聚集事後機率線性迴歸調適演算法應用於語音辨識
Aggregate *a Posteriori* Linear Regression for Speech Recognition

黃志賢 王奕凱 簡仁宗

國立成功大學資訊工程學系

{acheron,display}@chien.csie.ncku.edu.tw, jtchien@mail.ncku.edu.tw

摘要

在本論文中，我們提出一套由聚集事後機率(aggregate *a posteriori*)為基礎之鑑別式線性回歸(linear regression)轉換矩陣參數調適演算法。在近幾年，由於鑑別式訓練的效果優越，於是出現使用鑑別式訓練法則進行轉換矩陣調適，稱為最小分類錯誤率線性迴歸(minimum classification error linear regression, MCELR)調適演算法。我們認為使用最小分類錯誤率準則進行線性迴歸調適時，若能再進一步考慮線性迴歸矩陣之事前機率分佈，則可以結合貝氏法則之強健性與最小分類錯誤率之鑑別性，以估測出更佳之轉換矩陣用於語者調適上。透過聚集事後機率與鑑別式訓練間之關連及適當之條件簡化，則可得到參數更新之封閉解(close form)型式以加速鑑別式訓練的參數估測。在實驗中，我們使用 TCC300 語料進行語音模型參數之訓練與迴歸矩陣之事前機率分佈之參數估測，而在調適及測試時，則使用公共電視台所錄製之電視新聞語料，進行轉換矩陣估測強健性之評估與其他轉換矩陣參數調適效能之比較，在不同調適語料之實驗結果發現我們提出之聚集事後機率線性回歸可以有效達到鑑別式語者調適的效果。

1. 緒論

在語音辨識的相關研究中，常常需要面對的問題是用於訓練時的語料與測試時語料的語者或環境常常大不相同。每個人的聲學特質都不相同，而不同環境所產生的背景雜訊也都不同。如何有效地將訓練所得的語音模型配合測試時所使用的語料特性進行適當的語者調適，以有效地消除這兩者之間的不匹配情形，是許多學者研究的課題。

語音模型的參數必須在訓練時使用大量語料進行估測，最普遍使用的模型訓練準則為最大相似度估測(maximum likelihood estimate, MLE)[19]，在此種方法中，當模型與所收集之訓練語料的相似度最大時，即可求得在此估測準備下最佳的語音模型參數。由於語音模型參數的估測，有所謂不完整資料(incomplete data)的問題，所以皆利用 EM(Expectation-Maximization)演算法[6]進行理論推導。

除了使用最大相似度作為參數估測準則之外，另一個也常被用於作為參數估測的是基於貝氏理論的最大事後機率(maximum *a posteriori*, MAP)估測法則[8]。貝氏估測法則認為參數為一隨機變數，可以機率分佈表示之。利用根據所給定的訓練語料而使得對應的模型參數之事後機率最大之特性，即可求得基於此方法之最佳參數。在最大事後機率訓練法則之訓練機制下，一般不可直接最大化模型參數之事後機率，而常根據貝氏法則，將之拆解為語料與模型間相似度與模型參數事前機率之組合，所以可利用事前資訊對模型參數加以限制，可以改善訓練資料稀疏所產生的錯誤訓練問題。

除了前述兩者參數估測準則之外，鑑別式訓練(discriminative training)[3]則提供了在模型訓練上的另一種選擇。由較早的 multilayer perceptron(MLP)[17]、learning vector quantization(LVQ)[18]，到近來的最小分類錯誤(minimum classification error, MCE)[11]、最大相互資訊(maximum mutual information, MMI)[20]，有許多不同的理論方法。鑑別式訓練與其它模型訓練方法最大的不同是，除了考慮樣本與本身模型的相似度之外，還額外考慮樣本與其它模型之間的相似度，這種作法可以避免模型訓練時，原本就相似的語音模型產生互相混淆的情況。

Qi Li [15]在 2002 年提出一般化最小錯誤率(generalized minimum error rate, GMER)，由事後機率的角度出發，定義聚集事後機率(aggregate *a posteriori*, AAP)，並將事後機率改寫為具鑑別性形式的誤辨率(misclassification measure)函式。在訓練模型參數上，不使用一般的廣義機率遞減法則(generalized probabilistic descent, GPD)，透過一些條件假設，即可推導出模型參數估測的封閉解形式。

在語者調適的研究上，最廣為使用的有最大相似度線性迴歸(maximum likelihood linear regression, MLLR)調適[7][14]與最大事後機率調適兩大類方法。在本研究中我們將使用前者作為調適的主要架構，透過所估測出之線性迴歸矩陣對語音模型參數進行調適。由於考慮到使用語料量稀少易造成調適效果失準的情況，引入線性轉換矩陣之事前分佈資訊，以強健化調適效能外，也將由鑑別式訓練之角度出發，嘗試找出不同於傳統以貝氏法則為準之最大化

聚集事後機率線性迴歸(aggregate *a posteriori* linear regression, AAPLR)演算法。故我們會針對文獻中所提過之以線性迴歸為主之調適演算法作回顧。除了最大相似度線性迴歸調適演算法之外，主要有最大事後機率線性迴歸(MAPLR)[21]、考慮到漸進式(sequential)學習的近似貝氏線性迴歸(quasi-Bayes linear regression, QBLR)[5]與最小分類錯誤線性迴歸(minimum classification error linear regression, MCELR)[4][9][10]。

我們將提出的語音模型參數調適演算法，使用連續語音辨識系統進行與其他調適演算法的效能評估。接下來，我們先回顧近年來鑑別式訓練的相關研究文獻與前述幾種以轉換為主之語音模型調適演算法及將聚集事後機率應用在鑑別式聲學模型參數估測上的方法。其次，說明我們將一般化最小錯誤率應用在語音模型參數調適及在調適時考慮轉換矩陣的事前機率分佈，最後得到估測的轉換矩陣參數封閉解之相關理論內容。接著說明實驗設定與進行方式並由實驗所得結果進行討論。在結尾部份則簡單歸納本論文的主要重點與結論，並說明未來繼續研究的方向與課題。

2. 鑑別式訓練及線性回歸調整

最大相似度參數估測法則是最普遍用來訓練隱藏式馬可夫模型參數的方法，它利用 EM 演算法估測模型參數非常有效率；最大相似度的缺點是模型參數只利用屬於本身模型的資料來估測，和其它模型的參數估測基本上是獨立的。最小分類錯誤和最大交互資訊，是近來較廣為利用的鑑別式訓練方法，除了訓練語音模型外，還用在語言模型(language model)的訓練上[13]、語者辨識模型訓練、特徵參數擷取。使用鑑別式訓練估測模型參數時，除了本身模型的資料外，還考慮與其它模型參數之鑑別性，所以可以更正確地估測出所需的模型參數內容。在[15][16]中，作者提出了另一種鑑別式訓練方法，稱作一般化最小錯誤率，從事後機率出發，定義與最大事後機率相似的目標函式，並且改寫為鑑別式訓練的形式，以下分別簡介這三種鑑別式訓練法則。

