Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7922943

HYBRID CNN-LSTM BASED MODEL FOR SENTIMENT ANALYSIS IN TWEETS

M.Srisankar
Research Scholar, Department of computer science, Government Arts college, udumalpet, Tamilnadu, India, shankar276@gmail.com,

Dr.K.P.Lochanambal
Assistant Professor, Department of computer science, Government Arts college, udumalpet, Tamilnadu, India, kplmca@rediffmail.com

Abstract: Deep Learning (DL) schemes offer solutions to a wide variety of problems. Mainly, two kinds of Neural Networks (NNs) namely, Convolutional Neural Networks (CNN) and Bi-directional Long Short Term Memory (LSTM) networks are used. CNN+Bi-a LSTM based DL scheme is used along with pre-trained scheme for automatically determining features to explore sentiments and categorise reviews as well as opinions based on positive or negative polarities. The proposed model offers better performance on standard datasets. The performance of proposed mechanism is compared with standard DL-based schemes. CNN+Bi-LSTM scheme is proposed for Sentiment Analysis (SA) of Twitter data. The performance of different Word Embedding (WE) systems like Word2Vec and global vectors (GloVe) for word representation are compared based on best scoring values. The proposed scheme offers 93% accuracy for categorizing tweets into negative as well as positive cases on benchmark dataset. The scheme uses LSTM-based NN with limited factors offering modest outcomes and outdid other ML schemes in terms of Accuracy, Precision, Recall and F1-Score.
Keywords: Sentiment Analysis, Deep Learning, Convolutional Neural Network (CNN), Bi-directional Long Short Term Memory (LSTM), word embedding, Twitter data

1. INTRODUCTION

With dramatic increase in the usage of social media, Sentiment Analysis (SA) has gained much popularity amid people with varying interests as well as motivations. As users express their views about diverse subjects related to politics, culture, education, commercial products, travel or subjects of overall interest, gaining understanding from data is important. In addition to information about sites visited by users and buying preferences, expressing their feelings through messages in different platforms plays a dominant role in estimating people's view of a specific subject.

A common technique is classification of text polarity based on satisfaction, dissatisfaction or neutrality. Polarity varies with labelling or quantity of levels ranging from positive to negative, expressing the feelings changing from happy to unhappy moods. There are several approaches involved in based on diverse techniques of Natural Language Processing (NLP) and Machine Learning (ML) schemes for determining necessary attributes and categorising text in suitable polarity tags. With popularity of Deep Learning (DL) techniques, several Deep Neural Networks (DNNs) are successfully used. Predominantly, Convolutional Neural Networks (CNNs) as well as Long Short-Term Memory (LSTM)

networks can be employed in SAs. Numerous studies have shown their efficiency when applied alone or in grouping. In NLP, among approaches used for determining attributes from words, Word2Vec as well as Global Vectors (GloVe) for word demonstration are highly popular.
The accuracy attained with above schemes is increased but not acceptable, thus making SA an open area of research. Several techniques are designed or current ones are improved. As present approaches have a variety based on network configuration, fine-tuning research on assessment of already employed techniques is essential to have accurate idea about limits as well as challenges in SA. In this paper, standard Deep Learning (DL) schemes along with configurations depending on acceptable dataset about SA are built from Twitter data in testing framework. Section 2 details about the work done by diverse authors related to SA. Section 3 shows the method and diverse NN configuration which are applied. In section 4, results are discussed which compares diverse approaches amid them. Section 5 gives the conclusion.

