

퍼지 클러스터링을 이용한 심전도 신호의 라벨링에 관한 연구

공인욱, 이정환, 이상학, 최석준, 이명호
연세대학교 공과대학 전기공학과

A Study on Labeling of ECG Signal using Fuzzy Clustering

I. W. Kong, J. W. Lee, S. H. Lee, S. J. Choi, and M. H. Lee

Department of Electrical Engineering, College of Engineering, Yonsei University

ABSTRACT

This paper describes ECG signal labeling based on Fuzzy clustering, which is necessary at automated ECG diagnosis. The NPPA(Non parametric partitioning algorithm) compares the correlations of wave forms, which tends to recognize the same wave forms as different when the wave forms have a little morphological variation. We propose to apply Fuzzy clustering to ECG QRS Complex labeling, which prevents the errors to mistake by using If-then comparison.

The process is divided into two parts. The first part is a parameters extraction process from ECG signal, which is composed of filtering, QRS detection by mapping to a phase space by time delay coordinates and generation of characteristic vectors. The second is fuzzy clustering by FCM(Fuzzy c-means), which is composed of a clustering, an assessment of cluster validity and labeling.

1. 서론

심전도 자동진단 시스템은 심전도 파형을 수집하는 하드웨어, 필터링, QRS 컴플렉스의 검출, 군(group) 단위로 라벨링, 대표비트를 이용한 진단 파라미터 추출, 진단 알고리즘의 적용으로 나눌 수 있다. 이들 과정에서 군단위로 라벨링하는 과정은 파라미터의 추출을 위한 대표 비트(beat)의 선택, 신뢰성 있는 리듬 분석(부정맥 진단), 24시간 홀터 모니터링 시스템에서의 비정상 비트 검출에 필수적으로 필요하다. 기존의 라벨링 방법이 IF-THEN에 의한 이진 논리를 이용한 방법이었던 때문에 같은 파형이지만 약간의 형태변화가 있는 경우, 일부는 같은 군으로, 일부는 다른 군으로 분류되는 결과가 나온다. 이와 같은 문제점을 해결하기 위하여 본 연구에서는 퍼지 클러스터링(Fuzzy clustering)을 이용하여 이러한 이진 구조의 문제점을 해결하고자 한다.

원 심전도 신호에서 Integer filter를 사용하여 근전도 등의 고주파 잡음 성분을 제거한다. 파라미터 추출을 위한 기준점으로 R파를 이용한다. R파를 검출

하기 위해서, 심전도 신호를 2차원 시간 지연 좌표계를 사용하여 재구성한 뒤, 위상궤적이 최대의 변위량을 가지는 지점으로부터 설정된 구간내에서 R파를 검색한다.

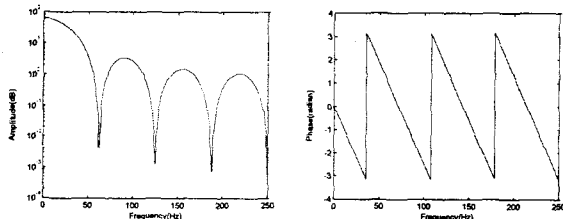
구해진 심전도 비트로부터 파라미터를 추출하여 각각의 비트에 대응되는 특성벡터를 생성한다. 생성된 특성벡터는 비슷한 형태를 가진 비트끼리는 위상 공간에서 동일한 군을 형성하게 되고, 비트의 형태가 다르다면 다른 군을 형성한다. 퍼지 클러스터링(Fuzzy c-means)을 이용하여 각 비트의 각 군에 대한 M.F.(membership function)와 군의 중심, 개수를 알 수 있게 된다. 위상공간 내에 존재하는 군의 종류를 알기 위해서 클러스터의 개수가 실제 특성벡터에 얼마나 적합한지를 나타내는 적합도(cluster validity)를 사용한다.

