ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 3. Issue 1 (2023)
https://doi.org/ 10.62019/abbdm.v3i1.140

ASIAN BULLETIN OF BIG DATA MANAGEMENT
http://abbdm.com/
ISSN (Print): 2959-0795
ISSN (online): 2959-0809

A Machine Learning Approach for Textual Sentiment Classification

Huma Tauseef, Tehmina Shahid, Naveed Iqbal*, Aqsa Shabbir, Sahar Zia, Ahmad Faisal, Sajjad Rabbani

Chronicle

Article history

Received: Aug 11, 2023
Received in the revised format: Sept 4, 2023
Accepted: Sept 17, 2023
Available online: Sept 19, 2023

Huma Tauseef & Tehmina Shahid are currently affiliated with the Computer Science Department, Lahore College for Women University, Lahore, Pakistan.
Email: huma.tauseef@lcwu.edu.pk
Email: tehminashahid45@gmail.com

Naveed Iqbal is currently affiliated with the Department of Mass Communication, Lahore College for Women University, Lahore, Pakistan.
Email: naveed.iqbal@lcwu.edu.pk

Aqsa Shabbir & Sajjad Rabbani are **currently** affiliated with the Department of Electrical Engineering, Lahore College for Women University, Lahore, Pakistan.
Email: aqsa.shabbir@lcwu.edu.pk
Email: sajjadra94@gmail.com

Sahar Zia currently affiliated with the Department of Geography, Lahore College for Women University, Lahore, Pakistan.
Email: sahar.zia@lcwu.edu.pk

Ahmad Faisal currently affiliated with the Department of Electrical Engineering, National University of Science & Technology, Islamabad, Pakistan.
Email:amirza.bee21seecs@seecs.edu.pk

Corresponding Author*

Abstract

The proliferate use of microblogs and social networks have become valuable sources to determine individuals' opinion about an entity, product, topic, events, and politics etc. Main challenges implicated by the literature review include; contextual understanding and domain adaptation. Another challenge is that models trained on one domain may not generalize well to other domains due to differences sentiment expressions, or domain-specific terms. Therefore, textual sentiment analysis has become hotspot for research purpose. The paper proposes deep learning neural network models: LSTM and Bidirectional LSTM aims to automatically predict the sentiment polarity from given user posted text review into positive or negative class. Further, feature vectors are formed for each input sentence using Global vector for word representation (Glove) algorithm. Our proposed model utilizes 300-dimensional Glove for feature embedding. This high dimensional pre train vector contains semantically closer words. The impact of varying nature of datasets on the performance of both models for sentiment analysis is also investigated. Experiment is conducted on two Amazon product reviews datasets. The proposed research concluded that BLSTM achieved higher accuracy than Single LSTM and also outperforms the state -of-the-art models on document level reviews. Both models excel in capturing the temporal dependencies and linguistic structures. Highest accuracies of 94% and 96% are achieved on Amazon food reviews and Mobile reviews respectively using the BLSTM based model. The proposed model proved robust to changing sentiment trends or evolving language use as the datasets used are real-time datasets of thousands of users from different countries having variation in sentiment expression. The proposed model grasps the contextual understanding and domain adaptation very well.

Keywords: Amazon reviews, Bidirectional LSTM, Glove, LSTM, Sentiment analysis.

INTRODUCTION

Nowadays internet is used extremely as a discussion medium. Web has turned into a very helpful and effective communication way where people express their suggestions, share useful information and promote their products by posting a text through social media websites. These days a huge number of individuals use social websites like LinkedIn, Facebook, Twitter, amazon, daraz.pk and so forth which are creating sentiment rich information in the form of status, comments, posts, reviews, tweets etc. In the era of microblogging, social sites are considered to be a treasure of opinions. Textual Sentiment analysis have great importance for the companies, customer, Politics, detection of sarcasm and domestic violence. The research field of sentiment analysis refers to extract attitudes, behaviors, opinions and emotion of people embedded in the textual content published by user. People extremely depend upon online published user content to make any decision. Sentiment analysis is important for business as it is essential to have feedback to make products better and for political parties to know their voting rates (Kannangara, 2018). The process of extracting the high value information and predicting the behavior or attitude expressed in the blog post data is a very hot and interesting topic for researchers nowadays. A classification algorithm is applied to predict whether a piece of text has positive or negative implication. However, it's not easy to predict human feelings from raw data, there are some challenges that have to be faced in sentiment analysis. The limitation of text length and the informal type of the text makes sentiment analysis more challenging. Another challenge that must be faced is, in Sentiment extraction because of the unique characteristics of social channels. Noise in text, stop words and negation are some of the challenges that are confronted generally in text sentiment mining (Giachanou & Crestani, 2016).

