The 2018 Conference on Computational Linguistics and Speech Processing
ROCLING 2018, pp. 31-45

使用性別資訊於語者驗證系統之研究與實作

A study and implementation on Speaker Verification System using Gender Information

蘇俞睿　Yu-Jui Su
國立臺灣大學資訊工程學系
Department of Computer Science and Information Engineering
National Taiwan University
shikari.su@mirlab.org

張智星　Jyh-Shing Roger Jang
國立臺灣大學資訊工程學系
Department of Computer Science and Information Engineering
National Taiwan Normal University
jang@csie.ntu.edu.tw

詹博丞 Po-Cheng Chan
中華電信研究院
Chunghwa Telecom Laboratories, Taoyuan, Taiwan
cbc@cht.com.tw

摘要

在語者驗證領域中，在不改變聲學模型架構之前提下，以男性與女性之語料分別訓練的性別相關模型取代性別不相關模型，是常見的提升系統辨識率作法之一。然而，在實際運用情形中，由於測試語者的性別是未知的，因此性別分類器在此流程下便扮演了非常重要的角色，其準確度更會直接影響語者驗證系統的表現；而確保系統面對不同性別之仿冒者皆能正確拒絕，亦是此作法相當重要的一項訴求。為探討不同的「語者性別資訊運用方法」對於語者驗證系統所產生的影響，本論文實作了以 i-向量與機率性線性判別分析模型為語者特徵與評分器之語者驗證系統，與 2 種以 i-向量為基礎的性別分類器。本論文在分析一般使用性別相關模型之語者驗證系統的弱點後，分別於「性別分類器表現良好」與「性別分類器表現不良」之兩大狀況下提出其他不同的性別資訊應用方法，並分析各方法在不同的仿冒者性別組成下之表現，最後亦達成了在各種情況下皆能讓系統表現超越傳統作法之目標。

Abstract

For speaker verification task, one way to improve system's accuracy without changing the algorithm of acoustic model is to use gender-dependent model instead of gender-independent one. However, since test speakers' genders are not available, gender classifier plays an important role since its accuracy directly affects the performance of the speaker verification system overall. Furthermore, ensuring that the system can maintain good performance under different gender composition of test speakers is also an important issue. To explore the impact of different gender information's usage on speaker verification system, this paper implemented a speaker verification system using i-vector and PLDA model as speaker feature and scoring model respectively, and 2 i-vector-based gender classifier. After analyzing the weakness of speaker verification system using gender-dependent model in a general way, we proposed different methods for the application of gender information under the conditions when gender classifier has good and poor performance respectively. Moreover, we analyze the performance of each method under different gender composition of test speakers as well. Finally, we reached the goal of improving our system to achieve better performance than tradition practice under different circumstances.

關鍵詞：語者驗證、性別資訊、性別分類器、i-向量、機率性線性判別分析

Keywords: Speaker Verification, Gender Information, Gender Classifier, i-vector, PLDA

一、緒論

語者辨識，或稱自動語者辨識（automatic speaker recognition），是一種將人類的語音透過電子設備轉換成電子訊號後，利用演算法分析不同語者之間的聲紋特性差異，來進行語者身份辨識的工作。語者辨識的應用領域相當廣泛，除了最常見的個人設備存取控制之外，舉凡門禁系統、信用交易、犯罪偵測、電子會議輔助、刑事案件舉證等皆屬於其應用範疇；且由於其需求不斷擴大，如何增進語者辨識技術的準確度與或提升系統演算法之速度，至今仍吸引學界與業界投入大量研究心力。本論文以語者驗證作為研究主軸，主要方向為探討「語者的性別資訊」對於語者驗證系統所產生的影響，探討的項目包括以下 3 點：第一，本論文將探討前人如何利用性別資訊提升語者驗證系統之辨識率，並分析該作法之弱點；第二，本論文將提出其他性別資訊應用方法之改良，並分析各種方法適用之情況；最後，本輪文將實作性別分類器，並實際將其應用於語者驗證系統中。

二、實驗語料與資料配置

本論文使用了 NIST SRE 2010 與 TIMIT 兩份語料庫分別進行實驗，以下將簡介此二語料庫與資料配置。NIST SRE 2010 [1] 是一場在 2010 年由美國國家標準技術研究所（National Institute of Standards and Technology, NIST）所舉辦的語者辨識（Speaker Recognition Evaluation, SRE）比賽，本論文取其訓練資料中的 3 種資料類型：「10sec」、「core」 與「8conv」，並除去重複的語者後進行實驗，其中「10sec」為長度約 10 秒的電話對話錄音，一共有 2,257 位（男 959／女 1,298）語者，每位語者提供 1 份錄音檔；「core」為長度約 3 至 15 分鐘的電話對話錄音或麥克風錄製的面試錄音，一共有 4,004 位（男 1,817／女 2,187）語者，每位語者提供 1 份錄音檔；「8conv」為長度約 2 至 3 分鐘的電話對話錄音，一共有 445 位（男 194／女 251）語者，每位語者提供 8 份錄音檔。3 種資料類型之音檔取樣頻率為 8,000 赫茲、樣本位元數皆為 16。資料配置方面，訓練資料是以 3 種資料類型各取 2／3 語者數量，一共 4,469 位（男 1,965／女 2,504）語者的集合；測試資料則是剩下的 1／3 語者數量，一共 2,237 位（男 1,005／女 1,232）語者的集合。而關於測試資料中註冊語料與驗證語料的分配，每位語者皆僅會註冊 1 組語者模型，其中「8conv」的每組模型皆以 7 份音檔註冊而成，剩下 1 份音檔作為驗證音檔；而由於「10sec」與「core」的每位語者皆只有 1 份音檔，因此這裡將每份音檔分割為 4 等份，每組模型皆以 3 份音檔註冊而成，剩下 1 份音檔作為驗證音檔。測試資料的每一位語者皆會自驗證語料中挑選 1,000 份音檔來測試自己的語者模型，1,000 份音檔中其中 1 份是會是正樣本（屬於此語者的聲音），而其他 999 份則是隨機挑選的負樣本（不屬於此語者的聲音），因此所有欲評測的樣本一共有 $2{,}237 \times 1{,}000 = 2{,}237{,}000$ 份。

