

改良語句模型技術於節錄式語音摘要之研究

Improved Sentence Modeling Techniques for Extractive Speech Summarization

劉士弘 Shih-Hung Liu, 陳冠宇 Kuan-Yu Chen,
王新民 Hsin-Min Wang, 許聞廉 Wen-Lian Hsu
中央研究院資訊科學研究所
{journey, kychen, whm, hsu}@iis.sinica.edu.tw

陳柏林 Berlin Chen
國立臺灣師範大學資訊工程學系
berlin@ntnu.edu.tw

摘要

由於網際網路的蓬勃發展與海量資料時代的來臨，近幾年來自動摘要(Automatic Summarization)已儼然成為一項熱門的研究議題。節錄式(Extractive)自動摘要是根據事先定義的摘要比例，從文字文件(Text Documents)或語音文件(Spoken Documents)中選取一些能夠代表原始文件主旨或主題的重要語句當作摘要。在相關研究中，使用語言模型(Language Modeling)結合庫爾貝克-萊伯勒離散度(Kullback-Leibler Divergence)的架構來挑選重要語句之方法，已初步地被驗證在文字與語音文件的自動摘要任務上有不錯的成果。基於此架構，本論文探究語句明確度(Clarity)資訊對於語音文件摘要任務之影響性，並進一步地藉由明確度的輔助來重新詮釋如何能在自動摘要任務中適當地挑選重要且具代表性的語句。此外，本論文亦針對語句模型的調適方法進行研究；在運用關聯性(Relevance)的概念下，嘗試藉由每一語句各自的關聯性資訊，重新估測並建立語句的語言模型，使其得以更精準地代表語句的語意內容，並增進自動摘要之效能。本論文的語音文件摘要實驗語料是採用公視廣播新聞(MATBN)；實驗結果顯示，相較於其它現有的非監督式摘要方法，我們所發展的新穎式摘要方法能提供明顯的效能改善。

關鍵詞：節錄式自動摘要、語言模型、庫爾貝克-萊伯勒離散度、語句明確度、關聯性

一、緒論

隨著海量資料時代的來臨，巨量的文字及多媒體影音資訊被快速地傳遞並分享於全球各地，資訊超載(Information Overload)的問題也因此產生。如何能讓人們快速且有效率地瀏覽與日俱增的文字資訊或多媒體影音資訊，已成為一個刻不容緩的研究課題。在眾多的研究方法中，自動摘要(Automatic Summarization)被視為是一項不可或缺的關鍵技術[16]。自動摘要之目的在於擷取單一文件(Single-Document)或多重文件(Multi-Document)中的重要語意與主題資訊，藉此讓使用者能更有效率地瀏覽與理解文件的主旨，以便快速地獲得其所需的資訊，避免花費大量時間在審視文件內容。另一方面，語音是多媒體文件中最具資訊的成分之一；如何透過語音(文件)摘要技術來自動地、有效率地處理具時序性的多媒體影音內容，例如：電視新聞、廣播新聞、郵件、電子郵件、會議及演講

錄音等[25]，更是顯得非常重要。其關鍵原因在於多媒體影音內容往往長達數分鐘或數小時，使用者不易於瀏覽和查詢，而必須耐心地閱讀或聽完整份多媒體影音內容，才能理解其中所描述的語意與主題，這違反人們講求方便、有效率的資訊獲取方式。

雖然對於含有語音訊號的多媒體影音，我們可透過自動語音辨識(Automatic Speech Recognition, ASR)技術自動地將其轉換成易於瀏覽的文字內容，再藉由文字文件摘要的技術來做處理，以達到摘要多媒體影音或其它語音文件之目的。但就現階段語音辨識技術的發展，語音文件經語音辨識後自動轉寫成文字的結果，不僅存在辨識錯誤的問題，也缺乏章節與標點符號，使得語句邊界定義不清楚而失去文件的結構資訊；除此之外，語音文件通常含有許多口語語助詞、遲疑、重覆等內容，這都使得語音文件摘要技術的發展面臨更多的挑戰。

一般來說，自動摘要研究可從許多不同面相來進行探討，包括了來源、需求、方式、用途以及模型技術，以下將簡述各個不同面相的相關議題[22]：

1. 來源：根據文件來源，可以分為單一文件摘要與多重文件摘要[3]；單一文件摘要是依據事先定義好的摘要比例，選取能夠代表文件的句子當作摘要；而多重文件摘要是收集多篇相似的文件，需要移除文件間彼此冗餘性(Redundancy)的資訊[4]，考慮文件描述事件發生的先後順序(Causality)[12]，並且確認文件之間的因果關係，經由這些資訊希望能產生有連貫性的文件摘要。

2. 需求：依據使用者需求不同，摘要內容可區分為具有資訊性(Informative)、指示性(Indicative)、以及評論性(Critical)。具有資訊性的摘要是用來表達文件描述的主旨內容與核心資訊；具指示性的摘要是希望將文件中的主題內容做簡單的描述，並將文件分成不同的主題，例如：政治性、學術性、體育性和娛樂性文件，因此所產生的摘要不要求傳達詳細的原始文件內容；具評論性的摘要提供文件正面與反面的觀點(Positive and Negative Sentiments)[9]。

3. 方式：可概分為二大類，節錄式(Extractive)摘要與抽象式(Abstractive)摘要(或重寫式摘要)。前者主要是依據特定的摘要比例，從最原始的文件中選取重要的語句來組成摘要；而後者是在完全理解文件內容之後，重新撰寫產生摘要來代表原始文件的內容，其所使用之語彙或慣用語不一定是全然地來自於原始文件，此種摘要方式是最為貼近人們日常撰寫摘要的形式。然而抽象式摘要需要複雜的自然語言處理(Natural Language Processing, NLP)技術，如資訊擷取(Information Extraction)、對話理解(Discourse Understanding)及自然語言生成(Natural Language Generation)等[26][34]，因此，近年來節錄式摘要之研究仍為主流。

4. 用途：依摘要用途可分為一般性(Generic)摘要與以查詢為基礎(Query-focused)的摘要。前者是從整篇文件中萃取出能夠突顯整篇文件全面性主題資訊的語句，期望摘要產生的內容可以涵蓋整篇文件所有重要的主題；後者透過使用者或特定的查詢來產生與查詢相關的摘要。

5. 模型技術：簡單分成三大類，(i)以簡單的語彙(Lexical)與結構(Structural)特徵做為判斷摘要語句的模型技術[38]，(ii)監督式機器學習(Supervised Machine Learning)以及(iii)非監督式機器學習(Unsupervised Machine Learning)[20]之模型技術。雖然非監督式機器學習的方法在一般的情況下其效能沒有監督式機器學習方法來的好，但非監督式機器學習方法不需要事先準備大量人工標記的訓練資料，以及具有容易實作(Easy-to-Implement)的特性，仍吸引許多學者進行研究與發展，本論文主要也是採用非監督式機器學習的方式來完成自動摘要之任務。

綜觀上述各個面向，本論文主要探究一般性、單一文件節錄式語音摘要問題，並發展和改進非監督式機器學習模型技術。基於近年來，語言模型結合庫爾貝克-萊伯勒離

散度之非監督式模型技術運用在資訊檢索研究上已有非常好的成果[18]，並已初步被應用於語音文件摘要之研究上[36]，本論文將延續此一研究主軸且提出兩個研究貢獻。其一為初步探究使用語句明確度(Clarity)[11]在語音文件摘要任務中之效用，並同時檢視明確度的內部組成成份(即語句與被摘要文件之非相關資訊的交互亂度和語句本身資訊複雜度)；藉由明確度的輔助來重新詮釋如何能在自動摘要任務中適當地挑選重要且具代表性的語句。其二為有鑑於關聯性(Relevance)的概念在資訊檢索領域中已有不錯的發展成果[14]，本論文嘗試結合關聯性資訊來重新估測並建立語句的語言模型，使其得以更精準地代表語句的語意內容，期望可增進自動摘要之效能。

