朝陽科技大學

資訊管理系碩士在職專班一年級

91(下)Neural Network Final Project

類神經網路於語音四聲的辨識

指導老師：李麗華教授

學生姓名：朱 孝 國

學號：９１５４６１０

中 華 民 國 九 十 二 年 六 月

摘要

本文以倒傳遞類神經網路(Backpropagation Neural Network ；BPN) **為架構，設計一針對特定字彙、特定語者、非連續語音之國語發音四聲辨識系統。**

對於同樣的字，依照發聲而有不同語意的情況在語音辨識上是相當常見且必須要處理的問題，雖然語音辨識相關研究頗多，但中文語系並未像國外有 Festival 等開放原始碼之軟體可供研究，故本文輔以原始碼與相關 GPL 工具之結合，實有降低技術門檻，理論與實務結合之重大社會意義。

關鍵字：語音辨識，類神經網路，開放原始碼

簡介

隨著人們對語音信號的研究，今天我們已經能夠憑藉著語音和電腦溝通了，也就是說，電腦可以分辨出人類所說的話。因此想要操作電腦除了可以經由鍵盤或滑鼠輸入命令外，最直接的方式便是由嘴巴說出指令囉！譬如說『芝麻開門』，門就開了，說『音樂』，音響就啟動了，如果能夠有這樣的一套系統，生活不是就更方便了嗎？

所以電腦語音辨認就是讓電腦聽得懂人們所講的話，其基本架構包含 2 大部分，一個是訓練的部分，另一個是辨認的部分。

一般語音辨認在分類上有下列分野

1 依辨認字彙多寡，數量愈多辨識的困難也愈高
  - 1.1 特定字彙：特定的單字、詞或片語
  - 1.2 少量字彙：100 個單字左右
  - 1.3 大量字彙：例如辨識全部的中文字
2 依使用者的限制
  - 2.1 特定語者(speaker dependent)：使用前系統必須先學習，也就是將使用者的語音參考樣本存入比對的資料庫。
  - 2.2 不特定語者(speaker independent)：使用系統前不需先學習，系統以內存有眾多語者的資料可供比對，但是要做到這樣的規格，又有不錯的辨識率，實在不容易。
3 依說話方式的連續性
  - 3.1 非連續語音：所說的每一個字必須分開。
  - 3.2 連續語音：以一般自然流利的方式說話，但是這種連音問題很難辨認的相當的好。

語音辨認本身是屬於樣本分類(Pattern Classification)的問題。目前至少有四種方法是運用來解決此方面的問題。

1. 樣本比對：使用動態規劃原理來做樣本比對。
2. 統計模式：如使用 HMM(Hidden Markov Modeling)。
3. 知識基礎系統：就是專家系統。
4. 類神經網路(Artificial Neural Network;ANN)：類神經網路具有高速的平行計算與學習、記憶、容錯等能力，因此應用在複雜且資料量大的語音問題便相當的適合。

本研究擬採用類神經網路的方式來做國語語音的四聲辨識，應用其中網路類型之一的倒傳遞類神經網路(Backpropagation Neural Network;BPN)來擔任樣本分類訓練與辨識的角色。

系統架構

1. BPN 的訓練過程

1.1. 語料錄製

語料的選擇以特定語者，就是我本人啦，包含四個音的詞，選定為『類神經網路』五個字，各發四種音共 20 個音，以 GoldWave 軟體錄音,選擇 u-law,mono 型態存檔，即錄製取樣頻率為 8000Khz，16Bit，單聲道的音波檔如下：

	類	神	經	網	路
一聲	ㄌㄟ	ㄕㄣ	ㄐㄧㄥ	ㄨㄤ	ㄌㄨ
二聲	ㄌㄟˊ	ㄕㄣˊ	ㄐㄧㄥˊ	ㄨㄤˊ	ㄌㄨˊ
三聲	ㄌㄟˇ	ㄕㄣˇ	ㄐㄧㄥˇ	ㄨㄤˇ	ㄌㄨˇ
四聲	ㄌㄟˋ	ㄕㄣˋ	ㄐㄧㄥˋ	ㄨㄤˋ	ㄌㄨˋ

