

웨이브렛 변수화 기반의 부정맥 분류 알고리즘 최적화

김진권, 이병우, 이명호  
연세대학교 전기전자공학부

Optimization on arrhythmia classification algorithm using wavelet parameterization

Jinkwon Kim, Byoungwoo Lee, MyoungHo Lee  
Dept. of Electrical and Electronic Engineering, Yonsei University

**Abstract** - ECG 기반의 부정맥 자동 분류에 관한 연구는 지난 수십 년간 다양한 방법으로 연구되어 왔다. 많은 연구들이 부정맥을 구별해 낼 수 있는 특징 벡터를 찾아내기 위해 연구하였으나, 피험자의 ECG 특징이 각각 다르기 때문에 부정맥으로 인한 차이와 개인 간 차이를 구별하기 어려웠다. 생체데이터는 그 특성상 서로 다른 특징을 갖고 있으며, 다양한 특징을 가진 사람들에게 적용하기 위한 범용성과 부정맥 검출의 정확성 사이에 교환적 관계를 갖게 된다. 특히 ECG 데이터의 경우 사람 식별 데이터로 사용하고자 하는 연구가 있을 정도로 개인 간 편차가 분명하다. wavelet 분석방법은 다양한 mother wavelet을 사용할 수 있다는 점을 큰 장점으로 가지고 있으며, wavelet parameterization 기법을 사용하여 임의의 직교 wavelet basis를 발생시킬 수 있다. 본 논문은 wavelet parameterization을 사용하여 개인 간의 ECG 파형의 차이를 상쇄시키고, 부정맥의 차이만을 부각시킴으로써 ECG 기반의 부정맥 자동 분류 성능을 높이고자 하는데 목적이 있다.

2.1.1 Wavelet parameterization

Mallat의 Multi-resolution analysis(MRA) 연구에 따르면, scaling function  $\Phi$ 와 wavelet function  $\Psi$  사이에 two-scale recursive 형태로 아래의 식과 같은 관계를 갖고 있다.

$$\Phi(t/2) = \sqrt{2} \sum_n h[n] \Phi(t-n) \tag{1}$$

$$\Psi(t/2) = \sqrt{2} \sum_n g[n] \Phi(t-n) \tag{2}$$

위 식에서  $g$ 와  $h$ 는 FIR 필터와 같은 형태로 나타나며, 서로  $g[k] = (-1)^{1-k} h[1-k]$ 의 관계를 갖는다. 결국, wavelet의 결정에 있어  $h$ 로 나타내지는 필터 계수만을 결정하면,  $g, \Phi, \Psi$ 가 모두 결정이 된다. 계수의 수가  $L$ 인  $h$ 의 경우 scaling function의 존재와 직교성을 보장하기 위해 다음의 조건을 만족해야 한다.

$$\sum_{n=1}^L h[n] = \sqrt{2} \tag{3}$$

$$\sum_{n=1}^L h[n] \cdot h[n-2k] = 0, \quad \text{for } 1, \dots, L/2-1 \tag{4}$$

위의 조건을 만족시키는  $L/2-1$  개의 계수를 결정함으로써 위의 조건을 만족하는 wavelet을 결정할 수 있게 된다. 이 계수들을 각도로 표현하여 wavelet의 결정을 변수화 할 수 있으며, 이를 통해  $-\pi$ 에서  $\pi$ 까지 탐색하여 모든 wavelet의 변형을 나타낼 수 있다. 특히 이 논문에서 사용하고자 하는  $L=6$ 개인 경우,  $\alpha$ 와  $\beta$ 로 나타낼 수 있으며, 주어진  $\alpha, \beta$ 에 대해  $h$ 는 아래의 식과 같이 결정된다[4].

$$\begin{aligned} h(0) &= [(1 + \cos(\alpha) - \sin(\alpha))(1 - \cos(\beta) - \sin(\beta)) + 2\sin(\beta)\cos(\beta)] / (4\sqrt{2}) \\ h(1) &= [(1 - \cos(\alpha) + \sin(\alpha))(1 + \cos(\beta) - \sin(\beta)) - 2\sin(\beta)\cos(\beta)] / (4\sqrt{2}) \\ h(2) &= [1 + \cos(\alpha - \beta) + \sin(\alpha - \beta)] / (2\sqrt{2}) \\ h(3) &= [1 + \cos(\alpha - \beta) - \sin(\alpha - \beta)] / (2\sqrt{2}) \\ h(4) &= 1/\sqrt{2} - h(0) - h(2) \\ h(5) &= 1/\sqrt{2} - h(1) - h(3) \end{aligned}$$