本論文後續安排如下：第二章扼要地介紹現今自動摘要模型技術的相關研究與發展；第三章首先介紹使用語言模型於節錄式語音摘要任務之原理，然後闡述如何將明確度運用至摘要語句之挑選，並且說明如何藉助語句關聯性資訊來改進語句模型之估測，使其得以更精準地代表語句的語意內容；第四章介紹實驗語料與設定以及摘要評估之方法；第五章說明實驗結果及其分析；最後，第六章為結論與未來研究方向。

二、自動摘要模型技術

本論文將過去摘要研究所陸續發展出的自動摘要模型技術大略地歸納成三大類[22]：

1. 以簡單詞彙與結構特徵為基礎之自動摘要模型技術：在 1950 年代，有學者提出使用詞頻(Frequency)來評量每一個詞的重要性與計算文件中每一個語句的顯著性(Significance Factor)[21]。在實作上，可以對每一個詞進行詞幹分析(Stemming)，將其還原成詞根(Root Form)，同時移除停用詞(Stop Word)的影響並計算實詞(Content Word)的重要性等，最後將語句依其顯著分數進行排序(由高至低)，再根據特定的摘要比例來進行節錄式摘要的產生。後來，有學者利用自然語言分析(Natural Language Analysis)技術對文件結構進行剖析，根據文法結構(Grammar Structure)與語言機制(Linguistic Devices)來決定不同語段的凝聚關係(Cohesion)，例如：首語重複(Anaphora)、省略(Ellipsis)、結合(Conjunction)，或同義詞(Synonymy)、上義詞(Hypernym)等語彙關係(Lexical Relation)，並以此結果進行文件自動摘要。相關研究包括使用語彙鏈(Lexical Chain)[1]、宏觀語段結構(Discourse Macro Structure)[30]、修辭結構(Rhetorical Structure)[38]等。另有學者在審視 200 篇科技文件後，發現有 85%的重要語句出現在文件中的第一段，7%的重要語句出現在最後一段[2]。因此，提出了語句在文件中的位置(Position)資訊是進行摘要語句選取時的一項關鍵線索。

2. 以非監督式機器學習為基礎之自動摘要模型技術：非監督式機器學習通常將自動摘要任務視為如何排序並挑選具代表性語句之問題，其方法通常是計算出一種摘要特徵供語句排序使用，常見的特徵有：語句與文件相關性[10]、語句所形成的語言模型生成文件之機率等[5]、語句間之相關性[23][32]、或語句與文件在潛藏主題空間中的距離關係[17]等。

3. 以監督式機器學習為基礎之自動摘要模型技術：監督式機器學習通常將自動摘要之任務視為二元分類問題(Binary Classification)，亦即將語句區分為摘要語句或非摘要語句。我們必須事先準備好一些訓練文件以及其對應的人工標註摘要資訊，然後透過各種分類器的學習機制，進行分類模型的訓練。對於尚未被摘要之文件，此類方法將文件裡的每個語句進行二元分類，即可依其結果產生出摘要。此類方法較著名的相關研究包括簡單貝氏分類器(Naïve-Bayes Classifier)[13]、高斯混合模型(Gaussian Mixture Model, GMM)[24]、隱藏式馬可夫模型(Hidden Markov Model, HMM)[8]、支援向量機(Support Vector Machines, SVM)與條件隨機場域(Conditional Random Fields, CRF)[28]等。監督式模型可同時結合多種摘要特徵來表示每一語句(通常是由上述以詞彙或結構為基礎之摘要方法、或是各式非監督式摘要模型針對語句所輸出的分數或機率值)，綜合各種摘要

特徵所形成的特徵向量將被用來做為監督式摘要模型判斷語句是否屬於摘要語句的依據[17]。

此外，文字文件所要強調的是怎麼說(What-is-said)，而語音文件擁有許多純文字文件所沒有的資訊，通常除了怎麼說，更強調的是如何說(How-is-said)[27]，明顯地，語音是多媒體內涵中最具資訊的成分之一，也因此語音文件摘要的相關研究通常從多媒體語音訊號中萃取豐富的韻律資訊(Prosodic Information)來判斷語句的重要性，如：音調(Intonation)、音高(Pitch)、音強(Power)、語者發聲持續時間(Duration)、語者說話速率(Rate)、語者(Speaker)、情感(Emotion)和說話時場景(Environment)等資訊，這些都是從事語音文件摘要時可以善加利用的語句特徵資訊[20]。

三、使用語言模型於語音文件摘要

語言模型的研究與發展最早是源自於語音辨識及自然語言處理。語言模型旨在描述語言中的所有詞彙之間共同出現與相鄰資訊的關係。其假設人類語言生成(Human Language Generation)是一個隨機過程，而語言模型就是在模擬如何由詞彙構成片語、語句、段落或者文件之過程的機率模型，故又稱為生成式語言模型(Generative Language Modeling)[36]。最簡單的語言模型為單連語言模型(Unigram Language Model, ULM)，它不考慮詞彙之間的順序關係，只個別考慮每一個詞本身出現的機率。較為複雜且常被使用的語言模型為 N-連語言模型，通常 N 為 2 或 3 (即二連或三連語言模型)，其考慮兩個詞彙或三個詞彙之間共同出現與緊連的順序關係。值得一提的是，單連語言模型和 N-連語言模型的主要優點之一是：它們僅需使用訓練語料來估測每一個詞本身出現的機率分佈，或者詞彙之間共同出現與鄰近關係的機率分佈，並不需要額外的人工標記資訊，因此語言模型是屬於基於非監督式機器學習之模型技術。

在過去幾年中，語言模型在資訊檢索任務中已被廣泛地應用且有不錯的實務成效[36]；但就我們所知，在語音文件摘要的任務上，關於使用語言模型的研究是相對較少的。本論文將藉由語言模型的使用來進行摘要語句選取，其基本方法主要可分為兩種，第一為使用語句語言模型生成文件的文件相似度量值(Document Likelihood Measure, DLM)[5]，第二為使用庫爾貝克-萊伯勒離散度量值(Kullback-Leibler Divergence Measure, KL)[17][18]。此外，本章第 3 小節我們將闡述如何額外地考量使用明確度量值於輔助摘要語句之選取，並在第 4 小節提出使用基於關聯性資訊來改進語句模型之估測，使其得以更精準的代表語句的語意內容。

1、文件相似度量值

我們可以把語音文件摘要任務視為是資訊檢索的問題。一般來說，資訊檢索(Information Retrieval, IR)旨在尋找相關文件(Relevant Document)來回應使用者所送出的查詢(Query)或資訊需求(Information Need)。同樣地，在從事語音文件摘要時，我們可將每一篇被摘要文件視為是查詢，而文件中的語句(Sentence)視為候選資訊單元(Candidate Information Unit)；據此，我們可以假設在被摘要文件中，與其愈相關的語句愈有可能是可用來代表文件主旨或主題之摘要語句。

