The 2016 Conference on Computational Linguistics and Speech Processing
ROCLING 2016, pp. 325-341

基於增強式深層類神經網路之語言辨認系統

Reinforcement Training for Deep Neural Networks-based Language Recognition

蕭硯文　Yen-Wen Hsiao
Department of Electronic Engineering, National Taipei University of Technology
eric6300224@gmail.com

劉翊睿　Hung-Jui Liu
Department of Electronic Engineering, National Taipei University of Technology
rayliu0116@gmail.com

廖元甫　Yuan-Fu Liao
Department of Electronic Engineering, National Taipei University of Technology
yfliao@ntut.edu.tw

摘要

本論文之目標要建立一個基於增強式學習之語言辨認系統，並參與 NIST LRE2015 評比。語言辨認常受到其他相似的語系(out of set, OOS)使效能下降。為了能解決目標語言與 OOS 極為相似與常用的訓練準則與實際應用情境偏離的情況，因此本論文提出新的考慮 OOS 的 DNN 架構並使用 reinforcement learning (RL) 來做訓練，系統特色在於先把 OOS 做細分，包括建立一個可同時辨認目標語言與所有 OOS 的 DNN 架構；以及將整個任務分解成兩個輸出相乘的 DNNs，一個負責語言分群，一個負責區分目標與非目標語言。所提出的系統皆以 LRE2015 規定的代價函數(越低越好)進行實驗比較，根據 LRE2015 評分結果，官方給定的 LDA 語言辨識系統，其分數為 39.033，使用傳統 DNN 其分數為 30.136，而使用本論文所提出兩種新 DNN+reinforcement 其分數分別為 20.899

分與 19.384 分，結果可以發現採用本論文所提出的 DNN+reinforcement 能有最佳的辨識表現。

關鍵詞：語言辨認、網路學習、Q-learning、機率線性鑑別分析、The 2015 Language Recognition i-Vector Machine Learning Challenge

一、 簡介

美國國家標準技術研究所(National Institute of Standards and Technology, NIST)以往每兩年都會舉辦語言辨識評估(Language Recognition Evaluation, LRE)競賽[1]，主要的目的就是藉由世界各個專家或學者的力量來解決語言辨識技術層面的問題，因此參與語言辨識評比不但能讓各個研究團隊了解自我實力的落點，同時也能在評比之後吸收大家的優缺點以作為往後技術的研究方向。

但LRE2015跟之前比賽相當不同，主要是，2015比賽只給予音檔的特徵函數i-Vector，來做目標語言的辨認，而且以往LRE都只要求辨識約20類目標語言，但在LRE2015需辨識的語言種類卻多達50種之多，此外語料裡還包含非目標語言(out of set, OOS)，且非目標語言與目標語言兩者語系十分相近0，更糟糕的是LRE2015的評分標準對OOS的辨認錯誤率(false alarm)給予相當大的處罰，因此在處理LRE2015語料上我們將會面臨三大問題分別為：

- 需要辨識更多語料種類。
- 目標語言與OOS極為相似，不易做線性區分。
- 傳統的語言模型訓練準則，會有訓練結果可能與實際應用情境評分標準偏離的情況。

傳統的語言辨識系統通常使用LDA(Linear Discriminant Analysis)[2]、PLDA(Probabilistic Linear Discriminant Analysis)[4]或是深度類神經網路(Deep Neural Network, DNN)，前兩者語言系統主要為線性分類，對於只辨識區分比較明確的目標語

言還可以達到一定的水準，不過面臨到重疊性很嚴重的OOS與目標語言相近的問題時，單單使用線性分類來處理是不夠的。於是有學者提出使用深度類神經網路(Deep Neural Network, DNN)架構的非線性分類的特性來解決有OOS語料的情況，傳統上DNN是用交叉熵法(cross-entropy)[5]，不過在使用交叉熵法來訓練模型實際上與比賽要求的最終目標不一致，因為比賽設定的false alarm和missing的權重並不一樣，但cross-entropy對此兩種錯誤卻是一視同仁，因此使用cross-entropy可能會有偏離預期目標，造成錯誤判斷等情況。

由於目標語系與OOS語系相當接近，本論文提出了兩個新的DNN架構利用其非線性的特性來解決語料重疊嚴重的情況，並把重點放在把目標語言與OOS再細分處理上，因此建立兩種新的DNN架構包括：一為可同時辨認目標語言與所有OOS的DNN架構；以及一為可將整個任務分解成兩個輸出相乘的DNNs，一個負責語言分群，一個負責區分目標與目標語言。

