

數位典藏多媒體系統使用隱藏式馬可夫檢索模型之研究

陳怡婷¹ 陳柏琳¹ 林順喜¹ 梅士杰² 廖康任²

¹國立台灣師範大學資訊工程研究所

²國立歷史博物館

g93470070@csie.ntnu.edu.tw

摘要

本論文探討隱藏式馬可夫模型(Hidden Markov Model, HMM)於數位典藏多媒體檢索系統之研究。在檢索模型訓練上我們採用了非監督式(Unsupervised)的訓練方法；同時，也研究一系列的不同參數調整方式。另一方面，當查詢是以語音的方式輸入時，經由語音辨識處理有可能會產生辨識錯誤，為了降低語音辨識錯誤對於檢索效能的影響，我們使用了以音節(Syllable)為基礎的索引特徵，比較它與以詞(Word)為基礎的索引特徵在檢索上的表現，並且嘗試同時結合這兩種索引特徵。最後，我們也比較了隱藏式馬可夫檢索模型與其它傳統檢索模型，諸如向量空間模型(Vector Space Model, VSM)與潛藏語意分析模型(Latent Semantic Analysis, LSA)等，在檢索效能上的表現。本論文是以國立歷史博物館的數位典藏資料作為實驗的文件集，以語音輸入為實驗的測試查詢。我們發現在各種不同檢索條件下，隱藏式馬可夫模型的檢索精確率都優於其它的檢索模型。尤其當在檢索系統被要求僅能回傳少許文件的情況下，實驗結果顯示隱藏式馬可夫模型的精確率遠優於一般傳統檢索模型。所以，我們認為隱藏式馬可夫檢索模型是非常適合於如PDA等手攜式設備之小介面檢索環境使用。

關鍵詞

數位典藏、語音辨識、隱藏式馬可夫模型、潛藏式語意分析模型

1. 序論

隨著電腦科技的快速演進與網際網路的興盛，數位典藏多媒體資訊快速成長，內容亦趨於多樣化，為了讓使用者從大量龐雜的數位典藏資料中快速且正確找尋出所需的內容，除了要為龐雜的數位典藏文物提供一個親切且便捷的資訊檢索環境外，如何快速且正確的找到使用者所需之資訊更為當前最迫切的需求。行政院國科會在這幾年所執行的數位典藏國家型計畫正是扮演著火車頭角色，推動國內數位典藏技術的研發與整合創新。

國立歷史博物館擁有相當豐富的典藏資料，目前也逐步正在進行一系列數位化的工作與加值應用。因此，需要一個完整的多媒體系統來整合與呈現這些多樣性數位化典藏資料。有鑑於此，台灣師大資工所與史博館共同合作開發了一套適用於館內作為PDA導覽

的無線網路環境下以語音為基礎的多媒體資訊檢索介面[1]。此系統能以口語為輸入介面，並可以隨時偵測使用者所在之展覽場地資訊以即時顯示相關內容，當使用者想查詢相關資訊時僅需對PDA導覽系統以語音輸入的方式下達查詢，系統便會及時地將欲查詢內容回傳，透過PDA導覽系統介面呈現給使用者完整的相關資訊[2]。

本論文探討隱藏式馬可夫模型[3]使用於數位典藏多媒體檢索系統之成效。由於PDA導覽系統的輸出介面不像電腦螢幕般大尺寸，並不能同時呈現太多筆檢索回傳資料；若採用多頁顯示，使用者也較難在PDA的小螢幕上做太多操作，加上多數的使用者只是希望於導覽過程中找到最相關的資料，並不想花太多時間在大量資料的瀏覽與閱讀上。所以，PDA導覽系統之檢索應著重於結果的精確率(Precision)而非召回率(Recall)。經由一系列實驗的驗證，我們發現與其他檢索模型諸如向量空間模型、潛藏語意分析模型相較，隱藏式馬可夫模型的確可以達到較高的精確率，同時也可以經由非監督方式訓練，在實際使用時逐漸改進其檢索的效能。

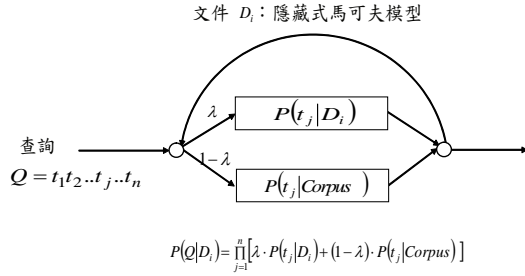
本論文接下來的安排如下：第二章我們將說明隱藏式馬可夫模型的概念及訓練方式，第三章我們介紹其它目前被廣泛使用的檢索模型，第四章描述我們的實驗環境及設定，然後於第五章呈現我們的實驗結果及分析，在第六章中將為我們所發展的數位典藏PDA導覽多媒體資訊檢索系統作一個概述，最後第七章為結論。

