National Ilan University

國立宜蘭大學工程學刊第 13 期，30-42 頁，2019 年 12 月

Bulletin of College of Engineering National Ilan University No. 13, PP. 30-42, December 2019

DOI:10.6176/BCE.201912_(13).0003
Bulletin of College
of Engineering
http://engineering.niu.edu.tw

應用注意力機制於深度學習之行為辨識

陳文輝 [1] 張瑀軒 [2]

[1] 國立台北科技大學 教授 [2] 國立台北科技大學 研究生

摘要

近年來，人工神經網路取得突破性的發展，許多研究利用循環神經網路來對序列資料進行編碼，取得上下文的關聯。本研究以行為辨識為主軸，利用長短期記憶網路加上注意力機制，讓模型能聚焦在重要的輸入上，使得編碼後的上下文向量更具代表性。本文提出的方法和未使用注意力機制的模型相比，提升了 3.8%的辨識率。

關鍵詞: 注意力機制、循環神經網路、深度學習、行為辨識、行動裝置

*通訊作者 **Email：t105618011@ntut.edu.tw**

National Ilan University

國立宜蘭大學工程學刊第 13 期，30-42 頁，2019 年 12 月

Bulletin of College of Engineering National Ilan University No. 13, PP. 30-42, December 2019

DOI:10.6176/BCE.201912_(13).0003
Bulletin of College
of Engineering
http://engineering.niu.edu.tw

Human Activity Recognition Using Attention Mechanism on Deep Learning

Wen-Hui Chen[1], Yu-Xuan Zhang[2]

[1]Professor, National Taipei University of Technology, Graduate Institute of Automation Technology
[2]Graduate Student, National Taipei University of Technology, Graduate Institute of Automation Technology

ABSTRACT

In recent years, deep neural networks have made breakthroughs. Many studies use recurrent neural networks (RNNs) to encode sequence data and obtain context correlation. In this study, behavior recognition is the main axis, and the model can focus on the important input by using the long short-term memory (LSTM) and attention mechanism, which overcomes the problem of model performance degradation caused by too long sequence data, and makes the context vector more representative. Compared with the model without attention mechanism, the proposed method improves the recognition rate by 3.8% on average.

Keywords: Attention Mechanism, Recurrent Neural Network, Deep learning, Activity Recognition, Mobile Devices

*通訊作者 **Email：t105618011@ntut.edu.tw**

一、導論

由於微機電系統、無線通訊的蓬勃發展，使智慧型手機成為人們生活中不可或缺的科技裝置。近年來，深度學習的興起，許多應用都藉由深度學習取得了前所未有的成果，其中有卷積神經網路(convolutional neural network, CNN)、自動編碼器(auto-encoder, AE)和循環神經網路等等幾個經典的模型架構。

循環神經網路擅長對序列資料進行關聯分析，其神經元內部有個記憶位置，可以將過去輸入過的資訊儲存在其中。然而，在訓練循環神經網路的過程中，只是將一段序列資料依序輸入到循環神經網路中，對於前期輸入的資料來說，雖然記憶會被儲存在循環神經網路的內部記憶位置中，但是隨著時間拉長，記憶的效果會漸漸消失，導致重要特徵的記憶可能會因為時間步階的長度而被覆蓋過去。

本文利用注意力機制來克服的記憶消失的問題，利用注意力機制對序列資料進行偵查，得到每個時間點的貢獻分數，利用加權和得到最終的向量表示，使得重要的特徵不會因為序列資料的長度過長而造成記憶消失的問題，也讓模型有選擇性地對輸入資料進行學習。本文使用 UCI (University of California, Irvine)的公開資料集進行實驗，該資料集為 6 種人類日常活動的動作，資料由 3 軸加速度計和 3 軸角加速度計組成。實驗結果顯示，附有注意力機制的模型可以掌握序列資料中更多重要的特徵，也使得原始模型效能有所提升。

二、模型架構

2-1. 長短期記憶

循環神經網路是經典的深度學習架構之一，在翻譯系統[1]、語音辨識[2]和情緒分析[3]等等的應用上都有亮眼的成果。其內部有個記憶位置，可以將先前輸入過的資料記憶在該神經元裡，這種網路的架構，可以因應序列資料的些許差異，進而產生不同的輸出結果。長短期記憶[4]為循環神經網路的一種變化型，其內部多了閘門控制開關，對該時間點的資料進行選擇，相比於原始循環神經網路的結構，長短期記憶網路在訓練時可以減緩梯度消失的問題。

