

## 第1章 序論

隨著時代的演進，如家庭個人電腦的普及、網際網路服務的興起、無線數位通訊的發展，電腦、手機等日新月異的電子產品已經改變我們的日常生活。這些設備的功能越來越多樣化，加上裝置有迷你化的趨勢，手動操作的確有諸多不便。對人類來說，語音是最自然且直接的溝通方式，如果我們能直接透過語音操作電腦設備，且電腦能夠分析理解我們的語音要求，做因應的處理，將會節省許多人力與時間。例如智慧型問答與對話系統(Intelligent Question-Answering and Dialogue System)是一個理想的環境。要達到這樣的目標，我們一般的處理步驟是先對使用者描述問題的語音進行連續語音辨識，轉換成文字。有了初步的文字結果後，再透過各種技術進一步分析語句的語意(Semantics)。得到最合適的結果後，再透過文字轉語音(Text-to-Speech, TTS)的語音合成(Speech Synthesis)技術輸出語音，回應使用者。可想而知，首要的語音辨識正確率對後續的語意分析工作影響甚鉅，如果辨識錯誤太多，使用者的語意可能沒辦法正確地透過文字呈現，甚至變成截然不同的語意。

除了對話系統外，語音辨識亦有許多應用層面，例如語音文件檢索與摘要(Spoken Document Retrieval and Summarization)，都需要先經過語音辨識系統辨識，得到對應的轉寫文字(Transcription)資訊後，再進行相關的分析與處理。所幸，因為資訊科技的發展，電腦晶片處理的速度與日俱增，對於辨識技術及其應用的整合與實現而言，的確是一大助益。我們亦相信，透過專家學者的共同努力，語音辨識於現實生活的應用將會更加成熟。

本章主要針對研究需要，先對語音辨識做一簡介，然後說明本論文研究內容與方向。

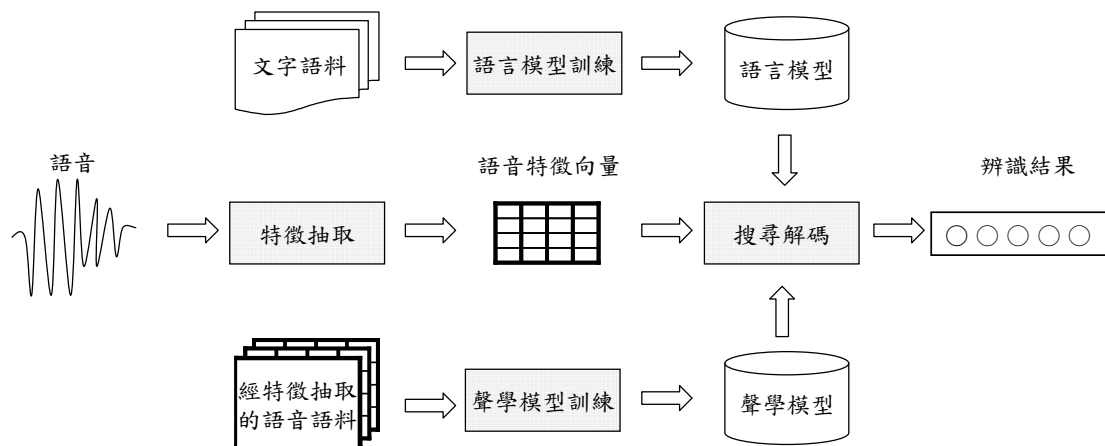


圖 1-1 自動語音辨識流程圖

### 1.1 研究背景-語音辨識

自動語音辨識(Automatic Speech Recognition, ASR)系統架構為給定一段語音  $X$ ，欲得到一句最可能符合要求的詞序列  $W^*$ ，通常我們會使用統計式模型及最大事後機率(Maximum a Posteriori, MAP)法則進行語音辨識[Jelinek 1999]：

$$\begin{aligned}
 W^* &= \arg \max_w P(W | X) \\
 &= \arg \max_w \frac{P(W)p(X | W)}{p(X)} = \arg \max_w P(W)p(X | W)
 \end{aligned}
 \tag{1-1}$$

