

# 基於文本分析之認知偏誤類型辨識技術

<sup>1</sup>阮鈺鈞 <sup>2</sup>李祐誠 <sup>2</sup>黃靖婷 <sup>1</sup>陳欣進 <sup>1</sup>鄧閔鴻 <sup>2</sup>林維暘

<sup>1</sup>國立中正大學心理學系

<sup>2</sup>國立中正大學資訊工程學系

wl02659021@gmail.com, syw2352@gmail.com

## 摘要

本篇論文主要為了認知治療的分析，利用類神經網路將認知偏誤分為六大類：「極端化」、「災難化」、「讀心術」、「應該必須語句」、「其他負向類型」、「正向語句」。本實驗語料為本校心理系提供，是依據不同認知偏誤類型所模擬生成可能會出現的句子，有七十六筆訓練資料及二十八筆測試資料。

研究方法為建立一個三層的類神經網路模型，包含九個節點的輸入層、八個節點的隱藏層和六個節點的輸出層，首先以人工的方式將語句進行編碼後作為輸入，運用該方法將語句歸類為這六類的其中一類。實驗結果的準確率達78.5%，主要是極端化這一類被分錯，但因實驗語料較少，還有待探討。

關鍵詞：認知行為治療(CBT)、認知偏誤、類神經網路、分類、文本分析

## 1. 前言

在過去的這幾百年內，人類文明的大幅度進步，造就了很多的便利。在20世紀的時候，由於網路的發明，使得人類之間的溝通變得不再具有空間限制，就算你和另一個人相隔很遠，也可以進行通訊。而時至今日，網路已經成為我們日常生活不可或缺的事物了。

隨著網際網路通訊日漸的發展，一些潛藏的危機也隨之浮現。人因為網路通訊的便捷，使得我們彼此之間的互動變少了，以前總是一起聊天說話，現在取而代之的是利用電腦螢幕或者是手機來使用通訊軟體，與他人進行溝通，而這也慢慢地衍生出一些問題，像是憂鬱症以及其他的心理疾病。這些疾病並不如我們平常的感冒，他們是由於病毒入侵我們的身體所導致，他們是心理疾病，也就是因為一些潛在的因素，而這因素可能是來自於家庭方面，也可能是來自於社會環境的影響。而心理疾病它並不像感冒，我們可以憑藉著藥物的攝取，來使得我們迅速地恢復健康，而且簡單根除，它需要的是關懷，是需要心靈上得到一些幫助，才能得以治療，而我們所需要的就是這些心靈的醫生——心理治療師。

心理治療師對於這些精神疾病患者，會施予認知行為治療（Cognitive Behavioral Therapy，簡稱CBT）。CBT是一種心理治療的取向、一種談話治療，以目標導向與系統化的程序，解決喪失功能的情緒、行為與認知問題，這意味著它用來解決和特

定精神紊亂相關的問題，治療師的角色是幫助客戶發現和練習有效的機制來達到確定的目標並且降低紊亂的症狀。可是對於患者來說，在透過這樣的治療過程中，雖然有一定的效果，但是帶來的副作用也相對來說是很大的，最重要的是，對於患者來說，要曝光於生活中也是一個很大的挑戰。

每個人誕生在這世界上，都是獨一無二的，就算是雙胞胎，也會在個性上有些不同，因為如此，每個人所擁有的心理疾病也會不盡相同，而認知治療它會針對不同面向問題的人們，施以不同的治療方案。心理治療師在與個案(即接受治療者)進行對話時，通常會產生一個文本，而文本的內容是記錄整個對話的過程，我們可以從這文本中找到一些關鍵字(key word)，而這些關鍵字就可以讓我們判斷個案的類型，進而對他們施以相對應的療程。

心理學家所討論的「認知偏差」(cognitive bias)，是指一些非常普遍、阻礙確當思考的認知模式。這些認知模式，通常是基於深層的心理現象，並不容易改變和避免。認知偏差有很多種，包括決策上的偏差，有統計和或然率的一些常見的誤解，以及錯誤思考模式等等[1]。而有許多學派提出不同的認知偏誤分類，例如貝克的認知分類分為七種：(1)二分化的思考(2)選擇性摘錄(3)獨斷的推論(4)災難化(5)過度類化(6)標籤化(7)誇大或貶低。

