

植基於深度學習假新聞人工智慧偵測：台灣真實資料實作

Deep Learning Based Fake News AI Detection : Evidence From

Taiwan News Report

汪志堅 Chih-Chien Wang

國立台北大學資訊管理研究所

Graduate Institute of Information Management

National Taipei University

wangson@mail.ntpu.edu.tw

戴敏育 Min-Yuh Day

淡江大學資訊管理學系

Department of Information Management

Tamkang University

myday@mail.tku.edu.tw

胡林麓 Lin-Lung Hu

國立台北大學資訊管理研究所

Graduate Institute of Information Management

National Taipei University

s710636103@gm.ntpu.edu.tw

摘要

近年來透過網路媒介，使得假新聞得以快速流竄。而許多國家都受到假新聞嚴重影響，使得假新聞偵測變成一個重要課題。本研究蒐集兩個台灣闢謠網站資料，「新聞小幫手」以及「真的假的」。並運用深度學習的三種方法 Gated Recurrent Unit (GRU)、Long Short Term Memory (LSTM) 以及 Bidirectional Long Short Term Memory (BLSTM)，進行假新聞偵測。實驗結果發現，深度學習可以運用在台灣假新聞偵測當中，且 BLSTM 的方法最適合用於假新聞偵測。本研究實驗降低假新聞比例，模擬 25% 以及 5% 假新聞比例，讓研究樣本可以更趨近於真實情況。最後本研究使用交叉資料集測試，了解實務與學術上的差距。

Abstract

In recent years, because the Internet acts as a medium, fake news can be quickly spread. Many countries have been seriously affected by fake news. making fake news detection become an important issue. This study collects two Taiwan refute rumors sites. And use the three methods of deep learning for fake news detection., Gated Recurrent Unit (GRU), Long Short Term Memory (LSTM), and Bidirectional Long Short Term Memory (BLSTM). The experimental results show that deep learning can be used in Taiwan's fake news detection, and the BLSTM method works best. Research experiments reduce the proportion of fake news, simulating 25%

and 5% fake news ratios. Let the research sample be closer to the real situation. Finally, this study used a cross-data set test to understand the gap between practice and theory.

關鍵詞：假新聞、假新聞偵測、深度學習、人工智慧、雙向長短期記憶

Keywords: Fake News, Fake News Detection, Deep Learning, Artificial Intelligence, Bidirectional Long Short Term Memory

一、緒論

假新聞的定義為有意撰寫且誤導新聞讀者，但驗證是虛假的新聞[1]。近年來各國都受到假新聞嚴重影響，例如：2016 美國總統大選、台灣衛生紙之亂、台灣關西機場假新聞事件等等[2]。近年的假新聞都是透過網路為媒介。而要檢測網路上的內容的真偽，最原始的方式是透過「手動」方式[3]，使得 Facebook 與 Google 兩家網路巨頭公司都想讓假新聞偵測自動化。

過往多位學者曾應用機器學習或者深度學習來進行假新聞偵測，如：Support Vector Machines(SVM)、Naive Bayes、Convolutional Neural Network(CNN)、Logistic Regression(LR)、Neuro Linguistic Programming(NLP)、Recurrent Neural Network(RNN)、Gated Recurrent Unit(GRU)、Long Short-Term Memory(LSTM)。本研究也嘗試使用台灣的假新聞資料，探究以深度學習方式偵測假新聞的可能性。

二、文獻探討

深度神經網路(Deep Neural Network)，是多層的神經網路，可透過電腦找出特徵值，經由深入學習後，讓預測結果更有效，目前主要的神經網路模型有 20 幾種，其中，遞迴神經網路(Recurrent Neural Network, RNN)可用於解決帶有順序性的問題，例如：自然語言處理、語音辨識、手寫辨識等等[4]。

Hochreiter and Schmidhuber [5]提出長短期記憶(Long Short-Term Memory, LSTM)，是為了解決 RNN 所產生梯度消失及梯度爆炸而產生。LSTM 與 RNN 相比多了 3 個控制器，分別為輸入閘門(Input Gate)、輸出閘門(Output Gate)、忘記閘門(Forget Gate)，當有了控制器閘門的機制 LSTM 就能夠將記憶長期記住。

