Sequential Models for Sentiment Analysis: A Comparative Study

Olaronke Oluwayemisi Adebanji, Irina Gelbukh, Hiram Calvo(✉), and Olumide Ebenezer Ojo

Instituto Politécnico Nacional, Natural Language and Text Processing Laboratory,
Centro de Investigacion en Computación, CDMX, Mexico
i.gelbukh@nlp.cic.ipn.mx, hcalvo@cic.ipn.mx

Abstract. Sentiment analysis has been a focus of study in Natural Language Processing (NLP) tasks in recent years. In this paper, we propose the task of analysing sentiments using five sequential models and we compare their performance on a Twitter dataset. We used the bag of words, as well as the tf-idf, and the Word2Vec embeddings, as input features to the models. The precision, recall, f1 and accuracy scores of the proposed models were used to evaluate the models' performance. The Bi-LSTM model with Word2Vec embedding performs the best against the dataset, with an accuracy of 84%.

Keywords: Sentiment analysis · Sequence modeling · Word embedding · Machine learning algorithm · Deep learning algorithm

1 Introduction

Sentiment analysis is an open research area in the field of Natural Language Processing (NLP). The use of sequential modeling techniques in sentiment analysis research has increased considerably. Many language models have been developed to help in the automatic processing and understanding of text in order to properly interpret the sentiments in text. These language models receive a sequence of data as input, examines each element of the sequence, then output the data in a sequence. Textual data has a structure that is based on the order of individual characters or words, which can be decoded as a sequence. There have been several examples of complex sequences being represented using machine learning algorithms. Sequence-to-sequence problems are being solved using a number of machine learning techniques which involves learning task-specific, nonlinear, and more abstract feature representations from raw data and has shown to be an excellent achievement for human language understanding.

Twitter is a rapidly developing and influential social media network that allows users to send and receive short messages known as tweets. As an integral part of the online community, it is one of the most widely used social media platforms in the world. Tweets contain opinions on a wide range of topics in

O. Pichardo Lagunas et al. (Eds.): MICAI 2022, LNAI 13613, pp. 227–235, 2022.
https://doi.org/10.1007/978-3-031-19496-2_17

different fields of life. Text in tweets is one of the most popular forms of sequence data. Although seen as a string of letters or a string of words, sequence data can be passed into sequential models for sentiment analysis. Previous sentiment analysis research looked at the efficacy of several classifiers [1,5,13,16], including transformer models [8], on a variety of datasets.

In this study, we analyze the sentiments in tweets using well-known machine learning and neural network techniques. More importantly, we publish the results of our own experiments on the same dataset using multiple approaches, allowing for direct comparison. The investigation was conducted in the context of processing sequential data by employing sentiment 140 datasets [5] containing 1,600,000 tweets extracted from Twitter where there are 800,000 tweets annotated as negative and 800,000 positive tweets. Following our inquiry, we evaluated the well-known algorithms in order to show that the proposed sequential models are effective in predicting sentiments in text. The following are the other sections that make up the paper: Sect. 2 gives a quick overview of sentiment analysis research, and Sect. 3 discusses the technique for cleaning and processing the datasets, as well as the algorithms used in this experiment. Section 4 provides a detailed explanation and analysis of our findings, while Sect. 5 summarizes the conclusion and future plans.

2 Literature Review

Different traditional machine learning techniques [13,14,19], as well as neural network models [1,3,6,11,15,16] have been utilized in the past to learn how to predict the sentiments in text. In the task of analysing sentiments, the Naive Bayes Algorithm, Decision Tree Classifier, Logistic Regression, Support Vector Machines and other machine learning classifiers in [13,14,19] with varied parameters (n-gram size, corpus size, number of sentiment classes, balanced vs. unbalanced corpus, multiple domains) performed well. Different pre-processing methods were applied, with the use of multiple classifiers in the experiments resulting in a more efficient evaluation than any single classifier.

