

新冠疫苗公眾立場的自動化檢測：一種基於深度學習的方法

背景和動機

香港新冠疫苗的低接種率反映了社會存在疫苗猶豫。據政府統計，截止至 2021 年 8 月 14 日，香港有 42.3% 的人口接種了第二針疫苗，距離達到群體免疫還有較大差距。因此，清楚了解公眾對接種疫苗的立場（以下簡稱為：（公眾）疫苗立場），無疑有助於政府作出相應決策，應對疫苗猶豫問題。

此前，立場檢測主要依靠調查和民意測驗。然而，這些傳統的立場檢測方法成本較高且不可持續、無法及時更新。如今，網絡社交平台構成了個人參與社會互動的一個主要組成部分。這些平台被認為是有效的資訊傳播工具，來表達意見和分享觀點，因此在網絡上每天都會產生大量的數據。相關研究可以運用這個新興且重要的渠道，通過分析社交平台數據對公眾立場進行檢測。隨著這種非結構化數據量的增加，對自動化智能立場檢測的需求也隨之顯著增加 [1]。

由於對龐大數量的非結構化數據進行人工處理幾乎不可能做到，因此本研究期望運用網絡社交平台的數據，探索對公眾立場進行自動檢測的方法。在這項研究中，我們用人工智能技術，基於香港幾個主要的社交平台收集和標註後的文本，建立和訓練了一個機器學習模型。具體來說，我們使用了被稱為深度學習 [2] 的機器學習技術，它可以識別這些文本數據體現的文意，並自動將立場分類。深度學習模型可以應用於許多不同的領域，如決策支持系統和政府情報機構；對政策制定者了解公眾對新冠疫苗相關事件或政策的觀點和反饋來說，該模型也具有實踐意義。

網絡社交平台數據收集和預處理

1. 數據收集

為了研究公眾對接種新冠疫苗的立場，我們通過與數行者科技合作研發的數據引擎從香港三大社交平台（香港討論區¹、高登討論區²及親子王國³）收集了評論流數據，市民在這些平台

¹ 香港討論區。 <https://www.discuss.com.hk>

² 高登討論區。 <https://forum.hkgolden.com>

³ 親子王國。 <https://www.baby-kingdom.com>

分享來自各種新聞媒體（如《香港 01》、微博、《第一財經日報》、《東方日報》和《文匯報》等）的資訊，並且互相交流意見。值得注意的是，在香港廣泛使用的所有社交平台中，這三個平台獲得與接種新冠疫苗相關的評論最多。研究使用表 1 所列的關鍵字，自動篩選出自 2020 年 12 月 23 日（當天，香港行政長官林鄭月娥首次宣布港府採購了 2,250 萬劑新冠疫苗，並頒布了相關規例）以來市民發表有關接種疫苗的評論，然後進行人工篩選，進一步剔除無關的評論。經過隨機採樣後，得出的數據集包含 10,722 篇評論。

表 1：用於評論檢索的關鍵字

與接種新冠疫苗相關的關鍵字	疫苗, 免疫, 科興, 復必泰, 北京生物, 武漢生物, 輝瑞, 莫德納, 克爾來福, 復星, 阿斯利康, 滅活, MRNA, 蛋白, 谷針, 一針, 1 針, 兩針, 2 針, 接種, 打針, 不良反應, 副作用, VAXX, VACCIN, IMMUNIZATION, IMMUNE, INOCULATION, IMMUNE, IMMUNOSUPPRESSED, MODERNA, PFIZER, SINOVAC, CORONAVAC, COMIRNATY, BIONTECH, ASTRAZENECA
---------------	---

2. 數據標註

在本研究中，我們將評論的立場分為表 2 所列的四類：*提倡接種*、*不鼓勵接種*、*質疑接種*和*評論接種* [3,4]，不屬於這些類別的評論被歸類為未知。我們將所收集的評論分配給幾個標註員，每篇評論至少由兩個標註員獨立標註。我們以 kappa 係數[5]來衡量這些標註的可靠性，當中平均 kappa 係數為 0.683，說明標註結果達到一定的可靠性，可以從而得出初步結論。研究採用粵語新冠疫苗立場數據集（CCVS-數據集），供後續分析。表 3 概括了此數據集的統計數據。

