

PPG와 ECG의 상관 관계에 기반한 심박 시계열 데이터 이상 상황 탐지 최적 모델 비교 연구[☆]

A Comparative Study on the Optimal Model for abnormal Detection event of Heart Rate Time Series Data Based on the Correlation between PPG and ECG

김진수¹ 이강윤^{1*}
Jin-soo Kim Kang-yoon Lee

요 약

본 논문은 , 가 , 가 가 가 가
PhysioNet MIT-BIH Arrhythmia Database
(ECG)

주제어 : 광혈류측정(PPG), 심전도(ECG), 이상상황 탐지, SVM, LSTM

ABSTRACT

This paper Various services exist to detect and monitor abnormal event. However, most services focus on fires and gas leaks. so It is impossible to prevent and respond to emergency situations for the elderly and severely disabled people living alone. In this study, AI model is designed and compared to detect abnormal event of heart rate signal which is considered to be the most important among various bio signals. Specifically, electrocardiogram (ECG) data is collected using Physionet's MIT-BIH Arrhythmia Database, an open medical data. The collected data is transformed in different ways. We then compare the trained AI model with the modified and ECG data.

keyword : Photoplethysmography(PPG), Electrocardiogram(ECG), Abnormal event detection, SVM, LSTM

1. 서 론

이상 상황이란 다양한 분야에 존재한다. 특히 사회적 약자인 독거노인과 중증 장애인들을 위한 이상 상황 서비스는 국가적 서비스로 제공되고 있다. 그 중에서 보건복지부의 응급안전알림서비스는 보건복지부 산하에 있는 사회안전보장원의 사회안전망 서비스 중 하나로 현재

도 서비스를 제공하고 있다. 하지만 응급안전알림서비스는 화재와 가스누출에만 초점이 맞춰져있으며, 사회적 약자인 독거노인의 고독사 혹은 중증 장애인들의 심정지 등 위급상황에 대해 사전예방과 실시간 대응이 불가능한 현실이다. 또한 이상 상황을 탐지하기 위해서는 다양한 신호에 대한 분석과 상황을 고려해야하고, 실시간으로 모니터링되는 과정이 필요하기 때문에 현실적으로 서비스가 진행되기 어렵다. 하지만 모든 이상 상황의 탐지는 어렵지만 대상의 사망이나, 생명의 직접적인 위협을 탐지하는 것은 인공지능 모델과 사람의 생체 데이터를 활용하면 극복할 수 있는 부분이다. 특히 인공지능 모델의 경우 사람을 대신하고 실시간으로 탐지가 가능한 장점이 있다. 그러나 인공지능 모델을 제작하기 위해서는 데이터가 준비되어야 한다. 이 부분을 보완하는 방법은 오픈 데이터

¹ Dept. of Computer Engineering, Gachon University, Seongnam, 13120 Korea.

* Corresponding author (keylee@gachon.ac.kr)

[Received 30 August 2019, Reviewed 23 September 2019(R2 14 November 2019), Accepted 6 December 2019]

☆ 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학 ICT 연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2019-2017-0-01630)

를 활용하는 방법이다. 전 세계적으로 잘 알려진 오픈 의료 데이터 포털인 PhysioNet을 이용하면 다양한 상황에서 얻어진 생체 데이터를 수집할 수 있다.

본 연구에서는 PhysioNet의 MIT-BIH Arrhythmia Database의 심전도(ECG)데이터를 이용하여 최적의 데이터 변형 방법과 인공지능 모델을 비교 연구한다.[1] 데이터 변형에서는 공통적으로 수집된 데이터를 최적의 윈도우 사이즈(Window Size)로 분리하고, 분리된 데이터를 각 인공지능 모델에 최적의 방법으로 데이터를 변형한 후 학습한 인공지능 모델과 기본 심전도 수치를 통해 이상상황 탐지 모델을 비교한다.[2][3][4]

2. 관련연구

심박 데이터는 접촉센서의 작동 원리에 따라 데이터의 형태가 나뉜다. 매우 널리 사용되는 2가지 방식은 광혈류측정 원리를 이용한 PPG 데이터와 심전도를 측정하는 ECG 데이터이다. 이 두 데이터는 예전부터 많은 상관관계에 대한 연구가 진행되어 왔으며 특히 심박변이도(HRV)에서 90%가 넘는 상관관계를 가진다는 연구 결과가 있다.[5]

인공지능 모델 역시도 다양한 이상 상황을 판단하기 위해 다양한 인공지능 모델들이 연구되었고, 특히 딥 러닝 분야의 연구가 활발히 진행되고 있다.