2.1 最小分類錯誤(MCE)訓練法則

在兩個類別 C_1, C_2 的分類器裡，假設 $\mathbf{x} \in C_1$，貝氏分類法則定義了最基本的誤辨值函式(misclassification measure)為

$$d(\mathbf{x}) = P(C_2 \mid \mathbf{x}) - P(C_1 \mid \mathbf{x}) \tag{1}$$

上式表示類別 C_1 的觀察資料 $\mathbf{x}$ 被分類器分類到類別 C_2 的可能性，在多個類別的分類器[12]裡，定義誤辨值函式

$$d_k(\mathbf{x}) = \sum_{i \in M_i} \frac{1}{m_k} \left[g_i(\mathbf{x};\Lambda) - g_k(\mathbf{x};\Lambda) \right] \tag{2}$$

其中 $g_i(\mathbf{x};\Lambda)$ 為觀察資料 $\mathbf{x}$ 對類別 C_i 的相似度，Λ 表示所有類別的模型參數，$M_k = \{ i \mid g_i(\mathbf{x};\Lambda) > g_k(\mathbf{x};\Lambda) \}$，代表一群對觀察資料 $\mathbf{x}$ 的相似度比類別 C_k 對觀察資料 $\mathbf{x}$ 相似度更具競爭性的類別集合，即混淆類別(confusing classes)或競爭類別(competing classes)的集合。

式子(2)中，S_k 並非是固定的集合，它隨著模型參數 Λ 和觀察資料 $\mathbf{x}$ 而改變，而且該式在 Λ 不連續[12]，這在最陡坡降法(gradient descent)裡並不適用，因此另外定義了一個連續性的誤辨值公式為

$$d_k(\mathbf{x}) = -g_k(\mathbf{x};\Lambda) + \left[\frac{1}{M-1} \sum_{j, j \neq k} g_j(\mathbf{x};\Lambda)^{\eta} \right]^{1/\eta} \tag{3}$$

其中 η 是一個正數，藉著改變 η 的值，可以改變式子裡具影響力的競爭類別數量，令 $\eta \to \infty$，一個極端的誤辨值公式為

$$d_k(\mathbf{x}) = -g_k(\mathbf{x};\Lambda) + g_i(\mathbf{x};\Lambda) \tag{4}$$

類別 C_i 是除了類別 C_k 外，和觀察資料 $\mathbf{x}$ 相似度最大的類別，$d_k(\mathbf{x}) > 0$ 代表發生分類錯誤，$d_k(\mathbf{x}) \leq 0$ 代表正確分類。為了更進一步完成目標函式的定義，把誤辨值公式代入 cost function

$$l_k(x;\Lambda) = l(d_k(\mathbf{x})) \tag{5}$$

cost function 一般為連續性，範圍為[0,1]的函式，最常用於 MCE 的為 sigmoid，

$$l(d_k) = \frac{1}{1 + \exp(-\gamma d_k + \theta)} \tag{6}$$

對於某個觀察資料 $\mathbf{x}$，我們可以 cost function 定義分類器的效率為

$$l(\mathbf{x};\Lambda)=\sum_{i=1}^{M} l_i(\mathbf{x};\Lambda)1(\mathbf{x}\in C_i) \tag{7}$$

最後利用廣義機率遞減(generalized probabilistic decent, GPD)演算法進行疊代運算以實現 MCE 法則。

$$\Lambda_{t+1}=\Lambda_t-\varepsilon_t U_t \nabla l(x;\Lambda)|_{\Lambda=\Lambda_t} \tag{8}$$

廣義機率遞減法則是應用很廣的演算法，利用反覆的計算，遞迴得到一收斂的值，缺點是收斂速度慢，而且式中的學習係數ε_t需對應不同的資料特性去調整。更進一步之相關參數估測過程與結果詳見[11]。

2.2 最大交互資訊(MMI)訓練法則

除了最小分類錯誤法則外，最大交互資訊也是普遍利用的鑑別式訓練式法則[1][20]，最大交互資訊較隱性的引入了觀察資料與其它類別的相似度，所以與一般化最小錯誤率較相似，在混合數高的情況下，最大交互資訊能訓練出比最小分類錯誤辨識率更高的模型參數[1]，由於最大交互資訊考慮了觀察資料和所有類別的相似度，因此比最小分類錯誤在實作上難度更高。為了快速計算隱藏式馬可夫模型和觀察資料$\mathbf{x}$的相似度，必須使用 forward-backward 演算法。透過 forward probability $\alpha_j(t)$ 與 backward probability $\beta_j(t)$ 的表示，類別C_m產生觀察資料$\mathbf{X}$的機率可寫為下式

$$P(X\mid C_m,\Lambda)=\sum_{t=1}^{T}\sum_{j=1}^{N}\alpha_j(t)\beta_j(t) \tag{9}$$

定義類別C_m與觀察資料$\mathbf{X}$的交互資訊為

$$\begin{aligned} I_\Lambda(C_m,\mathbf{X}) &= \log\frac{P(\mathbf{X}\mid C_m)}{P(\mathbf{X})} \\ &= \log P(\mathbf{X}\mid C_m)-\log P(\mathbf{X}) \\ &= \log P(\mathbf{X}\mid C_m)-\log\sum_{m'=1}^{M}P(\mathbf{X}\mid C_{m'})P(C_{m'}) \end{aligned} \tag{10}$$

其中$P(\mathbf{X},C_m)$代表類別C_m與$\mathbf{X}$同時出現的機率，即聯合相似度(joint likelihood)。由(10)式可看出，除了觀察資料$\mathbf{X}$與對應類別C_m的相似度之外，還加入了$\mathbf{X}$與其它類別的相似度作為參數估測的考量，因此它屬於鑑別式訓練的一種，以最大交互資訊法則得到的模型參數可使得觀察資料$\mathbf{X}$與類別C_m有較高的相依性，即$I_\Lambda(C_m,\mathbf{X})$較高。與最小分類錯誤相同，最大交互資訊也必須以廣義機率遞減演算法實現，即

$$\Lambda_{n+1}=\Lambda_n-\varepsilon\nabla I_\Lambda(C_m,\mathbf{X}) \tag{11}$$

在這裡以轉移機率、平均值向量、共變異矩陣作說明，而偏微的對象主要是最大交互資訊中的相似度函式

$$\frac{\partial}{\partial\Lambda}\log P(\mathbf{X}\mid C_m,\Lambda)=\frac{1}{P(\mathbf{X}\mid C_m,\Lambda)}\frac{\partial}{\partial\Lambda}P(\mathbf{X}\mid C_m,\Lambda) \tag{12}$$