當給予一篇被摘要文件 D 時，文件中每一語句 S 的事後機率 $P(S|D)$ 可以用來表示語句 S 對於文件 D 的重要性。當使用語言模型來計算 $P(S|D)$ 時，我們透過貝氏定理(Bayes' Theorem)將 $P(S|D)$ 展開成[5]：

$$P(S|D) = \frac{P(D|S)P(S)}{P(D)} \quad (1)$$

其中 $P(D)$ 是文件 D 的事前機率，由於 $P(D)$ 不影響語句的排序結果，故可省略不討論；

另一方面， $P(S)$ 是語句 S 的事前機率，可以使用各式非監督式方法或監督式方法來求得[5]。本論文的研究假設語句的事前機率為一個均勻分布(Uniform Distribution)，所以 $P(S)$ 亦可省略。最後， $P(D|S)$ 是語句 S 所形成的語言模型生成文件 D 之機率(或稱作文件相似度)，可以用來表示文件 D 與語句 S 之間的相似關係，如果語句 S 生成文件 D 的機率值愈高，代表語句 S 與文件 D 愈為相似(語句愈能代表文件 D)，即愈有可能是摘要語句。我們可以更進一步地假設文件 D 中詞與詞之間是獨立的，並且不考慮每一個詞在文件 D 中發生的順序關係(即詞袋假設(Bag-of-Word Assumption))，則語句 S 生成文件 D 的文件相似度量值(Document Likelihood Measure, DLM) $P(D|S)$ 可拆解成文件 D 中每一的詞 w 個別發生的條件機率之連乘積：

$$P(D|S) = \prod_{w \in D} P(w|S)^{C(w,D)} \quad (2)$$

此種方法是為語句 S 建立一個語句模型(Sentence Model) $P(w|S)$ ， w 是出現在文件 D 中的詞， $C(w,D)$ 是詞 w 出現在文件 D 中的次數。其中，我們可利用最大化相似度估測(Maximum Likelihood Estimation, MLE)的方式來建立每一個語句的語句模型：

$$P(w|S) = \frac{C(w,S)}{|S|} \quad (3)$$

在(3)中， $C(w,S)$ 表示詞 w 在語句 S 中出現的次數， $|S|$ 則表示語句 S 的總詞數。值得注意的是，由於語句 S 通常僅由少數字詞所組成，因此容易遭遇資料稀疏(Data Sparseness)的問題，這會使得語句模型使用最大化相似度估測時，不僅可能無法準確地估測每一個詞在語句中真正的機率分佈，也可能因為某些詞的條件機率值為零，導致語句 S 產生文件 D 的機率值為零。為了減輕上述的現象，本論文使用 Jelinek-Mercer 平滑化(Smoothing)技術藉由使用以大量文字語料訓練而成的背景單連語言模型(Background Unigram Language Model)來調適語句模型[35]，故 $P(D|S)$ 可進一步地表示成：

$$P(D|S) = \prod_{w \in D} [\lambda \cdot P(w|S) + (1-\lambda) \cdot P(w|B)]^{C(w,D)} \quad (4)$$

其中， $P(w|B)$ 是詞 w 在背景單連語言模型 B 中之機率值。

2、庫爾貝克-萊伯勒離散度量值

語言模型使用於文件摘要的研究中，除了可被用於計算語句生成文件的可能性外，另一種方式為藉由庫爾貝克-萊伯勒離散度量值(Kullback-Leibler Divergence Measure, KL)，來評估文件中每一個語句的重要性。當使用庫爾貝克-萊伯勒離散度量值於摘要任務中，被摘要文件 D 和 D 中的每一個語句 S 都將分別被描述為一個單連語言模型；當相對於被摘要文件 D 的文件模型(Document Model)，語句模型的離散度量值愈小時，則代表語句與文件愈相關，亦即語句 S 愈重要。在此摘要架構下，排序語句重要性的公式如下[18]：

$$KL(D||S) = \sum_{w \in V} P(w|D) \log \frac{P(w|D)}{P(w|S)} \quad (5)$$

其中， V (Vocabulary) 表示一個由語言裡所有可能的語彙所形成的集合。本論文的研究中，文件模型 $P(w|D)$ 的建立方式與語句模型相同(參照式(3))。當我們更進一步地對(5)作分析時，可以發現當文件模型僅使用最大化相似度估測(MLE)的前提下，採用庫爾貝克-萊伯勒離散度量值所得到的語句排序將與使用文件可能性(Document Likelihood)測量方式(即文件相似度量值)所得到的結果是相同的，其推導如下[6]：

$$\begin{aligned}
 -KL(D \| S) &= \sum_{w \in V'}^{rank} P(w|D) \log P(w|S) \\
 &= \sum_{w \in V'} \frac{C(w, D)}{|D|} \log P(w|S) \\
 &= \sum_{w \in V'}^{rank} C(w, D) \log P(w|S) \\
 &= \log P(D|S) \\
 &= P(D|S)
 \end{aligned} \tag{6}$$

由於使用庫爾貝克-萊伯勒離散度量值時，不僅語句被表示成語句模型，每一篇被摘要文件 D 亦被視為一個文件(機率)模型，而文件模型在經由各式語言模型調適與平滑化的技巧下，可以有系統地、適當地調適文件模型的機率分佈；因此相較於文件相似度量值(DLM)只能針對語句模型進行調適，庫爾貝克-萊伯勒離散度量值(KL)能透過不同模型參數估測技術的使用而獲得更佳的自動摘要效能。

3、明確度量值

文件相似度量值(DLM)與庫爾貝克-萊伯勒離散度量值(KL)皆著重於探討語句與文件之間的相似度，但在選取適當的語句作為摘要之任務上，我們認為亦可額外地考量由其它不同角度出發所擷取之線索；譬如，探討語句本身所蘊含的詞彙使用資訊，以及語句與非相關(Non-relevance)資訊(在這裡是指被摘要文件的非相關資訊)間的關係。基於此概念，本論文首先提出使用明確度(Clarity)[11]量值來輔助庫爾貝克-萊伯勒離散度量值進行摘要語句選取。同時，我們亦將深入地探討這兩種不同的量值(一為語句 S 與文件 D 的庫爾貝克-萊伯勒離散度，另一為語句 S 的明確度)對於選取重要且具代表性之語句的實際影響。

首先，我們將每一篇被摘要文件 D 中語句 S 的明確度量值定義如下：

$$\text{Clarity}(S) \stackrel{def}{=} KL(S \| N_D) = CE(S \| N_D) - H(S) \tag{7}$$

其中 $CE(S \| N_D)$ 為語句 S 與被摘要文件 D 的非相關資訊 N_D 之間的交互亂度(Cross Entropy, CE)：

$$CE(S \| N_D) = - \sum_{w \in V'} P(w|S) \log P(w|N_D) \tag{8}$$

我們認為每一篇被摘要文件 D 中可同時擷取出兩種不同面向的資訊，分別是相關(Relevance)與非相關(Non-relevance)資訊。我們將文件 D 中的相關資訊定義為是文件所欲表達的主旨或主題資訊；相反地，文件 D 的非相關資訊則是與該文件內容完全沒有關聯、甚至是背道而馳的主旨或主題資訊。因此，摘要語句模型應與由相關資訊所估測的模型(如文件模型)愈相似(接近)，而與非相關資訊所估測的模型愈不相似(遠離)。當假設對於每一篇被摘要文件而言，語料庫中絕大部分的文件都與其主旨或內容不相關時，則我們可藉由語料庫中大量文件所估測而得的背景單連語言模型(參照式(4)與其說明)來近似每一篇被摘要文件 D 之非相關資訊 N_D 所對應的模型 $P(w|N_D)$ 。簡言之， $CE(S \| N_D)$ 旨在描述語句 S 與被摘要文件 D 的非相關資訊 N_D 之間的相似關係，可視為是語句 S 的一種外在資訊(Extrinsic Information)。若語句 S 與被摘要文件的非相關資訊 N_D 的交互亂度值愈大(亦即語句 S 與被摘要文件 D 的非相關資訊之用字遣詞是大相徑

