The *Overcoming Vaccine Hesitancy in Hong Kong* Project
by Hong Kong Baptist University
Report Series No. 15
27-August-2021

香港浸會大學
HONG KONG BAPTIST UNIVERSITY

Towards Automatic COVID-19 Vaccination Stance Detection in Hong Kong: A Deep Learning-Based Approach

Background and Motivation

The low vaccination rate reflects the existence of vaccine hesitancy in Hong Kong society. According to government statistics, as of 14 August 2021, 42.3% of the city's population had been fully vaccinated which remains far from the rate required to reach herd immunity. To alleviate this situation, a clear understanding of the public's stance on vaccination can undoubtedly help the government make more relevant decisions and future policies.

Previously, stance detection mainly relied on public surveys and polls. However, these traditional stance detection methods cost a lot, are unsustainable and not up to date. Nowadays, online social platforms have constituted a major component of an individual's social interaction. These platforms are considered robust information dissemination tools to express opinions and share views. Consequently, huge amounts of data are generated every day on the Internet, which provides a newly developing but important channel to detect public stance through online social platform data. As the volume of such unstructured data increased, the request for automatic identification and extraction of stances grows significantly [1].

Our research is about the automatic public stance detection of online social platform data, as manual processing of sheer volume of such unstructured data is not feasible. In this research, we leveraged AI techniques, built and trained a machine learning (ML) model with massive text data collected and annotated from major social platforms in Hong Kong. Specifically, we used the ML algorithmic technique called *deep learning* [2] which can recognise the representation of these text

data and automatically classify the stance. The deep learning model can be applied in many different domains, such as decision support systems and government intelligence, and can be of particular interest for policy makers who seek public opinions and reactions to COVID-19 vaccine-related events or policies.

Online Social Platform Data Collection and Pre-processing

1) Data Collection

In order to study the public stance towards COVID-19 vaccination, we collected streaming data of comments through a database engine powered by HKBU and Datago company on three major social platforms in Hong Kong, namely HKDiscuss[1], HKGolden[2], and BabyKingdom[3], where people can share and exchange comments on information from various news media, e.g., HK01, Weibo, China Business News, Oriental Daily News, and Wen Wei Po. It is noteworthy that among all social platforms widely used in Hong Kong, these are the top three from where we acquired the largest number of COVID-19 vaccination-related comments. Using the keywords listed in Table1, we first filtered out the comments of interest posted since 23 December 2020 (i.e., the day when Carrie Lam, the Chief Executive of Hong Kong, announced for the first time that the Hong Kong government had purchased 22.5 million doses of COVID-19 vaccines and promulgated relevant regulations). Afterwards, a manual selection was conducted to further remove unrelated comments. After random sampling, the resulted dataset consists of 10,722 comments.

[1] HKDiscuss. https://www.discuss.com.hk

[2] HKGolden. https://forum.hkgolden.com

[3] BabyKingdom. https://www.baby-kingdom.com

Table 1: Keywords used for comment retrieval

COVID-19 vaccination related keywords	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, VACCINE, IMMUNIZATION, IMMUNE, INOCULATION, IMMUNE, IMMUNOSUPPRESSED, MODERNA, PFIZER, SINOVAC, CORONAVAC,COMIRNATY, BIONTECH, ASTRAZENECA

2) Data Annotation

In this study, we categorised the stance of comments into four classes as defined in Table 2, which are *promotional*, *discouraging*, *querying*, and *commenting* [3,4]. Texts that do not belong to any of these categories are regarded as *unknown*. We assigned the comments collected to a few annotators in a way such that each comment was annotated by at least two annotators independently. We used the kappa coefficient [5] as a measurement of the reliability of their annotations. In this work, the mean kappa coefficient is 0.683, which indicates that the annotation results achieve a reliability that allows tentative conclusions to be drawn. Therefore, we obtained the Cantonese COVID-19 Vaccine Stance Dataset (CCVS-Dataset) for subsequent analyses. Several statistics of CCVS-Dataset are summarised in Table 3.

Table 2: Definitions of Stance Categories

Promotional	**Discouraging**
• Describe public health benefits or safety of vaccination. • Describe risks of not getting vaccinated. • Encourage vaccination. • Refute the argument against vaccines.	• Describe invalidity or safety risks of vaccination. • Discourage vaccination. • Question the effectiveness/safety of vaccines.

