應用於對話式介面之語句用意分析

（一）摘要

隨著行動裝置的蓬勃發展，人們的日常與手機越來越形影不離，因此人機互動系統的介面設計變得十分重要。有的人機互動系統利用簡易的介面，來幫助使用者在短時間內上手、有的則利用易懂的圖標，讓使用者一眼就能掌握系統狀態，然而，多數介面都需要使用者的手動操作，在工作繁忙或手忙腳亂的當下，就顯得十分不便。如果能夠省去手動操作的步驟，並改以口語對答的方式來幫助使用者完成工作，那麼使用者將不再需要熟悉各種五花八門的介面，也不需要在百忙之中騰出手腳。如此一來，生活就能夠更加便利。

目前智慧語音助理雖然能夠實現日常的簡單功能，例如設定鬧鐘、天氣查詢、播放音樂等等，但若是其他組合式或更為複雜的功能，往往因為難以分析整句話的語意，而導致無法有效地處理。例如「把大專生專題計畫申請辦法的電子郵件轉寄給高禾」，系統必須判斷出此句語意包含搜尋及寄信兩個基本動作，又或「我想要找高禾在上個月寄給我的電子郵件」，系統必須準確分析出「找」對應的查詢功能，以及「高禾」、「上個月」對應的內容及收件時間等查詢條件，才能處理這個問題。

本研究將以 Gmail 應用程式為範例，利用實驗室過去在資訊擷取、語意理解的研究成果，開發可以透過對話互動查詢、寄送、讀取的電子郵件對話式互動介面，讓使用者能夠透過口語對話分析使用者的語意，並準確地提供相對應或複合式的功能。

（二）研究動機與研究問題

行動裝置的普遍，衍生了各種讓用戶的日常生活更便利的應用程式，而這些應用程式也仰賴其介面設計，讓使用者隨時可以執行所需要的服務。但這有兩個缺點，第一，每個應用程式多少都有其特殊的介面設計與使用方式，對新手或是年長的使用者來說，要熟悉各種不同的使用介面是非常吃力的；第二，許多功能隱藏於某個頁面底下，需要多個「點選」操作，導致不少情況無法方便使用，如煮飯中、開車中等，都是使用者無法騰出雙手使用服務的情境。因此語音的智慧助理是一個未來資訊科技發展相當重要的研究議題，也是讓人們更自然的應用口語對答方式來使用服務，進一步讓日常更加的便利。

現今已有如 Google Home, Amazon Echo 等產品，可以讓使用者透過語音就可以打電話或是設定鬧鐘。但這些智慧助理尚有相當大的進步空間，例如提供較複雜的功能如寄 email、查看行程等。因此我們希望能夠開發新的語意分析模組，來增進智慧助理的功能。

現有的智慧助理通常利用關鍵字比對來猜測使用者所想要使用的功能，再找出對應功能的參數傳入目標功能。另一種方式則是先利用廣泛的命名實體辨識模型判斷出語句中重要資訊的類型後，再利用分類器判斷出目標功能，以達到使用者的目的。

命名實體辨識技術(Named Entity Recognition)就是指把目標類型的詞語從句子中標記出來，最常使用的方法即是序列標記，例如利用條件隨機域(CRF)。近年來隨著深度學習的發展，有不少論文結合遞 RNN 或相關神經網路架構去做模型的正確率比較高，在下一個部分會仔細討論。然而命名實體辨識模型的訓練是需要不少已標記資料的，而網路上暫時找不到相關的公開資料以供使用。在本研究中我們預計參考 WIDM 實驗室發展的 Web NER Model Generation Tool，利用遠距監督方式(Distant Learning)去完成訓練資料的準備。

另外一個問題則是意圖或功能的判定。一般說來，分類器的部分如果有足夠的資料來做訓練是相對十分簡單的，有許多類神經網路的架構都是為了解決分類問題而提出的。但使用的輸入與實際目的，這是網路上面難以找到的資訊，也因此在此嘗試使用有限的資訊並利用增強式學習來訓練出分類器。

這次研究目的是讓使用者能夠以對答的方式達成 Gmail 與 Google 行事曆相關的功能操作，並簡化過程以便延伸發展。未來再結合其他程式的 API，增進智慧助理的功能。