庭的)，則表示語句 S 與文件 D 的非相關資訊 N_D 愈不相似；反之，若語句 S 與被摘要文件的非相關資訊 N_D 的交互亂度值愈小（亦即語句 S 與 N_D 的用字遣詞是差不多的），則語句 S 與被摘要文件的非相關資訊 N_D 愈相似。

在式(7)明確度量值中的 $H(S)$ 為語句 S 之本身的資訊複雜度(Sentence Entropy, SE)：

$$H(S) = -\sum_{w \in V} P(w|S) \log P(w|S) \quad (9)$$

$H(S)$ 是描述語句本身使用詞彙之集中性，因此語句 S 本身的資訊複雜度可視為是語句本身的一種內在資訊(Intrinsic Information)。當語句複雜度值 $H(S)$ 愈小時，表示語句所使用的不同詞彙之個數愈少或愈集中，語句 S 所呈現的主題也愈聚焦，即語句 S 愈具有獨特性(Specificity)；反之，當語句複雜度值 $H(S)$ 愈大時，表示語句 S 所使用的不同詞彙之個數愈多或愈發散，且各個詞彙出現的頻率相近，也就是語句中較無特別強調的詞彙，所以相較之下，語句 S 所蘊含的資訊可能較複雜，比較不具獨特性。綜觀以上分析，若語句 S 的明確度愈高，表示語句 S 與被摘要文件 D 的非相關資訊 N_D 之間的交互亂度愈大且語句 S 本身的詞彙使用複雜度愈小；換句話說，即此語句所蘊含的資訊不僅不同於被摘要文件 D 的非相關資訊，並且所欲表達的主題內容是較為明確且單純的。

由於語句的明確度是描述語句與文件之非相關資訊 N_D 間的關係以及語句本身的資訊，我們進一步的將語句與文件間相似度的資訊與明確度相結合，做為最終語句重要性排序之依據：

$$-KL(D \| S) + Clarity(S) \quad (10)$$

庫爾貝克-萊伯勒離散度量值愈小，表示語句與被摘要文件的相似度應將會愈大；語句明確度量值愈大，則愈有可能表示語句不僅具有獨特性且能明確呈現被摘要文件之主題。綜合這兩個面向，我們期望可以挑選出與被摘要文件相似度高並且言簡意賅的語句來形成摘要。再者，因明確度量值可區分為兩個部分，一為語句 S 之本身的資訊複雜度，另一為語句 S 與被摘要文件 D 的非相關資訊 N_D 之間的交互亂度，在實驗中我們將更進一步地探討庫爾貝克-萊伯勒離散度量值分別與這兩種成分相結合之摘要成效：

$$-KL(D \| S) - H(S) \quad (11)$$

$$-KL(D \| S) + CE(S \| N_D) \quad (12)$$

目前對於非相關資訊的取得與估測仍是一個值得討論的議題[33]，在本論文後續之實驗中，將初步使用背景單連語言模型來作為每一篇被摘要文件 D 的非相關資訊所對應的語言模型[7]。另一方面，明確度之概念也常被用於資訊檢索領域中，其目的是為了要預測檢索字串(Query)之難易度而衍生出來的概念[31]，本論文是首次使用明確度之概念用於(語音)文件摘要任務中。

4、使用關聯模型

除了結合語句明確度於語音文件摘要之研究外，本論文亦針對語句模型調適進行初步研究。通常，文件中的語句僅由少許的詞彙所組成，當語句模型使用最大化相似度估測時，容易遭遇資料稀疏的問題，藉由背景語言模型進行語句模型之調適為最常見的方法之一(參照式(4))。

雖然文件中的語句通常是簡短的，但我們認為每一語句 S 皆是被用來描述一個概念、想法或主題，我們稱之為語句的關聯類別(Relevance Class) R_S 。在本論文中，我們

的目標是想進一步地模型化關聯類別所代表的資訊，藉此來豐富語句模型所能傳達的語意內容或主題特性。然而，實際上每一語句 S 的關聯類別 R_S 是非常難以求得的；為此，我們透過虛擬關聯回饋(Pseudo Relevant Feedback, PRF)來尋找與關聯類別可能相關的一些文件，並藉由這些文件來近似關聯類別。更明確地，在實作上我們首先把每一語句 S 當作查詢(Query)，代表一個資訊需求(Information Need)，輸入到一個資訊檢索系統中，找出一些與語句 S 相關的關聯文件 $\mathbf{D}_{\text{Top}} = \{D_1, \dots, D_M\}$ ，稱之為虛擬關聯文件(Pseudo Relevant Documents)，用以代表關聯類別 R_S 。接著，透過檢視詞彙 w 與語句 S 在這些虛擬關聯文件中同時出現之關係，可計算出詞彙與語句的聯合機率[14]：

$$P_{\text{RM}}(w, S) = \sum_{D_j \in \mathbf{D}_{\text{Top}}} P(w, S | D_j) P(D_j) \quad (13)$$

當我們進一步地假設在給定某一篇虛擬關聯文件時，詞彙與語句是獨立的，並且語句內的詞彙也是獨立且不考慮其先後次序(即所謂的詞袋假設)，則透過虛擬關聯回饋所估測的語句模型為：

$$P_{\text{RM}}(w | S) = \frac{\sum_{D_j \in \mathbf{D}_{\text{Top}}} \prod_{w' \in S} P(w' | D_j) P(w | D_j) P(D_j)}{\sum_{D_{j'} \in \mathbf{D}_{\text{Top}}} \prod_{w'' \in S} P(w'' | D_{j'}) P(D_{j'})} \quad (14)$$

我們稱之為關聯模型(Relevance Model, RM)。關聯模型的優點在於藉由虛擬關聯文件的資訊，可以更清楚地知道語句所蘊含的資訊、所欲表達的內涵，所以相較於傳統使用最大化相似度估測的語句模型，可更準確地表達語句的語意內容或主題特性，以提升摘要的成效。運用此一關聯模型來調適語句模型時，庫爾貝克-萊伯勒離散度量值的公式(參照式(5))可進一步地表示成：

$$KL(D \| S) = \sum_{w \in V} P(w | D) \log \frac{P(w | D)}{\gamma \cdot P(w | S) + (1 - \gamma) \cdot P_{\text{RM}}(w | S)} \quad (15)$$

其中 $0 \leq \gamma < 1$ ，當 $\gamma = 0$ 代表使用關聯模型取代原本的語句模型。

四、實驗語料及評估方法

1、實驗語料

本論文實驗語料庫為公視新聞語料(Mandarin Chinese Broadcast News Corpus, MATBN)，是由中央研究院資訊科學研究所耗時三年與公共電視台合作錄製並整理的中文新聞語料，其錄製內容為每天一個小時的公視晚間新聞深度報導。我們抽取其中由 2001 年 11 月到 2002 年 8 月總共 205 則新聞報導，區分成訓練集(共 185 則新聞)以及測試集(共 20 則新聞)兩部分，其詳細的統計資訊如表一所示。全部 205 則語音文件長度約為 7.5 小時，我們先做人工切音，切出真正含有講話內容的音訊段落，再經由語音辨識器自動產生出的語音辨識結果稱之為語音文件(Spoken Document, SD)，因此語音文件中只包含有語音辨識錯誤之雜訊；另一方面，我們將此 205 則語音文件藉由人工聽寫的方式，產生出沒有辨識錯誤的正確文字語料，我們稱之為文字文件(Text Document, TD)，每則文字文件再經由三位專家標記摘要語句，我們將此標記的人工摘要做為語音文件與文字文件的正確摘要答案。藉由比較語音文件和文字文件的摘要效能，我們可以觀察語音辨識錯誤對於各種摘要方法之影響。本研究的背景語言模型訓練語料取材自

