線上新聞語音檢索系統

陳江村 羅瑞麟 張智星
國立清華大學 資訊工程系
新竹市光復路二段 101 號
E-mail : {jtchen,roro,jang}@wayne.cs.nthu.edu.tw
TEL: (03)5715131-3582

摘要：
在此報告中，我們實作了一個結合隱藏式馬可夫模型(Hidden Markov Model, HMM)為基礎的 HTK(HMM Toolkit)和網頁資料檢索技術的線上新聞語音資料檢索系統。一般的網頁資料檢索(如 google)須使用者輸入相關文字，才得以文字比對方式進行檢索。在此我們則嘗試加入語音辨識的技術讓使用者更易進行檢索。本系統分成新聞前處理及語音查詢兩階段。在辨識內容固定，高準確度的辨識結果下，本系統特別適用於手機、PDA、嵌入式系統等小型、不易以手操作輸入的裝置。本系統亦經清大盲友會的盲人朋友試用，反應十分良好。

關鍵詞:語音辨識、資料檢索、HMM、Viterbi Search、新聞檢索。

1 前言

目前的網際網路中，www.google.com[6]是每個人都不可或缺的工具，其提供的準確性和資料的可用性一直為人稱道，堪稱為文字檢索的翹楚，也因此，網際網路上的資訊成了一個無所不包的資料庫。而在此同時，相關的多媒體檢索技術也相繼發表[1]，顯示了多媒體方面的檢索需求。而由於語音的便利性和可用性(相對於以內容為主的影像檢索)，語音方面的檢索方法已成了多媒資訊檢索的重要研究。

語音檢索的方法可分為兩種，一為語音文件檢索(Spoken Document Retrieval)，一為語音文字檢索(Speech Recognition and Retrieval)。前者不考慮到語音模型，直接以語音的特徵參數，在另一語音文件中進行比對，希望找出最接近的語音內容。這樣的檢索方式雖可跨越語言模型，但在長時間的語音文

件中，辨識率並不高，並不符合高準確性的要求[11]。而語音文字檢索，則是在特定語音模型下，先進行語音辨識，再以辨識出來的文字進行檢索[3]。由於目前文字比對技術成熟，故此法的關鍵在於語音辨識的好壞，辨識文字內容則可十分廣泛。

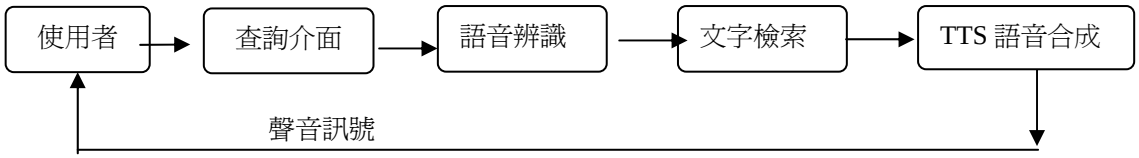
而所謂的語音辨識，主要是用來辨識出聲音的文字內容為何。一般來說，語音辨識的辨識成功率和辨識內容的範圍有很大關係。在大領域的文字辨識中，辨識結果往往出現相近但非正確的答案，這在目前仍是很難克服的問題。

目前此領域最有名的模型為 HMM。藉由特定語言語料的訓練，我們可以利用 HTK[7](Hidden Markov Model ToolKit)實作出某一特定領域的高準確度的語音辨識系統，比方說唐詩三百首的語音辨識，其準確率接近九成九[10]。