（三）文獻回顧與探討

3.1 分類器

對於任一應用程式而言，其提供的功能都是有限的，因此要判斷出使用者的用意，或者說該使用哪一個功能就需要經由分類模型去做判斷了。常見的分類模型包括 SVM(Support Vector Machine)、ANN(Artificial Neural Network)等。在此暫時決定使用 MLP(Mutilayer Perceptron)是因為需要分辨的功能非常廣泛，而 SVM 可能就需要多個模型(C_2^n)才能完成分類。分類內容也有可能並不是那麼單純的能夠理解成空間中的兩能分割區域，這時在使用 SVM 之前就需要經過正確與有效的投影才能有成果。另一個問題則是在新增分類類別時，MLP 可以直接在其隱藏層與輸出層增加類神經元，重置訓練係數，再次繼續訓練就可以更新模型，而 SVM 則需要新增已存在的分類類別數量，運算量相對大。

3.2 命名實體辨識

命名實體辨識(NER, Name Entity Recognition)，是一個自然語言處理的技術，主要目標是辨識出特定類型的實體。實體(Entity)的意義包含抽象的概念如時間、日期，也包含了具體的事物。如，小明明天要去公園。「小明」是人名，是具體的人物，屬於名為「人物」的實體，「明天」是時間，是特定的抽象概念，屬於名為「時間」的實體，而「公園」屬於「地點」。

CRF(Conditional Random Field)是一個常用於 NER 的機率統計模型，而隨著神經網路架構的發展與使用，有人容入了相關的概念並做出更好的成果。其中 RNN(Recurrent Neural Network)是一個如同有了記憶元件的神經網路架構，在自然語言的處理中，前後文的一同處理是非常重要的；LSTM(Long Short-Term Memory)是一個有增強功能的記憶元件的神經網路架構，在自然語言處理中大受歡迎的模型之一；而 CNN(Convolutional Neural Networks)雖然通常應用於圖像的處理與辨識，因為其模型能夠找出部分重要資訊，也有人應用於自然語言的處理。因此在此討論其相關演算法 CRF, RNN, LSTM 與 CNN。

- 條件隨機域

條件隨機域(CRF, Conditional Random Field)，是一個無向性的機率模型。每個頂點代表其中一個變數，而點間(線上)的數值則代表兩個變數所代表意義的關係遠近，近的話就代表兩者屬性與關聯都比較相近，可以視為同一類別。John Lafferty[1]於 2001 年發表了一篇使用 CRF 於 NER 的標記。利用這機率模型來判定某特定類型資料的開始與結束，並加以標記。

- 遞歸神經網路

遞歸神經網路(RNN, Recurrent Neural Network)，是一個輸出可能成為同一層輸入的類神經網路。優勢是可以找出一段較長文字之間的關係，其架構如圖 3.2，因為每個字的輸入都與前面所有字的輸入有關聯，也因此可以紀錄其中的關係。Kaisheng Yao[2]等人於 2013 年利用 RNN 去做 NER，部分成果如圖 1，發現利用適當的 RNN 架構出來的正確率(F_1)比 CRF 模型出來的還要高至少 2%。雖然 Vedran Vukotic[3]發表了一篇單純的 RNN 無法取代 CRF，但 Kaisheng Yao[4]同批研究人員又於 2014 年發表了一個先把輸入經過 RNN 模型後，再利用 RNN 的輸出送入 CRF 做判定，發現其正確率(F_1)比單純的 CRF 模型還要高 3%、RNN 模型高 1%。

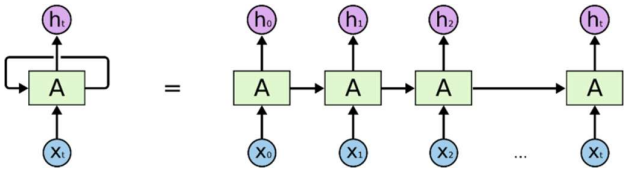

圖 1 RNN 架構圖

- 長短期記憶

長短期記憶(LSTM, Long Short Term Memory)是 Sepp Hochreiter[5]於 1997 年提出的一個類神經網路架構，如圖 2 利用三個 sigmoid(閘門)分別決定是否要記得、是否要忘記、與是否要輸出。「記得」的部分有點像是 RNN 資訊傳向前一層神經網路的機制，「忘記」則是判定所記得的資訊是否還有利用價值，而「輸出」的機制則是判定時機是否適合做出反應。其使用於語言處裡是由 Martin Sundermeyer[6]提出的，其用處與上一段的 RNN 有點相似，但是個增強挑選記憶

