中華民國計算語言學學會
通訊 ACLCLP
The Association for Computational Linguistics and Chinese Language Processing
第二十三卷第五期
發行人：許聞廉
主　編：王新民
執行編輯：黃　琪

E-Mail：aclclp@hp.iis.sinica.edu.tw
地 址：台北市研究院路二段128號中研院資訊所
Website : http://www.aclclp.org.tw
電 話：(02)2788-3799 ext:1502
傳 眞：(02)2788-1638
劃 撥：19166251

本 期 要 目

Information on upcoming conferences

ACL 2013

The 51st Annual Meeting of the Association for Computational Linguistics
- Conference Date: August 4-9, 2013
- Submission Deadline:
  Long paper: February 20, 2013
  Short paper: April 14, 2013
- Location: Sofia, Bulgaria
- http://acl2013.org/site/

ACM SIGIR 2013

The 36th Annual ACM SIGIR Conference
- Conference Date: July 28- August 1, 2013
- Submission Deadline: January 21, 2013
- Location: Dublin, Ireland
- http://sigir2013.ie/

ICASSP 2013

The 38th International Conference on Acoustics, Speech and Signal Processing
- Conference Date: May 26-31, 2013
- Submission Deadline: Closed
- Location: Vancouver, Canada
- http://www.icassp2013.com/

IJCNLP 2013

The 6th International Joint Conference on Natural Language Processing
- Conference Date: October 15-18, 2013
- Submission Deadline: TBD
- Location: Nagoya, Japan
- Website: Coming soon

INTERSPEECH 2013

The 14th Annual Conference of the International Speech Communication Association
- Conference Date: August 25-29, 2013
- Submission Deadline: March 18, 2013
- Location: Lyon, France
- http://www.interspeech2013.org/

MT Summit 2013

The 14th Machine Translation Summit
- Conference Date: September 2-6, 2013
- Submission Deadline: April 15, 2013
- Location: Nice, France
- http://www.mtsummit2013.info/

NAACL HLT 2013

The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies
- Conference Date: June 9-14, 2013
- Submission Deadline: Closed
- Location: Atlanta, GA, USA
- http://naacl2013.naacl.org/

PACLIC 2013

The 27th Pacific Asia Conference on Language, Information, and Computation
- Conference Date: November 22-24, 2013
- Submission Deadline: June 1, 2013 (tentative)
- Location: National Chengchi University, Taiwan
- http://paclic27.nccu.edu.tw/

Call for Papers

International Journal of Computational Linguistics and Chinese Language Processing

Special Issue on **Processing Lexical Tones in Natural Speech**

This special issue aims to address questions about how lexical tones are processed by humans and machines in the context of natural, continuous speech. Lexical tones in tone languages have been widely investigated in the fields of linguistics, psycholinguistics, computational linguistics, and language acquisition by applying a wide range of theoretical, empirical, and experimental approaches. As the phonetic representation of lexical tones which are produced in connected speech can differ considerably from that of lexical tones which are produced in isolation, research interests constantly grow in how lexical tones are produced, perceived, and processed in realistic speech data. This special issue aims to bring together methodologies from different research disciplines to extend our understanding of lexical tones used in real speaking situations. We welcome submissions addressing the following issues.

- **Modeling lexical tones**: Can lexical tones which are produced in natural speech be more accurately described and modeled by quantitative/gradient measures or by categorical systems? Is a hybrid approach possible? In what way can lexical tones be represented and analyzed by utilizing spoken corpora?
- **Human language processing**: What role do lexical tones play in the mental lexicon? How are lexical tones produced and perceived by native and non-native language users?
- **Language acquisition**: How are lexical tones acquired by typical developing children, hearing-impaired children, and second language learners? Are the phonological development patterns different from each other?
- **Speech technology**: What kind of information about lexical tones can be integrated into ASR and synthesis systems to improve system performances?
- **Other research results related to lexical tones in natural speech** are also welcome to contribute to this special issue.

Paper submission deadline: February 28, 2013
Notification of acceptance: May 31, 2013
Final paper due: August 31, 2013
Tentative publication date: December, 2013

All submitted papers should present original research work, which has not been published elsewhere. Submitted manuscripts will be peer-reviewed by at least two independent reviewers. For detailed submission guidelines, please visit the website of the *International Journal of Computational Linguistics and Chinese Language Processing* at http://www.aclclp.org.tw/journal/submit.php. Please also feel free to contact the Guest Editor of this special issue, Dr. Shu-Chuan Tseng, at tsengsc@gate.sinica.edu.tw, if you need any additional information.

遞迴式類神經網路語言模型應用額外資訊於語音辨識之研究

Recurrent Neural Network-based Language Modeling with Extra Information Cues for Speech Recognition

黃邦烜　Bang-Xuan Huang
國立臺灣師範大學資訊工程學系
699470204@ntnu.edu.tw

郝柏翰　Hank Hao
國立臺灣師範大學資訊工程學系
60047082s@ntnu.edu.tw

陳冠宇　Menphis Chen
中央研究院資訊科學研究所
kychen@iis.sinica.edu.tw

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

摘要

近年來類神經網路興起，其運用在語言模型領域有不錯的成效，如前饋式類神經網路語言模型。不同於傳統 N 連語言模型，前饋式類神經網路語言模型是將詞序列映射至連續空間來估測下一個詞出現的機率，以解決資料稀疏的問題。此外，更有學者使用遞迴式類神經網路來建構語言模型，期望藉由遞迴的方式將歷史資訊儲存起來，進而獲得長距離的資訊。

本論文根據遞迴式類神經網路的基礎，使用關聯資訊來捕捉長距離資訊；另外，也探討了根據語句的特性來動態地調整語言模型。實驗結果顯示，使用額外資訊於遞迴式類神經網路語言模型對於大詞彙連續語音辨識的效能有相當程度的提昇。

關鍵詞：語音辨識、語言模型、遞迴式類神經網路

一、 緒論

語音是人與人溝通的基本媒介，如果無法透過語音來對話，便無法正確表達彼此的想法。在對話中，人們藉由語調和詞句，了解對方的情緒與想法等諸多細節，而聽懂對方想表達的資訊；這些人所擁有的天賦，是目前資訊科技所無法達到而需研究的；因此，自動語音辨識的研究也變得更加重要。在自動語音辨識的過程中，我們需先透過特徵擷取(Feature Extraction)來處理語音訊號，得到可以代表此段語音訊號的特徵參數；接著，將所擷取的