就現有的文本分析來看，大部份的文本分析都是建立在網路文本、部落格文本、歌詞文本等之類的文字資料上，幾乎沒有以治療相關的文本作為研究方向，我們之所以會選認知治療文本來做分析，是因為：

1. 認知治療是目前比較被廣泛使用的一種治療方式：認知治療本身是一種以解決問題為取向的心理治療方式，在於幫助患者面對這問題時重心建立所該要有的情緒，通常這個是效果比較好的一種治療方法。

2. 認知治療本身是有系統性的：認知治療本身是一套系統性的治療方法，讓醫生在做跟患者治療時可以做出相對精確性的判斷，做出適當的回應，這樣所產生的文本也會比較有分析的方向和價值。

就如同上面兩點所述，做認知治療文本分析，可以讓我們有更好的方式去判斷患者情緒，也可以比較好辨識患者的關鍵字從中找出問題，並且在分析之後交由 chatbot 學習，可以對心理治療這一塊有很大的發展。

認知治療 chatbot 的出現，會對於這個社會有很大的影響，首先，對於某些患者來說，是非常重

要的，因為那些患者害怕人群，足不出戶，對他們來說，踏出家門就是一個很大的考驗，更不用提他會去找心理治療師做治療，也因為這樣，讓這些患者的情形更加惡化，現在如果有了這個東西，患者就有機會在家裡進行認知治療，透過在手機或電腦使用 chatbot 程式，藉由直接跟 chatbot 對話也可以對患者做到同樣的效果，藉由這樣的方式，讓患者可以接觸人群，之後再更進一步的找心理治療師做心理諮商。就目前的 chatbot 來看，大部分都是以從對話中截取關鍵字去進行演算法找出最適合的答案，可是如果是要進行認知治療的話，患者目前的情緒狀態也會影響答案，而且有可能會讓患者不知道 chatbot 所要表達的意思為何，造成兩邊的資訊不對等，這樣反而會對患者造成更多影響，認知治療 chatbot 不僅僅只是單一向量的考慮，要如何全部概括，並且讓自己的回覆可以給患者完整的答案，這些問題從我們在進行文本分析就會面臨到這些狀況，我們的分析方式和結果，還有精確程度，這就會對往後發展認知治療 chatbot 造成很大的影響，也是我們現在這專題的最大困難和挑戰。

## 2. 文獻探討

國外資料中最早出現的聊天機器人為 ELIZA [2]，是在西元1966年由麻省理工學院(MIT)開發。當時主要的問題是找出句子的關鍵字並如何採取回應，而當時的內容也只局限於人工的回覆腳本。聊天機器人之後的重大發展其中一項就是語料庫，也就是語言學中大量的文本經過整理，具有既定格式。文獻中利用自然語言處理(NLP)試圖進行兩項研究，分別是將聊天機器人產生出兩種比較少樣本的語言，以及將大量語言資料利用機器學習的方法進行處理，最後產生人工智慧標記語言(AIML) [3]。

近年來的多項文獻更是指出聊天機器人對於認知治療的重要性。其中包括聊天機器人是否能夠取代心理醫師[4]。實驗者在面對非人類的時候在負面狀態時願意表達自己的負面情緒的人數，相較於面對真人是兩倍[5]。另外也有文獻指出在面對機器人，使用者更感到安心並願意說出自己的心情[6]。Woebot 是由史丹佛大學的心理學家首創的輔導機器人。以一篇透過 Woebot 來做的研究為例，主要透過使用者傳遞的訊息辨識，以及利用程式(app)所做的認知心理治療[7]。結果顯示對於實驗年齡在18到28之間的人具有部份的治療效果，但是還是達到商品化的階段。相關困難例如語言生硬、只能針對特定表達方式做回應、效果不夠顯著，都是未來能夠改善的發展方向。

國內也有許多基於中文文本分析之分類，文本分析主要都是針對情緒正負的分類[8]，或是對於專業領域相關的分類[9]。在文本處理上有採用 jieba 之類的斷詞系統，再根據簡中或是繁中的情緒字典賦予語句權重。[8]應用上，能夠在與情境有相關領域發揮用處，例如遊戲的角色能夠依據目前情境做出反應[10]。