1.2. 語料波形特徵擷取

如果以 8000Khz 的取樣頻率對語音作曲樣，那麼一秒鐘長度的語音就有 8000 個取樣點，如果以這麼龐大的資料量來做模型的訓練與辨識的工作，可能很難達到即時辨識的目的。因此訓練模型的第一步，就是對語音作前處理以及求出語音資料的特徵值。語音辨識技術中使用的特徵值有許多種，本研究中所使用的是倒頻譜參數及轉移倒頻譜參數 2 種。從取出語音的取樣值到求出特徵值共有六個步驟：

1. 取出語音內的取樣資料。
2. 將語音分成一個個的音框。
3. 以能量量測的方法去除靜音的部分。
4. 對每個音框的語音資料作預強調及漢明窗的處理。
5. 對每個音框求出一組線性預估係數，也就是 LPC。
6. 以 LPC 來求出倒頻譜參數及轉移倒頻譜參數，也就是特徵值。

以上的步驟由洪朝貴老師撰寫為 C 語言的程式名為 sprec.c 結合相關的函式庫，取得之特徵值如下

訓練語料的原始特徵值(20 個音)	
ㄌㄟ	69 73 74 75 75 74 73 72 72 73 75
ㄕㄣ	74 75 74 74 73 73 73 73 73
ㄐㄧㄥ	71 73 73 121 73 74 74 75
ㄨㄤ	66 68 71 70 118 118 118 71 71 70 71
ㄌㄨ	71 73 72 72 72 72 71 71 71
ㄌㄟˊ	46 45 44 45 46 47 49 51 54 55 63 75 80
ㄕㄣˊ	77 50 36 24 52 51 51 51 50 50 99 52 54 59
ㄐㄧㄥˊ	28 23 49 49 50 50 49 49 50 52 53 58
ㄨㄤˊ	49 50 50 49 50 97 98 51 53 55 59 63
ㄌㄨˊ	52 51 52 51 51 53 56 62 69 75 83
ㄌㄟˇ	42 50 48 46 45 43 43 43 44 49 57 70 78 83
ㄕㄣˇ	116 50 63 41 47 43 44 44 43 43 45 47 49 52
ㄐㄧㄥˇ	58 0 44 44 45 44 41 40 45 48 51 60
ㄨㄤˇ	44 45 45 91 91 42 91 91 73 93 50 59
ㄌㄨˇ	47 44 42 42 42 41 43 46 51 62 77
ㄌㄟˋ	78 83 80 75 72 69 64 60 54
ㄕㄣˋ	79 79 77 75 73 68 112 58
ㄐㄧㄥˋ	128 127 77 75 73 68 62 56
ㄨㄤˋ	75 79 77 95 73 71 69 64 58
ㄌㄨˋ	77 79 77 74 70

1.3. 正規化處理

根據以上的特徵值，我們可以發現其長度不一致，且其值域落在[0~116]之間，很明顯的需要做正規化的動作以使資料可以被 BPN 的 Input Layer 接受，因此有了下列 2 種問題：

1．要用什麼方法來使特徵值長度一致

2．是否要將特徵值的值域縮小到[0~1]之間

1.4. 網路訓練

本來是很單純的訓練過程，但因為考量到正規化，於是我決定以試誤法來進行測試，看網路是否能夠收斂，決定的網路訓練模型皆以 2000 次學習循環，學習率為 1 進行下列四種網路的訓練：

1．將資料以 0 補足到最大長度，就進行訓練(特徵值最大長度為 14)

輸入層單元數	隱藏層單元數	輸出層單元數	結果
14	10	4	無法收斂

2．將資料以 0 補足到最大長度,且予以正規化/100

輸入層單元數	隱藏層單元數	輸出層單元數	結果
14	10	4	可收斂

3．將資料以內插法取最大長度的 2 倍-1 補到 27 個

輸入層單元數	隱藏層單元數	輸出層單元數	結果
27	18	4	無法收斂

4．將資料以內插法取最大長度的 2 倍-1 補到 27 個,且予以正規化/100

輸入層單元數	隱藏層單元數	輸出層單元數	結果
27	18	4	可收斂

就以上的測試，我們可以知道特徵值的縮小值域是收斂的關鍵，但究竟是要以候補零法或以內插法，在訓練過程時看不太出來差異，因此本研究保留第 2 項與第 4 項的方式移到辨識過程再做比較。