$P(W | X)$  即是給定語音  $X$  時，詞序列  $W$  的事後機率。經過貝氏定理轉換後，可以表示成三個機率， $p(X | W)$ 、 $P(W)$  與  $p(X)$ 。其中  $p(X | W)$  代表聲學模型 (Acoustic Model, AM) 產生語音  $X$  的機率密度 (Probability Density)，直接估測語音訊號  $X$  發生在詞序列  $W$  中的音節 (Syllable)、音素 (Phoneme) 或更小發音單位對應模型的相似度 (Likelihood)； $P(W)$  代表語言模型 (Language Model, LM) 計算詞序列  $W$  的機率，用於評估詞序列  $W$  於自然語言的合理性，可視為詞序列  $W$  的事前機率，輔助解決聲學上的混淆，使得最後選擇出的詞序列  $W^*$  能夠符合該語言的形式； $p(X)$  表示語音  $X$  的事前機率密度，然而，我們是針對某句語音  $X$  進行辨識，對於每一條詞序列  $W$  都一樣同除以  $p(X)$ ，故可以忽略。語音辨識亦可視為分類

(Classification)問題，透過聲學模型與語言模型的使用，從所有可能的詞序列 $W$ 集合中，指派最可能的詞序列 $W^*$ 給語音訊號 $X$ 。

語音辨識過程大致可分成幾個部份，特徵抽取(Feature Extraction)、聲學及語言模型訓練(Acoustic and Language Model Training)及搜尋解碼(Searching and Decoding)，如圖 1-1 所示。

(一) 特徵抽取：特徵抽取是將語音訊號 $X$ 轉換成語音辨識系統可容易分析、使用的資料型態，如特徵向量(Feature Vectors)。特徵抽取的目的是將語音特性表現於特徵向量上，例如被廣泛使用的梅爾倒頻譜係數(Mel-frequency Cepstral Coefficients, MFCC)，能夠模擬人類聽覺感知特性[Davis and Mermelstein 1980]；或是鑑別式特徵，如線性鑑別分析(Linear Discriminant Analysis, LDA)等，具備區分不同發音特性的能力[Duda and Hart 1973]。有了特徵之後，可以進一步研究因為環境變化或噪音干擾問題使得語音特徵向量表現能力受限的語音強健性(Speech Robustness)問題，例如倒頻譜平均消去法(Cepstral Mean Subtraction, CMS)[Atal 1974]、倒頻譜正規化法(Cepstral Normalization, CN)[Viikki and Laurila 1998]或統計圖等化法(Histogram Equalization, HEQ)[Korkmazsky et al. 2004; Lin et al. 2006]等語音強健方法在過去幾年間被提出。

(二) 聲學模型及訓練：一般而言，我們會使用隱藏式馬可夫模型(Hidden Markov Model, HMM)來建立聲學模型，且因為語音有時序性，所以通常是由左至右(Left-to-right)的模型[Rabiner 1989]。隱藏式馬可夫模型包含數個狀態(State)，對於模型中的每一個狀態而言，都有所屬的轉移機率分布(Transition Probability Distribution)，用以控制狀態停留或是轉移；每個狀態亦有各自的觀測機率分布(Observation or Emission Probability Distribution)，觀測向量即為語音特徵向量，通常每維度皆為實數連續值，所以我們採用多變量高斯混合分布(Multivariate Gaussian Mixture Distribution)來模擬特徵向量的機率分布，即高斯混合模型(Gaussian Mixture Model, GMM)。因此，此種聲學模型又稱為連續密度隱藏式馬

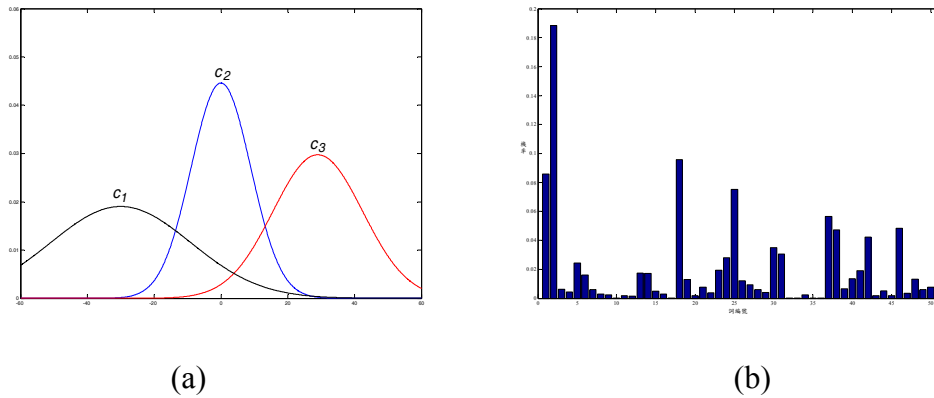


圖 1-2 (a) 高斯混合機率分布 (b) 多項式機率分布

可夫模型(Continuous Density Hidden Markov Model, CDHMM)。例如 圖 1-2(a)可視為是由三個高斯分布及其對應權重  $c_i$  組成的單一維度高斯混合模型。除了隱藏式馬可夫模型，近年來亦有學者提出使用最大熵值馬可夫模型(Maximum Entropy Markov Model)[Kuo and Gao 2004]及隱藏式條件隨機場(Hidden Conditional Random Field, HCRF)[Gunawardana et al. 2005]建立聲學模型。