2. 隱藏式馬可夫模型

隱藏式馬可夫模型視每一篇文件(Document)為一個機率式生成模型(Probabilistic Generative Model)，對於每一個索引(可以是詞或音節，或是數個詞或音節的組合)都有一個對應的機率分佈。輸入的查詢(Query)與某一篇文件的相似程度，是藉由查詢中所有的索引(Indexing Terms)在此篇文件發生可能性(Likelihood)來決定。也就是說當查詢的索引在文件的機率分佈值連乘積越高，則查詢與文件內容的相關度就越高，其公式如下：

$$P(Q|D_i) = \prod_{t_j \in Q} [\lambda \cdot P(t_j|D_i) + (1-\lambda) \cdot P(t_j|Corpus)] \quad (1)$$

其中 t_j 為查詢 Q 中一個的索引； $P(t_j|D_i)$ 為索引 t_j 在文



圖一、隱藏式馬可夫檢索模型。

件 D_i 發生的機率； $P(t_j|Corpus)$ 為索引 t_j 在整個文件集 (或語料) 發生的機率，是用來平滑化 (Smooth) 索引 t_j 在文件 D_i 發生的機率 $P(t_j|D_i)$ ，並同時表示索引 t_j 在語言裡的統計資訊； λ 是比重參數 (Weighting Parameter)。此檢索模型可表示成擁有兩個狀態 (States) 的隱藏式馬可夫模型，如圖一所示。

當有訓練查詢 (Training Queries) 以及它們與文件間的相關資訊 (Relevance Information) 時，隱藏式馬可夫模型可以透過監督式 (Supervised) 的方式進行模型訓練 [3]。但是在一般實用的情況下，訓練查詢以及它們與文件間相關資訊之取得非常不易，因此必須發展非監督式 (Unsupervised) 模型訓練方式 [4]。在執行非監督式模型訓練時，我們可以將每篇文件本身視作對應的相關查詢，用來訓練文件的隱藏式馬可夫模型。同時，比重參數 λ 也可共享 (Tied) 於文件模型間。藉由期望值最大化 (Expectation-Maximization, EM) 演算法 [5]，模型訓練公式可用下列式子表示：

$$\hat{\lambda} = \frac{\sum_{D_i \in \{Doc\}} \sum_{t_j} E(t_j, D_i)}{\sum_{D_i \in \{Doc\}} |D_i|} \quad (2)$$

$$\hat{P}(t_j|D_i) = \frac{E(t_j, D_i)}{\sum_{s=1}^S E(t_s, D_i)} \quad (3)$$

$$E(t_j, D_i) = c(t_j, D_i) \cdot \frac{\lambda \cdot P(t_j|D_i)}{\lambda \cdot P(t_j|D_i) + (1-\lambda) \cdot P(t_j|Corpus)} \quad (4)$$

其中 $\{Doc\}$ 為整個文件集； S 為所有不同索引的個數； $|D_i|$ 為文件 D_i 的全部索引個數； $c(t_j, D_i)$ 為索引 t_j 在文件 D_i 出現的次數。

在訓練初始化時， $P(t_j|D_i)$ 與 $P(t_j|Corpus)$ 可分別利用最大可能性 (Maximum Likelihood) 估測，依據索引 t_j 實際出現在文件 D_i 以及整個文件集發生的次數而求得。在本論文中我們實驗了三種非監督式期望值最大化模型參數訓練方式：第一種為僅使用式 (2) 來求取新的比重參數 $\hat{\lambda}$ ；第二種為僅使用式 (3) 來求取新的

$\hat{P}(t_j|D_i)$ 機率值；第三種訓練方式結合上述二種訓練方式。在本論文我們將執行一系列的實驗來求得最佳的隱藏式馬可夫模型模型參數，然後與其它經最佳化過後的檢索模型作比較，以驗證隱藏式馬可夫模型的優越性。

3. 其它檢索模型

在本章我們介紹兩種目前常被廣泛使用的檢索模型：向量空間模型與潛藏語意分析模型。

3.1 向量空間模型

向量空間模型將每一篇文件分別視為空間中的一個向量，向量的每一個維度代表某個索引 t_j 在文件 D_i 的統計資訊，可以表示成：

$$w_{ji} = [1 + \log(tf_{ji})] \cdot \log\left(\frac{N}{N_j}\right) \quad (5)$$

tf_{ji} 代表索引 t_j 在文件 D_i 出現的次數， $1 + \log(tf_{ji})$ 代表 t_j 在文件中的頻率 (Term Frequency)； $\log \frac{N}{N_j}$ 是反文件頻率 (Inverse Document Frequency, IDF)， N 是所有文件的數目， N_j 是出現索引 t_j 的文件數目，當 t_j 出現在越多文件中代表它越不重要。索引 t_j 在查詢 Q 中的統計資訊 w_{jQ} 也可以使用相同的方式以式 (5) 表示。最後，查詢 Q 與文件 D_i 的相關性可以用餘弦值來評估 (Cosine Measure) [6]：

