Sentiment Analysis of Political Posts on Hong Kong Local Forums Using Fine-Tuned mBERT

Guanrong Li†, Ziwei Wang†, Minzhu Zhao†, Yunya Song† Liang Lan‡
†Department of Journalism, Hong Kong Baptist University, Hong Kong SAR, China
‡Department of Interactive Media, Hong Kong Baptist University, Hong Kong SAR, China
‡Corresponding author. E-mail address: lanliang@hkbu.edu.hk

2022 IEEE International Conference on Big Data (Big Data) | 978-1-6654-8045-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigData55660.2022.10020704

***Abstract*—Sentiment analysis is an important and challenging task in natural language processing. It has been studied for a few decades. Recently, Bidirectional Encoder Representations from Transformer (BERT) model has been introduced to tackle this task and gain very promising results. However, most existing studies on fine-tuning BERT models for sentiment analysis focus on high-resource language (e.g., English or Mandarin). This paper studies the sentiment analysis of Cantonese political posts on Hong Kong local forums. We first collected and labeled posts related to Anti-Extradition Law Amendment Bill (Anti-ELAB) movement in Hong Kong discussion forums. We then examined the performance of dictionary-based sentiment analysis, traditional machine learning-based, fine-tuned BERT and fine-tuned multilingual BERT (mBERT) models. Our results show that fine-tuned mBERT model achieves the best performance on our collected and labeled Cantonese dataset.**

***Index Terms*—Sentiment Analysis, BERT, Fine-tuned BERT, multilingual BERT**

I. Introduction

Sentiment analysis is an important task in natural language processing (NLP). It aims to examine users' subjective sentiments toward a certain entity automatically. Sentiment analysis is also referred to as opinion mining, where opinions about an entity are extracted and analyzed automatically. Although there exist slight differences, sentiment analysis and opinion mining are used interchangeably. Because understanding sentiments and opinions is critical to human activities and behaviors, sentiment analysis has become one of the most studied research areas in NLP in recent years. Furthermore, due to the importance of sentiment analysis to the whole society, the related studies spread beyond computer science to other research fields including management and social science [1]. With the proliferating information volume on social media, sentiment analysis is vital in several real-life applications including commercial, political, and public security [2].

As an essential and challenging NLP task, sentiment analysis has been studied for a few decades using both supervised and unsupervised approaches in existing research [3]. For supervised sentiment analysis, the combination of hand-craft features and traditional machine learning (ML) models are utilized in early works. For unsupervised sentiment analysis, techniques with lexicons, grammar, and syntactic patterns have been explored in previous studies. Recently, deep learning (DL) has become a powerful technique and demonstrated outstanding performance in several application domains. The state-of-art deep learning models have been applied in sentiment analysis with promising results [1]. Very recently, fine-tuning a pre-trained Bidirectional Encoder Representations from Transformer (BERT) model for downstream text classification became a new paradigm in NLP and has demonstrated outstanding performance on several other NLP tasks. And the idea of fine-tuning BERT model for sentiment analysis is raised and recent studies have demonstrated promising results [4], [5].

As a prominent kind of Chinese dialect, Cantonese is used as a daily language by a wide population in southern Mainland China, Hong Kong SAR, and overseas Chinese communities. Cantonese is different from Chinese in terms of vocabulary, grammar, and unique characters, which raise a challenge for analyzing Cantonese sentiment. However, sentiment analysis on Cantonese text is less studied [6]. Most existing works for sentiment analysis based on BERT models focus on high-resource languages (e.g., English and Mandarin). Since the resource for fine-tuned Cantonese BERT models is limited, existing works for Cantonese sentiment analysis did not utilize the state-of-art BERT models.

To explore different sentiment analysis methods on Cantonese data, in this paper, we first collected and labeled posts related to Anti-ELAB movement in Hong Kong discussion forums. To examine the performance of existing sentiment analysis methods on Cantonese political-related posts, we extended the existing methods including the early dictionary-based, the traditional ML-based, fine-tuning BERT and multilingual BERT (mBERT) models to analyze Cantonese sentiment. The contribution of the work are summarized as follows:

- We collect and label a real-world Cantonese sentiment analysis dataset related to political discussion on Hong Kong local forums.
- We evaluated dictionary-based methods, ML-based models, and BERT-based models for Cantonese Sentiment analysis;
- We demonstrated the effectiveness of the fine-tuned

mBERT model for Cantonese Sentiment analysis using our collected dataset.