雙向長短期記憶(Bidirectional- Long Short-Term Memory, BLSTM)在 2005 年由 Graves and Schmidhuber [6]提出來，他是由前向長短期記憶與後向長短期記憶結合而成，是為

了解解決後向修飾前向的問題。

閘門循環單元(Gate Recurrent Unit, GRU)與長短期記憶類似。根據 Chung, et al. [7]研究指出閘門循環單元比起長短期記憶會使用更少的記憶、減少中央處理器(CPU)的運算時間以及加快收斂的速度，而閘門循環單元在較小的資料集的表現比長短期記憶來的更加優異。

(二) 假新聞資料庫蒐集

過去假新聞偵測研究中，大多數研究都是根據專家檢核、研究自行蒐集、群眾闢謠網站、社群網站提供以及競賽提供，五種方法所建立起資料庫。例如：BuzzFeed 記者 Silverman, et al. [8]對針對政治相關 Facebook 粉絲專頁，所發的文章做事實檢核、Wang [9]與加拿大維多利亞大學，透過 PolitiFact 事實檢核網站所蒐集的資料庫、The FEVER 1.0 Shared Task 競賽所提供 185,445 筆資料、Pérez-Rosas, et al. [10]透過 GossipCop 八卦檢核網站、Ma, et al. [11]. 微博(Weibo)管理中心提供假新聞報告資料集、Svård and Rumman [12]自行蒐集 201 篇美國新聞文章，其中有 120 是假新聞，81 篇是真新聞、Pérez-Rosas, et al. [10]自行蒐集假新聞，並請外包公司改寫。近年來台灣開始有事實檢核網站以及群眾闢謠網站的出現，但仍少有相關的研究成果被提出。

(三) 應用深度學習假新聞偵測研究

過去假新聞偵測使用的特徵大多是使用文字特徵，並使用機器學習的方式進行偵測，例如：使用 Support Vector Machines(SVM)、Unsupervised、Naive Bayes。而因深度學習發展優異，也開始有研究將深度學習應用在假新聞偵測中。不過，在台灣，這方面的研究成果仍較少被提出。本研究整理出過去研究所使用深度學習的假新聞偵測研究：

三、研究方法

(一) 資料集

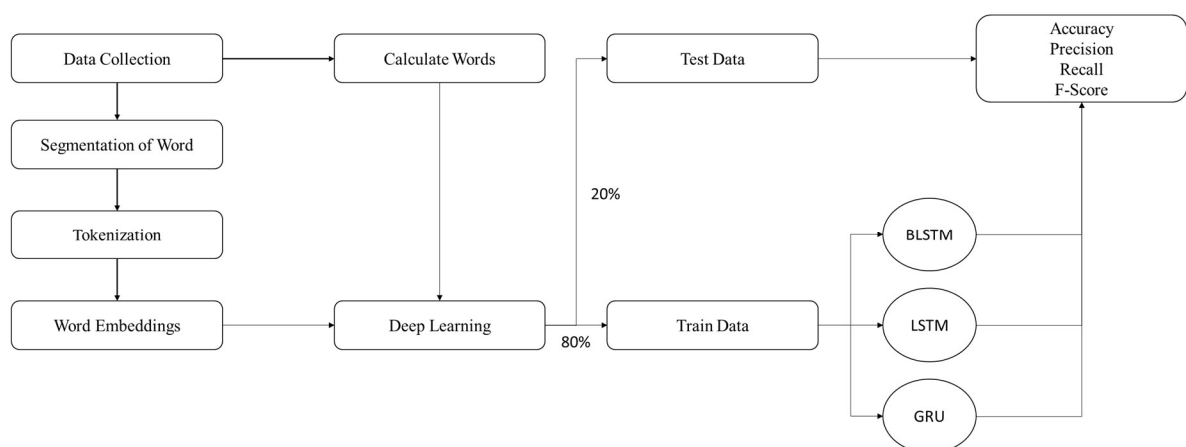
本研究蒐集資料集是台灣組織「g0v.tw 台灣零時政府」，所創立的兩個專案：

「Cofacts 真的假的」以及「新聞小幫手」。「Cofacts 真的假的」透過 Line Chatbot 來蒐集社群上流傳的假新聞，在藉由群眾舉報然後在群眾核對，讓大家一起打擊網路上的謠言以及假新聞。本研究將資料存至 SQL Server，至截稿為止資料庫真新聞有 6,107 筆資料，假新聞有 9,141 筆資料。

