

東海大學資訊管理研究所

碩士學位論文

連鎖型動物醫院之即時通訊機器人設計

The Design of Chatbot for Chain Veterinary Hospital

指導教授：陳澤雄博士

研究生：鄭建達撰

中華民國 107 年 1 月

東海大學資訊管理學系碩士學位  
考試委員審定書

資訊管理學系研究所 鄭建達 君所提之論文

連鎖型動物醫院之即時通訊機器人設計

The Design of Chatbot for Chain Veterinary Hospital

經本考試委員會審查，符合碩士資格標準。

學位考試委員會 召集人：陳澤龍 (簽章)  
委員：黃愉  
游厚龍  
鐘玉男  
陳澤雄

中華民國 107 年 01 月 17 日

## 銘謝

時光飛逝，轉眼兩年的碩士生涯就快到了結尾，在這研究所時期中學會了以前沒接觸過的工具、知識和社會上的眉角。首先要很感謝父母在背後無悔的支持，沒有他們相信跟幫忙，這研究所是肯定念不了的。再來要很感謝指導老師陳澤雄教授，在這兩年中所給的幫助，讓我在各方面都有所成長，其中包含計畫撰寫、市場分析、實習申請、生活支持和論文指導等等，在指導老師熱心的輔導下，碩士兩年時光沒有白流，每一天都相當充實，千言萬語也說不出我內心的感謝，我只能說，老師辛苦了，謝謝您。再來我想感謝的是黃映瑋在我忙碌的時候，總是在一旁協助我；在我壓力大或者低潮時，總是不厭其煩的陪著我，同時野處理我很多生活瑣事，就為了讓我有更多時間專注在論文或案子上面。最後我想感謝是研究所這兩年中陪伴我的所有老師和同學，因為有你們，讓這兩年更有色彩。

感謝口試委員在口試的時候提出所有的指教，讓我在格式上和整理論文流程有更上一層樓的理解，當我口試完看到平裝本上有老師們的細心備註，我頓時感到溫暖，讓我明白這一切都是有人在關心的。

兩年碩士生涯結束了，代表的是即將步入社會，也開始了我人生的新篇章。希望在畢業之後，能在工作上付出更多的熱忱跟努力，並且找到自己的興趣跟專業然後繼續精進，最後盡力把自己所會所學貢獻且回饋給社會國家。

論文名稱：連鎖型動物醫院之聊天機器人設計與實現

校所名稱：東海大學

畢業時間：2018年01月

研究生：鄭建達

指導教授：陳澤雄

## 摘要

近幾年，科技產業的蓬勃發展，使人類對於人工智慧的需求日益增加，而人工智慧與人類對話方面的研究，也有許多優秀的表現。然而鮮少有運用聊天機器人於動物醫院方面的研究，因此本研究結合人工智慧的技術與即時通訊的設計，希望可以得以使動物醫院服務更加完善。

目前聊天機器人為熱門的議題及應用，除了提供使用者溝通聊天的功能外，也可以有問與答的功能。而本論文以 Dcard 論壇中之文章為作訓練資料，將聊天機器人應用於動物醫院場域中，透過 Dcard 資料中所產生的對話，使用機器學習的方式及統計工具從中截取重要特徵，判斷磁性、找尋關鍵詞彙進而建立出配對的檢索式模型，讓使用者在輸入語句後便可以透過上述方法，將句子套入模型，找出相對應的回答回覆給使用者。

關鍵詞:人工智慧、聊天機器人、動物醫院、機器學習、檢索式模型

Title of Thesis : The Design of Chatbot for Chain Veterinary Hospital

Name of Institute : Tunghai University, Graduate Institute of Information Management

Graduation Time : (01/2018)

Student Name : ChienTa Cheng Advisor Name : TzerShyong Chen

### **Abstract**

Information technology has been booming for years, human relies on artificial intelligent(AI) more and more. In addition, AI gives great performance on nature language process such as chatbot. However, there are few chatbot research on the veterinary hospital field. Thus, this paper tends to develop an instant messenger with AI technology.

Chatbot has been one of the famous application lately. Indeed, it is also an interesting topic. It could not only provide the entertainment, but also offer the instant response function. This research tends to crawl the data from Dcard pet forum and store in MongoDB for model training purpose, implementing the machine learning method to extract certain features. Finally, we use certain packages in Python coding to analyze all sentences in MongoDB including post title, comments, and contents. At the end, whenever users or members input a question or sentence, the chatbot would trace back to the trained model and give the proper response.

Keywords: Artificial Intelligence, Chatting Machine, Veterinary Hospital, Machine Learning, Retrieve-based model

# 目次

摘要	I
Abstract	II
第一章、緒論	1
第一節 研究背景	1
第二節 研究目的	1
第三節 研究架構	2
第二章、文獻探討	4
第一節 遠距照顧	6
第二節 互動式智慧資訊助理	6
第三節 機器學習與聊天機器人定義	7
第四節 人工智慧與聊天機器人應用及關鍵技術	9
第五節 聊天機器人設計相關軟體套件	11
第六節 Facebook messenger 平台	13
第三章、研究方法	14
第一節 系統架構	14
第二節 開放式聊天機器人資料蒐集流程	15
第三節 固定式聊天機器人資料蒐集流程	17
第四節 開放式聊天機器人資料選用	17
第四章、系統設計	19
第一節 系統設計	19
第二節 資料爬蟲	21
第三節 產生對話	22
第四節 資料前處理	25
第五節 檢索式模型	27
第六節 API 服務器	29

第七節 臉書平台-----	29
第五章、結論與未來展望-----	31
第一節 結論-----	31
參考文獻-----	32



## 圖次

圖 1.3 研究流程圖	3
圖 3.1 系統架構圖	15
圖 3.2 Dcard 標記回覆	16
圖 3.3 開放式聊天機器人資料蒐集流程	17
圖 3.4 固定式聊天機器人資料樣本設定	18
圖 4.1.1 樣版式模型	19
圖 4.1.2 檢索式模型	20
圖 4.1.3 生成式模型	20
圖 4.2.1 Python 網路爬蟲	21
圖 4.2.2 爬蟲結果 JSON	21
圖 4.3.1 產生對話流程圖	22
圖 4.3.2 標題標籤處理	23
圖 4.3.3 留言標籤處理	23
圖 4.3.4 標題+內文配對圖	24
圖 4.3.5 標題+內文+最佳回覆配對圖	24
圖 4.3.6 留言配對圖	25
圖 4.5.1 檢索式模型流程圖	27
圖 4.5.2 Match 流程圖	28
圖 4.7.1 結果呈現圖一	30
圖 4.7.2 結果呈現圖二	30

# 第一章 緒論

## 第一節 研究背景

現代社會因科技創新和思想的進步，使人們的生活方式越來越多元，再加上物價及收入水平不對等的社經因素，導致現代人普遍晚婚及高齡化與少子化的問題出現。因此，在現代社會中視寵物為子並將寵物視為精神寄託的現象越來越普遍，顯現出寵物已逐漸成為許多人們生活中不可或缺的夥伴，例如：寵物已成為平時休閒娛樂的第一選擇與對象，或是成為人際關係溝通的橋樑與生活話題[1]。隨著寵物地位的提升，寵物的照護也受到重視，尤其在高齡化及少子化社會中顯得格外重要，因此寵物相關產業隨著社會型態的改變而快速發展[2]。

隨著寵物在飼主心中的地位提升，傳統的看診服務已經無法滿足飼主們的需求，使得各家動物醫院必須重新思考營運模式和提供更多元與創新的服務，對於潛在客戶群的拓展與寵物相關資訊的資料庫開發，勢必需要詳細思考與制定策略。飼主對於飼養寵物的需求提升，故傳統的動物醫院必須轉型並提供更精緻多元的服務。全國寵物連鎖醫院在 2004 年的營收規模已經突破億元，在台灣市場趨於飽和的情況下，近年仍保有 10% 以上的營收成長空間，顯見此市場還在持續的擴展，亦顯示現今社會對於寵物照護的重視有增長的趨勢[3]。

## 第二節 研究目的

近年來，人工智慧的技術蓬勃發展，在各領域和實務上均有卓越的效果，而聊天機器人也逐漸成為熱門的方向，隨著通訊平台使用者日漸增加，包含 Wechat、Line、Facebook Messenger 等，人與人之間的溝通也成為生活中很重要的一部分。而廣為人知的聊天機器人如 Apple Siri 等，透過自然語言處理技術，將日常聊天與服務結合，進而提供使用者許多的便利性。不論是在搜尋資料、找美食、查詢天氣等，都比傳統自行查詢或詢問還要方便且更即時。

聊天機器人分類為固定式及開放式，固定式只提供特定的主題的回答或互動，而開放式領域則是接受任何不同領域的話題，並且嘗試去互動，而上述的 Apple Siri 則屬於開放式。相較於固定式聊天機器人，Apple Siri 沒辦法回答更專業或個人化的問題如「我家的狗狗目前狀況如何?」、「我家貓咪需要補充甚麼營養?」等等問題。

