Soft Computing (2020) 24:8187–8197
https://doi.org/10.1007/s00500-019-04402-8

FOCUS

Exploration of social media for sentiment analysis using deep learning

Liang-Chu Chen[1] · Chia-Meng Lee[1] · Mu-Yen Chen[2]

Published online: 8 October 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract
With the rapid growth of web content from social media, such studies as online opinion mining or sentiment analysis of text have started receiving attention from government, industry, and academic sectors. In recent years, sentiment analysis has not only emerged under knowledge fusion in the big data era, but has also become a popular research topic in the area of artificial intelligence and machine learning. This study used the Militarylife PTT board of Taiwan's largest online forum as the source of its experimental data. The purpose of this study was to construct a sentiment analysis framework and processes for social media in order to propose a self-developed military sentiment dictionary for improving sentiment classification and analyze the performance of different deep learning models with various parameter calibration combinations. The experimental results show that the accuracy and F1-measure of the model that combines existing sentiment dictionaries and the self-developed military sentiment dictionary are better than the results from using existing sentiment dictionaries only. Furthermore, the prediction model trained using the activation function, Tanh, and when the number of Bi-LSTM network layers is two, the accuracy and F1-measure have an even better performance for sentiment classification.

Keywords Sentiment analysis · Social media · Deep learning · LSTM · Bi-LSTM

1 Introduction

The development of information technology and the widespread use of the Internet have given rise to the Web 2.0 trend (O'Reilly and Battelle 2004). Web 2.0 emphasizes interactive connection and collaborative sharing among users in the online world so that users can directly participate in the production of Web content and generate large amounts of text content on a wide variety of topics. Social networking is a Web 2.0-based platform and application service, with social media being an important part of social networking and content sharing (Asur and Huberman 2010). The emergence and popularity of such social media platforms as Twitter, Facebook, and Weibo have confirmed the attraction of the Web 2.0 generation.

With the rapid growth of messaging and data from social media, issues such as online opinions and reviews have received attention from the government, industry, and academia. In general, the scope of social media mining covers three directions, namely users, relationships, and content (Guellil and Boukhalfa 2015). Exploring and analyzing the textual content of social media to obtain valuable information is the most important support in the field of management and scientific research (Hangya and Farkas 2017). Various scholars have researched different application areas, such as forecasting election results (Sang and Bos 2012), travel reviews and recommendations (Xiang et al. 2018), and analysis of the news diffusion effect (Ahmed and Lugovic 2019). Research on the sentiment analysis of texts based on sentiment detection or opinion mining has also received extensive attention from scholars (He et al. 2015; Al-Mansouri 2016; Balikas et al. 2017; Mäntylä et al. 2018).

Communicated by Mu-Yen Chen.

✉ Liang-Chu Chen
nctuhorse@gmail.com

Chia-Meng Lee
krrgfu@gmail.com

Mu-Yen Chen
mychen.academy@gmail.com

[1] Department of Information Management, Management College, National Defense University, Taipei City, Taiwan

[2] Department of Information Management, National Taichung University of Science and Technology, Taichung City, Taiwan

Sentiment analysis using text mining is a method that combines natural language processing and computing mechanisms to detect such attitudes as opinion orientation, concept classification, and emotional response (Pang and Lee 2008; Araque et al. 2017). Designing an effective algorithm to improve the accuracy of sentiment or opinion classification is a key element of such research. In recent years, sentiment analysis has been widely applied to textual materials to judge mainstream opinions and voices and try to determine the public's thread of thoughts. Furthermore, since research on sentiment analysis involves cross-domain characteristics and covers multi-faceted technical knowledge, sentiment analysis is not only an emerging subject under knowledge fusion in the big data era, but also a hot research topic in the area of artificial intelligence and machine learning (Hangya and Farkas 2017; Chen and Chen 2019). Mäntylä et al. (2018), who collected data related to sentiment analysis from Scopus and published a total of 6996 academic papers from 2004 to 2016, found that sentiment analysis can be applied to such areas as movies, travel, health, elections, professional knowledge, and spam. Shayaa et al. (2018) also proposed applications of data-based sentiment analysis, covering health care, finance, sports, politics, hospitality and tourism, product marketing, etc. Based on the information collected from these studies, no literature has yet addressed sentiment analysis in the military field.

The national defense and military field are a rigorous discipline. However, in response to the recent and rapid spread of information during the Internet era, relevant units have paid serious attention to exchanges between the military and civilians and the development of public opinion analysis by not only publishing micro-films and related advocacy messages through social media, but also attempting to detect attitudes and trends regarding specific events or issues by analyzing netizens' messages, in order to master such information and provide instant responses. As an electronic bulletin board system built on Taiwan's academic network resources, Ptt.cc (PTT) is the most used and influential online forum space in Taiwan. PTT has a large number of users and articles; as a result, posts and feedback on this platform often become the focus of attention and materials among journalists. Therefore, relevant government departments attach great importance to the opinions expressed by the public on PTT.

Using posts on the Militarylife PTT board as research subjects, this study attempted to develop a community sentiment analysis process based on deep learning technology and different experimental designs using sentiment dictionaries and model parameter setting (including activation function and network layer selection) as the empirical basis for building a combination model for multiple types of sentiment analysis and exploring better learning mechanisms through effective evaluation indicators.

2 Literature review

2.1 Sentiment analysis and classification

Sentiment analysis, which is also known as opinion mining, primarily uses natural language processing and information extraction techniques to conduct text mining and analysis. The tendency of a particular text is judged based on the context and polarity obtained and may be a potential argument, opinion, or sentimental state of the text (Day and Teng 2017). The more textual data collected, the easier it is to find a significant correlation between text and sentiment type. This correlation can simultaneously be used to predict the sentiment orientation of different types of text.

