

心跳種類的分析與判斷：模糊 C-平均值演算法

葉雲奇

清雲科技大學 電子工程系

yunchi@cyu.edu.tw

林泓志

清雲科技大學 電子工程系

m9811001@cyu.edu.tw

摘要—本研究是以 Fuzzy C-Means 演算法為基礎，提出一個簡單有效的心跳種類的分析與判斷方法，稱為模糊 C 平均值分析法(Fuzzy C-Means Method, FCMM)。本研究能分析與判斷五種不同的心跳，包括正常的心跳(normal heartbeats) 與較常發生的四種不正常的心跳(包括“左束分支阻斷 (Left Bundle Branch Block)”、“右束分支阻斷 (Right Bundle Branch Block)”、“心室過早收縮 (Ventricular Premature Contractions)”、及“心房過早收縮 (Atrial Premature Contractions)”)。本研究共分成四個主要部份，分別是：(1) QRS extraction stage：尋找心電圖(ECG)信號中的 QRS 複合波、P 波及 T 波的位置；(2) qualitative features stage：選取主要特徵點；(3) Procedure-FCM：計算每一個群集的中心值向量；(4) Procedure-HCD：心跳種類的分析與判斷。最後再以 MIT-BIH 資料庫中的 ECG (electrocardiogram) 信號來測試本研究的效能。經實際的測試，平均正確判斷率可達 93.78%。

關鍵詞：ECG 信號；模糊 C 平均值演算法。

Abstract—This study proposes a simple and reliable method, termed the Fuzzy C-Means method (FCMM) for classifying the heartbeat cases from electrocardiogram (ECG) signals. The FCMM can accurately classify and distinguish the difference between normal heartbeats and abnormal heartbeats. Classifying the heartbeat cases from ECG signals consists of four main procedures: (i) QRS extraction stage for detecting QRS waveform using the Difference Operation Method; (ii) qualitative features stage for qualitative feature selection on ECG signals; (iii) Procedure-FCM is used to compute the cluster center for

each class; and (iv) Procedure-HCD is used to determine the heartbeat case for the patient. The available ECG records in the MIT-BIH arrhythmia database are utilized to illustrate the effectiveness of the proposed method. Experimental results show that the total classification accuracy was approximately 93.78%.

Keywords: ECG signal; Fuzzy C-Means.

一、前言

在傳統的集群分析方法中，假若把每一個觀測物均“嚴格”的劃分成「屬於」或「不屬於」某個集群，這種分析方式就稱為“硬式分群法 (Hard clustering)”。此方法的缺點是容易造成原始資料訊息的喪失，使得分群的結果不具意義。另一種方法稱為模糊集群分析法(Fuzzy cluster analysis)，亦即某一個觀測物它屬於某一個集群之程度是一個從 0 到 1 之間的數值(該數值稱為 membership value)，並且以該 membership value 的大小來決定該觀測物屬於某一個集群的大小程度。在模糊集群分析法中最常使用的是模糊 C 平均值(Fuzzy C-Means, FCM)演算法。FCM 演算法目前已廣泛的應用於模糊分群技術，它最早是由 Dunn 於 1973 年提出[1]，由 Bezdek 於 1981 年提出改良[2]。FCM 演算法現在已被廣泛的應用於各種領域，例如，圖形識別、影像處理、醫學工程等等。

心跳種類的分析與判斷，近年來，研究學者也提出了許多方法，如時域分析法[3]、頻域分析法[4]、小波轉換分析法[5]、類神經網路分

析法[6]、Cross-Distance 分析法[7],等等。本研究是以 FCM 演算法為基礎，提出一個簡單有效的心跳種類的分析與判斷方法，稱為模糊 C 平均值分析法(Fuzzy C-Means Method, FCMM)。本研究能判斷五種心跳種類，包括正常的心跳 NORM (normal heartbeats)與較常發生的四種不正常的心跳（包括 LBBB (left bundle branch block)、RBBB (right bundle branch block)、VPC (ventricular premature contractions)、及 APC (atrial premature contractions))。FCMM 共分成四個主要部份，分別是：(1) QRS extraction stage：尋找心電圖(ECG)信號中的 QRS 複合波、P 波及 T 波的位置；(2) qualitative features stage：選取主要特徵點；(3) Procedure-FCM：在本研究中，它的主要用途是計算每一個群集的中心值向量；(4) Procedure-HCD：心跳種類的分析與判斷。系統方塊圖如圖 1 所示。

