Bert Based Chinese Sentiment Analysis for Automatic Censoring of Dynamic Electronic Scroll

Zong-Yu He and Wei-Liang Lin
National Chung Hsing University, Taichung City, Taiwan, ROC
arthurhe1810@gmail.com and wllin@nchu.edu.tw

***Abstract*—This article uses BERT[1] algorithm to judge Chinese sentiment. The goal is to automatically censor short advertising words on advertising boards or message boards, and filter out some inappropriate speeches. The method is to collect various speeches on the internet, including online shopping reviews, meal reviews, store reviews, etc., as well as emotion dictionaries, and use these data to train the algorithm so that the algorithm can correctly identify the emotions of short sentences.**

I. INTRODUCTION

For a conventional electronic scroll system, the contents are censored by human and posted for a period, i.e., days or weeks. However, this is not efficient for targeting audience. A more precision way is to let ad investor to decide the contents in real time, and the posted period down to minute-long, for example, in a train station environment. Most trains stop for minutes long. Furthermore, within such an automatic ad system, time-sharing is possible, and advertising cost can be reduced. To reach these two goals, automatic contents-submit and censoring are necessary.

We implement the automatic contents-submit function through a web server. Therefore, content-submitter can submit through internet and specify ad time. Due to the electronic scroll nature, we limit the contents length to less than or equal to 15 characters. Our automatic censor system is based on BERT (Bidirectional Encoder Representations from Transformers) model, developed by Google in 2018, and used widely in NLP (Natural Language Processing) area. The initial model weights are from large amount of language data training, and in our application, the output tells whether the input ad contents positive sentiment or not. If the result is not positive, the system will say sorry to the submitter and direct to a manual-handle route with a higher ad price.

Due to that the ad may offend someone if the contents stimulate negative sentiment, and cause the ad system facing legal issue, the automatic system only accepts contents with positive sentiment with high confidence. This consideration affects our cost definition during BERT weight training. We provide experiment result compared with a commercial sentimental analysis tool. Our result is better for our specific application.

We have implemented the whole system including submitting web server, automatic censor algorithm, and a simple Arduino-controlled electronic scroll hardware.

II. RELATED WORK

Transformer [3]: Transformer is a sequence to sequence model [2], which uses many self-attention operations. After the Q, K, and V vector operations, the input feature vector can be obtained through soft-max.

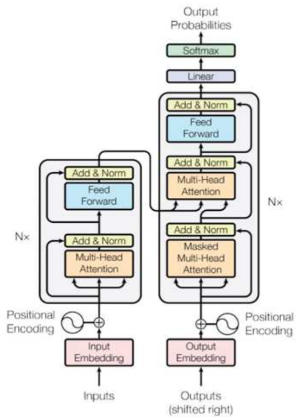

Figure 1. Transformer Architecture [3]

BERT: BERT is composed of an encoder based on Transformer, and its training will be divided into two stages, pre-train and fine-tune. In the pre-train stage, data without labels will be used for training, which is mainly divided into two training methods, Mask LM and Next Sentence Prediction. Mask LM refers to randomly masking some words, and then guessing what the word is, or randomly replacing the mask with a certain word, and then guessing whether it is correct. Next Sentence Prediction refers to entering two paragraphs of text, and then checking whether there is a contextual relationship between the two. Fine-tune, as shown in Figure 2, means that a classifier is connected to the pre-train model, and this classifier will be trained with labeled data. Also, Bert fine-tunes the parameters of the pre-train model during training.

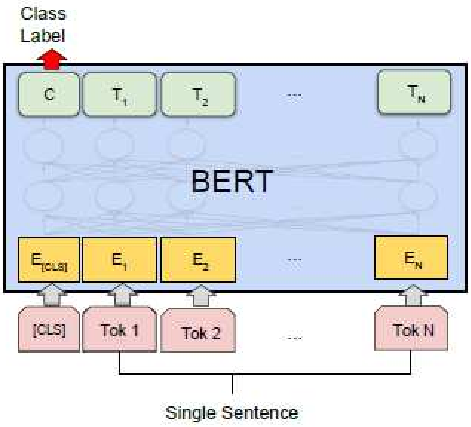

Figure 2. Fine-tuning BERT on single sentence classification task [1]

III. PROPOSED CENSOR ARCHITECTURE

Figure 3 is our proposed censor architecture. First, a short Chinese sentence will be input, and then the trained BERT model will make the first sentiment analysis. If it is judged as negative or neutral at this stage, it will be output directly. If the input is more than 7 characters long and judged as positive, it will also be output directly too, but if the input is less than 7 characters, it will enter the stage of word segmentation and then make additional judgments. The addition of word segmentation processing is to improve BERT's poor judgment of sarcastic sentences that contain both good and bad words.

