The 2015 Conference on Computational Linguistics and Speech Processing
ROCLING 2015, pp. 4-17

使用詞向量表示與概念資訊於中文大詞彙連續語音辨識之語言模型調適

Exploring Word Embedding and Concept Information for Language Model Adaptation in Mandarin Large Vocabulary Continuous Speech Recognition

陳思澄 Ssu-Cheng Chen, 洪孝宗 Hsiao-Tsung Hung, 陳柏琳 Berlin Chen
國立臺灣師範大學資訊工程學系
Department of Computer Science and Information Engineering
National Taiwan Normal University
{60247071S, 60047064S, berlin}@ntnu.edu.tw

陳冠宇 Kuan-Yu Chen
中央研究院資訊科學研究所
Institute of Information Science, Academia Sinica
Kychen@iis.sinica.edu.tw

摘要

近年來深度學習(Deep Learning)激起一股研究熱潮；隨著深度學習的發展而有分散式表示法(Distributed Representation)的產生。此種表示方式，不僅能以較低維度的向量表示詞彙，還能藉由向量間的運算，找出任兩詞彙之間的語意關係。本論文以此為發想，提出將分散式表示法，或更具體來說是詞向量表示(Word Representation)，應用於語音辨識的語言模型中使用。首先，在語音辨識的過程中，對於動態產生之歷史詞序列與候選詞改以詞向量表示的方式來建立其對應的語言模型，希望透過此種表示方式而能獲取到更多詞彙間的語意資訊。其次，我們針對新近被提出的概念語言模型(Concept Language Model)加以改進；嘗試在調適語料中以句子的層次做模型訓練資料選取之依據，去掉多餘且不相關的資訊，使得經由調適語料中訓練出的概念類別更為具代表性，而能幫助動態語言模型調適。另一方面，在語音辨識過程中，會選擇相關的概念類別來動態組成概念語言模型，而此是透過詞向量表示的方式來估算，其中詞向量表示是由連續型模型(Continue Bag-of-Words Model)或是跳躍式模型(Skip-gram Model)生成，希望藉由詞向量表示記錄每一個概念類別內詞彙彼此間的語意關係。最後，我們嘗試將上述兩種語言模型調適方法做結合。本論文是基於公視電視新聞語料庫來進行大詞彙連續語音辨識(Large Vocabulary Continuous Speech Recognition, LVCSR)實驗，實驗結果顯示本論文所提出的語言模型調適方法相較於當今最好方法有較佳的效用。

關鍵詞：語音辨識、語言模型、詞向量表示、概念模型

Abstract

Research on deep learning has experienced a surge of interest in recent years. Alongside the rapid development of deep learning related technologies, various

distributed representation methods have been proposed to embed the words of a vocabulary as vectors in a lower-dimensional space. Based on the distributed representations, it is anticipated to discover the semantic relationship between any pair of words via some kind of similarity computation of the associated word vectors. With the above background, this article explores a novel use of distributed representations of words for language modeling (LM) in speech recognition. Firstly, word vectors are employed to represent the words in the search history and the upcoming words during the speech recognition process, so as to dynamically adapt the language model on top of such vector representations. Second, we extend the recently proposed concept language model (CLM) by conduct relevant training data selection in the sentence level instead of the document level. By doing so, the concept classes of CLM can be more accurately estimated while simultaneously eliminating redundant or irrelevant information. On the other hand, since the resulting concept classes need to be dynamically selected and linearly combined to form the CLM model during the speech recognition process, we determine the relatedness of each concept class to the test utterance based the word representations derived with either the continue bag-of-words model (CBOW) or the skip-gram model (Skip-gram). Finally, we also combine the above LM methods for better speech recognition performance. Extensive experiments carried out on the MATBN (Mandarin Across Taiwan Broadcast News) corpus demonstrate the utility of our proposed LM methods in relation to several well-practiced baselines.

Keywords: speech recognition, language modeling, deep learning, word representation, concept language model

一、 緒論

語言模型(Language Models, LM)不僅在語音辨識中扮演重要的角色，還可以應用至資訊檢索、機器翻譯、手寫辨識以及文件摘要等不同任務之中，成為關鍵的組成[1, 2]。在語音辨識過程中，我們通常會透過語言模型來補足聲學模型經常不能充分應付同音異字或發音混淆的情況，並幫助語音辨識系統從眾多混淆的候選詞序列假設(Candidate Word Sequence Hypotheses)中找出最有可能的結果[3, 4]。*N* 連(*N*-gram)語言模型為語音辨識之中最為常見的統計式語言模型，用來估測每一個待預測詞彙在其先前緊鄰的 *N*-1 個詞彙已知的情況下出現的條件機率；假設每一個詞彙出現的機率僅與它緊鄰的前 *N*-1 個詞彙相關，可以透過多項式分布(Multinomial Distribution)來表示。然而 *N* 連語言模型僅能擷取短距離的詞彙規則資訊，而無法考慮長距離的語句或篇章資訊；當詞序列越長時參數量越多，使得 *N* 連語言模型會有維度詛咒的問題。另一方面，*N* 連語言模型亦容易面臨訓練語料與測試語料不匹配(Mismatch)而造成估測誤差。有鑑於此，近十幾年來許多動態語言模型調適技術被提出，用以發展有效的語言模型輔助並彌補傳統 *N* 連(*N*-gram)語言模型不足之處。常見的有快取模型(Cache Model)[5]，以及在資訊檢索領域的主題模型(Topic Model)[6]等。其中又以機率式潛藏語意分析(Probabilistic Latent Semantic Analysis, PLSA)[7]以及其延伸狄利克里分配(Latent Dirichlet Allocation, LDA)[8]最為普遍被使用。

