第 4 章 實驗語料與實驗

4.1 實驗語料

實驗語料蒐集自 News 98 新聞網 [News 98] ，包含 2001 年 8 月 1 日至 8 月 24 日中午 12:00~13:00 的 FM 廣播新聞，相關統計資料如表 4. 1 所示：

表 4. 1 News 98 廣播新聞相關統計

新聞時間	2001 年 8 月 1 日~2001 年 8 月 24 日
新聞數	200 則
總長度	1.61 小時
平均每則新聞長度	28.96 秒
總大小（人工轉寫）	約 31 仟字
平均每則新聞大小（人工轉寫）	約 157 字

這 200 則廣播新聞，有自動轉寫（Automatic Transcription）與人工轉寫（Human Transcription）兩種資料集，自動轉寫部分包含兩個辨識結果，相關資訊如表 4. 2 所示：

表 4. 2 自動轉寫相關資訊

語音辨識自動轉寫	
經詞圖（Word Graph）搜尋後的辨識結果，以 SP_WG 代表	加上非監度式語者調適（+MLLR）的辨識結果，以 SP_Adapt 代表
辨識率為 84.11%	辨識率為 84.64%

自動摘要評估的標準答案部分，請三位國立台灣大學文學院大三以上的學生，分別對 200 則廣播新聞的人工轉寫做摘要，摘要的結果分為句排名（Extraction）與重寫（Abstraction）兩種 [何遠 2003]。

(a) 以句排名標準答案為例：

（1）經發會兩岸組達成鬆綁戒急用忍共識

（2）行政院也將大幅開放兩岸經貿政策

（5）一向主張戒急用忍的經建會主委陳博志今天表示

（7）如果國內企業需要大陸市場來成為世界領導廠商的話

（8）他絕對樂觀其成

（9）不過陳博志也強調

（3）經發會分組會議達成的鬆綁共識不是毫無設限

（4）應該有總量管理的觀念

（6）他也認為對大陸投資不應該太多

此表示第一句重要性第 1、第二句重要性第 2、第三句重要性第 5 … 以此類推。此種人工標準的優點在於可容易的可選出任意比例的摘要結果做為標準答案。

(b) 以重寫標準答案為例：

經發會達成鬆綁戒急用忍共識，行政院也將大幅開放兩岸經貿政策。經建會主委陳博志表示鬆綁並非毫無設限，應有總量管理觀念。

此種人工標準答案可讀性較佳。

此外為了訓練隱藏式馬可夫模型與主題混合模型的參數，將此 200 則廣播新聞按時間先後順序分為發展集（Development Set）及測試集（Test Set）兩部分。其中發展集用來訓練參數，而測試集用以算出最後的摘要結果，並與其他摘要模型做比較，發展集與測試集的相關資訊如表 4.3 所示：

表 4.3 發展集與測試集大小

語料庫	新聞數
發展集（Development Set）	100 則
測試集（Test Set）	100 則
總共　200 則	

4.2 斷詞與統計式語言模型

斷詞（Tokenization）是利用詞典，將一個字串中的文字，比對詞典內的詞來當做斷詞的依據。舉例來說，就是將下列中字句：

“經發會兩岸組達成鬆綁戒急用忍共識”

斷為

“經發會”“兩岸”“組”“達成”“鬆綁戒急用忍”“共識”

在斷詞演算法，本實驗以長詞為優先，如“鬆綁”“戒急用忍”與“鬆綁戒急用忍”皆為詞典內之詞的話，則斷詞演算法將輸出“鬆綁戒急用忍”。

經由第 3 章的討論，在隱藏式馬可夫模型與主題混合模型中，需要訓練語料庫，用以得到統計式語言模型與文件-標題查詢範例的訓練集合，本實驗採用中央通訊社（Central News Agency）[中央通訊社] 在西元 2001 年 08 月且型態屬於故事（type="story"）的文字新聞做為訓練語料庫，每一篇新聞皆含有文件與標題兩部份，其內容包括國內外及大陸文教、交通、社會、財經、國會、影劇、醫藥衛生、體育及地方新聞，相關資訊如表 4. 4 所示：

表 4. 4 中央社文字語料相關統計

新聞時間	2001 年 8 月
新聞數	14,178 則
總大小	約 709 萬字
平均每則新聞長度	約 500 字

在訓練統計式語言模型時，一般第一步是將語料做斷詞，再用斷詞過的語料來訓練語言模型，最後得到以詞為基礎的語言模型。此種做法的優點在於字句的主旨由詞的詞義而非字的字義組成，因此以詞為基礎的語言模型可以得到以字為基礎的語言模型，更豐富的語言資訊。

然而斷詞也有其不足之處：其一，因為字組成詞的變化程度相當大，是以一個字句難免會有多種斷詞方式；其二，未知詞（Out-of-Vocabulary, OOV）的問

題，例如專有名詞、地名、人名等，不在詞典裡的機率通常很高。這兩個不足之處是目前以斷詞為基礎的研究所必需解決的問題。

語言模型的訓練部份本實驗使用 SRI 國際（SRI International）語音科技及研究實驗室（Speech Technology and Research Laboratory）所提供的 SRI 語言模型工具箱（SRI Language Modeling Toolkit）[SRI Toolkit]。 SRI 語言模型工具箱是一套用來訓練或應用統計式語言模型（Statistical Language Model）的工具程式組合。

本實驗所使用的命令參數如下：

ngram-count -text 斷詞後的文字語料 -lm 輸出語言模型檔

語言模型的平滑化（Smoothing）演算法則是使用預設的 Good-Turning 演算法。

4.3 自動摘要評估

本實驗的評估方式有三種：第一為餘弦（Cosine）評估：藉由計算自動摘要與人工摘要的相關度為標準；第二為 ROUGE（Recall-Oriented Understudy for Gisting Evaluation）評估 [Lin 2003]：其為一召回率導向的自動摘要評估方式；第三為平均精確度（Mean Average Precision）評估：藉由計算在相同摘要比例下，自動摘要與人工摘要的平均精確度為標準，然因自動轉寫與人工轉寫斷句的不同，造成自動摘要與人工摘要的字句無法判斷是否相關（即所摘要出的字句，在人工句排名內），是以本論文提出以改良型字錯誤率（modified Character Error Rate, m-CER）為基礎的相關判斷準則，只要符合此一準則就視自動摘要的字句與人工摘要的字句是相關而予以計算精確度，茲將其分別說明如下各小節。

