낙상 감지 및 활동 인식을 위한 다중 헤드 주의 메커니즘 기반 경량 시간적 합성곱 신경망

손창식*, 강원석

대구경북과학기술원 지능형로봇연구부

*changsikson@dgist.ac.kr, wskang@dgist.ac.kr

A lightweight temporal convolutional network with multi-head self-attention mechanism for fall detection and activity recognition

Chang-Sik Son*, Won-Seok Kang

DGIST, Division of Intelligent Robot

요 약

본 논문은 다중 센서 신호를 기반으로 낙상 및 신체 활동을 효과적으로 인식할 수 있는 다중 헤드 주의 메커니즘 기반의 경량 시간적 합성곱 신경망을 제안한다. 특히, 시간적 합성곱 블록을 통해 인과적 정보와 장기 시간 의존성을 효과적으로 포착하면서도 훈련 파라미터 수를 줄이는 구조적 효율성을 달성하였다. 또한, 다중 헤드 주의 메커니즘은 시간과 채널 간의 상호 관계를 학습하여 중요한 시점의 특징을 강조하고 불필요한 정보는 억제하도록 설계하였다. 제안된 모델은 UMAFall과 UCI-HAR 공개 데이터셋에서 3가지 낙상 유형과 6가지 일상생활 동작 인식에 적용되었으며, ResNet18 기반 경량 모델 (LRN) 대비 파라미터 수 (0.077M)를 약 3배 감소시키면서도, 낙상 감지에서 1.5%, 일상생활 동작 인식에서 0.6% 향상된 매크로 F1 점수를 기록하여 성능과 효율성 모두에서 우수함을 입증하였다.

Ⅰ. 서 론

인간 활동 인식은 사람이 수행하는 특정 움직임이나 행동을 감지하고 해석하는 데 중점을 둔 역동적이고 광범위한 연구 분야이다. 최근에는 환자의 일상 활동과 움직임을 추적하거나, 노인의 낙상을 감지하거나, 도난 및 강도와 같은 사건을 실시간 탐지하는 시스템 등 다양한 분야에 적용되고 있다. 딥러닝 기술을 활용한 인간 활동 인식 모델이 활발히 연구되고 있으나, 여전히 실시간 처리와 인식 성능 간의 균형을 효율적으로 해결하는 방법이 요구된다. 본 연구에서는 다중 센서 신호를 기반으로, 낙상 및 일상생활 동작을 효과적으로 인식하기 위한 경량의 시간적 합성곱 구조를 제안한다. 제안하는 모델은 다중 헤드 주의 메커니즘을 적용하여 다양한 동작의 특성을 보다 정확하게 포착할 수 있도록 설계하였다.

Ⅱ. 본론

낙상 감지 및 신체 활동 인식을 위한 다중 헤드 주의 메커니즘 기반의 경량 시간적 합성곱 신경망 구조는 그림 1에 제시되어 있다. 제안된 구조는 합성곱 임베딩, 시간적 합성곱, 트랜스포머 인코더, 추론의 네 가지 주요 블록으로 구성된다.

2.1. 경량 시간적 합성곱 신경망 구조

합성곱 임베딩 블록은 1차원 합성곱 (Conv1D), 배치 정규화, ReLU 활성화 함수로 구성되며, 가속도 및 각속도와 같은 다중 센서 신호로부터 시간적 지역 특징을 효과적으로 추출하는 데 사용된다. 이어서, 시간적 합성곱 블록은 이러한 지역 특징으로부터 인과적 정보 [1]와 더 긴 시간적 의존성 [2] 을 포착할 수 있도록 설계되었다. 트랜스포머 인코더 블록에서는 다중 헤드 주의 메커니즘 [3]을 적용하여, 입력 시퀀스의 모든 시간적 지점과 채널 간의 상호 관계를 동시에 고려함으로써 낙상 또는 특정 신체 활동이 발생한 시점의 중요한 특징 정보를 강조하고, 보다 표현력 있는 특징을 추출할 수 있도록 하였다. 마지막으로, 전역 평균 풀링 계층은 다양한 지역 특징의 차원을 효과적으로 축소하는 데 활용되었다.