本論文旨在於發展新穎動態語言模型調適技術，用以輔助並彌補傳統 *N* 連(*N*-gram)語言模型不足之處。首先，我們提出將分散式表示法之詞向量表示(Word

Representation or Embedding)應用於語音辨識的語言模型中使用。在語音辨識的過程中，對於動態產生之歷史詞序列(Word History)與候選詞(Candidate Word)改以詞向量表示的方式來建立其對應的語言模型，希望透過詞向量表示而能獲取到更多詞彙間的語意資訊。其次，我們針對新近被提出的概念語言模型(Concept Language Model)加以改進；嘗試在調適語料中以句子的層次做模型訓練資料挑選之依據，去掉多餘且不相關的資訊，使得經由調適語料中挑選出的概念類別更為具代表性，而能幫助動態語言模型調適。另一方面，在選擇相關的概念類別來動態組成概念語言模型時，而此是透過詞向量表示的方式來估算，其中詞向量表示是由連續型模型(Continue Bag-of-Words Model)或是跳躍式模型(Skip-gram Model)生成，希望藉由詞向量表示記錄每一個概念類別內詞彙彼此間的語意關係。最後，我們嘗試將上述兩種語言模型調適技術做結合。本論文是基於公視電視新聞語料庫來進行中文大詞彙連續語音辨識(Large Vocabulary Continuous Speech Recognition, LVCSR)實驗，比較本論文所提出語言模型調適技術與其它當今常用語言模型調適技術之效能。

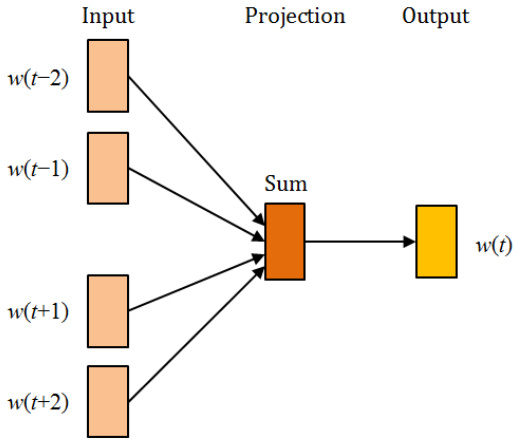
本論文的後續安排如下：第二節介紹詞向量表示法以及本論文嘗試將詞向量表示應用於詞圖搜尋中之方法；第三節介紹將詞向量表示資訊融入概念語言模型；第四節介紹實驗語料、實驗設定以及實驗結果分析；第五節則是結論及未來展望。

二、 詞向量表示法應用於詞圖搜尋之中

在自然語言中，最常見也是最為直覺的詞表示方式為 One-hot Representation，亦即將每個詞表示成一個很長的 N 維向量。其中 N 為詞彙的大小，而向量中僅有其中一維的值為 1，用來表示當前的詞，其餘則表示為 0。此種表示方式是採用稀疏的方式來儲存，並假設兩兩詞彙間彼此獨立，所以從此向量中並無法找出兩兩詞彙之間的關係。

因此於 1986 年時，Hinton 提出了分散式表示法(Distributed Representation) [9]做為詞的表示法，是透過前饋式類神經網路(Feed-Forward Neural Network)訓練而成。這種向量表示是將詞表示成一個較低維度的實數向量。每個詞彙之間的關係可以利用餘弦或是歐式距離計算找出兩個詞向量間的語意相似度，我們將這些詞向量稱為詞表示法(Word Representation or Embedding)。

有鑑於使用傳統類神經網路語言模型來訓練詞向量會造成訓練時間過長，Tomas Mikolov 等人[10]於是提出所謂的連續型詞袋模型(Continuous Bag-of-Words Model, CBOW)與跳躍式模型(Skip-Gram Model, SG)，這兩種模型使用階層軟式最大化(Hierarchical Soft-max, HS)[10]以及負例採樣(Negative Sampling, NS) [11]方法來提高訓練的速度並改善訓練後詞向量的表示能力。



