

부정맥 증상을 자동으로 판별하는 Random Forest 분류기의 정확도 향상을 위한 수정 알고리즘에 대한 연구

이 현 주^{*} · 신 동 규^{**} · 박 희 원^{***} · 김 수 한^{****} · 신 동 일^{*****}

요 약

생체신호의 한 분야인 심전도는 분류알고리즘을 사용한 실험이 일반적이다. 심전도를 실험한 논문에서 사용된 분류알고리즘은 대부분 SVM(Support Vector Machine), MLP(Multilayer Perceptron) 이었으나, 본 실험은 Random Forest 분류기를 시도하였다. 실험방법은 Random Forest 알고리즘을 실험데이터의 신호의 특징에 기반하여 분석하도록 수정하였고, 분류기의 수정된 알고리즘 성능을 규명하기 위하여 SVM과 MLP 분류기와 정확도를 비교 분석하였다. 실험에서는 심전도 신호의 R-R interval을 추출하여 시행하였으며 또한 동일한 데이터를 사용한 타 논문의 결과와 본 실험의 결과를 비교 분석하였다. 결과는 수정된 Random Forest 분류기가 SVM, MLP 분류기, 그리고 타 실험의 결과보다 정확도 부분에서는 우수한 결과를 도출하였다. 본 실험의 전처리 과정에서는 대역통과필터를 사용하여 R-R interval을 추출하였다. 그러나 심전도 실험에서는 대역통과 필터 뿐 아니라, 웨이블릿 변환, 메디안 필터, 유한 임펄스 필터 등으로 실험하는 경우가 많다. 따라서 향후에는 전처리과정에서 기저선 잡음(baseline wandering)을 효율적으로 제거하는 필터의 선택이 필요하며, R-R interval을 정확하게 추출할 수 있는 방법에 대한 연구가 필요하다고 사려된다.

키워드 : 심전도, Random Forest, SVM, MLP, R-R Interval, 분류기, 정확도

Research on the modified algorithm for improving accuracy of Random Forest classifier which identifies automatically arrhythmia

HyunJu Lee^{*} · DongKyo Shin^{**} · HeeWon Park^{***} · SooHan Kim^{****} · DongIl Shin^{*****}

ABSTRACT

ECG(Electrocardiogram), a field of Bio-signal, is generally experimented with classification algorithms most of which are SVM(Support Vector Machine), MLP(Multilayer Perceptron). But this study modified the Random Forest Algorithm along the basis of signal characteristics and comparatively analyzed the accuracies of modified algorithm with those of SVM and MLP to prove the ability of modified algorithm. The R-R interval extracted from ECG is used in this study and the results of established researches which experimented co-equal data are also comparatively analyzed. As a result, modified RF Classifier showed better consequences than SVM classifier, MLP classifier and other researches' results in accuracy category. The Band-pass filter is used to extract R-R interval in pre-processing stage. However, the Wavelet transform, median filter, and finite impulse response filter in addition to Band-pass filter are often used in experiment of ECG. After this study, selection of the filters efficiently deleting the baseline wandering in pre-processing stage and study of the methods correctly extracting the R-R interval are needed.

Keywords : ECG, Random Forest, SVM, MLP, R-R Interval, Classifier, Accuracy

1. 서 론

심전도(ECG: Electrocardiogram)는 심장활동 시에 나타나는 전기적 신호로서 심장 상태와 질환을 알아 볼 수 있는 중요한 자료로 쓰인다[1]. 심전도 신호는 P, Q, R, S, T의 다섯 파형으로 구성되며, 파형의 높낮이와 간격의 특징에 따라 신호를 분류하고, 심장의 상태와 질병의 유무를 판단하여 심전도데이터를 구성 할 수 있다. 심전도 신호로 알 수 있는 질환으로 부정맥이 있는데 일반적으로 심장박동의

* 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2009년도 문화콘텐츠 산업기술지원사업의 연구결과로 수행되었음.

† 준 회 원 : 세종대학교 컴퓨터공학과 박사과정

†† 중 신 회 원 : 세종대학교 컴퓨터공학과 교수

††† 정 회 원 : 삼성전자 VD사업부 상무

†††† 정 회 원 : 삼성전자 VD 사업부 수석연구원

††††† 중 신 회 원 : 세종대학교 컴퓨터공학과 교수(교신저자)

논문접수 : 2011년 6월 8일

수정일 : 1차 2011년 8월 1일, 2차 2011년 9월 6일

심사완료 : 2011년 9월 20일

리듬이 비정상적으로 빠르거나 너무 느린 상태를 말한다[2]. 이 부정맥의 연구를 위하여 공개된 데이터베이스로는 MIT-BIH Arrhythmia Database가 존재한다.