TIMIT [2] 是由德州儀器（Texas Instruments, TI）、麻省理工學院（Massachusetts Institute of Technology, MIT）與斯坦福國際研究院（SRI International）聯合提供的語料庫，其一共包含了 630 位（男 438／女 192）語者。這些語者的腔調包含了美國 8 種主要地方方言腔調，每位語者皆提供 10 份錄音檔，每份錄音檔之取樣頻率為 16,000 赫茲、樣本位元數為 16，錄音內容則為指定內容、包含豐富音素（phoneme）的一句話，每句話平均長度約為 3 秒，最長不超過 8 秒。資料配置方面，訓練資料是以男性與女性各取 2／3 語者數量，一共 420 位（男 292／女 128）語者的集合；測試資料則是剩下的 1

／3 語者數量，一共 210 位（男 146／女 64）語者的集合。而關於測試資料中註冊語料與驗證語料的分配，每位語者皆會註冊 2 組語者模型，每組模型皆以 4 份音檔註冊而成，而剩下的 2 份音檔則作為測試音檔。測試資料的每一位語者皆會自驗證語料中挑選 100 份音檔來測試自己的語者模型，100 份音檔中其中 1 份是會是正樣本，而其他 99 份則是隨機挑選的負樣本，因此所有欲評測的樣本一共有 $210 \times 2 \times 100 = 42{,}000$ 份。

三、實驗一：對照組

實驗一將介紹本論文設定的對照組流程與實驗結果，此實驗為完全不使用語者性別資訊的傳統語者驗證方法，亦為本論文之後所有實驗之比較依據。本實驗之流程架構如圖 1 所示：

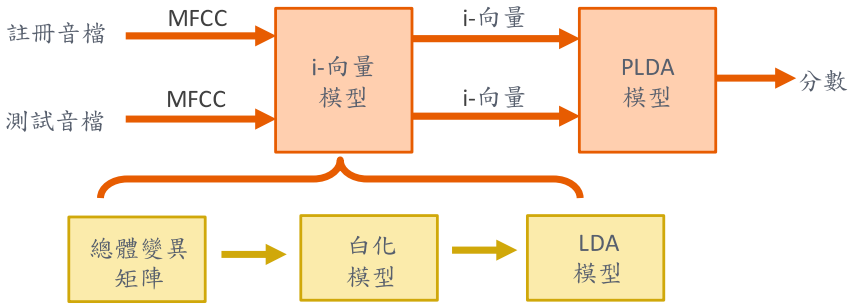


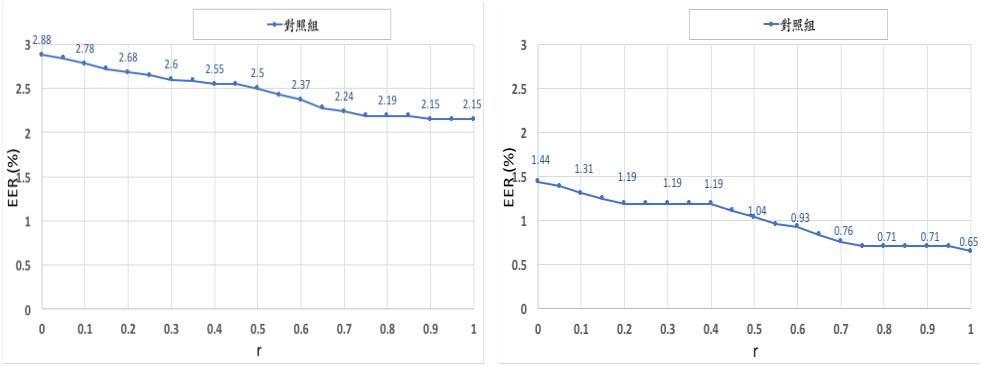
圖 1、對照組之流程架構圖

流程圖中之各項模型以 Bob toolbox [3]進行實作。由圖中可見，註冊音檔與測試音檔分別抽取 MFCC 特徵後，會經由 i-向量模型各自抽取 i-向量 [4] [5] [6]，最後經由 PLDA 模型算出分數 [7] [8] [9]；其中，此處所述之 i-向量模型裡除了包含原本用以將超向量轉化為 i-向量的總體變異矩陣之外，亦包涵了白化轉換模型 [10] 與線性判別分析模型 [11]。表 1 為對照組流程所使用的模型參數，而本論文其他語者驗證相關流程之模型參數設定亦與對照組相同：