4.3.1 餘弦評估

在此自動摘要的評估方式上，是以計算與人工摘要結果的相關度做為標準。由於人工摘要的產生有兩種。第一種為摘錄式(Extractive)摘要(或稱為句排名方式)：對於每則新聞中的字句依其對於整則新聞的重要性作排名，因此很容易的可選出

任意比例的摘要結果，如 20%、30%、50%、70% 等摘要比例；第二種為非摘錄式（Non-Extractive or Abstractive）摘要（或稱為重寫摘要的方式）：其摘要比例就無法作任意的修改。

假設 A_d 代表對某篇文件 d 自動摘要結果、$E_{n,d}$ 代表第 n 個人對文件 d 以摘錄式摘要結果、$R_{n,d}$ 代表第 n 個人對文件 d 重寫摘要結果、以及 $m\%$ 為可能的摘要比例，則對所有新聞自動摘要的正確率被定義為 [Orasan 2002; 何遠 2003]：

$$ACC(m\%) = \frac{1}{D}\frac{1}{N}\sum_{d=1}^{D}\sum_{n=1}^{N}\frac{SIM\left(A_d(m\%), E_{n,d}(m\%)\right) + SIM\left(A_d(m\%), R_{n,d}\right)}{2} \tag{4. 1}$$

其中，$SIM(\cdot,\cdot)$ 為評估自動摘要與人工摘要結果的相關度公式，此公式使用向量空間模型做為相關度評估的方法：將自動摘要與人工摘要結果以向量形式表現，並計算向量間夾角的餘弦值來得到相關度，例如式中的 $SIM\left(A_d(m\%), E_{n,d}(m\%)\right)$ 可表示成：

$$SIM\left(A_d(m\%), E_{n,d}(m\%)\right) = \frac{\vec{V}_{A_d(m\%)} \cdot \vec{V}_{E_{n,d}(m\%)}}{\left|\vec{V}_{A_d(m\%)}\right|\left|\vec{V}_{E_{n,d}(m\%)}\right|} \tag{4. 2}$$

上式中 $\vec{V}_{A_d(m\%)}$ 與 $\vec{V}_{E_{n,d}(m\%)}$ 分別為自動摘要結果 $A_d(m\%)$ 與人工摘錄式摘要結果 $E_{n,d}(m\%)$ 的向量表示式，而向量表示式中的每一維度則是代表著某一個索引特徵 t，t 在摘要中的重要性（或稱權重）$w(t)$，是以它在摘要中出現的詞頻（Term Frequency, TF）與反文件頻率（Inverse Document Frequency, IDF）乘積來表示:

$$w(t) = \left(\frac{f_t}{\max_h f_h}\right)\log\left(N/n_t\right) \tag{4. 3}$$

上式中 f_t 為索引特徵 t 在摘要中出現的次數，$\log(N/n_t)$ 是反文件頻率，N 是所有廣播新聞則數，n_t 是索引特徵 t 出現的廣播新聞則數。

4.3.2 ROUGE 評估

ROUGE（Recall-Oriented Understudy for Gisting Evaluation）是召回率導向的自動

摘要評估方式 [Lin 2003]。此評估計算自動摘要與人工摘要重疊單位元的次數，單位可為 *N*-連語言模型（*N*-gram Language Model）、詞順序（Word sequences）與詞對（Word pairs）。

如 *N*-連語言模型單位元的評估，如下所示：

$$ROUGE-N=\frac{\sum_{S\in\{人工摘要標準答案\}}\sum_{N-連語言模型\in S}Count_{match}(N-連語言模型)}{\sum_{S\in\{人工摘要標準答案\}}\sum_{N-連語言模型\in S}Count(N-連語言模型)} \tag{4. 4}$$

$Count_{match}$ 代表 *N*-連語言模型共同出現在自動摘要與人工摘要標準答案的最大次數，在人工摘要標準答案部份本實驗僅使用句排名。此外本實驗以 ROUGE-2 ，即雙連語言模型為評估標準。

4.3.3 平均精確度評估

在此摘要評估方式上，是以不同摘要比例 *m*% 下的平均精確度（Mean Average Precision, MAP）來評估。假設 A_d 代表對某篇文件 *d* 自動摘要的句排名、$E_{n,d}$ 代表第 *n* 個人對文件 *d* 做的句排名以及 *m*% 為可能的摘要比例（用以保留前 *m*% 分數較大的字句來計算平均精確度），平均精準度計算公式如下：

$$\mathrm{mAP}(\mathrm{m\%})=\frac{1}{D}\frac{1}{N}\sum_{d=1}^{D}\sum_{n=1}^{N}\frac{1}{R_d}\sum_{s=1}^{R_d}\frac{s}{r_{d,s,n}} \tag{4. 5}$$

R_d 為 A_d (m%) 與 $E_{n,d}$ (m%) 相關的字句個數，$r_{d,s,n}$ 為 A_d (m%)與 $E_{n,d}$ (m%) 相關的第 *s* 句從 A_d 前頭（Top）數來的位置。

在自動轉寫上，因斷句的不同而無法與人工轉寫的字句做一對一對應，是以無法決定所摘要出的字句在人工標準答案的字句上，也就是說式(4.5)中無法決定相關，是以本論文定義相關判斷準則來解決此一問題。

首先介紹 Levenshtein Distance（LD），LD 為二個字串之間最短的編輯距離。最早由俄國科學家 Vladimir Levenshtein 在 1966 所提出 [Levenshtein 1966]，又稱為編輯距離（Edit Distance）。假設 *I* 是測試字串，*R* 是正確字串，並以單字

（Character）為索引。LD 演算法計算字串 *I* 轉換為 *R* 所需的刪除、插入、替代編輯次數。例如：

- *I* 是”陳水扁” 且 *R* 是 “陳水扁” 則 LD(*I*,*R*)=0 ，因在此範例中無任何轉換
- *I* 是”陳水篇” 且 *R* 是 “陳水扁” 則 LD(*I*,*R*)=1 ，因在此範例中只需一個替代（改變篇到扁）就足以將 *I* 轉換到 *R*

是以經由 LD 演算法，可以得到<插入，刪除、替代、正確>的個數，且愈大的 LD，代表字串的相異度愈大。目前 LD 演算法，已大量運用於拼字檢查 [黃上銘等 2003]、語音辨識、DNA 分析等應用上。