2.1 광 혈류 측정(PPG)

광 혈류 측정 원리를 이용한 PPG는 사람의 피부에 LED 빛을 비추고, 반사된 빛을 가시도 포토다이오드를 사용하여 혈류로 인한 반사된 빛의 세기를 측정함으로써 데이터를 얻는다. PPG는 심박 뿐만 아니라 동맥 혈압, 경직도, 맥박 전이 시간, 맥박 파동 속도, 심장 출력, 동맥 순응성, 주변 저항 등 심혈관계와 관련한 정보도 역시 제공한다.[6][7]

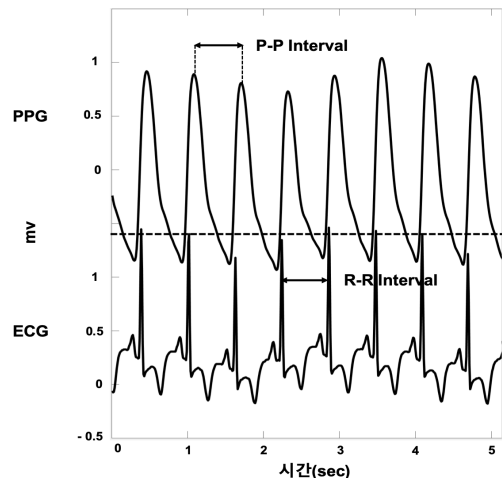
2.2 심전도(ECG)

심전도 ECG는 심장박동의 주기 중에서 일어나는 심장의 전기적 활동 상태를 그래프에 나타낸 것이다. 심전도는 불규칙한 심장박동인 부정맥을 판단할 뿐만 아니라 심근장애, 심방 심실의 비대, 확장, 폐순환 장애 등 심장 질환 및 연관성 질환 진단에 유용한 지표이다.

2.3 PPG와 ECG의 상관관계

심박변이도(HRV)는 연구 및 임상 분야에서 심혈관계 분석에 도움을 주는 지표로 심박 연구에 널리 활용된다.[8][9] 기존의 심박변이도는 ECG의 R-R Interval을 통해서 얻었지만 보다 측정과 수집이 편리한 PPG 데이터를 활용하기 위한 비교 연구가 많이 진행되었고, ECG와 PPG를 비교하는 다양한 지표들을 비교해 본 결과 지표 중 가장 큰 오차는 2.46%에 불과하다는 연구결과를 통해 강력한 상관관계를 가지고 있다는 것을 알 수 있다.[10]

본 연구에서도 ECG의 R-R Interval 수치를 이용해 심박변이도(HRV)의 여러 기준 수치를 파악하고 그 수치를 인공지능 모델에 활용하여 결과를 비교하였다.[11]

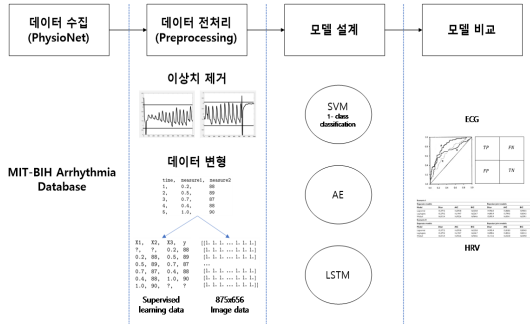


(그림 1) PPG/ECG 비교 그래프
(Figure 1) PPG/ECG Comparison graph

3. 심박 이상 상황 탐지 모델 프로세스

심박 이상 상황 탐지를 위한 프로세스는 (1) 데이터수집, (2) 데이터 전처리, (3) 모델 설계, (4) 모델 비교 순으로 이루어진다. (1) 데이터 수집 단계에서는 PhysioNet의 다양한 데이터를 통해 연구를 진행하며 가장 적합한 데이터 셋을 선정하였다. 데이터 전처리 프로세스에서는 오픈 데이터의 노이즈가 들어간 이상치를 제거하고, 데이터 베이스의 데이터를 학습에 최적화할 수 있는 윈도우 사이즈(Window Size)로 나눈다. 분리된 데이터를 각각 분류 모델에 맞춰 2가지 방식으로 적용한다.

