

Article

Valence and Arousal-Infused Bi-Directional LSTM for Sentiment Analysis of Government Social Media Management

Yu-Ya Cheng ¹, Yan-Ming Chen ¹, Wen-Chao Yeh ²  and Yung-Chun Chang ^{1,*} 

¹ Graduate Institute of Data Science, Taipei Medical University, Taipei 106, Taiwan; i906108009@tmu.edu.tw (Y.-Y.C.); m946109002@tmu.edu.tw (Y.-M.C.)

² Institute of Information Systems and Applications, National Tsing Hua University, Hsinchu 300044, Taiwan; s109065801@m109.nthu.edu.tw

* Correspondence: changyc@tmu.edu.tw

Abstract: Private entrepreneurs and government organizations widely adopt Facebook fan pages as an online social platform to communicate with the public. Posting on the platform to attract people's comments and shares is an effective way to increase public engagement. Moreover, the comment functions allow users who have read the posts to express their thoughts. Hence, it also enables us to understand the users' emotional feelings regarding that post by analyzing the comments. The goal of this study is to investigate the public image of organizations by exploring the content on fan pages. In order to efficiently analyze the enormous amount of public opinion data generated from social media, we propose a Bi-directional Long Short-Term Memory (BiLSTM) that can model detailed sentiment information hidden in those words. It first forecasts the sentiment information in terms of Valence and Arousal (VA) values of the smallest unit in a text, and later fuses this into a deep learning model to further analyze the sentiment of the whole text. Experiments show that our model can achieve state-of-the-art performance in terms of predicting the VA values of words. Additionally, combining VA with a BiLSTM model results in a boost of the performance for social media text sentiment analysis. Our method can assist governments or other organizations to improve their effectiveness in social media operations through the understanding of public opinions on related issues.

Keywords: sentiment analysis; valence-arousal; social media analytics



Citation: Cheng, Y.-Y.; Chen, Y.-M.; Yeh, W.-C.; Chang, Y.-C. Valence and Arousal-Infused Bi-Directional LSTM for Sentiment Analysis of Government Social Media Management. *Appl. Sci.* **2021**, *11*, 880. <https://doi.org/10.3390/app11020880>

Received: 2 December 2020

Accepted: 14 January 2021

Published: 19 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

From entrepreneurs to government organizations, Facebook fan pages are widely adopted as an online social platform to communicate with the public. Releasing posts to attract people to comment and share is an effective way to establish public engagement. Anyone can post on a fan page, and the comment functions allow users who have read the posts to respond and express their thoughts. Hence, it enables us to understand the users' emotional feelings regarding that post by analyzing the comments. The goal of this study is to employ natural language processing (NLP) technologies to process the fan pages of large organizations and determine the public image of them. In order to efficiently analyze the enormous amount of public opinion data from social media, our research discovers the differences between actual sentiments and the selected emoji used by readers of government fan pages. For example, crying face emojis are typically used to represent negative sentiments such as sadness. However, by observing the comments left by the same viewers, we may find that they cried because they were moved, which is actually a positive emotion. Figure 1 shows the examples of one of the posts by the Facebook fan page of the Ministry of Health and Welfare of Taiwan, along with its comments. This figure displays the important fact that the emoji and text contents of the comments need in-depth analysis in order to precisely understand the emotion of the responses. Therefore, we propose a model that can efficiently identify viewers' emotional reactions after reading

posts on social media pages of a government. Our approach can support organizations in establishing more interactions with fans and creating favorable images.

The figure displays a central Facebook post from the Ministry of Health and Welfare (衛生福利部) dated April 25, 2020. The post reports 1 new confirmed case (Dunmu ship crew member) and 0 local cases. Surrounding the post are several comments expressing gratitude, support, and encouragement.

Post Content:

衛生福利部
4月25日 · 公
#衛福編編報報 發文時間：2020/04/25
! 新增1例確診，為敦睦艦隊磐石艦實習生!
#指揮中心快訊

新增確診案例通報 | 2020年4月25日

新增確診案例
1
案 429

敦睦遠訓：1例
本土個案：0例
境外移入：0例

Yu-Pei Chen和其他3萬人 852則留言 484次分享

Comments:

- The Taiwan government and people became united against the virus. 全國上下一條心.....
- Dunmu navy has worked hard! Please take care of yourself. Thank you for spending efforts for our nation, thank you. 敦睦戰士辛苦了，請安心養病好好照顧自己，謝謝您們為國家辛苦了，謝謝
- No confirmed indigenous cases in 13 consecutive days. 13天無本土確診
- Taiwan, fight on! 台灣加油
- Thanks to the command center, medical staff, and the crew of the Pangshih ship, I wish you to recover and be released from the quarantine soon. 謝謝指揮中心，醫護人員，磐石的海軍弟兄加油，希望早日康復，解除隔離
- Medical care in Taiwan is the best, it's wonderful to have you! 台灣醫療最棒，有你真好
- Thanks to Minister and command center for their hard work~ Trust and follow your epidemic prevention steps. 感謝辛苦的部長及指揮中心~信任並跟隨你們的防疫腳步
- The crew of the navy really have to cooperate well with the epidemic prevention unit! Looking forward to reaching the target of having zero local cases. 各位弟兄真的要好好配合檢疫單位! 期待000的目標

Figure 1. Examples of a post and its comments found on the Facebook fan page of the Ministry of Health and Welfare.

Sentiment analysis is one of the most active research areas among the NLP fields. In this domain, sentiments are commonly represented by several categories (e.g., positive and negative). Most previous studies address sentence- and document-level sentiment analysis. However, a huge number of informal messages are posted every second on social media, which are mostly in short text form. Due to the heavy omission of information, most machine comprehension models have difficulty in processing this kind of data compared to complete paragraphs in articles.