表 1、語者驗證系統模型參數設定

Model Parameter		Value
MFCC extraction	Frame size	512 sample points
	Hop size	128 sample points
UBM model	Number of mixture	128
	Number of training iteration	10
i-vector model	Dimension of i-vector	300
	Number of training iteration	5
LDA model	Reduced Dimension of i-vector	50
PLDA model	Dimension of speaker factor	50
	Dimension of channel factor	50
	Number of training iteration	10

對照組流程的實驗結果如圖 2 所示：



NIST SRE 2010　　TIMIT

圖 2、對照組之流程架構圖

圖之左側與右側分別為 NIST SRE 2010 語料庫與 TIMIT 語料庫之實驗結果，兩張圖的縱軸為語者驗證系統之等錯誤率，單位為百分比；橫軸為 r，此 r 值代表的是「與註冊語者不同性別的仿冒者比例」，以下將簡單說明此實驗設計之意義。在前一章節中提及，本論文之資料配置形式為每一位註冊語者配對 1 份自己的音檔作為正樣本，並搭配許多其他仿冒者的聲音作為負樣本，而挑選仿冒者的一項重要考量，便是這些語者的性別。今假設語者驗證系統架設於使用者之行動裝置上，則可以考慮兩種情景，如圖 3 所示：

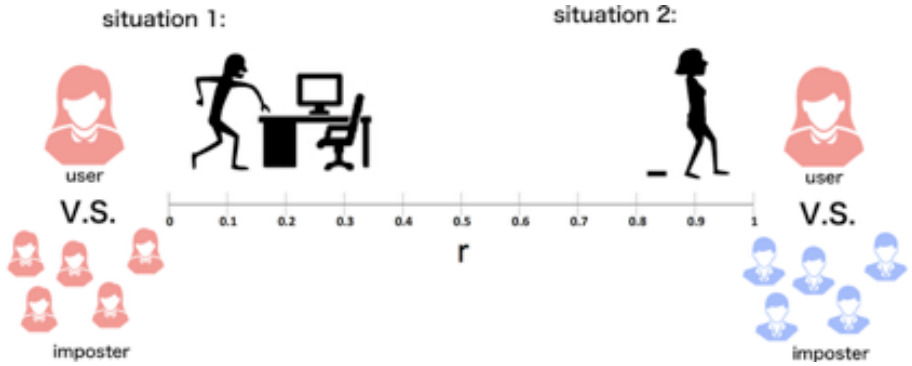


圖 3、不同的 r 值與系統應用情境

第一，當使用者到了教室或是工作場所時，若有同事或同學意圖在未經該使用者同意下存取其行動裝置，由於他們皆熟悉該使用者的聲音，因此勢必會找聲音與該使用者相像的人—至少會與該使用者同性別—來解鎖其語者驗證系統，在此情境下，我們可以假設絕大部分的仿冒者皆會和使用者同性別（事實上，NIST SRE 2010 比賽亦是設計成每個樣本的註冊方與測試方的性別皆相同，因此與此情境相同）。第二，當使用者的行動裝置（被陌生人）偷竊或不慎遺失時，由於拿到其行動裝置的人並不知曉該使用者的性別或聲音特性，因此此種情境下便不能假設仿冒者會和該使用者之性別相同，仿冒者和該使用者的性別相同的機率在此時只有 1／2，甚至更低。綜合以上兩種情境，我們便必須

確保語者驗證系統在不同的 r 值，也就是不同的「仿冒者與使用者不同性別的機率」的情況下，皆應有良好的表現。因此，本論文所有的語者驗證相關實驗，皆會在不同的 r 值之下測量系統的表現。由圖 3 中可見，雖然對照組流程於 TIMIT 語料與 NIST SRE 2010 語料上的表現有明顯差距，但兩者的變化趨勢是相同的：從最左側的 r 等於 0 慢慢往最右側的 r 等於 1 移動時，系統的等錯誤率皆呈現近似線性下降的趨勢。造成此下降趨勢之原因其實非常直觀：與使用者不同性別的仿冒者，相較於同性別者，大多會與該使用者在聲紋特性上有較大的差別，造成其被系統拒絕的機率也相對高出許多，因此當 r 值越大，正樣本所得到的分數與其他負樣本相比便會相對高出許多，進而得到較低的等錯誤率。

四、實驗二：性別資訊應用方法 — 當性別分類器表現良好

實驗二將介紹若干種將性別資訊應用於語者驗證系統以提升其辨識率之方法，包括一般最常見作法，以及本論文提出的針對前者弱點所進行之改良方法。另外，本實驗之目的在於證明在性別分類器表現良好之前提下，藉由有效運用性別資訊，便能提升系統辨識率；因此本實驗中對於所有原本應假定為未知性別的測試語者，皆以其真實性別代替性別分類器的預測結果。