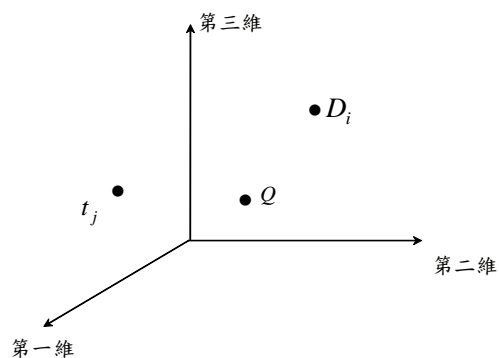
$$SIM(Q, D_i) = \frac{\sum_j w_{jQ} \cdot w_{ji}}{\sqrt{\sum_j w_{jQ}^2} \cdot \sqrt{\sum_j w_{ji}^2}} \quad (6)$$

當餘弦值越大時代表查詢 Q 與文件 D_i 的相關性越大。

3.2 潛藏式語意分析模型

隱藏式馬可夫模型與向量空間模型基本上是以逐字比對 (Literal Term Matching) 方式來評估查詢與文件的相關性。在實際的情況下，如果當查詢與文件是在闡述相同的主題，因為查詢與文件的用字遣詞不同，例如「國科會」與「國家科學委員會」、「陳水扁」與「中華民國總統」，這可能會使得某些與查詢極相關的文件無法被檢索到。因此，以概念比對 (Concept Matching) 的檢索模型如潛藏語意分析模型便被提出以解決此一問題 [7][8]。

潛藏式語意分析模型以“索引—文件”矩陣 (Term-Document Matrix) W 來描述索引與文章之間的關係，矩陣中的元素 w_{ji} 代表索引 t_j 在文件 D_i 的統計資訊，同樣也可以用式 (5) 來表示。同時，每個索引可以表示成 W 中的一個列向量 (Row Vector)，而每篇文件則可以表示成 W 中的一個行向量 (Column Vector)。然後經由對矩陣 W 執行奇異值分解 (Singular Value Decomposition, SVD) 後，可以把所有的索引向量與文件向量共同投射到一個低維度 (R 維度) 的潛藏語意空



圖二、潛藏語意空間與潛藏式語意分析檢索模型之示意圖。

間，可表示成：

$$W \approx \hat{W} = U\Sigma V^T \quad (7)$$

其中 \hat{W} 是 W 的近似矩陣， U 是左奇異矩陣， Σ 是 $R \times R$ 的對角矩陣， V 是右奇異矩陣， T 是矩陣轉置。每個索引向量因此可用 U 的某一行向量 \bar{u}_j (R 維度) 來表示，每個文件向量可用 V 的某一行向量 \bar{v}_i (R 維度) 來表示。其中， \bar{u}_j 與 \bar{v}_i 的每一維度都有一對一的對應關係，代表者某種潛藏的語意(Semantic)或主題(Topic)[7]。對於查詢以及其它新的文件也可以透過摺入(Fold-in)的方式，在此潛藏語意空間以向量形式表示，而查詢與文件的相關性也可以用餘弦值來評估。圖二為潛藏語意空間與潛藏式語意分析檢索模型之示意圖。

4. 實驗設定

本論文使用史博館約 2,800 則的數位典藏說明文件與詮釋資料為檢索的目標文件集；而測試查詢共有 16 句，其對應的語音查詢則分別由兩男兩女每人使用 PDA 各錄 16 句查詢之辨識結果，詳細的統計資訊參見表一。語音查詢的辨識是使用中文大詞彙連續語音辨識器[9]，辨識結果如表二所示。語音的聲學模型是使用 8.5 小時的 PDA 語音語料所訓練而成[10]，由於我們是使用一般文字新聞語料來訓練語言模型，並沒有特別以數位典藏的文字語料庫來訓練，因此導致語音查詢的正確率在音節、字、詞上分別僅有 72.69%、58.90%與 43.30%。另一方面，檢索評估方法是以平均精確度在不同的文章切斷值 k (Mean Average Precision at a Given Document Cutoff Value k) 來評估，文章切斷值 k 為保留前 k 名排序較大的文章來計算平均精確度[11]，公式表示如下：

$$mAP_{cutoff\ k} = \frac{1}{L} \sum_{l=1}^L \frac{1}{N_l} \sum_{s=1}^{N_l} \frac{s}{r_{l,s}} \quad (8)$$

L 為查詢輸入的個數， N_l 為在前 k 篇檢索出的文章中與查詢輸入 Q_l 有關的文章個數， $r_{l,s}$ 為在前 k 篇檢索出的文章中與查詢輸入 Q_l 有關的第 s 篇文章從排序過後檢索結果序列(Ranked List)前面(Top)數來的位置。如本論文一開始所提及的，我們的 PDA 檢索系統