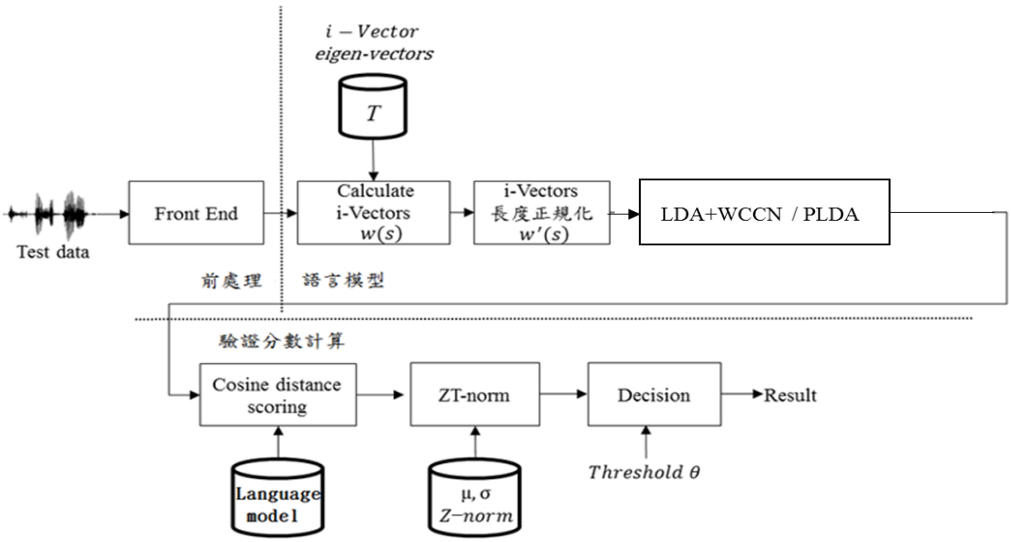
此外為了避免訓練結果與預期目標偏離的現象，本論文使用增強式訓練系統（Reinforcement Learning ,RL）[6]來做訓練，由於RL是直接依據比賽最終目標來不斷更新語言模型的訓練方向，以達到自我學習的效果，因此能夠解決Cross-Entropy對missing與false alarm一視同仁的問題，使其訓練結果貼近LRE 2015要求的評估方式。

二、 相關研究

傳統方法對於語言辨識會使用 LDA 或 PLDA，近幾年來則常使用 DNN，彼此間各有特色，本章節將針對 LRE2015 來探討以上三種辨識系統。

（一） LDA 與 PLDA 語言辨識系統[2]

LDA / PLDA 語言辨認系統架構如圖一所示：



圖一、LDA / PLDA 語言辨認系統架構

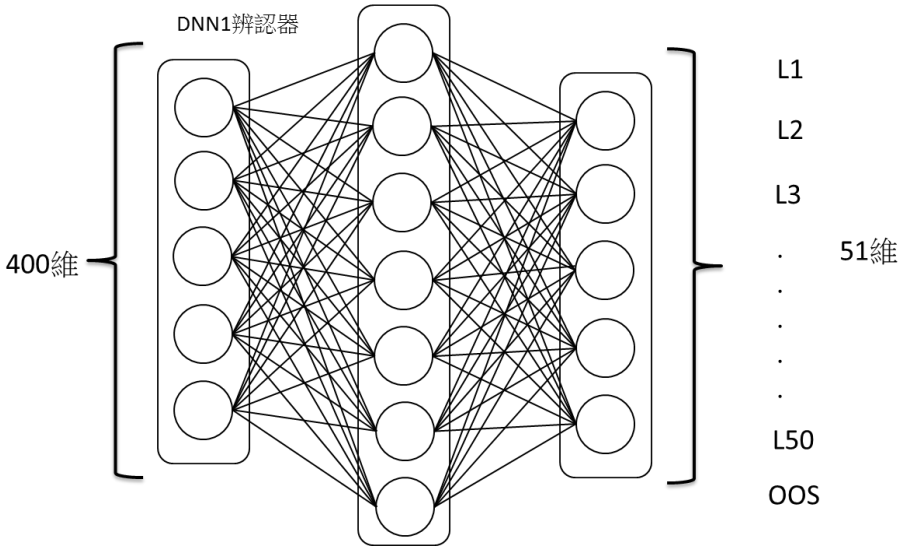
LDA 與 PLDA 系統一開始都會將所有的語料先經過前處理產生強健的聲特徵參數，然後使用訓練語料建構出 i-Vector 的空間基底 T 。之後使用線性識別分析(Linear Discriminant Analysis; LDA)來處理 i-Vector 的投影量 w(s)，LDA 能將資料群由高維度的空間中投影到低維度的空間，因此會先找出一組基底向量來進行線性座標轉換，轉換後的資料群使用類別內散佈矩陣(within-class scatter matrix, WCCN)加以處理，使得同一類的資料群藉由投影後可以愈拉近愈好，不同一類的資料群藉由投影後可以愈分開愈好。相對於 LDA，PLDA 則未使用 WCNN，但依舊可以達成讓同種語言間的變異量變小，不同種語言間的變異量變大。最後驗證分數的部分，LDA 與 PLDA 都將目標語言和測試語言進行 Cosine Distance Scoring 來計算答案，以完成有效的訓練與分類。