綜合以上觀點，我們實作了一個結合 HTK 和新聞網頁內容的檢索技術，希望能達到一個以語音為基礎的 News Google。

2　語音新聞檢索理論背景

本系統使用了語音辨識、文字比對和語音合成三種技術。流程圖如下：

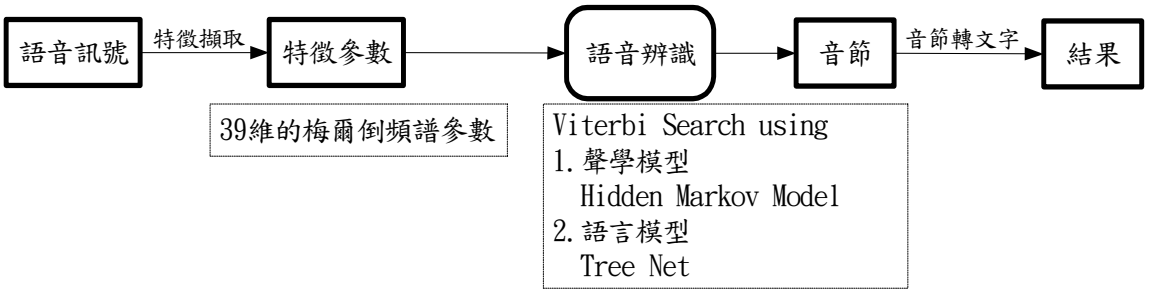


圖表 1　語音新聞檢索服務流程

文字比對部份，由我們只對新聞標題部份作比對，因此以下以語音辨識和合成作為介紹重點。

2.1 語音辨識部份

一般而言，語音辨識的演算法流程如下圖所示：

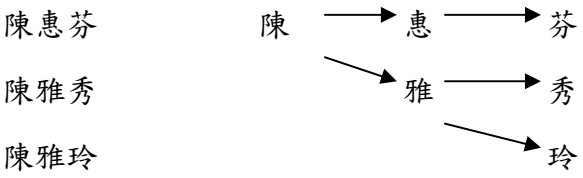


圖表 2 語音辨識的演算法

其中所採用的方法和理論如下：

1. 語言模型 Tree Net

把每個單音節視為一個節點，節點和節點間相連關係的樹狀結構。圖例如下：

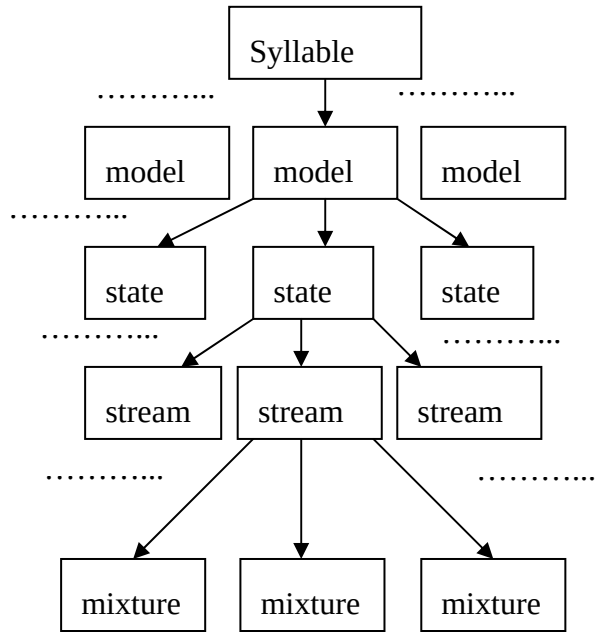


圖表 3 Tree Net 檔示意

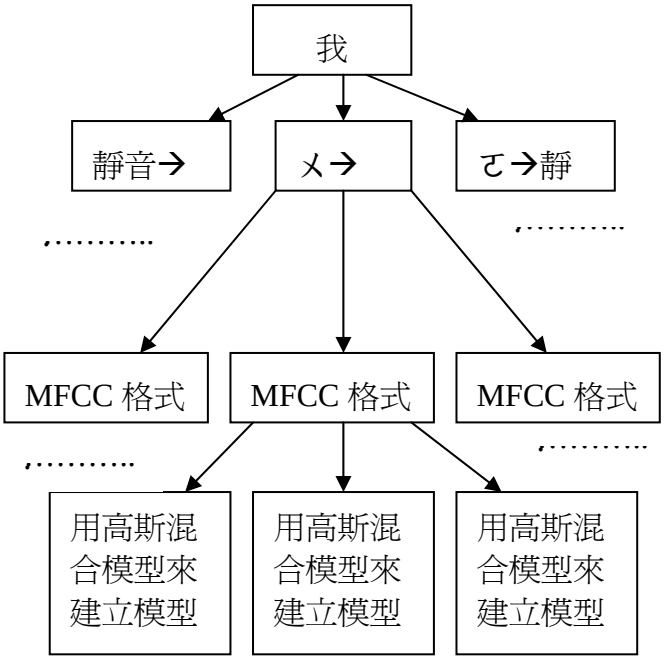
2. 聲學模型 Hidden Markov Model

隱藏式馬可夫模型基本上是一種雙重隨機過程，而之所以稱為隱藏式是因為其中有一組隨機過程是隱藏的，看不見的，在語音中就如同人類在發聲的過程中其發聲器官狀態變化是看不見的，好比喉嚨、舌頭與口腔的變化是不可能從可觀測的語音訊號序列看出來的。