相似度函式可由 forward-backward probability 表示

$$\begin{aligned} P(\mathbf{X}\mid C_m,\Lambda) &= \sum_{t=1}^{T}\sum_{j=1}^{N}\alpha_j(t)\beta_j(t) \\ &= \sum_{t=1}^{T}\sum_{j=1}^{N}\left\{\sum_{i=1}^{N}\alpha_i(t)a_{ij}\right\}b_j(\mathbf{x}_t)\beta_j(t) \end{aligned} \tag{13}$$

進一步之參數估測過程與結果，請詳見[20]。

2.3 一般化最小錯誤率(GMER)

一般化最小錯誤率是由 *Qi Li* 在 2002 年所提出，以下簡介一般化最小錯誤率的精神和作法。在一個具有M個類別的分類問題裡面，令觀察資料$\mathbf{X}$屬於類別C_m，α_i表示將$\mathbf{X}$分類到類別C_i的動作，則可定義一 loss function 為

$$l(\alpha_i\mid C_m)=\begin{cases}0 & i=m \quad i,m=1,\ldots,M\\ 1 & i\neq m\end{cases} \tag{14}$$

將分類錯誤指定一個單位的 loss，若分類正確則不指定 loss，代表分類錯誤的風險(risk)，且定義對觀察資料 $\mathbf{X}$ 採取動作 α_i 的分類錯誤機率為

$$R(\alpha_i \mid \mathbf{X}) = \sum_{j=1}^{M} l(\alpha_i \mid C_j) P(C_j \mid \mathbf{X}) = 1 - P(C_m \mid \mathbf{X}) \tag{15}$$

$P(C_m \mid \mathbf{X})$ 代表 $\mathbf{X}$ 屬於類別 C_m 的事後機率，貝氏法則告訴我們，令 $P(C_m \mid \mathbf{X})$ 最大可降低分類錯誤的機率，稱作最小錯誤率(minimum error rate, MER)，$P(C_i \mid \mathbf{X})$ 一般以一組定義好的模型參數 λ_i 來計算，即 $P(C_i \mid \mathbf{X}) = P_\lambda(C_i \mid \mathbf{X})$，由於模型參數與類別有一對一的關係，因此簡化表示為 $P(C_i \mid \mathbf{X}) = P(\lambda_i \mid \mathbf{X})$。在訓練方面，首先定義聚集事後機率(aggregate *a posteriori*, AAP)

$$J = \frac{1}{M} \sum_{m=1}^{M} \sum_{n=1}^{N_m} \frac{P(\mathbf{X}_{m,n} \mid \lambda_m) P_m}{P(\mathbf{X}_{m,n})} \tag{16}$$

$\mathbf{X}_{m,n}$ 代表模型 m 的第 n 個訓練資料，長度為 T_n，即 $\mathbf{X}_{m,n} = \{\mathbf{x}_{m,n,t}\}_{t=1}^{T_n}$，$P_m$ 為類別 m 的事前機率，假設訓練資料分佈為 independent, identically distributed (i.i.d)，因此 $\mathbf{X}_{m,n}$ 與 λ_m 的相似度可表示為 $P(X_{m,n} \mid \lambda_m) = \prod_{t=1}^{T_n} P(x_{m,n,t} \mid \lambda_m)$。為了具有鑑別式訓練的形式，將(16)式改寫為

$$\tilde{J} = \frac{1}{M} \sum_{m=1}^{M} \sum_{n=1}^{N_m} l(d_{m,n}) = \frac{1}{M} \sum_{m=1}^{M} \sum_{n=1}^{N_m} l_{m,n} \tag{17}$$

其中 l 為(6)式 sigmoid function。

$$d_{m,n} = \log P(\mathbf{X}_{m,n} \mid \lambda_m) P_m - \log \sum_{j \neq m} P(\mathbf{X}_{m,n} \mid \lambda_j) P_j \tag{18}$$

為了讓正確類別與競爭類別佔有不同的百分比，在(18)式裡第二項乘上 L，$0 < L \leq 1$，當 $L = 1$ 時，代表正確類別與競爭類別具同樣重要性，同時令(6)式 sigmoid function 內 $\gamma = 1$，$\theta = 0$ 時，$\tilde{J} = J$，$P(\mathbf{X}_{m,n} \mid \lambda_m)$ 為 GMM 函式，為了令 $\tilde{J}$ 為最大，因此對 $\tilde{J}$ 取 gradient 並令為零可得到

$$\begin{aligned} \nabla_{\theta_{mi}} \tilde{J} = & \sum_{n=1}^{N_m} \sum_{t=1}^{T_n} \Omega_{m,i}(\mathbf{x}_{m,n,t}) \nabla_{\theta_{mi}} \log P(\mathbf{x}_{m,n,t} \mid \lambda_{m,i}) \\ & - L \sum_{j \neq m} \sum_{n=1}^{N_j} \sum_{t=1}^{T_n} \Omega_{j,i}(\mathbf{x}_{j,n,t}) \nabla_{\theta_{mi}} \log P(\mathbf{x}_{j,n,t} \mid \lambda_{m,i}) = 0 \end{aligned} \tag{19}$$

其中

$$\Omega_{m,i}(\mathbf{x}_{m,n,t}) = l_{m,n}(1 - l_{m,n}) \frac{c_{m,i} P(\mathbf{x}_{m,n,t} \mid \lambda_{m,i})}{P(\mathbf{x}_{m,n,t} \mid \lambda_m)} \tag{20}$$

$$\Omega_{j,i}(\mathbf{x}_{j,n,t}) = l_{j,n}(1 - l_{j,n}) \frac{c_{m,i} P(\mathbf{x}_{j,n,t} \mid \lambda_{m,i}) P_m}{\sum_{k \neq j} P(\mathbf{x}_{j,n,t} \mid \lambda_k) P_k} \tag{21}$$

為了得到模型參數的封閉解(close-form solution)，這裡假設(20)與(21)式與模型參數獨立，若欲求平均值向量，將(19)式 $\log P(\mathbf{x}_{m,n,t} \mid \lambda_{m,i})$ 對平均值向量取偏微分後可得

$$\nabla_{\mu_{m,i}} \log P(\mathbf{x}_{m,n,t} \mid \lambda_{m,i}) = \Sigma_{m,i}^{-1}(\mathbf{x}_{m,n,t} - \mu_{m,i}) \tag{22}$$

將上式代入(19)式移項後可得平均值向量的解為

$$\mu_{m,i} = \frac{\sum_{n=1}^{N_m} \sum_{t=1}^{T_n} \Omega_{m,i}(\mathbf{x}_{m,n,t}) \mathbf{x}_{m,n,t} - L \sum_{j \neq m} \sum_{n=1}^{N_j} \sum_{t=1}^{T_n} \Omega_{j,i}(\mathbf{x}_{j,n,t}) \mathbf{x}_{j,n,t}}{\sum_{n=1}^{N_m} \sum_{t=1}^{T_n} \Omega_{m,i}(\mathbf{x}_{m,n,t}) - L \sum_{j \neq m} \sum_{n=1}^{N_j} \sum_{t=1}^{T_n} \Omega_{j,i}(\mathbf{x}_{j,n,t})} \tag{23}$$