첫 번째 방법은 데이터를 이미지로 변형하여 오토인코더에 학습 데이터로 활용하는 방법이다. 그리고 두 번째 방법은 LSTM에 활용될 수 있도록 시계열 데이터로 변형하는 방법이다. 각 방법에 대한 데이터 변형 후 기본 심전도 데이터를 학습한 인공지능 모델과 변형된 데이터로 학습한 인공지능 모델을 비교한다.



(그림 2) 이상 상황 탐지 모델 프로세스

(Figure 2) Abnormal event Detection Model process

3.1 광 혈류 측정(PPG)

PhysioNet은 1999년 국립 보건원(NIH)의 후원으로 건립되었으며, 수많은 임상 데이터와 데이터에 접근 가능성이 편리한 소프트웨어 등 많은 오픈 소스를 제공하고 있다. 특히 전 세계의 의료 기관들이 임상연구를 위해 혹은 다른 이유를 위해 수집한 데이터를 PhysioNet을 통하여 오픈 데이터로써 배포하고 있고 지속적으로 업데이트 및 챌린지를 진행하고 있다.

본 연구에서 사용한 데이터는 PhysioNet의 의료 데이터베이스 중 많이 사용되는 MIT-BIH Arrhythmia Database를 이용하였다. 이 오픈데이터는 48명의 환자를 약 30분 간 ECG 수치를 측정한 데이터이다.

3.2 데이터 전처리

MIT-BIH Arrhythmia Database의 데이터는 시계열 ECG 데이터이다. 오픈 데이터에는 이상치가 존재하는데 본 데이터베이스에서 알려진 이상치는 102번, 104번, 232번 환자이다. 수집된 데이터 중 노이즈가 포함된 이상치를 모델 학습의 퀄리티를 높이기 위해 제거한 후 데이터를 학습에 적합한 윈도우 사이즈로 나눈다. 이 전처리는 인공지능 모델 학습에 큰 영향을 미치며, 윈도우 사이즈에 따

라 정상 파형과 이상 파형을 나누는 기준이 되기도 한다. 본 연구에서는 다양한 윈도우 사이즈를 통해 최적의 윈도우 사이즈를 판단하였고 R-R Interval 10개의 윈도우로 분리하였다.

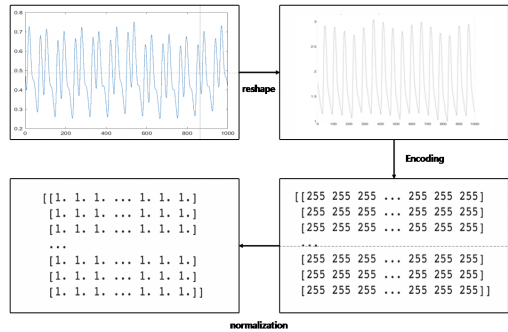
첫 번째 데이터 변형은 연속적인 시계열 데이터를 다음 상황 예측을 위한 지도학습 형태로 변형하는 데이터 변환이다. 아래 그림과 같은 방법으로 데이터 변형을 진행하는데 변형된 데이터는 LSTM 모델에서 기본 수치를 이용해 예측하는 LSTM 모델과 결과를 비교하기 위해 사용된다.

