第三章 特徵擷取與分類

本文提出一套特徵擷取與分類的方法對腦電波做分類，先以對角化主要成份分析法，從包含了時間、頻率，與空間域資料的高維度矩陣中擷取出維度較小的主要特徵矩陣，再以本文定義的絕對貢獻度與相對貢獻度當做選取分量的標準，在以改良式最近鄰居分類法將特徵矩陣分類時，只用絕對貢獻度較高的分量來計算歐氏距離。接下來將依照資料分析流程，介紹我們所使用與提出的演算法。

3.1 混合時間、頻率與空間域特徵的初始資料矩陣

不同事件所產生的腦電波，在空間域、時間域與頻率域，會有不同的特徵，其中空間域與時間域的資料，可以由量測到的數值直接得到。因為腦電波是屬於時間序列的訊號，所以頻率域的資料需要經過時間—頻率的轉換才能獲得，本研究使用快速傅立葉轉換(Fast Fourier Transform, FFT)來求得腦電波訊號的頻譜值。

快速傅立葉轉換，是為改善離散傅立葉轉換所提出的快速頻譜分析方法，在介紹之前，先從傅立葉級數開始說明。傅立葉級數的基本觀念，就是任何一個週期函數，都可表示成不同頻率的弦波函數的線性組合，即一個週期訊號$x(t)$，可以表示為傅立葉級數，如(3.1)式所示。

$$x(t)=\frac{a_0}{2}+\sum_{n=1}^{\infty}\left(a_n\cos\frac{2\pi n}{T}t+b_n\sin\frac{2\pi n}{T}t\right),\ 0\le t\le T \tag{3.1}$$

其中

$$a_0=\frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}f(t)dt$$

$$a_n=\frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}}f(t)\cos\frac{2\pi n}{T}tdt\ ,\ n=1,2,3...$$

$$b_n = \frac{2}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)\sin\frac{2\pi n}{T} t dt\ ,\ n = 1,2,3...$$

，T為取樣週期。

以複數型式表示傅立葉級數，將更為簡潔。整理(3.1)可得複數型式的表示式為：

$$x(t) = \sum_{n=-\infty}^{\infty} c_n \exp\left(\frac{j2\pi n}{T} t\right)\ ,\ 0 \le t \le T \tag{3.2}$$

其中

$$c_n = \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} f(t)\exp\left(\frac{-j2\pi n}{T} t\right) dt\ ,\ n = 1,2,3...$$

由(3.2)式可知，傅立葉級數所使用的條件是對週期為T的信號進行分析。若取樣週期不是信號週期的整數倍，或信號本身即是非週期性時，將會造成分析的誤差。

傅立葉級數僅適合對連續的週期性訊號做頻譜的分析，而我們所要處理的腦電波訊號，則是非週期性的訊號，因此，當我們將一段有限長度的類比腦電波訊號，轉換成數位訊號以輸入到個人電腦中做後續訊號處理的時候，則必須使用離散傅立葉轉換，才得以準確分析其頻譜。假設一段長度為N個取樣點的數位訊號，則離散傅立葉轉換可表示為：

$$x(n) = \sum_{m=0}^{N-1} X(m)\exp\left(\frac{j2\pi mn}{N}\right)\ ,\ n = 0,1,2,..., N-1 \tag{3.3}$$

其中

$$X(m) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)\exp\left(\frac{-j2\pi mn}{N}\right)\ ,\ m = 0,1,2,..., N-1$$

在實際分析腦電波訊號時，取樣點數N往往非常大，因而造成很大的計算量，使得計算速度緩慢，在用於即時系統的時候，這個問題就顯得特別嚴重。因此，在實際應用的時候，通常會採用Cooley與Tukey所提出的快速傅

立葉轉換[30]，來取代原始的離散傅立葉轉換，以降低計算量，增加系統的速度。我們先將(3.3)式的離散傅立葉轉換，重新表示如(3.4)式的形式：

$$X(m) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)W^{mn} \ , \ m = 0,1,2,...,N-1 \tag{3.4}$$