심전도 신호는 일반적으로 R-R interval과 QRS-Complex등의 간격을 추출하여 실험한다. Tsipouras, Fotiadis 그리고 Sideris[3]는 R-R interval의 신호에서 발생하는 심장박동의 특징에 따라 부정맥을 탐지하고 분류하였다. 그들은 첫째, 부정맥 신호에서 발생하는 beat를 분류하여 신호를 탐지하였고 둘째, 부정맥신호에서 발생하는 6개의 특징을 가지고 단계별로 부정맥 추출작업을 실행하였다. SVM(Support Vector Machine)분류기를 응용한 Song[4]의 실험은 정상리듬(NSR)분류기, 심실빈맥(VT)분류기, 심실세동(VT)분류기 그리고 부정맥분류기와 같은 SVM다원 분류기로 실험하였고, 각각의 분류기는 입력되는 신호의 특징에 따라서 부정맥이 검출 될 시에 양의 값을 출력하였다. 이와 같이 심전도 실험에서는 SVM분류기와 MLP(Multilayer Perceptron)분류기가 가장 많이 사용되었다.

그런데 부정맥 실험결과와 정확도를 향상시키기 위한 노력의 일환으로 SVM과 MLP와 알고리즘이 다른 Random Forest 분류기의 성능을 실험할 필요가 있다고 본다. 따라서 본 고는 Random Forest의 분류기의 성능을 알기 위해서 SVM과 MLP 분류기의 정확도를 비교 분석 하였다. 또한 동등한 데이터와 R-R interval을 추출하여 실험한 타 논문들의 결과와 본 실험의 결과를 비교 분석하였다. R-R interval을 MIT-BIH Arrhythmia Database에서 제공한 beat annotation을 참고하여 신호를 분류하고 데이터를 구성하였으며 분류기 알고리즘의 수정도 시도하였다.

2. 이론적 배경

2.1 MIT-BIH Arrhythmia Database

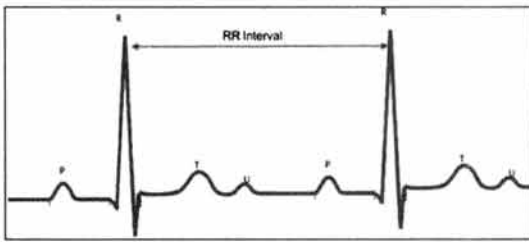
MIT-BIH(The Massachusetts Institute of Technology - Beth Israel Hospital) Arrhythmia Database(부정맥 데이터베이스)[5]는 BIH(Beth Israel Hospital) 부정맥 실험실에서 1975년 이래로 Boston's Israel Hospital 과 MIT의 지원을 받아서 부정맥의 분석과 관련된 주제들을 연구한 데이터이다. MIT-BIH Arrhythmia는 최초로 부정맥 탐지와 평가를 위한 테스트 데이터를 위해서 일반적으로 사용이 가능해진 부정맥 데이터이며, 전체 데이터의 기록은 시간과 채널당 360 샘플에서 디지털화된 기록이다. 이 기록은 1975년과 1979년 사이에 BIH 부정맥 실험실에서 연구된 것이다. 이 데이터는 총 48개로 구성되었는데, 23개의 기록은 40%의 외래환자와 60%의 입원환자 두 집단에서 선택되었고, 이는 4000개의 데이터 집합으로부터 임의로 수집된 것이다. 나머지 25개의 데이터는 23개의 데이터와 동일한 집단에서 수집된 데이터이지만, 임상적으로 중요한 부정맥 신호를 가지고 있다<표 1 참조>. 실험데이터의 class파일(속성과파일)은 annotation symbol을 바탕으로 구성하였다.

2.2 R-R interval 특징 추출

R-R interval은 R파의 한 Peak에서 그 다음에 측정되는 Peak까지의 시간을 의미하며, 각각의 interval은 한 번의 cardiac cycle로 구성된다. (그림 1)은 R-R interval을 나타낸 것이다[6].

<표 1> MIT-BIH 부정맥 데이터베이스의 beat annotation

annotation symbol	Type of arrhythmia	Classification
N	Normal beat	Normal (N)
P	Paced beat	
f	Fusion of paced and normal beat	
x	Non-conducted P-wave (blocked APB)	
L	Left bundle branch block beat	
R	Right bundle branch block beat	
Q	Unclassifiable beat (beat annotations)	
v	Premature ventricular contraction (beat annotation)	Premature ventricular contraction (PVC)
A	Atrial premature beat	Atrial premature (A)
[Start of ventricular flutter/fibrillation	Ventricular flutter/fibrillation (VF)
!	Ventricular flutter wave	
]	End of ventricular flutter/fibrillation (beat annotations)	
(BII	28 heart block (rhythm annotation)	28 heart block (BII)
/	Paced beat	
L	Left bundle branch block beat	



(그림 1) R-R interval

R-R interval은 연속된 시간의 형태로 반복하여 지속적으로 발생되는데, 심전도 신호에서 QRS 검출기를 적용할 때 R-R interval의 sequence가 변형된다[7].