In light of this, we propose a Bi-directional Long Short-Term Memory (BiLSTM)-based model to perform fine-grained sentiment analysis of words to improve the efficiency in dealing with short texts. It first forecasts the sentiment information in terms of Valence and Arousal (VA) values of the smallest unit, i.e., words, in the text. Later, it fuses these features into a deep learning model to analyze the sentiment of the whole text. Additionally, combining VA with a BiLSTM model results in superb performance for social media text sentiment analysis. By using this method, we are able to classify the comments as positive or negative and also to comprehensively calculate the proportions of positive and negative comments for each post before determining the sentiment trend of public opinions for the post. Finally, we use a word cloud to visualize the public sentiments towards the post. In this way, the government can improve its effectiveness in social media operations through understanding its public image and opinion on related issues with the support of our prediction model. In recent year, similar methods have been applied to analyze other Facebook fan pages [1,2].

2. Related Work

2.1. Sentiment Analysis

Given the ubiquity of the Internet as well as the popularity of social networking platforms in recent years, the web has already become an essential field for information sharing and delivery. People also express personal opinions or discuss ongoing public issues through social networking platforms such as Facebook and Instagram. The detection of consequential information and key expression has become a popular topic in data science research. Sentiment analysis techniques are widely used to discover and extract subjective messages (e.g., point of view and manner) from the texts of social media in various fields, including especially the areas of politics, products, and movies [3]. In the political field, Shakeel and Karim [4] introduced a sentiment classification model with multiple cascades for informal short texts. The dataset under their investigation is MultiSenti, in which tweets during the 2018 general election of Pakistan were collected in order to discover the general sentiment of the public, particularly about the election process and results. Experimental results of their proposed method were compared to three previous models of multilingual sentiment classifier, and they demonstrated that their approach exceeds others overall. In addition, in the field of film, Yenter and Verma [5] proposed a CNN-LSTM (Convolutional Neural Networks and Long Short-Term Memory) model. Their model was capable of predicting the sentiment polarity of reviews from the Internet Movie Database (IMDb) dataset with accuracy above 89%. Singh et al. [6] integrated SentiWordNet features with a support vector machine to predict the polarity of movie reviews and blog posts. The result demonstrated that the vector space-based model can benefit from fusing lexical features with sentiment information. In addition, Zhang et al. [7] proposed the Weakness Finder, an expert framework that can help manufacturers identify the weakness of products from Chinese feedback by using aspect-based sentiment analysis. Using the morpheme-based method and the HowNet-based similarity measure, it extracted and grouped the features. Then, they defined and categorized the implicit features for each aspect with the collocation selection method. To evaluate the polarity of each aspect in phrases, they used the sentence-based sentiment analysis approach. The method has been used to help a manufacturer of body wash to recognize potential weaknesses of their product, and its performance was outstanding.

Based on the structural units of grammar, Turney [8] initiated a method for sentiment classification on the document level in 2002. Subsequently, classification of smaller units, such as sentences [9,10], phrases [11,12], and words [13], were developed by other researchers. It can be inferred that sentiment analysis exhibits flexibility at various levels of units in grammar. Our method manipulates the smallest one, i.e., “words”, to conduct sentiment prediction experiments, which we then extend to sentiment classification tasks for each comment on social media so as to explore the sentiment of the public towards the government. Our method is unique in that we solve the problem of directly using the emoji ratio as the overall sentiment of a post. Therefore, we can delicately detect the sentiment and perception of the public on a post. We observe that the need to understand the public image and communication of the government to the people is very prominent, so we applied this approach in the real world to analyze the novel task of analyzing posts by the Ministry of Health and Welfare.

2.2. Valence and Arousal

Yu et al. [14] proposed the approach of using Valence and Arousal for Chinese sentiment analysis. Valence and Arousal can be represented by a two-dimensional vector space. Valence is a measure of the degree of sentiment, and it ranges from 1 to 9, which represents negative to positive emotion, respectively. Arousal is the level of emotional agitation represented by values also from 1 to 9, with higher ones representing more agitated emotion and lower ones representing calm. Yu et al. [15] proposed a weighted graph model that considers both the relations of multiple nodes and their similarities as weights to automatically determine the VA ratings of affective words. Experiments on Chinese

affective lexicons show that this method yielded a smaller error rate on VA prediction than the linear regression, kernel method, and PageRank algorithms used in previous studies.

Subsequently, the Dimensional Sentiment Analysis for Chinese Phrases (DSAP) shared task was held at the 8th International Joint Conference on Natural Language Processing (IJCNLP 2017). It consisted of two subtasks, including the evaluations of words and phrases. There were 2802 words in the training dataset and 750 in the testing dataset, and they have been merged into the Chinese Valence-Arousal Words (CVAW) 3.0 (<http://nlp.innobic.yzu.edu.tw/resources/cvaw.html>) corpus. Figure 2 shows a scatter plot of 3552 words in CVAW 3.0 and their corresponding Valence and Arousal values. For example, with the word “joy”, which has V: 7.4 and A: 6.2, a Valence value of 7.4 shows that the word’s tendency is towards the positive emotions, and an Arousal value of 6.2 further indicates that the emotion is not overly agitated but is also not completely calm. In another example, the word “lost”, with V: 3 and A: 3.3, signals a negative and subtle mood. To predict Valence and Arousal ratings, Wu et al. [16] suggested an approach using a densely linked LSTM network and word features. Notably, they used word embedding along with part-of-speech (POS) and word clusters as additional features to train the LSTM. The findings indicated that the system’s ability in predicting valence and arousal dimensions for Chinese words is outstanding. Zhou et al. [17] suggested a framework that primarily employs multi-layer neural networks, with a stack of input features such as word embedding, POS-tags (POST), word clustering, prefix, character embedding, and cross sentiment information. The model was optimized with AdaBoost. This method was ranked 2nd in the final round. Li et. al. [18] combined three models, namely, E-HowNet-based, embedding-based, and character-based models, to construct a model for this competition.

