IJARIIT INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078

(Volume 8, Issue 2 - V8I2-1346)
Available online at: https://www.ijariit.com

Sentiment classification with the BERT procedure based on deep learning

P. Susmitha
sushmita.pathipati@gmail.com
Annamacharya Institute of Technology and Sciences, Rajampet, Andhra Pradesh

V. Swetha
swethavalluru7654@gmail.com
Annamacharya Institute of Technology and Sciences, Rajampet, Andhra Pradesh

S. Udaya
udayasoftware70@gmail.com
Annamacharya Institute of Technology and Sciences, Rajampet, Andhra Pradesh

K. Anusha
anu.anusha.2208@gmail.com
Annamacharya Institute of Technology and Sciences, Rajampet, Andhra Pradesh

P. Nagendra
pnr2830@gmail.com
Annamacharya Institute of Technology and Sciences, Rajampet, Andhra Pradesh

ABSTRACT

The need for Sentiment classification is a critical step in determining how people feel about a product, service, or issue. To solve the sentiment categorization problem, many natural language processing models have been developed. The majority of them, however, have concentrated on binary sentiment categorization. In this study, we tackle the fine-grained sentiment categorization task using BERT, a powerful deep learning model. Experiments show that without complicated architecture, our model outperforms other popular models in this job. In the process, we also demonstrate the utility of transfer learning in natural language processing.

Keywords: *Sentiment Classification, Natural Language Processing, Deep Learning*

1. INTRODUCTION

The Sentiment classification is a type of text classification in which a piece of text is assigned to one of several predetermined sentiment categories. It's a challenge of supervised machine learning. Positive and negative sentiment are the two available classes in binary sentiment categorization. There are five classes in fine-grained sentiment classification (very negative, negative, neutral, positive, and very positive).

The input to the sentiment classification model, like any other machine learning model, must be a fixed-size vector of numbers. As a result, we must convert a text—a sequence of words encoded in ASCII or Unicode—into a fixedsized vector that contains the text's relevant information. For this purpose, many statistical and deep learning NLP models have been proposed. There has been a recent surge in the development of NLP and other deep learning architectures.

While transfer learning (pretraining and finetuning) has become the industry standard in computer vision, NLP has yet to completely embrace the concept. The transfer learning revolution in NLP has begun with neural language models like as word vectors [1], paragraph vectors [2], and GloVe [3]. BERT (Bidirectional Encoder Representations from Transformers) [4], a deep bidirectional language model based on the Transformer architecture [5], was recently published by Google researchers, and it enhanced the state-of-the-art in several popular NLP tasks.

2. RELATED WORK

Sentiment classification is one of the most prominent NLP problems, and as a result, there has been a lot of research and advancement in accurately handling this task. The majority of the algorithms have concentrated on binary sentiment classification, owing to the availability of huge public datasets such as the IMDb movie review dataset [6]. Only a few significant deep learning NLP methods used to sentiment categorization are discussed in this section.

The embedding process converts a text into a fixed-size vector, which is the first stage in sentiment classification. Researchers first faced the difficulty of learning word embeddings since the amount of terms in the lexicon after tokenization and stemming is restricted. Mikolov et al. [1] proposed the first promising language model. They used enormous amounts of unlabeled text to train continuous semantic representations of words that could be fine-tuned for downstream tasks. To learn semantic word embeddings effectively, Pennington et al. [3] employed a co-occurrence matrix and exclusively trained on nonzero entries. For a lower vocabulary size and faster training, Bojanowski et al. [7] split words into character n-grams.

Next, a variable number of word vectors are combined into a single fixed-size document vector. Taking the total or the average is a simple method, but it loses the ordering information of words and hence produces poor results. By exploiting the innate tree structure of natural language phrases, Tai et al. [8] employed recursive neural networks to generate vector representations of sentences. For enhanced interaction between child nodes in recursive networks, Socher et al. [9] presented a tensor-based compositionality function. The Stanford Sentiment Treebank (SST) dataset was also introduced for fine-grained sentiment classification. Tai et al. [10] used several types of long short-term memory (LSTM) networks to classify sentiment, while Kim [11] used convolutional neural networks (CNN).