The research in this paper aims to achieve the following objectives:

- To investigate the effectiveness of Long Short-Term Memory (LSTM) Bidirectional LSTM (BLSTM) in capturing contextual information for sentiment analysis tasks.
- To explore the impact of varying nature of datasets on the performance of LSTM and BLSTM models for sentiment analysis.
- To compare the performance of LSTM and BLSTM with other machine learning architectures, such as CNN-BLSTM or SR-LSTM models, for sentiment analysis tasks.

Two Neural network models: LSTM and BLSTM are evaluated in this paper. Sentiment Analysis is carried out to predict the sentiment of the product reviews. This research aims to investigate techniques for improving model performance in multiple product domains. The proposed models also prove robust to changing sentiment trends or evolving language use as the datasets used are real-time datasets of thousands of users from different countries having variation in sentiment expression. Different dimensional word vectors are used to extract features from the text. Further, the performance of both models is compared with the existed state of art models using different performance measure attributes. This paper is managed in following sections. Section 2 comprises the related work in the respective field of sentiment analysis. Section 3 describes the propose methodology used for this work. Research experiments and results are list out in section 4 and the last section 5 concluded the final results and propose future works.

LITERATURE REVIEW

This section describes the research work done in literature on textual sentiment analysis using latest techniques. Further it overviews the existing Machine learning (ML) and Deep learning approaches for text classification and feature extraction. In the last few years, a variety of new ideas have been developed in the research field for textual sentiment analysis. Sentiment analysis is also known as opinion mining. Textual sentiment classification can be performed at three different levels: Document level, Sentence level and feature level. At Document level, sentiment analysis refers to classify the opinion either into positive or negative class expressed in the whole document. In Sentence level sentiment classification, each sentence is classified into positive, negative or neutral class. Feature level SA refers to extract features and identify the sentiment of a specific entity from the source of data (Suganthi, & Geetha, 2017). There has been created many algorithms that predict whether a published text content is targeted or contain any opinion and whether text showed positive or negative sentiment. There are four approaches used to implement sentiment analysis namely: Machine learning, Lexicon, Hybrid and Graph. Machine learning based approach employs various machine learning algorithms either known as unsupervised algorithms (Terrana et al.,2014a) or supervised algorithms (Xia et al., 2015) to predict sentiment of text post.

The Lexicon based approach uses manually or automatically created dictionary of positive and negative words to derive the sentiment polarity(Medhat et al.,2014). Hybrid based approach uses a combination of machine learning and lexicon based approaches to improve the performance of classification(Ghiassi, Skinner, & Zimbra, 2013). In graph based approach, labels are distributed at the node and this approach can be applied to content contained any type of social relation (Terrana et al.,2014b). In the survey of sentiment analysis, the research is carried out on different opinion mining approaches for social media data. The survey underlines that ML and Deep learning methods have gained a prominence progress in the field of artificial intelligence and started to be implemented for sentiment analysis of textual data (Li et al., 2019). In another proposed framework, sentiment classification has conducted on unstructured product and service reviews using combination of two approaches of ML and evaluation models (Zablith & Osman, 2019).

While another paper on sentiment analysis addressed the Natural Language Processing problem using a three-layer structure organized as: syntactics, semantics, and pragmatics. It also provided the guideline to sentiment analysis including word sense disambiguation, part of speech tagging, text chunking, lemmatization, concept extraction, subjectivity detection, aspect extraction, and polarity detection issues (Cambria, Poria, Gelbukh, & Thelwall, 2017). While another study focused on the combined sentiment analysis of text and image using two separate Convolution Neural Network (CNN) based architectures for extracting textual and visual features which can be joint as input of another CNN architecture for exploiting the internal relation between text and image (Cai & Xia, 2015).

In the (Zhou et al., 2016) proposed framework, text classification is performed using combination of two models. First, it utilized the LSTM to convert long sequence sentences into vectors and then 2D max pooling operation is applied to get most meaningful features. The integration of 2D max pooling into BLSTM helped to get fix length feature vectors. While in a study, true authorship of text posted on social media has been analyzed using binary n-gram features and forensic linguistics techniques. They viewed text as ASCII codes then create n-grams (n=1-15) feature vectors from them stored as a profile for each text. Forensic linguistics method applied to get the information about authorship of text. Then they employed Euclidian distance algorithm k nearest neighbors (KNN) and outlier classifier IQR to predict the profiles of real authors. Their research concluded that KNN performs better with single n-gram features comparatively to Interquartile Range (IQR) (Peng, Choo, & Ashman, 2016). In (Das & Kolya, 2017) this paper, author proposed Naive Bayes algorithm to analyze the sentiment of tweets containing opinions about General sales tax (GST). They used dictionary approach to make feature set. Their research testing results showed the overall percentage of people's opinion about GST. While in a paper (Bansal & Srivastava, 2018), random forest algorithm showed better classification results on amazon reviews to get it classified. To extract the sentiment, they used very famous Continuous Bag of Words (CBOW) and skip gram algorithms of word2vec tool with 400 dimensional vectors. They applied multiple machine learning models from which random forest performs well amongst all by achieving the accuracy of 91%. Main challenges implicated by the literature review include; contextual understanding and domain adaptation. The proposed research improves the contextual understanding by grasping the context in which words or phrases are used using LSTM and BLSTM. Models trained on one domain may not generalize well to other domains due to differences sentiment expressions, or domain-specific terms. Both proposed models adapt well to multiple domains (food, cell phones).

