

# 智慧語音情緒識別之自我檢測系統

## Self-detection System of Intelligent Speech Emotion Recognition

陳佳鈞<sup>1</sup> 徐豐明<sup>2</sup>

CHEN, CHIA CHUN<sup>1</sup> SHYU, FONG MING<sup>2</sup>

<sup>1</sup>國立臺中科技大學 多媒體設計系研究所 研究生

<sup>1</sup>National Taichung University of Science and Technology Department of Multimedia Design Student

E-mail : [s1810722018@gms.nutc.edu.tw](mailto:s1810722018@gms.nutc.edu.tw)

<sup>2</sup>國立臺中科技大學 多媒體設計系研究所 副教授

<sup>2</sup>National Taichung University of Science and Technology Department of Multimedia Design Associate Professor

E-mail : [fms@nutc.edu.tw](mailto:fms@nutc.edu.tw)

### 摘要

現今生活壓力越來越大常常壓的人喘不過氣，因壓力過大引發精神疾病而自殺的事件頻傳，許多研究也開始轉往在人工智慧的應用當中，藉此解決或探討某些議題，聊天機器人在未來也有很大的市場，但沒有能判定情緒的類型，本研究使用機器學習結合自然語言處理的技術進行語音情緒識別，當自我能做檢測就可先行評估狀況，並得知後續如何進行處理方式，成效是想達到透過自我檢測系統了解自我的情緒狀況以及後續的處理方式。

**關鍵字：**機器學習、語音情緒識別、自然語言處理、聊天機器人

### Abstract

Nowadays, the pressure of life is increasing, and those who often feel stressed cannot breathe. Suicidal behavior caused by mental illness is usually caused by excessive stress. Many studies have also begun to transfer to the application of artificial intelligence to solve or carry out certain discussions. Chatbots will also be a big market in the future, but chatbots cannot understand emotions. This research combines machine learning and natural language processing techniques for speech emotion recognition. When the self is able to detect it, it can first assess the situation and then assess its handling. The benefit is to understand the emotional state of the self through the self-detection system and subsequent processing methods.

**Keywords :** Machine Learning, Speech Emotion Recognition, Natural Language Processing, Chatbot

## 壹、前言

現今生活壓力越來越大，家庭、工作、同儕關係等瑣事常常壓的人喘不過氣，因壓力過大引發精神疾病而自殺的事件頻傳，而科技的進步可以為生活帶來便利，卻沒有任何一個科技可以完全消除焦慮與煩惱，這些壓力如果沒有一個好的抒發管道，在後期很容易產生精神疾病，但這個社會中往往有些人不太敢去尋求精神科的幫助，然而需要一個聆聽者陪伴與聊天，所以有一個抒發管道是求助於社工師或諮商師，導致社工師、諮商師常常有人手不足的狀況，透過人工智慧結合不同的技術輔助改善其問題。

在隨著台灣老年人口比數逐年增加，長期照護的需求越來越高。據內政部統計顯示，家庭仍是大部分老人照護的主力。照護者的負擔日漸沉重，社福團體的人員相對不足。近年來，這問題不光是只有台灣的困擾，周圍的國家都有類似的問題。為了減輕護理人員在日常生活中對弱勢族群安全保障的負擔，有許多議題開始轉向人工智慧照顧或陪伴上。

由於人工智慧蓬勃的發展，人工智慧所應用到的成面相當廣泛，例：醫療業、金融業、生產業、服務業等等，在人工智慧的影響下影像處理、語音處理、物聯網等等都有相當的成就，近幾年人工智慧結合語音辨識的聊天機器人成為熱門的研究議題，讓人工智慧理解人類的情緒並結合語音辨識，能讓人智慧更清楚明白人類所想要表達的事情，人工智慧理解人類的情緒時，可根據與使用者聊天的狀況，去分析數據資料來回應適當地答案，運用於社工師能輔助來與需要陪伴或聊天的人進行對談，經過這些數據綜合分析後，可得知使用者的情緒或者壓力是否有異常或過大或有憂鬱的傾向，透過這些不同技術的結合可讓生活更加便利與改善現今社會壓力無法宣洩的問題。

許多研究也開始轉往在人工智慧的應用當中，藉此解決或探討某些議題，在關於人工智慧與情緒辨識相關的議題是值得探討並進行研究的，人工智慧與語音情緒識別會是接下來相關議題間的趨勢。

硬體選擇也是影響使用者體驗的重要因素，不同的設備以及外觀能提供的感覺之間也會有差距，在人工智慧的世界裡，有一個很著名的理論就是恐怖谷理論，當這個東西的外觀太像真實的人類時，反而人類是會感到害怕、排斥反應，恐怖谷現象可以用以下想法解釋，如果一個實體「不夠擬人」，那它像人的特徵就會顯眼並且容易辨認，產生移情作用。在另一方面，要是一個實體「足夠擬人」，那它不太像人的特徵就會成為顯眼的部份，在人類的眼中產生一種古怪的感覺。

M. Mori (1970)的假設指出，由於機器人與人類在外表、動作上相似，所以人類亦會對機器人產生正面的情感；直到一個特定程度，他們的反應便會突然變得極為負面。哪怕機器人與人類只有一點點的差別，都會顯得非常顯眼刺眼，讓整個機器人顯得非常僵硬恐怖，使人有面對殭屍的感覺。可是當機器人和人類的相似度繼續上升，相當於普通人之間的相似度的時候，人類對他們的情感反應會再度回到正面，產生人類與人類之間的移情作用。

「恐怖谷」一詞用以形容人類對自己的相似度達到特定程度之機器人的排斥反應。而「谷」就是指在研究裡「好感度對相似度」的關係，如：圖 1，在相似度臨近 100%前，好感度突然墜至反感水平，回升至好感前的那段範圍。這也是為什麼目前市面上機器人的外觀要做成動物的外觀或者是較可愛外觀的原因。