不過 LDA 與 PLDA 系統都是使用線性分類來做辨識，對於處理重疊性較小的語言類別上都能有不錯的辨識效果，但面對 LRE2015 裡每一種目標語言與相近的 OOS 語系狀況時就可能無法分割目標與 OOS 語系，可能造成的辨識結果不佳。

（二） 傳統 DNN 語言辨識架構+交叉熵 Cross-Entropy

深度類神經網路(Deep Neural Network, DNN)是一種具備至少一個隱藏層的類神經網路，因此擁有多節點及多層的結構特性，這對於在訓練及分析數據特徵上有很大得益處。DNN 不同於 LDA 或是 PLDA 辨識系統僅止於線性分類，在 DNN 上則是使用像 Rectified Linear Unit (RLU)或是 Sigmoid Function 等非線性激活函數來辨識語言。

由於 LRE2015 的目標是分類目標語言(50 類)與非目標語言，於是傳統 DNN 在輸入端輸入 i-Vector，輸出端則依據 LRE2015 目標，直覺的分為 50 類目標語言以及 1 類 OOS 共 51 類。其架構通常如下圖二所示(以下以 DNN1 代表)：



圖二、傳統 DNN 針對 LRE2015 架構(DNN1)

此外傳統訓練方法常使用交叉熵 (Cross-Entropy)，Cross-Entropy 是一個常見的成本函式。Cross-Entropy 產生於信息論裡面的信息壓縮編碼技術，後來演變成為從博弈論再到機器學習等，並成為其他領域裡的重要技術手段。其定義如下：

$$Cost_{y'}(y) = -\frac{1}{n}\sum_{i}[y_i' ln y_i + (1-y_i') ln(1-y_i)] \qquad (1)$$

其中y是我們預測的概率分佈，y'是實際的分佈。簡單來說交叉熵用來衡量我們的預測，評估整體訓練的好壞。

傳統 DNN 在面對 LRE2015 目標語言與非目標語言相似的問題時，可以用非線性分類來做辨識，相對於線性分類而言這能達到更好的辨識結果；但使用傳統 DNN 架構來處理 LRE2015 的語料還是不夠成熟，主要在於傳統 DNN 在分類上是把全部 OOS 歸為同一類來與 50 類目標語言來做辨識，但 LRE2015 中的 OOS 並不是都同語系的，若只把 OOS 當成一類其實是把很多不同東西放在一起，就會讓 DNN 很難學習。

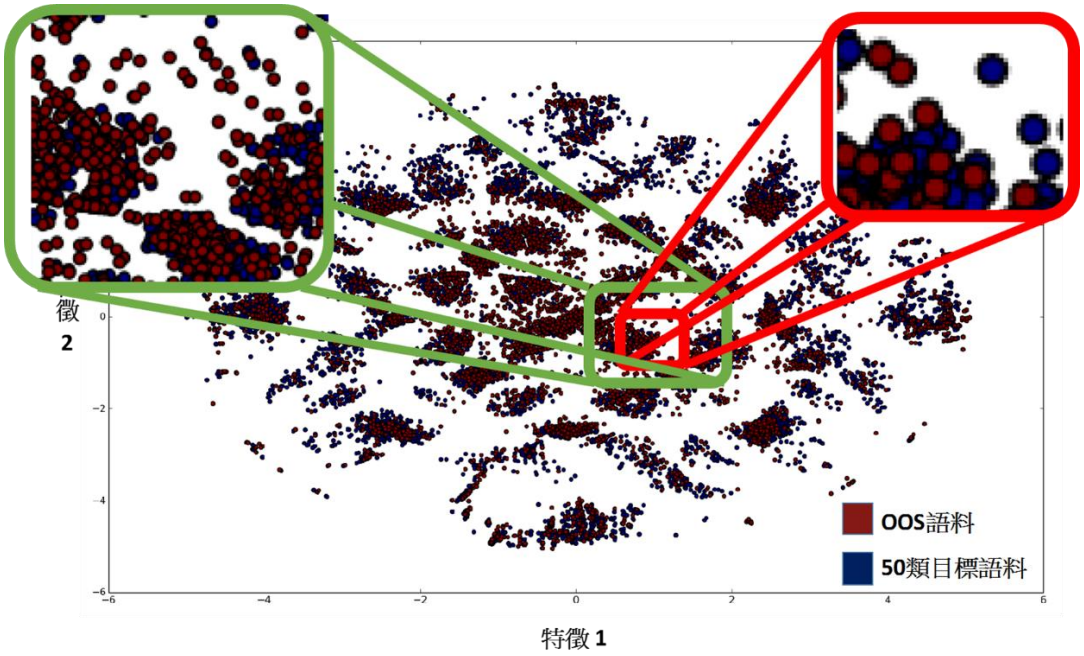
此外在訓練上使用 Cross-Entropy 可能會產生訓練結果與 LRE2015 預期目標偏離的現象，原因主要在於 Cross-Entropy 在面對 OOS 與 50 類目標語言時，兩者重視的比例程度是一致的，這可能導致最後辨識結果偏離比賽要求的目標。

