JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지 Vol. 23, No. 2: 117~126, Feb. 2019

Auto Regressive모델링 기반의 특징점 추출과 Support Vector Machine을 통한 조기수축 부정맥 분류

조익성[1] · 권혁숭[2*] · 김주만[2] · 김선종[2]

Feature Extraction based on Auto Regressive Modeling and an Premature Contraction Arrhythmia Classification using Support Vector Machine

Ik-sung Cho[1]· Hyeog-soong Kwon[2*]· Joo-man Kim[2]· Seon-jong Kim[2]

[1]Assistant Professor, Daepartment of Creative Integrated General Studies, Daegu University, Gyungsan 38453, Korea
[2*]Professor, Department of IT Engineering, Pusan National University, Miryang 50463, Korea

요 약

부정맥 분류를 위한 기존 연구들은 분류의 정확성을 높이기 위해 신경망, 퍼지, 시계열 주파수 분석, 비선형 분석법 등이 연구되어 왔다. 이러한 방법들은 분류율를 향상시키기 위해 정확한 특징점과 많은 양의 신호를 처리해야 하기 때문에 데이터의 가공 및 연산이 복잡하며, 다양한 부정맥을 분류하는데 어려움이 있다. 본 연구에서는 AR(Auto Regressive) 모델링 기반의 특징점 추출과 SVM(Support Vector Machine)을 통한 조기수축 부정맥 분류 방법을 제안한다. 이를 위해 잡음을 제거한 ECG 신호에서 R파를 검출하고 QRS와 RR 간격의 특정 파형 구간을 모델링하였다. 이후 최적 세그먼트 길이($n1, n2$), 최적 차수($p1, p2$)의 4가지 AR 모델링 변수를 추출하고 SVM을 통해 Normal, PVC, PAC를 분류하였다. 연구의 타당성을 입증하기 위해 MIT-BIH 부정맥 데이터베이스를 대상으로 한 R파의 평균 검출 성능은 99.77%, Normal, PVC, PAC 부정맥은 각각 99.23%, 97.28, 96.62의 평균 분류율을 나타내었다.

ABSTRACT

Legacy study for detecting arrhythmia have mostly used nonlinear method to increase classification accuracy. Most methods are complex to process and manipulate data and have difficulties in classifying various arrhythmias. Therefore it is necessary to classify various arrhythmia based on short-term data. In this study, we propose a feature extraction based on auto regressive modeling and an premature contraction arrhythmia classification method using SVM., For this purpose, the R-wave is detected in the ECG signal from which noise has been removed, QRS and RR interval segment is modelled. Also, we classified Normal, PVC, PAC through SVM in realtime by extracting four optimal segment length and AR order. The detection and classification rate of R wave and PVC is evaluated through MIT-BIH arrhythmia database. The performance results indicate the average of 99.77% in R wave detection and 99.23%, 97.28%, 96.62% in Normal, PVC, PAC classification.

키워드 : AR 모델, QRS, R파, RR 간격, 조기 수축

Key word : AR model, Premature contraction , QRS, R wave, RR interval

Received 17 October 2018, **Revised** 5 November 2018, **Accepted** 5 November 2018

*** Corresponding Author** Hyeog-soong Kwon(E-mail:hskwon@pusan.ac.kr Tel:+82-55-350-5411)
Professor, Department of IT Engineering, Pusan National University, Miryang 50463, Korea

Open Access http://doi.org/10.6109/jkiice.2019.23.2.117 print ISSN: 2234-4772 online ISSN: 2288-4165

ⓒThis is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/li-censes/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright Ⓒ The Korea Institute of Information and Communication Engineering.

Ⅰ. 서 론

일반적으로 심전도(ECG : electrocardiography) 신호를 해석하는 단계는 인체로부터의 신호 획득, 잡음 제거, QRS의 검출 및 진단을 포함하며, 신호의 분석을 토대로 심장질환의 이상유무를 판단하는데 많이 활용되고 있다. 부정맥은 심박동수가 비정상적이거나 심장의 리듬이 불규칙한 상태를 의미한다. 부정맥의 형태는 정상인의 가벼운 리듬 장애부터 생명을 위협하는 장애까지 다양한 형태로 나타나며, 이는 심장세포의 자동성 장애나 전도장애로 인해 초래된다. 부정맥은 심장박동이 불규칙한 상태로 심전도 신호 분석을 통하여 정상신호와 이상 신호를 검출함으로써 분류할 수 있다[1-3]. 특히 심실조기수축(PVC : Premature Ventricular Contraction)은 가장 흔한 부정맥으로 심장질환이 있었던 환자의 경우 심실빈맥(VT : Ventricular Tachycardia)과 같은 위험한 심장질환을 유발하며, 심방조기수축(PAC : Premature Atrial Contraction)은 뇌경색이나 전신 색전증의 합병증을 일으켜 질병 이환률과 사망률을 증가시킬 수 있다. 따라서 PVC, PAC와 같은 부정맥의 조기 검출은 매우 중요하다.

부정맥을 분류하기 위한 방법으로는 시계열 주파수 분석, 신경망, 퍼지, 비선형 분석법 등이 연구되어 왔다. 하지만 기존 연구들은 분류를 위해 많은 양의 심전도 데이터를 처리해야 하는 불편함과 여러 가지 특징점 검출로 인해 계산시간이 길어지는 단점이 있으며, 이와 동시에 다양한 부정맥을 분류하는데 어려움을 갖고 있다[4-7]. 이러한 문제점을 극복하기 위해서는 단기간의 입력 데이터를 통한 특징점 추출이 필요하며 적은 수의 학습 데이터만으로 다양한 부정맥을 정확하게 분류할 수 있는 방법이 필요하다.