2. BACKGROUND

With increase in social media popularity and different platforms permitting people to show opinion on diverse subjects, SA and opinion mining has become a subject which attracts attention of researchers. Pang & Lee (2009) have detailed several SA methods used presently. DNNs are highly used in SA tasks. CNNs (Forsyth et al 1999) and Recurrent Neural Networks (RNNs) (Rumelhart et al 1985) are widely employed as CNN responds to dimensionality reduction. LSTM networks, a class of RNN are capable of handling temporal as well as sequential data (Hochreiter & Schmidhuber 1997). Chen (2015) has shown that CNN frameworks may be used in classification of sentences. Furthermore, it is seen that CNN offers considerable performance better than the traditional schemes (Ray & Chakrabarti 2022). Lai et al (2015) have shown the efficacy of RNN as they outperform the standard schemes. CNN and LSTM networks are implemented showing noteworthy merits of using 2 NNs. GRU networks (Chung et al 2014) is also used to perform efficient SA. Several DL schemes are used in SA (Zhang et al 2018), it is seen that Word Embedding (WE) involves 2 methods, Word2Vec (Mikolov et al 2013) or GloVe (Pennington et al 2014).Currently, Twitter is an influencing social media platforms that aids in information sharing (Kwak et al 2010). Public opinions are obtained from tweets about several subjects, measuring impact of diverse events or categorizing sentiments is of increased interest. Early works for SA using diverse approaches for determining features depending on unigrams, bi-grams and Part-of-Speech (PoS) particular polarity attributes. ML classifiers like Support Vector Machines (SVMs) or Bayesian networks (Kouloumpis et al 2011) are used in SA. Nowadays, DL schemes are efficient in gaining increased scores in competitions using diverse blends of NNs and several configurations of WE features. Related to SA in Twitter dataset, a quantity of studies are used in determining performance.Deriu et al (2016) have propounded diverse CNN configurations using Word2Vec as well as GloVe where outcomes are joined in a Random Forest (RF) classifier. Rouvier & Favre (2016) have used embeddings trained on lexical, PoS as well as sentiment embeddings that initialise input of deep CNN framework. Baziotis et al (2017) have propounded 2 configurations using LSTM networks that are bi-directional in nature. WE is done using GloVe. Cliche (2017) has propounded a method that is a blend of CNN and LSTM networks. Bojanowski et al (2107) have employed Word2Vec, GloVe as well as FastText models and have shown that GloVe does not offer improved performance. A modified CNN called RCNN

(Lei et al 2015) is used successfully (Yin et al 2017). It is comparatively challenging to assess role of dataset, network configuration or particular setup as well as tuning. A framework is designed to compare the schemes and clarify merits and limits of every specific configuration.

3. PROPOSED SCHEME

Dataset, WE models with different configurations and diverse DNN configurations are employed in the study. GRU networks and RCNNs are not incorporated as they give comparable outcomes along with LSTM networks as well as CNN.

3.1 Dataset and Pre-processing

Tweets are pre-processed so as to improve system's training performance. Additional pre-processing task is performed to remove and alter some characters. It involves conversion of letters to lowercase, removal of special characters along with emoticons and URL tagging.

3.2 Word Embedding (WE)

WE models like Word2Vec as well as GloVe are used. Word2Vec is used in creating Word Vectors (WVs) of size 25 depending on the given dataset. Word2Vec is configured by using Continuous Bag of Words (CBOW) model. In addition, words < 5 times are cast-off.
Lastly, maximum skip length amid words is assigned 5. GloVe is employed with pre-trained WVs are 25-dimensional and are built from tweets (2 billion) that establishes a considerably greater training dataset. Normalisation of vectors is done using subsequent equations:

$$\mathrm{v}_\mathrm{i}' = \frac{v_i - \mathrm{v}_{\min}}{\mathrm{v}_{\max} - \mathrm{v}_{\min}} \tag{1}$$

Where,

v_i' - Normalised i-value of 25-D vector

$\mathrm{v}_{\min}$ - Least value of vector

$\mathrm{v}_{\max}$ - Highest value of vector

- **Sentence Vectors:** They are built by combining tweet WVs to create a distinct vector. On considering diverse lengths, sentences of 40 words length are created. As tweets have diverse lengths, they have more number of words, and additional words are removed. In case of tweets involving sizes less than 40, words are repeated until anticipated size is attained. An alternate technique is to employ zero padding for filling missing words. In this paper, zero padding is done if the words are not present in vocabulary.
- **Sentence Areas:** The WVs of sentence found in regions are divided to conserve information along with long-distant dependence through sentences during prediction (Wang et al 2016). The split is done based on punctuation marks present in a sentence. In current configuration, every region includes 10 words with each sentence having 8 regions. If words or regions are absent, then zero padding is performed.