三、 基於增強式學習的語言辨識系統

為了解決 LRE2015 目標語言語與 OOS 相似，及訓練模型結果與實際目標可能偏差等問題，本篇論文採取 DNN 的語言辨識系統，並提出新的 DNN 語言辨識架構，來促使有效分類 50 類目標語言及 OOS，此外我們也搭配使用增強式學習(Reinforcement Learning, RL)來訓練語言模型，藉以提升辨識的效果。

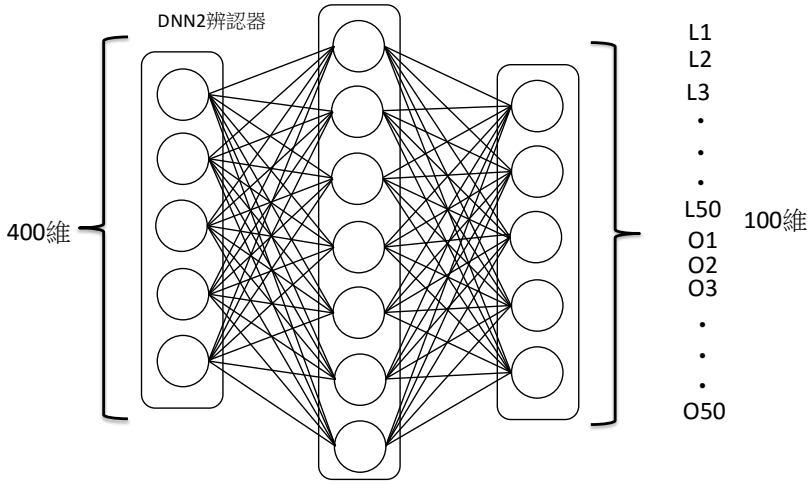
（一） 新的 DNN 語言辨識架構

為了了解 OOS 與目標語言實際嚴重重疊的情況，在前置實驗中我們實作(T-Distributed Stochastic Neighbor Embedding, T-SNE)[8]，把 400 維 i-Vector 共 21431 筆語料(15000 筆 50 類目標語言+6431 筆 OOS)與 50 類目標語言+OOS 的標記丟入 T-SNE 系統來做處理，輸出 2 維特徵矩陣共 21431 筆語料，其結果為 2 維特徵語料分布圖，如下圖三所示，