$$\begin{aligned}
 (t_1, p_1, e_1) &\rightarrow (p_1, e_1), (p_2, e_2) \\
 (t_2, p_2, e_2) &\rightarrow (p_2, e_2), (p_3, e_3) \\
 (t_3, p_3, e_3) &\rightarrow (p_3, e_3), (p_4, e_4) \\
 (t_4, p_4, e_4) &\rightarrow (p_4, e_4), (p_5, e_5) \\
 &\vdots \\
 &\vdots \\
 (t_n, p_n, e_n) &\rightarrow (p_n, e_n), (p_{n+1}, e_{n+1})
 \end{aligned}$$

$$t = \text{time}, p = \text{ppg}, e = \text{ecg}$$

(그림 3) 시계열 데이터 변형

(Figure 3) Time series data Transformation



(그림 4) 이미지 데이터 변형

(Figure 4) Image data Transformation

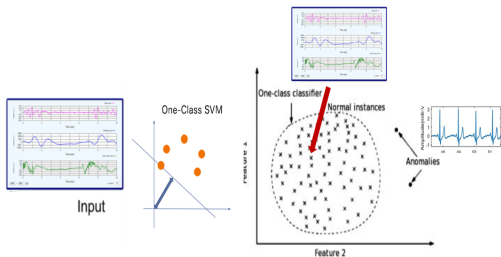
두 번째 데이터 변형은 오토인코더 모델 비교에 사용되는 데이터 변형 방법이다. 시계열 데이터를 활용하여 이미지 데이터로 변형하는 방법은 다음과 같다. 현재 수집된 데이터의 측정시간이 서로 다르기 때문에 데이터를 최소 측정시간을 기준으로 데이터를 재구성한다. 이후 데이터를 파형 그래프 형태로 이미지화 시킨다. 그리고 파

형 이미지 데이터를 학습 전 이미지 특징을 가지며 연산을 최소화할 수 있는 크기로 수정한다. 수정한 이미지 데이터를 학습이 가능하도록 숫자형 벡터로 읽은 후 정규화한다.

심전도의 R-R Interval을 통해 심박변이도를 구하고 심박변이도를 통한 비교 또한 진행하였다. 관련연구에서 설명한 것처럼 심박변이도는 심전도 분석에 영향력 있는 수치로 심박변이도를 측정하는 다양한 기준이 존재한다. 본 연구에서는 18개의 심박변이도 기준을 통해 인공지능 모델을 학습하고 비교하는 연구도 진행하였다.

3.2 모델 설계

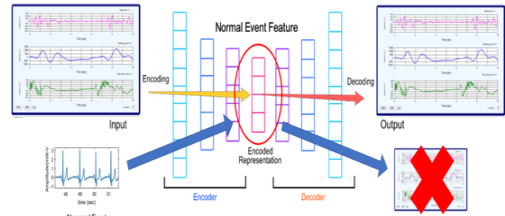
본 연구에서 진행할 모델은 3가지 방법이다. 첫 번째 SVM모델 비교에서는 기존 SVM분류와 One-class SVM모델을 비교한다. 비정상 데이터보다 정상데이터의 비중이 높은 경우 One-class SVM을 통해 학습을 진행하면 기존 SVM모델과 비교하여 좋은 결과를 도출할 수 있다.



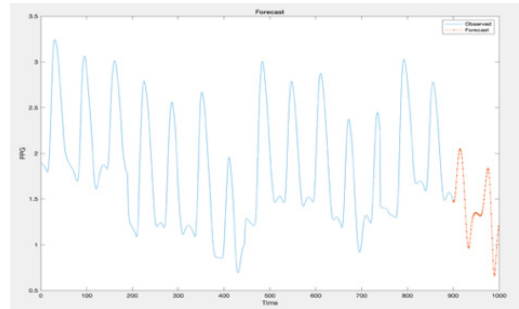
(그림 5) One-class Classification SVM 모델
(Figure 5) One-class Classification SVM model

두 번째는 오토인코더 모델을 학습한다. 기본 심전도 데이터를 활용해 정상 데이터를 학습시켜 평균 제곱 오차를 구하고 전처리의 첫 번째 방법인 이미지 변형 데이터를 활용하여 정상 데이터의 평균 제곱 오차를 구하여 정상 상황 탐지 정확도를 구한다.

세 번째 LSTM모델은 기존 시계열 심전도 데이터를 통해 예측값을 구한 후 평균 제곱오차를 통해 기준치를 정하였다. 이후 기준치를 역치(Threshold)로 사용하여 정확도를 계산하였다. 모델 비교는 심전도데이터와 시계열 변형 데이터로 모델을 각각 학습시키고 테스트 셋을 이용하였다.



