없는 문장으로 학습을 진행했고, 그 결과 F0.5 점수 0.86의 성능을 달성하였다. 또한 G2P 노이즈만을 사용하여 디노이징 사전 학습을 추가한 모델과 편집거리 기반 노이즈만을 사용하여 디노이징 사전 학습을 추가한 모델의 성능을 측정한 결과 모두 F0.5 점수 0.86의 수준에 머물렀다. 따라서 언어학적 지식 없이 무작위로 생성한 편집거리 기반 노이즈만으로 이루어진

데이터셋으로 사전 학습한 모델이 전문가의 언어학적 지식과 상대적으로 더 많은 노력이 필요한 G2P 노이즈만으로 이루어진 데이터셋으로 사전 학습한 모델과 비슷한 수준의 성능을 보이는 것을 확인할 수 있었다.

2. 관련 연구

기존의 한국어 GEC 시스템은 규칙 기반 혹은 통계 기반의 모델 [3]로 문법 교정이 진행되었다. 규칙 기반 혹은 통계적 방법을 활용하여 문법 교정 시스템을 제작할 경우 전문가의 지식이 필수적이다. 하지만 복잡한 규칙들을 포함하는 시스템을 제작하는 것은 쉬운 일이 아니다. 따라서 최근 연구에서는 인공 신경망을 활용하는 방법이 대두되고 있다. [4]에서는 한국어 문법 오류 교정을 위해 Encoder-Decoder 기반의 인공 신경망을 사용했고, [5]에서는 Transformer [6] 기반의 모델을 사용하여 한국어 대화체 문장을 교정하는 방법을 제시하였다. 또한 [7]에서는 Transformer 기반의 모델과 Copying Mechanism을 같이 사용하는 등 인공 신경망을 활용한 문법 교정 연구가 이어지고 있다.

최근의 영어 GEC 시스템의 경우 부족한 맞춤법 교정 데이터셋을 보강하기 위해 합성된 데이터셋을 생성하여 최대한 많은 양의 데이터로 모델을 훈련시키는 방법 [8]을 사용하기도 한다. 또한 최근에는 GEC 시스템을 오류가 있는 문장에서 오류가 없는 문장으로의 기계 번역 시스템의 일종으로 보는 관점의 연구 [9][10]도 있다. 그리고 [2]에서는 기존 영어 문단 단위 데이터셋인 C4 데이터셋을 문장 단위로 분할하여 T5 모델을 사전 학습시키는 단계를 추가하였다. 그 결과 영어 GEC 모델 중 최고의 성능을 달성하였다. 따라서 본 연구에서는 한국어 데이터셋을 문장 단위로 분할하여 T5 모델을 사전 학습하는 실험을 진행하였다.

3. 방법론

3.1 한국어 문법 오류 노이즈

영어 GEC 시스템의 디노이징 학습 단계 [2]에서는 영어 사용자들의 빈번한 문법 오류를 노이즈로 사용하여 문법 교정 모델이 자주 발생하는 문법 오류를 더 잘 교정할 수 있도록 하였다. 따라서 한국어 GEC 시스템의 디노이징 학습 단계에서도 한국어 사용자들이 빈번하게 실수하는 문법 오류를 노이즈로 사용하여 성능을 향상하고자 하였다. 본 연구에서는 한국어 사용자들이 빈번하게 사용하는 오류를 모방하기 위해 G2P 노이즈와 편집거리 기반 자모 단위 노이즈를 사용하였다. 또한 노이즈를 추가하기 전 데이터셋에는 노이즈가 최대한 적고 문법 오류가 없어야 모델의 디노이징 학습이 제대로 이루어질 수 있기 때문에 신문 데이터셋을 사용하였다.

3.1.1 G2P 노이즈

G2P 노이즈란 Grapheme to Phoneme 노이즈로, 자소를 발음되는 대로 음소처럼 적은 노이즈를 의미한다. [11]에 따르면 한국어 문법 오류의 대부분은 문장을 발음되는 대로 적는 경우가 빈번하다. 따라서 G2P 노이즈를 사용하여 GEC 모델이 빈번한 문법 오류에 좋은 성능을 보일 수 있도록 하였다. G2P 노이즈는 편집거리 기반 자모 단위 노이즈보다 인간의 노력이 더 필요하며, 발음과 관련된 언어학적 지식을 바탕으로 생성된다. 원본 문장에 G2P 노이즈를 추가한 예시는 다음과 같다.