All of the methods described above are context-free, in the sense that they generate a single word embedding for each word in the lexicon. "Bank" is represented similarly in "bank deposit" and "river bank," for example. Contextual embeddings have been the subject of recent language model research. From a left-to-right and right-to-left LSTM-based language model, Peters et al. [12] retrieved context-sensitive characteristics. To train deep bidirectional representations from unlabeled texts, Devlin et al. [4] presented BERT (Bidirectional Encoder Representations from Transformers), an attention-based Transformer architecture [5]. Because it is not reliant on sequential or recurrent connections, their design not only achieves state-of-the-art outcomes on many NLP tasks, but also allows for a significant degree of parallelism.

3. METHODOLOGY

Sentiment classification takes a natural language text as input and returns a sentiment score of 0, 1, 2, 3, or 4 as a result. From the input sentence to the output score, system comprises three stages. To create a sentiment classifier, we employ a pre-trained BERT model.

A. BERT

BERT (Bidirectional Encoder Representations from Transformers) is an embedding layer that uses both left and right context conditioning in all layers to train deep bidirectional representations from unlabeled texts. It is pretrained using the following objectives from a large unsupervised text corpus:

• Masked word prediction: In this task, 15% of the words in the input sequence are masked out, then the complete sequence is given to a deep bidirectional Transfomer [5] encoder, which then trains the model to predict the masked words.

• Next sentence prediction: BERT takes two sentences A and B as inputs and learns to categorise whether B genuinely follows A or is just a random sentence.

B. PROPOSED METHOD

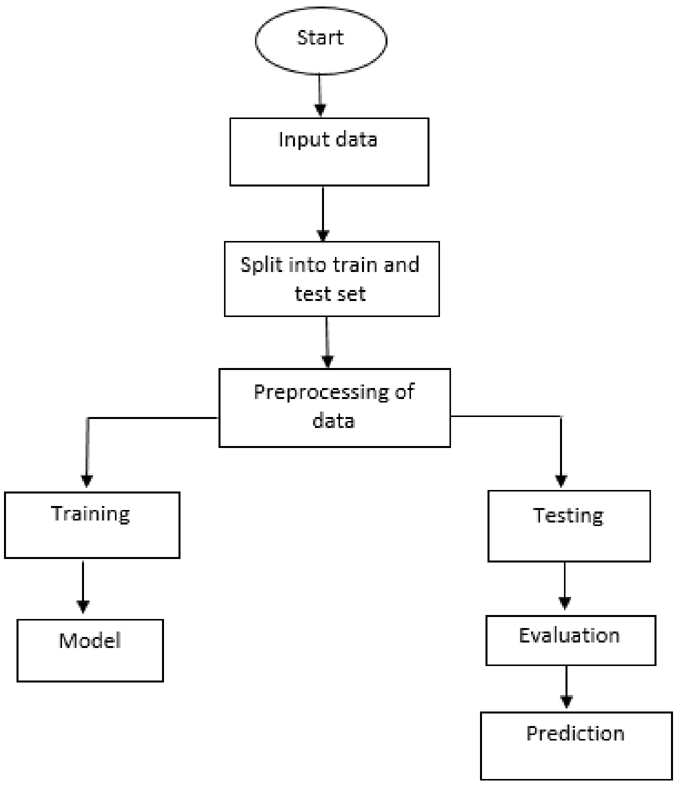


Fig 1: Block diagram model for proposed method

- **Input Format**

BERT requires a specific format for its input token sequence. Every sequence should start with a [CLS] token (classification token), and each sentence should end with a [SEP] token (separation token).