在中文語音辨識裡，中文音節結構可視為由兩個次音節(Subsyllable)，即一個聲母(Consonant)與一個韻母(Vowel)所組成，所以需要分別建立對應的聲學模型，我們稱為INITIAL與FINAL模型。有了模型之後，最常用的訓練方法為最大化相似度(Maximum Likelihood, ML)訓練法則[Bahl et al. 1983]，或是鑑別式訓練法則，如最大化交互資訊(Maximum Mutual Information, MMI)[Bahl et al. 1986]、最小化分類錯誤(Minimum Classification Error, MCE)[Juang and Katagiri 1992]及近年來被提出且相當有效的最小化音素錯誤(Minimum Phone Error, MPE)[Povey 2004]等。

(三) 語言模型及訓練：在語言模型方面，詞序列機率  $P(W) = P(w_1, \dots, w_n)$  是採用離散樣本(Discrete Samples)的機率分布， $w_i$  表示詞序列  $W$  的第  $i$  個詞。不像聲學模型使用機率密度函數(Probability Density Function, PDF)，語言模型直接估測詞序列的機率質量機率函數(Probability Mass Function, PMF)，例如 圖 1-2(b)是詞典大小為 50 且僅考慮一個詞的多項式機率分布  $P(w)$ ，對於X軸上的每個詞  $w$ ，Y

軸都有對應的機率值。如果針對每一種可能的詞序列建立模型，所需要的參數空間為 $|V|^n$ 成指數成長， $|V|$ 為詞典大小， $n$ 為詞序列長度，所以我們需要限制語言模型的規模。目前最廣泛使用的是 $N-1$ 階馬可夫假設( $N-1$  order Markov Assumption)，即 $N$ 連( $N$ -gram)詞語言模型。假設詞 $w_i$ 只與其前 $N-1$ 個詞有關，則詞序列機率可表示成：

$$P(w_1, \dots, w_n) = \prod_{i=1}^n P(w_i | w_1, \dots, w_{i-1}) \approx \prod_{i=1}^n P(w_i | w_{i-N+1}, \dots, w_{i-1}) \quad (1-2)$$

除了 $N=1$ 的單連(Unigram or Monogram)詞模型之外，目前最常用的是 $N=2$ 的二連(Bigram or Digram)及 $N=3$ 的三連(Trigram)詞模型。在文獻中也說明了三連語言模型能夠簡單使用且有效地提升辨識效果[Goodman 2001]。近年來因為電腦硬體發展，四連(Fourgram or Quadrigram or Tetragram)或五連(Fivegram or Quintagram)詞語言模型也慢慢地被普遍使用。

$N$ 連詞模型的基本訓練方式為最大化相似度估測(Maximum Likelihood Estimation, MLE)，可以直接由訓練語料的 $N$ 連詞頻數(Word Count)估測 $N$ 連詞機率分布，以三連詞為例：

$$P(w_i | w_{i-2}, w_{i-1}) = \frac{C(w_{i-2}, w_{i-1}, w_i)}{C(w_{i-2}, w_{i-1})} \quad (1-3)$$

$C(\cdot)$ 表示詞序列出現的詞頻數。然而使用最大相似度估測會遇到資料稀疏(Data Sparseness)的問題，因為訓練語料的不足，沒辦法收集到所有可能的 $N$ 連詞頻數。在樣本不足的情況下，估測得到的機率不夠可靠，例如在計算一段詞序列機率時，若當中有一個 $N$ 連詞條件機率值為0，會使得整個詞序列機率為0，導致辨識錯誤。為了解決資料稀疏的問題，有許多平滑化(Smoothing)技術陸續被提出，例如Katz平滑化法[Katz 1987]、Witten-Bell平滑化法[Witten and Bell 1991]或是Kneser-Ney平滑化法[Kneser and Ney 1995]等，透過詞頻數統計資訊，使得在訓練語料中有出現的 $N$ 連詞折扣(Discount)部分次數給在訓練語料中未出現過的 $N$ 連