METHODOLOGY

This section entails the proposed methodology implemented to perform sentiment analysis.

Preprocessing

Amazon reviews datasets for food and mobile reviews is collected. Amazon reviews datasets offer a rich and diverse source of data that can lead to the development of effective sentiment analysis models with real-world applications. The data is converted to lower case. Then it is cleaned by removing stop words, punctuation and special characters. Afterwards, stemming and lemmatization is done. Natural Language Toolkit (NLTK) version 3.4.5 is employed for data cleaning and further preprocessing of the text. Lastly, the ratings are converted into binary labels 0 or 1 to predict positive and negative sentiment of reviews. Proposed Model framework is extension to the baseline LSTM and BLSTM neural networks.

MODELS

LSTM

LSTM is the particular type of Recurrent neural networks, introduced by Hochreiter and Schmidhuber to remedied the gradient vanishing problem of Recurrent Neural Networks (RNNs). As RNNs has their own memory capable to persist previous inputs in internal state. RNN decays when its new input overwrites the activations and forget the previous inputs. LSTMs are more powerful to handle long term dependencies and variable length sequences. LSTM can add or delete information for a long period of time. Input weights are randomly assigned which then feed to feature vector embedding layer. The LSTM layer contains memory units which can be varies. To calculate inputs and their sums each layer uses activation function according to the classification problem. Hidden layers mostly used activation function Rectified linear unit (RELU). For binary classification output layer mostly uses Sigmoid function to predict the class label either 0 or 1.

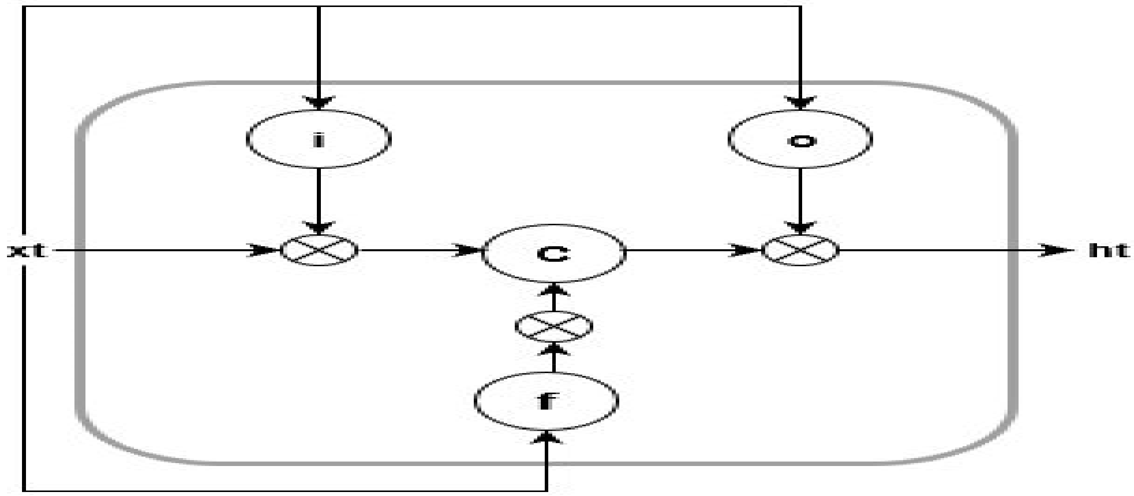


Figure 1.
Basic structure of LSTM unit

Table 1.
consists of hyperparameters setting for LSTM model.