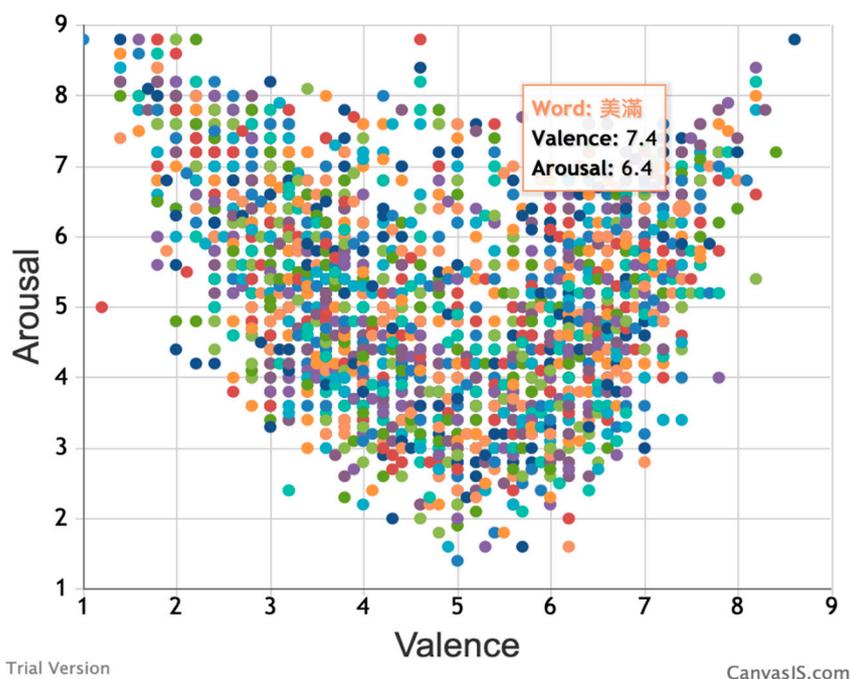


Figure 2. Scatter plot of VA values of words in the Chinese Valence-Arousal Words (CVAW) 3.0 dataset.

Some recent research has improved the effectiveness of sentiment analysis; for example, Chang et al. [19] described an approach that can strengthen the model by offering more refined emotional knowledge to improve the effectiveness of film recommendation systems. Wang et al. [20] devised a tree-structured regional CNN-LSTM network made up of two components: regional CNN and LSTM. Experimental findings on various datasets indicated that this approach exceeded the lexicon- and regression-based models, as well

as the different neural networks previously proposed. To sum up, Valence and Arousal measures can more clearly detect the emotional content of each word, as well as effectively improve sentiment analysis models. Thus, this study builds on previous work and employs a VA-infused Bi-directional LSTM network for sentiment analysis of government social media content.

3. Materials and Methods

In this research, we propose a novel method for sentiment analysis of government social media content that considers finer emotions of words through the VA representation model and deep neural network. Figure 3 illustrates our framework. First, we collected the posts and corresponding comments from the Facebook fan page of the Ministry of Health and Welfare as the target of our experiments. After preprocessing, we performed Valence and Arousal prediction on every word. These predictions were subsequently incorporated into a BiLSTM network to form the classification model. Finally, the data visualization component was adopted to summarize the sentiment trend of posts on social media. We will explain the implementation detail of each component in the following sections.

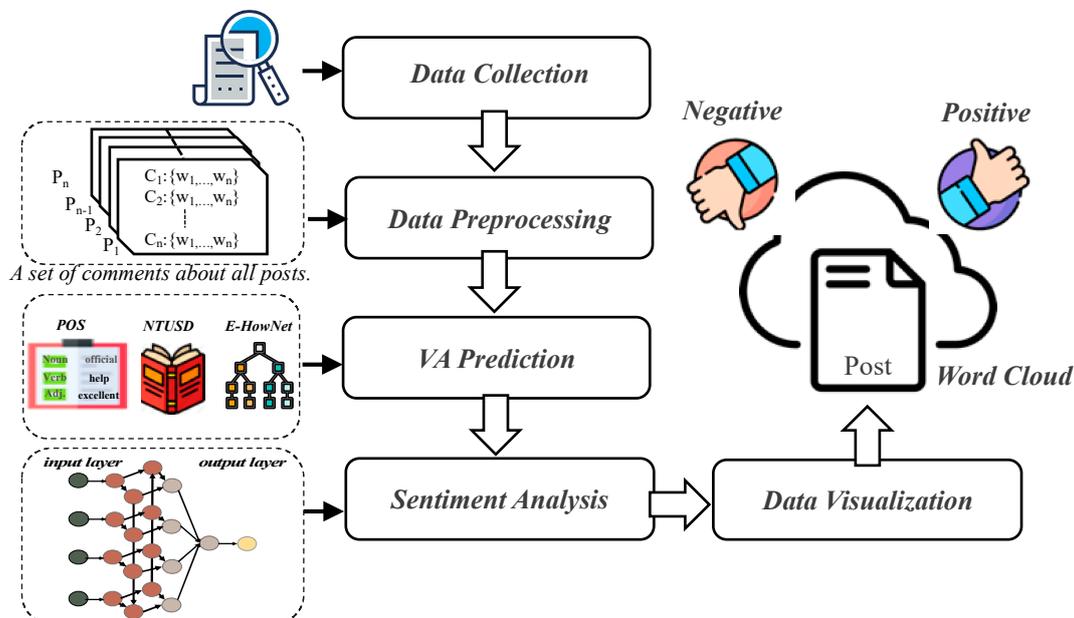


Figure 3. Overall workflow of the research.