$$RR_i = t_i - t_{i-1} \quad (\text{식 1})$$

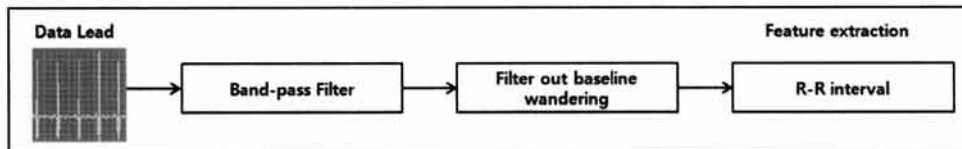
R-R interval의 sequence는 시간의 연속성에 기초하여 형성되며, 즉각적인 심장비율에 상응하는 해당 sequence는 아래의 식으로 정의된다[7].

$$f_i = \frac{1}{RR_i} \quad (\text{식 2})$$

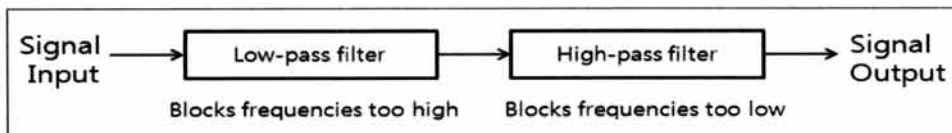
(그림 2)는 R-R interval의 특징추출을 나타낸 것이다. 전처리 과정에서는 필터를 사용하여 데이터의 기저선 잡음을 걸러낸다. 필터는 원하지 않는 성분(주파수 성분)을 제거하며, 거리, 속도, 가속도, 온도, 힘 등을 계측하여 전기적 신호로 변환하는 장치이다. 필터는 메디안필터(Median Filter), 유한 임펄스필터(finite impulse response), 웨이블릿 변환(Wavelet transform), 푸리에 변환(Fourier transform)과 대역통과필터(Band-pass Filter)등이 존재한다.

본 실험에서는 NI Labview(National Instrument Labview)에서 제공되는 Biomedical Startup Kit 3.0을 추출작업에 사용하였으므로, Kit에서 제공하는 대역통과필터를 사용하였다(그림 3 참조). 대역통과필터는 single filter 안에서 low-pass와 high-pass를 조합하여 잡음을 걸러내도록 디자인된 필터이다[8].

그리고 대역통과 필터의 low-pass를 10Hz, high-pass는 25Hz로 설정하였다. 설정된 필터는 데이터 신호의 잡음을 제거하고, 제거된 신호를 통해 R-R interval을 추출한다. 추출된 R-R interval은 분류기 실험에서 사용된 Weka에서 실험할 수 있는 형태로 디자인되었다. 디자인된 데이터는 Weka에서 제공되는 분류기의 하나인 Random Forest 분류



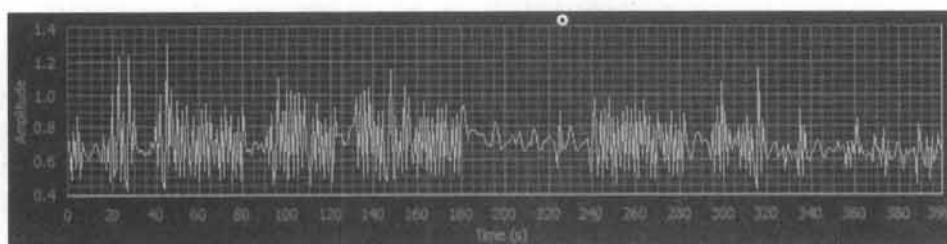
(그림 2) R-R interval의 특징추출



(그림 3) 대역통과필터의 입·출력 모드



(그림 4) 특징 추출전의 부정맥 데이터



(그림 5) 특징 추출후의 부정맥 데이터

기로 실험하였다. (그림 4)와 (그림 5)는 특징추출전과 추출 후의 부정맥 데이터를 나타낸 것으로, 추출된 신호는 신호의 간격과 높낮이에 따라서 정상적인 신호와 부정맥 신호로 분류된다.