特徵參數轉換成語音特徵向量，以利語音辨識系統使用或分析。另一部分，則使用語音語料和文字語料分別建構出聲學模型(Acoustic Model)和語言模型(Language Model)，用以表示語音與文字之間的對應關係以及代表語言中各種詞彙的出現情形。再根據聲學模型、語言模型、詞典和特徵向量所提供的資訊以進行語言解碼(Linguistic Decoding)，獲得最後辨識結果。爲了達到電腦能理解人類的語音的目標，本論文研究語音辨識中的語言模型，希望藉由語言模型能捕捉語言的規律性。*N* 連語言模型是較常見的語言模型之一，它易於產生且容易使用的特性引發許多學者研究與使用。但此語言模型有資料稀疏與缺乏長距離資訊等問題，因此有不同類型的語言模型被發展出並期望解決這些問題，前饋式類神經網路語言模型(Neural Network Language Models, NNLM)則是其中之一。它將歷史詞序列的資訊投影到連續空間，借以解決資料稀疏的問題，但對於長距離資訊的取得仍不盡理想。因此，爲了獲得長距離的資訊，有所謂遞迴式類神經網路語言模型(Recurrent Neural Network Language Models, RNNLM)被提出。1994 年有研究[1]指出，遞迴式類神經網路較難取得更長距離的資訊，其理由是當句子越長時，越遠距離的資訊透過機率相乘所得到的值會趨近於零。本論文延續先前對於遞迴式類神經網路語言模型之研究，嘗試使用額外的資訊來增進遞模型的預測能力，期望在大詞彙連續語音辨識中有相當程度的改善。

本論文的安排如下：第二章簡介類神經網路語言模型；第三章介紹遞迴式類神經網路語言模型於自動語音辨識之使用，並且說明遞迴式類神經網路語言模型相關理論及架構；第四章探索遞迴式類神經網路語言模型之改進；第五章介紹實驗語料、實驗設定以及實驗結果分析；第六節則是結論及未來展望。

二、 類神經網路語言模型於自動語音辨識之使用

(一) 類神經網路簡介

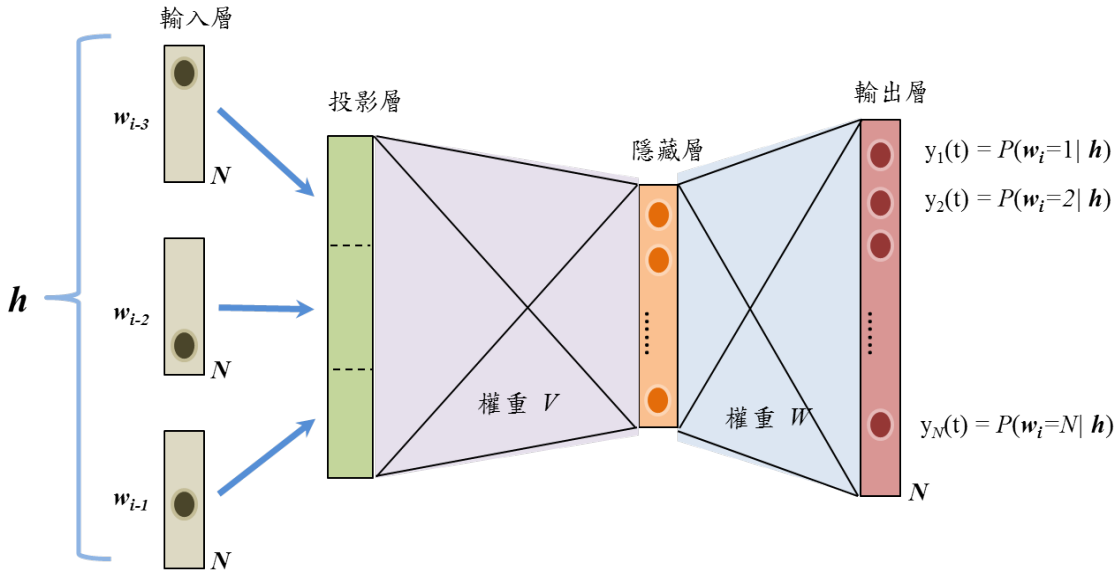
類神經網路(Neural Networks)起源於人工智慧(Artificial Intelligence)，又可稱爲人工類神經網路(Artificial Neural Networks, ANN)。自 1940 年開始科學家開始模仿神經元(Neuron)的運作模式，認爲如果兩個神經元同時被觸發，則它們之間的連結就會獲得增強。直到近年來，類神經網路結合了各項領域，如資訊、金融甚至心理學等都有不錯的成效，其中；像是感知器演算法(Perceptron)是第一個實踐出類神經網路的創舉。

目前類神經網路主要被用於分類及預測上，在影像處理方面，如圖案的辨識或雜訊的處理等，而在語音處理中則有語言模型、語音合成與強健性語音辨識等；另外則是氣象預測、電腦輔助教學、手寫辨識以及超大積體電路的應用。本論文則是探討語音辨識裡的語言模型部份；以下介紹目前所發展出來各種常見的類神經網路語言模型。

(二) 類神經網路語言模型

將類神經網路與語言模型結合則可表示成圖一，稱爲前饋式類神經網路語言模型[2]。其主要架構包括輸入層(Input Layer)、隱藏層(Hidden Layer)和輸出層(Output Layer)；有時會額外加入一層投影層(Projection Layer)，用來將歷史詞序列的資訊投影至此連續空間，並降低輸入層的維度。

不同於 *N* 連語言模型會有資料稀疏的問題，投影層可以接受所有可能的詞序列組合；並且，詞序列中的每個詞能各自貢獻出權重值來估測下一個詞出現的可能性。層跟層之間的神經元靠著突觸(Synapse)來傳遞訊息。各層可以向量表示，層之間的突觸則以矩陣表示。接下來先介紹各層所代表的意義：



圖一、類神經網路語言模型架構

- 輸入層與隱藏層

輸入層爲欲預測詞的歷史資訊，其中歷史資訊以 ***h*** 表示。每個詞使用 one-of-*N* 方式進行編碼，例如詞 w_i 爲詞典中的*第 l* 個詞，在長度爲 *N* 的向量中，詞 w_i 只有第 *l* 維是 1 其餘爲 0。輸入層、隱藏層及輸出層中包含了許多節點，輸入層中的節點以變數 *i* 表示，隱藏層的節點以變數 *j* 表示，輸出層的節點則以變數 *k* 表示。各層的節點結合後可形成一個向量，式(1)爲各節點在向量中的表示方式。$x_i(t)$表示輸入層中於 *t* 時間點第 *i* 個節點，其中 *i* 爲 1 到 *N* 之間表示輸入層大小，*N* 爲詞彙的數量。

$$x_i(t) = \begin{cases} 1 & \text{if } i \in \text{對應的詞} \\ 0 & \text{其他} \end{cases} \tag{1}$$

