DOI: 10.54254/2755-2721/6/20230415

Improving the BiLSTM model performance for tweet sentiment analysis

Jiaheng Dong

Computing and Software system, The University of Melbourne, Melbourne 3010, Australia

Jiahengd@student.unimelb.edu.au

Abstract. Twitter is a microblogging website where users can publish brief entries known as tweets. These tweets can occasionally reveal the users' attitudes and feelings. This paper analyses three emoticon processing methods with the BiLSTM model to discover the efficiencies of different methods in helping deep learning models classify tweet sentiments. Firstly, the simply removing method, the replacing with description text method, and the replacing with predefined sentiment method are established. Then the BiLSTM model is used to train and test with different methods on the Sentiment140 dataset. The performances of all models are evaluated by accuracy, F1 score, precision score, and recall score. The experimental results show that the replacing with predefined sentiments method provides the highest accuracy which is 0.84. The simply removing method also produces the testing accuracy as 0.84, but it performs worse in the last epoch, the training and validation accuracy, and the training and validation loss. The replacing with description text method produces the worst accuracy which is 0.83. It indicates that predefining the most possible sentiments of the popular emoticons has a reliable efficiency in optimizing the performance of deep learning models when the tweets with emoticons take a small proportion.

Keywords: emoji, emoticon, sentiment analysis, Twitter.

1. Introduction

Globally, people utilize social media platforms to openly express their feelings through text-based messages and photographs. Twitter is one of the most popular social platforms, which allows users to communicate, create networks, and share thoughts easily and promptly [1]. More specifically, users can send and read 140-character messages for each tweet. The audience on the Twitter platform grows continuously data which significantly improves the efficiency of collecting training and testing data for sentiment analysis. Sentiment analysis aims to identify the sentiments expressed in the text source [2]. It makes monitoring online chats more efficient for a business to comprehend the social sentiment surrounding its brand, product, or service. Recent developments in deep learning have greatly increased algorithms' capacity to analyze text in social media. Therefore, in some domains, such as assessing personal preferences and personalities, the creative use of data processing methods is crucial in capturing and predicting the feelings of tweets.

Based on previous research, sentiment analysis is mainly applied in the NLP fields, and the NLP tasks are trained to be language specific, while emojis and emoticons are a language in themselves [3]. The traditional sentiment analysis is mainly using machine learning models, such as Naïve Bayes [4], SVM [5], logistic regression [6], and deep learning models, such as RNN [7] and transformer [8]. All of them are sensitive to the language, and to some extent, the English text with emotions is a multilingual problem. Specifically, for the Z-generation, chatting with plain emojis or emoticons is popular and time-saving. Due to the unclear and changeable expressions of emojis and emoticons, the proper processing methods of them are still a challenge when the set of tweets which contains emoticons or emojis takes a small proportion of the whole training dataset.

This paper aims to analyze tweets sentiments in the Sentiment140 dataset with optimized BiLSTM to correctly classify and predict the tweets with emoticons. Firstly, the emoticons are processed by each emoticon processing method. In the simply removing method, all the emoticons are deleted, producing plain text to the model. The replacing with description text method takes the emoticons as unique words. The replacing with predefined sentiments method translates the emoticons to their popular sentiment. Secondly, the BiLSTM model is used to track in order to evaluate three emoticon preprocessing methods, which could enhance the effectiveness of deep learning models in analyzing and forecasting the sentiments from the English tweet. Thirdly, to facilitate the model's training and testing, the Sentiment140 dataset is used, which consists of 1.6 million labeled instances, which have been classified as "positive" or "negative" and evenly distributed.

We evaluate our method with accuracy, F1 score, recall score, and precision score. The experimental results show that the replacing emoticons with predefined sentiments method and the simply removing method provide the highest accuracy, which is 0.84. The replacing with the description text method provides an accuracy score of 0.83. However, the simply removing method performs worse in the last epoch of the training and validation accuracy, and training and validation loss. Therefore, predefining sentiments of popular emoticons optimizes the deep learning model the most in correctly classifying and predicting the tweets with emoticons.