	最小值	最大值	平均值
文件長度(以詞為單位)	4	17,657	215.77
文字查詢長度(以詞為單位)	2	9	5.50
語音查詢長度(以秒為單位)	1.60	3.60	2.494
與查詢相關文件數	4	274	41.44

表一：實驗文件及查詢的相關統計資訊。

	正確率
音節	72.69%
字	58.90%
詞	43.30%

表二、語音查詢之辨識率。

所需要的是精確檢索結果，因此我們希望所找到的前 k 篇文章即為正確的相關文章，並不強調要檢索系統傳回大量的相關文章，因為 PDA 的螢幕顯示無法容納大量的檢索結果列表，所以在此我們僅針對回傳前 k ($k = 3, 5, 10, 15$) 篇文章的情況作評估。

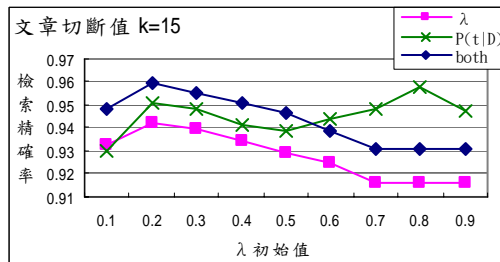
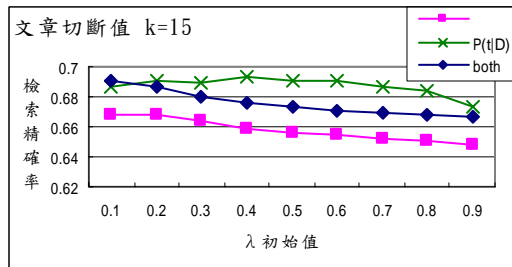
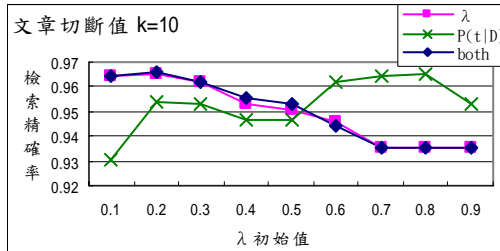
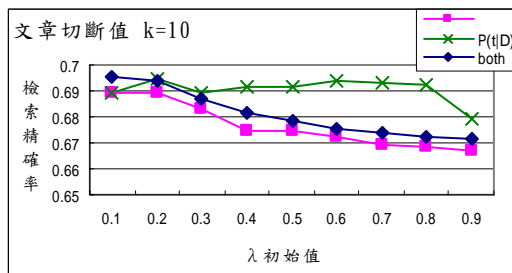
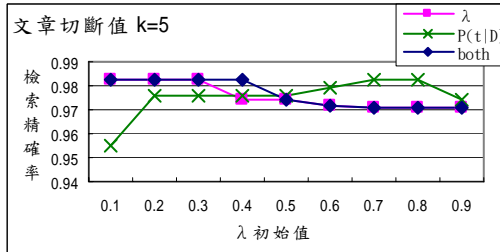
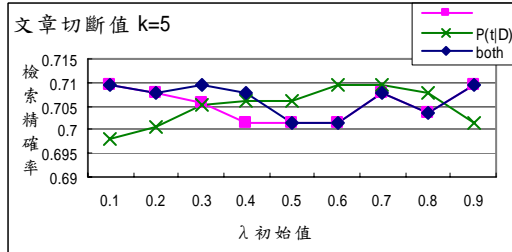
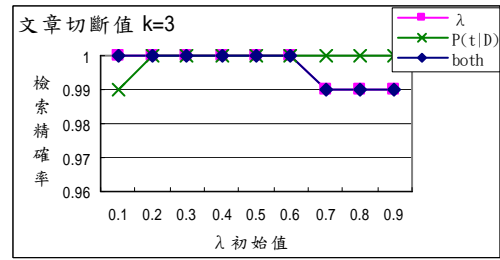
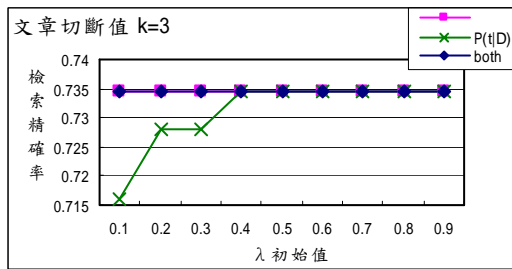
5. 實驗結果