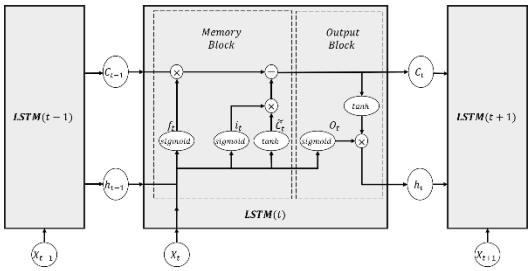

圖 1 長短期記憶網路結構圖

F為遺忘門(forget gate)，如式(1)，為決定哪些訊息需要被捨棄，其中σ為 Sigmoid 函數、X為輸入資料、h為隱藏層輸出。

$$F_t = \sigma\left(W_f X_t + U_f h_{t-1} b_f\right) \tag{1}$$

I為輸入門(input gate)，如式(2)，為決定哪些要加入當前的記憶進行更新。

$$I_t = \sigma(W_i X_t + U_i h_{t-1} + b_i) \tag{2}$$

$\tilde{C}$為當前新的狀態值，如式(3)，結合輸入門與遺忘門來對原本的內部記憶C進行更新，如式(4)。其中⊙為逐元素相乘符號

$$\widetilde{C_t} = \tanh(W_c X_t + U_c h_{t-1} + \quad b_c) \tag{3}$$

$$C_t = f_t \odot C_{t-1} + I_t \odot \widetilde{C_t} \tag{4}$$

最後經由輸出門O(output gate)來對更新後的內部記憶C進行選擇，成為長短期記憶的隱藏層輸出h，如式(5)、式(6)。

$$O_t = \sigma(W_o X_t + U_o h_{t-1} + b_o) \tag{5}$$

$$h_t = O_t \odot \tanh(\ C_t\ ) \tag{6}$$

2-2. 注意力機制

D. Bahdana 等人[5]為了解決文本翻譯的問題，提出了共同學習(jointly learning)的概念。圖 2 為編碼器-解碼器的模型架構，常用的編碼器-解碼器結構模型為 LSTM-LSTM，循環神經網路的特點是能將一段有前後相關的資料進行連結。長短期記憶網路在每個時間點的輸出狀態為h_i，輸入資料$X = [X_1, X_2, \dots, X_T]$依序被輸入至長短期記憶網路中，則可以得到每個時間點的輸出狀態$[h_1, h_2, \dots, h_T]$，而h_T為模型看過整段序列資料後所得到的狀態，我們將此h_T做為上下文向量C，作為輸入資料輸入至解碼端的長短期記憶網路進行預測目標語句。

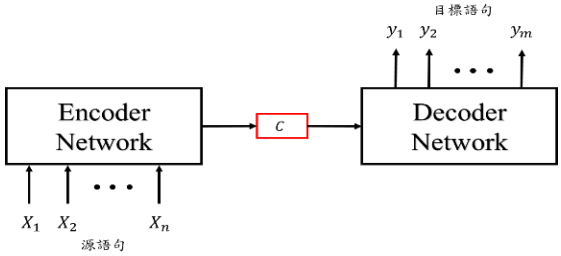

圖 2 編碼器-解碼器模型架構

但是這樣是不合理的，在一段冗長的句子中，各個字詞所代表的重要程度並不相同，而利用長短期記憶網路得到的上下文向量，並沒辦法透徹地應用在解碼器上。每個經過長短期記憶網路得到的字詞，應該要使用不同的上下文向量，如圖 3。

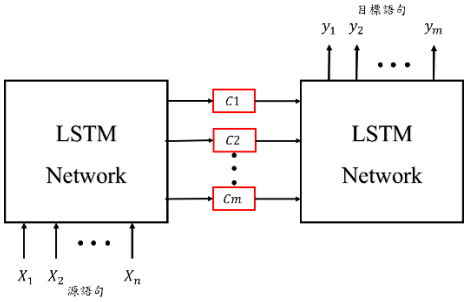

圖 3 編碼器-解碼器結構圖(注意力機制)

各個上下文向量C為由長短期記憶網路的各時間輸出狀態搭配不同權重 α 所組成，如式(7)。

$$C_i = \sum_i^n \alpha_{ij} h_j \tag{7}$$

每個時間點的權重 α 為相似分數 e 經由 *softmax* 函數求得，如式(8)。S為解碼器端的長短期記憶網路隱藏層輸出，若要從解碼器端求得下一個預測的字詞，必須由編碼器端當前字詞以前的輸入求得，如式(9)，最後再經過 f 函數得到相似度的分數，常見的 f 函數為內積。

$$\alpha_{ij} = \frac{e_{iq}}{\sum_{q=1}^{n} e_{iq}} \tag{8}$$

$$e_{ij} = f(S_{i-1}, h_j) \tag{9}$$

而這些上下文向量則代表著在各個目標輸出語句裡，由各個輸入字詞的集合所組成，對於生成特定的字詞，在輸入的字詞選擇上就應該要有不同比重，這就是注意力機制的前身。

本文將使用注意機制於長短期記憶網路，讓模型能夠對輸入的序列資料進行選擇性地學習，並給予每個時間點的狀態對應的權重值，最後使用加權和得到這段序列資料的特徵表示。如此一來，序列資料不會因為時間步階長度的問題，而導致儲存在內部的記憶被遺忘，最終降低模型的辨識效能。