以圖一爲例，***h*** 代表了前三個詞(t=3)的歷史資訊，結合了三個歷史詞向量，成爲另一個長度爲 3*N* 的輸入層向量 ***x***，可視爲一個四連的類神經網路語言模型。在傳統前饋式類神經網路裡，向量 ***x*** 則會透過權重 *V* 來傳遞，而權重 *V* 會以矩陣的方式來表示，權重 *V* 中包含了許多輸入層節點和隱藏層節點間的鏈結權重值。式(2)爲輸入層各節點傳遞至隱藏層中的節點 *j*。v_{ji} 是第 *j* 個隱藏層節點對第 *i* 個輸入層節點的鏈結權重值，θ_j 爲第 *j* 個隱藏層節點的偏權值，$net_j(t)$爲第 *j* 個隱藏層節點淨輸入值，$y_j(t)$則爲第 *j* 個隱藏層節點。

$$\begin{aligned} net_j(t) &= \sum_i v_{ji} x_i(t) + \theta_j \\ y_j(t) &= f(net_j) \end{aligned} \tag{2}$$

其中， $f(net_j)$爲網路的活化函數(Activation Function)。爲了保證輸出值能介於 0 到 1 之間，本論文中所使用的是雙彎曲函數(Sigmoid Function)，如式(3)所表示：

$$f(x) = \frac{1}{1+e^{-x}} \tag{3}$$

- 隱藏層與輸出層

如同輸入層與隱藏層，隱藏層的各個節點會透過權重 *W* 傳遞給輸出層，可以透過式(4)來表示：

$$\begin{aligned} net_k(t) &= \sum_j w_{kj} y_j(t) + \theta_k \\ y_k(t) &= g(net_k) \end{aligned} \tag{4}$$

其中，w_{kj} 是第 *k* 個隱藏層節點對第 *j* 個輸入層節點的鏈結權重值，θ_k 爲第 *k* 個隱藏層節點的偏權值，$net_k(t)$爲第 *k* 個隱藏層節點淨輸入值，$y_k(t)$爲第 *k* 個輸出層節點。爲了使輸出層各節點的值總和爲 1，最後的 $g(net_k)$爲軟化最大值活化函數(Softmax Activation Function)，也是轉移函數(Transfer Function)的一種。如式(5)來表示：

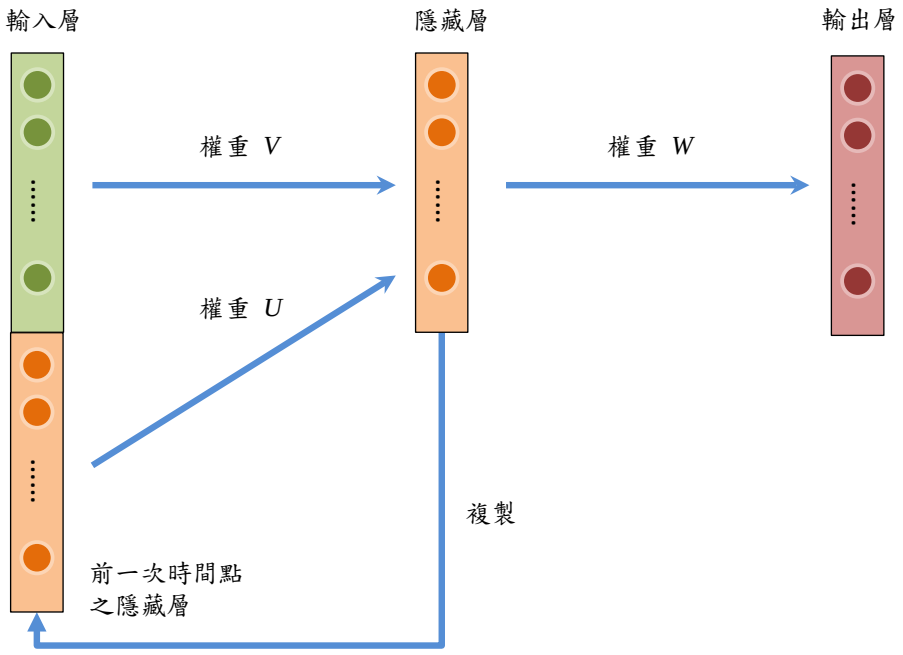
$$g(net_k) = \frac{e^{net_k}}{\sum_k e^{net_k}} \tag{5}$$

最後輸出層的結果可視爲一個 *N* 維的向量，其第 *l* 維的意義是在歷史詞序列 ***h*** 發生的情況下，目前預測的詞 w_i 發生的機率，其數學表示式爲 $P(w_i = l \mid h)$ 。

三、 遞迴式類神經網路語言模型

有別於傳統類神經網路，遞迴式的類神經網路更能帶來較好的訓練能力，一般常見的是於 1990 年由 Elman 所發展的艾爾曼網路(Elman Networks)[3]。其概念是將隱藏層的輸出當作下一次時間點隱藏層的輸入，而根據不同的需求也有許多不同的網路形成，如喬丹網路(Jordan Networks)[4]是將輸出層的輸出再傳遞給下一時間點的隱藏層、雙向遞迴式類神經網路(Bi-directional RNN)[5]利用歷史資訊和未來資訊來做預測，使用的是兩個遞迴式類神經網路來做結合及階層遞迴式類神經網路(Hierarchical RNN)等。本論文則是以艾爾曼網路來進行探討。

遞迴式類神經網路語言模型和傳統類神經網路語言模型主要的差別除了少了投影層、增加了前一時間點的隱藏層外，另一個差別就是輸入層部分。在訓練過程時，輸入層是一次以一個詞來表示並訓練，每一個詞的表示方法則與傳統前饋式類神經網路語言模型相同。其網路的結構是把輸入層加大，將上一時間點的隱藏層預先儲存起來，若以時間方式來階層展開的話，將會更清楚看出其遞迴的概念，如圖二所示。由於遞迴式類神經網路具有時序處理(Temporal Processing)的能力，而一般來評估此類型的網路常會注意它們的穩定性(Stability)、可控性(Controllability)及可觀察性(Observability)。穩定性注重的是隨著時間改變，網路輸出結果需是受侷限的且輸出後的調整量不可過於劇烈，例如網路中輸出的部份或權重。可控性在意的是「是否能夠控制的動態行爲」，如果在有限的步驟中，一個初始狀態是可控制至任何期望的狀態，則此遞迴式網路可被稱爲具有可控性的。可觀察性關注的是「是否可觀察出控制應用的結果」，如果網路的狀態可以確定從一組有限的輸入或輸出測量，則稱做此網路有可觀察性。