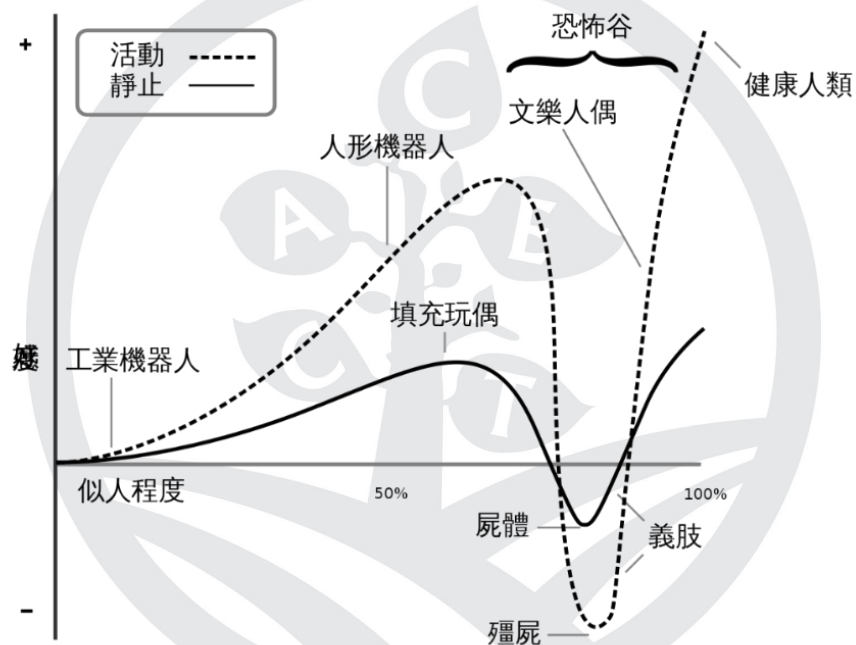


圖 1. 恐怖谷好感度對相似度

(資料來源：<https://reurl.cc/K6oNej>)

(Adang Suwandi Ahmad and Arwin Datumaya Wahyudi Sumari, 2017, July)

## 一、 研究動機與目的

人生總是面臨許多壓力，來自工作、家庭、生活轉變、甚至健康狀況都可能讓你很煩、覺得壓力很大，人工智慧理解人類的情緒再結合語音辨識，可讓人工智慧的應用提升到另外一個層級，當人工智慧結合語音辨識並且理解人類的情緒，可根據與使用者聊天的狀況，去分析數據資料並回應適當地答案，輔助社工師來減輕社工師的工作量以及與需要陪伴或聊天的人進行對談，經過這些聊天過程的數據綜合分析後，可得知使用者自我的情緒是否異常或者壓力是否有異過大。

目前人工智慧結合各項技術的應用是現在大家勇於嘗試的，在人工智慧結合語音辨識的應用，是目前廣泛應用的技術類型，因此聊天機器人的發展性也極為的廣泛，但大多數的聊天機器人並沒有情緒，現今的聊天機器人可以達到控制家電、簡易的服務、回答使用者問題、根據使用者的需求去找尋最佳的答案等，但結合情緒識別上相對來說並沒有，所以這是研究的動機之一。

事實上目前也有相關的研究，但在實際應用層面還是有可發展性，人工智慧的應用中，若人工智慧能夠理解情緒，也可結合更多不同的技術，在台灣這一個逐漸高齡化的社會中，照護型機器人在未來也會有很大的市場，但人類是一個具有豐富情感的動物，只有單純的聊天、陪伴會讓人類感覺沒有溫度，所以讓人工智慧學會理解人類的情緒其中一個很重要的因素之一，然而得知人類的情緒在未來擴展性也非常的大，能得知使用者對東西的喜好，透過喜好推薦使用者感興趣的商品，或者能透過現在的情緒給予使用者最好的回覆，在現今的客服人若人工智慧理解情緒後是可以取代掉的一個市場。

本研究目的是讓使用者達到自我檢測情緒狀況的系統，可透過本系統的檢測能得知本身的情緒指標為何，也能得知下一步該怎麼做的建議，透過自我檢測系統能發現自己該不該去找諮商師談談，如果可先得知自身狀況再去做決定的話，能有效的運用資源，並減少社工師、諮商師的工作量，自我檢測的系統有提供聊天的功能，能提供給使用者一個訴說事情、抒發壓力的一個管道，透過本系統可以解決諮商師與社工師人力不足的狀況以及讓需要找人聊天的人有一個能傾訴心情跟事情的地方。

總結研究目的是讓使用者達到自我檢測情緒狀況的系統，可透過本系統的檢測能得知本身的情緒指標為何，也能得知下一步該怎麼做的建議，並提早知道自己情緒狀況為何，如果有異常發生能盡快地找尋相關的管道尋求協助，可降低精神疾病的產生。

## 二、 研究範圍與限制

為了使研究的目標相對明確並在整理測試結束後較容易進行分析統整，因而本研究的範圍與限制可大致分為三個部分：系統設定部分、體驗類型與數據收集類型。

### (一) 系統設定部分

由於情緒感受屬於主觀感覺的一部分，每個人對自己情感的表達並不一樣，相對的每個人對情緒的認知也有所不同，因此要用客觀的感受去界定情緒異常的範圍，是一件不簡單的事，而在系統設定部分，數據異常的範圍，為採用大部人的情緒感受為基準，而非在系統當中初始即設定何謂異常範圍，以此增加系統判定情緒異常範圍的準確性。

### (二) 問題類型

本論文是聊天抒發壓力以及檢測自我情緒壓力值體驗為主要測試內容，此類型的體驗可修改內容豐富多變化，並有許多現有的資料可作為參考，設計問題的面向非常的重要，問題類型有的較刺激、有的較微平淡，設計問題也必須要有效度跟信度。

聊天體驗當中，由於問題的類型對使用者有不同的感受，每個問題的設計都會影響到數據的結果，因此在設計問題時，會找尋相關領域的專家學者進行合作來確保問題的可靠性，最後再將設計的問題拿出來做測試以及評估效度與信度，當達到一定的標準才會套用在機器人身上，當使用者回答問題以後，機器人所反饋的內容也是相當重要，機器人會根據使用者所回答的內容去判斷如何反饋最佳的回覆，因此在設計問題與設計回答方式的內容是非常重要的，在設計反饋回答的內容，也會找尋相關領域的專家學者協助設計，最後會在進行縝密的評估與測試。

### (三) 數據收集類型

將數據資料綜合分析出結果，減少資料誤差值與盡可能提升精準度，並收集每個回答的狀態，再把每個回答的數據進行運算，才不會因為一個問題的情緒起伏過大而直接判定使用者的情緒為異常，情緒感受為主觀的狀況下，數據進行統整分析，將得到的結果，去與大部分人的情緒感受為正常狀態進行比較，可增加系統判定情緒異常範圍的準確性。

### 三、 研究流程與架構

本研究分為三個階段來進行敘述，如圖 2 所示，依序為緒論、相關文獻探討、體驗內容規劃分析。

- (一) 緒論：主要內容為探討研究背景與動機、確認研究目的，簡明研究範圍以及此研究方法的限制與如何解決的方法，並大致說明本研究的流程與架構。
- (二) 相關文獻探討：本研究包含人工智慧結合語音情緒識別給予使用者的體驗影響，以情緒識別與心理狀態的反應為首要主題，探討相關的文獻內容，參考或修改或其優化研究方法與系統設計，並敘述如何將其應用在本研究當中。
- (三) 體驗內容規劃分析：本章節將介紹聊天體驗內容的具體規劃以及相關操作方法，規劃出聊天體驗的流程架構，並說明為何許選擇此方式的原因，以及描述具體操作及體驗內容的需求進行規畫上的內容分析討論。