3. 본 론

3.1 Random Forest 분류기의 수정 알고리즘과 수식

3.1.1 Random Forest 수정 알고리즘

1998년 Ho는 한 논문에서 Random Subspace는 각 트리에서 성장한 subset을 사용하여 임의로 선택하는 방법이라고 언급하였고, 1년 후 Breiman은 임의로 추출되는 새로운 분석 데이터를 설계하여 원본분석 데이터에서 사용하였다 [9]. Random Forest는 임의의 벡터선택으로 사용되는 알고리즘으로 Decision Tree 분류기를 위해 특별히 설계된 앙상블 기법이다. 각각의 Decision Tree는 일정한 확률분포로부터 생성된 임의의 벡터를 사용한다. 임의의 벡터는 트리가 성장할 때 각 노드의 분할을 위해 무작위로 선택된 F개의 입력 특징을 전체적으로 조사하는 대신에 선택된 F개의 특징으로부터 결정된다[10]. 그리고 Forest-RI와 Forest-RC라는 입력 특징을 갖는데 Forest-RI는 RI의 벡터를 무작위로 선택하여 입력하는 방법이다. Forest-RC는 F개의 입력특성이 일반적인 선형조합에 도달하면 입력한 데이터를 최적의 상태로 분리한다[9]. 수정된 알고리즘에서 Forest-RI는 가장 많이 나타나는 신호를 먼저 선택하고, Forest-RC는 부정맥 신호를 추적하여 분류하도록 고안하였다. 그리고 Decision Tree 대신에 Best-First decision tree(B-F Tree)를 적용하였다. 일반적으로 Decision Tree는 주어진 데이터를 분류하는 목적으로 사용되는 알고리즘으로 목표변수를 기준으로 분류한다. 또한 트리는 목표변수와 상관도가 높은 변수를 선택하여 그 변수의 범주를 조합하고 상관도가 높은 범주를 분리하는데 이는 데이터의 특성에 따른 한계를 갖는다. 그러므로 데이터의 특성이 특정 변수에 수직적으로 구분되지 못할 때 분류율이 떨어지고 트리가 복잡해지므로 최적의 정확도를 보장하지 못한다. 따라서 이러한 단점을 보완하기 위해서 B-F Tree를 대신 적용하여 알고리즘을 수정하였다. B-F Tree는 고정된 명령 대신에 best-first 명령으로 node들을 확장하는 방법으로 각 단계에서 최적으로 분리된 node를 더해주며 분리가 필요한 모든 node에서 발생하는 오류를 최소화 한다. 각 단계에서 트리는 최적화된 subset를 선택하여 확장한다. 그리고 축소된 프로세스는 모든 node들이 특정한 수 또는 pure node에 도달할 때까지 확장된다[11]. B-F Tree는 가지치기(Pruning) 부분에서 pre-pruning과 post-pruning 두 방법을 실행할 수 있다. 첫 번째 Pre-pruning은 트리가 성장할 때 분리하는 데이터가 실제적이지 않으면 성장을 중지한다. 그리고 더 나아가서 분할 할 때는 평균 오류건적을 증가시켜 최종 확장한 수를 가지고 이전에 확장된 수를 선택하는 방법이다. 두 번째 Post-pruning은 모든 트리들이 완전히 확장될 때까지 노드

를 지속적으로 확장하는 방법으로, 확장 한 수와 가지의 평균 오류 건적을 최소화 하여 선택한다[11]. 두 사례는 최종 트리의 모든 데이터와 확장되어 선택한 수에 기반하여 만들어졌다.

3.1.2 수식

Random Forest 분류기로 실험하여 TP(True Positive), TN(True Negative), FP(False Positive) 그리고 FN(False Negative)을 도출 한 후에 Accuracy(정확도), Sensitivity(민감도), Specificity(특이도), DER(Detection Error Rate)과 PPR(Positive Predictive Rate)를 측정하였다. 수식은 다음과 같다.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{Correctly\ Classified}{Total\ Instance} \quad (식\ 3)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (식\ 4)$$

$$Specificity = \frac{TN}{TN + FP} \quad (식\ 5)$$

$$DER = \frac{FN + FP}{TP + FN} \quad (식\ 6)$$

$$PPR = \frac{TP}{TP + FP} \quad (식\ 7)$$

만일 식의 값을 %로 구하고자 한다면, 각 식의 값을 구한 뒤 100을 곱해 주면 된다.