2. BPN 的辨識過程

因本研究採取特定語者與特定特定字彙的因素，因此在辨識時的語料也發相同的字音，但於不同的時間錄製以確保與訓練時的語料不同。

2.1. 補零法

辨識過程畫面擷取

```
D:\perl\neural\sprec1>perl bpn.pl -t tonetest.txt -u bpn.ini
-----------------------------------------------------------------
    1/5 Max=1    value=4, Second=0    value=1, Third=0    value=1
    2/5 Max=1    value=4, Second=0.58 value=3, Third=0    value=1
    3/5 Max=0.99 value=1, Second=0    value=1, Third=0    value=1
    4/5 Max=0.95 value=3, Second=0    value=1, Third=0    value=1
    5/5 Max=1    value=4, Second=0.15 value=3, Third=0    value=1
-----------------------------------------------------------------
Begin time:2003/8/16  3:49:46
End time  :2003/8/16  3:49:46
innode=14, hidden node=10, out node=4, lr=1, epochs=2000
=================================================================
```

發音	判定	結果
類(ㄌㄟˋ)	4 聲	正確
神(ㄕㄣˊ)	4 聲	錯誤
經(ㄐㄧㄥ)	1 聲	正確
網(ㄨㄤˇ)	3 聲	正確
路(ㄌㄨˋ)	4 聲	正確

由辨識結果來看，補零法在判斷 2 聲時較行不通。

2.2. 內插法

辨識過程畫面擷取

```
D:\perl\neural\sprec1>perl bpn.pl -t tonetest.txt -u bpn.ini
------------------------------------------------------------------
    1/5 Max=1    value=4, Second=0.01 value=3, Third=0    value=1
    2/5 Max=0.22 value=2, Second=0.11 value=1, Third=0    value=1
    3/5 Max=0.99 value=1, Second=0    value=1, Third=0    value=1
    4/5 Max=0.98 value=2, Second=0.23 value=1, Third=0    value=1
    5/5 Max=1    value=4, Second=0    value=1, Third=0    value=1
------------------------------------------------------------------
Begin time:2003/8/16  3:43:7
End time  :2003/8/16  3:43:7
innode=27, hidden node=18, out node=4, lr=1, epochs=2000
==================================================================
```

發音	判定	結果
類(ㄌㄟˋ)	4聲	正確
神(ㄕㄣˊ)	2聲	正確
經(ㄐㄧㄥ)	1聲	正確
網(ㄨㄤˇ)	2聲	錯誤
路(ㄌㄨˋ)	4聲	正確

由辨識結果來看，內插法在判斷 3 聲時較行不通。

實驗結果

本實驗證明應用類神經網路於語音辨識的可行性，但由實驗結果來看，學生有下列心得提出來做參考：

1. 對於如何將資料予以數位化成可供類神經網路參考有了更清楚的瞭解，特別是對於資料有不同的正規化方式。
2. 錄音時要注意與麥克風之間的雜音問題,還有語者本身的換氣聲，都會干擾特徵值，特別是一開始錄音與發音時，會有單一條的波的振幅特別大，之前沒有注意它會造成特徵值不夠平滑。
3. 適當的決定音框的門檻值很重要,假設發 5 個音，若門檻值從 0.2~0.9 都可正確切割為 5 個音框的話，那要選 0.9，因為選擇小的門檻值會造成特徵值的前後有不合理的數值出現，而且發一樣的音量，但二聲與三聲的門檻值會特別低，因此我想訓練的樣本應該要逐字決定其門檻值才能有較完美的結果。
4. 當我將特徵值算出來時，其實我有點傻眼了，因為看起來每個音的值都差異頗大，我還有點沒信心以類神經網路可以訓練的出來呢，後來以 BPN 的作業程式拿來做一些修改，發覺居然可以耶，真是非常的高興。
5. 做完這次的期末作業，可以感受到事先的蒐集資料，進行步驟與方式與真正訓練與辨識的時間相比真是多出很多，因此我也可以瞭解 ANN 之所以被當作工具來使用的意涵了，難怪這麼多人的研究都可以和它沾上邊。