表一、實驗語料統計資訊

	訓練集	測試集
語料時間	2001/11/07-2002/01/22	2002/01/23-2002/08/22
文件個數	185	20
文件平均持續幾秒	129.4	141.2
文件平均詞個數	326.0	290.3
文件平均語句個數	20.0	23.3
文件平均字錯誤率 (Character Error Rate, CER)	28.8%	29.8%
文件平均詞錯誤率 (Word Error Rate, WER)	38.0%	39.4%

2001 到 2002 年的中央社新聞文字語料(Central News Agency, CNA)，並且以 SRI 語言模型工具[29]訓練出經平滑化的單連語言模型，我們假設此單連語言模型為明確度中的非相關資訊之來源。另外，本論文蒐集 2002 年中央通訊社的 101,268 則同時期新聞文字文件做為建立關聯模型時的檢索標的[6]。

2、評估方法

自動摘要的評估方法主要有兩種，一為主觀人為評估，另一為客觀自動評估；前者為請幾位測試人員來為系統所產生的摘要做評估，給分的範圍為 1-5 分，後者則是預先請幾位測試者依據事先定義好的摘要比例挑選出適合的摘要語句，系統所產生的摘要句子將與測試者所挑選出的句子計算召回率導向的要點評估(Recall-Oriented Understudy for Gisting Evaluation, ROUGE)[19]。由於主觀人為評估非常耗時耗力，所以目前多數自動摘要方法皆採用召回率導向的要點評估做為文件摘要的評估方式，本論文亦採用此種評估方式。ROUGE 方法是計算自動摘要結果與人工摘要之間的重疊單位元(Units)數目占參考摘要(Reference Summary)長度(單位元總個數)的比例。估計的單位可以是 N -連詞(N -gram)、詞序列(Word Sequences)，如：最長相同詞序列或詞成對(Word Pairs)。由於此方法是採用單位元比對的方式，不會產生語句邊界定義的問題，並且適合於多份人工摘要的評估。其評估的分數有三種，ROUGE-1(Unigram)、ROUGE-2(Bigram)和 ROUGE-L(Longest Common Subsequence)分數，ROUGE-1 是評估自動摘要的訊息量，ROUGE-2 是評估自動摘要的流暢性，ROUGE-L 是最長共同字串，本論文希望觀察摘要的流暢性，因此，實驗數據主要是以 ROUGE-2 分數為主。本論文所設定的摘要比例為 10%，其定義為摘要所含詞彙數占整篇文件詞彙數的比例，也就是以詞彙做為判斷摘要比例的單元。在挑選摘要語句過程中，若選到某語句中的某個詞彙時就已經剛好達到摘要比例，為了保持語句語意完整性，此語句剩下的詞彙也會被挑選成為摘要。

五、實驗結果

本論文主要著重在非監督式摘要方法之發展與改進，因此所比較的對象以非監督式摘要方法為主；除此之外，本論文亦嘗試與現今最被廣為使用的監督式機器學習方法做比較，即支持向量機(SVM)[37]。

1、基礎實驗

首先，我們比較庫爾貝克-萊伯勒離散度(KL)與數個非監督式摘要方法之摘要成效，包

表二、基礎實驗結果

		F-score (%)		
		ROUGE-1	ROUGE-2	ROUGE-L
TD	LS	22.5	09.8	18.3
	LEAD	31.0	19.4	27.6
	VSM	34.7	22.8	29.0
	KL	41.1	29.8	36.1
SD	LS	18.1	04.4	13.8
	LEAD	25.5	11.7	22.1
	VSM	34.2	18.9	28.7
	KL	36.4	21.0	30.7

含有最長語句摘要(Longest Sentence, LS)、首句摘要(LEAD)[27]以及向量空間模型(Vector Space Model, VSM)[36]。一般來說，文件中長句可能蘊含有較豐富的主題資訊，因此依據文件中語句長度做排序後，依序選取最長語句做為摘要結果是一種簡單的摘要方法。除此之外，也有學者研究發現，文件常以開門見山法的方式來提點出主題，因此文件開頭的前幾個語句經常是具代表性的語句，首句摘要即是以此概念出發，選取前幾句語句來形成整個文件的摘要。最長語句摘要(LS)及首句摘要(LEAD)都僅適用在一部分具有特殊結構的文件上，因此它們的缺點就是有其侷限性。另外，向量空間模型是把文件和語句分別視為一個向量，並使用詞頻-反文件頻(TF-IDF)特徵來計算每一維度的權重值，文件與語句間的關聯性是藉由餘弦相似度量值來估測，當語句分數較高時，則越有機會成為此文件的摘要[36]。

表二為本論文之基礎實驗結果。首先，在 TD 的實驗中，KL 的摘要效果比 LS、LEAD 及 VSM 等非監督式摘要方法來得好些；因 LS 與 LEAD 僅適用於特殊文章結構上，所以若被摘要文件不具有某種特殊的文章結構，其摘要效能就會有限。相較之下，KL 是較具一般性的摘要方法，因此比較不會受限於文章的結構之影響，故摘要效能比 LS 以及 LEAD 來得彰顯。KL 與 VSM 皆使用淺層的詞彙(詞頻)資訊，但由於 KL 是計算語句模型與文件模型之間的距離關係，對於代表語句與文件的語言模型，我們較容易透過各種技術來進行模型的估計與調適，進而獲得較好的摘要成果。另一方面，在 SD 的實驗中，KL 同樣較優於 LS、LEAD 之摘要方法，但 VSM 的結果則稍微較 KL 好一點，我們認為這可能是因為 VSM 比較不受到語音辨認錯誤的影響。

通常語音文件主要會有語音辨識錯誤和語句邊界偵測錯誤的問題，但我們有先經人工切音，因此摒除了語句邊界偵測錯誤的問題，藉由比較 TD 與 SD 之實驗結果，我們可以觀察語音辨識錯誤率對摘要結果的影響性。比較各式方法，SD 比 TD 下降了 1.9%~8.8%的 ROUGE-2 摘要效能，由此可知語音辨識錯誤率對摘要效能是有顯著的影響性。為了減緩語音辨認錯誤的問題，在未來我們將嘗試使用音節(Syllable)為單位來建立語句以及文件模型；或利用詞圖(Word Graph)、混淆網路(Confusion Network)來含括更多的可能正確候選詞彙以裨益模型估測；更可利用韻律資訊(Prosodic Information)等聲學線索來輔助減緩語音辨認錯誤對摘要效能的影響。

2、使用明確度量值之實驗

接著，我們探討語句明確度量值於語音文件摘要之成效。實驗結果如表三所示，使用語句明確度量值(KL+Clarity, 參照式(10))來輔助挑選摘要語句確實較單純使用 KL

表三、考量明確度量值之實驗結果

		<i>F</i> -score (%)		
		ROUGE-1	ROUGE-2	ROUGE-L
TD	KL	41.1	29.8	36.1
	KL+CE	41.1	29.9	36.2
	KL+SE	44.0	32.6	38.6
	KL+Clarity	44.7	33.5	39.3
SD	KL	36.4	21.0	30.7
	KL+CE	36.4	21.9	31.2
	KL+SE	39.6	25.3	34.7
	KL+Clarity	40.3	26.1	35.4