式中 $W = \exp\left(\frac{-j2\pi}{N}\right)$，$N$必須為2的次方。若$N$=8，則$X(0)$~$X(7)$可表示為：

$$\begin{bmatrix} X(0) \\ X(1) \\ X(2) \\ X(3) \\ X(4) \\ X(5) \\ X(6) \\ X(7) \end{bmatrix} = \frac{1}{N}\begin{bmatrix} W^0 & W^0 & W^0 & W^0 & W^0 & W^0 & W^0 & W^0 \\ W^0 & W^1 & W^2 & W^3 & W^4 & W^5 & W^6 & W^7 \\ W^0 & W^2 & W^4 & W^6 & W^0 & W^2 & W^4 & W^6 \\ W^0 & W^3 & W^6 & W^1 & W^4 & W^7 & W^2 & W^5 \\ W^0 & W^4 & W^0 & W^4 & W^0 & W^4 & W^0 & W^4 \\ W^0 & W^5 & W^2 & W^7 & W^4 & W^1 & W^6 & W^3 \\ W^0 & W^6 & W^4 & W^2 & W^0 & W^6 & W^4 & W^2 \\ W^0 & W^7 & W^6 & W^5 & W^4 & W^3 & W^2 & W^1 \end{bmatrix}\begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ x(3) \\ x(4) \\ x(5) \\ x(6) \\ x(7) \end{bmatrix} \tag{3.5}$$

如(3.5)式，原始的離散傅立葉轉換的計算過程，需要作N^2次的複數乘法，與$N(N$-1)次的複數加法。快速傅立葉轉換是把原來離散傅立葉轉換中N維的方陣，將其因子分解(Factorization)成r ($r = \log_2 N$)個均為N維方陣的矩陣因子，在分解時，將零引入矩陣內，使得每一個矩陣因子的每一列中，只有兩個非零的數。所以快速傅立葉轉換的計算，只需要$(N \times r)/2$次複數乘法，和$(N \times r)$次的複數加法運算，節省了很多計算所需的時間。

若以信號流程圖表示，則可得如圖3.1所示的流程圖[31]：

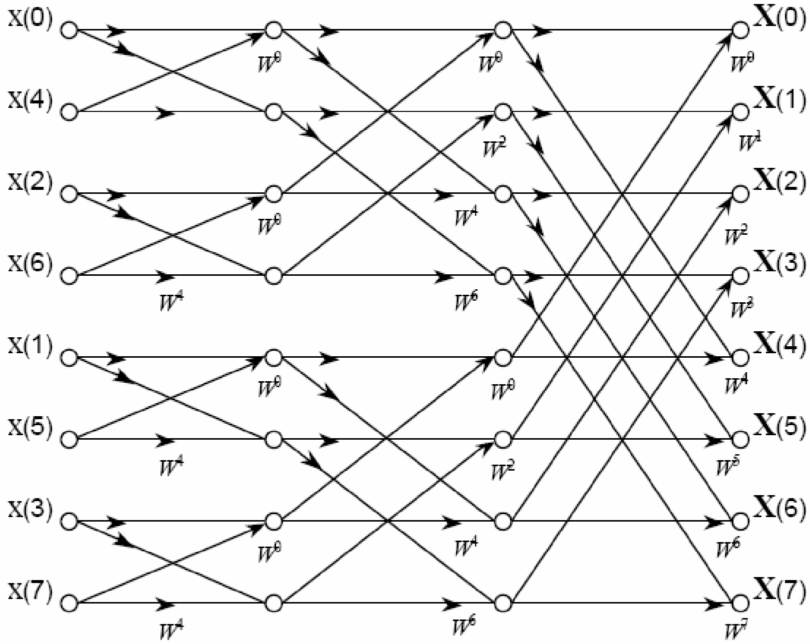

圖3.1　FFT 信號流程圖[31]

3.2 以對角化主要成份分析法(DiaPCA)擷取主要特徵

在得到腦電波空間域、時間域與頻率域的資料以後，我們所要做的，就是將這三個域的資料所組合成的高維度資料矩陣降低維度，只保留其中主要的成份。我們一開始之所以使用腦電波在三個域的資料，是因為不同人做想像動作所產生的腦電波，表現出差異的特徵也會略有不同，為了不遺失能夠表現出特徵差異的資料，才將所有資料包含進去。但是資料量並不是越多越好，過多的資料，不但增加了後續運算所需的時間，也可能包含了錯誤的特徵，而誤導了辨識的結果。因此，許多團隊採用主要成份分析法(Principal Component Analysis, PCA)來降低這三個域的資料所組合成的高維度矩陣之維度，只保留其中主要的特徵成份。