2. Method

This section describes the establishment of the BiLSTM model and the processing details of the three emoticon cleaning methods. Firstly, the qualified words are chosen, which could be utilized in feature selections to reduce the noises and unmeaningful feature dimensions before implementing different emoticon processing methods (Sec. 2.1). Secondly, the emoticon processing methods are used to clear the data in advance, the three methods, which are the simply removing method, the replacing emoticons with description method, and the replacing emoticons with predefined sentiments method (which three) are illustrated (Sec. 2.2). Thirdly, the BiLSTM model is implemented with a specific hyper-parameter which might provide local maximum performance (Sec. 2.3).

2.1. Data pre-processing

The dataset used in this paper is Sentiment140 [9]. Five typical and most popular emoticons are used to stand for the majority of emoticons, which are the heart, the smiling face, the sad face, the neutral face, and the LOL face [10]. As for the training dataset, there are 31729 tweets that include at least one of these five emoticons.

We do not take into consideration of the emoticons and emojis since the emoticons in the Sentiment140 dataset mainly appear at the end of the tweet as an ending phrase. These five emoticons in each tweet are removed. The tweets which included strongly emotional emoticons and neutral textual content took a small proportion in current years. The plain texts of most of the tweets target specific emotions, and emoticons and emojis are mainly used as attitude enhancement strategies. Moreover, emoticons have their own language. Among the Z-generations, using plain emoji text is common and understandable. In the language-specific text, the emojis and emoticons produce some certain bias in predicting and training for most of the deep learning models, such as BiLSTM. Specifically, emoticons and emojis in different countries have different meanings, and meanings are also different in each generation. Such as the smiling face, which is one of the most popular emojis. In China, the new generation is mainly using it to present an awkward and speechless emotion, which is the opposite meaning compared to the old generation.

DOI: 10.54254/2755-2721/6/20230415

*2.1.1. Replacing emoticons and emojis with description. A*ll the emoticons are shown in Table 1.The r'<3' is replaced with “heart”, the r"[8:=;]['`\-]?[)d]+" is replaced with “smileface”, the r"[8:=;]['`\-]?p+" is replaced with “lolface”, the r"[8:=;]['`\-]?\(+" is replaced with “sadface” and the r"[8:=;]['`\-]?[\/|l*]" is replaced with “neutralface”

Table 1. Examples of translating emoticons into descriptions.

Emoticons in python expression	Examples in the real world	Description
r’<3’	‘<3’	heart
r”[8:=;][‘`\-]?[]d]+”	‘:)’, ‘:d’	smileface
r”[8:=;][‘`\-]?p+”	‘:P’	lolface
r”[8:=;][‘`\-]?\(+”	‘:(’	sadface
r"[8:=;]['`\-]?[\/|l*]"	‘:|’	neutralface

To some extent, simply removing the emoticons and emojis will lose confidence and accuracy in modifying and predicting sentiments [11]. Another way is to treat the emoticons and emojis as regular tokens [12] just like the textural name of each emoticon above. This method relies on the number of instances in each tweet, which will provide more training chances for the deep learning models [13].

2.1.2. Replacing with predefined sentiments. The “heart” is replaced with “happiness”, the “smileface” is replaced with “happy, awkward”, the “lolface” is replaced with “extremely happy”, the “sadface” is replaced with “upset, exciting” and the “neutralface” is replaced with "okay, speechless". Each emoticon or emoji has its own major sentiment in each period, instead of replacing the emoji with long descriptions, which is shown in figure 1. It is better to show the attitude directly within two words [14], because some of the descriptions of different emojis may overlap. Such as "face with tears of joy 😂” has the same word “tears” as “face with tears of grievance 😣", the "tears" in the common plain text might lead to negative prediction, so it is better to replace “😂” with “extremely happy”. The current popular sentiment of each emoticon is focused. By updating the meaning of each emoticon or emoji, the misunderstanding rate of each emoticon will be reduced. Moreover, for some specific aspects of products which need to use sentiment analysis, the sentiments of different emoticons or emojis can be set differently, based on the customer habits. Therefore, it can be used to handle common tweets, and by using different meaning packages, it also can increase the accuracy in predicting some specific aspects of tweets.