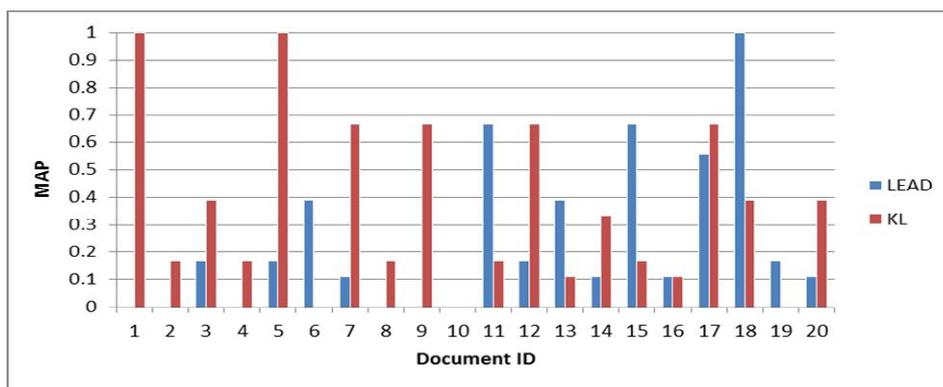
可以獲得更好的摘要效果，這是因為庫爾貝克-萊伯勒離散度量值愈小，表示語句與被摘要文件的相似度應將會愈大；語句明確度量值愈大，則愈有可能表示語句不僅具有獨特性且能明確呈現被摘要文件之主題，綜合這兩個面向後，可挑選出與被摘要文件相似度高並且言簡意賅的語句來形成摘要。我們同時分析了 KL、KL+Clarity 及人工所選的平均摘要語句長度分別為 22.7、20.3 以及 17.2 個詞彙，由此可知 KL+Clarity 所選的摘要語句會比較接近人工所挑選摘要語句之長度，也可看出使用 KL+Clarity 相對於 KL 會比較偏好簡短的語句。

另外，本論文亦針對明確度量值中的內在資訊－語句資訊複雜度(KL+SE，參照式(11))以及外在資訊－語句與被摘要文件的非相關資訊之交互亂度(KL+CE，參照式(12))進行探討。實驗結果顯示 KL+SE 會比 KL 以及 KL+CE 來得好，這個結果說明了語句資訊複雜度在摘要語句的選取上是相當重要的，因為它可以表現出語句本身的獨特性，使語句更能呈現文件所要表達的主題。然而，KL+CE 的實驗結果不論在 TD 或 SD 中皆與 KL 相差不多，對此我們認為可能的原因是因為本論文使用背景單連語言模型來作為被摘要文件的非相關資訊的對應模型，因此在單獨使用的情況下成效不彰。實際上，每一篇被摘要文件的非相關資訊應該都要有所不同，但我們在摘要實驗中先簡單假設每一篇被摘要文件的非相關資訊都是同一個(即背景單連語言模型)，如何為每一篇被摘要文件建立其真正的非相關資訊模型將是我們未來重要的研究課題。總結而言，明確度量值除了考慮語句本身的複雜度資訊外，也考量到使用被摘要文件的非相關資訊來幫助選取摘要語句，所以結合使用明確度量值之方法來輔助挑選重要且具代表性的語句對摘要效能的提升是非常有助益的。

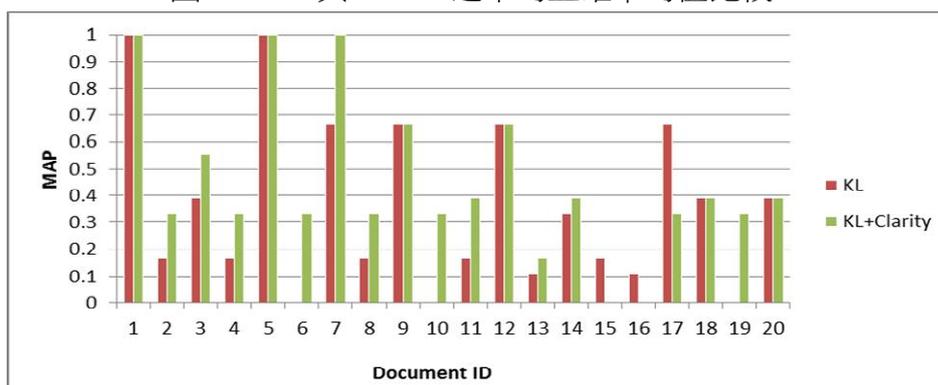
3、語句明確度之分析

為了更進一步、嚴格地分析 KL+Clarity 的摘要能力，本小節以平均正確率均值(Mean Average Precision, MAP)來比較 KL、LEAD 和 KL+Clarity 的摘要能力。相較於 ROUGE 是較寬鬆地計算兩語句間詞彙重疊數目比例做為評量標準，平均正確率均值是嚴格的以每一篇被摘要文件所選出的摘要語句，是否與人工參考摘要語句相同作為評分標準。本實驗中，我們使用各種方法分別計算文件中每一語句後依據各自的分數排序，選取排名前 3 高的語句來計算平均正確率均值(MAP)；另外，由於 SD 中會有語音辨識錯誤等雜訊的干擾，故我們選擇 TD 做為分析之語料。

首先比較 KL 與 LEAD 的平均正確率均值，如圖一所示，KL 在大部分的文件中



圖一、KL 與 LEAD 之平均正確率均值比較



圖二、KL 與明確度之平均正確率均值比較

其平均正確率均值都大於 LEAD，唯有少數幾篇文件的平均正確率均值低於 LEAD，我們觀察那幾篇文件後發現其文章結構是以開門見山法的形式來呈現，因此 LEAD 在這幾篇文件可獲得一定程度的摘要結果。接著，我們比較 KL 與 KL+Clarity 之平均正確率均值，由圖二中可觀察到 KL+Clarity 的平均正確率均值在測試集的大多數文件中皆會高於 KL，只有少數幾篇文件(第 15、16 及 17 篇)會低於 KL 的平均正確率均值。我們認為其原因是可能是因為本論文使用背景單連語言模型來作為所有被摘要文件的非相關資訊的對應模型，而這幾篇文件的非相關資訊可能與背景單連語言模型較不相近，因此造成其摘要結果不如預期。

4、考量關聯模型之實驗

使用關聯模型於語句模型之建立時，需要做一次的資訊檢索來為每個語句找出虛擬關聯文件，本論文採用文件相似度量值 $P(S \parallel D)$ [36]，由同時期的新聞文字文件(共 101,268 篇)中為每一語句選取出 15 篇虛擬關聯文件來進行關聯模型之估測與相關實驗[6]。由於文件中的語句通常相對簡短，因此當使用最大化相似度估測建立語句模型時，容易遭遇資料稀疏的問題，不容易獲得精準的模型，故我們期望考慮額外的關聯資訊於語音文件摘要，亦即藉由虛擬關聯文件來重新估測並建立語句的語言模型，能獲得進一步地摘要成效。重新估測後的關聯模型則可與原本的語句模型相結合或取代之，相結合的參數調整在本實驗中是採用經驗設定(Empirical Setting)。實驗結果如表四所示，在 TD 與 SD 之摘要成效上，使用關聯模型(KL+RM)相較於 KL 在 ROUGE-2 的結果上能有 3.7%與 3.5%的改進。

接著，我們更進一步地結合本論文所探討之語句明確度量值以及關聯模型，實驗結果如表四所示。首先，結合關聯模型與明確度量值(KL+Clarity+RM) 相較於 KL+RM 在 TD 及 SD 的 ROUGE-2 結果分別有 3.8%與 2.1%的進步率。總而言之，結合了庫爾貝克