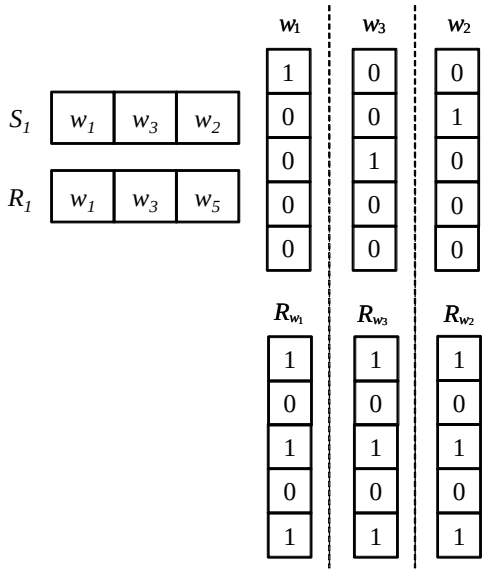
圖二、遞迴式類神經網路架構

但 Bengio 等學者[6]發現，利用梯度下降法(Gradient Descent Method)於遞迴式類神經網路中，對於學習長距離的資訊是十分困難的。而要獲得長距離資訊必須要具有任意時間學習並且擁有抵抗其它資訊干擾的能力。但因爲隨著時間變化，距離較遠的資訊會被每一次時間點的輸入資訊所干擾，反而降低了遞迴式結構的好處。因此下一章將討論如何將遞迴式類神經網路語言模型做進一步的改進。

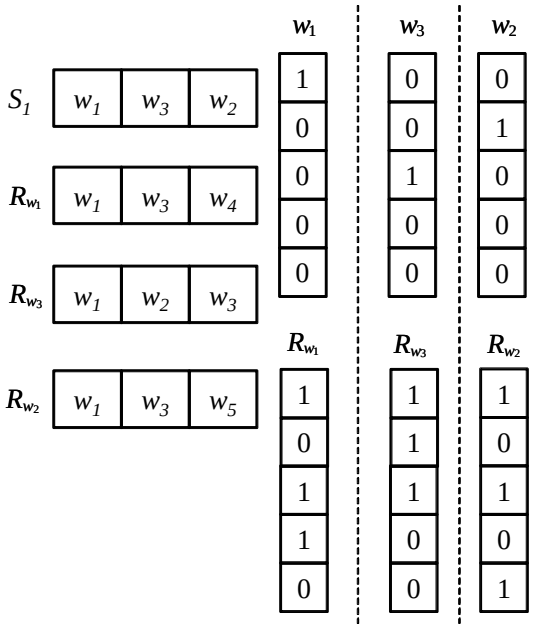
四、 探索遞迴式類神經網路語言模型之改進

(一) 結合關聯資訊於遞迴式類神經網路語言模型

傳統統計式 N 連語言模型是容易使用而且是目前常見的方法之一，但此模型仍有缺乏表示長距離資訊的能力與以及會有資料稀疏的問題。使用類神經網路語言模型能有效解決資料稀疏之問題，可惜在長距離資訊上仍稍嫌不足；因此，遞迴式類神經網路語言模型的發展是希望能取得更多長距離資訊。許多國外研究也顯示遞迴式類神經網路語言模型的確能比一般類神經網路語言模型帶來更好的成效，這也是本論文使用遞迴式類神經網路語言模型來探討之原因。遞迴式類神經網路語言模型中回饋的方式是使用時序性倒傳遞演算法；然而，此方法被證明出此模型無法有效獲得長距離的資訊。因此，本論文將探討透過加入關聯資訊(Relevance Information)來幫助預測下一個詞的可能性。



圖三、語句關聯資訊概念圖



圖四、詞關聯資訊概念圖

關聯資訊則以向量來表示，大小如同原本的輸入層一樣，因此本論文將輸入層擴增爲兩倍，前半段爲原本訓練資料的資訊，後半段爲對應訓練資料的關聯向量。關聯向量主要又分爲兩種，一種爲句子間的關聯，稱做語句關聯資訊(Sentence Relevance Information)，

另一種爲詞跟詞之間的關聯，稱爲詞關聯資訊(Word Relevance Information)。語句關聯資訊的產生，是將欲檢索的句子放進訓練語料中進行一次檢索，檢索完可得知對所有訓練語句的關聯分數，根據此關聯分數我們可以決定要使用多少關聯語句來當作關聯向量。圖三爲語句關聯資訊的概念圖，以句子 S_1 爲例， S_1 中依序包含了詞 w_1 、 w_3 和 w_2 ， R_1 則爲對應 S_1 的語句關聯資訊。但在訓練模型時是以詞爲單位進行訓練，因此賦予每個詞的關聯資訊皆爲此句的語句關聯資訊。 R_{w_1} 、 R_{w_3} 和 R_{w_2} 則爲詞 w_1 、 w_3 和 w_2 所對應的關聯資訊。而詞關聯資訊則是避免使每個詞對應到相同的關聯資訊，其產生的方式是從訓練語料中收集每一個詞左右相鄰文段中，相隔一定距離內其它詞出現的頻率，結果會得知數個關聯的詞，每一個詞的關聯資訊長度也皆不同，其概念如圖四所示。

另外，本論文將關聯資訊以三種不同方式來表示，分別爲詞頻數、正規化後及值設定爲 1 或 0 來進行探討，觀察關聯資訊帶來之影響。以詞頻數表示則代表我們使用實際的次數來增加詞與詞之間的關聯性，具有較真實的資訊。以正規化表示使所有關聯資訊的總和爲 1，此表示法以較公平的方式來給予值，貢獻大的，也就是次數較多的則值越高；反之，貢獻低的則值較低。而以設定爲 1 或 0 來表示，則是將有次數出現的維度以 1 來表示；反之，沒有出現爲 0，因此此種表示法代表詞與關聯詞之間，其關聯程度均相同。

除此之外，也發展了動態詞關聯的方式，由於每個詞所對應的關聯向量是固定的，因此我們將歷史資訊中的關聯資訊做結合，得到新的關聯資訊。其中，因爲每個詞的歷史資訊大部分皆不同，所以結合出來的關聯向量也皆不同。而根據歷史資訊的遠近，分別使用不同權重來做結合，越遠的歷史資訊則隨著時間越來越小，我們可以用遞迴的方式以式(6)來表示。

$$\begin{cases} if \ \ t = 0, & R_t^{'} = R_0 \\ if \ \ 0 < t \le L, & R_t^{'} = (1-\alpha)\cdot R_{t-1} + \alpha \cdot R_t \end{cases} \tag{6}$$

R_t 爲在 t 時間點之原始關聯資訊， $R_t^{'}$ 爲新獲得的關聯資訊， L 爲在該語句所含詞的數目， α 則爲可調控之參數。

一開始時，詞的關聯資訊爲原始的關聯資訊；而當時間點大於 0 且小於等於句子長度時，會與所有歷史詞的關聯資訊做線性結合。因此距離越遠的詞，其關聯資訊的權重就越小，相反地，距離越近的詞，其關聯資訊的權重就越大，因而達到動態效果的詞關聯資訊。本論文的 α 值爲 0.6。以圖五爲例，句子 S_1 中含有五個詞，而圖中爲詞 w_4 的關聯資訊。可看出其關聯資訊爲所有歷史詞的加總，權重部份則取決於詞的距離。