Sentiment analysis can be divided into two major categories, namely sentiment classification and feature-based opinion mining. Sentiment analysis can be further classified into two approaches, namely the corpus-based approach and the dictionary-based approach (Shelke et al. 2012). Based on the corpus-based approach, sentiment words can be placed into a corpus for learning purposes in order to obtain sentiment scores for words, while the dictionary-based approach involves the extraction of sentiment words from text, as well as the use of lexical databases (e.g., WordNet) to obtain the sentiment scores of words. On the other hand, feature-based opinion mining consists of using feature engineering and learning algorithms to transform text units into feature vectors in order to perform sentiment analysis. Three other approaches, namely the lexicon-based approach, the machine learning approach, and the hybrid approach, are used by other scholars as important classification methods for sentiment analysis (Medhat et al. 2014; Jain and Singh 2019), where machine learning is the most popular approach among them (Piryani et al. 2017).

By compiling recent studies related to sentiment analysis, areas of application include product marketing, health care, finance, elections, politics, sports, film reviews, and tourism and hospitality (Shayaa et al. 2018; Mäntylä et al. 2018). Various studies aim to extract mainstream opinions and voices regarding specific issues or explore the tread of thoughts of specific ethnic groups based on a variety of online textual data. For example, Zavattaro et al. (2015) analyzed the sentiments of Internet users from tweets on Twitter to judge their degree of participation in politics, while Day and Lin (2017) collected reviews from the Google Play mobile app to learn about consumer opinions and comments. On the other hand, Chen and Chen (2019) used finance blogs and news contents to predict the

direction of financial stock markets. The aforementioned are all examples of applications of sentiment analysis in different fields.

2.2 Deep learning

Deep learning is a branch of machine learning, as well as the mainstream trend in the development of machine learning (Zhang et al. 2018). In deep learning, both linear and nonlinear transformations can be carried out through multilayer neural networks to extract the features of data so that the computer can observe, learn, and react to complex scenarios (Deng and Yu 2014; Day and Teng 2017). The most representative example in this case is AlphaGo. Common deep learning methods include convolutional neural networks (CNNs) for computer vision and image recognition (Krizhevsky et al. 2012) and recurrent neural networks (RNNs) for machine translation services based on natural language processing and statistical techniques (Cho et al. 2014).

Although deep learning has been proven to produce good results in many applications, various problems still require improvement, such as exploding and vanishing gradient, difficulties in model interpretation, related parameter settings, increasingly complex model training as the number of network layers increases, and how to maintain a certain accuracy rate to improve training speed. These problems need to be studied and overcome in the field of deep learning.

Long Short Term Memory (LSTM) is an extension of RNN architecture and was proposed by Hochreiter and Schmidhuber (1997). LSTM not only covers the basic structure of RNN, but also comprises three components, namely input gate, output gate, and forget gate (Zhang et al. 2018). These control gates are turned on or off according to the received signal and have their own weights. Furthermore, these gates filter input signals and decide whether to allow these signals to pass based on their strength and imported content. The forget gate of LSTM can select remembered data and forgotten data to overcome the inability of RNN to learn due to vanishing and exploding gradients (Day and Teng 2017). LSTM has been proven to be particularly useful in learning a variety of sequence modeling tasks that involve unknown lengths because it can keep long-term memory (Zhang et al. 2016).

However, LSTM only considers context messages in a single direction. Therefore, in 2005, Graves applied the Bidirectional Long Short Term Memory (Bi-LSTM) architecture to extract more refined features and ultimately improve the performance of the traditional LSTM model. This architecture uses two LSTMs in opposite directions, namely a LSTM forward layer and a LSTM backward layer, which run from the beginning and end of the sequence in order to obtain the forward and reverse context messages. Each output is the final outcome arising from the addition of two LSTMs by combining the outputs of the forward layer and backward layer.

LSTM-based applied research covers issues in different directions, including speech classification (Lehner et al. 2015), scene recognition (Chen et al. 2017), analysis of stock market price fluctuations (Di Persio and Honchar 2016), time series forecasting (Karim et al. 2017), healthcare monitoring (Verma and Kumar 2019), and human behavior and motion recognition (Fok et al. 2018).

2.3 Activation function

Activation function is an important parameter in the deep learning model and is a rule that converts the input values and weights of each neuron into a nonlinear one. Without a nonlinear activation function, the effect of deep learning does not differ from that of a general neural network. The commonly used nonlinear activation functions are Sigmoid, Tanh, ReLU, etc. (Day and Lin 2017).

Sigmoid is a common activation function in RNN and can convert input values from 0 to 1. It is monotonically continuous, has a limited output range, and can be optimized for stability. However, it often has a vanishing gradient, thus leading to training issues. Its calculation method is expressed in Formula (1). Tanh is a variant of the Sigmoid function, which converts input values from -1 to 1. It converges faster than the Sigmoid function but may still cause the vanishing gradient. Its calculation method is expressed in Formula (2). The ReLU function can convert all values into values greater than 0. If the input value is less than 0, it will output 0; otherwise, it will directly output the input value. Its calculation is also simple, as shown in Formula (3). Therefore, ReLU can effectively improve training speed and reduce the vanishing gradient issue. In recent years, ReLU has been the most commonly used activation function in deep learning research (Zhang et al. 2018).

$$\text{sigmoid}(x) = \frac{1}{1 + \mathrm{e}^{-x}} \tag{1}$$

$$\tan h(x) = \frac{\mathrm{e}^{x} - \mathrm{e}^{-x}}{\mathrm{e}^{x} + \mathrm{e}^{-x}} \tag{2}$$

$$ReLU(x) = \max(0, \mathrm{x}), \quad \text{that is}$$
$$\begin{cases} \sigma(x) = 0, & x \le 0 \\ \sigma(x) = x, & x > 0 \end{cases} \tag{3}$$

However, different deep learning models may produce different results through the use of the activation function. Therefore, exploring different types of neuron activation functions and network architectures is a common theme of many studies (Deng et al. 2013).