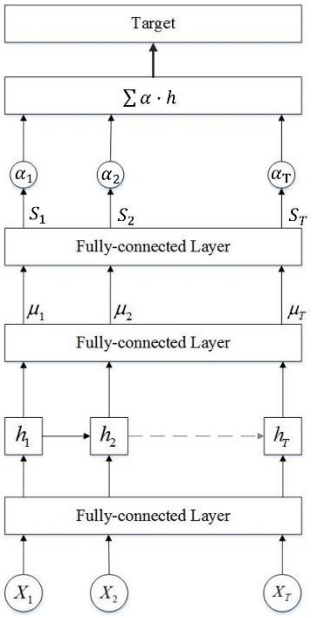

圖 4 注意力模型架構圖

圖 4 為本文使用的注意力模型架構，X為輸入的序列資料，其中X＝$[X_1, X_2, \ldots, X_i, \ldots, X_T]$。首先將$X_i$經過一層線性轉換層進行第一步的特徵抽取，為了避免在訓練網路時產生梯度消失的現象，選擇 ReLU 函數作為激勵函數。接著依序輸入至長短期記憶網路，然後將長短期記憶網路每個時間點的輸出h_i蒐集起來，利用一層全連接層做線性轉換得到u_i，如式(10)。

$$u_i = \text{ReLU}(W_a h_i + b_a) \tag{10}$$

下個步驟為再搭建一層全連接層，為了求出每個u_i所得到的分數，這層隱藏層的神經元只有一顆，如此一來，將u_i映射成S_i時，進行內積運算的權重為一個向量，此向量我們稱為u_c，如式(11)。u_c的功能為找出哪個時間點的輸入值是重要的，而S_i為u_c和每個u_i做內積後所得到的分數，再透過softmax函數將所有的分數壓縮在 0~1 之間，得到每個時間點的重要程度比例 α，如式(12)。

$$S_i = u_c \cdot u_i \tag{11}$$

$$\alpha_i = \frac{\exp(S_i)}{\sum_i^T \exp(S_i)} \tag{12}$$

接著將α_i及h_i進行加權和的運算，得到這段訊號的特徵表示，如式(13)。最後再連接一層全連接層來做動作預測的分類，本次使用的行為辨識數據集有 6 個動作，所以在 Target 這層的神經元設定為 6 個。

$$\text{Final} = \sum_i^T \alpha_i h_i \tag{13}$$

三、實驗結果與分析

本文使用 UCI 公開數據庫的行為辨識數據集進行實驗與分析，實驗中會使用兩個模型對資料進行訓練，第一個模型為長短期記憶網路，第二為增加注意力機制在長短期記憶網路上的注意力模型。實驗結果由 3 個部分組成，第一為比較兩個模型的效能，第二為將注意力模型的注意力分數視覺化，第三為本文提出的模型與相關文獻比較。

UCI 公開數據庫的行為辨識數據集[6]，其動作類型為日常生活活動(activity of daily living, ADL)。該研究者找來 30 位年齡由 19-48 歲的自願受測者，將智慧型手機放置在受測者腰包中，以 50Hz 的取樣頻率紀錄 3 軸直線加速度計與 3 軸角加速度計的資料，共蒐集 6 種行為動作，動態的為走路、上樓梯、下樓梯，靜態的為站立、坐著、平躺。本文將資料以時間步階 100 的長度進行切割，接著將資料以 8:2 分成訓練資料與測試資料，如表 1。

表 1 裁切後資料量

動作	訓練資料量(時間步階100)	測試資料量(時間步階100)
走路	920	242
上樓	857	222
下樓	780	202
坐著	988	224
站立	1064	253
平躺	1042	270
總計	5651	1413

