4. 摘錄式語音文件摘要使用機率生成式摘要模型

本論文探討研究機率生成式模型於摘錄式語音文件摘要之研究，提出使用關聯性模型來提昇隱藏式馬可夫摘要模型(HMM)的摘要正確率；另一方面，也提出以詞層次主題混合摘要模型來摘錄文件中重要的文句。本章前三個小節中，將詳細的說明與介紹機率生成模型的概念及本論文所提出的摘要方法；此外，我們亦於第四小節中嘗試結合各種有利於選取重要文句的摘要特徵，作為文句的事前資訊(事前機率，Prior Probability)，例如辨識信心度、聲學特徵、語言學特徵及文章結構(文句位置)，研究文句事前機率值的估測對於摘要結果的影響。

4.1 機率生成模型

在本論文中，吾人將摘錄式摘要視為一個“文句重要性排序”的問題，希望依序選出最能表達整篇文件主題或概念的文句。因此，我們以最大事後機率(Maximum A Posteriori Probability, MAP)法則為基礎，首先提出以機率生成模型的方式來表示文句與文件之間的關係。當給一篇欲摘要之文件 D 時，希望找出與文件 D 最為相似的文句 S_i，可以下面的式子來表示：

$$\arg\max_{S_i} P(S_i|D) \tag{4-1}$$

透過貝式定理(Bayes' rule)轉換，式(4.1)可表示為：

$$\arg\max_{S_i} P(S_i|D) = \arg\max_{S_i} \frac{P(D|S_i)P(S_i)}{P(D)} \tag{4-2}$$

由於在排序時文件機率 $P(D)$ 對文件中所有文句 S_i 的影響均是相同的，可將其忽略不看，故式(4-2)可進一步表示成：

$$\arg\max_{S_i} P(S_i|D) = \arg\max_{S_i} P(D|S_i)P(S_i) \tag{4-3}$$

其中機率值$P(D|S_i)$為文句S_i生成文件D的可能性(Likelihood)；$P(S_i)$為文句S_i本身的事前機率(Prior Probability)，我們可簡單地假設其為一致性分佈，使每一文句S_i的事前機率相同，亦可經由文句本身的語言學分數、重要性分數、文章結構或是聲學特徵給予不同的機率分佈。

下面各小節中，將闡述本論文中提出的關聯性模型以及詞層次主題混合摘要模型。

4.2 關聯性模型與隱藏式馬可夫模型

4.2.1 背景簡介：關聯性模型(Relevance Model)

關聯性模型是由Lavrenko與Croft所提出並應用於文字文件之資訊檢索[Lavrenko and Croft 2001; Croft and Lafferty 2003]。資訊檢索是根據使用者所下的查詢(Query)，從大量文件集中找出與查詢相關的文件子集合回傳給使用者，以幫助使用者快速、正確地找到所需資訊。通常，資訊檢索可視為文件集中每一文件與查詢之間相關程度排名的問題，當文件的主題、內容與使用者查詢的越相關(相似)，則相關性排名越前面[Chen et al. 2005]。但是，對於如何正確地的估測文件集中每一文件與查詢之間的相關度，尚存在著某種程度上的困難。一般來說，使用者所下的查詢大多為簡短的關鍵詞、片詞或是文句，鮮少會完整或詳盡地描述問題，這可能是因為使用者不知道該如何定義問題，亦可能因使用者沒有充裕的時間或不想太費心思。因此，若僅以這樣有限的線索(使用者查詢)，希望正確地找出最相關的文件內容，可能會因同義字及問題定義不明確而發生語意或主題上的混淆，無法找到最相關資訊。

Lavrenko 與 Croft 二位學者提出關聯性模型，透過建立與查詢有關的”關聯”分佈來彌補查詢內容的不足，強健使用者查詢的內容，以正確找出相關文件。如此，當欲檢索與查詢相關的文件時，不再侷限於查詢本身的資訊，而以代表查詢

內容分佈的關聯性模型(Relevance Model)，來評估文件與查詢間的相關程度[Lavrenko and Croft 2001; Croft and Lafferty 2003]。

然而，該如何定義與建立關聯性模型呢？通常，建立關聯性模型需有與使用查詢相關的資訊、文件，例如與查詢相關的相關文件子集(Relevant Document Subset)，才能正確的估測查詢的機率分佈。但是，這樣的相關資訊，通常需經由人工標註(Tagging or Labeling)或使用者回饋(User Feedback)才能取得。然而，相關文件集的標註是費時、費力又高成本的，而透過與使用者互動所取得的資訊(使用者回饋)品質不一，其中可能含有主觀與可靠度的因素，同時亦增加使用者的負擔與時間。因此，相關文件集的取得並不容易。

「局部性回饋(Local Feedback)」[Baeza-Yates and Ribeiro-Neto 1999]技術可被使用來建立關聯性模型。由於與查詢相關的文件集取得不易，因此需要透過自動的方式來取得相關的資訊，而局部性回饋是最為常見的方法之一。其作法為利用使用者所下的查詢先對一個大的文件集進行初步的文件檢索，找出與查詢最為相關的文件子集。如此，便可獲得與查詢最相關的前 L 篇文件作為查詢的相關文件子集。接著，透過這些相關文件子集，用來建立查詢的關聯性模型：

$$P(w|M_R)=\sum_{l=1}^{L}P(D_l|R)P(w|D_l) \tag{4-4}$$

其中 $P(w|M_R)$ 為詞 w 於關聯模型(M_R)中的機率分佈，表示相關文件子集中詞 w 的發生機率；L 為檢索系統所回傳的相關文件數；機率值 $P(D_l|R)$ 表示第 l 篇相關文件於相關文件子集 R 中的分佈機率值；機率值 $P(w|D_l)$ 為第 l 篇相關文件中詞 w 的機率分佈值，通常每一文件 D_l 可表示成一個多項式模型(Multinomial Model)，用以描述每一個詞 w 於文件 D_l 中的機率分佈。如此，文件 D_l 件的可能性(Likelihood)定義如下：

$$L(D_l)=\prod_{w\in D_l}P(w|D_l)^{D_l(w)} \tag{4-5}$$

其中，$D_l(w)$ 為詞 w 在文件 D_l 中出現的次數。機率值 $P(w|D_l)$ 的估測可以最大相似度估測法(Maximum Likelihood Estimation, MLE)來計算：

$$P(w \mid D_l) = \frac{D_l(w)}{|D_l|} \tag{4-6}$$

$|D_l|$表示文件D_l的長度，可表示為：

$$|D_l| = \sum_{w_n \in D_l} D_l(w_n) \tag{4-7}$$

最後，進一步的檢索結果可透過所建立的關聯性模型來找出與查詢最相關的文件集，亦可用於調適(Adapt)原來的查詢模型(Query Model)，以彌補原查詢分佈估測的不足，或是利用相關文件集的資訊進行查詢擴增(Query Expansion)。

4.2.2 使用關聯性模型提昇文句估測

如論文中 2.3.3 小節所提，隱藏式馬可夫摘要模型將欲摘要文件D中每一文句S_i視為一個機率生成模型(HMM)，其中包含二個 N-連詞機率分佈－文句模型(Sentence Model)與文件集模型(Collection Model)。在使用單連詞(Unigram)機率分佈情況下可表示成：

$$P(D \mid S_i) = \prod_{w \in D} \left[\lambda \cdot P(w \mid S_i) + (1-\lambda) \cdot P(w \mid C)\right]^{n(w,D)} \tag{4-8}$$

$n(w, D)$為詞w於D出現的次數，λ是權重參數(Weighting Parameter)，$P(w_i \mid S_i)$為文句S_i產生詞w的機率值。$P(w \mid C)$為詞w在整個文件集(或語料)發生的機率，用來平滑化(Smooth) 詞w於文句S_i發生的機率，同時表示詞w在語言裡的統計資訊。重要文句的選取以其生成文件D中所有的詞w的可能性(Likelihood)乘積來決定。本論文中所探討的隱藏式馬可夫模型皆採用單連詞機率分佈。

在隱藏式馬可夫模型中，文句模型與文件集模型的機率分佈皆可使用最大相似度估測法估測。例如機率$P(w \mid S_i)$的計算方式可表示如下：

$$P(w \mid S_i) = \frac{S_i(w)}{|S_i|} \tag{4-9}$$

其中$S_i(w)$表示詞w在文句S_i出現的次數；$|S_i|$為文句S_i的長度。由式(4-9)可看出文句S_i中詞w的機率分佈是由詞w於文句出現的次數來估測。但是，在文件中通

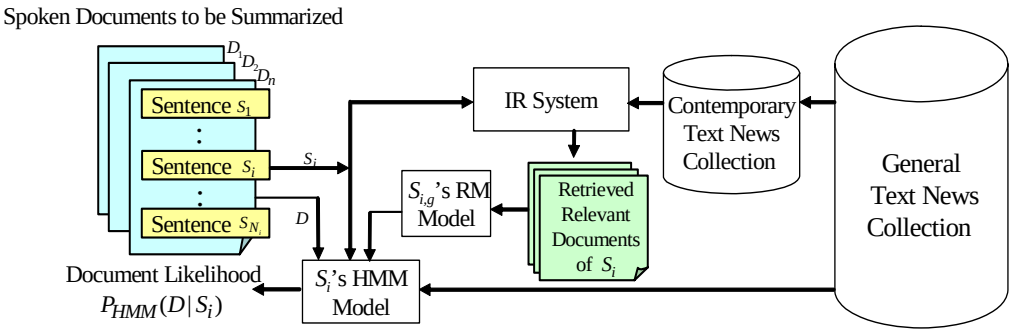


圖 4.1 每一文句的機率生成模型及關聯性模型示意圖

常每一文句都為簡短的詞串，僅包含少量的文字內容與資訊，每一個詞可能最多僅出現一次，造成文句 S_i 的機率模型估測不夠健全與可靠，使得所估測之文句模型無法表示文句“真正”的文句模型(True Sentence Model)。因此，我們希望提昇對文句模型估測的精確性，正確地表示每一索引於於文句 S_i 中的分佈機率值，以提高文句 S_i 生成文件可能性的估測，提昇自動語音文件摘要的正確率。

在本論文中，吾人將關聯性模型(Relevnace Model)延伸應用於自動語音文件摘要，以改善因文句的內容簡短與資料量的不足，造成文句模型估測的偏差及無法精確地計算文件與文句間的相似度。我們將每一文句 S_i 的關聯性模型視為文句模型參數的事前機率(Prior Probability)分佈，用來調適原來僅使用最大相似度估測法所估測的文句機率模型，提高對於文句 S_i 機率模型分佈的估測可靠性。希望當使用隱藏式馬可夫摘要模型進行文件摘要時，藉由透過每一文句所對應的關聯性模型的調適，帶來更多相關的資訊以彌補文句資訊量的不足，對於文句模型的機率分佈的估測能夠更精確及可靠。

每一文句 S_i 所對應之關聯性模型的建立，通常需取得與文句 S_i 相關的文件集資訊。但是，在實際應用中並沒有這樣的資訊可以運用。因此，我們採用局部性回饋技術自動地建立每一文句所關聯性模型。如圖 4.1 所示。首先，我們將摘

要文件中每一文句 S_i 視為一個查詢，作為資訊檢索系統的輸入，接著經由資訊檢索系統對於與摘要文件同一時期的文件集進行檢索後，找出與文句 S_i 最相關的前 L 篇文件，視為相關的文件集；然後，利用此相關文件集來建立文句 S_i 的關聯性模型，並用於調適原文句 S_i 模型的機率分佈。每一文句 S_i 所對應的關聯性模型可以下面公式近似求得表示：

$$P(w|R_{S_i}) = \sum_{l=1}^{L} P(D_l|S_i)P(w|D_l) \tag{4-10}$$

$P(w|R_{S_i})$ 為詞 w 於文句 S_i 所對應的關聯性模型(R_{S_i})中的分佈機率值；L 為檢索系統所回傳的相關文件數；$P(w|D_l)$ 為第 l 篇相關文件中詞 w 的機率分佈，計算方式如式(4-6)所示；$P(D_l|S_i)$ 表示文句 S_i 生成第 l 篇相關文件的可能性，可透過貝式定理(Bayes' rule)表示成：

$$P(D_l|S_i) = \frac{P(S_i|D_l)P(D_l)}{P(S_i)} = \frac{P(S_i|D_l)P(D_l)}{\sum_{j=1}^{L} P(S_i|D_j)P(D_j)} \tag{4-11}$$

如假設文件 D_l 的事前機率為一致性分佈，式(4-11)可化簡為：

$$P(D_l|S_i) = \frac{P(S_i|D_l)}{\sum_{j=1}^{L} P(S_i|D_l)} \tag{4-12}$$

$P(D_l|S_i)$ 可視為一個權重值，表示文件 D_l 對於文句 S_i 的重要性。若檢索系統是採用機率生成模型(例如，隱藏式馬可夫模型)進行文件的檢索，$P(D_l|S_i)$ 則可直接由檢索系統的輸出結果取得。

本論文將每一文句 S_i 所對應的關聯性模型視為文句 S_i 模型參數的事前機率分佈，以狄利克雷分佈(Dirichlet Distribution)表示，經調適後之文句 S_i 模型的機率分佈如下所示：

$$P^*(w|S_i) = \sum_{l=1}^{L} P(D_l|S_i)\frac{S_i(w) + mP(w|D_l)}{|S_i| + m} \tag{4-13}$$

其中，關聯性模型為狄利克雷混合分佈(Dirichlet Mixtures Distribution)，表示文句 S_i 多項式分佈模型參數的事前機率分佈。m 為等量取樣大小(Equivalent Sample Size)，通常 m 值的估測採經驗法則，不同論文中設定不一，太多介於 500~3000

之間。式(4-13)，經由下面的公式推導，可看出文句模型與關聯生模型的線性插補法(Linear Interpolation)公式其實為狄利克雷混合分佈的一個特例(Special Case)：

$$
\begin{aligned}
P^*(w|S_i) &= \sum_{l=1}^{L} P(D_l|S_i)\frac{S_i(w)+mP(w|D_l)}{|S_i|+m} \\
&= \left[\sum_{l=1}^{L} P(D_l|S_i)\frac{mP(w|D_l)}{|S_i|+m}\right] + \left[\sum_{l=1}^{L} P(D_l|S_i)\frac{S_i(w)}{|S_i|+m}\right] \\
&= \left[\sum_{l=1}^{L} P(D_l|S_i)\frac{mP(w|D_l)}{|S_i|+m}\right] + \frac{S_i(w)}{|S_i|+m} \\
&= \left[\sum_{l=1}^{L} P(D_l|S_i)P(w|D_l)\frac{m}{|S_i|+m}\right] + \frac{|S_i|}{|S_i|+m}\cdot\frac{S_i(w)}{|S_i|} \\
&= \sum_{l=1}^{L} P(D_l|S_i)P(w|D_l)\left(1-\frac{|S_i|}{|S_i|+m}\right) + \frac{|S_i|}{|S_i|+m}\cdot P(w|S_i) \\
&= \alpha\cdot\sum_{l=1}^{L} P(D_l|S_i)P(w|D_l) + (1-\alpha)\cdot P(w|S_i) \qquad \text{where } \alpha = \left(1-\frac{|S_i|}{|S_i|+m}\right)
\end{aligned}
$$

(4-14)

在本論文的實驗中，由於m值的範圍很大(1~∞)，為了參數調整的方便，我們在下面有關於關聯性模型於文句模型調適的實驗裡，皆以調動參數α為主。最後，文句S_i的隱藏式馬可夫摘要模型，由調適後之文句模型及集合模型所組成，估測每一文句S_i生成文件的可能性D：

$$P(D|S_i) = \prod_{w\in D}\left[\lambda\cdot P^*(w|S_i) + (1-\lambda)\cdot P(w|C)\right]^{n(w,D)} \tag{4-15}$$

直觀來看，由每一文句所對應的關聯性模型分佈，可有以下二點觀察：

1. 當語音文件中文句S_i的內容包含有較重要概念或豐富資訊內容時，其所對應的關聯性模型分佈亂度(Entropy)較小且資訊較集中，對於概念或主題上不相關詞，機率值亦相對較小。同時越能代表文句本身的機率分佈。如此，可簡單的由關聯性分佈來看出文句S_i內容是否較具鑑別力及內容獨特性。

2. 反之，若文件中某一文句 S_i 的內容過於簡短，或是內容無意義或不具代表性，其所對應的關聯性模型的機率分佈亂度較大，不相主題與概念的詞於分佈中機率值會無明顯差異，則當此文句被視為查詢進行檢索時，檢索系統將會找到許多主題不相關文件。如此，亦表示關聯性模型對於文句模型的調適，將會把此文句模型機率分佈的估測偏差變得更大；這也代表此文句對於文件的發生可能性會降低，以結果來說，似乎可達到我們預期的目的。

從初步實驗結果來觀察，當文句愈具代表性，所對應的關聯性模型對於文句模型的調適亦愈有正向的貢獻；反之，當所對應的關聯性模型可能因文句本身內容無重要的資訊，而無法對文句分佈有一個好的估測，在文句模型的摘適將會帶來負向的貢獻。

4.2.3 關聯性模型於文句模型調適之實驗結果

本論文分別採用二種測試語料來觀察吾人所提之方法在摘錄式摘要的表現。同時，使用二種評估方法來進行實驗的分析與討論，評估方法的設定與基礎實驗相同。如此，我們可觀察關聯性模型於文句模型調適，在二種不同測試語料與評估方法下的結果，並且驗證所提之方法有助於語音文件摘要結果的提昇。以下實驗我們希望達到：

1. 加入關聯性模型於文句模型之調適，可以提昇原 HMM 的摘要結果，因此在摘要正確率上，本實驗結果應較原 HMM 來得高。

2. 加入關聯性模型於文句模型之調適，可以提昇機率式摘要模型的摘要正確率，也就是本實驗結果應較論文中的基礎實驗結果好。

下面，分別呈現在不用的測試語料、評估方法下，關聯性模型的調適於人工轉寫文件與自動轉寫文件摘要上的實驗結果，其中隱藏式馬可夫模型參數 λ 於

DataSet1 設定為 0.05，於 DataSet2 設定為 0.01，與基礎實驗設定相同。第三章基礎實驗中的隱藏式馬可夫模型結果，為本實驗的基礎實驗(Baseline)，觀察關聯性模型的調適是否可有效提昇隱藏式馬可夫模型的摘要結果；然後，再將最後的參數設定之摘要結果，與基礎實驗中各種摘要方法的摘要正確率比較。

測試語料 DataSet1 實驗結果：

在本實驗中，我們有三個主要的參數需要作設定與調整：式(2-10)中參數L、式(2-14) 中參數α值、及檢索系統的搜尋範圍(Search Range)設定。其中，α值是用來平衡每一文句的關聯性模型與原文句模型機率分佈的權重值。L值是用來決定建立關聯性模型的文件數，L值設定越大，表示關聯性模型中可能包含有更多與文句相關的資訊，但是相對地也可能表示關聯性模型中帶有更多雜訊，會為文句模型帶來混淆的資訊內容，因此，我們觀察不同L值設定下的實驗結果。搜尋範圍設定是用來限制檢索系統對檢索文件集的搜尋範圍；也就是說，對於每一篇語音新聞文件中的每一文句，我們對整個文字文件訓練語料集中所有文件進行檢索，而是只考慮檢索文字文件集訓練集中，與此新聞事件(文件)發生日期的同一天與發生前幾Dy天的新聞文件。因此，搜尋範圍設定即是決定要搜尋文字文件訓練集中，與每一篇文件發生日期同一天及前幾天中的所有文件，進行檢索並回傳相關文件。這樣做的目的為：

1. 整個文字文件訓練集包含有大量文件數，直接對文字文件集進行檢索，可能會造成摘要系統的效率不佳，我們希望加入關聯性模型的調適並不會增加系統太大的負擔；

2. 在實際且線上即時系統的應用上，當給定一篇欲摘要的新聞文件，我們可以取得的相關資訊(由網路上或是過去收集的資料庫中)只有過去已發生的新聞事件內容，並沒有得到未來的新聞事件。因為一個新聞事件通常會持續一段時間，所以從過去幾天的新聞中所找到相關的資訊，可以作為每一文句的關聯性模型一個很好的估測資訊。

下面實驗我們將觀察在發展集每一參數於不同設定下，摘要的結果，同時嘗

試找出一組最好的實驗設定，得到最佳的摘要正確率。最後，我們將相同的實驗設定在測試集中進行觀察與分析。本實驗中，為了增加每一文句所對應的關聯性模型的正確性，檢索系統採用雙音節(Bi-Syllables)為索引進行檢索。

討論參數α值設定：

我們先以檢索系統搜尋每一篇摘要文件中，與每一文句相關的前五篇文件(L=5)來建立關聯性模型，搜尋範圍Dy設為5，作為初步的實驗設定。然後，觀察不同的α值下之摘要結果。表 4.1 為 DataSet1 發展集中，人工轉寫文件分別在餘弦評估及 Rouge-2 評估下的結果。表中每一行表示在α值為 0、0.1、…、1.0 時，不同摘要比例下的摘要正確率。若α=0 代表原來隱藏式馬可夫模型的結果，α=1 代表僅使用每一文句的關聯性模型進行文件摘要處理。

由表 4.1 的結果來看，在α值介於 0.1~1.0 的設定中，二種評估方法結果均顯示出當α=0.1 時，有最好的摘要結果；同時，正確率於摘要比例 20%、30%、70%下，皆較原來隱藏式馬可夫模型的結果好。特別是在 Rouge-2 評估下摘要比例 30%時，有 20%以上的進步。經由初步的實驗結果，可以證明所提的方法可能是有用的，雖然結果於摘要比例 10%時卻有明顯地下降。

表 4.2 為 DataSet1 發展集中，自動轉寫文件分別在餘弦評估及 Rouge-2 評估下的結果。表中每一行分別表示在不同α值設定時，不同摘要比例下的摘要正確率。同樣地，α=0 代表原來隱藏式馬可夫模型的結果，α=1 為只使用關聯性模型進行文件摘要處理的結果。由表 4.2 我們可以觀察出不同α值設定，α值介於 0.1~1.0 的設定中，當α=0.1 時有最好的摘要結果，而且隨著α值的遞增，不同摘要比例下的正確率有遞減的情形，與人工轉寫文件有一致性的結果。當α=0.1 時，在餘弦評估的結果中，摘要比例 30%、40%、50%時正確率有些微的提昇，然而 Rouge-2 評估結果中，在摘要比例 10%、20%、30%、70%下正確率都有小幅度的進步。

雖然，於表 4.1 與 4.2 結果中關聯性模型於原文句模型的調適結果，對

於正確率的提昇並不明顯。但是，由初步結果可以觀察出關聯性模型於隱藏式馬可夫模型上的應用，正確率確實有些微進步，對隱藏式馬可夫模型的摘要結果帶來某種程度上的提昇。

表 4.1 DataSet1,發展集：關聯性模型於隱藏式馬可夫模型調適結果－α 值分析 (人工轉寫文件)

Cosine	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*1*
10%	0.4203	0.4073	0.4031	0.4016	0.4002	0.3933	0.3886	0.3871	0.3733	0.3361	0.3030
20%	0.4657	0.4663	0.4640	0.4661	0.4566	0.4607	0.4580	0.4538	0.4353	0.3992	0.3752
30%	0.5178	0.5293	0.5205	0.5234	0.5191	0.5119	0.5082	0.5130	0.5073	0.4935	0.4640
50%	0.6271	0.6257	0.6235	0.6202	0.6200	0.6184	0.6162	0.6135	0.6165	0.6068	0.5864
70%	0.6996	0.7025	0.6999	0.6974	0.6947	0.6946	0.6948	0.6942	0.6947	0.6926	0.6729
Rouge-2	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*1*
10%	0.4157	0.4064	0.3920	0.3860	0.3739	0.3719	0.3648	0.3615	0.3351	0.2911	0.2655
20%	0.4591	0.4628	0.4556	0.4599	0.4380	0.4443	0.4337	0.4242	0.3911	0.3508	0.3425
30%	0.4947	0.5202	0.5030	0.5112	0.5011	0.4926	0.4863	0.4920	0.4745	0.4590	0.4190
50%	0.6527	0.6511	0.6470	0.6336	0.6294	0.6290	0.6291	0.6261	0.6290	0.6107	0.5761
70%	0.7845	0.7857	0.7806	0.7710	0.7631	0.7608	0.7572	0.7539	0.7519	0.7409	0.6995

表 4.2 DataSet1,發展集：關聯性模型於隱藏式馬可夫模型調適結果－α 值分析(自動轉寫文件)

Cosine	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*1*
10%	0.3415	0.3330	0.3149	0.3246	0.3246	0.3204	0.3161	0.3163	0.3070	0.2945	0.2666
20%	0.3789	0.3788	0.3646	0.3661	0.3624	0.3617	0.3573	0.3544	0.3474	0.3309	0.3186
30%	0.4281	0.4338	0.4176	0.4163	0.4131	0.4057	0.3985	0.3965	0.3950	0.3883	0.3812
50%	0.5159	0.5182	0.5165	0.5136	0.5126	0.5085	0.5095	0.5101	0.5061	0.5078	0.4958
70%	0.5847	0.5872	0.5868	0.5877	0.5883	0.5877	0.5868	0.5857	0.5823	0.5824	0.5687
Rouge-2	*0*	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*1*
10%	0.3084	0.3131	0.2890	0.3035	0.3035	0.2928	0.2904	0.2865	0.2777	0.2578	0.2207
20%	0.3467	0.3531	0.3376	0.3368	0.3329	0.3282	0.3254	0.3200	0.3155	0.2958	0.2738
30%	0.3734	0.3846	0.3591	0.3548	0.3509	0.3367	0.3268	0.3184	0.3234	0.3225	0.3109
50%	0.4768	0.4750	0.4765	0.4698	0.4684	0.4601	0.4563	0.4520	0.4392	0.4415	0.4188
70%	0.5631	0.5661	0.5656	0.5644	0.5653	0.5630	0.5596	0.5556	0.5509	0.5453	0.5185

表 4.3 DataSet1,發展集：關聯性模型於隱藏式馬可夫模型調適結果－ *Dy* 值分析(人工轉寫文件)

Cosine	*Baseline*	*5*	*10*	*15*	*20*	*25*	*30*
10%	0.4203	0.4073	0.4163	0.4155	0.4151	0.4189	0.4230
20%	0.4657	0.4663	0.4709	0.4681	0.4677	0.4716	0.4716
30%	0.5178	0.5293	0.5347	0.5341	0.5340	0.5342	0.5334
50%	0.6271	0.6257	0.6278	0.6270	0.6267	0.6247	0.6248
70%	0.6996	0.7025	0.7022	0.7041	0.7042	0.7022	0.7016
Rouge-2	*Baseline*	*5*	*10*	*15*	*20*	*25*	*30*
10%	0.4157	0.4064	0.4234	0.4187	0.4175	0.4224	0.4313
20%	0.4591	0.4628	0.4692	0.4646	0.4634	0.4683	0.4683
30%	0.4947	0.5202	0.5307	0.5272	0.5276	0.5284	0.5280
50%	0.6527	0.6511	0.6539	0.6534	0.6512	0.6434	0.6444
70%	0.7845	0.7857	0.7838	0.7880	0.7869	0.7858	0.7861

