반지형 웨어러블 인터페이스를 이용한 1D CNN-BiLSTM 기반의 수화 단어 인식

김예린, 홍동진, 장원두*

*부경대학교 인공지능융합학과

yerin2919@pukyong.ac.kr, djhong91@pknu.ac.kr, *chang@pknu.ac.kr

1D CNN-BiLSTM based Sign Language Recognition Using a Ring-Type Wearable Device

Ye-Rin Kim, Dong-Jin Hong, Won-Du Chang*

Division of Computer Engineering and Artificial Intelligence, Pukyong National University

요 약

본 연구에서는 일상생활에서 착용할 수 있는 반지형 웨어러블 인터페이스를 이용하여 수화 단어를 인식하는 모델을 제안한다. 반지형 웨어러블 인터페이스는 내장된 가속도 및 자이로 센서를 이용하여 손가락 움직임을 3차원 좌표값으로 추적하고 기록한다. 제안하는 모델은 수화 동작의 국소 패턴과 시간적 의존성을 동시에 학습할 수 있도록 1D CNN과 BiLSTM을 결합한 하이브리드 모델로 설계되었다. 실험은 20명의 사용자로부터 50개의 단어를 각 5회 반복 수행하여 수집된 총 5,000개의 시계열 데이터셋을 대상으로 수행되었다. 실험 결과, 제안하는 1D CNN-BiLSTM 하이브리드 모델은 1D CNN, BiLSTM 대비 정확도가 각각 12.93%p, 5.43%p 향상됨을 확인하였다. 본 연구는 반지형 웨어러블 인터페이스를 이용한 수화 인식 시스템의 실용적 가능성과 사용자 일반화 성능을 확보하기 위한 기반이 될 것이다.

Ⅰ. 서 론

수화 인식 기술은 청각장애인의 의사소통 지원 및 인간-컴퓨터 상호작용(HCI) 관점에서 중요한 응용 분야로 주목받고 있다[1]. 수화 인식 기술은 크게 카메라 기반의 비전 방식과 웨어러블 센서 기반 방식으로 구분된다. 기존의 비전 기반 수화 인식은 사용자가 외부 장치를 착용하지 않아도 된다는 장점이 있으나, 조명 변화나 배경의 영향이 크며, 인식 시스템을 보유한 청자와만 대화가 가능하다는 한계가 있다[2]. 웨어러블 센서 기반 방식은 이러한 한계를 극복할 수 있는 대안으로 제시되고 있으나, 기존 연구에서 사용된 데이터 글러브나 굽힘 센서 기반 장치는 착용 부담이 크고 촉각을 제한하여 일상적인 활동에서 사용하기 어렵다는 문제가 있다[3]. 본 연구에서는 이러한 한계로부터 비교적 자유로운 반지형 웨어러블 인터페이스를 이용한 수화 단어 인식 모델을 제안한다.

Ⅱ. 본론

1. 데이터 수집

본 연구에서는 수화 단어 데이터를 수집하기 위해 TapStrap2를 사용하였다. TapStrap2는 양손의 모든 손가락으로 착용하는 웨어러블 기기로서, 각 손가락 모듈에는 가속도 센서가 내장되어 있으며 엄지손가락 모듈에는 자이로 센서가 추가로 내장되어 있다. 각 센서는 3차원 공간에서의 이동을 추적하기 위해 3축의 좌표 이동에 관한 정보를 저장하고 있다. 따라서 TapStrap2를 양손에 착용하여 손가락의 움직임을 추적할 때는 각 수집 프레임마다 10개의 3축 가속도 정보와 2개의 3축 자이로 정보를 더하여 총 36채널의 신호가 기록된다. 그림 1은 실험에 TapStrap2의 외형을 나타낸 그림이다.

그림 1. TapStrap2 웨어러블 기기

2. 모델 구조

수집된 IMU 기반 수화 시계열 데이터를 효과적으로 인식하기 위해 그림 2와 같이 1D CNN과 BiLSTM을 결합한 하이브리드 구조로 설계하였다. 먼저 입력 시퀀스는 커널 크기가 1인 Conv1D를 통해 초기 특징 시퀀스로 변환된 후, 세 개의 커널 크기가 7인 Conv1D를 통과하여 축약된 시계열 특징 시퀀스로 변환된다. 커널 크기가 7인 각 Conv1D층의 출력이 다음 층에 전달되기 전 단계에서는 특징 벡터의 정규화를 수행하고 비선형성을 갖도록 배치 정규화와 ReLU 활성화 함수를 공통으로 적용한다. 또한 커널 크기가 7인 Conv1D 사이에는 ReLU 활성화 함수를 적용한 후, 커널 크기가 2인 Max Pooling을 이용해 시간축을 단계적으로 축소한다.

제안하는 모델은 모든 Conv1D를 통과한 특징 시퀀스를 BiLSTM을 통해 양방향으로 시간적 문맥 정보를 학습하게 하고, BiLSTM의 출력을 Mean Pooling과 Max Pooling에 각각 전달한 후 연결하여 최종 특징 시퀀스를 만든다. 최종 특징 시퀀스는 Dense와 Softmax 연산을 통해 분류되어 인식된 단어를 출력한다.