本次實驗使用 F1 度量(F1-measure)作為模型效能評估的方式，如式(14)，F1 度量為精確率(Precision, P)和召回率(Recall, R)的調和均值，避免資料量不均勻的狀況影響辨識率的正確性。

$$\frac{2}{F1} = \frac{1}{P} + \frac{1}{R} \tag{14}$$

當模型訓練完成後，接著輸入測試資料進行預測，而預測結果會使用混淆矩陣進行觀察與分析，表 2 為混淆矩陣中會出現的四種情況:分別是正確正例(True Positive, TP)、錯誤正例(False Positive, FP)、錯誤負例(False Negative, FN)、正確負例(True Negative, TN)。精確率為所有被判定為正例的結果中，正確正例所涵蓋的比例，如式(15);召回率為所有被預測為正確的結果中，正確正例所涵蓋的比例，如式(16)

表 2 混淆矩陣

種類		預　測	
		正確	錯誤
真實	正例	正確正例(TP)	錯誤正例(FP)
	負例	錯誤負例(FN)	正確負例(TN)

$$P = \frac{TP}{TP+FP} \tag{15}$$

$$R = \frac{TP}{TP+FN} \tag{16}$$

表 3 為訓練完成的長短期記憶網路模型和注意力模型效能。原始的長短期記憶網路對於感測器資料已經取得不錯的辨識率，但是對於長久以前的輸入而已，其內部記憶可能無法將資訊傳遞至後面的網路。加上注意力機制後的模型則可以找出每個時間步階的輸入貢獻值，運用加權和將記憶傳遞至最後的特徵表示，因此模型不會因為時間步階太長的關係，導致記憶被遺忘。而注意力模型的效能比原來的長短期記憶模型提升了 3.8% 的辨識率。

表 3 模型效能比較表

模型	長短期記憶網路模型	注意力模型
F1-score	90.5%	94.3%

長短期記憶網路模型和注意力模型的預測結果，我們運用混淆矩陣來呈現，如表4-表5，表格中對角線的數字為測試資料預測正確的結果，但坐著和站立的動作有較多預測錯誤的情況，因為資料集在蒐集時受測者將手機放置腰包進行實驗，故腰包的方向為一致，使得模型有些誤判。接著將混淆矩陣以圖形的方式呈現，如圖5-圖6，對角線為預測正確的結果，所以這部分的圖形顏色較深。

表 4 長短期記憶網路模型混淆矩陣

種類		預測結果					
		走路	上樓	下樓	坐著	站立	平躺
真實結果	走路	222	1	19	0	0	0
	上樓	31	191	0	0	0	0
	下樓	20	0	182	0	0	0
	坐著	1	0	1	188	34	0
	站立	3	0	1	22	227	0
	平躺	0	0	0	0	0	270

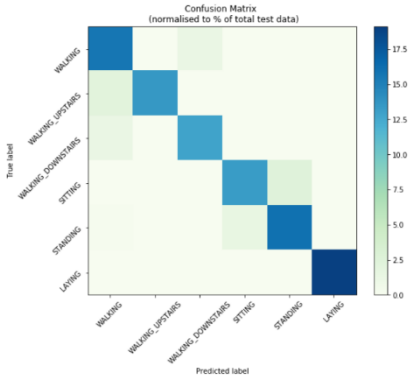

圖 5 長短期記憶網路模型混淆矩陣視覺化

表 5 注意力模型混淆矩陣

種類		預測結果					
		走路	上樓	下樓	坐著	站立	平躺
真實結果	走路	226	1	15	0	0	0
	上樓	2	218	2	2	0	0
	下樓	3	0	199	0	0	0
	坐著	0	1	4	194	25	0
	站立	2	1	0	24	226	0
	平躺	0	0	0	0	0	270

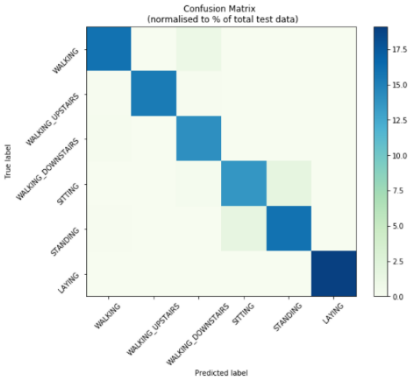

圖 6 注意力模型混淆矩陣視覺化

Y. Li 等人[7]認為以往抽取資料特徵的方式必須透過人工的方式一一陳列，然而這個世界的科學還有人類未及探索的，所以還是會存在疏漏的特徵項目。作者利用堆疊的稀疏自動編碼器(sparse auto-encoder, SAE)，讓模型自行抽取原始資料的特徵，利用更少的維度來表示輸入的資料，最後使用支持向量機進行分類。Y. Zhao 等人[8]使用殘差雙向長短期記憶網路(deep residual bidirectional LSTM)對感測器資料進行實驗，利用雙向循環神經網路的特點，將序列資料由前至後的方向看過一次，另一個方向為由後往前，讓神經網路在同一筆資料上得到兩個不同視野的特徵。作者在模型架構上使用了殘差結構，讓神經網路在訓練的過程中減緩梯度消失的現象。相關文獻比較表格如表 5。