- 參數 *Dy* 值設定：

延續前面的實驗，我們想要進一步探討不同的檢索範圍，對於關聯性模型結果的影響；也就是研究檢索系統所搜尋的文件數量大小，於關聯性模型實驗中的正確率結果。同樣地，我們仍以檢索系統所回傳的前五篇文件(L=5)，來建立關聯性模型，但是檢索的範圍 *Dy* 值嘗試採用 5、10、15、20、25、30 等設定。結果如表 4.3、表 4.4 所示，其中 α 值設定為 0.1。

表 4.3、表 4.4 分別為發展集中，人工轉寫文件與自動轉寫文件在不同 *Dy* 值的結果；表格上半部分為餘弦評估結果，下半部分為 Rouge-2 評估結果。首先，由人工轉寫文件的結果來觀察(表 4.3)，可以發現於摘要比例 10%s 時，摘要正確率與 *Dy* 的值呈現出某種程度上的線性關係，也就是說當 *Dy* 增加正確率亦有些微的提昇。但是，於其他摘要比例下就無此一線性關係存在。餘弦評估結果中，不同 *Dy* 設定為何，在摘要比例為 20%、30%及 70%時，正確率皆有小幅度的進步；而當 *Dy* =30 時，低摘要比例

10%的正確率亦有所進步。Rouge-2 評估的結果中，除了於摘要比例 50%下部分的摘要結果有些微的下降外，其他摘要比例結果在不同 *Dy* 值設定下，均有明顯的提昇；當摘要比例為 30%時，甚至有 2~3%的絕對進步。

表 4.4 DataSet1,發展集：關聯性模型於隱藏式馬可夫模型調適結果－ *Dy* 值分析(自動轉寫文件)

Cosine	*Baseline*	*5*	*10*	*15*	*20*	*25*	*30*
10%	0.3415	0.3330	0.3369	0.3375	0.3375	0.3375	0.3370
20%	0.3789	0.3788	0.3787	0.3787	0.3787	0.3787	0.3771
30%	0.4281	0.4338	0.4327	0.4322	0.4322	0.4322	0.4301
50%	0.5159	0.5182	0.5196	0.5201	0.5194	0.5192	0.5190
70%	0.5847	0.5872	0.5881	0.5880	0.5890	0.5904	0.5888
Rouge-2	*Baseline*	*5*	*10*	*15*	*20*	*25*	*30*
10%	0.3084	0.3131	0.3180	0.3208	0.3208	0.3208	0.3224
20%	0.3467	0.3531	0.3549	0.3549	0.3549	0.3549	0.3550
30%	0.3734	0.3846	0.3838	0.3827	0.3827	0.3827	0.3842
50%	0.4768	0.4750	0.4774	0.4778	0.4769	0.4768	0.4769
70%	0.5631	0.5661	0.5673	0.5667	0.5686	0.5702	0.5701

接著，由表 4.4 實驗結果中，可以看出關聯性模型對於自動轉寫文件文句模型調適的結果。餘弦評估結果顯示在摘要比例 30%、50%及 70%時，正確率在不同 *Dy* 設定值下皆有進步，但是於低摘要比時 10%正確率是下降的。然而，從 Rouge-2 評估的結果，我們可以發現無論是在低摘要比例或是高摘要比例下，正確率均有所進步，特別是在摘要比例 10%、20%、30%時，正確率均有 0.5%～1.4%的絕對進步。

這部分的實驗結果顯示不同 *Dy* 值的設定對於正確率的影響並不會太大，但是 *Dy* 值設定為 30 似乎有最佳的結果。在 Rouge-2 評估結果中，實驗結果皆呈現出所提之方法有不錯的進步；而在餘弦評估中，部分結果顯示出小幅度的進步。

- 參數L值設定：

接下來的實驗，我們分析建立關聯性模型所使用的文件數L，對於摘要正確率的影響。實驗中，α值設定為 0.1，Dy值為 30。實驗結果如表 4.5、表 4.6 所示，我們很驚訝地發現在不同文件數L設定值下的摘要結果很相似，尤其當摘要比例越低時，不同文件數L的結果是相同的。這樣的實驗結果，讓我們重新思考實驗中的每一個環節，尋找合理的理由來解釋這個現象。我們發現可能的原因為：

1. 公式(4-10)中，機率值$P(D_l \mid S_i)$的影響。當機率值$P(D_l \mid S_i)$對式(4-10)的機率值$P(w \mid R_{S_i})$影響很大時，若所回傳的相關文件中，相關性排名越前面的機率值$P(D_l \mid S_i)$很大，而越後面的相關性文件之機率值$P(D_l \mid S_i)$很小，那麼關聯性模型中機率值$P(w \mid R_{S_i})$便會受最前面幾篇相關文件的影響很大，不論L值設定為多少，關聯性模型中詞的機率分佈只會被少數相關文件所支配(Dominate)。

表 4.5 DataSet1,發展集：關聯性模型於隱藏式馬可夫模型調適結果－L值分析(人工轉寫文件)

Cosine	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.4230	0.4230	0.4230	0.4230	0.4230	0.4229
20%	0.4716	0.4716	0.4716	0.4716	0.4716	0.4716
30%	0.5334	0.5334	0.5334	0.5334	0.5334	0.5334
50%	0.6248	0.6248	0.6248	0.6248	0.6248	0.6248
70%	0.7016	0.7018	0.7020	0.7021	0.7021	0.7021
Rouge-2	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.4313	0.4313	0.4313	0.4313	0.4313	0.4313
20%	0.4683	0.4683	0.4683	0.4683	0.4683	0.4683
30%	0.5280	0.5280	0.5284	0.5280	0.5280	0.5280
50%	0.6444	0.6447	0.6447	0.6447	0.6447	0.6447
70%	0.7861	0.7868	0.7869	0.7867	0.7867	0.7867

表 4.6 DataSet1,發展集：關聯性模型於隱藏式馬可夫模型調適結果－L 值分析(自動轉寫文件)

Cosine	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.3370	0.3370	0.3370	0.3370	0.3370	0.3370
20%	0.3771	0.3771	0.3771	0.3771	0.3771	0.3771
30%	0.4301	0.4301	0.4301	0.4301	0.4301	0.4301
50%	0.5190	0.5181	0.5181	0.5181	0.5173	0.5173
70%	0.5888	0.5887	0.5887	0.5887	0.5887	0.5887
Rouge-2	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.3224	0.3224	0.3224	0.3224	0.3224	0.3224
20%	0.3550	0.3550	0.3550	0.3550	0.3550	0.3550
30%	0.3842	0.3842	0.3842	0.3842	0.3842	0.3842
50%	0.4769	0.4759	0.4759	0.4759	0.4756	0.4756
70%	0.5701	0.5703	0.5703	0.5703	0.5703	0.5703

2. 若所回傳的 30 篇相關文件資訊很集中，不會因為文件數多而造成所建立的關聯性模型包含太多雜訊，而使正確率下降的情況下，可能會發生此情形。

因此，針對上述原因 1，我們嘗試減少機率值 $P(D_l \mid S_i)$ 對關聯性模型的影響，對機率值 $P(D_l \mid S_i)$ 加上一個次方項，重新進行實驗分析。表 4.7 與表 4.8 為機率值 $P(D_l \mid S_i)$ 開 0.01 次方後的結果。由表 4.7 結果可以看出在不同 L 值設定下之摘要結果有較明顯的不同，但是結果均較表 4.5 中的結果較差。因此，對於人工轉寫文件而言，機率值 $P(D_l \mid S_i)$ 的次方項設定為 1 會有較佳的結果；同時與第三章所列之所有摘要方法結果比較，於餘弦評估結果中，摘要比例 20%、30%的結果較好，於 Rouge-2 評估結果中，不同摘要比例的結果皆較所有基礎實驗方法好。表 4.7 結果顯示於餘弦評估結果中，當 L 值為 30 時於低摘要比例 10%、20%下較基礎實驗隱藏式馬可夫模型的結果好，於 Rouge-2 評估結果中，不同的 L 值設定下於摘要比

例 10%、20%、30%時較基礎實驗隱藏式馬可夫模型的結果好，並且於 L 值為 30 時摘要比例 10%、20%的結果較第三章中所有摘要方法結果好。比較表 4.6 與表 4.8 的結果，可以看出於自動轉寫文件結果中，當機率值 $P(D_l \mid S_i)$ 的次方項設定為 0.01 時有更高的正確率，同時於餘弦評估方法下低摘要比例 10%、20%均較基礎實驗隱藏式馬可夫模型的結果好，並且當 L 值為 30 時摘要比例 10%、20%及 30%的結果較第三章中所有摘要方法結果好；於 Rouge-2 評估不同 L 值設定結果中皆較第三章中所有摘要方法結果好。

在 DataSet1 發展集所進行的實驗中，我們嘗試了三種參數的調整，並且試著分析實驗結果。然而，對於影響正確率的因素還有許多，例如上面所提到的機率值 $P(D_l \mid S_i)$ 設定問題。最後，我們使用與發展集實驗相同的實驗設定於測試

表 4.7 DataSet1,發展集：關聯性模型於隱藏式馬可夫模型調適結果－
調整機率值 $P(D_l \mid S_i)$ 次方項 (人工轉寫文件)

Cosine	*Baseline*	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.4203	0.4195	0.4180	0.4180	0.4189	0.4193	0.4216
20%	0.4657	0.4723	0.4661	0.4701	0.4709	0.4714	0.4714
30%	0.5178	0.5283	0.5159	0.5175	0.5174	0.5191	0.5175
50%	0.6271	0.6282	0.6268	0.6276	0.6275	0.6271	0.6263
70%	0.6996	0.7006	0.7002	0.7002	0.7016	0.7012	0.7008
Rouge-2	*Baseline*	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.4157	0.4159	0.4195	0.4195	0.4233	0.4233	0.4261
20%	0.4591	0.4678	0.4649	0.4688	0.4726	0.4726	0.4726
30%	0.4947	0.5127	0.4974	0.5045	0.4992	0.5013	0.4984
50%	0.6527	0.6515	0.6479	0.6457	0.6452	0.6459	0.6456
70%	0.7845	0.7836	0.7847	0.7841	0.7841	0.7847	0.7846

表 4.8 DataSet1,發展集：關聯性模型於隱藏式馬可夫模型調適結果－
調整機率值 $P(D_l \mid S_i)$ 次方項 (自動轉寫文件)

Cosine	*Baseline*	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.3415	0.3465	0.3511	0.3477	0.3459	0.3459	0.3462
20%	0.3789	0.3837	0.3872	0.3827	0.3827	0.3823	0.3826
30%	0.4281	0.4253	0.4289	0.4227	0.4255	0.4296	0.4307
50%	0.5159	0.5188	0.5165	0.5169	0.5168	0.5166	0.5171
70%	0.5847	0.5877	0.5855	0.5850	0.5847	0.5847	0.5849
Rouge-2	*Baseline*	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.3084	0.3369	0.3377	0.3231	0.3179	0.3179	0.3165
20%	0.3467	0.3757	0.3667	0.3527	0.3527	0.3526	0.3512
30%	0.3734	0.3725	0.3757	0.3614	0.3682	0.3745	0.3753
50%	0.4768	0.4779	0.4744	0.4750	0.4754	0.4748	0.4750
70%	0.5631	0.5652	0.5628	0.5634	0.5631	0.5630	0.5629

表 4.9 DataSet1,測試集：關聯性模型於隱藏式馬可夫模型調適結果 (人工轉寫文件)

Cosine	*Baseline*	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.3841	0.3945	0.4055	0.4046	0.4056	0.4023	0.4002
20%	0.4064	0.4054	0.4209	0.4150	0.4157	0.4148	0.4179
30%	0.4855	0.4782	0.4805	0.4807	0.4850	0.4874	0.4890
50%	0.6037	0.5982	0.5990	0.6011	0.6016	0.6042	0.6050
70%	0.6820	0.6771	0.6802	0.6816	0.6815	0.6830	0.6830
Rouge-2	*Baseline*	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.3714	0.3922	0.4087	0.4090	0.4105	0.4049	0.3980
20%	0.3892	0.3903	0.4166	0.4072	0.4085	0.4063	0.4092
30%	0.4669	0.4491	0.4575	0.4608	0.4670	0.4727	0.4739
50%	0.6308	0.6286	0.6252	0.6312	0.6323	0.6349	0.6353
70%	0.7660	0.7650	0.7658	0.7648	0.7657	0.7678	0.7678

表 4.10 DataSet1,測試集：關聯性模型於隱藏式馬可夫模型調適結果(自動轉寫文件)

Cosine	*Baseline*	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.3155	0.3213	0.3213	0.3218	0.3218	0.3218	0.3218
20%	0.3381	0.3403	0.3388	0.3402	0.3402	0.3404	0.3410
30%	0.3993	0.4033	0.4010	0.4001	0.3991	0.3989	0.3984
50%	0.5027	0.5049	0.5046	0.5031	0.5021	0.5023	0.5027
70%	0.5601	0.5611	0.5629	0.5622	0.5622	0.5613	0.5630
Rouge-2	*Baseline*	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.2932	0.3182	0.3169	0.3168	0.3168	0.3168	0.3168
20%	0.3191	0.3264	0.3252	0.3240	0.3240	0.3249	0.3251
30%	0.3705	0.3671	0.3720	0.3731	0.3689	0.3692	0.3692
50%	0.4732	0.4774	0.4767	0.4745	0.4745	0.4754	0.4763
70%	0.5595	0.5596	0.5609	0.5609	0.5609	0.5609	0.5617

集，結果如表 4.9 與表 4.10 所示。其中，α=0.1，Dy=30，機率值$P(D_l \mid S_i)$次方項為 0.01。

由測試集的實驗結果(表 4.9 與表 4.10)觀察，人工轉寫文件的餘弦評估結果中，當L值設定為 25、30 正確率較隱藏式馬可夫模型的結果(Baseline)好，甚至較第三章中所有摘要方法結果來得好；同樣地，Rouge-2 評估結果中，當L值設定為 20、25 及 30 時，正確率亦較隱藏式馬可夫模型的結果(Baseline)好，甚至比所有摘要方法結果好。在自動轉寫文件的結果中，於二種評估方法下正確率均要較隱藏式馬可夫模型的結果(Baseline)好，特別是於 Rouge-2 評估結果中有明顯的進步，例如摘要比例 10%有 2%以上的絕對進步；與基礎實驗中的所有摘要方法結果比較，表 4.10 的餘弦評估結果於摘要比例 10%時較機率生成式摘要方法 HMM 與 TMM 好，但是沒有較其他摘要方法好，於 Rouge-2 評估結果亦沒有較其他摘要方法好。我們認為是因發展集與測試集不匹配的原因。在關聯性模型於隱藏式馬可夫模型調適的實驗中，關聯性模型的參數(α 、 Dy 、 L)或隱藏式馬

可夫模型參數λ於發展集中所調整的最佳參數設定，於測試集中並一定會得到最佳的實驗結果，這一點我們可從測試集其他參數設定的結果證明。但是，這樣的問題也可能發生於其他摘要方法上，因此對於所提方法之參數的設定，或許未來我們可以進一步使用自動的方法來決定。

測試語料 DataSet2 實驗結果：

我們亦於 DataSet2 進行關聯性模型於隱藏式馬可夫模型的摘要實驗，其中參數 α 、L 及 Dy 的設定同樣先以 DataSet2 的發展集作調整，然後使用於測試集。由於 DataSet1 的實驗中發現機率值 $P(D_l \mid S_i)$ 對於摘要結果的影響，因此於 DataSet2 的實驗中機率值 $P(D_l \mid S_i)$ 亦可嘗試不同次方項的結果，此部分的實驗我們對於不同次方項的摘要結果不作進一步的探討，以下的實驗結果皆為次方項設定為 0.05 下的摘要結果。

- 參數α值設定：

  實驗先將參數 L 及 Dy 設定 5，觀察不同的α值(0.1~0.9)的摘要結果。表 4.11

表 4.11 DataSet2,發展集：關聯性模型於隱藏式馬可夫模型調適結果－α值分析 (人工轉寫文件)

Cosine	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*1*
10%	0.4534	0.4588	0.4605	0.4600	0.4542	0.4547	0.4513	0.4500	0.4415	0.4231
20%	0.5776	0.5816	0.5904	0.5911	0.5967	0.5964	0.5964	0.5922	0.5843	0.5779
30%	0.6429	0.6476	0.6533	0.6502	0.6553	0.6633	0.6571	0.6534	0.6508	0.6376
50%	0.7434	0.7468	0.7478	0.7483	0.7459	0.7478	0.7439	0.7430	0.7407	0.7289
70%	0.8005	0.8016	0.8032	0.8022	0.8029	0.8035	0.8028	0.8002	0.7984	0.7871
Rouge-2	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*1*
10%	0.2926	0.2989	0.3045	0.3017	0.2871	0.2931	0.2929	0.2935	0.2987	0.2901
20%	0.4052	0.4100	0.4183	0.4151	0.4296	0.4269	0.4301	0.4238	0.4180	0.4175
30%	0.4859	0.4950	0.5093	0.4998	0.5112	0.5199	0.5047	0.4921	0.4888	0.4901
50%	0.6645	0.6661	0.6635	0.6610	0.6529	0.6541	0.6486	0.6420	0.6327	0.6276
70%	0.8001	0.8007	0.7980	0.7929	0.7920	0.7906	0.7891	0.7851	0.7743	0.7612

表 4.12 DataSet2,發展集：關聯性模型於隱藏式馬可夫模型調適結果－α 值分析 (自動轉寫文件)

Cosine	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*1*
10%	0.4185	0.4183	0.4135	0.4078	0.4085	0.4071	0.4046	0.3990	0.3882	0.3637
20%	0.5179	0.5195	0.5168	0.5184	0.5172	0.5169	0.5157	0.5105	0.5005	0.4632
30%	0.5623	0.5639	0.5661	0.5677	0.5664	0.5629	0.5605	0.5577	0.5494	0.5234
50%	0.6200	0.6205	0.6196	0.6206	0.6192	0.6193	0.6164	0.6127	0.6052	0.5777
70%	0.6341	0.6341	0.6350	0.6337	0.6336	0.6330	0.6321	0.6301	0.6261	0.6094
Rouge-2	*0.1*	*0.2*	*0.3*	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*1*
10%	0.2265	0.2282	0.2230	0.2158	0.2148	0.2122	0.2140	0.2156	0.2130	0.2128
20%	0.2797	0.2788	0.2746	0.2784	0.2769	0.2780	0.2767	0.2717	0.2647	0.2469
30%	0.3167	0.3182	0.3194	0.3184	0.3132	0.3108	0.3111	0.3075	0.2985	0.2898
50%	0.3966	0.3960	0.3938	0.3937	0.3926	0.3917	0.3876	0.3837	0.3755	0.3559
70%	0.4320	0.4321	0.4329	0.4303	0.4284	0.4272	0.4245	0.4217	0.4153	0.4002

為發展集人工轉寫文件於餘弦評估及 Rouge-2 評估下的結果。從實驗結果中顯示 α 值於 0.1~0.6 之間皆有不錯的摘要結果，同時於不同評估方法下皆較隱藏式馬可夫模型的結果好。由表 4.11 中可以發現當 α 值為 3 或 4 時有最好的結果。表 4.12 為發展集自動轉寫文件於餘弦評估及 Rouge-2 評估下的結果，當 α 值介於 0.1~0.3 之間摘要結果皆較隱藏式馬可夫模型好。由表 4.11 及表 4.12 結果發現最佳的 α 值設定並不相同，我們進而觀察不同的 L 及 Dy 值設定下不同 α 值的摘要結果，發現於其他 L 及 Dy 值設定下，大部分實驗結果均顯示 α 值於人工轉寫文件中設定為 0.4 時有最佳的結果，於自動轉寫文件中設定為 0.2 時有最佳的結果。

- 參數 Dy 值設定：

  接下來，我們探討不同的檢索範圍的摘要結果。實驗中仍以檢索系統所回傳的前五篇文件(L=5)來建立關聯性模型，於人工轉寫文件之 α 值設定為 0.4，自動轉寫文件之 α 值設定為 0.2，Dy 值嘗試以 5、10、15、20、25、30 等設定。結果如表 4.13、表 4.14 所示。

表 4.13 為人工轉寫文件於不同評估方法下的摘要結果，從結果可以看出於不同 *Dy* 值下正確率均較隱藏式馬可夫模型高；同時於二種評估方法下，當 *Dy* 值為 15 及 20 時均有較好的結果。進一步比較 *Dy* 值 15 及 20 時於低摘要比例 10%、20%與 30%的結果，當 *Dy* =20 可得到不錯的結果。表

表 4.13 DataSet2,發展集：關聯性模型於隱藏式馬可夫模型調適結果－ *Dy* 值分析(人工轉寫文件)

Cosine	*5*	*10*	*15*	*20*	*25*	*30*
10%	0.4600	0.4590	0.4687	0.4684	0.4644	0.4628
20%	0.5911	0.5836	0.5831	0.5891	0.5914	0.5923
30%	0.6502	0.6534	0.6512	0.6536	0.6522	0.6533
50%	0.7483	0.7497	0.7488	0.7509	0.7499	0.7487
70%	0.8022	0.8029	0.8019	0.8034	0.8041	0.8031
Rouge-2	*5*	*10*	*15*	*20*	*25*	*30*
10%	0.3017	0.3027	0.3214	0.3209	0.3134	0.3101
20%	0.4151	0.4085	0.4156	0.4207	0.4228	0.4266
30%	0.4998	0.5037	0.5044	0.5103	0.5028	0.5059
50%	0.6610	0.6609	0.6608	0.6635	0.6607	0.6617
70%	0.7929	0.7945	0.7942	0.7985	0.7994	0.8000

表 4.14 DataSet2,發展集：關聯性模型於隱藏式馬可夫模型調適結果－ *Dy* 值分析(自動轉寫文件)

Cosine	*5*	*10*	*15*	*20*	*25*	*30*
10%	0.4183	0.4159	0.4163	0.4217	0.4212	0.4223
20%	0.5195	0.5176	0.5199	0.5167	0.5185	0.5183
30%	0.5639	0.5641	0.5644	0.5629	0.5637	0.5628
50%	0.6205	0.6211	0.6198	0.6206	0.6205	0.6202
70%	0.6341	0.6339	0.6348	0.6348	0.6351	0.6348
Rouge-2	*5*	*10*	*15*	*20*	*25*	*30*
10%	0.2282	0.2263	0.2273	0.2308	0.2328	0.2387
20%	0.2788	0.2742	0.2803	0.2794	0.2766	0.2739
30%	0.3182	0.3165	0.3174	0.3157	0.3170	0.3154
50%	0.3960	0.3962	0.3945	0.3968	0.3965	0.3961
70%	0.4321	0.4316	0.4328	0.4324	0.4314	0.4312

4.14 為自動轉寫文件於不同評估方法下的結果，於餘弦評估結果中不同 *Dy* 值設定之正確率皆較隱藏式馬可夫模型高，當 *Dy* 值為 25 及 30 時有最好結果；於 Rouge-2 評估結果中，不同 *Dy* 值下低摘要比例 10%的正確率皆較隱藏式馬可夫模型高，當 *Dy* 值為 25 及 30 時有最佳的結果。進一步比較自動轉寫文件結果中，*Dy* 值為 25 及 30 時於低摘要比例 10%、20%與 30%的結果，當 *Dy* =25 可得到不錯的結果。

- 參數 L 值設定：

  最後，我們探討使用於建立關聯性模型的文件數 L，在 DataSet2 中對於摘要正確率的影響。同樣地，α 值於人工轉寫文件設定為 0.4，於自動轉寫文件設定為 0.2。*Dy* 值的設定採用前面實驗的結果，於人工轉寫文件設為 20，於自動轉寫文件設為 25。實驗結果如表 4.15、表 4.16 所示。如表 4.15 所示，當 L 值為 2 時，人工轉寫文件於不同評估方法下皆有最高的的摘要正確率；同時，在低摘要比例 10%、20% 時，亦較隱藏式馬可夫模型及

表 4.15 DataSet1,發展集：關聯性模型於隱藏式馬可夫模型調適結果－ L 值分析(人工轉寫文件)

Cosine	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.4684	0.4597	0.4541	0.4577	0.4601	0.4578
20%	0.5891	0.5897	0.5863	0.5863	0.5853	0.5858
30%	0.6536	0.6514	0.6508	0.6489	0.6506	0.6494
50%	0.7509	0.7499	0.7477	0.7456	0.7454	0.7452
70%	0.8034	0.8013	0.8016	0.8018	0.8016	0.8007
Rouge-2	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.3209	0.3063	0.3029	0.3107	0.3151	0.2998
20%	0.4207	0.4174	0.4170	0.4175	0.4148	0.4146
30%	0.5103	0.5010	0.5017	0.4949	0.4949	0.4934
50%	0.6635	0.6633	0.6609	0.6609	0.6603	0.6586
70%	0.7985	0.7933	0.7951	0.7952	0.7974	0.7947

表 4.16 DataSet1,發展集：關聯性模型於隱藏式馬可夫模型調適結果－L 值分析(自動轉寫文件)

Cosine	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.4212	0.4215	0.4146	0.4112	0.4119	0.4122
20%	0.5185	0.5196	0.5184	0.5193	0.5182	0.5176
30%	0.5637	0.5614	0.5589	0.5612	0.5611	0.5610
50%	0.6205	0.6200	0.6203	0.6190	0.6198	0.6198
70%	0.6351	0.6344	0.6345	0.6344	0.6341	0.6340
Rouge-2	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.2328	0.2330	0.2221	0.2196	0.2193	0.2199
20%	0.2766	0.2782	0.2791	0.2777	0.2762	0.2759
30%	0.3170	0.3148	0.3117	0.3155	0.3156	0.3156
50%	0.3965	0.3963	0.3972	0.3957	0.3968	0.3965
70%	0.4314	0.4314	0.4319	0.4316	0.4311	0.4315

其他基礎實驗的摘要方法之結果好。表 4.16 為自動轉寫文件於不同評估方法的摘要結果，當 L 值為 10 時有最高的摘要正確率，並且於在低摘要比例 10%、20% 時較隱藏式馬可夫模型的結果好。比較與其他基礎實驗中不同摘要方法的結果，當 L =10 時在餘弦評估方法下於摘要比例 20%、30% 有較好的結果；當 L =15 時在 Rouge-2 評估方法下於低摘要比例 10%、20% 有較好的結果。

由發展集的實驗結果可以證明關聯性模型於隱藏式馬可夫模型調適的方法對摘要正確率是有幫助的，透過建立每一文句的關聯性模型，可有效地提昇隱藏式馬可夫模型的摘要結果，甚至較其他摘要方法的結果好。接著，我們使用相同的參數設定於測試集中，實驗結果表 4.17 及表 4.18 所示。其中，於人工轉寫文件 α 值為 0.4，於自動轉寫文件 α 值為 0.2； Dy 值於人工轉寫文件設定為 20，於自動轉寫文件設為定 25；機率值 $P(D_l \mid S_i)$ 次方項為 0.05。表 4.17 為人工轉寫文件於不