• Contain both promotional and discouraging information, but express support subjectively.	• Contain negative attitude/arguments against vaccination. • Contain both promotional and discouraging information, but express opposition subjectively.
Querying	**Commenting**
• Contain indecision and uncertainty about the risks or benefits of vaccination. • Contain questions about effectiveness/safety or possibility of side-effects.	• Contain no elements of uncertainty, and no promotional and discouraging content but its post does somehow relate to vaccine. • Include factual recommendations about whether one should get vaccinated under different circumstances.

Table 3: Overview of the CCVS-Dataset

Platforms	Promotional	Discouraging	Querying	Commenting	Unknown	Total
HKDiscuss	2290	185	7	890	1834	5206
HKGolden	423	591	12	1974	1793	4793
Baby Kingdom	182	19	3	164	355	723
Overall	2895	795	19	3028	398	10722

Research Problem Statement

We treat stance detection of vaccine hesitancy as a multi-class classification problem. Let x_i be a comment text, consisting of n words $\{W_0, W_1, \ldots, W_{n-1}\}$. Each comment text's stance can be *promotional*, *discouraging*, or *others* (including *commenting*, *querying* and *unknown*).

Given a set of n labeled text instances $\boldsymbol{D} = \{\boldsymbol{x}_i, \boldsymbol{y}_i\}$ with $\boldsymbol{X} = \{\boldsymbol{x}_i\}$ denoting the set of feature vectors of the instances and $\boldsymbol{Y} = \{\boldsymbol{y}_i\}$ being their corresponding true labels in one-hot representation. We intend to train the model by minimising the Cross-Entropy loss [6] defined in Equation (1):

$$Loss_{\text{CE}} = \sum_{\boldsymbol{x}_i, \boldsymbol{y}_i \in D} \sum_{j=0}^{c-1} y_{i,j} \, log \, \boldsymbol{p}_j(\boldsymbol{x}_i), \quad (1)$$

where $\boldsymbol{p} = [\boldsymbol{p}_0(\boldsymbol{x}_i), \boldsymbol{p}_1(\boldsymbol{x}_i), \ldots, \boldsymbol{p}_{c-1}(\boldsymbol{x}_i)]$ is a probability distribution, and each element $\boldsymbol{p}_j(\boldsymbol{x}_i)$ represents the probability that the data sample $\boldsymbol{x}_i$ belongs to class j; $\boldsymbol{y}_i = [y_{i,0}, y_{i,1}, \ldots, y_{i,c-1}]$ is the one-hot representation of the sample's label, which means $y_{i,j} = 1$ when the sample belongs to class j, otherwise $y_{i,j} = 0$; and c is the number of stance classes.

Stance detection, as we defined here, is a three-class text classification task. Notably, it is a more challenging task than some other text categorisation problems, as the stance can be expressed in various complicated ways in which, to identify the overall position, inference and background information might be required. Also, analyses of online social platform data present a unique set of challenges for technologies designed for more formal and organised text, as the comments considered in this research are usually short, informal, and have special markers such as hashtags and emoticons. In addition, the lack of resources of natural language processing (NLP) techniques for the Cantonese language and a large amount of labeled data further complicates the design space to train deep learning models, making learning difficult. Thus, we utilise the *data augmentation* technique to facilitate model training.

Data Augmentation

It can be seen from Table 3 that there is a significant imbalance between the data of different stance labels. In the model training, the prediction will be biased towards the classification with many samples, which would greatly reduce the generalisation of the model [7]. The lack of a large amount of training data and the imbalance between categories seriously affect the performance of

the model. Therefore, we *augmented the training set* as an effective method to expand the size of data samples. This operation can not only increase the amount of training data to improve the generalization ability of the model, but also increase the noise of the data to improve the robustness of the model [8, 9]. In our work, we used the original texts in the small dataset as blueprints, and then modify them with our heuristics, a process analogous to image distortion. Specifically, we randomly performed the following operations.

1) **Masking**. With probability p_{mask}, we randomly replace a word with [MASK], which corresponds to an unknown token in our models and the masked word token in BERT [10]. Intuitively, this rule helps clarify the contribution of each word towards the label, e.g., the network produces less confident logits for the instance "相信政府相信科[MASK]係每個港民應[MASK]嘅責任" than for the instance "相信政府相信科學係每個港民應擔嘅責任" .
2) **N-gram sampling** [11]. With probability p_{NG}, we randomly sample an *n*-gram from the example, where *n* is randomly selected from $\{2, 3, 4, 5, 6\}$. This rule is conceptually equivalent to dropping out all other words in the example, which is a more aggressive form of masking.