2.2 제안 알고리즘의 평가

제안하는 알고리즘을 평가하기 위해 MIT-BIH arrhythmia database를 기반으로 평가 Dataset을 구성하였다. MIT-BIH DB 중 한 대상에게서 정상 심박과 조기심실수축이 각각 100 회 이상 나타난 사람을 추출하였으며, 이중 가장 특징이 상이한 4명의 대상을 아래 표 1과 같이 선정하였다. 여러 부정맥 중 조기심실수축을 선정한 이유는 가장 위험한 심실성 부정맥의 주요 지표가 되는 심박 종류이며, 정상심박과 모양이 흡사한 심방조기수축과 구별하기 위해서는 morphology 분석을 수행하여야 하기 때문에 본 알고리즘이 제안되었을 때 가장 필요성이 높고, 성능 평가에 적합하기 때문이다. 선정된 dataset은  $h$ 의  $L=6$ 인 wavelet parameterization 방법으로  $\alpha, \beta$ 를 변화시켜가며 optimal wavelet을 선정하였다.  $\alpha, \beta$ 는 각각  $-\pi$ 에서  $\pi$ 까지 360회 sampling 하였다.

<표 1> 연구의 평가 Dataset의 구성

subject file 번호	정상 beat 수	PVC beat 수	특징
106	1507	520	일반적 Normal beat 형태, multifocal 형 PVC
119	2302	109	일반적 Normal beat 형태, QRS complex가 매우 넓은 unifocal 형 PVC
203	2529	444	QRS complex가 넓은 Normal beat 형태, multifocal 형 PVC
215	3196	164	일반적 Normal beat 형태, QRS complex가 좁은 unifocal 형 PVC

optimal wavelet의 선정은 아래의 표 2에 제시된 기준에 따라 선정된다.

1. 서 론

ECG 기반의 심장 부정맥 판별 알고리즘은 지난 수십 년간 신호처리 기법의 발전에 힘입어 많은 성능 향상을 이루어왔다. 하지만 심장 부정맥 판별 알고리즘은 아직 실용으로 널리 사용되고 있지 않으며, 많은 연구들이 이루어지고 있다. 이와 같은 결과는 현재 심장 부정맥 판별 알고리즘이 아직 다양한 대상에 대해 강한 성능을 내지 못하다는 데 원인이 있으며, 최근 이와 같은 성능을 갖춘 알고리즘을 제시하는 연구들이 수행되었다. 기존의 일반적인 제안 알고리즘들이 학습 ECG dataset과 평가 ECG dataset에서 같은 대상의 데이터를 사용함으로써 매우 높은 성능을 낸 반면[1], 최근 발표된 일부 알고리즘의 경우 학습 ECG dataset과 평가 ECG dataset에 같은 대상이 속하지 않게 함으로써 이전보다 낮은 성능이나 높은 평가를 받고 있다[2]. 이는 실제 개발된 알고리즘의 적용 상황에서 모든 대상의 데이터를 알고리즘에 학습시킬 수 없기 때문에 후자의 연구가 보다 실제 상황에 적합한 평가를 거친 것으로 해석할 수 있다. 학습 ECG dataset과 평가 ECG dataset에 같은 대상의 데이터가 섞이지 않게 했을 때 낮은 성능을 나타내는 원인은 생체신호의 특징 때문이다. 생체신호는 대상에 따라 그 특징이 일정하게 나타나지 않으며, 특히 ECG의 경우 대상인식 연구가 수행되고 있을 정도로 대상간의 차이가 분명하다. 이런 대상간의 차이는 심장 부정맥 판별 알고리즘 연구에 있어서는 큰 장애물로 작용하는데, 알고리즘에게 ECG 데이터 중에서 어느 특징이 정상 심박의 대상 간 차이이고, 어느 특징이 부정맥과 정상 심박의 차이인지 학습을 시켜야하기 때문이다. 이 부정맥의 특징은 환자의 질병 원인과 상태에 따라 많은 차이를 보이기 때문에 부정맥과 정상 심박 그룹간의 구별을 더욱 어렵게 하고 있다. 이와 같은 문제를 극복하기 위해 대상에 적응적인 알고리즘(patient adaptive algorithm)을 제안하는 연구가 다양하게 수행되어왔다. 본 논문은 wavelet의 특징을 이용하여 대상에 적응적인 심장 부정맥 판별 알고리즘을 제안하고자 한다.