Hyperparameters	Values
Layers	6
Input length	50
Maximum sequence length	50
Embedding dim	50
LSTM unit	100
Drop_out	0.50
Hidden layer function	RELU
Output layer function	Sigmoid
Optimizer	Adam
Epoch	8
Batch size	128

BLSTM

Bidirectional LSTM is extension to the baseline LSTM model with improved classification performance, based on the idea that at each time the output should be dependent on the previous input sequence and also on next input sequence. For example, it has to look both forward and backward to predict a missing input in a sequence. Simple LSTMs are unidirectional as they train only one LSTM on input sequences but BLSTM can train two LSTMs on input sequence which are stacked on top of each other. The first one on input sequence as it is in original order and the second one in the reverse direction of input sequence. Output of BLSTM is then computed which is based on hidden states of both LSTMs. Bi directional property makes LSTM faster and provides better training in sequence classification.

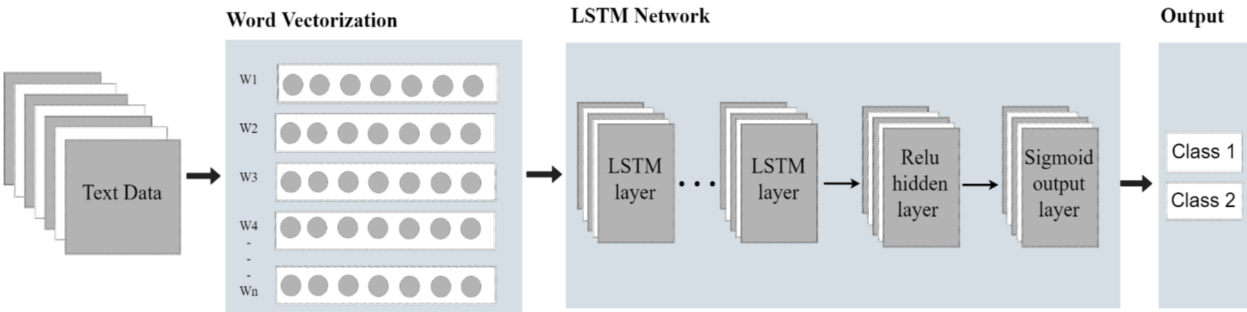


Figure 2.
The process of sentiment analysis using LSTM with Word vector representation

Table 2.
Hyperparameters setting for BLSTM model.

Hyperparameters	Values
Layers	4
Input length	50
Maximum sequence length	50
Embedding dim	300
LSTM unit	128
Drop_out	0.2
Hidden layer function	RELU
Output layer function	Sigmoid
Optimizer	Adam
Epoch	16
Batch size	128

Word Embeddings

Word embeddings is a technique belongs to Natural language processing (NLP) for feature learning also known as distributed representation of words. It aims to map each word into geometric space contained semantic meaning. The technique transforms each word into a real valued and low dimensional vector such that conceptually similar or related words become closer in the vector space. Each dimension of vector shows the idle feature of a word. Word embeddings are learned by two most commonly used neural network language models; word2vec and glove which learns embedding from text. Both models also published pre-trained word vectors that are trained on large corpus of text. For our experiment glove method is used to train word embeddings.

Glove Method

The technique used here for word embedding is Global Vectors for word representation (Glove) method which was developed at Stanford (Pennington, Socher, & Manning, 2014). Glove is a pre trained word embedding algorithm that efficiently learned word vectors by aggregating word to word co-occurrence matrix on non-zero entries using global statistics from a text dataset. Glove is extension to Word2vec method that outperforms other previous methods used for word vector embedding.

Experimental Results and discussion

In this section, the results obtained from the experiments are discussed. Sentiment analysis is done by implementing two deep learning models LSTM and BLSTM on different datasets. The experiments consist of two phases: one in which reviews are converted into word vector representation and second in which classification algorithm is used.

Dataset

The datasets used in the experiments to evaluate the proposed model are taken from the Kaggle website available publicly. Datasets consists of reviews contained following fields: id, product id, user id, profile name, helpfulness votes, rating score, time, summary and review text. However, we used only three fields that are rating score, summary and review text and rest of unnecessary fields are removed from dataset. To perform binary classification, rating scores are converted into two types of labels: positive and negative. Reviews that have scores from 1-2 assigned the label '0' and from 4-5 assigned label '1', indicating negative and positive sentiment respectively. Reviews with score 3 are considered as neutral so we didn't take them into account. Statistics of both datasets are shown in Table 3. Holdout Validation method was used by splitting the dataset into training, validation, and test sets. The model is trained on the training set, hyperparameters are tuned using the validation set, and the final evaluation is performed on the test set. The whole dataset was split into 67% for training and 33% for validation and testing.

Table 3.
Statistical, training, validation and testing set information of datasets used.