本論文之目的在建構一個結合開放式與固定式的聊天機器人系統，利用爬蟲技術的方式讓使用者在 Facebook Messenger 可以談論寵物相關領域的話題，讓聊天機器人可以與使用者進行基本的對談，並且與後台資料庫結合，讓飼主能夠詢問更專業或個人化的問題。

### 第三節 研究架構

本研究的內文共有五章，分述如下：第一章緒論，包含「研究背景」、「研究目的」與「研究架構」。第二章文獻探討，包含「遠距照顧」、「互動式智慧資訊助理」、「機器學習與聊天機器人定義」、「人工智慧與聊天機器人應用及關鍵技術」、「聊天機器人設計相關軟體套件」與「Facebook messenger 平台」。第三章研究方法，包含「系統架構」、「開放式聊天機器人資料蒐集流程」、「固定式聊天機器人資料蒐集流程」與「開放式聊天機器人資料選用」。第四章系統設計，包含「系統設計」、「資料爬蟲」、「產生對話」、「資料前處理」、「檢索式模型」、「API 服務器」與「臉書平台」。第五章則為結論與未來展望，本研究的研究流程如圖 1.3 所示：

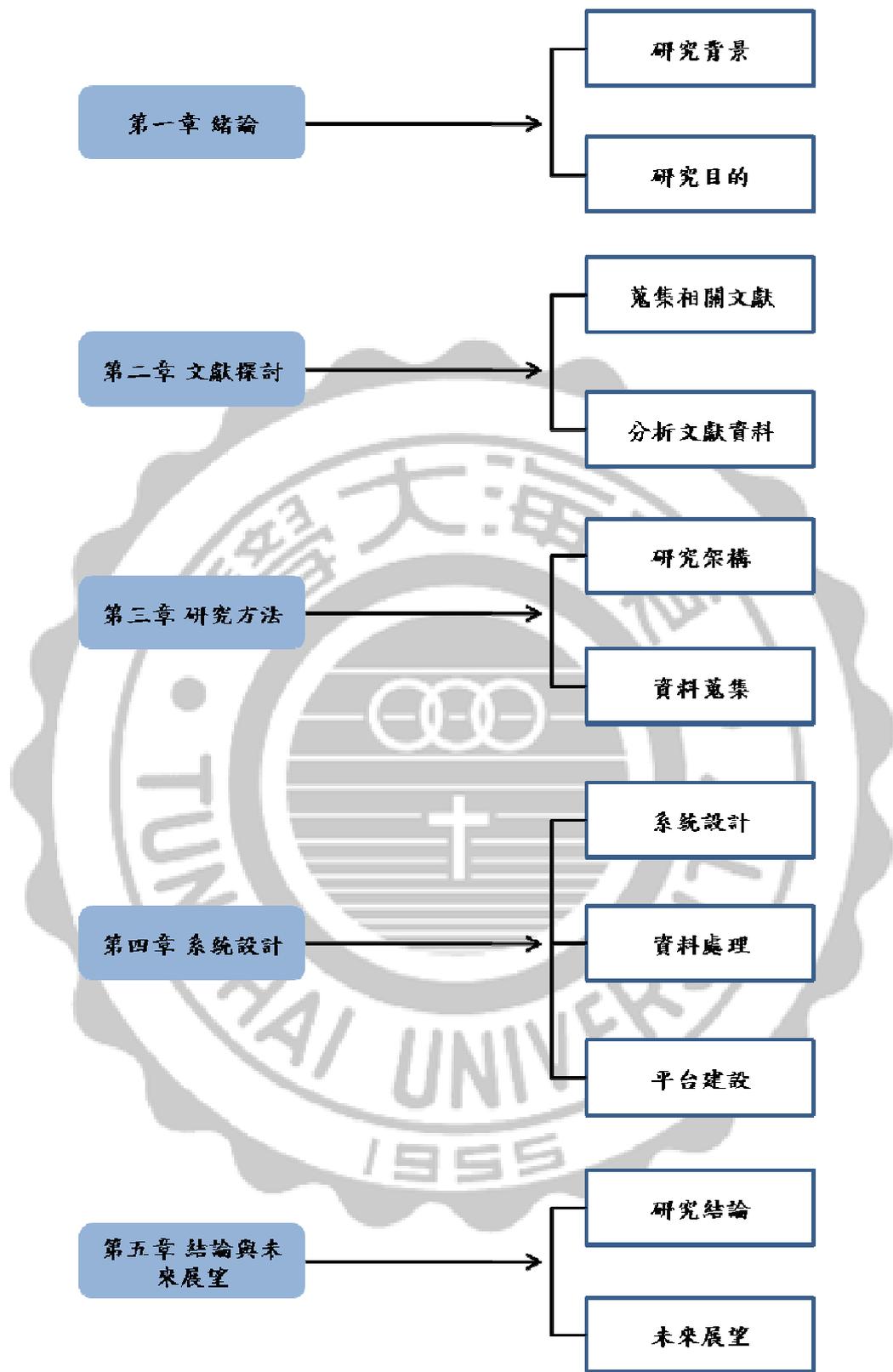


圖 1.3 研究流程圖

## 第二章 文獻探討

本章節將討論動物醫院產業、互動式智慧資訊助理、機器學習、與聊天機器人(Chatbot)等領域之國內外相關研究。在傳統的動物醫院，看診通常不會區分動物的種類與看診部位，大都是由同一位獸醫看診及治療。但隨著時代的進步與動物醫院的建立，飼主帶寵物上門求診時，期望有更貼心及細緻的照護選項，例如：飼主帶寵物到動物醫院求診時可以先依照寵物的種類，選擇具有專長的獸醫看診，例如：貓科、犬科等類別，更精確的看診可以分為心臟科、肝臟科、眼科、皮膚科、腎臟科等專門的單位。因此，現在動物醫院大都有提供分科看診的醫療服務。在學者鄒榮周於研究中[4]，將動物醫院的經營特色整理並分析：

- 一、動物醫院為服務業，其服務的對象為飼主及寵物。
- 二、注重醫療技術與熟練度，動物醫療人員必須不斷地自我充實及加強。
- 三、動物醫療的內容是具有彈性的，提供的服務會隨著飼主的要求而做調整。
- 四、動物醫療行業人力需求較為密集。
- 五、動物醫療人員因專業領域不同，需要與其他專業領域的寵物醫療人員互相合作，因此溝通能力、領導力及團隊合作極為重要。
- 六、動物醫院在寵物市場上極具競爭力，因獸醫師相關執照取得不易。隨著時代趨勢的變遷，學界也不斷擴增獸醫的相關學系，因此當越來越多人都有機會投入到寵物市場，相對的競爭力也就越來越大。

上述對動物醫院經營特色的整理可得知，動物醫院具有專業服務模式之性質，所以動物醫院必須提供能滿足顧客的服務，有效降低飼主的疑慮，故動物醫護人員的素質須特別加強，在不同的區域、時間、人員所提供的服務內容會產生不同的效果，也會導致動物醫院績效上的差異。依據經營的型態，目前台灣動物醫院市場的分類如下：

- 一、單店經營：領有獸醫執照資格的獸醫個人經營診所。
- 二、多店經營：由本店拓展出來的多個分店，開設的店名不一定相同，但經營者為同一人。
- 三、合資經營：多個具有相同背景的夥伴共同開設。
- 四、連鎖經營：具有制度、店名相同、有總公司與分公司的公司來經營。
- 五、複合式經營：多種寵物產業結合的經營模式，例如寵物美容、寵物用品等相關寵物產業之結合。

目前台灣動物醫院市場的型態多為單店經營，因為動物醫療的模式較為簡單。早期一位獸醫即可治療寵物的病症，但是隨著需求不斷的改變，飼主於寵物求診時若想購買寵物相關的商品，或兼做寵物美容及健康檢查及諮詢等服務，就需複合式寵物醫院，因此動物醫院的市場經營方向漸漸地轉為複合式經營模式。

由於現代科技發達，飼主接收寵物醫療市場資訊非常便利，動物醫院之間相對的競爭也較激烈，因此動物醫院的經營模式必須與時俱進，讓飼主感受到動物醫院的服務品質隨著醫療技術有所進步[5]。以往的動物醫院在服務上沒有現今方便，然而隨著網路通訊技術的進步，飼主不必到醫院就可以用網路預約看診，診療結束後動物醫院亦能透過電話或電子郵件詢問寵物的復原狀況[6]。本計劃為了提升飼主與動物醫院之間的便利與實用性，以建置動物醫療雲端智慧資訊助理為主要目標，建置雲端服務平台提供飼主查詢看診時間、地點、網路預約、及查詢寵物相關資訊等服務，同時使飼主得以遠距諮詢服務，系統將根據飼主的詢問給予適合的回饋，有別於以往的應答機制，系統將自動回饋給使用者較接近自然口語的回答，能大幅提升使用者整體滿意度與信任感。

## 第一節 遠距照護

Lewis 等學者認為遠距醫療可增加落後地域獲得專業醫療諮詢的機會[7]；Kuszler 給予遠端醫療一個簡單的定義，「簡單，但有用」，遠端醫療讓醫療人員可以透過遠端通訊設備來聯繫病患、診察、諮詢及提供治療方案。遠端醫療屬於新形態的醫療系統[8]，其主要功能有