2. 본 론

2.1 제안 알고리즘의 구성

본 논문은 wavelet의 주요한 장점인 mother wavelet을 응용에 따라 선택할 수 있다는 점을 활용하여 대상에 적응적인 부정맥 판별 알고리즘의 구현 가능성을 평가해보고자 한다. 이미 다양한 wavelet에 대한 부정맥 판별 알고리즘에 대한 연구가 수행되어 왔었으나 본 논문에서는 wavelet parameterization을 통하여 optimal wavelet을 대상에 적응적으로 적용하는 방식을 제안한다. wavelet parameterization은 다양한 orthogonal wavelet을 몇 개의 parameter를 조절하여 얻을 수 있도록 하는 기법을 말하며, 이와 같은 기법을 통하여 이미 응용분야에 적응적인 생체신호 압축 알고리즘[3] 등의 연구가 시도되어 왔다. 본 알고리즘에서는 ECG의 morphology만을 사용하여 정상 심박과 조기심실수축(Premature Ventricular Contraction beat)이 구별될 수 있도록 wavelet을 최적화 하는 방법을 제안한다. 이를 통하여 대상의 차이에 의해 발생한 데이터의 차이를 줄이고, 부정맥으로 인해 발생한 데이터의 차이를 부각시켰으며, Fisher's discriminant ratio를 통해 기존 연구들에서 사용한 특징들과 제안 방식을 통한 특징이 어느 정도 부정맥간 차이를 부각시키는 지에 대해 평가해 보았다.

일반적으로 정상심박이 조기심실수축에 비해 QRS complex의 주파수 구성 성분이 고주파수 성분을 더 많이 포함하고 있기 때문에 wavelet decomposition 후 특정 scale에 신호의 구성 성분을 가장 잘 집중하는 경우를 최적화된 경우로 정하였다. wavelet decomposition은 5 단계로 수행했다.

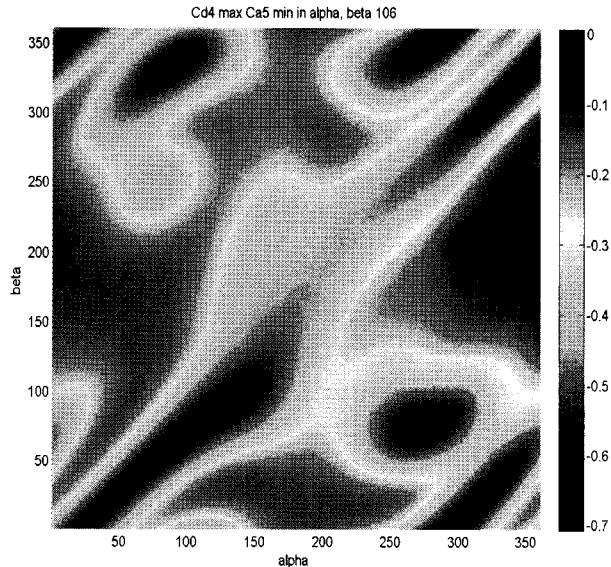
〈표 2〉 Wavelet parameterization의 최적화 기준

최적화 기준	설명
Cd3max	신호의 구성성분이 전체에 대해 scale 3의 detail에 가장 높은 비율을 차지하는 경우
Cd4max	신호의 구성성분이 전체에 대해 scale 4의 detail에 가장 높은 비율을 차지하는 경우
Cd5max	신호의 구성성분이 전체에 대해 scale 5의 detail에 가장 높은 비율을 차지하는 경우
Cd3max Ca5min	신호의 구성성분이 scale 3의 detail에 가장 높은 비율을 scale 5의 approximation에 가장 낮은 비율을 차지하는 경우
Cd4max Ca5min	신호의 구성성분이 scale 4의 detail에 가장 높은 비율을 scale 5의 approximation에 가장 낮은 비율을 차지하는 경우
Cd5max Ca5min	신호의 구성성분이 scale 5의 detail에 가장 높은 비율을 scale 5의 approximation에 가장 낮은 비율을 차지하는 경우