詞，以解決機率值為 0 的情況[Chen and Goodman 1999]。

(四) 搜尋解碼：進行搜尋解碼時，我們會同時使用到特徵向量、聲學模型及語言模型。藉由特徵向量在聲學模型上的相似度及對應詞序列的語言模型機率，我們可以找出最可能符合輸入語音的詞序列。一般我們會使用維特比動態規劃搜尋(Viterbi Dynamic Programming Search)[Viterbi 1967]，透過此搜尋法，可以較有效率地找出可能的詞序列。此外，為了降低搜尋空間，加快辨識速度，我們通常也會使用聲學或語言模型前看(Acoustic or Language Model Look-ahead)[Aubert 2002;Chen *et al.* 2004a]技術進行裁減(Pruning)的動作，將可能性較低的詞先行排除。

早期的語音辨識系統功能簡單，能夠辨識的詞彙較少，例如數字辨識(Digit Recognition)或是獨立詞辨識(Isolated Word Recognition)等，所以需要的運算量及搜尋空間都較小。近年來，大詞彙連續語音辨識已經是先進且炙手可熱的研究(State-of-the-art)之一；除了研究電話交談語音(Conversational Telephone Speech, CTS)及廣播新聞(Broadcast News)辨識，近來也針對自發性語音(Spontaneous Speech)、課程教學(Lecture)辨識或是語言夾雜(Code-Mixing)的語音辨識進行研究。經過語音辨識系統辨識後，我們可以將語音轉換成文字。透過文字的方式，我們可以較容易進一步地分析其語意。由先前介紹得知，在語音辨識的基本架構中，語言模型扮演重要的角色。值得一提的是，對於辨識後的文字資訊使用，語言模型仍可以再次被應用。換句話說，代表自然語言規則或語意的語言模型，可以應用於各種自然語言處理的領域中，例如資訊檢索(Information Retrieval)[Chen *et al.* 2004b]、文件摘要(Document Summarization)[Chen *et al.* 2007]、機器翻譯(Machine Translation)[Brown *et al.* 1993]等。而本論文主要研究各種層次語言資訊的語言模型應用於語音辨識。

## 1.2 研究內容-語言模型演進

語言模型裡最基本的是 $N$ 連詞模型，其想法可以追溯於Claude Elwood Shannon在資訊理論(Information Theory)研究中所提出的問題：給定一個字母序列，下一個字母最有可能會是什麼？這個問題可以藉由從訓練資料中求得給定歷史序列下所有字母的機率分布解決。明確地說，如果字母只與最後 $N-1$ 個字母有關，即歷史序列是為前 $N-1$ 個字母，則定義為 $N$ 連詞模型，這就是馬可夫假設，後來延伸成以詞為單位。 $N$ 連詞模型已被廣泛應用，但其缺點是僅能使用短距離的詞彙資訊，所以過去到現在陸續都有專家學者提出不同的模型來改善 $N$ 連詞模型。語言模型的演進如圖 1-3 所示。需要注意的是，有些模型可能最早被提出應用於其他領域，如資訊檢索等。在這邊，我們選擇呈現最早被應用在語音辨識相關研究上的時間。除此之外，有更多的研究是在於對已有的模型進行改善。

Kuhn在1988年提出了快取模型(Cache Model)來輔助 $N$ 連詞模型，並於1990年發表[Kuhn and Mori 1990]。快取模型的概念是如果我們講了一些詞，則一段時間內這些詞再次出現的機率會很高。以單連詞為例，一種作法是先訓練好一個常出現詞的單連模型，當辨識過程中又出現時，可以直接使用；另一種作法是記錄歷史詞序列曾經出現過的詞，當接下來的辨識過程中再次出現時，加重比重。1990年，Brown等人提出了 $N$ 連類別模型(Class-based  $N$ -gram Model)，並於1992年發表[Brown et al. 1992]。 $N$ 連類別模型是為了改善 $N$ 連詞模型的資料稀疏問題，透過詞的分類，可能性從詞典大小降為詞類別種類，可以降低 $N$ 連詞模型的參數。詞分類的方式主要是使用貪婪演算法(Greedy Algorithm)最大化相鄰詞類別交互資訊(Mutual Information)。

1992年，Rosenfeld於他的博士論文計畫書提出將最大熵值法(Maximum Entropy Principle, ME)應用於語言模型，並於1994年完成論文，且於1996年發表[Rosenfeld 1996]。首先需要知道，一個隨機變數的熵值越大，表示其機率分布越平均。最大熵值法主要是給定某些已知資訊特徵的限制下，希望模型機率分布

的熵值越大越好，即希望其他未知資訊的機率分布能夠越平均。訓練方式則可採用改善迭代調整法(Improved Iterative Scaling, IIS)或是廣義迭代調整法(Generalized Iterative Scaling, GIS)。透過最大熵值法架構，能夠整合不同的自然語言資訊特徵，例如在 1993 年Rosenfeld與Lau共同提出使用觸發對(Trigger Pair)資訊於最大熵值法模型的觸發對語言模型[Lau et al. 1993]。

