

# 진단 시뮬레이션을 위한 Conformer 기반의 질의응답 텍스트 모델 설계

김강민, 신영서, 최우열, 전찬준\*

조선대학교 컴퓨터공학과

{could714, ys070400}@chosun.kr, {wyc, cjchun}@chosun.ac.kr

## Conformer-based QA Text Model for Diagnosis Simulation

Kangmin Kim, Yeongseo Shin, Wooyeol Choi, Chunjun Chun\*

Department of Computer Engineering, Chosun University

### 요 약

임상실습 교육은 학생들에게 이론적 지식을 기초로 전문 역량 및 현장 대처능력을 개발시켜 전문적인 의료인으로서의 성장을 도모한다. 표준화 환자 기법은 임상실습 교육 중 하나로, 환자 역할을 하는 연기자를 활용하는 교육 방식이다. 이 기법은 의료인들에게는 실제 환자와의 면담 상황을 익숙해지도록 돕고 교육자들에게는 학생들의 임상수행 능력 평가를 돕는 도구이다. 그러나 표준화 환자를 이용한 실습은 연기자 활용 과정에서 소모되는 인건비와 그들을 교육할 전문 인력 부족의 문제와 마주하게 된다. 본 논문에서는 임상진단 시나리오의 질의응답 텍스트와 Conformer를 활용한 기계번역 기반의 QA 모델을 통해 연기자를 대체할 수 있는 AI-환자를 제안한다.

### 1. 서 론

임상실습 교육을 통해 예비 의료인들은 이론 교육에서 학습한 지식을 임상현장에 적용하고 현장에서의 대처능력을 습득하며 전문적인 의료인으로서 성장한다. 이를 위한 임상실습 교육 방식 중 하나가 시뮬레이션 기반 의학 교육이다. 시뮬레이션 기반 교육은 제한된 실습 환경에서 학생들에게 추가적인 실습의 기회와 안전한 실습 환경을 제공하기 위한 대안으로서 의학 및 간호 교육 분야에서 활용되고 있다. 강의식 교육과 비교해볼 때 시뮬레이션 교육의 최대 장점은 학습자의 지식과 기술을 향상시킬 뿐만 아니라 터득한 지식을 기반으로 문제해결 능력과 임상수행 능력을 효과적으로 높인다는 것이다 [1].

표준화 환자는 시뮬레이션 교육 방식 중 하나이며, 학생들의 임상 수행능력 교육 및 평가를 위해 실제 환자의 병력과 신체검사 결과를 재현하도록 훈련받은 연기자를 활용한다. 이 방식은 학생과 전공의(resident) 교육에 있어서 실제 환자와의 면담을 능숙하게 하고 그들의 임상 실행능력을 평가하는 데 있어 신뢰성 있는 도구로 평가된다 [2]. 이러한 이점으로 인해 표준화 환자 기법은 전 세계적으로 객관화 구조적 임상 시험에 적극적으로 활용되고 있다. 그러나 연기자를 활용한 표준화 환자는 지출되는 인건비뿐만 아니라 숙련된 표준화 환자 연기자의 부족과 연기자 교육을 위한 전담 인력의 부족 문제를 겪는다 [3].

이에 본 논문에서는 자연어 처리 분야에 커다란 영향을 미친 Transformer와 최근 음성 분야에서 두각을 나타내는 Conformer를 활용한 기계번역 기반의 QA 모델을 통해 치과 표준화 환자 임상진단에서 연기자를 대체하는 AI-환자를 구현함으로써 앞서 언급한 문제를 해결하고자 한다.

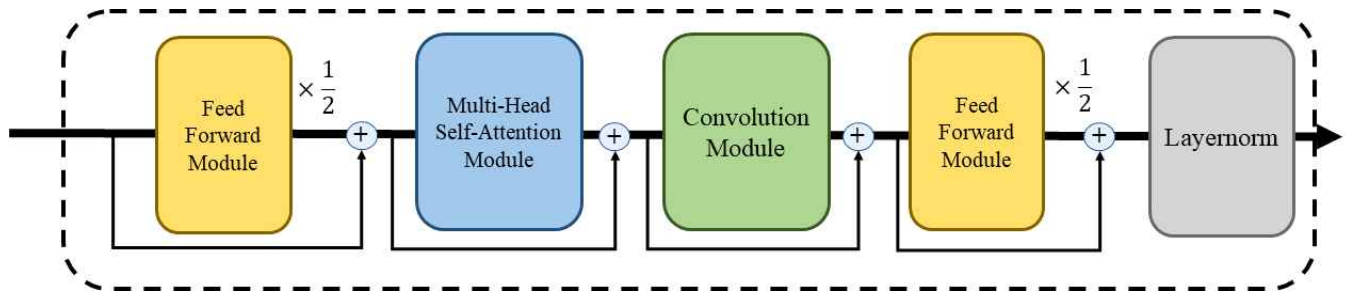
### 2. 한국어 임상진단 데이터와 학습모델 및 방법

의료 분야와 관련된 데이터의 특성상 치과 의사들의 실제 문진 데이터의 확보에는 어려움이 있다. 그러므로 본 논문에서는 표준화 환자 연기자에게 주어지는 몇 가지의 환자 진단 시나리오 대본을 기반으로 의사의 질문과 환자의 대답을 생성하는 한국어 데이터 생성기를 구현하여 AI-환자의 학습을 위한 데이터를 수집한다. 데이터 생성기의 질문 유형은 환자 확인, 임상 증상, 병력, 생활 습관, 진단으로 분류된다. 또한, 각 유형은 세부 유형이 있으며, 총 19가지의 세부 유형이 존재한다. 본 데이터 생성기는 모델에게 다양한 구문 표현을 학습시키기 위해 문장에 어휘적 데이터 증대 방식을 적용하며, 증대 방식으로는 접사 변환, 단어 변환, 독립어 추가가 사용된다.