主要成份分析法的目的，在於從初始資料矩陣中選出前d個對應的特徵值(Eigenvalue)較大的特徵向量(Eigenvector)，將原始資料投影在該特徵向量上，雖然新維度比原始資料維度低，但仍可保留住原始資料中較重要的訊息。在圖3.2中[32]，有一群二維空間的資料點，其中w1、w2是其兩個正交

的特徵向量。如果要以其中一個軸向的投影量，來取代原本二維的資料，則可以看到w1上的投影量，要比w2上的投影量，保留了更多原來資料點分布的特徵，因此w1是較佳的特徵向量。而主要成份分析法，就是將n個特徵向量依照其保留的資訊量，由高到低排列，讓使用者可以視情況，保留d個(d<n)較佳的特徵向量，以達到降低維度的效果，同時仍保留了原始資料的重要特徵。

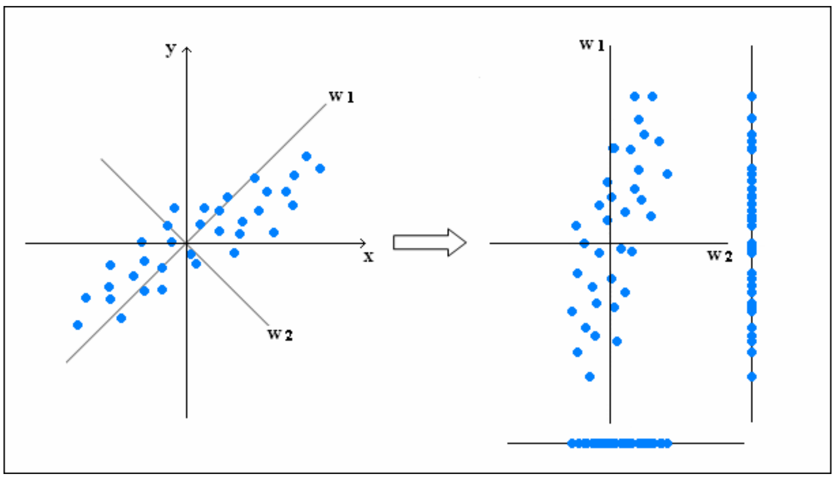

圖3.2　　主要成份分析法示意圖[32]

主要成份分析法看似能夠降低矩陣的維度，並且保留重要的特徵，但是在其計算的過程中，卻隱含一個計算上的盲點，使得矩陣列向量所帶有的特徵，在運算過程中遺失了。有許多學者針對主要成份分析法的這個缺點提出改善的方法，對角化主要成份分析法就是其中之一，而且在人臉辨識的應用中，辨識率有明顯的提升[16]，因此我們將它應用在腦電波的辨識。

為了解決主要成份分析法遺失矩陣列向量中特徵的缺點，對角化主要成份分析法將初始資料矩陣的元素以對角化的方式重新排列。假設腦電波的原始矩陣A為一個m×n的矩陣，對角化的示意圖如圖3.3所示(以m＝4,n＝5為例)[16]：

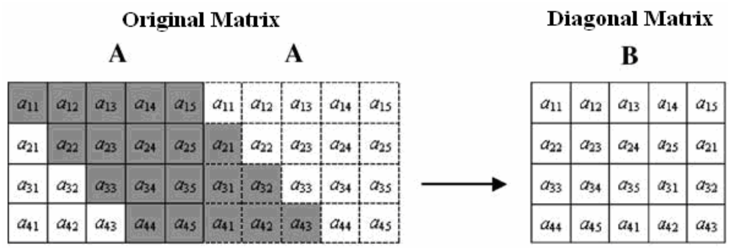

圖3.3　　原始矩陣對角化示意圖[16]