2.4 LSTM-based sentiment analysis research

Research on sentiment analysis includes many characteristics of multi-faceted technology and cross-disciplinary areas. Scholars in the big data and artificial intelligence fields have recently paid special attention to the development of sentiment analysis-related studies, especially those that target social network (or media) (Hangya and Farkas 2017; Piryani et al. 2017; Chen and Chen 2019). Zhang et al. (2018) investigated deep learning-based sentiment analysis studies from a comprehensive perspective and provided a complete literature survey from the introduction of fundamental concepts, the structure of learning models, the techniques of sentiment classification, and the applications in different fields. LSTM is one of the most popular methods for deep learning applied to research on sentiment analysis (Cliche 2017). Table 1 shows a list of studies on sentiment analysis based on LSTM-related models as compiled in this study.

Tang et al. (2015) extended LSTM to build a Target-Dependent Long Short Term Memory (TD-LSTM) model, which uses Twitter text to automatically consider target information. The results of this study showed that TD-LSTM can significantly improve prediction accuracy. Al-Mansouri (2016) combined clustering techniques in text mining and LSTM for deep learning to predict stock prices, which revealed that the prediction accuracy of a particular cluster with an effective time is 77%. Xu et al. (2016) classified a number of lengthy sentiment texts and found that it easily causes memory shortage; therefore, they proposed a Cached Long Short Term Memory (CLSTM) model that could further refine extracted sentiment features. Vo et al. (2017) constructed a model combining CNN and LSTM to investigate product evaluations using a Vietnamese corpus. This experiment confirmed that the accuracy of the combined model is better than that of such individual models as the support vector machine (SVM), LSTM, and CNN. Day and Lin (2017) targeted consumer reviews of smartphones and conducted an assessment using LSTM for deep learning in combination with an opinion dictionary to find out consumers' opinion tendency, while making comparisons with methods for machine learning, such as naive Bayes and SVM, respectively. Based on its experimental results, LSTM can effectively improve the accuracy of sentiment analysis, compared with the two other types of machine learning methods.

Balikas et al. (2017) proposed a multiplex learning framework with different levels of sentiment polarity, while using word vector conversion and a Bi-LSTM model to complete the determination of sentiment of the netizens' messages. Shen et al. (2017) constructed a CNN-BLSTM model using a word vector as the text input feature to analyze emotion recognition, which achieved better prediction effects in the sentiment analysis of film reviews. Yoon and Kim (2017) combined the features of CNN and Bi-LSTM to extract the high-dimensional and long-term dependent text features and conduct sentiment classification of Twitter text using the enriched dictionary feature of

Table 1 Studies on sentiment analysis based on LSTM-related models

Author	Sentiment classification[a]	Topic/object	Model[b]
Tang et al. (2015)	Binary	Text on Twitter	TD-LSTM
Al-Mansouri (2016)	Binary	Stock price forecasting on Google and Yahoo	LSTM
Xu et al. (2016)	1–10 rating (IMDB)1–5 star rating (Yelp)	Film reviews on IMDB and restaurant reviews on Yelp	CLSTM
Vo et al. (2017)	Ternary	Vietnamese product review using VS and Vietnamese corpus using VLSP	LSTM + CNN
Day and Lin (2017)	Binary	Consumer reviews via mobile app, Google Play	LSTM
Balikas et al.(2017)	Ternary and quinary	Text on Twitter	Bi-LSTM
Shen et al. (2017)	Binary	Film reviews	CNN + Bi-LSTM
Yoon and Kim (2017)	Ternary	Text on Twitter	CNN +Bi-LSTM
Xu et al. (2019)	Binary	Hotel reviews on travel service Web site, Ctrip	Bi-LSTM
Zhou et al. (2019)	Binary	Weibo comments	Stacked Bi-LSTM

[a]Binary (positive and negative); ternary (positive, neutral and negative); quinary (very positive, positive, neutral, negative and very negative)

[b]*TD-LSTM* Target-Dependent Long Short Term Memory, *CLSTM* Cached Long Short Term Memory