前人作法（一般常見作法）

本方法之流程架構如圖 4 所示：

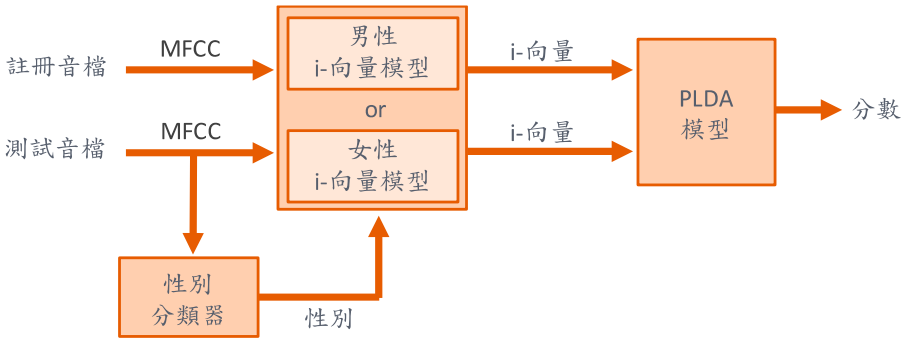
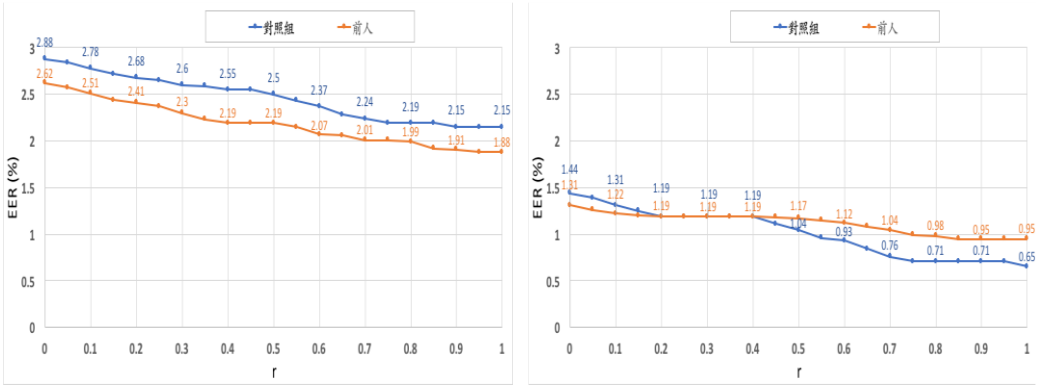


圖 4、前人作法之流程架構圖

其流程基本上與對照組流程類似，兩者不一樣的地方為：對照組之 i-向量模型是使用訓練資料中所有語者共同訓練而得來，而本作法之 i-向量模型則分為 2 個性別相關的模型，分別是使用訓練語者中的男性語者與女性語者進行訓練而得來，而所有註冊音檔或測試音檔，皆必須使用對應其語者性別之模型來抽取 i-向量。其中，註冊語者的性別可以假設是已知的，因為在實際運用情境中，我們可以要求所有使用者要註冊系統時提供自己的性別資訊；但是由於測試語者的性別是未知的，因此測試音檔必須先藉由性別分類器預測其性別，才能套入正確性別的 i-向量模型來抽取 i-向量。本做法的實驗結果如圖 5 所示：



NIST SRE 2010　　TIMIT

圖 5、前人作法之結果

同樣地，圖之左側與右側分別為 NIST SRE 2010 語料庫與 TIMIT 語料庫之實驗結果，兩張圖的藍線為方才對照組流程的結果，橘線則是本做法的結果。由圖中可以看到，在 NIST SRE 2010 語料中，本做法的結果從 r = 0 至 r = 1 始終都較對照組作法要進步，證明了其實用性；但反觀 TIMIT 語料，其實驗結果在最左側 r = 0 時尚較對照組進步一些，但慢慢往右側 r = 1 方向移動時，由於其等錯誤率的下降速率較對照組慢上許多，導致其結果逐漸被對照組超越，甚至被拉開距離。由於許多語者驗證比賽的樣本（如 NIST SRE 2010）並不包含註冊語者與測試語者性別不同的情形，因此許多以本作法實作的語者驗證系統，便沒能對此現象做進一部探討或進行補償措施。而本研究推測，造成上述現象的主因，是來自於本做法的一個缺陷：使用性別相關模型抽取的男性 i-向量與女性 i-向量，在空間分佈上可能存在重疊現象。圖 6 為 i-向量之理想二維點分佈圖：

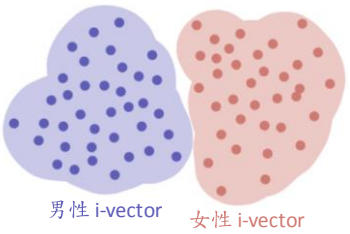


圖 6、i-向量之理想二維點分佈圖

在理想上，因為男性與女性的聲紋特性存在很大的差異，容易讓人以為 i-向量會在空間分佈上呈現男女各自群聚，且容易界定出兩群 i-向量之邊界的情形，正如同圖 6 所示。在此假設下，當一個樣本的註冊方與測試方擁有不同性別，因為兩者的 i-向量之間的距離很大，因此這個樣本便相當容易被系統拒絕。但實際情況中卻並非一定如此，尤其就男／女性 i-向量模型在訓練過程中並沒有接觸過女／男性的資料這個事實來看，我們更必須懷疑此假設之合理性。圖 7 左右兩圖是分別使用主成分分析（Principal Components Analysis, PCA）和 t-分佈隨機相鄰嵌入（T-distributed Stochastic Neighbor Embedding, TSNE），將訓練資料中的 i-向量降至二維平面所繪製的點分佈圖：