接著定義改良型字錯誤率（modified Character Error Rate, m-CER）如下，假設 *I* 是測試字串、*R* 是正確字串，並以單字（Character）為索引，<插入，刪除、替代、正確>為經由 LD 演算法所計算出的個數，則

$$m-CER=\frac{(\text{插入}+\text{刪除}+\text{替代})-(diff)}{\text{R的字數}} \tag{4.6}$$

$$diff=\begin{cases}\text{句長差} & \text{if 句長差} \le \text{正確個數}\\ \text{正確個數} & \text{if 句長差} > \text{正確個數}\end{cases} \tag{4.7}$$

其中 *diff* 主要用來解決，因句長差（$\text{abs}(|R|-|I|)$）所產生的錯誤個數。

例如：

I ：臺北電信網路展今天開幕

R ：臺北電信網路展今天開幕一連展出四天

LD = <插入，刪除，替代、正確> = < 6,0,0,11>

如不使用句長差，則原字錯誤率高達 $CER=\frac{(6+0+0)}{11}=\frac{6}{11}=0.5454$

然而經由計算 *diff* 後，則 *m-CER* =0.0

經由上述討論，式(4.5)評估自動轉寫與人工轉寫，字句相關的定義如下：

假設 A_d 代表對某篇文件 *d* 自動摘要的句排名、$E_{n,d}$ 代表第 *n* 個人對文件 *d* 做

的句排名、$m\%$ 為可能的摘要比例、$m-CER_{\min}$ 代表 $A_d(m\%)$ 中的某一句 I，對 $E_{n,d}(m\%)$ 每一字句 R ，計算 $m\text{-}CER$ 所得的最小值

如果 $m\text{-}CER_{\min}<0.3$ 則 相關

4.4 實驗結果

基礎實驗（Baseline）有三個，分別為 VSM（向量空間模型）、RM（相關評估）與 LSA（潛藏語意分析），本論文提出的模型計有 embedded LSA（嵌入式潛藏語意分析）、HMM-Type1（隱藏式馬可夫模型-型一）、HMM-Type2（隱藏式馬可夫模型-型二）、TMM(D|T)（主題混合模型，監督式訓練）、TMM(T|T)（主題混合模型，非監督式訓練）。

根據台大與中研院近年來在中文語音文件檢索（Spoken Document Retrieval）領域的研究成果 [Chen *et al.* 2002]，若使用音節（Syllable）或字（Character）為單位的索引特徵（Indexing Feature）組合，往往可以有比詞更好的檢索表現。因此，本實驗將探討不同的特徵單位（Feature Unit）對自動摘要的影響。首先，以所帶資訊的豐富程度來看，顯然詞所帶的資訊量是最多的，其次是字，然後為音節 [Lee *et al.* 2003]，如表 4. 5 所示：

表 4. 5 音節、字、詞所帶資訊的比較

單位	所帶資訊的比較
音節（Syllable）	低
字（Character）	中
詞（Word）	高

此外在音節中，因為中文的同音異義字（Homonym）很多，一個音節可能對應好幾個字，所以音節的混淆度（Ambiguity）很大。然而雖然音節所帶的資訊少、混淆度大，但是如果將兩個音節連結起來成為一個雙音節（Syllable Pair），那麼所帶的資訊就變得豐富很多 [Lee *et al.* 2003] ，綜合上述討論，本實驗除了使用單詞為特徵單位來做自動摘要外，並將實驗使用以雙字、雙音節組合為特徵

單位，來觀察自動摘要的正確率。

其中雙字、雙音節都是互相重疊（Overlapping）的。例如一字句由 10 個音節所組成：

$$\left(S_1, S_2, ..., S_{10}\right)$$

則其雙音節特徵即為

$$(S_1, S_2), (S_2, S_3), ..., (S_9, S_{10})$$

同樣地互相重疊的雙字，定義類似。

4.4.1 餘弦評估

在本小節所有表格中，TMM 欄內的數字代表所使用的潛藏主題數。如表 4. 6 TMM P(D|T) 64 即代表使用 64 個潛藏主題數。此外 HMM 與 TMM 的數據，均使用期望值最大化演算法，自動調整參數與訓練模型。

由表 4. 6~表 4. 8 以單詞為特徵單位的結果顯示，在低摘要比例（20%）時，TMM 與 HMM-Type2 其結果不論在人工轉寫或自動轉寫上，均優於其他摘要模型，並以 TMM 為較佳。進一步比較表 4. 6（人工轉寫）、表 4. 7（自動轉寫 SP_WG）可發現，所有模型的正確率均下降，這可由兩個方面來說明：其一，因為為語音辨識會產生錯誤的辨識結果，錯誤結果當然會影響摘要的正確率；其二，人工轉寫的斷句與標準答案（句排名）是一致的，而語音辨識結果的斷句目前主要是以靜音（silence）的長度為標準，因此最後斷開成的字句會與標準答案的字句有所差異，所以也會導致摘要正確率的下降。此外比較表 4. 7、表 4. 8（自動轉寫 SP_Adapt）可發現，隨著辨識率的上升，其結果也有所提升。

表 4.6 摘錄方法比較(評估方式：餘弦，特徵單位：詞，人工轉寫)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 64	TMM P(T|T) 32
20%	0.4473	0.4551	0.3599	0.4439	0.3639	0.4577	0.4631	0.4654
30%	0.5071	0.5108	0.4413	0.4983	0.4132	0.5278	0.5296	0.5265
50%	0.6365	0.6269	0.6098	0.6334	0.5918	0.6260	0.6302	0.6287
70%	0.6961	0.6956	0.6864	0.7051	0.6729	0.6951	0.6960	0.6952

表 4.7 摘錄方法比較(評估方式：餘弦，特徵單位：詞，自動轉寫 SP_WG)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 64	TMM P(T|T) 64
20%	0.3573	0.3586	0.2971	0.3652	0.3112	0.3719	0.3771	0.3750
30%	0.4074	0.4052	0.3487	0.4020	0.3570	0.4160	0.4188	0.4226
50%	0.5397	0.5130	0.4863	0.5102	0.5098	0.5285	0.5279	0.5289
70%	0.5839	0.5608	0.5540	0.5747	0.5738	0.5782	0.5777	0.5771