3.1. Valence-Arousal Analysis

This section introduces the features used in the Valence-Arousal analysis step. It can be divided into three parts: (1) The part-of-speech (POS) labels of Chinese morphemes. (2) Morpheme classification according to the National Taiwan University Sentiment Dictionary (NTUSD) [21] (3) Transform the morphemes into concept tags using the E-HowNet [22] and calculate the similarity to obtain the Valence and Arousal values as our features. Due to the characteristics of Chinese writing conventions, spaces are not used to separate words as they are in English. So, an important step in Chinese NLP is to first conduct tokenization. To do that, this study employed Multi-Objective NER POS Annotator (MONPA) [23], which was trained on (Traditional) Chinese corpus for this purpose. It provided segmentation, POS tags, and named entity recognition (NER) functionality specifically for Traditional Chinese. We first acquired the POS of each Chinese word in the dataset through the MONPA system, and supplied the POS features to the proposed model structure. Then, we generated 50-dimension vectors as the representation for word sentiment prediction.

Besides adding the POS feature vectors of each word, we employed a sentiment dictionary to further empower our model. The NTUSD is a sentiment dictionary that has been established by the Natural Language Processing Laboratory from National Taiwan University, and it includes 2810 positive and 8276 negative Chinese words. For example, positive words include “harmony”, “confidence”, “contentment”, etc.; negative words include “despiteful”, “insecurity”, “unrealistic”, etc. We used the proposed model to perform binary classification by introducing NTUSD and determined the Valence and Arousal values of words. Specifically, when a word appeared in the positive dictionary, we set “1” as its category; when it appeared in the negative dictionary, we set “−1” as its category. Moreover, we calculated the cosine similarity and sorted the values for those unknown words that were not included in NTUSD. We selected the top five words in NTUSD that were most similar to the unknown words, which allowed us to complete the word classification task based on the positive or negative labels in NTUSD. We employed this word classification method to capture more sentimental features.

Additionally, the E-HowNet, which is built upon the HowNet knowledge, was applied to convert the meaning of a word into more abstract concepts and to systematically express the word with a concept. For example, the concept of “indignant” is the same as “resentful”. Additionally, the concept of “faith” is identical to “trust”. Furthermore, we calculated the average of the Valence and Arousal values for the words that shared the same concept in order to obtain meaning features of the words. If there were no matching concepts that could be found, cosine similarity was computed to discover the closest concept, and the values of Valence and Arousal were represented by their mean value. In this way, two-dimensional features of abstractive semantics could be acquired.

Finally, embeddings from Bidirectional Encoder Representations from Transformers (BERT) [24] were incorporated with the features mentioned above, which were able to depict more complex sentiments of different words in the context. In the end, there were a total of 821 dimensions in the final representative vectors, which were a combination of BERT embedding of the vocabularies, one by one, and associated with the POS feature vectors, sentimental features in NTUSD, and E-HowNet.

Bi-directional Long Short-Term Memory can not only learn dependency within a greater distance, but can also better incorporate bi-directional information through the combination of two LSTMs [25], one from front to back and the other in the opposite direction. Furthermore, the Attention layer is capable of learning the different weights of each part of the input and extracting more crucial information [26]. In recent years, the ensemble learning approach has been widely used for boosting classification performance [27–29]. Its popularity is attributable to its outstanding performance in comparison to single learners, while it is also relatively easy to deploy in industrial applications. In light of this, an ensemble learning method was adopted to train the models and obtain representative sentiment outputs by assembling the models of different levels. The ensemble results from the 3rd to the 5th BiLSTM layers and the Attention layer were utilized for the Valence, and the ensemble results from the 1st to the 3rd BiLSTM layers and the Attention layer were employed for the Arousal. The structures of the predicted values of Valence and Arousal by the ensemble learning models are shown in Figure 4.

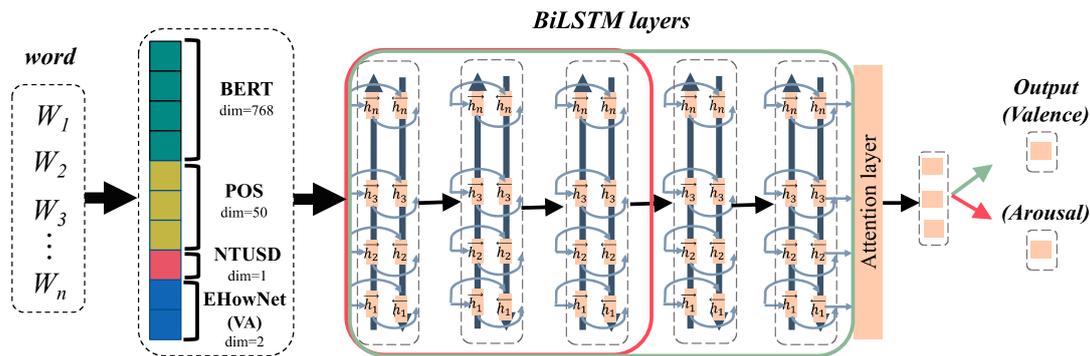


Figure 4. The sentimental Valence-Arousal prediction model based on deep neural networks and comprehensive semantic features.