S_1	w_1	w_2	w_3	w_4	w_5

$$R_{w4}^{'} = 0.064R_{w1} + 0.096R_{w2} + 0.24R_{w3} + 0.6R_{w4}$$

圖五、動態詞關聯資訊範例

(二) 語句相關之遞迴式類神經網路語言模型

本節探討藉由動態語言模型調整方式來增進語言模型預測能力；因此，對於不同測試語句使用不同的遞迴式類神經網路語言模型或是結合不同群的遞迴式類神經網路語言模型。

不同於上述的方法裡，所有測試語句皆使用由相同訓練語料訓練出的遞迴式類神經網路語言模型，本論文希望針對各句測試語句，以線性組合的方式結合不同訓練語料所訓練出的遞迴式類神經網路語言模型，期望找出較適合各句測試語句的遞迴式類神經網路語言模型。一開始，先將所有訓練語句的正確轉寫語句以單連詞向量(Unigram Word Vector)表示，並進行分群(Clustering)。本論文所使用的分群方法爲 *K* 平均演算法(*K*-means)[7]，在透過分群過後，我們可用各群所分好的訓練語句來訓練各群的遞迴式類神經網路語言模型，假設所有訓練語句可分爲 *S* 群。接著，對 *S* 群中每一群訓練語句分別算出各自的特徵權重向量，意即平均向量(Mean Vector)，此部分可由每一群中各句訓練語句的單連詞向量進行加總並取平均求得。

可以用式(7)來表示：

$$v_s = \frac{\sum_{k=1}^{L_s} v_{s,k}}{L_s} \tag{7}$$

其中 $v_{s,k}$ 爲第 *S* 群中第 *k* 句訓練語句的單連詞向量，v_s 是第 *S* 群的平均向量，L_s 爲 *s* 訓練語句群中所含訓練語句之句數。如此一來，當有一測試語句需要進行估測時，可以利用此測試語句之單連詞特徵向量和所算好的各群特徵權重向量來求取相似度(Similarity)，並選取欲使用的遞迴式類神經網路語言模型或結合不同群的遞迴式類神經網路語言模型。由於我們無法得知測試語句的正確轉寫語句。在此，測試語句之單連詞特徵向量皆以其 *M* 條最佳辨識結果中的第一名來表示。

本論文主要使用三種選取方式來選取，在計算相似度時，使用的是餘弦值來計算：

$$\cos(U_k, v_s) = \frac{u_k \cdot v_s}{\sqrt{u_k^{\,2}}\sqrt{v_s^{\,2}}} \tag{8}$$

其中 U_k 表示第 *k* 句測試語句，u_k 爲測試語句第 *k* 句中 *M* 條最佳辨識結果第一名之單連詞向量，v_s 爲使用第 *s* 群訓練語句的單連詞特徵向量。以下則介紹三種選取方式：

(1) 選取相似度最大權重法：此方法只選取和測試語句最相似的訓練語句群，也就是所謂相似度最大的。可用式(9)來表示：

$$RNNLM_{U_k} = \arg\max_s \cos(U_k, v_s) \tag{9}$$

其中， $RNNLM_{U_k}$ 代表所挑選出來的遞迴式類神經網路語言模型，因此式(9)則表示挑選 P 群中餘弦值最大之遞迴式類神經網路語言模型。

$$P(U_k) \approx P_{RNNLM_{U_K}}(U_k) \tag{10}$$

因此估測 U_k 的機率便可表示成式(10)。

(2) 相似度線性組合法：此方法之組合係數是經由計算測試語句與 S 群訓練語句間的相似度來求得，因此如果某測試語句和某群相似度較大則表示該群較符合測試語句之特性，也就是該群會有較大之貢獻。而測試語句與各訓練語句群的組合係數 $\gamma_{k,p}$ 爲：

$$\gamma_{k,s} = \frac{\cos(u_k, v_s)}{\sum_{s'=1}^{S} \cos(u_k, v_{s'})} \tag{11}$$

接著將各群之模型分數用此係數線性組合以獲得新的語言模型機率：

$$P(U_k) = \sum_{s=1}^{S} \gamma_{k,s} \cdot P_{RNNLM_s}(U_k) \tag{12}$$

而 $P(U_k)$ 爲測試語句經過線性結合之遞迴式類神經網路語言模型機率， $P_{RNNLMs}(U_k)$ 爲第 s 群之遞迴式類神經網路語言模型機率。

(3) 相似度均勻組合法：

該方法類似相似度線性組合法，將原本組合係數調整成均勻(Uniform)組合係數，因此各群的貢獻度將會相同，則各群均勻組合係數 $\beta_{k,s}$ 爲：

$$\beta_{k,s} = \frac{1}{S} \tag{13}$$

其中 S 爲分群個數。所以最後結合完的分數可以用式(14)來表示：

$$P(U_k) = \sum_{s=1}^{S} \beta_{k,s} \cdot P_{\mathrm{rnnlm}_s}(U_k) \tag{14}$$

我們可以將算好的分數結合原本背景語言模型所預測的機率，以此來得到更好的結果。

五、 實驗結果與分析

(一) 實驗語料

本論文使用之實驗語料是來自於公視新聞(Mandarin Across Taiwan-Broadcast News, MATBN)[8]。公視新聞語料是 2001 年至 2003 年間由中研院資訊所口語小組(SLG)與公共電視(PTS)合作錄製，共計 197 個小時之語音資訊與其內容標記。

選自公視新聞 2001 年至 2002 年外場採訪記者，分別爲訓練集語料 30,600 句(約 23 小時)、測試集語料 1,997 句(約 1.5 小時)及發展集語料 1,998 句(約 1.5 小時)。如表一所示。聲學模型訓練語料爲公視新聞 2001 至 2002 年外場採訪記者語料，共 30,632 句(約 23 小時)，其中包含了實驗訓練語料 30,600 句。

另外背景語言模型使用的訓練語料是來自 2001 至 2002 年中央通訊社(Central News Agency, CNA)的文字新聞語料，內含有約一億五千萬個中文字，經由斷詞之後約有八千萬個詞。此語言模型是使用 SRI Language Modeling Toolkit (SRILM)[9]訓練而得，採用 Katz Back-off 平滑化方法[10]來解決資料稀疏的問題。