2.4 線性迴歸語者調適

根據語音模型與語者間之相關性可分為語者獨立(speaker-independent, SI)語音模型及語者相依

(speaker-dependent, SD)語音模型。使用語者相依之語音模型，在辨識時，須先行指定或偵測要使用的語者模型組別，而語者獨立則不須，以此差別看來，語者相依之語音辨識系統，使用上較不便，且需儲存多組語音模型。相對來說，使用語者獨立語音模型時所需要的語音模型數量會較少且模型特性與每一位測試語者均不甚吻合。所以，辨識率會較差。一般而言，使用語者相依語音模型的辨識系統效能會比語者獨立之辨識系統效能高二至三倍[7]。

為了保留兩者優點，一般皆訓練出一組語者獨立的語音模型，取其模型總數量較少的優點，而以此模型為基礎，再利用一些由測試語者所錄得之調適語料，先調適出與該語者語音特性較相符的語音模型，即所謂的語者相依語音模型，可有效提升語音辨識率。不過用於調整的語料一般並不多，容易造成調適語料稀疏的問題，為了解決樣本數不足的問題，做法是將語音模型分群，為每一群的語音模型找出一個參數轉換矩陣，群集內的模型調整只要依照此轉換矩陣即可得到更新後參數。為了得到更新後的轉換矩陣，可以利用不同的法則，較常見的有最大相似度線性迴歸法則，最大事後機率線性迴歸法則，最小分類錯誤線性迴歸法則。

2.5 最大相似度線性迴歸(MLLR)

最大相似度線性迴歸的目標就是，對一群集 s，計算一轉換矩陣 W_s，使得群集內所有調適資料的相似度最大，最大相似度線性迴歸調適演算法的好處在於，調適語料不需要完全涵蓋所有模型，即使沒有調適資料的模型，也可以經由同類別的轉換矩陣進行調適。以調整平均值向量為例，在計算轉換矩陣之前，將平均值向量延展為

$$\xi_s = [1, \mu_1, \mu_2, \ldots, \mu_D]^T \tag{24}$$

其中，D 為向量維度，則更新後的平均值向量為

$$\hat{\mu}_s = \mathbf{W}_{r(s)} \xi_s \tag{25}$$

其中，$r(s)$ 代表狀態 s 所屬迴歸類別，$\mathbf{W}_{r(s)}$ 代表迴歸類別(regression class) $r(s)$ 的轉換矩陣，維度為 $D \times (D+1)$，則透過 EM 演算法，最後可以得到每一個迴歸類別的轉換矩陣之每一列計算方式如下

$$\mathbf{w}_i^T = \mathbf{G}^{(i)^{-1}} \mathbf{z}_i^T \tag{26}$$

$\mathbf{w}_i^T$ 和 $\mathbf{z}_i^T$ 分別代表 $\mathbf{W}$ 和 $\mathbf{Z}$ 的列向量[14]。

2.6 最大事後機率線性迴歸(MAPLR)

由於以最大相似度為主之線性轉換矩陣在計算上十分簡易，所以其應用十分普遍，然而，若調適語料過少，或語料特性不具代表性時，則可能導致得到的轉換矩陣仍舊無法符合測試語者的語音特性，於是，便考慮到引入轉換矩陣的事前分佈資訊。矩陣參數的事前分佈可以在估測轉換矩陣時限制參數可能的調適量，使得參數的估測更具強健性，由文獻實驗可看出，最大事後機率線性迴歸可達到比最大相似度線性迴歸更好的辨識率[21]。

2.7 最小分類錯誤線性迴歸(MCELR)

最小分類錯誤的鑑別式訓練方式在很多應用都能顯示出不錯的效能，不過最小分類錯誤一般以廣義機率遞減演算法實現，並沒有在理論上證明它能收斂到更好的模型，當訓練資料變少時，錯誤的收斂停止點更容易發生，因此將 MCE 應用在模型調適時，使用線性迴歸有其必要。*Chengalvarayan* 在 1998 年提出最小分類錯誤線性迴歸[4]，使用全域性的轉換矩陣並以廣義機率遞減演算法估測矩陣參數，實驗結果顯示出其調適效果比最大相似度線性迴歸演算法好。而在[10]中，更進一步使用多組迴歸類別的轉換矩陣進行調適，在同樣使用廣義機率遞減演算法下，可以有更好的調適效能改進。另外，在[9]中，作者不利用廣義機率遞減演算法實現最小分類錯誤線性迴歸調適演算法，而以一般化調適作法計算轉換矩陣，即轉換矩陣以群集為單位，將最小分類錯誤的目標函式改寫後，可以透過 EM 演算法以封閉解的方式計算轉換矩陣。

3. 聚集事後機率線性迴歸鑑別式調適法

在最小分類錯誤估測法則中，並不考慮類別的事前資訊，且使用廣義機率遞減演算法實現，在調適資料少時，更容易發生錯誤訓練的問題，因此，Beyerlin 將所有模型(語音模型、語言模型)組成一個事後機率的線性組合，利用鑑別式訓練估測出線性組合的係數[2]。由先前所介紹的一般化最小錯誤率[15][16]，從最大事後機率的角度出發，另外定義所謂聚集事後機率(AAP)，並將式子改寫為鑑別式訓練的形式，在所給定的部份假設下，可以得到鑑別式訓練的封閉解，相較於傳統使用的廣義機率遞減演算法，有較快的計算速度，而且不用調整學習速率(learning rate)和步進大小(step size)。由於調適時資料較少，於是將一般化最小錯誤率代入尋找轉換矩陣也應該相當合適。

3.1 聚集事後機率線性迴歸(AAPLR)與最小分類錯誤線性迴歸(MCELR)之關係

考慮到最大事後機率在少量訓練語料下可以得到比最大相似度較正確的模型參數，由前述的一般化最小錯誤率介紹中可以看出，它將事後機率中原本與模型參數無關的 $P(\mathbf{x}_{m,n})$ 表示成與模型相關，即具鑑別式訓練的形式，將原本最小分類錯誤中鑑別式函式為相似度函式改為事後機率函式，可以結合這兩種模型估測方式的優點，並利用封閉解的解法可以快速估測出模型參數，改善以往以廣義機率遞減法則實作時收斂太慢的缺點。

由於語音模型調適時資料量通常較少，因此將一般化最小錯誤率的方式導入將有助於參數的估測，我們將此調適的方式稱為聚集事後機率線性迴歸(AAPLR)調適演算法。為了以 AAPLR 的方式計算轉換矩陣，且加入轉換矩陣的事前資訊可以讓其估測較具強健性，因此將(16)式聚集事後機率改寫為

$$J=\sum_{m=1}^{M}\sum_{n=1}^{N_m}\frac{p(\mathbf{x}_{m,n}\mid\hat{\mathbf{W}}_{r(m)};\Lambda)P_m g(\hat{\mathbf{W}}_{r(m)})}{p(\mathbf{x}_{m,n})}, \tag{27}$$