3.2 실험방법

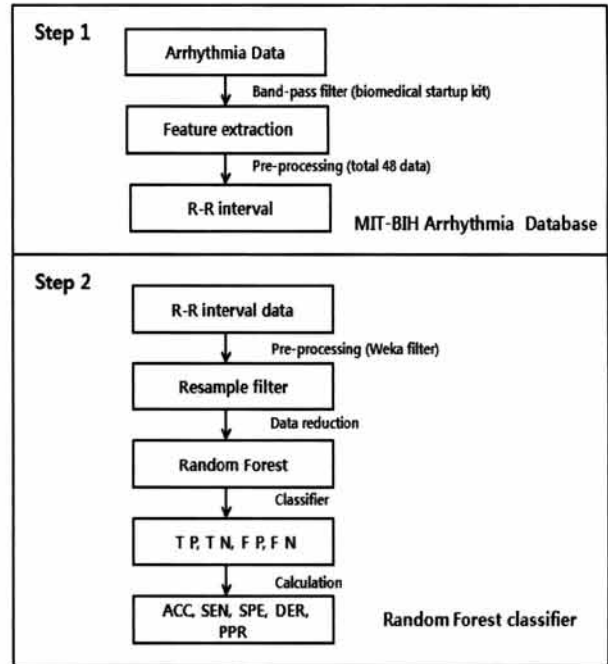
실험데이터는 대역통과필터(Band-pass filter)를 사용하여 Narrow와 Wide 두 파형으로 R-R interval을 추출하여 구성하였다. 데이터의 속성파일(class파일)은 2.1에서 설명한 바 처럼 beat annotation을 참고하여 만들었다. 구성된 부정맥 실험데이터에서 가장 많이 발생한 beat는 N(Normal)신호와 ~ (Change in signal quality)신호이다. N은 정상적인 파형을 뜻하고 ~는 파형이 현재에서 다른 파형으로 전환됨을 나타낸다. 부정맥을 의미하는 beat는 V(Premature Ventricular Contraction)신호와 A(Atrial Premature Beat)신호가 가장 많이 발생하였는데 본 실험은 빈번하게 발생하는 beat를 기반으로 알고리즘을 수정하였다. 수정한 알고리즘에서 Forest-RI는 N(Normal)신호와 ~ (Change in signal quality)신호를 먼저 선택하고, Forest-RC는 V(Premature Ventricular Contraction)신호와 A(Atrial Premature Beat)신호를 추적하여 부정맥을 분류하도록 고안하였다. 그리고 Forest-RI는 N신호와 ~신호를 분리한 후에 다른 신호들을 추적하여 분류하도록 설계하였다. 수정한 알고리즘(The Algorithm)은 아래와 같다.

The Algorithm:

1. Forest-RI는 F= N(Normal), F= ~ (Change in signal quality)을 우선순위로 분류
 - 1.1 N = 정상신호
 - 1.2 ~ = 신호를 변경
2. Forest-RI는 V(Premature Ventricular Contraction), A(Atrial Premature Beat)를 분류
3. Forest-RC는 V, A를 Arrhythmia로 분류
 - 3.1 V = Arrhythmia
 - 3.2 A = Arrhythmia
4. Forest-RI는 N, ~, V, A 외에 다른 신호를 추적하여 분류
5. Forest-RI와 Forest-RC 반복
6. end Forest-RI/ Forest-RC

본 실험에서는 Forest-RI와 Forest-RC의 수정 외에도 Decision Tree 대신에 Best-first decision tree(B-F Tree)를 적용하여 오류를 줄였다. B-F Tree는 고정된 명령대신 best-first 명령으로 node들을 확장하는 방법으로 분리에이터가 실제적이지 않으면 성장을 중지하고, 최종 확장한 수를 가지고 판단하며 확장한 수와 가지의 평균 오류 견적을 최소화 하므로 상대적으로 Decision Tree 보다 오류율을 줄일 수 있다. Forest-RI와 Forest-RC에게 신호를 선택하여 추적·분류하도록 하였을 때 실험결과는 TP(True Positive)가 수정 전보다 상대적으로 높은 값을 도출하였고, 반면에 FP(False Positive)의 값은 줄었다. 그리고 B-F Tree를 적용하였을 때의 오류율(Out-Of-Bag)이 적용전보다 상대적으로 줄어 정확도가 전보다 높게 도출되었다. B-F Tree의 성능이 Decision Tree 보다 우수함을 확인하기 위하여 한 데이터를 선택하여 실험하였는데 실험결과는 Decision Tree는 트리의 크기는 341, leave는 171로 정확도는 90.69%가 도출되었다. 반면에 B-F Tree는 트리의 크기는 567, leave는 284로 정확도는 93.37%가 도출되었다.

분류기실험은 Weka-3.6.2 버전에서 Random Forest분류기로 진행하였다(그림 6 참조). 실험은 전처리 단계에서 R-R interval을 추출하였고, 추출된 데이터를 Weka에서 실험할 수 있도록 데이터를 구성하여 시행하였다. Weka에서의 실험은 먼저 unsupervised에서 제공된 Re-sample 필터를 사용하여 데이터를 축소하였고, 축소된 데이터를 Random Forest 분류기로 실험하였다. Random Forest 분류기에서의 실험은 다중교차검증(k-fold cross-validation)방법으로 진행하였다. 다중교차검증(k-fold cross-validation)방법은 데이터를 k개의 동일한 크기 구획으로 분할하는 방법으로 구획들 중 하나가 시험용으로 선택되고 나머지는 훈련용으로 사용되는 방법이다[10]. 이러한 절차에 따라서 각 구획이 시험용으로 정확히 한번 만 사용되도록 k번 반복되며 총 오류는 k번의 모든 수행에 대한 오류들을 합해서 얻어진다. 본 실험에서는 k의 값을 10으로 설정하여 10-fold cross-validation으로 실험을 진행하였고, 도출된 TP, TN, FP, FN의 수치를 기반으로 Accuracy(정확도), Sensitivity