3.2. Sentiment Prediction of Social Media Comments

The workflow of the proposed framework for sentiment analysis on social media is displayed in Figure 5. The VA values of all words in a social media comment were estimated using the method mentioned in the previous section, and they were finally assembled to estimate the sentimental orientation of the whole comment. We introduce the workflow, step by step, in detail below. First of all, using MONPA, we carried out tokenization on the comments and removed stop words such as “Is”, “Instead of”, “To”, etc. Next, the words were represented by vectors using BERT [24]. The predicted values of Valence and Arousal ranged from 1 to 9. In order to better identify the sentiment of the messages, Log-Likelihood Ratio (LLR) estimation was introduced in this research to score the most representative words for the positive and negative categories. Specifically, LLR utilized Equation (1) to obtain the likelihood of the hypothesis, which states that the existence of a positive term in a sentence is beyond chance in a large training corpus of sentences. It is a promising method for the selection of crucial word features.

$$-2\log \left[\frac{p(w)^{N(w \wedge PS)} (1 - p(w))^{N(PS) - N(w \wedge PS)} p(w)^{N(w \wedge \neg PS)} (1 - p(w))^{N(\neg PS) - N(w \wedge \neg PS)}}{p(w|PS)^{N(w \wedge PS)} (1 - (w|PS))^{N(PS) - N(w \wedge PS)} p(w|\neg PS)^{N(w \wedge \neg PS)} (1 - (w|\neg PS))^{N(\neg PS) - N(w \wedge \neg PS)}} \right] \quad (1)$$

The PS in Equation (1) indicates the collection of positive sentences within the training samples (sentences). $N(PS)$ and $N(\neg PS)$ indicate the quantity of positive and negative samples. $N(w \wedge PS)$ denotes the number of positive sentences that contain the positive word w . We utilized maximum likelihood estimation to obtain probabilities $p(w)$, $p(w|PS)$, and $p(w|\neg PS)$. After this calculation, a term with a higher LLR can be considered to have a higher connection with a specific sentiment. We selected the top 200 highest ranked words as additional features for the input to the model.

Through the LLR calculation, the weight of each word in the vocabulary of the positive and negative categories were obtained, and the first 200 keywords for each category were used to represent the Valence and Arousal value of the original text in the following manner. The top three words with the highest weight were selected, and their Valence and Arousal were used to represent the sentence. If there were less than three, the words with the highest VA-value were selected as an alternative. Therefore, a six-dimensional feature vector could be acquired. The feature vectors were incorporated with the BERT embeddings vector, which is the representative vector at the sentence level, to obtain a 744-dimensional vector as the final representation of this sentence. As shown in Figure 5, the above-mentioned vectors were sent into the Bi-LSTM and Attention model to obtain the binary sentiment classification results.

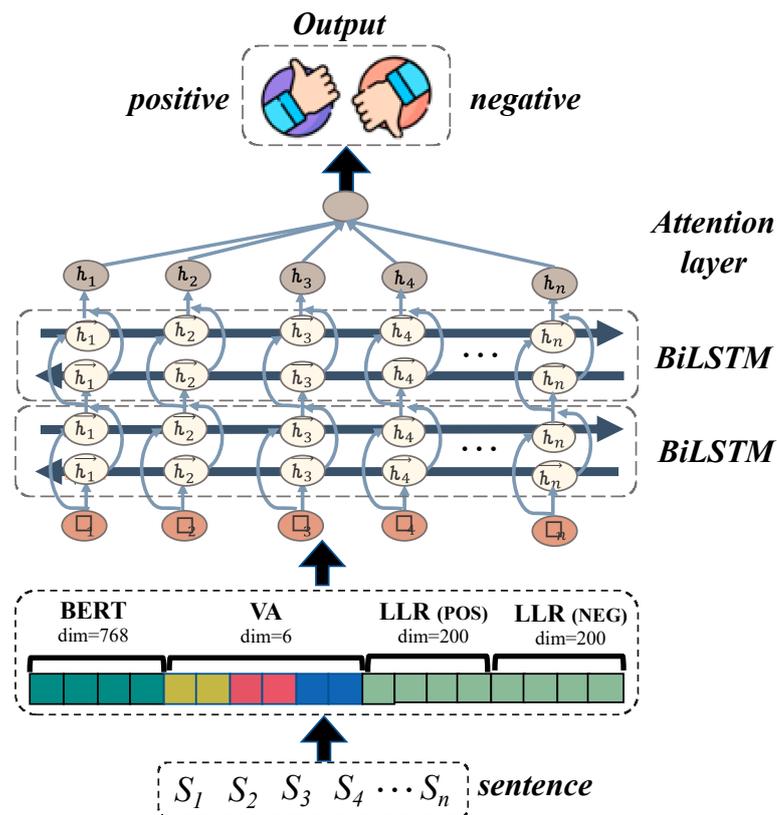


Figure 5. The sentiment prediction model based on deep neural network and comprehensive features.

4. Experimental Results and Discussion

We first assessed the efficiency of various models on predicting VA values of a word on the official data from the Dimensional Sentiment Analysis for Chinese Phrases (DSA_P) Competition, which included 2802 training and 750 testing instances. In our experiments, we used a dropout rate of 0.2, a batch size of 64, and the Adam optimizer with 0.01 learning rate. The VA prediction model was implemented using Keras (<https://keras.io/>). Mean Absolute Error (MAE) and Pearson Correlation Coefficient (PCC) were the two metrics employed to validate these approaches. MAE's definition is as listed in Equation (2). It aims at reflecting the overall difference between actual and estimated values. Therefore, a smaller MAE indicates a better estimate. PCC provides the correlation between these values, with a range of -1 to 1 . A PCC that is close to 1 indicates a higher correlation between the two numbers; if a value is within 0 to 0.09 , it indicates no correlation; 0.1 to 0.3 shows a low correlation; 0.3 to 0.5 a medium correlation; and greater than 0.5 indicates a clear correlation between these numbers. In Equation (3), we denote PCC as r . A_i refers to the correct response, P_i refers to the outcome of the model (valence or arousal values), and n refers to the count of test samples. \bar{A} and \bar{P} are the arithmetic mean, respectively, of the previous A_i and P_i , and σ is the standard deviation.