的版本。

因為 LSTM 能夠判斷出長斷句的關聯，用處與 RNN 相似，所以 Zhiheng Huang[7] 在 2015 利用 LSTM+CRF,BI-LSTM+CRF 來做 NER，而其正確率比其他演算法都要來得高。如圖 3，BI-LSTM 是指 bidirectional LSTM，與 LSTM 相似，運算量為兩倍，其中多的那一倍就是順序反過來運算。

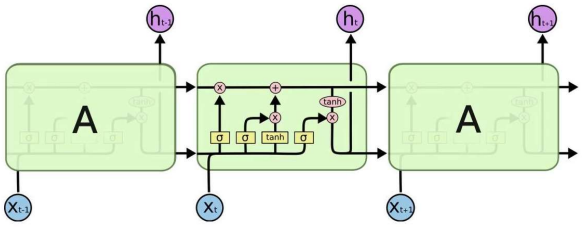

圖 2 LSTM 架構圖

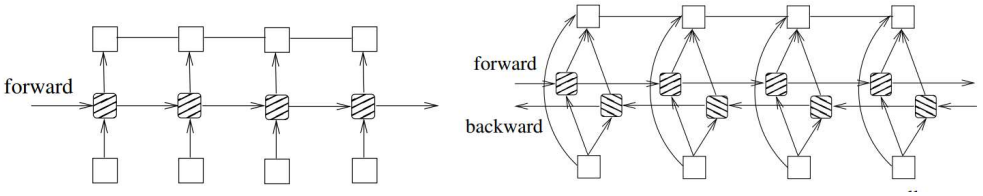

圖 3 左 LSTM-CRF、右 BI-LSTM-CRF，最底層為輸入、斜線為 LSTM、而最上層為 CRF

- 卷積神經網路

卷積神經網路(CNN, Convolutional Neural Network)是基本上利用多次的 Convolution 與 pooling 的步驟，之後再丟入神經網路做分類，通常是應用於圖片的處裡。Convolution 是把部分原輸入切出來並送出，這樣資料量就會大增許多，因此會接著做 pooling。Pooling 就是減少資訊量的意思，其中必須挑出重要的部分並保留，舉常見的 maxpooling 微例，會找尋範圍內最大的數值並以此當成那範圍的代表。圖片裡通常關鍵的只有一小部分，而我們所需要做的是找出那個地方並加以做判定。

近期開始有人應用於自然語言處裡(NLP)，如 Nal Kalchbrenner[8]於 2014 提出的利用 CNN 來做句型詞性分類，能夠代替部分剖析樹使用、Yoon Kim[9]於 2014 提出的利用 CNN 去做句子的分類，如圖 4。而其中較有趣的應用是英文單字的字根分析。一個英文單字可能只有一部份是有意義的，而 Xuezhe Ma[10]於 2016 利用這個概念結合了 BI-LSTM 與 CRF 做 NER。他先使用 CNN 做一層字詞代表的處理，在經由 BI-LSTM+CRF 設計出了一個 LSTM-CNNs-CRF 模型。