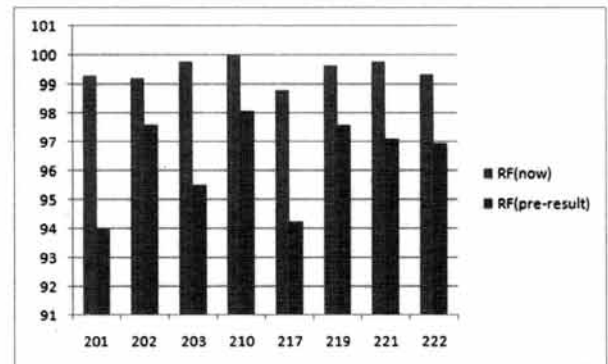
(그림 6) 실험과정

(민감도), Specificity(특이도), DER(Detection Error Rate)과 PPR(Positive Predictive Rate)를 측정하였다.

3.3 실험결과

3.3.1 Random Forest 분류기 알고리즘의 수정 전(pre-result)과 후(now)의 정확도 비교

RF(Random Forest)는 정확도의 향상을 위해서 본 실험에서는 우선 신호를 선택하여 추적하는 방법으로 수정하였다. (그림 7)은 Random Forest의 수정전과 후의 정확도를 나타낸 그래프이다.



(그림 7) Random Forest의 수정전과 후의 정확도 비교

데이터는 K&L[12]의 실험데이터를 참고하여 선택하였다. 그래프는 왼쪽에서부터 RF(now)-> RF(pre-result) 순서이다. now는 알고리즘 수정후의 실험결과이며 pre-result는 수정 전의 실험결과이다. 그래프를 살펴보면 수정 후의 알고리즘이 수정전보다 높은 정확도를 도출하였음을 알 수 있다.

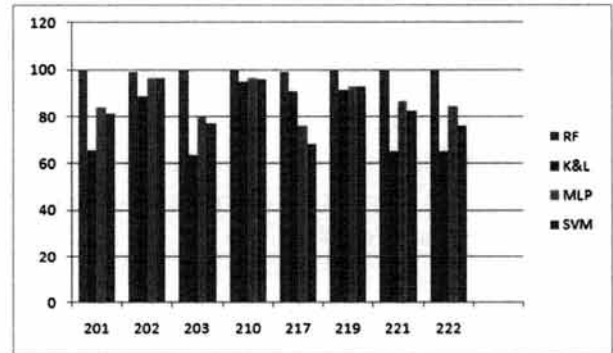
3.3.2 타 논문과 SVM, MLP와 Random Forest의 정확도 비교

<표 2>는 타 논문과 RF(Random Forest)에서 실험하였던 Sensitivity(민감도), Specificity(특이도), Accuracy(정확도)를 도출한 결과이다.

<표 2> Sensitivity, Specificity, Accuracy(정확도)측정

Record_no	Sensitivity (%)		Specificity (%)		Accuracy (%)	
	K&L	RF	K&L	RF	K&L	RF
201	96.44	99.71	39.52	88.23	65.46	99.28
202	80.79	99.62	94.64	72.73	88.72	99.2
203	81.36	99.89	21.92	71.43	63.37	99.77
210	96.89	100	0	0	94.73	100
217	72.78	99.51	94.22	27.27	90.86	98.81
219	96.86	99.81	64.20	89.66	91.39	99.63
221	92.25	99.86	50.66	94.87	65.24	99.76
222	92.25	99.78	50.66	91.94	65.24	99.32
Mean	88.63	99.77	51.98	67	78.13	99.47

<표 2>에서 K&L은 K. Tateno[12]의 결과이고, RF는 Random Forest로 본 실험의 결과이다. 정확도를 비교해보면 다음과 같다. K&L의 실험에서 210번 데이터가 결과가 가장 좋았다. 210번 데이터를 RF의 결과와 비교해보면 RF가 5.27% 높은 정확도를 도출하였음을 알 수 있다. RF의 정확도를 비교 분석하기 위해서 (그림 8)은 타 분류기인 SVM(Support Vector Machine)과 MLP(Multilayer Perceptron) 그리고 K&L과 RF의 정확도를 그래프로 나타낸 것이다. 데이터는 K&L의 실험데이터와 동일하게 선택하였다. 그래프는 RF-> K&L-> MLP-> SVM 순서이다. (그림 8)을 살펴보면 RF가 다른 실험에 비해서 100%에 근접해 있음을 알 수 있다.



(그림 8) RF와 K&L, MLP, SVM의 정확도 분석

<표 3>은 48개의 R-R interval 특징 데이터를 Random Forest로 실험한 결과를 나타낸 것이다. TP, TN, FP 그리고 FN의 값을 먼저 도출한 후에 DER(Detection Error Rate), PPR(Positive Predictive Rate), SEN(Sensitivity%), SPE(Specificity%), ACC(Accuracy%)를 계산하고 평균값(Mean)을 구하였다. <표 3>을 살펴보면 ACC에서 100% 결과를 도출한 데이터가 11개 존재한다. 103, 104, 107, 113, 115, 117, 121, 122, 123, 210, 228번 데이터이며 이를 포함한 전체 데이터는 90% 이상의 정확도를 도출하였다.