表四、考量關聯模型之實驗結果

		<i>F</i> -score (%)		
		ROUGE-1	ROUGE-2	ROUGE-L
TD	KL+RM	45.3	33.5	40.3
	KL+CE+RM	45.9	34.5	41.2
	KL+SE+RM	47.7	36.4	42.6
	KL+Clarity+RM	47.7	37.3	42.6
SD	KL+RM	39.3	24.5	34.1
	KL+CE+RM	39.1	26.2	34.7
	KL+SE+RM	40.1	26.4	35.2
	KL+Clarity+RM	40.0	26.6	35.4

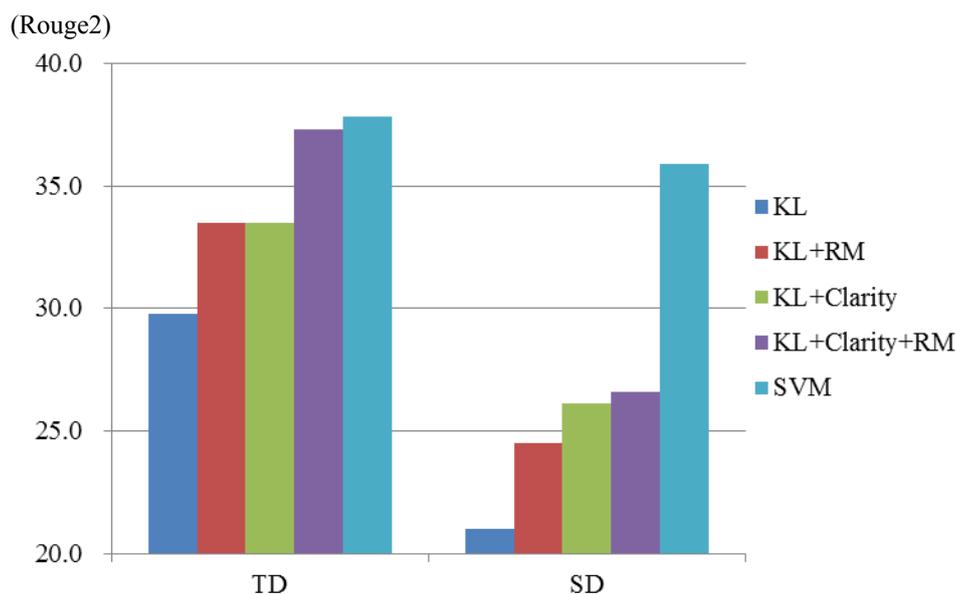
-萊伯勒離散度量值、明確度量值以及關聯模型，是由四個面向來挑選重要的摘要語句，一為語句與文件之相似度($KL(D \| S)$)，二是語句本身資訊複雜度($H(S)$)，第三是語句與被摘要文件的非相關資訊之交互亂度($CE(S \| N_D)$)，最後為語句與關聯文件之相關資訊($P_{RM}(w | S)$)，此實驗結果亦顯示，這四個面向之資訊可以相輔相成的使用，達到最佳的摘要成效。

在關聯模型的相關實驗中，語音辨識錯誤也是影響摘要效能非常嚴重，在 KL+Clarity+RM 的數據中，SD 比 TD 劇烈下降了 10.7% 的 ROUGE-2 摘要效能，在未來研究中，我們認為可以以次詞索引(Subword Indexing)的方式來建立關聯模型以減緩語音辨識錯誤之影響。

5、與監督式模型之比較

除了各式非監督式摘要方法外，本論文亦嘗試比較支持向量機(SVM)於文件摘要之成效。支持向量機是現今常見的監督式機器學習方法之一，近年來已有學者將其運用至文件摘要領域之中[37]。本論文使用訓練集的 185 篇文件進行支持向量機模型的訓練語料，我們為文件中的每一語句抽取 19 維特徵[17]，包括有韻律特徵(Prosodic Features)、語彙特徵(Lexical Features)、結構特徵(Structural Features)以及基本的模型特徵(Model Features)等資訊，其核心函數設定為半徑式函數(Radial Basis Function)，其中 SVM 的參數設定都是使用預設值。

實驗結果如圖三所示，一如預期地，SVM 與其他各式非監督式模型相比較，不論是在 TD 或 SD 的實驗上(其 ROUGE-2 分別為 37.8 及 35.9)都是表現最好的方法，這是由於監督式機器學習藉由使用人工標注的摘要句子進行模型之訓練，其使用的資訊較非監督式機器學習方法多且正確，因此其摘要的效果也較非監督式機器學習來的好。值得一提的是，將明確度與關聯模型(KL+Clarity+RM)互相結合之後，摘要之成效在 TD 上可逼近於監督式機器學習方法的 SVM，此一實驗結果令人感到驚訝，因為本論文所探討之各式摘要方法僅考慮了文件與語句中的單一種特徵值，即藉由詞彙分佈資訊來挑選語句，而支持向量機不僅使用了 19 種特徵值，更需要使用人工標注的正確答案進行模型的訓練。我們認為，此結果之原因可能是由於支持向量機之摘要技術是將摘要任務視為一個二元分類問題，在自動摘要的研究中或許可以達到某一程度的摘要成效，但未必是最好的解決方法。另一方面，在 SD 的實驗中，SVM 相較於其他方法能擁有特別突出的結果，其原因可能是因為我們所使用的實驗語料是經人工切音，因此 SD 中語句的



圖三、SVM 與其他非監督式摘要方法之比較

韻律特徵是很正確的資訊，又因為韻律特徵對語音文件摘要是具有相當程度的幫助，所以 SVM 在 SD 的實驗中才能表現得如此傑出。

六、結論與未來方向

本論文主要有兩個貢獻，第一為首次探究明確度於語音文件摘要上之效用，當與庫爾貝克-萊伯勒離散度相結合後，運用於語音文件摘要上具有加成作用之效果。我們亦同時檢視明確度的內部組成，將之區分成內在資訊(語句本身資訊複雜度)及外在資訊(語句與被摘要文件非相關資訊之交互亂度)兩個面向，來詮釋明確度如何輔助挑選文件中重要且具代表性的摘要語句。第二，基於所謂關聯性(Relevance)的概念，本論文嘗試使用虛擬關聯文件來重新估測並建立語句的語言模型，使其得以更精準地代表語句的語意內容，以增進自動摘要的效能。相較於其它現有的非監督式摘要方法，本論文所提出之摘要方法有明顯的效能改善，甚至可以逼近常見的監督式摘要方法。

未來，我們的研究將有三個主要的方向：首先，本論文所提出之語句明確度是由兩種資訊組合而成，而這兩種資訊在摘要語句挑選時扮演同等重要的角色，我們將進一步的研究是否可以針對不同的文件或不同的語句給予適當的權重調整，以期獲得更好的摘要成效；第二，目前關聯模型僅運用於重建語句的語言模型，我們將嘗試使用被摘要文件的關聯資訊來重新估測並建立文件的語言模型；最後，我們希望將明確度此一摘要特徵資訊結合於監督式機器學習方法(如 CRF 或深度類神經網絡(Deep Neural Network Learning, DNN)等)中，期望訓練後的模型能夠在文字文件摘要或語音文件摘要上獲得更好的表現。

致謝

本論文之研究承蒙教育部-國立臺灣師範大學邁向頂尖大學計畫(102J1A0800)與行政院國家科學委員會研究計畫(NSC 101-2221-E-003-024-MY3、NSC 101-2511-S-003-057-MY3、NSC 101-2511-S-003-047-MY3 和 NSC 102-2221-E-003-014-MY3)之經費支持，謹此致謝。