同評估方法的摘要結果；當 L=5 時，二種評估方法於低摘要比例 10%、20%及30%的摘要結果，皆較隱藏式馬可夫模型的結果好；於餘弦評估方法下，當 L=5時不同摘要比例的結果皆較其他基礎實驗摘要方法好，於 Rouge-2 評估方法下，當 L=20、25 及 30 時在低摘要比例 10%結果較其他基礎實驗摘要方法好。表 4.18為自動轉寫文件於不同評估方法的摘要結果；當 L=30 時，二種評估方法於低摘要比例 10%有最好的摘要結果。於餘弦評估方法下，當 L=10、15 時於不同摘要比例的結果均較隱藏式馬可夫模型好，但是結果較基礎實驗中 VSM 及 MMR 等相似度估測(餘弦估測)的摘要方法差；於 Rouge-2 評估方法下，當 L=25 時摘要比例 10%、20%及 30%的結果較隱藏式馬可夫模型好，同時亦較基礎實驗中非機率生成式摘要方法的正確率高；但是，較 TMM 結果差。

從 DataSet2 的實驗結果中，我們同樣發現了發展集與測試集不匹配情形，但是實驗結果仍然顯示出使用關聯性模型應用於文件摘要中，可以有效提昇摘要的正確率，特別是對於低摘要比例下的摘要結果。此外，於 Rouge-2 的實驗結果

表 4.17 DataSet2,測試集：關聯性模型於隱藏式馬可夫模型調適結果 (人工轉寫文件)

Cosine	*Baseline*	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.4835	0.5000	0.4936	0.4900	0.5046	0.5076	0.5044
20%	0.6166	0.6258	0.6162	0.6164	0.6154	0.6165	0.6187
30%	0.6579	0.6685	0.6684	0.6646	0.6642	0.6613	0.6609
50%	0.7338	0.7406	0.7398	0.7395	0.7403	0.7386	0.7386
70%	0.7873	0.7914	0.7919	0.7912	0.7913	0.7908	0.7910
Rouge-2	*Baseline*	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.3280	0.3473	0.3361	0.3301	0.3568	0.3592	0.3562
20%	0.4499	0.4590	0.4463	0.4459	0.4425	0.4424	0.4444
30%	0.5012	0.5031	0.5055	0.5013	0.5014	0.4986	0.4962
50%	0.6534	0.6520	0.6512	0.6503	0.6512	0.6487	0.6493
70%	0.7915	0.7908	0.7948	0.7940	0.7938	0.7926	0.7928

表 4.18 DataSet2,測試集：關聯性模型於隱藏式馬可夫模型調適結果 (自動轉寫文件)

Cosine	*Baseline*	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.3837	0.3847	0.3880	0.3876	0.3906	0.3911	0.3922
20%	0.4739	0.4793	0.4792	0.4742	0.4726	0.4739	0.4732
30%	0.5110	0.5129	0.5130	0.5135	0.5133	0.5128	0.5128
50%	0.5619	0.5667	0.5667	0.5671	0.5666	0.5661	0.5659
70%	0.5837	0.5853	0.5850	0.5852	0.5848	0.5848	0.5845
Rouge-2	*Baseline*	*L=5*	*L=10*	*L=15*	*L=20*	*L=25*	*L=30*
10%	0.2008	0.1998	0.2050	0.2044	0.2073	0.2085	0.2108
20%	0.2501	0.2542	0.2587	0.2510	0.2500	0.2506	0.2493
30%	0.2820	0.2850	0.2855	0.2855	0.2842	0.2847	0.2839
50%	0.3622	0.3652	0.3646	0.3654	0.3653	0.3656	0.3659
70%	0.4065	0.4056	0.4062	0.4061	0.4062	0.4058	0.4060

中呈現出所提之摘要方法對正確率有明顯的提昇，甚至較其他基礎實驗結果好。

經由一連串的實驗可以證明所提之方法確實對於正確率的提昇有所幫助，尤其是對於低摘要比例的摘要結果的提昇；不過，實驗結果仍有許多進步的空間，例如：由提昇檢索系統正確率及召回率方面著手，檢索系統可採用雙音節與詞的組合作為索引[Chen et al. 2005]，如此，不僅可增加對於辨識結果的強健性(雙音節索引)，亦結合了較具語意的資訊內容(詞索引)。另外，在摘要系統上，亦可考慮以不同索引特徵或是索引特徵的組合[Chen et al. 2006]，來增加摘要正確率。其他方面，像是上面所提及之關於參數的調整，或是進一步將文句本身的重要性納入考量，如對文句事前機率的估測(於後面小節中將有更詳細的說明)等。

4.2.4 以相關文件進行隱藏式馬可夫模型參數訓練之實驗結果

如同上面所述，當文件中每一文句視為一個查詢來進行相關文件檢索時，檢索系

統所回傳的相關文件子集，除了可用來建立關聯性模型作為文句模型的調適外，亦可將這些相關文件子集，視為每一文句所對應的訓練語料，作為訓練文句的隱藏式馬可夫模型參數使用。如第二章中關於隱藏式馬可夫模型的敘述表示，當有訓練文件集(文件及其對應的摘要資訊)時，隱藏式馬可夫摘要模型可以透過監督式方式進行模型的參數訓練，然而這樣的資訊通常取得不易，因此只能採用非監督式訓練來訓練參數。現在，透過局部性回饋技術的使用，可經由檢索的方式得到與文句相關的文件集作為訓練語料，利用文句與其對應相關文件的資訊來進行模型參數 λ 與每一文句 S_i 產生詞 w 的機率值 $P(w|S_i)$ 之監督式訓練：

$$\hat{\lambda} = \frac{\sum_{D_r \in \{R\}} \sum_{w \in D_r} n(w, D_r) P(S_i \mid w, D)}{\sum_{D_r \in \{R\}} \sum_{w \in D_r} n(w, D_r)} \tag{4-16}$$

$$\hat{P}(w \mid S_i) = \frac{n(w, S_i) P(S_i \mid w, D)}{\sum_{w_s \in S_i} n(w_s, S_i) P(S_i \mid w_s, D)} \tag{4-17}$$

$$P(S_i \mid w, D) = \frac{\lambda P(w \mid S_i)}{\lambda P(w \mid S_i) + (1-\lambda) P(w \mid C)} \tag{4-18}$$

其中，$\{R\}$ 為每一文句 S_i 所對應的相關文件子集，D_r 表示相關文件子集中某一文件，$n(w, D_r)$ 表示索引 w 於文件 D_r 中出現的次數，機率值 $P(w|C)$ 為某文件集產生詞 w 的機率。

下面實驗為使用每一文句之相關文件子集來訓練隱藏式馬可夫模型參數，實驗分別嘗試二種參數訓練的方式：1. 僅訓練模型參數 λ 值 (*Training_*λ)，如式(4-16)所示；2. 同時訓練模型參數 λ 值與機率值 $P(w|S_i)$ (*Training_All*)，如式(4-16)與式(4-17)所示。

測試語料 DataSet1 實驗結果：

實驗設定為：採用每一文句的前 10 篇相關文件作為訓練語料(L =10)， Dy =30；隱藏式馬可夫模型參數 λ 初始值為 0.05，訓練次數為 10。表 4.19 及表 4.20 為測

試語料 DataSet1 人工轉寫文件結果，其中表 4.19 為餘弦評估的摘要結果，表 4.20 為 Rouge-2 評估結果，TD-1 與 TD-2 分別表示 DataSet1 中發展集及測試集的實驗結果。基礎實驗(Baseline)為隱藏式馬可夫模型參數λ設定為 0.05，且沒有進行參數訓練的摘要結果。表中欄位 *Training_*λ、*Training_All* 分別表示二種訓練方式的摘要正確率。從人工轉寫文件的實驗結果(表 4.19 及表 4.20)，可以得到以下的結論：首先，與 Baseline 結果比較，所有的結果均顯示出於摘要比例為 10%、20%及 30%時，摘要正確率有明顯的提昇，於摘要比例為 50%、70%時有部分的結果亦有小幅度的上升；接著，比較二種訓練方式的摘要結果，在餘弦評估結果中顯示在同時訓練模型參數λ值與機率值$P(w|S_i)$(*Training_All*)下有較高的正確率，在 Rouge-2 評估結果中於低摘要比例為 10%、20%時有相同的結果。最後，將參數訓練之摘要結果與第三章中其他基礎實驗結果比較，TD-1 於摘要比例 20%、30%時 *Training_All* 結果中有較高的正確率，TD-2 的 *Training_All* 結果中

表 4.19 DataSet1,人工轉寫文件：隱藏式馬可夫模型監督式訓練參數實驗結果 (餘弦評估)

TD-1	*Baseline*	*Training_*λ	*Training_All*	**TD-2**	*Baseline*	*Training_*λ	*Training_All*
10%	0.4203	0.4242	0.4245	*10%*	0.3841	0.3943	0.4016
20%	0.4657	0.4705	0.4738	*20%*	0.4064	0.4204	0.4661
30%	0.5178	0.5288	0.5298	*30%*	0.4855	0.4910	0.5234
50%	0.6271	0.6210	0.6271	*50%*	0.6037	0.6177	0.6202
70%	0.6996	0.6960	0.6956	*70%*	0.6820	0.6763	0.6974

表 4.20 DataSet1,人工轉寫文件：隱藏式馬可夫模型監督式訓練參數實驗結果 (Rouge-2 評估)

TD-1	*Baseline*	*Training_*λ	*Training_All*	**TD-2**	*Baseline*	*Training_*λ	*Training_All*
10%	0.4157	0.4159	0.4194	10%	0.3714	0.3947	0.4160
20%	0.4591	0.4674	0.4731	20%	0.3892	0.4186	0.4434
30%	0.4947	0.5308	0.5249	30%	0.4669	0.4874	0.4909
50%	0.6527	0.6492	0.6603	50%	0.6308	0.6695	0.6557
70%	0.7845	0.7843	0.7831	70%	0.7660	0.7687	0.7527

於不同摘要比例下皆有較高的正確率，由結果可看到正確率最多有 5%～6%的進步。

表 4.21、表 4.22 為自動轉寫文件結果，SD-1 與 SD-2 分別表示 DataSet1 中發展集及測試集自動轉寫人件結果。基礎實驗(Baseline)為隱藏式馬可夫模型參數λ設定為 0.05。相較於 Baseline 的摘要結果，二種評估結果於低摘要比例下正確率皆有所提昇，特別是在 Rouge-2 評估結果下，於低摘要比例 10%及 20%正確率有大幅度的提升。與第三章中其他基礎實驗結果比較，在餘弦評估結果中於低摘要比例 10%、20%時，參數訓練後之摘要結果皆有較高的正確率，在 Rouge-2 評估結果中於摘要比例 10%、20%、30%及 50%下皆有較高的正確率，正確率最多有 5%～6%的進步。

表 4.21 DataSet1,自動轉寫文件：隱藏式馬可夫模型監督式訓練參數實驗結果 (餘弦評估)

SD-1	*Baseline*	*Training_*λ	*Training_All*	**SD-2**	*Baseline*	*Training_*λ	*Training_All*
10%	0.3415	0.3441	0.3437	*10%*	0.3155	0.3309	0.3475
20%	0.3789	0.3806	0.3827	*20%*	0.3381	0.3490	0.3622
30%	0.4281	0.4208	0.4193	*30%*	0.3993	0.3948	0.3983
50%	0.5159	0.5283	0.5256	*50%*	0.5027	0.5092	0.4991
70%	0.5847	0.5859	0.5864	*70%*	0.5601	0.5597	0.5528

表 4.22 DataSet1,自動轉寫文件：隱藏式馬可夫模型監督式訓練參數實驗結果 (Rouge-2 評估)

SD-1	*Baseline*	*Training_*λ	*Training_All*	**SD-2**	*Baseline*	*Training_*λ	*Training_All*
10%	0.3084	0.3528	0.3539	*10%*	0.2932	0.3316	0.3566
20%	0.3467	0.3852	0.3894	*20%*	0.3191	0.3412	0.3642
30%	0.3734	0.3842	0.3830	*30%*	0.3705	0.3739	0.3768
50%	0.4768	0.4952	0.4915	*50%*	0.4732	0.4880	0.4709
70%	0.5631	0.5746	0.5669	*70%*	0.5595	0.5610	0.5452

測試語料 DataSet2 實驗結果：

實驗設定為：採用每一文句的前 10 篇相關文件作為訓練語料(L=10)，Dy=20；隱藏式馬可夫模型參數λ初始值為 0.01，訓練次數為 50。表 4.23 至表 4.26 為測試語料 DataSet2 人工轉寫文件及自動轉寫文件於不同評估方法之結果，其中 TD-1 與 TD-2 分別表示發展集及測試集人工轉寫文件摘要結果，SD-1 與 SD-2 表示發展集及測試集自動轉寫文件摘要結果。基礎實驗(Baseline)為隱藏式馬可夫模型參數λ設定為 0.01，且沒有進行參數訓練的摘要結果。

從發展集人工轉寫結果(表 4.23 與表 4.24 之 TD-1 結果)來看，進行參數訓練後之餘弦評估結果較基礎實驗結果差，而 Rouge-2 評估於訓練參數λ之摘要比例 20%、30%及 50%結果，明顯較基礎實驗結果好。從測試集人工轉寫結果(TD-2)

表 4.23 DataSet2,人工轉寫文件：隱藏式馬可夫模型監督式訓練參數實驗結果 (餘弦評估)

TD-1	*Baseline*	*Training_*λ	*Training_All*	**TD-2**	*Baseline*	*Training_*λ	*Training_All*
10%	0.4427	0.3991	0.4029	*10%*	0.4835	0.4837	0.5050
20%	0.5740	0.5692	0.5513	*20%*	0.6166	0.6089	0.6208
30%	0.6346	0.6332	0.6387	*30%*	0.6579	0.6465	0.6481
50%	0.7376	0.7348	0.7331	*50%*	0.7338	0.7359	0.7375
70%	0.7985	0.7958	0.7929	*70%*	0.7873	0.7902	0.7899

表 4.24 DataSet2,人工轉寫文件：隱藏式馬可夫模型監督式訓練參數實驗結果 (Rouge-2 評估)

TD-1	*Baseline*	*Training_*λ	*Training_All*	**TD-2**	*Baseline*	*Training_*λ	*Training_All*
10%	0.2839	0.2417	0.2429	10%	0.3280	0.3153	0.3648
20%	0.3995	0.4205	0.3981	20%	0.4499	0.4732	0.4800
30%	0.4790	0.5026	0.5065	30%	05012	0.5214	0.5105
50%	0.6539	0.6813	0.6747	50%	0.6534	0.6862	0.6787
70%	0.8019	0.8129	0.8047	70%	0.7915	0.8203	0.8093

來看，不論於何種評估方法下，參數訓練(*Training_All*)結果於低摘要比例 10%及20%時皆較基礎實驗結果好，特別是在 Rouge-2 評估下正確率有很明顯的提昇，並且亦較其他摘要方法之結果(與第三章基礎實驗結果比較)好許多。

表 4.25 與表 4.26 為自動轉寫文件參數訓練之結果。在餘弦評估結果中，發展集(SD-1)與測試集(SD-2)自動轉寫文件的參數訓練結果並沒有較基礎實驗好；但是，但是， Rouge-2 評估結果中不同摘要比例之摘要結果皆較基礎實驗結果好(除 SD-1 摘要比例 10%之結果外)，並且於部分摘要比例下正確率亦較第三章基礎實驗中其他摘要方法好一些。

從 DataSet2 的結果來看，餘弦評估結果並沒有顯示出模型參數訓練有較好的摘要結果，但是由 Rouge-2 評估結果可明顯地觀察出經模型參數訓練後，可以達到更高的摘要正確率，甚至可較其他摘要方法之結果來得好。雖然，在參數訓練的實驗中，DataSet2 的實驗結果並不如 DataSet1 對於正確率有明顯的提昇，甚至於 DataSet2 在某些部份的摘要結果上會有變差的情形，這可能是因二套測試集中文件內容的差異使得實驗設定需再作調整，或是訓練語料與文句之間不匹配的問題。但是，我們仍然可以從實驗結果中看到參數訓練對於摘要正確率的影響；同時，也可以從結果中發現機率生成式摘要模型仍有許多可以改進及探討的空間

表 4.25 DataSet2,自動轉寫文件：隱藏式馬可夫模型監督式訓練參數實驗結果 (餘弦評估)

SD-1	*Baseline*	*Training_*λ	*Training_All*	**SD-2**	*Baseline*	*Training_*λ	*Training_All*
10%	0.4128	0.3865	0.3849	*10%*	0.3837	0.3698	0.3745
20%	0.5156	0.4998	0.5054	*20%*	0.4739	0.4623	0.4652
30%	0.5623	0.5501	0.5548	*30%*	0.5110	0.4995	0.5052
50%	0.6174	0.6149	0.6156	*50%*	0.5619	0.5573	0.5547
70%	0.6339	0.6322	0.6347	*70%*	0.5837	0.5811	0.5734

表 4.26 DataSet2,自動轉寫文件：隱藏式馬可夫模型監督式訓練參數實驗結果 (Rouge-2 評估)

SD-1	*Baseline*	*Training_* λ	*Training_All*	**SD-2**	*Baseline*	*Training_* λ	*Training_All*
10%	0.2141	0.2089	0.2011	*10%*	0.2008	0.2142	0.2086
20%	0.2747	0.2767	0.2886	*20%*	0.2501	0.2639	0.2581
30%	0.3156	0.3248	0.3219	*30%*	0.2820	0.2936	0.2953
50%	0.3959	0.4021	0.3955	*50%*	0.3622	0.3657	0.3609
70%	0.4315	0.4351	0.4344	*70%*	0.4065	0.4063	0.3963

4.3 詞層次主題混合模型(word Topical Mixture Model , wTMM)

本論文中提出另一種概念比對方式的機率式摘要模型－詞層次主題混合模型(wTMM)，我們可以將語言中每個詞 w_j 視為包含 K 潛藏主題的機率模型 M_{w_j} ，用來預測另一個詞 w 發生的機率：

$$P(w|M_{w_j}) = \sum_{k=1}^{K} P(w|T_k)P(T_k|M_{w_j}) \tag{4-19}$$

$P(w|M_{w_j})$ 表示詞 w_j 所對應的詞模型 M_{w_j} 生成詞 w 的機率，其中 $P(w|T_k)$ 為單連語言模型，表示詞 w 於潛藏主題 T_k 中的機率分佈， $P(T_k|M_{w_j})$ 為詞 w_j 的詞模型 M_{w_j} 對於潛藏主題 T_k 的權重機率值。也就是說透過詞於這些不同潛藏主題下的分佈，及詞與不同潛藏主題之間的關係，來建立詞與詞之間的關聯性。如圖 4.2 所示。機率值 $P(w|T_k)$ 及 $P(T_k|M_{w_j})$ 可經由期望值最大化訓練的方式來估測。

詞層次主題混合模型透過 K 個潛藏主題的資訊來表示詞生成另一個詞的機率 $P(w|M_{w_j})$ ；其中機率值 $P(w|T_k)$ 及 $P(T_k|M_{w_j})$ 可採用監督式與非監督式訓練來估測。若可取得包含有文字與文字之間關聯性的訓練語料，即可使用監督式訓練來訓練機率值 $P(w|T_k)$ 及 $P(T_k|M_{w_j})$ ，例如我們可以從網路上得到許多文字新

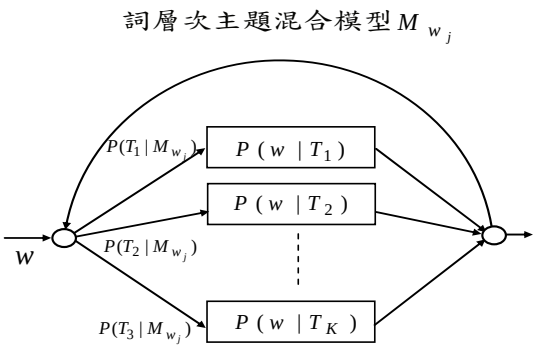


圖 4.2 詞層次主題混合模型。

聞集$\{D_C\}$，通常這些新聞文件都會有一句人工產生的標題，於是可以將文件集中每篇文件D_C與其標題H_C的對應關係$\{D_C, H_C\}$，視為標題H_C中每一個詞w_j的主題混合模型M_{w_j}與對應的文件D_C中每一個詞w的關聯性，以訓練所有詞層次主題混合模型M_{w_j}的機率分佈$P(w|T_k)$及$P(T_k|M_{w_j})$：

$$\hat{P}(w|T_k) = \frac{\sum_{D_C} n(w, D_C) P(T_k \mid w, H_C)}{\sum_{D_C'} \sum_{w_n \in D_C'} n(w_n, D_C') P(T_k \mid w_n, H_C')} \tag{4-20}$$

$$\hat{P}(T_k \mid M_{w_j}) = \frac{\sum_{D_C} \sum_{w \in D_C} n(w, D_C) P(M_{w_j} \mid w, M_{H_C}) P(T_k \mid w, M_{w_j})}{\sum_{D_C} \sum_{w' \in D_C} n(w', D_C) P(M_{w_j} \mid w', M_{H_C})} \tag{4-21}$$

其中，$P(T_k \mid w, H_C)$的計算方式為：

$$P(T_k \mid w, M_{H_C}) = \frac{P(w|T_k)\left[\sum_{w_j \in H_C} \alpha_{j,C} P(T_k \mid M_{w_j})\right]}{\sum_{l=1}^{K} P(w|T_l)\left[\sum_{w_j \in H_C} \alpha_{j,C} P(T_l \mid M_{w_j})\right]} \tag{4-22}$$

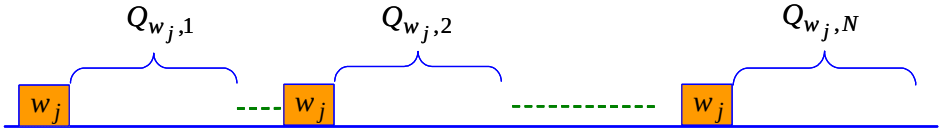
$P(T_k \mid w, M_{w_j})$可表示為:

$$P(T_k \mid w, M_{w_j}) = \frac{P(w \mid T_k) P(T_k \mid M_{w_j})}{\sum_{l=1}^{K} P(w \mid T_l) P(T_l \mid M_{w_j})} \tag{4-23}$$

$P(M_{w_j} \mid w, M_{H_C})$ 可表示為:

$$P(M_{w_j} \mid w, M_{H_C}) = \frac{\alpha_{j,C} P(w \mid M_{w_j})}{\sum_{w_l \in H_C} \alpha_{l,C} P(w \mid M_{w_l})} \tag{4-24}$$

如此，便可以透過任何的文字新聞訓練集來訓練所有詞 w_j 的主題混合模型。此外，我們亦可採用非監督式訓練來訓練機率值 $P(w \mid T_k)$ 及 $P(T_k \mid M_{w_j})$，如[Chiu and Chen 2007]中所提之方法，對於某文件中每一個詞 w_j 而言，我們可以簡單地假設在以其為中心的後前 N 個詞都是與其相關的；因此，透過收集文件集中每一個



$$Q_{w_j} = Q_{w_j,1}, Q_{w_j,2}, \cdots, Q_{w_j,N}$$

圖 4.3 詞與其鄰近詞的關係圖

詞 w_j 與其鄰近詞 Q_{w_j} 的關係作為訓練語料(如圖 4.3 所示)，訓練詞層次主題混合模型 M_{w_j} 的機率分佈 $P(w \mid T_k)$ 及 $P(T_k \mid M_{w_j})$：

$$P(w \mid T_k) = \frac{\sum_{w_j} n(w, Q_{w_j}) P(T_k \mid w, M_{w_j})}{\sum_{w_l} \sum_{w' \in Q_{w_l}} n(w', Q_{w_l}) P(T_k \mid w', M_{w_l})} \tag{4-25}$$

$$P(T_k \mid M_{w_j}) = \frac{\sum_{w_s \in Q_{w_j}} n(w_s, Q_{w_j}) P(T_k \mid w_s, M_{w_j})}{\sum_{w_l \in Q_{w_j}} n(w_l, Q_{w_j})} \tag{4-26}$$

$$P\left(T_k \mid w, M_{w_j}\right) = \frac{P(w \mid T_k) P\left(T_k \mid M_{w_j}\right)}{\sum_{l=1}^{K} P(w \mid T_l) P\left(T_l \mid M_{w_j}\right)} \tag{4-27}$$

其中， Q_{w_j} 為詞 w_j 所對應的鄰近詞。

當給定一篇欲摘要文件時，文件中每一文句 S_i 可表示成句中一連串詞主題混合模型的組合：

$$Model(S_i) = \sum_{w_j \in S_i} \alpha_{j,i} M_{w_j} \tag{4-28}$$

其中，每一個詞主題混合模型 M_{w_j} 於文句 S_i 中，都有一個權重值 $\alpha_{j,i}$ ，用以表示詞 w_j 於文句 S_i 中的重要性與貢獻程度。 $\alpha_{j,i}$ 最簡單的估測方式，可以根據 w_j 於文句 S_i 出現的頻率來決定，同時 $\alpha_{j,i}$ 須滿足 $\sum_{w_j \in S_i} \alpha_{j,i} = 1$ ，。而每一文句 S_i 生成文件 D 的可能性即可表示為：

$$P\left(D \middle| S_i\right) = \prod_{w \in D} \left[\sum_{w_j \in S_i} \alpha_{j,i} \sum_{k=1}^{K} P\left(w \middle| T_k\right) P\left(T_k \middle| M_{w_j}\right) \right]^{n(w,D)} \tag{4-29}$$

其中，機率值 $P(w \mid T_k)$ 及 $P(T_k \mid M_{w_j})$ 可直接使用文件與其對應之摘要內容的訓練語料，或使用與摘要文件同一時期或同一領域的文字新聞集 $\{D_C\}$ 與其標題 H_C 的對應關係進行監督式訓練，如式(4-20)至式(4-24)所示，亦或是採用非監督式訓練機率值，如式(4-25)至式(4-27)所示。圖4.4為詞層次主題混合模型的概念圖，圖中左邊部分表示以與摘要文件同一時期或同一領域的文字新聞集 $\{D_C\}$ 所訓練出來機率值 $P(w \mid T_k)$ 及 $P(T_k \mid M_{w_j})$ ，可直接被使用於詞層次主題混合摘要模型中，進行文件摘要處理。

詞層次主題混合模型與主題混合模型[Chen et al. 2006; 黃耀民 2005; 陳怡婷 et al. 2005]皆屬於概念比對方式(Concept Matching)的摘要方法。前者視每一個詞為一個主題混合模型，然後將文句視為一個詞層次主題混合模型的組合；後者