實驗的測試查詢分為文字查詢與語音查詢二種，前者以正確的文字輸入，後者則是語音輸入經由語音辨識後所產生的辨識結果。為了證實隱藏式馬可夫模型適用於我們的 PDA 多媒體導覽系統，因此，本論文的實驗主要是以使用語音查詢為主。一開始，我們先針對不同索引特徵的使用，為隱藏式馬可夫模型找出一組最佳的模型參數組合，來進行一連串的檢索實驗，並與其他檢索模型的檢索結果進行比較及分析。也就是說，我們使用語音查詢並依序以詞、雙音節、及詞與雙音節組合的索引特徵進行檢索實驗，觀察不同檢索模型在使用不同索引組合下的檢索表現。我們希望藉由這一系列的實驗來探討隱藏式馬可夫模型在語音資訊檢索上的效能。

5.1 隱藏式馬可夫模型參數訓練

如第 2 章所提及，在本論文中我們希望發展隱藏式馬可夫模型的非監督式訓練方式，在沒有訓練查詢與其相關文件資訊的情況下，將每篇文件本身視作對應的相關查詢來訓練文件的隱藏式馬可夫模型，如式(2)、式(3)、式(4)所示，我們將某篇文件 D_i 視為一查詢並以文件 D_i 本身為查詢之相關文件的方式來作為訓練，這樣的作法較能適合於真實的檢索系統使用情境。我們希望從不同的模型參數訓練設定中，找到一組最佳的訓練設定，然後據此以比較隱藏式馬可夫模型與其他二種檢索模型的檢索表現。

首先，我們使用語音查詢輸入以及詞索引特徵，探討三種非監督式期望值最大化模型參數訓練方式



圖三、使用語音查詢輸入以及詞索引特徵之隱藏式馬可夫模型檢索結果。

圖四、使用文字查詢輸入以及詞索引特徵之隱藏式馬可夫模型檢索結果。

(參見第 2 節)：(1)僅訓練比重參數 λ ；(2)僅訓練文件產生索引機率分佈 $P(t_j|D_i)$ ；(3)同時訓練比重參數 λ 與文件產生索引機率分佈 $P(t_j|D_i)$ 。同時，因為期望值最大化僅能找出區域最佳(Local Optimum) 模型參數，我們希望藉由設定比重參數 λ 不同的初始化值 (0.1~0.9)，訓練出較佳的模型參數組合。圖三為上述三種非監督式模型參數訓練方式在不同文章切斷值以及不同的比重參數 λ 初始化值設定下的檢索結果。由實驗結果來看，不同的模型參數訓練組合在某些比重參數 λ 初始化值下(例如 $\lambda = 0.3$)均有不錯的檢索表現。但是若比較四種不同文章切斷值 ($k = 3, 5, 10, 15$) 的檢索精確率可以發現，僅訓練文件產生索引機率分佈 $P(t_j|D_i)$ 普遍地可達較高之檢索精確率。(圖三與圖四中 λ 表示訓練比重參數 λ 且訓練次數為 1 之檢索結果、 $P(t_j|D_i)$ 表示訓練機率分佈 $P(t_j|D_i)$ 且訓練次數為 10 之檢索結果，both 表示同時訓練比重參數 λ 與文

件產生索引機率分佈 $P(t_j|D_i)$ 且訓練次數為 1 之檢索結果。)

此外，我們亦使用文字查詢輸入以及詞索引特徵來從事上述的資訊檢索實驗，其結果如圖四所示。雖說由表二我們可發現，語音輸入的詞錯誤率達 56% 以上(正確率 43.30%)，但平均而言對於不同模型參數訓練方式的檢索結果約造成不到 30% 的精確率下降(比較圖三與圖四)。而我們過去在從事以文字查詢輸入檢索語音新聞文件的實驗時卻發現[3]，當語音新聞文件有 40% 左右的詞錯誤率時，檢索結果僅會造成不到 5% 的精確率下降。因此我們可推論，對於語音資訊檢索而言，語音查詢對於檢索表現的影響是遠大於語音文件。因此，對於使用語音輸入的檢索系統，我們認為必須發展一套強健檢索機制，以降低語音辨識錯誤所會造成的影響。

5.2 隱藏式馬可夫模型與其他檢索模型之比較

接下來的實驗，我們將在以使用語音查詢輸入以及詞索引特徵的情況下，比較隱藏式馬可夫模型(HMM)與其它兩種常見的資訊檢索模型(向量空間模型(VSM)以及潛藏式語意分析模型(LSA))在檢索實驗的表現。對於隱藏式馬可夫模型是使用 5.1 節中所得到的最佳模型參數之檢索結果；而向量空間模型與潛藏式語意分析模型則同樣地是採用它們最佳模型設定下所得之檢索結果。圖五為三種檢索模型的檢索精確率比較，我們可以清楚地看出隱藏式馬可夫模型的精確率較其他二種檢索模型高出許多，特別是在文章切斷值較小(也就是系統回傳文件較少時)時更為明顯。此結果初步證明了隱藏式馬可夫模型的確可以比其它傳統檢索模型達到較佳的檢索精確率。