Our data augmentation procedure is as follows: given a training example $\{\boldsymbol{W}_0, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_{n-1}\}$, where $\boldsymbol{W}_i$ is a multi-dimensional vector, we iterate over the words, drawing from the uniform distribution $X_i \sim \text{UNIFORM}[0,1]$ for each $\boldsymbol{W}_i$. If $X_i < p_{\text{mask}}$, we apply masking to $\boldsymbol{W}_i$. After iterating over the words, with probability p_{NG}, we apply *n*-gram sampling to this entire synthetic example. The final synthetic example is appended to the augmented, unlabeled dataset. We apply this procedure n_{iter} times per example to generate up to n_{iter} samples from a single example, with any duplicates discarded.

Model Architecture

In this section, we demonstrate the architecture of our model. As shown in Figure 1, it contains three main components: 1) Word Encoder: input sentences and map each word into word embedding; 2) Convolutional Feature Extraction: use a convolutional neural network (CNN) [12] to extract the key features in word embedding [13]; and 3) Classifier: use a fully connected neural network to do classification. These components will be presented in detail in the following.

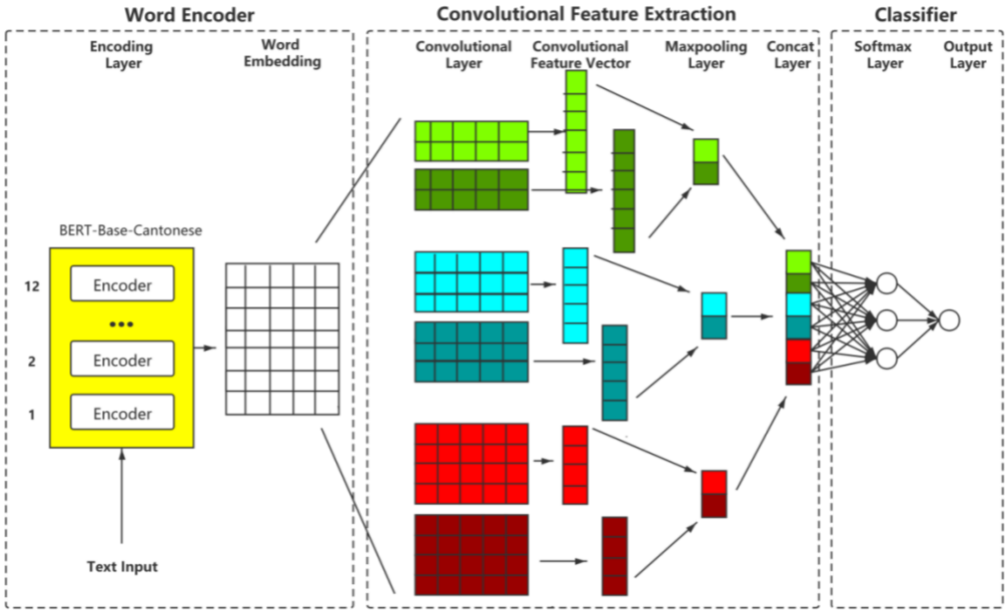

Figure 1: The architecture of our model

1) Word Encoder

We use the BERT model pre-trained by a multi-domain Cantonese corpus based on the collected Cantonese tweets from Twitter, the Hong Kong Cantonese corpus[4], and the Hong Kong mid-twentieth century Cantonese corpus[5]. It can learn the grammatical and semantic features of Cantonese. Specifically, the part of Word Encoder takes the pre-processed single comment text $\boldsymbol{T} = \{\boldsymbol{T}_0, \boldsymbol{T}_1, \ldots, \boldsymbol{T}_{n-1}\}$ as input, then passes it to the fine-tuned BERT model

[4]Hong Kong Cantonese corpus http://compling.hss.ntu.edu.sg/hkcancor/
[5]Hong Kong mid-twentieth century Cantonese corpus http://corpus.ied.edu.hk/hkcc/

for extracting word embedding after tokenization, and obtains high-quality semantic features $\boldsymbol{W}$ in the word embedding form $\boldsymbol{W} = \{\boldsymbol{W}_0, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_{n-1}\}$, where $\boldsymbol{W}_i$ is a 768-dimensional vector. Such word embedding features are then the inputs of the convolutional neural network (CNN).

