

# ECG 신호에서 단위패턴간 유사도분석을 이용한 부정맥 분류 알고리즘

배정현\*, 임승주\*, 김정주\*, 박성대\*, 김정도\*

\*호서대학교 전자공학과

e-mail:gowjdgus103@naver.com

## The classification of arrhythmia using similarity analysis between unit patterns at ECG signal

Junghyoun Bae\*, Seung-Ju Lim\*, Jeong-Ju Kim\*,

Sung-Dae Park\*, Jeong-Do Kim\*

\*Dept. of Electronic Engineering, Hoseo University

### 요 약

본 논문에서는 조기 심실 수축과 조기 심방 수축을 검출함에 있어 정밀한 QRS 구간의 폭, 정확한 P파와 T파의 크기 및 위치를 크게 요구하지 않고, 데이터의 가공과 복잡한 알고리즘의 사용에 의해 발생하는 ECG 데이터의 변형과 손실을 최소화할 수 있으며, 또한 개인차 때문에 발생할 수 있는 오류를 최소화하기 위한 알고리즘을 제안한다. 이를 위해 ECG 신호를 각각의 단위 파형으로 분리한 후, 정상 R-R 간격을 가지는 파형을 기준으로 기준파형을 만들어, 각 파형과 기준파형사이의 패턴 대조 및 유사도 분석을 통해 조기 심실수축과 조기심방수축을 검출할 수 있도록 하였다.

### 1. 서론

ECG 신호를 통해 부정맥을 검출하고 분류하기 위한 다양한 연구가 성공적으로 진행된 바 있다. Erik Zellmer는 QRS 데이터 셋과 T 데이터 셋을 사용하여 여러 가지의 특성을 만든 후 비선형 SVM으로 LBBB, RBBB, PAC, PVC 등을 분류하였고[1], Liang-Yu Shyu은 웨이블릿 변환을 이용하여 QRS 구간의 길이 외 여러가지 특징을 추출하여 Fuzzy Neural network를 사용하여 NB, PVC, LBBB, RBBB, APC 등을 분류하였다[2]. 그러나 이러한 방법을 비롯한 대부분의 방법들은 특징 추출을 위해 P-QRS-T 지점의 정확한 측정과 검출을 필요로 하기 때문에 고급 장비의 사용, 안정된 측정환경 등을 필요로 할 뿐만 아니라 데이터의 가공과 복잡한 알고리즘을 사용해야 한다는 문제점을 가지고 있다. 이 문제보다 더 큰 문제는, 정상 QRS 구간의 폭, P파와 T파의 크기 및 위치 등이 개인차를 가지고 있다는 것이다. 심한 경우 P파나 T파가 거의 구분할 수 없는 경우도 존재한다. 이러한 문제 때문에 심각한 부정맥을 측정하기 위해서는 안정된 환경에서의 측정, 장시간의 측정 및 고급장비 등이 필수적이다. 이러한 조건은 유비쿼터스 환경에서의 측정이 점점 요구되는 현대적 의료 경향에 상반되는 것들이다.

본 논문에서는 조기 심실 수축(PVC)과 조기 심방 수축(APC)을 검출함에 있어 정밀한 QRS 구간의 폭, 정확한 P파와 T파의 크기 및 위치를 크게 요구하지 않고, 데이터의 가공과 복잡한 알고리즘의 사용에 의해 발생하는 ECG 데이터의 변형과 손실을 최소화할 수 있으며, 또한 개인차

때문에 발생할 수 있는 오류를 최소화하기 위한 알고리즘을 제안한다. 이를 위해 ECG 신호를 각각의 단위 파형으로 분리한 후, 정상 R-R 간격을 가지는 파형을 기준으로 기준파형을 만들어, 각 파형과 기준파형사이의 패턴 대조 및 유사도 분석을 통해 조기 심실수축과 조기심방수축을 검출할 수 있도록 하였다. ECG 신호의 형태 비교 시 각 개인의 기준 ECG 파형을 사용하게 되는데, 이는 개인차 때문에 생기는 일반화의 어려움을 해결하기 위함이다. 제안된 방법의 유용성 입증을 위하여 MIT-BIH 데이터 베이스 중 13개의 레코드를 선택하여 실험이 이루어 졌으며, 성공적인 실험결과를 얻었다.