另一方面，潛藏式語意分析模型的檢索精確率為三者之中最差，我們推測可能是因語音辨識正確率太低使得語音查詢在潛藏語意空間所建構出的向量表示式，並不能正確地代表原始查詢的語意，而使得查詢與文件在潛藏語意空間比對時，並未找出真正語意相關的文件，造成檢索精確率的下降。

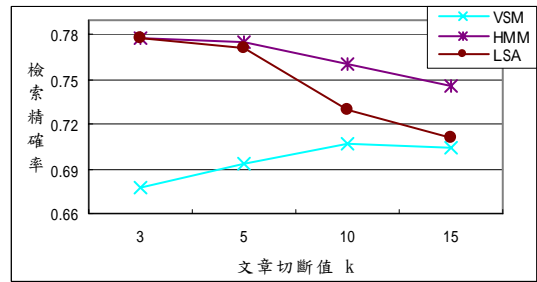
5.3 以雙音節為索引特徵之檢索模型比較

為了避免語音查詢輸入經語音辨識產生錯誤而對資訊檢索產生影響，我們在此嘗試以語句中任一對重疊雙音節(Overlapped Syllable Pair)為索引特徵。其理由是因音節的辨識率遠高於詞的辨識率，以及幾乎 90% 以上常用中文詞中都為雙音節詞，因此以雙音節為索引特徵一方面可在某種程度上代表語(文)句中語意資訊，另一方面又可以達到強健性索引的目的[8]。圖六為使用語音查詢輸入以及雙音節索引特徵之檢索結果，與圖五比較後我們可觀察到本論文所討論到的三種檢索模型在以雙音節為索引特徵時所得之精確率，均較以詞為索引特徵時提高許多，因此驗證了雙音節索引特徵的確能減輕語音辨識所造成的影響。

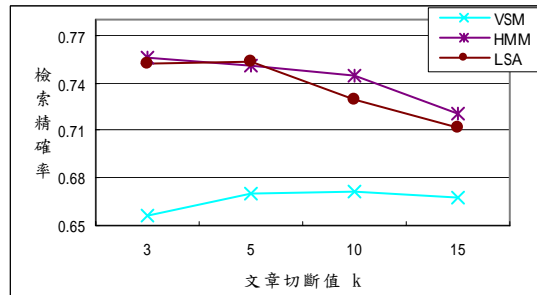
另一方面，對於潛藏式語意分析模型而言，它的檢索精確率在使用雙音節索引特徵時大幅度提高，除了優於向量空間模型外，甚至於文章切斷值 $k=3$ 時與隱藏式馬可夫模型有相同的精確率。從圖五與圖六結果來看，潛藏式語意分析模型的檢索表現很容易受到語音辨識正確率所影響。而對於隱藏式馬可夫模型而言，採用雙音節為索引特徵可以提高其精確率，但是不論是採用詞或雙音節的索引特徵，隱藏式馬可夫模型都表現出比其它兩種檢索模型較佳的結果。

5.4 以詞及雙音節組合為索引特徵之檢索模型比較

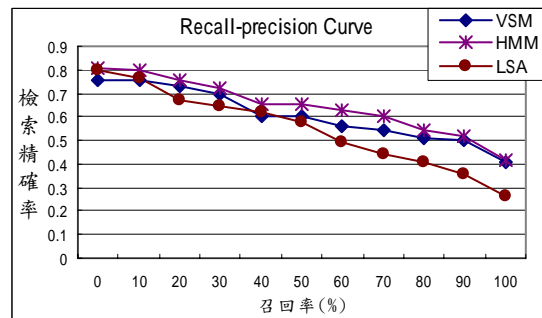
在上述實驗，不論是以詞或雙音節為索引特徵均證明了隱藏式馬可夫模型的檢索精確率都優於向量空間模型及潛藏語意分析模型。在本小節裡，我們更進一步嘗試使用詞及雙音節組合為索引特徵，觀察其檢索表現。我們分別以使用詞以及使用雙音節為索引特徵進行資訊檢索，再將每篇文件對於這兩種索引特徵所計算出的相似評估值以各佔 1/2 的權重比例加總後進行排序。雖說或許使用其它權重比例進行加總後，可能