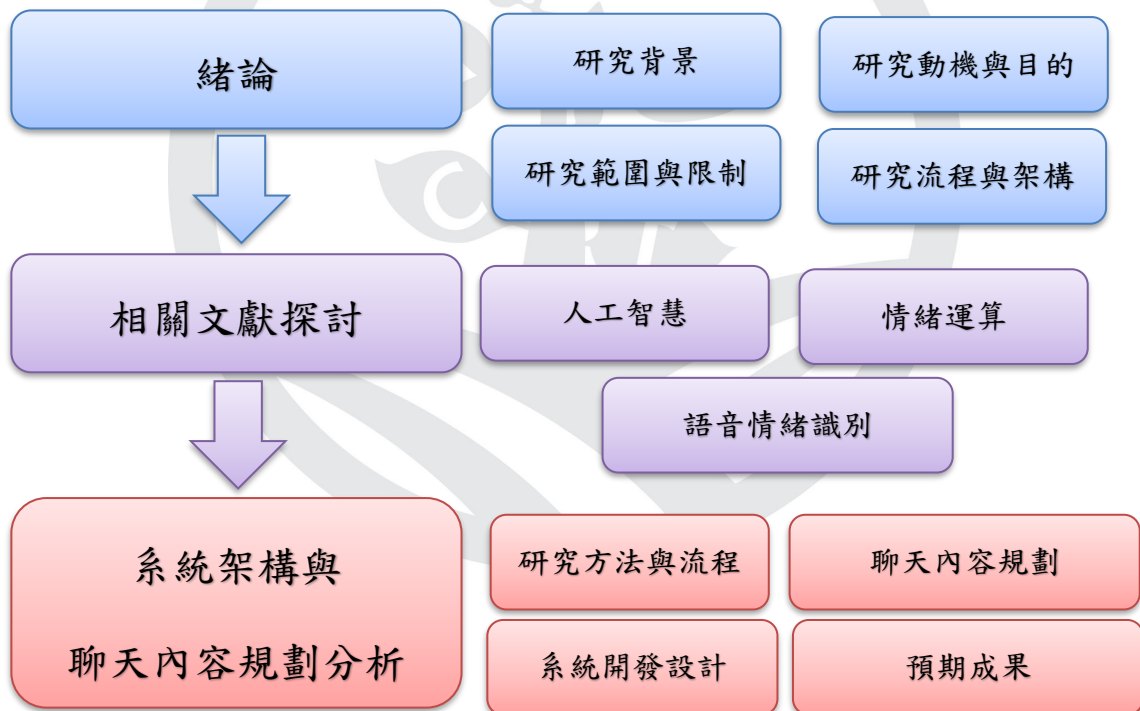


圖 2. 研究流程與架構

## 貳、文獻探討

本研究是以人工智慧與語音辨識以及情緒運算為主軸進行，著重在使用者聊天體驗完畢後分析的情緒指標與給予使用者建議。於文獻探討上，參考不同面向的論文與議題，在文獻探討中分為三個部分敘述：人工智慧、語音識別、情緒運算，其中以語音情緒識別為本研究重點，後述的議題，也將以討論人工智慧結合情緒運算為重點進行敘述。

人工智慧長期以來一直在研究智能代理即使這個領域的研究開始興起1990年代中期，(Adang Suwandi Ahmad and Arwin Datumaya Wahyudi Sumari, 2017, July)提出在現實生活中主體被定義為對情況造成重大影響，為了給這個情況一個解釋，代理必須具備能力，這裡的“能力”環境是管理任務執行時間的能力執行，知道要移動到哪裡，知道如何完成任務，知道所執行任務的成功程度，並且完成任務的後果。

人工智慧的發展與結合了語音辨識，開始了解人類的語言，讓生活越來越便利，可以幫助人類聲控家電、能當一個解說員、能陪伴人類聊天，但是人類不可取代的是有溫度的行為與語氣，本研究認為如果結合情緒辨識讓人工智慧理解，能應用到更多的成面，醫療資訊的提供與有溫度的關心、更具有人性化的服務、可透過分析得知每個人的喜好的東西。

## 四、人工智慧

「世界上已經出現一種機器，可以自行思想、自行發展、自行創造，而且這種能力即將快速發展，在可預見的未來，它們處理問題的範圍將擴大到人類心智所能處理的範圍。」此時人工智慧逐漸嶄露頭角，並漸漸應用於各行各業。

隨著資訊科技爆炸性的成長，高效能的資訊發展與通訊網路，提供了人工智慧廣大的發展空間，目前人工智慧的發展已將人類科技引領到另一高峰，人工智慧技術的應用也有進入人類生活的趨勢，且有相當不錯的成果對我們的資訊社會帶來不小的沖擊與影響。

### (一)人工智慧之定義

在探討人工智慧之前，讓我們先了解何謂「智慧」？電腦如果具有如人類一般能對事物認識、了解、感知、學習、選擇、判斷之能力，即可稱為人工智慧，而在人工智慧有分為二種如下：

1. 監督式學習:監督式學習顧名思義就是給予人工智慧樣本或是一組規則，讓它從規則裡面去學習，透過給予的規則來預測或分類，通常可以跟給予的樣本做比對驗證就可以知道誤差值大不大，所以監督式學習很取決於給予的資料多寡越多的資料相對就會越準確。
2. 無監督式學習:無監督式學習是給予人工智慧特徵來做判斷，透過給予的特徵由人工智慧分析關聯，再將得到的關聯性去做辨識，等於是自己慢慢地摸索關聯性相對的比較沒有限定住的框架，往往無監督式學習在實驗前半段都會比監督式學習成長來的慢，後半段無監督式學習會突破監督式學習的成長曲線。

## 五、 機器學習

機器學習是一種弱人工智慧(narrow AI)，它可以讓電腦尋找資料中得到複雜的函數(或樣本)來學習並創造演算法(或一組規則)，並利用它來做預測。同時並藉由驗證資料比對計算分類結果，來判定模型是否適合用來預測或分類。

目前多層次的人工神經網絡模型，如:圖 3，主要包含輸入層 (input layer)、隱層 (hidden layer) 與輸出層 (output layer) 分以下三個進行說明：