附錄

1．訓練語料的特徵值(補零法處理)

ㄉㄟ	0.69 0.73 0.74 0.75 0.75 0.74 0.73 0.72 0.72 0.73 0.75 0 0 0
ㄕㄣ	0.74 0.75 0.74 0.74 0.73 0.73 0.73 0.73 0.73 0 0 0 0 0
ㄐㄧㄥ	0.71 0.73 0.73 1.21 0.73 0.74 0.74 0.75 0 0 0 0 0 0
ㄨㄤ	0.66 0.68 0.71 0.7 1.18 1.18 1.18 0.71 0.71 0.7 0.71 0 0 0
ㄉㄨ	0.71 0.73 0.72 0.72 0.72 0.72 0.71 0.71 0.71 0 0 0 0 0
ㄉㄟˊ	0.46 0.45 0.44 0.45 0.46 0.47 0.49 0.51 0.54 0.55 0.63 0.75 0.8 0
ㄕㄣˊ	0.77 0.5 0.36 0.24 0.52 0.51 0.51 0.51 0.5 0.5 0.99 0.52 0.54 0.59
ㄐㄧㄥˊ	0.28 0.23 0.49 0.49 0.5 0.5 0.49 0.49 0.5 0.52 0.53 0.58 0 0
ㄨㄤˊ	0.49 0.5 0.5 0.49 0.5 0.97 0.98 0.51 0.53 0.55 0.59 0.63 0 0
ㄉㄨˊ	0.52 0.51 0.52 0.51 0.51 0.53 0.56 0.62 0.69 0.75 0.83 0 0 0
ㄉㄟˇ	0.42 0.5 0.48 0.46 0.45 0.43 0.43 0.43 0.44 0.49 0.57 0.7 0.78 0.83
ㄕㄣˇ	1.16 0.5 0.63 0.41 0.47 0.43 0.44 0.44 0.43 0.43 0.45 0.47 0.49 0.52
ㄐㄧㄥˇ	0.58 0 0.44 0.44 0.45 0.44 0.41 0.4 0.45 0.48 0.51 0.6 0 0
ㄨㄤˇ	0.44 0.45 0.45 0.91 0.91 0.42 0.91 0.91 0.73 0.93 0.5 0.59 0 0
ㄉㄨˇ	0.47 0.44 0.42 0.42 0.42 0.41 0.43 0.46 0.51 0.62 0.77 0 0 0
ㄉㄟˋ	0.78 0.83 0.8 0.75 0.72 0.69 0.64 0.6 0.54 0 0 0 0 0
ㄕㄣˋ	0.79 0.79 0.77 0.75 0.73 0.68 1.12 0.58 0 0 0 0 0 0
ㄐㄧㄥˋ	1.28 1.27 0.77 0.75 0.73 0.68 0.62 0.56 0 0 0 0 0 0
ㄨㄤˋ	0.75 0.79 0.77 0.95 0.73 0.71 0.69 0.64 0.58 0 0 0 0 0
ㄉㄨˋ	0.77 0.79 0.77 0.74 0.7 0 0 0 0 0 0 0 0 0

2.訓練語料的特徵值(內插法處理)