圖一、連續型詞袋模型示意圖

(一)、連續型模型

連續型詞袋模型(CBOW)與前饋式類神經網路類似，不同之處在於連續型詞袋模型將非線性隱藏層(Non-Linear Hidden Layer)移除，並且在輸入層的所有單詞皆共享隱藏層。如圖一所示，此模型包含三層，分別為輸入層、投影層、輸出層。已知當前詞 w_t 的上下文$w_{t-2},w_{t-1},w_{t+1},w_{t+2}$ 的情況下預測當前詞w_t出現的機率。在此目標函數為最大化訓練語料庫中所有詞彙平均的發生機率:

$$\frac{1}{T}\sum_{t=k}^{T-k} log\, P(w_t|w_{t-k},\ldots,w_{t+k}) \tag{1}$$

其條件機率可以透過 Softmax 函數轉換為:

$$P(w_t|w_{t-k},\ldots,w_{t+k})=\frac{e}{\sum_i e^{y_i}} \tag{2}$$

其中 $\mathbf{y}=\{y_1,\ldots,y_v\}$，而 $\mathbf{y}$ 中的每個 y_i 為對於每一個詞 w_i還未經過正規化的 *log* 機率值，計算如下式:

$$\mathbf{y}=\mathbf{b}+Uh(w_{t-k},\ldots,w_{t+k},X) \tag{3}$$

其中U、$\mathbf{b}$為 Softmax 的參數，h 是從矩陣 X 中的詞向量($\vec{w}_{t-k},\ldots,\vec{w}_{t+k}$)加總平均，$X$為根據每個詞$w_i$的向量所組成的矩陣。

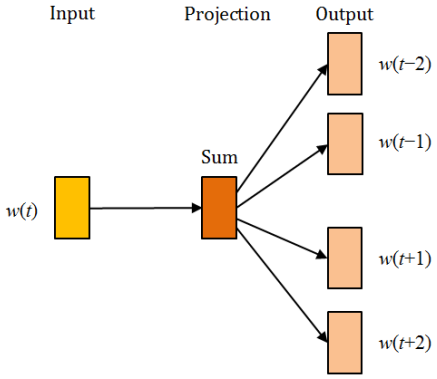
(二)、跳躍式模型

跳躍式模型(Skip-gram)與連續型詞袋模型(CBOW)相反，使用當前的詞來預測周圍的詞。在已知當前詞 w_t 的情況下，預測其上下文 $w_{t-2},w_{t-1},w_{t+1},w_{t+2}$ 的機率。給定一段詞序列 $w_1,w_2,w_3,\ldots,w_t$ ，在此最大化目標函數:

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-c\le k\le c,k\neq 0} \log P(w_{t+k}|w_t) \tag{4}$$

其中 c 為訓練上下文的窗口大小，T 為訓練的文字語料長度，$P(w_{t+k}|w_t)$ 表示在當

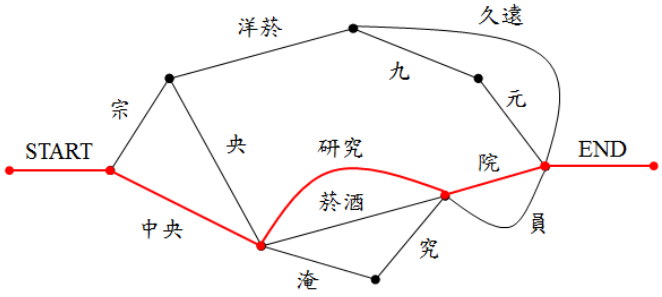
前詞 w_t 的條件下 w_{t+k} 出現的機率。計算在一個固定的窗口大小內兩兩詞彙之間的機率，可以用來找出在一段語句中詞彙彼此之間的相互關係。上下文的窗口越大，預測的結果越精準，相對的訓練時間亦會隨之增加。



圖二、跳躍式模型示意圖

(三)、將詞向量表示應用於詞圖搜尋

在語音辨識的過程中，每個音框會記錄語言模型的歷史詞序列、候選詞對應的開始與結束的音框、以及搜尋時聲學模型的解碼分數，來建立詞圖(Word Graph)，並在詞圖上使用三連詞(Trigram)或四連詞(Fourgram)等類似語言模型，在重新進行一次詞圖動態規劃搜尋(Word Graph Rescoring)中，找出一條最佳的辨識詞序列，如圖三所示。



圖三、詞圖搜尋示意圖

詞圖是由詞彙樹複製搜尋過後所建立的圖，而詞圖中的每個分支(Arc)表示經過裁剪過後所保留的詞段，每個詞段會記錄其聲學分數。接著針對每個詞段進行維特比(Viterbi)搜尋，並記錄與每個詞段相連且最有可能的下一個詞段(亦即前詞段之結束時間與下一詞段的開始時間相同並且維特比分數為最高者) 。然而從詞圖中所保留的詞段，在聲學模型中大多為同音異字或是混淆的，所以需要透過語言模型的輔助。