The effectiveness of deep neural network models of varied complexity was leveraged on in [1] to automatic detect aggression in social media posts. The trials were carried out using models ranging in complexity from CNN, LSTM, BiLSTM, CNN-LSTM, LSTM-CNN, CNN-BiLSTM, and BiLSTM-CNN to BiLSTM-CNN. The models were able to perform better in classes where there are more training examples, but not in other classes.

An experiment was conducted in [3] to detect the use of violent threat language in YouTube comments to individuals and groups. The investigation was conducted using two text representations: bag of words (BOW) and pretrained word embedding such as GloVe and fastText. Deep learning classifiers such as 1D-CNN, LSTM, and bidirectional LSTM (BiLSTM) were applied, and it was discovered that deep learning outperforms other methods.

S. Poria et al. [16] were successful in analyzing features from short texts using a deep Convolutional Neural Network (CNN) based on multiple kernel learning.

In this case, a faster variant of their technique was developed based on decision-level fusion, which includes assigning a weight to the classifier, and was able to improve the performance of the multimodal sentiment analysis framework. The authors in [11] employed a neural architecture based on recurrent neural networks to maintain the track of each individual party's status during a conversation and used that knowledge for emotion analysis. This scalable technique examines each incoming utterance in light of the speaker's attributes, thereby giving the utterance a richer context.

The interpretations in these studies were based on real dataset inputs that maximized each output with respect to an input sequence. These sentimental tweets are important in a range of fields, including health [2], economics [13, 14] and political campaign news [4]. We compared the performance of multiple machine learning and neural network models for sentiment analysis using well-known methods of processing sequential inputs.

3 Experimental Setup

We apply different machine learning and deep learning classifiers for the task of tweets classification into positive and negative classes. The accuracy of the various algorithms proposed were obtained and compared. We computed each model's performance scores and drew inferences from them.

Dataset. The proposed techniques were applied to the Sentiment140 dataset [5] to predict the positive and negative classes and evaluated using the precision, recall, f1 and accuracy scores. The Sentiment140 dataset was created using the sentiment140 corpus, which contains 1.6 million tweets stripped of emoticons, 50% of which are positive and 50% of which are negative. The data was cleaned by removing information that isn't relevant to the analysis and tokenized to turn raw data into usable data that can be digested by the models. The bag of words with the term frequency-inverse document frequency (tf-idf) method [17] was used to rescale the data. We also represented the document vocabulary using Word2Vec representations as an embedding, which aids the deep learning algorithm in automatically understanding word analogies. The pre-processed data is separated into two groups, with the training data accounting for 80% of the total dataset and the test data accounting for 20%. The statistics of the data is as shown in Table 1.

Table 1. Statistics of the dataset

Dataset	No. of tweets
Training data	1,280,000
Testing data	320,000

Models. In this section, we propose to explain the various sequential models used for sentiment analysis in this task. Word2Vec [12] is a model architecture for learning word embeddings from huge datasets. The bag of words (with tf-idf) and word embedding (with Word2Vec) features were extracted as input into the machine learning and deep learning algorithms used.

Some machine learning techniques that has performed well in some sentiment analysis tasks [7,9,10,13,20] were proposed, which includes the Naive Bayes Algorithm (NBA), Random Forest Classifier (RFC), Support Vector Machines (SVM), and Logistic Regression Model (LRM). The deep learning approach implemented is the bidirectional long-short term memory (Bi-LSTM) network [18] with the Word2Vec embedding used as input into it. These machine learning and deep learning models have been thoroughly evaluated for different sentiment analysis tasks and have shown consistently good results when working with a variety of dataset types. The Word2Vec-learned embeddings combined with the Bi-LSTM method outperformed the other learning algorithms in this natural language processing task of sentiment analysis.