제안하는 알고리즘의 평가는 각 대상에 optimized wavelet을 선정하여 해당 대상의 정상심박과 조기심실수축 심박을 decomposition하여 얻어진 특징들의 Fisher's discriminant ratio(FDR)을 평가하였다. 대조군은 기존의 연구에서 주로 사용되는 db3, sym3, coif1 wavelet으로 전체 대상을 decomposition하여 얻은 특징의 FDR이다. 평가 대상이 되는 특징은 기존의 wavelet을 이용하여 ECG의 morphology를 표현하는 많은 연구들이 decomposition 계수의 scale band의 energy를 특징으로 사용하였으므로, 다음의 6가지 특징을 선정하여 평가해 보았다. 세 가지는 전체 decomposition 계수의 제곱합을 기준으로 scale 5의 approximation 계수의 제곱합의 비(Ca5/Total), scale 5의 detail 계수의 제곱합의 비(Cd5/Total), scale 4의 detail 계수의 제곱합의 비(Cd4/Total)이며, 나머지 세 가지는 scale band 사이의 비율 표시하기 위해 scale 5의 approximation 계수의 제곱합을 기준으로 scale 5 detail 계수의 제곱합의 비(Cd5/Ca5), scale 4의 detail 계수의 제곱합의 비(Cd4/Ca5), scale 5의 detail 계수의 제곱합을 기준으로 scale 4의 detail 계수의 제곱합의 비(Cd4/Cd5)이다.

### 2.3 제안 알고리즘의 결과

제안 알고리즘의 평가 결과를 표 3에 요약하였다. 기존의 연구들에서 가장 보편적으로 사용되는 db, coif, sym wavelet family에서 L=6인 db3, sym3, coif1을 대조군으로 제시하였다. 대조군의 특징들을 보면, 주로 203번 대상은 전체 특징에 대해 FDR이 1 보다 상당히 많이 낮은 것으로 나타났다. 즉 기존의 일반적인 wavelet을 사용한 경우 일부 대상에 있어 판별 성능이 매우 낮게 나타날 수 있다는 것을 나타낸다. 반면 optimal wavelet의 경우 Cd3max Ca5min과 Cd4max Ca5min에서 Cd5/Total 특징이 전체 대상에 대해 모두 1 이상의 FDR 값을 갖는 것으로 보아 일반적인 wavelet에 비교해서 정상 심박과 조기심실수축 심박의 구별 성능을 높여줄 수 있었다.



〈그림 1〉 Cd4max Ca5min 최적화 기준을 만족시키는  $\alpha$ 와  $\beta$  탐색

### 3. 결 론

본 논문은 대상에 적용적인 알고리즘의 방법으로 wavelet parameterization을 제안 하였다. FDR의 측정을 통해 기존의 일반적 wavelet과 비교하여 정상심박과 조기심실수축의 차이를 확장시킬 수 있음을 보였다. 하지만 본 논문은 연구는 부정맥 판별의 전 단계인 특징추출에서 가능성을 보였을 뿐 획득된 특징을 이용하여 대상에 변화에 강한 특징을 갖는 연구를 지속해야 할 것이다.