在詞圖搜尋時，給定歷史詞序列　下預測當前詞 w_i 的機率可以由下式表示:

$$(\quad|\quad)=\sum_{\in}\qquad(\quad|\quad) \tag{5}$$

在此加入參數 α_j，並且假設參數 $_1$，$_2$,…，加總為 1，使得距離詞 越近的詞給予較大權重，亦即在歷史詞序列中越靠近當前詞 的詞越重要。(|)表示在給定歷史詞序列 H_i 中詞 下預測當前詞 的機率，可以由(6)式得到:

$$(w_i|w_m)=\frac{e^{\vec{w}_i\cdot\vec{w}_m}}{\sum_{W\in V}e^{\vec{w}_i\cdot\vec{W}}} \tag{6}$$

其中 $\vec{w}_i$ 為當前詞 w_i 的詞向量表示，$\vec{w}_m$ 為詞圖中的候選詞 w_m 的詞向量表示，而 W 為對於詞 w_i 的所有候選詞集合，最後透過 Softmax 函數將其轉換為機率的方式表示。

三、 將詞向量表示資訊融入概念語言模型

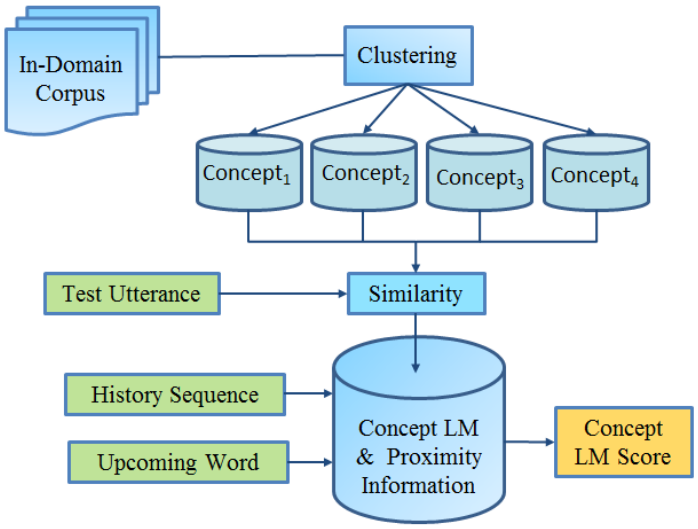
在 2014 年時，有學者[12]提出了概念語言模型(Concept Language Model, CLM)，其想法是認為一般人在表達一件事物時，所講的每一語句背後都隱含語者內心欲表達的概念，希望藉由近似語者欲表達的概念，預測在此概念下的詞彙使用分布資訊，並將不同程度的鄰近資訊(Proximity Information)融入概念語言模型以放寬詞袋(Bag-of-Words)假設的限制，達到動態語言模型調適的效果。

概念語言模型假設在調適語料的文件集內之文件可以由一組概念類別 **C** 來表示，藉由語者講欲表達的語言資訊 W 與這些概念類別的個別關聯程度來獲得語句可能的概念分布，並做為語言模型預測的根據，如式(7)所示:

$$\begin{aligned}P_{\mathrm{CLM}}(w_i\,|\,H_i,W)&=\frac{\sum_{C\in\mathbf{C}}P(w_i,H_i\,|\,C)P(C\,|\,W)}{\sum_{C'\in\mathbf{C}}P(H_i\,|\,C')P(C'\,|\,W)}\\&=\frac{\sum_{C\in\mathbf{C}}P(w_i\,|\,C)\prod_{l=1}^{L_i}P(h_l\,|\,C)P(C\,|\,W)}{\sum_{C'\in\mathbf{C}}\prod_{l'=1}^{L_i}P(h_{l'}\,|\,C')P(C'\,|\,W)}\end{aligned} \tag{7}$$

其中概念類別的求取可透過 *K*-Means 演算法[13]求得；$P(C\,|\,W)$ 可基於將語言資訊 W 與每一個概念類別 C 表示成向量形式，計算 W 與 C 之餘弦相似度求得；$P(w_i\,|\,C)$ 代表概念類別 C 預測詞彙 w_i 的單連語言模型機率，可透過最大化相似機率估測(Maximum Likelihood Estimation, MLE)。我們可以將式(7)中概念類別 C 預測詞彙 w_i 的語言模型延伸成為詞雙連(Word Bigram)或者詞三連(Word Trigram)語言模型，概念語言模型可以同時考慮詞彙間出現的先後規則性或是鄰近資訊(Proximity Information)，以免除詞袋(Bag-of-Words)假設的限制。例如，當使用雙連資訊時，所形成的概念語言模型(記作 BCLM)如式(8)所示:

$$\begin{aligned}&P_{\mathrm{BCLM}}(w_i\,|\,H_i,W)=\\&\quad\frac{\sum_{C\in\mathbf{C}}P(w_i\,|\,h_L,C)P(h_1\,|\,C)\prod_{l=2}^{L_i}P(h_l\,|\,h_{l-1},C)P(C\,|\,W)}{\sum_{C'\in\mathbf{C}}P(h_1\,|\,C')\prod_{l'=2}^{L_i}P(h_{l'}\,|\,h_{l'-1},C')P(C'\,|\,W)}\end{aligned} \tag{8}$$