短句語料	句數	長度(小時)
訓練集語料	30,600	約 23
發展集語料	1,998	約 1.5
測試集語料	1,997	約 1.5

表一、實驗語料統計資訊

(二) 基礎實驗結果

實驗的產生是從詞圖產生 100(M=100)句最佳候選詞序列，再經由訓練好的遞迴式類神經網路語言模型得到各句詞序列的分數，並加入聲學模型的資訊以此獲得語音辨識的總分數。透過所選出的第一名詞序列和正確答案去算出編輯距離(Edit distance)得到字正確率。遞迴式類神經網路語言模型則是使用 Mikolov 等學者[11]所發展的 Recurrent Neural Network Language Modeling Toolkit (RNNLM)訓練而得。本論文所提出的應用關聯資訊則是對其套件做修改，將關聯資訊加入於遞迴式類神經網路語言模型；語句相關之遞迴式類神經網路語言模型則是利用此套件來訓練遞迴式類神經網路語言模型。而實驗的目的，則是希望透過藉由遞迴式類神經網路語言模型來重新排序找出字正確率最高的詞序列。

語言模型設定方面，在訓練遞迴式類神經網路語言模型時隱藏層個數爲 100、類別層個數也爲 100 、遞迴的次數爲 4 且訓練及辨識過程中，句子和句子之間是獨立的，也就是說上一句的句子和目前訓練的句子是不相關聯的。

	發展集語言複雜度	發展集語言複雜度	發展集語料字正確率(%)	測試集語料字正確率(%)	絕對提昇率(%)	相對提昇率(%)
背景三連語言模型(BG)	450.93	459.06	84.73	83.61	-	-
RNN	607.07	623.50	82.31	82.41	-1.2	-7.32
RNN+BG	232.31	236.97	85.67	85.17	1.56	9.52
Oracle	-	-	93.22	92.66	-	-

表二、遞迴式類神經網路語言模型之基礎實驗結果

表二是關於遞迴式類神經網路語言模型的基礎實驗結果，從語言複雜度的角度來看，遞迴式類神經網路語言模型(RNN)的語言複雜度爲 623.50，再看到背景三連語言模型(BG)的部份，由於訓練語料較多，因此其效果會比遞迴式類神經網路語言模型來的好。而根據文獻中所看到的，遞迴式類神經網路語言模型在獨自使用時效果較不明顯，必須和其它模型做結合，才會有更好的表現。實驗中也可看到背景語言模型結合遞迴式類神經網路語言模型(RNN+BG)效果會來得最好；可以見得，遞迴式類神經網路語言模型仍然具有不錯的成效。

另一部分，我們可以看到 Oracle 部分(意即假使能正確選取到 100 句最佳詞序列中字錯誤率最低的詞序列)，字正確率可到達 93.22%，這意味著我們仍有很大的進步空間。而這部份的趨勢也和語言複雜度相同，單獨使用遞迴式類神經網路語言模型時，辨識率下降 1.2%，但透過與背景語言模型的結合，其絕對提昇率有 1.56%以及相對提昇率 9.52%。

在以下第(3)節和第(4)節的實驗中，我們將使用背景語言模型結合遞迴式類神經網路語言模型(RNN+BG)的語言複雜度 236.97 和辨識率 85.17%作爲 RNN 的基礎辨識率，以此來和我們所提出的方法做比較。

(三) 結合關聯資訊於遞迴式類神經網路語言模型之實驗結果

首先，我們使用語句關聯資訊來幫助遞迴式類神經網路語言模型做估測。在訓練模型時，語句關聯資訊是挑選最相關的訓練語句，由於挑選過多的關聯資訊會導致辨識率下降，因此只挑選最相關的部分。其中，因爲遞迴式類神經網路語言模型是以詞爲單位進行訓練，所以每個詞需要對應到一個關聯向量，此部分的詞的關聯向量爲此句的關聯向量。而關聯向量的表示方式分別使用了句子中詞出現的次數、將詞出現的次數做正規化及出現該詞的維度設爲 1。從表三中可看到語言複雜度部份，以句中詞出現的次數較好，而辨識率方面，則是使用正規化的表示較好，相較於基礎辨識率 85.17%小幅度的進步了 0.04%，而使用句中次數則下降了 0.08%，以及設定爲 1 的方法也下降了 0.14%。

表三是在結合語句關聯資訊中，使用三種表示法的辨識率結果，可看出使用正規化值的表示法較好，其餘兩種的辨識率則較 RNN 基礎辨識率來得低。探究其辨識率進步不大的原因應爲每一語句內，詞的關聯向量皆爲此句的關聯向量，因此關聯向量的重複率就等同於語句中詞的數量。這也成爲了辨識率降低的原因之一，另一個原因則是原本輸入的資訊可能被關聯資訊所干擾。因此我們嘗試將關聯資訊切得更細，使用詞關聯資訊來幫助估測。

關聯資訊表示方式	發展集語言複雜度	測試集語言複雜度	發展集語料字正確率(%)	測試集語料字正確率(%)	絕對提昇率(%)	相對提昇率(%)
RNN 基礎辨識率	232.31	236.97	85.67	85.17	-	-
句中詞出現的次數	223.63	229.01	85.63	85.09	-0.08	-0.56
正規化值	230.51	236.45	85.71	85.21	0.04	0.27
出現該詞則設爲 1 否則爲 0	226.04	231.19	85.56	85.03	-0.14	-0.95

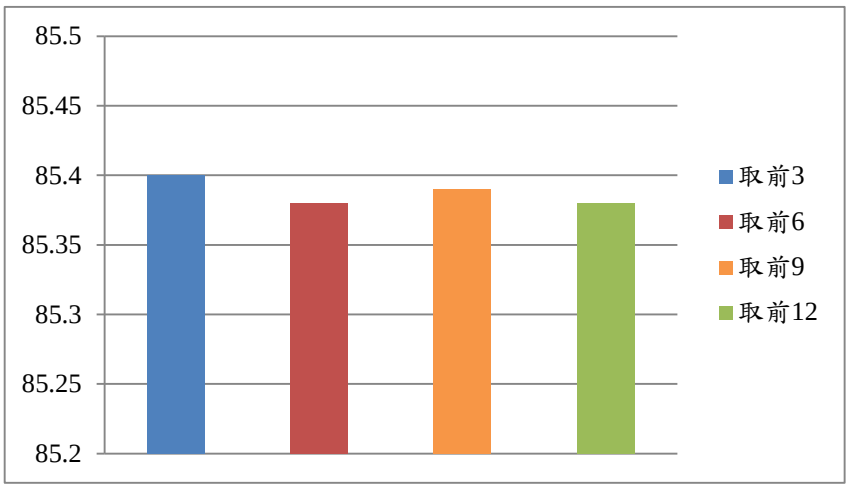
表三、結合語句關聯資訊之實驗結果

詞關聯資訊不同於語句關聯資訊之處，就是每一語句中的詞有各自的關聯資訊。詞關聯資訊的產生，是將訓練語料中，詞出現的地方找其相鄰詞作爲關聯資訊，出現相同的相鄰詞則會累加出現次數，本論文則是取左右距離爲 3。