表 4.8 摘錄方法比較(評估方式：餘弦，特徵單位：詞，自動轉寫 SP_Adapt)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 32	TMM P(T|T) 32
20%	0.3630	0.3603	0.3138	0.3726	0.3206	0.3712	0.3718	0.3738
30%	0.4269	0.4195	0.3592	0.4142	0.3697	0.4234	0.4243	0.4230
50%	0.5413	0.5132	0.4937	0.5176	0.5057	0.5296	0.5311	0.5302
70%	0.5874	0.5632	0.5558	0.5788	0.5760	0.5775	0.5791	0.5783

由表 4.10~表 4.11 以雙音節為特徵單位的結果顯示，在低摘要比例（20% 與 30%）時，TMM 與 HMM-Type2 其結果不論在人工轉寫或自動轉寫上，均優於其他摘要模型。此外比較表 4.10、表 4.11 可發現，隨著辨識率的上升，其結果也有所提升。

表 4. 9 摘錄方法比較(評估方式：餘弦，特徵單位：雙音節，人工轉寫)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 4	TMM P(T|T) 4
20%	0.4485	0.4425	0.3688	0.4457	0.3613	0.4618	0.4618	0.4525
30%	0.5176	0.5075	0.4578	0.5159	0.4352	0.5254	0.5283	0.5246
50%	0.6326	0.6300	0.6101	0.6413	0.5997	0.6386	0.6387	0.6380
70%	0.7054	0.7066	0.6882	0.7116	0.6692	0.7007	0.7014	0.7018

表 4. 10 摘錄方法比較(評估方式：餘弦，特徵單位：雙音節，自動轉寫 SP_WG)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 16	TMM P(T|T) 16
20%	0.3665	0.3686	0.3054	0.3567	0.3266	0.3811	0.3811	0.3833
30%	0.4128	0.4121	0.3608	0.4038	0.3783	0.4167	0.4150	0.4138
50%	0.5287	0.5125	0.4931	0.5268	0.5017	0.5318	0.5287	0.5329
70%	0.5851	0.5782	0.5640	0.5881	0.5633	0.5889	0.5879	0.5887

表 4. 11 摘錄方法比較(評估方式：餘弦，特徵單位：雙音節，自動轉寫 SP_Adapt)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 16	TMM P(T|T) 16
20%	0.3633	0.3675	0.2966	0.3646	0.3276	0.3833	0.3833	0.3833
30%	0.4157	0.4079	0.3555	0.4136	0.3824	0.4296	0.4268	0.4273
50%	0.5299	0.5185	0.4917	0.5315	0.5119	0.5391	0.5371	0.5407
70%	0.5854	0.5793	0.5625	0.5880	0.5694	0.5912	0.5899	0.5908

由表 4. 12~表 4. 14 以雙字為特徵單位的結果顯示，向量空間模型為基礎的摘要模型（VSM 與 RM）有較佳的結果。此外在高摘要比例（70%）時，embedded LSA 其結果均優於其他摘要模型。

表 4. 12 摘錄方法比較(評估方式：餘弦，特徵單位：雙字，人工轉寫)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 32	TMM P(T|T) 64
20%	0.4596	0.4538	0.3781	0.4381	0.3458	0.4460	0.4531	0.4489
30%	0.5191	0.5255	0.4487	0.5165	0.4343	0.5107	0.5211	0.5059
50%	0.6407	0.6335	0.6049	0.6430	0.5895	0.6386	0.6399	0.6390
70%	0.7008	0.7024	0.6803	0.7064	0.6669	0.7020	0.7011	0.7012

表 4. 13 摘錄方法比較(評估方式：餘弦，特徵單位：雙字，自動轉寫 SP_WG)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 64	TMM P(T|T) 4
20%	0.3862	0.3879	0.2974	0.3544	0.3231	0.3708	0.3741	0.3737
30%	0.4223	0.4101	0.3586	0.4109	0.3829	0.4168	0.4173	0.4207
50%	0.5331	0.5176	0.5037	0.5293	0.5118	0.5253	0.5258	0.5265
70%	0.5857	0.5729	0.5611	0.5884	0.5657	0.5819	0.5816	0.5821

表 4. 14 摘錄方法比較(評估方式：餘弦，特徵單位：雙字，自動轉寫 SP_Adapt)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 64	TMM P(T|T) 64
20%	0.3784	0.3779	0.3009	0.3508	0.3234	0.3749	0.3794	0.3759
30%	0.4335	0.4240	0.3671	0.4106	0.3864	0.4331	0.4357	0.4304
50%	0.5354	0.5189	0.5044	0.5299	0.5104	0.5297	0.5302	0.5319
70%	0.5882	0.5766	0.5551	0.5920	0.5710	0.5869	0.5869	0.5881

4.4.2 ROUGE 評估

在本小節所有表格中 TMM 欄內的數字，代表所使用的潛藏主題數。此外 HMM 與 TMM 的數據，均使用期望值最大化演算法，自動調整參數與訓練模型。

由表 4. 15~表 4. 17 以單詞為特徵單位的結果顯示，在低摘要比例（30%）時，TMM 與 HMM-Type2，其結果不論在人工轉寫或自動轉寫上，均優於其他摘要

模型，並以 TMM 為較佳。此外在高摘要比例（50% 或 70%）時，embedded LSA 其結果均優於其他摘要模型。

表 4. 15 摘錄方法比較(評估方式：ROUGE，特徵單位：詞，人工轉寫)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 64	TMM P(T|T) 32
20%	0.3697	0.3890	0.2722	0.4046	0.1797	0.3969	0.4070	0.4098
30%	0.4272	0.4623	0.3701	0.4480	0.2175	0.4694	0.4736	0.4676
50%	0.6142	0.6302	0.5892	0.6511	0.4824	0.6110	0.6161	0.6136
70%	0.7221	0.7503	0.7046	0.7666	0.6357	0.7277	0.7293	0.7260

表 4. 16 摘錄方法比較(評估方式：ROUGE，特徵單位：詞，自動轉寫 SP_WG)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 64	TMM P(T|T) 32
20%	0.2208	0.2284	0.1906	0.2758	0.1485	0.2538	0.2590	0.2581
30%	0.2532	0.2648	0.2259	0.2694	0.1762	0.2804	0.2817	0.2766
50%	0.3992	0.3905	0.3565	0.4059	0.3304	0.4004	0.3987	0.3994
70%	0.4665	0.4542	0.4405	0.4831	0.4268	0.4714	0.4694	0.4697