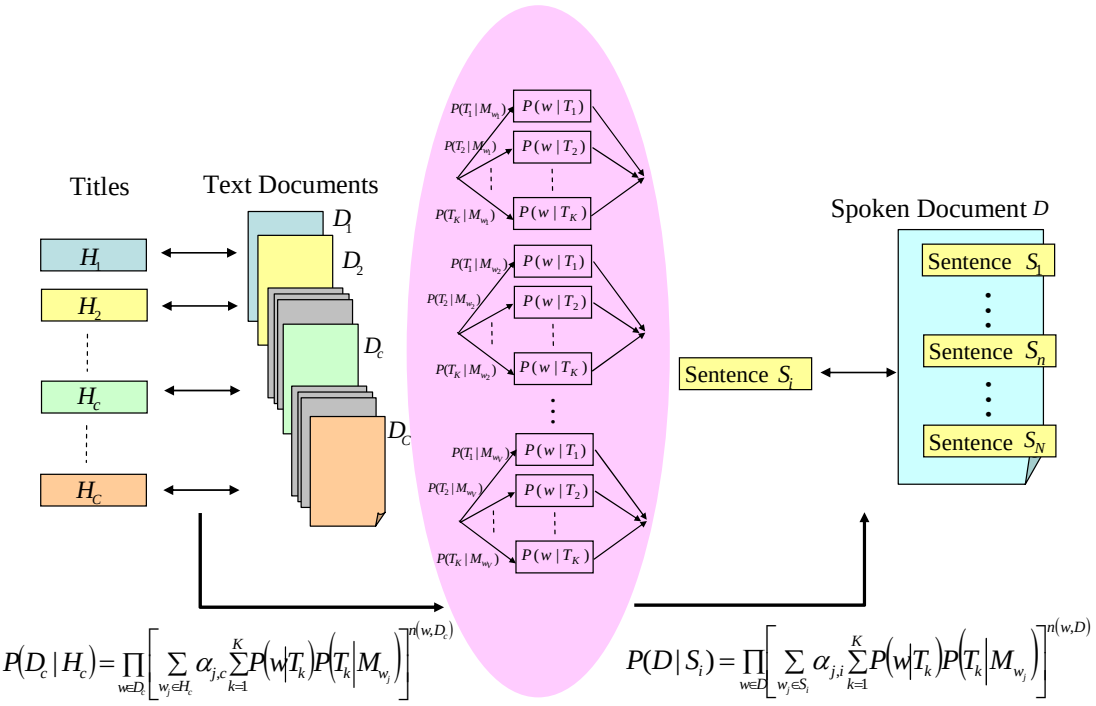


圖 4.4 詞層次主題混合摘要模型概念圖。

則以文件中每一文句 S_i 為一個主題混合模型，二者皆可採用監督式與非監督式訓練來估測機率值。但是，在主題混合模型中，雖然機率值 $P(w|T_k)$，可以直接使用同一時期的訓練文件集所訓練的機率值，如式(2-29)，不過對於機率值 $P(T_k|S_i)$ 可能無法取得相關的訓練資料可供訓練，這或許可使用4.2小節所提到之每一文句的相關文件集來訓練機率值 $P(T_k|S_i)$。然而，關於詞層次主題混合模型的機率值 $P(w|T_k)$ 及 $P(T_k|M_{w_j})$，我們可以較容易取得能夠表示詞與詞之間關聯性的文件集，如[Chiu and Chen 2007]所提的方式，或是由網路取得相關語料等。

4.3.1 詞層次主題混合模型實驗結果

使用詞層次主題混合模型於文件摘要時，我們同時考慮文件中每一個詞 w 於文句 S_i 的機率分佈 $P(w|S_i)$，並將其與式(4-29)結合為：

$$P(D|S_i)=\prod_{w\in D}\left[\beta\cdot\left(\sum_{w_j\in S_i}\alpha_{j,i}\sum_{k=1}^{K}P(w|T_k)P(T_k|M_{w_j})\right)+(1-\beta)\cdot P(w|S_i)\right]^{n(w,D)} \quad (4\text{-}30)$$

其中，β 是比重參數。在本實驗中，我們觀察詞層次主題混合模型於摘錄式文件摘要應用上的摘要結果；我們分別於二套測試語料中進行摘要實驗，經由測試語料中的發展集來調整不同主題數 K 的 β 值，嘗試找出最佳的參數設定；然後，使用於測試集。對於詞層次主題混合模型中機率值 $P(w|T_k)$ 及 $P(T_k|M_{w_j})$ 的訓練，我們使用發展集中每一文件與其所對應的其中一份參照摘要中 30%摘要比例的摘要結果進行監督式訓練，以及[Chiu and Chen 2007]所提之方法進行非監督式訓練，觀察不同機率值訓練方式下詞層次主題混合模型的摘要結果。

測試語料 DataSet1 實驗結果：

- 監督式訓練方式：

我們使用 DataSet1 發展集中每一篇文件與其三份參照摘要中某一份參照摘要的 30%摘要比例的摘要結果進行監督式訓練。訓練公式可參考式式(4-20)至式(4-24)所示，吾人將每一篇文件與其所對應的摘要內容，視為摘要內容中每一個詞 w_j 的主題混合模型 M_{w_j} 與其對應的文件中每一個詞 w 的關聯性，來訓練機率值 $P(w|T_k)$ 及 $P(T_k|M_{w_j})$。

表 4.27 為發展集人工轉寫文件於不同評估方法及不同主題數下之摘要結果，表 4.28 為發展集自動轉寫文件於不同評估方法及不同主題數下之摘要結果，

其中β值嘗試使用 0.89,0.90,0.91,…,1.0 等值，並且於表 4.27 及表 4.28 中呈現在不同主題數下最佳β值設定的結果，在不同主題數及不同評估方法下β值的設定皆不同。由於詞層次主題混合模型是採用發展集中的文件進行訓練，因此表 4.27 及表 4.28 之摘要結果皆屬內部測試(Inside Test)結果，於不同摘要比例時都有極高的摘要正確率。比較餘弦評估與 Rouge-2 的摘要結果，發現不論於人工轉寫摘要或自動轉寫摘要上，Rouge-2 評估下之摘要正確率皆高。

接著，將所訓練之詞層次主題混合模型使用於測試集，並且於不同主題數下β值的設定與發展集(表 4.27 及表 4.28)相同，進行文件摘要實驗。實驗結果如表 4.29 及 4.30 所示。表 4.29 為人工轉寫文件摘要結果，可以看出在低摘要比例 10%與 20%下，餘弦評估結果中主題數為 4 時有最好的結果，Rouge-2 評估結果中當主題數為 4 或 8 時有最好的結果。若與其他方法比較，當主題數為 4 時於二種評估方法下摘要比例 10%、20%及 30%之正確率皆較所有基礎實驗方法結果好；與表 4.9 中關聯性模型於隱藏式馬可夫模型調適後結果比較，於餘弦評估結果中摘要比例 10%、20%及 30%時，詞層次主題混合模型之正確率略高一些，於 Rouge-2

表 4.27 DataSet1,發展集：監督式訓練之詞層次主題混合模型實驗結果 (人工轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.4542	0.4599	0.4758	0.4673	0.4701	0.4419
20%	0.5065	0.5226	0.5326	0.5374	0.5411	0.5251
30%	0.5552	0.5718	0.5922	0.6139	0.6248	0.6258
50%	0.6419	0.6425	0.6459	0.6550	0.6499	0.6535
70%	0.7083	0.7084	0.7015	0.7085	0.7013	0.7029
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.4586	0.4834	0.5043	0.5075	0.5247	0.4611
20%	0.5147	0.5643	0.5753	0.5867	0.5951	0.5607
30%	0.5464	0.6041	0.6469	0.6738	0.6926	0.6922
50%	0.6540	0.6706	0.6637	0.6690	0.6679	0.6692
70%	0.7684	0.7790	0.7427	0.7616	0.7491	0.7587

表 4.28 DataSet1,發展集：監督式訓練之詞層次主題混合模型實驗結果（自動轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3697	0.3802	0.4005	0.3919	0.3949	0.3930
20%	0.4021	0.4243	0.4476	0.4551	0.4472	0.4611
30%	0.4461	0.4698	0.4980	0.5080	0.5064	0.5128
50%	0.5258	0.5395	0.5494	0.5520	0.5493	0.5466
70%	0.5846	0.5861	0.5874	0.5927	0.5888	0.5881
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3662	0.3850	0.4190	0.4114	0.4219	0.4128
20%	0.4070	0.4301	0.4628	0.4791	0.4758	0.4830
30%	0.3963	0.4249	0.4683	0.4773	0.4825	0.4891
50%	0.4817	0.4810	0.4874	0.4944	0.4918	0.4896
70%	0.5603	0.5408	0.5369	0.5546	0.5455	0.5471

評估結果中摘要比例 20%及 30%時亦有些微較好的結果。

表 4.30 為自動轉寫文件摘要結果，當主題數為 8 時，二種評估方式於低摘要比例 10%有最佳的摘要結果。比較基礎實驗中摘要方法與詞層次主題混合模型之摘要結果，於餘弦評估結果中，當主題數為 4 時低摘要比例 10%及 20%之正確率較基礎實驗中機率生成式模型 HMM 與 TMM 結果好，但是較部分的非機率生成式模型摘要方法的結果來得差一點；於 Rouge-2 評估結果中，當主題數為 8 時低摘要比例 10%之摘要結果亦較其他基礎實驗好(除了較 eLSA 有極些微的差距)。若與表 4.10(關聯性模型於隱藏式馬可夫模型調適後結果)中最好的結果比較，可以看出於二種評估方法下低摘要比例 10%之結果均較關聯性模型的調適結果好，於餘弦評估結果中摘要比例 20%及 30%之結果亦較表 4.10 結果來得好；但是，於 Rouge-2 評估結果摘要比例 20%、30%、50%及 70%下皆較關聯性模型調適的摘要結果差。

表 4.29 DataSet1,測試集：監督式訓練之詞層次主題混合模型實驗結果 (人工轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3998	0.4049	0.3857	0.3889	0.3620	0.3714
20%	0.4205	0.4232	0.4119	0.4048	0.3950	0.4002
30%	0.4785	0.4996	0.4926	0.4903	0.4702	0.4731
50%	0.6114	0.5999	0.5977	0.6009	0.6156	0.6094
70%	0.6815	0.6782	0.6802	0.6803	0.6788	0.6775
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3920	0.4014	0.4096	0.3811	0.3457	0.3599
20%	0.3977	0.4169	0.4167	0.3849	0.3791	0.3855
30%	0.4472	0.4843	0.4554	0.4723	0.4477	0.4571
50%	0.6130	0.6208	0.6222	0.6235	0.6446	0.6414
70%	0.7393	0.7531	0.7395	0.7589	0.7579	0.7640

表 4.30DataSet1,測試集：監督式訓練之詞層次主題混合模型實驗結果 (自動轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3204	0.3219	0.3268	0.3205	0.3143	0.3028
20%	0.3394	0.3431	0.3423	0.3466	0.3336	0.3238
30%	0.4042	0.4071	0.4076	0.4006	0.3961	0.3903
50%	0.5045	0.4984	0.4988	0.4979	0.4987	0.4936
70%	0.5603	0.5606	0.5596	0.5637	0.5612	0.5585
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3094	0.3058	0.3222	0.3028	0.2989	0.2822
20%	0.3244	0.3149	0.3206	0.3161	0.3157	0.3007
30%	0.3753	0.3663	0.3681	0.3675	0.3598	0.3511
50%	0.4716	0.4508	0.4502	0.4643	0.4675	0.4610
70%	0.5579	0.5398	0.5380	0.5574	0.5616	0.5611

➢ 非監督式訓練方式：

表 4.31 至表 4.34 為非監督式訓練方式之發展集及測試集的摘要結果。詞層次主題混合模型的訓練方式如[Chiu and Chen 2007]所提，我們使用 CNA 西元 2002 年 8 月至 10 月的文字新聞語料中每一個詞與其後一個鄰近詞的關係來訓練機率值。表 4.31 與表 4.32 為發展集中人工轉寫文件及自動轉寫文件的實驗結果，其中 β 值的設定，我們嘗試了 0.89,0.90,…,1.0 之間的數值，並於不同主題數下找出最佳的 β 值設定，相同的設定使用於測試集中(如表 4.33 及表 4.34)。

表 4.31 為不同主題數下人工轉寫文件摘要結果。由結果來看，當主題數為 32 時二種評估方法摘要結果中皆顯示於低摘要比例下有最高的正確率。與基礎實驗中非機率生成式摘要方法比較，二種評估方式之結果顯示詞層次主題混合模型於低摘要比例 10%、20%時有較好的結果；與機率生成式摘要方法比較，二種評估方法於低摘要比例 10%時，詞層次主題混合模型之摘要結果較基礎實驗中 HMM 及 TMM 好，同時亦較表 4.7 中關聯性模型於隱藏式馬可夫模型調適後結果來得好。表 4.32 為不同主題數下自動轉寫文件摘要結果。同樣地，我們可以

表 4.31 DataSet1,發展集：非監督式訓練之詞層次主題混合模型實驗結果 (人工轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.4262	0.4216	0.4221	0.4235	0.4307	0.4279
20%	0.4718	0.4689	0.4721	0.4645	0.4701	0.4684
30%	0.5215	0.5194	0.5198	0.5171	0.5218	0.5210
50%	0.6220	0.6224	0.6211	0.6233	0.6206	0.6268
70%	0.7002	0.6957	0.7010	0.6937	0.6959	0.6945
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.4242	0.4102	0.4114	0.4114	0.4268	0.4268
20%	0.4683	0.4659	0.4691	0.4587	0.4727	0.4681
30%	0.5054	0.5009	0.5000	0.4982	0.5067	0.5072
50%	0.6502	0.6446	0.6468	0.6382	0.6494	0.6555
70%	0.7842	0.7718	0.7788	0.7660	0.7786	0.7756

表 4.32 DataSet1,發展集：非監督式訓練之詞層次主題混合模型實驗結果 (自動轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3486	0.3502	0.3563	0.3494	0.3605	0.3587
20%	0.3830	0.3839	0.3854	0.3863	0.3883	0.3884
30%	0.4222	0.4215	0.4197	0.4253	0.4235	0.4207
50%	0.5137	0.5145	0.5161	0.5165	0.5182	0.5130
70%	0.5839	0.5817	0.5836	0.5823	0.5833	0.5844
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3245	0.3334	0.3428	0.3302	0.3459	0.3493
20%	0.3522	0.3620	0.3651	0.3611	0.3634	0.3698
30%	0.3509	0.3593	0.3540	0.3684	0.3595	0.3646
50%	0.4577	0.4568	0.4599	0.4741	0.4680	0.4633
70%	0.5486	0.5479	0.5497	0.5598	0.5486	0.5534

發現在餘弦評估結果中，當主題數為 32 時於低摘要比例 10%有最好的摘要結果；在 Rouge-2 評估結果中，主題數為 64 於低摘要比例 10%及 20%有最高的正確率。與基礎實驗中所有摘要方法比較，詞層次主題混合模型於二種評估方式下，低摘要比例 10%、20%的摘要結果皆較其他摘要方法好；同時，於低摘要比例 10%時正確率亦較關聯性模型的調適結果(表 4.8)略高一些。

表 4.33 及表 4.34 為測試集中人工轉寫文件與自動轉寫文件摘要結果。從人工轉寫文件結果(表 4.33)中找出一組正確率最高的摘要結果，發現於二種評估方法下主題數為 4 有最好的結果；同時，於低摘要比例 10%、20%及 30%時皆較基礎實驗中其他摘要方法的摘要結果好。進一步與關聯性模型的調適結果(表 4.9)比較，發現關聯性模型的調適結果於低摘要比例 10%下有較高的正確率。在自動轉寫文件結果表(4.34)中，當主題數為 2 時，二種評估方法的摘要結果皆有較高的正確率；雖然從自動轉寫文件的結果來看，詞層次主題混合模型於餘弦評估結果中，正確率並沒有較基礎實驗中非機率生成式摘要方法高，但是於 Rouge-2 評

表 4.33 DataSet1,測試集：非監督式訓練之詞層次主題混合模型實驗結果 (人工轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3805	0.3973	0.3922	0.3800	0.3754	0.3818
20%	0.4093	0.4244	0.4172	0.4192	0.4032	0.4108
30%	0.4893	0.4933	0.4913	0.4893	0.4831	0.4910
50%	0.5994	0.6012	0.6037	0.6104	0.6053	0.6051
70%	0.6817	0.6828	0.6797	0.6842	0.6765	0.6791
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3674	0.3926	0.3837	0.3699	0.3589	0.3685
20%	0.3960	0.4209	0.4102	0.4130	0.3868	0.3991
30%	0.4747	0.4767	0.4761	0.4708	0.4627	0.4767
50%	0.6272	0.6251	0.6328	0.6351	0.6366	0.6361
70%	0.7638	0.7641	0.7596	0.7597	0.7631	0.7605

表 4.34 DataSet1,測試集：非監督式訓練之詞層次主題混合模型實驗結果 (自動轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3307	0.3247	0.3256	0.3113	0.3255	0.3210
20%	0.3492	0.3429	0.3435	0.3372	0.3444	0.3357
30%	0.4146	0.4057	0.4049	0.4066	0.4058	0.4045
50%	0.5027	0.4963	0.5001	0.5026	0.4997	0.5013
70%	0.5597	0.5591	0.5560	0.5614	0.5578	0.5610
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3248	0.3227	0.3235	0.2939	0.3139	0.3113
20%	0.3324	0.3286	0.3324	0.3200	0.3276	0.3176
30%	0.3816	0.3696	0.3657	0.3752	0.3670	0.3704
50%	0.4581	0.4514	0.4541	0.4673	0.4549	0.4590
70%	0.5424	0.5437	0.5372	0.5571	0.5379	0.5542

估結果中於摘要比例 10%、30%下有較高的正確率。與關聯性模型的調適結果(表 4.10)比較，詞層次主題混合模型於摘要比例 10%、20%及 30%下皆有較高的正確率。

比較監督式訓練及非監督式訓練之摘要結果。在人工轉寫文件摘要結果，雖然非監督式訓練中最高的摘要正確率不及監督式訓練的摘要結果，但是在部分主題數下非監督式訓練的摘要正確率較監督式訓練結果好；而於自動轉寫文件摘要結果中，非監督式訓練的摘要正確率較監督式訓練還要好。因此，從實驗結果發現若可以取得任何描述關於詞與詞之間關係的語料，即可透過非監督式訓練來訓練詞層次主題混合模型，並且得到不錯的結果。

測試語料 DataSet2 實驗結果：

- 監督式訓練方式：

我們同樣使用 DataSet2 發展集中每一篇文件與其三份參照摘要中某一份參照摘要的 30%摘要比例的摘要結果進行監督式訓練。表 4.35 與表 4.36 為發展集人工轉寫文件及自動轉寫文件的摘要結果，其中β值亦嘗試了 0.89,0.90,0.91,…,1.0 等設定值，於表 4.35 及表 4.36 中呈現出不同主題數下最佳β值設定的結果，並且相同的設定使用於測試集中。

由於表 4.35 及表 4.36 屬於內部測試(Inside Test)的結果，因此於不同評估方法及不同摘要比例下皆有很高的正確率，並且較基礎實驗結果好；同時，似乎隨著主題數變大，正確率亦有遞增的情形。將監督式訓練後之詞層次混合模型使用於測試集的文件摘要中，結果如表 4.37 及表 4.38 所示。在測試集人工轉寫文件結果中(表 4.37)，低摘要比例 10%下餘弦評估結果於主題數 32 有最高正確率，Rouge-2 評估結果於主題數 64 有最高正確率；比較餘弦評估結果中主題數為 2 時之摘要結果與基礎實驗中所有摘要方法結果比較，詞層次主題混合模型於不同摘要比例下皆有較高的正確率，於 Rouge-2 評估結果中主題數為 16 時，詞層次主題混合模型於低摘要比例 10%、20%及 30%下亦有較高的正確率。

表 4.35 DataSet2,發展集：監督式訓練之詞層次主題混合模型實驗結果 (人工轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.4685	0.4894	0.4910	0.4894	0.4905	0.4902
20%	0.6157	0.6376	0.6545	0.6660	0.6626	0.6635
30%	0.6836	0.7101	0.7347	0.7370	0.7450	0.7519
50%	0.7597	0.7735	0.7812	0.7856	0.7837	0.7858
70%	0.8117	0.8124	0.8142	0.8156	0.8173	0.8147
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3124	0.3381	0.3623	0.3658	0.3763	0.3720
20%	0.4552	0.4903	0.5138	0.5467	0.5507	0.5644
30%	0.5446	0.5788	0.6272	0.6407	0.6619	0.6601
50%	0.6719	0.6786	0.6851	0.7023	0.6935	0.6973
70%	0.8062	0.7952	0.7917	0.8002	0.7992	0.7966

表 4.36 DataSet2,發展集：監督式訓練之詞層次主題混合模型實驗結果（自動轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.4211	0.4304	0.4388	0.4456	0.4500	0.4430
20%	0.5292	0.5421	0.5510	0.5609	0.5658	0.5634
30%	0.5734	0.5887	0.6015	0.6159	0.6173	0.6193
50%	0.6265	0.6349	0.6429	0.6476	0.6499	0.6537
70%	0.6401	0.6434	0.6492	0.6529	0.6545	0.6541
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.2345	0.2463	0.2454	0.2479	0.2682	0.2640
20%	0.2844	0.3023	0.3316	0.3380	0.3538	0.3522
30%	0.3335	0.3510	0.3661	0.3924	0.3975	0.4025
50%	0.4101	0.4155	0.4223	0.4201	0.4265	0.4279
70%	0.4403	0.4449	0.4477	0.4454	0.4497	0.4508

表 4.37 DataSet2,測試集：監督式訓練之詞層次主題混合模型實驗結果 (人工轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.5092	0.5077	0.4982	0.4968	0.5097	0.5082
20%	0.6287	0.6306	0.6208	0.6228	0.6250	0.6213
30%	0.6670	0.6682	0.6653	0.6617	0.6563	0.6596
50%	0.7416	0.7423	0.7384	0.7367	0.7394	0.7378
70%	0.7909	0.7901	0.7886	0.7872	0.7870	0.7895
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3534	0.3481	0.3312	0.3589	0.3531	0.3598
20%	0.4583	0.4613	0.4555	0.4832	0.4884	0.4764
30%	0.5025	0.5068	0.4996	0.5173	0.5106	0.5146
50%	0.6542	0.6562	0.6535	0.6666	0.6695	0.6759
70%	0.7910	0.7884	0.7886	0.8038	0.8067	0.8121

表 4.38 DataSet2,測試集：監督式訓練之詞層次主題混合模型實驗結果 (自動轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3886	0.3917	0.3910	0.3869	0.3835	0.3814
20%	0.4696	0.4733	0.4800	0.4736	0.4762	0.4710
30%	0.5096	0.5078	0.5111	0.5069	0.5142	0.5107
50%	0.5634	0.5638	0.5668	0.5643	0.5653	0.5662
70%	0.5847	0.5855	0.5882	0.5857	0.5876	0.5885
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.2160	0.2184	0.2171	0.2045	0.2207	0.2184
20%	0.2552	0.2609	0.2574	0.2466	0.2658	0.2678
30%	0.2883	0.2937	0.2854	0.2821	0.2991	0.3024
50%	0.3682	0.3695	0.3682	0.3651	0.3716	0.3714
70%	0.4102	0.4124	0.4121	0.4094	0.4151	0.4165

從測試集自動轉寫文件結果中(表 4.38)，可以看出於低摘要比例 10%下餘弦評估於主題數 4 時有最高正確率，Rouge-2 評估於主題數 32 時；與基礎實驗中所有摘要方法結果比較，詞層次主題混合模型於 Rouge-2 評估方法下不同摘要比例之結果皆有較高的正確率。

最後，比較詞層次主題混合模型(表 4.37 與表 4.38)與關聯性模型調適(表 4.17 與表 4.18)的摘要結果，詞層次主題混合模型於 Rouge-2 評估下之正確率有明顯較高的結果。這可能是因關聯性模型於隱藏式馬可夫模型的調適上有較多影響摘要結果的因素及較多參數需要調整，因此關聯性模型調適的摘要結果較詞層次主題混合模型來得略差一些。

- 非監督式訓練方式：

我們同樣使用經 CNA 西元 2002 年 8 月至 10 月的文字新聞語料所訓練的詞層次主題混合模型於測試語料 DataSet2，即是採用與 DataSet1 非監督式訓練方式實驗中相同的詞層次主題混合模型。表 4.39 至表 4.42 為 DataSet2 之非監督式訓練方式的實驗結果，表 4.39 與表 4.40 為發展集人工轉寫文件及自動轉寫文件的摘要結果，表中呈現出於不同主題數下最佳的β值設定，相同的設定亦使用於測試集中(表 4.41 與表 4.42)。從表 4.39 的摘要結果中，可以看出於二種評估方法下主題數為 64 時有最高的正確率，與基礎實驗中其他摘要方法的結果比較，非監督式訓練下之詞層次主題混合模型的摘要正確率很明顯較低，特別是餘弦評估方法下之結果。表 4.40 為發展集自動轉寫文件摘要結果，當主題數為 2 時二種評估結果之低摘要比例 10%有最高的正確率，若與基礎實驗中其他摘要方法的結果比較，非監督式訓練之詞層次主題混合模型於主題數為 16 的 Rouge-2 評估結果下有較好的摘要結果。

表 4.41 與表 4.42 為測試集人工轉寫文件及自動轉寫文件摘要結果。比較非監督式訓練與監督式訓練之摘要結果，發現人工轉寫文件於監督式訓練下有明顯較高的摘要正確率，而且非監督式訓練下人工轉寫文件正確率(表 4.41)亦較基礎

表 4.39 DataSet2,發展集：非監督式訓練之詞層次主題混合模型實驗結果 (人工轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.4326	0.4364	0.4332	0.4354	0.4308	0.4442
20%	0.5672	0.5598	0.5660	0.5579	0.5620	0.5691
30%	0.6351	0.6324	0.6291	0.6291	0.6318	0.6365
50%	0.7363	0.7385	0.7392	0.7391	0.7381	0.7368
70%	0.8014	0.7988	0.7976	0.7987	0.7989	0.8007
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.2842	0.2865	0.2850	0.2872	0.2790	0.2979
20%	0.3962	0.3913	0.3983	0.3888	0.3885	0.3899
30%	0.4868	0.4849	0.4846	0.4823	0.4963	0.4813
50%	0.6587	0.6638	0.6702	0.6660	0.6707	0.6465
70%	0.8084	0.8076	0.8073	0.8034	0.8059	0.7927

表 4.40 DataSet2,發展集：非監督式訓練之詞層次主題混合模型實驗結果 (自動轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.4206	0.4154	0.4133	0.4102	0.4061	0.4084
20%	0.5128	0.5136	0.5158	0.5100	0.5070	0.5084
30%	0.5611	0.5579	0.5606	0.5541	0.5544	0.5592
50%	0.6207	0.6196	0.6205	0.6211	0.6195	0.6206
70%	0.6316	0.6337	0.6328	0.6361	0.6377	0.6380
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.2223	0.2229	0.2156	0.2246	0.2171	0.2203
20%	0.2735	0.2835	0.2807	0.2809	0.2863	0.2832
30%	0.3194	0.3210	0.3245	0.3267	0.3251	0.3225
50%	0.4000	0.4069	0.4059	0.4069	0.4085	0.4089
70%	0.4343	0.4389	0.4389	0.4417	0.4413	0.4416