(그림 6) 오토인코더 모델
(Figure 6) Autoencoder model



(그림 7) LSTM 예측 그래프
(Figure 7) LSTM Prediction Graph

3. 모델 비교 결과

모델 비교는 각각 모델의 정확도와 오분류율, 평균 제곱의 오차를 통해 비교하였다. 각 모델에 대한 정확도는 아래 표1과 같다. 모델의 비교 결과 SVM 모델은 기존 이진분류 SVM보다 One-class SVM의 결과가 더 높은 정확도를 보여주었다.

오토인코더의 결과를 보면 기존 데이터를 통한 분류 결과보다 이미지 변형 데이터를 통해 학습시킨 오토인코더가 더 좋은 정확도를 보여주었다.

LSTM의 결과는 좋은 결과를 보여주지 못하였는데, 이러한 결과를 보여준 이유는 시퀀스 데이터의 차원이 저차원이기 때문에 좋은 결과를 보여주지 못하였다. 하지만 시계열 데이터로 변형한 학습의 결과가 기존 데이터보다 높은 성능 보여주었다. 대부분의 결과 기존 심전도 수치를 이용한 것 보다 높은 정확도를 보여주었다. 하지만 예외로 One-class SVM에서는 더 낮은 결과를 보여주었다. 대부분의 모델에서 심박변이도(HRV)를 통한 모델 학습의 결과가 높은 이유는 심박변이도로 변형 시 모델에 Feature가 증가하고 기존 심전도 수치보다 뚜렷한 수치를 보여주었기 때문으로 판단된다.

본 연구에서 사용한 심전도 윈도우 사이즈는 다양한 수치로 나누어 진행하였다. 10, 15, 20, 25, 30 등 다양한 윈도우 사이즈로 데이터를 나누어 모델을 학습시킨 후 모델의 정확도를 비교하였다. 학습 데이터의 양의 차이에서 윈도우 사이즈 10을 선정하였다.

5. 결 론

본 연구에서 생체 데이터 중 생명과 위중하다가 판단 되는 심박 데이터를 이용하여 심박 이상 상황 탐지 모델을 비교 연구하였다. 데이터의 수집은 오픈 데이터를 제공하고 'PhysioNet'을 통해 수집하였으며, 많이 사용되는 MIT-BIH Arrhythmia Database를 사용하였다. 심전도 데이터를 이용하여 연구 중 최적의 윈도우 사이즈인 10으로 구분하여 데이터를 분리하고 기본 심전도 데이터와 데이터 변형을 한 시계열 데이터, 이미지 데이터로 인공지능 모델을 학습하여 비교하였다. 연구의 결과로 인공지능 모델에 맞춰 데이터 변형을 통해 학습한 인공지능 모델들이 결과가 높게 나왔고, SVM의 One-class SVM이 가장 높은 정확도를 보였다. 심박변이도로 기본 심전도를 변환하여 인공지능 모델을 학습시킨 결과도 대부분이 기본 수치보다 더 높은 정확도를 보였지만, SVM에서는 예외의 결과를 보였다. 현재 PPG와 ECG 같은 접촉식센서 측정에서 발전되어 마이크로웨이브를 이용한 비접촉센서 연구가 이루어진다면 접촉식센서의 데이터 수집에 단점이 되는 센서 부착의 거부감을 제거하고 접촉식센서의 활용 모델에 적용할 수 있을 것이다.

참고문헌(Reference)