- 一、增加居家安全性
- 二、系統主動提醒
- 三、監測活動安全
- 四、監測生理資訊
- 五、傳送即時訊息
- 六、異常提醒

Ward 等學者也認為輕微的疾病透過遠端醫療進行診療，可解決城鄉差距的問題[9]。

## 第二節 互動式智慧資訊助理

互動式智慧資訊助理是期望人類能使用語言、文字達成與智慧科技裝置之間能相互溝通的系統模式，協助人類完成許多事情[10]。這類型的代表系統如 APPLE Siri[11]、Google Now、微軟 Cortana、智慧對話系統 Alice[12]等，尤其在 Siri 帶動全球個人智慧資訊助理的開發風潮後，智慧資訊助理發展以聊天機器人 (Chatbot) 為其核心，將使用者談話內容或是需求跟智慧資訊助理講述，而智慧資訊助理即可透過語音辨識生成相對應的回覆回饋給使用者。早期的智慧資訊助理傾向問答機器，通常是以資料庫內的資料作為回應，這種對話機制缺乏彈性、豐富度與情緒[13]。然而隨著現今人工智慧、資料探勘的技術蓬勃發展，互動式系

統透過機器學習來建置，使其能理解使用者的自然對話內容[14]，並管理對話以及自動生成符合自然語言的對話內容回應。

### 第三節 機器學習及聊天機器人定義

#### 一、機器學習

機器學習是實現人工智慧的其中一種方法，除了能讓電腦系統從歷史資料中學習一套模式進而發展至各領域，例如：影像辨識或聊天機器人，還能結合統計、機率及其他資訊理論等，機器學習流程為輸入一串訓練資料、演算學習以及輸出，而使用機器學習的系統透過不斷反覆的演算來提升能力[15]，最著名的例子就是打敗各界圍棋好手的圍棋機器人 AlphaGo。依照機器學習學習過程可分為監督式與非監督式兩種方法[16]。

##### (一). 監督式學習 (Supervised Learning)

監督式學習透過反覆的輸入與輸出中機器會自行調整，並找到最好的演算或計算過程，在訓練結束後便擁有判斷與預測的能力，依據學習時建立的模型為使用者找出一個最佳的解答。

##### (二). 非監督式學習 (Unsupervised Learning)

以語意判別為例，非監督式學習是在已知句子要表達的意思後，分辨句子中的主詞、動詞、受詞，一開始就給予機器簡單的學習或判斷方式、規則等概念，在輸入資料給機器後讓機器自行去判斷學習正確答案。

#### 二、聊天機器人 (Chatbot)

聊天機器人是一種通過機器學習去模擬人類的自然語言，達成與使用者對話的程序，一般大都設在特定的裝置上，例如電腦或行動智慧設備，而軟體是

其核心部分[17]。聊天機器人是一種特殊的自動問答系統，其特點是模仿人類的語言習慣，有時甚至能有情緒的回答問題。透過機器學習訓練出來的模式與方法來找尋使用者問題的最佳回覆，而較高階的 Chatbot 應對的結果能以接近人類自然語言的口語進行回答，加強人與機器之間的溝通能力。聊天機器人以任務與技術面可分為兩種類別[18]。

#### (一). 任務式 (Task Oriented) 與非任務式 (Non-Task Oriented)

任務式的聊天機器人是具有明確的目標或是服務對象，例如：個人智慧資訊助理、語音客服機器人等，大多具有特定功能；然而，非任務式的聊天機器人則是沒有特定的領域範圍以及服務對象，多以聊天或是娛樂性質為主。

#### (二). 探索式與形成式的聊天機器人

探索式機器人是從既有的資料庫中去找尋對應的答案，但如果對話資料庫內的資料量不夠，容易造成搜尋不到答案的情況，且此類型屬於早期的問答機制，機器本身並沒有理解使用者所輸入的語言與意義，回答大多生硬且制式化。形成式聊天機器人則是透過人工智慧或機器學習的方式去學習一套應答模型，當使用者輸入語音對話後，機器會通過自然語言理解、對話管理、以及自然語言生成三個核心技術，整合出完整的對話並回應給使用者。

隨著時代的變遷，寵物醫院也在不斷的轉型，與此同時，社會型態的轉變也讓寵物的地位越來越重要，而寵物醫院的需求也隨之增加，如何提供更好的服務來連結醫院與客戶的關係也是重要的關鍵。在資訊普及的時代，如何有效利用新興技術並且於眾多服務當中脫穎而出，是很重要的議題。因此，本研究提出之智慧資訊助理資訊系統將能有效提升服務效率及品質，利用機器學習的方式讓人工智慧更加人性化且足以應對各類問題，有效解決客戶在任何時間提出的任何諮詢並給予專業的意見。

#### 第四節 人工智慧與聊天機器人應用及關鍵技術

人工智慧是指以人工的方式製造出來的系統所表現出來的智慧，是指通過電腦程式的手段實現的類人智慧的技術[19]。人工智慧主要是在解決具體問題，主要是透過使用各種不同工具來建置特定的應用。另一方面，聊天機器人則是一種經過人工訓練而能模擬人類對話的程序，通常運作在特定的軟體平台上，如 PC 平台或者移動終端設備平台，而硬體機械體則不是必須的承載設備。另外，聊天機器人也是一種即時性的自動問答系統，配合類人類的語言習慣，透過匹配的方式來尋找問題最合適的回覆或解答。聊天機器人對於已知道答案的問題，往往回答的比較正確，而對於不知道的問題，則通過猜測、轉移話題，或回答不知道的方式給出答案[20][21]。

##### 一、聊天機器人實際應用

隨著技術的成熟，聊天機器人在各領域的應用也層出不窮。可以分為線上客服、教育、個人助理和娛樂等 4 大種類。

- (一). 線上客服聊天機器人系統主要功能是與使用者進行基本溝通並即時且自動回覆相關問題，此功能不但能降低企業營運支出，也能提升用戶體驗，其使用裝置大部份都是在利用網站或手機 app。
- (二). 在教育方面，聊天機器人系統會根據教育內容建置數位互動式的教學環境，例如幫助使用者學習語言、指導用戶學習專業知識等。
- (三). 個人助理系統則大多透過語音或文字的方式與使用者互動(應該是與使用者互動?)，安排個人化的日常事務服務如查詢、設定行程、智慧搜索等等，其使用裝置大多為手機。
- (四). 在娛樂方面，聊天機器人系統主要功能是與用戶進行開放式主題互動，進而實現心理陪伴等功能。其使用裝置大多為社交媒體為主。

## 二、 聊天機器人關鍵技術結構

機器人關鍵技術結構的組成元素有三個，分別是自然語言理解、對話管理系統和自然語言生成系統。在接受到使用者語音或文字的輸入後，自然語言理解會將之統一轉換成文字形式進行處理，理解使用者所輸入的語意後，再導入至對話管理系統。對話管理模塊則負責協調各個模組的調用及維護當前對話狀態。最後，自然語言生成系統負責選擇特定的回覆方式，並轉成文字或語音輸出給使用者 [22]。

### (一). 自然語言理解

自然語言理解主要為解析使用者所輸入的對話內容，透過監督式機器學習的方式來判斷該句子的意圖和實質概念，讓智慧資訊助理在接受到使用者的句子之後，可以根據先前分類且訓練過的資料，判斷該句子屬性，同時透過訓練資料來擷取使用者的具體目標。

### (二). 對話管理系統

對話管理系統主要是根據目前已知的資訊來決定多次來回對話的方式，此系統是運用深度學習技術，更準確的分析與判斷句子。且系統又可細分為對話狀態追蹤 (DST) 及決策模塊 (Policy)，決策模塊的目的是根據當前狀態來選擇輸出方式。另一方面，先利用監督式學習的方式，將從網路上的對話力用爬蟲存入資料庫後，透過訓練產出一個對話模型，再設計一個使用者模擬器來加強智慧資訊助理的訓練，透過人工對話及校準的方式來提高準確率。使用者模擬器會在每段對話中產出一個對話目標，例如預約、諮詢等目的。之後會針對智慧資訊助理的回答來反饋加權值，透過此方式校準類神經訊號參數，如果回答正確即為正分，反之則為負分。透過更新回合不斷提高智慧資訊助理的判斷準確率，降低發生錯誤之機率。

### (三). 自然語言生成系統

為了避免在與聊天機器人對話過程中產生過於呆板和制式化的回覆，自然語言生成系統設定回應規則，讓同一個句子可以用不同的方式表達，使聊天機器人能更貼近人性與生活化。近年來，在聊天機器人系統上的對話生成主要包含檢索式和生成式兩種技術[23]。

#### 1. 檢索式對話

檢索式對話式是透過深度機器學習的方式，在已有的對話資料庫中，以匹配的方式來找尋最佳回覆的內容。但是此方法的限制較多，與使用者間的互動只能以固定模式進行。

#### 2. 生成式對話

生成式對話是以深度機器學習將已有的資料庫學習完，再透過與目前機器翻譯類似常用的編譯碼，以逐字生成的方式完成一則完整的回覆。利用此技術所產稱的回覆雖然更具有多樣性，但是缺點為許多以此種方式產生的回覆語句無法被人類所理解。