### 2. ECG 신호의 특성과 R-peak의 검출

조기 심실 수축이 발생하면 급격하게 R-R 간격이 감소하고 ECG 신호가 정상 신호에 비해 일그러진 형태를 보이며 QRS구간 구간의 길이가 120ms 이상으로 증가한다 [3][4]. 조기 심방 수축 역시 발생 시 급격하게 R-R 간격이 감소 하지만 ECG신호의 형태와 QRS구간 구간의 길이가 정상신호와 유사하다. R-R 간격이 급격히 감소하는 지점의 ECG 신호를 검출하고 검출된 ECG 신호와 정상 ECG 신호의 형태를 비교한다면 조기 심실 수축과 조기 심방 수축의 분류는 가능하다. 이를 위해서는 먼저 R-peak의 검출이 필요하고 R-R 간격의 변이를 나타내는 HRV(Heart Rate Variability)가 필요하다.

R-peak 검출 방법으로는, 널리 사용되고 있는 Pan & Tompkins 알고리즘[5]을 필터링 부분만 변형하여 사용하

였다. 그림 1은 전처리 과정의 블록 다이어그램이다.

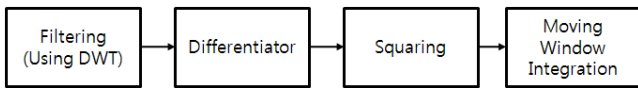


그림 1 R-peak 검출을 위한 전처리 과정

높은 R-peak 검출율을 위해서는 먼저 노이즈를 제거하는 작업이 수행되어야 한다. 노이즈 제거를 위한 방법으로 이산 웨이브렛 변환(Discrete Wavelet Transform : DWT)을 사용하였다[6][7]. 이산 웨이브렛 변환은 노이즈를 효과적으로 제거함과 동시에 원신호의 왜곡을 최소화할 수 있어 전처리 과정에 효과적으로 사용될 수 있는 최적의 도구이다. 본 실험에서는 실험 데이터로 MIT-BIH를 사용하였다. 그림 2는 MIT-BIH 123번 레코드의 원신호(a)와 필터링된 신호(b)를 비교한 것이다. (a)의 기저선 노이즈가 (b)에서는 제거된 것을 볼 수 있다. 필터링이 완료된 신호는 미분, 제곱, 이동평균을 차례로 수행하였다.

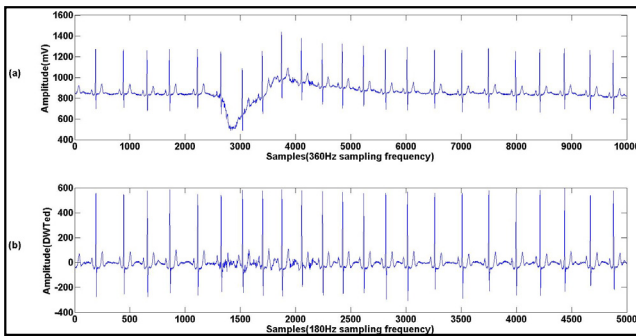


그림 2 MIT-BIH 123번의 원신호와 필터링된 신호

효율적인 검증을 위해 R-peak 검출율이 99.5% 이상인 것 중에서 조기 심방 수축만을 가지는 레코드, 조기 심실 수축만을 가지는 레코드, 두 가지 모두 가지는 레코드, 두 가지 모두 가지지 않는 레코드를 각각 선정하였다.

표 1 MIT-BIH 레코드의 R-peak 검출 결과

레코드	총 비트	Pan & Tompkins	제안된 방법	검출율(%)
100	2273	2273	2273	100
102	2187	2187	2187	100
117	1535	1533	1534	99.93
119	1987	1986	1982	99.75
122	2476	2474	2476	100
123	1518	1518	1518	100
124	1619	1619	1619	100
212	2748	2748	2747	99.96
220	2048	2048	2048	100
221	2427	2427	2426	99.96
230	2256	2255	2255	99.96
231	1573	1573	1570	99.81
233	3079	3078	3065	99.55
total	27726	27719	27700	99.92

표1은 13개 레코드에 대한 R-peak 검출 실험결과와 Pan & Tompkins의 QRS 검출 결과의 비교이다. 제안된 방법이 기존의 Pan & Tompkins의 방법과 거의 비슷한 검출

률을 보이고 있으나, 3장의 단위 패턴 유사도 분석을 위해서 제안된 필터링 방법이 필수적으로 사용되어야 한다.

