

생성형 AI를 활용한 스마트워치 PPG 신호 증강 및 유효성 검증 연구

김하정, 김지민, 문정윤, 유재현*
성신여자대학교 AI 융합학부

20221346@sungshin.ac.kr, 20231342@sungshin.ac.kr,
20221356@sungshin.ac.kr, jhyoo@sungshin.ac.kr

A Study on Smartwatch PPG Signal Augmentation and Validation using Generative AI

Ha Jeong Kim, Ji Min Kim, Jung Yoon Moon, Jaehyun Yoo
School of AI Convergence, Sungshin Women's University

요약

본 논문은 스마트워치를 통해 수집한 PPG 데이터를 이용하여 위험 감지 모델을 학습하기 위한 목적으로, 데이터의 라벨링 및 생성형 모델을 이용한 데이터 증강 방법을 제시한다. 실제 착용 환경에서 수집된 PPG 데이터에는 노이즈나 경계적 신호로 인한 오분류 가능성이 존재하므로 평균 피크 값을 활용한 리라벨링 기법을 적용하여 신뢰성 있는 학습 데이터를 구축한다. 또한, 트랜스포머 기반의 생성 모델을 이용하여 위험 데이터를 증강함으로써 데이터 불균형 문제를 해결한다. 생성된 데이터와 실제 수집 데이터를 통합하여 위험 추론 분류 모델을 학습한 결과 94.38%의 정확도로 위험 상황을 분류할 수 있음을 검증하였다.

I. 서론

최근 웨어러블 디바이스의 발전으로 생체 신호 기반 감정 인식 연구가 활발히 진행되고 있다. 특히 PPG 신호는 혈류 변화를 광학적으로 측정하는 방식으로 심리적 상태를 간접적으로 추론할 수 있어 활용도가 높다 [1].

그러나 실제 환경에서 수집된 PPG 데이터는 움직임, 착용자의 생리적 특징과 같은 요인과 주변 환경 요인 등으로 인해 데이터 라벨링 오류가 크다. 특히 감정 인식을 목적으로 데이터를 수집할 때, 사용자의 감정에 대한 정확한 정답을 알기 어렵다. 따라서 본 논문에서는 데이터의 정답을 annotation 하기 위해 사용자의 조사 결과에만 의존하지 않고, 데이터 처리 단계에서 리라벨링을 제안한다. PPG는 심장 박동에 의한 주기적인 펄스로 이루어진 신호이다. 펄스의 피크값들의 평균값의 대소비교를 통해 리라벨링 기법을 도입하여 데이터 annotation 을 하여 유의미한 신호를 강조하고 노이즈 데이터를 배제하도록 유도한다.

또한 실제 환경에서는 고품질 위험 상황 데이터가 상대적으로 희소하고 수집 비용도 높기 때문에 데이터의 양적·질적 한계를 보완할 필요가 있다. 이를 위해 본 연구는 트랜스포머 기반 시계열 생성 모델인 HeartGPT를 활용하여 위험 구간에 대해 증강 데이터를 생성하고 이를 실제 수집된 PPG 신호와 통합하였다. 최종적으로 이렇게 보완된 데이터셋을 기반으로 1D-CNN 분류기를 학습시켜 심박 신호만으로 위험 상황을 정밀하게 탐지하는 시스템을 제안한다.

II. 장에서는 데이터 전처리와 리라벨링 기법, 데이터 증강 방법, 위험 감지 추론 모델을 소개하고, III. 장에서는 실험 결과, IV. 장은 결론을 맺는다.

II. 본론

2.1 데이터 전처리 및 리라벨링

본 연구에서는 실험을 활용해 PPG 데이터를 수집하였다. 긍정·부정 영상 실험에서는 감정유형에 따라 데이터를 분리한 뒤 긍정 구간에는 0, 부정 구간에는 1의 라벨을 부여하였다. VR 공포 게임 실험에서는 수집된 PPG 신호의 전반부를 비위험 구간, 후반부를 위험 구간으로 간주하여 각각 라벨 0과 1을 부여하였다. 수집된 PPG 신호는 피험자별로 통합한 후 500 샘플 단위로 청크화하였다. PPG 센서는 외부 환경이나 움직임에 민감하므로 0.5~8Hz 대역 통과 필터(Band-pass filter)를 적용하여 노이즈를 제거하고 주요 심박 성분만을 남겼다. 이후 HeartPy 피크 탐지를 통해 실질적 피크 수가 30bpm 이상이고 심박수가 정상 범위에 해당하는 경우를 유효 청크로 간주하였다. 이후 훈련 데이터와 테스트 데이터를 일정 비율로 나누었다.