Dataset is transformed 2 times thus creating 2 separate datasets, including non-regional as well as regional-based sentences. Initially, the size of input is taken as 1250, wherein every sentence has 50 words of size 25, while the second is 2400, wherein every sentence is split into 8 regions with 12 words of size 25.

3.3 NEURAL NETWORKS (NNs)

CNN as well as LSTM networks are proposed for assessing twitter data. Moreover, SVM classifier is employed. The networks are verified with regional as well as non-regional datasets. Totally 8 configurations are offered.

RCNN as well as GRU networks are not employed as they offer comparable performance with respect to CNN as well as LSTM networks respectively. The networks are trained using 300 epochs using sigmoid activation functions.

- **CNN:** In this network, 1D CNN layer is employed. Sentence vector is convolved involving 12 kernels of size 1 × 3. It performs better in contrast to kernel configuration. The max pooling layer is of size 1 × 3. CNN factors are the same for ensuing CNN configurations. Lastly, 3D output forecasts polarity based on positive, negative as well as neutral response.
- **LSTM network:** A LSTM layer is employed involving a dropout of 20%. Output is of size 1 × 3 for predicting polarity (positive, neutral or negative).
- **CNN with LSTM Networks:** Outputs of CNN as well as LSTM networks are taken and their results are assessed together. Soft voting depending on network outputs selects the prediction response. CNN as well as LSTM networks have similar settings as 2 former configurations. CNN involves 12 kernels of size 1 × 3 with max pooling layer also of the same size.
- **3-Layer CNN with LSTM Networks:** It uses a 3-layer 1D CNN with LSTM network involving one layer. In this configuration, input is sent to 3-layered CNN. Input is of size 1250 in case it depends on words (non-regional) or 2000 in case it depends on regions (regional).
- **Numerous CNNs with LSTM Networks:** Input is split into fundamental elements, words (non-regional) along with regions (regional). The elements are given as input to CNN, the output of which is sent to a LSTM network. Based on kind of input, there are 50 or 8 respective CNNs. Each CNN network uses 12 kernels.
- **CNN with 3-Layers and Bi-directional LSTM (Bi-LSTM) Network:** This involves a configuration same as in the previous case with BiLSTM Network. The efficiency of this kind of LSTM networks is compared with conventional LSTM networks.
- **Several CNNs and BiLSTM Network:** This setup involves a configuration similar to previous one involving BiLSTM network.

3.4 METHODOLOGY

LSTM architecture (Hochreiter & Schmidhuber 1997) is used extracting features automatically. The proposed DL-based scheme is detailed in the ensuing section. Annotated Sentiment140 Twitter dataset is taken from Stanford University is used [23]. It is built using Twitter API and includes 1.6 million annotated tweets, either Negative or Positive. It is pre-processed using text normalization. Needless features like usernames, URLs and re-tweets are removed. Stemming is applied and stop words, additional white spaces, numbers and punctuation are removed, and text is converted to lowercase.

