

長者日常對話與認知執行功能關係探討：使用詞向量與迴歸模型

Discussion on the relationship between elders' daily conversations and cognitive executive function: using word vectors and regression models

Ming-Hsiang Su, Yu-An Ko and Man-Ying Wang
Soochow University, Taiwan
huntfox.su,kelen850408, mywang.scu@gmail.com

摘要

隨著國人平均壽命攀升，老年人口的健康照護問題也更加多元，而且長期照護的需求也日漸增加，因此如何幫助高齡長者擁有良好生活品質與尊嚴之維持，是我們需要思考的。本研究欲藉由深度模型進一步探討正常老化者自然語言的特性。首先以焦點團體的方式收集資料，使長者在過程中自然地與其他參與者互動。接著透過詞向量模型與迴歸模型建立基於對話資料的EF預測模型，幫助了解EF的退化軌跡與建立預警。

Abstract

As the average life expectancy of Chinese people rises, the health care problems of the elderly are becoming more diverse, and the demand for long-term care is also increasing. Therefore, how to help the elderly have a good quality of life and maintain their dignity is what we need to think about. This research intends to explore the characteristics of natural language of normal aging people through a deep model. First, we collect information through focus groups so that the elders can naturally interact with other participants in the process. Then, through the word vector model and regression model, an executive function prediction model based on dialogue data is established to help understand the degradation trajectory of executive function and establish an early warning.

關鍵字：詞向量、迴歸模型、認知執行功能
Keywords: word vector, regression model, cognitive executive function

1 Introduction

國家發展委員會指出台灣在 2018 年邁入高齡的社會，而在 2020 年時，超高齡人口已經占 10.3%，進而推估在 2025 年台灣將邁入超高齡社會（國家發展委員會，2020）。隨著台灣人民的壽命延長，對於銀髮族的健康照護問題也逐漸受到重視，對於銀髮族長期居家照護的需求也日漸增加。如何幫助銀髮族不僅在醫療層面受到照護，也能擁有優質生活品質與生命尊嚴，是一重要課題。

學者們的研究指出阿茲海默症 (Alzheimer's Disease, AD) 和輕度認知障礙 (Mild Cognitive Impairment, MCI) 患者在在疾病前驅期，它們的語言處理過程便出現缺陷 (Taler & Phillips, 2008)。如何識別早期跡象和症狀的能力對於干預措施的發展相當重要。藉由認知和思維障礙在說話方式和說話內容中的表現，我們可以藉由語言分析去洞悉神經功能，進而使用自動化或半自動化的語音和語言分析方法進行更深入的研究，以找出患者的語言與聲音特徵。

Calzà 等人 (Calzà et al., 2021) 建立一個用於早期診斷和篩查的自動系統，旨在量化和描述由於認知能力下降導致的語言特徵變化。他們招募 48 名健康對照者和 48 名受損受試者。每個受試者接受了簡短的神經心理學篩查，並通過三個啟發任務收集了半自發語音產生的樣本。他們利用支持向量 (SVC) 和隨機森林分類器 (RFC) 來區分健康對照和 MCI 受試者。他們的實驗結果顯示他們所提出的方法是識別癡呆臨床前階段的一種很有前途的方法。Fraser 等人 (Fraser et al., 2019) 考慮一種級聯方法來組合來自多種語言任務的數據。26 名 MCI 參與者和 29 名健康對照組完成圖片描述、默讀和大聲朗讀三種語言任務。研究成

果指出最好的分類結果是通過組合任務級別的數據來實現的 (AUC=0.88, 準確度=0.83)。這優於基於神經心理學測試分數 (AUC = 0.75, 準確度 = 0.65) 以及多模態分類的“早期融合”方法 (AUC = 0.79, 準確度 = 0.70) 訓練的分類器。Wang 等人 (Wang et al., 2019) 的研究旨在檢驗情景圖片描述和自發自我介紹任務中的相關語音生成是否可用於區分具有 MCI 心理測量證據的個體和認知完整的個體。他們的研究揭示 MCI 參與者語義內容和句法複雜性的線性下降趨勢, 以及明顯更大的不流暢跡象和語音生成減少。這些發現擴展了文獻中的報導, 並對疑似 MCI 的篩查和診斷具有重要意義。