圖四、概念語言模型流程圖

(一)、結合詞向量表示與概念資訊於語言模型

本論文將詞向量表示法融入概念語言模型中，並以式(8)所示的詞雙連概念語言模型(BCLM)為例。首先，在調適語料文件集內之文件由一組概念類別 **C** 來表示，以群聚之間的相似度近似語句概念表達的涵意。在調適語料中以句子的層次做模型訓練資料選取之依據，將具有相似語意或是相同概念的語句歸為同一個類別中，使得經由調適語料中訓練出的概念類別更為具代表性。其中 *W* 代表語者所講語句欲表達的語言資訊，在此以語音辨識初步所產生的詞圖(Word Graph)來近似。

而 $P(C|W)$是透過語言資訊 W 與每一個概念類別 C ，以詞向量表示(Word Embedding)的方式，先將詞轉換成向量的形式，接著計算其餘弦相似度而得。其中詞向量表示是由連續型模型(Continue Bag-of-Words Model)或是跳躍式模型(Skip-gram Model)生成。 (|)表示概念類別 *C* 預測詞彙 的單連語言模型機率，可以透過最大化相似機率估測而得。

四、 實驗設定與結果討論

(一)、實驗語料

本研究所進行之語音辨識實驗是使用台師大所自行研發的大詞彙連續語音辨識系統(詞典大小約為 7 萬 2 千詞)[14]以及公視電視新聞語音語料庫(Mandarin Across Taiwan Broadcast News, MATBN)[15]。此新聞語音語料庫是由中央研究院資訊所口語小組耗時三年(2001~2003)與公共電視台[PTS]合作錄製完成。我們初步選擇外場採訪記者語料作為實驗題材，將其中約 25 小時收錄於 2001 年 11 月至 2002 年 12 月期間的語料作為最小化音素錯誤(Minimum Phone Error, MPE)聲

學模型訓練的語料來建立聲學模型(Acoustic Models)[16]。本論文以 2003 年所蒐集的語料中挑選約 1.5 個小時，包含 292 句語句。

在語言模型的估測上，我們使用自 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)做為背景語料庫用來訓練三連語言模型(Trigram Language Model)，此語言模型是使用 SRI Language Modeling Toolkit (SRILM)[17]訓練而得，採用 Good-Turning 平滑化方法來解決資料稀疏的問題。另一方面，我們亦蒐集同為公視電視新聞語料庫中的同領域文件做為調適語料庫，用來估測本論文所探討的各式做為調適之用的語言模型，總共約三千六百四三句語句。本論文實驗所使用之語音語料庫以及文字語料庫的扼要統計資訊分別如表一與表二所示。

表一 語音辨識實驗使用之語音語料統計資訊

	詞典大小	句數	長度(小時)	說話速度
語料	約 72000 詞	292	約 1.5	8.52 字/秒

表二 語言模型估測所使用背景文字語料以及調適文字語料統計資訊

語料	詞數	句數
調適語料	約 1,000,000	3,643
背景語料	約 80,000,000	2,068,991

(二)、基礎實驗結果

在基礎實驗部分，首先僅使用背景語言模型於中文大詞彙連續語音辨識，觀察其字辨識錯誤率(Character Error Rate, CER)，我們亦比較同領域語料訓練的語言模型結合背景語言模型的字錯誤率。另外，我們以詞圖最佳解碼(Oracle)作為語音辨識效能的上界；詞圖中最佳解碼是利用動態規劃方式，找出詞圖中字錯誤率最低之路徑。基礎實驗於測試集之字辨識率結果如表三所示。

表三、語音辨識基礎實驗之字辨識率(%)結果

	字錯誤率(%)
背景單連語言模型(UBG)	34.30
背景雙連語言模型(BBG)	22.24
背景三連語言模型(TBG)	20.22
同領域雙連語言模型+TBG	19.12
同領域三連語言模型+TBG	19.04
詞圖中最佳解碼(Oracle)	7.72

(三) 將詞向量表示應用於詞圖搜尋之實驗結果

本論文希望利用詞向量表示找到詞彙間彼此的語意關係，利用詞向量表示於語音辨識的詞圖搜尋中，希望藉此能達到提升辨識率的效果。表四為比較不同維度以及不同詞向量表示(Skip-gram, CBOW)於詞圖搜尋的字錯誤率結果，在此維度設定以 10 至 50 作為實驗之比較，以較小維度之差異比較，減少其計算複雜度。