따라서 Random Forest 분류기는 전체 데이터 결과에서 효율적인 정확도를 도출하였다.

4. 결론 및 연구방향

본 논문에서는 SVM과 MLP 분류기의 정확도와 Random Forest 분류기의 정확도를 MIT-BIH Arrhythmia Database를 사용하여 비교 분석하였다. R-R interval 데이터 추출(전 처리 과정)에는 Biomedical Startup Kit을 이용하였고, Random Forest 분류기는 Weka에서 제공된 것을 수정하여 활용하였다.

<표 3> Random Forest 분류기로 실험한 R-R interval 특징데이터 실험결과

record	TP	TN	FP	FN	DER	PPR	SEN	SPE	ACC
100	2257	0	0	4	0.1561	100	99.73	100	99.82
101	1793	5	1	0	0.0557	99.94	100	83.33	99.94
102	2085	27	64	0	3.0695	97.02	100	29.67	97.06
103	1947	0	0	0	0	100	100	0	100
104	2104	0	0	0	0	100	100	0	100
105	1570	17	4	0	0.2547	99.74	100	81	99.74
106	1570	17	4	0	0.2547	99.74	100	81	99.75
107	2120	0	0	0	0	100	100	0	100
108	1808	7	4	4	0.4415	99.77	99.77	63.63	99.56
109	2500	0	34	0	1.36	98.65	100	0	98.65
111	2100	13	0	3	0.1426	100	99.85	0	99.85
112	2528	8	4	0	0.1582	99.84	100	66.66	99.84
113	1772	0	0	0	0	100	100	0	100
114	1856	9	4	4	0.4301	99.78	99.78	69.23	99.57

115	1936	0	0	0	0	100	100	0	100
116	2345	0	21	12	1.4	99.11	99.49	0	98.61
117	1531	2	0	0	0	100	100	100	100
118	2209	18	24	19	1.93	98.92	99.15	42.85	98.106
119	1321	49	3	1	0.3026	99.77	99.92	94.23	99.71
121	1855	0	0	0	0	100	100	0	100
122	2460	0	0	0	0	100	100	0	100
123	1512	0	0	0	0	100	100	0	100
124	1542	13	1	2	0.1943	99.93	99.87	92.85	99.81
200	1852	59	9	12	1.1266	99.516	99.35	86.76	98.91
201	1748	60	8	5	0.7415	99.54	99.71	88.23	99.28
202	2081	24	9	8	0.8137	99.57	99.62	72.73	99.2
203	2972	10	4	3	0.2352	99.86	99.89	71.43	99.77
205	2596	33	20	1	0.8086	99.23	99.96	62.26	99.2
207	2138	18	4	4	0.3734	99.81	99.81	81.81	99.63
208	1997	63	17	7	1.1976	99.15	99.65	78.75	98.84
209	2877	12	135	27	5.5785	95.51	99.07	8.16	94.69
210	2352	25	0	0	0	100	100	0	100
212	2644	33	10	32	1.5695	99.62	98.8	76.74	98.45
213	2885	343	9	55	2.1768	99.68	98.13	97.44	98.05
214	1933	7	6	6	0.619	99.69	99.69	53.84	99.38
215	2556	23	10	13	0.8952	99.61	99.49	69.7	99.11
217	2246	6	16	11	1.1962	99.29	99.51	27.27	98.81
219	1600	26	3	3	0.3742	99.81	99.81	89.66	99.63
220	2015	6	14	9	1.1363	99.31	99.55	30	98.87
221	2066	37	2	3	0.2416	99.9	99.86	94.87	99.76
222	2355	137	12	5	0.7203	99.49	99.78	91.94	99.32
223	2261	0	4	4	0.3532	99.82	99.82	0	99.64
228	1983	0	0	0	0	100	100	0	100
230	2224	0	2	2	0.1821	99.91	99.91	0	99.82
231	941	0	1	1	0.2123	99.89	99.89	0	99.78
232	1725	27	21	6	1.5597	98.79	99.65	56.25	98.48
233	2195	0	2	2	0.1821	99.91	99.91	0	99.82
234	2690	39	20	2	0.8172	99.26	99.93	66.1	99.2
Mean	2076	24.43	10.541	5.625	0.6929	99.55	99.75	46	99.326

두 집단 간의 차이를 뚜렷이 하기 위하여 Random Forest 분류기의 알고리즘을 다음과 같이 세 단계로 수정하였다. 첫 번째, 신호를 무작위 선택 방법에서 고 빈도 신호를 우선 선택하는 방법으로 수정하였다. 두 번째, 최적 분리에서 부정맥 신호를 탐지하여 분류하도록 수정하였다. 세 번째, Decision Tree를 사용하는 대신에 Best-first decision tree를 적용하였다. 그 결과 Random Forest 분류기 정확도를 최대화할 수 있어 SVM과 MLP 분류기보다 성능이 우수하였고 오류를 줄일 수 있었다. 그리고 K. Tateno의 연구 결과와 비교하여 정확도 면에서 수정된 알고리즘이 높은 것을 확인할 수 있었다.