이를 위해 AR모델링 기반의 특징점 추출과 SVM (Support Vector Machine)을 통한 조기수축 부정맥 분류 방법을 제안한다. 먼저 잡음이 제거된 심전도 신호에서 R파를 검출한 후 QRS와 RR간격 구간을 추출하여 모델링함으로써 최적 세그먼트 길이($n1, n2$), 최적 차수($p1, p2$)의 4가지 AR모델링 변수를 추출하고 SVM을 통해 정상, PVC, PAC를 분류하였다. 제안한 방법의 우수성을 입증하기 위해 MIT-BIH 부정맥 데이터베이스를 대상으로 정상, PVC, PAC의 분류율을 비교하였다. 본 논문의 구성은 다음과 같다. 2장에서는 제안한 기법에 대하여 논하고, 3장에서는 실험결과에 대하여 논하고 4장에서 결론으로 마무리한다.

Ⅱ. 제안한 기법

AR모델을 통한 최적 특징점 검출과 SVM을 통한 부정맥 분류 방법의 전체 구성도는 그림 1과 같은 단계로 나누어진다. 먼저 ECG신호로부터 전처리 과정을 통해 잡음을 제거한 후 최적 문턱치 설정을 통해 R파를 검출하는 단계, QRS와 RR간격 구간을 추출하는 단계, 추출된 구간을 AR모델링하여 변수값을 추출하는 단계, SVM을 통해 Normal과 PVC, PAC를 분류하는 단계로 나누어진다.

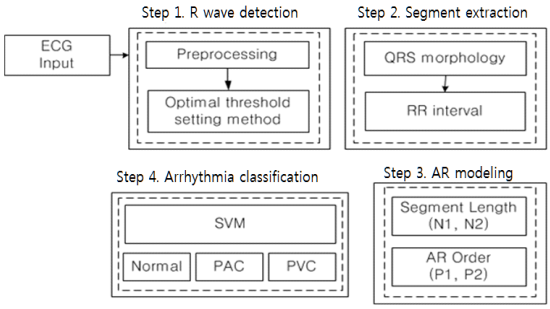


Fig. 1 System configuration

2.1. 전처리 및 R파 검출

정확한 진단을 위해서는 심전도 신호의 잡음 제거를 통해 정확한 R파를 검출하는 것이 필요하다. 본 연구에서는 샘플링 주파수에 따른 전처리 기법과 적응형 문턱치에 의한 최적치를 통해 R파를 검출하였다[8].

샘플링 주파수에 따른 성능은 분류성능에 영향을 끼치지 않는 최적의 샘플율을 정하기 위함이다. 일반적으로 MIT-BIH 데이터베이스의 경우 샘플링 주파수는 360Hz, 이동 평균의 범위는 8, 미분 파라미터의 범위는 7, 피크 문턱치는 6.6이라는 것을 확인할 수 있다. 이는 실험을 통한 경험 수치이며, 최적화된 R파를 검출하기 위해서는 프로세서의 성능에 맞는 샘플율의 선택이 중요하다는 것을 확인할 수 있었다.

2.2. 특징 구간 추출

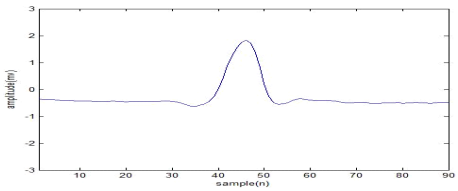
부정맥 분류를 위해 QRS와 RR간격의 특징 구간을 사용한다. 특징 구간은 특정 위치를 의미하는 것이 아니라, AR모델링의 최적 범위를 설정하기 위한 정상과 부정맥의 범위를 모두 포함하는 샘플의 수를 의미한다. 샘플 수는 AR모델링을 위해 선택된 세그먼트에 영향을 미치며 최소 한 번의 심장 사이클이 되도록 정확하게 모델링하여 진단한다. 일반적으로 정상신호의 QRS와 RR간격은 대부분 같은 형태를 가지고 있지만 부정맥은 매우 다양한 형태를 가진다.

그림 2는 MIT-BIH부정맥 데이터베이스 레코드를 대상으로 정상(normal)과 3개의 다른 형태의 PVC 그리고 PAC의 QRS형태를 나타내고 있다. 그림 2(b), 2(c), 2(d)에서와 같이 PVC간격은 그림 2(a)의 normal과 그림 2(e)의 PAC와 비교했을 때 그 간격이 매우 넓은 것을 확인할 수 있다. 일반적으로 정상 신호의 QRS간격은 60ms를 넘지 않는 경우가 대부분이며 PVC부정맥의 경우에는 그 간격이 120ms ~150ms에 이른다. 이를 샘플링 주파수가 360Hz인 MIT-BIH부정맥 데이터베이스를 기준으로 환산하였을 경우 60ms는 21.6, 120ms는 43.2, 150ms는 54 샘플 수로 각각 나타낼 수 있다.

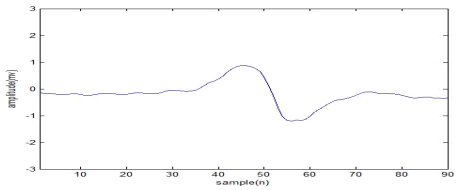
그림 3은 MIT-BIH데이터베이스의 부정맥을 대상으로 각 신호에 따른 RR시간 간격의 변화를 분석한 화면을 나타내고 있다.

그림 3(a)에 나타난 하나의 PAC 비트의 경우 이전 RR간격은 314ms, 이후는 508ms이며, 그림 3(b)에 나타난 PVC의 경우에는 이전 RR간격은 약 490 ~ 520ms, 이후는 1170 ~ 1300ms 간격으로 세 비트 모두 정상 비트에 비교했을 때 그 간격이 작아졌다가 2배 정도 급격히 증가하는 것을 확인할 수 있다. 그림 3(c)에 나타난 세 개 의 조기심실수축(PVC) 비트의 경우에는 이전 RR간격은 약 500 ~ 520ms, 이후는 920 ~ 1050ms이며, 그림 3(d)에 나타난 PVC의 경우 이전 RR간격은 약 540 ~ 560ms, 이후는 1300 ~ 1340ms로 세 비트 모두 정상에 비해 매우 작아졌다가 2배 정도 급격히 증가하는 것을 확인할 수 있다. 일반적으로 정상 RR 간격은 0.6s ~ 1.2s의 범위를 가지지만 부정맥의 경우에는 2.0s에 이르는 경우도 있다. MIT-BIH부정맥 데이터베이스를 기준으로 시간 간격을 샘플수로 환산하면 0.6s는 210.6, 1.2ms는 430.2, 2s는 720 샘플 수로 나타낼 수 있다.