在繼續推導聚集事後機率線性迴歸演算法前，我們將透過 EM 演算法，發掘最大事後機率線性迴歸與使用最小分類錯誤準則之參數估測演算法之間的差異。給定一語音觀察樣本序列 $\mathbf{X}=\{\mathbf{x}_1,\ldots,\mathbf{x}_T\}$，其長度為 T，且存在線性轉換矩陣集合 $\mathbf{W}=\{\mathbf{w}_1,\ldots,\mathbf{w}_R\}$，其中共有 R 組類別。則在給定觀察樣本序列 $\mathbf{X}$ 時，線性轉換矩陣的事後機率可表示如下

$$g(\mathbf{W}\mid\mathbf{X};\Lambda) \tag{28}$$

其中，Λ 表示用於相似度計算之語音模型集合。而上述之事後機率又可以透過貝氏法則轉換如下之相似度與事前機率之組合

$$\begin{aligned} g(\mathbf{W}\mid\mathbf{X};\Lambda)&=\frac{p(\mathbf{X}\mid\mathbf{W};\Lambda)g(\mathbf{W})}{p(\mathbf{X})}\\ &=\prod_{t=1}^{T}\frac{p(\mathbf{x}_t\mid\mathbf{W};\Lambda)g(\mathbf{W})}{p(\mathbf{x}_t)} \end{aligned} \tag{29}$$

此處之 $g(\mathbf{W})$ 代表線性轉換矩陣 $\mathbf{W}$ 之事前分佈機率。再進一步將(29)式對數化可得

$$\log g(W\mid X;\Lambda)=\sum_{t=1}^{T}\log\frac{p(x_t\mid W;\Lambda)g(W)}{p(x_t)} \tag{30}$$

在 EM 演算法中之 E-step 即用於計算以下之輔助函式

$$\begin{aligned} R(\hat{\mathbf{W}}\mid\mathbf{W})&=E\left\{\log\frac{p(\mathbf{X},\mathbf{q}\mid\hat{\mathbf{W}};\Lambda)g(\hat{\mathbf{W}})}{p(\mathbf{X})}\middle|\mathbf{X},\mathbf{W}\right\}\\ &=\sum_{i=1}^{M}\sum_{t=1}^{T}p(q_t=i\mid\mathbf{x}_t,\mathbf{W};\Lambda)\log\frac{p(\mathbf{x}_t,q_t=i\mid\hat{\mathbf{W}};\Lambda)g(\hat{\mathbf{W}})}{p(\mathbf{x}_t)} \end{aligned} \tag{31}$$

其中，$\mathbf{q}=(q_1,q_2,\ldots,q_T)$ 表示給定之觀察序列 $\mathbf{X}$ 之每一時間點所對應之狀態序列。$\hat{\mathbf{W}}$ 表示透過 EM 演算法欲估測之新轉換矩陣參數，而 $\mathbf{W}$ 則是現有透過 EM 演算法在前一次 M 步驟中所估測出之最佳轉換矩陣參數。令 $\gamma_i(\mathbf{x}_t)=p(q_t=i\mid\mathbf{x}_t,\mathbf{W};\Lambda)$ 用以表示在第 t 個時間點，觀察樣本 $\mathbf{x}_t$ 停留於第 i 個狀態之機率，則(31)式之輔助函式可簡單表示為

$$\begin{aligned} R(\hat{\mathbf{W}}\mid\mathbf{W})&=E\left\{\log\frac{p(\mathbf{X},\mathbf{q}\mid\hat{\mathbf{W}};\Lambda)g(\hat{\mathbf{W}})}{p(\mathbf{X})}\middle|\mathbf{X},\mathbf{W}\right\}\\ &=\sum_{i=1}^{M}\sum_{t=1}^{T}\gamma_i(\mathbf{x}_t)\log\frac{p(\mathbf{x}_t,q_t=i\mid\hat{\mathbf{W}};\Lambda)g(\hat{\mathbf{W}})}{p(\mathbf{x}_t)} \end{aligned}. \tag{32}$$

在此，我們使用維特比(Viterbi)近似法則來簡化我們的式子。於是，我們使用最佳的狀態序列來取代原有需考慮所有可能性之表示法，同時上述之狀態停留機率 $\gamma_i(\mathbf{x}_t)$ 則簡化如下

$$\gamma_i(\mathbf{x}_t)=\begin{cases}0 & q_t=i\\ 1 & q_t\neq i\end{cases} \tag{33}$$

此外，為了與以下之聚集事後機率比較，我們將使用下列定義之符號重新表示(32)式。使用 m 來表示原有之狀態標示 i，即將狀態視為語音模型類別；使用 $\mathbf{x}_{m,n}$ 取代原有之 $\mathbf{x}_t$。因為原有之觀察樣本 $\mathbf{x}_t$ 在經過維特比解碼器對應出最佳之狀態後，即可以明確知道兩者間之關連。所以用 $\mathbf{x}_{m,n}$ 來表示原有觀察樣本 $\mathbf{x}_t$ 為對應至第 m 類模型之第 n 個觀察樣本。則(32)式可以重新表示為

$$R(\hat{\mathbf{W}} \mid \mathbf{W}) = \sum_{m=1}^{M}\sum_{n=1}^{N_m} \log \frac{p(\mathbf{x}_{m,n}, m \mid \hat{\mathbf{W}}_{r(m)}; \Lambda) g(\hat{\mathbf{W}}_{r(m)})}{p(\mathbf{x}_{m,n})}. \tag{34}$$

其中，$\hat{\mathbf{W}}_{r(m)}$ 表示該線性轉換矩陣是用於轉換第 m 類語音模型參數之用。一般而言，線性轉換矩陣是根據所有語音模型參數中具相似特性之分群結果而分為數個類別，如分為 R 群，被分於同群之語音模型是共用同一組轉換矩陣進行轉換。於是在給定語音模型類別 m 後，即可以透過上述之關係，得到對應之轉換矩陣類別。另外，我們是以 $r(m)$ 表示第 r 類轉換矩陣與第 m 類語音模型之關係。

從另一方面來看，遵循上述變數、標示之定義，則轉換矩陣 $\mathbf{W}$ 之聚集事後機率定義即為(27)式，從(27)式與(34)式比較可知，在使用 EM 演算法對語音模型或是此處所考慮之轉換矩陣之參數進行估測時，是將第(34)式之 $R(\hat{\mathbf{W}} \mid \mathbf{W})$ 針對所欲估測之參數予以偏微分後，而透過封閉解來得到更新的參數內容。而在聚集事後機率的定義式中，則是將各個類別之事後機率全部加總起來，於是在文獻中接下來的推導過程中，才可朝所謂的最小分類錯誤之鑑別式參數估測之同理性進行推導，並經一些假設設定後，得以使用封閉解的方式進行參數內容之更新。

