

相似度比率式鑑別分析應用於大詞彙連續語音辨識

Likelihood Ratio Based Discriminant Analysis for Large Vocabulary Continuous Speech Recognition

李鴻欣 Hung-Shin Lee

國立臺灣師範大學資訊工程學系

Department of Computer Science and Information Engineering

National Taiwan Normal University

hungshin@live.com

陳柏琳 Berlin Chen

國立臺灣師範大學資訊工程學系

Department of Computer Science and Information Engineering

National Taiwan Normal University

berlin@ntnu.edu.tw

摘要

在近十年來所發展出的自動語音辨識(automatic speech recognition, ASR)技術中，仍有許多研究者嘗試僅藉由前端處理來產生具有鑑別性的語音特徵，而獨立於後端模型訓練與分類器特性。本論文即在此思維下提出嶄新的鑑別式特徵轉換方法，稱為普遍化相似度比率鑑別分析(generalized likelihood ratio discriminant analysis, GLRDA)，其旨在利用相似度比率檢驗(likelihood ratio test)的概念尋求一個維度較低的特徵空間。在此子空間中，我們不僅考慮了全體資料的異方差性(heteroscedasticity)，即所有類別之共變異矩陣可被彈性地視為相異，並且在分類上，因著我們也將類別間最混淆之情況（由虛無假設(null hypothesis)所描述）的發生率降至最低，而達到有助於分類正確率提升的效果。同時，我們也證明了傳統的線性鑑別分析(linear discriminant analysis, LDA)與有名的異方差性線性鑑別分析(heteroscedastic linear discriminant analysis, HLDA)可被視為我們所提出之普遍化相似度比率鑑別分析(GLRDA)的兩種特例。此外，為了增進語音特徵的強健性，我們所提出的方法更可進一步地與辨識器所提供的實際混淆資訊結合，而獲得在中文大詞彙連續語音辨識的實驗中，相較於以上兩種傳統方法更高的辨識正確率。

關鍵詞：語音辨識、特徵擷取、相似度比率、鑑別分析、混淆資訊

一、緒論

為了降低計算量與模型的複雜度，語音特徵轉換(feature transformation)在自動語音辨識(automatic speech recognition, ASR)中扮演了很重要的角色。它的目標在於尋求一個

線性轉換 $\Theta \in \mathcal{R}^{n \times d}$ ，將原有在 n 維空間的聲學特徵向量，投影至 d 維的子空間($d < n$)，使得新的特徵在資料類別間具有較好的鑑別力[1]。而在實務上，語音特徵轉換的技術可被分為兩種範疇[2]：相依於分類器(classifier-dependent)與獨立於分類器(classifier-independent)。在相依於分類器的範疇中，如某些基於最小音素錯誤(minimum phone error, MPE)[3]與最小分類錯誤(minimum classification error, MCE)[4]的鑑別式方法，轉換矩陣是結合聲學模型(acoustic models)中的參數估測或是分類器所掌握的分類規則一併求得。相對地，獨立於分類器的範疇則是基於各種不同的類別分離度標準，特別是幾何分離度，在聲學模型訓練之前就依據既有的類別統計資訊求出轉換矩陣。例如，線性鑑別分析(linear discriminant analysis, LDA)即試圖最大化類別間的平均馬氏距離平方(squared Mahalanobis distance)[5]；而做為線性鑑別分析(LDA)的普遍化，異質性鑑別性分析(heteroscedastic linear discriminant analysis, HLDA)則在最大化相似度(maximum likelihood)的框架下，處理每一類別具有相異之共變異矩陣的情形[6]。另外，異質性鑑別分析(heteroscedastic discriminant analysis, HDA)個別地考慮了每一類別的分佈，而產生新的目標函式[7]；為了保留比線性鑑別分析(LDA)和異質性鑑別性分析(HLDA)更多的鑑別資訊，最大化交互資訊(maximum mutual information, MMI)和最小分類錯誤(MCE)標準也被引入此範疇中使用[8]。此外，最近的研究者開始將成對經驗錯誤率(pairwise empirical error rate)列入考量，期望辨識器在前端處理與後端分類階段的不一致性能夠降低至一定程度[9-11]。

對於獨立於分類器的範疇來說，雖然在類別分離度與辨識結果之間仍存有較大的差距，也就是較高的類別分離度，並不必然保證有較低的辨識錯誤率。但在本論文中，我們仍將研究重點聚焦於此，原因在於：當語音特徵擷取完全與後端聲學模型分離，對於較複雜的自動語音辨識系統，聲學模型訓練模式的改變，就較不會影響到前端的訊號處理，使得此系統較易於被分析解構。而當某些系統的聲學模型機制是固定的，或是以硬體方式呈現，那麼我們就能在無法更動硬體的情況下，對前端訊號處理進行研究或改善[12]。更重要的是，我們相信在此範疇中所設計出的方法，能夠更廣泛地應用在其他圖型辨識(pattern recognition)的領域，如人臉辨識等，而不侷限於系統模型較為複雜的語音辨識。

在本論文中，我們提出了一個嶄新的鑑別式特徵擷取方法，稱為普遍化相似度比率鑑別分析(generalized likelihood ratio discriminant analysis, GLRDA)，其旨在利用相似度比率檢驗(likelihood ratio test, LRT)的概念來尋求一個維度較低的特徵空間。在此子空間中，我們不僅考慮了全體資料的異方差性(heteroscedasticity)，即所有類別母體之共變異矩陣可被彈性地視為相異，並且在分類上，因著我們也將類別間最混淆之情況（由虛無假設(null hypothesis)所描述）的發生率降至最低，而達到有助於分類正確率提升的效果。此外，若我們假設所有類別母體均遵循高斯分佈(Gaussian distribution)，且針對其共變異矩陣給予不同的限制，則普遍化相似度比率鑑別分析(GLRDA)可被化約至傳統的線性鑑別分析與有名的異方差性線性鑑別分析。而為了增進聲學特徵的強健性，我們的方法更可進一步地與辨識器所提供的經驗混淆資訊結合。

二、普遍化相似度比率鑑別分析

(一) 背景

根據統計式假設檢定(statistical hypothesis testing)的定義[13]，相似度比率檢定(LRT)是一種廣為使用的方法，藉著它我們可獲得虛無假設(null hypothesis) H_0 與完全普遍化之對立假設(alternative hypothesis) H_1 間相互比較的檢定統計量。在本論文中，虛無假設 H_0 通常表示不利於我們的目標設定，或我們不願見到的情況，在鑑別式特徵擷取上，即為使類別母體不具鑑別性的情況。值得一提的是，虛無假設和對立假設之聯集(union)恰為完整的參數空間。

若 Ω 表示完整的參數空間(parameter space)，而 ω 表示被虛無假設 H_0 所限制的參數子空間，則相似度比率檢定針對虛無假設 H_0 和對立假設 H_1 之間的標準為

$$LR = \frac{\sup L_{\omega}}{\sup L_{\Omega}} \quad (1)$$

其中， L 表示訓練樣本(sample)資料的相似度， $\sup L_S$ 則表示以 S 為參數子空間時的最大相似度。由式(1)可看出，相似度比率檢定是由兩個部分組成：最大相似度與比率。使用最大化相似度估計法(maximum likelihood estimation, MLE)的用意在於找出最適合兩個統計假設或最具代表性的參數估計量。而相似度比率其背後的邏輯則在於，若我們不考慮任何信心度量測(confidence measure)且虛無假設 H_0 絕對為真(true)，則在完整參數空間 Ω 中的最大相似度參數估測必定發生在參數子空間為 ω 的情況；因此， $\sup L_{\omega}$ 與 $\sup L_{\Omega}$ 必定會非常接近，使 LR 趨近於1。反之，若 H_0 絕對為假(false)，則最大相似度發生的參數空間必定不是 ω ；因此， $\sup L_{\omega}$ 將會遠小於 $\sup L_{\Omega}$ 。