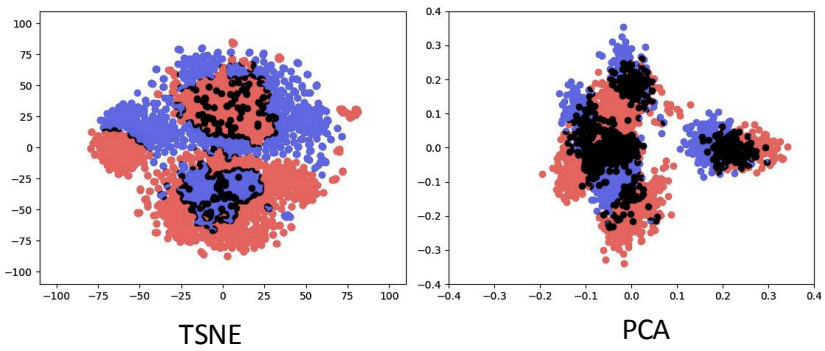


圖 7、i-向量之實際二維點分佈圖—以 TSNE 與 PCA 將 i-向量降為二維

由兩張圖我們能清楚看到，兩個性別的 i-向量群在空間分佈上確實存在重疊的情形。這種現象會造成語者驗證系統的錯誤率上升，尤其在當欲測試的樣本中含有大量註冊方與測試方不同性別的情形，即 r 有較大的值時更為顯著。為方便討論，此處假設系統之評分器僅以註冊方與測試方 i-向量之間的歐氏距離作為判斷是否為同一語者之唯一標準。在不分男、女資料所訓練出來的混合 i-向量模型，在訓練資料足夠的前提下，我們可以假定其會盡量將不同語者的 i-向量在空間上彼此分開；但是，正如方才所述，由於男／女性 i-向量模型在訓練過程中並沒有接觸過女／男性的資料，因此便很有可能產生特定男性語者與女性語者的 i-向量距離過度相近的現象，使得系統容易將他們誤判為同一

語者。此現象在訓練流程中，會使得系統的評分器—機率線性判別分析模型容易被混淆；在測試流程中，也容易會造成不同性別間語者被錯誤判斷成同一語者的比例較對照組流程還要高的現象。因此，針對因不同性別 i-向量空間重疊現象所造成的上述缺陷，本論文接下來將提出一種簡單的改良機制，以令結果獲得改善。

方法一：拒絕與註冊者性別不符之測試者

此改良方法之流程架構如圖 8 所示：

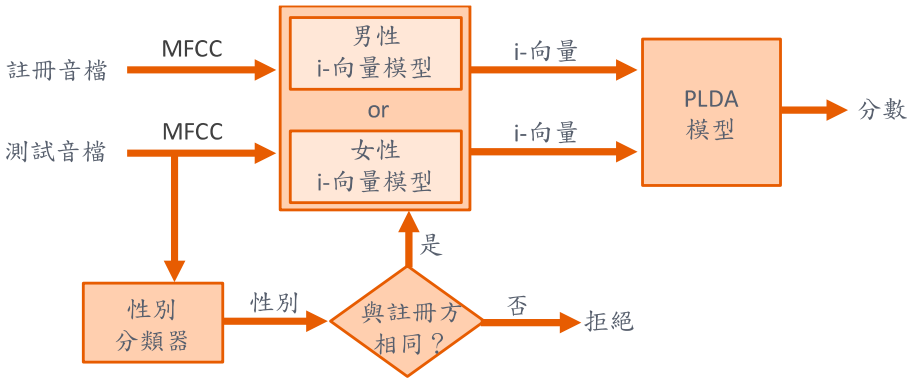


圖 8、方法一之流程架構圖

此版本之流程與前人作法類似，不同之處在於其多出了一個步驟：在以性別分類器判斷出測試語者之性別後，須檢視其是否與註冊語者之性別相同，若相同，則繼續進行原本的流程—以性別相關模型抽出雙方的 i-向量之後，以機率性線性判別分析模型計算出分數；若不同，則直接拒絕此測試語者（打上非常低的分數），因為不同性別的聲音不可能屬於同一位語者。如此一來，便能消除在前人作法中因不同性別的 i-向量在空間分佈上的重疊現象而造成的錯誤接受率攀升之問題。方法一流程的實驗結果如圖 9 所示。同樣地，藍線與橘線分別為對照組流程與前人作法流程之實驗結果，而新增的紅線則是