表 4. 17 摘錄方法比較(評估方式：ROUGE，特徵單位：詞，自動轉寫 SP_Adapt)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 32	TMM P(T|T) 32
20%	0.2291	0.2306	0.2041	0.2831	0.1588	0.2609	0.2612	0.2618
30%	0.2745	0.2833	0.2382	0.2849	0.1789	0.2862	0.2889	0.2840
50%	0.4060	0.3929	0.3697	0.4133	0.3248	0.4049	0.4060	0.4052
70%	0.4680	0.4527	0.4441	0.4873	0.4304	0.4707	0.4722	0.4720

由表 4. 18~表 4. 20 以雙音節為特徵單位的結果顯示，在低摘要比例（20% 或 30%），TMM 與 HMM-Type2，其結果不論在人工轉寫或自動轉寫上，大致均優於其他摘要模型。此外在高摘要比例（50% 或 70%）時，embedded LSA 其結果均優於其他摘要模型。

表 4. 18 摘錄方法比較(評估方式：ROUGE，特徵單位：雙音節，人工轉寫)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 4	TMM P(T|T) 16
20%	0.3834	0.3902	0.2941	0.4168	0.1352	0.4273	0.4273	0.4151
30%	0.4493	0.4664	0.3671	0.4848	0.2224	0.4843	0.4864	0.4801
50%	0.6110	0.6284	0.5623	0.6492	0.4728	0.6310	0.6310	0.6334
70%	0.7395	0.7605	0.6828	0.7646	0.6176	0.7398	0.7408	0.7421

表 4. 19 摘錄方法比較(評估方式：ROUGE，特徵單位：雙音節，自動轉寫 SP_WG)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 4	TMM P(T|T) 16
20%	0.2513	0.2590	0.1990	0.2559	0.1730	0.2812	0.2812	0.2826
30%	0.2790	0.2874	0.2216	0.2863	0.1883	0.2933	0.2919	0.2884
50%	0.4041	0.3988	0.3548	0.4248	0.3170	0.4142	0.4135	0.4154
70%	0.4770	0.4779	0.4465	0.4909	0.4087	0.4852	0.4849	0.4855

表 4. 20 摘錄方法比較(評估方式：ROUGE，特徵單位：雙音節，自動轉寫 SP_Adapt)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 4	TMM P(T|T) 16
20%	0.2503	0.2574	0.1962	0.2674	0.1701	0.2809	0.2809	0.2809
30%	0.2839	0.2837	0.2281	0.3030	0.1902	0.3068	0.3027	0.3027
50%	0.4066	0.4065	0.3560	0.4314	0.3245	0.4235	0.4228	0.4253
70%	0.4780	0.4830	0.4458	0.4917	0.4095	0.4843	0.4841	0.4846

由表 4. 21～表 4. 23 以雙字為特徵單位的結果顯示，在低摘要比例（20%）時，向量空間模型為基礎的摘要模型（VSM 與 RM）有較佳的結果。此外在高摘要比例（50% 或 70%）時，embedded LSA 其結果大致均優於其它摘要模型。

表 4. 21 摘錄方法比較(評估方式：ROUGE，特徵單位：雙字，人工轉寫)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 64	TMM P(T|T) 64
20%	0.4132	0.4194	0.2910	0.4091	0.1474	0.4037	0.4110	0.4110
30%	0.4733	0.4979	0.3553	0.5000	0.2343	0.4787	0.4706	0.4712
50%	0.6271	0.6452	0.5575	0.6506	0.4559	0.6418	0.6402	0.6401
70%	0.7361	0.7627	0.6817	0.7586	0.6047	0.7470	0.7403	0.7472

表 4. 22 摘錄方法比較(評估方式：ROUGE，特徵單位：雙字，自動轉寫 SP_WG)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 64	TMM P(T|T) 4
20%	0.2767	0.2804	0.1735	0.2569	0.1673	0.2518	0.2549	0.2578
30%	0.2811	0.2765	0.2065	0.2910	0.1934	0.2814	0.2880	0.2840
50%	0.4134	0.4099	0.3604	0.4215	0.3298	0.4122	0.4126	0.4143
70%	0.4772	0.4800	0.4403	0.4921	0.4143	0.4783	0.4773	0.4786

表 4. 23 摘錄方法比較(評估方式：ROUGE，特徵單位：雙字，自動轉寫 SP_Adapt)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T) 64	TMM P(T|T) 4
20%	0.2685	0.2680	0.1852	0.2589	0.1641	0.2626	0.2651	0.2626
30%	0.2905	0.2931	0.2215	0.2971	0.1969	0.2979	0.3057	0.2979
50%	0.4160	0.4111	0.3647	0.4252	0.3244	0.4170	0.4178	0.4190
70%	0.4798	0.4863	0.4339	0.4953	0.4157	0.4829	0.4829	0.4829

4.4.2 平均精確度評估

在本小節所有表格中，每一欄的 HMM 與 TMM 數據又可分成三列：第一列為在發展集中找的較佳參數，第二列為直接在測試集找的較佳參數，第三列為使用期望值最大化演算法，自動調整參數與訓練模型所得的數據。

例如表 4. 24 中 HMM-Type2 0.40 / 0.95 ，分別代表 λ 為 0.40(發展集)與 0.95

（測試集）；TMM P(D|T) 0.40,16 / 0.95,32 / 64 分別代表 λ 為 0.40 和潛藏主題數為 16（發展集）、λ 為 0.95 和潛藏主題數為 32（測試集）與潛藏主題數為 64。

由表 4. 24~表 4. 26 以單詞為特徵單位的結果顯示（HMM 與 TMM 使用期望值最大化演算法，即第三列數據），在低摘要比例（20% 或 30%）時，TMM 與 HMM-Type2，其結果不論在人工轉寫或自動轉寫上，均優於其它摘要模型，並以 TMM 為較佳。

此外，比較 HMM 與 TMM，於發展集與測試集的數據（即第一列、第二列數據）顯示，因發展集與測試集不一致使得參數的變化不一，在發展集較佳的參數拿到測試集使用時，其結果有所降低。另一方面，期望值最大化演算法所得到的結果，大致與測試集所得到的較佳結果不分軒輊。