(二) 基本概念

相似度比率檢定在語音處理上的應用並不廣泛，近年來它常被用於評估音素間的混淆程度(phonetic confusions)[14]或是語音活動偵測(voice activity detection, VAD)[15]。在鑑別式語音特徵擷取技術中，我們並不打算緊密地遵照相似度比率檢定的過程，而目標也不在依據統計量來檢定虛無假設是真是假。我們的目標在於尋找一個投影子空間，使得虛無假設在此子空間中盡可能不會為真。為了使所有類別母體在此子空間中具有鑑別性，我們設計了以下的鑑別式統計假設：

$$\begin{cases} H_0 : \text{所有類別母體均相同} \\ H_1 : \text{所有類別母體均相異} \end{cases}$$

因此，我們所找到的子空間 $\Theta \in \mathfrak{R}^{n \times d}$ ，必須盡可能地推翻不具鑑別性的虛無假設 H_0 ，也就是使其相似度最小。普遍化相似度比率鑑別分析(generalized likelihood ratio discriminant analysis, GLRDA)目標函式便可寫成

$$J_{\text{GLRDA}}(\Theta) = LR_{\text{GLRDA}}(\Theta) = \frac{\sup L_{\text{所有類別母體均相同的參數子空間}(\Theta)}}{\sup L_{\text{完整的參數空間}(\Theta)}} \quad (2)$$

轉換矩陣 Θ 便可藉由最小化 $J_{\text{GLRDA}}(\Theta)$ 求得。

(三) 同方差性(Homoscedasticity)

一般來說，我們會以類別母體之期望值向量的估計量所形成的空間作為判斷所有類

別母體是否相同的參數空間。若所有類別母體具同方差性(homoscedasticity)，也就是每一類別母體具有相同的共變異矩陣，則令 $\boldsymbol{\mu}_i$ 為每一類別母體 C_i 的期望值向量， $\boldsymbol{\Sigma}_i$ 為每一類別母體 C_i 的共變異矩陣， H_0^{homo} 和 H_1^{homo} 可設定為：

$$\begin{cases} H_0^{\text{homo}} : \text{對於每一類別 } C_i, \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, \boldsymbol{\mu}_i = \boldsymbol{\mu}。 \\ H_1^{\text{homo}} : \text{對於每一類別 } C_i, \text{ 且 } \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}, \boldsymbol{\mu}_i \text{ 不受任何限制。} \end{cases}$$

H_0^{homo} 代表了一種極端的情況，若它為真，則所有類別母體幾近完全重疊，也就沒有任何鑑別性。因此，普遍化相似度比率鑑別分析(GLRDA)的任務即在 H_0^{homo} 最不可能為真的情況下，找出最合適的投影子空間。

以下的命題證明了若每一類別母體均遵循高斯分佈(Gaussian distribution)，則線性鑑別分析(LDA)轉換矩陣，等同於將 H_0^{homo} 與 H_1^{homo} 置於普遍化相似度比率鑑別分析(GLRDA)的框架下求解。

命題一：若每一類別母體 C_i 都具有高斯分佈，則最小化普遍化相似度比率鑑別分析(GLRDA)的目標函式

$$J_{\text{GLRDA}}^{\text{homo}}(\boldsymbol{\Theta}) = \frac{\sup L_{H_0^{\text{homo}}}(\boldsymbol{\Theta})}{\sup L_{H_1^{\text{homo}}}(\boldsymbol{\Theta})} \quad (3)$$

等同於最大化線性鑑別分析(LDA)的目標函式

$$J_{\text{LDA}}(\boldsymbol{\Theta}) = \frac{|\boldsymbol{\Theta}^T \mathbf{S}_B \boldsymbol{\Theta}|}{|\boldsymbol{\Theta}^T \mathbf{S}_W \boldsymbol{\Theta}|} \quad (4)$$

其中， $\mathbf{S}_B \in \mathfrak{R}^{n \times n}$ 和 $\mathbf{S}_W \in \mathfrak{R}^{n \times n}$ 分別代表類別間散佈矩陣(between-class scatter matrix)與類別內散佈矩陣(within-class scatter matrix)[16]。

證明：為了方便起見，我們先將式(3)取對數，這並不影響 $\boldsymbol{\Theta}$ 的求解：

$$\log J_{\text{GLRDA}}^{\text{homo}}(\boldsymbol{\Theta}) = \sup \log L_{H_0^{\text{homo}}}(\boldsymbol{\Theta}) - \sup \log L_{H_1^{\text{homo}}}(\boldsymbol{\Theta}) \quad (5)$$

而 $\log L_{H_0^{\text{homo}}}(\boldsymbol{\Theta})$ 和 $\log L_{H_1^{\text{homo}}}(\boldsymbol{\Theta})$ 可被進一步分別表示為樣本所有資料 \mathbf{x}_i^N 在所屬於高斯分佈之類別母體下的相似度：

$$\begin{aligned} \log L_{H_0^{\text{homo}}}(\boldsymbol{\Theta}) &= \log p(\mathbf{x}_i^N, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Theta}) \\ &= g(N, d) - \sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}})^T \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}) + \text{trace}(\tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\boldsymbol{\Sigma}}| \right) \end{aligned} \quad (6)$$

$$\begin{aligned} \log L_{H_1^{\text{homo}}}(\boldsymbol{\Theta}) &= \log p(\mathbf{x}_i^N, \{\boldsymbol{\mu}_i\}, \boldsymbol{\Sigma}, \boldsymbol{\Theta}) \\ &= g(N, d) - \sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_i)^T \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_i) + \text{trace}(\tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\boldsymbol{\Sigma}}| \right) \end{aligned} \quad (7)$$

其中， $g(N, d) = (-Nd/2) \log(2\pi)$ ， N 為樣本所有資料總數， n_i 為類別 C_i 的資料數， C 為類別總數， d 為投影後之特徵子空間的維度（或特徵數）； $\tilde{\mathbf{m}}_i$ 和 $\tilde{\mathbf{S}}_i$ 分別為經過 $\boldsymbol{\Theta}$ 轉換

後的樣本期望值向量與共變異矩陣，而 $\tilde{\boldsymbol{\mu}}$ 、 $\{\tilde{\boldsymbol{\mu}}_i\}$ 和 $\tilde{\boldsymbol{\Sigma}}$ 則是根據統計假設 H_0^{homo} 與 H_1^{homo} 而設定之經過 Θ 轉換後的母體期望值向量與共變異矩陣，即我們所要估計的參數。