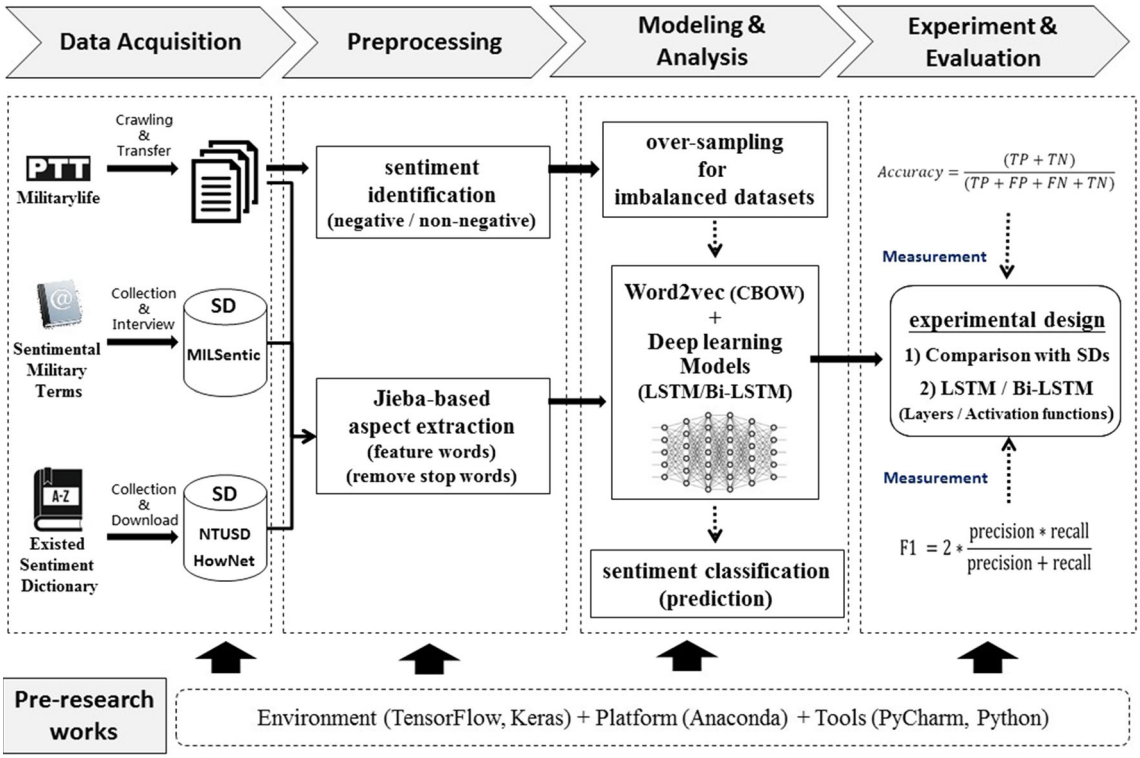

Fig. 1 Research framework and processes

the multi-channel method. Xu et al. (2019) collected contents related to hotel reviews on the travel service network Ctrip and judged sentiment classifications using the Bi-LSTM method with TD-IDF as the lexicon weighing scheme, showing good results in the process. Zhou et al. (2019) proposed a stacked Bi-LSTM learning model and integrated the lexicon-based vectorization scheme for continuous bag-of-words (CBOW) to analyze the sentiment polarity of user comments on the Chinese Web site Weibo.

From the aforementioned literature, various studies related to sentiment analysis where the LSTM model is applied have achieved diverse results.

3 Research methodology and process design

Based on the implementation of the sentiment analysis or opinion mining recommended by Guellil and Boukhalfa (2015) and Hemmatian and Sohrabi (2017), this study proposes a sentiment analysis framework for a social network based on deep learning models, including pre-research works, data acquisition, preprocessing, modeling and analysis, experimenting, and evaluation.

The subject of this study is the Militarylife board on PTT, the Web site for the largest online communities in Taiwan. In this study, sentiment training for text mining and deep learning is carried out using a systematic method with the support of a self-developed military sentiment dictionary, while an effective analysis model is built by calibrating different parameters. Figure 1 shows the research framework and process of this study, and the relevant steps are described below.

I. Pre-research works:

This step includes the installation and setup of a platform, tools, and software (Anaconda, PyCharm, and Python), as well as the deep learning environment (TensorFlow and Keras modules).

II. Data acquisition

This stage includes two main tasks: collection of community posts and development of military sentiment dictionary.

A web crawler was written with Python and released into the MilitaryLife PTT board to extract the content of posts and messages. The total number of registered users on PTT is 1.5 million people, while more than 150,000 users go online on PTT during peak periods. PTT has more than 20,000 different bulletin boards with different themes. More than 20,000 new articles and 500,000 posts are uploaded every day. In other words, PTT is the most used online forum in Taiwan. For this study, the data collection

Table 2 The positive and negative lexical data and sources in our study

Sentiment dictionary	#Positive	#Negative	Ratio (*P*:*N*)	#Total	Source
NTUSD	2810	8276	1:3	11,085	[1]
HowNet	4566	4370	1:1	8936	[2]
MILSentic	53	73	3:4	126	[3]

[1] http://nlg.csie.ntu.edu.tw/download.php

[2] http://www.keenage.com/html/c_bulletin_2007.htm

[3] Self-developed military sentiment dictionary (by terms collection and interview)

period lasted from January 2015 to February 2019, extracting a total of 17,819 articles.

In addition to using two Chinese sentiment dictionaries, namely the National Taiwan University Sentiment Dictionary (NTUSD) and HowNet, as the basis, this study also develops one set of a preliminary military sentiment dictionary, MILSentic, by compilating special military sentiment words and interviewing military professionals familiar with the community's language. The content of the dictionary covers commonly seen positive and negative sentiment or evaluative words, including 53 positive words (e.g., lean and united) and 73 negative words (e.g., bruise and heavenly soldiers). In the subsequent model analysis, these sentiment dictionaries were used to assist learning and prediction. Table 2 shows the positive and negative lexical data in each dictionary and the reference sources.

III. Preprocessing

The preprocessing stage can be divided into two tasks: Jieba-based aspect extraction and sentiment identification. The former focuses on sentiment dictionary-supported Jieba word segmentation processing, while the latter emphasizes the discrimination of the sentiment polarity of articles.

The Jieba Chinese word segmentation system offers three modes, namely full mode, precise mode, and search engine mode. Due to the large number of articles and messages on PTT and the uneven length of the various content, this study uses the precise mode, which is suitable for text analysis when conducting Chinese word segmentation and can generate representative and more accurate sentiment lexicon in each article after removing stop words using Python, with the goal of facilitating subsequent analyses.

After analyzing different social media channels, Kumar et al. (2018) found that positive and negative news can attract users' attention, and the popularity of such news is much greater than that of neutral news. Posts on the Militarylife PTT board mostly express and freely discuss current affairs regarding the national army or related personal experience. In terms of content, negative critical articles are more significant than positive affirmative articles; therefore, this study combined positive news and neutral news as nonnegative news. Furthermore, this study used binary (nonnegative and negative) relationship as the basis for conducting feature learning and sentiment prediction.

Nonnegative or negative articles in the originally collected data were classified by human means. In addition to the authors of this paper, this study also invited two assistants to aid in the judgment of sentiment polarity. In the case of inconsistent judgment of sentiment polarity, a third person was invited to conduct classification, and the sentiment polarity of an article was determined by vote. A total of 11,631 negative articles and 6188 nonnegative articles were obtained after conducting this procedure.

IV. Modeling and analysis

The processing of training data and construction of the model are the two important works in this stage.

Ertekin (2013) mentioned that in a training data set, the amount of one type of data will affect the accuracy of the prediction results. To resolve this issue, scholars have proposed two methods, namely undersampling and oversampling (Amin et al. 2016). Undersampling involves reducing the amount of categorical data with high data volume and balancing it with the amount of categorical data with low data volume, while oversampling involves increasing the amount of categorical data with low data volume and balancing it with the amount of categorical data with high data volume. Based on Ertekin's (2013) study, oversampling can improve classification performance more significantly than undersampling for complex data types. Since the number of negative articles collected in this study is much larger than that of nonnegative articles, this study adopted the oversampling method by increasing the number of nonnegative articles to achieve data balance and ensure accuracy. After increasing nonnegative articles, 80% of the data was classified as training data, while the remaining 20% was classified as validation data in order to carry out the operation of subsequent learning models.