2.2. 데이터셋

제안된 경량 합성곱 신경망의 성능을 평가하기 위해, 두 가지 공개 데이터셋인 UMAFall [4]과 UCI-HAR [5]을 사용하였다. UMAFall은 총 19명의 참가자로부터 수집되었으며, 가슴, 허리, 손목, 발목의 4개 신체 부위에 부착된 센서 태그와 오른쪽 주머니에 위치한 스마트폰을 통해 데이터를 기록하였다. 해당 데이터셋은 3가지 낙상 유형 (전방, 후방, 측면 낙상)과 12가지 일상생활 동작 (activities of daily living, ADL)을 포함한다. 본 연구에서는 15가지 활동 중, 가슴 부위의 센서 태그에서 수집된 9채널 신호 (3축 가속도계, 3축 자이로스코프, 3축 지자기 센서; 20Hz)를 활용하였으며, 이를 기반으로 낙상 여부를 탐지하는데 활용하였다. UCI-HAR 데이터셋은 30명의 지원자를 대상으로 수집되었으며, 각 지원자는 스마트폰을 허리에 착용한 채 6가지 일상생활 활동 (걷기, 계단 오르기/내리기, 앉기, 서기, 누워 있기)을 수행하였다. 스마트폰에 내장된 3축 가속도계와 3축 자이로스코프를 통해 선형 가속도와 각속도가 50Hz의 일정한 샘플링 속도로 측정되었다.

2.3. 데이터 전처리 및 분할

본 연구에서는 낙상 감지 및 신체 활동 인식의 성능을 보다 객관적으로 평가하기 위해, Z-점수 표준화와 슬라이딩 윈도우 기반 시퀀스 분할이라는 최소한의 전처리 기법만을 적용하였다. UMAFall 데이터셋에서는 가슴 부위에 부착된 센서 태그로부터 수집된 9채널 신호를 고정된 윈도우 길이 1초 (20개 샘플), 이동 간격 0.5초 (10개 샘플)를 이용하여 시퀀스로 분할하였다. 훈련과 실험용 데이터는 무작위로 70%와 30% 비율을 가지

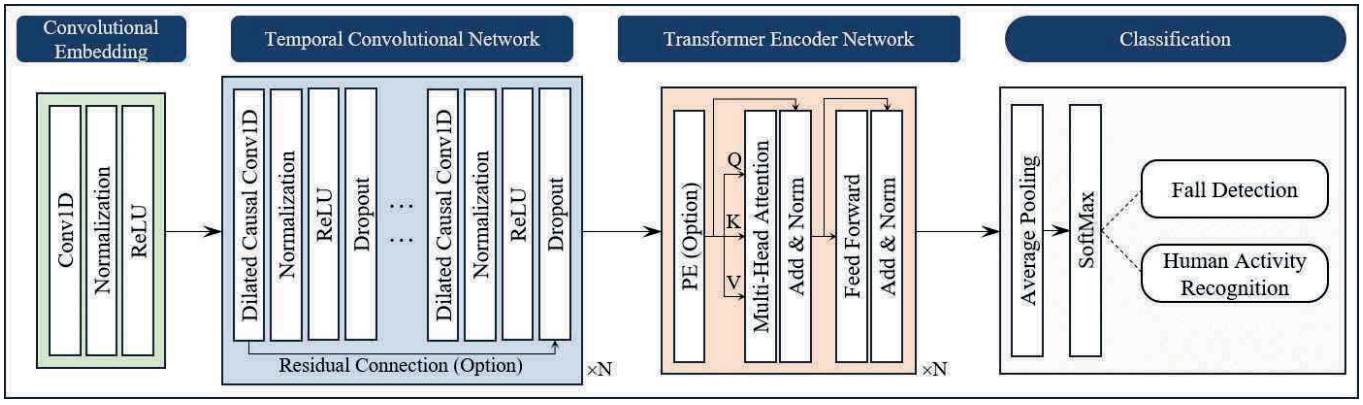

그림 1. 경량 시간적 합성곱 신경망 구조

도록 하였고, 검증용 데이터는 훈련 데이터에서 무작위로 10% 추출하였다. UCI-HAR 데이터셋은 선형 가속도와 각속도 신호를 고정된 윈도우 길이 2.56초 (128개 샘플)와 이동 간격 1.28초 (64개 샘플)를 이용하여 시퀀스로 분할하였다. 이 데이터셋은 참가자를 기준으로 무작위로 분할된 훈련용 (70%)과 테스트용 (30%)을 제공함으로, 훈련 데이터에서 무작위로 10% 추출된 시퀀스를 검증용으로 사용하였다.