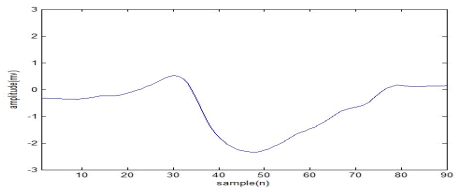
즉 RR간격은 200 ~ 800 샘플 수 범위 안에 나타난다. 그러므로 Normal, PVC, PAC는 QRS간격이 20~60 사이의 샘플 수를 가지며, RR간격은 200~800 사이의 샘플 수를 가진다는 것을 알 수 있으며, 샘플링 구간의 범위가 부정맥의 차이를 판단할 수 있는 중요한 변수가 된다는 것을 확인할 수 있다.



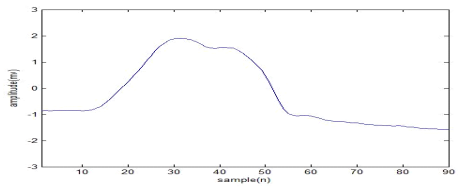
(a) 101, 103 record(Normal)



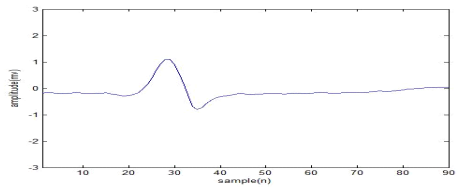
(b) 105, 106, 114, 116, 213 record(PVC)



(c) 200, 233 record(PVC)

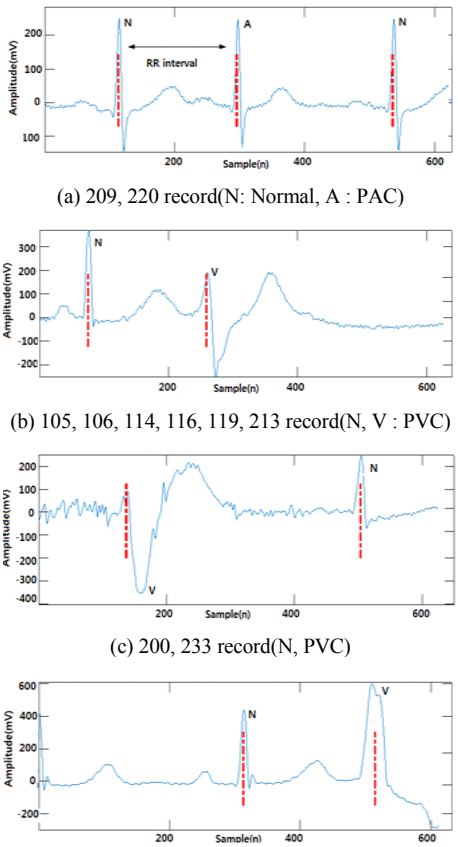


(d) 119, 223 record(PVC)



(e) 209, 220 record(PAC)

Fig. 2 Variability of QRS morphology by each arrhythmia



(a) 209, 220 record(N: Normal, A : PAC)

(b) 105, 106, 114, 116, 119, 213 record(N, V : PVC)

(c) 200, 233 record(N, PVC)

(d) 119, 223 record(N, PVC)

Fig. 3 Variability of RR interval by each arrhythmia

2.3. AR모델링

일반적으로 부정맥 분류를 위한 입력 데이터의 양은 처리 속도에 영향을 미치므로 신호 데이터의 차원을 감소시키는 것이 필요하다. 본 연구에서는 추출된 신호의 요소를 직접 처리하는 대신 계산량을 줄이고 신호의 모델링이 가능한 AR기법을 적용하여 부정맥 분류를 위한 입력 변수로 사용한다. AR모델링 기법은 ECG신호에서 이전과 현재의 신호 파형에 자기상관을 적용하여 계수를 추출함으로써 부정맥 신호들 간의 차별성 있는 특징을 파악할 수 있는 방법이다[9].

AR모델링은 심박변이도(HRV : Heart Rate Variability)를 모델링하고 ECG신호의 전력스펙트럼 추정을 위해 폭넓게 사용된다[10]. AR모델링으로부터 검출된 변수들은 다른 특징들과 함께 결합되어 부정맥 분류를 위해 사용된다[10].

AR모델링은 과거 출력의 피드백에 의하여 현재의 출력이 결정되는 방법으로 시계열 $\{x_n\}$에 대하여 과거 p 점의 값 x_{n-1}, • •, x_{n-p}으로 현재 값 x_n을 예측한다. 이를 나타내면 수식 (1)과 같다.

$$x_n = -\sum_{i=1}^{p} a_i x_{n-i} + e_n \tag{1}$$

여기서 x_n는 심전도의 시계열 신호이며 QRS세그먼트, RR간격 세그먼트 정보를 포함하고 있다. e_n는 예측오차, a_i는 AR계수, p는 AR차수이다. 신호를 정확히 모델링하기 위해서는 e_n을 최소화 하는 것이 필요하다. AR모델링에서 중요한 것은 최적의 차수이며 이를 통해 각 계수가 결정된다. 계수 추정을 위해서는 입력 x_n에 대한 자기상관 함수 사이의 관계를 알아야 한다. 이를 나타낸 것이 *Yule-Walker* 방정식이다.