### 3. 단위 패턴 유사도 분석

R-R의 시간 간격은 ECG신호의 이상 유무를 판단하는 유용한 정보가 된다[8]. 특히, 조기 심실 수축과 조기 심방 수축 두 가지 부정맥은 R-R 간격이 급격히 짧아졌다가 길어지는 공통점이 있다. 차이점으로는 조기 심실 수축의 경우는 ECG 파형 중 QRS구간이 많이 일그러짐에 비해 조기 심방 수축은 정상 파형과 QRS구간이 거의 유사하다. 본 논문에서는 HRV에 문제가 있는 지점을 먼저 검출하고, 문제가 있는 R-R 간격 근처의 정상파형 20개를 단위 패턴으로 분리 한 후, 이 20개의 정상 단위패턴을 이용하여 기준패턴을 만든다. 이 기준패턴과 문제가 있는 이상 단위패턴의 형태를 비교하여 기준패턴과 이상단위패턴이 유사한 파형이면 조기 심방 수축, 그렇지 않으면 조기 심실 수축으로 분류하였다. 그림 3은 제안된 단위 패턴 유사도분석법의 구성도이다.

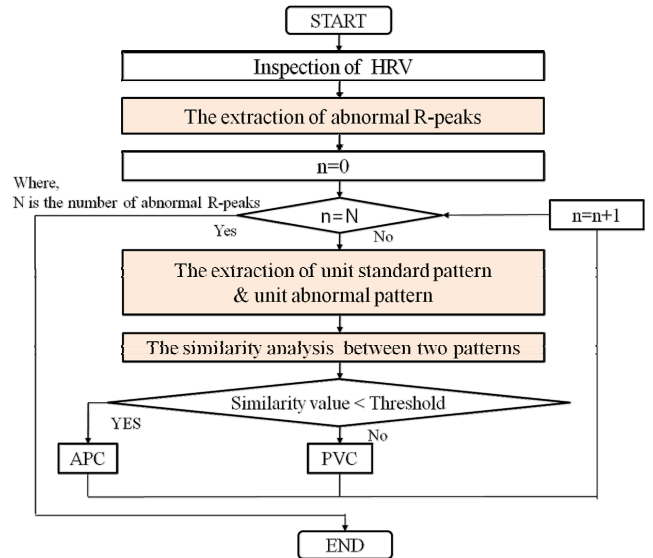


그림 3 부정맥 분류를 위한 단위패턴 유사도 분석 알고리즘

본 논문에서 제안된 단위패턴 유사도 분석법은 크게 두 단계로 나누어진다. 첫 번째, HRV를 조사하여 이상 R-peak를 추출한다. 이를 위해, HRV의 흐름 중 심박 수가 급격히 증가하는 지점을 조기 심실 수축이나 조기 심방 수축 의심 지점으로 검출하였다. 아래의 두 조건을 모두 만족하는 HRV가 부정맥의 가능성이 있는 지점이다.

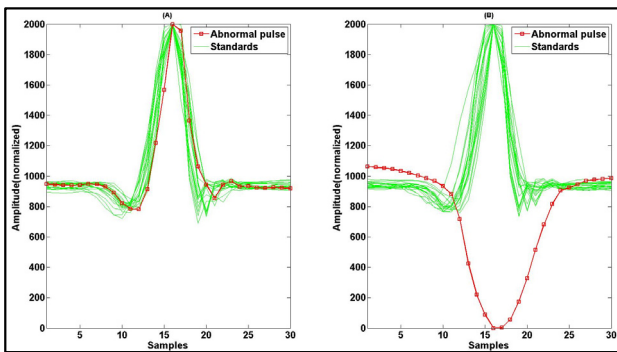
- 조건1) 전체 HRV의 평균의 120%가 넘는 HRV 지점
- 조건2) 조건1)을 만족하지 않는 이전 HRV 중 가장 가까운 지점의 HRV에 대해 120%를 넘는 HRV 지점

두 번째, 단위 기준패턴과 단위 이상 패턴을 추출한 후 유사도 분석을 통해 조기심실수축과 조기심방수축을 분리하는 과정이다. R-peak 지점을 중심으로 전·후 각각 15개, R-peak 지점 포함 총 31개의 샘플링된 값을 단위패턴으로 선택하였다. 360Hz로 샘플링 주파수는 이산 웨이브

렛 변환 과정을 통하여 180Hz가 되고, 31개의 샘플은 약 172ms동안의 신호를 나타낸다. 이는 일반적인 QRS구간의 폭인 60~120ms를 충분히 포함하도록 하기 위함이다[9].