表四、應用詞向量表示於詞圖搜尋中之字錯誤率(%)比較表

維度大小	跳躍式模型(Skip-gram)	連續型詞袋模型（CBOW)
10	19.85	19.86
20	19.85	19.87
30	**19.83**	**19.84**
40	19.85	19.86
50	19.85	19.84

由表四中可以看出融入詞向量表示的資訊於詞圖搜尋中，我們可以很明顯地觀察出，加入詞向量的資訊對於語音辨識準確率的提升有幫助。不論是用跳躍式模型(Skip-gram)還是使用連續型詞袋模型(CBOW)所訓練得到的詞向量表示，將其應用於語音辨識的詞圖搜尋之中，字錯誤率從原本只使用詞圖搜尋時之字錯誤率 20.2 下降至 19.83 (使用 Skip-gram)，獲得不錯的效能提升。

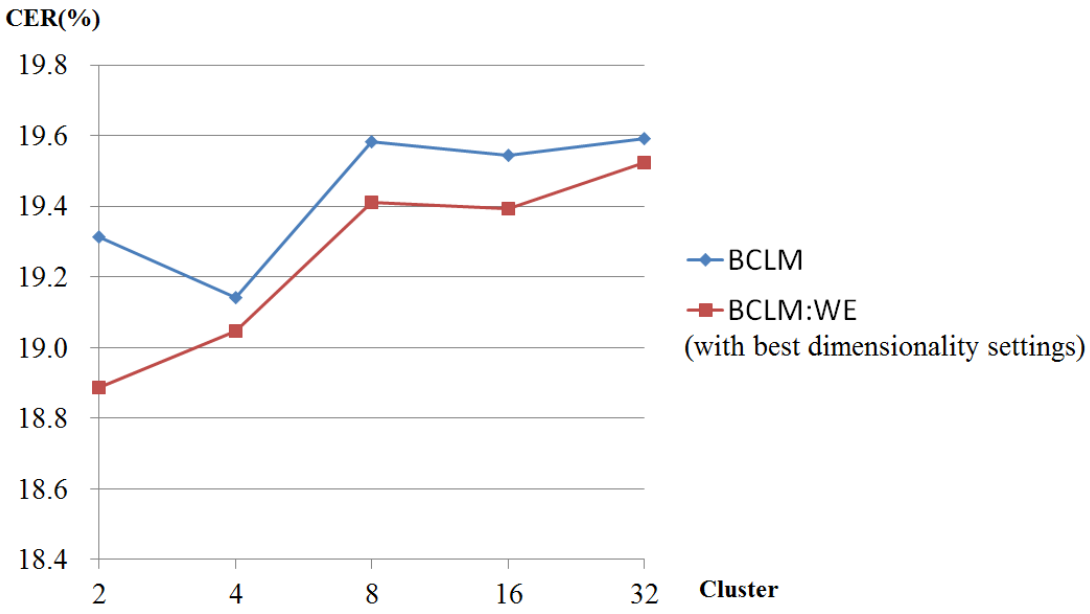
(四) 結合詞向量表示資訊於概念語言模型之實驗結果

本論文嘗試將詞向量資訊應用於概念語言模型之中，在本實驗中，我們將調適語料以句子為單位，利用 *K*-means 分群法將調適語料中的語句分為多個概念類別。

另外在計算測試語句與概念群聚相似度部分，我們使用詞向量表示並透過餘弦方式計算其相似度。本實驗比較傳統概念語言模型(BCLM)與結合詞向量表示於概念語言模型(簡稱為 BCLM:WE)皆作用於不同群聚數目之字錯誤率結果;上述兩種方法皆與背景三連語言模型做線性結合。本實驗採用跳躍式模型(Skip-gram)作為詞向量訓練，相較於連續型模型(CBOW)有較佳實驗結果。其中BCLM:WE(10)表示使用跳躍式模型訓練維度為 10 之詞向量，結合概念語言模型的實驗結果。實驗結果如表五所示，圖五以折線圖方式呈現其實驗結果。

表五、結合詞向量資訊於概念模型之不同群聚數的字錯誤率(%)比較表

群聚個數	2	4	8	16	32
BCLM	19.31	19.14	19.58	19.54	19.59
BCLM:WE(10)	18.89	19.05	**19.40**	19.39	**19.52**
BCLM:WE(20)	18.90	19.05	19.40	**19.39**	19.52
BCLM:WE(30)	18.89	**19.04**	19.40	19.39	19.52
BCLM:WE(40)	**18.88**	19.05	19.40	19.39	19.52
BCLM:WE(50)	18.88	19.04	19.40	19.39	19.52