입력 데이터(x_n)에 대한 자기상관함수 $R_{xx}(i)$는 수식 (2)와 같이 나타난다.

$$\begin{aligned} R_{xx}(i) &= E\left[x_{n+i} x_n^*\right] \\ &= E\left[x_n^*\left(-\sum_{l=1}^{p} a_l x_{n+i-l} + e_{n+i}\right)\right] \\ &= -\sum_{l=1}^{p} a_l R_{xx}(i-l) + E\left[e_{n+i} x_n^*\right] \quad l = 1,2,\ldots\ldots,p \end{aligned} \tag{2}$$

여기서 $E[\,\bullet\,]$는 입력 데이터의 기대값을 의미한다. $E\left[e_{n+i} x_n^*\right]$ 값은 $i=0$일 때는 σ^2, $i>0$일 때는 0을 나타내므로 수식 (2)는 수식 (3)과 같이 나타낼 수 있다.

$$R_{xx}(i) = \begin{cases} -\sum_{l=1}^{p} a_l R_{xx}(i-l), & i>0 \\ -\sum_{l=1}^{p} a_l R_{xx}(-l) + \sigma^2, & i=0 \end{cases} \tag{3}$$

차수(p)를 선택하여 수식(3)으로부터 $i>0$인 경우에 대해 AR계수$\{a_1, a_2 \cdots, a_p\}$을 구할 수 있으며, 분산(σ^2)은 $i=0$에 대한 식으로부터 구할 수 있다.

수식 (3)은 행렬을 이용하여 수식(4)와 같이 표현할 수 있다.

$$\begin{bmatrix} R_{xx}[0] & \cdots R_{xx}(p-1) \\ \vdots & \ddots \quad \vdots \\ R_{xx}(p-1) \cdots & R_{xx}[0] \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R_{xx}(1) \\ \vdots \\ R_{xx}(p) \end{bmatrix} \tag{4}$$

여기서 자기상관 행렬 R_{xx}는 허미트(Hermite) 행렬이다. 이는 (i,j)의 변수가 (i,j)에만 의존하는 특수한 구조를 갖는 정칙행렬을 나타낸다. 수식 (4)를 분산(σ^2)을 포함한 식으로 확장하면 수식 (5)와 같이 표현할 수 있다.

$$\begin{bmatrix} R_{xx}[0] \cdots R_{xx}(p) \\ \vdots \quad \ddots \quad \vdots \\ R_{xx}(p) \cdots R_{xx}[0] \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ \vdots \\ 0 \end{bmatrix} \tag{5}$$

본 연구에서는 수식 (5)에 의한 계수를 계산하기 위해 AR모델링 방법 중 가장 빠르고 우수한 성능을 가지고 있는 Burg알고리즘을 사용하였다. Burg알고리즘은 Levinson의 재귀 관계식을 이용하여 오류 신호들 사이의 재귀관계를 얻고 모델링 차수 p에서 AR계수 $\{a_{11}, a_{12}, \sigma_1^2\}, \{a_{21}, a_{22}, \sigma_2^2\}, \cdots \{a_{p1}, a_{p2}, \sigma_p^2\}$ 집합을 반복해서 구하는 방식이다. 즉, p가 1일 때, a_1을 구하고, p가 2일 때, a_1, a_2, p가 n일 때 모델차수 n만큼 $a_1 \ldots a_n$을 반복해서 구한다. 수식 (1)에 의해 예측 오차(e_n)는 수식(6)과 같이 나타낼 수 있다.

$$e_n = x_n + \sum_{i=1}^{p} a_i x_{n-i} \tag{6}$$

AR계수(a_i)는 수식 (7)과 같이 최소 제곱법(MSPE : mean squared prediction error)을 사용하여 제곱된 예측 오차의 에너지가 최소가 되도록 함으로써 결정된다.

$$MSPE = E[(e_n)^2] = E\left[\left(x_n + \sum_{i=1}^{p} a_i x_{n-i}\right)^2\right] \tag{7}$$

모델링 차수를 증가시켜 신호를 추정하는 것이 더 정확한 신호 추정을 나타내는 것은 아니기 때문에 최적값을 결정하는 것이 매우 중요하다.

2.4. AR모델링 변수 추출

AR모델링 변수로써 QRS와 RR간격을 이용했다. 이는 정상과 부정맥을 결정하고 해석하는데 가장 중요하게 쓰이는 특징이며, 분류의 정확도를 높이기 위함이다. 여기서는 AR모델링 변수로서 QRS구간의 길이($n1$), QRS구간의 차수($p1$), RR간격 구간의 길이($n2$), RR간격 구간의 차수($p2$)에 대한 최적치를 설정하였다. 이 값들은 SVM의 입력 인자로 사용되며, 부정맥의 성능개선에 중요한 인자가 된다. 이는 MIT-BIH부정맥 데이터베이스 48개 레코드에 대한 반복적인 실험을 통해 최적의 값을 선정한 것이다. 최적 R파 검출을 위한 파라미터는 4개의 벡터 값으로 수식 (8)과 같다.

$$AR_{th} = [n1, p1, n2, p2] \tag{8}$$

SVM입력을 위한 최적 모델링 변수를 선정하기 위해 QRS구간의 길이($n1$)의 범위는 20 ~ 60샘플 수까지, RR간격 구간의 길이($n2$)는 200~800 샘플 수까지 2의 간격으로 QRS구간의 차수($p1$)과 RR간격 구간의 차수($p2$)는 2에서 9까지 1의 간격으로 변화시키면서 SVM의 입력 벡터로 적용시켜 실험하였다. 먼저 $n1$은 임시값으로 고정 후 $p1$값을 변경하여 평균 분류율이 제일 높은 $p1$를 선택 고정한 후, $n1$을 변경시키면서 평균 분류율이 제일 높은 $n1$을 선택하는 과정으로 진행된다. $n2$, $p2$도 위와 같은 교차 검증(cross validation)을 통해 최적값을 검출한다.