본 연구에서는 제안한 방법의 적합성을 검증하기 위해 CSE(Common standard for quantitative electrocardiography)데이터 베이스를 이용하였고, 이 데이터 베이스의 심전도 파형을 이용하여 모든 알고리즘에 필요한 상수를 조정하였다.

2. 특성벡터의 생성

근잡음등의 고주파 노이즈를 제거하기 위해서, 0° , $\pm 60^\circ$, $\pm 90^\circ$, $\pm 120^\circ$, 180° 에 위치한 극점을 이용하여 모든 계수가 정수이고, 극점과 영점을 상쇄시켜 연산부담을 줄인 Integer 필터를 이용한다.

$$H(z) = \left(\frac{1 - z^{-8}}{1 - z^{-1}} \right)^2 \quad (1)$$

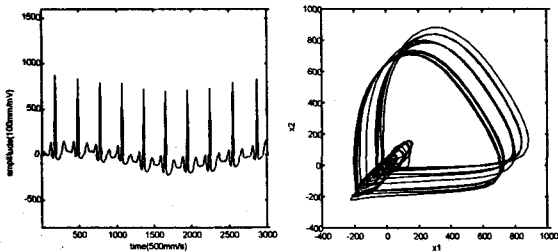


(a) 진폭응답 (b) 위상응답
그림 1 저역통과 필터의 진폭응답과 위상응답
Fig. 1 Frequency response of the low pass filter

위상공간상의 재구성 방법은 F, Takens에 의해 제안된 상태변수를 지연시간으로 하는 시간 지연 좌표계를 사용한다. 스칼라 시계열 데이터 $\xi(t_k)$ 로 부터 위상공간에 재구성할 경우 차원이 d_E 이고, 지연시간이 τ 라면 d_E 차원 공간에서의 상태 벡터는 식(2)와 같이 주어진다.

$$x(t_s) = \begin{bmatrix} \xi(t_s) \\ \xi(t_s + \tau) \\ \vdots \\ \xi(t_s + \tau \cdot (d_E - 1)) \end{bmatrix} \quad (2)$$

심전도 신호에 가장 적합한 차원과 지연시간을 위상공간 이용율을 산출하는 Fill-factor 알고리즘으로 구해보면, Fill-factor는 차원 d_E 에 의해서는 거의 변화를 보이지 않지만, 지연시간 τ 에 의해서는 많은 변화를 보인다. 그래서, Fill-factor에 많은 변화를 주지 않는 차원 d_E 는 연산부담을 줄이기 위해 2차로 결정하고, 지연시간 τ 는 최적의 공간이용도를 가지는 것으로 계산된 20ms를 사용한다. 이렇게 재구성된 위상공간상의 궤적은 그림 2와 같고, 기저선 동요로 인해 궤적의 위치는 변하더라도, 궤적의 모양은 변하지 않음을 확인 할 수 있다.



(a) 심전도 신호 (b) 위상공간상의 궤적
그림 2 심전도 신호와 위상공간상의 궤적
Fig. 2 ECG & it's portrait on the phase space

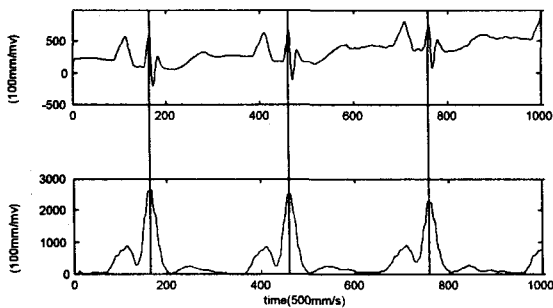


그림 3 심전도 신호와 위상공간의 궤적 변위량
Fig. 3 ECG & the displacement of portrait on the phase space

신호과형중 피크는 위상공간상에서 하나의 루프를 형성하는데, R파의 극대점에서 궤적변위량이 가장 크게 된다. 그림 3은 심전도 신호와 궤적변위량의 관계를 나타낸 것인데, 궤적 변위량이 R파에서 가장 큰 값을 가짐을 확인할 수 있다. 식(3)은 본 연구에서 제