라벨 0과 1에 해당하는 데이터를 시각화한 결과 신호의 진폭 차이가 두 집단을 구별하는 주요 특성으로 판단되었고 이 차이를 보다 명확히 하기 위해 추가적인 기준을 설정하였다. 평균 피크 진폭 값이 분포의 평균(μ) + n *표준편차(σ)를 threshold n 으로 정의하고, 부정으로 라벨링된 청크 중 threshold를 초과하는 경우에만 1로 유지하고 그렇지 않은 경우에는 0으로 재조정하는 리라벨링을 수행하였다. 피크 탐지 실패나 기준 미달로 유효성 조건을 충족하지 못한 청크는 -1로 처리하여 학습에 사용하지 않았다. 이와 같은 기준 재정의는 라벨의 신뢰도를 높인데 목적이 있으며 최종적으로 정제된 청크들은 HeartGPT [2] 기반 데이터 증강 과정에 활용되었다.

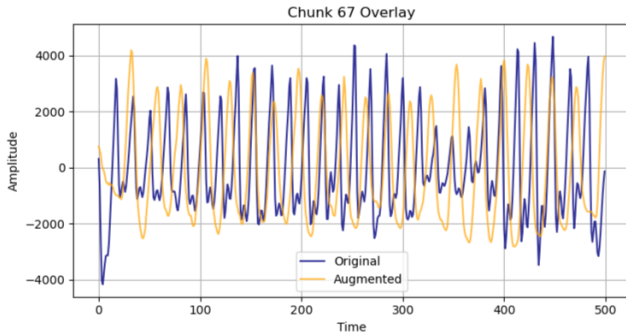


Figure 1 원본 데이터와 HeartGPT 로 증강한 데이터 비교

2.2 데이터 증강

본 연구에서는 PPG 시계열 데이터를 기반으로 트랜스포머 구조의 생성 모델인 HeartGPT 를 활용하여 데이터 증강을 수행하였다. HeartGPT 는 정규화된 PPG 신호를 정수형 토큰 시퀀스로 변환한 후, 자기회귀 방식으로 다음 시점을 예측하며 새로운 시계열을 생성하는 언어 모델 기반 구조를 따른다. 특히 반복적인 심장 주기에서 중요한 특징을 주의(attention) 메커니즘으로 포착하며 문맥 기반의 토큰 표현을 통해 주기적 패턴의 의미를 효과적으로 학습한다. 이 과정에서 모델은 과거 심장 박동에서 비슷한 위치를 참고해 다음 신호를 예측하고 특정 어텐션 구조는 PPG 신호에서 dicrotic notch 처럼 중요한 생리적 특징에 특히 민감하게 반응한다. 또 같은 값이라도 심장 주기 안에서 등장하는 위치에 따라 다르게 해석되도록 학습해 생성된 시계열이 실제 생체 신호의 흐름과 잘 맞도록 만들어준다.

부정 감정 데이터는 평균 피크 값이 사전에 정의한 threshold 를 초과하는 경우에만 1 로 유지되고 그 외에는 0 으로 재조정되었기 때문에 최종적으로 1 로 라벨링된 데이터의 수가 상대적으로 적어져 데이터 불균형이 발생하였다. 이에 따라 각 threshold 설정에 따라 긍정·부정 간 비율을 고려해 증강 배수를 달리 설정하고, 부족한 부정 감정 샘플에 대해 HeartGPT 를 활용한 시계열 생성을 수행하였다. 이 때 생성된 토큰열은 역정규화 과정을 통해 원래 PPG 신호의 스케일로 복원하였다. 이를 통해 긍정 및 부정 감정 간의 데이터 분포를 균형 있게 조정하여 학습 데이터의 편향을 완화하였다.