- 輸入層 (Input layer)：主要是負責接受大量的向量的訊息。
- 輸出層 (Output layer)：訊息在神經元中進行的輸出向量結果。
- 隱藏層 (Hidden layer)，簡稱「隱層」，是介於輸入層與輸出層之間的一個媒介，是可以連結神經元和鏈接並組成各個層面，因此隱層可以擁有相當多層，但是習慣上還是只會用於一層。

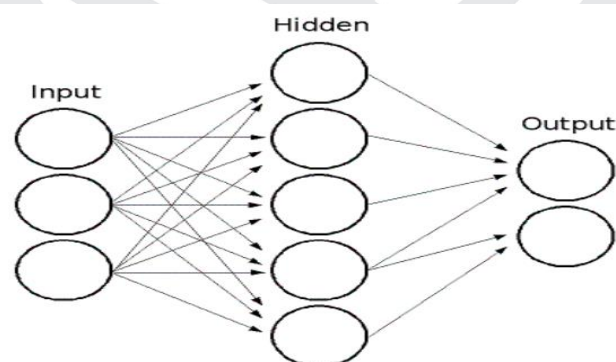


圖 3：類神經網路結構圖 (Xingbao Gao and Li-Zhi Liao, 2010,JUNE)

## (一) 隨機森林

在機器學習中，隨機森林（英語:Random Forest, RF）是一個包含多個決策樹的分類器，其輸出的類別是由個別樹輸出的類別的眾數而定，而隨機森林是屬於整合學習的一種組合分類演算法，整合學習的核心思想就是將若干個弱（基）分類器組合起來，得到一個分類效能顯著優越的強分類器。如果各弱分類器之前沒有強依賴關係、可並行生成，就可以使用隨機森林演算法。

隨機森林利用自主抽樣法 (Xuefeng Zhang, Wenyong Wang, Xuanxuan Zheng, Yongxue Ma, Yao Wei, Ming Li and Yu Zhang, 2019)從原資料集中有放回地抽取多個樣本，對抽取的樣本先用弱分類器—決策樹進行訓練，然後把這些決策樹組合在一起，通過投票得出最終的分類或預測結果。

## (二) 長短期記憶法

長短期記憶（英語:Long Short-Term Memory, LSTM）是一種時間遞歸神經網絡（RNN），論文首次發表於 1997 年。由於獨特的設計結構，LSTM 適合於處理和預測時間序列中間隔和延遲非常長的重要事件。

LSTM 是一種含有 LSTM 區塊（blocks）或其他的一種類神經網路，(S.Dhananjay Kumar and Subha DP, 2019)文獻中或其他資料中 LSTM 區塊可能被描述成智慧型網路單元，因為它可以記憶不定時間長度的數值，區塊中有一個 gate 能夠決定 input 是否重要到能被記住及能不能被輸出 output。

LSTM 的經典模型，如:圖 4，四個 S 函數單元，最左邊函數依情況可能成為區塊的 input，右邊三個會經過 gate 決定 input 是否能傳入區塊，左邊第二個為 input gate，如果這裡產出近似於零，將把這裡的值擋住，不會進到下一層。左邊第三個是 forget gate，當這產生值近似於零，將把區塊裡記住的值忘掉。第四個也就是最右邊的 input 為 output gate，他可以決定在區塊記憶中的 input 是否能輸出。

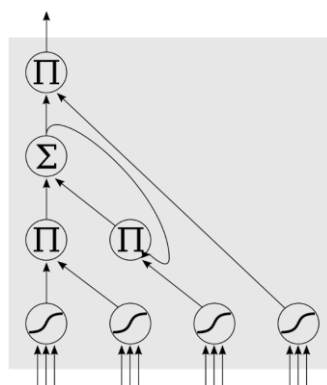


圖 4：LSTM 的經典模型

(圖片參考來源：[https://commons.wikimedia.org/wiki/File:Lstm\\_block.svg](https://commons.wikimedia.org/wiki/File:Lstm_block.svg))

### (三) k 近鄰分類

KNN 演算法又稱 k 近鄰分類(英語:k-nearest neighbor classification)演算法，它是根據不同特徵值之間的距離來進行分類的一種簡單的機器學習方法，它是一種簡單但是懶惰的演算法，它的訓練資料都是有標籤的資料，即訓練的資料都有自己的類別。KNN 演算法主要應用領域是對未知事物進行分類，即判斷未知事物屬於哪一類，判斷思想是基於歐幾里得定理，判斷未知事物的特徵和哪一類已知事物的特徵最接近。它也可以用於迴歸通過找出一個樣本的 k 個最近鄰居，將這些鄰居的屬性的平均值賦給該樣本，就可以得到該樣本的屬性。

在利用 KNN 演算法判斷類別時 K 的取值很重要。KNN 演算法主要依據鄰近的 k 個樣本來進行類別的判斷。然後依據 k 個樣本中出現次數最多的類別作為未知樣本的類別。這也就是我們常常說到的”物以類聚，人以群分”、”近朱者赤，近墨者黑”。

### (四) 極限梯度提升

XGboost 又稱為極限梯度提升(英語:eXtreme Gradient Boosting)，該算法思想就是不斷地添加樹，不斷地進行特徵分裂來生長一棵樹，每次添加一個樹，其實是學習一個新函數，去擬合上次預測的殘差。當我們訓練完成得到 k 棵樹，我們要預測一個樣本的分數，其實就是根據這個樣本的特徵，在每棵樹中會落到對應的一個葉子節點，每個葉子節點就對應一個分數，最後只需要將每棵樹對應的分數加起來就是該樣本的預測值。

(Tianqi Chen and Carlos Guestrin, 2016)等人說目標函數由兩部分構成，第一部分用來衡量預測分數和真實分數的差距，另一部分則是正則化項。正則化項同樣包含兩部分，T 表示葉子結點的個數，w 表示葉子節點的分數。 $\gamma$  可以控制葉子結點的個數， $\lambda$  可以控制葉子節點的分數不會過大，防止過擬合。

XGBoost 廣泛用於數據科學競賽和工業界，因為它有以下六項優點：

1. 使用許多策略去防止過擬合，如：正則化項、Shrinkage and Column Subsampling 等。
2. 目標函數優化利用了損失函數關於待求函數的二階導數。
3. 支持並行化，這是 XGBoost 的閃光點，雖然樹與樹之間是串行關係，但是同層級節點可並行。具體的對於某個節點，節點內選擇最佳分裂點，候選分裂點計算增益用多線程並行，訓練速度快。
4. 添加了對稀疏數據的處理。
5. 交叉驗證、early stop，當預測結果已經很好的時候可以提前停止建樹，加快訓練速度。
6. 支持設置樣本權重，該權重體現在一階導數 g 和二階導數 h，通過調整權重可以去更加關注一些樣本。