首先，將原始矩陣A(Original Matrix)複製且橫向並排，其中a_{ij}(i=1~4, j=1~5)為矩陣中的元素，然後只保留圖中深色部分的元素而白色部分去除掉，也就是只保留5條對角線的元素，接著再將深色部分的像素重新排列整齊，得到對角化的矩陣B(Diagonal Matrix)。經過這樣的步驟以後，對角化後矩陣B相對於原始矩陣A，並沒有增加或減少元素，只是重新排列，而且是依照原始矩陣的對角線順序排列，因此B稱為對角化矩陣。而B矩陣每一行向量的元素，其實就包含了原始矩陣A在行向量與列向量上的資料，因此，將主要成份分析法使用在對角化矩陣的時候，雖然只運用到B矩陣行向量上的資料，但是等於同時運用到了原始矩陣A在行向量與列向量的資料。

Zhang的實驗的結果顯示，在人臉辨識的應用中，使用對角化主要成份分析法比原來沒有對角化步驟增加了5%的辨識率。因此，我們將對角化主要成份分析法應用在腦電波的辨識上，希望同時保留時間域、頻率域與空間特徵，並測試它使用在腦電波辨識的效果。接下來介紹對角化主要成份分析法擷取主要腦電波特徵的後續步驟，也是跟主要成份分析法相同的部分。

假設我們要將腦電波分為M類，則先將已知類別的訓練資料，依照類別的不同分開來求對角化的矩陣的平均值，得到M個類別的平均矩陣，分別記為$B_1, B_2 \ldots, B_M$，再求對角化的類別間的總體平均矩陣$\overline{B}$(m×n)：

$$\overline{B} = \frac{1}{M}\sum_{k=1}^{M} B_k \tag{3.6}$$

接著求其共變異數矩陣 G(n×n)，定義如下：

$$G = \frac{1}{M}\sum_{k=1}^{M}\left(B_k - \overline{B}\right)^T\left(B_k - \overline{B}\right) \tag{3.7}$$

(3.7)式將對角化的類別平均矩陣 B_k 減總體平均矩陣 $\overline{B}$ 得到的矩陣先轉置再相乘，在這個運算過程中，只有 $\left(B_k - \overline{B}\right)$ 的行向量在運算，列向量在運算過程中被忽略了，這也就是主要成份分析法計算上的盲點。求得 G 之後，計算它的特徵向量 X，與特徵值 λ：

$$GX = \lambda X \tag{3.8}$$

再將這 n 個特徵向量依照其對應的特徵值的大小來排序，對應的特徵值越大，則該特徵向量就越有利於辨識。假設 $X_d=[x_1,\ldots,x_d]$ 為 $X=[x_1,\ldots,x_n]$ 中，前 d 個最佳的特徵向量($d<n$)所組成的矩陣，則把待分類的測試資料的初始矩陣 A_{test} 乘上 X_d，就得到測試資料的特徵矩陣 C_{test} (m×d)：

$$C_{test} = A_{test} X_d \tag{3.9}$$

此時，C_{test} 矩陣所包含的，就是經過對角化主要成份分析法降低維度後的主要成份，將此特徵矩陣，再經由後續的分類演算法，就可以辨識出腦電波所屬的類別。

3.3 以改良式最近鄰居分類法分類腦電波

(3.8)式中，從共變異數矩陣 G 求出特徵向量 X，並保留其中對應到較大特徵值的 d 個特徵向量，記為 X_d。接下來我們將各類別未經對角化的平均矩陣 A_k 分別乘上 X_d，以得到各類別腦電波的標準特徵矩陣 C_k，如(3.10)式所示。

$$C_k = A_k X_d \tag{3.10}$$

接著，就可以用最近鄰居分類法，將測試資料A_{test}的特徵矩陣C_{test}分類。最近鄰居分類法藉由計算C_{test}與各個類別的標準特徵矩陣C_k(k＝1~M)的歐氏距離，找出與C_{test}距離最近的標準特徵矩陣之後，就把測試資料歸類到該類別，完成對角化主要成份分析法最後的分類步驟，歐氏距離的數學式如(3.11)式所示：

$$d\left(C_{test},C_k\right)=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{d}\left(C_{test}^{(i,j)}-C_k^{(i,j)}\right)^2} \tag{3.11}$$

歐氏距離的計算，可以視為是把兩個向量在各個分量上垂直投影後的距離平方，相加，再開平方。然而，我們發現這樣的計算方式，因為把特徵矩陣在每個分量的距離都計算進去，容易被少數幾個不利於辨識的分量所影響，造成辨識錯誤的機會增加，如果以二維的向量來當作例子，可以圖3.4來說明：