안한 궤적변위량이다.

$$dvec(i) = \sum_{i=\tau-10}^{i+\tau} |(x(i), x(i+\tau)) - (x(i+1), x(i+1+\tau))| \quad (3)$$

궤적 변위량은 R파 부근에서 가장 큰 값을 가진다는 특성을 이용하여, 그 값이 문턱치 이상인 곳을 R파(또는 QS형태)로 검색한다. 문턱치는 후방검색(back search)기법을 이용하여, 예상 RR간격내에 설정된 문턱치 이상의 궤적 변위량이 존재하지 않을 경우 문턱치를 낮추어 다시 검색하도록 해서, 궤적변위량이 갑자기 작아져 첫번째 검색시 실패하는 경우에 대비한다. 이 경우 문턱치의 초기값, 문턱치의 최소치, 문턱치의 변화량, 최소 RR간격, 예상되는 RR간격에 관한 상수를 정의해야 한다.

기저선 동요에 영향을 받지 않는 시간지연좌표계상의 궤적변위량으로부터 식(4)와 식(5)를 사용하여 2차원 특성벡터 (SF1, SF2)를 구한다. q=심전도 비트의 번호이고, r(q)는 q에 해당하는 R파의 위치이다.

$$sf_1(q) = \frac{\sum_{i=r(q)-100}^{r(q)} dvec(i)}{\sum_{i=r(q)-100}^{r(q)+100} dvec(i)} \quad (4)$$

$$sf_2(q) = \frac{\sum_{i=r(q)-100}^{r(q)+100} dvec(i)}{\max_j \sum_{i=r(q)-100}^{r(q)+100} dvec(i)} \quad (5)$$

3. 심전도 신호의 라벨링

Fuzzy c-Means 알고리즘은 각 특성벡터에 대한 각 클러스터(cluster)의 소속도로서 M.F.인 $u_i(x)$ 를 클러스터링에 이용한다. $X=(x_1, \dots, x_n)$ 으로 n개의 R^d 차원의 벡터들의 있을때, X를 c개의 클러스터로 나누게 되면 S_1, \dots, S_c 의 부분집합으로 나누어지고, X를 c개의 클러스터로 나누게 되면 $u_i(x)$ 는 M.F.가 될 수 있도록 식(6)을 만족시켜야 한다.

$$\sum_i u_i(x) = 1 \quad i = 1, \dots, c, \quad x \in X \quad (6)$$

FCM(Fuzzy c-means)은 식(7)의 평가함수(objective function)를 최소화 시키도록 u_{ik} 와 클러스터의 중심 v_i 를 결정한다. u_{ik} 는 x_k 가 i번째 클러스터에의 M.F.를 나타낸다.

$$U = \sum_i \sum_k (u_{ik})^m |x_k - v_i|^2 \quad (7)$$

여기서, $m > 1$; m은 실수

식(7)의 평가함수는 각각의 데이터(x_k)의 유사성을 벡터들 사이의 거리로써 측정한다고 가정한다. v_1, \dots, v_c 는 클러스터 중심(cluster center)을 나타내는 벡터이다. 위상공간상의 임의의 위치로 초기치를 가지지만, FCM 알고리즘에 의해 U가 최소값을 가지게 되면, v_i 는 군의 중심에 해당하는 값을 가진다. 식(7)의 평가함수는 v_i 에 가까운 x_k 의 소속도 u_{ik} 가 큰 값을

가지도록 하면 평가함수가 적은값을 나타낸다.

U의 m(exponent weight)이 크면 클러스터 중심에서 멀리 떨어진 데이터가 U에 적게 관여하게 하므로 노이즈를 제거시키는 기능을 할 수 있다. 대개의 경우 이상 심전도 비트는 그 수가 적으므로 본 연구에서는 m=1.5로 설정하여 개체 수가 적은 클러스터도 평가되도록 한다.