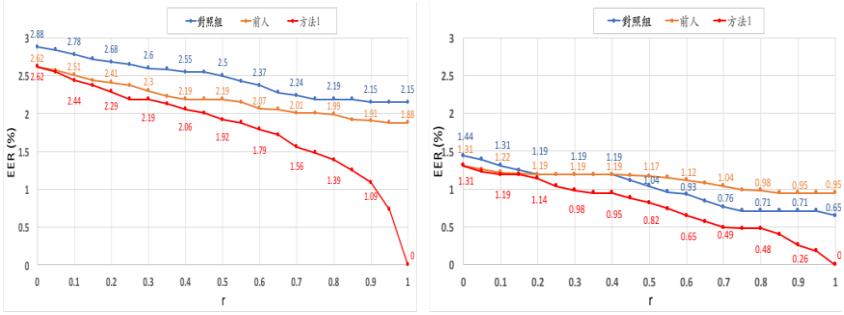


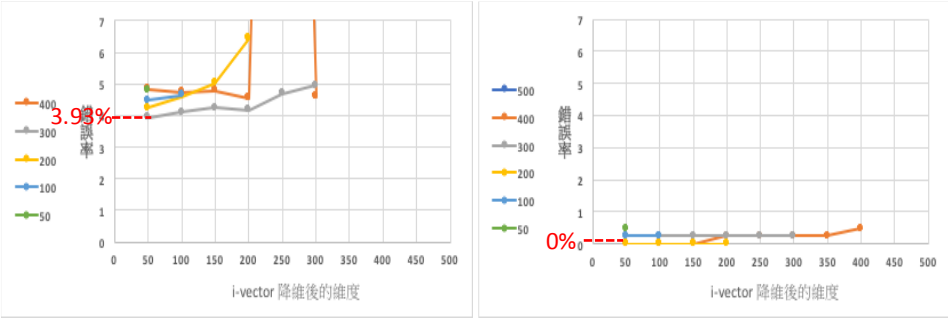
圖 9、方法一流程之結果

本方法一流程之實驗結果。在圖中，我們可以看到在最左側 r = 0 的時候，因為沒有註冊語者與測試語者性別不同的樣本，因此方法一與前人做法的表現是相同的；但慢慢往右側 r = 1 方向移動時，由於註冊方與測試方性別不同的樣本皆被系統拒絕，因此方法一在兩種語料的等錯誤率始終低於對照組，甚至越往右便越與對照組的結果拉開距離；到最後 r = 1 時，由於所有錯誤樣本皆成功被系統拒絕，等錯誤率便可達到 0%。此實驗結果除證明了前人做法確實存在著本論文所預測的缺陷與造成它的原因，亦證明了方法一流程在改善此種缺陷上的效果。

五、實驗三：性別分類器實作

支持向量機性別分類器

此分類器同樣以性別不相依模型產生的 i-向量作為特徵，並以支持向量機作為二元分類器，判斷音檔的性別。本分類器之實驗結果如圖 10 所示：



NIST SRE 2010　　TIMIT

圖 10、支持向量機性別分類器之結果

本分類器之實驗與前者相同，調整了 i-向量模型中的 i-向量之初始維度與其降維後的維度，兩者分別在圖 10 中以不同顏色的線與圖之橫軸表示；此外，亦調整了支持向量機中最重要的 2 個參數：cost 與 gamma。在圖中，每一個數據點皆是將 cost 與 gamma 分別在集合{0.1, 1, 10, 100, 1000}與{0.01, 0.1, 1, 10, 100}內進行排列組合，並以網格搜尋法（grid search）所尋找到的最佳參數組合之結果。同樣地，圖中的紅色虛線所標記的是兩種語料庫中最佳參數所表現出的結果，NIST SRE 2010 語料庫的錯誤率可達 3.93%，而 TIMIT 語料庫則可以達到完全分類正確，即 0%錯誤率的結果。

神經網路性別分類器

此分類器同樣以性別不相依模型產生的 i-向量作為特徵，並以中間 4 層的神經網路模型作為二元分類器，判斷音檔的性別。而神經網路模型的參數（keras API）如表 2 所示：

模型參數	層參數
batch _size = 128	kernel_initializer='glorot_normal'
epochs = 200	bias_initializer='zeros'
loss = binary_crossentropy	BatchNormalization()
optimizer = Adam(lr=3e-4)	Activation('relu')
metrics = 'accuracy'	Dropout(0.35)

表 2、神經網路性別分類器之模型參數

其中，模型的每層節點數量由輸入至輸出端線性減少，舉例來說，若 i-向量維度數為 n，神經網路維中間層數為 4 層，則模型之各層結點數則如圖 11 所示：

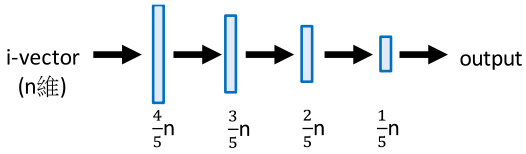


圖 11、神經網路模型之形狀

本分類器之實驗結果如圖 12 所示：

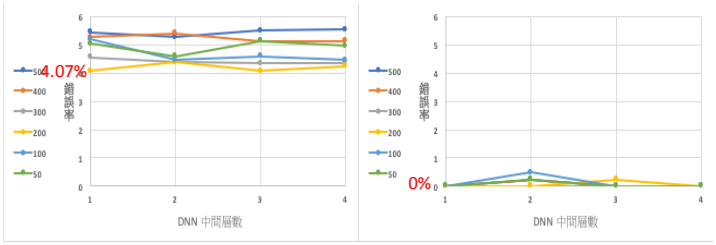


圖 12、神經網路性別分類器之結果

與前一個分類器不同的是，本實驗所使用的 i-向量一律維持初始維度，而不使用線性判別分析模型進行降維，而本分類器之實驗過程調整的兩個參數則改為是 i-向量之初始維度與神經網路模型的中間層數，兩者分別在圖 12 中以不同顏色的線與圖之橫軸表示。同樣地，圖中的紅色虛線所標記的是兩種語料庫中最佳參數所表現出的結果，NIST SRE 2010 語料庫的錯誤率可達 4.07%，而 TIMIT 語料庫則與支持向量機性別分類器一樣，可以達到完全分類正確，即 0%錯誤率的結果。