결과적으로 정확도 향상에서는 좋은 결과를 얻었으나, 전

처리과정에서의 데이터 차원축소와 사용한 대역통과필터(Band-pass Filter)의 한계극복과 효율적인 R-R interval 구간분리에서는 다른 실험의 과정보다는 미흡한 결과를 보였다. 그러므로 향후에는 전처리과정에서 기저선 잡음(baseline wandering)을 효율적으로 제거할 수 있는 필터의 선택과 R-R interval을 정확하게 추출할 수 있는 방법에 대한 연구가 필요하다.

참 고 문 헌

- [1] K. S. Park, B. H. Cho, D. H. Lee, S. H. Song, J. S. Lee, Y. J. Chee, I. Y. Kim, and S. I. Kim, "Hierarchical Classification

of ECG Beat Using Higher Order Statistics and Hermite Model," J Kor Soc Med Informatics, Vol.15, pp.117-131, 2009.

- [2] Korean Heart Rhythm Society : <http://www.k-hrs.org/>
- [3] M. G. Tsipouras, D. I. Fotiadis and D. Sideris, "An arrhythmia Classification system based on the RR-interval signal," Artificial Intelligence in Medicine, Vol.33, pp.237-250, 2005.
- [4] M. H. Song, J. Lee, S. F. Cho and K. J. Lee, "SVM Classifier for the Detection of Ventricular Fibrillation," The Institute of Electronics of Korea, Vol.42, pp.261-268, 2005.
- [5] PhysioBank : <http://physionet.mit.edu/physiobank/database/mitdb/>
- [6] NI Biomedical Startup Kit 3.0 : <http://decibel.ni.com/content/docs/DOC-12646>
- [7] G. D. Clifford, F. Azuaje and P. E. McSharry, "Advanced Methods and Tools for ECG Data Analysis," Artech House, pp.101-102, 2006.
- [8] All About Circuits : http://www.allaboutcircuits.com/vol_2/chpt_8/4.html
- [9] L. Breiman, 'Machine Learning', Kluwer Academic Publishers in Netherlands, 2001.
- [10] P. N. Tan, M. Steinbach and V. Kumar, 'Introduction to Data Mining', 1-st Addison-Wesley, 2006.
- [11] H. Shi, "Best-first Decision Tree Learning," The University of Waikato in NewZealand, pp.3-5, 2007.
- [12] K. Tateno and L. Glass, "A Method for Detection of Atrial Fibrillation Using RR Intervals," Computers in Cardiology(IEEE), Vol.27, pp.391-394, 2000.



이 현 주

e-mail : nedkelly@gce.sejong.ac.kr
 2011년 세종대학교 컴퓨터공학과(석사)
 2011년 9월~현재 세종대학교 컴퓨터공학과 박사과정
 관심분야: HCI, ECG, BCI, Mobile Programing, Bio-information etc.



신 동 규

e-mail : shindk@sejong.ac.kr
 1986년 서울대학교 계산통계학과(학사)
 1992년 M.S. in Computer Science, Illinois Institute of Technology
 1997년 Ph.D in Computer Science, Texas A&M University

1986년 2월~1991년 8월 한국국방연구원, 연구원
 1997년 8월~1998년 2월 현대전자 멀티미디어 연구소, 차장(책임연구원)
 1998년 3월~현재 세종대학교 컴퓨터공학과 교수
 관심분야: XML, 보안, 전자상거래, MPEG etc.



박 희 원

e-mail : heewonpark@samsung.com
 1982년 서강대학교 전자공학과(학사)
 2000년 독일 Ruhr 전자공학과(박사)
 관심분야: 멀티미디어, Broadcasting, 전자회로 etc.



김 수 한

e-mail : ksoohan@samsung.com
 1992년 전남대학교 신소재공학과(학사)
 1993년~현재 삼성전자주식회사 영상디스플레이 사업부 재직
 관심분야: N-Screen, 멀티미디어, Cloud 서비스 etc.



신 동 일

e-mail : dshin@sejong.ac.kr
 1988년 연세대학교 전산학과(이학사)
 1993년 Computer Science, Washington State University(M.S.)
 1997년 Computer Science, University of North Texas(Ph.D)

1997년 9월~1998년 2년 시스템공학연구소 선임연구원
 1998년 3년~현재 세종대학교 컴퓨터공학과 교수
 관심분야: HCI, 무선인터넷, 게임엔진, CSCW etc.