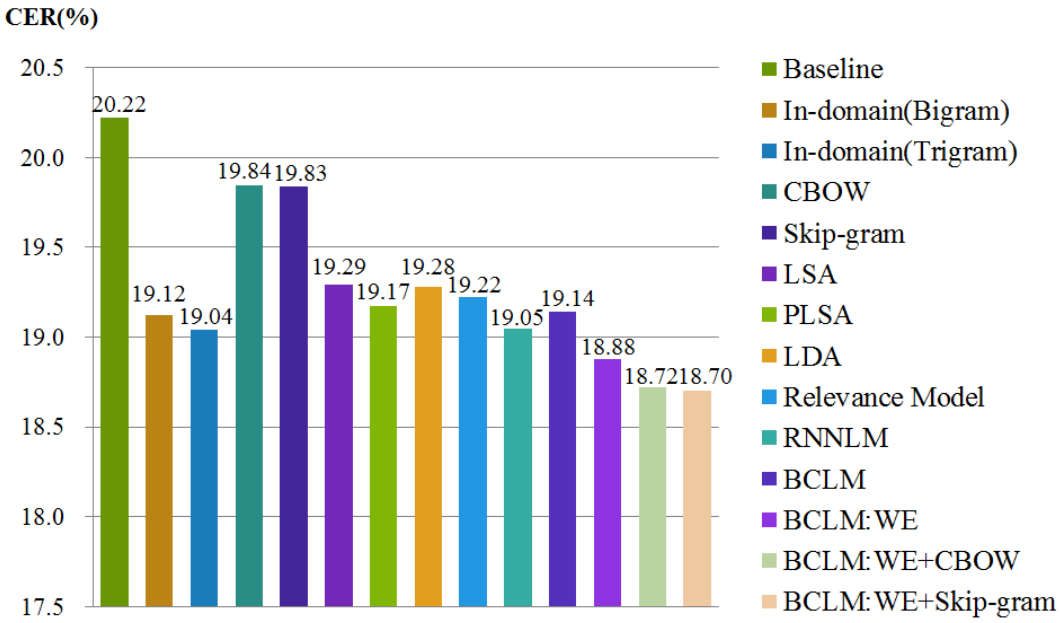
圖五、結合詞向量資訊於概念模型之不同群聚數的字錯誤率(%)比較圖

由圖五我們可以看出結合詞向量表示於概念語言模型(BCLM:WE)中之字錯誤率相較於傳統概念語言模型(BCLM)皆有較好的表現，當群聚數目為 2 時，使用跳躍式模型(Skip-gram)訓練得到的詞向量表示於概念語言模型(BCLM:WE(40))當維度為 40 時，字錯誤率可降低至 18.88。另外，亦可由圖五中看出當群聚數目增加時有利於模型的描述，但是由於分群數過多會導致每群資料

量過少而無法描繪出其概念的特性，因此群聚的數目亦是會影響辨識結果的重要關鍵。

(五) 各式語言模型之實驗結果比較

圖六為各式語言模型與背景三連語言模型(TBG)結合後之字錯誤率結果比較，其中 Baseline 為詞圖搜尋(Word Graph Rescoring)僅使用背景三連模型結果，其字錯誤率為 20.22;而 CBOW 與 Skip-gram 為本論文所提出將詞向量表示應用於詞圖搜尋之實驗結果，相較於沒有使用詞向量表示於詞圖搜尋結果有 0.39 絕對字錯誤率下降。接著，我們比較潛藏語意分析 (Latent Semantic Analysis, LSA)[18]、機率式潛藏語意分析(Probabilistic Latent Semantic Analysis, PLSA)[7]、狄利克里分配(Latent Dirichlet Allocation, LDA)[8]、關聯模型(Relevance Model, RM)[19, 20]、遞迴式類神經網路語言模型(Recurrent Neural Network, RNN)[21]、概念模型(Bigram Concept Language Model, BCLM)[12]以及本論文提出結合詞向量表示於概念語言模型(BCLM:WE)之實驗結果。最後，BCLM:WE+CBOW 與 BCLM:WE+Skip-gram 為本論文所提出的兩種方法結合(亦即第二節以及第三節所提出語言模型調適方法之結合)，實驗果顯示，兩者結合過後效果為最好，字錯誤率可下降至 18.70。由圖六結果觀察得知，本論文提出將詞向量表示應用於語言模型中，對語音辨識的提升確實有幫助。



圖六、各式語言模型之字錯誤率(%)結果比較圖

五、 結論與未來展望

近年來深度學習(Deep Learning)激起一股研究熱潮；隨著深度學習的發展而有分散式表示法(Distributed Representation)的產生。此種表示方式，不僅能以較低維度的向量表示詞彙，還能藉由向量間的運算，找出任兩詞彙之間的語意關係。本論文以此為發想，提出將分散式表示法應用於語音辨識的語言模型中使用。主要

