

한국어 문장형 수학 문제 풀이를 위한 Conformer 기반의 심층 산술 풀이모델

김강민¹, 박현우¹, 박시웅², 전찬준^{1,*}

¹조선대학교 컴퓨터공학과

²한국전자통신연구원 호남권연구센터

could714@chosun.kr, 20213120@chosun.kr, swp@etri.re.kr, cjchun@chosun.ac.kr

I. 서론

문장형 수학 문제(Math Word Problem, MWP)는 수식과 같은 수학적 언어가 아닌 자연어로 서술된 수학 문제를 의미한다. '5 + 3' 과 같이 주어진 수식을 풀어 내는 단순 계산문제와 달리 문장형 수학 문제에서는 주어진 상황을 이해하고, 이를 토대로 대응하는 풀이식을 세워야 한다. 이로 인해 문장형 수학 문제 풀이모델은 문제 해결력에 더불어 독해력과 추론력이 요구된다.

Conformer는 음성 데이터의 지역적 정보와 전역적 정보를 유기적으로 활용하기 위해 합성곱 신경망과 Transformer를 결합한 모델이다. 해당 모델은 음성 인식과 화자 분리와 같은 음성 분야에서 Transformer보다 뛰어난 성능을 보이며 활발히 연구되고 있다 [1, 2].

Conformer의 데이터 특징 정보 활용 방식은 자연어 처리에서도 문장 내 단어 간의 대응 관계 및 특징 학습에 효과를 보일 수 있을 것으로 예상된다. 이에 본 논문에서는 Conformer를 문장형 수학 문제 풀이에 적용하여, 모델 구조로 인한 독해력과 추론력 향상을 Transformer와 비교 분석한다.

II. 한국어 데이터 세트와 학습모델 및 방법

수학 문제 풀이 시스템 설계를 위한 연구는 국내보다 국외에서 활발하기에, 이를 위한 공개된 벤치마크 데이터는 주로 영어와 중국어로 구성되어 있다. 해당 문제로 인해 이전에 진행된 한국어 기반의 수학 문제 풀이 연구에서는 영어 데이터 세트를 한국어로 번역한 후, 의미권 표현을 수정하는 과정을 거친다 [3]. 본 논문에서는 이러한 과정 대신, [4]에서 제안한 한국어 문장형 수학 문제 데이터 생성기를 활용해 데이터를 수집한다. 해당 데이터 생성기에는 산술 연산, 순서 정하기, 수 찾기, 도형으로 구성된 초등교과 수준의 문제 유형이 존재한다. 또한, 접사 및 피연산자 변환과 같은 데이터 변환 방식이 존재하기 때문에 다양한 문장 표현과 피연산자를 갖는 데이터를 필요한 만큼 생성할 수 있다.

실험에 쓰인 모델은 Transformer와 Conformer이다 [1, 5]. Conformer 모델은 Transformer의 기본적인 구조에서 인코더만 Conformer 인코더로 변경한 모델을 의미한다. 이는 Conformer가 지역적, 전역적 특징 추출

에 특화되어 있으므로, 정보 추출 역할을 맡는 인코더 부분만 변경하는 것이 적합하다. 최적화에는 Adam을 사용하며, 임베딩 방식은 FastText를 적용한다 [6]. 각 모델의 임베딩과 은닉 상태 및 배치의 크기는 256이며, 학습률은 5e-4를 적용한다. 드롭아웃은 0.4로 설정하며, 멀티 헤드는 8개를 사용한다. 모델에 적용되는 FFN의 크기는 512로 설정하며, 모델의 인코더와 디코더 층수는 3개씩 쌓는다. Conformer에는 깊이별 합성곱 연산이 추가로 요구되기에 이를 위한 커널 크기를 31로 설정한다.

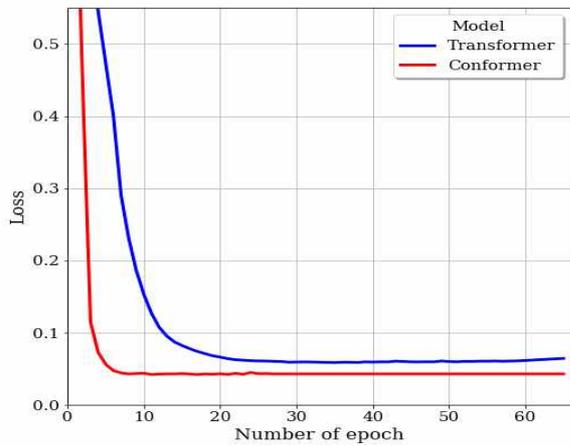
III. 실험 결과

모델의 성능 측정을 위한 평가지표로 BLEU (Bilingual Evaluation Understudy) 점수와 정확도를 사용한다. BLEU 점수는 모델의 예측 수식이 정답 수식과 얼마나 유사한지를 나타내며, 정확도는 예측 수식으로 얼마나 정답을 맞혔는지를 나타낸다.