六、實驗四：性別資訊應用方法 — 當性別分類器表現不良

實驗四同樣將介紹若干種性別資訊應用於語者驗證系統之方法，但不同於實驗二以其真實性別代替性別分類器的預測結果，本實驗將使用在實驗三中實作的性別分類器，並特別針對當性別分類器表現不良的情況，介紹其他不同於實驗二的性別資訊應用方法。

實際套用性別分類器所預測性別

在實驗三中，因 TIMIT 語料庫在性別分類器上可以達到零錯誤率的效果，故其於語者驗證流程中套用實驗三所實作的性別分類器之結果，將與實驗二無異，因此本章節之實驗將僅包含 NIST SRE 2010 語料的部分。圖 13 是對照組流程與將在實驗三中於 NIST SRE 2010 語料中表現最好的支持向量機性別分類器套用於實驗二各流程之實驗結果：

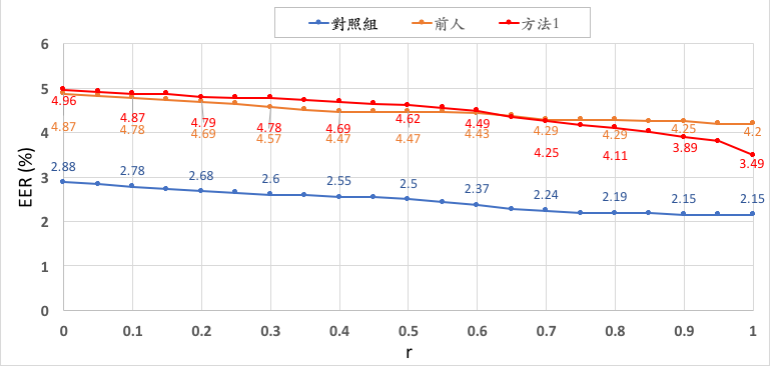


圖 13、各項語者驗證流程使用支持向量機性別分類器之實驗結果

同樣地，藍線、橘線、紅線分別代表對照組、前人作法與方法一的結果，由圖中可以見到，在實際套用真實性別分類器之後，實驗二中的所有流程在不同 r 值下的實驗結果產生了明顯的退步，甚至比對照組流程之結果更差。雖支持向量機性別分類器之錯誤率僅有 3.93%，但這些少量的錯誤卻會造成語者驗證系統之表現造成巨大影響，其原因一樣可分為錯誤接受與錯誤拒絕兩種情形來探討：我們可以想像，當註冊語者與測試語者為不同人時，即使性別分類器誤判了測試語者的性別，系統錯誤接受此負樣本的機率仍然不高。但當註冊語者與測試語者為同一人，一旦測試語者被誤判性別，使用前人作法的系統會分別以不同性別之模型為註冊方與測試方抽取 i-向量，此種情形幾乎必然會造成系統為此正樣本評出較非常低的分數；而使用方法一流程之系統，在檢測出註冊語者與

測試語者分屬不同性別後，便會對此樣本予以拒絕；最後，使用方法二流程之系統，由於訓練資料中同一語者的 i-向量中用以代表性別的最後一維特徵皆有相同的數值，因此當評分器見到註冊語者與測試語者在該維特徵之數值相異，自然同樣會評出非常低的分數。綜上所述，實驗二的各種方法皆會為大量正樣本評出低落的分數，進而使得等錯誤率跟著大幅提升。

方法二：檢查測試音檔的男女性別機率差

在先前的實驗中，我們可以見到圖 13 與圖 9 之差距來自於少數遭誤判性別的測試語者被實驗二之各流程系統錯誤拒絕。倘若我們能將這些測試語者遭誤判性別的樣本與其他樣本分別交由對照組流程與實驗二之改良流程進行判別，理論上便能得到近似於實驗二之效果。若欲實際執行此概念，要如何預測性別分類器能否正確預測測試語者之性別，必然是此方法必須面對的第一個問題，而本流程將提供一個較簡單的預測方法：以測試音檔的性別機率來預測性別分類器能否正確預測其性別。實驗三所實做的性別分類器，除了能預測一位語者之性別，更可以精確地計算出該語者屬於男性或女性的機率，而若一份音檔的兩個機率值相差非常多，代表性別分類器對此音檔之性別預測結果有較大的信心，因此便將此樣本交由方法一流程進行判別，反之，若一份音檔的兩個機率值相差過小，則代表性別分類器的判斷有較大的機會出錯，因此便將此樣本交由對照組流程進行判別。依此概念，便可以衍生出如圖 14 之流程：