表 4.41 DataSet2,測試集：非監督式訓練之詞層次主題混合模型實驗結果 (人工轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.4910	0.4958	0.4872	0.4723	0.4856	0.4822
20%	0.6175	0.6117	0.6137	0.6076	0.6012	0.6041
30%	0.6578	0.6601	0.6509	0.6507	0.6528	0.6554
50%	0.7346	0.7349	0.7346	0.7335	0.7305	0.7323
70%	0.7862	0.7861	0.7872	0.7839	0.7827	0.7864
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3339	0.3430	0.3227	0.3264	0.3212	0.3099
20%	0.4659	0.4626	0.4733	0.4554	0.4710	0.4371
30%	0.5079	0.5129	0.5115	0.4935	0.5056	0.4874
50%	0.6640	0.6648	0.6666	0.6620	0.6633	0.6484
70%	0.7983	0.7965	0.8031	0.7991	0.8044	0.7881

表 4.42 DataSet2,測試集：非監督式訓練之詞層次主題混合模型實驗結果 (自動轉寫文件)

Cosine	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.3828	0.3842	0.3875	0.3854	0.3833	0.3794
20%	0.4718	0.4728	0.4691	0.4670	0.4652	0.4642
30%	0.5098	0.5075	0.5079	0.5066	0.5047	0.5022
50%	0.5625	0.5629	0.5619	0.5644	0.5639	0.5632
70%	0.5836	0.5845	0.5843	0.5861	0.5848	0.5870
Rouge	*2*	*4*	*8*	*16*	*32*	*64*
10%	0.2091	0.2194	0.2204	0.2231	0.2218	0.2247
20%	0.2557	0.2675	0.2673	0.2688	0.2718	0.2696
30%	0.2876	0.2993	0.2985	0.3015	0.3009	0.3014
50%	0.3656	0.3688	0.3693	0.3692	0.3696	0.3690
70%	0.4111	0.4113	0.4102	0.4138	0.4112	0.4105

實驗中其他摘要方法及關聯性模型調適的摘要結果低；但是，在自動轉寫文件的結果中，監督式訓練與非監督式訓練之摘要結果的差異並不大，同時於 Rouge-2 評估結果中，摘要結果亦較基礎實驗及關聯性模型調適的摘要結果好。因此，雖然於非監督式訓練下人工轉寫文件的摘要結果並不好，但是經由實驗結果證明，採用非監督式訓練之詞層次主題混合模型，適用於自動轉寫文件摘要的處理。

從 DataSet1 與 DataSet2 的非監督式訓練及監督式訓練實驗結果，不論於 DataSet1 或 DataSet2 的監督式訓練之詞層次主題混合模型的摘要實驗，皆有相當不錯的摘要結果；同時，實驗結果亦顯示出非監督式訓練之詞層次主題混合模型適用於自動轉寫文件中，並且可以得到很好的摘要結果，摘要正確率甚至較監督式訓練之摘要結果來得好。因此，在實際應用上，便可使用非監督式訓練的方式來訓練詞層次主題混合模型。

最後，比較本論文中所提出之關聯性模型於隱藏式馬可夫模型的調適及詞層次主題混合模型的摘要結果，詞層次主題混合模型似乎有些微較高的正確率，其中可能的原因如前面討論所提，可能因前者存在有較多影響摘要結果的因素及較多參數需要調整。但是，兩種方法的摘要結果差異並不大，將大部分的實驗結果與基礎實驗作比較，於低摘要比例下皆有較好的摘要正確率，證明所提之方法確實可有效提昇摘要正確率。

4.4 摘要特徵於機率生成式摘要模型之初步應用

文字文件內容，除了提供文字訊息作為重要文句選取的依據，文句中更包含了文法、語意及結構等資訊，可視為重要的摘要資訊(Prosodic Information)；然而，語音文件內容可能因辨識錯誤、文句邊界定義等問題，而使得文字訊息、文法及語意的資訊相對較缺乏。但是，語音文件卻帶來豐富的聲韻特徵，例如：人在說話時，(1)發音的長短快慢，以及(2)語氣的抑揚頓挫、高低起伏等。因此，若能

將這些語言學、聲韻學及文件結構等資訊善加使用，相信可以提昇摘錄式語音文件摘要的精確率。

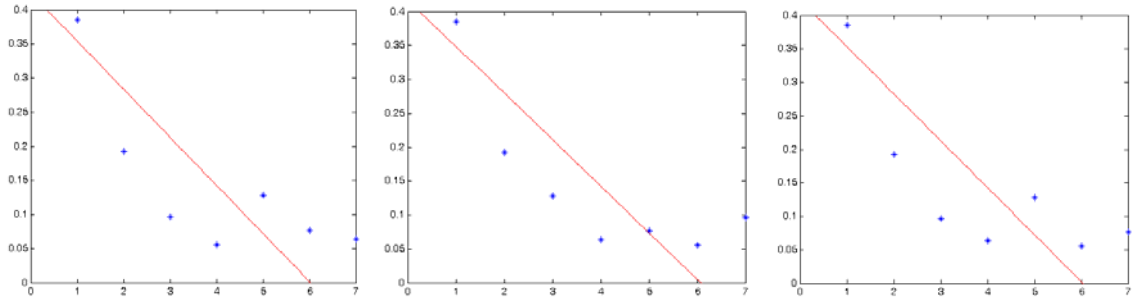
關於摘要特徵的相關研究，已經有許多學者提出許多研究成果[Koumpis et al. 2005; Maskey et al. 2005]。然而，大部份皆是將摘要特徵運用於摘要文句分類的作法上[Koumpis et al. 2005; Maskey et al. 2005]，使用每一文句的各種摘要特徵作為文句是否重要或歸納為摘要內容的依據；亦或是將摘要特徵值用於計算文句重要性分數[Kikuchi et al. 2003]。本小節中，將探討如何將不同摘要特徵使用於機率生成式摘要模型，每一文句的摘要特徵被視為文句本身的重要性資訊，用以估測文句事前機率。

如第一節所述，我們以最大事後機率法則為基礎，提出機率生成式摘要模型，來表示文句與文件之間的關係。其中，透過貝式定理與化簡，可表示為式(4-3)所示，以文句 S_i 生成文件 D 的可能性 $P(D|S_i)$ 及文句 S_i 的事前機率 $P(S_i)$ 之乘積，作為重要文句選取的依據。在無法得知文句 S_i 的事前資訊時，我們可以簡單地假設文件中所有文句為一致性分佈，使每一文句 S_i 的事前機率值相同，如同前面小節的作法：僅考慮文句生成文件的可能性作為文句重要性的依據。但是，在實際情況下，文件中每一文句的重要性與分佈並非是相同的。本小節中將嘗試使用文句中各種摘要特徵，或是其於文句中的分佈來估測每一文句的事前機率，同時分析使用不同摘要特徵所估測之事前機率，對於摘要結果是否有所提昇。

為了方便觀察與分析使用每一種摘要特徵所估測之文句事前機率，對於摘要結果的影響，我們將式(4-3)修改為：

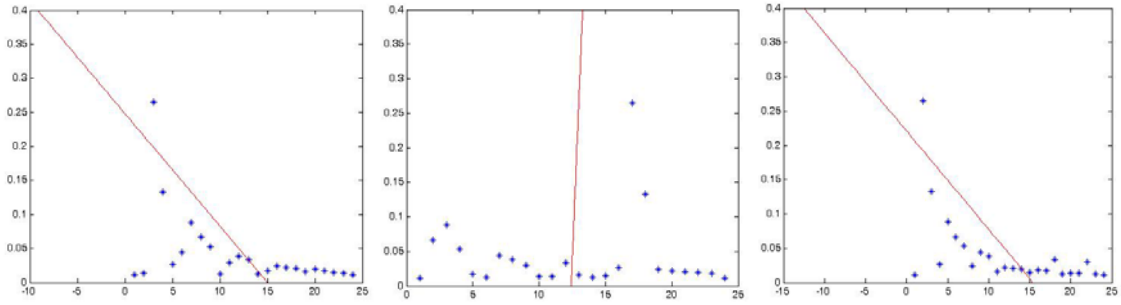
$$\max_{S_i} P(S_i|D) = \max_{S_i} P(D|S_i)P(S_i)^{\gamma} \tag{4-31}$$

γ 值用來強調或不強調事前機率 $P(S_i)$ 的重要性，當 γ 值越大代表越重視事前機率值的資訊。若將 γ 值設定為零，表示摘要結果僅根據文句生成文件的可能性 $P(D|S_i)$ 來選擇重要的文句；若 γ 值為無窮大（∞），表示摘要結果僅以文句事前機率 $P(S_i)$ 作為選取重要文句的依據。下面的實驗將嘗試不同 γ 值的調整，來分



a.第一份參照摘要之文句重要性與文句位置關係圖　b.第二份參照摘要之文句重要性與文句位置關係圖　c.第三份參照摘要之文句重要性與文句位置關係圖

圖 4.5 測試語料 DataSet1:參照摘要文件(N200108011200-01.txt)中文句重要性與文句位置關係圖（其中，橫軸表示文件中每一文句於原文件的位置，縱軸為參照摘要中文句的重要性分數）



a.第一份參照摘要之文句重要性與文句位置關係圖　b.第二份參照摘要之文句重要性與文句位置關係圖　c.第三份參照摘要之文句重要性與文句位置關係圖

圖 4.6 測試語料 DataSet2:參照摘要文件(PTSND20011107_1.txt)中文句重要性與文句位置關係圖（其中，橫軸表示文件中每一文句於原文件的位置，縱軸為參照摘要中文句的重要性分數）

析不同摘要特徵對實驗結果的影響；其中γ值的範圍我們設定為0~5每次遞增0.2、5~100每次遞增2、100~1000每次遞增50、1000~1000每次遞增500，以及無窮大(∞)等，來觀察於不同設定值下摘要結果的變化。

初步的實驗中，將嘗試使用十種摘要特徵於文句事前機率的估測，同時分別使用測試語料DataSet1與DataSet2進行一連串的實驗分析，觀察於不同文件內容長度的測試語料下之摘要結果。實驗以隱藏式馬可夫模型為文句生成模型，用以

計算文句生成文件的可能性$P(D|S_i)$，觀察結合不同摘要特徵所估測之文句事前機率後的摘要結果，其中隱藏式馬可夫模型參數λ於DataSet1及DataSet2皆設為0.05。相信由初步的實驗結果，可以發現不同摘要特徵所估測的文句事前機率與文句生成模型之間的關係，以及文句事前機率的估測對文件摘要正確率的影響。本小節中所有實驗結果採用Rouge-2評估方法[Lin 2003]，Rouge-2評估為目前語音文件摘要研究中最常用的評估方式，許多摘要研究文獻[Maskey et al. 2005; Hirohata et al. 2005] 皆使用此種評估方法。

4.4.1 文句位置（Location）

對於廣播新聞來說，主播通常會在播報一則新聞時，先描述新聞的大綱或主題以吸引聽眾或觀眾的注意。經由對新聞測試語料的觀察，我們發現廣播新聞文件中文句的位置資訊是一個很重要的摘要特徵；當文句位於原始文件中越前面的位置，其重要性越高。從參照摘要文件中分析文句重要性排名與文句位於原始文件中的位置之間的關係，如圖4.5與圖4.6所示，可觀察兩者之間是否具有線性關係。其中，依據文句重要性排名可計算其重要性排名分數$\mathrm{Im}(S_i)$(如圖4.5與圖4.6中縱軸的分數)：

$$\mathrm{Im}(S_i) = \frac{1/SenRank(S_i)}{\sum_{S_z \in D} 1/SenRank(S_z)} \tag{4-32}$$

是以文句S_i在參照摘要中人工所標註的文句排名$SenRank_{S_i}$之倒數，然後經正規化(Normalization)處理，使文件D中所有文句S_i的重要性排名分數之和為1。圖4.5與圖4.6中分析圖a、b、c分別表示DataSet1中文件“N200108011200-01.txt”及DataSet2中“PTSND20011107_1.txt”於三份不同參照摘要內容中，每一文句重要性與其位於文件的位置之間的關係圖。從圖4.5中a、b、c三小圖可看出文句重要

性與其位置之間，確實存在著些微的線性關係，當文句於文件中的位置越前面，其重要性越高；從圖4.6中a、c小張圖亦可以看出些微線性關係。

因此，本論文嘗試將文件中每一文句於文件的位置資訊，用來估測文句事前機率：

$$P(S_i) = \frac{L(S_i)}{\sum_{S_z \in D} L(S_z)} \tag{4-33}$$

$$L(S_i) = \frac{TotalSentence_D}{SentenceLocation_{S_i}} \tag{4-34}$$

其中，$L(S_i)$為文句S_i於文件D中的重要性分數，是以文句於文件的位置資訊來計算，用以表示文句於文件中的位置越前面者越重要；$TotalSentence_D$為文件D的總文句數，$SentenceLocation_{S_i}$表示文句S_i位於文件D的位置，例如文句S_i若是文件D內容的第一個文句，則其位置為1。

測試語料 DataSet1 實驗結果：

圖4.7與圖4.8為發展集中人工轉寫文件與自動轉寫文件，於不同γ值及不同摘要比例下摘要結果的曲線圖，其中不同的曲線代表不同摘要比例之摘要結果。圖中，橫軸表示式(4-31)參數γ的值，縱軸表示摘要正確率。透過這樣的曲線圖可以清楚地觀察所估測之文句事前機率，對於摘要結果的影響。當γ值為0時（圖中最左側的點），為僅使用文句之隱藏式馬可夫模型生成文件的可能性來產生句排名的摘要結果；當γ值為無窮大時（圖中最右側的點），為僅以文句事前機率產生句排名的摘要結果。如此，當我們可以從圖中任一條曲線中找出其中一點，其所對應的正確率高於最左側與最右側兩點的正確率，即表示當文句生成模型(隱藏式馬可夫模型)與文句事前機率作適當的結合(某個γ值設定)時，可以提昇摘要正確率。

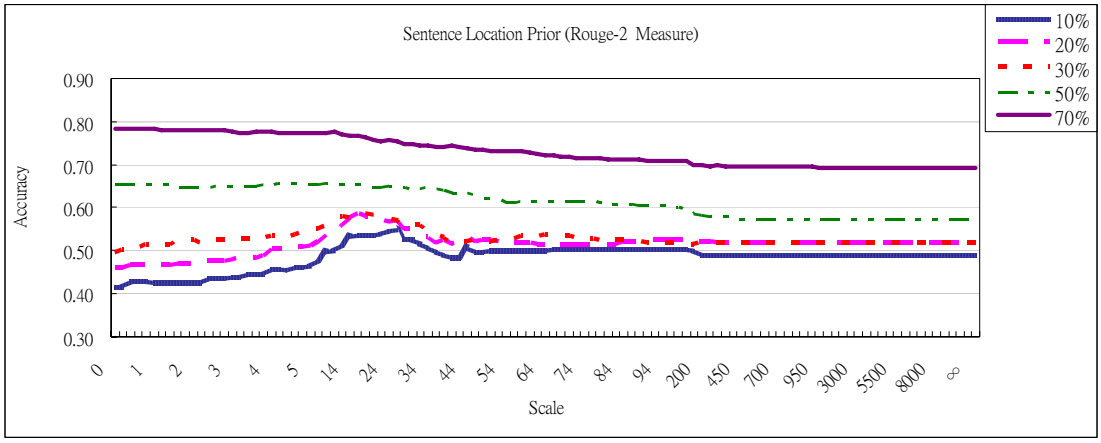


圖 4.7 隱藏式馬可夫模型結合文句位置資訊的摘要結果(DataSet1,發展集人工轉寫文件)

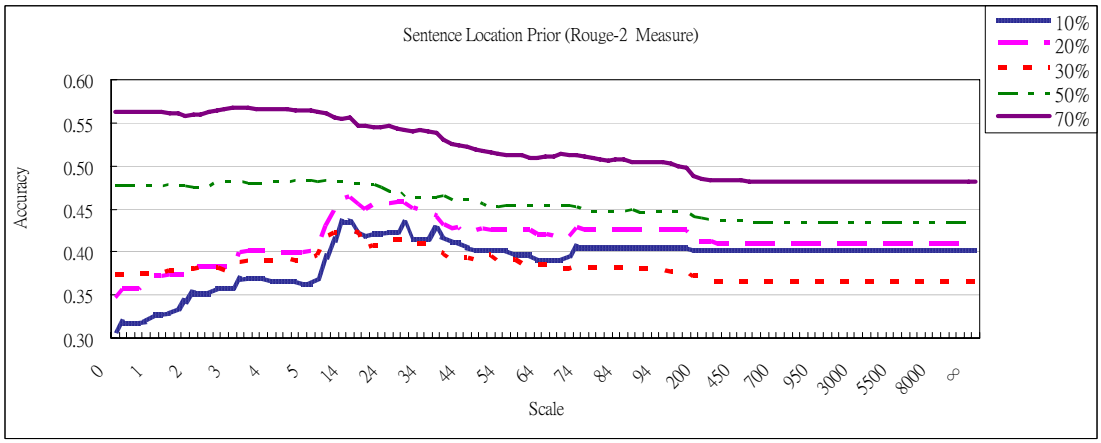


圖 4.8 隱藏式馬可夫模型結合文句位置資訊的摘要結果(DataSet1,發展集自動轉寫文件)

從圖 4.7 與圖 4.8 結果中發現：當摘要比例為 10%、20%及 30%時，使用文句位置資訊所估測之事前機率，可有效地提昇人工轉寫文件及自動轉寫文件的正確率。觀察僅以位置資訊所估測之文句事前機率來產生文句重要性排名的結果，於低摘要比例為 10%、20%及 30%皆有相當不錯的摘要結果，甚至較隱藏式馬可夫模型的摘要結果還要好。但是，文句生成模型與事前機率之結合可得到更高的正確率。在人工轉寫文件結果中(圖 4.7)，當摘要比例為 10%及 γ 值為 26 時，正確率由 0.4157 提昇至 0.5493，有 13%以上的進步；當摘要比例為 20%及 γ 值為 16 時，正確率由 0.4591 提昇至 0.5840，正確率亦有 12%以上的進步。在自動轉

寫文件中(圖 4.8)，當摘要比例為 10%及γ值為 14 時，正確率由 0.3084 提昇至 0.4352，有 12%以上的進步；當摘要比例為 20%及γ值為 14 時，正確率由 0.3467 提昇至 0.4642，正確率亦有近 12%的進步。

測試語料 DataSet2 實驗結果：

同樣地，圖 4.9 與圖 4.10 為發展集人工轉寫文件與自動轉寫文件於不同γ值的摘要結果。觀察僅使用文句事前機率來選取重要文句之摘要結果，人工轉寫文件及自動轉寫文件結果於低摘要比例為 10%、20%及 30%時，較隱藏式馬可夫模型的摘要結果好，尤其是自動轉寫文件摘要比例為 10%時結果最明顯。但是，文句生成模型與事前機率的結合有更高的正確率。

由人工轉寫文件結果(圖 4.9)來看，當摘要比例為 10%及γ值為 46 時，正確率由 0.2525 提昇至 0.5064，有 25%以上的進步；當摘要比例為 20%及γ值為 56 時，正確率由 0.4215 提昇至 0.5817，正確率亦有 16%以上的進步。在自動轉寫文件(圖 4.10)結果中，當摘要比例為 10%及γ值為 40 時，正確率由 0.2089 提昇至 0.3600，有 15%以上的進步；當摘要比例為 20%及γ值為 72 時，正確率由 0.2789 提昇至 0.4091，正確率亦有近 13%的進步。

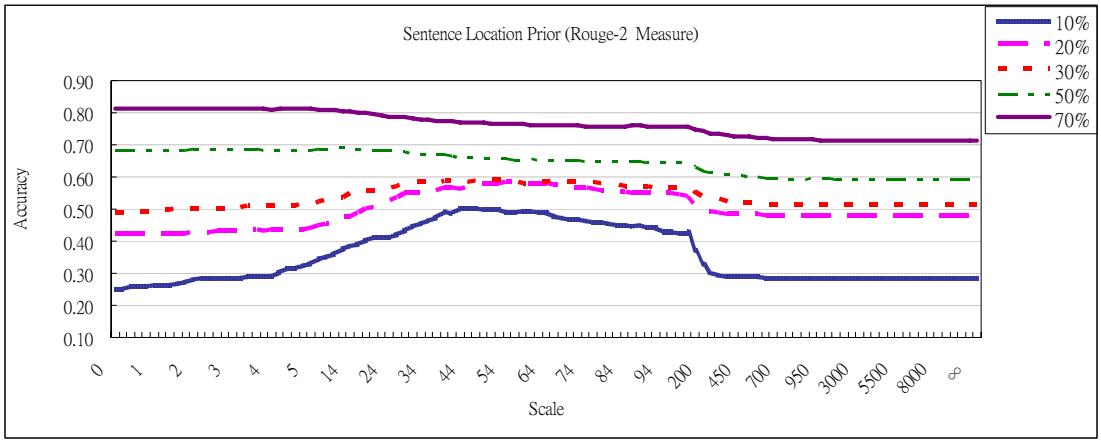


圖 4.9 隱藏式馬可夫模型結合文句位置資訊的摘要結果(DataSet2,發展集人工轉寫文件)

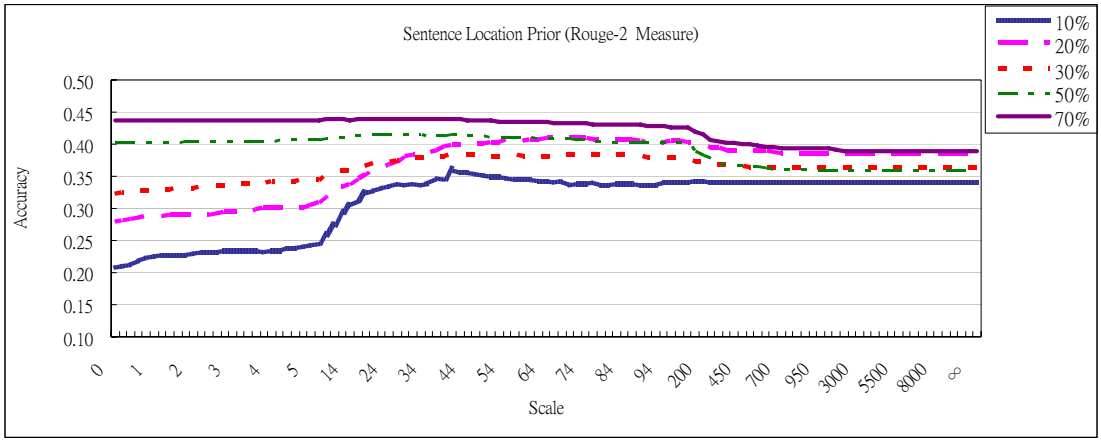


圖 4.10 隱藏式馬可夫模型結合文句位置資訊的摘要結果(DataSet2,發展集自動轉寫文件)

從 DataSet1 及 DataSet2 的摘要結果觀察，文句的位置資訊為新聞語料中重要的摘要特徵，對於人工轉寫文件及自動轉寫文件的正確率有明顯的提昇。將文句生成模型與以位置資訊所估測之事前機率作適當的結合，可以得到更好的摘要結果，較單獨使用文句生成模型或事前機率所產生之摘要結果好。實驗結果亦證明了若對文句事前機率作適當的估測，可有效提昇摘要之正確率。

4.4.2 文句長度（Sentence Length）

使用文句的長度作為摘要特徵。以詞為單位元計算文句中單位元的個數作為文句的長度，然後以文句長度佔整篇文件長度的比例作為文句之事前機率：

$$P(S_i) = \frac{|S_i|}{|D|} \tag{4-35}$$

用以表示當文句的長度越長，可能包含較多或越重要的內容，因此越為重要。

測試語料 DataSet1 實驗結果：

從圖 4.11 與圖 4.12 為人工轉寫文件及自動轉寫文件摘要結果，當摘要比例為 10%、20%及 30%時，此摘要特徵可提昇人工轉寫文件及自動轉寫文件的正確率。此外，由結果發現，在摘要比例為 10%、20%及 30%時，比較僅以文句長度所估測之文句事前機率結果，與僅以文句生成模型－隱藏式馬可夫模型所產生的摘要結果，文句生成模型的摘要結果有較高的正確率；但是，在摘要比例為 50%

及 70%時，僅以文句長度所估測之文句事前機率的摘要結果，較隱藏式馬可夫模型的摘要結果好。

在人工轉寫文件結果中，當摘要比例為 10%及γ值為 1 時，正確率由 0.4157 提昇至 0.4285；當摘要比例為 20%及γ值為 44 時，正確率由 0.4591 提昇至 0.4720，正確率皆有 1%以上的進步。在自動轉寫文件的摘要結果中，當摘要比例為 10%及γ值為 14 時，正確率由 0.3084 提昇至 0.3163；當摘要比例為 20%及γ值為 20 時，正確率由 0.3467 提昇至 0.3624。於自動轉寫文件上正確率提昇的幅度較小一些。

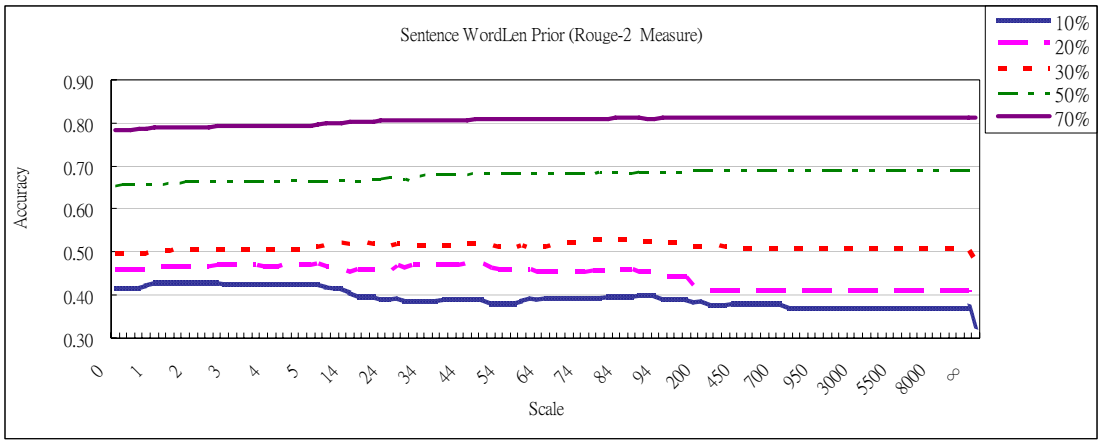


圖 4.11 隱藏式馬可夫模型結合文句長度之摘要結果(DataSet1,發展集人工轉寫文件)