欲求得在假設 H_0^{homo} 下的最大相似度估計量 $\tilde{\boldsymbol{\mu}}_0^{\text{homo}}$ 和 $\tilde{\boldsymbol{\Sigma}}_0^{\text{homo}}$ ，可將式(6)分別對 $\tilde{\boldsymbol{\mu}}$ 和 $\tilde{\boldsymbol{\Sigma}}$ 偏微分，並令其為 0，可得：

$$\begin{aligned} \frac{\partial \log L_{H_0^{\text{homo}}}(\Theta)}{\partial \tilde{\boldsymbol{\mu}}} &= \frac{\partial \left(-\sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}})^T \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}) + \text{trace}(\tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\boldsymbol{\Sigma}}| \right) \right)}{\partial \tilde{\boldsymbol{\mu}}} \\ &= -\sum_{i=1}^C n_i (\tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}})) = 0 \\ \Rightarrow \tilde{\boldsymbol{\mu}}_0^{\text{homo}} &= \sum_{i=1}^C \frac{n_i}{N} \tilde{\mathbf{m}}_i = \tilde{\mathbf{m}} \end{aligned} \quad (8)$$

$$\begin{aligned} \frac{\partial \log L_{H_0^{\text{homo}}}(\Theta)}{\partial \tilde{\boldsymbol{\Sigma}}} &= \frac{\partial \left(-\sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{homo}})^T \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{homo}}) + \text{trace}(\tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\boldsymbol{\Sigma}}| \right) \right)}{\partial \tilde{\boldsymbol{\Sigma}}} \\ &= -\sum_{i=1}^C n_i \tilde{\boldsymbol{\Sigma}}^{-1} + \sum_{i=1}^C n_i \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\mathbf{S}}_i \tilde{\boldsymbol{\Sigma}}^{-1} + \sum_{i=1}^C n_i \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}}) \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}})^T \tilde{\boldsymbol{\Sigma}}^{-1} = 0 \\ \Rightarrow \tilde{\boldsymbol{\Sigma}}_0^{\text{homo}} &= \left(\sum_{i=1}^C \frac{n_i}{N} \tilde{\mathbf{S}}_i \right) + \left(\sum_{i=1}^C \frac{n_i}{N} (\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}}) (\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}})^T \right) = \tilde{\mathbf{S}}_W + \tilde{\mathbf{S}}_B = \tilde{\mathbf{S}}_T \end{aligned} \quad (9)$$

其中 $\mathbf{S}_T \in \mathfrak{R}^{n \times n}$ 為全體散佈矩陣(total scatter matrix)[16]。將 $\tilde{\boldsymbol{\mu}}_0^{\text{homo}} = \tilde{\mathbf{m}}$ 和 $\tilde{\boldsymbol{\Sigma}}_0^{\text{homo}} = \tilde{\mathbf{S}}_T$ 代入式(6)，可得在假設 H_0^{homo} 下的最大對數相似度：

$$\begin{aligned} \sup \log L_{H_0^{\text{homo}}}(\Theta) &= \max \log p(\mathbf{x}_1^N, \tilde{\boldsymbol{\mu}}_0^{\text{homo}}, \tilde{\boldsymbol{\Sigma}}_0^{\text{homo}}, \Theta) \\ &= g(N, d) - \frac{N}{2} \log |\tilde{\mathbf{S}}_T| - \frac{Nd}{2} \end{aligned} \quad (10)$$

同理，欲求得在假設 H_1^{homo} 下的最大相似度估計量 $\{\tilde{\boldsymbol{\mu}}_{1,i}^{\text{homo}}\}$ 和 $\tilde{\boldsymbol{\Sigma}}_1^{\text{homo}}$ ，可將式(7)分別對 $\tilde{\boldsymbol{\mu}}_i$ 和 $\tilde{\boldsymbol{\Sigma}}$ 偏微分，並令其為 0，可得：

$$\begin{aligned} \frac{\partial \log L_{H_1^{\text{homo}}}(\Theta)}{\partial \tilde{\boldsymbol{\mu}}_i} &= \frac{\partial \left(-\sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_i)^T \tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_i) + \text{trace}(\tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\boldsymbol{\Sigma}}| \right) \right)}{\partial \tilde{\boldsymbol{\mu}}_i} \\ &= n_i (\tilde{\boldsymbol{\Sigma}}^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_i)) = 0 \\ \Rightarrow \tilde{\boldsymbol{\mu}}_{1,i}^{\text{homo}} &= \tilde{\mathbf{m}}_i \end{aligned} \quad (11)$$

$$\begin{aligned}
\frac{\partial \log L_{H_1^{\text{homo}}}(\Theta)}{\partial \tilde{\Sigma}} &= \frac{\partial \left(-\sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_{1,i}^{\text{homo}})^T \tilde{\Sigma}^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_{1,i}^{\text{homo}}) + \text{trace}(\tilde{\Sigma}^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\Sigma}| \right) \right)}{\partial \tilde{\Sigma}} \\
&= -\sum_{i=1}^C n_i \tilde{\Sigma}^{-1} + \sum_{i=1}^C n_i \tilde{\Sigma}^{-1} \tilde{\mathbf{S}}_i \tilde{\Sigma}^{-1} = 0 \\
\Rightarrow \tilde{\Sigma}_1^{\text{homo}} &= \left(\sum_{i=1}^C \frac{n_i}{N} \tilde{\mathbf{S}}_i \right) = \tilde{\mathbf{S}}_W
\end{aligned} \tag{12}$$

將 $\tilde{\boldsymbol{\mu}}_{1,i}^{\text{homo}} = \tilde{\mathbf{m}}_i$ 和 $\tilde{\Sigma}_1^{\text{homo}} = \tilde{\mathbf{S}}_W$ 代入式(7)，可得在假設 H_1^{homo} 下的最大對數相似度：

$$\begin{aligned}
\sup \log L_{H_1^{\text{homo}}}(\Theta) &= \max \log p(\mathbf{x}_1^N, \{\tilde{\boldsymbol{\mu}}_{1,i}^{\text{homo}}\}, \tilde{\Sigma}_1^{\text{homo}}, \Theta) \\
&= g(N, d) - \frac{N}{2} \log |\tilde{\mathbf{S}}_W| - \frac{Nd}{2}
\end{aligned} \tag{13}$$

最後，將式(10)與式(13)代入式(5)，可得在同方差性假設下的對數相似度比率：

$$\begin{aligned}
&\log J_{\text{GLRDA}}^{\text{homo}}(\Theta) \\
&= \left(g(N, d) - \frac{N}{2} \log |\tilde{\mathbf{S}}_T| - \frac{Nd}{2} \right) - \left(g(N, d) - \frac{N}{2} \log |\tilde{\mathbf{S}}_W| - \frac{Nd}{2} \right) \\
&= \frac{N}{2} (\log |\tilde{\mathbf{S}}_W| - \log |\tilde{\mathbf{S}}_T|) = \frac{N}{2} (\log |\tilde{\mathbf{S}}_W| - \log(|\tilde{\mathbf{S}}_B| + |\tilde{\mathbf{S}}_W|)) \\
&= \frac{N}{2} \log \frac{|\tilde{\mathbf{S}}_W|}{|\tilde{\mathbf{S}}_B| + |\tilde{\mathbf{S}}_W|} = \frac{N}{2} \log \frac{1}{\frac{|\tilde{\mathbf{S}}_B|}{|\tilde{\mathbf{S}}_W|} + 1}
\end{aligned} \tag{14}$$

Θ 可經由最小化式(14)來求出。因為對數函數為單調遞增(monotonically increasing)函數，所以 Θ 也可藉由最大化式(14)中的 $|\Theta^T \mathbf{S}_B \Theta| / |\Theta^T \mathbf{S}_W \Theta|$ 求得，而此項正好是線性鑑別分析(LDA)的目標函式（式(4)）。 ■

(四) 異方差性(Heteroscedasticity)

現在，我們考慮異方差性的統計假設[17]：

$$\begin{cases} H_0^{\text{heter}} : \text{每一類別 } C_i \text{ 均呈高斯分布，且 } \boldsymbol{\mu}_i = \boldsymbol{\mu}, \boldsymbol{\Sigma}_i \text{ 不受任何限制。} \\ H_1^{\text{heter}} : \text{每一類別 } C_i \text{ 均呈高斯分布，且 } \boldsymbol{\mu}_i \text{ 與 } \boldsymbol{\Sigma}_i \text{ 均不受任何限制。} \end{cases}$$

仿照命題一的方式，普遍化相似度比率鑑別分析(GLRDA)的目標函式可寫成：

$$\log J_{\text{GLRDA}}^{\text{heter}}(\Theta) = \sup \log L_{H_0^{\text{heter}}}(\Theta) - \sup \log L_{H_1^{\text{heter}}}(\Theta) \tag{15}$$