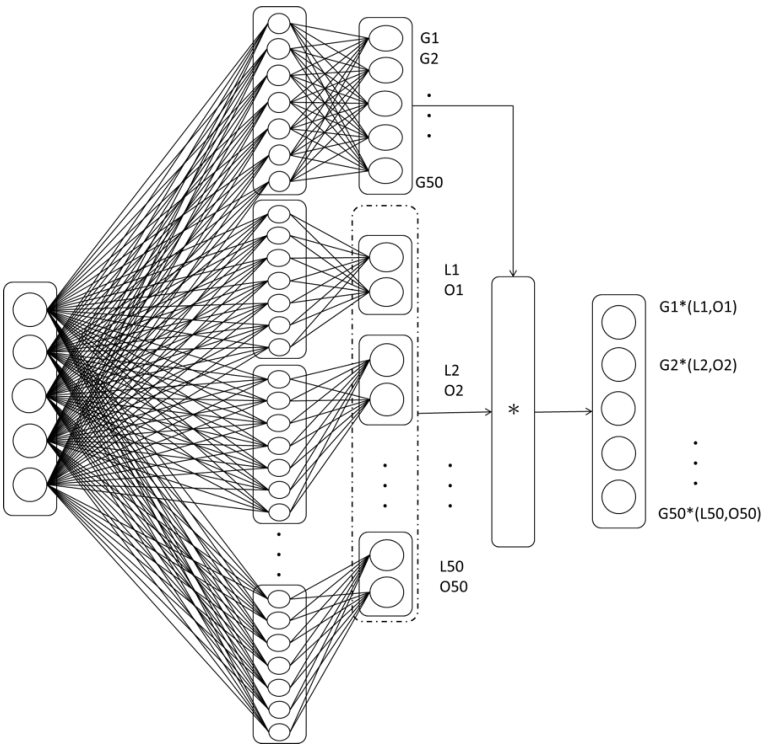


圖三、T-SNE 二維特徵圖

從 T-SNE 二維特徵圖可以看到非目標語言與 50 類目標語言幾乎重疊在一塊(由上圖的局部放大圖可知)，但有趣的是從整體來看其中 OOS 的分布也有 50 類的現象，因此我們認為要解決 OOS 與目標語言重疊的問題，必須要先將目標與其相似非目標分成多個相似語言群，再進一步對每一群相似語言做細分成目標與非目標語言，因此，我們提出兩種新的 DNN 架構，以下我們會以分別以 DNN2 與 DNN3 來代表之。新的 DNN 主要改變在於我們將 OOS 仔細區分為 50 類來做辨識，於是新的 DNN2 與 DNN3 需要辨識 100 種語言類別，其架構如下圖四與圖五所示，其中 DNN2 在 OOS 的細分上是直接分為 50 類然後與 50 類目標語言一起來做辨識；DNN3 將整個任務分解成兩個輸出相乘的 DNNs，一個負責語言分群，一個負責區分目標與非目標語言，兩者分工。



圖四、基於增強式的語言辨識系統之 DNN2 架構圖

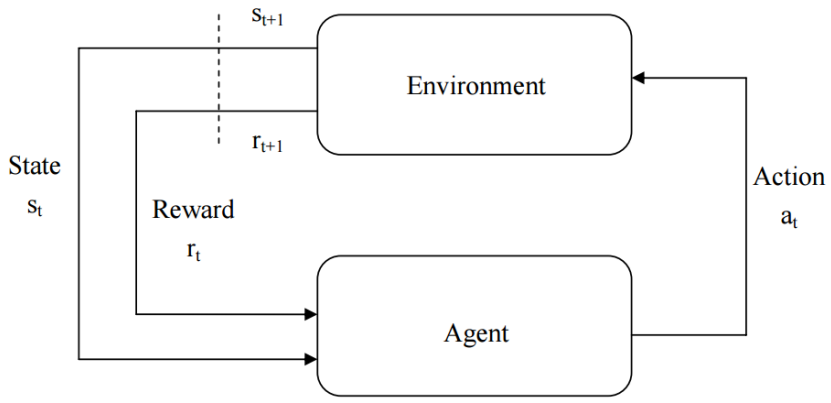


圖五、基於增強式的語言辨識系統之 DNN3 架構圖

（二） 增強式學習(Reinforcement Learning, RL)訓練方法。

增強式學習是從系統與環境的互動中，不斷地嘗試不同的行動，來找尋最佳的策略的一種學習方式。在 RL 裡會有一個學習代理人(agent)會根據現在所處的 state 採取對應的 action 。在一開始沒有任何辨別認知的基礎下，agent 可任意選擇一項動作，environment 接收到此動作後，會根據此 action 回饋給代理人一個 reward，讓 agent 得知

執行此 action 是好還是壞。當得到 reward 的同時，environment 也會提供下一個 state 給 agent，之後不斷重複下去，直到代理人學會如何對每 state 採取正缺的 action，其架構如下圖六所示。



圖六、增強式學習法關係圖

在 Q-learning 裡，我們定義一個函數$Q(s,a)$代表針對系統能得到最大利益，其定義如下公式(2)。

$$Q(s_t, a_t) = \max R_{t+1} \tag{2}$$

$Q(s,a)$就好比在s下，執行什麼a，最後能得到回應最佳的反應，並稱為 Q-function，其值則為 Q-value。

那如何來更新 Q-value 呢? 我們可以使用 Bellman 方程式迭代更新 Q-value。其表示如公式(3)。

$$Q(s,a) = r + \gamma max_{a'} Q(s', a') \tag{3}$$

由 Bellman 方程式(3)延伸出以下更新 Q-value 的公式(4)。

$$Q(s,a) = Q(s,a) + \alpha[r + \gamma max_{a'} Q(s', a') - Q(s,a)] \tag{4}$$

α 為學習率(learning rate)，做為預判未更新前的 Q-value 和新提出的 Q-value 之間可能的差異。特別是，當 α= 1，表示完全學習新的 Q-value，不考慮舊 Q-value。使用$max_{a'} Q(s', a')$來更新 Q-value 在早期學習階段可能是完全錯誤的，但經過多次迭代後就能獲得良好的

結果[7]，因此如果我們執行此更新次數足夠的話，Q 函數將漸漸收斂並得到正確的 Q-value。

以下我們提出一個使用 Q-Learning 來訓練語言模型的演算法，如圖七的 pseudo code 所示：

Initialize $Q(s,a)$ arbitrarily
Observe initial state s
Repeat
Select and carry out action a
Observe reward r and new state s'
$Q(s,a) = Q(s,a) + \alpha[r + \gamma max_{a'} Q(s',a') - Q(s,a)]$
$s = s'$
Until terminal