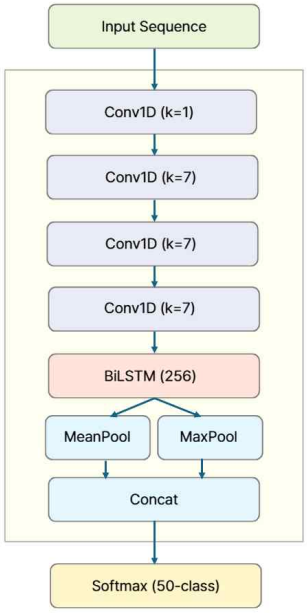

그림 2. 제안하는 수화 단어 인식 모델 구조

Ⅲ. 실험

1. 데이터셋

실험에 사용된 데이터셋은 20명의 피실험자에게 TapStrap2를 착용한 상태에서 일상 생활에서 자주 사용되는 50가지의 단어를 수어로 표현한 데이터로 구성되었다. 단어마다 데이터를 5회 수집하였으며, 3초 이내에 표현하도록 제한하였다. 데이터셋은 이러한 과정을 통해 총 5,000개의 데이터로 만들어졌다.

수집된 데이터에는 사용자별 인터페이스 착용 위치나 동작의 크기 차이에 따라 데이터 분포가 달라지는 것을 보완하기 위한 전처리 과정이 수행되었다. 전처리 과정에서는 양손 가속도 및 자이로 신호의 공통 시간 구간을 추출한 뒤, 이를 200Hz 기준으로 보간하여 시간축을 통일하였다. 이후 채널별 z-score 정규화를 적용하여 신호의 스케일을 보정하였다.

5,000개의 데이터 중 3,500개의 데이터를 모델의 훈련에 사용하였고, 검증과 테스트에는 각각 750개의 데이터가 사용되었다. 이때, 동일 사용자의 데이터가 훈련, 검증 및 테스트 데이터에 동시에 포함되지 않도록 사용자 독립 분할 방식으로 검증하였다.

2. 실험 결과

제안된 수화 인식 모델의 인식 성능이 얼마나 향상되었는지 확인하기 위해 기반 모델인 1D CNN과 BiLSTM을 동일한 조건으로 학습한 후 정확도와 Macro-F1을 기준으로 평가하였다. 모델의 학습 조건은 표 1에 나타내었으며, 모든 비교 모델에 공통으로 적용되었다.

실험 결과, 표 2와 같이 제안된 1D CNN-BiLSTM 하이브리드 모델이 모든 평가 지표에서 가장 우수한 성능을 보였다. 제안된 모델은 86.53%의 정확도와 0.8615의 Macro-F1를 달성하여 1D CNN 모델 대비 12.93%p, BiLSTM 모델 대비 5.43%p의 정확도 향상을 확인할 수 있었으며, Macro-F1은 각각 0.1198, 0.0505만큼 향상되었음을 알 수 있었다. 제안된 모델은 단순한 형태의 시계열 인식 모델보다 시간적 문맥과 지역적 패턴을 학습하는 능력이 높아 실용적인 반지형 웨어러블 인터페이스를 이용한 수화 인식 시스템의 기반 모델이 될 수 있음을 보여주었다. 특히 1D CNN이 센서 신호의 국소적인 특징을 정교하게 추출하고, BiLSTM이 해당 특징의 장기 시간 의존성을 반영하여 문맥 정보를 보완함으로써 단일 모델 대비 더 높은 구분 성능을 달성한 것으로 판단된다.

표 1. 비교 모델들의 학습 조건

항목	값
Epoch	70
Batch size	64
Optimizer	AdamW
Learning rate	5×10^{-4}
Loss function	Weighted Cross-entropy + Label Smoothing
Augmentation Methods	Time Masking (ratio=0.12, p=0.7) Channel Dropout (rate=0.15, p=0.7) Gaussian Noise (std=0.01, p=0.7) Mix-up (alpha=0.4, p=0.5)

표 2. 제안된 모델과 기반 모델의 수화 단어 인식 성능 비교

비교 모델	정확도(%)	Macro-F1
1D CNN	73.60	0.7417
BiLSTM	81.10	0.8110
제안된 모델	86.53	0.8615

Ⅳ. 결론

본 연구는 반지형 웨어러블 인터페이스를 이용한 1D CNN-BiLSTM 기반의 수화 단어 인식 모델을 제안하였다. 제안된 모델은 입력된 시계열 신호로부터 지역적인 특징을 추출하는 1D CNN과 장기 시간 의존성을 학습하는 BiLSTM을 결합한 하이브리드 구조를 통해 20명의 피실험자로부터 수집된 50가지의 수화 단어를 86.53%의 정확도로 인식할 수 있었고, 0.8615의 Macro-F1을 달성하여 모든 클래스에 대하여 비교적 균일하게 높은 인식 성능을 보여주었다.

본 연구는 착용성이 우수한 반지형 웨어러블 인터페이스 기반의 수화 인식 시스템의 실용 가능성을 제시하였으며, 교차 사용자 환경에서의 사용자 일반화 성능 향상을 실험적으로 검증하였다. 향후 연구에서는 연속적으로 입력되는 수화 단어 인식 모델로 확장하고, 개인화 적응 기법, 모델 경량화 등을 수행하여 실사용 환경에서의 적용성을 개선할 것이다.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00334159),

참 고 문 헌

[1] K. Kudrinko et al., “Wearable sensor-based sign language recognition: A comprehensive review,” IEEE Reviews in Biomedical Engineering, vol. 14, pp. 82 - 97, 2020.

[2] Wang, X., et al., “Sign Language Translation Using Frame and Event Stream: Benchmark Dataset and Algorithms,” arXiv, arXiv:2503.06484, 2025.

[3] Mummadi, C. K. et al. "Real-time and embedded detection of hand gestures with an IMU-based glove," Informatics, vol. 5, no. 2, pp. 28, 2018.