평가함수 U가 일정 이하의 오차를 가질때까지 u_i 와 v_i 를 식(8),(9)에 의해 계산한다. 식(8)은 한번에 하나의 클러스터에 대해서 계산하여야만 올바르게 U가 수렴하게 된다.

$$v_i = \frac{\sum_k (u_{ik})^m x_k}{\sum_k (u_{ik})^m} \quad (8)$$

$$u_{ik} = (1/|x_k - v_i|^2)^{1/(m-1)} / \sum_j (1/|x_k - v_j|^2)^{1/(m-1)} \quad (9)$$

클러스터의 적합도(cluster validity)는 FCM에 의한 클러스터링 결과가 지금의 특성벡터들에 얼마나 적합한지를 평가할 경우에 사용된다. 이는 클러스터의 수를 측정할 경우 주로 사용되는데, 최적의 적합도를 가지는 파라미터 c가 바로 클러스터 수가 된다.

적합도를 평가하는 함수는 v_i 와 x_k 는 동일한 클러스터에 소속될때는 그 거리가 가까울 수록 좋은 적합도를 가져야 한다. 또한, 하나의 클러스터를 나눠서 v_i 와 x_k 가 가까워 지게 하는 오류를 방지해야 한다.

$$V_{XB} = \frac{\sum_{k=1}^n \sum_{i=1}^c u_{ik}^2 |x_k - v_i|^2}{n(\min_{i \neq j} |v_i - v_j|^2)} \quad (10)$$

식(10)의 분자항은 FCM에서의 평가함수 U와 유사하므로, 최적의 클러스터링이 되었다면 보다 작은 값을 가진다. 하지만, 하나의 클러스터를 둘로 나누어서 식(10)의 분자항이 적어져 V_{XB} 가 적어지는 오류를 막기 위해서 분모항이 필요하다. 만약 하나의 클러스터를 둘로 나누게 된다면 식(10)의 분모항은 작아지게 되어, 결국 V_{XB} 는 커지게 되어 필요 이상의 클러스터링을 제한한다.

또한, 식(10)은 적은 수의 데이터로 구성된 군에 대해서도 하나의 클러스터로써 평가되도록 식 내부에 특성벡터와 클러스터 중심 v_i 와의 거리에 관한 항이 포함되어 있다. 독립된 클러스터를 가지면서도 그 데이터의 개수가 적은 경우에도 전체 평가식 V_{XB} 에서 그 클러스터가 차지하는 비중이 거리에 관한 항때문에 커지게 된다. 이는 적은 수의 이상 심전도 파형도 하나의 군으로 평가 될 수 있다는 것을 의미한다.

라벨링을 위해서는 클러스터링된 특성벡터를 일반적인 개념의 클러스터링으로 변환해야 하는데, 식(11)에과 같은 역퍼지화(defuzzyfication)과정이 필요하다.

$$S_i = \{x \in X : u_i(x) \geq u_j(x); j=1, \dots, c\} \quad (11)$$

심전도 비트가 2종류 이상이 존재할 경우 어느 비트가 지배 형태인지 결정하여야 한다. 보통 지배형태

는 가장 많은 형태를 가지는 QRS컴플렉스형태가 되지만, QRS 심실성 이단현상(ventricular bigeminy)의 경우 비정상 비트가 정상 비트보다 많아질 수도 있으므로, 지배형태를 결정할때는 가장 많이 나타나는 형태로 결정해서는 않된다. 본 연구에서는 가장 많이 발생하는 형태가 두번째 많이 발생하는 형태보다 3배 이상 크지 않고, 그 비트 직후의 RR간격을 비교하여 가장 많이 발생하는 것이 150ms이상 길다면 두번째 많이 발생한 형태를 지배 형태로 결정한다.

4. 실험 및 결과 고찰

제안된 FCM과 V_{XB} 를 이용한 퍼지 클러스터링의 유용성을 평가하기 위해 5개 지점에 반경 0.1의 균을 난수를 이용하여 생성하되 각 클러스터내의 데이터를 40, 20, 10, 5, 3으로 불균일하게 한다.