4 Results and Discussion

In this paper, we present a comparison of multiple sequential models in a sequence labeling task applied on Twitter dataset. The logistic regression model (LRM), support vector machine (SVM), naive bayes algorithm (NBA), and random forest classifier (RFC) were the machine learning techniques applied to the dataset. The Bi-LSTM model with Word2Vec embedding was used to analyse the sentiments in the data for better and quicker decision making as a deep learning method, and we were able to compare the output of all the approaches proposed. The datasets were used as input into the algorithms in order to predict sentiments and classify them accordingly. Table 2 shows the features applied and the class of the dataset. The precision, recall, f1 and accuracy scores of the different models used were also presented. Among the models, the Bi-LSTM model with Word2Vec embedding performs the best with an accuracy of 84%. The accuracy score was our assessment metric since our dataset has an equal amount of positive and negative tweets. We also plotted the Confusion Matrix to see how our model performed against the dataset. Our analysis was able to show that the word embeddings was able to learn bigger dimensions quickly from enormous corpora of text (Figs. 1, 2, 3, 4 and 5).

Table 2. Precision, Recall, F1 and Accuracy Scores of the five proposed models.

Model	Features	Class	Precision	Recall	F1	Accuracy
LRM	Bag of Words+TFIDF	Negative	0.76	0.73	0.74	75%
		Positive	0.74	0.77	0.75%	
SVM	Bag of Words+TFIDF	Negative	0.74	0.75	0.74	74%
		Positive	0.75	0.73	0.74%	
NBA	Bag of Words+TFIDF	Negative	0.74	0.77	0.75	75%
		Positive	0.76	0.72	0.74%	
RFC	Bag of Words+TFIDF	Negative	0.74	0.62	0.68	70%
		Positive	0.67	0.78	0.72%	
Bi-LSTM	Word2Vec	Negative	**0.84**	**0.86**	**0.85**	**84%**
		Positive	**0.85**	**0.83**	**0.84%**	