〈표 3〉 Wavelet parameterization을 이용한 최적 wavelet과 기존 연구들의 사용 wavelet의 특징별 FDR 평가

subject 번호		FDR			
		106	119	203	215
db3 case	Cd5/Total	0.192	4.34	0.139	0.066
	Cd4/Total	1.579	5.34	0.088	0.403
	Ca5/Total	1.571	0.108	0	1.131
	Cd5/Ca5	0.393	3.147	0.069	0.211
	Cd4/Cd5	1.724	17.642	0.132	0.312
	Cd4/Ca5	1.201	3.546	0.052	0.509
coif1 case	Cd5/Total	0.086	5.29	0.742	0.681
	Cd4/Total	2.087	5.546	0.231	0.343
	Ca5/Total	2.07	0.524	0.002	1.25
	Cd5/Ca5	0.308	4.168	0.429	0.656
	Cd4/Cd5	1.609	14.376	0.259	0.283
	Cd4/Ca5	1.495	3.807	0.118	0.473
sym3 case	Cd5/Total	0.192	4.34	0.139	0.066
	Cd4/Total	1.579	5.34	0.088	0.403
	Ca5/Total	1.571	0.108	0	1.131
	Cd5/Ca5	0.393	3.147	0.069	0.211
	Cd4/Cd5	1.724	17.642	0.132	0.312
	Cd4/Ca5	1.201	3.546	0.052	0.509
Cd3max case	Cd5/Total	1.24	15.104	1.471	8.344
	Cd4/Total	1.009	10.925	0.372	3.13
	Ca5/Total	0.322	1.991	0.043	0.793
	Cd5/Ca5	0.039	7.089	0.049	2.288
	Cd4/Cd5	0.304	7.505	0.479	1.02
	Cd4/Ca5	0.002	5.177	0.001	1.239
Cd4max case	Cd5/Total	0.041	9.379	1.009	4.977
	Cd4/Total	2.688	2.235	0.125	1.067
	Ca5/Total	2.187	0.573	0.004	0.075
	Cd5/Ca5	0.504	6.332	0.164	1.922
	Cd4/Cd5	0.529	93.045	1.82	17.663
	Cd4/Ca5	1.094	0.615	0.049	0.23
Cd5max case	Cd5/Total	0.405	0.099	1.005	0.075
	Cd4/Total	1.516	0.496	0.291	1.554
	Ca5/Total	0.977	0.408	0.951	0.288
	Cd5/Ca5	0.577	0.078	0.752	0.139
	Cd4/Cd5	0.628	0.016	0.027	0.916
	Cd4/Ca5	0.976	0.275	0.299	0.198
Cd3max Ca5min case	Cd5/Total	1.036	32.942	1.486	1.073
	Cd4/Total	0.786	19.408	0.366	7.606
	Ca5/Total	0.291	4.976	0.043	0.733
	Cd5/Ca5	0.03	10.175	0.049	0.198
	Cd4/Cd5	0.301	9.627	0.532	0.669
	Cd4/Ca5	0.013	5.647	0.001	0.531
Cd4max Ca5min case	Cd5/Total	0.997	13.193	1.162	5.186
	Cd4/Total	0.797	0.469	0.014	0.685
	Ca5/Total	0.285	1.285	0.02	0.216
	Cd5/Ca5	0.028	8.055	0.073	1.744
	Cd4/Cd5	0.266	64.56	1.056	8.653
	Cd4/Ca5	0.014	0.084	0.001	0.05
Cd5max Ca5min case	Cd5/Total	0.234	21.163	0.256	0.527
	Cd4/Total	1.047	5.961	0.014	7.901
	Ca5/Total	0.382	3.286	0.221	0.073
	Cd5/Ca5	0.055	8.371	0.344	0.215
	Cd4/Cd5	0.129	12.64	0.15	4.774
	Cd4/Ca5	0.036	2.446	0.213	1.567

### Acknowledgement

본 연구는 지식경제부 및 정보통신연구진흥원의 IT핵심기술개발사업의 일환으로 수행하였음. [2005-S-093-03, 생체신호처리기반의 Implantable System 개발]

### 참 고 문 헌

- [1] Jinkwon Kim, et. al., "Algorithm for Classifying Arrhythmia using Extreme Learning Machine and Principal Component Analysis", 2007. EMBS, pp. 3257 - 3260, 22-26 Aug. 2007
- [2] Philip Brechet et. al., "Automatic classification of heartbeats using ECG morphology and heartbeat interval features", Biomedical Engineering, IEEE Trans., Volume 51, Issue 7, PP. 1196 - 1206, 2004
- [3] Laurent Brechet et. al., "Compression of Biomedical Signals With Mother Wavelet Optimization and Best-Basis Wavelet Packet Selection", Biomedical Engineering, IEEE Trans., Volume 54, Issue 12, PP. 2186 - 2192, 2007
- [4] Burrus, C.S., Gopinath, R.A., and Guo, H., "Introduction to wavelets and wavelet transforms", Prentice Hall, pp.53-66, 1997