而 $\log L_{H_0^{\text{heter}}}(\Theta)$ 和 $\log L_{H_1^{\text{heter}}}(\Theta)$ 可被分別進一步表示為

$$\begin{aligned}
\log L_{H_0^{\text{heter}}}(\Theta) &= \log p(\mathbf{x}_1^N, \boldsymbol{\mu}, \{\boldsymbol{\Sigma}_i\}, \Theta) \\
&= g(N, d) - \sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}})^T \tilde{\boldsymbol{\Sigma}}_i^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}) + \text{trace}(\tilde{\boldsymbol{\Sigma}}_i^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\boldsymbol{\Sigma}}_i| \right)
\end{aligned} \tag{16}$$

$$\begin{aligned}
\log L_{H_1^{\text{heter}}}(\Theta) &= \log p(\mathbf{x}_1^N, \{\boldsymbol{\mu}_i\}, \{\boldsymbol{\Sigma}_i\}, \Theta) \\
&= g(N, d) - \sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_i)^T \tilde{\boldsymbol{\Sigma}}_i^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_i) + \text{trace}(\tilde{\boldsymbol{\Sigma}}_i^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\boldsymbol{\Sigma}}_i| \right)
\end{aligned} \tag{17}$$

欲求得在假設 H_0^{heter} 下的最大相似度估計量 $\tilde{\boldsymbol{\mu}}_0^{\text{heter}}$ 和 $\tilde{\boldsymbol{\Sigma}}_{0,i}^{\text{heter}}$ ，可將式(16)分別對 $\tilde{\boldsymbol{\mu}}$ 和 $\tilde{\boldsymbol{\Sigma}}_i$ 偏微分，並令其為 0，可得：

$$\begin{aligned}
\frac{\partial \log L_{H_0^{\text{heter}}}(\Theta)}{\partial \tilde{\boldsymbol{\mu}}} &= \frac{\partial \left(- \sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}})^T \tilde{\boldsymbol{\Sigma}}_i^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}) + \text{trace}(\tilde{\boldsymbol{\Sigma}}_i^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\boldsymbol{\Sigma}}_i| \right) \right)}{\partial \tilde{\boldsymbol{\mu}}} \\
&= - \sum_{i=1}^C n_i \left(\tilde{\boldsymbol{\Sigma}}_i^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}) \right) = 0
\end{aligned} \tag{18}$$

$$\Rightarrow \tilde{\boldsymbol{\mu}}_0^{\text{heter}} = \left(\sum_{i=1}^C n_i \tilde{\boldsymbol{\Sigma}}_i^{-1} \right)^{-1} \sum_{i=1}^C n_i \tilde{\boldsymbol{\Sigma}}_i^{-1} \tilde{\mathbf{m}}_i \approx \left(\sum_{i=1}^C n_i \tilde{\mathbf{S}}_i^{-1} \right)^{-1} \sum_{i=1}^C n_i \tilde{\mathbf{S}}_i^{-1} \tilde{\mathbf{m}}_i$$

$$\begin{aligned}
\frac{\partial \log L_{H_0^{\text{heter}}}(\Theta)}{\partial \tilde{\boldsymbol{\Sigma}}_i} &= \frac{\partial \left(- \sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{heter}})^T \tilde{\boldsymbol{\Sigma}}_i^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{heter}}) + \text{trace}(\tilde{\boldsymbol{\Sigma}}_i^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\boldsymbol{\Sigma}}_i| \right) \right)}{\partial \tilde{\boldsymbol{\Sigma}}_i} \\
&= - \frac{1}{2} n_i \left(- \tilde{\boldsymbol{\Sigma}}_i^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{heter}}) (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{heter}})^T \tilde{\boldsymbol{\Sigma}}_i^{-1} - \tilde{\boldsymbol{\Sigma}}_i^{-1} \tilde{\mathbf{S}}_i \tilde{\boldsymbol{\Sigma}}_i^{-1} + \tilde{\boldsymbol{\Sigma}}_i^{-1} \right) = 0 \\
\Rightarrow \tilde{\boldsymbol{\Sigma}}_{0,i}^{\text{heter}} &= (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{heter}}) (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{heter}})^T + \tilde{\mathbf{S}}_i = \tilde{\mathbf{B}}_i + \tilde{\mathbf{S}}_i
\end{aligned} \tag{19}$$

其中， $\tilde{\mathbf{B}}_i = (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{heter}}) (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{heter}})^T$ 。值得一提的是，在式(18)中， $\tilde{\boldsymbol{\mu}}$ 的形式中含有尚未估計出的 $\tilde{\boldsymbol{\Sigma}}_i$ ，因此我們只能先令 $\tilde{\boldsymbol{\Sigma}}_i = \tilde{\mathbf{S}}_i$ ，得到 $\tilde{\boldsymbol{\mu}}$ 的近似估計量。將式(18)和式(19)中估計出的 $\tilde{\boldsymbol{\mu}}_0^{\text{heter}}$ 和 $\tilde{\boldsymbol{\Sigma}}_{0,i}^{\text{heter}}$ 代入式(16)，可得在假設 H_0^{heter} 下的最大對數相似度：

$$\begin{aligned}
\sup \log L_{H_0^{\text{heter}}}(\Theta) &= \max \log p(\mathbf{x}_1^N, \tilde{\boldsymbol{\mu}}_0^{\text{heter}}, \{\tilde{\boldsymbol{\Sigma}}_{0,i}^{\text{heter}}\}, \Theta) \\
&= g(N, d) - \frac{Nd}{2} - \sum_{i=1}^C \frac{n_i}{2} \log |\tilde{\mathbf{B}}_i + \tilde{\mathbf{S}}_i|
\end{aligned} \tag{20}$$

欲求得在假設 H_1^{heter} 下的最大相似度估計量 $\tilde{\boldsymbol{\mu}}_i^{\text{heter}}$ 和 $\tilde{\boldsymbol{\Sigma}}_{1,i}^{\text{heter}}$ ，可將式(17)分別對 $\tilde{\boldsymbol{\mu}}_i$ 和 $\tilde{\boldsymbol{\Sigma}}_i$ 偏微分，並令其為 0，可得：

$$\begin{aligned}
\frac{\partial \log L_{H_1^{\text{heter}}}(\Theta)}{\partial \tilde{\boldsymbol{\mu}}_i} &= \frac{\partial \left(- \sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_i)^T \tilde{\boldsymbol{\Sigma}}_i^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_i) + \text{trace}(\tilde{\boldsymbol{\Sigma}}_i^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\boldsymbol{\Sigma}}_i| \right) \right)}{\partial \tilde{\boldsymbol{\mu}}_i} \\
&= n_i \tilde{\boldsymbol{\Sigma}}_i^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_i) = 0 \\
\Rightarrow \tilde{\boldsymbol{\mu}}_{1,i}^{\text{heter}} &= \tilde{\mathbf{m}}_i
\end{aligned} \tag{21}$$

$$\begin{aligned}
\frac{\partial \log L_{H_1^{\text{heter}}}(\Theta)}{\partial \tilde{\Sigma}_i} &= \frac{\partial \left(-\sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_{1,i}^{\text{heter}})^T \tilde{\Sigma}_i^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_{1,i}^{\text{heter}}) + \text{trace}(\tilde{\Sigma}_i^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\Sigma}_i| \right) \right)}{\partial \tilde{\Sigma}_i} \\
&= -\frac{n_i}{2} \left(-\tilde{\Sigma}_i^{-1} \tilde{\mathbf{S}}_i \tilde{\Sigma}_i^{-1} + \tilde{\Sigma}_i^{-1} \right) = 0 \\
&\Rightarrow \tilde{\Sigma}_{1,i}^{\text{heter}} = \tilde{\mathbf{S}}_i
\end{aligned} \tag{22}$$