Fraser 等人 (Fraser et al., 2016) 展示從圖片描述任務引發的簡短敘述樣本中自動識別阿爾茨海默病的最先進的準確性, 並通過統計因素分析揭示突出的語言因素。藉由 DementiaBank 語料庫訓練分類器, 用以區分患有 AD 的參與者和健康對照組。為了檢查 AD 中語言障礙的異質性程度, 他們對這些言語和語言的測量進行了探索性因素分析, 並提供了對結果因素的解釋。在圖片描述任務中根據他們語言的短樣本, 區分患有 AD 的個體和沒有 AD 的個體, 獲得了超過 81% 的最先進的分類準確率。他們認為機器學習和語言分析在疑似 AD 的評估和聚類中非常有用。Asgari 等人 (Asgari et al., 2017) 認為輕度認知障礙 (MCI) 的指標可能存在於老年人口語內容中, 並且有助於區分 MCI 患者和認知完好的患者。他們對參與臨床試驗的 MCI 參與者和具有完整認知的參與者的口語進行了語言分析, 分類準確率為 84%, 遠高於 60% 的機會。Polsinelli 等人 (Polsinelli et al., 2020) 調查了執行功能 (EF) 的可變性如何體現在日常用詞模式中。他們使用語言查詢和字數統計對捕獲的話語的逐字記錄進行了文本分析。EF 使用經過驗證的測試電池組進行評估, 測量 WM、移位和抑制控制。他們發現較高的整體 EF, 尤其是工作記憶, 與更多的冠詞和介詞、更長的單詞和更多的數字語言相關。

本研究期望探討台灣正常銀髮族的語言與認知執行功能之間的關係, 藉由詞向量與迴歸模型建置, 搭配彩色路徑描繪測驗 (CTT) 檢測銀髮族的執行功能, 並且建立預測模型。

2 Dataset

本研究使用團體聊天方式來引導銀髮族語言使用的自然狀態。參與者為 53-74 歲的退休銀髮族 (36 位女性與 15 位男性), 參與者平均教育年數為 12.5 年。每場團體聊天人數為 2-5 人, 他們會討論日常生活經常從事的活動, 並且會針對有高共通性的議題進行深入討論, 藉以了解其執行步驟、並接受彩色路徑描繪測驗, 以測量執行功能與日常生活活動調查。

3 Word Embedding

本研究將參與者的語音資料轉成文字檔, 並使用 CKIPAPI 進行斷詞處理, CKIP 為中研院中文語言小組開發, 為中文自然語言處理提供相關的研究資料, 包括中文詞知識庫、語料等。接著本研究使用 Word2vec 進行詞向量特徵抽取。Word2vec 是 Google 所提供的一個用來產生詞向量的模型, 透過訓練可以將文本中的內容簡化成 n 維向量空間中的向量運算, 此向量空間上的相似度可以用來表示文本語意上的相似度。訓練 Word2vec 方式可分為兩種: 連續詞袋模型 (continuous bag-of-words, CBOW) 及 Skip-Gram 模型。在 CBOW 的方法裡, 訓練目標是用一個詞的鄰近詞去預測該詞的機率; 而 Skip-Gram 則是跟 CBOW 相反, 訓練目標是用一個詞去預測該詞鄰近詞的機率。另外為了比較不同的詞向量對於系統效能的差異, 本研究另外也使用 Doc2Vec 模型和 FASTText 模型進行詞向量模型訓練。

4 Regression Model

數據收集技術的進步大大增加了在科學與商業領域中預測變數的數量, 此時就需要模型選擇的方法來找出重要的預測變數, 傳統的模型選擇方法, 如逐步 (stepwise) 和向前 (forward) 雖簡單, 但被證實產生的模型準確性較低, 特別是當預測變數的數量很大時 (Desboulets, 2018)。且變數大於觀測值個數時也可能造成過度適配 (overfitting) 的問題, 因在樣本內的平方誤差和會越低, 此時可使用正規化 (regularization)。

Lasso 是最小絕對收縮選擇和算子的簡稱, 是一種採用了 L1 正則化 (L1-regularization) 的線性迴歸方法, 採用了 L1 正則會使得部分學習到的特徵權值為 0, 從而達到稀疏化和特徵

選擇的目的。在考慮一般的線性迴歸問題，給定 n 個數據樣本點 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ ，其中每個 $x_i \in R^d$ 是一個 d 維的向量，即每個觀測到的數據點是由 d 個變量的值組成的，每個 $y_i \in R$ 是一個實值。現在要做的是根據觀察到的數據點，尋找到一個映射 $f: R^d \rightarrow R$ ，使得誤差平方和最小，優化目標為：

$$\beta^*, \beta_0^* = \operatorname{argmin}_{\beta, \beta_0} \frac{1}{n} \sum_{i=0}^n (y_i - \beta^T x_i - \beta_0)^2 \quad (1)$$

$$\beta^* = \operatorname{argmin}_{\beta} \frac{1}{n} \sum_{i=0}^n ((y_i - \bar{y}) - \beta^T (x_i - \bar{x}))^2 \quad (2)$$

$$\beta^* = \operatorname{argmin}_{\beta} \frac{1}{n} \|y - X\beta\|_2^2$$

由於有 d 個變量，所以稱之為 Multiple Linear Regression。一般來說，迴歸問題是一個函數擬合的過程，那麼希望模型不要太複雜，否則很容易發生過擬合現象，所以要加入正則化項，而不同的正則化項就產生了不同的迴歸方法，其中以 Ridge Regression 和 Lasso 最為經典，前者是加入了 L2 正則化項，後者加入的是 L1 正則化項。本研究將使用 Multiple Linear Regression、Ridge Regression 和 Lasso 進行實驗。