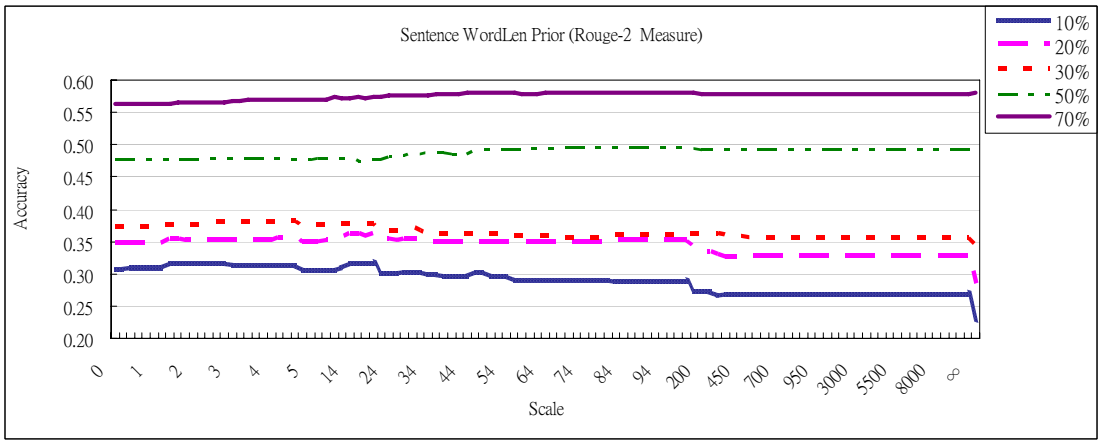


圖 4.12 隱藏式馬可夫模型結合文句長度之摘要結果(DataSet1,發展集自動轉寫文件)

測試語料 DataSet2 實驗結果：

同樣，亦可由 DataSet 的人工轉寫文件與自動轉寫文件(圖 4.13 與圖 4.14)中，發現相同的現象：於低摘要比例 10%、20%及 30%時，正確率有些許的提昇；但是，於摘要比例 50%及 70%時，僅以文句長度所估測之文句事前機率的摘要結果，較隱藏式馬可夫模型的摘要結果好。從人工轉寫文件結果來看，當摘要比例為 10%且於γ值為 3.8 時，正確率由 0.2525 些微地提昇至 0.2603。正確率於人工轉寫文件上提昇的幅度並不大，但是可對隱藏式馬可夫模型的摘要結果作微調。在自動轉寫文件摘要結果中，不同摘要比例結果中摘要比例為 10%時，正確率有較明顯的進步，正確率由 0.2089 提昇至 0.2159，約 0.7%的進步，但是提昇的幅度同樣亦不大。

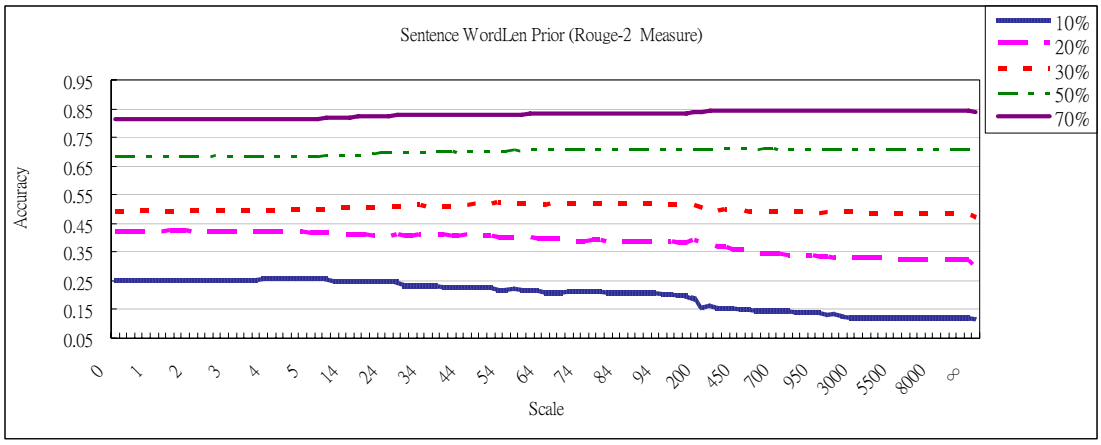


圖 4.13 隱藏式馬可夫模型結合文句長度之摘要結果(DataSet2,發展集人工轉寫文件)

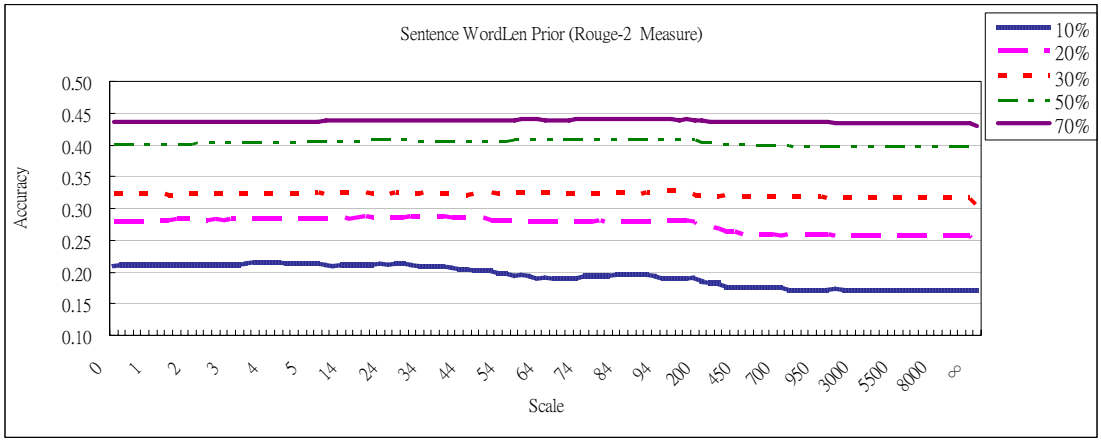


圖 4.14 隱藏式馬可夫模型結合文句長度之摘要結果(DataSet2,發展集自動轉寫文件)

4.4.3 語言學特徵－雙連詞語言模型分數（Bigram Language Model Score）

自然語言中包含了前後文資訊、語意資訊以及文法資訊等內容，好的語言模型估測可將這些資訊正確地表示出來。當給定一段詞序列$W = w_1, w_2, ..., w_n$，其語言模型機率$P(W)$，可表示成：

$$P(W) = P(w_1, w_2, ..., w_n) \tag{4-36}$$

可以利用連鎖律分解成：

$$P(W) = P(w_1)P(w_2 \mid w_1)...P(w_n \mid w_1, w_2, ..., w_{n-1}) = \prod_{i=1}^{n} P(w_i \mid h_i) \tag{4-37}$$

然而，當n很大時並沒有辦法直接計算機率值$P(w_n \mid w_1, w_2, ..., w_{n-1})$。因此，通常使用$N-1$階馬可夫模型假設，計算$N$連語言模型。其中，詞雙連(Word Bigram)語言模型為在自然語言中某個詞w_{n-1}後面緊接著詞w_n的可能性，以機率$P(w_n|w_{n-1})$表示，通常以大量的文件語料集來估測此機率值，可視為簡單的文法或語義概念的表示。當兩個詞w_{n-1}與w_n於一個語言中經常同時被使用，或是其屬於片語或慣用語時，機率值$P(w_n|w_{n-1})$便會較大；反正，若兩個詞不會一起出現於同一文句敘述中，通常也表示這兩個詞於文句中相連著使用是不合乎文法或語意的，機率值$P(w_n|w_{n-1})$會較小。

在包含辨識錯誤內容的自動轉寫文件中，辨識錯誤的文字可能造成文句內容

的不流暢；因此，我們使用文句的雙連詞語言模型的分數來估測文句事前機率值，認為平均雙連詞語言模型分數越大的文句越重要。文句的平均雙連詞語言模型分數的計算方式為：

$$avgLg(S_i)=\frac{1}{N_i}\log\big(P(w_1\mid Sil)\big)\sum_{n=2}^{N_i}\log\big(P(w_n\mid w_{n-1})\big) \tag{4-38}$$

其中，N_i 為文句 S_i 的長度，機率值 $P(w_1\mid Sil)$ 為文句開頭的第一個詞為 w_1 的機率值。文句事機率機為經正規化之文句平均雙連詞語言模型分數：

$$P(S_i)=\frac{avgLg(S_i)}{\sum_{S_z\in D}avgLg(S_z)} \tag{4-39}$$

測試語料DataSet1實驗結果：

圖4.15與圖4.16分別表示測試語料DataSet1發展集中，人工轉寫文件與自動轉寫文件的摘要結果；從摘要結果中觀察，文句平均雙連詞語言模型分數對摘要結果的提昇並不顯著。於人工轉寫文件結果中僅在低摘要比例10%及20%時，正確率有些微提昇，當摘要比例為10%及 γ 值為8時，正確率僅由0.4157提昇至0.4227，當摘要比例為20%時及 γ 值為36時，正確率由0.4591些微提昇至0.4651，而在高摘要比例50%及70%時正確率沒有進步的現象；於自動轉寫文件結果中，低摘要比例10%及20%時正確率有較明顯的提昇，當摘要比例為10%及 γ 值為28時，正確率由0.3084提昇至0.3275，當摘要比例為20%時及 γ 值為28時，正確率由0.3467些微提昇至0.3665，皆有接近2%的進步。

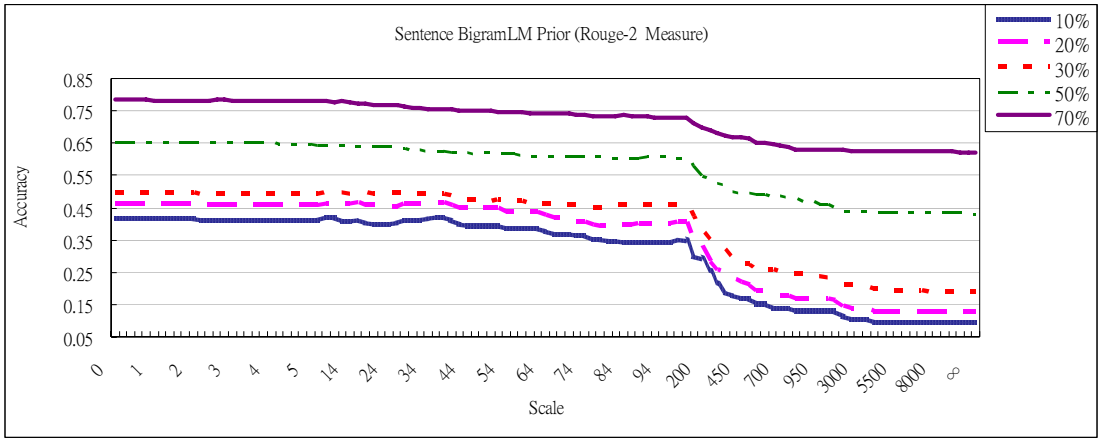


圖 4.15 隱藏式馬可夫模型結合詞雙連分數之摘要結果(DataSet1,發展集人工轉寫文件)

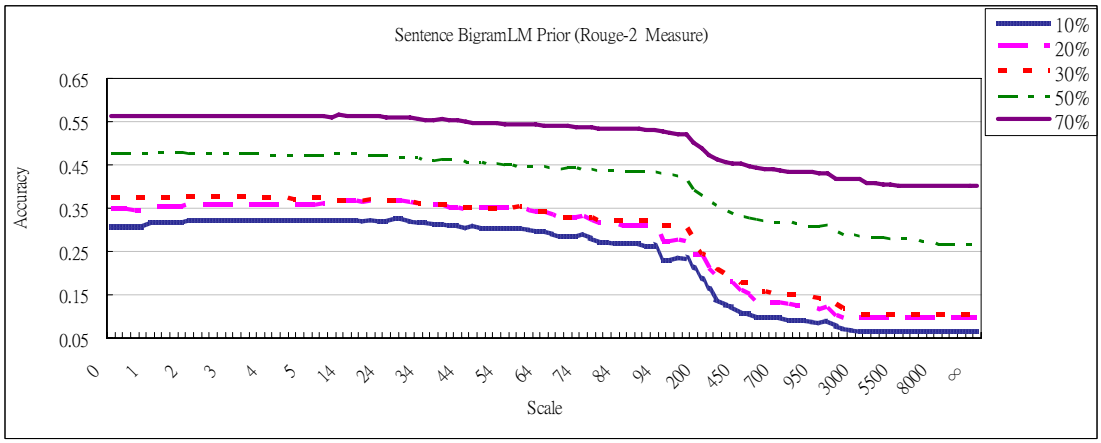


圖 4.16 隱藏式馬可夫模型結合詞雙連分數之摘要結果(DataSet1,發展集自動轉寫文件)

測試語料 DataSet2 實驗結果：

圖 4.17 與圖 4.18 為 DataSet2 中發展集人工轉寫文件與自動轉寫文件之摘要結果。從圖 4.17 及圖 4.18 的結果來看，不論是於人工轉寫文件或自動轉寫文件的摘要結果，皆顯示出使用文句平均雙連詞語言模型分數所估測之事前機率，對正確率提昇的幫助並不大。從人工轉寫文件結果中，發現僅於低摘要比例為 10%時正確率有不到 1%的進步，而在自動轉寫文件結果中，不論γ值設定為多少，正確率幾乎是沒有提昇

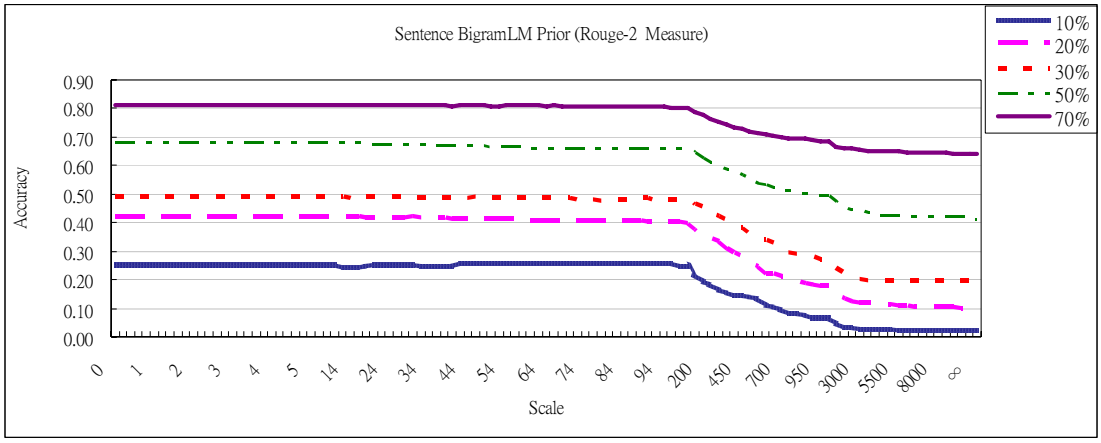


圖 4.17 隱藏式馬可夫模型結合詞雙連分數之摘要結果(DataSet2,發展集人工轉寫文件)

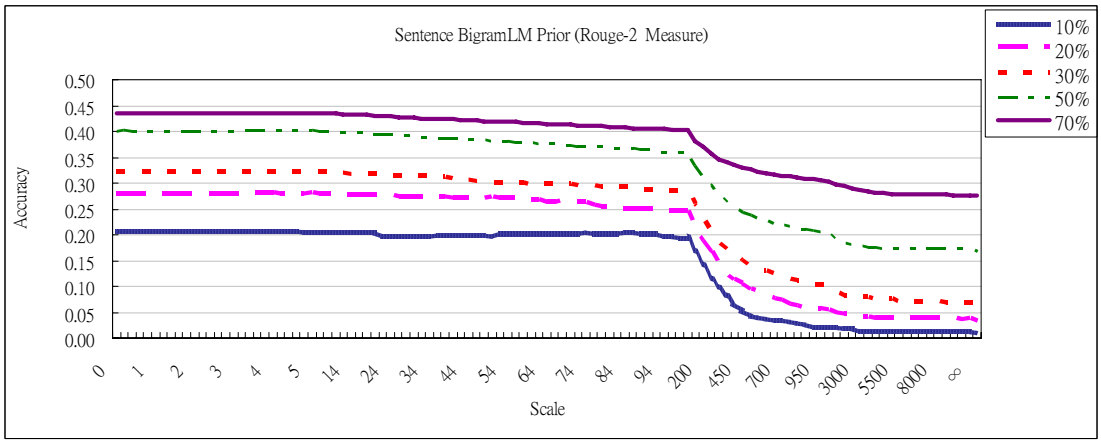


圖 4.18 隱藏式馬可夫模型結合詞雙連分數之摘要結果(DataSet2,發展集自動轉寫文件)

在使用詞雙連分數來估測文句事前機率進行摘要實驗的結果中，DataSet1 及 DataSet2 的摘要結果並沒有一致性，從 DataSet1 結果中顯示出此摘要特徵於低摘要比例 10%、20%下，對人工轉寫文件與自動轉寫文件的摘要正確率都有些微的提昇；但是，在 DataSet2 結果中顯示此摘要特徵對正確率的提昇是沒有幫助的。因此，此摘要特徵所估測之文句事前機率可能較適用於文件內容較少的文件，同時對於不同的測試集，其摘要結果可能會不一致。對文句事前機率的估測而言，似乎並不是一個好的摘要特徵。

4.4.4 信心度分數（Confidence Score）

辨識信心度分數，是對語音辨識結果的正確性與可靠度的估測，通常為一個介於0~1 的機率值，機率值越大表示對辨識結果越有信心、越相信辨識結果。由於環境、語者之間的差異，以及聲學模型、語言模型訓練語料的不足、訓練與測試語料的不匹配等問題，語音辨識的正確率無法達到百分之百。因此，許多學者研究及提出不同信心度評估(Confidence Measure)的方法，用來自動判斷辨識結果的可靠程度，可應用於許多領域上。

本論文中將辨識信心度分數視為一種摘要特徵，用以選取辨識信心度較高的文句，希望所摘錄之摘要內容能包含較少辨識錯誤文字。其中，每一個詞的辨識信心度分數為詞之辨識事後機率值（Posterior Probability）[Kemp and Schaaf 1997; 陳燦輝 2006]，是以計算語音辨識所產生的詞圖中，包含詞 w_n 的所有詞序列事後機率與詞圖中所有詞序列事後機率之比例，如式(2-13)。每一文句的平均信心度分數可表示為：

$$avgConf(S_i) = \frac{1}{N}\sum_{n=1}^{N_i} C(w_n) \tag{4-40}$$

其中，N_i 為文句 S_i 的長度，$C(w_n)$ 為詞 w_n 的辨識信心度分數。將式(4-39)經正規化之機率值作為文句事機率機：

$$P(S_i) = \frac{avgConf(S_i)}{\sum_{S_z \in D} avgConf(S_z)} \tag{4-41}$$

測試語料DataSet1實驗結果：

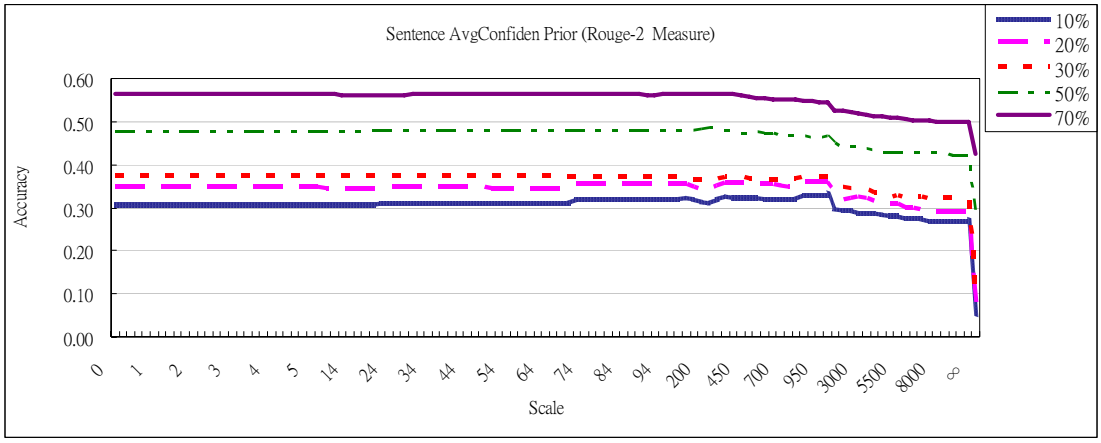


圖 4.19 隱藏式馬可夫模型結合信心度分數之摘要結果(DataSet1,發展集自動轉寫文件)

圖4.19為發展集自動轉寫文件之摘要結果，可以看出於低摘要比例10%、20%及30%下摘要正確率有較明顯的提昇。當摘要比例為10%及γ值為56時，正確率由0.3084提昇至0.3372；當摘要比例為20%時及γ值為28時，正確率由0.3467提昇至0.3777，正確率皆有將近3%的進步；當摘要比例為30%時及γ值為6時正確率由0.3734提昇至0.3807，亦有0.7%的進步。可以看出辨識信心度對於自動轉寫文件摘要正確率的影響與提昇。

測試語料 DataSet2 實驗結果：

圖4.20為發展集自動轉寫文件之摘要結果，結果顯示出於不同摘要比例下，摘要正確率皆有明顯的提昇，高摘要比例50%及70%時正確率有較明顯的進步。當摘要比例為10%及γ值為90時，正確率由0.2089提昇至0.2232；當摘要比例為20%時及γ值為86時，正確率由0.2789提昇至0.2973，正確率有1.5%的進步；而當摘要比例為70%時及γ值為6時，正確率可由0.4364提昇至0.4649，有2.8%的進步。同樣可以從實驗結果中看出辨識信心度對於自動轉寫文件摘要正確率提昇。

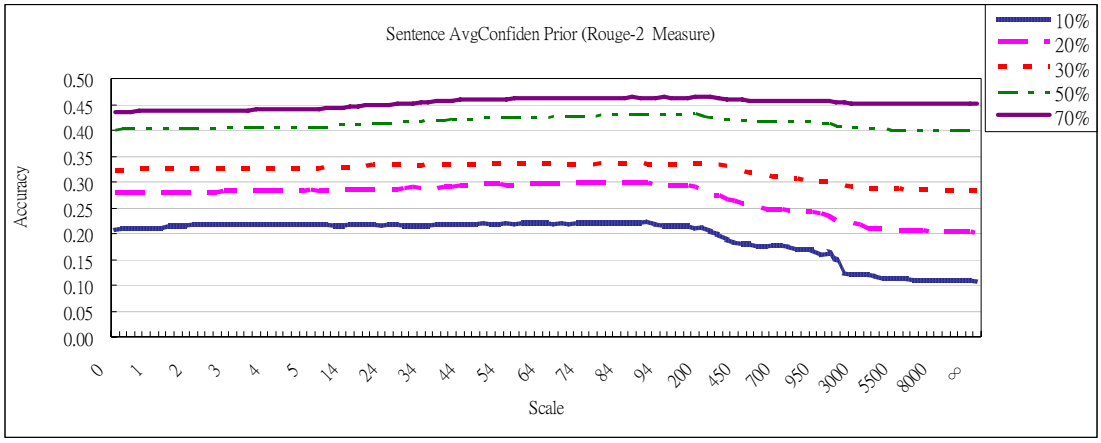


圖 4.20 隱藏式馬可夫模型結合信心度分數之摘要結果(DataSet2,發展集自動轉寫文件)

我們從 DataSet1 及 DataSet2 的實驗中可以看到一致性的摘要結果。證明信心度分數對於自動轉寫文件摘要而言，是一個可用的摘要特徵，能有效地提昇摘要正確率。

4.4.5 聲學特徵：音高（Pitch）

音高可用於判斷一個人講話的高低起伏。通常，人在敘述一件事情時，都會以說話的高低起伏、音揚頓挫，來強調或不強調說話的內容，以及吸引聽眾的目光。因此，音高可視為一種語音中重要的資訊，亦可用來估測文句的事前機率。論文中，音高的分析是使用 Snack 軟體[Snack]，求取每一個訊號音框(Frame)的音高值。其中，每一個詞所對應的音高值，可參考第二章式(2-16)與式(2-17)的計算方式[Huang et al. 2001]。以下的實驗，由於人工轉寫文件並沒有每個詞於語音訊號的起始時間與結束時間的資訊，吾人將語音與人工轉寫文件經對齊(Alignment)後，計算文件中每一個詞所對應的音高值。

實驗中，我們分別採用下列三種方式來估測文句事前機率：

a. 文句平均音高(Average Pitch)：計算每一文句的平均音高，如下所示：

$$AvgPitch(S_i) = \frac{1}{N_i}\sum_{n=1}^{N_i} Ph(w_n) \qquad (4\text{-}42)$$

其中，$Ph(w_n)$ 為詞 w_n 的音高值；N_i 為文句 S_i 的長度。如此，文句的事前機率可表示為：

$$P(S_i) = \frac{AvgPitch(S_i)}{\sum_{S_z \in D} AvgPitch(S_z)} \qquad (4\text{-}43)$$

表示對於同一個語者而言，當某一文句的平均詞音值較大，可能代表此文句較為重要，因此說話時特意提高此句的音高。

測試語料 DataSet1 實驗結果：

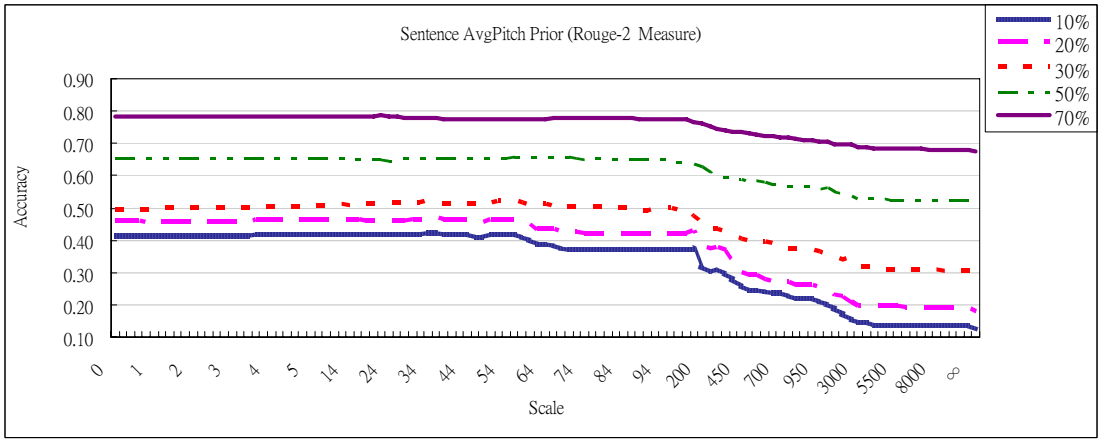


圖 4.21 隱藏式馬可夫模型結合平均音高值之摘要結果(DataSet1,發展集人工轉寫文件)

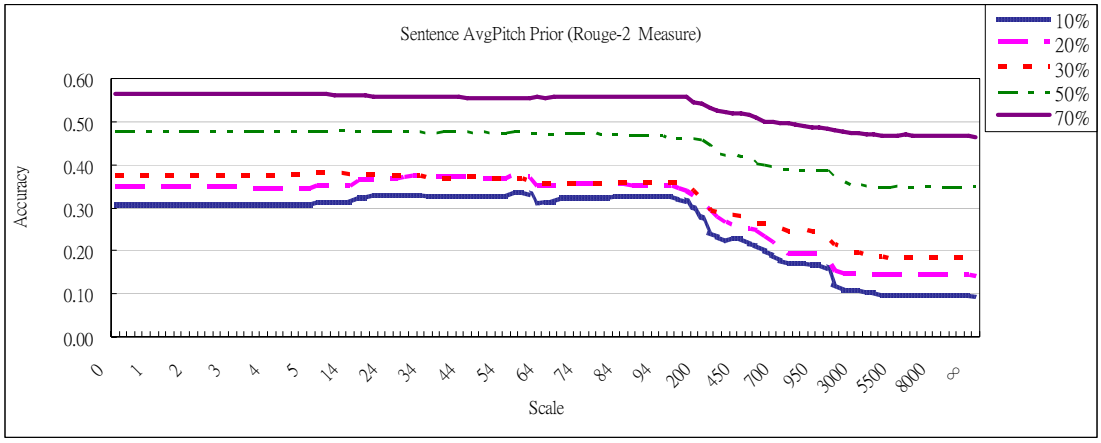


圖 4.22 隱藏式馬可夫模型結合平均音高值之摘要結果(DataSet1,發展集自動轉寫文件)