圖七、Q-Learning 訓練語言模型之 pseudo code

我們定義 Environment 表示 LRE2015 的訓練語料與其相對應的語言種類標準答案；Agent 表示語言辨識系統，會對訓練語料中取樣出來的測試語句(state)做出反應(action)，即猜測測試語句的語言種類，並其會依據 Environment 給予的 reward 調整辨識系統；其中 Reward 表示 Environment 將猜測答案與該語料的實際語言標準答案做比較後所得到的加分與扣分結果，若結果正確則保持原狀，錯誤則修正系統。

Q-Learning 的流程如下，第一次訓練會先隨機初始化語言模型，之後進行迭代，首先對給予的語料猜測其最有可能語言，然後觀察猜測語言之 reward 結果，再根據以上參數更新語言模型，接著換下一個語料，直到獲得最佳語言辨識效果。

四、 實驗結果

本章節我們將實作官方給定 LDA(baseline)、傳統 DNN 以及本論文所提出的兩種 DNN 語言辨識系統，在 DNN 方面我們會分別使用 Cross-Entropy、加權式代價函數法(Cost-function)及 Reinforcement Learning 來訓練語言模型，並使用官方給定的評分標準評估以上辨識系統。

（一） 實驗資料

NIST LRE2015 提供 400 維 i-Vector，當中包含訓練語料 15000 筆、6500 筆未做標示的測試語料以及 6431 筆非目標語言(OOS)之語料，如表一所示，其所包含語言如下表二所示。

表一、NIST2015 全部語料數量

語料種類	Data(400 維)數量
ivec15_lre_train_ivectors	15000
ivec15_lre_test_ivectors	6500
ivec15_lre_dev_ivectors	6431

表二、NIST LRE2015 50 類分類語言

Target Languages(train_ivectors)				
Amharic	Dari	Kazakh	polish	Tagalog
Arabic	English	Khmer	Portuguese	Tajik
Armenian	Farsi	Korean	Punjabi	Tatar
Azerbaijani	French	Kosovo	Romanian	Thai
Bengali	Georgian	Kurdish	Russian	Tibetan
Bosnian	Greek	Kyrgyz	Shona	Turkish
Burmese	Hausa	Laotian	Slovak	Ukrainian
Cantonese	Hindi	Mandarin	Somali	Urdu
Creole	Indonesian	Oromo	Spanish	Uzbek
Czech	Japanese	Pashto	Swahili	Zulu
Out of Target Languages(dev_ivectors)				
out_of_set				

（二） 實驗評分方法

主辦方所提出的評分標準如下：

$$Cost = \frac{(1 - P_{oos})}{n} * \sum_{k}^{n} P_{error}(k) + P_{oos} * P_{oos}(oos) \qquad (5)$$

$$P_{error} = \left(\frac{\#errors\_class\_k}{\#trials\_class\_k}\right), n = 50, and P_{oos} = 0.23 \qquad (6)$$

其中 $P_{oos}(oos)$ 表示為 OOS 但被歸類在 50 類的錯誤率；P_{oos}＝0.23 表示比例權重參數，因此 LRE2015 對 OOS 錯誤率的處罰相當嚴厲。

（三） 實驗設定

先利用 LRE2015 官方所給定的 LDA 語言辨識系統，算出基礎分數，之後再與利用深層網路學習和強化式語言辨認系統訓練出來的分數做分析與比較，以下是各系統的設定。

1. 基礎語言辨識系統 LDA

LDA 的實驗設定方面：

輸入部分：400 維 i-Vector 共 21431 筆語料(15000 筆 50 類目標語言+6431 筆 OOS)與 50 類目標語言+OOS 的 Label。

輸出部分：50 類目標語言+1 類 OOS = 51 維。

2. DNN 語言辨識系統

DNN 的實驗設定方面，會設計 3 種 DNN 架構包含 DNN1 表示傳統 DNN(圖二)、DNN2 表示圖四新提出 DNN 架構、DNN3 表示圖五新提出的 DNN 架構，並分別採用 Cost-function、Cross-Entropy 以及 Reinforcement Learning 來做訓練。

（1） DNN1：目標語言 50 類 + OOS 1 類

在 DNN1 方面有 1 層隱藏層，輸入 400 維 i-Vector，輸出則為 51 維(50 類目標語言+1 類 OOS)。

（2） DNN2：目標語言 50 類 + OOS 50 類

在 DNN2 方面有 1 層隱藏層；輸入 400 維 i-Vector，輸出則為 100 維(50 類目標語言+50 類 OOS)。

（3） DNN3：目標語言 50 類 + OOS 50 類

在 DNN3 方面有一層語言分群 DNN，一層 80 個目標與非目標分群 DNN 再搭配 1 個乘法 Gate；輸入 400 維 i-Vector，輸出則為 100 維(50 類目標語言+50 類 OOS)。