## 六、 語音辨識

語音辨識（英語:Speech Recognition）技術，也被稱為自動語音辨識（英語：Automatic Speech Recognition, ASR）、電腦語音識別（英語：Computer Speech Recognition）或是語音轉文字識別（英語：Speech To Text, STT），其目標是以電腦自動將人類的語音內容轉換為相應的文字，與說話人辨識及說話人確認不同，後者嘗試辨識或確認發出語音的說話人而非其中所包含的詞彙內容。

語音識別技術的應用包括 (Matthew A. Siegler and Richard M. Stem, 1995) 等人提出語音撥號、語音導航、室內設備控制、語音文檔檢索、簡單的聽寫數據錄入等，語音識別技術與其他自然語言處理技術如機器翻譯及語音合成技術相結合，可以構建出更加複雜的應用，例如語音到語音的翻譯。語音識別技術所涉及的領域包括：信號處理、模式識別、機率論和資訊理論、發聲機理和聽覺機理、人工智慧、演講文字紀錄等等。

### (一) 自然語言處理技術

自然語言處理(英語:Natural Language Processing, NLP)是人工智慧和語言學領域的分支學科，此領域探討如何處理及運用自然語言；自然語言處理包括多方面和步驟，基本有認知、理解、生成等部分，它的主要覆蓋的內容是：以一種智能與高效的方式，對文本數據進行系統化分析、理解與信息提取的過程，(Partha Mukherjee, Soumen Santra, Subhajit Bhowmick, Ananya Paul, Pubali Chatterjee and Arpan Deyasi, 2018) 等人提到自然語言處理是一種廣泛使用的技術，系統可以通過該技術來理解文字或語音的指令，通過使用自然語言處理 (NLP) 分析和處理文字，然後使用數字信號處理 (DSP) 技術將文字處理為語音合成器，開發了將文字轉換為語音的文字合成器，如:圖 5。

NLP 中理解語義分析的方法：

- 分佈式：它利用機器學習和深度學習的大規模統計策略。
- 框架式：句法不同，但語義相同的句子在數據結構（幀）中被表示為程式化情景。
- 理論式：這種方法基於的思路是句子指代真正的詞，結合句子的部分內容可表達全部含義。

- 交互式（學習）：它涉及到語用方法，在交互式學習環境中用戶教電腦一步一步學習語言。

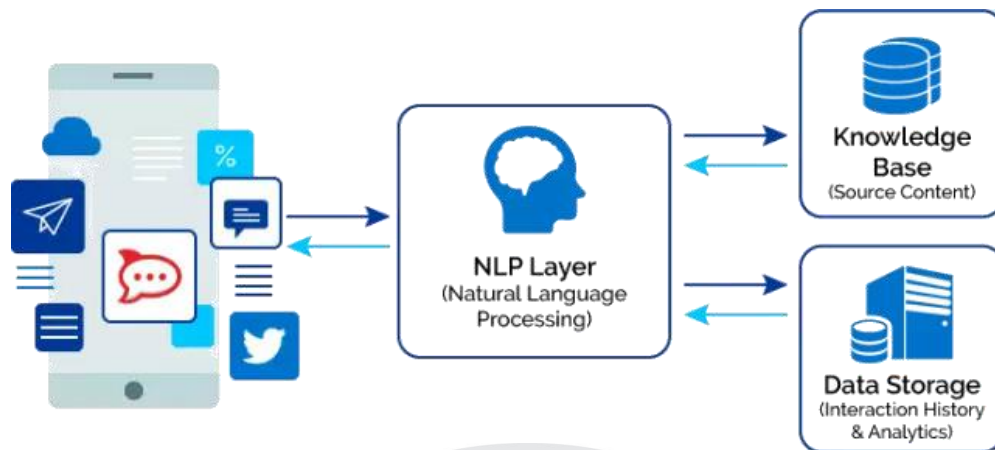


圖 5：自然語言處理結構圖(來源：機器之心)(黃小天、李亞洲, 2018)

## (二) Jieba

Jieba 是個 Python Based 的開源中文斷詞程式，(Fukuball, 2017)提出 Jieba 中文斷詞所使用的演算法是基於 Trie Tree 結構去生成句子中，中文字所有可能成詞的情況，然後使用動態規劃（Dynamic programming）算法來找出最大機率的路徑，這個路徑就是基於詞頻的最大斷詞結果。對於字典詞庫中不存在的詞，則使用了 HMM 模型（Hidden Markov Model 及 Viterbi）算法來辨識出來，基本上這樣就可以完成具有斷詞功能的程式了。

## 七、 聊天機器人

關於人工智慧聊天機器人是計算機程序，它們使用自然語言處理技術通過人與機器之間的聽覺或資料信息進行互動，其中大多數技術都基於模式匹配。(Wen Zhang, Heng Wang, Kaijun Ren and Junqiang Song, 2016, July)提到聊天機器人會識別人類的音頻並將其翻譯為文字，然後將文字與預先存儲在數據庫中的句子進行匹配，在聊天機器人的應用非常的廣泛，在醫療上面也有相關的貢獻，世界衛生組織（WHO）報告說第二大死因是癌症，癌症患者們擁有複雜的許多情緒，他們周圍的人也發現很難應付，他們需要不斷的支持來討論他們的問題並向他們提供事實的訊息。

(Belfin RV, Ashly Ann Mathew, Shobana AJ, Blessy Babu and Megha Manilal, 2019)等人介紹了一種潛在的解決方法，可為他們提供他們正在尋找的聊天機器人，提供的聊天機器人是僅針對癌症患者設計的癌症聊天機器人，人們可以詢問有關癌症症狀、治療、生存等的所有信息，該機器人通過從各種癌症論壇收集的

信息進行培訓，這些論壇提供了有關癌症的廣泛信息，也運用情緒分析識別使用者的情緒，以便該機器人可以給出類似人的行為並使他們感到舒適。

## 八、情緒的分類

情緒是指人有喜、怒、哀、樂、懼等心理體驗，這種體驗是人對客觀事物的態度的一種反映，情緒具有肯定和否定的性質，能滿足人的需要的事物會引起人的肯定性質的體驗，如快樂、滿意等；不能滿足人需要的事物會引起人的否定性質的體驗，如憤怒、憎恨、哀怨等；與需要無關的事物，會使人產生無所謂的情緒和情感。積極的情緒可以提高人的活動能力，而消極的情緒則會降低人的活動能力。