CNN+Bi-LSTM Framework

Glove 6B 300D are used as WE model in initial layer wherein, embedding learns words are taken from the dataset. Vocabulary size is taken as 15k with embedding layers' dimension as 30 x 300 matrix wherein, batch is of length 30. The features are passed from embedding to spatial drop out layer at a rate of 0.2 to circumvent overfitting. Output is fed into initial 1D CNN layer and filter size is 5 x 5. 64 filters are defined. This permits to train 64 diverse features on next layer of network. The second NN layer's output is a 26×64 matrix which is sent to Bi-LSTM layer (size 64) to determine extended range dependencies to find features and send to Fully Connected (FC) dense layer of size 128×512 neurons. Dense layer's output drops out with rate of 0.5 for dropping some arbitrary matrix weights. FC dense layer involving sigmoid function generates vector input to forecast units (positive, negative). The vector output is of length 512×1 neurons.Binary Cross Entropy (BCE) loss functions along with Adam optimizer include a learning rate of 1×10^{-4} for training model using 10-epochs of 1024 batch size.

4. RESULTS AND DISCUSSION

To annotated tweets, equal number of tweets is categorised as positive as well as negative. Python Keras is used with TensorFlow as backend. CNN-LSTM method of Sckit-learn toolkit is used. The training dataset includes 75% tweets with 10-epochs, and 25% is used for measuring performance after tuning.

The proposed CNN + Bi-LSTM offers 25%, 20%, 14% and 8% improved Accuracy in contrast to CNN, LSTM, CNN + LSTM and CNN + LSTM + SVM (Figure 2).

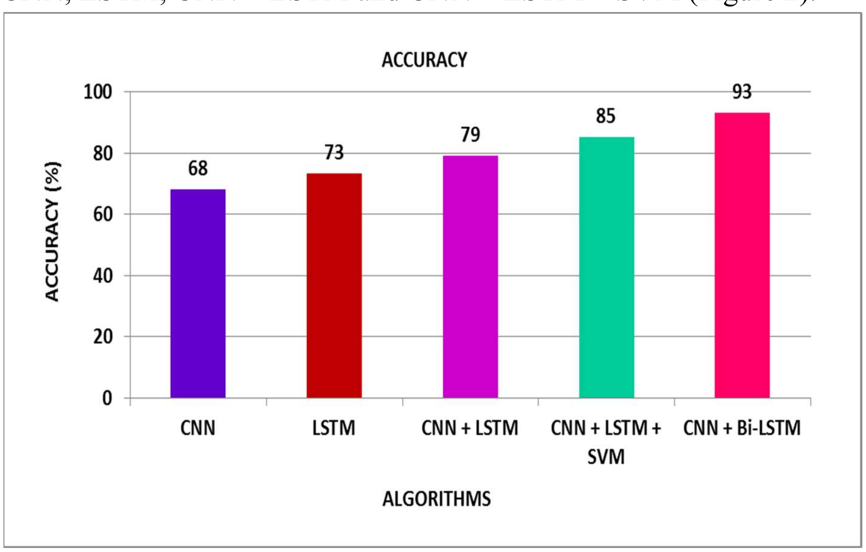

Figure 2: Accuracy

The proposed CNN + Bi-LSTM offers 29%, 22%, 18% and 11% better Precision in contrast to CNN, LSTM, CNN + LSTM and CNN + LSTM + SVM (Figure 3).