3.2 聚集事後機率線性迴歸(AAPLR)參數估測

接下來將利用(27)式進行模型參數的估測，與一般化最小錯誤率一樣，同樣可將(27)式改寫為一目標函式為

$$\tilde{J} = \frac{1}{M}\sum_{m=1}^{M}\sum_{n=1}^{N_m} l(d_{m,n}) = \frac{1}{M}\sum_{m=1}^{M}\sum_{n=1}^{N_m} l_{m,n} \tag{35}$$

$$d_{m,n} = \log p(\mathbf{x}_{m,n} \mid \lambda_m) P_m g(\mathbf{W}_{r(m)}) - \log \sum_{j \neq m} p(\mathbf{x}_{m,n} \mid \lambda_j) P_j g(\mathbf{W}_{r(m)}) \tag{36}$$

其中，$g(\mathbf{W}_{r(m)})$ 為轉換矩陣 $\mathbf{W}_{r(m)}$ 的事前機率分佈，$r(m)$ 代表模型 m 的迴歸類別，$g(\mathbf{W}_{r(m)})$ 為一矩陣版本高斯分佈，稱作 elliptically symmetric distribution 或 matrix variate normal distribution。

$$g(\mathbf{W}_{r(m)}) \propto |\Delta|^{-\nu/2} \cdot q\left(\sum_{d=1}^{D} (\mathbf{w}_{r(m)d} - \mathbf{m}_{r(m)d}) \Sigma_d^{-1} (\mathbf{w}_{r(m)d} - \mathbf{m}_{r(m)d})^T \right) \tag{37}$$

$\mathbf{q}$ 為一個 $[0,\infty)$ 的函式，$\mathbf{w}_{r(m)d}$ 和 $\mathbf{m}_{r(m)d}$ 分別代表轉換矩陣和平均矩陣的第 d 列向量，維度為 $1\times(D+1)$，Δ 為一維度 $D(D+1)\times D(D+1)$ 的區塊對角化共變異矩陣(block diagonal covariance matrix)，每一區塊由 $(D+1)\times(D+1)$ 的 Σ_d 組成。為了簡化最佳轉換矩陣，首先將加入轉換矩陣的高斯分佈改寫為一單變量形式如下

$$N(\mathbf{x}_{m,n} \mid \xi_{m,i}, \Sigma_{m,i}) = \frac{1}{(2\pi)^{D/2} |\Sigma_{m,i}|^{1/2}} \exp\left[-\frac{1}{2} \sum_{d=1}^{D} \frac{(\mathbf{x}_{m,n,d} - \mathbf{w}_{r(m)d} \xi_{m,i})^2}{\sigma_{m,i,d}^2} \right] \tag{38}$$

$\mathbf{w}_{r(m)d}$ 代表轉換矩陣 $\mathbf{W}_{r(m)}$ 的第 d 列向量。將(36)和變更過的高斯分佈(38)代入(35)式得到 AAPLR 的目標函式並對欲求的轉換矩陣第 d 列 $\mathbf{w}_{r(m)d}$ $(d=1,\ldots,D)$ 取偏微分得

$$\begin{aligned}
\nabla_{\mathbf{w}_{r(m)d}} J = & \sum_{m=1}^{M_c}\sum_{n=1}^{N_m} l(d_{m,n})(1 - l(d_{m,n})) \\
& \times \left[\begin{array}{l}
\displaystyle\sum_{i=1}^{I_m} \frac{c_{m,i} N(\mathbf{x}_{m,n} \mid \lambda_m, \mathbf{W}_{r(m)})}{P(\mathbf{x}_{m,n} \mid \lambda_m, \mathbf{W}_{r(m)})} \left(\frac{\mathbf{x}_{m,n,d} - \mathbf{w}_{r(m)d} \xi_{m,i}}{\sigma_{m,i,d}^2} \right) \xi_{m,i}^T \\
\quad + 2(\mathbf{w}_{r(m)d} - \mathbf{m}_{r(m)d}) \Sigma_{r(m)d}^{-1} \\
\quad - \displaystyle\sum_{\substack{m' \in \mathbf{W}_{r(m)} \\ m' \neq m}} \frac{P(\mathbf{x}_{m,n} \mid \lambda_{m'}, \mathbf{W}_{r(m)}) g(\mathbf{W}_{r(m)})}{\displaystyle\sum_{j \neq m} P(\mathbf{x}_{m,n} \mid \lambda_j, \mathbf{W}_{r(j)}) g(\mathbf{W}_{r(j)})} \\
\quad \times \left[\displaystyle\sum_{i=1}^{I_{m'}} \frac{c_{m',i} N(\mathbf{x}_{m,n} \mid \lambda_{m'}, \mathbf{W}_{r(m)})}{P(\mathbf{x}_{m,n} \mid \lambda_{m'}, \mathbf{W}_{r(m)})} \left(\frac{\mathbf{x}_{m,n,d} - \mathbf{w}_{r(m)d} \xi_{m',i}}{\sigma_{m',i,d}^2} \right) \xi_{m',i}^T \right. \\
\quad \left. + 2(\mathbf{w}_{r(m)d} - \mathbf{m}_{r(m)d}) \Sigma_{r(m)d}^{-1} \right]
\end{array} \right]
\end{aligned} \tag{39}$$

令上式為零，移項後可得 $\mathbf{w}_{r(m)d}$ 的解為

$$\mathbf{w}_{r(m)d}\left(\begin{array}{l}\sum_{m=1}^{M_c}\sum_{n=1}^{N_m}\sum_{i=1}^{I_m} l(d_{m,n})(1-l(d_{m,n}))\Omega_{m,i}(\mathbf{x}_{m,n})\frac{1}{\sigma_{m,i,r}^2}\xi_{m,i}\xi_{m,i}^T \\ -2\sum_{m=1}^{M_c}\sum_{n=1}^{N_m} l(d_{m,n})(1-l(d_{m,n}))\Sigma_{r(m)d}^{-1} \\ -\sum_{m=1}^{M_c}\sum_{n=1}^{N_m}\sum_{\substack{m'\in\mathbf{W}_c \\ m'\neq m}}\sum_{i=1}^{I_{m'}} l(d_{m,n})(1-l(d_{m,n}))\omega_{m',n,i}(\mathbf{x}_{m,n})\Omega_{m',i}(\mathbf{x}_{m,n})\frac{1}{\sigma_{m',i,d}^2}\xi_{m',i}\xi_{m',i}^T \\ +2\sum_{m=1}^{M_c}\sum_{n=1}^{N_m}\sum_{\substack{m'\in\mathbf{W}_c \\ m'\neq m}} l(d_{m,n})(1-l(d_{m,n}))\omega_{m',n,i}(\mathbf{x}_{m,n})\Sigma_{r(m)d}^{-1}\end{array}\right)$$