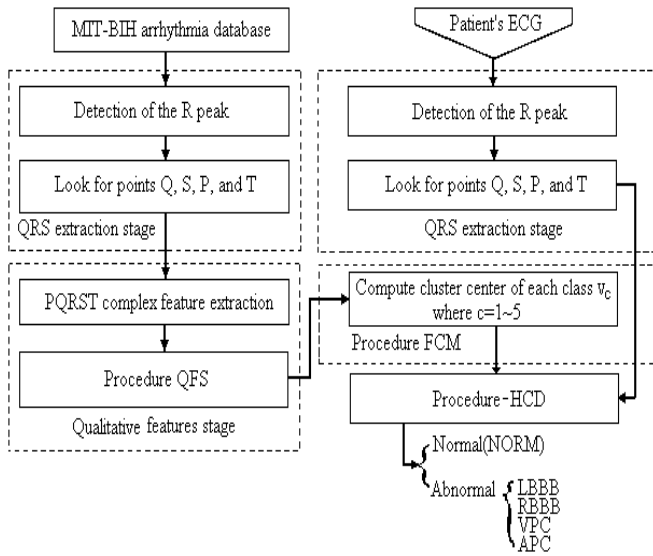


圖 1. 系統方塊圖

二、方法描述：模糊 C 平均值分析法 (Fuzzy C-Means Method, FCMM)

FCMM 共分成四個主要部份，分別是：(1)

QRS extraction stage：以差分運算法(Difference Operation Method)尋找心電圖(ECG)信號中的 QRS 複合波、P 波及 T 波的位置；(2) qualitative features stage: 從許多特徵點當中選取最重要的主要特徵點，其目的是要減少資料分群判斷的運算時間及減少記憶體儲存空間；(3) Procedure-FCM；(4) Procedure-HCD。其中的前兩大部份，請參考我們先前所提出的文獻[9]。後兩大部份說明如下：

(一). Procedure-FCM (Fuzzy C-Means)

在本研究中，Procedure-FCM 之主要用途是計算每一個群集的中心值向量。執行步驟如下所述：[10]

Step 1-1：已知資料 $X = \{X_1, X_2, \dots, X_N\}$ ，其內的每一個元素 $X_k \in R^p$ 。要分群的群集數 c 事先已選定，且滿足 $2 \leq c \leq (N - 1)$ 。歸屬度矩陣 $U^{(0)}$ 的初始值是隨意給定一個數值，如式子(1)。但歸屬隸屬度 $u_{i,k}$ 的值要滿足式子(2)的二個條件，而 $u_{i,k}$ 是表示某個元素 X_k 在群集 i 中的歸屬度。

$$U^{(0)} = \left(u_{i,k}^{(0)} \right)_{c \times N}$$

$$= \begin{bmatrix} u_{1,1}^0 & u_{1,2}^0 & u_{1,3}^0 & \cdots & u_{1,N}^0 \\ u_{2,1}^0 & u_{2,2}^0 & u_{2,3}^0 & \cdots & u_{2,N}^0 \\ u_{3,1}^0 & u_{3,2}^0 & u_{3,3}^0 & \cdots & u_{3,N}^0 \\ u_{4,1}^0 & u_{4,2}^0 & u_{4,3}^0 & \cdots & u_{4,N}^0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{c,1}^0 & u_{c,2}^0 & u_{c,3}^0 & \cdots & u_{c,N}^0 \end{bmatrix}_{c \times N} \quad (1)$$

$$\begin{cases} \text{條件 1: } 0 \leq u_{i,k} \leq 1, \\ \text{and} \\ \text{條件 2: } \sum_{i=1}^c u_{i,k} = 1, \end{cases} \quad (2)$$

其中， $1 \leq i \leq c, 1 \leq k \leq N$

Step 1-2: 選定模糊度參數 m (其中， $1 < m < \infty$)，

及設定“疊代次數” $t=0$ 。

Step 1-3：“疊代次數”的值增加 1；即是

$$t \leftarrow t+1 \quad (3)$$

Step 1-4：計算(或更新) c 個群集的中心值向

量 $v_i^{(t)}$ ，如式子(4)。

$$v_i^{(t)} = \frac{\sum_{k=1}^N (u_{i,k}^{(t-1)})^m (X_k)}{\sum_{k=1}^N (u_{i,k}^{(t-1)})^m}, \quad 1 \leq i \leq c \quad (4)$$