단위 패턴은 단위기준패턴과 단위이상패턴으로 구분되는데, 단위기준패턴은 정상적인 단 패턴 20개를 이용하여 만들어 진다. 문제가 있는 HRV 지점이라고 판단된 지점 이후에 나타난 정상 단위패턴 20개를 이용하여 단위 기준패턴을 만들게 된다. 두 번째 과정을 정리하면 다음과 같다.

1. 전체 R-peak  $R_p^1, R_p^2, \dots, R_p^T$  중에서 HRV의 조사를 통해 이상 R-peak  $R_{ab}^1, R_{ab}^2, \dots, R_{ab}^N$ 을 구한다.  
여기서,  $T$ 는 총 R-peak의 수,  $N$ 은 이상 R-peak의 수
2.  $n = 1$ 로 초기화한다.
3. 만약  $n = N$ 이면 12번으로 이동한다.
4.  $i = 1, m = 1$ 로 놓는다.
5.  $n$ 번째 이상 R-peak  $R_{ab}^n$  부근의 31개 샘플을 이용하여  $n$ 번 단위 이상패턴  $S_{ab}^n$ 을 만든다.
6.  $R_{ab}^n$ 에 해당하는 R-peak를  $R_p^j$ 라 가정하고,  
만약  $R_p^{j+i}$ 가 정상 R-peak이면, 31개의 샘플로 구성된 단위 패턴  $S_m^n$ 을 만들고  $m = m + 1$ 로 놓는다.  
만약  $R_p^{j+i}$ 가 이상 R-peak이면, 7번으로 이동한다
7.  $i = i + 1$ 로 놓는다.
8. 만약  $m \leq 20$ 이면, 6번으로 이동  
그렇지 않으면, 9번으로 이동
9. 저장된  $s_1^n, s_2^n, \dots, s_{20}^n$ 을 이용하여, 평균 패턴  $s_{av}^n$ 을 구한 후,  $s_{av}^n$ 과  $s_1^n, s_2^n, \dots, s_{20}^n$  사이의 유사도 분석을 행한다. 유사도 분석은 cosine similarity 방법을 이용한다. 만약 유사도가 정해진 기준안에 포함되지 않는 패턴이 존재하면 이 패턴을 제거한 후, 새로운 평균을 구한후, 이를  $n$ 번째 단위 기준 패턴  $s_{std}^n$ 으로 정한다.
10.  $S_{ab}^n$ 와  $s_{std}^n$  사이의 유사도 분석을 통해, 기준값 미만이면 APC, 이상이면 PVC로 결정한다.
11.  $n = n + 1$ 로 놓은 후에, 3번으로 이동한다.
12. 분석을 마친다.



(a) APC형 이상단위패턴과 (b) PVC형 이상단위패턴과 20개의 정상단위패턴  
그림 4 MIT-BIH 100에서 검출된 APC와 PVC

유사도 분석시의 오류를 방지하기 위해, R-peak는의 크기가 동일한 크기를 가질 수 있도록 일반화시켰다.

또한, 전체 ECG 측정신호의 뒤쪽에서는  $s_1^n, s_2^n, \dots, s_{20}^n$ 를 구할 수 없는 경우가 존재한다. 만약 20번  $s_{ab}^n$ 까지 구할 수 없다면,  $s_{std}^{n-1}$ 의 단위기준패턴을 사용하도록 하였다.