從不同的角度可以將情緒分為以下六類，第一類是原始的基本情緒，往往具有高度的緊張性，如快樂、憤怒、恐懼、悲哀。快樂是盼望的目的達到後，緊張被解除時的情緒體驗；憤怒是願望目的不能達到、一再受阻、遭受挫折後積累起來的緊張的情緒體驗；恐懼是在準備不足、不能處理和應付危險可怕事件時產生的情緒體驗；悲哀是與所追求、熱愛的事物的喪失，所盼望的事物的幻滅有關的情緒體驗。

第二類是與感覺刺激有關的情緒，如疼痛、厭惡、輕快等。第三類是與自我評價有關的情緒，主要取決於一個人對於自己的行為與各種標準的關係的知覺，如成功感與失敗感、驕傲與羞恥、內疚與悔恨等。第四類是與別人有關的情緒，常常會凝結成為持久的情緒傾向與態度，主要是愛與恨。第五類是與欣賞有關的情緒，如驚奇、敬畏、美感和幽默。第六類是根據所處狀態來劃分的情緒，如心境、激情和應激狀態等。

## 九、語音情緒識別系統

一段連續之語音信號，透過類比/數位轉換（A/D Converter）之後，轉換成一個離散時間序列，再進入特徵分析（Feature Analysis）的階段，在此階段中透過預處理和特徵提取，提取出語音特徵參數向量（Feature Vectors），作為語音訓練階段（Training 階段），以及語音識別階段（Recognition Phase）的輸入向量。（張柏雄, 2002)提到在訓練階段，特徵向量會先作向量量化，編碼成固定長度之向量，通過樣本訓練（Pattern Training）得到識別階段中所需的參考樣本（Reference Patterns）；在識別階段中，將測試樣本與參考樣本分別批准樣本比對演算法（模式比較法），將比對結果決定性法則（Decision Rule）處理後，得到最後的識別結果。

### (一) 資料平衡化

在機器學習以及大數據的領域中，說分類問題一直以來都是一個重要的研究方向，在資料分佈大致平衡的假設狀況為降低之下，目前許多模型都能獲得良好的結果(史柏元, 2016)，許多的分類問題都是不平衡的，當資料集產生嚴重的類別不平衡問題時，會導致分類器在訓練模型的過程中受到極大影響，而不平衡學習的困難之處主要在於不同類別樣本數量的懸殊差異。

### (二) 基準特徵集

在 INTERSPEECH 2009 Emotion Challenge 中，使用 openSMILE 2.0 工具擷取官方設定的標準特徵，對於每一個音檔擷取 16 個低階參數(Low-level descriptors, LLDs)，包含每一音框的均方根能量(Root Mean Square(RMS) frame energy)、過零率(Zero-Crossing-Rate, ZCR)、12 個梅爾倒頻譜係數(MFCCs)、諧音噪音比(Harmonics-to-Noise, HNR)、音高頻率(Pitch Frequency)，加上每個低階參數的一階係數差(Deltas)(陳嘉穎, 2016)，最後在低階參數上取其 12 個泛函，包含平均值(Mean)、標準差(Standard Deviation)、峰值(Kurtosis)、偏斜(Skewness)、最小值(Maximum Value)、最大值(Minimum Value)、相對位置(Relative Position)、和(Range)以及兩個線性回歸係數(Linear Regression Coef-ficients)並取其均方誤差(Mean Square Error)。

## 參、研究實施與設計

依據文獻探討設計本研究之系統與實驗內容規劃，以下由第一小節開始說明研究流程與研究方法和系統架構與流程，以及介紹體驗具體事項，在依據此流程進行設計與開發。

### 一、研究方法

研究主題分成多個面相，智慧語音情緒識別之自我檢測系統經由多個部分完成，而每個部分代表不同研究階段，故使用漸層模式進行研究，將研究方法分成四個階段：

- (一) 系統分析與前置作業準備階段
- (二) 系統研發設計階段
- (三) 功能開發與測試階段
- (四) 實驗測試
- (五) 評估準確度

#### 1. 系統分析與收集相關的文獻與資料

本研究會收集所需的相關文獻，從國外論文和著作與校內外圖書館館藏中，尋找有關於本研究的相關資料，整理出對此有益的資訊，關於系統分析的部分，採用人工智慧語音情緒識別的技術，並與心理諮商領域的專家學者設計問題與回答在進行情緒運算為本研究主要核心技術。

#### 2. 系統研發設計階段

本系統由三個主要架構所組成—語音情緒識別、機器學習、情緒運算，結合三種技術針對目標族群為大眾都可對自我情緒做一個檢測的系統，首先設計問題類型與機器人最佳的回答方式，根據使用者聊天中所回答的內容、語氣、聲道、頻率，利用機器學習的正規化處理與隨機森林（英語:Random Forest, RF）、長短期記憶法（英語:Long Short-Term Memory, LSTM）與極限梯度提升（英語:eXtreme Gradient Boosting）等演算法，去決策最佳化的回答為何，最後透過情緒運算去計算情緒指標為多少，並判斷是否有異常再給予使用者適當的建議。

#### 3. 功能開發與測試階段

本系統將透過 Google AI 與 Google Cloud Speech-to-Text 以及 AI Platform Notebooks，透過這三種開發環境進行研究，將資料庫的數據匯入至 AI Platform Notebooks 進行人工智慧模型的訓練，嵌入 Google Speech API 的服務能增加語音辨識的功能性與實用性，在測試階段會讓使用者實際體驗與機器人聊天，再根據聊天的狀況收集數據做分析。

#### 4. 實驗測試

此階段為系統模型的測試工作，此測試的目的為根據收集到的數據資料分析，聊天體驗過後數據是否出現誤差，經過分析後如果誤差值過大或者是精準度太低，會再進行參數的調整與修正，並了解本系統是否滿足使用者的需求以及達到預期期望之效果。

#### 5. 評估準確度

評估準確性的步驟，會讓使用者做使用性測驗，將得到的使用性量表數據統整分析，來評估情緒指標是否準確。因情緒為主觀感受，所以用使用性量表來評估準確度，根據得到的準確度再優化系統，最後進行第二次使用性測驗，期望將準確度提升到一定水準。

## 二、 研究工具

系統的開發環境採用 Google AI 與 Google Cloud Speech-to-Text 以及 AI Platform Notebooks，程式語法則是採用 Python 撰寫，依照研究方向設計出所需的機器學習系統作為基底，並將相關技術與概念結合，站在使用者的角度會希望得知自己情緒的指標以及把想說的話告訴機器人，站在開發者的角度，本系統將跟使用者進行簡單的聊天對談，並將其聊天對談的資料做為參考數據傳送至系統作分析，並給予使用者的自我情緒指標與最好的建議。