其中群集中心值向量 $V = \{v_1, v_2, \dots, v_c\}$ 。

Step 1-5：更新 $U^{(t)}$ 。其中 $U^{(t)}$ 是一個大小為 $c \times N$

的歸屬度矩陣(如式子(5))，而 $u_{i,k}^{(t)}$ 之計算如式子(6)。

$$U^{(t)} = \left(u_{i,k}^{(t)} \right)_{c \times N} = \begin{bmatrix} u_{1,1}^t & u_{1,2}^t & u_{1,3}^t & \cdots & u_{1,N}^t \\ u_{2,1}^t & u_{2,2}^t & u_{2,3}^t & \cdots & u_{2,N}^t \\ u_{3,1}^t & u_{3,2}^t & u_{3,3}^t & \cdots & u_{3,N}^t \\ u_{4,1}^t & u_{4,2}^t & u_{4,3}^t & \cdots & u_{4,N}^t \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{c,1}^t & u_{c,2}^t & u_{c,3}^t & \cdots & u_{c,N}^t \end{bmatrix}_{c \times N} \quad (5)$$

$$u_{i,k}^{(t)} = \frac{\left[\frac{1}{\|X_k - v_i^{(t)}\|^2} \right]^{1/(m-1)}}{\sum_{q=1}^c \left[\frac{1}{\|X_k - v_q^{(t)}\|^2} \right]^{1/(m-1)}}, \quad (6)$$

其中 $1 \leq i \leq c$ 及 $1 \leq k \leq N$

式子(6)中的 $\|X_k - v_i^{(t)}\|^2$ ，是表示元素 X_k 到群集中心 v_i 的歐氏距離(Euclidean distance)。

Step 1-6：條件 $\|U^{(t)} - U^{(t-1)}\| < \varepsilon$ ，若有滿足

則跳到 Step 1-7。若沒有滿足，則回到 Step 1-3。其中 ε 為預先設定之很小的正數。

Step 1-7：判斷元素 X_k 屬於 class- z ，條件是它最後的 $u_{i,k}^{(t)}$ 的值，在群集 i 中是最大的值，如式子(7)。

$$z = \arg \left\{ \text{Max} \left(u_{i,k}^{(t)} \right), i = 1, 2, \dots, c \right\} \quad (7)$$

(二). Procedure-HCD (Heartbeat Case Decision)

假設每一個樣本點 X 有四個特徵值(經 Procedure-QFS 所選的特徵值)，表示如下:[9]

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} \text{QRS} - \text{dur} \\ \text{QTP} - \text{int} \\ \text{Ratio} - \text{RR} \\ \text{Area} - \text{R'ST}' \end{bmatrix} \quad (8)$$

本單元分成兩個部份說明，一是初值的計算，另一個是心跳種類的分析與判斷。

1. 初值的計算：

(1) 計算 Mean vector.

$$\bar{X}_k = \frac{1}{N_k} \sum_{i=1}^{N_k} X_{k,i}, \quad K=1, 2, \dots, 5 \quad (9)$$

其中 $X_{k,i}$ 是第 k 群(假若有 5 群)的第 i 個樣本， N_k 是第 k 群的樣本總個數。

(2) 計算 Sample covariance matrix.

$$S_k = \frac{1}{N_k - 1} \sum_{i=1}^{N_k} (X_{k,i} - \bar{X}_k)(X_{k,i} - \bar{X}_k)^T, \quad k=1, 2, \dots, 5 \quad (10)$$

其中 S_k 是 4×4 的矩陣。

(3) 計算 Pooled sample covariance matrix S .

$$S = \frac{\sum_{k=1}^5 (N_k - 1) \cdot S_k}{N_T - k} \quad (11)$$

$$= \frac{(N_1 - 1)S_1 + (N_2 - 1)S_2 + (N_3 - 1)S_3 + (N_4 - 1)S_4 + (N_5 - 1)S_5}{N_1 + N_2 + N_3 + N_4 + N_5 - 5}$$