將式(21)和式(22)中估計出的 $\tilde{\boldsymbol{\mu}}_1^{\text{heter}}$ 和 $\tilde{\Sigma}_{1,i}^{\text{heter}}$ 代入式(17)，可得在假設 H_1^{heter} 下的最大對數相似度：

$$\begin{aligned}
\sup \log L_{H_1^{\text{heter}}}(\Theta) &= \max \log p(\mathbf{x}_1^N, \{\tilde{\boldsymbol{\mu}}_1^{\text{heter}}\}, \{\tilde{\Sigma}_{1,i}^{\text{heter}}\}, \Theta) \\
&= g(N, d) - \sum_{i=1}^C \frac{n_i}{2} \left((\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}}_i)^T \tilde{\Sigma}_i^{-1} (\tilde{\mathbf{m}}_i - \tilde{\mathbf{m}}_i) + \text{trace}(\tilde{\Sigma}_i^{-1} \tilde{\mathbf{S}}_i) + \log |\tilde{\Sigma}_i| \right) \\
&= g(N, d) - \sum_{i=1}^C \frac{n_i}{2} (d + \log |\tilde{\Sigma}_i|) \\
&= g(N, d) - \frac{Nd}{2} - \sum_{i=1}^C \frac{n_i}{2} \log |\tilde{\Sigma}_i|
\end{aligned} \tag{23}$$

最後，將式(20)與式(23)代入式(15)，經過整理，可得：

$$\begin{aligned}
\log J_{\text{GLRDA}}^{\text{heter}}(\Theta) &= \left(g(N, d) - \frac{Nd}{2} - \sum_{i=1}^C \frac{n_i}{2} \log |\tilde{\mathbf{B}}_i + \tilde{\mathbf{S}}_i| \right) - \left(g(N, d) - \frac{Nd}{2} - \sum_{i=1}^C \frac{n_i}{2} \log |\tilde{\Sigma}_i| \right) \\
&= -\sum_{i=1}^C \frac{n_i}{2} (\log |\tilde{\mathbf{B}}_i + \tilde{\mathbf{S}}_i| - \log |\tilde{\Sigma}_i|) = -\sum_{i=1}^C \frac{n_i}{2} \log |\mathbf{I}_{(p \times p)} + \tilde{\Sigma}_i^{-1} \tilde{\mathbf{B}}_i| \\
&= -\sum_{i=1}^C \frac{n_i}{2} \log |\mathbf{I}_{(p \times p)} + \tilde{\Sigma}_i^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{heter}}) (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{heter}})^T| \\
&= -\sum_{i=1}^C \frac{n_i}{2} \log \left(1 + (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{heter}})^T \tilde{\Sigma}_i^{-1} (\tilde{\mathbf{m}}_i - \tilde{\boldsymbol{\mu}}_0^{\text{heter}}) \right)
\end{aligned} \tag{24}$$

因此，我們可得到異方差性之普遍化相似度比率鑑別分析(GLRDA)的目標函式：

$$G_H(\Theta) = -\sum_{i=1}^C \frac{n_i}{2} \log \left(1 + (\Theta^T \mathbf{m}_i - \Theta^T \boldsymbol{\mu}_0^{\text{heter}})^T (\Theta^T \mathbf{S}_i \Theta)^{-1} (\Theta^T \mathbf{m}_i - \Theta^T \boldsymbol{\mu}_0^{\text{heter}}) \right) \tag{25}$$

爲了使用梯度下降等遞迴式的最佳化技術求解 Θ ，式(25)對 Θ 的一階偏導數可寫成：

$$\frac{\partial G_H(\Theta)}{\partial \Theta} = -\sum_{i=1}^C n_i \frac{(-\mathbf{S}_i \Theta \tilde{\Sigma}_i^{-1} \tilde{\mathbf{B}}_i + \mathbf{B}_i \Theta) \tilde{\Sigma}_i^{-1}}{1 + \text{trace}(\tilde{\Sigma}_i^{-1} \tilde{\mathbf{B}}_i)} \tag{26}$$

其中， $\mathbf{B}_i = (\mathbf{m}_i - \boldsymbol{\mu}_0^{\text{heter}})(\mathbf{m}_i - \boldsymbol{\mu}_0^{\text{heter}})^T$ ， $\tilde{\mathbf{B}}_i = \Theta^T \mathbf{B}_i \Theta$ ， $\tilde{\Sigma}_i = \Theta^T \mathbf{S}_i \Theta$ 。

(五) 討論與比較

表一、普遍化相似度比率鑑別分析(GLRDA)在不同假設下的統計量歸納表

統計假設	期望值向量 估計量	共變異矩陣 估計量	含重要項之 最大對數相似度
$H_0^{\text{homo}} \begin{cases} \Sigma_i = \Sigma \\ \mu_i = \mu \end{cases}$	$\tilde{\mathbf{m}}$	$\tilde{\mathbf{S}}_T$	$-\frac{N}{2} \log \tilde{\mathbf{S}}_T $
$H_1^{\text{homo}} \begin{cases} \Sigma_i = \Sigma \\ \mu_i: \text{無限制} \end{cases}$	$\tilde{\mathbf{m}}_i$	$\tilde{\mathbf{S}}_W$	$-\frac{N}{2} \log \tilde{\mathbf{S}}_W $
$H_0^{\text{heter}} \begin{cases} \Sigma_i: \text{無限制} \\ \mu_i = \mu \end{cases}$	$\left(\sum_{i=1}^C n_i \tilde{\mathbf{S}}_i^{-1} \right)^{-1} \sum_{i=1}^C n_i \tilde{\mathbf{S}}_i^{-1} \tilde{\mathbf{m}}_i$	$\tilde{\mathbf{B}}_i + \tilde{\mathbf{S}}_i$	$-\sum_{i=1}^C \frac{n_i}{2} \log \tilde{\mathbf{B}}_i + \tilde{\mathbf{S}}_i $
$H_1^{\text{heter}} \begin{cases} \Sigma_i: \text{無限制} \\ \mu_i: \text{無限制} \end{cases}$	$\tilde{\mathbf{m}}_i$	$\tilde{\mathbf{S}}_i$	$-\sum_{i=1}^C \frac{n_i}{2} \log \tilde{\mathbf{S}}_i $

表一歸納了普遍化相似度比率鑑別分析(GLRDA)在不同假設下的統計量。它是一個較大之相似度比率的框架，不僅是線性鑑別分析(LDA)的普遍化形式（其轉換矩陣可由 H_0^{homo} 與 H_1^{homo} 的相似度比率得到），以下命題亦證明了它也是異方差性線性鑑別分析(HLDA)的普遍化形式。

命題二：異方差線性鑑別分析(HLDA)轉換矩陣可由普遍化相似度比率鑑別分析(GLRDA)中的 H_0^{homo} 與 H_1^{heter} 的相似度比率得到。也就是說，普遍化相似度比率鑑別分析(GLRDA)是異方差線性鑑別分析(HLDA)的普遍化形式。

證明：根據[6]，HLDA 目標函式為

$$J_{\text{HLDA}}(\Theta) = -\frac{N}{2} \log |\Theta_{(n-d)}^T \mathbf{S}_T \Theta_{(n-d)}| - \sum_{i=1}^C \frac{n_i}{2} \log |\Theta_d^T \mathbf{S}_i \Theta_d| + N \log |\Theta| \quad (27)$$