表 4. 24 摘錄方法比較(評估方式：MAP，特徵單位：詞，人工轉寫)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1 0.90 / 0.75	HMM-Type2 0.40 / 0.95	TMM P(D|T) 0.40,16 / 0.95,32 / 64	TMM P(T|T) 0.35,08 / 0.95,16 / 32
20%	0.3517	0.3606	0.2653	0.3756	0.1828 0.1928 0.1828	0.3739 0.3877 0.3856	0.3739 0.3888 0.3906	0.3772 0.3881 0.3989
30%	0.5139	0.5185	0.4260	0.5132	0.3158 0.3269 0.3055	0.5264 0.5498 0.5435	0.5264 0.5500 0.5504	0.5283 0.5494 0.5486
50%	0.7456	0.7575	0.7214	0.7518	0.6191 0.6203 0.6135	0.7550 0.7637 0.7621	0.7591 0.7626 0.7589	0.7568 0.7644 0.7591
70%	0.8328	0.8437	0.8201	0.8490	0.7382 0.7403 0.7307	0.8427 0.8487 0.8428	0.8443 0.8489 0.8466	0.8437 0.8502 0.8448

表 4. 25 摘錄方法比較(評估方式：MAP，特徵單位：詞，自動轉寫 SP_WG)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1 0.70 / 0.80	HMM-Type2 0.70 / 0.10	TMM P(D|T) 0.55,64 / 0.10,64 / 64	TMM P(T|T) 0.55,32 / 0.10,64 / 32
20%	0.2683	0.2683	0.2317	0.3050	0.1928 0.2011 0.2011	0.3000 0.3406 0.3272	0.3122 0.3439 0.3289	0.3122 0.3472 0.3306
30%	0.3853	0.3787	0.3430	0.3956	0.3028 0.3067 0.3058	0.4100 0.4592 0.4459	0.4308 0.4597 0.4457	0.4292 0.4607 0.4450
50%	0.6845	0.6706	0.6240	0.6876	0.6088 0.6089 0.6036	0.6978 0.6910 0.6959	0.7025 0.6928 0.7003	0.7020 0.6964 0.6957
70%	0.7674	0.7542	0.7364	0.7690	0.7392 0.7379 0.7289	0.7706 0.7716 0.7771	0.7747 0.7747 0.7818	0.7730 0.7742 0.7774

表 4. 26 摘錄方法比較(評估方式：MAP，特徵單位：詞，自動轉寫 SP_Adapt)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1 0.70 / 0.95	HMM-Type2 0.45 / 0.4	TMM P(D|T) 0.55,64 / 0.55,64 / 64	TMM P(T|T) 0.40,04 / 0.40,04 / 32
20%	0.2742	0.2769	0.2350	0.2994	0.1944 0.2028 0.2044	0.3050 0.3050 0.3117	0.3117 0.3050	0.3117 0.3117
30%	0.4100	0.4097	0.3505	0.3996	0.3078 0.3089 0.3067	0.4322 0.4339 0.4359	0.4322 0.4312	0.4355 0.4356
50%	0.6930	0.6861	0.6485	0.6828	0.6093 0.6082 0.6007	0.7061 0.7059 0.7045	0.7070 0.7052	0.7090 0.7030
70%	0.7749	0.7612	0.7474	0.7708	0.7455 0.7341 0.7312	0.7870 0.7872 0.7843	0.7871 0.7856	0.7878 0.7842

由表 4. 27~表 4. 29 以雙音節為特徵單位的結果顯示（HMM 與 TMM 使用期望值最大化演算法，即第三列數據），在低摘要比例（20%）時，TMM 與 HMM-Type2，其結果不論在人工轉寫或自動轉寫上，均優於其它摘要模型，並以 TMM 為較佳。此外 HMM-Type2 與 TMM，其在發展集所得的參數 λ 均為 1.00，是以均退化為僅使用每一字句產生索引特徵的機率值，$p(w|S_i)$。

表 4. 27 摘錄方法比較(評估方式：MAP，特徵單位：雙音節，人工轉寫)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1 0.10 / 0.05	HMM-Type2 1.00 / 1.00	TMM P(D|T) 1.00 / 1.00 / 04	TMM P(T|T) 1.00 / 1.00 / 16
20%	0.3753	0.3725	0.2994	0.4081	0.1761 0.1782 0.1475	0.4369 0.4153	0.4369 0.4156	0.4369 0.4036
30%	0.5459	0.5389	0.4450	0.5721	0.3383 0.3406 0.3118	0.5794 0.5658	0.5794 0.5678	0.5794 0.5611
50%	0.7550	0.7582	0.6752	0.7733	0.6382 0.6440 0.6125	0.7834 0.7752	0.7834 0.7769	0.7834 0.7786
70%	0.8370	0.8491	0.7854	0.8506	0.7285 0.7386 0.7253	0.8552 0.8508	0.8552 0.8516	0.8552 0.8508

表 4. 28 摘錄方法比較(評估方式：MAP，特徵單位：雙音節，自動轉寫 SP_WG)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1 0.20 / 0.20	HMM-Type2 1.00 / 0.30	TMM P(D|T) 1.00 / 0.30,16 / 04	TMM P(T|T) 1.00 / 0.20,02 / 16
20%	0.3122	0.3133	0.2367	0.2883	0.2322 0.2306	0.2983 0.3467 0.3500	0.2983 0.3467 0.3500	0.2983 0.3467 0.3522
30%	0.4326	0.4209	0.3472	0.4148	0.3742 0.3623	0.4270 0.4584 0.4642	0.4270 0.4584 0.4642	0.4270 0.4554 0.4579
50%	0.7272	0.7141	0.6328	0.7137	0.6507 0.6317	0.7053 0.7288 0.7295	0.7053 0.7297 0.7287	0.7053 0.7281 0.7272
70%	0.7949	0.7853	0.7287	0.7931	0.7328 0.7264	0.7909 0.7893 0.7907	0.7909 0.7894 0.7898	0.7909 0.7885 0.7894

表 4. 29 摘錄方法比較(評估方式：MAP，特徵單位：雙音節，自動轉寫 SP_Adapt)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1 0.05 / 0.05	HMM-Type2 1.00 / 0.30	TMM P(D|T) 1.00 / 0.30,16 / 04	TMM P(T|T) 1.00 / 0.20,02 / 16
20%	0.2942	0.2969	0.2267	0.2775	0.2317 0.2322	0.3092 0.3375 0.3350	0.3092 0.3375 0.3350	0.3092 0.3350 0.3350
30%	0.4268	0.4175	0.3399	0.4148	0.3861 0.3575	0.4448 0.4609 0.4639	0.4448 0.4609 0.4623	0.4448 0.4584 0.4571
50%	0.7180	0.7107	0.6217	0.7187	0.6612 0.6256	0.7102 0.7250 0.7318	0.7102 0.7256 0.7282	0.7102 0.7269 0.7300
70%	0.7953	0.7951	0.7297	0.8009	0.7518 0.7211	0.7972 0.7969 0.7979	0.7972 0.7961 0.7960	0.7972 0.7957 0.7969