In this stage, the Genism suite in Python was employed to convert words into word vectors using Word2vec and calculate the relevance and word frequency between words. Word2vec is the most popular and efficient algorithm for

extracting the low-dimensional vector representation of words (Mikolov et al. 2013; Araque et al. 2017). This study inputs training data and uses the CBOW algorithm to establish a word vector model in order to facilitate pre-judgment and analysis through deep learning models.

V. Experiment and evaluation:

This research experiment aimed to verify the effectiveness of the self-developed military sentiment dictionary, MILSentic, and compare the impact of existing sentiment dictionaries (NTUSD + HowNet) and of adding the self-developed military sentiment dictionary (NTUSD + HowNet + MILSentic) on the performance of the learning model. Furthermore, deep learning was based on two models, namely LSTM and Bi-LSTM, while training and learning were conducted using such methods as layer setting and activation function (Sigmoid, Tanh, and ReLU). A combination of parameter settings with better performance was selected using prediction accuracy. Therefore, this study was divided into three experiments.

- *Experiment 1* Verification of the self-developed military sentiment dictionary, MILSentic.
- *Experiment 2* Performance analysis using LSTM as the learning model in combination with the layer setting and activation function.
- *Experiment 3* Performance analysis using Bi-LSTM as the learning model in combination with the layer setting and activation function.

This study validated the performance of the training model by using two indicators, namely accuracy and F1-measure (Hemmatian and Sohrabi 2017; Wang et al. 2019), where the F1-measure was the weighted average of precision and recall. Table 3 shows the relationship between these parameters and all the samples.

- *Accuracy (A)* This refers to the ratio of all correct predictions (actually positive and predicted to be positive + actually negative and predicted to be negative) to all samples, which is expressed using the following formula: A = (TP + TN)/(TP + FP + FN + TN).
- *Precision (P)* This refers to the ratio of correct predictions that are positive (actually positive and predicted to be positive) to all predictions that are positive (actually positive and predicted to be positive + actually negative but predicted to be positive), which is expressed using the following formula: P = TP/(TP + FP).
- *Recall (R)* This refers to the ratio of correct predictions that are positive (actually positive and predicted to be positive) to all actually positive samples (actually positive and predicted to be positive + actually positive but predicted to be negative), which is expressed using the following formula: R = TP/(TP + FN).
- *F-measure* This refers to the weighted average of precision (P) and recall (R), which is expressed using the formula: F1-measure = 2PR/(P + R).

Table 3 The relationship between evaluation parameters

Predicted	Actually	
	Positive	Negative
Positive	True positive (TP)	False positive (FP)
Negative	False negative (FN)	True negative (TN)

4 Experimental results and analysis

4.1 Parameter settings for the experimental model

The experimental model in this study obtained the optimal values for each parameter through trial and error, where the vector dimension of the embedding layer was set to 100, and the hidden layer of LSTM/Bi-LSTM used 64 neurons, while the Adam optimizer used had a default learning rate of 0.001. When the initial value was set to 0.01 after testing, the model had a better recognition rate. The dropout layer was set to 0.5 and placed between the hidden layer and the output layer of LSTM. To prevent model overfitting, a fully connected layer was then added, which used 25 neurons. The last part of the output layer used two neurons and output the results using the softmax function. Furthermore, with regard to the experiment increasing the number of network layers, not only was a hidden layer added to LSTM/Bi-LSTM in each stage, but a dropout layer was also added simultaneously.

4.2 Experimental results and analysis

Regarding the content of word segmentation in negative and nonnegative articles, the Word2vec algorithm was used to learn the vector representation of words. The parameter threshold was set to 10, meaning that after filtering article lexicon, words with a word frequency of less than 10 (min_count < 10) were ignored. Lexicon that was set through the threshold and met the specific sentiment of the article was converted into a word vector model, thus performing sentiment analysis of the text using two deep learning models (LSTM and Bi-LSTM). The experimental results are described as follows.

4.2.1 Performance validation for MILSentic

The combinations of sentiment dictionaries used in Experiment 1 were divided into two types: (1) existing sentiment dictionaries (NTUSD + HowNet) and (2) the combination of existing sentiment dictionaries and the self-developed military sentiment dictionary (NTUSD + HowNet + MILSentic). Regarding the model for learning sentiment polarity, a basic prediction model was established using LSTM combined with the Sigmoid activation function. Figure 2 shows the experimental results.

According to the data in Fig. 2, the accuracy and F1-measure of the model for polarity prediction using sentiment dictionaries (NTUSD + HowNet) was 82.60% and 81%, respectively, while the accuracy and F1-measure of the model for polarity prediction using the combination of existing sentiment dictionaries and the self-developed military sentiment dictionary (NTUSD + HowNet + MILSentic) was 84.10% and 82.40%, respectively, 1.5% and 1.4% greater than the former. Both indicators exhibited an increasing trend, indicating that the addition of the MILSentic sentiment dictionary can improve the accuracy of polarity classification prediction.

4.2.2 Performance validation for LSTM/Bi-LSTM parameter calibration

In Experiments 2 and 3, LSTM and Bi-LSTM were used as learning models in combination with the setting of network layers and three types of activation functions (i.e., Sigmoid, Tanh, and ReLU) to conduct performance analysis and comparison. Table 4 shows the combined results of both experiments.

Overall, accuracy and F1-measure exhibited an increasing trend when the number of network layers was increased. However, as the number of network layers increased to a certain level, the model performance started decreasing, even in its accuracy. Therefore, the higher the number of layers does not necessarily mean the better the prediction effect. Based on the data in Table 4, except that Sigmoid and ReLU in LSTM mode exhibited a better effect in the first layer, the remaining model combinations achieved better results in the second layer. The data also demonstrated that when the Bi-LSTM learning model was used in combination with two network layers, the activation effect of the Tanh function and the accuracy of the prediction model trained were optimal, with 92.68% accuracy and the F1-measure of 88.41%.