其中 S 是 4×4 的矩陣。

2. 心跳種類的分析與判斷：

執行步驟如下：

Step 2-1: 設 $M_k = 0$, M_k 是各種心跳種類的總數，而 $k = 1, 2, 3, 4, 5$ 。

Step 2-2: 從病人身上取一個心跳 Y_i ，假設每一個心跳有四個特徵值(經 Procedure-QFS 所選的特徵值)，表示如下:[9]

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{bmatrix} = \begin{bmatrix} \text{QRS} - \text{dur} \\ \text{QTP} - \text{int} \\ \text{Ratio} - \text{RR} \\ \text{Area} - \text{R'ST}' \end{bmatrix} \quad (12)$$

Step 2-3: 測量 Y_i 與 \bar{X}_k 之間的馬氏距離

$$d^2(Y_i, k) = (Y_i - \bar{X}_k)' S^{-1} (Y_i - \bar{X}_k), \quad (13)$$

其中, $k = 1, 2, 3, 4, 5$ 。

Step 2-4: 心跳 Y_i 被指定是第 k^* 群，條件是 Y_i 的馬氏距離 $d^2(Y_i, k)$ 在 k 群中是最小的值。

$$d^2(Y_i, k^*) = \text{Min}\{d^2(Y_i, k)\}, \quad (14)$$

其中, $k = 1, 2, 3, 4, 5$ 。

Step 2-5: $M_{k^*} = M_{k^*} + 1$ (15)

Step 2-6: 判斷是否要取下一個心跳 Y_i 。

首先是執行 $i = i + 1$ 。假若 $i \leq n$ 能滿足, 則回到 Step 2-2 取下一個心跳 Y_i 作

測試。否則, 繼續執行 Step 2-7。其中 n 是測試的心跳總個數。

Step 2-7: 病人的心跳種類被判斷是第 k^* 群,

條件是它的 $M_{k^*} \neq 0$ 。

三、性能評估

評估所使用的參數有 sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), 及 total classification accuracy (TCA)。這些參數的定義如下:[3][9]

$$Se_i = \frac{TP_i}{TP_i + FN_i} \quad (16)$$

$$Sp_i = \frac{TN_i}{TN_i + FP_i} \quad (17)$$

$$PPV_i = \frac{TP_i}{TP_i + FP_i} \quad (18)$$

$$NPV_i = \frac{TN_i}{TN_i + FN_i} \quad (19)$$

$$TCA = \frac{\text{the number of correct diagnosis beats}}{\text{the number of total beats}} = \sum_{i=1}^5 \frac{TP_i}{T_r} \quad (20)$$

其中, TP_i (true positives) 是表示能正確的判斷是心跳病症分類 i 的總心跳個數 (即是將 NORM 正確的判斷是 NORM), FN_i (false negatives) 是表示不能正確的判斷是心跳病症分類 i 的總心跳個數 (即是 NORM 不能正確的判斷是 NORM), TN_i (true negatives) 是表示將不是心跳病症分類 i 的心跳, 判斷成不是心跳病症分類 i 的總心跳個數 (即是將 LBBB、RBBB、VPC、及 APC 等, 判斷成不是 NORM), FP_i (false positives) 是表示將本來就不是心跳病症分類 i 的心跳, 判斷成是心跳病症分類 i 的總心跳個數 (即是將 LBBB、RBBB、VPC、及 APC 等, 判

斷成是 NORM)。T_r 是表示所有待測試之心跳總個數。最後，以 MIT-BIH 資料庫中的部份檔案，驗證本研究所提出的 FCMM 的效能。經實際的測試，心跳病症分類 NORM 是 95.06%，LBBB 是 91.03%，RBBB 是 90.50%，VPC 是 92.63%，APC 是 93.77%，平均正確判斷率可達 93.78%。表 1 是性能比較表，它是本文所提的 FCMM 法與現有且性能評估方式類似的文獻[6, 11]的性能比較表，由此表中可看出 FCMM 法與其它方法一樣有很高的正確判斷率，但 FCMM 法是一個方法簡單不須要複雜數學計算的方法。

四、結論

使用 FCMM 的最大好處是方法簡單，不需要複雜的數學式子，以及有高的辨識率。經實際的測試，Sensitivity 如下：心跳分類 NORM 是 95.06%，LBBB 是 91.03%，RBBB 是 90.50%，VPC 是 92.63%，APC 是 93.77%，平均正確判斷率可達 93.78%。測試 30 分鐘長的 MIT-BIH 中的 ECG 信號，所需要的測試時間少於 1 分鐘。