### (一) 開發可能遇到的問題

開發時可能遇到的問題，語音資料庫的建置與收集，資料庫的語言必需是中文的才有辦法做出比對，根據語音資料庫的資料跟使用者的語音頻率與回答問題的內容進行比對與分析，所以在語音資料庫的建置與收集方面是需要花費很多的時間，為了讓情緒識別具有一定的信度與效度，需要心理諮商專業領域的學者共同合作設計問題與回答的最佳方式，再開發後也需要找尋受測者進行測試，來佐證一下準確度，總結遇到的困難分成三個重點：

1. 中文語音資料庫的建置與收集
2. 需要心理諮商專業領域的學者共同合作
3. 需要找尋一定的受測者進行測試

### 三、系統流程

本研究開始會收集相關的文獻與資料，去得知目前專家學者研究後碰到的問題或有什麼優缺點進一步的分析後，會找心理諮商領域的學者進行設計問題與設計機器人回答的最佳方式內容，在設計問題與回答內容時可以建立系統的架構，當擬定好問題與回答也有架構時就會進行系統的開發，根據系統架構擬定開發完後會得到系統模型，重複經過各項功能測試與細修最後將系統模型嵌入機器人中，以下是簡易流程，如：圖 6。

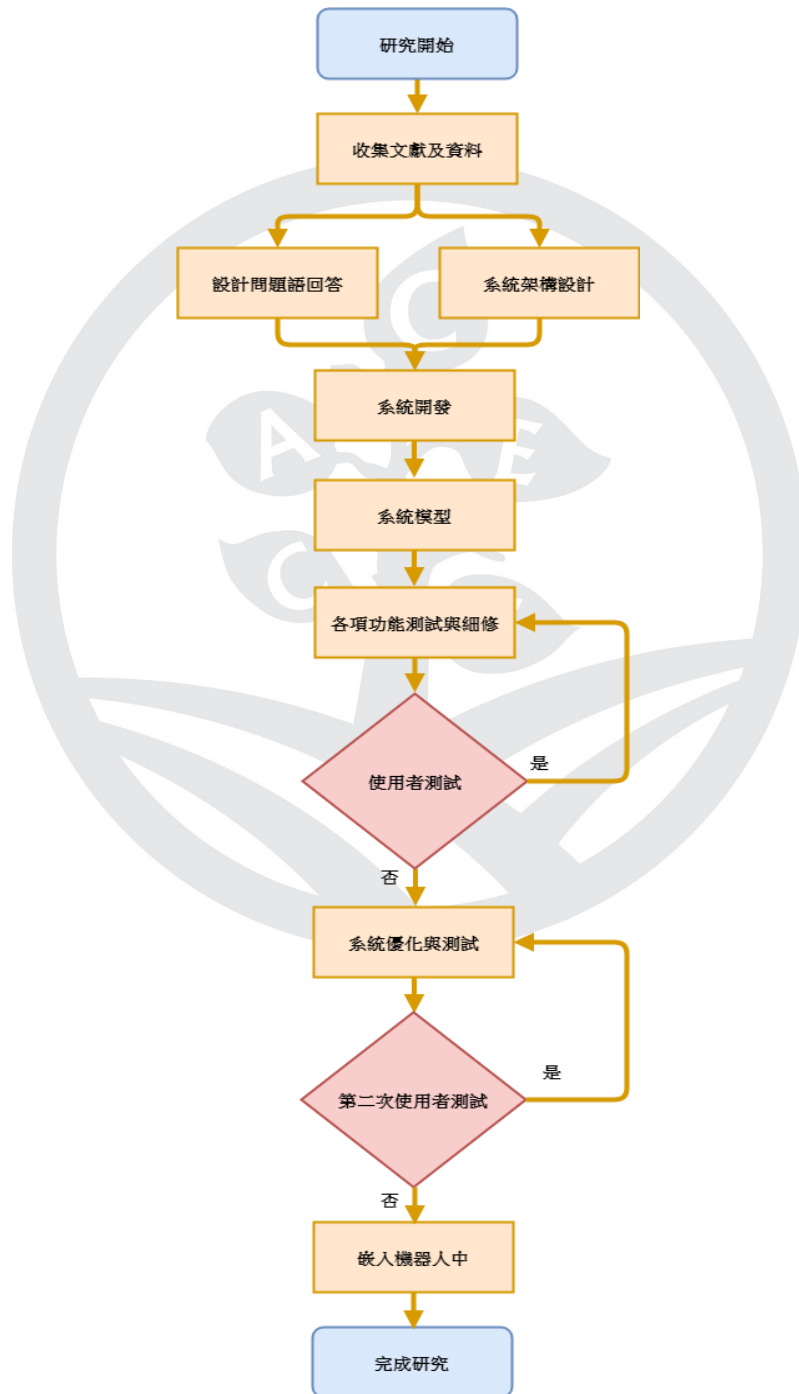


圖 6：簡易研究流程圖

#### 四、系統架構

本系統架構由 Google AI 與 Google Cloud Speech-to-Text 以及 AI Platform Notebooks 三種開發的環境去執行，撰寫的程式的語法是 Python，主要透過隨機森林(RF)、長短期記憶法(LSTM)、自然語言處理(NLP)等演算法，進行主要的人工智慧運算出情緒，並給予適當的回應與情緒指標與綜合建議，以下是初步系統架構圖，如:圖 7。