由表 4. 30~表 4. 32 以雙字為特徵單位的結果顯示（HMM 與 TMM 使用期望值最大化演算法，即第三列數據），向量空間模型為基礎的摘要模型（VSM 與 RM）有較佳的結果。此外在低摘要比例（20%）時，TMM 其結果大致均優於其它摘要模型；在高摘要比例（50%）時，VSM 其結果均優於其它摘要模型。

表 4. 30 摘錄方法比較(評估方式：MAP，特徵單位：雙字，人工轉寫)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1 0.05 / 0.05	HMM-Type2 1.00 / 0.95	TMM P(D|T) 1.00 / 0.9,04 / 64	TMM P(T|T) 1.00 / 0.9,64 / 64
20%	0.3983	0.3992	0.2924	0.4017	0.2049 0.1742	0.3944 0.4003 0.3917	0.3944 0.4103 0.4125	0.3944 0.4103 0.4053
30%	0.5683	0.5744	0.4435	0.5696	0.3981 0.3550	0.5653 0.5753 0.5599	0.5653 0.5786 0.5612	0.5653 0.5683 0.5567
50%	0.7798	0.7760	0.6944	0.7701	0.6504 0.6401	0.7748 0.7735 0.7714	0.7748 0.7758 0.7679	0.7748 0.7740 0.7726
70%	0.8520	0.8598	0.7883	0.8490	0.7367 0.7288	0.8488 0.8508 0.8498	0.8488 0.8513 0.8479	0.8488 0.8488 0.8483

表 4. 31 摘錄方法比較(評估方式：MAP，特徵單位：雙字，自動轉寫 SP_WG)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1 0.05 / 0.25	HMM-Type2 1.00 / 1.00	TMM P(D|T) 1.00 / 1.00 / 64	TMM P(T|T) 1.00 / 1.00 / 64
20%	0.3122	0.3089	0.2217	0.2833	0.2333 0.2650 0.2322	0.3106 0.2872	0.3106 0.2956	0.3106 0.2839
30%	0.4232	0.4104	0.3482	0.4152	0.4080 0.4239 0.3742	0.4421 0.4113	0.4421 0.4163	0.4421 0.4080
50%	0.7282	0.7049	0.6478	0.7102	0.6693 0.6728 0.6632	0.7132 0.7144	0.7132 0.7161	0.7132 0.7153
70%	0.8059	0.7984	0.7407	0.7936	0.7660 0.7565 0.7473	0.7977 0.7893	0.7977 0.7873	0.7977 0.7892

表 4. 32 摘錄方法比較(評估方式：MAP，特徵單位：雙字，自動轉寫 SP_Adapt)

摘要比例	VSM	RM	LSA	embedded LSA	HMM-Type1 0.05 / 0.25	HMM-Type2 1.00 / 1.00	TMM P(D|T) 1.00 / 1.00 / 64	TMM P(T|T) 1.00 / 1.00 / 64
20%	0.2950	0.2867	0.2150	0.2750	0.2272 0.2711 0.2350	0.3056 0.2933	0.3056 0.3072	0.3056 0.2939
30%	0.4245	0.4120	0.3339	0.4146	0.3890 0.4261 0.3853	0.4471 0.4309	0.4471 0.4390	0.4471 0.4293
50%	0.7215	0.7043	0.6416	0.7085	0.6677 0.6738 0.6461	0.7110 0.7184	0.7110 0.7196	0.7110 0.7186
70%	0.8068	0.8026	0.7366	0.7997	0.7711 0.7577 0.7424	0.8020 0.7962	0.8020 0.7972	0.8020 0.7995

4.5 綜合比較

由表 4. 33 以單詞為特徵單位的結果顯示（使用期望值最大化演算法），對於主題混合模型而言，隨著潛藏主題的增加其結果有所提升。

表 4. 33 不同潛藏主題個數比較(評估方式：MAP，特徵單位：詞，人工轉寫)

摘要比例	2	4	8	16	32	64
20%	0.3856	0.3856	0.3856	0.3856	0.3856	0.3906
30%	0.5435	0.5441	0.5441	0.5418	0.5428	0.5504
50%	0.7609	0.7618	0.7614	0.7603	0.7575	0.7589
70%	0.8427	0.8435	0.8441	0.8425	0.8436	0.8466

由表 4. 34~表 4. 36 的結果顯示（使用期望值最大化演算法），不論在人工轉寫或自動轉寫上，雙音節為特徵單位的結果大致均較其它特徵單位來得好，且模型中以 HMM-Type2 及 TMM 為較佳。此外，雙音節、雙字其結果均較以單詞為特徵單位來得好。

表 4. 34 不同特徵單位比較(評估方式：MAP，摘要比例：20%，人工轉寫)

特徵單位	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T)	TMM P(T|T)
W	0.3517	0.3606	0.2653	0.3756	0.1828	0.3856	0.3906	0.3989
S2	0.3753	0.3725	0.2994	0.4081	0.1475	0.4153	0.4156	0.4036
C2	0.3983	0.3992	0.2924	0.4017	0.1742	0.3917	0.4125	0.4053

表 4. 35 不同特徵單位比較(評估方式：MAP，摘要比例：20%，自動轉寫 SP_WG)

特徵單位	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T)	TMM P(T|T)
W	0.2683	0.2683	0.2317	0.3050	0.2011	0.3272	0.3289	0.3306
S2	0.3122	0.3133	0.2367	0.2883	0.2306	0.3500	0.3500	0.3522
C2	0.3122	0.3089	0.2217	0.2833	0.2322	0.2872	0.2956	0.2839

表 4. 36 不同特徵單位比較(評估方式：MAP，摘要比例：20%，自動轉寫 SP_Adapt)