「新聞小幫手」是一個闢謠的網站，主要是讓民眾對於可疑新聞來此詢問，由其他民眾蒐集資料後闢謠，但目前於 2018/10/24 停止服務，但資料依舊會保留在網站上面。因「新聞小幫手」沒有真新聞的標記資料，本研究蒐集了真新聞資料，主要來源是蘋果日報，再透過相似度分析將以被舉報假新聞的資料給剔除。至截稿為止真新聞的資料一共蒐集了 11,015 筆，假新聞的資料一共有 5,795 筆，經資料整理後，使用了 2,014 筆。

(二) 研究流程

本研究先將兩個資料集蒐集後，進行文字的預處理。將中文字先進行斷詞後，在將中文字進行 Tokenization 以及文字的詞數計算，以便深度學習處理。然後將預處理資料以 80-20 法則分成訓練集與測試集。本研究實驗三種深度學習的方法，Gated Recurrent Unit (GRU)、Long Short Term Memory (LSTM) 以及 Bidirectional Long Short Term Memory (BLSTM)。然後使用 Accuracy、Precision、Recall 及 F-Score 對模型進行評估。最後將此上述經過繪製成以下流程圖：



圖一、研究流程圖

(三) 深度學習架構

本研究會將三個資料集都以 80-20 法則分成 80%訓練集與 20%的測試集，在讀取資料後建立出 Token 字典，字典依據出現次數的多寡以降冪方式排序，接下來使用 Token 把新聞文字轉成數字陣列。由於新聞字數並不一定，為了讓轉換後的字數相同。本研究將採用截長補短的方式，利用前面所計算的詞數，設定一固定長度為 n ，若新聞文字長度大於 n ，則截去後面的數字;反之如果新聞文字若小於 n ，則會補 0 直到 n 的長度。之後便加入 BLSTM、LSTM 與 GRU 模型，先建立出線性堆疊模型，然後再加入平坦層、隱藏層、輸出層，再以五種不同大小的 Dropout、三種不同的損失函數、三種不同的激活函數以及三種不同的優化器，總共 135 個排列組合，找出最佳組合的模型。在資料經過一連串迭代之後，使 Accuracy、Precision、Recall、F-score 檢測模型。

四、實驗結果

(一) 訓練與測試筆數

本研究將真新聞與假新聞比例成 1:1，假新聞的比例將會佔 50%。接著以 80-20 法則分成訓練與測試筆數，下表為「新聞小幫手」與「真的假的」訓練與測試筆數：

表二、假新聞比例 50%訓練與測試筆數

資料集	訓練筆數 (假：真)	測試筆數 (假：真)
新聞小幫手	1611:1611	403:403
真的假的	4886:4886	1221:1221

(二) 深度學習結果

本研究將實驗中得出的三種方法最佳配置如表二所示。最後以 Accuracy、Precision、Recall 以及 F-Score 對各種模型做評估如表三所示：

表三、50%假新聞比例最佳配置

資料集	方法	激活 函數	損失 函數	優化器	Dropout
新聞 小幫手	GRU	sigmoid	MSLE	Adam	0.6
	LSTM	sigmoid	binary_crossentropy	AdaMax	0.6
	BLSTM	relu	MSE	RMSprop	0.5
真的假的	GRU	sigmoid	binary_crossentropy	Adam	0.4
	LSTM	sigmoid	binary_crossentropy	RMSprop	0.4
	BLSTM	relu	MSLE	AdaMax	0.5

表四、50%假新聞比例模型評估

資料集	方法	Loss	Accuracy	Precision	Recall	F-Score
新聞 小幫手	GRU	0.053	0.881	0.885	0.876	0.88
	LSTM	0.045	0.888	0.881	0.898	0.889
	BLSTM	0.091	0.898	0.898	0.898	0.898
真的假的	GRU	0.104	0.677	0.635	0.836	0.722
	LSTM	0.108	0.685	0.641	0.844	0.728
	BLSTM	0.218	0.683	0.638	0.846	0.728