國內也有較為完整的情緒類型的分類，不過分類輸入是採用部落格文本搭配表情符號[11]，或是

英文文本作為輸入[12]。至於認知偏誤類型的分類國內還沒有相關研究。

## 3. 研究方法

本研究嘗試建立一個如圖1的「認知偏誤辨識系統」，藉由輸入認知治療中的個案對話語句，判斷所屬認知偏誤類型。系統開發上則使用 python 語言。

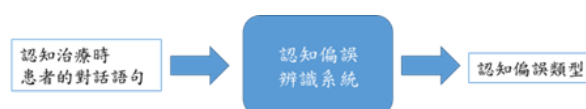


圖 1. 系統流程

認知治療之中會針對個案分析其認知偏誤(Cognitive Bias)類型，其中「極端化」、「災難化」、「讀心術」、「應該必須語句」為分類較為明確之類型，因此實驗中將明確分類這四種類型，其它認知偏誤歸類為「其它負向類型」，其餘為「正向語句」。表1為測試資料之範例。

表 1. 測試資料

語句	判斷類型
世上的人不在乎我	極端化
我完全沒有貢獻	極端化
我的未來沒有希望了	災難化
我好像每件事都做不好	災難化
我覺得他可能不喜歡我	讀心術
他認為我很爛	讀心術
我不應該失敗	應該必須
我應該要買農藥	應該必須
我覺得我很失敗	其他負向
我曾自殺	其他負向
我有很多朋友	正向
我喜歡跟自己相處	正向

「極端化」是指語句中帶有全有全無或是非黑即白的概念，例如：「他一定不信任我」。本研究中有11個訓練樣本，2個測試樣本。

「災難化」則會對某個錯誤過大的渲染，例如：「我未來可能無法成功」。本研究中有10個訓練樣本及5個測試樣本

「讀心術」認為自己知道他人的想法。例如：「他應該覺得我很糟」。本研究中共有9個訓練樣本

及4個測試樣本。

「應該必須語句」則是包含絕對性的語句。例如：「我應該要在期限之前完成的」。本研究中有10個訓練樣本和5個測試樣本。

不在前述四項的認知偏誤本研究中歸類為「其它負向類型」。例如：「我覺得我很失敗」。共有28項訓練樣本及8項測試樣本。而非負向類型則在「正向語句」，訓練樣本及測試樣本分別為8項及4項。表2為整理過後之表格。

表 2.分類類型

	訓練	測試
極端化	11	2
災難化	10	5
讀心術	9	4
應該必須語句	10	5
其他負向類型	28	8
正向語句	8	4
共計	76	28

輸入目前為人工編碼，如表3所示，將語句依照主詞、否定詞、助動詞、一般動詞、感覺動詞、補充語句、時間點、發生可能性、自殺風險分析成九項數值。以「我的未來沒有希望了」作為範例，主詞為1、否定詞為1、助動詞為0、一般動詞為0、感覺動詞為0、補充語句為1、時間點為1、發生可能性為0.5、自殺風險為0.2。

表3. 編碼方式

類別	層次	編碼
主詞	自己	1
	與我有關事物	0.5
	他者	0
否定詞	否定	1
	肯定	0
助動詞	必須	1
	應該	0.9
	可能	0.4
	不可能	0.8
一般動詞	正性	1

感覺動詞	中性	0
	負性	-1
補充語句	覺得、認為	1
	其他	0
時間點	正性	1
	中性	0
	負性	-1
發生可能性	未來	1
	現在	0
	過去	-1
自殺風險	可能性高	0.9
	機率一半	0.5
	可能性低	0.1
	意念	1
	工具	2
	未遂	3

在「認知偏誤辨識系統」中採用如圖2的類神經網路(Artificial Neural Network)架構，其中隱藏層的節點數有8個，第一層和第二層的啟動函數(Activation Function)皆使用 Sigmoid 函數，而成本函數(Cost Function)則使用 crossentropy。