II. Methodology

A. Dataset Collection and Labeling

We collect Cantonese posts related to the Anti-ELAB movement in three Hong Kong discussion forums (HKDiscuss[1], HKGolden[2], LIHKG[3]) and recruit three coders to label sentiment (positive, neutral, negative). All coders major in communication and have intensive experience. The final label for each post is determined by a majority vote. The confusing one with all three labels different is eliminated. Finally, the data contains 1096 data samples. Among them, 331 are labeled positive, 649 are labeled neutral, and 116 are labeled negative.

B. Sentiment Analysis Methods

To examine the performance of the existing methods on our collected data, we evaluate the following methods that contain dictionary-based method, traditional machine learning technique, fine-tuned BERT, and multilingual BERT (mBERT) models.

1) Linguistic Inquiry and Word Count (LIWC): . LIWC is an early and transparent method for sentiment analysis. For a given text, it compares each word in the text to the word list in the positive and negative dictionary and calculates the percentage of the total words that match each of the dictionaries [7].

2) Term Frequency-Inverse Document Frequency (TF-IDF) + forum: : Generally speaking, a traditional machine learning-based method for sentiment analysis contains two modules: text feature extraction and classification model. One of the popular text feature extraction methods is Bag-of-Word based on TF-IDF features. TF-IDF calculates a normalized count where each word count is divided by the number of documents in which this word appears. TF-IDF is one of the most popular term-weighting schemes to transform words into numbers that can be fed to machine-learning algorithms. We trained logistic regression based on our collected data by using TF-IDF as the features. Here, We used the term "forum" to denote that the classification model is trained based on our collected data from local Hong Kong discussion forums.

3) ChineseBERT(CBERT) + SMP: : BERT is one of the state-of-art machine learning methods for neural language representation. For all BERT-based or mBERT-based methods mentioned subsequently, We use a pre-trained model to get the representation of text, a dropout layer to prevent overfitting, and finally a dense layer for sentiment analyses task. Here, we fine-tuned ChineseBERT model based on a publicly available labeled simplified Chinese dataset named SMP[4]. When applying this model to do sentiment prediction on our collected Cantonese dataset, we need to translate the Cantonese (i.e., traditional Chinese) to simplified Chinese first and then feed our translated text into the model for prediction. During the translation, we simply ignore Cantonese slang if it can not be translated into simplified Chinese.

4) mBERT + SMP: . It is similar to *CBERT + SMP*, but we replaced CBERT with mBERT pre-trained on the top 104 languages (including Traditional Chinese and Simplified Chinese) with the largest Wikipedia. Then the translation process is not needed for sentiment prediction on our collected Cantonese data.

5) CBERT + forum: . We adopt the approach of *CBERT + SMP*, except the fine-tuned data was changed from SMP to our collected forum data. We also adopt the translation process to fit the requirement of CBERT. We use our collected Cantonese data to fine-tune CBERT by translating traditional Chinese into simplified Chinese.

6) mBERT + forum: . Since there is no pre-trained BERT model on a large-scale Cantonese corpus, we fine-tune the pre-trained mBERT using our collected Cantonese forum data directly. Then the fine-tuned mBERT can be used for sentiment analysis on our collected Cantonese data without translation.

III. Experiment

A. Experimental Setting

To obtain a robust and accurate evaluation, we use the five-fold cross-validation method to evaluate the performance of each model. In specific, we shuffle the dataset and split it into 5 groups. We take each unique group as the test set and the remaining groups are divided into the train set and validation set according to the ratio of 9:1. For all of the methods, we select the optimal version by the validation set in the training process and then verify the effectiveness of each method in the test set.

The publicly available simplified Chinese dataset SMP was released from a sentiment analysis competition and contained six categories in the raw data. To fine-tune a BERT model on this simplified Chinese dataset for sentiment analysis on our forum data with three outputs (i.e., positive, neutral and negative), we keep only happy, neural, and sad labels to fit the labels of forum data. Totally, there are 5,379 happy, 4990 sad, and 5,749 neutral.

As the dictionary-based method LIWC does not need to be trained, we directly use it on the test set. We adopt the one-vs-rest scheme and Limited-memory BFGS algorithms to obtain the three-class prediction from logistic regression used in the TF-IDF methods. For the BERT model and BERT multilingual model, We utilize the dropout technique, and Adam optimizer with a learning rate 10^{-4}, and all hyper-parameters are manually tuned on the validation set. We use four widely used classification metrics, i.e., Accuracy, Precision, Recall, and F1-Score, for performance evaluation.