원본 문장: 어제는 날씨가 맑았는데, 오늘은 흐리다.

노이즈 문장: 어제는 날씨가 맑았는데, 오노른 흐리다.

3.1.2 편집거리 기반 자모 단위 노이즈

편집거리 기반 자모 단위 노이즈란 편집거리 알고리즘을 기반으로 하여 문장에 자모 단위 삽입, 삭제, 대체를 수행한 노이즈를 의미한다. 편집거리 알고리즘이란 두 문자열의 유사도를 계산하는 알고리즘으로, 한 문자열에 삽입, 삭제, 대체 연산을 몇 번 수행했을 때 다른 문자열과 같아지는지를 계산하는 알고리즘이다. 편집거리 알고리즘을 기반으로 문장에 노이즈를 추가하면, 문장에 무작위로 자모 단위 삽입, 삭제, 대체를 수행하며 이 또한 G2P 노이즈와 마찬가지로 한국어 사용자들이 자주 하는 실수를 모방한다. 편집거리 기반 자모 단위 노이즈는 언어학적 지식 없이 무작위로 쉽게 생성할 수 있다. 문장에 자모 단위 삽입, 삭제, 대체 노이즈를 추가한 예시는 다음과 같다.

원본 문장: 어제는 날씨가 맑았는데, 오늘은 흐리다.

삽입 문장: 어제는 날씨가 맑았는데, 오늘은 흐리당.

삭제 문장: 어제는 날씨가 맑았는데, 오늘은 흐리다.

대체 문장: 어제는 날씨가 맑았는데, 오늘은 흐리다.

3.2 데이터셋

노이즈가 비교적 적은 신문 데이터셋을 문장 단위로 분할했고, 각 문장에 G2P 노이즈와 편집거리 기반 자모 단위 노이즈를 추가하여 데이터셋을 구축하였다. 전체 데이터셋을 세 부분으로 분할하여 첫 번째 부분은 노이즈가 추가되지 않은 원본 문장 쌍으로, 두 번째 부분은 G2P 노이즈가 추가된 문장과 원본 문장 쌍으로, 세 번째 부분은 편집거리 기반 노이즈가 추가된 문장과 원본 문장 쌍으로 구성하였다. 비율을 조절하며 14개의 데이터셋을 생성했고, 구체적인 비율은 표 1과 같다.

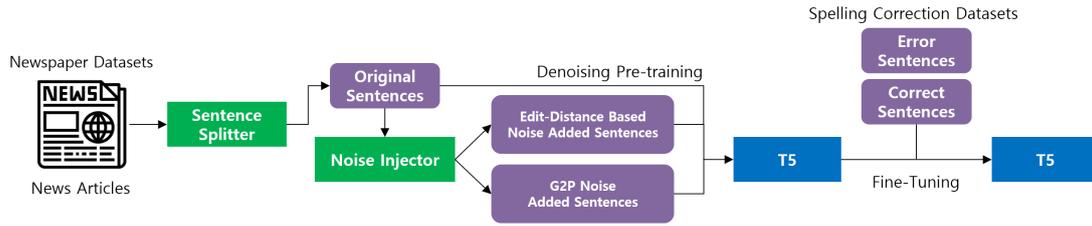


그림 1. 전체 실험 흐름도

원본-원본	G2P-원본	편집거리-원본
100%	0%	0%
80%	20%	0%
80%	0%	20%
60%	40%	0%
60%	0%	40%
40%	60%	0%
40%	0%	60%
20%	80%	0%
20%	60%	20%
20%	40%	40%
20%	20%	60%
20%	0%	80%
0%	100%	0%
0%	0%	100%