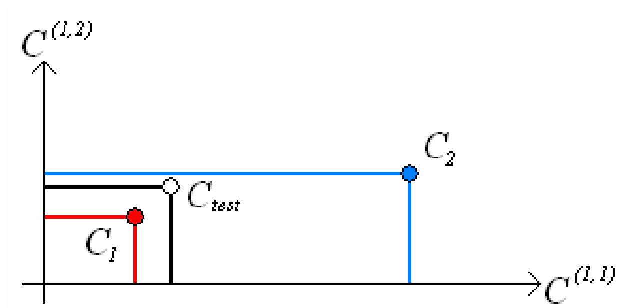

圖3.4　　二維向量座標軸的絕對正向貢獻度說明圖

圖中C_1與C_2分別是1類與2類資料的標準特徵矩陣，$C^{(1,1)}$與$C^{(1,2)}$則是代表二維向量的兩個互相正交的座標軸。我們要求測試資料的特徵矩陣C_{test}與C_1、C_2的距離，從圖中可以很明顯的看出C_{test}與C_1的距離比較近，所以C_{test}應該要被分類為1類。但是仔細觀察可以發現，把這三個向量分別投影到$C^{(1,1)}$與$C^{(1,2)}$來計算距離，結果兩個座標軸的結果不一樣。在$C^{(1,1)}$軸上，C_{test}的確與C_1距離較近，因此$C^{(1,1)}$可以說是對於分類有正向的貢獻；而在$C^{(1,2)}$軸上，C_{test}卻是距離C_2比較近，因此$C^{(1,2)}$對於分類是呈現負向的貢獻。

由於腦電波的特徵矩陣維度很高，如果提供負向貢獻的分量太多，會導致辨識的錯誤。因此，我們先去評估各個分量的正向貢獻度，並由高到低排列，在使用最近鄰居分類法分類的時候，我們只用正向貢獻度較高的幾個分量來計算歐氏距離，這就是我們提出的改良式最近鄰居分類法的精神。

至於如何評估每個分量的正向貢獻度，我們定義了兩種參數，其中一種我們稱它為「絕對貢獻度」，另一種則稱做「相對貢獻度」。

1.絕對(正、負向)貢獻度：

絕對貢獻度是以該分量在單獨被用來辨識時造成的正確率，來評估其對分類的貢獻度。以本論文所使用的腦電波資料集I為例，該資料集總共有278筆想像動作的腦電波資料(分為抬起左手小指與吐舌頭兩類)，我們所定義的單一分量的絕對貢獻度，其計算步驟如下：

(1) 以對角化主要成份分析法從277筆訓練資料中，求出兩個類別的標準特徵矩陣C_1,C_2 (維度是64×1)。

(2) 對剩下的一筆已知類別的測試資料A_{test}計算其主要特徵矩陣C_{test}。

(3) 一次只使用特徵矩陣裡的一個分量來計算C_{test}與C_1,C_2的歐氏距離，並將C_{test}分類，64個分量各分類一次。

(4) 比較使用64個分量分類的結果與該測試資料的已知類別是否相同，若相同(辨識結果為正確)，則將該分量的絕對正向貢獻度+1；反之，辨識結果與已知類別不同，則該分量的絕對負向貢獻度-1，