Dataset	Dataset size	Train set	Validate set	Test set
Food reviews	332474	222757	54858	54859
Mobile reviews	121943	81701	20121	20121

Evaluation Metrics

Proposed models are evaluated for performance using four metrics; accuracy, precision, recall, and Area Under the Curve (AUC). Accuracy is the most common measure to evaluate the performance of a classification algorithm. Eq. (1-4) show the formulas of these metrics.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

where:

TP: True Positives (correctly predicted positive instances)

TN: True Negatives (correctly predicted negative instances)

FP: False Positives (incorrectly predicted positive instances)

FN: False Negatives (incorrectly predicted negative instances)

$$AUC = \Sigma[0 \text{ to } n-1] \left((TPR(i) + TPR(i+1)) * (FPR(i+1) - FPR(i)) / 2 \right) \quad (4)$$

where:

TPR(i): True Positive Rate at threshold i

FPR(i): False Positive Rate at threshold i

n: Number of threshold points in the ROC curve

EXPERIMENTS

Experiment 1

We compare the performance of single LSTM model on two datasets. In this experiment, the embedding layer used glove pre trained vectors of 50 dimension to extract semantic features and LSTM layer used 100 hidden units. Other hyperparameters used in hidden layers are a dropout of 0.5 and average function RELU. Output layer used activation function sigmoid. The maximum input length used is 50, Adam optimizer and loss function binary_crossentropy. The model gained accuracy on each dataset is respectively 95% and 92%. The proposed model achieves better accuracy on mobile review dataset.

Table 4.
Experiment 1 results on both datasets.

LSTM	Vec Dim	Acc	Prec	Rec	AUC
Mobile reviews	50	0.95	0.96	0.98	96.7
Food reviews	50	0.92	0.95	0.94	96.9

*Vec Dim = Vector Dimension, Acc= Accuracy, Prec= Precision, Rec= Recall, AUC= Area under the curve

Experiment 2

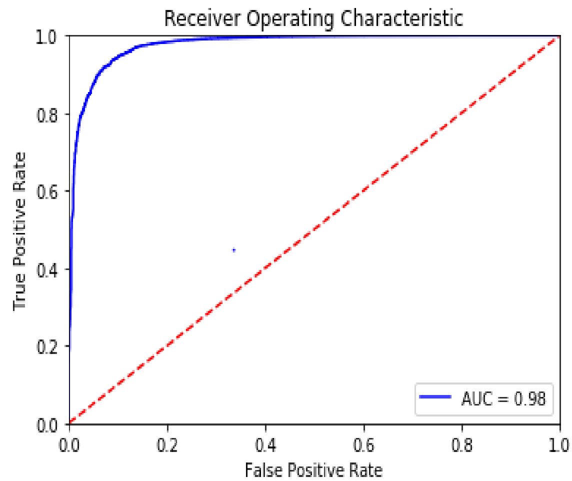
The BLSTM model was implemented on two datasets and its performance was compared. The model contains a BLSTM layer with 128 hidden units of output and a recurrent dropout of 0.2. The embedding layers used glove pre trained vectors of 300 dimension to extract semantic orientation information. An average activation function RELU was used in other hidden layer and output layers used sigmoid function to combine the output with kernel initialization to uniform. The model was trained in 16 epochs with training batch size of 128, loss of binary_crossentropy and adam optimizer.

The proposed model achieved 96% accuracy on mobile review dataset and 94% on food review dataset.

Table 5.
Experiment 2 result on both datasets.

BLSTM	Vec Dim	Acc	Prec	Rec	AUC
Mobile reviews	300	0.96	0.97	0.99	0.98
Food reviews	300	0.94	0.96	0.97	0.98

*Vec Dim = Vector Dimension, Acc= Accuracy, Prec= Precision, Rec= Recall, AUC= Area under the curve



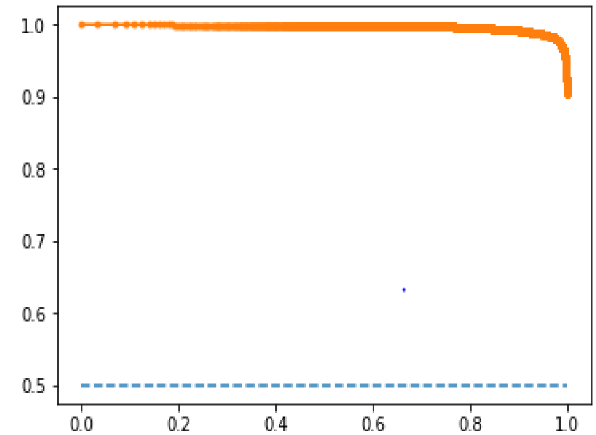


Figure 3.
ROC and AUC for BLSTM model

Figure 4.
Precision Recall graph for BLSTM

Comparison to baseline models

In this section, the results obtained from our experiments compared with other competitive baseline models in terms of their classification accuracies. Results of all models are reported in Table 7. First compare the traditional and famous deep learning model Convolutional neural network CNN and its two variants: single layer CNN and three-layer CNN. Now we could compare the single LSTM model with 50-dimensional glove and BLSTM model with 300-dimension glove. The experimental results showed that BLSTM has successfully achieved higher accuracy than Single LSTM. It is observed that as dimension increases the classification results got more accurate which could be because high dimensional pre train vector contains semantically closer words. BLSTM performed better classification also because of its property of forward and backward feed of input output.