而另一組隨機過程稱為觀測序列(observation sequence)，它是由狀態觀測機率(state observation probability)來描述在每個狀態下觀測到各種語音特徵參數的機率分佈。HMM 的狀態觀測機率函式$b_j(o_t)$是採用高斯混合密度函數或稱高斯混合模型(Gaussian Mixture Model, GMM)來計算連續機率密度，因此每一個聲音單元(Model)皆有一組 Continuous HMM 參數。

圖表 4 為 Model, State, Stream 和 Mixture 的階層示意圖，圖表 5 則以”我”此一 syllable 為例，示範 CHMM 的建立方式。



圖表 6　Model, State, Stream 和 Mixture 示意圖



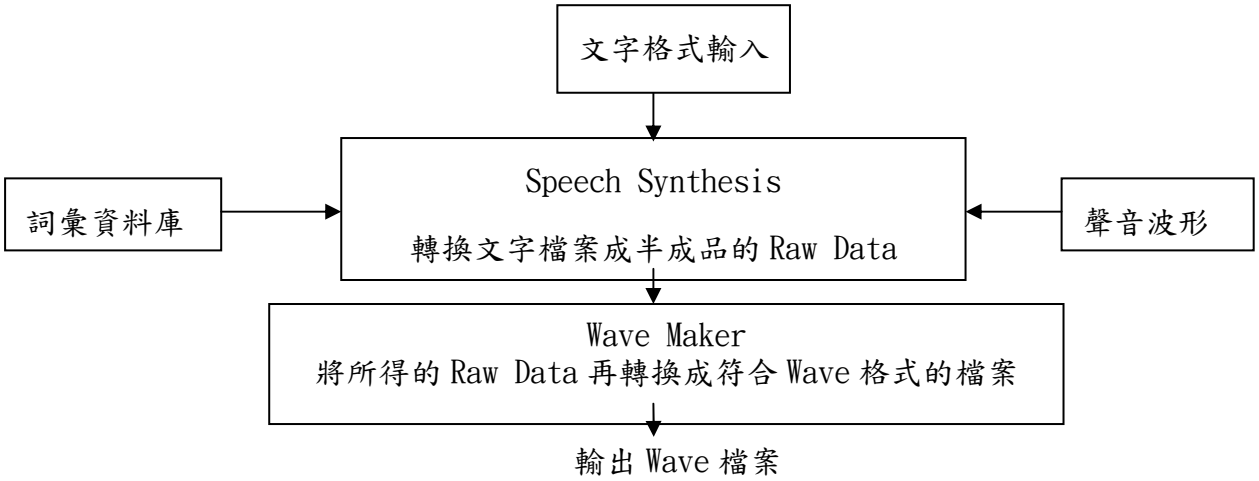
圖表 7　以 GMM 建立 syllable 的 CHMM 流程示意圖

3.　辨識方法

我們根據 Tree Net 的路徑進行以 Viterbi Search，以辨識出機率最高的路徑。其中我們也加上了 Beam Search(光束搜尋法)的作法以進行加速[2]。光速搜尋法在搜尋過程中會慢慢丟棄低機率的搜尋目標，使得愈後面的比對速度會愈加快，此法可有效減少搜尋時間且不會犧牲太多準確性[5]。