5 Experiment Settings and Results

本研究將斷詞後的資料集分別進行 Doc2Vec，FASTText 和 Word2Vec 詞向量模型訓練。在 Doc2Vec 詞向量模型中，文本向量維度為 300 維。而 Multiple Linear Regression 的實驗結果為訓練資料的 MSE: 1.20e-09；測試資料的 MSE: 1951.29；訓練資料的 R-squared: 1.00；測試資料的 R-squared: -1.94。另外我們進行 Multiple Linear Regression、Ridge Regression 和 Lasso 模型比較，如 Figure 1 所示，在 300 維詞向量中，我們可以發現 Ridge Regression 模型有較佳的結果。其中 Lasso 模型中， $\alpha = 1$ 時，僅使用了 28 個特徵（係數的非零值）； $\alpha = 0.01$ 時，僅使用了 37 個特徵；而 $\alpha = 0.00001$ 時，則使用了 300 個特徵。

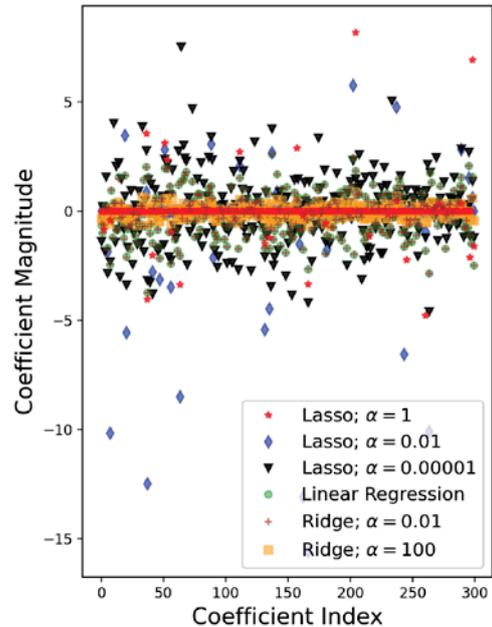


Figure 1: 不同 Regression 模型效能比較

在 FASTText 詞向量模型中，詞向量維度為 50 維。為了避免每個人的文本向量長度不同，所以我們採取每個人的文本只取 296 個詞，而整個文本向量維度為 50×296 維。而 Multiple Linear Regression 的實驗結果為訓練資料的 MSE: 2.82e-09；測試資料的 MSE: 702.15；訓練資料的 R-squared: 1.00；測試資料的 R-squared: -0.06。另外我們進行 Multiple Linear Regression、Ridge Regression 和 Lasso 模型比較，如 Figure 2 所示，在 300 維詞向量中，我們可以發現 Ridge Regression 模型有較佳的結果。其中 Lasso 模型中， $\alpha = 1$ 時，僅使用了 29 個特徵（係數的非零值）； $\alpha = 0.01$ 時，僅使用了 41 個特徵；而 $\alpha = 0.00001$ 時，則使用了 346 個特徵。

在 Word2Vec 詞向量模型中，詞向量維度為 50 維。為了避免每個人的文本向量長度不同，所以我們採取每個人的文本只取 296 個詞，而整個文本向量維度為 50×296 維。而 Multiple Linear Regression 的實驗結果為訓練資料的 MSE: 2.62e-09；測試資料的 MSE: 809.47；訓練資料的 R-squared: 1.00；測試資料的 R-squared: -0.22。另外我們進行 Multiple Linear Regression、Ridge Regression 和 Lasso 模型比較，如 Figure 3 所示，在 300 維詞向量中，我們可以發現 Ridge Regression 模型有較佳的結果。其中 Lasso 模型中， $\alpha = 1$ 時，僅使用了 31 個特徵（係數的非零值）； $\alpha = 0.01$ 時，僅