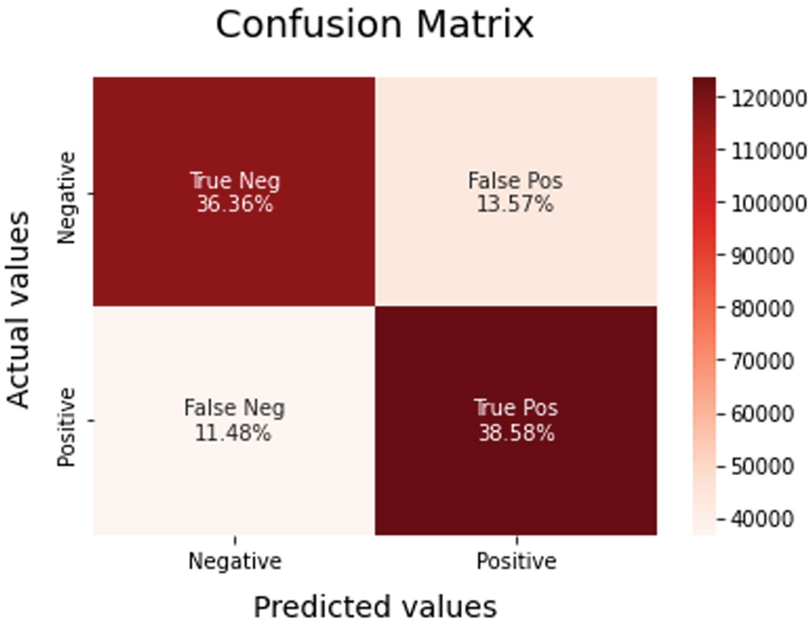

Fig. 1. Confusion matrix of the logistic regression model

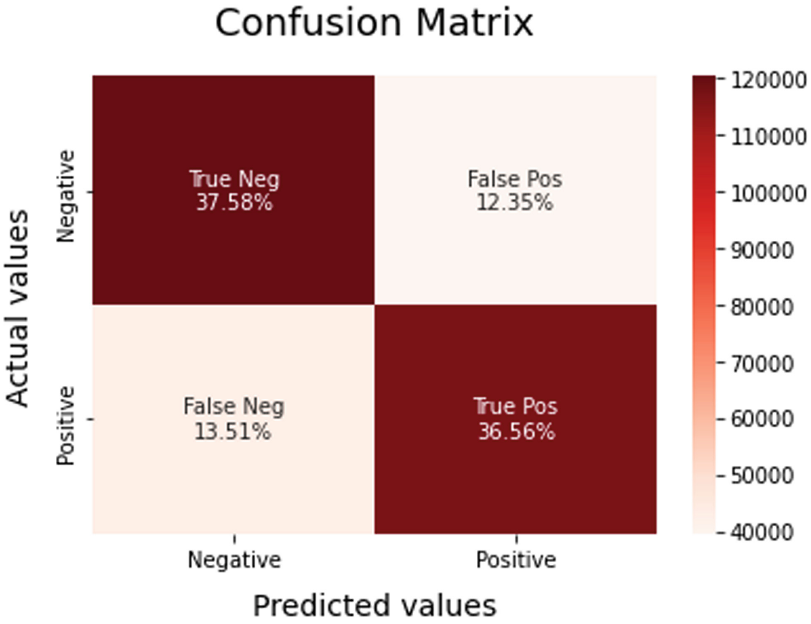

Fig. 2. Confusion matrix of the support vector machine

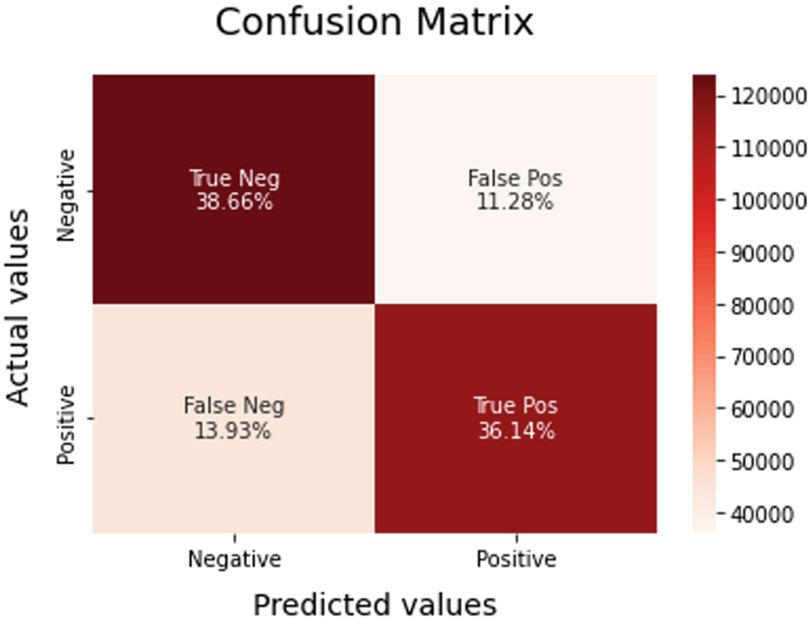

Fig. 3. Confusion matrix of the Naive Bayes algorithm

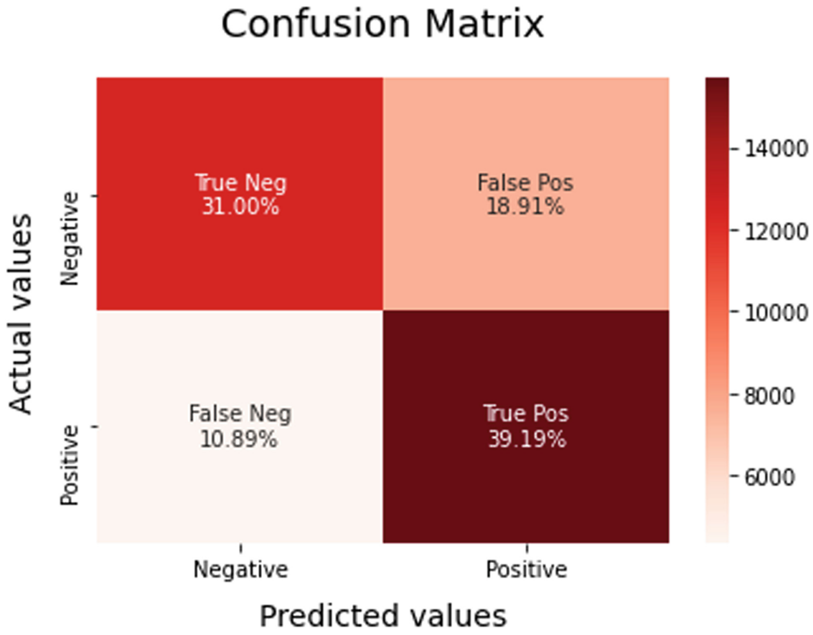

Fig. 4. Confusion matrix of the random forest classifier

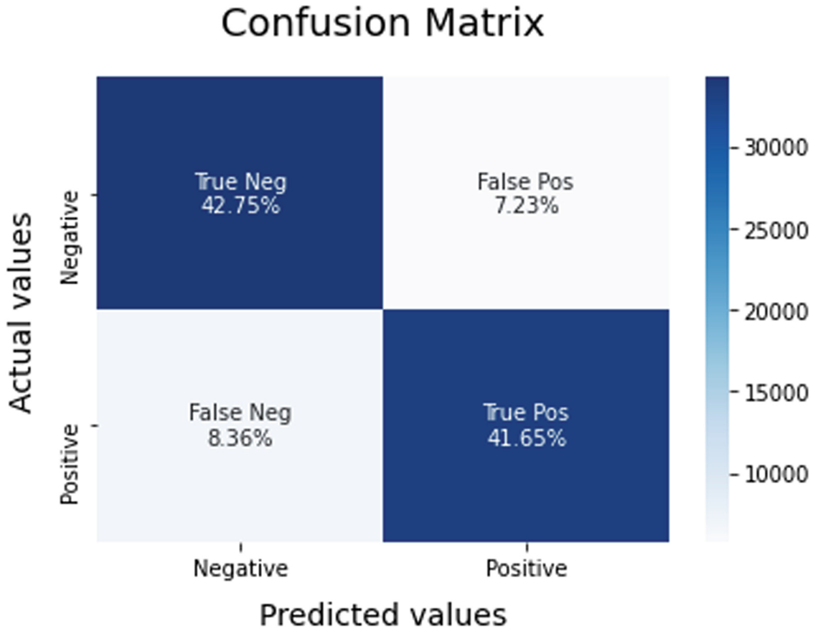

Fig. 5. Confusion matrix of the Bi-LSTM model

5 Conclusion

We investigated the performance of sequential models in the task of sentiment analysis of tweets using the sentiment140 dataset. It was discovered that the performance of a classifier is influenced by the feature representation and the type of model used. We primarily employed bag-of-word (with tf-idf), and word embedding (with Word2Vec) as features in this study, and discovered that the Word2Vec feature with deep learning method had an average gain in accuracy of 10% above the bag-of-words approach with traditional machine learning methods. The most significant advantage of employing deep learning techniques is that they incrementally learn high-level features from data. In the future, we plan to use other word embedding techniques and apply different transformer models to this task.

Acknowledgement. The work was done with partial support from the Mexican Government through the grant A1-S-47854 of the CONACYT, Mexico, and by the Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico, under Grants 20211884, 20220859, and 20220553, EDI; and COFAA-IPN.