因為在此， Θ 為全秩矩陣(full-rank matrix)，且 $\Theta_{(n \times n)} = [\Theta_d, \Theta_{(n-d)}]$ ，我們可證明[18]

$$\begin{aligned} |\Theta^T \mathbf{S}_T \Theta| &= |\Theta_d^T \mathbf{S}_T \Theta_d| \times |\Theta_{(n-d)}^T \mathbf{S}_T \Theta_{(n-d)}| \\ \Rightarrow \log |\Theta^T \mathbf{S}_T \Theta| &= \log |\Theta_d^T \mathbf{S}_T \Theta_d| + \log |\Theta_{(n-d)}^T \mathbf{S}_T \Theta_{(n-d)}| \\ \Rightarrow \log |\Theta_{(n-d)}^T \mathbf{S}_T \Theta_{(n-d)}| &= \log |\Theta^T \mathbf{S}_T \Theta| - \log |\Theta_d^T \mathbf{S}_T \Theta_d| \end{aligned} \quad (28)$$

$$\begin{aligned} N \log |\Theta| - \frac{N}{2} \log |\Theta^T \mathbf{S}_T \Theta| \\ = N \log |\Theta| - \frac{N}{2} \log |\Theta| - \frac{N}{2} \log |\mathbf{S}_T| - \frac{N}{2} \log |\Theta| \\ = -\frac{N}{2} \log |\mathbf{S}_T| \end{aligned} \quad (29)$$

將式(28)代入式(27)，且考慮式(29)所推導出的結果，我們可得出異方差線性鑑別分析(HLDA)的另一種目標函式表示法：

表二、MATBN 訓練語料之音素辨識中前 10 組最易混淆之音素模型配對

K	類別 (音素) 配對	(RCD 模型)	錯誤音框數
1	in (ㄨㄣ)	ing (ㄨㄥ)	66,353
2	an (ㄤ)	eng (ㄥ)	42,550
3	i (ㄧ)	sil (靜音)	31,796
4	u (ㄨ)	sil (靜音)	29,082
5	sic_e (ㄛ的空聲母)	sil (靜音)	26,134
6	sic_i (ㄨ的空聲母)	sil (靜音)	25,709
7	ing (ㄨㄥ)	sil (靜音)	21,629
8	g_u (ㄍㄨ)	sil (靜音)	19,197
9	ian (ㄧㄢ)	ie (ㄧㄝ)	17,212
10	sic_i (ㄨ的空聲母)	i (ㄧ)	17,022

$$J_{\text{HLDA}}(\Theta) = - \underbrace{\sum_{i=1}^C \frac{n_i}{2} \log |\Theta_d^T \mathbf{S}_i \Theta_d|}_{\text{sup log } L_{H_1^{\text{heter}}}(\Theta)} - \left(- \frac{N}{2} \log |\Theta_d^T \mathbf{S}_T \Theta_d| \right) - \frac{N}{2} \log |\mathbf{S}_T| \quad (30)$$

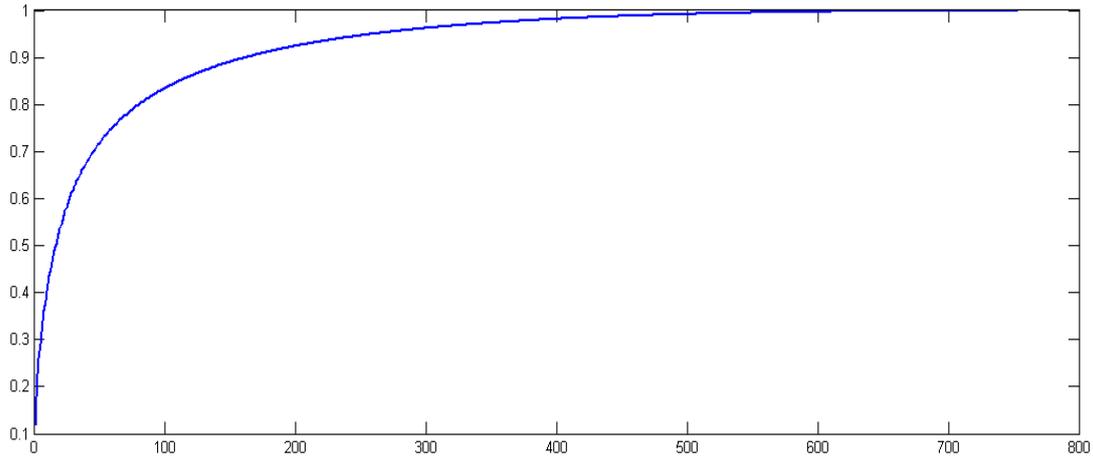
很明顯地，若不考慮常數 $(-N/2) \log |\mathbf{S}_T|$ ，則式(30)等同於普遍化相似度比率鑑別分析 (GLRDA) 中， H_0^{homo} 與 H_1^{heter} 含重要項之最大對數相似度比率（見表一）。 ■

由命題二可看出，異方差線性鑑別分析(HLDA)與異方差性之普遍化相似度比率鑑別分析(GLRDA)的主要差別在於虛無假設 H_0 的設定，異方差線性鑑別分析(HLDA)的虛無假設 H_0^{homo} 較為嚴格，相對地要使它盡可能不發生的難度也較低。反之，異方差性之普遍化相似度比率鑑別分析(GLRDA)的虛無假設 H_0^{heter} 所產生的參數空間就比較大，應會使得找出的投影子空間在類別鑑別性上就較為強健(robust)。

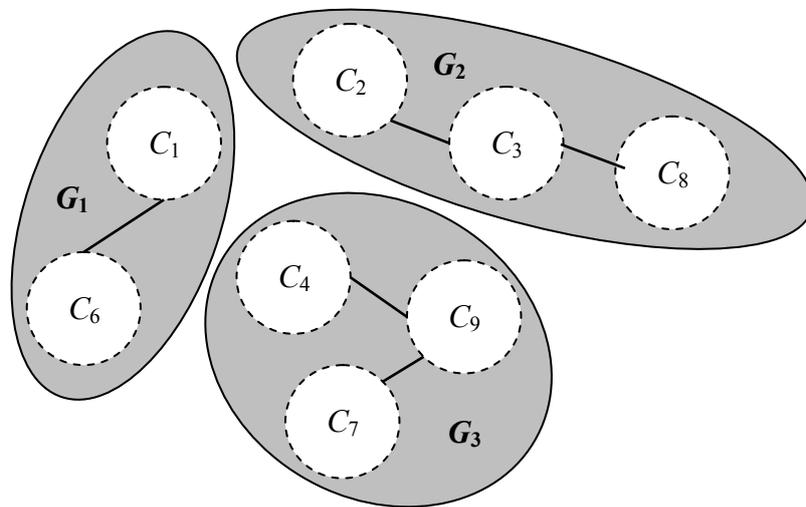
三、混淆資訊的延伸

由異方差性之普遍化相似度比率鑑別分析(GLRDA)的虛無假設 H_0^{heter} 來看，它設想每一類別母體的期望值向量在投影子空間中幾乎重疊在一起，但事實上這個假設的設定未必精確。在使用線性鑑別分析(LDA)作為聲學特徵擷取方法之 MATBN 訓練語料（參見第四節）的音素辨識結果中，我們將所有類別配對，依其混淆程度由大至小排序，可得到前 K 個類別配對，如表二。舉例來說，在表二中，最易於混淆的音素類別配對是(in, ing)，其錯誤音框數為 66,353，表示原本屬於音素 in 和 ing，卻分別被辨識器錯分至音素 ing 和 in 的音框總數。也就是說，以經驗資訊產生的事實來看，我們最不願意見到的假設，就是使這些易於混淆之類別配對的期望值向量幾乎重疊在一起，而非更廣泛的假設全體類別母體之期望值向量重疊，因為這些易於混淆之類別配對才是全體錯誤的主要來源。再者，從圖一中我們可以發現約前 10% 易於混淆之類別配對主導了約 80% 的錯誤音框總數。