表 5 相關文獻比較

參考文獻	方法	辨識率
Y. Li et al. [7]	Sparse Auto-Encoder	92.16%
Y. Zhao et al. [8]	Deep Residual Bidirectional LSTM	93.57%
本研究	Attention model	94.3%

輸入資料經過注意力機制後得到各個時間點的貢獻分數，圖 7 為將各個時間點的貢獻分數視覺化，圖 7 上為原始的六軸感測器訊號，圖 7 下的長條圖部分為感測器訊號被模型認為重要程度的權重值，從圖中可以看到，訊號在不同的時間點有相對應的關注程度。在訓練注意力模型的過程中，模型依照數據集中的資料進行分析，而在式(11)所提到的權重向量u_c，被訓練成可以針對不同動作的訊號，給予每個時間點的資料不同的權重值，對於輸入資料有策略性地選擇，和原始將資料視為相等重要程度的長短期記憶網路相比，注意力模型較有效地保留特徵，不會因為前期的輸入訊號過於久遠，而導致重要的特徵無法得到訓練，最終造成模型的誤判。

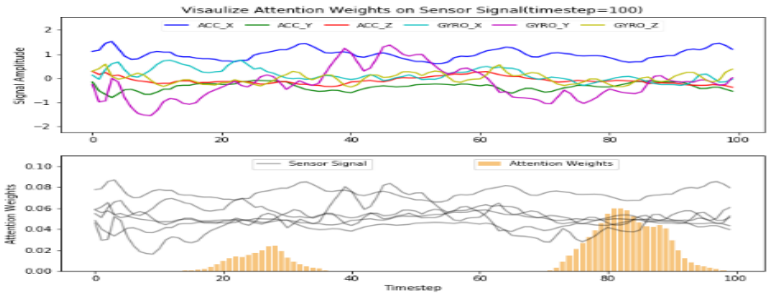

圖 7 視覺化注意力係數

四、結論與未來展望

本文運用注意力機制提升循環神經網路的效果，讓其模型更具強健性。傳統的循環神經網路架構為依序輸入訊號資料給模型，但是這並不是一個明智的方法，時序資料中，不是每筆資料都具有明確的特徵，如果直觀地將資料依序輸入進模型，通通交給模型進行取捨，則會讓模型限制在框架中無法真正地理解資料。透過注意力機制的輔助，讓模型選擇性的進行學習，不但讓模型聚焦在重要時刻的輸入，也可以減緩時間步階太長而造成的記憶消失問題。

附有注意力機制的模型使模型架構的強健性增加了不少，但缺點是會增加額外的計算資源，未來可以使用其他的深度學習架構來降降低模型的參數量，使整體模型能夠達到強健性高且參數量少的目標。

參考資料

[1]. M. Auli, M.Galley, C.Quirk and G. Zweig, “Joint Language and Translation Modeling with Recurrent Neural Networks,” In *EMNLP*, 2013.

[2]. Z. Meng, S. Watanabe, J. Hershey and H. Erdogan, “Deep Long Short-Term Memory Adaptive Beamforming Networks for Multichannel robust Speech Recognition,” *International Conference on Acoustics, Speech and Signal Processing, IEEE* , Mar. 2017, pp.271-275.

[3]. L. Li and Z.L. Zhang, “Emphasizing essential words for sentiment Classification Based on Recurrent Neural Networks,” *Journal of Computer Science and Technology 32*, July 2017, pp.785-795.

[4]. F. Gers, “Long Short-Term Memory in Recurrent Neural Networks,” Unpublished PhD dissertation, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2001.

[5]. D. Badhanau, K. Cho and Y. Dengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” arXiv preprint arXiv:1409.0473, 2014.

[6]. J. L. Reyes-Ortiz, L. Oneto, A. Sama, X. Parra and D. Anguita, “Transition-Aware Human Activity Recognition using Smartphones,” Neurocomputing 171, 2015, pp. 754-767.

[7]. Y. Li, D. Shi, Bo Ding and D. Liu, “Unsupervised Feature Learning for Human Activity Recognition Using Smartphone Sensors,” *Mining Intelligence and Knowledge Exploration*, 2014, pp. 99-107.

[8]. Z. Yu, Y. Rennong, C .Guillaume and G. Maoguo, “Deep Residual Bidir-LSTM for Human Activity Recognition using Wearable Sensore,” *CoRR*, vol. abs/1708.08989.