2.3 위험 감지 추론 모델

증강된 데이터와 실제 데이터를 합친 데이터셋을 기반으로 1D-CNN 모델을 학습하였다. 모델은 다음과 같은 구조를 가진다. 마지막 출력 layer 를 제외한 모든 layer 에서 ReLU 활성화 함수를 사용하였다. 마지막 출력 값은 감정 정도를 측정하기 위해 sigmoid 함수를 사용한다.

III. 실험 결과

본 연구에 사용된 데이터는 긍정 및 부정 영상 시청 실험과 VR 공포 게임 플레이 실험을 통해 삼성 갤럭시 워치 5 를 참가자의 손목에 착용한 상태에서 자체 개발한 애플리케이션을 통해 수집되었다. 긍정 및 부정 영상 실험에는 20 대 여성 12 명이 참여하였으며, 각 참가자는 약 1 시간 30 분 분량의 긍정 영상과 세트당 10~20 분으로 구성된 총 5 세트의 부정 영상을 시청하였다. VR 공포 게임 실험에는 총 7 명이 참여하였으며, 각 참가자는 전반부 10 분은 비위협 자극,

후반부 10 분은 위협 자극으로 구성된 20 분 길이의 공포 게임을 VR 기기를 착용하고 플레이하였다.

수집된 PPG 신호는 일련의 전처리 과정을 거쳐 총 3,838 개의 유효 청크로 정제되었으며, 이 중 긍정 감정에 해당하는 청크는 2,096 개(54.6%), 부정 감정 청크는 1,742 개(45.4%)로 구성하였다. 전체 데이터 중 무작위로 20%를 분리하여 테스트 데이터로 활용하였다.

전처리를 마친 후 별도로 분리한 테스트 데이터에 대해 라벨링의 효과를 정량적으로 검증하였다. 평균 피크값 기반 threshold 에 따라 데이터를 재구성한 결과, 라벨링을 적용하지 않은 경우 정확도는 0.51~0.57 수준으로 낮게 유지되었다. 그러나 동일한 조건에서 라벨링을 적용한 경우 정확도는 최대 0.94 까지 크게 향상되었다. 이는 평균 피크값을 기준으로 위협 청크를 재분류한 라벨링 과정이 위협 상황에 해당하는 패턴을 더욱 명확하게 인식할 수 있음을 의미한다.

Table 1 라벨링 임계값에 따른 정확도 비교

Threshold (평균 + $n \times$ 표준편차)	Accuracy (w/o Relabeling)	Accuracy (with Relabeling)
$n=0$	0.57	0.79
$n=0.5$	0.54	0.88
$n=1$	0.51	0.93
$n=1.5$	0.53	0.94

IV. 결론

본 논문에서는 PPG 신호를 활용한 위협 감지 정확도를 향상시키기 위해 라벨링과 트랜스포머 기반 데이터 증강 기법을 결합한 학습 데이터 구성 방식을 제안하였다. 이러한 과정은 학습 데이터의 품질과 균형을 동시에 확보함으로써 분류 모델의 일반화 성능을 효과적으로 향상시켰다.

향후에는 생성 모델의 품질을 더욱 고도화하여 감정 신호의 다양성과 현실성을 높이고 감정의 이산적 분류를 넘어 연속적인 감정 스펙트럼이나 위협 정도와 같은 정량적 예측으로 발전시킬 예정이다.

ACKNOWLEDGMENT

이 논문은 경찰청이 지원한 ‘사회적 약자 보호 강화 기술 개발(www.kipot.or.kr)’의 지원을 받아 수행된 연구결과입니다. [과제명: 저전력 복합측위, 근접탐색 기술 기반 범죄피해 안전조치 대상자 위치추적 통합관계 플랫폼 개발/과제번호: RS-2023-00236101]

참 고 문 헌

- [1] S. Yoo, G. Hwang, J. Yoo. "Detection of Psychological Risk for Protected Individuals by Using PPG Signals from Smartwatch", J-KICS, vol. 50, no. 4, pp. 572-577, 2025.
- [2] Devies, H. J., Mandic, M. J. D.P. "Interpretable PreTrained Transformers for Heart Time-Series Data", arXiv preprint arXiv:2407.20775
https://doi.org/10.48550/arXiv.2407.20775 (2024).