但是，若我們只考慮這些類別配對之期望值向量重疊的假設，則會發生以下情況：



圖一、前 K 組易於混淆之類別配對與累積錯誤音框比率圖
(橫軸為 $K/10$ ，縱軸為累積錯誤音框比率)



圖二、類別配對與群聚形成示意圖

若類別配對 C_1 與 C_2 為最混淆之配對，則虛無假設可設定為 $\mu_1 = \mu_2$ ；而類別配對 C_2 與 C_3 為次混淆之配對，則虛無假設可增加 $\mu_2 = \mu_3$ 。因此，整個虛無假設可合併為 $\mu_1 = \mu_2 = \mu_3$ 。也就是說，我們必須在類別配對集合中找到所有相關的類別配對以組成混淆群聚(confusable cluster)，如圖二。若我們把所有類別視為圖形(graph)中的點(vertex)，而由易於混淆之類別配對所建立的關係視為兩點之間的邊(edge)，則混淆群聚的產生可被視為尋找圖形(graph)中所有的連通子圖(connected subgraph)。所以，我們可以使用一些圖論中的演算法，如滿水填充演算法(flood fill algorithm)[19]，來解決這個問題。

因此，我們可以將異方差性之普遍化相似度比率鑑別分析(GLRDA)改良成基於混淆資訊之普遍化相似度比率鑑別分析(confusion information based GLRDA, CI-GLRDA)：令 $G: \{G_k\}$ 為所有根據前 K 組易於混淆之類別配對，並利用滿水填充演算法求出之群聚的集合，則其虛無假設與對立假設可設定如下：

表三、異方差性之普遍化相似度比率鑑別分析在不同期望值估計下之正確率(%)

GLRDA	Without MLLT	With MLLT
權重平均(weighted mean)	62.34	74.88
算術平均(arithmetic mean)	58.68	74.45

$$\begin{cases} H_0^{Cl} : \text{每一類別 } C_i \text{ 均呈高斯分布，且 } \Sigma_i \text{ 不受任何限制，而若 } C_i \in G_{l_i} \text{，則 } \mu_i = \mu_{l_i} \text{。} \\ H_1^{Cl} : \text{每一類別 } C_i \text{ 均呈高斯分布，且 } \mu_i \text{ 與 } \Sigma_i \text{ 均不受任何限制。} \end{cases}$$

其中， l_i 為群聚編號，用來標示類別 C_i 所屬的群聚。

類似於上一節的最大化相似度估計法，我們可得到基於混淆資訊之普遍化相似度比率鑑別分析(confusion information based GLRDA, CI-GLRDA)目標函式：

$$G_{CI}(\Theta) = - \sum_{i=1, G_{l_i} \in G}^C \frac{n_i}{2} \log \left(1 + (\Theta^T \mathbf{m}_i - \Theta^T \mu_{l_i}^{Cl})^T (\Theta^T \mathbf{S}_i \Theta)^{-1} (\Theta^T \mathbf{m}_i - \Theta^T \mu_{l_i}^{Cl}) \right) \quad (31)$$

而式(31)對 Θ 的一階偏導數亦可表示成：

$$\frac{\partial G_{CI}(\Theta)}{\partial \Theta} = - \sum_{i=1, G_{l_i} \in G}^C n_i \frac{(-\mathbf{S}_i \Theta \tilde{\mathbf{S}}_i^{-1} \tilde{\mathbf{B}}_{l_i} + \mathbf{B}_{l_i} \Theta) \tilde{\mathbf{S}}_i^{-1}}{1 + \text{trace}(\tilde{\mathbf{S}}_i^{-1} \tilde{\mathbf{B}}_{l_i})} \quad (32)$$

其中， $\mathbf{B}_{l_i} = (\mathbf{m}_i - \mu_{l_i}^{Cl})(\mathbf{m}_i - \mu_{l_i}^{Cl})^T$ ， $\tilde{\mathbf{B}}_{l_i} = \Theta^T \mathbf{B}_{l_i} \Theta$ ， $\tilde{\mathbf{S}}_i = \Theta^T \mathbf{S}_i \Theta$ 。

四、實驗結果與分析

(一) 實驗設定

本論文主要使用的語料庫為 MATBN 中文電視新聞語料[20]，內含外場記者的語料總共約 27 小時，其中 24.5 小時 (5,774 句，再切成 34,672 個短句供聲學模型訓練之用) 做為聲學模型訓練的語料，1 小時 (230 句) 為辨識評估的資料，另有 1.5 小時 (292 句) 則為發展集(developing set)，用來決定特殊參數的調整，如梯度下降等最佳化方法的遞迴次數或在上節中提到的 K 。

在本論文中使用梅爾倒頻譜係數(Mel-frequency cepstral coefficients, MFCCs)作為最基本的語音特徵參數。在聲學模型部分，我們將每個中文字視為由一個聲母與一個韻母組成，採用傳統式由左至右的隱藏式馬可夫模型(hidden Markov models, HMMs)，分別為聲母及韻母建立 INITIAL 與 FINAL 模型，並且考慮聲母會受右相連韻母影響其發音特性，所以採用右相關聯模型，(right-context-dependent model, RCD model)，加上一個靜音(silence)模型，總共有 151 個聲學模型[21]。聲學模型會透過 EM 演算法(expectation-maximization algorithm)，經過 10 次遞迴的最大化相似度訓練而得。而在語言模型方面，我們使用了詞二連以及詞三連語言模型(word bigram and trigram language models)，並以從中央通訊社(Central News Agency, CNA)2001 與 2002 年所收集到的約一億七千萬個中文字語料作為背景語言模型訓練時的訓練資料。本論文中的語言模型訓

表四、各種特徵擷取方法在中文大詞彙辨識系統下之詞正確率(%)

方法	Without MLLT	With MLLT
線性鑑別分析(LDA)	71.46	74.33
異方差性線性鑑別分析(HLDA)	70.28	74.88
異方差性鑑別分析(HDA)	71.36	74.53
異方差性之普遍化相似度比率鑑別分析 (Heteroscedastic GLRDA)	62.34	74.88
基於混淆資訊之普遍化相似度比率鑑別 分析(CI-GLRDA)	63.62	75.26

練工具採用 SRI Language Modeling Toolkit (SRILM)[22]。就以梅爾倒頻譜係數(MFCCs)為基礎實驗(baseline)而言，其詞辨識正確率(character accuracy)為 72.23%。

(二) 實驗結果

本論文中所提到的重要特徵抽取方法，如線性鑑別分析(LDA)、異方差性線性鑑別分析(HLDA)、異方差性鑑別分析(HDA)，以及我們所提出的普遍化相似度比率鑑別分析(GLRDA)，都是作用在 162 維($n = 162$)的超級向量(super-vector)上進行降維處理。此超級向量是由連續 9 個音框之梅氏濾波器組(Mel-frequency filterbank)所輸出的 18 維特徵向量串接而成，目的在於捕捉音框間的動態資訊。而目標維度則設定為 39 維($d = 39$)，其目的則在於使我們能在子空間維度固定的情況下，定性地比較各種方法的優劣。分類的最小單位則是以隱藏式馬可夫模型(HMMs)中的狀態(state)為主，並經由一個辨識效果較高的系統針對每一訓練語句進行強制校準(forced alignment)，從而產生語句中的類別（音素和狀態）分界。而混淆資訊的獲得即以此為正確答案，針對每一音框進行類別比對而得。由於這些特徵轉換方法所擷取出的特徵向量並不會使得各個類別的共變異矩陣為對角化，會造成後端隱藏式馬可夫模型(HMMs)參數的估計失真，所以我們嘗試在這些方法後各自加上最大化相似度線性轉換(maximum likelihood linear transformation, MLLT)[23]。