Table 6.
Comparison of accuracies achieved from both experiments.

Model	Amazon mobile reviews	Amazon food reviews
LSTM	95%	92%
BLSTM	96%	94%

Table 7.
Comparison of classification accuracies of proposed model against baseline models

	Accuracies (%)	
Models	**Mobile review**	**Food review**
Single layer CNN	96	92
Three layer CNN	94	84
GRU	94.9	93.71
RCNN	95.7	93.42
LSTM	95	92
BLSTM	**96**	**94**

It is evident from Table 8 that our model outperforms all state of art models. Table 8 shows the comparative analysis of classification accuracies with competitive models. First, we compare the results of two machine learning models BLR and NB (Barnaghi, Ghaffari, & Breslin, 2016). It can be found that BLR has better performance on twitter data than NB. They achieved 74% and 66% classification accuracy respectively which is lesser than our results. They have used multiple feature extraction techniques: Uni gram, bi-gram, External Lexicon and TF-IDF for their models. While in other work NB, bagging (SVM), single SVM and Majority voting have applied on three datasets (Catal & Nangir, 2017). Amongst all model, majority voting got high classification accuracy of 86, 83 and 79 on books, movies and shopping reviews datasets respectively. They have also chosen TD-IDF and bi-grams features. Majority performed well but our model still performs better than machine learning models because neural networks are faster and they don't need a lot of dataset for training, also LSTM are popular for learning sequences of data.

Now we could compare the performance of proposed deep neural network model 1-layer CNN- BLSTM on IMDB movie review dataset which performed quite better than machine learning algorithms (Shen, Wang, & Sun, 2017). Their model achieved 89.7% accuracy using 50-dimensional glove embedding. BLSTM are more powerful that they alone can successfully learn large data with better accuracy if embedding dimension would be increases as our experiment shows. In Recent proposed work (Rao, Huang, Feng, & Cong, 2018), SR-LSTM and SSR-LSTM have performed on IMDB, Yelp 2014 and 2015 datasets on which they got same accuracies of 63% for both models. They have used glove embedding of 300 dimensional with the idea of sorted input of sentences but their accuracies are less than our BLSTM because their approach used fix number of sentences in a document.

In another work (Okada, Yanagimoto, & Hashimoto, 2018) gated CNN model with max pooling via SPP (spatial pyramid pooling) has proposed which divides the convolutional output into sub partitions and conducts Max pooling for each partition. Their model showed a good performance on amazon product review dataset with accuracy of 91% with simple SPP means simple max pooling and results got poor as SPP levels upgraded. It can be observed that gated mechanism of LSTM proved to be a good choice in sentiment analysis that's why in our case BLSTM showed a high performance on amazon reviews. NgramCNN (Çano & Morisio, 2018) was designed to perform sentiment analysis on song lyrics, IMDB and amazon mobile reviews datasets using word2vec pretrained vectors for text features selection. It can be studied that their model performs good on IMDB reviews by getting 91.2% classification accuracy but showed poor performance on amazon reviews because dataset is very large.

Tuning of hyperparameters would lead to good performance accuracy as in our experiment was achieved. We also compared our results with proposed Neural bag of words Ngram (Jing, Li, & Duan, 2018) model that has been evaluated on Amazon and twitter datasets. It can be found that their model performed better on amazon dataset than twitter dataset with accuracy of 91.47 and 80.13% respectively. They first made n grams of text and then glove embeddings of 100 dimension was used to form word vectors. But in our case, we used 300-dimensional word embedding that showed prominent increase in classification accuracy as compared to them. In comparison to other proposed LSTM (Chen et al., 2019) with LDA topic clustering in which their model performed better on amazon mobile dataset with classification accuracy of 86.7%. It can be observed that the performance of their method is still less than our model as they used LDA to divide the training dataset into multiple topic sub datasets, then each topic sub-dataset used to train corresponding topic base learner. In our work, proposed Single LSTM and Bidirectional LSTM models achieved noticeable high accuracy score on Amazon mobile review dataset and food reviews dataset. BLSTM performs better than single LSTM on both datasets due to its property that it can feed the input back and forward which takes less time and performs better classification. With high dimension of 300 word Embedding BLSTM achieved 96% and 94% accuracy on Amazon mobile and food reviews datasets respectively.