- [1] Moody, G. B., & Mark, R. G., "The impact of the MIT-BIH Arrhythmia Database," *IEEE Engineering in Medicine and Biology Magazine*, 20(3), pp. 45 - 50, 2011.
<https://doi.org/10.1109/51.932724>
- [2] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, *A Practical Guide to Support Vector Classification*, 2016.
<https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- [3] BALDI, Pierre. Autoencoders, unsupervised learning, and deep architectures. In: *Proceedings of ICML workshop on unsupervised and transfer learning*, pp. 37-49, 2012.
<https://doi.org/10.1.1.296.3397>
- [4] HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. "Long short-term memory," *Neural computation*, 9.8: pp. 1735-1780, 1997.
<https://doi.org/10.1162/neco.1997.9.8.1735>
- [5] KACHUEE, Mohammad, et al., "Cuffless blood pressure estimation algorithms for continuous health-care monitoring," *IEEE Transactions on Biomedical Engineering*, 64.4: pp. 859-869, 2016.
<https://doi.org/10.1109/tbme.2016.2580904>
- [6] DE SÁ FERREIRA, Arthur, et al., "Three-section transmission-line arterial model for noninvasive assessment of vascular remodeling in primary hypertension," *Biomedical Signal Processing and Control*, 4.1: pp. 2-6, 2009.
<https://doi.org/10.1016/j.bspc.2008.07.001>
- [7] DAVIES, Justine I.; STRUTHERS, Allan D., "Beyond blood pressure: pulse wave analysis - a better way of assessing cardiovascular risk?," *FUTURE CARDIOLOGY*, Vol.1, No.1, 2005.
<https://doi.org/10.1517/14796678.1.1.69>
- [8] CHUANG, Chiung-Cheng, et al., "Photoplethysmography variability as an alternative approach to obtain heart rate variability information in chronic pain patient," *Journal of clinical monitoring and computing*, 29.6: pp. 801-806, 2015.
<https://doi.org/10.1007/s10877-015-9669-8>
- [9] GIL, Eduardo, et al., "Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions," *Physiological measurement*, 31.9: 1271, 2010.
<https://iopscience.iop.org/0967-3334/31/9/015/>
- [10] LU, Sheng, et al., "Can photoplethysmography variability serve as an alternative approach to obtain heart rate variability information?" *Journal of clinical monitoring and computing*, 22.1: pp. 23-29, 2008.
<https://doi.org/10.1007/s10877-007-9103-y>
- [11] SHAFFER, Fred; GINSBERG, J. P., "An overview of heart rate variability metrics and norms," *Frontiers in public health*, 5: 258, 2017.
<https://doi.org/10.3389/fpubh.2017.00258>
- [12] VAN RAVENSWAALJ-ARTS, Conny MA, et al., "Heart rate variability," *Annals of internal medicine*, 118.6: pp. 436-447 1993.

- <https://doi.org/10.7326/0003-4819-118-6-199303150-00008>
- [13] SAYKRS, B. McA., "Analysis of heart rate variability," *Ergonomics*, 16.1: pp. 17-32, 1973.
<https://doi.org/10.1080/00140137308924479>
- [14] THAYER, Julian F., et al., "A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health," *Neuroscience & Biobehavioral Reviews*, 36.2: pp. 747-756, 2012.
- [15] Moody, G. B., & Mark, R. G., "The impact of the MIT-BIH Arrhythmia Database," *IEEE Engineering in Medicine and Biology Magazine*, 20(3), pp. 45 - 50, 2001.
<https://doi.org/10.1016/j.neubiorev.2011.11.009>
- [16] HIRSCH, Judith Aa; BISHOP, Beverly, "Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate," *American Journal of Physiology-Heart and Circulatory Physiology*, 241.4: pp. H620-H629, 1981.
<https://doi.org/10.1152/ajpheart.1981.241.4.H620>

● 저 자 소 개 ●



김 진 수(Jin-soo Kim)

2018년 가천대학교 컴퓨터공학과(공학사)
2019년~현재 가천대학교 대학원 IT융합공학과(석사과정)
관심분야 : 인공지능, 딥 러닝, 머신러닝, 온톨로지 etc.
E-mail : wlstn25092303@gmail.com



이 강 윤(Kang-yoon Lee)

1986년 연세대학교 전자공학과(공학사)
1996년 연세대학교 대학원 전자계산학과(공학석사)
2010년 숭실대학교 대학원 IT정책경영학과(공학박사)
1991년~2016년 한국 IBM 연구소(연구소장)
2016년~현재 가천대학교 컴퓨터공학과 교수
관심분야 : 인공지능, IoT, 빅데이터, 클라우드, 커그너티브컴퓨팅, etc.
E-mail : keylee@gachon.ac.kr