Emoji	Description
😂	Face with tears of joy
😘	Face blowing a kiss
😄	Grinning face with smiling eyes
😌	Relieved face
😝	Squinting face with tongue
😥	Sad but relieved face
😠	Angry face
😭	Loudly crying face
😓	Downcast face with sweat
😰	Anxious face with sweat

Figure 1. Detailed description of some emojis [15].

2.2. The BiLSTM model

Besides having the typical functionality of a sequence processing model, the bidirectional LSTM [16] consists of two LSTMs, which receive inputs in opposite directions. The architecture of BiLSTM is

DOI: 10.54254/2755-2721/6/20230415

shown in figure 2. BiLSTMs effectively expanding the network's information pool, enhancing the context supplied to the algorithm [17]. There are seven layers have been used to form the BiLSTM model, which are the embedding layer, two Bidirectional layers, one convolutional layer, two dense layers and one max pooling layer.

Firstly, in the embedding layer pre-setting process, we set the dimension to 100, which is large enough for this dataset. As for the word2vec model [18][19], the minimum count is set to 5, which means the words which appear less than 5 times will be cleaned. The rare words are always informative, which might create more bias. And the input length is set to 60, and the total number of vocabulary is set to 60000, which slightly exceeds the true vocabulary length of 51808. In the embedding layer, the input dimension is the vocabulary length, the output dimension is the embedding dimension, and the weights are evaluated by using the embedding matrix.

Secondly, the bidirectional layers using LSTM with 100 memory units and with a 0.2 drop ratio which are mainly used to implement the biLSTM learning process. Then the convolutional layer is used with the "relu" activation function, which transfers the linear relations to the non-linear correlation. The second last dense layer is built by setting the dimensionality of the output space as 16 unis and also use the "relu" as the activation function. The last dense layer is using the "sigmoid" as the activation function.

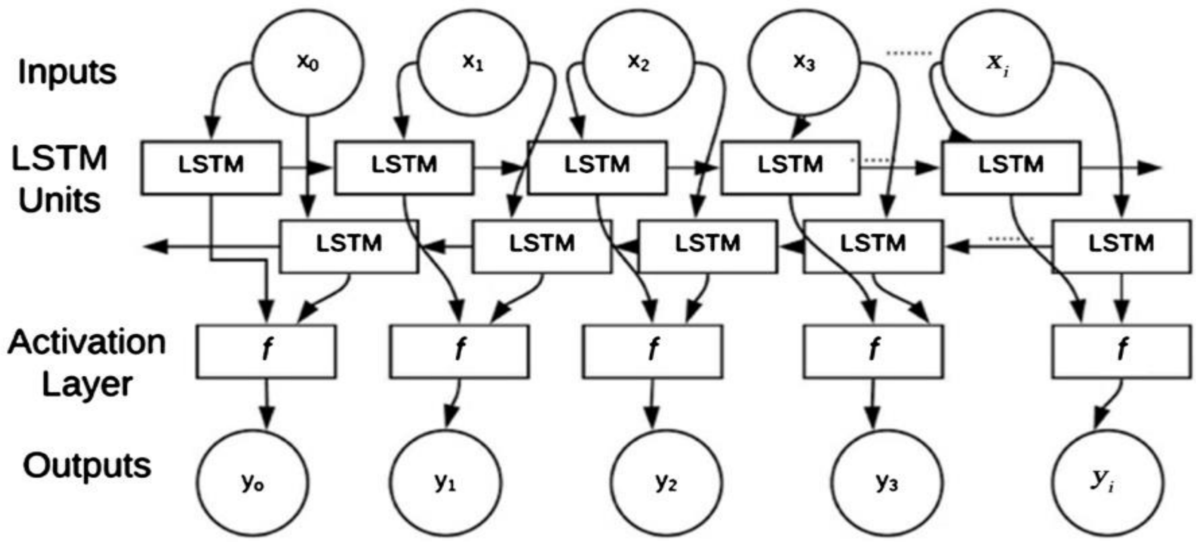

Figure 2. BiLSTM model with activation function [11].