그림 4, 5는 MIT-BIH부정맥 데이터베이스를 대상으로 한 QRS와 RR구간에 대한 최적 길이와 차수를 각각 나타내고 있다. MIT-BIH 데이터베이스의는 기준 샘플링 주파수가 360Hz이므로 이를 기준으로 적용하였을 때 실험결과 $AR_{th} = [n1, p1, n2, p2]$의 값은 각각 $AR_{th} = [24, 9, 600, 4]$일 때 최적의 검출율을 얻는 것을 확인할 수 있었다.

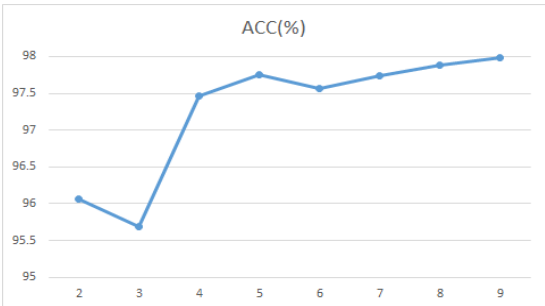


Fig. 4 Accuracy through QRS egment(n1=24), AR order (p1 : 2 ~ 9)

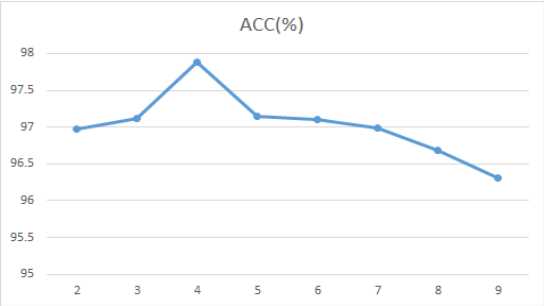


Fig. 5 Accuracy through RR segment(n2=600), AR order(p2 : 2 - 9)

Ⅲ. SVM을 통한 부정맥분류

선형 SVM은 가중치 벡터 w와 바이어스 b로 구성되는 $w \cdot x+b=0$의 최적 분할 경계면을 갖도록 데이터 $(x_i, y_i)^{N_{i-1}}$를 학습시키는 것이다. 여기서 x_i는 i번째 입력 패턴이고, y_i는 원하는 출력이다. 입력 패턴을 최적으로 분할하는 최적 분할 경계면은 비용함수($\theta(w)$)를 최소화하며 비용함수는 $\theta(w)=\frac{1}{2}\|w\|^2$로 표현된다. 따라서 수식 (9)와 같이 위의 제약조건을 가진 최적의 라그랑지 승수(α_i)를 찾으면 최적의 가중치 벡터($\hat{w}$)를 계산할 수 있으며, 최적의 바이어스($\hat{b}$)는 서포트 벡터로부터 계산된다.

$$\hat{w}=\sum_{i=1}^{N}\alpha_i x_i y_i \ , \ \hat{b}=-\frac{1}{2}\hat{w}x_r x_s \tag{9}$$

여기서 x_r과 x_s는 서포트 벡터를 의미한다.

따라서 학습 데이터$(x_i, y_i)^{N_{i-1}}$가 입력으로 주어졌을 때 최적가중치 벡터($\hat{w}$)와 최적 바이어스($\hat{b}$)를 구하게 된다. 또한 선형적으로 분류가 불가능 한 경우에는 일반화된 최적 분리 경계면을 위해 음수가 아닌 스칼라 변수 $\zeta_i \geq 0$를 사용한다. 여기서 $\zeta_i \geq 0$는 잘못된 변수와 관계된 오차의 척도를 의미하는 슬랙변수(slack variable)이다. 이러한 경우 슬랙변수(ζ_i)를 포함한 제약조건과 비용함수($\theta(w,\zeta)$)는 수식 (10)과 같다.

$$y_i(w \cdot x)+b \geq 1-\zeta, \tag{10}$$
$$\theta(w)=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{N}\zeta_i \quad i=1,2.....N$$

여기서 C는 양의 상수값으로 학습오차와 일반화 능력 사이의 상관관계를 조절하는 변수를 나타낸다. 즉 데이터 샘플들이 다른 클래스에 놓이는 것을 허용하는 정도를 결정하고, 이 값은 작을수록 경계의 마진이 넓어지고, 클수록 좁아지면서 세분화된다.

본 논문에서는 앞서 2장에서 언급한 바와 같이 QRS과 RR간격 기반의 두 개의 AR모델을 각각 구간 길이($n1$, $n2$)와 차수($p1$, $p2$) 로 나누고, SVM의 입력 벡터 $(x_i, y_i)^{N_{i-1}}$로 사용한다.

SVM의 특징은 최적 분할 경계면(optimal hyperplane)을 찾기 위하여 전체 데이터를 다루지 않고, 분할 경계면 주면의 몇 개의 서포트 벡터만을 이용함으로 훈련 시간을 현저하게 줄일 수 있다는 장점이 있다[11-12].

3.1. 파라미터에 따른 SVM성능

최적의 분류 성능을 얻으려면 C와 γ를 잘 조정해야 한다. 일반적으로 SVM알고리즘 중에서 가장 널리 사용되는 함수는 가우시안 RBF(Radial Basis Function)커널 함수($K(x_i, x_j)$)이며 수식 (11)과 같이 나타난다.

$$K(x_i, x_j)=exp(-\gamma\|x_i-x_j\|^2) \tag{11}$$

여기서 x_i,x_j는 데이터 포인터이며, γ는 가우시안 커널의 폭을 제어하는 파라미터로서 결정 경계의 곡률을 나타낸다.