## 第五節 聊天機器人設計相關軟體套件

### 一、Jieba

Jieba 是中國開發者 Sun Junyi 所開發的中文斷詞函式庫，實作於多種程式語言，主要算法是使用 trie-tree 資料結構結構來進行字典配對，並將配對中剩下的單字以 hmm 隱馬可夫模型來組合產生出新詞[24]。

### 二、Gensim

Gensim 是一個開源的 python 函式庫，主要是提供主題模型的建立，當中也實現了多種算法，如統計中的加權技術-TF-IDF、主題模型分類的無監督式模型-

LDA、詞的空間向量算法- Word2Vec[25]。

### 三、Deepnlp

Deepnlp 是基於 TensorFlow 開發的一個 Python 版本 NLP 開源函式庫，主要是將 TensorFlow 平台上的一些機器學習模型結合 NLP 算法，並實現如斷詞、詞性標註、自動抽取摘要、關鍵句子抽取等自然語言處理功能[26]。

### 四、Web Crawler

網路爬蟲(Web Crawler) 是一自動爬取網頁的程式，它能夠透過搜尋引擎從網路中下載網頁，為搜尋引擎的重要組成元件。網路爬蟲可將自己所需的頁面內容儲存下來，以便事後資料分析或相關應用[27]。

### 五、MongoDB

MongoDB 是 10gen 開發出來的 NoSQL 資料庫，是用來處理大數據級的資料庫，而他是以文本的方式儲存的資料庫，儲存的方式與 JSON 格式完全相同，另外多了很多的靈活性，允許每一筆文件的欄位型態不同[28]。

### 六、JSON

JSON(Javascript Object Notation)是一種數據交換格式，以純文字儲存和傳送簡單的結構資料，可以透過特定的格式儲存任何資料(字串、數字、陣列、物件)，也可以透過物件或陣列來傳送較複雜的資料，易於閱讀和編寫，也易於機器解析和生成，這些特性使 JSON 成為較理想的數據交換語言。

## 第六節 Facebook messenger 平台

Facebook messenger platform 是聊天開發平台由 Facebook 提供給開發者，開發者可以在平台上提供服務，這個平台支援多種傳輸，例如:文字、圖片及影片等，同時也提供以 API 的方式傳遞文件。此外，在聊天的過程中，還提供多種互動按

鈕功能，遵守平台開發規定後即可提交審核，通過後便可以發布應用程式，使用者即可在 Facebook 平台上找到此應用程式[29]。



## 第三章 研究方法

本研究將針對一般常見的寵物健康建置一個開放式聊天機器人系統，不論使用者是否為該動物醫院的客戶都可以與其互動，聊天內容廣泛較不受侷限如動物習慣、寵物飲食、異常狀態諮詢和趣味分享等。而在專業領域上，將透過一個醫院資訊系統來記錄客戶寵物的個別情況，這種固定式的聊天機器人能夠更精準且即時回覆關於寵物的問題，同時也能提供預約、門診提醒、飲食攝取通知和備忘錄等的服務。

### 第一節 系統架構

由於資訊科技的運用，對於寵物相關資訊的取得不再如以往只能透過與獸醫人員面對面諮詢。本研究根據現有的機制及環境需求設計聊天機器人，運用機器學習與類神經網路等技術，使寵物飼主可以獲得最即時的寵物醫療資訊。讓使用者可以與其進行一般寵物相關對談，也進而提供會員更專業及個人化的服務，整體流程如下圖 3.1 所示。

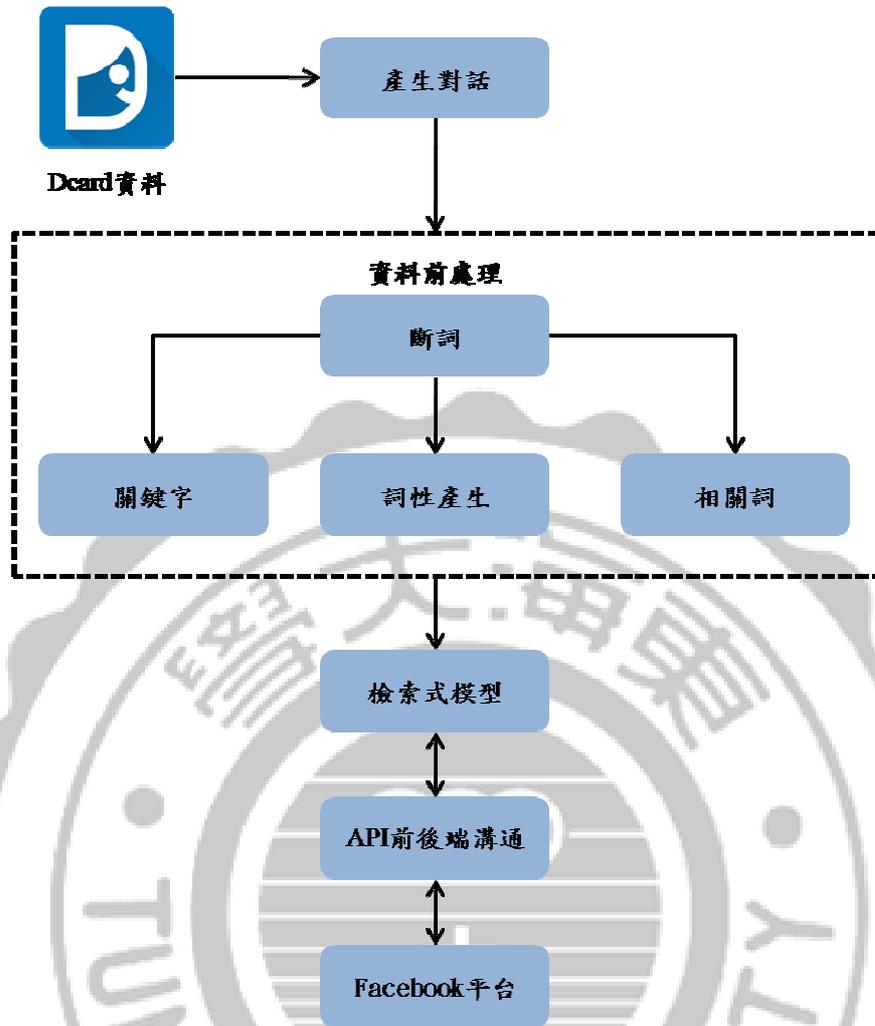


圖 3.1 系統架構圖

## 第二節 開放式聊天機器人資料選用

本論文於建置聊天機器人前，使用 Dcard 論壇之寵物區的資料做為訓練樣本。

Dcard 為台灣一個提供社群網路服務的網站，僅開放台灣跟海外部分大學的學生註冊[30]。此社交平台有以下 3 個特性：

### 一、身分驗證

使用者必須以大學學校信箱註冊。

### 二、匿名

使用者在平台上討論時不得交換個資。

### 三、標記回覆

貼文中的留言常以 Bx(x 為數字)作為代號，並且與先前的留言者動，如下圖 3.2。



圖 3.2Dcard 標記回覆

匿名的特性讓用戶在討論上可以更活躍，由於必須使用大學信箱註冊，因此樣本資料的同質性會比其他社交平台高，而最後的標記回覆則是可明白使用者們句子間的呼應，適合做為對話訓練樣本。

### 第三節 開放式聊天機器人資料蒐集流程

在蒐集流程上，首先必須設置爬蟲(Crawler)，抓取 Dcard 寵物版的新文章儲存至資料庫，並且給予一個辨識號(ID)，而後每天更新同辨識碼文章內的留言至資料庫。關於爬蟲的設定則是使用排程方式，每 6 小時定期抓取新資料樣本，詳細流程如下圖 3.3 所示。

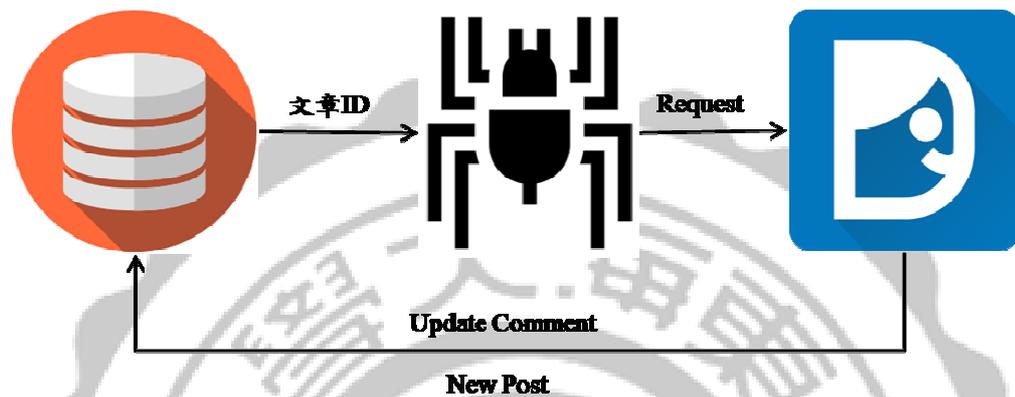


圖 3.3 開放式聊天機器人資料蒐集流程

#### 第四節 固定式聊天機器人資料蒐集流程

有別於開放式聊天機器人資料樣本收集方式，固定式會針對較個人化或專業領域的問題做互動，本研究將透過人工的方式定義使用者意圖(Intent)和實體概念(Entity)。其中實體概念是根據各別動物醫院所提供的服務定義，本研究將服務分類為「門診」、「就醫紀錄」、「醫囑」、「檢驗報告」和「主動式提醒服務」等五大類，而在這五大類當中，都有個別的價值(Value)，例如門診預約底下就存在「內科」和「外科」的價值。另外一方面，本研究以人工的方式設計使用者的可能性問句，並歸類其語句的「意圖」，這邊分為三大類「預約」、「查詢」和「提醒」。於設計時也預先將飼主較常使用的文句，利用不同語句表達，例如在預約意圖中可能的詢問方式包含「我想要預約門診」、「我想要掛號」、「下星期內/外科有空嗎？」等方式，資料樣本設計如下圖 3.4 所示。