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - P_i| \quad (2)$$

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{A_i - \bar{A}}{\sigma_A} \right) \left(\frac{P_i - \bar{P}}{\sigma_P} \right) \quad (3)$$

As shown in Table 1, the proposed method is compared with the top three competitors in the DSA_P word Valence and Arousal prediction task [30], i.e., THU_NGN, AL_I_NLP, and CKIP. Our method is outstanding in predicting the dimensional sentiment of Chinese words, achieving a comparable performance with the highest-ranking competitor (i.e.,

THU_NGN). For the valence values, MAE is reduced by 0.543 and PCC increased to 88.7%. This indicates a very high correlation between the model prediction and the correct values. Similar observations can be made regarding the arousal value prediction task. The MAE is considerably lowered by 0.855, and PCC is increased to 68.9%.

Table 1. Performance comparison of our method and the top-ranking teams in the Valence and Arousal value prediction task in the Dimensional Sentiment Analysis for Chinese Phrases (DSA_P) Competition, 2017.

Group	Valence		Arousal	
	MAE	PCC	MAE	PCC
THU_NGN	0.509	0.908	0.864	0.686
Our method	0.543	0.887	0.855	0.689
AL_I_NLP	0.545	0.892	0.857	0.678
CKIP	0.602	0.858	0.949	0.576

In the second experiment, we utilized the outputs from the prediction of valence-arousal model and integrated them into a hybrid Deep Neural Networks (DNN) for the classification of the overall sentiment of a comment on social media. Because the goal of this research is to recognize public opinion to assist government social media management, only coarse-grained sentiment categories (positive and negative) were considered in this experiment. Moreover, in order to thoroughly prove the effectiveness of the proposed model, we conducted two experiments. First, we validated the dataset from the Natural Language Processing and Chinese Computing (NLPC) 2014 competition (http://tcci.ccf.org.cn/conference/2014/pages/page04_sam.html) of sentiment classification for Chinese product reviews, which contained multiple domains such as books, DVDs, and electronics. There were 10,000 examples for training and 2500 for testing. In this experiment, we compared the proposed model with the best teams in this competition, which is denoted as NNLM [31]. In order to demonstrate the generalization ability of the proposed model, we conducted the second experiment using the E-commerce service review dataset (ECSR) (<https://github.com/renjunxiang/Text-Classification>). This dataset consists mainly of the review comments for TV products and distribution services collected from several E-commerce websites. In this dataset, the average length of review comments was 72 words. Each review was given a sentiment tag: *positive or negative*. The data contained a total of 4212 reviews, in which 1883 were positive and 2329 were negative. We performed 10-fold cross validation to examine the performance. Here, we set the dropout rate as 0.25 and the batch size as 32, and we used the Adam optimizer with a learning rate of 0.01. Our proposed BiLSTM-based sentiment classification approach was implemented using Keras. We used precision, recall, and F₁-measure for our evaluations [32]. Furthermore, we calculated the macro average of these metrics for the overall comparison. Precisely, letting C_i be the corpus in our studies, we calculated precision, recall, and F₁-measure P(C_i), R(C_i), F₁(C_i), and micro-average F^μ, as in Equations (4)–(7).

$$P(C_i) = \frac{TP(C_i)}{TP(C_i) + FP(C_i)} \quad (4)$$

$$R(C_i) = \frac{TP(C_i)}{TP(C_i) + FN(C_i)} \quad (5)$$

$$F_1(C_i) = \frac{2 \times P(C_i) \times R(C_i)}{P(C_i) + R(C_i)} \quad (6)$$

$$F^\mu(C_i) = \frac{\sum_{i=1}^n 2 \times P(C_i) \times R(C_i)}{\sum_{i=1}^n P(C_i) + R(C_i)} \quad (7)$$

where $TP(C_i)$ indicates the quantity of correct positive cases and $FP(C_i)$ denotes the number of false positives (namely, negative cases that are wrongly classified as positives). Analogously, $TN(C_i)$ and $FN(C_i)$ indicate the number of true negatives and false negatives, respectively. To systematically assess the relative effectiveness of the compared methods, the F_1 value is also used.

Next, we evaluated the performance of the embeddings to demonstrate the effectiveness of our novel text representation method. Table 2 shows the gain in performance after applying LLR and VA, denoted as $Em_{BERT+LLR} + BiLSTM_Att$ and $Em_{BERT+LLR+VA} + BiLSTM_Att$, respectively. To provide an all-inclusive performance evaluation, we compared our method to the state-of-the-art system (denoted as NNLM) of the NLPCC 2014 dataset. As shown in this table, the $Em_{BERT} + BiLSTM_Att$ can achieve about 74% and 87% F_1 -score on NLPCC 2014 and ECSR datasets, respectively. By using LLR features, we further improved the system performance because it successfully discriminates words that are highly correlated with a certain emotion, thereby boosting the BiLSTM's ability to find representative lexicons of sentiment. Moreover, the VA features improve the F_1 score significantly. It indicates that when the model considers both valence value as the polarity and arousal value as the strength of sentiment, it can greatly improve the effectiveness of sentiment analysis. Notably, our method outperforms the comparison in each and every category. This is because our method infuses emotion-specific VA features to BiLSTM with the attention mechanism, thereby effectively enhancing its ability to correctly identify the sentiment of Chinese product reviews. According to the above experiment results, our method can indeed improve the performance by providing more detailed emotional knowledge to enhance the effectiveness of sentiment classifiers, and thus it achieves remarkable performances in different types of sentiment classification datasets. The source codes of the proposed method and comparisons can be found in GitHub (<https://github.com/yuyachengg/sentiment>).

Table 2. Performances of sentiment analysis models on different datasets.