C는 수식 (10)에서 설명한 것과 같이 양의 상수값으로 학습오차와 일반화 능력 사이의 상관관계를 조절한다. 두 값 모두 커질수록 알고리즘의 복잡도는 증가하고, 작아질수록 복잡도는 낮아진다. 최적의 C와 γ의 매개변수 쌍을 검출하기 위해 grid search를 통하여 진행하였다. 매개변수 값에 따른 분류율은 표 1과 같이 매개변수 C와 γ값의 조합에 따른 성능 차이를 보여주고 있다. 표와 같이 실험을 통해 C값의 증가에 따라 성능이 향상되고 여백을 희생하는 대신 분할 영역 안에 존재하는 훈련데이터 또는 틀린 훈련데이터의 수를 줄여주고 있다. 결과적으로 최적의 매개변수는 C=200, γ=1일 때라는 것을 확인할 수 있었다.

Table. 1 Comparison of classification rates by parameters

γ \ C	10	50	100	200
0.1	70.00	81.71	84.76	85.01
1	95.34	97.32	97.65	98.06
2	95.03	96.61	97.52	97.82
3	85.75	95.14	97.05	97.87

3.2. R파 검출율

MIT-BIH 부정맥 데이터베이스를 통하여 R파 검출에 대한 성능 평가가 수행되었다. MIT-BIH 부정맥 데이터베이스를 통하여 R파 검출에 대한 성능 평가가 수행되었다. 검출률 계산을 위한 수식은 (12)와 같다. 아래 식에서 FP(False Positive)는 검출알고리즘이 R파를 검출하였으나, MIT-BIH부정맥 데이터베이스에는 존재하지 않을 경우이고, FN(False Negative)은 MIT-BIH부정맥 데이터베이스에는 있으나 알고리즘이 이를 검출하지 못한 경우이다. 전체 비트는 MIT-BIH에 존재하는 R파의 총 개수이다. 전체 레코드를 대상으로 한 평균 R파 검출율은 99.77%의 우수한 성능을 나타내었다.

$$Accuracy = \frac{total\,beat - (FP + FN)}{total\,beat} \times 100\% \qquad (12)$$
$$(FP\text{: } False\,Positive, FN\text{: } False\,Nagative)$$

3.3. 부정맥 분류 성능 비교

제안된 AR모델링을 통한 특징점 추출과 SVM을 통한 부정맥 분류 알고리즘의 성능을 평가하기 위하여 MIT-BIH부정맥 데이터베이스를 대상으로 그 효용성을 실험하였다. 검출율의 계산은 식 (12)와 같으며, MIT-BIH부정맥 데이터베이스 39개의 레코드를 대상으로 부정맥을 분류한 결과를 표 2에 나타내었다. 111번, 112, 113, 115, 122, 212, 214, 232번 레코드의 경우 Noraml, PAC, PVC중 한 종류의 신호만 있는 경우로 분류의 신뢰성을 위해 본 실험에서는 제외하였다.

Normal, PAC, PVC의 평균 분류율은 각각 99.23, 96.62, 97.28%의 우수한 특징을 나타내었다. 특히 분류한 심장질환 중 가장 큰 위험도를 가지고 있는 PVC에 대한 분류율을 기존 알고리즘과 비교하여 표 3에 나타내었다. 제안한 방법은 기존 신경망을 통한 분류방법(ANN)보다 적은 학습 데이터의 양으로 분류의 정확도와 처리속도를 개선시킬 수 있음을 확인 할 수 있었다. 또한 ECG morphology를 이용한 방법과 비교하였을 때 QRS와 RR간격을 분류 변수로 사용한 것은 동일 하지만 제안한 방법은 AR모델링을 통한 세그먼트 특징을 추출과 최적 변수를 사용함 분류의 정확도를 높일 수 있었다. 결과적으로 제안된 AR모델링 기반의 특징점 추출과 SVM을 통한 조기수축 부정맥 분류 방법은 복잡도를 줄임과 동시에 우수한 분류율을 나타내었다.

그러나 그림 6(a)와 같이 233번 레코드는 다른 레코드에 비해 PVC가 normal로 잘못 검출 되는 경우가 나타났는데, 이는 그림 6(a)의 3번째, 6번째 비트에 나타난 것과 같이 PVC임에도 불구하고 normal 이전과 이후가 서로 다른 종류의 PVC가 반복됨으로써 RR간격이 normal과 유사한 형태를 가지고 있기 때문에 정확한 검출이 이루어지지 않았다. 또한 그림 6(b)의 3번째 비트와 같이 간혹 223번과 230번 레코드의 경우 PAC가 normal로 잘못 검출 되는 결과가 발생하였는데 이는 현재 R피크의 진폭이 높은 PAC를 나타내며 RR간격의 변동이 크지 않아 AR모델 변수가 이를 충족시키지 못하는 경우가 발생하였다고 판단된다. 이러한 경우는 R파의 진폭과 같은 간단한 특징점을 입력 벡터로 사용하면 문제가 해결되리라고 판단된다. 따라서 제안된 부정맥 분류 방법은 복잡도를 줄임과 동시에 우수한 검출율을 나타내는 것을 확인할 수 있었다.

Ⅳ. 결 론

본 연구에서는 최적의 R파를 검출하고 이를 통해 최소한의 R피크 기반의 특징점만 추출함으로써 연산의 복잡도를 줄이고, 실시간으로 정확하게 부정맥을 분류하기 위해 AR모델링 기반의 특징점 추출과 SVM을 통한 조기수축 부정맥 분류 방법을 제안하였다. 이를 위해 전처리를 통해 잡음이 제거된 심전도 신호에서 R파를 검출 한 후 QRS와 RR간격의 특정 파형 구간을 추출하여 모델링함으로써 최적 세그먼트 길이($n1, n2$), 최적 차수($p1, p2$)의 4가지 AR모델링 변수를 추출하고 SVM을 통해 정상, PVC, PAC를 분류하였다. 제안한 방법의 우수성을 입증하기 위해 MIT-BIH부정맥 레코드를 대상으로 한 R파의 평균 검출율은 99.77%의 성능을 나타내었으며, Normal, PVC, PAC 부정맥은 각각 평균 분류율은 각각 99.23, 96.62, 97.28%의 평균 분류율을 나타내었다.