[1]https://www.discuss.com.hk/
[2]https://www.hkgolden.com/
[3]https://lihkg.com/
[4]https://smp2020ewect.github.io

TABLE I: Comparison of Different Sentiment Analysis Methods

Method Type	Method	Accuracy	Precision	Recall	F1-score
Dictionary-based	LIWC	0.47±0.04	0.43±0.05	0.43±0.05	0.39±0.04
Zero-shot BERT	CBERT + SMP	0.58±0.03	0.34±0.08	0.34±0.01	0.28±0.03
	mBERT + SMP	0.58±0.02	0.27±0.06	0.33±0.01	0.26±0.01
Traditional ML	TF-IDF + forum	0.60±0.03	0.44±0.13	0.35±0.02	0.29±0.03
Fine-tuned BERT	CBERT + forum	0.70±0.01	0.67±0.04	0.57±0.04	0.59±0.04
	mBERT + forum	**0.72±0.03**	**0.70±0.05**	**0.58±0.07**	**0.60±0.07**

B. Experimental Results

The average results and standard deviations based on cross-validation for different sentiment analysis methods are shown in TableI.

Different types of methods. Firstly, we investigate the effectiveness of the different types of methods. Not surprisingly, there are notable differences in model types between the dictionary-based method, tradtional machine leanring based method, and BERT(mBERT) based methods. The dictionary-based method's method is not good as other machine learning-based methods. In this particular dataset, zero-shot BERT gets worse performance than traditional machine learning-based methods and fine-tuned BERT methods. The fine-tuned BERT methods get the highest performance.

The role of multilingual BERT. The fine-tuned multilingual BERT model gets a slightly better result than fine-tuned CBERT model. The performance gap between CBERT and the corresponding mBERT model is very small.

The role of the training dataset. From TableI, it is obvious that the performance of the BERT or mBERT model trained on SMP is significantly lower than that trained on our forum dataset. Whether we need to translate(CBERT) or not(mBERT), the results of training a model using a dataset from different domains are poor. The BERT-based model trained on our forum data significantly outperforms those trained on the SMP dataset, suggesting that directly using the existing fine-tuned BERT model on other datasets is not effective with our forum data. However, BERT still excels in sentiment analysis, and even basic fine-tuning is able to obtain better results than LIWC and can gain competitive results with traditional machine learning models on our forum data (i.e., TF-IDF + forum).

Overall speaking, the results show that the machine learning-based methods perform better than the dictionary-based method. Among the BERT-based methods, The mBERT trained on our collected forum data significantly outperforms those fine-tuned on the SMP dataset. It suggests that directly using a fine-tuned CBERT model based on publicly available simplified Chinese label data with translation under a zero-shot setting is ineffective in our problem.

IV. Conclusion

In this paper, we study sentiment analysis on Cantonese political posts related to the Anti-ELAB movement in Hong Kong discussion forums. We evaluated the performance of dictionary-based sentiment analysis, traditional machine learning-based, fine-tuned BERT and multilingual BERT (mBERT) models based on our collected dataset. Our experimental results show that the fine-tuned mBERT model achieves the best performance.

Acknowledgment

The research was supported by a General Research Fund grant (HKBU 12605520) of the Research Grant Council of Hong Kong, the Public Policy Research Funding Scheme (2021.A2.047.21B) from Policy Innovation and Co-ordination Office of the Hong Kong Special Administrative Region Government, and the Interdisciplinary Research Clusters Matching Scheme (IRCMS/19-20/D04), the Initiation Grant for Faculty Niche Research Areas (RC-FNRA-IG/21-22/COMF/01) of HKBU, and the AIS Scheme (AIS 21-22/01) and CMC Scheme (CMC/21-22/003) from School of Communication, HKBU, and the Natural Science Foundation Council of China (61906161).

References

[1] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1253, 2018.

[2] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowledge and Information Systems*, vol. 60, no. 2, pp. 617–663, 2019.

[3] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press, 2020.

[4] M. G. Sousa, K. Sakiyama, L. de Souza Rodrigues, P. H. Moraes, E. R. Fernandes, and E. T. Matsubara, "Bert for stock market sentiment analysis," in *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2019, pp. 1597–1601.

[5] M. Singh, A. K. Jakhar, and S. Pandey, "Sentiment analysis on the impact of coronavirus in social life using the bert model," *Social Network Analysis and Mining*, vol. 11, no. 1, pp. 1–11, 2021.

[6] R. Xiang, Y. Jiao, and Q. Lu, "Sentiment augmented attention network for cantonese restaurant review analysis," in *Proceedings of WISDOM'19: Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM'19)*, 2019.

[7] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: Liwc 2001," *Mahway: Lawrence Erlbaum Associates*, vol. 71, no. 2001, p. 2001, 2001.