五、參考文獻

- [1] J.C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact Well-Separated clusters", *Journal of Cybernetics*, Vol. 3, pp. 32-57, 1973
- [2] J.C. Bezdek, "Pattern recognition with fuzzy objective function algorithms", New York, Plenum Press, 1981
- [3] I. Christov, G. Gómez-Herrero, V. Krasteva, I. Jekova, "A. Gotchev K. Egiazarian, Comparative study of morphological and time-frequency ECG descriptors for heartbeat classification", *Med. Eng. Phys.*, Vol. 28, pp.876-887, 2006
- [4] V.X. Afonso, W.J. Tomkins, T.Q. Nguyen, S. Luo, "ECG beat detection using filter banks", *IEEE Trans. Biomed. Eng.*, Vol. 46, pp. 192-202, 1999
- [5] C.W. Li, C.X. Zheng, C.F. Tai, "Detection of ECG characteristic points using wavelet transforms", *IEEE Trans. on Biomed. Eng.*, Vol. 42, pp. 21-28, 1995
- [6] Z. Dokur, T. Olmez, "ECG beat classification by a Novel Hybrid neural network", *Computer Methods and Programs in Biomedicine*, Vol. 66, pp. 167-181, 2001
- [7] M. Shahram, K. Nayebi, "ECG beat classification based on a Cross-Distance analysis", *International Symposium on Signal Processing and its Applications, Malaysia*, pp.234-237, ISSPA-2001
- [8] MIT-BIH database distribution, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, 1998.
- [9] Yeh, Y.C., Wang, W.J., Chiou, C.W.: "Cardiac arrhythmia diagnosis method using Linear Discriminant Analysis on ECG signals", *Measurement*, Vol.42, pp. 778-789, 2009.
- [10] W. C. Chen, M. S. Wang, "A fuzzy c-means clustering-based fragile watermarking scheme for image authentication", *Expert System with Application*, Vol. 36, pp.1300-1307, 2009
- [11] I.Jekova, G.Bortolan, I.Christov, "Assessment and comparison of different methods for heartbeat classification", *Medical Engineering and Physics*, Vol. 30, pp. 248-257, 2008.

表 1. 性能比較表

			NORM	LBBB	RBBB	VPC	APC	
FCMM		Se (%)	95.06	91.03	90.50	92.63	93.77	93.78
		PPV (%)	97.79	91.96	89.95	82.89	80.22	
		TCA (%)						
Knn-NN [11]	Knn	Se (%)	94.80	58.10	88.50	88.80	74.50	--
		PPV (%)	98.09	74.36	78.86	54.79	78.49	
		TCA (%)						
	FL	Se (%)	94.23	64.26	75.93	85.30	65.16	--
		PPV (%)	94.33	63.93	90.24	58.78	93.10	
		TCA (%)						
	DA	Se (%)	88.84	85.56	81.38	70.96	80.52	--
		PPV (%)	97.81	44.76	76.28	68.65	96.78	
		TCA (%)						
	NN	Se (%)	86.54	64.94	58.59	85.98	62.82	--
		PPV (%)	95.57	54.32	57.32	48.30	64.84	
		TCA (%)						
RCE Network [6]	DWT	Se (%)	93.2	39.1	78.4	86.5	63.0	60.0
		PPV (%)	86.0	72.2	73.9	54.5	96.3	
		TCA (%)						
	DFT	Se (%)	86.3	60.8	47.0	77.6	85.0	50.1
		PPV (%)	73.1	38.0	73.6	60.6	92.2	
		TCA (%)						
InS Network [6]	DWT	Se (%)	100	94.6	98.6	91.3	100	95.7
		PPV (%)	96.7	91.0	94.2	93.5	98.0	
		TCA (%)						
	DFT	Se (%)	77.3	78.0	82.0	66.6	94.6	78.2
		PPV (%)	74.3	70.4	70.2	75.7	83.5	
		TCA (%)						
MLP Network [6]	DWT	Se (%)	100	48.0	74.6	98.6	99.3	87.6
		PPV (%)	92.6	96.0	99.1	81.3	78.8	
		TCA (%)						
	DFT	Se (%)	89.3	81.3	88.6	92.0	96.0	84.7
		PPV (%)	93.7	77.2	73.4	76.2	91.3	
		TCA (%)						