IF USER SAYS SOMETHING SIMILAR TO:	BOT REPLIES WITH TEXT ▾ <input checked="" type="checkbox"/> RANDOM
<input type="text" value="下星期內/外科有空嗎?"/>	<input type="text" value="好的，馬上幫您查看"/>
	OR TEXT ▾ <input type="text" value="請稍等，馬上為您查看"/>
	OR TEXT ▾ <input type="text" value="沒問題!馬上為您安排"/>

圖 3.4 固定式聊天機器人資料樣本設定



## 第四章 系統設計

### 第一節 系統設計

現今常見到的聊天機器人實作模型，其建置方式有以下三種：

#### 一、Rule-based model -樣板式模型

此類模型是透過建立規則來讓機器人找到對應的回覆，例如當使用者輸入「肚子餓」時，聊天機器人便會回覆「吃飯」，這類的模型建置通常需要以人工的方式來定義規則。

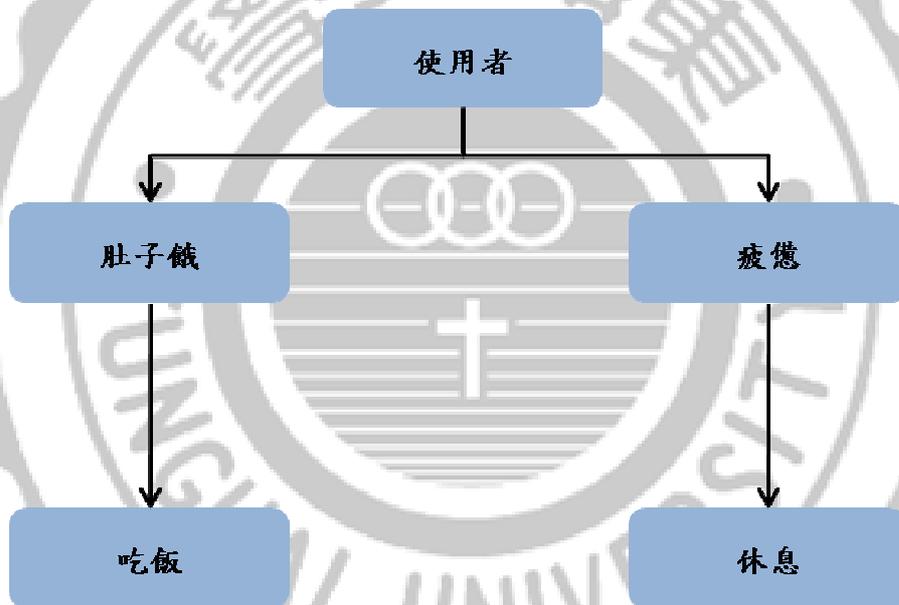


圖 4.1.1 樣板式模型

#### 二、檢索式模型 -檢索式模型

此模型的特性類似頁面搜索引擎的問答系統，在設計使用者意圖與實體關鍵句型或字後，再將使用者輸入的問題參數帶入，並尋找權重最高的意圖及實體，及以相似度最高的答案做為回覆。

Question set =  $Q_1, Q_2, Q_3, \dots, Q_n$

Answer set =  $A_1, A_2, A_3, \dots, A_n$

圖 4.1.2 檢索式模型

### 三、Generative model - 生成式模型

此模型目前作法是 sequence to sequence[31]來實作，sequence to sequence 是串接兩個 RNN[32]或 LSTM[33]，一個當作編碼器，把句子轉換成表示式，另一個當作解碼器用來當作轉譯，sequence to sequence 的邏輯是將中文句子編碼再解碼成英文句子，而 sequence to sequence 的聊天機器人則會將輸入句子編碼再解碼成下一句對應的句子。

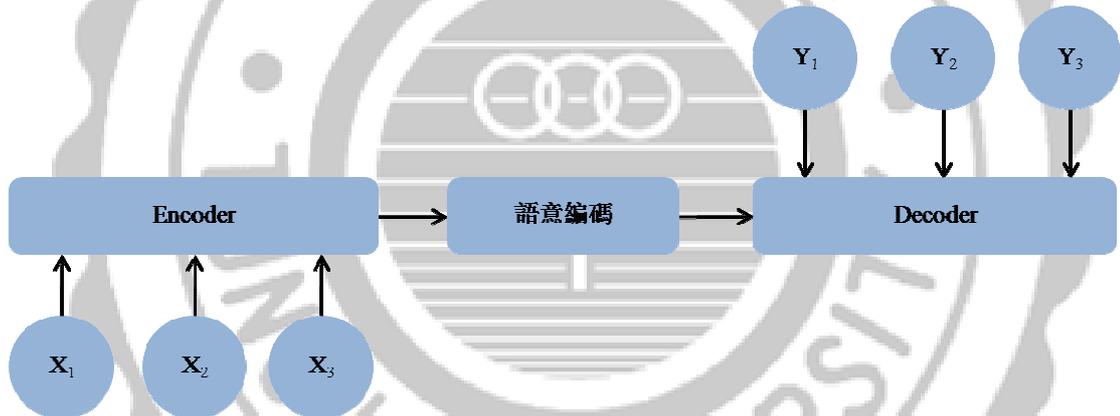


圖 4.1.3 生成式模型

生成式模型與檢索式模型都是偏向開放式聊天機器人模型，其中檢索式模型可用於找部分實作，如同義詞擴展的模糊搜尋等。在比較之後，本文將採用檢索式模型來建置固定式聊天機器人，將爬蟲所蒐集到的論壇寵物版資料做為訓練樣本，並將對話進行資料前處理包含斷詞、相關詞、和詞性判斷，最後將特徵導入訓練模組，達成更好的匹配。整體成果會在 Facebook messenger 呈現，在固定聊天機器人的建置下，本研究將導入第三方資料庫來實現，下面會說明其建置邏輯。另一方面，固定式聊天機器人的將利用樣板式模型方式建置，透過人工方式設計資料樣本加以訓練，回覆將會根據使用者的輸入去尋找是否有規則建立，若無則

回覆「我不清楚您的意思」。

## 第二節 資料爬蟲

是用網路爬蟲技術以及 Dcard 所提供的 API 網址，可以有效的將寵物版相關的文章儲存至後台資料庫。本研究使用 Python 做為爬蟲程式撰寫的語言，目的在於抓取文章標題、內容與留言，部分程式語言之撰寫及結果，如圖 4.2.1 及圖 4.2.2 所式。

```
1 #引入函示庫
2 import requests, json
3 import matplotlib.pyplot as plt
4 #取得文章列表資料
5 url="https://www.dcard.tw/_api/forums/pet/posts?popular=false"
6 reqs = requests.get(url)
7 #利用json.loads()解碼JSON
8 reqsjson = json.loads(reqs.text)
```

圖 4.2.1 Python 網路爬蟲

```
"id": 228099788,
"title": "#貓#圖#貓的世界",
"excerpt": "大家的貓咪都很喜歡看窗外嗎？查理每天都想再多認識這個世界，一開始會不懂窗外有什麼好看的，到後面發現自己也會陪著他一起看，好像更享受這個世界",
"anonymousSchool": false,
"anonymousDepartment": false,
"pinned": false,
"forumId": "7ee21581-1307-4ba9-b9dc-82028bdcca49",
"replyId": null,
"createdAt": "2018-01-10T09:50:01.416Z",
"updatedAt": "2018-01-10T09:50:01.416Z",
"commentCount": 0,
"likeCount": 8,
"withNickname": true,
"tags": [...], // 20 items
"forumName": "寵物",
"forumAlias": "pet",
"gender": "F",
"school": "查理媽",
"department": "mizuki_26",
"replyTitle": null,
"reportReason": "",
"hidden": false,
"withImages": true,
"withVideos": false,
"media": [...], // 8 items
```