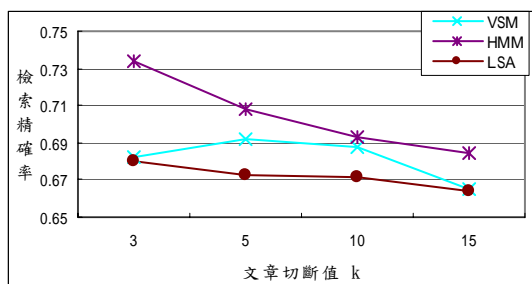
圖六、使用語音查詢輸入以及雙音節索引特徵之檢索模型比較。



圖七、使用語音查詢輸入以及詞與雙音節索引特徵組合之檢索模型比較。



圖八、使用語音查詢輸入以及詞與雙音節索引特徵組合之檢索模型比較(11 點差補精確率-召回率曲線)。



圖五、使用語音查詢輸入以及詞索引特徵之檢索模型比較。

會得到較佳的檢索精確率，但在本論文我們僅以 1/2 的權重比例來觀察以詞及雙音節組合為索引特徵之檢索表現。

圖七為詞及雙音節組合為索引特徵的語音查詢檢索結果，我們發現隱藏式馬可夫模型及潛藏式語意分析模型的檢索精確率相較於以詞為索引特徵(圖五)所得之精確率高，但是較以雙音節為索引特徵(圖六)之精

確率低一些，這兩種模型在以詞及雙音節組合為索引特徵時的精確率均介於以詞為索引特徵及以雙音節為索引特徵的結果之間。從傳統向量模型來看，此詞及雙音節組合為索引特徵的檢索方式呈現出較差的精確率，可能是因以詞為索引特徵取得之文件排名與以雙音節為索引特徵取得之文件排名並不一致，因此當採用詞及雙音節組合為索引特徵時，二者結果的合併反而降低其結果。



圖九、數位典藏 PDA 導覽系統。

除了上述以平均精確度在不同的文章切斷值 k 的評估方式外，為了進一步分析及了解三種檢索模型於不同召回率之平均精確度，我們於圖八呈現以雙音節為索引特徵的“11 點差補精確率-召回率曲線”(11-Point Interpolated Recall-Precision Curve)[12]，由圖可以觀察出，對於低召回率時，隱藏式馬可夫模型與潛藏式語意分析模型均有不錯的精確率，但是隱藏式馬可夫模型普遍優於其他檢索模型，甚至於高召回率時隱藏式馬可夫模型的精確率亦優於其他的檢索模型。

6. 國立歷史博物館之多媒體 PDA 導覽系統

我們結合國立歷史博物館豐富的數位典藏文物，完成一套以 PDA 為平台的多媒體導覽系統[2]，希望提供使用者一個自然且有效率的檢索與瀏覽環境。它主要包括了大詞彙連續語音辨識、資訊檢索、圖形檢索及室內定位等四部分，我們透過這些技術的整合來建構多功能型的 PDA 導覽系統，讓使用者可以非常便利與簡單的操作小型介面的 PDA 導覽系統。

使用者可以透過 PDA 在無線網路環境下以人性化的語音輸入方式或文字輸入的方式來檢索歷史博物館

的數位典藏內容。當使用者以語音輸入檢索時，所下達之查詢經過 802.11 無線網路傳送到傳到大詞彙連續語音辨識伺服器，產生語音辨識結果的詞串，然後由資訊檢索伺服器以辨識結果為查詢去檢索史博館的數位典藏多媒體文件，把檢索結果傳回 PDA，最後於 PDA 系統介面顯示相關數位典藏文物的資訊。圖九為 PDA 多媒體導覽系統之檢索介面及相關結果的圖示。我們的 PDA 多媒體導覽系統介面允許使用者以語音及文字二種方式進行查詢，待檢索系統產生結果後回傳於 PDA 畫面，使用者可依需求選擇欲查看的內容，例如 (a) 相關圖片及 (b) 相關文字等。

除了上述所描述的以文字或語音為基礎之數位典藏文物檢索系統外，我們亦建立以圖形內涵為基礎的檢索系統及無線網路環境下室內定位系統。圖形檢索系統讓使用者在使用多媒體導覽系統時可採用以圖找圖（顏色、紋路、形狀）的檢索方式來找到相關的圖片；而室內定位系統則可提供使用者即時的展場資訊，隨時讓使用者了解所在之位置，顯示展覽室相關資訊及簡介。目前我們的室內定位系統主要是以展區為主的定位，一開始我們針對展區進行規劃、定點定位，然後透過室內定位伺服與這些展區位置之間的溝通之連結後，即可隨時偵測多媒體導覽系統於館內任一展場的位置。

7. 結論

由於我們的 PDA 多媒體導覽系統以語音為輸入介面，因此希望能夠採用具有強健性的檢索模型使用於所開發的多媒體資訊檢索系統中，以期能避免語音辨識錯誤所產生的影響，甚至是更進一步彌補語音辨識所產生錯誤。此外，PDA 的小型介面並不能顯示大量的文件內容，而且配合史博館內導覽的需求，顯示大量的文件內容並不符合一般使用者的需要，因此我們亦希望採用可以達到高精確率的檢索模型於我們的系統中。基於上述的種種需求，我們選擇以隱藏式馬可夫模型使用於我們的數位典藏多媒體系統。我們以一連串的實驗證明隱藏式馬可夫模型可以達到極佳的精確率，同時也證明了隱藏式馬可夫模型非常適用於 PDA 的小型介面。