LSTM and BLSTM architectures for textual sentiment analysis have several advantages over traditional methods. The performance metrics provide implications on the transformative potential of these neural network models in understanding and analyzing sentiment in text data. The traditional approaches often rely on handcrafted features and simplistic algorithms while LSTMs and BLSTMs leverage sophisticated mechanisms to capture long-term dependencies and contextual information present in textual data. Furthermore, the LSTM and BLSTM models handle sequential data features to adapt to the inherently dynamic nature of text. By maintaining contextual information over extended sequences, these models excel in capturing the temporal dependencies and linguistic structures. In contrast, traditional methods struggle to capture the intricate interplay of words and phrases in a piece of text.

Table 8.
Comparison of results obtained from proposed model with state of art models.

Reference	Model	Dataset	Size	Training	Testing	Features	Accuracy (%)
2016 (Barnaghi, Ghaffari, & Breslin, 2016)	BLR, NB	Twitter dataset	4162	3746	416	Uni gram, bi-gram, External Lexicon, TF-IDF	74, 66
2016 (Catal & Nangir, 2017)	NB, Bagging, SVM Majority voting	Books, Movies, Shopping reviews	20623, 13156, 51879	19075, 10908, 46623	1548, 2248, 5256	TF-IDF, bi-grams	(85,83,82, 86) (82,80,81, 83) (79,76,79, 79)
(2017) (Shen, Wang, & Sun, 2017)	CNN-BLSTM	IMDB movie review dataset	50000	40000	10000	Glove 50d: Pre-trained word embedding vectors	89.7
(2018) (Rao, Huang, Feng, & Cong, 2018)	SR-LSTM, SSR-LSTM	IMDB, Yelp 2014 & 2015 datasets	208418 219701 76538	183019 194360 67426	25399, 25341, 9112	Glove 300d	SR-LSTM: 63.9 SSR-LSTM: 63.9
2018 (Okada, Yanagimoto, & Hashimoto, 2018)	Gated CNN	Amazon product reviews	20000	18000	2000	Spatial pyramid pooling	91

2018 (Çano & Morisio, 2018)	NgramCNN	IMDB movie, amazon mobile reviews.	50000, 232000	35000 162400	15000 69600	Word2vec	91.2
2018(Jing, Li, & Duan, 2018)	Neural Bag of words Ngram (NBOWN)	Amazon mobile, twitter airline	413827	351753	62074	Glove	91.47 80.13
2019 (Chen et al., 2019)	LSTM + ensemble learners	Amazon mobile, Amazon electronic, IMDB review	48000	40000	8000	LDA topic Cluster	86.7, 83.6, 82.1
Proposed Model	LSTM, BLSTM	Amazon food reviews, Amazon Mobile reviews	332474 121943	277616 101822	54858 20121	Glove	(92, 95) (**94, 96**)

CONCLUSION

In this paper, the proposed LSTM framework introduced to solve the sentiment polarity problem in different domains. Single and Bi-directional LSTM models are analyzed on users' reviews datasets. The research method employs different dimensional word embedding by utilizing pre trained word vector algorithm Glove. Each word maps to its feature vector and feed to the LSTM for classification. Single LSTM uses 50-dimensional Glove and BLSTM uses 300-dimensional Glove for feature embedding. The performance of both models has been evaluated by conducting experiments on two amazon datasets. BLSTM shows significantly excellent performance on both datasets as it achieved 94% and 96% prediction accuracy. In conclusion, the implications of using BLSTM for textual sentiment analysis underscore a paradigm shift towards more sophisticated, context-aware, and data-driven approaches in understanding and analyzing sentiment in text data. These models offered advantages in capturing long-term dependencies, handling sequential data effectively, and leveraging bidirectional context information, making them well-suited for textual sentiment analysis tasks compared to traditional methods or simpler neural network architectures. In future work, we will experiment with other deep learning models and feature reduction techniques to achieve more classification accuracy. Also, we would plan to build models with other datasets from different domains for sentiment analysis.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the supervisors and their different affiliations.

Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.

Availability of data and material: In the approach, the data sources for the variables are stated.

Authors' contributions: Each author participated equally to the creation of this work.

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

Consent to Participate: Yes

Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

Bansal, B., & Srivastava, S. J. P. c. s. (2018). Sentiment classification of online consumer reviews using word vector representations. 132, 1147-1153.

Barnaghi, P., Ghaffari, P., & Breslin, J. G. (2016). Opinion mining and sentiment polarity on twitter and correlation between events and sentiment. Paper presented at the 2016 IEEE second international conference on big data computing service and applications (BigDataService).