=

$$\left(\begin{array}{l}\sum_{m=1}^{M_c}\sum_{n=1}^{N_m}\sum_{i=1}^{I_m} l(d_{m,n})(1-l(d_{m,n}))\Omega_{m,i}(\mathbf{x}_{m,n})\frac{\mathbf{x}_{m,n,d}}{\sigma_{m,i,d}^2}\xi_{m,i}^T \\ -2\sum_{m=1}^{M_c}\sum_{n=1}^{N_m} l(d_{m,n})(1-l(d_{m,n}))\mathbf{m}_{r(m)d}\Sigma_{r(m)d}^{-1} \\ +\sum_{m=1}^{M_c}\sum_{n=1}^{N_m}\sum_{\substack{m'\in\mathbf{W}_c \\ m'\neq m}}\sum_{i=1}^{I_{m'}} l(d_{m,n})(1-l(d_{m,n}))\omega_{m',n,i}(\mathbf{x}_{m,n})\Omega_{m',i}(\mathbf{x}_{m,n})\frac{\mathbf{x}_{m,n,d}}{\sigma_{m',i,d}^2}\xi_{m',i}^T \\ +2\sum_{m=1}^{M_c}\sum_{n=1}^{N_m}\sum_{\substack{m'\in\mathbf{W}_c \\ m'\neq m}} l(d_{m,n})(1-l(d_{m,n}))\omega_{m',n,i}(\mathbf{x}_{m,n})\mathbf{m}_{r(m)d}\Sigma_{r(m)d}^{-1}\end{array}\right) \tag{40}$$

其中

$$\Omega_{m,i}(\mathbf{x}_{m,n}) = \frac{c_{m,i}N(\mathbf{x}_{m,n}\mid\lambda_m,\mathbf{W}_{r(m)})}{P(\mathbf{x}_{m,n}\mid\lambda_m,\mathbf{W}_{r(m)})} \tag{41}$$

$$\omega_{m',n,i}(\mathbf{x}_{m,n}) = \frac{P(\mathbf{x}_{m,n}\mid\lambda_{m'},\mathbf{W}_{r(m)})g(\mathbf{W}_{r(m)})}{\sum_{j\neq m}P(\mathbf{x}_{m,n}\mid\lambda_j,\mathbf{W}_{r(j)})g(\mathbf{W}_{r(j)})} \tag{42}$$

令(40)式的等號左側為 $\mathbf{w}_{r(m)d}\cdot\mathbf{L}$，$\mathbf{L}$ 為一維度 $(D+1)\times(D+1)$ 的方陣，且令等號右側為 $\mathbf{r}$，維度 $1\times(D+1)$，(40)式變為 $\mathbf{w}_{r(m)d}\cdot\mathbf{L}=\mathbf{r}$，可得轉換矩陣 $\mathbf{W}_{r(m)}$ 的第 d 列為

$$\mathbf{w}_{r(m)d} = \mathbf{r}\cdot\mathbf{L}^{-1} \tag{43}$$

重覆上述步驟，可求得所有迴歸類別的轉換矩陣。

4. 實驗與討論

4.1 實驗環境與語音參數、模型設定

在實驗的硬體方面，我們所使用的是 Pentium 4 2.0GHz 個人電腦，搭配 256MB 的記憶體容量，使用的作業系統為 Windows XP Professional 中文版，並以 Microsoft Visual C++ 6.0 作為軟體開發工具。在語音特徵參數求取部份，使用 HTK 中的 HCopy 指令取出語音特徵參數，每一音框的特徵參數皆為 26 維，其中包括 12 階的 MFCC，12 階的 delta MFCC，1 階 log energy 以及 1 階 delta log energy，關於 HCopy 的詳細介紹使用方法以及求取特徵參數的設定檔和 HTK 的其他命令使用方法等，請參考[22]。本實驗所使用的辨識系統是以連續密度隱藏式馬可夫模型(continuous density hidden Markov model, CDHMM)為架構，以中文之聲母與韻母作為 HMM 之基本單元，在聲母部份使用 3 個狀態表示，而韻母則使用 5 個狀態來表示。同時，混合數數量則根據各個狀態所分配到的音框數量來決定，但最大混合數不得超過 32 個。

4.2 實驗語料

在本實驗中，我們分別使用兩種語料以進行實驗，其一是 TCC300 麥克風語料庫，用於語者獨立之語音模型參數訓練；另一個則是公視晚間新聞語料，用於調適後之辨識效能改進評估。以下是這兩個語料庫的資料說明。

TCC 台大/成大/交大麥克風語音資料庫是由國立台灣大學、國立成功大學、國立交通大學各自擁有之語料庫集合而成，各校錄製之目的是為語音辨認研究，屬於麥克風朗讀語音。其中台大語料庫主要包含詞及短句，內容經過仔細設計，考慮了音節及其相連出現機率，由 100 人錄製而成；成大及交大語料庫主要包含長文語料，文章由中研院提供之 500 萬詞詞類標示語料庫中選取，每篇文章包含數百字，再切割成 3-4 段，每段含至多 231 字，由 200 人朗讀錄製，每人所讀文章皆不相同。在本論文的實驗中，我們所取的部份為交通大學及台灣大學所錄製的音檔。語音訊號取樣頻率為 16kHz，語音訊號量化析度為 16 位元。

公視晚間新聞語料是由中央研究院與公共電視臺共同錄製，主要是公視晚間新聞語音，錄製期間由 2000 年 1 月 11 日到 2000 年 2 月 9 日，總共 120 小時的新聞語料。

4.3 實驗結果與討論

首先，我們使用 TCC300 語料庫進行語者獨立之語音模型訓練，語料數共約 14000 句。訓練所得之語音模型，我們使用 TCC300 中另外未拿來訓練之語料進行測試共 900 句，其語音辨識率為 67.5%。另一份我們所使用之公視晚間新聞語料，初步已整理出三小時語料。所以在接下來的效能測試部份，我們使用此三小時語料進行實驗。我們將使用不同的調適語料量，分別進行最大相似度線性迴歸(MLLR)、最大事後機率線性迴歸(MAPLR)、最小分類錯誤線性迴歸(MCELR)與本論文所提之聚集事後機率線性迴歸(AAPLR)之效能評估。調適之語料量由最短之 2 句，至最長之 30 句，而調適之轉換矩陣類別為 2 類，進行效能之評估，實驗結果如下表所示。

調適方法	調適句數	矩陣類別數	辨識率(%)	調適時間(分鐘)
Baseline	-	-	44.9	-
MLLR	2	2	46.3	2
	5	2	54.1	3
	30	2	56.6	10
MAPLR	2	2	51.5	2
	5	2	54.1	3
	30	2	56.3	10
MCELR	2	2	48.2	2
	5	2	54.1	4
	30	2	56.8	13
AAPLR	2	2	51.5	2
	5	2	54.6	3
	30	2	57.1	11