關聯資訊表示方式	發展集語言複雜度	測試集語言複雜度	發展集語料字正確率(%)	測試集語料字正確率(%)	絕對提昇率(%)	相對提昇率(%)
RNN 基礎辨識率	232.31	236.97	85.67	85.17	-	-
詞出現的次數	230.05	234.93	85.88	85.36	0.19	1.26
正規化值	230.13	234.75	85.83	85.34	0.17	1.16
出現該詞則設爲 1 否則爲 0	230.12	234.83	85.77	85.40	0.23	1.52

表四、結合詞關聯資訊之實驗結果

表四使用詞關聯資訊於遞迴式類神經網路語言模型的實驗結果，由於部分詞的詞關聯資訊相當多，包含了關聯詞和非關聯的詞，因此，我們試著去調整詞關聯資訊的使用程度，其中詞關聯資訊的長度是根據發展集中最好的結果來設定。語言複雜度方面，詞出現次數、正規化值及出現該詞則設爲 1 否則爲 0 也都有進步；而字正確率方面，詞出現次數、正規化值及出現該詞則設爲 1 否則爲 0 均有提昇，絕對提昇率分別爲 0.19%、0.17%及 0.23%，相對提昇率則有 1.26%、1.16%及 1.52%的進步。比較三種表示法的辨識率，可看到出現該詞設爲 1 否則爲 0 的辨識結果較好，因爲此方法對於每個關聯詞的關聯度較公平，大家皆設定爲 1。而詞出現的次數和正規化值，因爲每個關聯詞之間的歧異度較高，尤其是詞出現的次數，次數的差距很大，導致有些關聯詞的貢獻被埋沒。



圖七、使用不同長度之詞關聯資訊辨識結果

另外，此部份實驗結果也顯示，使用出現該詞設爲 1 且調整詞關聯資訊的長度較好；於是我們進一步去觀察使用此表示法在不同詞關聯資訊長度上的比較。圖七則是其辨識結果，可以看出使用過多的資訊反而會導致效果減弱。

雖然從實驗結果中得知，使用詞關聯資訊的確提升了辨識率，但我們仍希望可以突破目前的瓶頸。於是我們發現到，雖然詞關聯資訊解決了語句關聯資訊的問題，但是仍有類似語句關聯資訊的缺點存在，其缺點則是語料中的每個詞所對應到的詞關聯資訊仍然一樣，造成在訓練中重複使用同樣的詞關聯資訊。此作法的詞關聯資訊是比較屬於全域(Global)的，也就是是針對所有訓練語料中，獲得詞與詞的關聯度。而在訓練遞迴式類神經網路語言模型中，我們也需要區域性的資訊，因爲相同的詞在不同的句子可能代表著不同的意思，所以我們希望藉由區域性的資訊來獲得上下文或句子中的資訊，使得在預測下一個詞時能夠符合句子中的意思。因此，我們希望詞的關聯資訊必須是會變動的或是動態的，如此一來才能包含全域性的資訊和區域性的資訊。爲此，本論文提出了動態詞關聯資訊來做更進一步的改進，此部分是先將所有歷史詞的關聯資訊做結合，其結合依據遠近來給予權重。因此，歷史詞與目前的詞距離越遠，則該歷史詞的關聯資訊貢獻越小；反之，歷史詞與目前的詞距離越近，則該歷史詞的關聯資訊越大。表五則是使用動態詞關聯資訊的實驗結果，結果顯示使用動態詞關聯資訊效果與一般的詞關聯資訊較差一點，辨識率大

約爲 85.34%左右；而其語言複雜度表現則較詞關聯資訊好。探究其原因，應爲關聯資訊中常包含關聯與非關聯的資訊，因此我們難以準確知道越近距離的詞關聯資訊有較相關的資訊，造成使用的關聯資訊無法正確代表與該詞相關。

根據本論文所提出的結合關聯資訊於遞迴式類神經網路語言模型的確有助於辨識率的提升，但是其效果仍有限且不是那麼的明顯，其原因大概歸類爲三種，其一是關聯資訊可能會對輸入層所要傳遞的資訊造成干擾，使得輸入層所要傳遞的資訊減弱，而關聯資訊被成爲主要傳遞的資訊;其二是關聯資訊結合輸入層，也可能只是將其表示方式做了延伸，而關聯資訊的表示法可能有更佳的表示方法；其三則是難以準確的決定關聯資訊，導致效果無法彰顯。

關聯資訊表示方式	發展集語言複雜度	測試集語言複雜度	發展集語料字正確率(%)	測試集語料字正確率(%)	絕對提昇率(%)	相對提昇率(%)
RNN 基礎辨識率	232.31	236.97	85.67	85.17	-	-
詞出現的次數	229.72	234.35	85.84	85.26	0.09	0.58
正規化值	231.42	236.19	85.83	85.34	0.17	1.14
出現該詞則設爲 1 否則爲 0	229.98	234.62	85.86	85.34	0.17	1.16

表五、結合動態詞關聯資訊之實驗結果

(四) 語句相關之遞迴式類神經網路語言模型之實驗結果

本節主要探究利用語句相關之遞迴式類神經網路語言模型之實驗結果，以下表六是將訓練語料分爲兩群用於發展集語料之結果，表七是將發展集語料的最佳設定用在測試語料的辨識結果；表八與表九，分別爲將訓練語料分爲四群的發展集語料與測試語料測試結果。

表七爲分成兩群後的字正確率，分別爲兩群使用選取相似度最大權重法、使用相似度線性組合法與使用相似度均勻組合法的字正確率。此部分我們會額外使用一個全部訓練語料所訓練出的遞迴式類神經網路語言模型來做輔助，稱爲遞迴式類神經網路背景語言模型。首先利用三種權重組合方式將各群做結合，接著再將結合完的結果加入遞迴式類神經網路背景語言模型，最後才結合背景語言模型。比較表六可以看到利用選取相似度最大權重法會有較好的結果與基礎遞迴式類神經網路語言模型相比，絕對提昇率有 0.24%的進步。

RNNLM 權重參數	選取相似度 最大權重法	相似度線性組合法	相似度均勻組合法
0	84.29	84.29	84.29
0.1	85.64	85.68	85.58
0.2	85.87	85.88	85.94
0.3	85.86	85.92	85.94
0.4	85.90	85.91	85.81
0.5	85.81	85.78	85.74
0.6	85.70	85.57	85.44
0.7	85.46	85.43	85.29
0.8	85.05	84.97	84.99
0.9	84.60	84.55	84.52
1	82.56	82.62	82.59

表六、發展集語料字正確率之兩群辨識結果

	字正確率(%)	絕對提昇率(%)	相對提昇率(%)
選取相似度最大權重法	85.41	0.24	1.60
相似度線性組合法	85.24	0.07	0.46
相似度均勻組合法	85.29	0.12	0.82