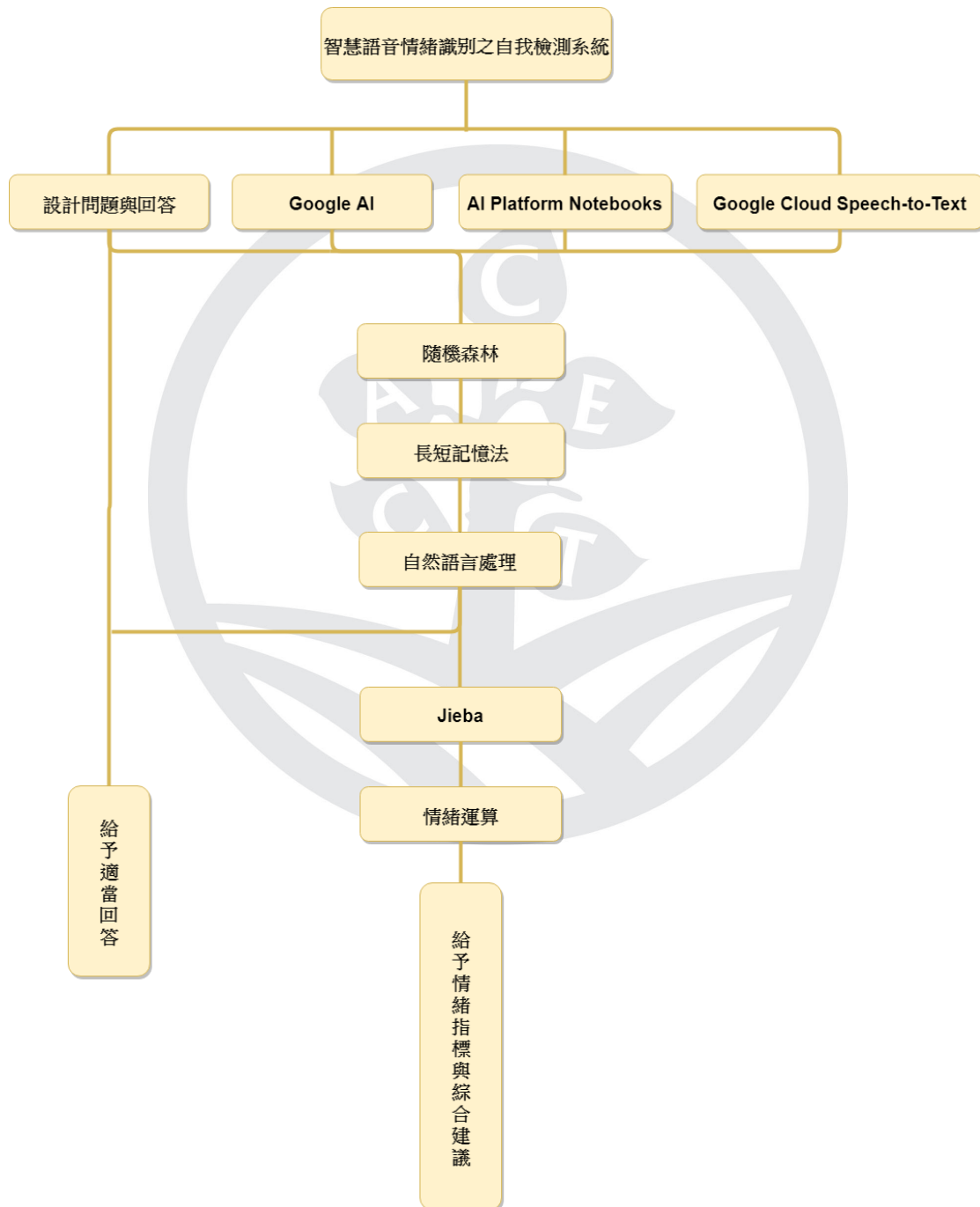


圖 7：系統架構圖

## 肆、預期結果與未來展望

本研究的預期成效想要增加對自我的壓力與精神狀況的認知，以及減少心理諮商師與社工師人力不足的問題，由於自我能做檢測並可獲得評估狀況，以及後續該如何進行下一步的處理方式，可達到有效的使用社會資源避免浪費與更了解自己，在此透過與機器人聊天後有些人能使心情比較愉快，有些人需要傾訴對象能訴說事情時，那機器人也可擔任其角色，當機器人理解情緒時，可讓整個聊天過程有了溫度，能讓使用者有更舒適以及更與像真人聊天的體驗，總結成效要達到透過自我檢測系統了解自我的情緒狀況以及後續的處理方式。



## 參考文獻

### 一、中文部分

1. 黃小天、李亞洲 (2018)。【機器之心】人工智能與自然語言處理概述：AI 三大階段、NLP 關鍵應用領。檢自 xenonstack : <https://kknews.cc/zh-tw/tech/8m44yj4.html>
2. Fukuball (2017)。如何使用 jieba 結巴中文分詞程式。檢自 jieba : <https://coderwall.com/p/38wtgw/jieba>
3. 張柏雄(2002)。中文語音情續之自動辨識。未出版之碩士論文，國立成功大學工程科學系研究所，臺南市。
4. 史柏元(2016)。應用創新數據偏斜強健性類神經網絡於語音情緒識別。未出版之碩士論文，國立中山大學計算機科學與工程系研究所，高雄市。
5. 陳嘉穎(2016)。應用因素分析與識別向量於語音情緒辨識。未出版之碩士論文，國立中山大學計算機科學與工程系研究所，高雄市。

### 二、英文部分

1. M. Mori (1970). The uncanny valley. *Energy*, vol. 7, no. 4, pp. 33–35.
2. Adang Suwandi Ahmad and Arwin Datumaya Wahyudi Sumari (2017,July). Cognitive Artificial Intelligence: Brain-Inspired Intelligent Computation in Artificial Intelligence. *Computing Conference* (pp. 18-20).
3. Belfin RV, Ashly Ann Mathew, Shobana AJ, Blessy Babu and Megha Manilal (2019). A Graph Based Chatbot for Cancer Patients. *International Conference on Advanced Computing & Communication Systems*.
4. Matthew A. Siegler and Richard M. Stem. (1995). ON THE EFFECTS OF SPEECH RATE IN LARGE VOCABULARY SPEECH RECOGNITION SYSTEMS. *Department of Electrical and Computer Engineering School of Computer Science Carnegie Mellon University Pittsburgh Pennsylvania*.
5. Partha Mukherjee, Soumen Santra, Subhajit Bhowmick, Ananya Paul, Pubali Chatterjee and

Arpan Deyasi. (2018). Development of GUI for Text-to-Speech Recognition using Natural Language Processing. *2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology*.

6. S.Dhananjay Kumar and Subha DP. (2019). PREDICTION OF DEPRESSION FROM EEG SIGNAL USING LONG SHORT TERM MEMORY(LSTM). *Proceedings of the Third International Conference on Trends in Electronics and Informatics*.
7. Tianqi Chen and Carlos Guestrin. (2016). XGBoost: A Scalable Tree Boosting System.
8. Wen Zhang, Heng Wang, Kaijun Ren and Junqiang Song. (2016,July). Chinese Sentence Based Lexical Similarity Measure for Artificial Intelligence Chatbot. *ECAI 2016 - International Conference – 8th Edition Electronics, Computers and Artificial Intelligence 30 June -02*.
9. Xingbao Gao and Li-Zhi Liao. (2010,JUNE). A New One-Layer Neural Network for Linear and Quadratic Programming. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, (pp. VOL. 21, NO. 6).
10. Xuefeng Zhang, Wenying Wang, Xuanxuan Zheng, Yongxue Ma, Yao Wei, Ming Li and Yu Zhang. (2019). *A Clutter Suppression Method Based on SOM-SMOTE Random Forest*. Nanjing Institute of Electronic Technology.