表 2：立場類別的定義

提倡接種：	不鼓勵接種：
<ul style="list-style-type: none"> • 描述接種疫苗對公共衛生的好處或其安全性。 • 描述不接種疫苗的風險。 • 鼓勵接種疫苗。 • 駁斥反對接種的論點。 • 包含提倡和不鼓勵接種的資訊，但主觀上表示支持接種疫苗。 	<ul style="list-style-type: none"> • 描述接種疫苗的無效性或安全風險。 • 勸阻接種疫苗。 • 質疑疫苗的有效性／安全性。 • 包含反對接種疫苗的消極態度／論點。 • 包含提倡和不鼓勵接種的資訊，但主觀上表示反對接種疫苗。
質疑接種：	評論接種：

<ul style="list-style-type: none"> 對接種疫苗的風險或好處存在猶豫不決及不確定性。 質疑疫苗的有效性／安全性或產生副作用的可能性。 	<ul style="list-style-type: none"> 沒有對接種疫苗表達不確定性，亦沒有包含提倡和不鼓勵接種的內容，但該評論在某程度上確實與疫苗有關。 包含關於在不同情況下市民應否接種疫苗的實際建議。
--	--

表 3：粵語新冠疫苗立場數據集（CCVS-數據集）之概述

社交平台	提倡接種	不鼓勵接種	質疑接種	評論接種	未知	評論總數
香港討論區	2290	185	7	890	1834	5206
高登討論區	423	591	12	1974	1793	4793
親子王國	182	19	3	164	355	723
總計	2895	795	19	3028	3985	10722

研究問題闡述

我們將疫苗猶豫的立場監測視作一個多分類問題。令 x_i 表示評論文本，其由 n 個詞 $\{W_0, W_1, \dots, W_{n-1}\}$ 組成。每則評論文本的立場可以是提倡、不鼓勵或者其他（包括評論、質疑或未知）。

給定一組（ n 個）經過標記的文本實例 $D = \{x_i, y_i\}$ ，其中 $X = \{x_i\}$ 表示實例中的特徵向量集合， $Y = \{y_i\}$ 表示實例的正標籤（以獨熱編碼表示）。我們將最小化方程(1)中定義的交叉熵損失[6]，以訓練模型。

$$Loss_{CE} = \sum_{x_i, y_i \in D} \sum_{j=0}^{c-1} y_{i,j} \log p_j(x_i), \quad (1)$$

其中， $p = [p_0(x_i), p_1(x_i), \dots, p_{c-1}(x_i)]$ 是一個概率分布，每個元素 $p_j(x_i)$ 則代表數據樣本 x_i 屬於 j 類別的概率， $y_i = [y_{i,0}, y_{i,1}, \dots, y_{i,c-1}]$ 是樣本標籤的獨熱編碼表示，當樣本屬於 j 類別時 $y_{i,j} = 1$ ，否則 $y_{i,j} = 0$ ； c 則代表立場類別的數量。

依據我們的定義，立場監測是一個三元文本分類任務。值得注意的是，這項任務比起其他一些文本歸類問題更具挑戰性，因為立場可以通過多樣而複雜的方式表現，為了確定整體立場，可能需要推理和背景信息。並且，在使用專為較正式且有組織的文本設計的技術分析網絡社交平台數據時，會遇到一系列特別的挑戰，因為本研究中的評論常常是較為簡短且非正式的文本，並且帶有特殊標記（例如標籤和表情符號）。此外，由於缺乏針對粵語的自然語言處理技術資源，同時缺乏大量標記數據，使得訓練深度學習模型的設計空間更加複雜、學習變得困難。因此我們運用了數據增強技術來支持模型訓練。

數據增強

從表 3 可以看出，不同的立場標籤之間存在著明顯的不平衡。在模型訓練之中，預測有可能會偏向於樣本較多的分類，這將大大降低模型的泛化能力。訓練數據的大量缺失以及不同類別之間的不平衡嚴重影響了模型的表現。因此，我們通過增強訓練集來有效擴大數據樣本的規模，該操作不僅可以增加訓練數據量以提高模型的泛化能力 [7]，也可以增加數據中的噪聲來提高模型的魯棒性[8, 9]。本研究中，我們使用小數據集中的原始文本作為樣本，進而做啟發式的修改，這個過程類似於圖像畸變。具體來說，我們隨即進行了以下操作：