The word segmentation process uses the python package of jieba[4], which can segment Chinese sentences, and then re-judge each segmentation words with our trained BERT model. If the words are still positive after this judgment, the input will be judged as positive, and if there are negative words in the judgment of these word segmentations, the input will be judged as neutral, and it will be reviewed by humans in the future.

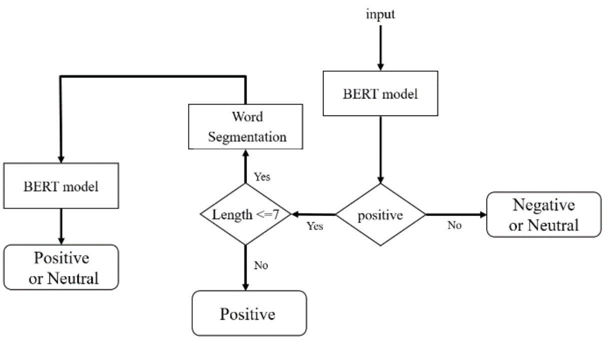

Figure 3. Proposed Censor Architecture

There is a Chinese pre-training model [6]. The overall fine tune process takes about 2 to 3 hours. Parameters are shown in TABLE I. The training data set has 28,410 samples, the validation set has 7,131 samples. The test set contains daily short sentences such as news headlines and google restaurant messages, a total of 204 samples, including 59 positives, 61 Neutrals and 84 Negatives. We focus on how many of the predicted positives are really positive.

TABLE I Parameters

Parameter	Value
Pre-process model	bert_zh_preprocess[5]
Pre-train model	bert_zh_L-12_H-768_A-12[6]
optimizer	adamw
Batch size	16
Epochs	8
Learning rate	2e-5

Figure 4 is the training accuracy curve, red is the training set, blue is the validation set, the training set starts to converge at the 8th epochs, and the validation set has maintained an accuracy of around 0.86.

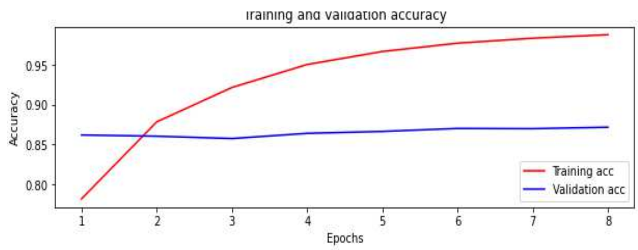

Figure 4. Training and validation set training accuracy

TABLE II and TABLE III are confusion matrices with and without word segmentation. Table IV is the result from a technology company in Taiwan that develops NLP applications. It provides a free sentiment analysis API [7] for open usage.

TABLE II. Confusion matrix with word segmentation

Confusion matrix (Unit: number)		Predict: Positive	Predict: Neutral	Predict: Negative
Ground Truth	Positive	***46***	13	0
	Neutral	***14***	42	5
	Negative	***1***	17	66

TABLE III Confusion matrix without word segmentation

Confusion matrix (Unit: number)		Predict		
		Positive	Neutral	Negative
Ground Truth	Positive	***48***	11	0
	Neutral	***15***	41	5
	Negative	***2***	16	66

TABLE IV. Confusion Matrix by Droidtown Linguistic Tech

Confusion matrix (Unit: number)		Predict		
		Positive	Neutral	Negative
Ground Truth	Positive	***38***	21	0
	Neutral	***13***	40	8
	Negative	***3***	38	43

Since we focus on how many of the predicted positives are really positive, the precision of predicted positive is important and emphasized in bold in TABLE V. The precision definition is as follows.

$$\text{Precision} = \frac{y_t}{y_t + y_0}$$

y_t means the predicted quantity is same as the ground truth quantity, y_0 means the predicted quantity is different from the ground truth quantity.