圖 4.2.2 爬蟲結果 JSON

## 第三節 產生對話

本研究在 Dcard 論壇資料蒐集了 120 筆文章，為了擁有訓練樣本，因此必須從蒐集到的資料中產出對話樣本。產生對話的流程圖，如圖 4.3.1 所示。

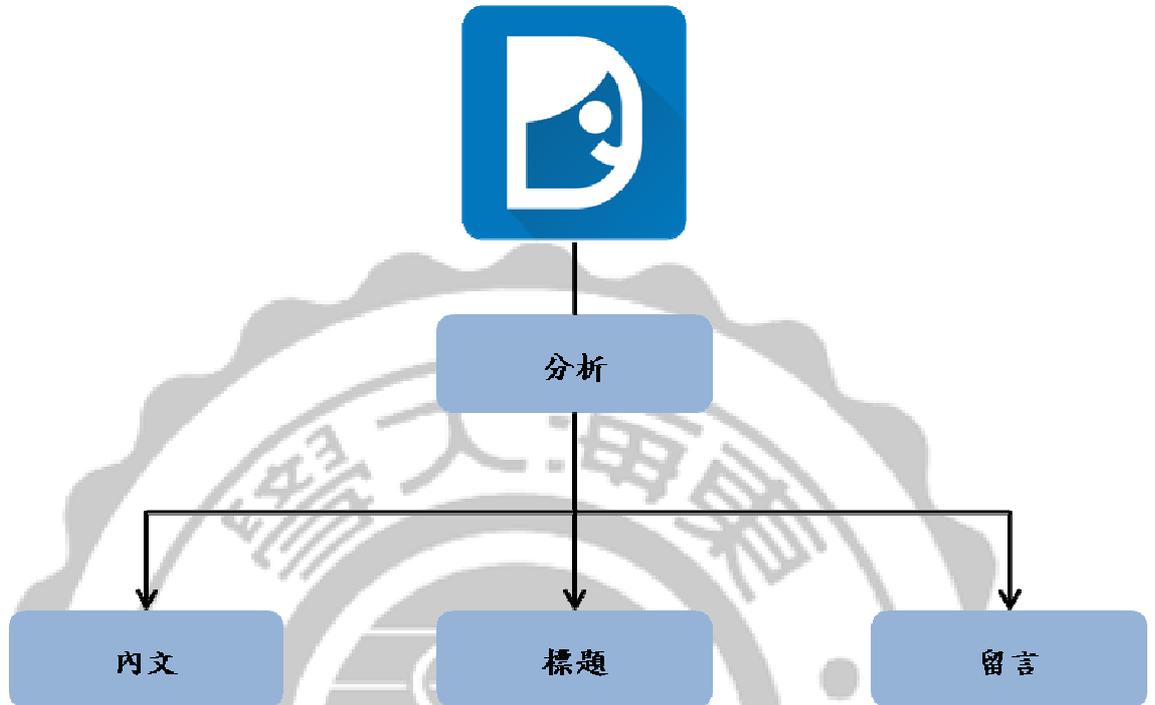


圖 4.3.1 產生對話流程圖

#### 一、Parse Dcard 資料

在產生對話前先對內容做簡單的分析：

##### (一). 圖片萃取(Extract image)

Facebook messenger 支援 Dcard 論壇中的圖片儲存模式，本文將圖片檔用儲存於 Imgur 的雲端空間。

##### (二). 標籤處理(Detag)

Dcard 在內文或回覆中常見使用者用#做為 tag 的前綴詞，透過標籤處理的方式可以抽取內文獲回覆中的重要特徵，。

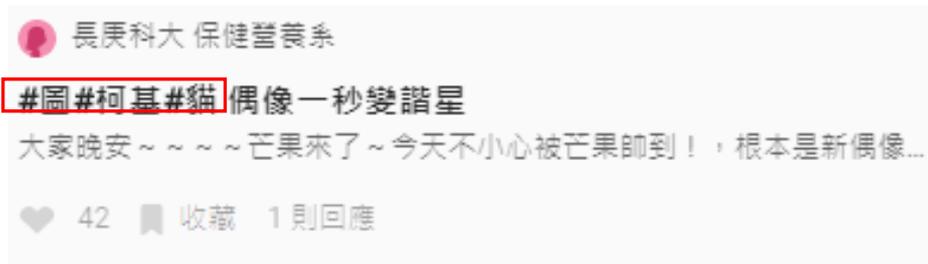


圖 4.3.2 標題標籤處理

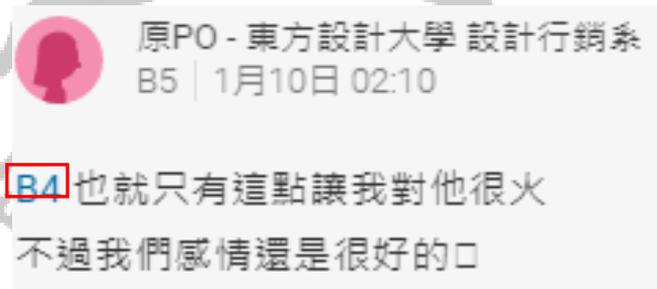


圖 4.3.3 留言標籤處理

## 二、產生對話

將資料分析後，產生配對，配對將分成以下三種的組合，所產生的對話配對當作樣本訓練用：

### (一). 標題+內文配對

第一種對話產生方式是將標題與內文做配對，這種方式具有延伸的效應，標題經常是內文的摘要，但是如果內文字數過多，造成回覆給使用者過長的句子，因此內文字數超過 250 的將不會產生出此種配對。

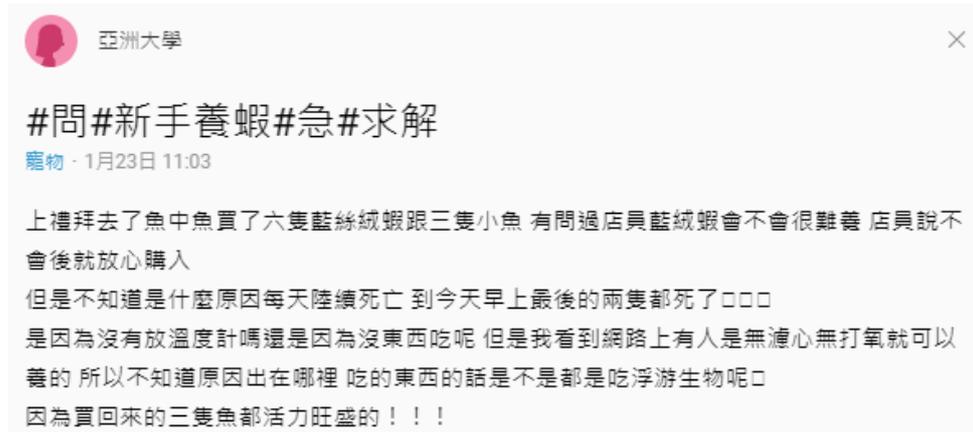


圖 4.3.4 標題+內文配對圖

(二). 標題+內文+最佳回覆配對

延續上一個方式，但若內文是問題式標題的延伸，換句話說，若內文跟標題一樣都是在表達某些問題的時候，則需要利用留言中按讚數最高的做為對話樣本配對。

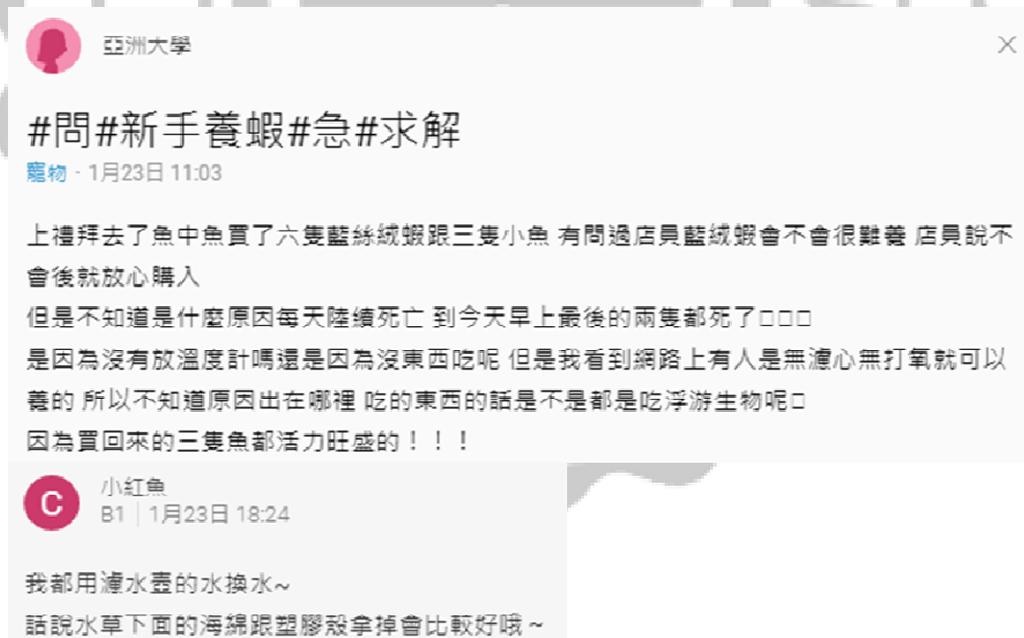


圖 4.3.5 標題+內文+最佳回覆配對圖

(三). 留言間配對

使用者習慣在回覆中用 Bx(x 為數字)來標記別人的回覆，將此標記取出可得知使用者欲回覆的對話，便可以將句子形成配對。



圖 4.3.6 留言配對圖

#### 第四節 資料前處理

本文將分為三個部份，依序為斷詞與關鍵詞權重、詞性標記和相關詞。

##### 一、斷詞與關鍵詞權重

對話樣本中，必須將句子精簡，將當中的文字資料進行斷詞，分割成較小且有意義的單位，將句子分割成詞彙後，詞可以提供給檢索式模型建立索引幫助計算相似度，也可以利用詞彙計算在所有文章中的空間分布，計算出關鍵詞的權重，進而找出句子中的重要詞彙。