- 1) **設置掩碼**。我們用[MASK]以 p_{mask} 的概率隨機替換一個詞，這個標記在 BERT [10] 模型中被用於表示未知或被遮蓋的詞彙。直觀來說，這個規則有助於釐清每個詞對標籤的貢獻，例如神經網絡對實例「相信政府相信科[MASK]係每個港民應[MASK]嘅責任」所產生的輸出的信息含量比對實例「相信政府相信科學係每個港民應擔嘅責任」所產生的要少。
- 2) **N-gram 採樣** [11]。我們以概率 p_{NG} 在數據集中做隨機採樣得到 n -gram，其中 n 從 {2, 3, 4, 5, 6} 中選擇。在概念上，這個規則等同於丟棄樣本中所有其他詞，是一種更為激進的掩碼形式。

在這個研究中，數據增強的步驟為：給定一個訓練樣本 $\{W_0, W_1, \dots, W_{n-1}\}$ ，其中 W_i 是一個多維向量，並在單詞上進行迭代，以均勻分布 $X_i \sim UNIFORM[0,1]$ 取得每個 W_i 。如果 $X_i < p_{mask}$ ，我們則對掩蓋 W_i 。在對單詞進行迭代以後，我們以 p_{NG} 的概率，對整個合成樣本進行 n -gram 採樣。最後的合成樣本被添加到增強的、未標記的數據集。我們在每個例子中應用這個程式 n_{iter} 次，從一個例子中產生多達 n_{iter} 個樣本，任何重複的樣本都將被丟棄。

模型架構

本節將展示本研究模型的架構。如圖 1 所示，它主要包含三個部分：1) 詞編碼器：輸入句子並將每個單詞映射到詞嵌入中；2) 捲積特徵提取：使用捲積神經網路 (Convolutional Neural Network, CNN) [12] 來提取詞嵌入中的關鍵特徵 [13]；以及 3) 分類器：使用全連接神經網路來做分類。以上模型組件將在下文中詳細介紹。

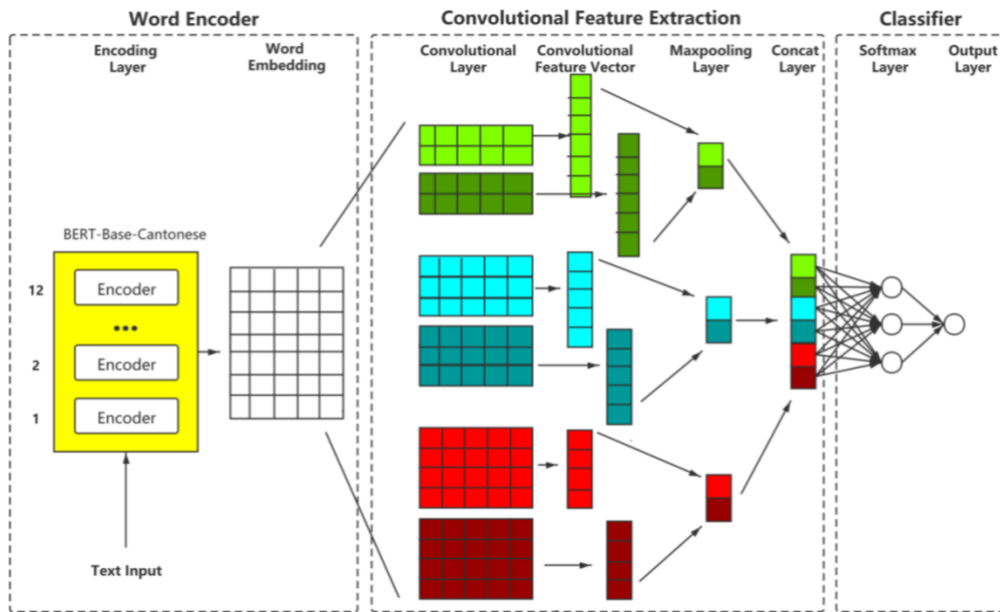


圖 1: 研究模型架構

1. 詞編碼器

本研究採用基於多領域語料的粵語語料庫（包括推特上所收集的粵語推文、香港粵語語料庫以及香港 20 世紀中期粵語語料庫）完成預訓練的 BERT 模型。該模型可以學習廣東話的語法和語義特徵。具體來說，詞編碼器部分將預處理後的單一評論文本 $T = \{T_0, T_1, \dots, T_{n-1}\}$ 作為輸入，然後將其輸精調後的 BERT 模型，得到分詞後的詞嵌入表示 $W = \{W_0, W_1, \dots, W_{n-1}\}$ ，其中 W_i 是一個 768 維的向量。該種詞表示可以保留高質量的語義特徵。接下來這些詞嵌入特徵將作為卷積神經網路 (CNN) 的輸入。