(5) 全部278筆資料輪流當作測試資料，並重覆(1)~(4)的步驟之後，各個分量的絕對(正、負向)貢獻度就被評估出來了。

單一分量的絕對(正、負向)貢獻度計算流程圖如圖3.5所示。在我們的實驗中，絕對正、負向貢獻度的範圍小於等於±278(訓練資料的數量)。

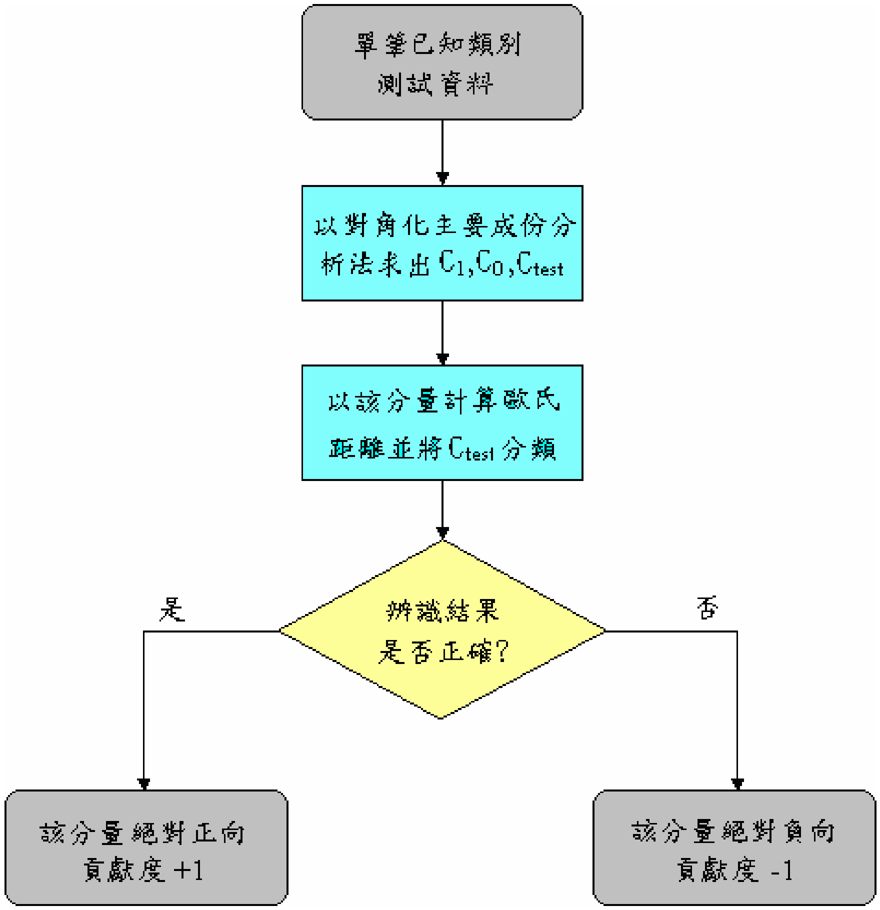

圖3.5　評估單一分量的絕對貢獻度流程圖

2.相對(正、負向)貢獻度

相對貢獻度是以該分量被從64個分量中剔除掉以後，對辨識的結果造成什麼樣的影響，來評估其對分類的貢獻度。同樣以本論文所使用的腦電波資料集I為例，我們所定義的單一分量的相對貢獻度，其計算步驟如下：

(1) 以對角化主要成份分析法從277筆訓練資料中，求出兩個類別的標準特徵矩陣C_1,C_2 (維度是64×1)。

(2) 對剩下的一筆已知類別的測試資料A_{test}計算其主要特徵矩陣C_{test}。

(3) 使用特徵矩陣全部64個分量來計算C_{test}與C_1,C_2的歐氏距離，並將C_{test}分類。

(4) 一次去除掉一個分量，以剩下的63個分量來計算C_{test}與C_1,C_2的歐氏距離，並將C_{test}分類，64個分量分別被測試過少了該分量的辨識結果。

(5) 比較去除掉該分量後的辨識結果是否與使用64個分量分類的結果相同，若相同(辨識結果不變)，則將該分量的相對貢獻度不變；若原本辨識正確，但去除掉該分量後辨識結果變成錯誤，表示該分量有助於辨識，所以該分量的相對正向貢獻度+1；反之，如果原本辨識錯誤，但去除掉該分量後辨識結果變成正確，表示該分量有害於辨識，所以該分量的相對負向貢獻度-1。