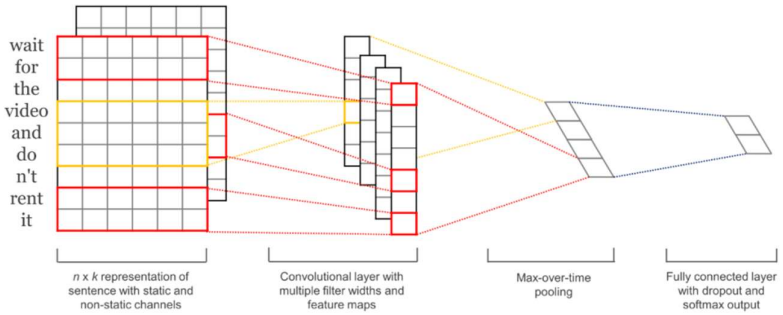

圖 4 利用 CNN 架構達成句子分類

3.3 命名實體辨識的前處理

在抓取句中重要資訊的類型之前，字詞在輸入之前的處理也是十分重要的。如果兩字相關性非常大，那麼他們的輸入也就應該要比較相近，也因此需要使用詞嵌入概念。

- 詞嵌入

詞嵌入是讓所有的字投影到向量座標中，而意義與類型較相近的字應該要在座標系統中較靠近的位置。這樣的話，這個字所投影到的向量就可以代表這個字。轉換方式通常是來自於大量的統計，而因為純粹做統計會導致運算負荷過大，所以會使用機率分布、類神經網路等來達到與類似於處理所有資料的結果。

Tomas Mikolov[11]等人於 2013 年提出了連續詞袋模型(CBOW, Continuous Bag of Words)與跳躍式模型(SG, Skip-Gram)，都是常用的詞嵌入模型，如圖 5。CBOW 模型利用某一關鍵詞前後文的字詞(圖中示意為前後各兩個字)當成訓練資料，期望輸出為關鍵詞，再做倒傳遞修正隱藏層的權重，目標是能夠利用前後文猜出中間的字詞，實驗結果為較適合常出現的字詞分析。而 Skip-Gram 模型則剛好相反，把某關鍵詞的的向量當成輸入，期望輸出最有可能的前後字詞，目標是能夠預測關鍵詞的前後文，架構理念不同導致就算是稀有字詞也十分精確，但速度相對慢。兩模型的第二層都是隱藏層，而其訓練後的權重值保有字詞的部分意義與關係，也因此可以拿來代替原字詞來做處理，也就是詞向量(Word2Vec)。

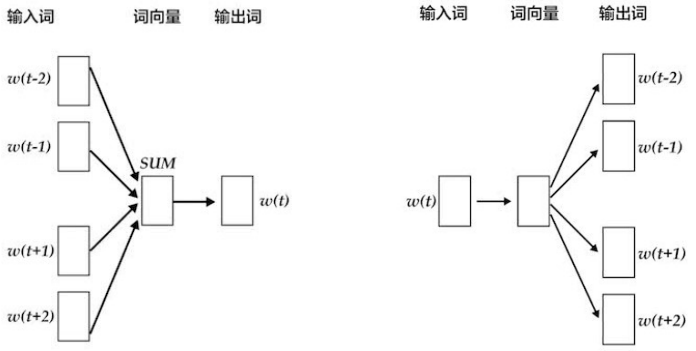

圖 5 左 CBOW、右 Skip-Gram

（四）研究方法及步驟

第一個部分為詞向量化的工具使用，嘗試簡易的 NER 演算法看是否能訓練成功，再利用不同的前處理看是否能有所改善。

第二個部分則是 NER 的主體。NER 的模型雖然也是有十分多種，但更困難的地方在於其訓練資料的生成。在網路上沒有所需標記類型的中文資料庫，而不同的功能更是需要各種類型的標記，也因此這是一個困難的步驟。

WIDM 實驗室發展的 Web NER Model Generation Tool 能夠利用種子實體，自動標記訓練資料，再利用 CRF 序列標記 package 訓練出一個標記某特定類型資訊的 NER 模型。其概念是利用已知的實體從網路上找出相關的句子，再應用自動標記做為監督式學習所需的訓練資料。

第三個部分是利用類神經網路做判定是屬於哪一個功能需要做反應。訓練資料更為難以準備，要以一些已知的條件從網路上找相似的語句，加以做監督式學習。如果訓練資料不足夠，則嘗試利用增強式學習訓練出分類器模型。

4.1 研究步驟

利用網路上的論壇，爬出其中較完整的語句當成測試資料，並經判斷出哪一個詞嵌入方法最適合。這步驟非常重要，因為如果詞向量沒有很準確的代表字詞意義的話，後面的步驟都很有可能會有比較差的表現。