3. Results

This section describes the training and testing results of the three emoticon processing methods by using the BiLSTM model. Firstly, the data distribution will be discussed, which shows a major influence on the baseline model result. The baseline result is mainly focusing on the majority voting method, it will be used as the accuracy lower threshold for validation accuracy of the BiLSTM model (Sec. 3.1). Secondly, the results of using different emoticon cleaning methods will be illustrated in three fields individually, which are the detailed training and testing processing stages, the trend of loss and accuracy in training and validation sets, and the evaluation matrices, which consist of the accuracy score, precision score, F1 score and recall score (Sec. 3.2-3.4).

3.1. Data distribution and baseline result

Figure 3 shows the distribution of the Sentiment140 dataset based on the sentiment of each tweet. The dataset is evenly distributed, as the result, there are 0.8 million negative tweets and 0.8 million positive tweets shown in figure 2.

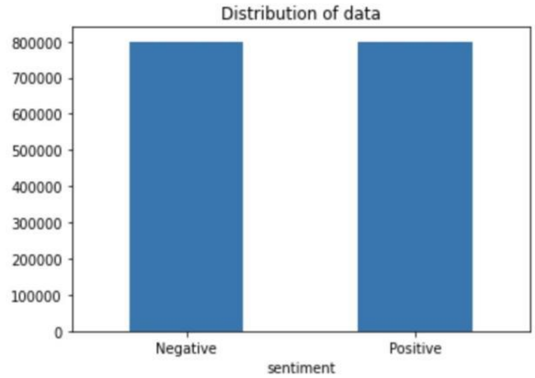

Figure 3. The distribution of two labels.

The accuracy of the baseline model is 0.5, because it is a balanced dataset, and the baseline model uses majority voting to calculate the accuracy.

3.2. *BiLSTM with simply removing*

The accuracy of BiLSTM with the simply removing method has an improvement from the baseline model. The model has been trained for 12 epochs, and the specific accuracy and cross-entropy loss are increasing as shown in figure 4 and 5. The final evaluation matrices which contain the precision score, recall score, F1 score, and accuracy score are shown in table 3.

Table 2. The specific training accuracy and cross-entropy loss of BiLSTM by removing all emoticons.

Epoch	Running time	Speed	Loss	Accuracy	Validation loss	Validation accuracy
1	59 s	66ms/step	0.4396	0.7937	0.4016	0.8160
2	47s	64ms/step	0.4023	0.8155	0.3858	0.8252
3	47s	64ms/step	0.3886	0.8229	0.3806	0.8277
4	47s	64ms/step	0.3793	0.8280	0.3767	0.8303
5	47s	64ms/step	0.3724	0.8317	0.3728	0.8333
6	48s	64ms/step	0.3665	0.8351	0.3696	0.8347
7	48s	64ms/step	0.3612	0.8380	0.3756	0.8333
8	47s	64ms/step	0.3572	0.8400	0.3667	0.8363

Table 2. (continued).

9	48s	64ms/step	0.3525	0.8424	0.3713	0.8358
10	48s	64ms/step	0.3486	0.8446	0.3692	0.8364
11	48s	64ms/step	0.3453	0.8464	0.3662	0.8372
12	47s	64ms/step	0.3419	0.8483	0.3676	0.8372