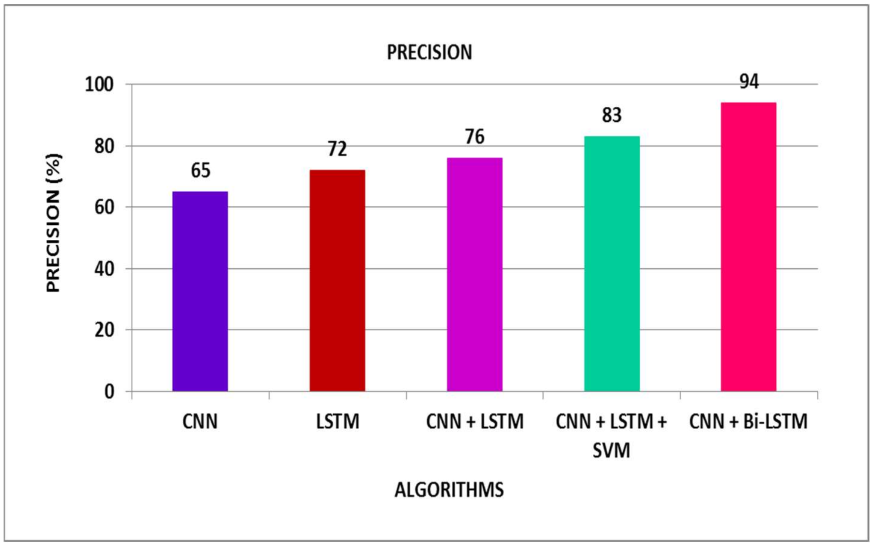

Figure 3: Precision

The proposed CNN + Bi-LSTM offers 26%, 17%, 14% and 5% improved Recall in contrast to CNN, LSTM, CNN + LSTM and CNN + LSTM + SVM (Figure 4).

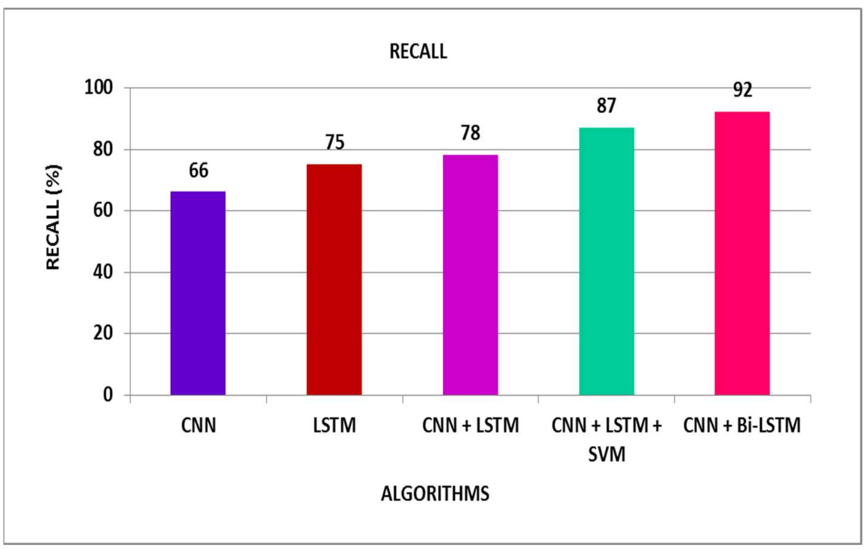

Figure 4: Recall

The proposed CNN + Bi-LSTM offers 26%, 20%, 15% and 9% increased F1-Score in contrast to CNN, LSTM, CNN + LSTM and CNN + LSTM + SVM (Figure 5).