參考文獻

- [1] R. Barzilay and M. Elhadad, *Using Lexical Chains for Text Summarization*, Proceedings of Workshop on Intelligent Scalable Text Summarization, pp. 10-17, 1997
- [2] P. Baxendale, *Machine-made Index for Technical Literature – an Experiment*, IBM Journal of Research and Development, Vol. 2, No. 4, pp. 354-361, 1958
- [3] X.-Y. Cai, and W.-J. Li, *Ranking through Clustering: An Integrated Approach to Multi-Document Summarization*, IEEE Transactions on Audio, Speech and Language Processing, Vol. 21, No. 7, pp.1424-1433, 2013
- [4] J. Carbonell and J. Goldstein, *The Use of MMR Diversity-based Reranking for Reordering Documents and Producing Summaries*, Proceedings of the 21th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 335-336, 1998
- [5] Y.-T. Chen, B. Chen and H.-M. Wang, *A Probabilistic Generative Framework for Extractive Broadcast News Speech Summarization*, IEEE Transactions on Audio, Speech and Language Processing, Vol. 17, No. 1, pp. 95-106, 2009
- [6] B. Chen, H.-C. Chang, K.-Y. Chen, *Sentence Modeling for Extractive Speech Summarization*, Proceeding of International Conference on Multimedia & Expo (ICME), 2013
- [7] B. Chen, K.-Y. Chen, P.-N. Chen, Y.-W. Chen, *Spoken Document Retrieval With Unsupervised Query Modeling Techniques*, IEEE Transactions on Audio, Speech and Language Processing, 20(9):2602-2612, 2012
- [8] J.-M. Conroy and D.-P. O’Leary, *Text Summarization via Hidden Markov Models*, Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 406-407, 2001
- [9] M. Galley, K. McKeown, J. Hirschberg, and E. Shriberg, *Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies*, Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL), pp. 669-676, 2004
- [10] Y. Gong and X. Liu, *Generic Text Summarization using Relevance Measure and Latent Semantic Analysis*, Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 19-25, 2001
- [11] S. Hummel, A. Shtok, D. Carmel, *Clarity Re-visited*, Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp.1039-1040, 2012
- [12] J.-J. Kuo and H.-H. Chen, *Multi-document Summary Generation using Informative and Event Words*, Journal of ACM Transactions on Asian Language Information Processing, Vol. 7, No.1, pp. 550-557, 2006
- [13] J. Kupiec, *A Trainable Document Summarizer*, Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 68-73, 1995

- [14] V. Lavrenko and W.-B. Croft, *Relevance -based Language Models*, Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 120-127, 2001
- [15] J.-H. Lee, S.-Y. Kong, Y.- C. Pan, Y. S. Fu, and Y.-T. Huang, *Multilayered Summarization of Spoken Document Archive by Information Extraction and Semantic Structuring*, Proceedings of the 7th Annual Conference of the International Speech Communication Association (Interspeech), pp. 1539-1542, 2006
- [16] S.-H. Lin and B. Chen, *A Survey on Speech Summarization Techniques*, The Association for Computational Linguistics and Chinese Language Processing Newsletter, Vol. 21, No. 1, pp. 4-16, 2010
- [17] S.-H. Lin and B. Chen, *Improved Speech Summarization with Multiple-hypothesis Representations and Kullback-Leibler Divergence Measures*, Proceeding of the 10th Annual Conference of the International Speech Communication Association (Interspeech), pp. 1847-1850, 2009
- [18] S.-H. Lin, Y.-M. Yeh and B. Chen, *Leveraging Kullback-Leibler Divergence Measures and Information-rich Cues for Speech Summarization*, IEEE Transactions on Audio, Speech and Language Processing. Vol. 19, No. 4, pp. 871-882, 2011
- [19] C.-Y. Lin, *ROUGE: Recall-oriented Understudy for Gisting Evaluation*. 2003 [Online]. Available: <http://haydn.isi.edu/ROUGE/>.
- [20] Y. Liu, and D. Hakkani-Tur, *Speech Summarization*, Chapter 13 in Spoken Language Understanding: System for Extracting Semantic Information from Speech, G. Tur and R. D. Mori (Eds), New York, Wiley, 2011
- [21] P. Luhn, *The Automatic Creation of Literature Abstracts*, IBM Journal of Research and Development, Vol. 2, No. 2, pp.159-165, 1958
- [22] I. Mani and M.-T. Maybury, *Advances in Automatic Text Summarization*, Cambridge: MIT Press, 1999
- [23] R. Mihalcea and P. Tarau, *TextRank Bringing Order into Texts*, Proceedings of Empirical Method in Natural Language Processing (EMNLP), pp. 404-411, 2004
- [24] G. Murray, S. Renals, and J. Carletta, *Extractive Summarization of Meeting Recordings*, Proceedings of the 6th Annual Conference of the International Speech Communication Association (Interspeech), pp. 593-596, 2005
- [25] M. Ostendorf, *Speech Technology and Information Access*, IEEE Signal Processing Magazine, Vol. 25, No. 3, 2008
- [26] C.-D. Paice, *Constructing Literature Abstracts by Computer Techniques and Prospects*, Journal of Information Processing and Management, Vol. 26, No. 1, pp. 171-186, 1990
- [27] G. Penn and X. Zhu, *A Critical Reassessment of Evaluation Baselines for Speech Summarization*, Proceedings of Annual Meeting of the Association for Computational Linguistics, pp. 470-478, 2008
- [28] D. Shen, J.-T. Sun, H. Li, Q. Yang, and Z. Chen, *Document Summarization using Conditional Random Fields*, Proceedings of International Joint Conference on Artificial Intelligence (IJCAI), pp. 2862-2867, 2007

- [29] A. Stolcke, SRI Language Modeling Toolkit, <http://www.speech.sri.com/projects/srilm/>.
- [30] T. Strzalkowski, J. Wand, and B. Wise, *A Robust Practical Text Summarization*, Proceedings of AAAI Conference on Artificial Intelligence Spring Symposium on Intelligent Text Summarization, pp. 26-33, 1998
- [31] S.-C. Townsend, Y. Zhou, W.-B. Croft, *Predicting Query Performance*, Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 299-306, 2002
- [32] X. Wan and J. Yang, *Multi-document Summarization using Cluster-based Link Analysis*, Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 299-306, 2008
- [33] X.-H. Wang, H. Fang, C.-X. Zhai. *A Study of Methods for Negative Relevance Feedback*, Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 219-226, 2008
- [34] M. Witbrock and V. Mittal, *Ultra Summarization: a Statistical Approach to Generating Highly Condensed Non-extractive Summaries*, Proceedings of the 22th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 315-316, 1999
- [35] C.-X. Zhai and J. Lafferty, *A Study of Smoothing Methods for Language Models Applied to Ad Hoc Information Retrieval*, Proceedings of the Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), pp. 334-342, 2011
- [36] C.-X. Zhai, *Statistical Language Models for Information Retrieval: A Critical Review*, Foundations and Trends in Information Retrieval, 2(3), pp.137-213, 2008
- [37] J. Zhang and P. Fung, *Speech Summarization without Lexical Features for Mandarin Broadcast News*, Proceedings of NAACL HLT, Companion Volume, pp. 213–216, 2007
- [38] J.-J. Zhang, H.-Y. Chan and P. Fung, *Extractive Speech Summarization using Shallow Rhetorical Structure Modeling*, IEEE Transactions on Audio, Speech and Language Processing, Vol. 18, No. 6, pp. 1147-1157, 2010