同樣在 1993 年，略詞模型(Skipping Model)被提出[Huang et al. 1993]。 $N$ 連詞模型需要以連續前  $N-1$  個詞為條件，估測目前的詞，當  $N$  越大，訓練語料出現的  $N$  連詞組合越少。略詞模型可視為是長距離  $N$  連詞(Long-distance  $N$ -gram)模型，即允許略過某個距離內的詞，使用更早出現的詞，再訓練新的  $N$  連詞模型，如Huang分析不同略詞距離的情況下所訓練出的略二連詞模型。

1997 年，聚合式馬可夫模型(Aggregate Markov Model, AMM)與混合階層馬可夫模型(Mixed-order Markov Model)被提出[Saul and Pereira 1997]。聚合式馬可夫模型亦可視為是一種二連類別(Word Class)語言模型，有別於  $N$  連類別模型需要將詞一對一對應到某詞類別，聚合式馬可夫模型允許每個詞根據不同的比重屬於不同的類別。透過這種柔性(Soft)分類，聚合式馬可夫模型比  $N$  連類別模型更有彈性。混合階層馬可夫模型可視是略詞二連模型的延伸，除了使用略詞二連模型捕捉不同距離的詞與詞關係之外，也同時估測不同略詞距離的影響。

同樣在 1997 年，結構化語言模型被提出[Chelba 1997]。結構化語言模型主要是使用語句結構資訊。首先針對歷史詞序進行剖析(Parsing)，得到詞剖析字首(Word-parse Prefix)，每個詞剖析字首頂點代表那個段落的主導詞(Headword)。根據主導詞與歷史詞，可以進一步估測目前詞的機率。除此之外，使用輸入層、隱藏層及輸出層的類神經網路模型(Neural Network Model)亦於 2000 年應用於語言模型，並在 2003 年被應用於結構化語言模型，其主要概念是使用連續空間建立模型[Bengio et al. 2003; Schwenk and Gauvain 2002; Xu et al. 2003]。

潛藏語意分析(Latent Semantic Analysis, LSA) 原先是用於資訊檢索，在 1997 年時被應用於語言模型上[Bellegarda 1997]。潛藏語意分析主要是對詞與文件建

立潛藏的語意關係，並使用線性代數方法中的奇異值分解(Singular Value Decomposition, SVD)達到此目的。由於線性代數方法不易調整參數，所以 1999 年時，Hoffmann提出了機率式潛藏語意分析(Probabilistic Latent Semantic Analysis, PLSA)用於資訊檢索，後來也用於語言模型上[Glidea and Hofmann 1999]。與潛藏語意分析相比，機率式潛藏語意分析使用較有彈性的機率架構建立模型。此外，在潛藏語意分析中，奇異值分解後的矩陣包含負數，對於語意的定義不明確，所以使用不同矩陣分解方式的非負矩陣分解(Non-negative Matrix Factorization, NMF)於 2001 年被應用於語言模型[Novak and Mammone 2001]。

2003 年，分散式馬可夫模型(Distributed Markov Model, DMM)被提出[Blitzer *et al.* 2003]。分散式馬可夫模型主要是改進聚合式馬可夫模型(AMM)只能使用二連詞的限制，透過將參數設定成二元值(Binary)向量，簡化模型所需的參數量，進而延伸歷史詞長度。2004 年，另一種觸發對模型被提出[Troncoso *et al.* 2004]。有別於使用最大熵值法整合資訊，這類模型先收集出現在同一窗內的詞配對(Word Pairs)統計值，然後直接估測觸發對的條件機率。再透過一些選擇觸發對的機制，如詞頻數與反文件頻數(Term Frequency and Inverse Document Frequency)保留部分觸發對，最後使用插補法(Interpolation)與 $N$ 連詞模型結合。2005 年，原本於資訊檢索使用的潛藏狄利克雷分配(Latent Dirichlet Allocation, LDA)被應用於語言模型[Tam and Schultz 2005]。潛藏狄利克雷分配可視為是潛藏語意分析(PLSA)的延伸。潛藏狄利克雷分配使用了額外的超參數(Hyper-parameters)來當作模型參數的事前機率，並且使用貝氏估測(Bayes Estimation)。換句話說，潛藏語意分析是使用最大事後機率估測潛藏狄利克雷分配的模型參數，且其模型參數事前機率為平坦(Flat)的特例。