Regarding the performance of the learning model, the Bi-LSTM-based training model was found to be better than the traditional LSTM model in terms of accuracy and F1-measure. In the process of calibrating activation functions, the Tanh function exhibited the best performance, differing from the general perception of deep learning (most learning models employ the ReLU function, which can solve the vanishing gradient problem). This finding may be due to the use of the LSTM learning model because the LSTM architecture can solve the problem of the vanishing gradient. The second possible reason is that when the number of LSTM network layers is low, the effect of using Tanh as the activation function is better, which has been proven by the experimental results and is consistent with other research results (Tsai et al. 2019).

In order to better understand the optimal calibration performance of different models and parameter combinations, the optimal prediction results were extracted and are shown in Fig. 3, where the value in the bracket is the optimal number of network layers. For example, this link, LSTM − Acc(1) = 0.841, F1(2) = 0.829 − Sigmoid, represents the activation function of Sigmoid as the LSTM learning model. When the number of network layers was 1, the predicted accuracy of sentiment polarity was 0.841; when the number of network layers was 2, the predicted F1-measure of sentiment polarity was 0.829, which was the best performing parameter matching situation in the experiment. The remaining connection paths were similarly interpreted.

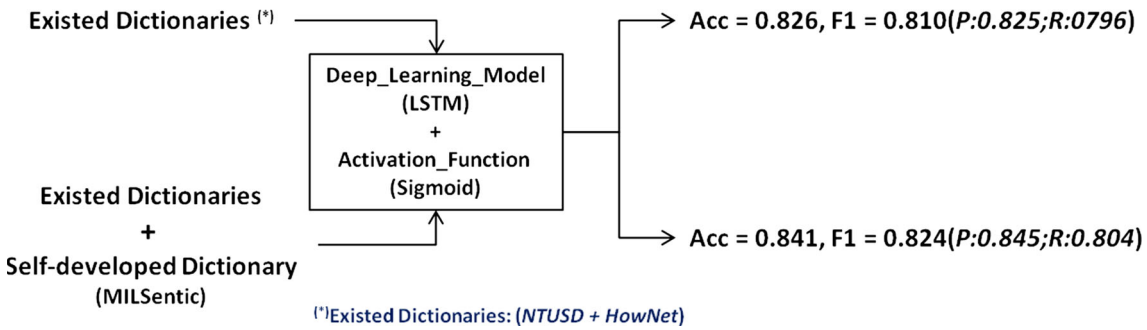

Fig. 2 Comparison of learning performance used for different sentiment dictionaries

Table 4 Comparison of different models and performance of parameter calibration

Model	Layer	Accuracy			F1-measure		
		Sigmoid (%)	Tanh (%)	ReLU (%)	Sigmoid (%)	Tanh (%)	ReLU (%)
LSTM	1	**84.08**	89.42	**83.27**	82.41	85.81	**81.39**
	2	83.75	**90.11**	82.75	**82.92**	**86.65**	80.31
	3	–	88.85	–	–	86.12	–
Bi-LSTM	1	83.70	90.46	84.04	82.99	87.52	83.28
	2	**85.40**	**92.68**	**84.94**	**84.31**	**88.41**	**84.43**
	3	84.42	91.81	84.27	84.02	88.22	83.90

Bold values represent the best performance in each model and parameter portfolio

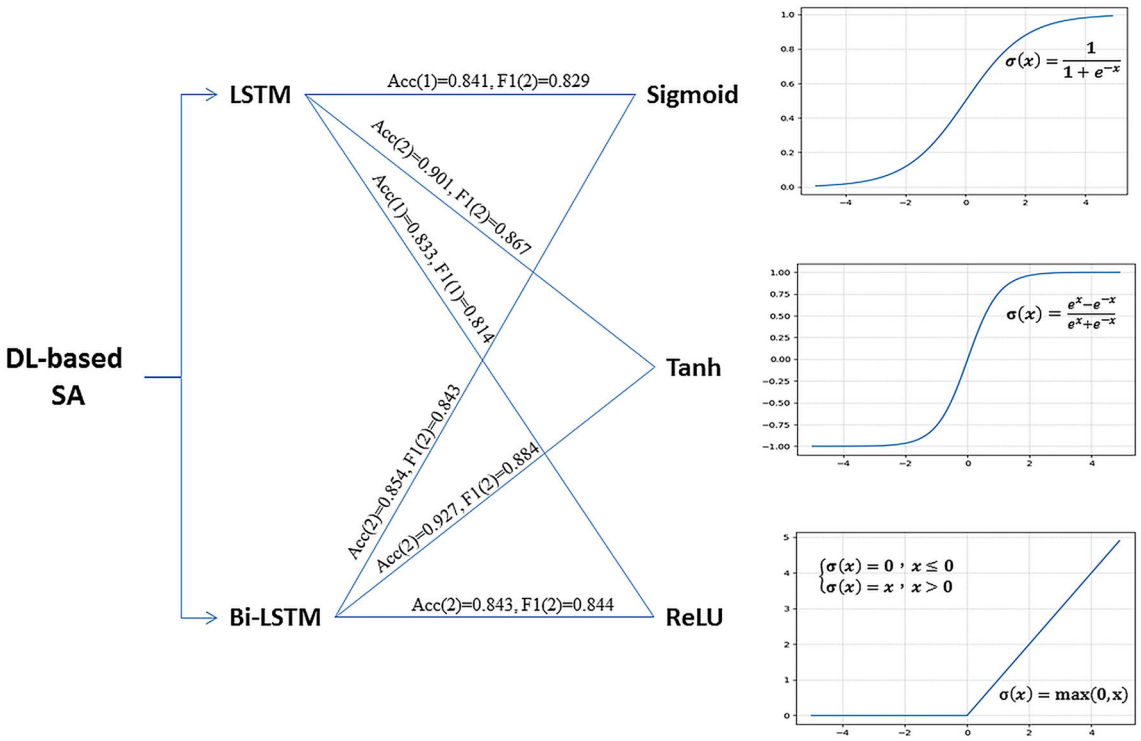

Fig. 3 Optimal performance under different learning models and the calibration of activation functions