圖 5.21 與圖 5.22 為發展集中人工轉寫文件與自動轉寫文件的摘要結果。從人工轉寫文件的結果來看，於低摘要比例 10%、20%及 30%時正確率有些微的提昇。當摘要比例為 10%且γ值為 34 時，正確率由 0.4157 些微提昇至 0.4229；當摘要比例為 20%時且γ值為 36 時，正確率由 0.4591 提昇 1%至 0.4691。從自動轉寫文件結果觀察，於低摘要比例 10%及 20%時正確率有較明顯的提昇，當摘要比例為 10%且γ值為 56 時，正確率由 0.3084 提昇至 0.3372，有接近 3%的進步，當摘要比例為 20%且γ值為 56 時，正確率由 0.3467 提昇至 0.3777，亦有 3%以上的進步。摘要結果於自動轉寫文件上有較明顯的提昇。

測試語料 DataSet2 實驗結果：

從圖 4.23 的人工轉寫文件摘要結果來看，僅在於摘要比例 10%及 30%時，摘要正確率有些微的提昇。於摘要比例 10%摘要結果中，當γ值為 20 時會有最高的正確率 0.2681，較僅以文句生成模型－隱藏式馬可夫模型之摘要正確率 0.2525 有 1%以上進步；於摘要比例 30%摘要結果中，當γ值為 100 時正確率可由 0.4880 提昇至 0.5134。圖 4.24 為自動工轉寫文件摘要結果。從

結果來看，其正確率的提昇並不如人工轉寫文件明顯，但是於不同摘要比例下正確率皆有些微不到 1%的提昇，例如於摘要比例 10%且當γ值為 18 時，正確率由 0.2089 提昇至 0.2095；於摘要比例 20%且當γ值為 10 時，正確率由 0.2789 提昇至 0.2830。

雖然從 DataSet2 的實驗結果中，正確率提昇的情形並不明顯，只有些微的進步，但是對於文件摘要結果仍有所幫助。由 DataSet1 及 DataSet2 的實驗結果，可以看出文句平均音高為一個可利用的摘要特徵。

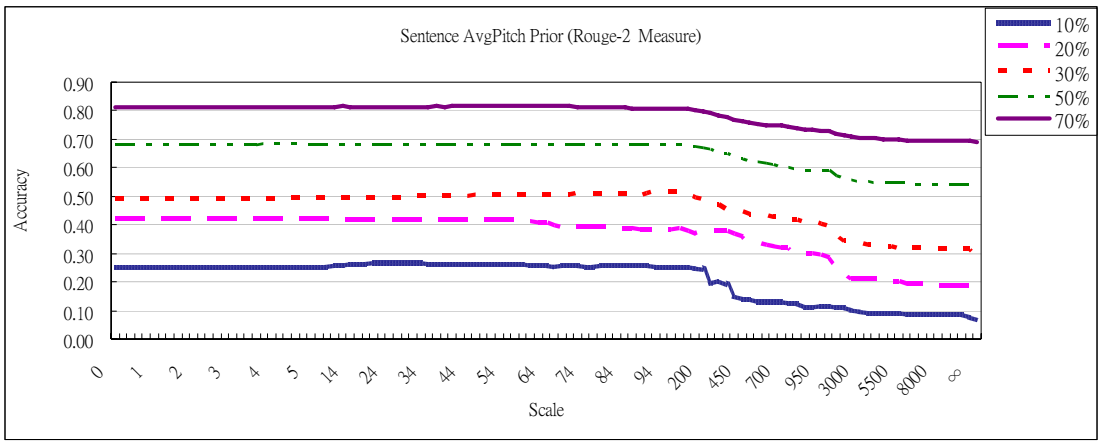


圖 4.23 隱藏式馬可夫模型結合平均音高值之摘要結果(DataSet2,發展集人工轉寫文件)

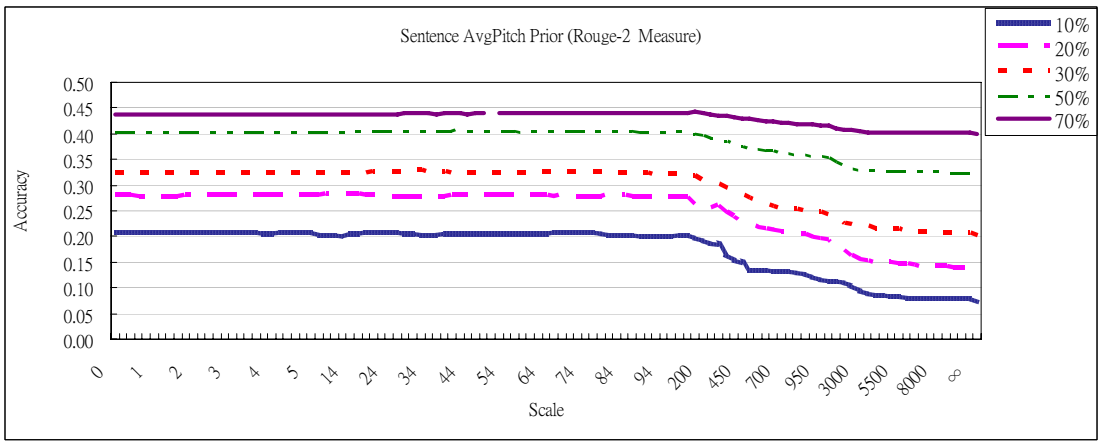


圖 4.24 隱藏式馬可夫模型結合平均音高值之摘要結果(DataSet2,發展集自動轉寫文件)

b. 文句音高熵值（Pitch Entropy）：熵值用於表示某個機率分佈的亂度：

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x) \tag{4-44}$$

X 為某個機率分佈中的所有事件，$p(x)$ 為某一事件 x 的機率值。當此分佈中每一個事件發生的機率越平均，代表此分佈的亂度越高，其熵值越大；反之，若此分佈中事件發生的機率越極端，代表此機率分佈的亂度越低，其熵值越小。吾人透過文句的熵值估測來表示文句中音高值的分佈情形。若文句有高低起伏的轉折，則其文句中的音高值就會有高有低，音高值分佈就會較亂，所計算的文句音高熵值便會較大；而當文句的描述語氣平穩無音揚頓挫，則其文句中的音高值可能差異不大且集中於某個值上下，所計算的文句音高熵值便會小。因此，即可透過每一文句的音高熵值表示文句的重要性。

首先，吾人先對文句中每個詞的音高值作量化的處理，如圖 4.25 所示。

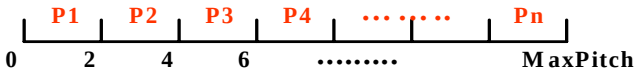


圖 4.25 音高值作量化處理之示意圖

先設定一個區間值，以 2 為例，從數值 0 開始，0~2 為屬同一個區間編號為 P1，2~4 為一個區間編號為 P2，範圍從 0 至文件中最大音高值為止，然後統計文句中每一個詞的音高值落於不同區間的次數，計算其機率值。如此，若將文句中每一個詞音高值落於某個區間視為一個事件 x，其機率值 $p(x)$，便可使用式(4-44)來計算文句的音高熵值。每一文句的正規化熵值(Normalized Entropy)為：

$$\varepsilon\_Pitch(S_i) = \frac{H(X)}{\log N} \tag{4-45}$$

最後，以文句正規化熵值所估測之文句事前機率可表示為：

$$P(S_i) = \frac{\varepsilon\_pitch(S_i)}{\sum_{S_z \in D} \varepsilon\_pitch(S_z)} \tag{4-46}$$

測試語料 DataSet1 實驗結果：

圖 4.26 與圖 4.27 分別為發展集人工轉寫文件及自動轉寫文件的摘要結果。從人工轉寫文件結果中，於摘要比例 10%、20%及 30%下正確率有極些微的提昇，約 0.5%以下的進步。但是，自動轉寫文件於摘要比例 10%及 20%下，正確率有較明顯的提昇；當摘要比例為 10%且γ值為 14 時，正確率由 0.3084 提昇至 0.3254，有 1.5%以上的進步，當摘要比例為 20%且γ值為 14 時，正確率由 0.3467 提昇至 0.3641，正確率亦有 1.5%以上的進步。摘要結果於自動轉寫文件上有較明顯的提昇。比較以文句平均音高及文句正規化熵值所估測之事前機率的摘要結果，前者可得到較高的正確率。

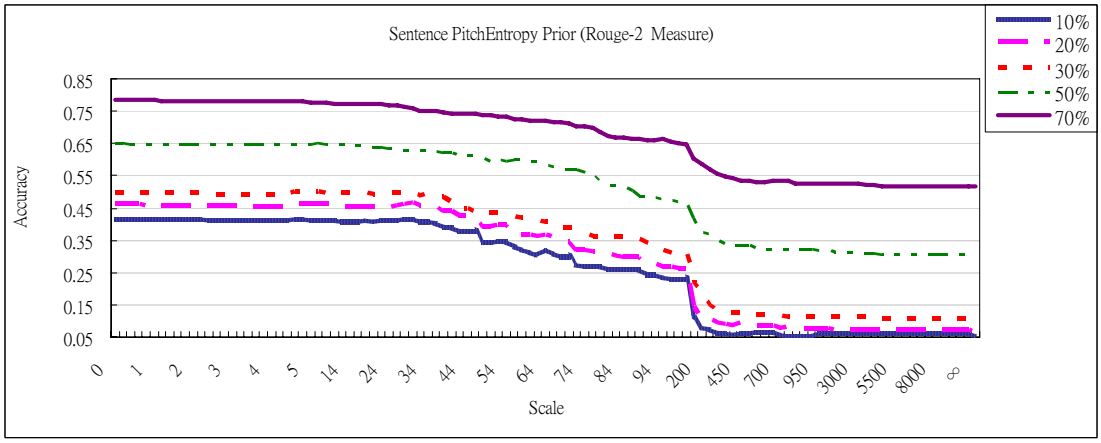


圖 4.26 隱藏式馬可夫模型結合音高熵值之摘要結果(DataSet1,發展集人工轉寫文件)

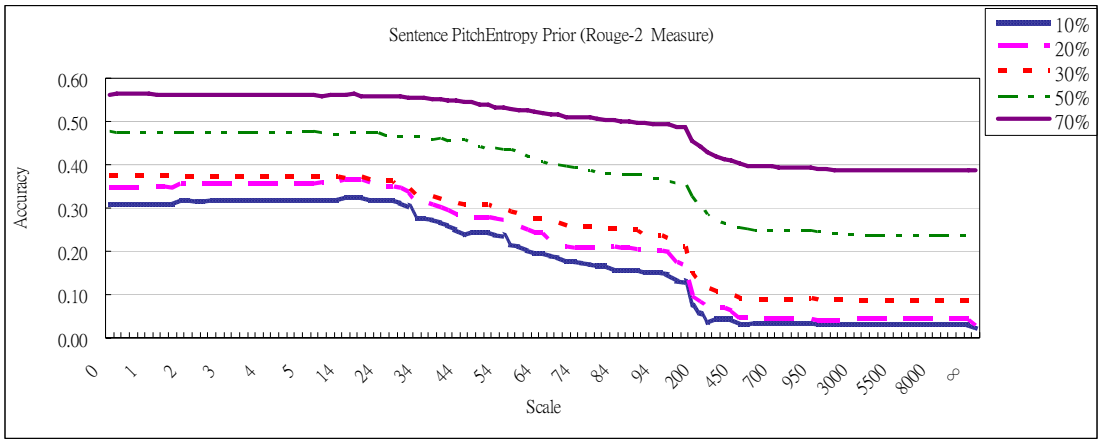


圖 4.27 隱藏式馬可夫模型結合音高熵值之摘要結果(DataSet1,發展集自動轉寫文件)

測試語料 DataSet2 實驗結果：

圖 4.28 為發展集中人工轉寫文件摘要結果，可以看出於低摘要比例 10%時，正確率有較明顯的提昇，由 0.2525 提昇至 0.2814 將近 3%的進步。若與圖 4.29 的自動轉寫文件結果比較，人工轉寫文件於摘要比例 10%下有較明顯的進步。從自動轉寫結果來看，同樣是於低摘要比例 10%時可看出正確率有較明顯的提昇，不過僅有約 0.7%的進步(由 0.2089 提昇至 0.2156)。

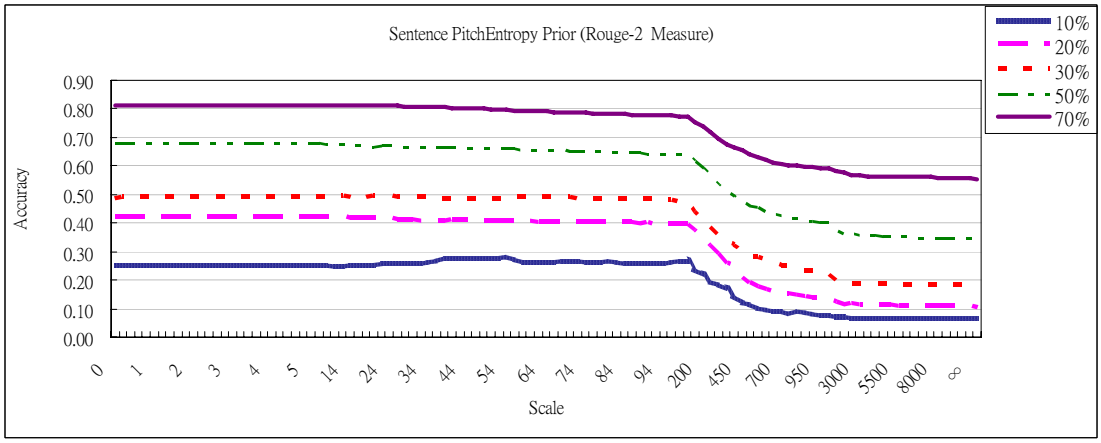


圖 4.28 隱藏式馬可夫模型結合音高熵值之摘要結果(DataSet2,發展集人工轉寫文件)

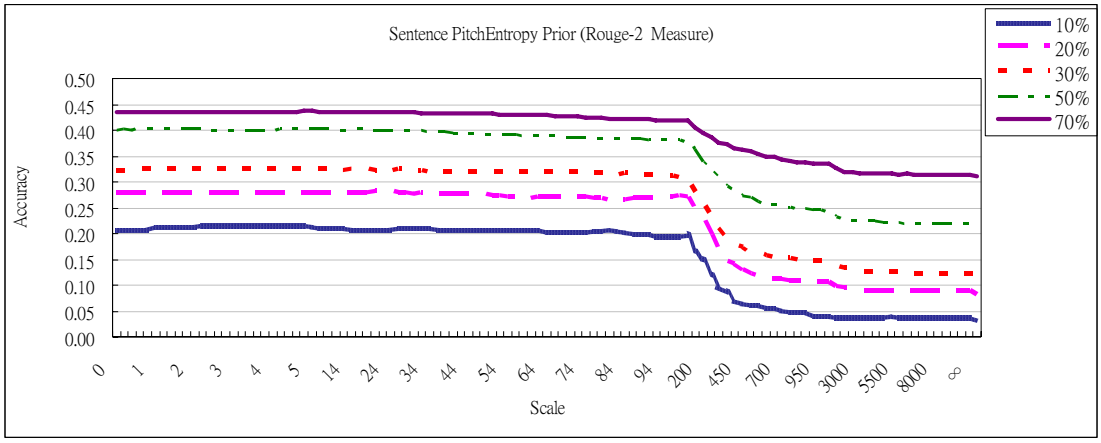


圖 4.29 隱藏式馬可夫模型結合音高熵值之摘要結果(DataSet2,發展集自動轉寫文件)

比較 DataSet1 與 DataSet2 的實驗結果，DataSet1 於自動轉寫文件中正確率有較明顯的提昇，而在 DataSet2 結果中正確率於人工轉寫文件上有較明顯的提昇，但是提昇幅度並不大。似乎只能說以文句正規化熵值所估測之文句事前機率，對文件內容較多的測試語料 DataSet2 摘要結果作了微調，使得其正確率有些微地進步。

c. 文句音高變異量（Pitch Variance）：在概念上與文句熵值的意義相似，用來表示文句中詞的音高值分散的情形(變異度)，是以計算文句中每一詞音高與其文句中平均詞音高的離散情形：

$$PitchVar(S_i) = \frac{1}{N_i} \sum_{w_n \in S_i} \left(Pitch(w_n) - AvgPitch_{S_i} \right)^2 \qquad (4\text{-}47)$$

$PitchVar(S_i)$ 表示文句 S_i 的音高變異量，$Pitch(w_n)$ 為文句 S_i 中詞 w_n 的音高值，$AvgPitch_{S_i}$ 為文句 S_i 的平均音高值，N_i 為文句 S_i 的長度。吾人認為：當某一文句的音高變異量較大，表示此文句中可能有高低起伏的語氣轉折，因此音高值較為分散；反之，可能表示語者說話的語氣平穩，則此文句較不重要。最後，文句的事前機率可表示為：

$$P(S_i) = \frac{PitchVar(S_i)}{\sum_{S_z \in D} PitchVar(S_z)} \quad (4\text{-}48)$$

測試語料 DataSet1 實驗結果：

圖 4.30 為發展集人工轉寫文件摘要結果。從結果來看，摘要正確率於低摘要比例 10%及 20%時並沒有提昇，考慮以文句音高變異量來估測事前機率的結果，顯示出對於低摘要比例之正確率沒有幫助。但是，於摘要比例 50%及 70%時，正確率有些微的提昇(例如於摘要比例 70%正確率由 0.7845 提昇至 0.7903)。圖 4.31 為發展集自動轉寫文件的結果，可以看出當摘要比例為 10%及 20%時，正確率有些許的提昇。摘要比例為 10%及γ值為 0.2 時，正確率由 0.3084 提昇至 0.3205；當摘要比例為 20%且γ值為 4 時，正確率由 0.3467 提昇至 0.3628，正確率有 1%以上的進步。因此，可以從初步的實驗分析中發現，文句音高變異量對於自動轉寫文件摘要而言，為可利用的摘要特徵。

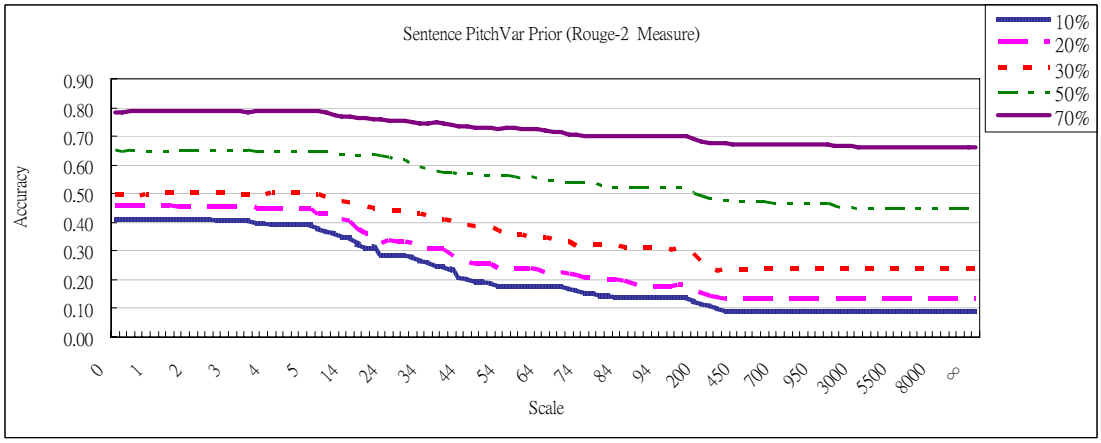


圖 4.30 隱藏式馬可夫模型結合音高變異量之摘要結果(DataSet1,發展集人工轉寫文件)

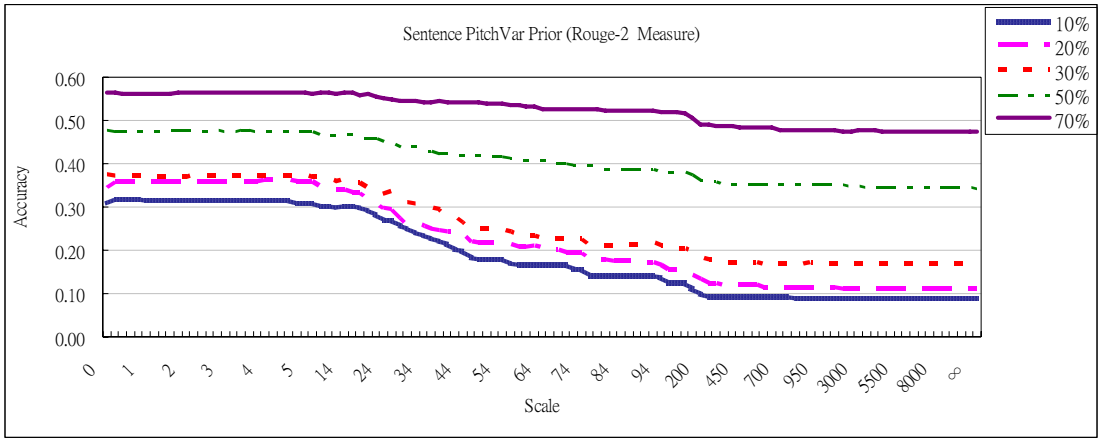


圖 4.31 隱藏式馬可夫模型結合音高變異量之摘要結果(DataSet1,發展集自動轉寫文件)

測試語料 DataSet2 實驗結果：

圖 4.32 及圖 4.33 為發展集中人工轉寫文件及自動轉寫文件的結果。由圖 4.32 及圖 4.33 的結果來看，於低摘要比例 10%下正確率皆有約 1%以上的提昇。

觀察 DataSet1 與 DataSet2 下的摘要結果，發現此摘要特徵(文句音高變異量)較適用於自動轉寫文件中，並且可有效地提昇低摘要比例 10%及 20%下之摘要結果。

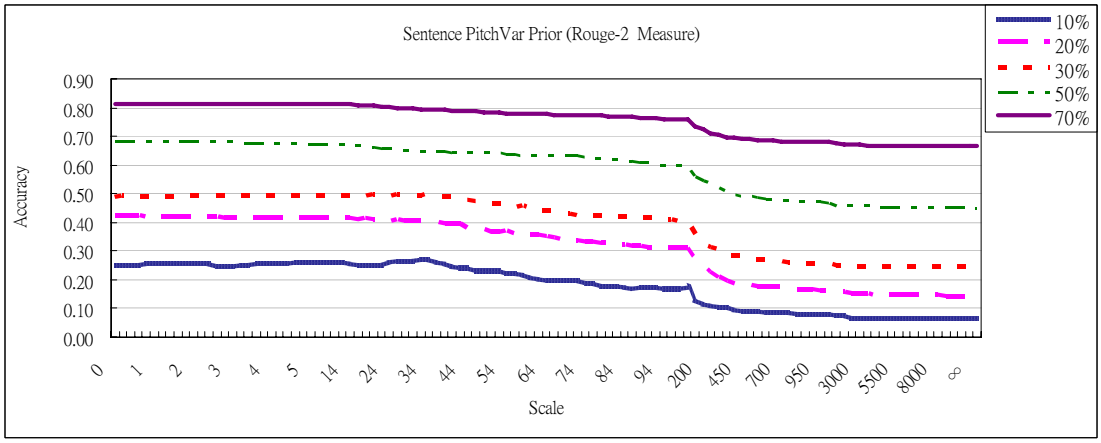


圖 4.32 隱藏式馬可夫模型結合音高變異量之摘要結果(DataSet2,發展集人工轉寫文件)

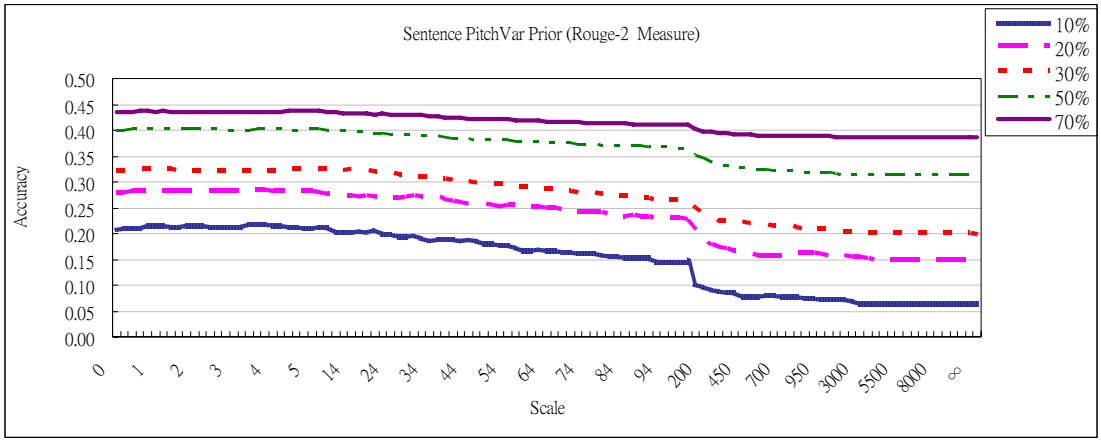


圖 4.33 隱藏式馬可夫模型結合音高變異量之摘要結果(DataSet2,發展集自動轉寫文件)

4.4.6 聲學特徵：能量（Energy）

能量可用來表示一個人說話的音量（說話的大、小聲），常被視為一種可利用的重要資訊。因為，當語者在敘述一件情事時，通常會刻意地提高音量來表示強調關鍵字或是說話的內容。論文中，能量分析是使用 Snack 軟體[Snack]，求取每一個訊號音框的對數能量（Log-energy）。每一個詞所對應的能量值，可參考第二章式(2-15)。

以下的實驗，由於人工轉寫文件並沒有每個詞於語音訊號的起始時間與結束時間的資訊，吾人將語音與人工轉寫文件經對齊(Alignment)後，計算文件中每一個詞所對應的音高值。

以下的實驗，人工轉寫文件中每一個詞所對應的能量值，是將語音與人工轉寫文件經對齊(Alignment)後而得到。下面，我們分別採用下列三種方式來估測文句事前機率：

a. 文句平均能量：是以計算每一文句的平均能量作為文句的前事機率。平均能量的計算方式如下：

$$AvgEnergy(S_i) = \frac{1}{N_i}\sum_{n=1}^{N_i} e(w_n) \quad (4\text{-}49)$$

其中，$e(w_n)$ 為詞 w_n 的能量值；N_i 為文句 S_i 的長度。文句的事前機率可表示為：

$$P(S_i) = \frac{AvgEnergy(S_i)}{\sum_{S_z \in D} AvgEnergy(S_z)} \quad (4\text{-}50)$$