의사와 환자 간의 대화에서는 진단을 위한 대화뿐만 아니라, 의사가 환자의 긴장감을 풀어주기 위해 간단한 잡담 또한 오갈 수 있다. 그러므로 본 논문에서는 AI-환자에게 간단한 잡담에 대한 적절한 응답을 학습시키기 위해 AI-Hub의 민원 상담 데이터에서 수집한 보편적인 질의응답 데이터를 추가적인 학습 데이터로 활용한다.

본 실험에는 Transformer와 Conformer 모델이 사용된다 [4, 5]. Conformer는 음성 분야에서 음성 데이터의 지역적 정보와 전역적 정보를 종합적으로 활용하기 위해 합성곱 신경망과 Transformer의 인코더를 결합한 모델이다. 해당 모델은 화자 분리와 음성 인식 분야에서 Transformer보다 우수한 성능을 보이며 활발히 연구되고 있다 [5, 6]. Conformer의 데이터 특징 활용 방식은 음성과 같이 순차적 데이터를 다루는 자연어 처리 분야에서 단어 특징 벡터 학습과 문장 내 단어 간의 대응 관계 학습에도 효과

\* 교신저자



(그림 1) Conformer 모델의 인코더 구조

를 보일 수 있을 것으로 전망된다. 다음 그림 1은 Conformer 모델의 인코더 구조를 나타낸다. Conformer 모델은 기존의 Transformer 모델에서 인코더 부분만 Conformer 인코더로 변경한 모델이다. Conformer는 입력 데이터의 특징을 파악에 특화되어 있으므로 디코더는 Transformer의 구조를 따른 채 인코더만 변경하였다.

모델에 적용된 초매개변수로 배치와 임베딩 및 은닉 상태의 크기는 두 모델 모두 256으로 적용되며, 멀티 헤드는 8, FFN의 크기도 512로 설정하였다. 드롭아웃과 학습률은 각각 0.4와 5e-4로 학습하였으며, Transformer의 인코더와 디코더는 각각 5개의 층으로 구성하였으며, Conformer는 각각 4개의 층을 활용하였다. Conformer에는 Transformer와 달리 1D-깊이별 합성곱 연산이 이루어지기에, 이를 위한 커널 크기를 31로 설정하였다. 최적화는 Adam을 적용하며, 임베딩 방식은 FastText [7]를 활용하였다.

(표 1) 데이터 세트에 따른 각 모델의 정량적 성능 결과

Model	검증		평가	
	BLEU(%)	WER(%)	BLEU(%)	WER(%)
Transformer	94.79	5.14	92.21	8.08
Conformer	96.40	1.04	96.13	1.34

### 3. 실험 결과 및 결론

각 모델의 성능을 측정하기 위해 본 논문에서는 BLEU (Bilingual Evaluation Understudy) 점수와 WER (Word Error Rate)를 사용하였다. BLEU 점수는 모델의 대답과 실제 대답의 전체적인 유사성을 계산하며, WER은 예측과 실제 대답 간의 세부적인 단어 수준의 오류율을 측정한다.

본 논문에서는 AI-환자의 학습을 위해 AI-Hub에서 수집한 67,124개의 보편적인 질의응답 데이터와 한 명의 표준화 환자 정보로 생성된 18,378개의 개인화 데이터를 사용하였다. 학습 데이터 세트는 보편 데이터 전체와 개인화 데이터의 60%를 합친 78,150개로 구성되었으며, 모델이 주된 목표는 진단 질문에 대한 응답이므로 검증 및 평가 데이터 세트는 개인화 데이터의 남은 20%씩 할당되어 각각 3,676개의 데이터로 이루어졌다.

표 1은 검증과 평가 데이터 세트에서 각 모델의 정량적 성능을 제시한다. Transformer의 평가 결과는 검증 데이터 세트의 결과보다 BLEU score는 2.58% 낮아지고 WER은 2.94% 높아져 각각 92.21%, 8.08%를 기록하였다. Conformer 또한 평가 시에 검증 데이터 세트의 결과보다는 성

능 감소가 있었지만, 최종적으로 BLEU score와 WER에서 각각 96.13%, 1.34%를 달성하여 Transformer보다 3.92%, 6.74% 개선된 성능을 확인하였다.

본 논문에서는 치과 임상진단을 위해 연기자를 대체할 Conformer 기반의 AI-환자를 제안하였다. 실험 결과 Conformer 기반의 AI-환자 모델은 Transformer 모델보다 우수한 성능을 보인 것을 확인하였다.

### 감사의 글

본 연구는 2023년도 연구개발특구진흥재단의 '기술사업화 협업 플랫폼' 사업으로 수행되었음 (과제고유번호: 1711177250)

### 참고 문헌

- [1] S. Cassidy, "Subjectivity and the valid assessment of pre-registration student nurse clinical learning outcomes: Implications for mentors," *Nurse Education Today*, vol. 29, pp. 33-39, Jan. 2009.
- [2] H. S. Barrows, "An overview of the uses of standardized patients for teaching and evaluating clinical skills. AAMC," *Academic Medicine*, vol. 68, no. 6, pp. 443-451, June 1993.
- [3] S. Shorey, E. Ang, J. Yap, E. D. Ng, S. T. Lau, and C. K. Chui, "A virtual counseling application using artificial intelligence for communication skills training in nursing education: Development study," *Journal of Medical Internet Research*, vol. 21, no. 10, pp. e14658, Oct. 2019.
- [4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. 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.
- [5] 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*, pp. 5036-5040, Oct. 2020.
- [6] 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)*, pp. 5749-5753, June 2021.
- [7] 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.