2007 年，Afify提出了高斯混合語言模型(Gaussian Mixture Language Model, GMLM)[Afify *et al.* 2007]。高斯混合語言模型與類神經網路模型相同，都是使用連續空間建立模型，且高斯混合語言模型更容易訓練與調適。高斯混合語言模型已經廣泛用於聲學模型，而用於語言模型上也是類似的訓練方式，主要是將預測

詞視為一個模型，而不同的歷史詞序列先透過降維技術降成可接受的特徵向量，當作是詞模型的訓練觀測向量，再使用最大相似度訓練法則估測模型參數，惟在語音辨識實驗裡，高斯混合語言模型進步幅度不大。

其他的整合資訊模型，如整句最大熵值法(Whole-Sentence Maximum Entropy Model, WSME)於 1997 年提出[Rosenfeld 1997]。傳統的最大熵值法估測條件機率，採用預測詞與歷史詞序列相關的特徵，所以模型所需的正規化因子(Normalization Factor)會相依於歷史詞序列，不同的歷史詞序列都需要計算，所以計算量很大，而整句最大熵值法是使用整個語句的特徵，並直接估測語句機率，所以正規化因子對所有語句皆相同。潛藏最大熵值法(Latent Maximum Entropy Principle, LME)亦於 2001 年提出[Wang et al. 2001]。潛藏最大熵值法將隱藏變數引入最大熵值法之中，用來表示不確定性的特徵，如語意特徵、文法特徵等。也因為有隱藏變數，需要先使用期望值最大化法(Expectation Maximization, EM)[Dempster *et al.* 1977]估測隱藏變數，再使用傳統最大熵值法的訓練方法。

我們可以發現，隨著時間的演進，語言模型的研究從早期的詞層次到詞類別層次，再到語句結構層次，最後是到文件主題層次，從具體的文字到抽象的語意，越來越高階的資訊被使用，也表示語言模型的研究發展目的是希望能夠呈現人類的思維。