特徵單位	VSM	RM	LSA	embedded LSA	HMM-Type1	HMM-Type2	TMM P(D|T)	TMM P(T|T)
W	0.2742	0.2769	0.2350	0.2994	0.2044	0.3117	0.3050	0.3117
S2	0.2942	0.2969	0.2267	0.2775	0.2322	0.3350	0.3350	0.3350
C2	0.2950	0.2867	0.2150	0.2750	0.2350	0.2933	0.3072	0.2939

4.6 隱藏式馬可夫模型與主題混合模型進一步實驗

對於隱藏式馬可夫模型-型一，可由式(3.5) 做句子擴充。首先每一測試文件與表 4. 4 語料庫，藉由向量空間表示法估測餘弦分數，並依分數大小選取前 50 名最相關的文件。接著由測試文件中的各字句，與此前 50 名文件的各字句依同樣方式，找到前 k 句相關的字句組成 $\widehat{S_i}$ ，以做句子擴充。

由表 4. 37 可知，以文件為模型時，擴展觀測（Observation）的字句，對自動摘要正確率有顯著的提升。

表 4. 37 HMM-Type1 字句擴充比較(評估方式：MAP，特徵單位：詞，人工轉寫)

摘要比例	λ=0.9	λ=0.9 k=23(加入 23 句)
20%	0.1828	0.3353
30%	0.3158	0.5168
50%	0.6191	0.7528
70%	0.7382	0.8313

對於隱藏式馬可夫模型-型二，可由式(3.10) 做字句移除。

由表 4. 38 可知，以字句為模型時，移除觀測（Observation）中模型本身的字句。於人工轉寫文件上，有助於自動摘要正確率的提升；然而在自動轉寫文件上，因辨識錯誤及斷句不同，抵銷了其結果，但是可以發現在辨識率上升時（SP_Adapt 較 SP_WG 好），其結果有所提升。

表 4. 38 HMM-Type2 字句移除比較(評估方式：MAP，特徵單位：詞，人工轉寫)

摘要比例	人工轉寫		SP_WG		SP_Adapt	
	HMM-Type2 λ=0.4	HMM-Type2 λ=0.4 Sentence Removal	HMM-Type2 λ=0.7	HMM-Type2 λ=0.7 Sentence Removal	HMM-Type2 λ=0.45	HMM-Type2 λ=0.45 Sentence Removal
20%	0.3739	0.3844	0.3000	0.2906	0.3050	0.3067
30%	0.5264	0.5420	0.4100	0.3800	0.4322	0.4128
50%	0.7550	0.7635	0.6978	0.6910	0.7061	0.6997
70%	0.8427	0.8457	0.7706	0.7604	0.7870	0.7774

對於主題混合模型，可由式(3.21) 做字句移除。

由表 4. 39~表 4. 40 可知，以字句為模型的主題混合模型，在移除觀測（Observation）中模型本身的字句。於人工轉寫文件上，有助於自動摘要正確率的提升；然而在自動轉寫文件上，因辨識錯誤及斷句不同，抵銷了其結果，但是可以發現在辨識率上升時（SP_Adapt 較 SP_WG 好），其結果有所提升。

表 4. 39 TMM p(D|T) 字句移除比較(評估方式：MAP，特徵單位：詞，人工轉寫)

摘要比例	Txt		SP_WG		SP_Adapt	
	TMM P(D|T) 0.40, 16	TMM P(D|T) 0.40,16 Sentence Removal	TMM P(D|T) 0.55,64	TMM P(D|T) 0.55,64 Sentence Removal	TMM P(D|T) 0.55,64	TMM P(D|T) 0.55,64 Sentence Removal
20%	0.3739	0.3828	0.3122	0.2872	0.3117	0.2967
30%	0.5264	0.5420	0.4308	0.3783	0.4322	0.4078
50%	0.7591	0.7651	0.7025	0.6891	0.7070	0.6962
70%	0.8443	0.8466	0.7747	0.7594	0.7871	0.7734

‘

表 4. 40 TMM p(T|T) 字句移除比較(評估方式：MAP，特徵單位：詞，人工轉寫)

摘要比例	txt		SP_WG		SP_Adapt	
	TMM P(T|T) 0.35,08	TMM P(T|T) 0.35,08 Sentence Removal	TMM P(T|T) 0.55,32	TMM P(T|T) 0.55 ,32 Sentence Removal	TMM P(T|T) 0.40,04	TMM P(T|T) 0.40,04 Sentence Removal
20%	0.3772	0.3828	0.3122	0.2922	0.3117	0.3067
30%	0.5283	0.5420	0.4292	0.3806	0.4355	0.4128
50%	0.7568	0.7662	0.7020	0.6905	0.7090	0.6992
70%	0.8437	0.8472	0.7730	0.7608	0.7878	0.7801

在主題混合模型中，進一步比較初始化 $p(T_k \mid S_i)$ 時，使用均勻分佈（Uniform）是否較好；換句話說，即假設字句 S_i 產生每一潛藏主題的機率皆一致，即

$$p(T_k \mid S_i) = \frac{1}{K} \tag{4. 8}$$

並且比較 S_i 迭代次數，即式(3.15)、(3.16) 的執行次數。

由表 4. 41 可知，使用均勻分佈在低摘要比例（20%）下，有不錯的表現。

表 4. 41 TMM p(D|T) 初始方式比較(評估方式：MAP，特徵單位：詞，人工轉寫)

摘要比例	1 次	100 次	Uniform 1 次	Uniform 100 次
20%	0.3828	0.3828	0.3844	0.3828
30%	0.5420	0.5424	0.5420	0.5424
50%	0.7651	0.7627	0.7640	0.7636
70%	0.8466	0.8454	0.8458	0.8463

4.7 本章小結

經由實驗結果顯示，於摘要模型比較上：使用隱藏式馬可夫模型或主題混合模型其結果較其它常見方法有顯著的提升，同時主題混合模型在幾乎所有情況下均較隱藏式馬可夫模型來得佳；於特徵單位的比較上：使用雙音節與雙字時，其結果優於使用詞為特徵單位。

此外對於隱藏式馬可夫模型-型一使用字句擴展，其結果有顯著的提升。對於隱藏式馬可夫模型-型二與主題混合模型，使用字句移除，在人工轉寫文件上均有效提升摘要結果；然而在自動轉寫上，因辨識錯誤及斷句不同，抵銷了其結果，但是可以發現在辨識率上升時，其結果有所提升。

最後在主題混合模型的初始化使用均勻分佈，在低摘要比例下，有不錯的表現。