(三) 不同假新比例結果

本研究認為假新聞佔樣本 50%並不符合真實情況，本研究又實驗了假新聞比例 25%以及 5%，並照一樣的研究方法來實驗。本研究在實驗「新聞小幫手」時發現，在 5%假新聞比例之下，可以再增加資料筆數。所以本研究又實驗了在相同假新聞比例在 5%情況下，會有什麼樣的結果。以下是本研究訓練與測試筆數以及實驗結果：

表五、假新聞比例 25%與 5%訓練與測試筆數

資料集	假新聞比率	訓練筆數 (假：真)	測試筆數 (假：真)
新聞小幫手	25%	1600:4800	400:1200
	5%	252:4800	63:1200
	5%*	421:8000	105:2000
真的假的	25%	1600:4800	400:1200
	5%	252:4800	63:1200

表六、假新聞比例 25%與 5%最佳配置

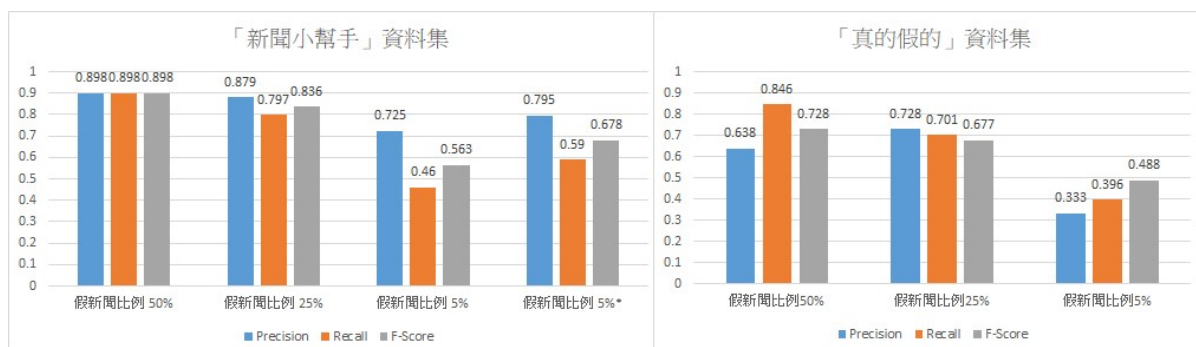
資料集	假新聞比例	最佳方法	激活函數	損失函數	優化器	Dropout
新聞小幫手	25%	BLSTM	sigmoid	binary_crossentropy	Adam	0.3
	5%	BLSTM	tanh	binary_crossentropy	Adam	0.4
	5%*	GRU	Tanh	MSLE	Adam	0.6
真的假的	25%	BLSTM	relu	binary_crossentropy	AdaMax	0.4
	5%	LSTM	tanh	binary_crossentropy	AdaMax	0.6

表七、假新聞比例 25%與 5%實驗結果

資料集	最佳方法	假新聞比例	Loss	Accuracy	Precision	Recall	F-Score
新聞小幫手	BLSTM	25%	0.053	0.922	0.879	0.797	0.836
	BLSTM	5%	0.045	0.964	0.725	0.46	0.563
	GRU	5%*	0.091	0.972	0.795	0.59	0.678
真的假的	BLSTM	25%	0.134	0.845	0.728	0.701	0.677
	LSTM	5%	0.616	0.948	0.333	0.396	0.488

最後以圖二可以發現，雖然降低假新聞比例，可以讓樣本可以更趨近於真實情況，但可以明顯發現 Precision、Recall 以及 F-Score 下降。甚至在「真的假的」在 5%假新聞比例四個指標皆未達到基準線。而在「新聞小幫手」的假新聞 5%實驗當中可以發現，相