2. 卷積特徵提取

立場分類的關鍵是準確提取文本的中心思想，而提煉中心思想的方法是提取文本的關鍵詞作為特徵，並根據這些特徵來訓練分類器 [14]。CNN 的卷積和最大池化過程是一個特徵提取的過程。利用 CNN 提取關鍵詞的特徵，比直接用預訓練的詞嵌入做立場分類更準確。

我們描繪了三種濾波器區域大小：2、3、4。濾波器對句子矩陣進行卷積，並生成特徵映射；對每個特徵映射進行 1-最大池化，即記錄每個特徵圖中最大的數字。因此，由三個特徵映射分別產生一個特徵向量，這三個向量被串聯起來，形成倒數第二層的特徵向量。

3. 分類器

最後的分類器，接收特徵向量作為輸入值，並使用它來對文本進行分類。在這一層面，本研究應用“Dropout” [15]作為正則化的手段。這是將權重向量中的數值隨機設置為零，以減少過擬合的程度。

4. 參數設置

表 4 列出了本研究實驗中使用的主要參數設置。

表 4: 參數設置

批大小	128
最大單句長度	32
Adam 學習率	0.00005
迭代次數	15
提前停止	100
濾波器個數	256
Dropout rate	0.1
p_{mask}	15%
p_{NG}	25%

對於預訓練的模型，本研究採用 BERT-base-uncased 模型。關於預訓練的 BERT 模型的參數，請參考 [10]。

5. 基準方法

本研究選擇了在文本分類領域常用的兩種精調的 BERT 模型之方法：

- a) BERT+FC: 直接將 BERT 模型連接到全連接網絡，基於預訓練模型產生的語義向量進行分類。
- b) BERT+Bi-LSTM: 雙向長短期記憶網絡 (Bi-LSTM) [16]常用於類似的工作中 [17,18]。它可以捕捉到分詞的序列信息。

實驗結果

本研究使用該模型檢測公眾對新冠疫苗接種之立場，並通過實驗來評估其性能。所有的實驗均使用 Tesla-T4 16G GPU，並在谷歌 Colab 上進行。實驗中使用 CCVS 數據集。我們以 3:1:1 的比例將其隨機分成訓練、驗證和測試集。每個實驗都獨立地重覆 10 次，取平均值作為最終結果。研究使用 F1 分數（宏平均、微平均和加權平均）對三種立場類別的模型進行評估，以滿足各種應用場景的需要。

表 5 顯示了我們的模型和基準之間的性能比較。我們可以看到，我們的模型在三類 F1 分數上都取得了最好的結果。具體來說，較高的 F1-宏平均分表明我們的模型可能有更好的能力來處理泛化問題。原因可能是它捕獲了文本的多個不同的 n-gram 特徵，對於一個 n-gram 特徵，它有許多不同的濾波器來從不同方面提取有效信息，從而不容易造成過擬合。另外，我們還進行了游融實驗來研究數據增強的貢獻。從表 5 中我們可以看出，這種操作在所有三個指標上都改善了我們的模型。本研究最終的實驗結果可以接近近期立場檢測的研究結果，例如 [19, 20, 21]；但是，面向粵語的自然語言處理使得本研究更富挑戰性。

表 5: 實驗結果

Models	F1-宏觀	F1-微觀(準確性)	F1-加權平均
BERT+FC	.473	.722	.688
BERT+Bi-LSTM	.395	.705	.629
BERT+CNN (本研究採用的模型)	.508	.740	.704
BERT+CNN (為進行數據增強)	.450	.733	.686