使用了 49 個特徵；而 $\alpha = 0.00001$ 時，則使用了 532 個特徵。

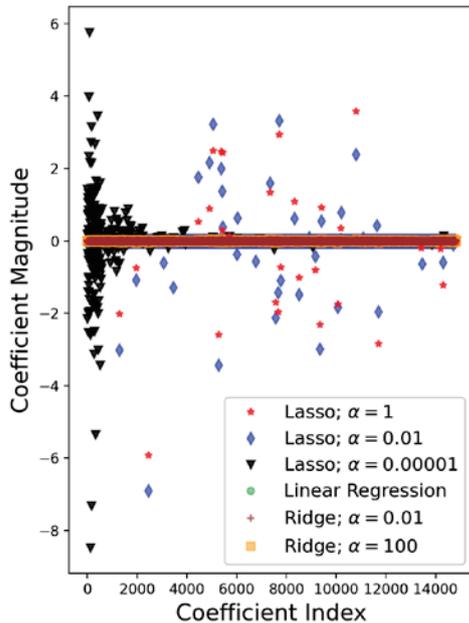


Figure 2: 不同 Regression 模型 FASTText 文本
向量效能比較

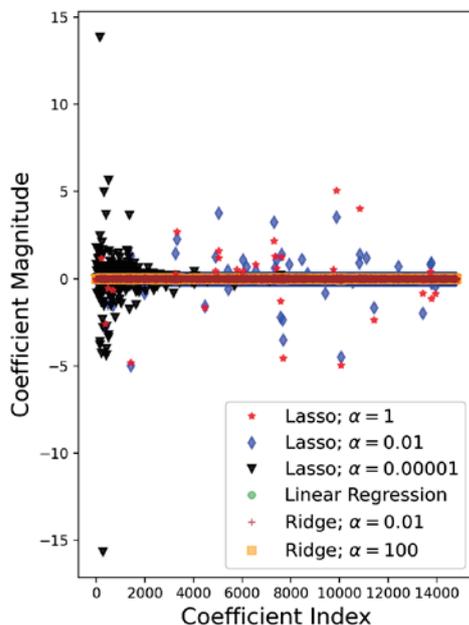


Figure 3: 不同 Regression 模型 Word2Vec 文本
向量效能比較

6 Conclusion

本研究欲藉由深度模型進一步探討正常老化者自然語言的特性。首先以焦點團體的方式收集資料，使長者在過程中自然地與其他參與者互動。接著透過詞向量模型與迴歸模型

建立基於對話資料的 EF 預測模型，幫助了解 EF 的退化軌跡與建立預警。實驗結果顯示，當詞向量使用 Word2Vec 模型而迴歸模型使用 Ridge 模型有較佳的預測效能。

在未來的研究中，本研究會持續增加資料集數量，另外也希望能加入 BERT 和 LSTM 等深度學習模型，以期能得到更好的結果。

References

- [1] 國家發展委員會. 2020. 中華民國人口推估 (2020 至 2070 年) 報告. 取自國家發展委員會, 人口推估查詢系統網址 <https://popproj.ndc.gov.tw/download.aspx?uid=70&pid=70>
- [2] Taler, Vanessa, and Natalie A. Phillips. 2008. Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review. *Journal of clinical and experimental neuropsychology*, 30(5): 501-556. <https://doi.org/10.1080/13803390701550128>.
- [3] Laura Calzà, Gloria Gagliardi, Rema Rossini Favretti, and Fabio Tamburini. 2021. Linguistic features and automatic classifiers for identifying mild cognitive impairment and dementia. *Computer Speech & Language*, 65: 101113. <https://doi.org/10.1016/j.csl.2020.101113>.
- [4] Kathleen C. Fraser, Kristina Lundholm Fors, Marie Eckerström, Fredrik Öhman, and Dimitrios Kokkinakis. 2019. Predicting MCI Status From Multimodal Language Data Using Cascaded Classifiers. *Frontiers in aging neuroscience*, 11, 205. <https://doi.org/10.3389/fnagi.2019.00205>.
- [5] Tianqi Wang, Chongyuan Lian, Jingshen Pan, Quanlei Yan, Feiqi Zhu, Manwa L. Ng, Lan Wang, Nan Yan. 2019. Towards the Speech Features of Mild Cognitive Impairment: Universal Evidence from Structured and Unstructured Connected Speech of Chinese. In *Proceedings of INTERSPEECH*. <https://doi.org/10.21437/Interspeech.2019-2414>.
- [6] Fraser, Kathleen C., Jed A. Meltzer, and Frank Rudzicz. 2016. Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, 49(2): 407-422. <https://doi.org/10.3233/JAD-150520>.
- [7] Asgari, Meysam, Jeffrey Kaye, and Hiroko Dodge. 2017. Predicting mild cognitive impairment from spontaneous spoken utterances. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 3(2): 219-228. <https://doi.org/10.1016/j.trci.2017.01.006>.
- [8] Angelina J. Polsinelli, Suzanne A. Moseley, Matthew D. Grilli, Elizabeth L. Glisky, and

Matthias R. Mehl. 2020. Natural, everyday language use provides a window into the integrity of older adults' executive functioning. *The Journals of Gerontology: Series B*, 75(9): e215-e220.

- [9] Desboulets, Loann David Denis. 2018. A Review on Variable Selection in Regression Analysis. *Econometrics*, 6(4): 45. <https://doi.org/10.3390/econometrics6040045>.