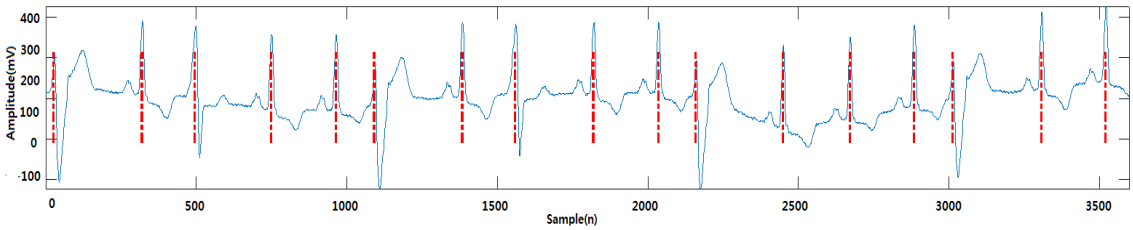
Table. 2 Arrhythmia classification rate

Rec	Total Bit no.	Normal beat	PVC beat	PAC beat	Normal classification Rate(%)	PVC classification Rate(%)	PAC classification Rate(%)
100	1902	1872	1	29	99.95	100	99.98
101	1525	1518	0	3	99.94	-	99.91
102	1822	1	3	0	99.95	96.92	-
103	1729	1727	0	2	99.94	-	99.83
104	1878	98	1	0	99.73	97.2	-
105	2208	2121	29	0	94.05	96.68	-
106	1699	1236	460	0	99.88	96.92	-
108	1601	1460	13	4	95.48	96.39	94.44
109	2102	0	32	0	99.95	99.47	-
112	2117	2109	0	2	99.95	-	99.93
114	1607	1559	30	10	99.56	97.23	98.4
116	2020	1918	98	1	98.96	98.75	98.91
117	1285	1283	0	1	99.92	-	99.85
118	1933	0	13	86	99.95	95.25	98.88
119	1662	1297	364	0	99.94	96.28	-
121	1565	1558	1	1	99.87	99.68	99.78
123	1269	1266	3	0	100	98.04	-
124	1368	0	47	2	99.93	99.1	99.52
200	2185	1438	700	28	99.77	93.47	98.12
201	1525	1185	198	30	98.22	96.88	97.55
202	1874	1800	15	36	96.95	93.61	95.78
203	2520	2103	373	0	99.93	86	-
205	2201	2123	65	2	100	99.72	99.86
207	1611	0	4	107	99.78	92.78	99.28
208	2445	1308	824	0	100	96.45	99.55
209	2522	2146	1	372	99.87	92.66	97.77
210	2207	2010	164	0	100	97.37	-
213	2700	2209	195	24	100	98.35	99.18
215	2795	2661	131	2	100	96.26	98.13
217	1848	235	140	0	99.86	99.86	-
219	1773	1715	51	7	94.57	96.57	94.57
220	1694	1601	0	93	99.9	-	99.9
221	2020	1704	316	0	100	99.9	-
222	2120	1695	0	208	98.75	-	93.75
223	2201	1656	455	65	99.92	89.57	94.75
228	1735	1398	302	3	97.48	95.48	97.48
230	1861	1858	1	0	98.18	98.18	-
233	2562	1859	692	4	99.97	98.9	99.44
234	2291	2238	3	0	100	98.34	-
total	75982	55965	5725	1122	99.23	97.28	96.62

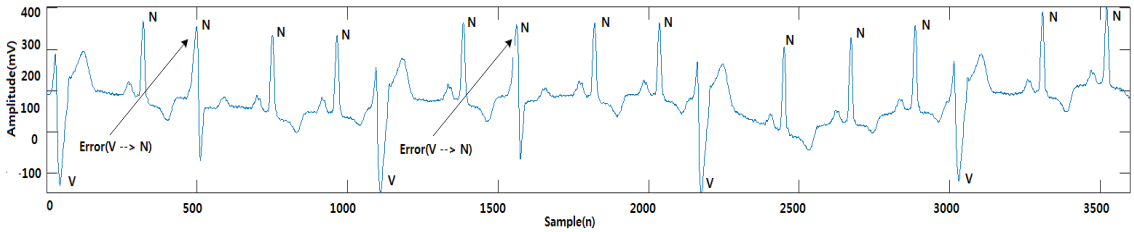
따라서 본 연구에서 제안한 방법은 연산 복잡도를 줄임과 동시에 실시간으로 정확한 부정맥을 진단하는 IoT 헬스케어 시스템에 적용될 수 있을 것으로 판단된다.

Table. 3 Performance comparison between the proposed algorithm and state of the art papers

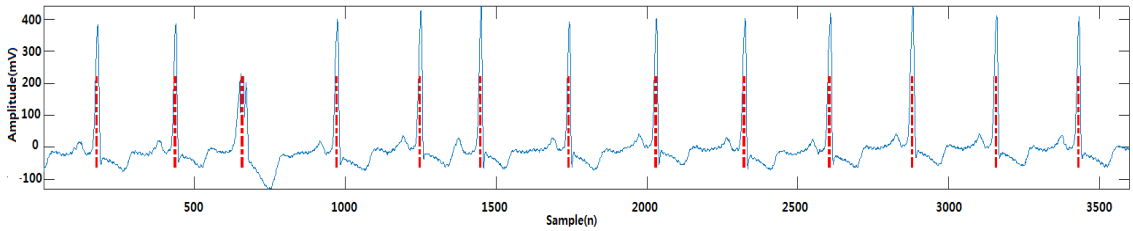
Classification method	Art papers	Avg(%)
ANN	Shyu et al.[8]	97.04
ECG morphology	de Chazal et al.[9]	94.11
Our approach	proposed	97.28