在異方差性之普遍化相似度比率鑑別分析(GLRDA)的部分，由於在目標函式(25)中， μ_0^{heter} 的部分是根據近似估計來的，我們在此採取了兩種近似方式，一種就是式(18)所提到的權重平均(weighted mean)，表示如下：

$$\mu_0^{\text{heter}} = \left(\sum_{i=1}^C n_i \mathbf{S}_i^{-1} \right)^{-1} \sum_{i=1}^C n_i \mathbf{S}_i^{-1} \mathbf{m}_i \quad (33)$$

而另一種則是全體資料的算術平均(arithmetic mean)，表示如下：

$$\mu_0^{\text{heter}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \quad (34)$$

表三顯示了異方差性之普遍化相似度比率鑑別分析(GLRDA)在不同期望值估計量下之詞正確率。我們可以看出，以權重平均作為 μ_0^{heter} 估計量的效果較好，因此，在之後基於混淆資訊的普遍化相似度比率鑑別分析(CI-GLRDA)，我們也會採用同樣的設定。此外，

我們也可以發現未經過 MLLT 處理的異方差性之普遍化相似度比率鑑別分析 (GLRDA)，辨識率仍偏低。這是因為根據其標準函式求出的轉換矩陣，並沒有使得每一類別的共變異矩陣趨於對角化的效果。

表四顯示出各種特徵擷取方法在大詞彙連續語音辨識之詞正確率。其中，異方差性鑑別分析(HDA)的目標函式可表示為：

$$J_{\text{HDA}}(\Theta) = \sum_{i=1}^C n_i \log |\Theta^T \mathbf{S}_i \Theta| + N \log |\Theta^T \mathbf{S}_B \Theta| \quad (35)$$

我們可以發現，各種針對線性鑑別分析(LDA)的改進方法均在辨識率上有些微進步。值得注意的是，在未加 MLLT 技術的情況，以線性鑑別分析(LDA)為基礎的方法並不會優於梅爾倒頻譜係數 (MFCCs)，這也驗證了在大詞彙語音辨識工作上，以線性鑑別分析(LDA)為基礎的方法會因著語料的不同而有或好或差的結果[24-25]。而在加入了 MLLT 技術之後，虛無假設內具有異方差性的方法，如異方差性線性鑑別分析(HLDA)與異方差性之普遍化相似度比率鑑別分析(Heteroscedastic GLRDA)，會有一致的實驗結果 (74.88%)，也都能看出打破傳統線性鑑別分析(LDA)中同方差性的效果。當我們進一步地加入混淆資訊後，並由發展集語料決定出最佳的 K 值 ($K = 100$)，辨識效果更好 (75.26%)，也證明了以經驗資訊作為輔助，根據類別的特性設計出適當的虛無假設是有助於分類的。

五、結論與未來展望

本文的主要貢獻在於，基於相似度比率檢驗(LRT)的另一種涵義下，提出了嶄新的、更普遍化的框架進行鑑別式特徵擷取。我們的方法不僅可以使用在語音處理上，也能夠應用在其他需要特徵擷取的領域中，而獲得更具鑑別性的特徵。此外，我們的方法也能與類別混淆資訊結合，使其成為理論與經驗兼具的方法。

普遍化相似度比率鑑別分析(GLRDA)的確具有進一步研究的空間：未來我們會嘗試各種將混淆類別分群的方法，找出屬於語音資料中，最混淆或使得辨識率最差的虛無假設表示法，使其更具分類上的代表性。

六、致謝

本研究承蒙國科會研究計畫 NSC 98-2221-E-003-011-MY3、NSC 96-2628-E-003-015-MY3、NSC 97-2631-S-003-003 的部分補助，僅此致謝。

參考文獻

- [1] B. D. Ripley, *Pattern Recognition and Neural Networks*. New York: Cambridge University Press, 1996.
- [2] X. Wang and K. K. Paliwal, "Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition," *Pattern Recognition*, vol. 36, pp.

- 2429-2439, 2003.
- [3] D. Povey, *et al.*, "fMPE: discriminatively trained features for speech recognition," in *Proc. ICASSP*, 2005, pp. 961-964.
 - [4] X.-B. Li, *et al.*, "Dimensionality reduction using MCE-optimized LDA transformation," in *Proc. ICASSP*, 2004, pp. 137-140.
 - [5] R. A. Fisher, "The statistical utilization of multiple measurements," *Annals of Eugenics*, vol. 8, pp. 376-386, 1938.
 - [6] N. Kumar and A. G. Andreou, "Heteroscedastic discriminant analysis and reduced rank HMMs for improved speech recognition," *Speech Communication*, vol. 26, pp. 283-297, 1998.
 - [7] G. Saon, *et al.*, "Maximum likelihood discriminant feature spaces," in *Proc. ICASSP*, 2000, pp. 1129-1132.
 - [8] K. Demuynck, *et al.*, "Optimal feature sub-space selection based on discriminant analysis" in *Proc. Eurospeech*, 1999, pp. 1311-1314.
 - [9] H.-S. Lee and B. Chen, "Linear discriminant feature extraction using weighted classification confusion information," in *Proc. Interspeech*, 2008, pp. 2254-2257.
 - [10] H.-S. Lee and B. Chen, "Improved linear discriminant analysis considering empirical pairwise classification error rates," in *Proc. ISCSLP*, 2008, pp. 149-152.
 - [11] H.-S. Lee and B. Chen, "Empirical error rate minimization based linear discriminant analysis," in *Proc. ICASSP*, 2009.
 - [12] X. Cui, *et al.*, "Stereo-based stochastic mapping with discriminative training for noise robust speech recognition," in *Proc. ICASSP*, 2009, pp. 2933-2936.
 - [13] W. J. Krzanowski, *Principles of Multivariate Analysis: A User's Perspective*. New York: Oxford University Press, 1988.
 - [14] Y. Liu and P. Fung, "Acoustic and phonetic confusions in accented speech recognition," in *Proc. Interspeech*, 2005, pp. 3033-3036.
 - [15] J. M. Górriz, *et al.*, "Generalized LRT-based voice activity detector," *IEEE Signal Processing Letters*, vol. 13, pp. 636-639, 2006.
 - [16] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic Press, 1990.
 - [17] N. A. Campbell, "Canonical variate analysis with unequal covariance matrices - generalizations of the usual solution," *Mathematical Geology*, vol. 16, pp. 109-124, 1984.
 - [18] M. Sakai, *et al.*, "Linear discriminant analysis using a generalized mean of class covariances and its application to speech recognition," *IEICE Trans. Information and Systems*, vol. E91-D, pp. 478-487, 2008.
 - [19] J. D. Foley, *et al.*, *Computer Graphics: Principles and Practice in C*, 2nd ed.: Addison-Wesley, 1995.
 - [20] H.-M. Wang, *et al.*, "MATBN: A mandarin Chinese broadcast news corpus," *International Journal of Computational Linguistics and Chinese Language Processing*, vol. 10, pp. 219-235, 2005.
 - [21] B. Chen, *et al.*, "Lightly supervised and data-driven approaches to mandarin broadcast news transcription," in *Proc. ICASSP*, 2004, pp. 777-780.
 - [22] A. Stolcke, *SRI Language Modeling Toolkit (Version 1.5.2)*.
 - [23] R. A. Gopinath, "Maximum likelihood modeling with Gaussian distributions for classification," in *Proc. ICASSP*, 1998, pp. 661-664.
 - [24] L. Wood, *et al.*, "Improved vocabulary-independent sub-word HMM modelling," in *Proc. ICASSP*, 1991, pp. 181-184.
 - [25] G. Yu, *et al.*, "Discriminant analysis and supervised vector quantization for continuous speech recognition," in *Proc. ICASSP*, 1990, pp. 685-688.