（四） 實驗結果

我們實驗所使用的語料庫以及 NIST LRE2015 評比的實驗環境設定，計算評比項目中的結果，且整理所有結果如表三，並針對評比結果做出實驗分析。

表三分為兩大部分，分別將訓練語料(Train)及測試語料(Test)丟入經過訓練的語言模型之後記錄其結果。訓練語言模型的方式有 7 種，分別為：官方給定的 LDA 語言辨識系統、DNN1+Cross-Entropy、DNN1+加權式代價函數法(Cost)、DNN1+增強式學習(reinforcement)、DNN2+Cross-Entropy、DNN2+加權式代價函數法(Cost)、DNN2+增強式學習(reinforcement)、 DNN3+Cross-Entropy、DNN3+加權式代價函數法(Cost)、DNN3+增強式學習(reinforcement)。

計分部分分為 Correct (%)與 Scores，Correct(%)表示成功辨識的語料占所有語料多少百分比，數值越高越好；Scores 表示將辨識正確的語料數量與錯誤的語料數量丟入 LRE2015 官方所給定的計分標準所算出的結果，數值越低其辨識系統表現越好。其中 LRE2015 以 Scores 為準，Correct(%)只是我們用來作為參考。

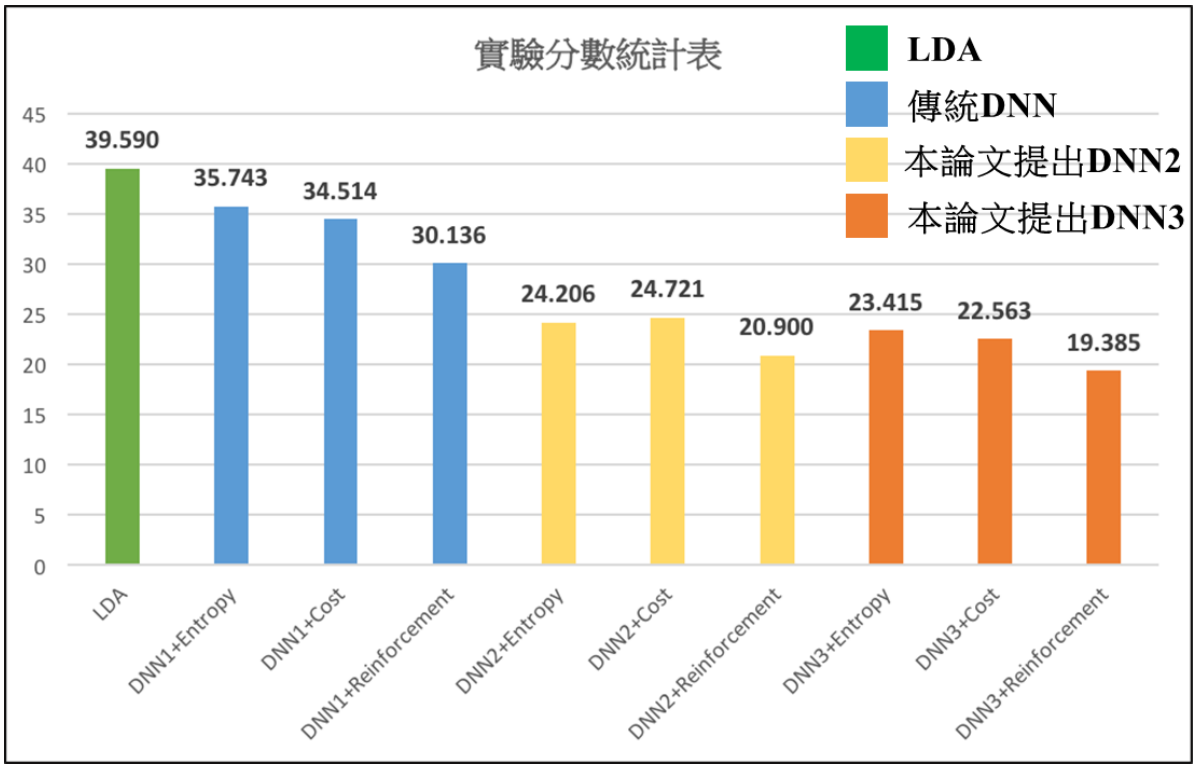
表三、語言辨識系統評估分數統計表

		Train		Test	
		Correct (%)	Scores	Correct (%)	Scores
baseline	LDA			60.3544	39.590
DNN1	entropy	85.5076	21.8589	64.2769	35.7433
	cost	97.7935	23.8572	65.4308	34.5135
	reinforcement	97.1935	24.0021	69.8308	30.1356
DNN2	entropy	97.9335	24.5862	69.6462	24.2055
	cost	97.7202	24.7397	72.7692	24.721
	reinforcement	97.7135	24.6911	73.8462	20.8996
DNN3	entropy	99.8497	23.4157	71.5837	23.4153
	cost	99.7592	23.5721	72.6493	22.5627
	reinforcement	99.9453	23.6834	74.5771	19.3847