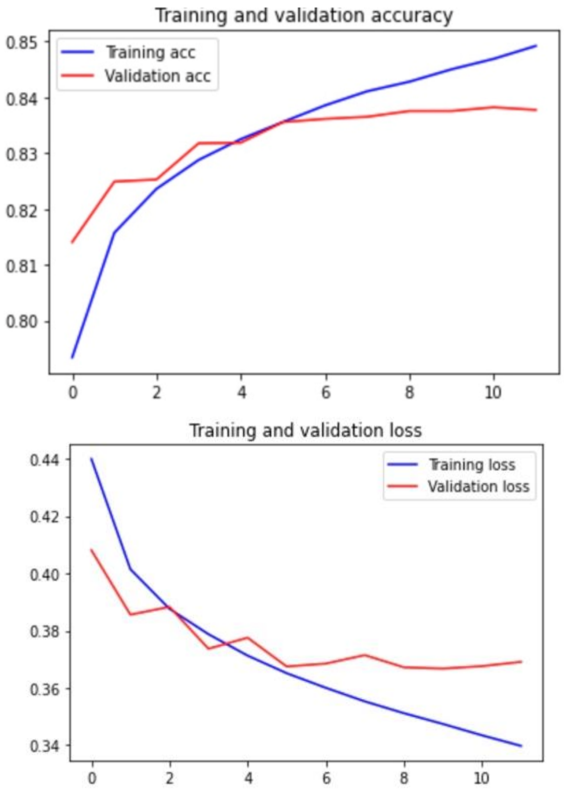

Figure 4. The trend of accuracy and loss of BiLSTM by removing all emoticons.

Table 3. Evaluation matrices of BiLSTM by removing all emoticons.

	precision	**Recall**	**F1-score**	**support**
0	0.85	0.82	0.83	39989
1	0.83	0.85	0.84	40011
Accuracy			0.84	80000
Macro avg	0.84	0.84	0.84	80000
Weighted avg	0.84	0.84	0.84	80000

DOI: 10.54254/2755-2721/6/20230415

The simply removing method provides an even rate for True Positive and Negative in figure 5. There is no emoticons to influence the weight of each sentiment, thus, the data distribution dominates the result, which is even.

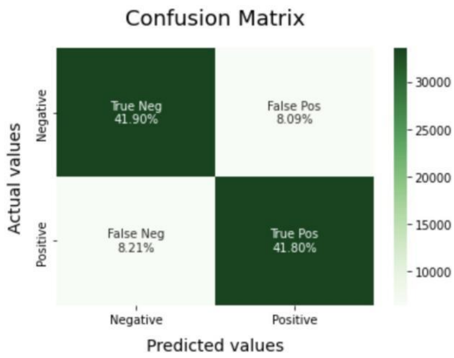

Figure 5. The confusion matrix of simply removing.

3.3. BiLSTM with replacing emoticons and emojis with descriptions

Table 4 and Figure 6 show the training and testing processes of replacing emoticons with descriptions by using the BiLSTM model. From table 4 and figure 6 we can see that this method helps the BiLSTM reduce some overfitting. The validation accuracy is greater than the training acracy, and the validation loss is lower than the training loss.

Table 4. The specific training accuracy and cross-entropy loss of BiLSTM by replacing emoticons and emojis with descriptions.

Epoch	Running time	Speed	Loss	Accuracy	Validation loss	Validation accuracy
1	57 s	66ms/step	0.4742	0.7709	0.4195	0.8069
2	47s	64ms/step	0.4350	0.7957	0.4023	0.8163
3	48s	64ms/step	0.4222	0.8037	0.4000	0.8182
4	47s	64ms/step	0.4142	0.8084	0.3885	0.8239
5	47s	64ms/step	0.4083	0.8118	0.3855	0.8268
6	48s	64ms/step	0.4046	0.8134	0.3826	0.8287
7	47s	64ms/step	0.4006	0.8160	0.3804	0.8304
8	48s	64ms/step	0.3977	0.8176	0.3808	0.8312

Table 4. (continued).

9	48s	64ms/step	0.3951	0.8189	0.3772	0.8326
10	47s	64ms/step	0.3929	0.8203	0.3748	0.8335
11	48s	64ms/step	0.3905	0.8213	0.3736	0.8340
12	47s	64ms/step	0.3892	0.8223	0.3756	0.8343