- **Pre-Processing**

Text preprocessing is a technique for cleaning up text data before feeding it to a machine-learning model. Text data comprises a wide range of noise, including emotions, punctuation, and text capitalised differently. Because machines cannot understand language, they want numbers, this is merely the beginning of the issues we will confront. As a result, we must devise a method for converting text to numbers that is both quick and efficient.

Pre-processing in the typical or conventional manner is a time-consuming and user-centric process. Under the hood of normal pre-processing processes, the following stages are carried out:

1. Lower casing the corpus
2. Removing the punctuation
3. Removing the stopwords

4. LITERATURE SURVEY

The usefulness of current lexical resources and linguistic factors for performing SA on Twitter messages and similar microblogging posts was highlighted by Kouloumpis et al. [16].

In comparison to part-of-speech traits and those pertaining to an established sentiment lexicon, the researchers believe that microblogging is more relevant and appropriate. Microblogging elements are unlikely to improve training data, according to the authors.

Another study [17] integrating rule-based classification and supervised learning procedures indicated that hybrid classification was more relevant and yielded promising results. Since the turn of the century, SA has risen to prominence as one of the most explored subjects in natural language processing (NLP) [18,19].

To streamline SA approaches, researchers have been continually evaluating sentences of various sorts. Mudinas et al. proposed a concept level SA system (psenti), which has been found to be more effective than a lexicon-based approach [20]. According to the findings of their study, the hybrid strategy is far more efficient and yields more accurate results.

Tripathy et al. developed four machine learning techniques for sentiment classification in their study [31]. Nave Bayes, Maximum Entropy, Stochastic Gradient Descent, and Support Vector Machine were the methods used. Their research showed that progressive classification can help reach accuracy. IMDB, a popular movie review website, was used to perform their investigation. Utilizing a mix of TF-IDF and count vectorizer techniques, the authors were able to obtain consistent high accuracy using an n-gram approach. To address the issue of words and punctuation marks that, while humanly understandable, lack a formal definition in the English lexicon, the authors created a new list of such words to aid with SA. In their research work, Mitchel et al. modelled sentiment detection. Their research focused on demonstrating the applicability of sentiment recognition as a problem involving sequence tagging [21].

Later research [22] looked at word embedding and automated feature combination using neural networks, which built on their findings. Arun et al. then conducted a SA on tweets about the Indian economy's demonetization.

Their method of investigation [23] consisted of extracting data from Twitter and converting it to text. This text was supposed to be the input dataset. Following the removal of stop words, SA was done in order to determine the polarity of the words and, as a result, the real tweets could be classified as positive or negative.

For SA, a new strategy was proposed to deal with demonetization, its consequences, and the public outcry it sparked. In the study, bigrams, data cleansing, polarity, sentiment scores, and graphical approaches were all applied. Researchers [12] looked analysed a collection of tweets in Spanish to conduct a comparative investigation of several algorithms for SA and topic detection.

Tweets, according to the authors, are difficult to judge because of the absence of context and the unusual shortness of the material involved. The authors of another e-commerce and online product review study [24] noted that public opinion and the opinions of buyers of products sold on online retail platforms will have far-reaching implications in terms of profitability and economic feasibility.