5 Conclusions

With the widespread use of social networks and social media, users find it relatively easy to use the relevant platforms to post messages or comments, resulting in rapid accumulation and huge amounts of community data. Therefore, the sentiment analysis of text has become an important task in the Internet and social media. This study used the Militarylife PTT board of Taiwan's largest online forum as the source of experimental data for sentiment analysis in the military field. At the same time, after combining the Jieba system and sentiment dictionaries to conduct Chinese word segmentation, training was carried out on two types of learning model, namely LSTM and Bi-LSTM, using the Word2vec vector conversion mechanism. The experimental results show that the accuracy and F1-measure of the model that combined existing sentiment dictionaries and the self-developed military sentiment dictionary, MILSentic, was 84.08% and 82.41%, respectively, which were better than the results from using just the existing sentiment dictionaries. Furthermore, the prediction model trained using the activation function Tanh and when the number of Bi-LSTM network layers was two, the accuracy and F1-measure were 92.68% and 88.41%, respectively, proving that Bi-LSTM demonstrated better performance than LSTM, as well as that the associated Tanh activation function can better improve the effect of sentiment classification.

This study confirms that such indicators as accuracy and F1-measure can reach a certain level when learning and training have been conducted using different parameter calibrations through the deep learning-based sentiment

analysis of social networks combined with the self-developed sentiment dictionary, MILSentic. The results of this study can be provided to government or military-related agencies to screen sentiment polarity of community articles in this social media era to rapidly understand the public evaluation of major military issues or public perception of the image of the national army, thus rapidly responding and referring them to the appropriate institutions for adjustment or policy revision.

This study offers the following contributions: (1) The prediction results from conducting sentiment analysis of the military board under Taiwan's PTT online forum, which has generally received less attention, can effectively provide the government or military-related organization systems with observations of social media opinions. (2) The self-developed military sentiment dictionary, MILSentic, proposed in this study, which can be applied to sentiment analysis, can effectively identify the sentiment types of posts and comments in military communities while improving the accuracy of sentiment polarity judgment. (3) This study introduces two types of deep learning models (Bi-LSTM vs. LSTM) for comparison and analysis. The results confirm that the Bi-LSTM model's forward and backward combination exhibits better performance than the traditional LSTM model for all indicators. (4) This study proposes different calibration and validation models using different parameter combinations (activation functions and number of network layers). The experiment results confirm that while increasing the number of network layers can improve accuracy, over-increasing the number of network layers may result in a decreasing effect, which is not cost-effective. Furthermore, the results also verify that sentiment polarity classification exhibits the best effects under the activation function Tanh.

In this study, sentiment analysis focuses only on the Militarylife PTT board. In the future, in addition to cross-platform integration on other military-related social media, further research can be used to improve model performance in combination with other effective learning features and attempt to expand calibration combinations with different models and parameters.

Relevant scholars are recommended to carry out the follow-up studies provided below:

1. The challenges of sentiment lexicon include such issues as differential judgment of sentiment words in different areas, inadequateness of non-English sentiment dictionaries, and insufficiency of specific sentiment lexicon in specific areas. Apart from the continuous expansion of the sentiment lexicon corpus in the military field, how to extract sentiment words in the new era in response to rapidly changing online terms to improve the accuracy of model prediction is worthy of detailed study.
2. This study only conducts sentiment analysis of text. However, as social media has a great demand for images and videos, as well as possesses cross-language and cross-cultural characteristics, detailed research on sentiment polarity classification can be conducted from the perspective of multimedia fusion.

Acknowledgements This research was partially sponsored by the Ministry of Science and Technology (MOST), Taiwan under Grant No: 107-2410-H-606-006

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Ahmed W, Lugovic S (2019) Social media analytics: analysis and visualisation of news diffusion using NodeXL. Online Inf Rev 43(1):149–160

Al-Mansouri E (2016) Using artificial neural networks and sentiment analysis to predict upward movements in stock price. Doctoral dissertation, Worcester Polytechnic Institute

Amin A, Anwar S, Adnan A, Nawaz M, Howard N, Qadir J, Hawlah A, Hussain A (2016) Comparing oversampling techniques to handle the class imbalance problem: a customer churn prediction case study. IEEE Access 4:7940–7957

Araque O, Corcuera-Platas I, Sanchez-Rada JF, Iglesias CA (2017) Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Syst Appl 77:236–246

Asur S, Huberman BA (2010) Predicting the future with social media. In: Proceedings of the 2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology vol 1, pp 492–499

Balikas G, Moura S, Amini MR (2017) Multitask learning for fine-grained twitter sentiment analysis. In: Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval, pp 1005–1008

Chen MY, Chen TH (2019) Modeling public mood and emotion: blog and news sentiment and socio-economic phenomena. Future Gener Comput Syst 96:692–699

Chen PJ, Ding JJ, Hsu HW, Wang CY, Wang JC (2017) Improved convolutional neural network based scene classification using long short-term memory and label relations. In: 2017 IEEE international conference on multimedia & expo workshops (ICMEW), pp 429–434

Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y (2014) Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078

Cliche M (2017) BB_twtr at SemEval-2017 task 4: twitter sentiment analysis with CNNs and LSTMs. arXiv preprint arXiv:1704.06125

Day MY, Lin YD (2017) Deep learning for sentiment analysis on google play consumer review. In: 2017 IEEE international conference on information reuse and integration (IRI), pp 382–388