Dataset	Evaluation (Precision/Recall/ F_1)	Method			
		NNLM	$Em_{BERT} +$ $BiLSTM_Att$	$Em_{BERT+LLR} +$ $BiLSTM_Att$	$Em_{BERT+LLR+VA} +$ $BiLSTM_Att$
NLPCC 2014	Positive	0.758/0.789/0.773	0.723/0.771/0.746	0.762/0.762/0.762	0.771/0.815/0.792
	Negative	0.780/0.748/0.764	0.754/0.701/0.726	0.764/0.763/0.764	0.782/0.764/0.773
	Macro-Avg	0.769/0.769/0.769	0.738/0.736/0.736	0.763/0.763/0.763	0.776/0.789/0.782
ECSR	Positive	-	0.853/0.853/0.853	0.901/0.913/0.907	0.932/0.903/0.917
	Negative	-	0.881/0.881/0.881	0.914/0.914/0.914	0.924/0.952/0.938
	Macro-Avg	-	0.867/0.867/0.867	0.907/0.914/0.911	0.928/0.928/0.928

The above experiments quantitatively evaluate the performance of the proposed method. To gain a deeper insight into the Facebook fan page of the Ministry of Health and Welfare, we carried out a case study specifically on the posts in Apr. 2020. During this period, there was no local confirmed case of COVID-19 for 13 consecutive days in Taiwan; however, one case was confirmed among the *Dunmu* navy crew employed by the Ministry of National Defense. The Ministry of Health and Welfare released 16 posts (among a total of 5744 posts) through its Facebook fan page, which attracted lively discussion from netizens. We first predicted the sentiment behind the text of posts and comments. Next, we used word clouds to visualize the categorized positive and negative keywords for each sentiment and color-coded them for clarity. It is intended for the readers to easily associate sentiments with their corresponding terms. The word cloud was built from the top fifty words with higher LLR values in each of the positive and negative sentiment categories. Moreover, we used larger font sizes for words with higher LLR weights. Figure 6 shows the resulting word cloud. We can observe that the polarity of sentiments in comments can be

on epidemic prevention measures have emerged among the community after the COVID-19 epidemic seemed to be calming in Taiwan. Moreover, we have also noticed negative words such as “Injury, Set sail, Mothership,” and “Diamond Princess.” This is because some people have accused the military personnel that participated in the Dunmu training mission held by the Ministry of National Defense: “There are already confirmed cases of the Diamond Princess, why the ship still insists on going out to sea?” On the other hand, many people wrote words of encouragement: “The navy has already done well, because of the epidemic prevention control on the shipboard, the virus did not cause serious damage on the ship. In addition, it has successfully completed the mission that has lasted more than a month, which was much better than two aircraft carriers belonging to the U.S. and French and also a group of cruise passengers.” After the above discussion, we have proven that the proposed method in this work can effectively and professionally analyze public opinion and can further understand the content of the speech in fine detail. It can go further in assisting the government’s management of its social media account, thereby improving the image of the government as well as building more favorable interactions with the people.

5. Conclusions

A set of sentiment analysis approaches consisting of BiLSTM and attention layers based on Valence-Arousal information was proposed in this research. It demonstrates the capability of predicting the sentiment behind the text on Chinese social media. This study verifies the efficacy of our method through the data from various competitions, and the experimental results show that our method is superior to the methods used by other teams in sentiment prediction of Valence-Arousal or Chinese short text. Furthermore, through visualization, it is possible to grasp the content and the trend of public opinion. Such results can even aid large organizations and governments in decision-making and management of their social media presence. Overall, our method is superb at forecasting word-level sentiment, and it can be used to observe social phenomenon where people in general have different points of view on the same issue. It is verified that our method can easily identify problems and analyze the image of government or business units. These results can help improve the management of social media accounts in the future and provide valuable insights.

Author Contributions: Conceptualization, Y.-C.C.; methodology, Y.-Y.C., Y.-C.C.; software, Y.-C.C.; validation, Y.-Y.C., Y.-M.C.; formal analysis, Y.-Y.C.; investigation, Y.-Y.C., Y.-M.C.; resources, Y.-C.C.; data curation, Y.-Y.C., Y.-M.C., W.-C.Y.; writing—original draft preparation, Y.-C.C., Y.-Y.C., Y.-M.C.; writing—review and editing, Y.-C.C., W.-C.Y.; visualization, Y.-M.C.; supervision, Y.-C.C.; project administration, Y.-C.C.; funding acquisition, Y.-C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ministry of Science and Technology of Taiwan under grant MOST 109-2410-H-038-012-MY2 and MOST 107-2410-H-038-017-MY3.

Data Availability Statement: The data of Valence and arousal can be found in website CVAW (<http://nlp.innobic.yzu.edu.tw/resources/cvaw.html>), and other data can be found in GitHub (<https://github.com/yuyachengg/sentiment>).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zavattaro, S.M.; French, P.E.; Mohanty, S.D. A sentiment analysis of US local government tweets: The connection between tone and citizen involvement. *Gov. Inf. Q.* **2015**, *32*, 333–341. [[CrossRef](#)]
2. Chen, Q.; Min, C.; Zhang, W.; Wang, G.; Ma, X.; Evans, R. Unpacking the black box: How to promote citizen engagement through government social media during the COVID-19 crisis. *Comput. Hum. Behav.* **2020**, *110*, 106380. [[CrossRef](#)] [[PubMed](#)]
3. Liu, B. Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* **2012**, *5*, 1–167. [[CrossRef](#)]
4. Shakeel, M.H.; Karim, A. Adapting deep learning for sentiment classification of code-switched informal short text. In Proceedings of the 35th Annual ACM Symposium on Applied Computing, Brno, Czech Republic, 30 March 2020; pp. 903–906.