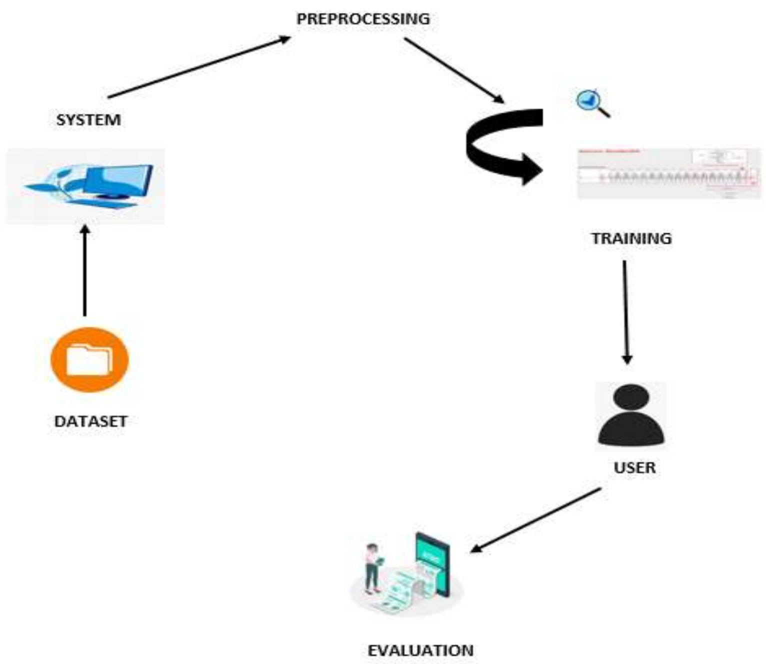
As a result, businesses and entrepreneurs invest in opinion mining and SA to develop strategies for staying ahead of the competition. The authors use the FRBAS R package to implement fuzzy rulebased systems (FRBS) with Mamdani and Takagi Sugeno Kang (TSK) models. The authors next compare these models to alternative categorization strategies, grading them on precision, recall, f-measure, and overall performance.

The authors [24] discovered a substantial dearth of resources to carry out SA in Arabic while investigating the feasibility and efficacy of SA in the context of Arabic tweets. The authors used Jordanian Arabic tweets in their study and used a variety of supervised machine learning algorithms to look at them. According to their findings, SVM classifiers with TF-IDF and bigrams outperformed Nave Bayesian classifiers. The authors gathered over a thousand Arabic tweets written in Jordanian dialect for their experiment.

The two machine learning methods, SVM and NB, were then applied to a set of ngrams with varied weighting schemes and stemming strategies.

When the authors [25] introduced a semisupervised recursive auto-encoding strategy for forecasting sentiment distributions without the need of sentiment lexica or rules involving polarity shifting, it received a lot of attention. The authors suggested the recursive neural tensor network (RNTN) in their study [27] to analyse and distinguish phrases and sentences of various lengths.
An alternate technique of sentiment categorization was presented in another research [28] for analysing sentences of various lengths and terms. While hyper parameter tweaking for word vectors has been the research of choice [29], the authors [30] have chosen deep recursive neural network (DRNN) for sentiment classification assignments. Surprisingly, the authors [32] believe that larger datasets including crowd sourcing and semi-supervised learning have significantly higher promise for improving classification outcomes.

5. ARCHITECTURE



6. CONCLUSION

Sentiment Analysis is now widely regarded as being crucial from a socioeconomic aspect. Understanding sentiment analysis and the approaches that can help achieve accuracy in a range of input forms is crucial for businesses, institutions, and individuals to survive and prosper. The applications and limitations of sentiment categorization in single and cross domains are discussed in this survey article. Subjectivity analysis, negation management, and polarity categorization must all be completed before sentiment analysis can begin.

7. REFERENCES

[1] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *CoRR*, vol. abs/1301.3781, 2013.
[2] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International Conference on Machine Learning*, 2014, pp. 1188–1196.
[3] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543.
[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *NAACLHLT*, 2018.
[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.
[6] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, June 2011, pp. 142–150.
[7] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.
[8] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2011, pp. 151–161.
[9] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2013, pp. 1631–1642.