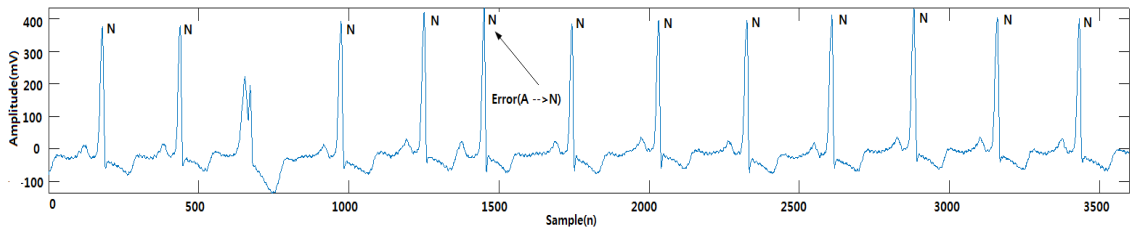
(a) R peak detection result(233 record)



(b) SVM classification result(223 record : PVC bit error)



(c) R peak detection result(233 record)



(b) SVM classification result(223, 230 record : PAC bit error)

Fig. 6 R peak detection & SVM classification result

REFERENCES

[1] S.-H. Liou, Y.-H. Wu, Y.-S. Syu, Y.-L. Gong, H.-C. Chen, and S.-T. Pan, "Real-time remote ECG signal monitor and emergency warning/positioning system on cellular phone," Intelligent Information and Database Systems, vol. 7198, Berlin, Germany: Springer-Verlag, 2012, pp. 336-345.

[2] C. Ye, "Heartbeat classification using morphological and dynamic features of ECG signals," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 10, pp. 2930-2941, October 2012.

[3] M. J. Rooijakkers, Rabotti C, De Lau H, Oei SG, Bergmans JW, and Mischi, "Feasibility Study of a New Method for Low-Complexity Fetal Movement Detection From Abdominal ECG Recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 5, pp. 1361-1368, Sept. 2016.

[4] Y.-L. Zheng, X-R Ding.Chung Y. Poon, B-P Lai Lo, Heye Zhang, X-L Zhou, G-Z Yang, Ni Zhao, Y-T Zhang, "Unobtrusive sensing and wearable devices for health informatics," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1538-1554, May 2014.

[5] S. Pal, and M. Mitra, "Empirical mode decomposition based ECG enhancement and QRS detection," *Computers in Biology and Medicine*, vol. 42, no. 1, pp. 83-92, Jan. 2012.

[6] P. Li, Y. Wang, P. Li, Y. Wang, J. He, L. Wang, Y. Tian, T.-S. Zhou, T. Li, J.-S. Li, "High-Performance Personalized Heartbeat Classification Model for Long-Term ECG Signal," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 78-86, Jan. 2017.

[7] H. Xia, I. Asif, and X. Zhao, "Cloud-ECG for real time ECG monitoring and analysis," *Journal of Computer Methods and Programs in Biomedicine*, vol. 110, no. 3, pp. 253-259, Jun. 2013.

[8] I. S. Cho, and H. S.Kwon "Optimal Threshold Setting Method for R Wave Detection According to The Sampling Frequency of ECG Signals," *Journal of Korea Institute of Information and Communication Engineering*, vol. 21, no. 7, pp. 1420-1428, July 2017.

[9] E. Orosco, P. Diez, E. Laciar, V. Mut, C. Soria, F di Sciascio, "On the use of high-order cumulant and for muscular-activity detection," *Biomedical Signal Processing and Control*, vol. 18, pp. 325-333, February 2015.

[10] T. Niederhauser, W-B Thomas, A. Haeberlin, T. Marisa, and R. A Wildhber, "Graphics-Processor-Unit-Based Parallelization of Optimized Baseline Wander Filtering Algorithms for Long-Term Electrocardiography," *IEEE Transactions on Biomedical Engineering*, vol. 62. no. 6, pp. 1576-1584, February 2015.

[11] Z. Chen, Z. Chen, J. Luo, K. Lin, J. Wu, T. Zhu, X. Xiang, and J. Meng, "An Energy-Efficient ECG Processor With Weak-Strong Hybrid Classifier for Arrhythmia Detection," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 7, pp. 948-952, August 2017.

[12] Q. Li, Q. Li, C. Rajagopalan, and G. D. Clifford, "Ventricular Fibrillation and Tachycardia Classification Using a Machine Learning Approach," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1607-1613, July 2013.

조익성(Ik-Sung Cho)

2003.02 : 한국해양대학교 공학박사
2014.02 : 부산대학교 바이오정보전자공학과 공학박사
2014.03 ~ 2018. 03 : 경운대학교 항공정보통신공학과 조교수
2018.04 ~ 현재 : 대구대학교 창조융합학부 조교수
※관심분야 : ECG Preporcessing, 부정맥 분류 알고리즘

권혁숭(Hyeog-Soong Kwon)

1995.08: 영남대학교 전자공학과 공학박사
1996.04-2006.02 :밀양대학교 정보통신공학과 부교수
2002.02~2003.02: 미국 OSU 방문교수
2011.07~2012.07: 미국 조지아텍 방문교수
2006.03 - 현재 : 부산대학교 IT 응용공학과 정교수
※관심분야 : 생체신호처리, ECG 특징점 추출, CDMA, 통신시스템

김주만(Joo-Man Kim)

1985.01~2000.02 : 한국전자통신 연구원 운영체제 연구팀장(책임연구원)
2000.03~2006.02 : 밀양대학교 정보통신공학과 교수
2006.03~현재 부산대학교 IT 응용공학과 교수
※관심 분야 : 실시간 시스템 제어, 임베디드 시스템, 헬스케어시스템

김선종(Seon-Jong Kim)

1997.03~2006.02 :밀양대학교 컴퓨터공학과 부교수
2006.03 -현재 : 부산대학교 IT 응용공학과 정교수
※관심분야 : 신호및영상처리, 머신비전, 스마트 카메라