2) Convolutional Feature Extraction

The key to stance classification is to accurately extract the central idea of texts, and the method of refining the central idea is to extract the keywords of text as features and train the classifier based on these features [14]. The convolution and max-pooling process of CNN is a process of feature extraction. We can extract the features of keywords more accurately than directly doing stance classification with pre-trained word embedding.

We depict three filter region sizes: 2, 3, and 4. Filters perform convolutions on the sentence matrix and generate feature maps; 1-max pooling is performed over each map, i.e., the largest number from each feature map is recorded. Thus, a feature vector is generated from all three maps, and these three features are concatenated to form a feature vector for the penultimate layer.

3) Classifier

Finally, the classifier receives the feature vector as input and uses it to classify the text. At this layer, we apply the *dropout* [15] as a means of regularisation. This entails randomly setting values in the weight vector to zero to reduce the degree of overfitting.

4) Parameter Settings

Table 4 presents the settings of major parameters used in our experiments.

Table 4: Parameter settings

Batch size	128
Max sentence length	32
Adam Learning rate	0.00005
Number of epochs	15
Early stopping	100
Filter number	256
Dropout rate	0.1
p_{mask}	15%
p_{NG}	25%

For the pre-trained model, we use the BERT-base-uncased model. Please refer to [10] for details about the parameters of the pre-trained BERT model.

5) Baselines

We chose two methods of fine-tuning the BERT model commonly used in the text classification field：

a) BERT+FC: Directly connect the BERT model to a fully connected network, and the classification is directly based on the semantic vector classification produced by the pre-trained model.

b) BERT+Bi-LSTM: Bidirectional Long Short Term Memory Networks (Bi-LSTM) [16] is commonly used in similar works [17,18]. It can capture the sequential information in the tokens.

Experimental Results

We conducted experiments to evaluate the performance of the model for COVID-19 vaccination stance detection. All experiments were conducted on Google Colab with a Tesla-T4 16G GPU. The dataset used in experiments is the CCVS-Dataset. We randomly split it into training, validation, and testing sets with the ratio of 3:1:1. Each experiment was repeated 10 times independently and the average value was taken as the final result. We evaluated the models using the *F1 score* (*macro*, *micro*, and *weighted average*) over the three stance classes to meet the needs of various application scenarios.

Table 5 illustrates a comparison of the performances between our model and the baselines. We can observe that our model achieves the best results on all three types of F1 scores. Specifically, the higher F1-macro score indicates that our model may have a better ability to deal with generalisation. The reason can be that it captures multiple different *n*-gram features of the text, and, for an *n*-gram feature, it has many different filters to extract useful information from different aspects so that it is not easy to overfit. Also, we conducted *ablation experiments* to investigate the contributions of data augmentation. We can see from Table 5 that this operation improves our model in all the three metrics. Our final experimental results can approach the results of the recent studies of stance detection, e.g., [19, 20, 21]. However, our problem is more challenging due to the language context of Cantonese.

Table 5: Experiment Results

Models	F1-macro	F1-micro(accuracy)	F1-weighted average
BERT+FC	.473	.722	.688
BERT+Bi-LSTM	.395	.705	.629
BERT+CNN (our model)	**.508**	**.740**	**.704**
BERT+CNN (without data augmentation)	.450	.733	.686

Conclusion and Discussion

Initiated by the "Overcoming Vaccine Hesitancy in Hong Kong" project, this research studies the public stance towards COVID-19 vaccination using data from different online social platforms in Hong Kong. Specifically, with the social platform data we collected and labelled, our research proposed an approach to automatic COVID-19 vaccination stance detection in messages from online social platforms with deep learning techniques. Our experimental results demonstrate the effectiveness of our approach in the COVID-19 vaccination stance detection. To the best of our knowledge, this is the first to study the problem of automatic COVID-19 stance detection in the language context of Cantonese.

We believe that our research can help the policy makers make better use of the online social platform data to estimate and understand the real-time trend of the public stance towards vaccination, and then formulate relevant policies to overcome vaccine hesitancy in Hong Kong. The arduous race to increase vaccine uptake could benefit from such understanding. Compared with the traditional methods like surveys and polls, our approach has the advantages of real-time, multiple information sources, and lower cost.

Our study might shed some light on the academia–government cooperation framework of response to the pandemic, which could better understand the public stance towards relevant events using online big data on different social platforms. In future, in case of an epidemic, this framework can provide guidance to mobilise resources, facilitate efficient collaboration among government, academia and community, and inform policy makers to encourage people to take actions against the epidemic, for example, getting vaccinated.

References

[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 cmplg/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. 19291958, 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.