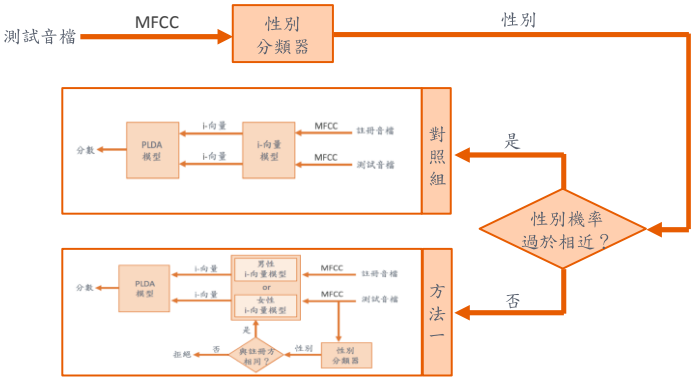


圖 14、方法二之流程架構圖

本流程使用實驗三中的支持向量機性別分類器為測試音檔進行性別預測，而系統之整體

表現與「測試音檔性別對數機率差之門檻值（t）」之關係則如圖 15 所示：

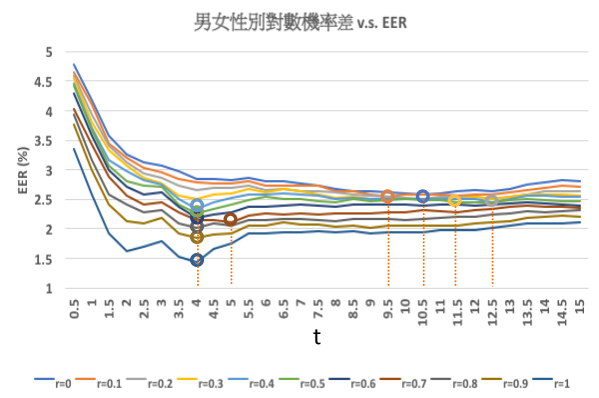


圖 15、方法二流程之選定測試音檔性別機率差門檻

由圖中可見，每條線皆呈現出兩側高、中間低的現象，而每條線的最低點位置（在圖中以圓圈標出）亦不相同，例如：當 r 值較小（0 ~ 0.3）時，這些線條之最低點出現在圖的右半部，顯示其最佳門檻值偏大；而當 r 值較大（0.4 ~ 1）時，這些線條之最低點出現在圖的左半部，顯示其最佳門檻值偏小。由此現象可以推斷，男女性別機率差之門檻並沒有最佳解，而是會隨著資料的 r 值而變動。圖 16 為方法二流程指定 t 值為 4 時之實驗結果：

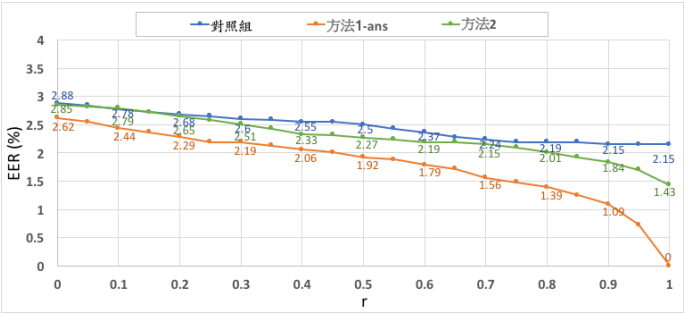


圖 16、方法二流程之結果

同樣地，藍線與橘線分別代表對照組與方法一流程使用真實性別代替性別分類器之結果，而新增的綠線則是本方法二流程之結果。我們可以看到，此方法二流程終於達到了最初的預期，也就是在性別分類器表現不良之前提下，仍然超越對照組流程的目標。

五、結論

本論文之各項實驗結果證實，無論性別分類器的表現或仿冒者的性別組成，皆會影響與者驗證系統的表現，因此不同的性別資訊運用方法，皆各有其適用的前提。以下整

理本論文所使用的各種方法與其對應的適用情形作為總結：

(1) 當我們能確保系統的使用者可以提供品質良好的錄音，則性別分類器有較大機會表現良好，因此可選擇方法一。

(2) 當我們不能保證使用者能提供優良品質之錄音，則性別分類器有較大機會表現不良，因此在這種情況下，方法二為較合適之選項。

(3) 傳統作法的優點為不必花費額外時間進行性別預測，因此當我們對於系統的整體運算速度有較高的需求，則可以選擇維持原來的傳統作法。

參考文獻

[1] https://www.nist.gov/itl/iad/mig/speaker-recognition-evaluation-2010

[2] Garofolo, John S., et al. TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. Web Download. Philadelphia: Linguistic Data Consortium, 1993.

[3] Anjos, André, et al. "Continuously reproducing toolchains in pattern recognition and machine learning experiments." (2017).

[4] Kenny, Patrick. "Joint factor analysis of speaker and session variability: Theory and algorithms." *CRIM, Montreal,(Report) CRIM-06/08-13* 14 (2005): 28-29.

[5] Kenny, Patrick, et al. "A study of interspeaker variability in speaker verification." *IEEE Transactions on Audio, Speech, and Language Processing* 16.5 (2008): 980-988.

[6] Dehak, Najim, et al. "Front-end factor analysis for speaker verification." *IEEE Transactions on Audio, Speech, and Language Processing* 19.4 (2011): 788-798.

[7] Prince, Simon JD, and James H. Elder. "Probabilistic linear discriminant analysis for inferences about identity." Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, 2007.

[8] Kenny, Patrick. "Bayesian speaker verification with heavy-tailed priors." Odyssey. 2010.

[9] Garcia-Romero, Daniel, and Carol Y. Espy-Wilson. "Analysis of i-vector length normalization in speaker recognition systems." *Twelfth Annual Conference of the International Speech Communication Association*. 2011.

[10] Mika, Sebastian, et al. "Fisher discriminant analysis with kernels." Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop.. Ieee, 1999.

[11] Lyu, Siwei, and Eero P. Simoncelli. "Nonlinear extraction of independent components of natural images using radial gaussianization." *Neural computation* 21.6 (2009): 1485-1519.