표 1 인위적인 데이터에 대한 FCM을 적용한 결과
table 1 Results applying FCM to artificial data

ID	c	x_{1c}	x_{2c}	Fc-M: x_{1c}	Fc-M: x_{2c}
1	40	0.3	0.7	0.2980	0.6847
2	20	0.3	0.3	0.2986	0.3100
3	10	0.5	0.5	0.4777	0.4975
4	5	0.7	0.7	0.7147	0.6830
5	3	0.7	0.3	0.6984	0.2760

표 2 인위적인 데이터 집합에 대한 적합도 V_{XB}
table 2 Validity function V_{XB} of artificial data

클러스터 #	Fc-M 평가함수 U	적합도 V_{XB}
2	1.15629153	0.13560435
3	0.51221316	0.06786655
4	0.30870119	0.06013958
5	0.19728114	0.03812922
6	0.12960826	0.26851741

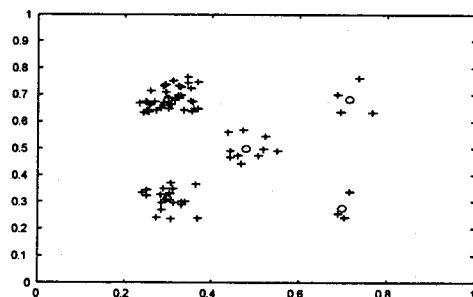


그림 4 인공 데이터에 F-cM 알고리즘의 적용
Fig. 4 FCM Algorithm on the artificial data

그림 4의 'o'표시의 군의 중심과 표 1을 보면 생성된 인위적인 데이터에 대해서 본 연구에서 제안한 방법에 의해 효과적으로 클러스터링 되었음을 알 수 있다.

표 2에서는 $c=5$ 인 경우 V_{XB} 가 최소값을 가지므로 클러스터의 수가 5임을 확인할 수 있다.

그림 5는 본 연구에서 제안된 퍼지 클러스터링을 이용한 라벨링 알고리즘을 구현한 프로그램이다.

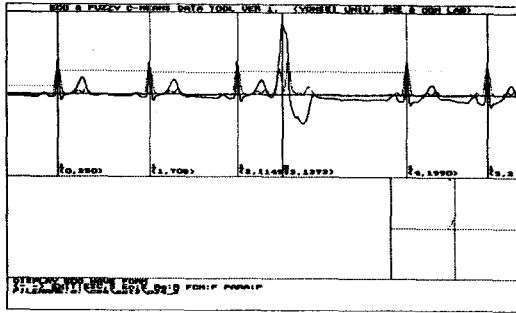


그림 5 제작된 심전도 라벨링 프로그램
Fig. 5 A constructed ECG labeling program

실제 심전도 신호를 가지고 FCM에 의해 클러스터링한 결과는 그림 6과 같고, 그 결과를 이용해서 라벨링한 결과는 그림 7과 같다. 그림 5은 전체 파형의 일부분을 표시한 것인데 지배형태인 5,8,9번이 지배형태임을 뜻하는 1번으로 라벨링되었고, 형태가 다른 두개의 심전도 신호는 각각 다르게 라벨링 되었음을 확인할 수 있다.