Cai, G., & Xia, B. (2015). Convolutional neural networks for multimedia sentiment analysis. Paper presented at the Natural Language Processing and Chinese Computing: 4th CCF Conference, NLPCC 2015, Nanchang, China, October 9-13, 2015, Proceedings 4.

Cambria, E., Poria, S., Gelbukh, A., & Thelwall, M. J. I. I. S. (2017). Sentiment analysis is a big suitcase. 32(6), 74-80.

Çano, E., & Morisio, M. (2018). A deep learning architecture for sentiment analysis. Paper presented at the Proceedings of the international conference on geoinformatics and data analysis.

Catal, C., & Nangir, M. J. A. S. C. (2017). A sentiment classification model based on multiple classifiers. 50, 135-141.

Chen, Z., Teng, S., Zhang, W., Tang, H., Zhang, Z., He, J., . . . Fei, L. (2019). LSTM sentiment polarity analysis based on LDA clustering. Paper presented at the Computer Supported Cooperative Work and Social Computing: 13th CCF Conference, ChineseCSCW 2018, Guilin, China, August 18–19, 2018, Revised Selected Papers 13.

Das, S., & Kolya, A. K. (2017). Sense GST: Text mining & sentiment analysis of GST tweets by Naive Bayes algorithm. Paper presented at the 2017 third international conference on research in computational intelligence and communication networks (ICRCICN).

Ghiassi, M., Skinner, J., & Zimbra, D. J. E. S. w. a. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. 40(16), 6266-6282.

Giachanou, A., & Crestani, F. J. A. C. S. (2016). Like it or not: A survey of twitter sentiment analysis methods. 49(2), 1-41.

Jing, C., Li, J., & Duan, X. (2018). Learning Word Sentiment with Neural Bag-Of-Words Model Combined with Ngram. Paper presented at the Intelligence Science II: Third IFIP TC 12 International Conference, ICIS 2018, Beijing, China, November 2-5, 2018, Proceedings 2.

Kannangara, S. (2018). Mining twitter for fine-grained political opinion polarity classification, ideology detection and sarcasm detection. Paper presented at the Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining.

Li, Z., Fan, Y., Jiang, B., Lei, T., Liu, W. J. M. T., & Applications. (2019). A survey on sentiment analysis and opinion mining for social multimedia. 78, 6939-6967.

Medhat, W., Hassan, A., & Korashy, H. J. A. S. e. j. (2014). Sentiment analysis algorithms and applications: A survey. 5(4), 1093-1113.

Okada, M., Yanagimoto, H., & Hashimoto, K. (2018). Sentiment Classification with Gated CNN and Spatial Pyramid Pooling. Paper presented at the 2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI).

Peng, J., Choo, K.-K. R., Ashman, H. J. J. o. N., & Applications, C. (2016). Bit-level n-gram based forensic authorship analysis on social media: Identifying individuals from linguistic profiles. 70, 171-182.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. Paper presented at the Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP).

Rao, G., Huang, W., Feng, Z., & Cong, Q. J. N. (2018). LSTM with sentence representations for document-level sentiment classification. 308, 49-57.

Shen, Q., Wang, Z., & Sun, Y. (2017). Sentiment analysis of movie reviews based on cnn-blstm. Paper presented at the Intelligence Science I: Second IFIP TC 12 International Conference, ICIS 2017, Shanghai, China, October 25-28, 2017, Proceedings 2.

Suganthi, D., Geetha, A. J. I. J. o. S. R. i. C. S., Engineering, & Technology, I. (2017). Twitter Sentiment Analysis on GST tweets using R tool. 2(5), 793-796.

Terrana, D., Augello, A., & Pilato, G. (2014a). Automatic unsupervised polarity detection on a twitter data stream. Paper presented at the 2014 IEEE International Conference on Semantic Computing.

Terrana, D., Augello, A., & Pilato, G. (2014b). Facebook users relationships analysis based on sentiment classification. Paper presented at the 2014 IEEE International Conference on Semantic Computing.

Xia, R., Xu, F., Zong, C., Li, Q., Qi, Y., Li, T. J. I. t. o. k., & engineering, d. (2015). Dual sentiment analysis: Considering two sides of one review. 27(8), 2120-2133.

Zablith, F., & Osman, I. H. J. A. M. M. (2019). ReviewModus: Text classification and sentiment prediction of unstructured reviews using a hybrid combination of machine learning and evaluation models. 71, 569-583.

Zhou, P., Qi, Z., Zheng, S., Xu, J., Bao, H., & Xu, B. J. a. p. a. (2016). Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling.