

透過語音特徵建構基於堆疊稀疏自編碼器演算法之婚姻治療中夫妻互動行為量表自動化評分系統
Automating Behavior Coding for Distressed Couples Interactions based on Stacked Sparse Autoencoder Framework using Speech-acoustic Feature

陳柏軒 (Po-Hsuan Chen)，李祈均 (Chi-Chun Lee)

國立清華大學電機工程學系 (Department of Electrical Engineering, National Tsing Hua University)

人與人之間交談互動，常透過語言傳達彼此的想法，並在這交談過程中得知雙方的行為反應。利用人為觀察來分析雙方行為反應，這種人為分析方式最早常應用在心理學和精神醫學方面 [2]。人為行為觀察已經相當的成功用於研究親密關係 [3][4]，因為夫妻的互動行為是影響親密關係程度的重要因素之一。然而使用人為觀察行為的方式長年存在根本問題，一方面太消耗時間，另一面也主觀。如果能透過電腦工程的方式來幫忙人為觀察將大大提升效率：即透過低層描述映射高層描述來預測與分析人類行為 [5]。這項研究領域是一個新興的領域。人類行為信號處理 (Behavioral Signal Processing, BSP) 目的在幫助連接信號處理技術與行為分析的跨領域學科，建立在傳統的信號處理研究，如語音識別，面手部追蹤等等。相關顯著 BSP 研究已發產於以人為中心的提取音頻，視頻信號，來分析高階人類行為甚或是情感方面 [6][7]。本論文利用 BSP 的基本思路應用在婚姻治療資料庫上面 [8]，婚姻治療資料庫會詳細說明在第二章。這個資料庫紀錄了夫妻在一段對話中談述了他們所選擇婚姻中的問題。觀察評分者在根據他們一段話的種種行為根據兩分精神醫療行為量表進行評分(例如：幽默行為、悲傷行為展現程度等等)。此篇論文延續上篇論文的研究內容來自動化分析夫妻一段對話的中個別行為分數[1]。一段語音經過訊號預處理，之後進行聲音特徵擷取(acoustic feature extraction)，再使用機器學習來作分類辨識，得到最後的準確率。其中，特徵擷取和機器學習的算法都會影響最後的準確率，思考如何改進這些影響因素，對整體準確率的提升是一大重要的課題，也是我們提出這篇論文重點。在特徵擷取方面，我們沿用三種低階語音特徵(Low Level Descriptors, LLDs)，語韻 (prosodic) LLDs、頻譜(spectrum) LLDs、和音質(voice quality) LLDs。切割三種說話者說話區間(speaker domain)，丈夫說話區間、太太說話區間、和不分人說話區間。再來對應各區間提取 20%語句，進行 7 種統計函數(functionals)，產生 2940 種低接原始特徵值。最後我們利用非監督深度學習的做法來降維找出相對關鍵的主要特徵值表現。深度學習在機器學習領域裡面是最近熱門的話題 [9]。深度學習可看成是一種資訊的表達方式，利用多層神經網絡，第一層輸入的數據學習之後，產生新的組合輸出，輸出值為第二層的輸入值，再經由學習產生新的輸出值，依此類推重覆把每層的資訊堆疊下去，透過這樣多層學習，可以得到對一個目標值好的特徵表示，相對準確率就能有所提升。至今存在多種深度學習框架如深度神經網路(DNN)、深度信念網路(DBN)和卷積神經網路(CNN)已被應用在語音 [10]、影像辨識 [11]和手寫識別 [12]等等。我們利用深度學習中的堆疊稀疏自編碼器(stacked sparse autoencoder, SSAE)，降低特徵值維度，提升特徵值整體相關性，最後利用簡單 LR 辨識行為分數高低。此初期研究結果顯示整體行為平均準確率 75%較之前研究使用 40479 維特徵值結合支持向量器 (support vector machine) [1]提升了 0.9%。

參考文獻

- [1] M. Black, A. Katsamanis, B. Baucom, C. Lee, A. Lammert, A. Christensen, P. Georgiou and S. Narayanan, 'Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features', *Speech Communication*, vol. 55, no. 1, pp. 1-21, 2013.
- [2] O'Brian, M., John, R.S., Margolin, G., Erel, O., 'Reliability and diagnostic efficacy of parent's reports regarding children's exposure to marital aggression', vol. 9, pp. 45-62, 1994
- [3] Karney, B.R., Bradbury, T.N., 'The longitudinal course of marital quality and stability: A review of theory, methods, and research. *Psychol' Bull*, vol. 118, pp. 3-34, 1995.
- [4] Gonzaga, G.C., Campos, B., Bradbury, 'Similarity, convergence, and relationship satisfaction in dating and married couples', *J. Personal. Soc. Psychol.*, vol. 93, pp. 34-48, 2007.
- [5] Schuller, B., Batliner, A., Seppi, D., Steidl, S., Vogt, T., Wagner, J., Devillers, L., Vidrascu, L., Amir, N., Kessous, 'The relevance of feature type for automatic classification of emotional user states: Low level descriptors and functionals', In: *Proc. Interspeech*, Antwerp, Belgium, pp. 2253-2256, 2007.
- [6] Burkhardt, F., Polzehl, T., Stegmann, J., Metze, F., Huber, R, 'Detecting real life anger', in: *Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing*, Taipei, Taiwan , pp.4761-4764, 2009.
- [7] Devillers, L., Campbell, N., ' Special issue of computer speech and language on affective speech in real-life interactions', *Comput. Speech Lang.*, vol. 25, pp. 1-3, 2011.
- [8] Christensen, A., Atkins, D.C., Yi, J., Baucom, D.H., George, W.H., 'Couple and individual adjustment for 2 years following a randomized clinical trial comparing traditional versus integrative behavioral couple therapy', *J. Consult. Clin. Psychol*, vol. 72, pp. 176-191, 2004.
- [9] G. Hinton, 'Reducing the Dimensionality of Data with Neural Networks', *Science*, vol. 313, no. 5786, pp. 504-507, 2006.
- [10] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath and B. Kingsbury, 'Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups', *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82-97, 2012.
- [11] E. Smirnov, D. Timoshenko and S. Andrianov, 'Comparison of Regularization Methods for ImageNet Classification with Deep Convolutional Neural Networks', *AASRI Procedia*, vol. 6, pp. 89-94, 2014
- [12] Y. Perwej and A. chaturvedi, 'Machine recognition of Hand written Characters using neural networks', *International Journal of Computer Applications*, vol. 14, no. 2, pp. 6- 9, 2011.