表七、測試語料字正確率之兩群辨識結果

而選取相似度均勻組合法與相似度線性組合法則略差於基礎辨識率，探究其原因應爲資料較偏向某一群，因此使用相似度均勻組合法與相似度線性組合法則較差，只要使用相似度最大的那群就能有較好的效果。

表九則是分成四群。將訓練語料分四群來訓練的實驗中，也可得知使用相似度最大權重法是較佳的。但是跟兩群的實驗相比，雖然與基礎辨識率相比仍有進步，三種選取方法還是較差一點。探討其原因應爲訓練語料不足的關係。由於分群數目提高，則每群中的訓練語料則隨之減少。因此無法訓練出學習能力較佳的遞迴式類神經網路語言模型，導致字正確率的下降。而實驗中也可看出，結合完各群結果後的辨識率仍不好，需要加入遞迴式類神經網路背景語言模型來輔助，以得到更好的辨識率。

RNNLM 權重參數	選取相似度 最大權重法	相似度線性組合法	相似度均勻組合法
0	84.29	84.29	84.29
0.1	85.65	85.55	85.59
0.2	85.73	85.73	85.71
0.3	85.86	85.86	85.78
0.4	85.87	85.81	85.69
0.5	85.79	85.77	85.60
0.6	85.61	85.50	85.52
0.7	85.39	85.35	85.35
0.8	85.15	85.13	85.05
0.9	84.71	84.66	84.63
1	82.75	83.39	83.31

表八、發展集語料字正確率之四群辨識結果

	字正確率(%)	絕對提昇率(%)	相對提昇率(%)
選取相似度最大權重法	85.33	0.16	1.06
相似度線性組合法	85.31	0.14	0.94
相似度均勻組合法	85.32	0.15	1.02

表九、測試語料字正確率之四群辨識結果

六、 結論與未來展望

傳統 N 連語言模型是目前語言模型當中常見的方法之一，但是卻難以捕捉到長距離的語句資訊，加上擁有資料稀疏和維度的詛咒之特性，長期以來一直難以突破。近年國外學者的研究發現類神經網路語言模型有不錯的成效，不僅能擁有 N 連語言模型的特性也能解決資料稀疏缺點，爲語音辨識與語言模型帶來嶄新的視野；然而類神經網路語言模型也仍存在一些缺點，例如外詞彙問題(out-of-vocabulary, OOV)、缺乏長距離資訊、運算的時間複雜度過高以及詞的表示方式缺少了詞的特性等問題。因此，也有學者針對類神經網路的變形，使用了具有遞迴能力的類神經網路來建構語言模型，而效果也比一般類神經網路語言模型好。

本論文針對遞迴式類神經網路語言模型做了更進一步的改善，期望使用關聯資訊和動態調整語言模型來輔助機率的估測。從實驗結果中可以看出使用關聯資訊的確能帶來幫助，但是效果仍不夠明顯，其原因應爲輸入層或前一時間點的資訊被關聯資訊所干擾，導致成效有限。而實驗中也發現到減少部分關聯資訊能提升辨識率，因此關聯資訊或其他資訊的表示法在未來研究上也是值得注意的部分。另一部分，本論文藉由將訓練語料分群並訓練各群的遞迴式類神經網路語言模型，期望藉由動態的調整語言模型來達到更好的辨識率。此部分實驗結果也顯示分兩群時，使用相似度線性組合法有較佳的成效。但分成四群時，由於各群中的訓練語料不足，因此無法訓練出學習能力較佳的遞迴式類神經網路語言模型。

在未來的研究裡，可以根據遞迴式類神經網路語言模型無法有效學習長距離資訊之缺點來進行改善，如加入不同的特徵或其他資訊來幫助估測，抑或是針對時序性倒傳遞演算法的缺點進行結構上的改進。而隨著時代的變遷，語言也不斷地在進化，許多以前沒有的詞語也不停出現，因此用不同平滑化的方法來處理外詞彙問題也是相當重要的議題。另外，與現行的語言模型結合，如主題模型或鑑別式語言模型等，使語言模型更具有一般性能力、適應性能力，甚至鑑別性能力也是將來值得探討的部分。由於鑑別式語言模型的概念和類神經網路語言模型相當的像，差別在於前者是監督式的，後者是非監督式的。而倘若將類神經網路語言模型改良成監督式的方法，則辨識率應該會有更好的提升，期望在未來能將此兩種語言模型做結合，並進一步的獲得更好的辨識結果。

致謝

本論文之研究承蒙教育部－國立台灣師範大學邁向頂尖大學計畫（101J1A0900 和 101J1A0901）與行政院國家科學委員會研究計畫(NSC 101-2221-E-003 -024 -MY3 和 NSC 99-2221-E-003 -017 -MY3)之經費支持，謹此致謝。

參考文獻

[1] Y. Bengio, P. Frasconi, and P. Simard, “The problem of learning long-term dependencies in recurrent networks,” in *Proc. IEEE International Conference on Neural Networks*, Vol. 3, pp. 1183-1188, 1993.

[2] Y. Bengio, R. Ducharme, P. Vincent, C. Jauvin, J. K, T. Hofmann, T. Poggio, and J. Shawetaylor. A neural probabilistic language model. In Journal of Machine Learning Research, 2003.

[3] J. L. Elman, “Finding structure in time,” *Cognitive Science*, Vol. 14, No. 2, pp. 179-211, 1990.

[4] M. L. Jordan, “Attractor dynamics and parallelism in a connectionist sequential machine,” in *Proc. the eighth annual conference of the cognitive science society*, pp.531-546, 1986

[5] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, Vol. 45, No. 11, pp. 2673-2681, 1997.

[6] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transaction on Neural Networks*, Vol. 5, No. 2, pp. 157-166, 1994.

[7] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability*, pp. 281–297.

[8] H.-M. Wang, B. Chen, J.-W. Kuo and S.-S. Cheng, "MATBN: A Mandarin Chinese broadcast news corpus," *International Journal of Computational Linguistics & Chinese Language Processing*, Vol. 10, No. 2, pp. 219-236, 2005.

[9] Stolcke, Andreas. Srilm - an extensible language modeling toolkit. In Proceedings of the International Conference on Spoken Language Processing, Denver, Colorado, September 2002.

[10] S. M. Katz, "Estimation of probabilities from sparse data for the language model component of a speech recognizer," in *Proc. IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-35, No. 3, pp. 400, 1987.

[11] T. Mikolov, S. Kombrink, A. Deoras, L. Burget and J. Černocký, "RNNLM - Recurrent neural network language modeling toolkit," in *Proc. IEEE workshop on Automatic Speech Recognition and Understanding*, 2011