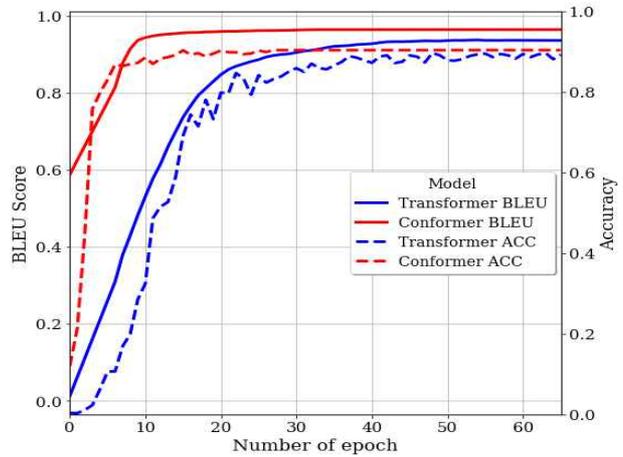
본 논문에서는 생성기를 통해 학습 데이터 15만 개와 검증 데이터 3만 개를 생성하였다. 평가 데이터는 모델의 성능뿐만 아니라 생성 데이터에 과적합되었는지 평가를 위해 시중에서 판매되는 수학 문제집에서 수집한 311개의 데이터를 활용했다.

그림 1은 검증 데이터 세트에서의 각 모델의 손실 값과 평가지표에 대한 성능을 보여준다. 그림 1-(a)에서는 손실 값의 수렴을 통해 모델의 학습 속도를 파악할 수 있다. Transformer는 66번의 학습 끝에 과적합이 일어나 종료됐지만, Conformer는 10번 이전에 Transformer보다 낮은 손실 값을 기록했으며, 28번째에 학습이 종료되어 Transformer보다 빠른 학습 속도를 보였다. 그림 1-(b)에서는 평가지표를 통한 각 모델의 성능을 개괄적으로 확인할 수 있다. 본 그림을 통해 Conformer가 모든 평가지표에서 Transformer보다 높은 성능을 보임을 확인할 수 있어, 학습 속도뿐만 아니라 성능적으로도 Transformer보다 뛰어남을 확인할 수 있다. 검증 데이터 세트에서의 정량적 성능은 다음 표 2에서 확인할 수 있다.

표 1은 검증 및 평가 데이터 세트에서 각 모델의 정량적 성능 결과를 나타낸다. 생성된 검증 데이터 세트에서는 Conformer가 BLEU 점수에서 95.91%, 정확도는



(a) 각 모델의 손실 값 결과



(b) 각 모델의 BLEU 점수, 정확도 성능 결과

그림 1. 검증 데이터 세트에서의 각 모델의 손실 값과 평가지표에 따른 성능 결과

표 1. 검증 및 평가 데이터 세트에서의 각 모델의 정량적 성능 결과

모델	검증		평가	
	BLEU (%)	정확도 (%)	BLEU (%)	정확도 (%)
Transformer	93.61	89.05	23.23	34.72
Conformer	95.91	89.61	26.76	36.01

89.61%를 기록하며 각각 Transformer보다 2.3%, 0.56% 우수한 성능을 보였다. 일반화된 풀이 능력을 측정하기 위한 평가 데이터 세트에서는 이러한 성능 차이가 더욱 두드러졌다. 학습에 등장하지 않은 문제의 유형과 단어로 인해 두 모델 모두 검증 데이터 세트에서보다 성능 하락이 있었지만, Conformer는 BLEU 점수에서 26.76%, 정확도에서는 112개를 맞춘 36.01%를 달성하며 Transformer보다 3.53%, 1.29% 더 우수한 성능을 보였다. 이러한 성능의 차이를 통해 Conformer의 지역적, 전역적 정보를 반영하는 모델의 구조가 학습과 수식 추론에 더 효율적임을 확인하였다.

IV. 결론

본 논문에서는 생성기 기반의 한국어 문장형 수학 문제 데이터 세트를 활용하여 Conformer 모델과 Transformer 모델의 성능을 비교하였다. 데이터 세트에는 초등학교 수준의 산술 연산, 순서 정하기, 수 찾기, 도형 유형을 포함하고 있다. 또한, 시중에 유통되는 수학 문제집에서 수집한 311개의 데이터를 평가 데이터를 활용하였다. 두 모델의 성능을 평가한 결과, Conformer 기반의 모델이 Transformer 모델보다 실제 문제에서도 비교적 강건한 성능을 보임을 확인하였다.

ACKNOWLEDGMENTS

본 연구는 2023년도 연구개발특구진흥재단의 지원을 받아 지역의 미래를 여는 과학기술 프로젝트의 일환으로 수행된 연구임 (인공지능 기반 메타버스 구현을 위한 융·복합 문화 가상 스튜디오)

참고문헌

- [1] A. Gulati, J. Qin, C-C. Chiu, N. Parmar, Y. Zhang, J. Yu, W. Han, S. Wang, Z. Zhang, Y. Wu, and R. Pang, "Conformer: Convolution-augmented Transformer for speech recognition," in *Proc. Interspeech 2020*, pp. 5036-5040, Oct. 2020.
- [2] S. Chen, Y. Wu, Z. Chen, J. Wu, J. Li, T. Yoshioka, C. Wang, S. Liu, and M. Zhou, "Continuous speech separation with Conformer," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pp. 5749-5753, June 2021.
- [3] K. S. Ki, D. S. Lee, and G. Gweon, "KoTAB: Korean template-based arithmetic solver with BERT," in *Proc. 2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, pp. 279-282, Dec. 2020.
- [4] K. Kim and C. Chun, "Synthetic data generator for solving Korean arithmetic word problem," *Mathematics*, vol. 10, no. 19, pp. 3225-3241, Sep. 2022.
- [5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Advanced in Neural Information Processing Systems(NeurIPS)*, vol. 30, pp. 5998-6008, Dec. 2017.
- [6] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135-146, June 2017.