(6) 全部278筆資料輪流當作測試資料，並重覆(1)~(5)的步驟之後，各個分量的相對(正、負向)貢獻度就被評估出來了。

單一分量的相對(正、負向)貢獻度流程圖如圖3.6所示，而相對正向貢獻度與相對負向貢獻度相加，則定義為相對貢獻度。

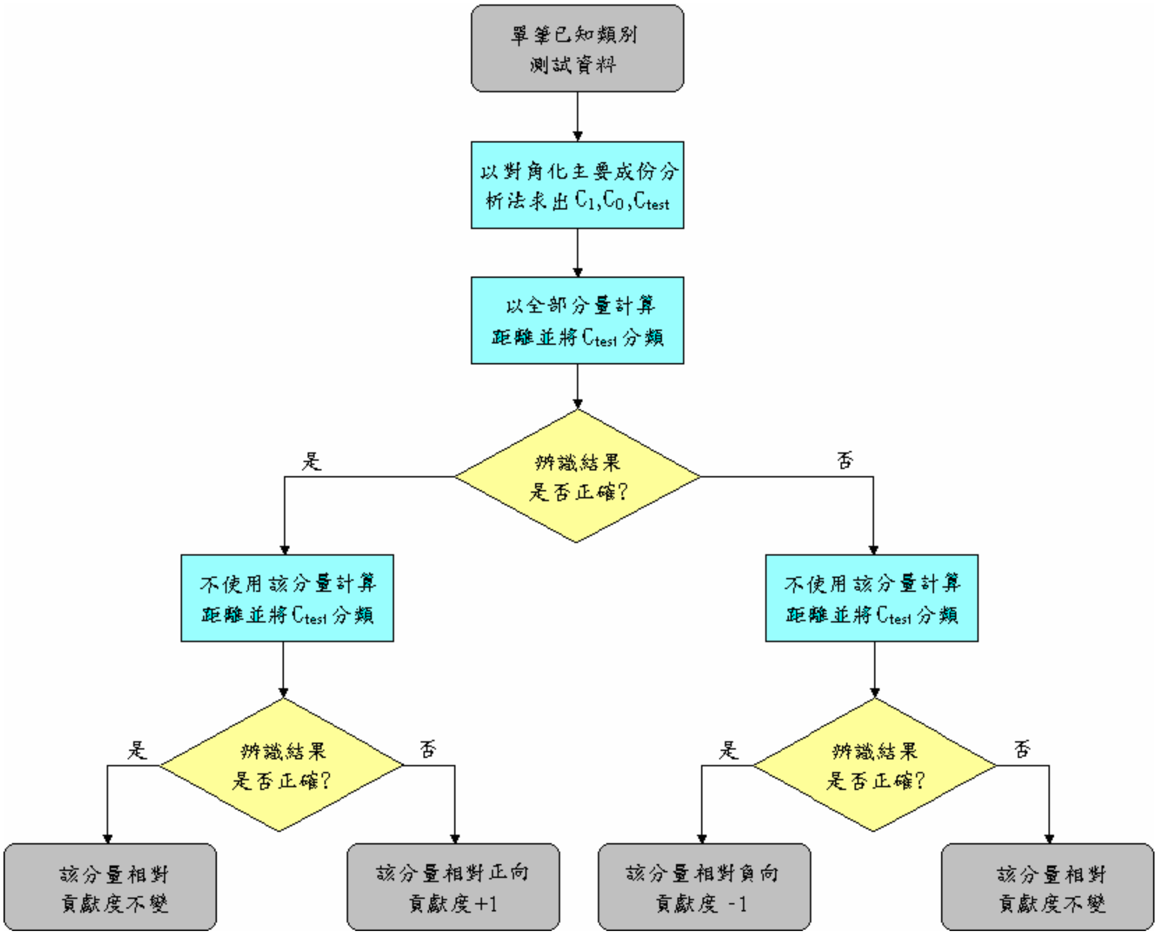

圖3.6　評估單一分量的相對貢獻度流程圖

我們所提出的方法與原本的最近鄰居分類法的差別，主要在於針對計算特徵矩陣之間歐氏距離的公式，定義了「絕對貢獻度」與「相對貢獻度」兩個參數，來評估各分量對於分類正確的貢獻度，依此來決定分類腦電波訊號的時候，要使用哪幾個分量來計算特徵矩陣的歐氏距離。

本章所介紹的這一套新的方法，命名為「改良式對角化主要成份分析法」，對特徵擷取的部分而言，其特點是使用了對角化主要成份分析法，使得特徵矩陣的維度降低，而且保留了有利於辨識的特徵；在分類方法的部分，則是定義了選取計算距離所採用的分量的標準(絕對與相對貢獻度)，使得最近鄰居分類法在計算特徵矩陣之間的歐氏距離時，可以排除可能導致辨識錯誤的分量，以提高辨識準確率。