同比例下增加訓練與測試筆數，可以增加模型的效率。



圖二、不同假新聞比例效率統計圖

五、結論

本研究實驗兩個中文資料集，發現深度學習確實可以用於偵測假新聞。本研究也模擬了3種假新聞比例的實驗，50%、25%以及5%。在「新聞小幫手」資料集當中，50%的F-Score來到0.898，25%的F-Score來到0.836以及5%的F-Score來到0.563。本研究又實驗了將5%的訓練與測試比數拉高，5%的F-Score從原本的0.563增長到0.678發現訓練與資料測試筆數增加，可以讓深度學習的效率更好。在「真的假的」資料集當中50%的F-Score在來到0.728，25%的F-Score來到0.701，而5%的F-Score只有0.396。相較於「新聞小幫手」資料集，「真的假的」資料集效率較差，甚至在5%假新聞比例完全無法偵測。本研究認為「真的假的」文章長度都短於「新聞小幫手」，資料集筆數不足導致深度學習的效率不佳。

本研究實驗了三種方法，BLSTM、LSTM以及GRU，發現大多數的情況下，BLSTM效果都是最佳。除了在「新聞小幫手」5%假新聞比例的GRU效果最好以及「真的假的」5%假新聞比例無法偵測之外，尚餘5種皆是BLSTM方法效果最佳。

而除此之外，本研究又為了進一步實驗將「新聞小幫手」做交叉資料集測試，發現如要將模型實務所需的發現。

在50%、25%與5%模型測試中也發現一樣的結果，Recall也偏高在0.93以及0.87，這也發現50%、25%的模型也能夠清楚了解假新聞脈絡，並且成功預測正確，但一樣對多種真新聞來源脈絡很不熟悉，常常真新聞錯當成假新聞。這說明了在訓練的資料

集當中，真新聞的來源並不夠多，導致模型常將真新聞錯當成假新聞。而在 5%的模型當中 Recall 有大幅度下降到 0.4，本研究認為是 5%所放入假新聞減少，導致準確抓出假新聞下降。

綜上所述，如要將模型進一步的實務化，除了假新聞資料蒐集外，必須要蒐集更多來源的真新聞，讓模型在面對各種不同真新聞之時，不會誤將新聞判別成為假新聞。

(三) 研究限制

本研究認為目前假新聞偵測最大限制還是假新聞的資料不足，無法確認此資料集是否能代表整個真實世界的狀況。資料集當中真新聞是否還有假新聞在其中或者假新聞當中有真新聞被誤報為假新聞，以及假新聞比例在真實資料集當中比例為何。本研究認為深度學習已經相當成熟，可以運用在假新聞偵測當中，但最大限制還是語料庫不足的問題，本研究認為假新聞偵測需要有更進一步的分類，以及更多標籤(Label)以及更多特徵(Feature)去標記。本研究認為可以再進一步嘗試蒐集更多真新聞，嘗試更低的假新聞比例，可以讓模型更能運用在實務上。相信有天假新聞偵測有天可以運用在各種裝置上，避免導致訊息接受者被誤導，造成社會的恐慌。

參考文獻

- [1] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of economic perspectives*, vol. 31, no. 2, pp. 211-36, 2017.
- [2] 汪志堅、陳才, *假新聞：來源、樣態與因應策略*. 台北: 前程文化, 2019.
- [3] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22-36, 2017.
- [4] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*, 2013: IEEE, pp. 6645-6649.
- [5] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [6] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5-6, pp. 602-610, 2005.
- [7] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated

- recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.
- [8] C. Silverman, L. Strapagiel, H. Shaba, Ellei Hall, and J. Singer-Vin. "Hyperpartisan Facebook Pages Are Publishing False And Misleading Information At An Alarming Rate." *buzzfeednews*. <https://www.buzzfeednews.com/article/craigsilverman/partisan-fb-pages-analysis> (accessed.
- [9] W. Y. Wang, "' liar, liar pants on fire": A new benchmark dataset for fake news detection," *arXiv preprint arXiv:1705.00648*, 2017.
- [10] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," *arXiv preprint arXiv:1708.07104*, 2017.
- [11] J. Ma *et al.*, "Detecting rumors from microblogs with recurrent neural networks," in *Ijcai*, 2016, pp. 3818-3824.
- [12] M. Svärd and P. Rumman, "COMBATING DISINFORMATION: Detecting fake news with linguistic models and classification algorithms," ed, 2017.