表一、MLLR, MAPLR, MCELR, AAPLR 在不同調整句數下之辨識率與調整時間比較

首先我們從不同方法的調適效果來比較，可以發現所提出之 AAPLR 與其他調適方法相較，無論給定多少調適語料，均可達到最佳之效能。而與 MCELR 之比較，可以發現最大之效能差距約有 3.3%。另外，由調適時間來比較，可以發現，AAPLR 雖然算是屬於鑑別性調適法則，但是在調適時間上，由於其參數估測有封閉解的存在，可以一次就將調適之最佳參數估測出，所以較同類型之 MCELR 花更短的時間在調適上。另外，由表上可以發現的是，當使用了 30 句調適語料時，所有方法的調適效果並沒有相當大的改進，推測原因應是出在轉換矩陣類別數量上的問題。由於使用之語料數量已不少，但是類別數量還是只有固定在 2 個，過少的轉換矩陣類別數，會使得調適語料無法發揮針對不同模型參數而估測出專屬之轉換矩陣，而失去大量調適語料應有之調適效能改進率。最後，在此初步實驗中，我們直接將 TCC300 所訓練出之語音模型，使用公視語料進行少量語料之調適效能實驗，而未考慮到兩種語料所具備之文句內容與語者分佈的差異。在此實驗結果中，不易區分出調適之效能是來自於針對文句內容的調適效能抑或是來自語者的調適效能。這是在未來我們將會再進行修正之處。

5. 結論與未來工作

在本研究中，我們提出一套具鑑別性訓練特性之快速調適演算法，聚集事後機率線性迴歸(AAPLR)調適演算法。根據最小錯誤率之原則，我們由事後機率出發，定義聚集事後機率函式進行線性迴歸矩陣參數之調適。此調適演算法之優點在於整合最大事後機率的調適演算法則與鑑別式訓練的精神。既可獲得鑑別式訓練必須考量其他類別

與所估測類別參數間之鑑別性法則而得到較傳統最大相似度估測更好的分類正確率，又從推導最終結果之封閉解而可獲得快速調適的效能。在實驗中，我們可以看到無論在任何調適資料量之下，所提出之調適演算法之效能可以比其他同樣基於線性迴歸調適為主之演算法有更好的效能表現。

在本論文中，一般化最小錯誤率中之類別機率以一常數表示，在模型參數估測中較不具參考價值，或許嘗試以真正的機率分佈來代表，可以推導出更完整之結果。此外，我們也將再深入由最基本之理論出發，將此一調適演算法演繹得更加完整。未來我們也將嘗試利用近似貝氏的方法進行理論推導以尋求漸進式調適之效能。此外，除了我們也將增加線性轉換矩陣的類別數，進行更多的實驗以驗證調適效能之外，也要採行先針對訓練與測試語料之文句內容差異進行所謂的 task 調適，以先去除此一因素，再行針對語者調適之效能進行實驗評估，我們也將增加回歸類別數目以及調整語料句數以更有效提高電視新聞語音辨識率。

6. 參考文獻

[1] L. Bahl, P. Brown, P. de Souza and R. Mercer, "Maximum mutual information estimation of hidden Markov model parameters for speech recognition", in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 11, April 1986, pp. 49-52.

[2] P. Beyerlin, "Discriminative model combination", in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 1, 1998, pp. 481-485.

[3] P. C. Chang and B.-H. Juang, "Discriminative training of dynamic programming based speech recognizers", *IEEE Trans. Speech and Audio Processing*, vol. 1, no. 2, pp. 135-143, April 1993.

[4] R. Chengalvarayan, "Speaker adaptation using discriminative linear regression on time-varying mean parameters in trended HMM", *IEEE Trans. Signal Processing Letters*, vol. 5, pp. 63-65, March 1998.

[5] Jen-Tzung Chien, "Quasi-Bayes linear regression for sequential learning of hidden Markov models", *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 5, pp. 268-278, July 2002.

[6] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society (B)*, vol. 39, pp. 1-38, 1977.

[7] M. J. F. Gales and P. C. Woodland, "Mean and Variance adaptation within the MLLR Framework," *Computer Speech and Language*, Vol. 10, pp. 249-264, 1996.

[8] J.-L. Gauvain and C.-H. Lee, "Maximum *a posteriori* estimation for multivariate Gaussian mixture observation of Markov chains", *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, pp. 291-298, April 1994.

[9] X. He, W. Chou, **"**Minimum classification error linear regression for acoustic model adaptation of continuous density HMMs", in *Proc. Int. Conf. Multimedia and Expo (ICME)*, vol. 1, 2003, pp. 6-9.

[10] W. Jian, H. Qiang, "Supervised adaptation of MCE-trained CDHMMs using minimum classification error linear regression**,"** in *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing (ICASSP)*, vol. 1, 2002, pp. I-605 - I-608.

[11] B.-H. Juang, W. Hou and C.-H. Lee, "Minimum classification error rate Methods for Speech Recognition", *IEEE Trans. Speech and Audio Processing*, vol. 5, no. 3 , pp. 257-265, May 1997.

[12] B.-H. Juang and S. Katagirl, "Discriminative learning for minimum error classification," *IEEE Trans. Signal Processing*, vol. 40, pp. 3043-3054, December 1992.

[13] H.-K.J. Kuo, E. Fosle-Lussier, H. Jiang and C.-H. Lee, "Discriminative training of language models for speech recognition", in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 1, 2002, pp. I-325-I-328.

[14] C. J. Leggeter and P.C. Woodland, "Maximum Likelihood Linear Regression for Speaker Adaptation of Continuous Density Hidden Markov Models," *Computer Speech and Language*, 1995, pp. 171-185.

[15] Q. Li, B.-H. Juang, "A new algorithm for fast discriminative training", in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 1, 2002, pp. 97-100.

[16] Q. Li, B.-H. Juang, "Fast discriminative training for sequential observations with application to speaker identification", in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, vol. 2, 2003, pp. 397-400.

[17] R. P. Lippmann, "An intorduction to computing with neural nets", *IEEE ASSP Mag*., pp. 4-22, April 1987.

[18] E. McDermott and S. Katagiri, "Shift-invariant multi-category phoneme recognition using kohonen's LVQ2," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1989, pp. 81-84.

[19] A. Sankar and C.-H. Lee**, "**A maximum-likelihood approach to stochastic matching for robust speech recognition", *IEEE Trans. Speech and Audio Processing*, vol. 4, no. 3, pp. 190-202, May 1996.

[20] R. Schlüter, W. Macherey, B. Müller and H. Ney, "A combined maximum mutual information and maximum likelihood approach for mixture density splitting", in Proc. *EUROSPEECH*, vol. 4, 1999, pp. 1715-1718.

[21] O. Siohan, C. Chesta, and C.-H. Lee. "Hidden Markov model adaptation using maximum a posteriori linear regression." in *Workshop on Robust Methods for Speech Recognition in Adverse Conditions*, Tampere, Finland, 1999.

[22] S. Young, J. Jansen, J. Odell, D. Ollason, and P Woodland. *The HTK Book (Version 2.0)*. ECRL, 1995.