5. Yenter, A.; Verma, A. Deep CNN-LSTM with combined kernels from multiple branches for IMDb review sentiment analysis. In Proceedings of the 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York, NY, USA, 19–21 October 2017; pp. 540–546.
6. Singh, V.K.; Piryani, R.; Uddin, A.; Waila, P. Sentiment analysis of Movie reviews and Blog posts. In Proceedings of the 2013 3rd IEEE International Advance Computing Conference (IACC), Ghaziabad, India, 22–23 February 2013; pp. 893–898.
7. Zhang, W.; Xu, H.; Wan, W. Weakness Finder: Find product weakness from Chinese reviews by using aspects-based sentiment analysis. *Expert Syst. Appl.* **2012**, *39*, 10283–10291. [[CrossRef](#)]
8. Turney, P.D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *arXiv* **2002**, arXiv:cs/0212032.
9. Kim, S.M.; Hovy, E. Determining the sentiment of opinions. In Proceedings of the 20th International Conference on Computational Linguistics, Geneva, Switzerland, 23–27 August 2004; pp. 1367–1373.
10. Hu, M.; Liu, B. Mining and summarizing customer reviews. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 22–25 August 2004; pp. 168–177.
11. Wilson, T.; Wiebe, J.; Hoffmann, P. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Vancouver, BC, Canada, 6–8 October 2005; pp. 347–354.
12. Agarwal, A.; Biadys, F.; Mckeown, K. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. In Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), Athens, Greece, 30 March–3 April 2009; pp. 24–32.
13. Sayeed, A.; Boyd-Graber, J.; Rusk, B.; Weinberg, A. Grammatical structures for word-level sentiment detection. In Proceedings of the 2012 Conference of the North American Chapter of the Association for computational Linguistics: Human Language Technologies, Montreal, QC, Canada, 3–8 June 2012; pp. 667–676.
14. Yu, L.C.; Lee, L.H.; Hao, S.; Wang, J.; He, Y.; Hu, J.; Lai, K.R.; Zhang, X. Building Chinese affective resources in valence-arousal dimensions. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 540–545.
15. Yu, L.C.; Wang, J.; Lai, K.R.; Zhang, X.J. Predicting valence-arousal ratings of words using a weighted graph method. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Beijing, China, 26–31 July 2015; pp. 788–793.
16. Wu, C.; Wu, F.; Huang, Y.; Wu, S.; Yuan, Z. Thu_ngn at ijcnlp-2017 task 2: Dimensional sentiment analysis for chinese phrases with deep lstm. In Proceedings of the IJCNLP 2017, Shared Tasks, Taipei, Taiwan, 27 November–1 December 2017; pp. 47–52.
17. Zhou, X.; Wang, J.; Xie, X.; Sun, C.; Si, L. Alibaba at IJCNLP-2017 Task 2: A Boosted Deep System for Dimensional Sentiment Analysis of Chinese Phrases. In Proceedings of the IJCNLP 2017, Shared Tasks, Taipei, Taiwan, 27 November–1 December 2017; pp. 100–104.
18. Li, P.H.; Ma, W.Y.; Wang, H.Y. CKIP at IJCNLP-2017 Task 2: Neural Valence-Arousal Prediction for Phrases. In Proceedings of the IJCNLP 2017, Shared Tasks, Taipei, Taiwan, 27 November–1 December 2017; pp. 89–94.
19. Chang, Y.C.; Yeh, W.C.; Hsing, Y.C.; Wang, C.A. Refined distributed emotion vector representation for social media sentiment analysis. *PLoS ONE* **2019**, *14*, e0223317. [[CrossRef](#)] [[PubMed](#)]
20. Wang, J.; Yu, L.C.; Lai, K.R.; Zhang, X. Tree-structured regional CNN-LSTM model for dimensional sentiment analysis. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2019**, *28*, 581–591. [[CrossRef](#)]
21. Ku, L.W.; Chen, H.H. Mining opinions from the Web: Beyond relevance retrieval. *J. Am. Soc. Inf. Sci. Technol.* **2007**, *58*, 1838–1850.
22. Chen, W.T.; Lin, S.C.; Huang, S.L.; Chung, Y.S.; Chen, K.J. E-HowNet and automatic construction of a lexical ontology. In Proceedings of the Coling 2010: Demonstrations, Beijing, China, 23–27 August 2010; pp. 45–48.
23. Hsieh, Y.L.; Chang, Y.C.; Huang, Y.J.; Yeh, S.H.; Chen, C.H.; Hsu, W.L. MONPA: Multi-objective named-entity and part-of-speech annotator for Chinese using recurrent neural network. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), Taipei, Taiwan, 27 November–1 December 2017; pp. 80–85.
24. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
25. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
26. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, 12 June 2017; pp. 5998–6008.
27. Moreno, J.G.; Boros, E.; Doucet, A. TLR at the NTCIR-15 FinNum-2 Task: Improving Text Classifiers for Numeral Attachment in Financial Social Data. In Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies, Tokyo, Japan, 8–11 December 2020.
28. Gomes, H.M.; Barddal, J.P.; Enembreck, F.; Bifet, A. A survey on ensemble learning for data stream classification. *ACM Comput. Surv. (CSUR)* **2017**, *50*, 1–36. [[CrossRef](#)]
29. Dong, X.; Yu, Z.; Cao, W.; Shi, Y.; Ma, Q. A survey on ensemble learning. *Front. Comput. Sci.* **2020**, *14*, 241–258. [[CrossRef](#)]
30. Yu, L.C.; Lee, L.H.; Wang, J.; Wong, K.F. IJCNLP-2017 Task 2: Dimensional Sentiment Analysis for Chinese Phrases. In Proceedings of the IJCNLP 2017, Shared Tasks, Taipei, Taiwan, 27 November–1 December 2017; pp. 9–16.

-
31. Wang, Y.; Li, Z.; Liu, J.; He, Z.; Huang, Y.; Li, D. Word vector modeling for sentiment analysis of product reviews. In *Communications in Computer and Information Science, Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Shenzhen, China, 5–9 December 2014*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 168–180.
 32. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, MA, USA, 2008.