[10] K. S. Tai, R. Socher, and C. D. Manning, “Improved semantic representations from tree-structured long short-term memory networks,” in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2015, pp. 1556–1566.
[11] Y. Kim, “Convolutional neural networks for sentence classification,” *arXiv preprint arXiv:1408.5882*, 2014.
[12] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, “Deep contextualized word representations,” *arXiv preprint arXiv:1802.05365*, 2018.
[13] D. Chen and C. Manning, “A fast and accurate dependency parser using neural networks,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 740–750.
[14] M. Schuster and K. Nakajima, “Japanese and korean voice search,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 5149–5152.
[15] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *The Journal of Machine Learning Research*, vol. 15, no. 1, pp.1929–1958, 2014.
[16] Kouloumpis, T. Wilson, and J. Moore, ―Twitter sentiment analysis: The good the bad and the omg!‖, In Proceedings of the Fifth International AAAI conference on weblogs and social media, Barcelona, Catalonia, Spain, pp. 538-541, July 2011.
[17] R. Prabowo, and M. Thelwall, ―Sentiment analysis: A combined approach‖, Journal of Informetrics, Vol. 3, No. 2, pp. 143-157, 2009.
[18] B. Liu, ―Sentiment analysis: Mining opinions, sentiments, and emotions‖, Cambridge University Press, June 2015.
[19] K. Ravi, and V. Ravi, ‖A survey on opinion mining and sentiment analysis: tasks, approaches and applications‖, Knowledge-Based Systems, ELSEVIER, Vol. 89, pp. 14-4, November 2015.
[20] Mudinas, D. Zhang, and M. Levene, ―Combining lexicon and learning based approaches for conceptlevel sentiment analysis‖, In Proceedings of the first international workshop on issues of sentiment discovery and opinion mining, Beijing China, pp. 1-8, August 2012, doi:10.1145/2346676.2346681.
[21] M. Mitchell, J. Aguilar, T. Wilson, and B. Van Durme, ―Open domain targeted sentiment‖, In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, Washington, USA, pp. 1643-1654, October 2013.
[22] M. Zhang, Y. Zhang, and D.T. Vo, ―Neural networks for open domain targeted sentiment‖, In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 612-621, September 2015.
[23] K. Arun, A. Srinagesh, and M. Ramesh, ‖Twitter sentiment analysis on demonetization tweets in India using R language‖, International Journal of Computer Engineering In Research Trends, Vol. 4, No. 6, pp. 252-258, June 2017.
[24] S.M. Basha, Y. Zhenning, D.S. Rajput, N. Iyengar, and D.R. Caytiles, ―Weighted fuzzy rule based sentiment prediction analysis on tweets‖, International Journal of Grid and Distributed Computing, Vol. 10, No. 6, pp. 41-54, June 2017.
[25] K.M. Alomari, H.M. ElSherif, and K. Shaalan, ―Arabic tweets sentimental analysis using machine learning‖, In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, pp. 602-610, Springer, June 2017, doi:10.1007/978-3-319-60042-0_66.
[26] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, & C. D. Manning, ―Semi-supervised recursive autoencoders for predicting sentiment distributions‖, In Proceedings of the conference on empirical methods in natural language processing, pp. 151– 161, , July 2011.
[27] R. Socher, A. Perelygin, J. Wu, J. Chuang, C.D. Manning, A.Y. Ng, and C. Potts, ―Recursive deep models for semantic compositionality over a sentiment treebank‖, In Proceedings of the 2013 conference on empirical methods in natural language processing, Seattle, Washington, USA, pp. 16311642, October 2013.
[28] . N. Kalchbrenner, E. Grefenstette, and P. Blunsom, 2014. ―A convolutional neural network for modeling sentences‖, In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, Maryland, pp. 655–665, June 2014, doi: /10.3115/v1/P14-1062.
[29] Y. Kim, ‖Convolutional neural networks for sentence classification.‖, In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) , Doha, Qatar, August 2014, doi:/10.3115/v1/D14-1181.
[30] O. Irsoy, and C. Cardie, ―Opinion mining with deep recurrent neural networks‖, In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 720–728, October 2014.
[31] S. ChandraKala, and C. Sindhu, ―Opinion mining and sentiment classification: A survey‖, ICTACT journal on soft computing, vol. 3, No. 1, pp. 420-425, October 2012.
[32] L. Qu, G. Ifrim, and G. Weikum, , ― The bag-ofopinions method for review rating prediction from sparse text patterns‖, In Proceedings of the 23rd international conference on computational linguistics, Coling 2010, Beijing ,Association for Computational Linguistics, pp. 913-921, August 2010.