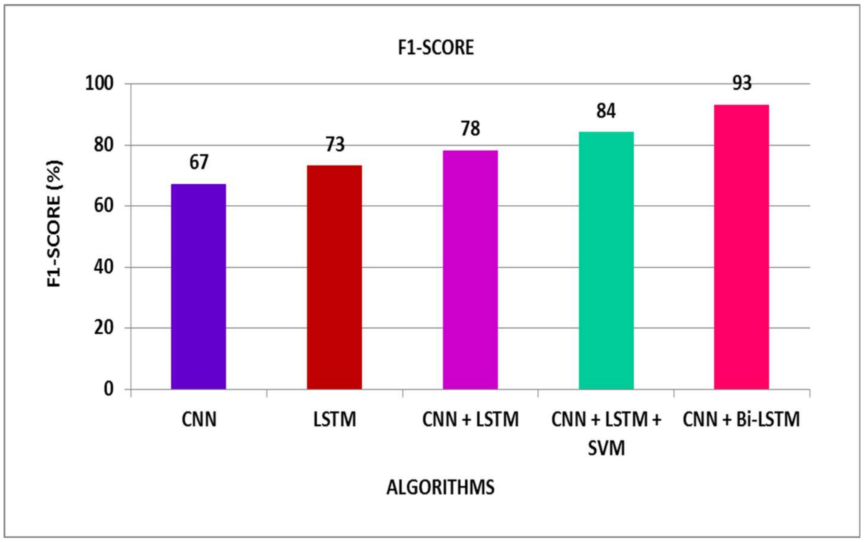

Figure 5: F1-Score

5. CONCLUSION

In this paper, CNN+Bi-LSTM based DL mechanism along with pre-trained model is proposed for SA, thus facilitating efficient learning that automatically finds features for determining sentiments and classifying reviews along with opinions depending on positive or negative polarities. It is seen that the proposed scheme offers improved performance on standard datasets in terms of Accuracy, Precision, Recall and F1-Score. The performance of WE systems like Word2Vec as well as GloVe for word representation are considered depending on best scores. The proposed scheme offers 93% classification accuracy for labelling tweets into negative as well as positive ones.

REFERENCES

1. Pang, B., & Lee, L. (2009). Opinion mining and sentiment analysis. Comput. Linguist, 35(2), 311-312.
2. Forsyth, D. A., Mundy, J. L., di Gesú, V., Cipolla, R., LeCun, Y., Haffner, P., ... & Bengio, Y. (1999). Object recognition with gradient-based learning. Shape, contour and grouping in computer vision, 319-345.
3. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). Learning internal representations by error propagation. California Univ San Diego La Jolla Inst for Cognitive Science.
4. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
5. Chen, Y. (2015). Convolutional neural network for sentence classification (Master's thesis, University of Waterloo).
6. Ray, P., & Chakrabarti, A. (2022). A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis. Applied Computing and Informatics, 18(1/2), 163-178.

7. Lai, S., Xu, L., Liu, K., & Zhao, J. (2015, February). Recurrent convolutional neural networks for text classification. In Proceedings of the AAAI conference on artificial intelligence (Vol. 29, No. 1).
8. Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.
9. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.
10. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
11. Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).
12. Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World wide web (pp. 591-600).
13. Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg!. In Proceedings of the international AAAI conference on web and social media (Vol. 5, No. 1, pp. 538-541).
14. Deriu, J. M., Gonzenbach, M., Uzdilli, F., Lucchi, A., De Luca, V., & Jaggi, M. (2016, June). Swisscheese at semeval-2016 task 4: Sentiment classification using an ensemble of convolutional neural networks with distant supervision. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) (pp. 1124-1128).
15. Rouvier, M., & Favre, B. (2016, June). SENSEI-LIF at SemEval-2016 Task 4: Polarity embedding fusion for robust sentiment analysis. In SemEval@ NAACL-HLT (pp. 202-208).
16. Baziotis, C., Pelekis, N., & Doulkeridis, C. (2017, August). Datastories at semeval-2017 task 4: Deep lstm with attention for message-level and topic-based sentiment analysis. In Proceedings of the 11th international workshop on semantic evaluation (SemEval-2017) (pp. 747-754).
17. Cliche, M. (2017). BB_twtr at SemEval-2017 task 4: Twitter sentiment analysis with CNNs and LSTMs. arXiv preprint arXiv:1704.06125.
18. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the association for computational linguistics, 5, 135-146.
19. Lei, T., Joshi, H., Barzilay, R., Jaakkola, T., Tymoshenko, K., Moschitti, A., & Marquez, L. (2015). Semi-supervised question retrieval with gated convolutions. arXiv preprint arXiv:1512.05726.
20. Yin, Y., Song, Y., & Zhang, M. (2017, August). Nnembs at semeval-2017 task 4: Neural twitter sentiment classification: a simple ensemble method with different embeddings. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017) (pp. 621-625).

21. Wang, J., Yu, L. C., Lai, K. R., & Zhang, X. (2016, August). Dimensional sentiment analysis using a regional CNN-LSTM model. In Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers) (pp. 225-230).
22. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
23. Available online: http://help.sentiment140.com/ site-functionality (accessed on 12 June 2020).