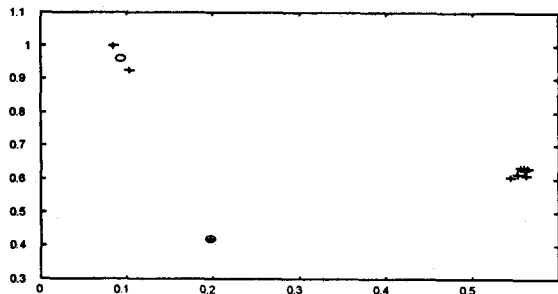


그림 6 CSE set3 #28의 F-cM 결과
Fig. 6 The FCM result of CSE set3 #28

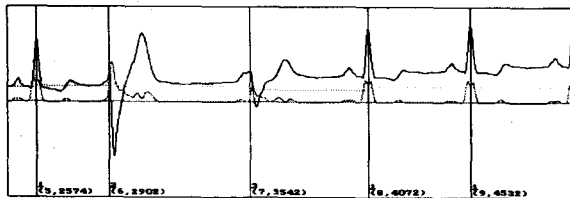


그림 7 CSE set3 #28의 라벨링 결과
Fig. 7 The labeling result of CSE SET3 #28

5. 결론

본 연구에서는 심전도 신호를 재구성하여 특성벡터를 구하고, FCM을 이용하여 심전도 신호를 라벨링할 것을 제안하고, 이를 통해 다음과 같은 결론을 얻었다.

- (1) 설계된 필터를 이용하여 시간지연좌표계상의 궤적자체의 모양을 변화시키는 고주파 잡음을 신호의 왜곡없이 효과적으로 제거할 수 있었다.
- (2) 제안된 특성벡터 생성 알고리즘은 그 기반을 2차원 시간지연좌표계상의 궤적변위량에 두고 있으므로, 기저선 동요등의 저주파 잡음의 영향을 무시할 수 있었다.
- (3) 특성 벡터들로 구성된 위상공간에서 F-cM 방법을 이용하여 클러스터링을 행할 수 있었다. 심전도 비트의 약간의 형태변화가 있을 경우 위상공간상에서도 약간의 변이를 보이게 된다. 퍼지 클러스터링에 의한 방법은 그 변이가 발생하더라도 동일한 군으로 인식하였다.
- (4) 본 연구에서 사용한 클러스터링 적합도 평가식 V_{XB} 는 그 개체수가 적지만 독립된 클러스터인 경우도 하나의 클러스터로 고려되어야 하는 심전도 신호의 클러스터링에 우수한 성능을 보였다.
- (5) 심실성 이단 현상시 지배형태를 RR간격을 이용하여 효과적으로 구분할 수 있었다.
- (6) 라벨링의 결과로 M.F.를 이용할 수 있으므로 퍼지추론을 이용한 진단에 효과적으로 사용될 수 있다.

참 고 문 헌

- [1] 정석현, 이정환, 이병채, 이명호, "시간 지연 2차원 좌표계를 이용한 실시간 QRS 검출에 관한 연구", 대한의용생체공학회 춘계학술대회 논문집, Vol. 17, No. 2, pp. 277-280, 1995
- [2] J.C.Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York:Plenum Press, 1981
- [3] X. L. Xie, G. A. Beni, "Validity measure for fuzzy clustering", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. 3, No. 8, pp. 841-846, 1991
- [4] Nikhil R. Pal, James C. Bezdek, "On Cluster Validity for the Fuzzy c-Means model", *IEEE Trans. on Fuzzy systems*, Vol. 3, No. 3, pp. 370-379, 1995
- [5] Willis J. Tompkins, *Biomedical Digital Signal Processing*, New Jersey: Prentice Hall, 1993
- [6] P. A. Lynn, "Online digital filters for biological signals: some fast designs for a small computer", *Med. & Biol. Eng. & Comput.*, Vol. 15, pp. 534-540, 1977
- [7] 장재호, "Fill-factor 알고리즘을 이용한 생체카오스의 어트랙터 재구성에 관한 연구", 연세대학교 석사 학위 논문, 1995.2
- [8] Michael P. Windham, "Cluster Validity for Fuzzy c-Means Clustering Algorithm", *IEEE Trans. Pattern Anal. Machine Intell.*, Vol. PAMI4, No. 4, pp. 357-363, 1982
- [9] Rui-Ping Li, Masao Mukaidono, "A Maximum-entropy approach to Fuzzy clustering", *Fourth IEEE International Conference on Fuzzy Systems*, Vol. 4, pp. 2227-2232, 1995