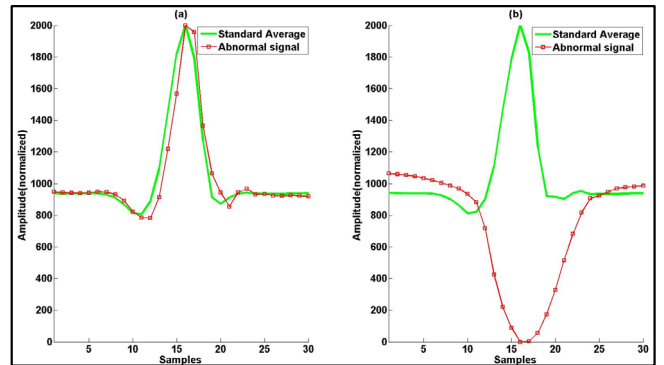
그림 4는 MIT-BIH 100번 레코드의 조기 심방 수축과 조기 심실 수축의 비교 결과를 보여준다. 실선으로 그려진 파형은 20개의 정상단위패턴이고, 사각형선으로 그려진 파형은 문제가 발생된 지점의 이상단위패턴이다. (a)의 이상단위패턴은 조기 심방 수축으로 분류할 수 있고, (b)의 이상단위패턴은 조기 심실 수축으로 분류할 수 있다.

코사인 유사도를 사용하여 HRV 문제 지점과 기준파형의 최종 평균의 유사도를 통하여 조기 심실 수축과 조기 심방 수축으로 분류하였다. 대표적인 유사도 구하는 방법은 K-means와 Fuzzy C-means등이 있다. 본 연구에서 코사인 유사도를 사용한 이유는 시간지연을 반영하여 유사도를 구하는 최적의 방법이기 때문이다. 코사인 유사도는 식 (1)과 같다[10]. 여기서 A와 B는 유사도를 구하기 위하여 서로 비교하게 될 31개 샘플로 이루어진 단위패턴의 벡터이다.

$$\cos(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad \text{식 (1)}$$

문제가 있는 지점의 기준파형들은 문제지점과의 비교를 위해 파형의 평균을 구하여 1개의 파형으로 재구성된다. 이때, 문제가 있는 지점이지만 HRV 검사에서 검출되지 않은 샘플이 기준파형 중 섞여 있을 수 있다. 이를 회피하기 위하여 20개 정상단위패턴의 평균과 20개 기준파형 각각의 유사도를 구하여 결과 값이 기준값 이상인 기준파형을 제외하고 나머지 기준파형의 최종 평균을 구성하였다.

유사도의 결과 값이 0.14 이하면 조기 심방 수축, 이상이면 조기 심실 수축으로 분류하였다. 그림 5의 (a)와 (b)는 MIT-BIH 100번의 문제 지점 중 26, 27번째 지점의 기준파형의 최종 평균과 문제 지점의 비교를 각각 보여준다.



(a) APC형 이상단위패턴과 (b) PVC형 이상단위패턴과 단위 기준패턴  
그림 5 MIT-BIH 100번의 26, 27번째 문제 지점의 파형

4. 실험결과

먼저, HRV만을 이용한 조기 심실수축이나 조기 심방수축 의심지점은 총 1597개 지점이 검출되었으며 FP(false positive)가 17개, FN(false negative)이 236개 발생했다. 4개의 FP는 231번 레코드에서 발생하였는데, 이레코드의 경우 비정상적인 형태의 HRV를 갖고 있었다. 이러한 경우 FP가 발생할 확률이 매우 높다. 221번 레코드에서 발생한 12개의 FP는 정상 신호이지만 애매한 HRV를 갖는 신호를 문제 지점으로 오 검출 하였다. 표 2는 조기 심실수축과 조기 심방 수축 분류 실험에 대한 결과이다.

표 2 MIT-BIH 와 실험 결과의 비교 (SEN(%))

record	MIT-BIH		Experiment			
	APC beat	PVC beat	APC		PVC	
			beat	SEN	beat	SEN
100	33	1	32	97.06	1	100
102	0	4	0	-	4	100
117	1	0	1	100	0	-
119	0	444	0	-	442	99.55
122	0	0	0	-	0	-
123	0	3	0	-	3	100
124	2	47	1	66.67	43	92.16
212	0	0	0	-	0	-
220	94	0	72	81.03	1	-
221	0	396	12	-	376	95.19
230	0	1	0	-	1	100
231	1	2	5	100	2	100
233	7	831	8	100	645	81.71
Total	138	1729	131	85.19	1518	76.8

신뢰성 평가는 Sensitivity(SEN)에 의해 이루어졌다. Sensitivity는 실제 부정맥 비트 검출에 관한 알고리즘의 능력을 나타낸다. 식(2)에서 TP는 true pulse이다.