##### (1) 斷詞

本論文將使用 Jieba 資料庫進行斷詞。

## (2) 關鍵詞的權重

tf-idf( term frequency-inverse document frequency)是本論文用來計算關鍵字的權重方式，tf-idf 是一種用於資訊檢索與文字探勘的常用加權技術，為一種統計方法，用來評估單詞對於文件的集合或詞庫中一份文件的重要程度，本論文用此方法來計算每個詞在於一個句子中的重要程度，tf-idf 的公式如下：

$$\text{weight}_{i,j} = \text{tf}_{i,j} * \log\left(\frac{N}{\text{df}_i}\right)$$

$\text{weight}_{i,j}$  : i 詞於 j 句子得到的分數

$\text{tf}_{i,j}$  : i 詞於 j 句子中出現的頻率

$\text{df}_i$  : 有多少句子中出現 i 詞

N : 總共有多少個句子

## 二、詞性標記

PoS tagging (Part-of-Speech tagging) 是處理自然語言中的一個重要議題，主要是根據型態與句法功能定義詞的分類，了解詞彙的分佈特徵、型態特徵及無關於含意，可用來作意圖分析。本研究使用 Deepnlp 建立詞性標記模型，更進一步尋找輸入與輸出之間的關係。

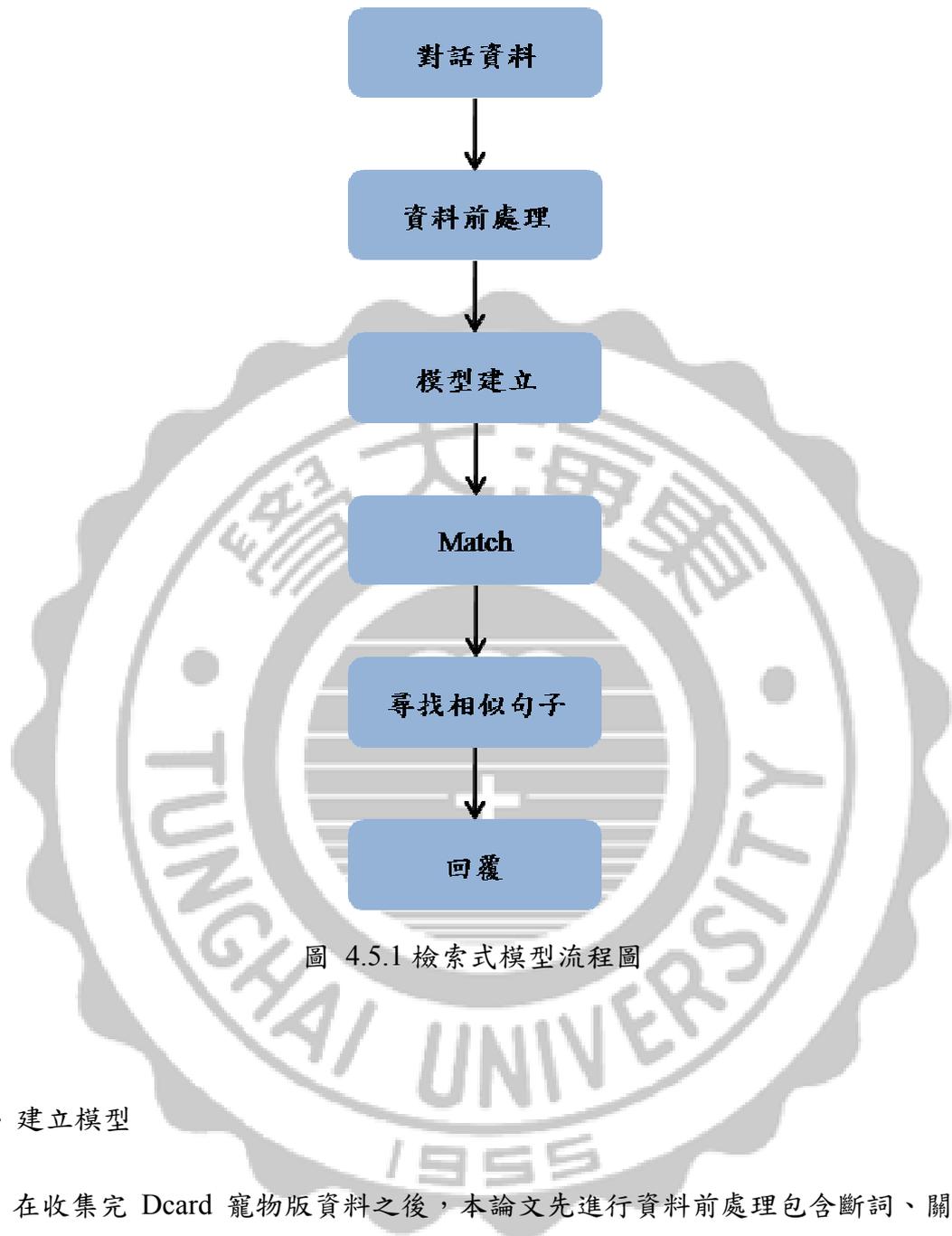
## 三、相關詞

本論文使用 Gensim 工具了解詞彙分布狀況。Gensim 中的 Word2Vec 為 Tomas Mikolov 於他的論文「Efficient Estimation of Word Representation in Vector Space, 2013」[34]中所提出。

## 第五節 檢索式模型

檢索式模型可分成三個階段：模型建立、配對(Match)、回覆(Response)三部

分，流程圖如圖 4.5.1 所示。



### 一、建立模型

在收集完 Dcard 寵物版資料之後，本論文先進行資料前處理包含斷詞、關鍵字權重計算和相關詞分布，根據抽取的特徵及分數進一步訓練模型，另外一方面，透過建立詞彙的索引，能有效提升尋找的速度。

### 二、Match

系統會將使用者輸入的句子進行斷詞分析，利用索引表中尋找有關聯詞彙的句子，從索引表中得到該句子辨識碼後，便會對這些句子進行計算分數，最

後得到依分數排序好的辨識碼。

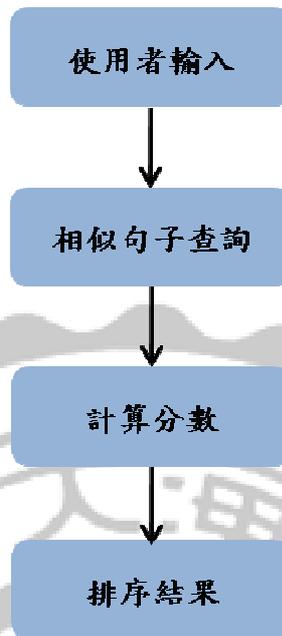


圖 4.5.2 Match 流程圖

### 三、Response

從 Match 步驟中，會得到依分數排序好的句子辨識碼，回覆(Response)會觀察分數，先對分數進行分群，並依序從高分群中挑出答案，放入回答清單，再對回答清單進行排除不可能的結果，最後從中挑選一則最為相符作為答覆。

### 第六節 API 服務器

API 服務器負責前後端的溝通，當使用者透過 Facebook messenger 發送了訊息，使用者的行為會被傳送到 API 服務器，API 服務器會根據使用者的行為，向檢索式模型要求對應的資料，API 服務器收到從 messenger 傳來的訊息後會先做檢查，再向後端要求服務。

### 第七節 Facebook messenger platform

Facebook messenger platform 是 Facebook 提供給開發者的工具，開發者可透過創立粉絲專業，將粉絲專業的聊天室綁定後端程式，提供 Chatbot 的服務，以

下將簡單介紹 Facebook messenger platform 所提供的功能:

### 一、傳送訊息

傳送訊息給使用者，可傳送五種類型的訊息：text、image、audio、videos、file。

### 二、按鈕

在訊息中附加按鈕，可定義按鈕的屬性，其中有網址按鈕、回傳按鈕、通話按鈕。

### 三、快速回覆

類似按鈕的功能，放在下方，可以加上圖示，欲使用需在當下點擊按鈕，一開始對話時便會洗掉快速回覆的按鈕。

### 四、公開

未經過審核不能公開，未公開期間只有開發人員與測試人員能傳送 messenger，公開後則 Facebook 的使用者都能到粉絲專業發送訊息。

圖 4.7.1 和圖 4.7.2 為在 Facebook messenger 平台上之畫面。

全國毛小孩醫院  
0人說這讚  
非政府組織 (NGO)

2017-12-27 0:39

好的，馬上幫您查看

我想要預約門診

沒問題!馬上為您安排

我想要預約門診

9:12

柯雞無敵啊!