結論

本研究由「助減香港市民疫苗猶豫」項目發起，運用香港不同網絡社交平台的數據，研究公眾對新冠疫苗的立場。具體來說，通過我們收集和標記的社交平台數據，本研究提出了一種運用深度學習技術在網絡社交平台的訊息中自動檢測有關新冠疫苗接種立場的方法。實驗結果表明該方法在新冠疫苗立場檢測上的有效性。據我們所知，這是首個在粵語語境下自動檢測有關新冠疫苗立場的研究。

我們相信該研究可以幫助政策制定者更好地運用網絡社交平台的數據，來預估和了解公眾對疫苗接種立場的實時趨勢，進而制定相關政策，克服疫苗猶豫，幫助取得疫苗推廣的成功。與問卷調查等傳統方法相比，該方法具有實時性、多信息源、低成本等優點。

本研究運用不同社交平台上的線上大數據，了解公眾對相關事件的立場，為學術界與政府共同應對傳染病大流行的合作框架提供啟示。如果未來面臨類似挑戰，該框架可以為調動資源提供指導，促進政府、學術界和社區之間的有效合作，並為政策制定提供參考，以鼓勵人們採取包括接種疫苗在內的積極行動抗擊疫情。

參考文獻

- [1] P. Sobhani, “Stance detection and analysis in social media,” Ph.D. dissertation, University of Ottawa, 2017.
- [2] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [3] S. Martin, E. Kilich, S. Dada, P. E. Kummervold, C. Denny, P. Paterson, and H. J. Larson, ““vaccines for pregnant women. . .?! absurd”–mapping maternal vaccination discourse and stance on social media over six months,” *Vaccine*, vol. 38, no. 42, pp. 6627–6637, 2020.
- [4] S. Mohammad, S. Kiritchenko, P. Sobhani, X. Zhu, and C. Cherry, “Semeval-2016 task 6: Detecting stance in tweets,” in *Proc. of SemEval*, 2016.
- [5] J. Carletta, “Assessing agreement on classification tasks: the kappa statistic,” *arXiv preprint cmlg/9602004*, 1996.
- [6] S. Mannor, D. Peleg, and R. Rubinstein, “The cross entropy method for classification,” in *Proc. of ICML*, 2005.
- [7] A. Ali, S. M. Shamsuddin, and A. L. Ralescu, “Classification with class imbalance problem,” *Int. J. Advance Soft Compu. Appl*, vol. 5, no. 3, 2013.
- [8] R. Tang, Y. Lu, L. Liu, L. Mou, O. Vechtomova, and J. Lin, “Distilling task-specific knowledge from bert into simple neural networks,” *arXiv preprint arXiv:1903.12136*, 2019.
- [9] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
- [10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.
- [11] P. F. Brown, V. J. Della Pietra, P. V. Desouza, J. C. Lai, and R. L. Mercer, “Class-based n-gram models of natural language,” *Computational linguistics*, vol. 18, no. 4, pp. 467–480, 1992.
- [12] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation applied to handwritten zip code recognition,” *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [13] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *arXiv preprint arXiv:1301.3781*, 2013.
- [14] Y. Zhang and B. Wallace, “A sensitivity analysis of (and practitioners’ guide to) convolutional neural networks for sentence classification,” *arXiv preprint arXiv:1510.03820*, 2015.

- [15] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929-1958, 2014.
- [16] Z. Huang, W. Xu, and K. Yu, "Bidirectional lstmcrf models for sequence tagging," *arXiv preprint arXiv:1508.01991*, 2015.
- [17] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using bilstm-crf and cnn," *Expert Systems with Applications*, vol. 72, pp. 221-230, 2017.
- [18] D. AlBatayha, "Multi-topic labelling classification based on lstm," in *Proc. of IEEE ICICS*, 2021.
- [19] M. Mohtarami, R. Baly, J. Glass, P. Nakov, L. M'arqu and A. Moschitti, "Automatic stance detection using end-to-end memory networks," *arXiv preprint arXiv:1804.07581*, 2018.
- [20] B. Riedel, I. Augenstein, G. P. Spithourakis, and S. Riedel, "A simple but tough-to-beat baseline for the fake news challenge stance detection task," *arXiv preprint arXiv:1707.03264*, 2017.
- [21] A. Hanselowski, A. PVS, B. Schiller, F. Caspelherr, D. Chaudhuri, C. M. Meyer, and I. Gurevych, "A retrospective analysis of the fake news challenge stance detection task," *arXiv preprint arXiv:1806.05180*, 2018.