當某一文句 S_i 的平均能量較文件中其他文句高，表示此文句可能是較重要的。

測試語料 DataSet1 **實驗結果：**

圖 4.34 與圖 4.35 為發展集中人工轉寫文件與自動轉寫文件的摘要結果。由人工轉寫文件與自動轉寫文件的結果來看，於低摘要比例 10%、20%時，正確率有些許的提昇。在自動轉寫文件中，當摘要比例為 10%且 γ 值為 150 時，正確率可由 0.3084 提昇至 0.3212，有 1%以上的進步；當摘要比例為 20%且 γ 值為 150 時，正確率可由 0.3467 提昇至 0.3666，亦約有 2%的進步。

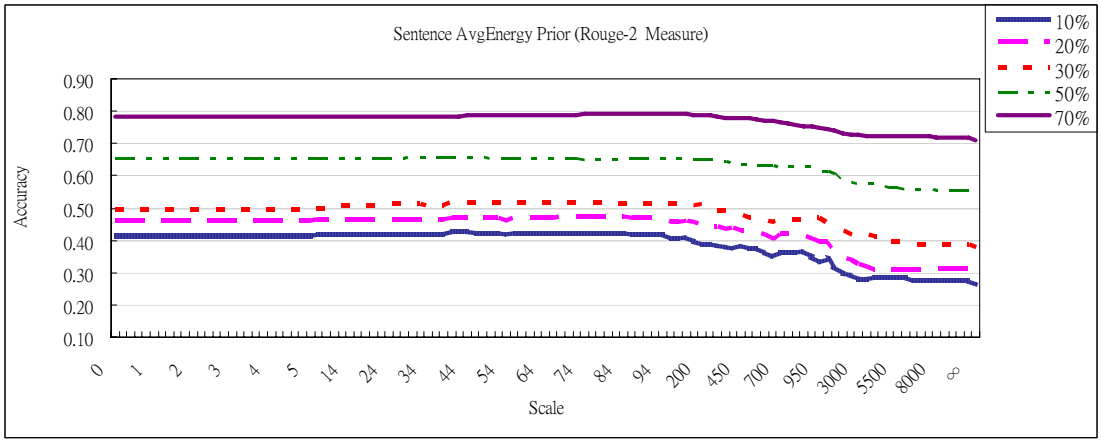


圖 4.34 隱藏式馬可夫模型結合平均能量值之摘要結果(DataSet1,發展集人工轉寫文件)

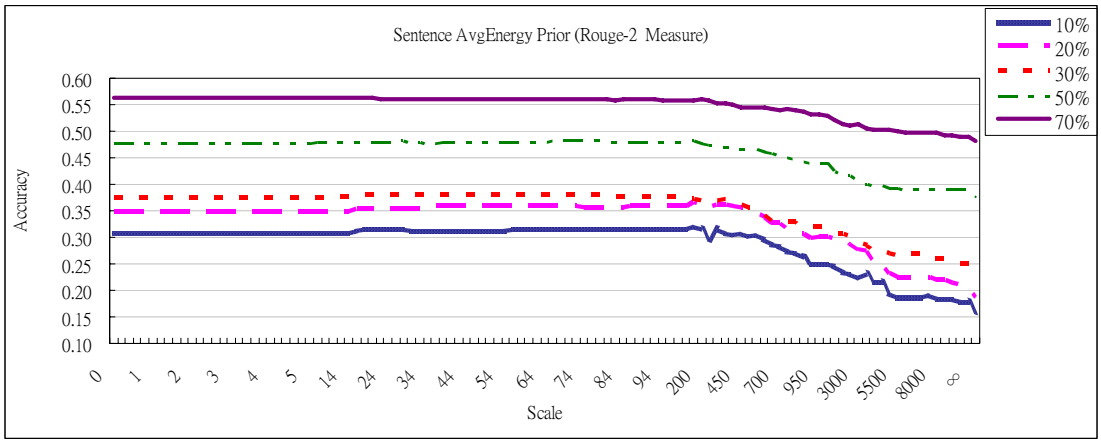


圖 4.35 隱藏式馬可夫模型結合平均能量值之摘要結果(DataSet1,發展集自動轉寫文件)

測試語料 DataSet2 實驗分析：

圖 4.36 及圖 4.37 為人工轉寫文件及自動轉寫文件結果。由結果中可以看出不論於人工轉寫文件或自動轉寫文件上，不同摘要比例之正確率皆有些微的提昇，但是於低摘要比例 10%時，正確率有較明顯的進步(自動轉寫文件於γ值為 78 時，正確率由 0.2089 提昇至 0.2165)。同時，比較圖 4.36 及圖 3.37 中摘要正確率提昇的幅度，人工轉寫文件的摘要結果之正確率有比較明顯的提昇。

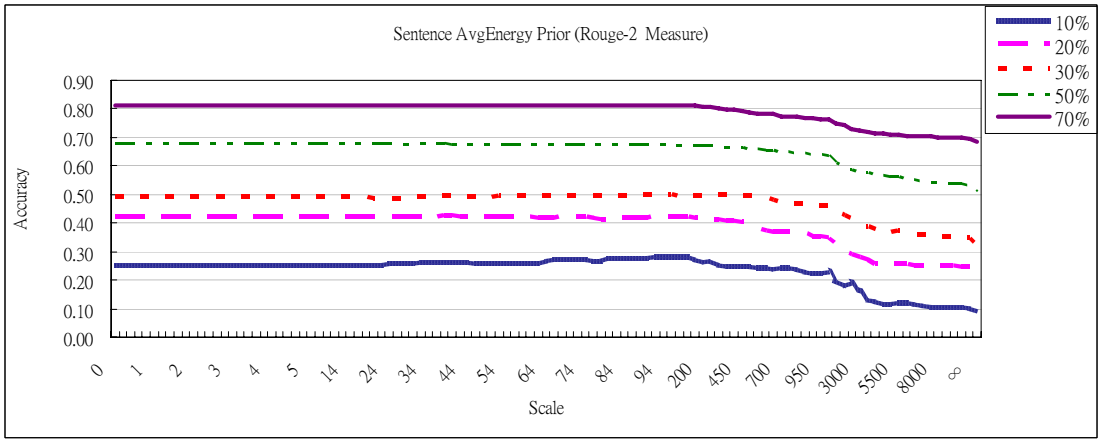


圖 4.36 隱藏式馬可夫模型結合平均能量值之摘要結果(DataSet2,發展集人工轉寫文件)

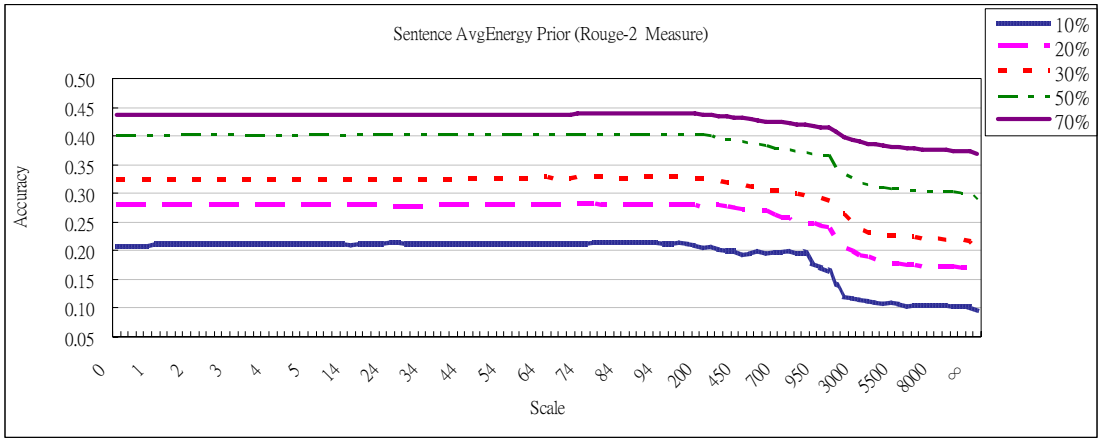


圖 4.37 隱藏式馬可夫模型結合平均能量值之摘要結果(DataSet2,發展集自動轉寫文件)

從 DataSet1 及 DataSet2 的實驗結果，可初步地驗證此摘要特徵對於人工轉寫文件與自動轉寫文件之摘要結果的提昇有是有幫助的。

b. 文句的能量熵值（Energy Entropy）：計算方式與 4.4.5 小節中音高熵值的計算相同。利用熵值來表示每一文句中能量的分佈，當某一文句中能量值的亂度較大，表示文句中可能含有重要的詞或關鍵字，因此以音量的高低來強調它；反之，當某一文句中能量值的亂度較小，表示文句音量無高低的變化，可能為較不重要的文句。

文句能量熵值的計算如前面所述，同樣先對句中每個詞的能量值進行量化的處理，然後以式(4-44)及式(4-45)來計算正規化能量熵值 $\varepsilon\_Power(S_i)$。最後，文句的事前機率可表示為：

$$P(S_i)=\frac{\varepsilon\_Power(S_i)}{\sum_{S_z\in D}\varepsilon\_Power(S_z)} \tag{4-51}$$

測試語料 DataSet1 實驗結果：

圖 4.38 與圖 4.39 為人工轉寫文件及自動轉寫文件的摘要結果。從圖 4.38 的結果來看，此摘要特徵對於人工轉寫文件摘要結果並沒有明顯的提昇。但是

在自動轉寫文件中低摘要比例 10%、20%時，正確率明顯有些許的提昇。在自動轉寫文件中，當摘要比例為 10%且γ值為 6 時，正確率可由 0.3084 提昇至 0.3251；當摘要比例為 20%且γ值為 5 時，正確率可由 0.3467 提昇至 0.3614，正確率皆有 1%以上的進步。

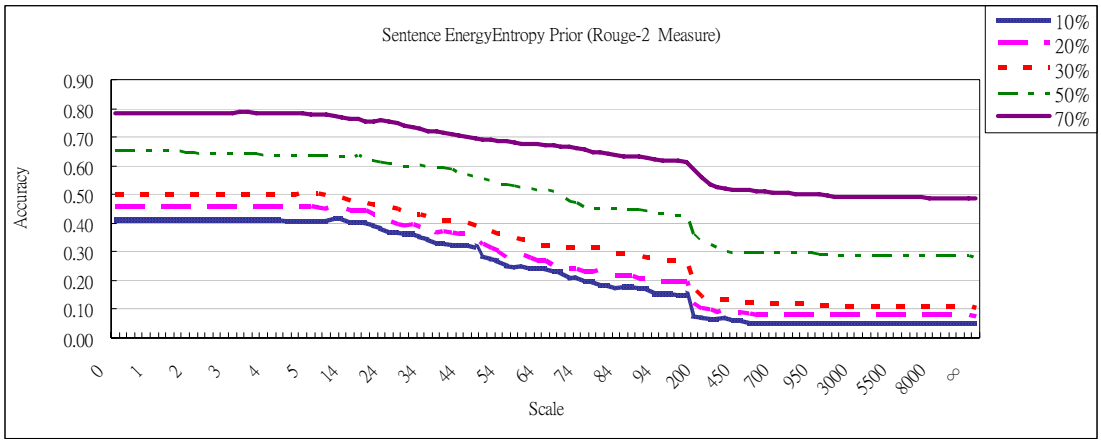


圖 4.38 隱藏式馬可夫模型結合能量熵值之摘要結果(DataSet1,發展集人工轉寫文件)

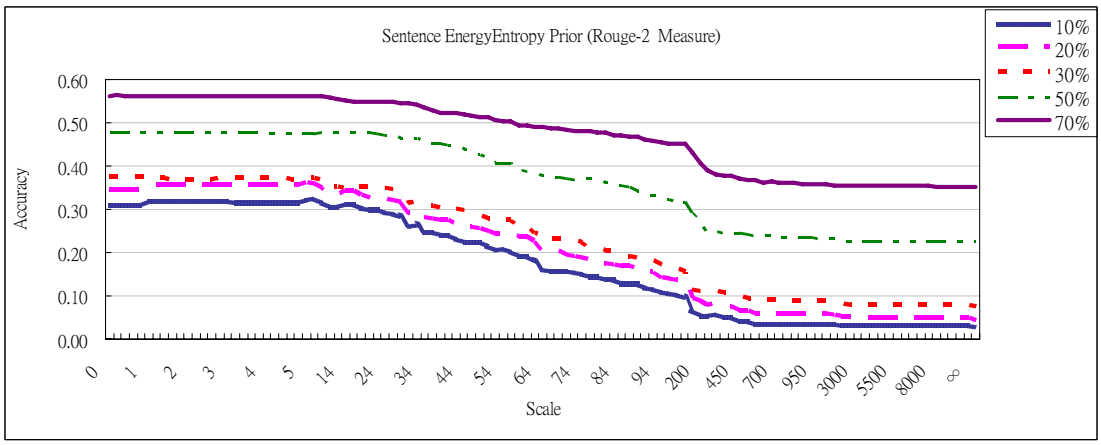


圖 4.39 隱藏式馬可夫模型結合能量熵值之摘要結果(DataSet1,發展集自動轉寫文件)

測試語料 DataSet2 實驗分析：

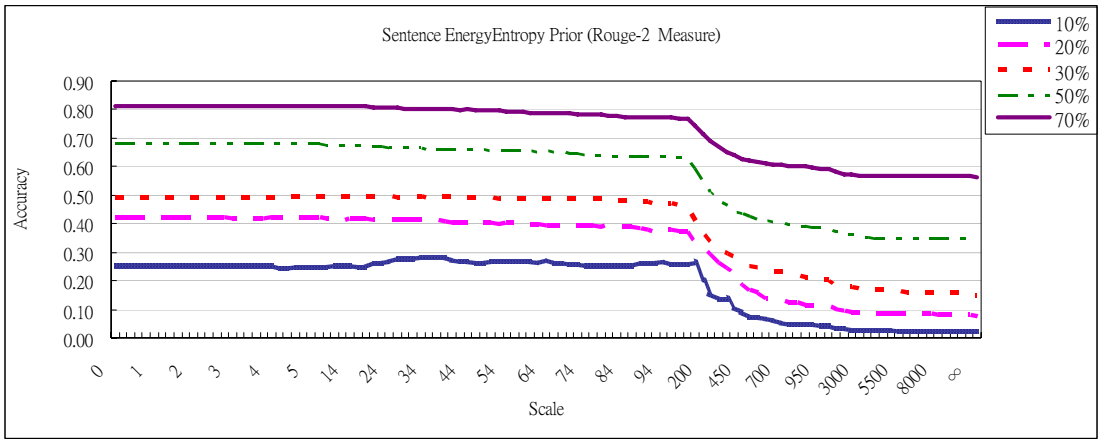


圖 4.40 隱藏式馬可夫模型結合能量熵值之摘要結果(DataSet2,發展集人工轉寫文件)

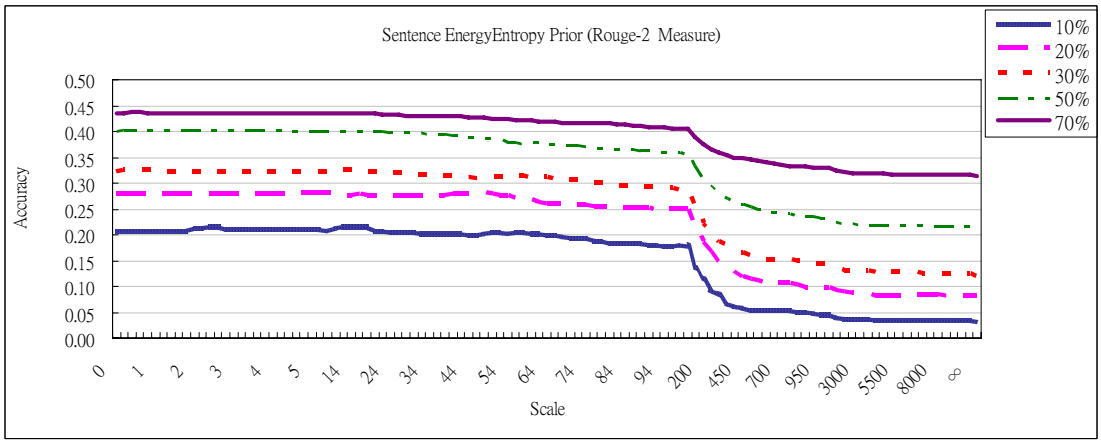


圖 4.41 隱藏式馬可夫模型結合能量熵值之摘要結果(DataSet2,發展集自動轉寫文件)

圖 4.40 為人工轉寫文件摘要結果，可以看出於低摘要比例 10%時，正確率有較明顯的提昇，於γ值為 32 時由 0.2525 提昇至 0.2836，約 3%的進步。圖 4.41 為自動轉寫文件結果，同樣可以看出於低摘要比例 10%及 20%下有較明顯的進步。當摘要比例 10%時，於γ值為 12 正確率由 0.2089 提昇至 0.2163；當摘要比例為 20%且γ值為 5 時，正確率由 0.2778 提昇至 0.2810。

比較以平均能量值與能量熵值來估測文句事前機率的結果，於正確率的提昇上有類似的結果，皆於低摘要比例 10%下有最明顯的提昇，同時於 DataSet2 同樣在人工轉寫文件較自動轉寫文件之正確率提昇的幅度較大一些。

c. 文句的能量變異量（Variance）：與 4.4.5 小節中音高變異量的概念及計算方法相同。以文句中能量的變異量來表示文句能量的分散情形，當作文句是否被強調或不被強調的資訊。每一文句的能量變異量計算方式為：

$$EnergyVar(S_i)=\frac{1}{N_i}\sum_{w_n\in S_i}\left(e(w_n)-AvgEnergy_{S_i}\right)^2 \qquad (4\text{-}52)$$

$EnergyVar(S_i)$ 表示文句 S_i 的能量變異量，$e(w_n)$ 為詞 w_n 的能量值，$AvgEnergy_{S_i}$ 為文句 S_i 的平均能量值。

測試語料 DataSet1 實驗結果：

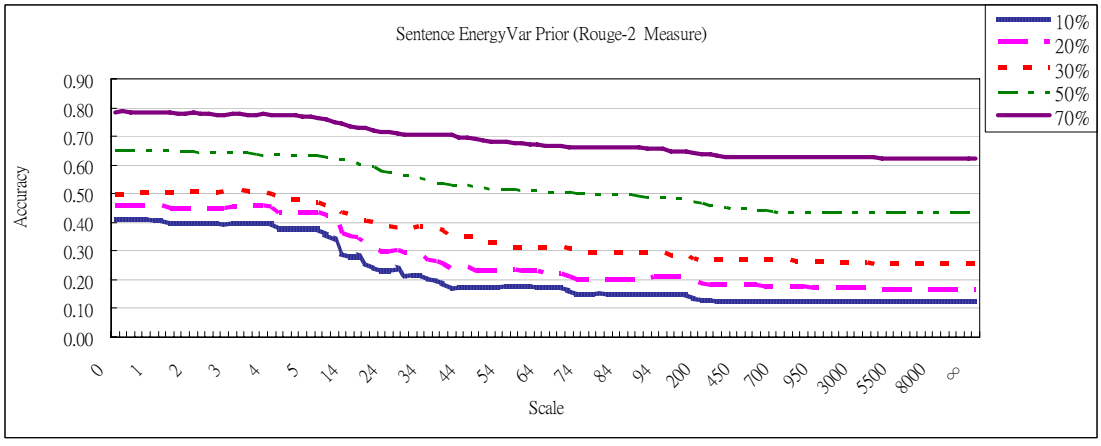


圖 4.42 隱藏式馬可夫模型結合能量變異量之摘要結果(DataSet1,發展集人工轉寫文件)

圖 4.42 與圖 4.43 為人工轉寫文件及自動轉寫文件的摘要結果。從人工轉寫文件結果，發現使用此摘要特徵來估測文句事前機率，對於人工轉寫文件低摘要比例結果的提昇沒有幫助。不過，自動轉寫文件的低摘要比例 10%、20%結果，正確率有些許的提昇，當摘要比例為 10%時，正確率由 0.3084 提昇至 0.3231；當摘要比例為 20%時，正確率可由 0.3467 提昇至 0.3584，正確率有 1%以上的進步。

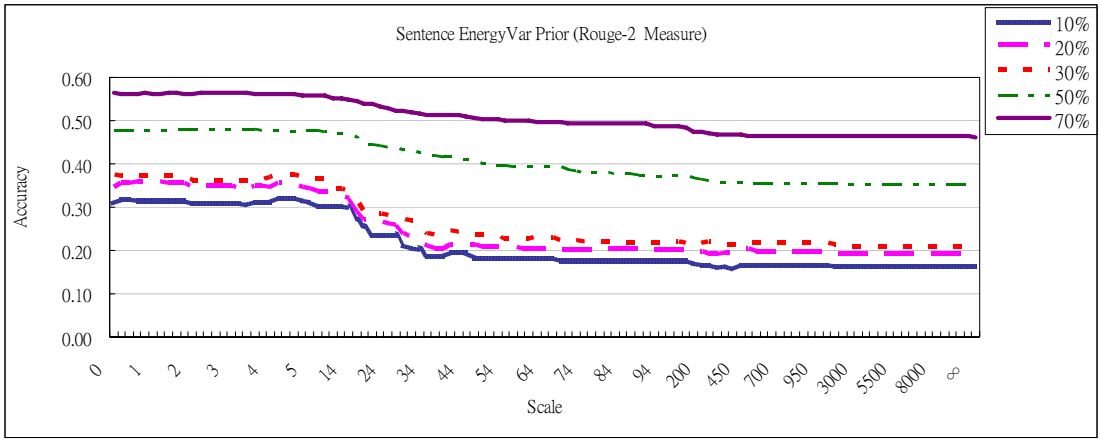


圖 4.43 隱藏式馬可夫模型結合能量變異量之摘要結果(DataSet1,發展集自動轉寫文件)

測試語料 DataSet2 實驗分析：

圖 4.44 為人工轉寫文件摘要結果，摘要結果於低摘要比例 10%時，正確率於γ值為 28 時由 0.2525 提昇至 0.2843，有 3%以上的明顯提昇。但是於其他摘要比例下，正確率僅有極些微的提昇或甚至沒有提昇。圖 4.45 為自動轉寫文件結果，可以於低摘要比例 10%及 20%下發現較明顯的進步。當摘要比例為 10%時，於γ值為 3.2 正確率由 0.2089 提昇至 0.2178；當摘要比例為 20%且γ值為 3 時，正確率僅由 0.2778 提昇至 0.2844。

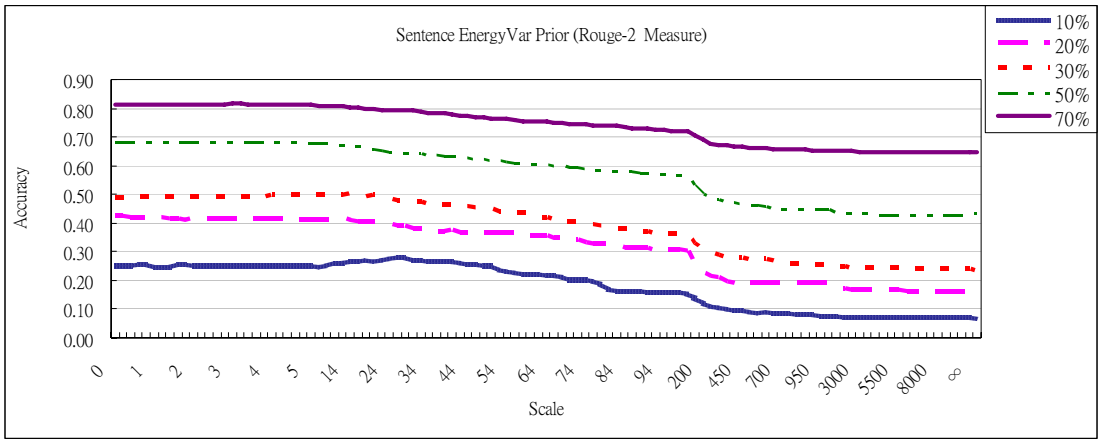


圖 4.44 隱藏式馬可夫模型結合能量變異量之摘要結果(DataSet2,發展集人工轉寫文件)

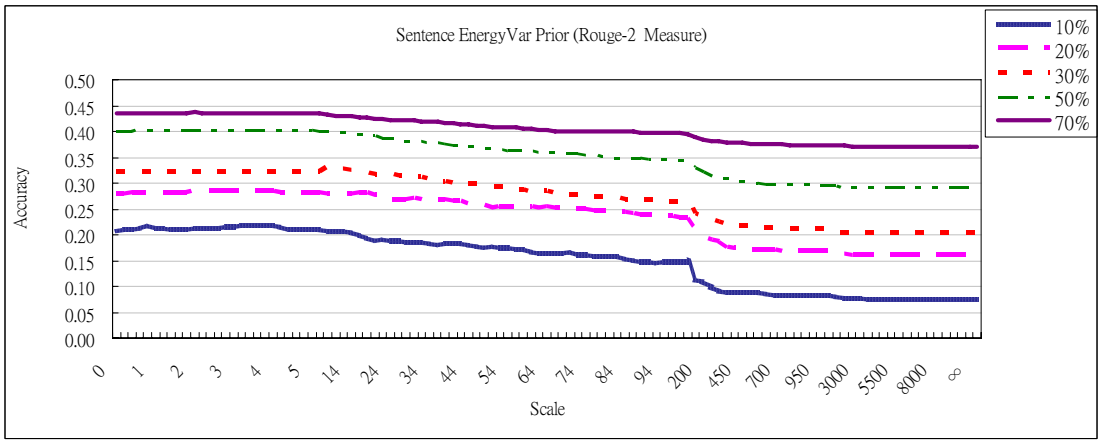


圖 4.45 隱藏式馬可夫模型結合能量變異量之摘要結果(DataSet2,發展集自動轉寫文件)

4.4.7 結論

本小節中，嘗試使用了十種不同的摘要特徵來估測文句的事前機率，以實驗的摘要結果進行一連串的分析與討論。從實驗結果的觀察有下列三點的結論：1. 文句結構特徵－文句位置非常適用於新聞測試語料中，以其來估測文句的事前機率可大幅提高摘要正確率，正確率較僅使用隱藏式馬可夫模型或文句事前機率所產

生的結果要好許多。2. 辨識信心度及聲韻特徵(音高與能量)於自動轉寫文件低摘要比例下，摘要結果有明顯的幫助，甚至於人工轉寫文件上也有很好的效果。3. 文句的事前機率估測可有效的提昇摘要的正確率，尤其是低摘要比例的摘要結果。因此，如何使用不同的摘要特徵來估測文句事前機率，以及如何調整最佳的 γ 值來結果文句事前機率與文句生成模型，都是未來可以再進一步研究的！

本論文初步探討摘要特徵於機率生成式摘要模型的應用，使用不同的摘要徵估測文句事前機率，並與文句生成模型結果觀察對於摘要正確率的影響。吾人於新聞測試語料中進行實驗結果分析與討論，相同的摘要特徵於其他不同的測試語料中，可能會有不一樣的結果。