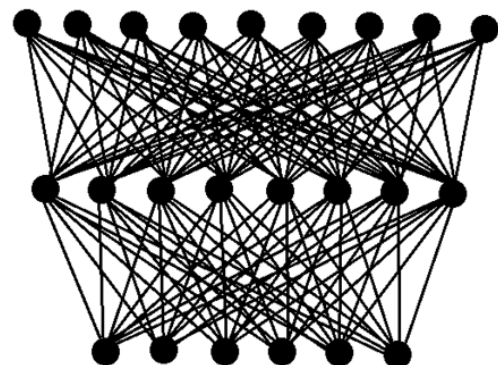


圖 2. 模型

$$S(x) = \frac{1}{1+e^{-x}} \quad (1)$$

$$C(y, \hat{y}) = -y \log \hat{y} - (1 - y) \log(1 - \hat{y}) \quad (2)$$

公式(1)為 sigmoid function，公式(2)為 cost function。

圖 3 為在不同更新率下，cost 隨著迭代次數增加而減少。

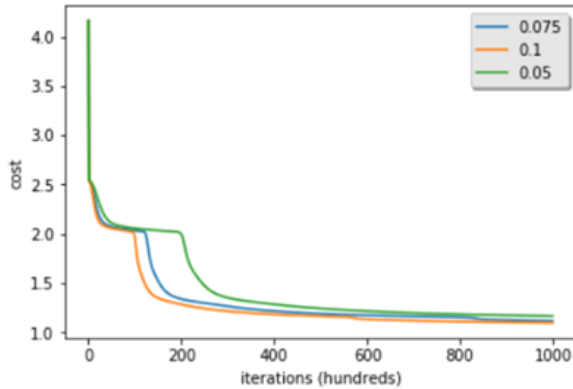


圖 3. Cost function

#### 4. 實驗結果

我們使用此模型輸出各個判斷類型的權重，再以權重最高的作為我們真正的判斷結果。實驗過程中有嘗試改變迭代次數以及更新率。只有輸出完全一致才會記入正確結果。結果最佳的訓練結果為60/76，而測試結果為22/28。

#### 5. 結論

本研究對四種較為明確的認知偏誤類型作判斷，其中所使用的資料為人工判斷的權重，未來期望能夠加入文本資料的預處理系統，讓實驗資料筆數增加以及減少工作量。

實驗結果的準確率仍有待加強，其中幾點目前已知可改進的方法。第一點是判斷錯誤主要為第一類，也就是無法判斷出極端化，推斷是特徵值在預處理階段沒有成功取出，但結果仍有被判斷在其他負向類型當中。如果預處理改變，模型層數以及節點勢必要跟著修正。第二點是還有許多偏誤類型尚未加入判斷，未來可以將之從其它偏誤類型分出。

#### 6. 參考文獻

- [1] Philosophy.hku.hk/think/chi/bias.php
- [2] Joseph Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine", *Communications of the ACM Volume 9 Issue 1*, Jan. 1966.
- [3] Shawar, Bayan Abu, and Eric Steven Atwell. "Using corpora in machine-learning chatbot

- systems." *International journal of corpus linguistics* 10.4 (2005): 489-516.
- [4] Tantam, Digby. "The machine as psychotherapist: impersonal communication with a machine." *Advances in Psychiatric treatment* 12.6 (2006): 416-426.
- [5] Miner, Adam, et al. "Conversational Agents and Mental Health: Theory-Informed Assessment of Language and Affect." *Proceedings of the Fourth International Conference on Human Agent Interaction*. ACM, 2016.
- [6] Cameron, Gillian, et al. "Towards a chatbot for digital counselling." *Proceedings of the 31st British Computer Society Human Computer Interaction Conference*. BCS Learning & Development Ltd., 2017.
- [7] Fitzpatrick, Kathleen Kara, Alison Darcy, and Molly Vierhile. "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial." *JMIR mental health* 4.2 (2017).
- [8] 楊邵為, Facebook 文章情緒分類器設計與實作, 碩士論文, 國立中正大學, 2017.
- [9] 陳禔多, 基於歌詞文本分析技術探討音樂情緒辨識之方法研究, 碩士論文, 國立政治大學, 2015.
- [10] 范國光, 人工情緒決策系統設計與其在遊戲設計的運用, 科技部研究計畫, 2016.
- [11] 黃信華, 以 FACEBOOK 塗鴉牆為文本分析情緒文字的關係, 碩士論文, 國立台南大學, 2013.
- [12] Huang, Yen-Hao, Lin-Hung Wei, and Yi-Shin Chen. "Detection of the Prodromal Phase of Bipolar Disorder from Psychological and Phonological Aspects in Social Media." *arXiv preprint arXiv:1712.09183* (2017).