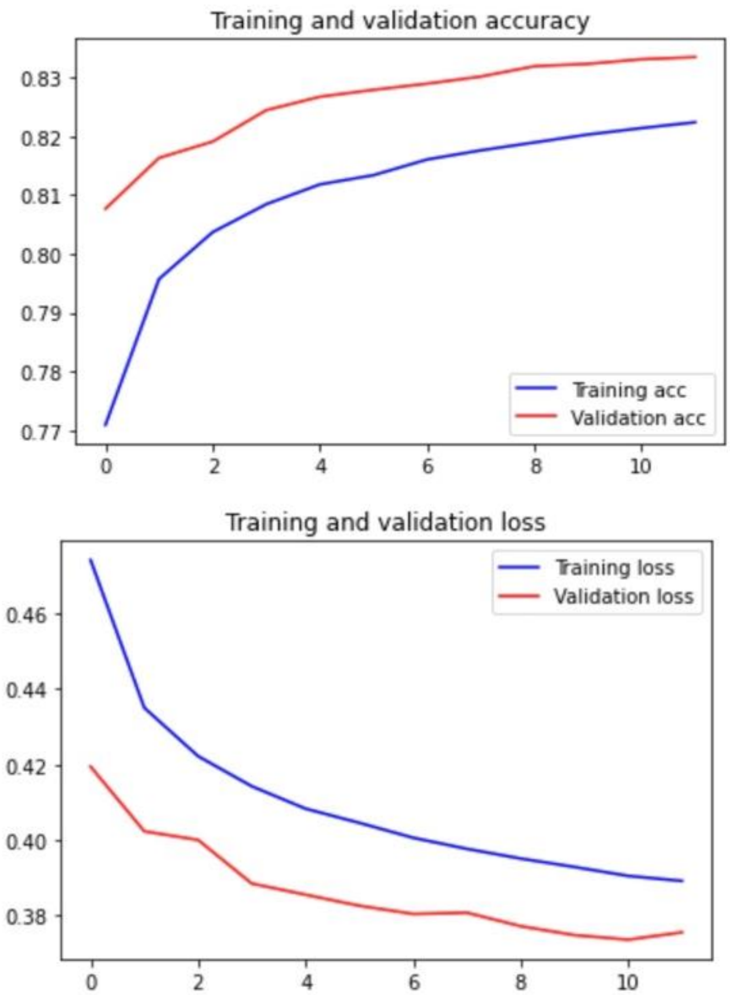

Figure 6. The trend of accuracy and loss.

The testing accuracy is lower than the simply removing method, which is shown in table 5. To some extent, the lower accuracy of replacing with description means that treating emoticons as tokens will provide the neural network with an unclear and misunderstanding emoticon meaning when the emoticon tweets are not the majority.

The replacing with the description method provides more accuracy in True Negative in figure 7. It indicates that the description does not provide the value of sentiment, it has the same ability as a unique token. Their meanings are easily influenced by the whole sentence, which might add more weight to the negative sentiment.

DOI: 10.54254/2755-2721/6/20230415

Table 5. Evaluation matrices of BiLSTM by replacing emoticons with descriptions.

	precision	Recall	F1-score	support
0	0.82	0.86	0.84	39989
1	0.83	0.85	0.83	40011
Accuracy			0.83	80000
Macro avg	0.83	0.83	0.83	80000
Weighted avg	0.83	0.83	0.83	80000

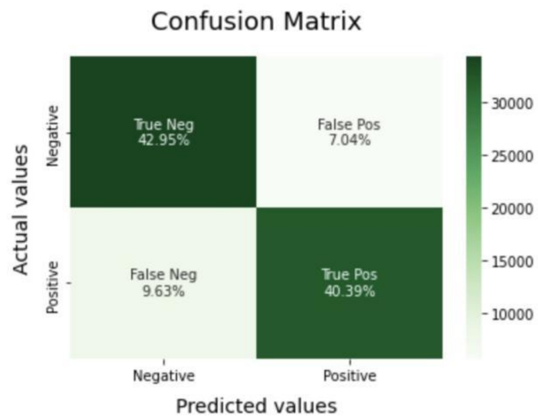

Figure 7. The confusion matrix of replacing with description method.

3.4. BiLSTM with replacing emoticons and emojis with predefined sentiments

Table 6 and Figure 8 show the training and testing processes of replacing emoticons with predefined sentiments by using the BiLSTM model. The training and validation accuracy is slightly higher than the simply removing method. And the training and validation loss is slightly lower in table 6. It has a similar trend as the simply removing method in figure 8.

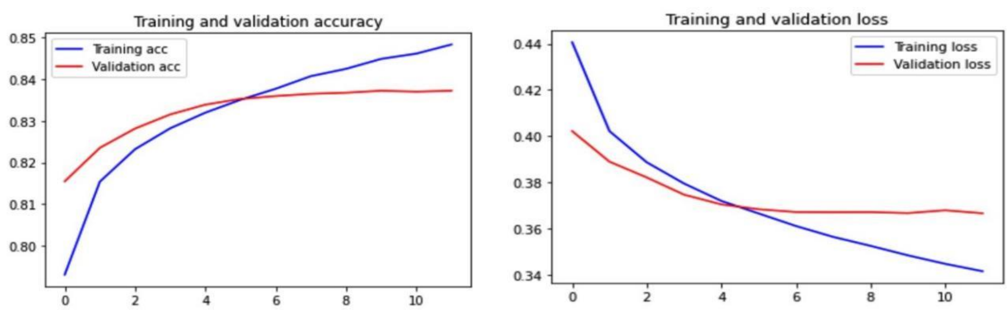

Figure 8. The trend of accuracy and loss.

The testing accuracy is the same as simply removing, which is shown in table 7. However, the third method tends to perform better in a tweet dataset where more tweets have emoticons. It has the best performance in training and validation accuracy, and it produces the lowest cross-entropy loss. Due to the Sentiment140 dataset only containing 2% tweets which have emoticons, the performance is limited.

DOI: 10.54254/2755-2721/6/20230415

Table 6. The specific training accuracy and cross-entropy loss.

Epoch	Running time	Speed	Loss	Accuracy	Validation loss	Validation accuracy
1	55 s	66ms/step	0.4400	0.7934	0.4081	0.8141
2	47s	64ms/step	0.4014	0.8157	0.3856	0.8249
3	48s	64ms/step	0.3876	0.8236	0.3883	0.8253
4	47s	64ms/step	0.3787	0.8288	0.3737	0.8318
5	47s	64ms/step	0.3713	0.8325	0.3775	0.8319
6	48s	64ms/step	0.3651	0.8356	0.3675	0.8356
7	48s	64ms/step	0.3600	0.8386	0.3685	0.8361
8	48s	64ms/step	0.3553	0.8411	0.3714	0.8365
9	47s	64ms/step	0.3512	0.8428	0.3672	0.8375
10	47s	64ms/step	0.3474	0.8450	0.3667	0.8376
11	48s	64ms/step	0.3434	0.8469	0.3676	0.8382
12	48s	64ms/step	0.3397	0.8492	0.3691	0.8378

Table 7. Evaluation matrices of BiLSTM by replacing the emoticons with the most correct sentiments.

	precision	Recall	F1-score	support
0	0.85	0.82	0.83	39989
1	0.83	0.85	0.84	40011
Accuracy			0.84	80000
Macro avg	0.84	0.84	0.84	80000
Weighted avg	0.84	0.84	0.84	80000

DOI: 10.54254/2755-2721/6/20230415

The replacing with sentiments method provides more accuracy in predicting True Positive in figure 9. Based on the 5 chosen emoticons and their popular sentiments, the positive sentiments dominate the expressions. Although some of the emoticons have opposite meanings, such as “smile face”. It is replaced with “happy, awkward”, the true meaning can be detected by consisting with other content of the tweet text. The predicational difference between True Negative and True Positive is acceptable for replacing emoticons with sentiments. It demonstrates that using suitable meaning for the same emoticons in different tweets using aspects, will potentially provide higher accuracy.

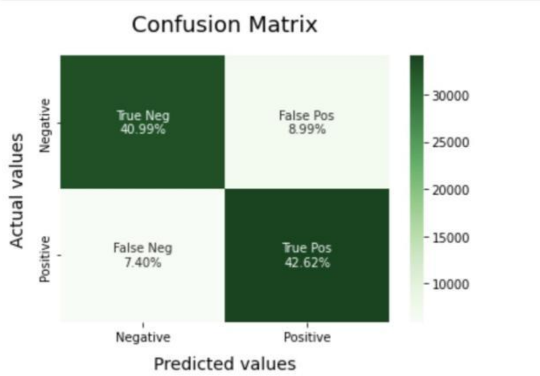

Figure 9. The confusion matrix of replacing with sentiments method.

4. Conclusions

This paper aims to discover the best way of processing emoticons and emojis. Based on the comparison results, the replacing emoticons with proper sentiments method is potentially the best way to help the BiLSTM model train and predict the tweet sentiments on the Sentiment140 dataset. By changing the meaning of the emoticons in different situations, it provides a stable result in increasing accuracy for a common deep learning method in different tweet aspects when the number of tweets with emoticons is small. Moreover, comparing with the description method, it is more straightforward for the neuron network to learn the true relations between text contents and the meanings of the emoticons, despite the predefined sentiments for the emoticon may be opposite. In the future, we will study on discovering a general form of automatically predefining the sentiments of different emoticons or emojis in different sentiment analysis fields, which will make the neural network train and test without experts in semi-supervised learning in different NLP questions.

References

[1] Mandloi L and Patel R 2020 Twitter Sentiments Analysis Using Machine Learninig Methods. 2020 International Conference for Emerging Technology (INCET). 1-5.
[2] Goyal G 2022 Twitter Sentiment Analysis|Implement Twitter Sentiment Analysis Model. Analytics Vidhya.
[3] R Inc 2022 Sentiment Analysis Challenges: Everything You Need to Know. Repustate.com.
[4] Rish I 2001 An empirical study of the naive Bayes classifier. In IJCAI 2001 workshop on empirical methods in artificial intelligence. 3 22 41-46.
[5] Cortes C and Vapnik V 1995 Support-vector networks. Machine learning. 20 (3) 273-97.
[6] Cramer J S 2002 The origins of logistic regression (Technical report). Tinbergen Institute. 119 167–78.
[7] Williams R J, Hinton G E and Rumelhart D E. 1986 Learning representations by back-propagating errors. Nature. 323 (6088) 533–6.
[8] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2017 Attention Is All You Need.
[9] Go A, Bhayani R and Huang L 2009 Twitter sentiment classification using distant

DOI: 10.54254/2755-2721/6/20230415

supervision. CS224N Project Report. Stanford. 1 12.
[10] 2022 The Most Used Emoji on Twitter in Every Country. Crossword-solver.io.
[11] Mahto P 2021 Text Preprocessing: How to handle Emoji ‘😄’ & Emoticon ‘ :-) ’?. Medium.
[12] Francesco B, Francesco R and Horacio S 2016 What does this emoji mean? A vector space skip-gram model for twitter emojis. In Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC 2016. Portoroz. Slovenia.
[13] Ayvaz S and Shiha M 2017 The Effects of Emoji in Sentiment Analysis. International Journal of Computer and Electrical Engineering. 9 1 360-9.
[14] Althobaiti M 2022 BERT-based Approach to Arabic Hate Speech and Offensive Language Detection in Twitter: Exploiting Emojis and Sentiment Analysis. International Journal of Advanced Computer Science and Applications. 13 5.
[15] Singh A, Blanco E and Jin W 2019 Incorporating Emoji Descriptions Improves Tweet Classification. Proceedings of the 2019 Conference of the North.
[16] 2022 Papers with Code - BiLSTM Explained. Paperswithcode.com.
[17] Karamitsos I, Afzulpurkar A and Trafalis T 2020 Malware Detection for Forensic Memory Using Deep Recurrent Neural Networks. Journal of Information Security. 11 02 103-120.
[18] Mikolov T, et al. 2013 Efficient Estimation of Word Representations in Vector Space.
[19] Mikolov T 2013 Distributed representations of words and phrases and their compositionality. Advances in Neural Information Processing Systems.