利用實驗室(WIDM)開發 PowerPOI 的工具，訓練出人名、時間、主旨等特定資訊類型的 CRF 模型，並嘗試加入 BI-LSTM 的架構比較其正確率。正確率的比較是基於所爬到的資料做正確率的計算，因此如果正確率出現異常則需要對爬文的工具做調整。

讓 NER 模型的訓練資料找尋與模型的調整能夠不斷的正常循環。嘗試使其避開極端測資的抓取，因為其字詞偏頗會導致語句的判定異常。

因為有多種類的 NER 模型需要訓練，因此要把上一個步驟改成自動化的執行，也就是去除人為調整的部分。

一句子的重點或使用者的用意是難以直接做分析的，也因此才要利用前面所做的 NER 抓取重要資訊與參數並送入類神經網路做合適功能的判斷。

資料量過少，或無法成功分類則利用強化式學習利用少筆的訓練資料做分類模型。

最後再把重要的資訊送入相關功能裡執行，主要以 Gmail 與 Google 行事曆為主。如果過程沒有明顯錯誤特徵，而最終還是沒有抓取到重點，則可以考慮使用 AM(attention mechanism)來對句子做前處裡。

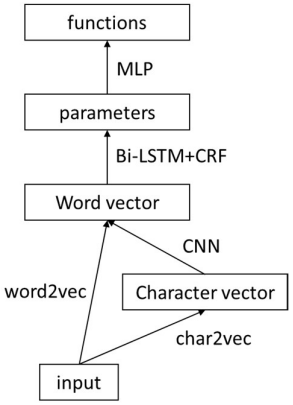

圖 6 模型示意圖

預計進度

2018/2019	七月	八月	九月	十月	十一月	十二月	一月	二月
詞嵌入的選定與完成	■							
NER 的測資抓取及模型訓練		■	■					
NER 訓練自動化				■				
功能的辨別					■	■		
與 google api 做結合							■	■

（五）預期結果

能夠從使用者的命令句中，判斷出重要的字詞與分析出其想要使用的功能。以 gmail、google calendar 的功能為展示，例如使用者輸入「幫我讀昨天小明寄來的信」，應該就能分析出搜尋日期為「昨天」、寄信者為「小明」、並且使用的功能為「讀」。

因為這項技術是基於市面上的多項智慧家庭商品做的改進，因此可以使用已知目的的語句來做比對，理論上無論是分辨用意能力還是字詞類型標記都應該會比較準確。

（六）參考文獻

[1] Lafferty J., McCallum A., Pereira F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data, Proc. 18th International Conf. on Machine Learning. Morgan Kaufmann: 282–289, 2001

[2] Yao K., Zweig G., Hwang MY., Shi Y, Yu D. Recurrent Neural Networks for Language Understanding Interspeech, 2013 - isca-speech.org
[3] Vukotic V., Raymond C., Gravier G. Is it time to switch to word embedding and recurrent neural networks for spoken language understanding?, InterSpeech, 2015
[4] Yao K., Peng B., Zweig G., Yu D., Li X.… RECURRENT CONDITIONAL RANDOM FIELD FOR LANGUAGE UNDERSTANDING, Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, 2014
[5] Hochreiter S., Schmidhuber J. Long short-term memory Neural computation, 1997
[6] Sundermeyer M., Schlüter R., Ney H. LSTM neural networks for language modeling, In INTERSPEECH-2012, 194-197, 2012
[7] Huang Z., Xu W., Yu K. Bidirectional LSTM-CRF models for sequence tagging, arXiv preprint arXiv:1508.01991, 2015
[8] Kalchbrenner N., Grefenstette E., Blunsom P. A convolutional neural network for modelling sentences, arXiv preprint arXiv:1404.2188, 2014.
[9] Kim Y. Convolutional neural networks for sentence classification, arXiv preprint arXiv :1408.5882, 2014.
[10] Ma X., Hovy E. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF, arXiv preprint arXiv:1603.01354, 2016
[11] Mikolov T., Sutskever I., Chen K., Corrado GS.… Distributed Representations of Words and Phrases and their Compositionality, Advances in Neural Information Processing Systems 26 (NIPS 2013), 2013

（七）需要指導教授指導內容

- 問題方向掌握
- 研究計畫實現方式
- 演算法與模型的理解與應用
- 報告書撰寫技巧