那狗狗好可愛

對呀!不要不要的

那狗狗可愛

柯雞無敵啊!

狗狗可愛

圖 4.7.1 結果圖示樣本一

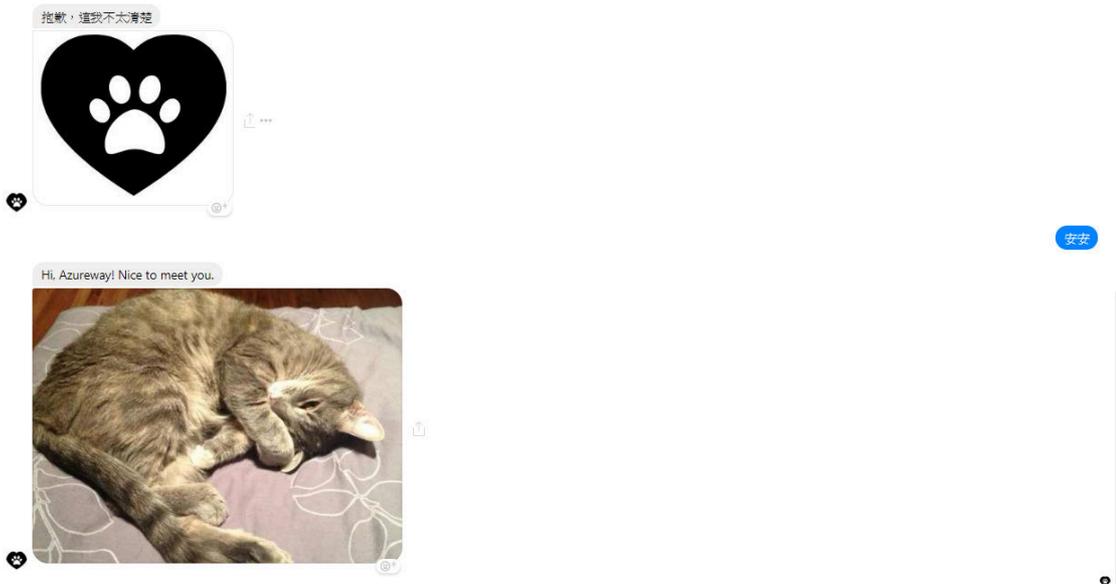


圖 4.7.2 結果圖示樣本二

## 第五章 結論與未來展望

### 第一節 結論

本論文實作了一個聊天機器人系統，並將系統包裝成應用程式取名為毛小孩動物醫院發佈於 Facebook messenger 上，讓使用者可以此系統滿足聊天的需求以及當遇上問題時能即時得到回應。

此系統在問答準確率方面，是透過找到最相似的句子，並將該句子的回覆作為答案給使用者，然而有些時候最相似的句子與使用者輸入仍然有極大的差距，因此使用者會感受到答非所問的疑惑感，而解決方式則是蒐集更多資料，或採用句型結構的判斷，而非找尋最相似句子的模型。另外一方面，在提升主動的部分，系統目前的特性是收到一句輸入問題，才會輸出一句答覆，然而此模式仍過於被動，未來期望能由聊天機器人主動詢問或開啟新話題。最後聊天機器人除了有聊天的功能，未來還能朝主動與使用者進行互動或推薦，增加使用者的黏著度。

## 參考文獻

- [1] 蔡岳廷(2009)，〈台灣民眾寵物飼養行為之研究：計畫行為理論之應用〉，來源：世新大學社會心理學研究所碩士論文。
- [2] 謝佳書(2009)，〈現代寵物飼養—互動論取向研究〉，來源：輔仁大學社會學系碩士班碩士論文。
- [3] 林洧楨(2012)，〈狗兒子、貓女兒，寵出五百億商機：從出生到墳墓什麼都能賣，生前契約也熱銷〉，財訊雜誌。
- [4] 鄒榮周(2007)，〈動物醫院服務品質與顧客滿意度、顧客忠誠度關係之研究〉，來源：國立嘉義大學管理學院碩士論文。
- [5] Dian A. H., Adhistya E. P., Silmi F., Ida F., "Management Information Systems Development for Veterinary Hospital Patient Registration Using First in First Out Algorithm," International Conference on Biomedical Engineering (IBIOMED), Yogyakarta, Indonesia, 2016.
- [6] 李易璋(2009)，〈建構寵物健康管理平台之研究〉，來源：臺北醫學大學醫學資訊研究所碩士論文。
- [7] Lewis E. R., Thomas C. A., Wilson M. L. and Mbarika, V. W., "Telemedicine in Acute-Phase Injury Management: a Review of Practice and Advancements," Telemedicine and e-Health, Vol.18, No.6, pp. 434-445, 2012.
- [8] Kuszler P. C., "Telemedicine and Integrated Health Care Delivery: Compounding Malpractice Liability," American Journal of Law and Medicine, Vol.25, No.2-3, pp.297-326, 1999 .
- [9] Ward M. M., Jaana M. and Natafqi, N., "Systematic Review of Telemedicine Applications in Emergency Rooms," International Journal of Medical Informatics, Vol.84, No.9, pp. 601-616, 2015.
- [10] Varsha B. and Potey M. A., "Question Answering System: A Heuristic Approach," International Conference on the Applications of Digital Information and Web Technologies (ICADIWT), Bangalore, India, 2014.
- [11] Hauswald J., Michael A. L., Yunqi Z., Cheng L., Austin R., Arjun K., Ronald G. D., Trevor M., Vinicius P., Lingjia T., Jason M., "Sirius: An Open End-to-End

- Voice and Vision Personal Assistant and Its Implications for Future Warehouse Scale Computers,”The International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), Istanbul, Turkey, pp.223-236, 2015.
- [12] Rui L. and Ming X.,“Development of Entertaining Robots: A Study into Intelligent Acquisition of Human Language in Robots,” International Conference on Information and Automation, Zhuhai, Macau, China, pp. 1368-1373, 2009.
- [13] Shamli D., Ritika B., Vijayalaxmi R., Asmita S.,“An Interactive Medical Assistant Using Natural Language Processing,” International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), Jalgaon, India, 2017.
- [14] 趙文銀(2015)，〈信息結構化方法與自然語言處理機制研究〉，來源：中國詞彙語義學研討會論文。
- [15] MacKay D. J. C. (2003)，〈資訊理論、推理和學習演算法〉，來源：劍橋大學出版社，ISBN 0-521-64298-1。
- [16] 張偉男，劉挺(2016)，〈聊天機器人技術的研究進展〉，中國人工智慧學會通訊第6卷第1期。
- [17] Francois-Lavet V., Fonteneau R., and Ernst D.,“How to Discount Deep Reinforcement Learning: Towards New Dynamic Strategies,” NIPS 2015 Workshop on Deep Reinforcement Learning, 2015.
- [18] Hado v. H.; Marco A. W.,“Reinforcement Learning in Continuous Action Spaces,” IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning, Honolulu, HI, USA,2007.
- [19] 維基百科  
<https://zh.wikipedia.org/wiki/%E4%BA%BA%E5%B7%A5%E6%99%BA%E8%83%BD>， 2017.12.
- [20] 張偉男，劉挺，“聊天機器人技術的研究進展”，中國人工智慧學會通訊，第6卷第1期，2016.
- [21] 寧長英，“智能聊天機器人的關鍵技術研究”，杭州電子科技大學碩士論文，2011.
- [22] 張偉男，劉挺，聊天機器人技術的研究進展， <http://chuansong.me/n/2217011>， 2016.1.

- [23] 趙文銀，“信息結構化方法與自然語言處理機制研究”，中國詞彙語義學研討會論文集，2015
- [24] Jieba - <https://github.com/fxsjy/jieba>
- [25] Gensim - <https://radimrehurek.com/gensim/>
- [26] Deepnlp - <https://github.com/rockingdingo/deepnlp>
- [27] Web Crawler -  
<https://zh.wikipedia.org/wiki/%E7%B6%B2%E8%B7%AF%E7%88%AC%E8%9F%B2>，2017.12.
- [28] MongoDB—<https://en.wikipedia.org/wiki/MongoDB>
- [29] Facebook messenger platform -  
<https://developers.facebook.com/docs/messenger-platform>
- [30] Dcard-<https://zh.wikipedia.org/wiki/Dcard>
- [31] Sutskever, Ilya, Vinyals, Oriol, and Le, Quoc VV. (2014). Sequence to sequence learning with neural networks. In NIPS, pp. 3104–3112.
- [32] Mikolov, T., Karafiat, M., Burget, L., Cernocký, J., and Khudanpur, S. (2010). Recurrent neural network based language model. In Proceedings of INTERSPEECH, vol. 2010, no. 9. International Speech Communication Association, 2010, pp. 1045–1048.
- [33] Hochreiter, S and Schmidhuber, J. (1997). Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780.
- [34] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In ICLR