2.2　語音合成

在輸出方面我們使用和黃紹華老師合作的語音合成技術。此合成方式是連接式的合成為基礎(Concatenation-Based)，基本流程如下：

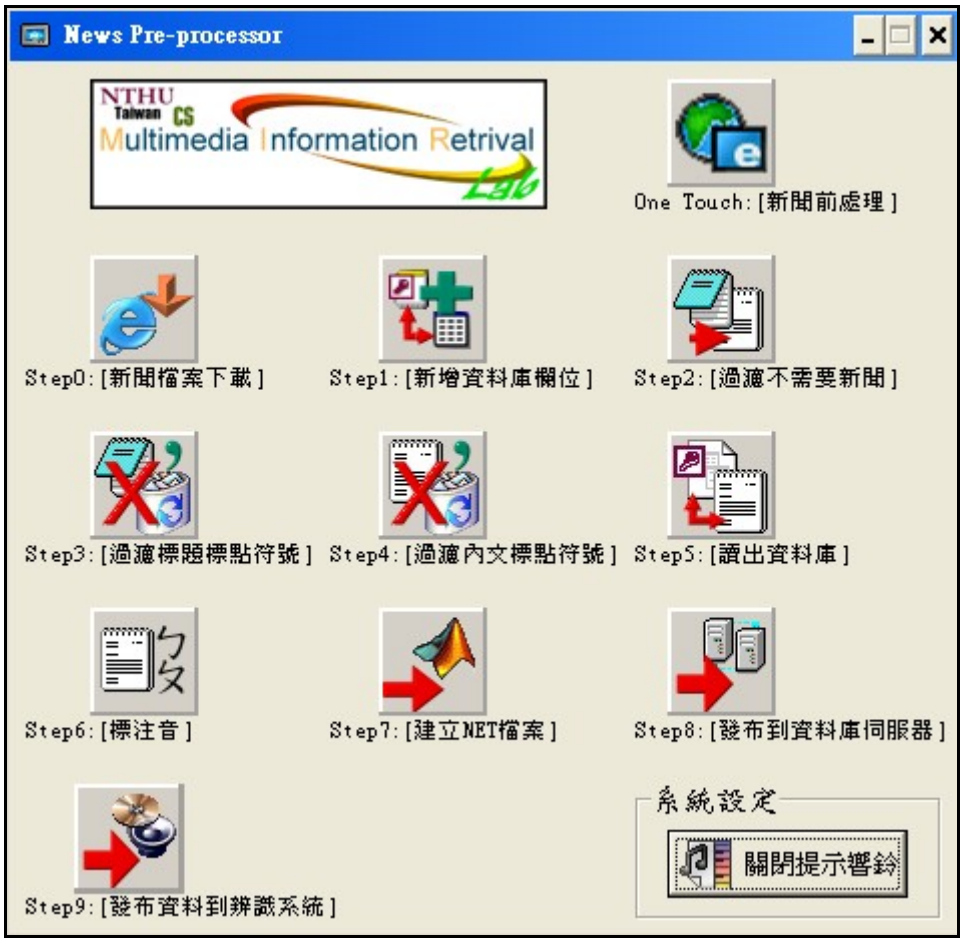


圖表 8 語音合成流程

3　語音新聞檢索系統架構

本系統分為兩大部份：新聞前處理及語音查詢新聞，分別介紹如下。

1. 新聞前處理



圖表 9 新聞前處理介面

如圖表 10 所示，本程式分為十大步驟.。【新聞前處理】按鈕則是按下後即可執行 Step0 ~ Step9。以下依序介紹各功能：

(1)【新聞檔案下載】

以 PERL 程式，從網路上抓取當日的新聞，目前系統預設值為抓取中國時報、台灣新生報、中央社新聞、新浪網新聞等四家網站的新聞。

(2)【新增資料庫欄位】

做完上一步驟抓取當日新聞完成後，在 Access 資料庫中新增兩個欄位，即 news_title_pure 及 news_content_pure，以便接下來的

處理。

(3)【過濾不需要新聞】

過濾漁業氣象這類型的新聞，如果不需要也可以省略這步驟。

(4)【過濾標題標點符號】

刪除標題標點符號，前處理系統才可以對純中文字進行標注音。將過濾完的文字存放於資料庫的 news_title_pure 欄位。

(5)【過濾內文標點符號】

刪除內文標點符號，前處理系統才可以對純中文字進行標注音。將過濾完的文字存放於資料庫的 news_content_pure 欄位。

(6)【讀出資料庫】

在進行標注音前將資料庫新增的兩欄位內的資料轉成文字檔。

(7)【標注音】

針對從資料庫轉出來的文字檔進行標注音。

(8)【建立 NET 檔案】

針對標好注音的檔案建立 Tree Net，以供語音辨識程式查詢。

(9)【發布到資料庫伺服器】

更新資料庫伺服器的資料。

(10)【發布資料到辨識系統】

更新辨識系統的辨識核心。

2. 語音查詢新聞

我們以 Borland C++ Builder 5.0 建構新聞語音查詢介面，如圖表 11。此介面分成標題查詢及內文查詢兩部份，顧名思義，標題查詢為找符合關鍵字的標題，而內文查詢則是只要內文有句子符合關鍵字即會顯示出來。以下介紹操作時大略的流程：

(1)按下一個檢索按鈕，系統會以語音的方式提示使用者準備錄音，錄音時間為三秒鐘。

(2)錄完音後辨識系統則開始辨識語音，而後將結果顯示在偏上方的白色區塊內。

(3)若使用者對查詢出來的新聞感興趣，則點選該條新聞後，偏下方的白色區塊即出現對應的新聞內容。

(4)使用者也可經由在下方白色區塊中按下滑鼠左鍵來聽取新聞的內容，該內容是以語音合成的方式即時產生的。

圖表 12 新聞語音查詢介面

4 結論

在本篇報告中，我們介紹了一個「線上新聞語音資料檢索系統」。歸納結果，在此列出此系統的特性：

1. 語音輸入：不鍵盤等須其他工具，即可將查詢內容輸入。
2. 快速檢索：藉由 offline 的標題索引和即時的語音辨識、文字比對，提供新聞標題的快速檢索。
3. 語音輸出：使用 Text-To-Speech 的語音合成，將查詢所得新聞進行播報。