표 1. 데이터셋의 구성 비율 표

전체 데이터셋은 총 160만 문장 쌍으로 이루어져 있으며, 약 5% 정도인 8만 문장 쌍을 검증 데이터셋으로 사용하였다. [2]와 마찬가지로 원본-원본 문장 쌍을 데이터셋에 포함해 모델을 훈련시킴으로써 모델이 문법적인 부분 또한 배울 수 있도록 하였다. 각 데이터셋을 생성하는 데에는 20초에서부터 2시간까지의 시간이 걸렸으며, G2P 노이즈의 비율이 높을수록 더 많은 시간이 소요됐다. G2P 노이즈를 생성하는 모듈이 형태소 분석기를 바탕으로 작동하기 때문에 편집거리 알고리즘 기반의 노이즈만 생성할 경우 시간이 더욱 적게 걸리며, 쉽게 생성된다. 원본-원본 문장 쌍으로만 이루어진 데이터셋을 통해 노이즈 없이 문장 단위 분할을 통한 성능 향상을 검증하려 했고, 원본-원본 문장 쌍의 비율을 20%로 고정한 채 G2P 노이즈-원본 문장 쌍과 편집거리 기반 노이즈-원본 문장 쌍의 비율만을 조절한 데이터셋들을 통해 두 노이즈 중 모델의 성능에 지배적인 영향을 주는 노이즈를 찾으려 하였다. 또한 편집거리 기반 노이즈-원본 문장 쌍의 비율을 0%로 고정한 채 원본-원본 문장 쌍의 비율과 G2P 노이즈-원본 문장 쌍의 비율만을 조절한 데이터셋들을 통해 G2P 노이즈를 사용할 때 원본-원본 문장

쌍의 최적 비율을 찾고자 했고, G2P 노이즈-원본 문장 쌍의 비율을 0%로 고정한 채 원본-원본 문장 쌍의 비율과 편집거리 기반 노이즈-원본 문장 쌍의 비율만을 조절한 데이터셋들을 통해 편집거리 기반 노이즈를 사용할 때 원본-원본 문장 쌍의 최적 비율을 찾고자 하였다. 또한 G2P 노이즈만을 사용한 데이터셋과 편집거리 노이즈만을 사용한 데이터셋을 통해 어느 노이즈가 문법 교정 모델에 효과적인지 알아보려고 하였다.

3.3 모델

T5[1]는 Text-To-Text Transfer Transformer의 약자로, 최근 자연어 처리 분야에서 많이 사용되는 Transformer[6]를 기반으로 작동한다. GEC 시스템의 경우, T5 모델은 대규모의 데이터셋을 통해 사전 학습한 뒤 맞춤법 교정 데이터셋으로 미세 조정하여 사용된다. 본 연구에서는 대규모의 한국어 데이터셋으로 사전 학습되어 배포되고 있는 KE-T5[12] 모델을 사용하였다. 약 57GB의 크기, 2천만 개의 신문 기사, 웹 문서, 메신저 데이터들로 이루어진 데이터셋으로 사전 학습된 모델을 사용하였다. 해당 모델을 한국어 맞춤법 교정 데이터셋으로 미세조정된 모델을 베이스라인으로 삼아 성능을 측정했고, 해당 모델에 디노이징 사전 학습을 추가로 진행한 뒤 한국어 맞춤법 교정 데이터셋으로 미세조정된 모델의 성능을 측정하여 비교하였다.

3.4 성능지표

GEC 시스템의 성능 측정 지표로는 F0.5 점수를 사용한다. F0.5 점수는 F-beta 점수의 한 종류로, F-beta 점수는 정확도 (Precision)와 재현율 (Recall)에 가중치를 두어 조화평균을 계산한 성능 지표이다. GEC 시스템의 경우 Recall보다 Precision에 비중을 두어 모델의 성능을 측정한다. F-beta 점수의 수식은 다음과 같다.

$$F_{\beta} = \frac{(\beta^2 + 1) * Precision * Recall}{\beta^2 * Precision + Recall}$$

위 수식의 β 에 0.5를 대입하고 모델의 정확도와 재현율을 Precision과 Recall에 대입한 점수를 F0.5 점수라 하며 해당 점수를 GEC 모델의 성능 평가 지표로 사용한다.