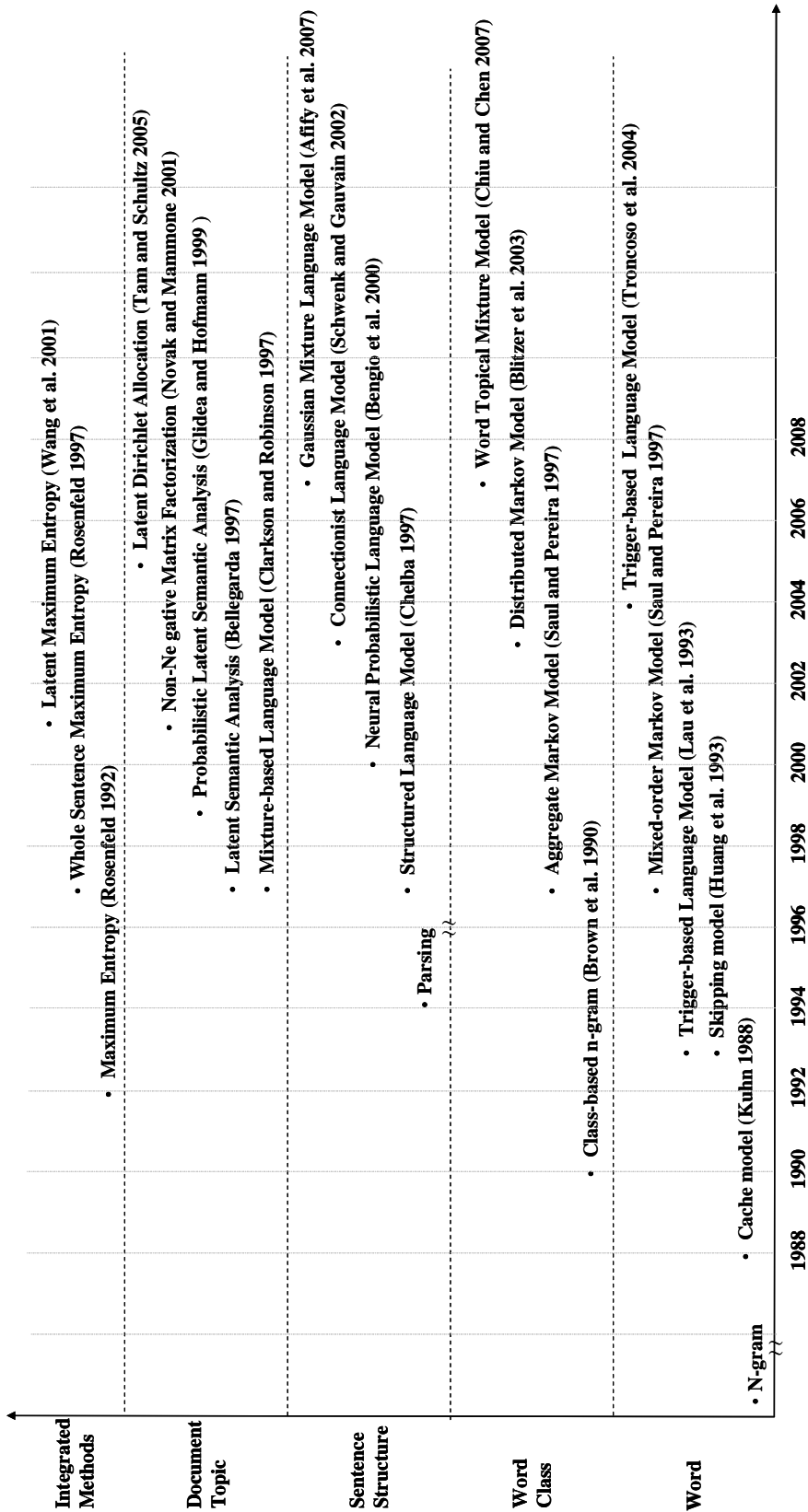


圖 1-3 語言模型之演進

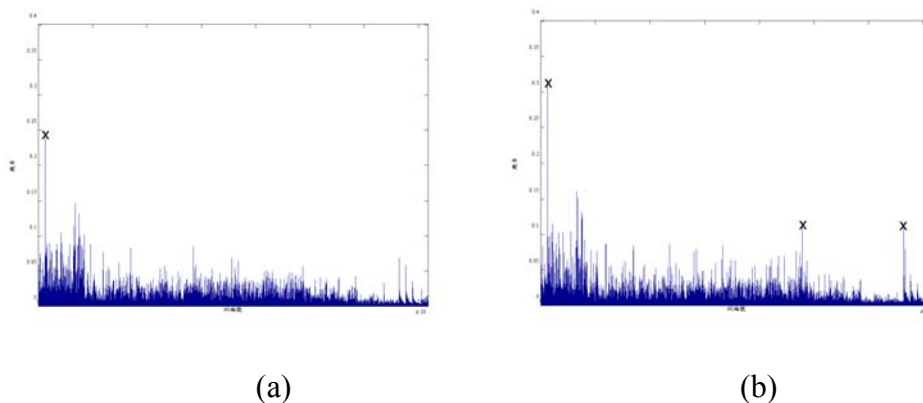


圖 1-4 單連詞機率分布於 (a) 訓練語料 (b) 調適語料

### 1.3 研究內容-語言模型調適

什麼是語言模型調適(Language Model Adaptation)? 簡單地說, 語言模型調適就是根據一些資訊或原則, 重新調整語言模型的機率分布。然而, 為什麼需要語言模型調適? 因為人類使用的自然語言具有高變化度的特性。首先, 隨著時間過去, 新詞不斷地被產生, 如生物或資訊科技使用的複合詞等, 所以詞典一直隨著時間的改變而增加; 而對於新聞事件來說, 不同時期發生的事件差異也很大。另外, 對於同一用詞而言, 用於不同領域也有不同的意義與出現頻率, 即一詞多義的情況; 對於不同的領域, 詞的基本特徵及語意用法亦不同, 例如體育與政治新聞, 常用術語差異很大。再者, 人們會針對不同的人或時機自然地調整語言的用法, 例如, 寫一般信件或正式的公文會採用不同的文法結構; 同學間聊天或跟上司報告, 用字遣詞也會不同[Bellegarda 2004]。因此, 對於語音辨識系統來說, 如果想辨識不同類型的語音, 適當的語言模型調適技術能夠有效地提高系統辨識正確率。

進行語言模型調適時, 語言模型使用的語料可分為兩部份, 訓練語料及調適語料。訓練語料通常是大量語料, 且主題性較為平均, 表示詞分布亦平均, 可能與測試語料較不相似。調適語料通常資料較少而與測試語料有一定的相關性, 可能是屬於某一個主題, 或是同一段時期。圖 1-4 是論文中所採用的訓練語料及調