貢獻可以分為兩個部分: 第一部分，本論文將詞向量表示資訊應用於詞圖搜尋之中，在語音辨識的過程中，對於動態產生之歷史詞序列與候選詞改以詞向量表示的方式來建立其對應的語言模型，透過此種表示方式而能獲取到更多詞彙間的語意資訊，以提升辨識的準確度。第二部分，我們針對新近被提出的概念語言模型(Concept Language Model)加以改進，在調適語料中以句子的層次做模型訓練資料選取之依據，去掉多餘且不相關的資訊，使得經由調適語料中訓練出的概念類別更為具代表性，而能幫助動態語言模型調適。另一方面，在語音辨識過程中，會選擇相關的概念類別來動態組成概念語言模型，而此是透過詞向量表示的方式來估算，藉由詞向量表示記錄每一個概念類別內詞彙彼此間的語意關係。最後，我們嘗試將上述兩種語言模型調適技術做結合。根據實驗結果顯示，本論文提出將詞向量表示(Word Representation)應用於語言模型中，對於語音辨識的準確率提升確實有幫助。

未來，我們希望將詞向量表示的資訊應用於其他的語言模型之中，例如應用於關聯模型、詞概念語言模型等。此外，我們希望依據詞圖搜尋的結果結合其他語言模型後，在第二階段的 *N* 條最佳結果(*N*-Best)重新排名時，使用長短期記憶類神經網路模型、遞迴式類神經網路等語言模型重新排序，希望藉由此方法達到辨識效能的提升。

參考文獻

[1] R. Rosenfeld, “Two decades of statistical language modeling: Where do we go from here,” *Proceedings of IEEE*, vol. 88, no. 8, 2000, pp. 1270–1278, 2000.

[2] J. R. Bellegarda, “Statistical language model adaptation: review and perspectives,” *Speech Communication*, vol. 42, no. 11, pp. 93–108, 2004.

[3] S. Furui, L. Deng, M. Gales, H. Ney and K. Tokuda, “Fundamental technologies in modern speech recognition,” *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 16–17, 2012

[4] D. O'Shaughnessy, L. Deng and H. Li, “Speech information processing: Theory and applications,” *Proceedings of the IEEE*, vol. 101, no. 5, pp 1034–1037, 2013.

[5] R. Kuhn, “Speech recognition and the frequency of recently used words: A modified Markov model for natural language,” in *Proceedings of* International *Conference on Computational Linguistics*, pp. 348–350, 1988.

[6] D. Blei and J. Lafferty, “Topic models,” in A. Srivastava and M. Sahami, (eds.), *Text Mining: Theory and Applications*, Taylor and Francis, 2009.

[7] T. Hofmann, “Probabilistic latent semantic indexing,” in *Proceeding of the ACM Special Interest Group on Information Retrieval*, pp. 50–57, 1999.

[8] D. M. Blei, A. Y. Ng and M. I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[9] G.E. Hinton, “Learning distributed representations of concepts,” in *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, pages 1–12, Amherst 1986, 1986. Lawrence Erlbaum, Hillsdale.

[10] T. Mikolov, K. Chen, G. Corrado and J. Dean, “Efficient estimation of word representations in vector space,” in *Proceeding of International Conference on Learning Representations*, 2013.

[11] A. Mnih and K. Kavukcuoglu, “Learning word embeddings efficiently with noise-contrastive estimation,” in *Proceeding of Advances in Neural Information Processing Systems*, pp. 2265–2273, 2013.

[12] 郝柏翰，“運用鄰近與概念資訊於語言模型調適之研究，” 國立臺灣師範大學資訊工程所碩士論文，2014。

[13] C. X. Zhai, “Statistical language models for information retrieval: A critical review,” *Foundations and Trends in Information Retrieval*, nol. 2, no. 3, 137–213, 2008.

[14] B. Chen, J.-W. Kuo and W.-H. Tsai, “Lightly supervised and data-driven approaches to Mandarin broadcast news transcription,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing*, 777–780, 2004.

[15] 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. 1, 219–235, 2005.

[16] S.-H. Liu, F.-H. Chu, S.-H. Lin, H.-S. Lee and Chen, “Training data selection for improving discriminative training of acoustic models,” in *Proceedings of IEEE workshop on Automatic Speech Recognition and Understanding*, 284–289, 2007.

[17] Stolcke, A. (2000). *SRI Language Modeling Toolkit*. Available at: http://www.speech.sri.com/projects/srilm/.

[18] J. R. Bellegarda, “A latent semantic analysis framework for large–span language modeling,” in *Proceedings of European Conference on Speech Communication and Technology*, pp.1451–1454, 1997.

[19] K.-Y. Chen and B. Chen, “Relevance language modeling for speech recognition,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 5568–5571, 2011.

[20] B. Chen and K.-Y. Chen, “Leveraging relevance cues for language modeling in speech recognition,” *Information Processing & Management*, Vol. 49, No 4, pp. 807–816, 2013.

[21] T. Mikolov, M. Karafiát, L. Burget, J. Černocký and S. Khudanpur, “Recurrent neural network based language model,” in *Proceedings of the Annual Conference of the International Speech Communication Association*, 1045-1048, 2010.