$$SEN = \frac{TP}{TP + FN} \tag{2}$$

제안된 알고리즘을 통한 분류 실험결과, 10개 레코드에서 조기 심실 수축과 조기 심방 수축이 99% 이상 분류되었다. 신뢰성 평가에서 231, 233번 레코드는 많은 오차를 보였다. 231번 레코드의 HRV는 다른 레코드들과는 다르게 정상 R-R 간격이 빠른 구간과 느린 구간이 반복적으로 나타났다. ECG 신호는 정상임에도 불구하고 느린 구간에서 빠른 구간으로 바뀌는 시점을 HRV 문제구간으로 잘못 검출하였다. 233번 레코드의 경우, 조기 심실 수축임에도 불구하고, 본 논문에서 정한 HRV 문제 지점 검사에서 R-R 간격에 문제가 없다고 판단한 경우가 약 180회 존재하여 정확한 검출이 이루어지지 못했다. 231번과 233번 레코드의 경우, 일반적인 PVC와 APC를 가지지 않는 특별한 경우에 해당한다. 이러한 레코드의 경우, R-R 간격, QRS Complex, P파 및 T파의 정확한 측정이 이루어질 경우에만 정확한 측정이 이루어질 수 있다. 본 연구에서는 복잡한 특성 여러 가지를 사용하지 않는다는 것을 연구 목적으로 하고 있기 때문에, 이를 감안한다면 두 레코드의

결과는 충분히 만족할 만한 것으로 생각된다.

5. 결론

본 논문에서는 복잡한 ECG의 특성 여러 가지를 복합적으로 사용하지 않으면서, 개개인간에 존재하는 ECG 신호 특성 차이에 의한 오류를 최소화 할 수 있는 새로운 형태의 알고리즘을 제안하였다. 실험 수행 결과 대부분의 조기 심실 수축과 조기 심방 수축을 검출할 수 있었다. 실제로, ECG는 몇 가지 특징만을 가지고 증상을 판별할 수 없다. P-QRS-T과 모두 개인별로 다른 특성을 지니기 때문에 문제의 ECG 전후의 신호도 중요한 정보가 될 수 있다. ECG에 문제가 발생한 시점 주변의 ECG 경향을 표준 ECG로 사용하는 것은 부정맥 뿐 아니라 다른 여러 가지 ECG 분석에도 활용될 수 있는 방법이다.

참고문헌

[1] Erik Zellmer, Fei Shang, Hao Zhang "Highly Accurate ECG Beat Classification based on Continuous Wavelet Transformation and Multiple Support Vector Machine Classifiers", Biomedical Engineering and Informatics, pp.1-5, 2009

[2] L. Y. Shyu, Y. H. Wu and W. C. Hu "Using wavelet transform and fuzzy neural network for VPC detection from the holter ECG", IEEE Trans. Biomed. Eng., vol. 51, p.1269, 2004.

[3] Karen M. Ellis "EKG Plain and Sample second Edition", p.16~17, p.29~30.

[4] 조익성, 권혁승, "PVC 분류를 위한 적응형 문턱치와 윈도우 기반의 R파 검출 알고리즘", 한국통신학회논문지, vol. 34, No. 11, pp.1289-1295, 2009.

[5] J. Pan and W. Tompkins, "A real time QRS detection algorithm", IEEE Trans. Biomed. Eng., vol. BME-32, no. 3, pp.230 - 236, 1985.

[6] S. Kadambe, R. Murray, and G. F. B. Bartels "Wavelet Transform based QRS complex detector", IEEE Trans Biomed Eng. 46(7):838-848, 1999.

[7] 박광리, 이경중, "웨이브렛 변환을 이용한 QRS 검출 알고리즘 개발", 대한전자공학회 강원지부 추계 학술대회 논문집 제2권 제2호 pp.19-22, 1996.

[8] 윤장혁, 정용주, "실시간 ECG 모니터링 시스템 개발을 위한 QRS 검출 알고리즘의 개발" 한국정보기술학회논문지 제8권 제5호, pp.153-163, 2010.

[9] S. Evans, H. Hastings, and M. Bodenheimer, "Differentiation of beats of ventricular and sinus origin using a self-training neural network", PACE, vol. 17, pp.611 - 626, 1994.

[10] Gerard Salton, Michael J.McGill, "Introduction to modern information retrieval", McGraw-Hill, p.120-126, 1983.