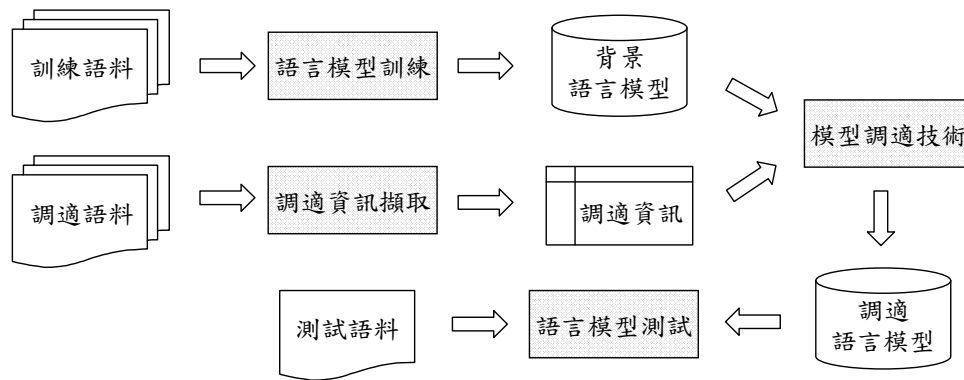


圖 1-5 語言模型調適流程圖

適語料未經平滑化的單連詞機率分布，X軸表示詞的編號，Y軸是對應的機率值。兩個分布大致相似，都有一個最高點，代表的是「的」這個功能詞(Function Word)機率。但是仔細觀察後可以發現，訓練語料的機率分布較為平均，而調適語料有比較多的起伏，表示某些詞機率比較大，代表某種主題性。例如，最後兩個較高點的分別代表「我們」與「這個」的機率。因為調適語料是廣播新聞語料，所以常出現這類口語。有了這兩套語料後，可以對訓練語料進行統計式估測產生背景語言模型分布，而對調適語料進行資訊擷取，找出可用的調適資訊；然後再透過一些語言模型調適技術將調適資訊適當地加入到背景語言模型，得到經調適後的語言模型，最後可以用測試語料驗證新模型的品質，流程如圖 1-5 所示。

根據語料時期性或領域性的不同，亦可分成幾種調適方式：同領域語言模型調適 (Within-Domain Language Model Adaptation)、跨領域語言模型調適 (Cross-Domain Language Model Adaptation) 及同時期語言模型調適 (Contemporary Language Model Adaptation)。同領域調適表示背景、調適及測試語料三者來源相同，時期不同；跨領域調適表示背景語料與測試語料來源不同，而調適語料與測試語料來源相同，例如背景語料是文字新聞而調適與測試語料為廣播新聞；同時期調適指的是調適與測試語料為同一時期，而來源不同。簡而言之，語言模型調適的目的就是要調整背景語言模型的機率分布，使其應用在測試語料上能有更好的效果[Rosenfeld 2000]。

## 1.4 研究成果

語言模型的應用層面很廣，所期望包含的資訊更是多樣化，本論文首先探討使用不同層次資訊的語言模型，如詞相關語言模型(Word-based Language Models)、詞類別相關語言模型(Word Class-based Language Models)與文件主題相關語言模型(Document Topics-Based Language Models)等，並應用於語音辨識中的語言模型調適。本論文主要提出詞主題混合模型(Word Topical Mixture Model, WTMM)[Chiu and Chen 2007]及位置相關語言模型(Position-Dependent Language Models)應用於語言模型調適。詞主題混合模型將詞視為一模型單位，透過隱藏的詞主題，預測另一個詞的機率。在語音辨識過程中，歷史詞序列可視為一個複合式詞主題混合模型(Composite WTMM)，透過線性結合歷史詞序列中的歷史詞模型，可以直接估測辨識詞的出現機率。位置相關語言模型主要探討詞在文件的位置與詞在語句位置的資訊，根據其所在位置不同，我們可以將位置資訊加入現有的模型，例如 $N$ 連詞模型或潛藏語意分析(PLSA)模型等。除了語言模型應用於語言辨識外，本論文也嘗試將語言模型技術應用於語音文件摘要[Chen et al. 2007]。我們將語音文件摘要視為語句排名的問題，並且可以用機率生成架構(Probabilistic Generative Framework)來表示。語音語句有許多統計資訊可以使用，如聲韻資訊(Prosodic Information)，辨識信心度分數(Confidence Score)，語言模型分數等，而本論文嘗試使用整句最大熵值(Whole Sentence Maximum Entropy, WSME)模型整合這些資訊，並估測語句事前機率。

## 1.5 論文架構

本論文接下來的章節概要如下：第二章先介紹相關的實驗架構，包括大詞彙連續語音辨識系統、實驗語料及評估方式等；第三章將回顧過去相關的語言模型研究，主要著重於使用不同語言資訊的統計式模型；第四章提出詞主題混合模型，探討詞與詞之間的關係，亦提出了位置相關語言模型，進一步呈現位置資訊；第

五章嘗試將語言模型技術應用於語音文件摘要，以機率生成架構為主，提出了以整句最大熵值模型作為語句事前機率模型；第六章則是結論與未來展望，探討未來可研究的方向。