我們一開始由不同的訓練方式與重覆訓練次數的實驗中找出一組最佳的隱藏式馬可夫模型的模型參數，接著再以此隱藏式馬可夫模型參數與向量空間模型、潛藏語意分析模型之檢索結果進行比較，證明隱藏式馬可夫模型之檢索正確率的確優於其他檢索模型。我們從實驗中發現隱藏式馬可夫模型於許多組模型參數設定下均有很好的結果，但是當訓練比重參數 λ 的迭代次數越多時 (λ 趨近於 1)，卻發現檢索結果呈現越差的情形，這可能是因為我們採用非監督式訓練方式，將每篇文件本身視作相關的查詢，用來訓練文件的隱藏式馬可夫模型，而當重覆次數越多時，式 (1) 計算查詢與文件的相關程度時就會使得 λ 值越大 (越信任文件本身的機率模型)，而使得 $P(t_j | Corpus)$ 的機率被忽略。因此我們希望能夠進一步使用其他方

式來訓練參數 λ ，我們相信這將會使得檢索結果達到更高的精確率。

雖然非監督式訓練之隱藏式馬可夫模型顯示出好的檢索表現，但當我們可以取得某些訓練資料—訓練查詢以及它們與文件間相關資訊，我們還是可經由監督式參數訓練方式而訓練出更好的隱藏式馬可夫檢索模型[3]。未來我們考慮將機率式概念比對的檢索模型(有別於傳統潛藏語意分析模型)[13]應用於系統中，希望我們的資訊檢索系統不僅僅只能以逐字比對的方式進行檢索，而能夠進一步檢索到許多相同主題或是相同概念的文件。此外，也希望針對檢索的文件集能夠以發展有組織及有系統的方式來作視覺化呈現[14][15]，自動地為文件產生摘要[8]，讓使用者可以很方便地對整個文件集進行瀏覽，直接找到所需之相關文件。

8. 參考文獻

- [1] S. S. Lin, B. Chen, and S. J. Huang, "Building a Multi-modal Multimedia Information Retrieval System for the National Museum of History," in *Proc. ICDAT 2004*.
- [2] 陳鴻彬、陳柏琳、林順喜, "語音辨識及資訊檢索技術於數位典藏多媒體文物之應用", 第三屆數位典藏技術研討會, 台北, 民國九十三年。
- [3] B. Chen, H. M. Wang, L. S. Lee, "A Discriminative HMM/N-Gram-Based Retrieval Approach for Mandarin Spoken Documents," *ACM Transactions on Asian Language Information Processing*, Vol. 3, No. 2, June 2004.
- [4] D. Hiemstra, S. Robertson, H. Zaragoza, "Parsimonious Language Models for Information Retrieval," in *Proc. SIGIR 2004*.
- [5] A.P. Dempster, N. M. Laird, D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm" *Journal of Royal Statistical Society B*, Vol. 39, No. 1, 1977.
- [6] G. Salton, M. J. McGill. *Introduction to Modern Information Retrieval*. New York: McGraw-Hill, 1983.
- [7] J. R. Bellegarda, "Latent Semantic Mapping: Dimensionality Reduction via Globally Optimal Continuous Parameter Modeling," *IEEE Signal Processing Magazine*, Vol. 22, No. 5, September 2005.
- [8] L. S. Lee, B. Chen, "Spoken Document Understanding and Organization for Efficient Retrieval/Browsing Applications," *IEEE Signal Processing Magazine*, Vol. 22, No. 5, September 2005.
- [9] B. Chen, J. W. Kuo, W. H. Tsai, "Lightly Supervised and Data-Driven Approaches to Mandarin Broadcast News Transcription," in *Proc. ICASSP 2004*.
- [10] B. Chen, Y. T. Chen, C. H. Chang, H. B. Chen, "Speech Retrieval of Mandarin Broadcast News via Mobile Devices," in *Proc. EUROSPEECH 2005*.
- [11] D. Harman, "Overview of the Fourth Text Retrieval Conference (TREC-4)," in *Proc. TREC-4*, 1995.
- [12] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, 1999
- [13] B. Chen, J. W. Kuo, Y. M. Huang, H. M. Wang, "Statistical Chinese Spoken Document Retrieval Using Latent Topical Information," in *Proc. ICSLP 2004*.
- [14] T. H. Li, M. H. Lee, B. Chen, L. S. Lee, "Hierarchical Topic Organization and Visual Presentation of Spoken Documents Using Probabilistic Latent Semantic Analysis (PLSA) for Efficient Retrieval/Browsing Applications," in *Proc. EUROSPEECH 2005*.
- [15] T. Kohonen, S. Kaski, K. Lagus, J. Salojvi, J. Honkela, V. Paatero and Saarela A, "Self organization of a massive document collection," *IEEE Transactions on Neural Networks*, Vol. 11, No. 3, 2000.