TABLE V Precision compare

	Positive	Neutral	Negative
With segmentation	***0.754***	0.583	0.929
Without segmentation	***0.738***	0.603	0.929
Droidtown Linguistic Tech	***0.703***	0.404	0.843

IV. RESULT AND CONCLUSIONS

Figure 5 is the prediction result of the test set, which does not contain sarcastic sentences, and it shows that the sentences of obvious praise or criticism can be judged correctly.

```
1   台灣好茶揚名國際(Taiwan's good tea is famous internationally)
1   荔枝風味大滿足(Satisfying the lychee flavor)
1   在地深耕打造好品質(Deep plowing on the ground to create good quality)
-1  換湯不換藥(Change the soup but not the medicine)
-1  科技部對中舉動遺憾(Ministry of Science and Technology regrets China's move)
-1  質疑學校作業疑點重重(Doubtful about school homework)
0   瀋陽首度跌出榜外(Shenyang fell out of the list for the first time)
0   室內外兩場免費活動暑假登場(Two free indoor and outdoor activities will debut in summer)
0   不要謙虛，不爭取就沒機會(Don't be humble, you won't have a chance if you don't fight)
```

Figure 5. Test result, 1 for positive, -1 for negative, 0 for neutral

TABLE VI is the different results before and after adding word segmentation, and results by Droidtown Linguistic Tech. It can be seen that some sarcastic sentences contain both positive and negative words and they are judged as neutral rather than positive in our approach with segmentation..

TABLE VI sarcastic sentences

	With segmentation	Without segmentation	Droidtown Linguistic Tech
阿不就好厲害 (That's just great)	Neutral	Positive	Negative
錢人終成眷屬 (If there all rich people, love will find find a way come together.)	Neutral	Positive	Positive
祝智障生日快樂 (happy birthday to retarded)	Neutral	Positive	Positive
愚人快樂 (happy fool)	Neutral	Positive	Positive

In addition to being able to judge some sarcastic sentences, this paper also has high precision for negative short sentences. However, there are still some areas to overcome, such as some comments that contain both good and bad words. For some high-end sarcastic comments, sometimes, even humans have to think a little bit. In such cases, it is still very difficult for machines to make correct judgments.

V. SYSTEM IMPLEMENTATION

The system consists of user interface, censor, and display three main blocks. The following describes the user interface briefly. The display block is made of an Arduino controller board and many LEDs, but not described here.

Through a web framework, Django, we implement a web server with a home page as shown in Figure 6.

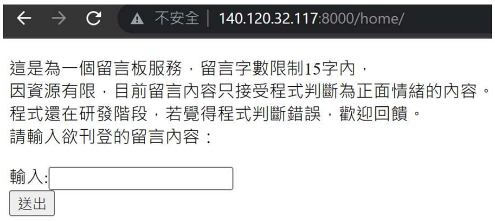

Figure 6. web framework

Once a user inputs an ad and hits submit, our censor operation will begin and return three possible results: “positive”, “negative”, and “neutral”

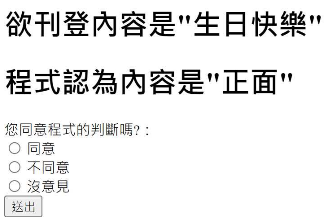

Figure 7. key in 生日快樂(happy birthday),result is positive.

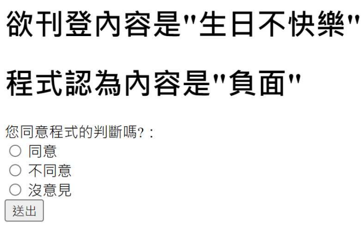

Figure 8. key in 生日不快樂(unhappy birthday),result is neutral.

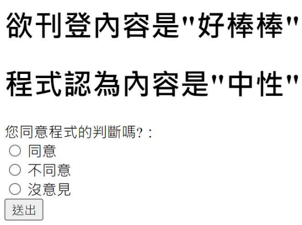

Figure 9. key in 好棒棒(very good),result is neutral.

ACKNOWLEDGEMENTS

The study is supported by Ministry of Science and Technology (MOST) under MOST 111B1275.

REFERENCES

[1] J. Ya J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, arXiv:1810.04805 [cs], May 2019, Accessed: Aug. 09, 2020. [Online]. Available: http://arxiv.org/abs/1810.04805.

[2] Harshitha Katpally; Ajay Bansal, “Ensemble Learning on Deep Neural Networks for Image Caption Generation”, In 2020 IEEE 14th International Conference on Semantic Computing (ICSC).

[3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, “Attention Is All You Need.”, In NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing.

[4] https://github.com/fxsjy/jieba

[5] https://hub.tensorflow.google.cn/tensorflow/bert_zh_preprocess/3

[6] https://tfhub.dev/tensorflow/bert_zh_L-12_H-768_A-12/4

[7] https://api.droidtown.co/#keymoji