ㄌㄟ	0.69 0.705385 0.720769 0.731538 0.735385 0.739231 0.743077 0.746923 0.75 0.75 0.75 0.747692 0.743846 0.74 0.736154 0.732308 0.728462 0.724615 0.720769 0.72 0.72 0.720769 0.724615 0.728462 0.734615 0.742308 0.75
ㄕㄣ	0.74 0.743077 0.746154 0.749231 0.747692 0.744615 0.741538 0.74 0.74 0.74 0.739231 0.736154 0.733077 0.73 0.73 0.73 0.73 0.73 0.73 0.73 0.73 0.73 0.73 0.73 0.73 0.73 0.73
ㄐㄧㄥ	0.71 0.715385 0.720769 0.726154 0.73 0.73 0.73 0.73 0.803846 0.933077 1.06231 1.19154 1.09923 0.97 0.840769 0.730385 0.733077 0.735769 0.738462 0.74 0.74 0.74 0.74 0.741923 0.744615 0.747308 0.75
ㄨㄤ	0.66 0.667692 0.675385 0.684615 0.696154 0.707692 0.706923 0.703077 0.736923 0.921538 1.10615 1.18 1.18 1.18 1.18 1.18 1.10769 0.926923 0.746154 0.71 0.71 0.709231 0.705385 0.701538 0.702308 0.706154 0.71
ㄌㄨ	0.71 0.721429 0.728571 0.722857 0.72 0.72 0.72 0.72 0.72 0.718571 0.712857 0.71 0.71 0.71 0.710.71 0.716154 0.722308 0.728462 0.727692 0.724615 0.721538 0.72 0.72 0.72 0.72 0.72 0.72 0.72 0.72 0.72 0.72 0.72 0.717692 0.714615 0.711538 0.71 0.71 0.71 0.71 0.71 0.71 0.71 0.71
ㄌㄟˊ	0.46 0.455385 0.450769 0.446154 0.441538 0.443077 0.447692 0.452308 0.456923 0.461538 0.466154 0.471538 0.480769 0.49 0.499231 0.508462 0.521538 0.535385 0.543077 0.547692 0.568462 0.605385 0.648462 0.703846 0.753846 0.776923 0.8
ㄕㄣˊ	0.77 0.635 0.5 0.43 0.36 0.3 0.24 0.38 0.52 0.515 0.51 0.51 0.51 0.51 0.51 0.505 0.5 0.5 0.5 0.745 0.99 0.755 0.52 0.53 0.54 0.565 0.59
ㄐㄧㄥˊ	0.28 0.258846 0.237692 0.3 0.41 0.49 0.49 0.49 0.493846 0.498077 0.5 0.5 0.499231 0.495 0.490769 0.49 0.49 0.491923 0.496154 0.500769 0.509231 0.517692 0.523077 0.527308 0.537692 0.558846 0.58
ㄨㄤˊ	0.49 0.494231 0.498462 0.5 0.5 0.498846 0.494615 0.490385 0.493846 0.498077 0.608462 0.807308 0.970769 0.975 0.979231 0.817308 0.618462 0.513846 0.522308 0.530769 0.539231 0.547692 0.562308 0.579231 0.596154 0.613077 0.63
ㄌㄨˊ	0.52 0.516154 0.512308 0.511538 0.515385 0.519231 0.516923 0.513077 0.51 0.51 0.51 0.514615 0.522308 0.53 0.541538 0.553077 0.569231 0.592308 0.615385 0.641538 0.668462 0.694615 0.717692 0.740769 0.768462 0.799231 0.83