4. 실험 및 결과

4.1 디노이징 사전 학습

한국어 데이터셋으로 사전 학습된 KE-T5[12] 모델에 3.2의 데이터셋을 사용한 디노이징 사전 학습 단계를 추가로 진행하였다. 전체 데이터셋을 6 epoch만큼 학습했고, A100 GPU 4개로 학습했을 때 1 epoch에 12시간이 소요됐으며, 사전 학습 전체에는 총 72시간이 소요됐다.

4.2 미세조정

디노이징 사전 학습이 완료된 후, 각 모델을 모두의 말뭉치 맞춤법 교정 데이터셋을 통해 미세조정을 진행하였다. 맞춤법 교정 데이터셋은 20만 쌍의 대화체 문장들로 이루어져 있으며, 문법 오류를 포함한 문장과 문법 오류가 교정된 문장이 쌍으로 구성되어 있다. 전체 20만 문장 쌍 중 약 1만 문장 쌍을 테스트 데이터셋으로 사용하였다. 학습 데이터셋은 디노이징 사전 학습 데이터셋의 12% 정도의 크기이며, 전체 데이터셋을 10 epoch만큼 학습하였다. A100 GPU 4개로 학습했을 때 1 epoch에 20분이 소요됐고, 미세조정 전체에는 총 3시간이 소요됐다. 사전 학습 데이터와 형식과 구성은 같으나, 사전 학습 데이터셋의 경우 대규모의 합성된 데이터셋이고 미세조정 데이터셋은 맞춤법 교정이 된 소규모의 데이터셋이라는 점이 다르다. 두 데이터셋의 대화 영역 또한 다르기 때문에 미세조정을 통해 모델의 훈련하는 것은 필수적이다.

4.3 실험 결과

3.2의 각 데이터셋마다 모델을 개별적으로 사전 학습시켰으며, 그 후 모두 맞춤법 교정 데이터셋으로 미세 조정된 뒤 성능 측정을 진행하였다. 총 15개의 모델을 훈련시킨 뒤 성능 측정을 진행했고, 데이터셋의 개수보다 하나가 더 많은 이유는 디노이징 학습 없이 문단 단위 사전 학습 후 맞춤법 교정 데이터셋으로 미세조정 시킨 모델을 포함하였기 때문이다. 전체 모델의 성능 표는 표2와 같다. 실험 결과 디노이징 사전 학습을 하지 않고 미세 조정은 똑같이 수행한 모델의 경우 F0.5 점수 0.797의 성능을 보였고, 원본 문장 쌍만으로 오토인코더와 같은 디노이징 사전 학습을 한 뒤 미세 조정된 모델의 경우 F0.5 점수 0.844의 성능을 보였다. 이는 문장 분할이 GEC 모델의 성능 향상에 영향을 미친다는 것으로 볼 수 있다. 또한 원본-원본 문장 쌍 20%, G2P 노이즈-원본 문장 쌍 40%, 편집거리

노이즈-원본 문장 쌍 40%의 데이터셋으로 디노이징 사전 학습한 모델의 성능이 F0.5 점수 0.868로 제일 높지만, 디노이징 학습을 수행한 다른 모델과 비교해 봤을 때 유의미한 성능 향상은 없다고 볼 수 있다. 따라서 언어학적 지식이 필요 없는 편집거리 노이즈만을 사용하더라도 언어학적 지식이 필요한 G2P 노이즈를 사용했을 때와 비슷한 성능을 보일 수 있음을 확인할 수 있다.

5. 결론

한국어로 사전 학습된 T5 모델에 추가로 문장 단위 디노이징 사전 학습을 수행하여 성능 향상을 보였다. 노이즈가 없는 데이터셋을 사용하여 문장 단위 분할의 효과를 검증하였으며, 노이즈를 추가한 데이터셋 또한 사용하여 디노이징 사전 학습의 효과를 검증하였다. 또한 G2P 노이즈를 사용하여 디노이징 사전 학습한 모델의 성능이 편집거리 기반 노이즈를 사용하여 디노이징 사전 학습한 모델의 성능과 비슷함을 보였다. 따라서 한국어 GEC 모델의 성능에 무작위로 쉽게 생성할 수 있는 편집거리 기반 노이즈가 상대적으로 많은 노력과 언어학적 지식이 필요한 G2P 노이즈에 필적하는 영향을 미치는 것을 확인하였다.