從表三來看辨識測試語料的部分，採用主辦方給定的 LDA 訓練系統其結果為 39.590 分，使用 DNN 架構方面分數都高於 LDA 系統分數，其中 DNN1 架構無細分 OOS 但採用 reinforcement 能得到最佳 30.1356 分，而在 DNN2 架構有細分 OOS 並採用 reinforcement 能得到最佳 20.8996 分，DNN3 架構細分 OOS，但採取兩個 DNN 分工並採用 reinforcement 能得到最佳 19.3847 分。由表三數據結果可以看出使用 DNN3 的分類方式明顯優於其他兩種架構，整體而言我們更可以看出使用 reinforcement 來訓練模型的最佳表現也遠高於使用 Cross-Entropy 和加權式代價函數法，且使用 DNN3 的架構來處理 OOS 又比 DNN2 好，所以最好的結果是使用 DNN3+reinforcement 語言辨識系統。

五、 結論

在本論文中我們所提出的新 DNN 架構搭配 reinforcement 的語言辨識方法，並參加 LRE 2015 評比，經比較 LDA，傳統 DNN，與我們提出的兩種新的 DNN 架構的效能，得到如下圖八的結果。



圖八、辨識系統評估分數統計表

其中由圖八數據總結得知，不管是 DNN1，DNN2 或者 DNN3，在測試語料的部分，DNN 使用 reinforcement 都比使用 Cross-Entropy 或是 Cost function 來的更好；此外考慮 OOS 的 DNN 架構也優於傳統 DNN，能有效的解決 LRE2015 目標語言與 OOS 語料相似的問題。因此我們提出的方法 DNN3+ Reinforcement Learning 的確對於 LRE 2015 評比能達到有效語言辨認結果。從表四 NIST LRE2015 官方評分結果排行榜，更可以發現我們所提出 DNN3+ Reinforcement Learning 辨識分數 19.385 分與前十名相比還算不錯。

表四、NIST LRE 2015 官方評分結果排行(前 10 名)

Rank	Name	Affiliation	Score on eval set
1	Hanwu Sun	Institute for Infocomm Research, Singapore	17.736
2	Konstantin Simonchik	individual	18.022
3	Kong Aik Lee	Institute for Infocomm Research, A*STAR, Singapore	17.802
4	Sergey Novoselov	individual	18.066
5	Haizhou Li	Institute for Infocomm Research, A*STAR, Singapore	17.758
6	Nguyen Trung Hieu	Institute for Infocomm Research	18.462
7	UTD-CRSS Team	University of Texas at Dallas (CRSS)	22.154
8	Qian Zhang	Center for Robust Speech Systems (CRSS),UT Dallas	23.011
9	Chengzhu Yu	University of Texas at Dallas (CRSS)	23.077
10	Chunlei Zhang	Center for Robust Speech Systems (CRSS),UT Dallas	24.308

致謝

本研究感謝教育部『大學以社教機構為基地之數位人文計畫』（A36 號）與科技部專題計畫（MOST 104-2221-E-027-079, 105-2221-E-027-119 and 103-2218-E-027-006-MY3）支持。

參考文獻

[1] NIST i-vector Machine Learning Challenge:

https://ivectorchallenge.nist.gov/

[2] Chengzhu Yu, Chunlei Zhang, Shivesh Ranjan, Qian Zhang, Abhinav Misra, Finnian Kelly, John H. L. Hansen," utd-crss system for the NIST 2015 language recognition i-Vector machine learning challenge". Available: http://www.utdallas.edu/~gxl083000/pdfs/2016_ICASSP_LRE_ivML.pdf

[3] Najim Dehak, Pedro A.Torres-Carrasquillo, Douglas Reynolds, Reda Dehak, "Language Recognition via Ivectors and Dimensionality Reduction", 2011. Available: https://groups.csail.mit.edu/sls/publications/2011/Dehak_Interspeech11.pdf

[4] Saad lrtza, "Scalable I-vector concatenation for PLDA based language identification system",IEEE ,2015.Available:

http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7415458&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D7415458

[5] Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, and Brian Kingsbury," Deep Neural Networks for Acoustic Modeling in Speech Recognition". Available: http://static.googleusercontent.com/media/research.google.com/zh-TW//pubs/archive/38131.pdf

[6] Tambet Matiisen, "Demystifying Deep Reinforcement Learning". Available: https://www.nervanasys.com/demystifying-deep-reinforcement-learning/

[7] Francisco S. Melo, "Convergence of Q-learning: a simple proof". Available: http://users.isr.ist.utl.pt/~mtjspaan/readingGroup/ProofQlearning.pdf

[8] Laurens van der Maaten , Geoffrey Hinton " Visualizing Data using t-SNE ".Available: https://lvdmaaten.github.io/publications/papers/JMLR_2008.pdf