Day MY, Teng HC (2017) A study of deep learning to sentiment analysis on word of mouth of smart bracelet. In: Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017, pp 763–770
Deng L, Yu D (2014) Deep learning: methods and applications. Found Trends Signal Process 7(3–4):197–387
Deng L, Hinton G, Kingsbury B (2013) New types of deep neural network learning for speech recognition and related applications: an overview. In: 2013 IEEE international conference on acoustics, speech and signal processing, pp 8599–8603
Di Persio L, Honchar O (2016) Artificial neural networks approach to the forecast of stock market price movements. Int J Econ Manag Syst 1:158–162
Ertekin Ş (2013) Adaptive oversampling for imbalanced data classification. In Information sciences and systems 2013. Springer, Cham, pp 261–269
Fok WW, Chan LC, Chen C (2018) Artificial intelligence for sport actions and performance analysis using recurrent neural network (RNN) with long short-term memory (LSTM). In: Proceedings of the 2018 4th international conference on robotics and artificial intelligence, pp 40–44
Guellil I, Boukhalfa K (2015) Social big data mining: a survey focused on opinion mining and sentiments analysis. In: 2015 12th international symposium on programming and systems (ISPS), pp 1–10
Hangya V, Farkas R (2017) A comparative empirical study on social media sentiment analysis over various genres and languages. Artif Intell Rev 47(4):485–505
He W, Wu H, Yan G, Akula V, Shen J (2015) A novel social media competitive analytics framework with sentiment benchmarks. Inf Manag 52(7):801–812
Hemmatian F, Sohrabi MK (2017) A survey on classification techniques for opinion mining and sentiment analysis. Artif Intell Rev 1–51
Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
Jain SK, Singh P (2019) Systematic survey on sentiment analysis. In: 2018 first international conference on secure cyber computing and communication (ICSCCC), pp 561–565
Karim F, Majumdar S, Darabi H, Chen S (2017) LSTM fully convolutional networks for time series classification. IEEE Access 6:1662–1669
Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
Kumar N, Nagalla R, Marwah T, Singh M (2018) Sentiment dynamics in social media news channels. Online Soc Netw Media 8:42–54
Lehner B, Widmer G, Bock S (2015) A low-latency, real-time-capable singing voice detection method with LSTM recurrent neural networks. In: 2015 23rd European signal processing conference (EUSIPCO), pp 21–25
Mäntylä MV, Graziotin D, Kuutila M (2018) The evolution of sentiment analysis—a review of research topics, venues, and top cited papers. Comput Sci Rev 27:16–32
Medhat W, Hassan A, Korashy H (2014) Sentiment analysis algorithms and applications: a survey. Ain Shams Eng J 5(4):1093–1113
Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781
O'Reilly T, Battelle J (2004) Opening welcome: state of the internet industry. California, San Francisco
Pang B, Lee L (2008) Opinion mining and sentiment analysis. Found Trends Inf Retr 2(1–2):1–135
Piryani R, Madhavi D, Singh VK (2017) Analytical mapping of opinion mining and sentiment analysis research during 2000–2015. Inf Process Manag 53(1):122–150
Sang ETK, Bos J (2012) Predicting the 2011 Dutch senate election results with Twitter. In: Proceedings of the 13th conference of the european chapter of the association for computational linguistics, pp 53–60
Shayaa S, Jaafar NI, Bahri S, Sulaiman A, Wai PS, Chung YW, Piprani AZ, Al-Garadi MA (2018) Sentiment analysis of big data: methods, applications, and open challenges. IEEE Access 6:37807–37827
Shelke NM, Deshpande S, Thakre V (2012) Survey of techniques for opinion mining. Int J Comput Appl 57(13):0975–8887
Shen Q, Wang Z, Sun Y (2017) Sentiment analysis of movie reviews based on CNN-BLSTM. Int Conf Intell Sci. Springer, Cham, pp 164–171
Tang D, Qin B, Feng X, Liu T (2015) Effective LSTMs for target-dependent sentiment classification. arXiv preprint arXiv:1512.01100
Tsai HC, Chiu CJ, Tseng PH, Feng KT (2019) Refined autoencoder-based CSI hidden feature extraction for indoor spot localization. In: 2018 IEEE 88th vehicular technology conference (VTC-Fall), pp 1–5
Verma, H., and Kumar, S. (2019) An accurate missing data prediction method using LSTM based deep learning for health care. In: Proceedings of the 20th international conference on distributed computing and networking, pp 371–376
Vo QH, Nguyen HT, Le B, Nguyen ML (2017) Multi-channel LSTM-CNN model for Vietnamese sentiment analysis. In: 2017 9th international conference on knowledge and systems engineering (KSE), pp 24–29
Wang R, Zhou D, Jiang M, Si J, Yang Y (2019) A survey on opinion mining: from stance to product aspect. IEEE Access 7:41101–41124
Xiang Z, Du Q, Ma Y, Fan W (2018) Assessing reliability of social media data: lessons from mining TripAdvisor hotel reviews. Inf Technol Tourism 18(1–4):43–59
Xu J, Chen D, Qiu X, Huang X (2016) Cached long short-term memory neural networks for document-level sentiment classification. arXiv preprint arXiv:1610.04989
Xu G, Meng Y, Qiu X, Yu Z, Wu X (2019) Sentiment analysis of comment texts based on BiLSTM. IEEE Access 7:51522–51532
Yoon J, Kim H (2017) Multi-channel lexicon integrated CNN-BiLSTM models for sentiment analysis. In: Proceedings of the 29th conference on computational linguistics and speech processing (ROCLING 2017), pp 244–253
Zavattaro SM, French PE, Mohanty SD (2015) A sentiment analysis of US local government tweets: the connection between tone and citizen involvement. Gov Inf Q 32(3):333–341
Zhang X, Lu L, Lapata M (2016) Top-down tree long short-term memory networks. In: Proceedings of NAACL-HLT, pp 310–320
Zhang L, Wang S, Liu B (2018) Deep learning for sentiment analysis: a survey. Wiley Interdiscip Rev Data Min Knowl Discov 8(4):e1253
Zhou J, Lu Y, Dai HN, Wang H, Xiao H (2019) Sentiment analysis of Chinese microblog based on stacked bidirectional LSTM. IEEE Access 7:38856–38866

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.