감사의 글

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2022-0-00223, (세부2) 자폐증 환자의 의사소통 능력 향상을 위한 디지털치료제 개발)

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2022-2020-0-01789)

참고문헌

- [1] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer." *J. Mach. Learn. Res.*, Vol. 21, No. 140, pp. 1-67, 2020.
- [2] S. Rothe, J. Mallinson, E. Malmi, S. Krause, and A. Severyn, "A simple recipe for multilingual grammatical error correction," *arXiv preprint arXiv:2106.03830*, 2021.
- [3] H.-C. Kwon, M.-y. Kang, and S.-J. Choi, "Stochastic korean word-spacing with smoothing using korean spelling checker," *International Journal of Computer Processing*

원본-원본	G2P-원본	편집거리-원본	사전 학습 단위	Precision	Recall	F0.5
0%	0%	0%	문단	0.806	0.764	0.797
100%	0%	0%	문장	0.851	0.818	0.844
80%	20%	0%	문장	0.865	0.839	0.860
80%	0%	20%	문장	0.868	0.841	0.862
60%	40%	0%	문장	0.867	0.840	0.861
60%	0%	40%	문장	0.870	0.844	0.865
40%	60%	0%	문장	0.869	0.843	0.863
40%	0%	60%	문장	0.870	0.846	0.865
20%	80%	0%	문장	0.867	0.843	0.862
20%	60%	20%	문장	0.870	0.848	0.865
20%	40%	40%	문장	0.872	0.848	0.868
20%	20%	60%	문장	0.872	0.849	0.867
20%	0%	80%	문장	0.871	0.847	0.866
0%	100%	0%	문장	0.868	0.844	0.863
0%	0%	100%	문장	0.871	0.866	0.866

표 2. 전체 모델 성능 표

- of Oriental Languages*, Vol. 17, No. 04, pp. 239–252, 2004.
- [4] 조승우, 권홍석, 정현영, and 이종혁, “Encoder-decoder 기반 한국어 문법 오류 교정을 위한 encoder 에서의 신경망 언어 모델 도입,” *정보과학회 컴퓨팅의 실제 논문지*, Vol. 24, No. 6, pp. 301–306, 2018.
- [5] 홍서영, 심미단, and 이대호, “Transformer 기반 한국어 채팅체 문장 교정 모델,” *한국정보과학회 학술발표논문집*, pp. 1409–1411, 2020.
- [6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, Vol. 30, 2017.
- [7] M. Lee, H. Shin, D. Lee, and S.-P. Choi, “Korean grammatical error correction based on transformer with copying mechanisms and grammatical noise implantation methods,” *Sensors*, Vol. 21, No. 8, p. 2658, 2021.
- [8] W. Zhao, L. Wang, K. Shen, R. Jia, and J. Liu, “Improving grammatical error correction via pre-training a copy-augmented architecture with unlabeled data,” *arXiv preprint arXiv:1903.00138*, 2019.
- [9] S. Chollampatt and H. T. Ng, “A multilayer convolutional encoder-decoder neural network for grammatical error correction,” *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32, No. 1, 2018.
- [10] M. Junczys-Dowmunt, R. Grundkiewicz, S. Guha, and K. Heafield, “Approaching neural grammatical error correction as a low-resource machine translation task,” *arXiv preprint arXiv:1804.05940*, 2018.
- [11] C. Park, S. Jeong, K. Yang, S. Lee, J. Joe, and H. Lim, “Korean spell correction based on denoising transformer,” *Annual Conference on Human and Language Technology*, pp. 368–372, 2019.
- [12] K. AIRC, “Ke-t5: Korean english t5,” Mar. 2021. [Online]. Available: <https://github.com/AIRC-KETI/ke-t5>