2.4. 실험환경 및 평가

경량 시간적 합성곱 신경망은 윈도우 10, 텐서플로우 2.5 프레임워크 상에서 구현하였고, AMD Ryzen 9 3900X 12-Core Processor 3.79 GHz, 128 GB 메모리, Nvidia Geforce RTX 2080 Ti 환경에서 실험하였다. 제안된 경량 모델의 성능 수준을 판단하기 위해, ResNet18의 변형된 경량 모델 LRN [6]과 비교하였다. 모든 실험은 배치크기 (64), 최대 에폭 수 (100)에서 범주형 교차 엔트로피 오차를 고려한 Adam 최적화 (초기 학습율, 1e-3)를 사용하여 학습하였다. 학습 동안에 Stochastic Gradient Descent with Warm Restarts (SGDR) [7]을 이용하여 점진적으로 학습율을 조정하였다. 이때 주기 증가 계수 (2), 최대 학습율 감소 계수 (0.5), 최저 학습율 비율 (0.01), 첫 번째 감쇠 주기 길이는 30 에폭 × 미니배치 크기로 설정하였다.

2.5. 결과

표 1은 제안된 경량 시간적 합성곱과 LRN과의 비교 결과를 보여준다. 제안된 방법은 LRN에 비해 훈련 파라미터 수를 약 3배가량 경량화 하면서, 3가지 낙상 유형에서 약 1.5%, 6가지 신체 활동 인식에서 약 0.6% 개선된 매크로 평균 F1 점수를 보였다.

표 1. 제안된 방법과 LRN의 성능 비교

평가척도	UMAFall		UCI-HAR	
	LRN	Proposed	LRN	Proposed
정확도	0.9594	**0.9746**	0.9477	**0.9559**
매크로 평균 F1	0.9593	**0.9744**	0.9489	**0.9553**
가중치 평균 F1	0.9594	**0.9746**	0.9479	**0.9556**
훈련 파라미터	0.234M	**0.077M**	0.234M	**0.077M**

Ⅲ. 결론

본 연구에서는 다중 센서 신호를 활용하여 낙상 감지와 신체 활동 인식이 가능한 경량 시간적 합성곱 신경망 구조를 개발하였다. 제안한 신경망은 다양한 지역적 특징으로부터 인과적 정보와 장기 시간 의존성을 효과적으로 포착하기 위해 팽창 인과적 합성곱을 활용하였다. 또한, 모든 시간 지점과 채널 간의 상호 관계를 동시에 고려하여 중요한 시점의 특징을 강조함과 동시에 불필요한 특징을 억제하기 위해 다중 헤드 주의 메커니즘을 적용하였다. 제안한 경량 모델은 공개 데이터셋 UMAFall과 UCI-HAR에서 LRN과의 비교를 통해 성능의 우수성을 보였다.

ACKNOWLEDGMENT

본 논문은 과학기술정보통신부에서 지원하는 DGIST 기관고유사업에 의해 수행 되었습니다(25-IT-02).

참 고 문 헌

[1] Lea, C., Flynn, M.D., Vidal, R., Reiter, A., and Hager, G.D., “Temporal convolutional networks for action segmentation and detection,” arXiv preprint arXiv:1611.05267, 2016.

[2] Yu, F., and Koltun, V., “Multi-scale context aggregation by dilated convolutions,” arXiv preprint arXiv:1511.07122, 2015.

[3] Vaswani, A., et al., “Attention is all you need,” Advances in Neural Information Processing Systems, 30, 2017.

[4] Casilari, E., Santoya-Ramon, J.A., and Cano-Garcia, J.M., “UMAFall: A multisensor dataset for the research on automatic fall detection,” Procedia Computer Science, 110, pp. 32-39, 2017.

[5] Anguita, D., Ghio, A., Oneto, L., Parra, X., and Reyes-Ortiz, J.L., “A public domain datasets for human activity recognition using smartphones,” Proceedings of European Symposium on Artificial Neural Networks (ESANN), pp. 437-442, 2013.

[6] Calatrava, F.M., and Mozos, O.M., “Light residual network for human activity recognition using wearable sensor data,” IEEE Sensors Letters, 7(10), 2023.

[7] Loshchilov, I., and Hutter, F., “SGDR: Stochastic gradient descent with warm restarts,” arXiv preprint arXiv:1608.03983, 2016.