4. 定時更新：每日固定時間更新網頁上即時新聞。

新聞語音查詢系統能讓網際網路的使用者有更多的方便。本系統的語音查詢有相當不錯的辨識率，而語音合成的部份表現也不會太糟，相信若用於PDA、手機等嵌入式的系統，會是個方便的工具。

線上新聞語音資料檢索系統雖然有許多優點，不過未來仍存在許多問題須要克服，例如解決斷詞的問題(文字辨識準確度)和分散式處理(大量使用者下的效率問題)等等，這些都是我們未來的工作。

5 參考資料

[1] J.-S. Roger Jang, Jiang-Chun Chen, Ming-Yang Kao,"MIRACLE: A Music Information Retrieval System with Clustered Computing Engines", International Symposium on Music Information Retrieval (MUSIC IR 2001)

[2] Jang, J. -S. Roger and Lin, Shiuan-Sung, "Optimization of Viterbi Beam Search in Speech Recognition", International Symposium on Chinese Spoken Language Processing, Taiwan, August 2002.

[3] Lawrence Rabiner, B.H Juang, Fundamentals of speech recognition, Prentice Hall, 1993.

[4] O' Shanughnessy, D., Speech Communication : human and machine, Addison-Wesley, 1987.

[5] Rabiner, L. and Juang, B.-W., Fundamentals of Speech Recognition. Prentice Hall PTR, Upper Saddle River, New Jersey, 1993

[6] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. In Ashman and Thistlewaite [2], pages 107-117.Brisbane,Australia.http://citeseer.nj.nec.com/brin98anatomy.html

[7] Steven Young, The HTK Book version 3, Microsoft Corporation, 2000.

[8] T.W. Parsons, Voice and Speech Processing, McGraw-Hill, 1986.

[9] 中文文句翻語音之韻律訊息合成，交大電信博士論文，黃紹華。

[10]林玄松,“Viterbi 搜尋的最佳化以及多語系辨識”, 清華大學碩士論文, 民國九十年。

[11]謝宏坤,“語音說明中搜尋任意定義之關鍵詞的研究”, 台灣科技大學碩士論文, 民國 89 年