ㄌㄟˇ	0.42 0.46 0.5 0.49 0.48 0.47 0.46 0.455 0.45 0.44 0.43 0.43 0.43 0.43 0.43 0.435 0.44 0.465 0.49 0.53 0.57 0.635 0.7 0.74 0.78 0.805 0.83
ㄕㄣˇ	1.16 0.83 0.5 0.565 0.63 0.52 0.41 0.44 0.47 0.45 0.43 0.435 0.44 0.44 0.44 0.435 0.43 0.43 0.43 0.44 0.45 0.46 0.47 0.48 0.49 0.505 0.52
ㄐㄧㄥˇ	0.58 0.334615 0.0892308 0.118462 0.304615 0.44 0.44 0.44 0.443846 0.448077 0.447692 0.443462 0.437692 0.425 0.412308 0.406538 0.402308 0.409615 0.430769 0.451154 0.463846 0.476538 0.489231 0.501923 0.523846 0.561923 0.6
ㄨㄤˇ	0.44 0.444231 0.448462 0.45 0.45 0.503077 0.697692 0.892308 0.91 0.91 0.796923 0.589615 0.457692 0.665 0.872308 0.91 0.91 0.875385 0.799231 0.737692 0.822308 0.906923 0.797692 0.615769 0.513846 0.551923 0.59
ㄌㄨˇ	0.47 0.458462 0.446923 0.436923 0.429231 0.421538 0.42 0.42 0.42 0.42 0.42 0.417692 0.413846 0.41 0.417692 0.425385 0.434615 0.446154 0.457692 0.475385 0.494615 0.518462 0.560769 0.603077 0.654615 0.712308 0.77
ㄌㄟˋ	0.78 0.795385 0.810769 0.826154 0.823077 0.813846 0.804615 0.792308 0.776923 0.761538 0.747692 0.738462 0.729231 0.72 0.710769 0.701538 0.692308 0.678462 0.663077 0.647692 0.633846 0.621538 0.609231 0.595385 0.576923 0.558462 0.54
ㄕㄣˋ	0.79 0.79 0.79 0.79 0.788462 0.783077 0.777692 0.772308 0.766923 0.761538 0.756154 0.750769 0.745385 0.74 0.734615 0.728077 0.714615 0.701154 0.687692 0.730769 0.849231 0.967692 1.08615 1.01615 0.870769 0.725385 0.58
ㄐㄧㄥˋ	1.28 1.27731 1.27462 1.27192 1.23154 1.09692 0.962308 0.827692 0.766923 0.761538 0.756154 0.750769 0.745385 0.74 0.734615 0.728077 0.714615 0.701154 0.687692 0.673077 0.656923 0.640769 0.624615 0.608462 0.592308 0.576154 0.56
ㄨㄤˋ	0.75 0.772857 0.787143 0.775714 0.821429 0.924286 0.855714 0.73 0.718571 0.707143 0.695714 0.675714 0.647143 0.614286 0.580.75 0.762308 0.774615 0.786923 0.785385 0.779231 0.773077 0.797692 0.853077 0.908462 0.933077 0.865385 0.797692 0.73 0.723846 0.717692 0.711538 0.705385 0.699231 0.693077 0.682308 0.666923 0.651538 0.635385 0.616923 0.598462 0.58
ㄌㄨˋ	0.77 0.773077 0.776154 0.779231 0.782308 0.785385 0.788462 0.788462 0.785385 0.782308 0.779231 0.776154 0.773077 0.77 0.765385 0.760769 0.756154 0.751538 0.746923 0.742308 0.736923 0.730769 0.724615 0.718462 0.712308 0.706154 0.7

3. 辨識語料的特徵值(補零法處理)

ㄉㄟˋ	0.79 0.81 0.76 0.72 0.67 0.62 0.56 0.5 0 0 0 0 0 0 0.743077 0.732308 0.721538
ㄕㄎˊ	0.47 0.49 0.51 0.53 0.54 0 0 0 0 0 0 0 0 0 0
ㄐㄧㄥ	0.74 0.75 0.74 0.75 0.77 0.77 0.77 0.75 0 0 0 0 0 0
ㄨㄤˇ	0.41 0.38 0.35 0.33 0.78 0.29 0.77 0.26 0.81 0.3 0 0 0 0
ㄉㄨˋ	0.84 0.86 0.82 0.71 0 0 0 0 0 0 0 0 0 0

4. 辨識語料的特徵值(內插法處理)

ㄉㄟˋ	0.79 0.795385 0.800769 0.806154 0.806154 0.792692 0.779231 0.765769 0.753846
ㄕㄎˊ	0.47 0.473077 0.476154 0.479231 0.482308 0.485385 0.488462 0.491538 0.494615
ㄐㄧㄥ	0.74 0.742692 0.745385 0.748077 0.749231 0.746538 0.743846 0.741154 0.741538
ㄨㄤˇ	0.41 0.399615 0.389231 0.378846 0.368462 0.358077 0.348462 0.341538 0.334615
ㄉㄨˋ	0.84 0.842308 0.844615 0.846923 0.849231 0.851538 0.853846 0.856154 0.858462