References

1. Aroyehun, S.T., Gelbukh, A.: Aggression detection in social media: using deep neural networks, data augmentation, and pseudo labeling. In: Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018), pp. 90–97. Association for Computational Linguistics, Santa Fe, New Mexico, USA, August 2018
2. Aroyehun, S.T., Gelbukh, A.: Detection of adverse drug reaction in tweets using a combination of heterogeneous word embeddings. In: Proceedings of the Fourth Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task, pp. 133–135. Association for Computational Linguistics, Florence, Italy, August 2019. https://doi.org/10.18653/v1/W19-3224, https://aclanthology.org/W19-3224
3. Ashraf, N., Mustafa, R., Sidorov, G., Gelbukh, A.F.: Individual vs. group violent threats classification in online discussions. In: Companion of The 2020 Web Conference 2020, Taipei, Taiwan, 20–24 April 2020, pp. 629–633. ACM/IW3C2 (2020). https://doi.org/10.1145/3366424.3385778
4. Clarke, I., Grieve, J.: Stylistic variation on the Donald Trump twitter account: a linguistic analysis of tweets posted between 2009 and 2018. PLOS ONE **14**, 1–27 (2019). https://doi.org/10.1371/journal.pone.0222062
5. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. Processing 1–6 (2009)
6. Han, W., Chen, H., Gelbukh, A., Zadeh, A., Morency, L.P., Poria, S.: Bi-bimodal modality fusion for correlation-controlled multimodal sentiment analysis. In: Proceedings of the 2021 International Conference on Multimodal Interaction, pp. 6–15. ICMI 2021. Association for Computing Machinery, New York, NY, USA (2021)
7. Hernández-Castañeda, A., Calvo, H., Gelbukh, A., Flores, J.J.: Cross-domain deception detection using support vector networks. Soft Comput. **21**(3), 585–595 (2017). https://doi.org/10.1007/s00500-016-2409-2

8. Hoang, T.T., Ojo, O.E., Adebanji, O.O., Calvo, H., Gelbukh, A.: The combination of BERT and data oversampling for answer type prediction. In: CEUR Workshop Proceedings, vol. 3119. CEUR-WS (2022)
9. Kolesnikova, O., Gelbukh, A.: Supervised machine learning for predicting the meaning of verb-noun combinations in Spanish. In: MICAI (2010)
10. Kolesnikova, O., Gelbukh, A.: A study of lexical function detection with word2vec and supervised machine learning. J. Intell. Fuzzy Syst. **39** (2020)
11. Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A., Cambria, E.: DialogueRNN: an attentive RNN for emotion detection in conversations. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33(01), pp. 6818–6825 (2019). https://doi.org/10.1609/aaai.v33i01.33016818
12. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. CoRR abs/1301.3781 (2013). http://dblp.uni-trier.de/db/journals/corr/corr1301.html#abs-1301-3781
13. Ojo, O.E., Gelbukh, A., Calvo, H., Adebanji, O.O.: Performance study of n-grams in the analysis of sentiments. J. Nigerian Soc. Phys. Sci. **3**(4), 477–483 (2021). https://doi.org/10.46481/jnsps.2021.201
14. Ojo, O.E., Gelbukh, A., Calvo, H., Adebanji, O.O., Sidorov, G.: Sentiment detection in economics texts. In: Advances in Computational Intelligence: 19th Mexican International Conference on Artificial Intelligence, MICAI 2020, Mexico City, Mexico, 12–17 October 2020, Proceedings, Part II, pp. 271–281. Springer-Verlag, Berlin, Heidelberg (2020). https://doi.org/10.1007/978-3-030-60887-3_24
15. Ojo, O.E., Hoang, T.T., Gelbukh, A., Calvo, H., Sidorov, G., Adebanji, O.O.: Automatic hate speech detection using CNN model and word embedding. Computación y Sistemas **26**(2) (2022)
16. Poria, S., Cambria, E., Gelbukh, A.: Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. In: EMNLP (2015)
17. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Inf. Process. Manage. **24**(5), 513–523 (1988)
18. Schuster, M., Paliwal, K.: Bidirectional recurrent neural networks. IEEE Trans. Signal Process. **45**(11), 2673–2681 (1997). https://doi.org/10.1109/78.650093
19. Sidorov, G., et al.: Empirical study of machine learning based approach for opinion mining in tweets. In: Batyrshin, I., González Mendoza, M. (eds.) MICAI 2012. LNCS (LNAI), vol. 7629, pp. 1–14. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37807-2_1
20. Sidorov, G., Velasquez, F., Stamatatos, E., Gelbukh, A., Chanona-Hernández, L.: Syntactic N-grams as machine learning features for natural language processing. Expert Syst. Appl. 41(3), 853–860 (2014). https://doi.org/10.1016/j.eswa.2013.08.015