

Sentiment Analysis Techniques for Deep Learning Classification and Comparison

Shuaiyu Chen^{1,a,*}

¹*School of Science and Technology, Beijing Normal University - Hong Kong Baptist University United International College, Zhuhai, Guangdong, China*
a. s230018009@mail.uic.edu.cn
**corresponding author*

Abstract: Sentiment analysis, a critical tool in natural language processing (NLP), has seen significant advancements with the rise of machine learning and deep learning techniques. This paper explores the evolution of sentiment analysis from traditional models, such as CNN and RNN, to more sophisticated approaches like LSTM, LSTM-CNN, and BERT. Using the IMDb dataset for binary classification of sentiment, we utilize metrics like F-Score, Precision, Accuracy, and Recall to compare the way these models perform. Despite the progress made, sentiment analysis still faces challenges, particularly in handling multiple languages, incorporating multimedia content, and addressing privacy and ethical concerns. This paper also discusses future directions for sentiment analysis, including the development of models for low-resource languages, real-time sentiment analysis, and the ethical implications of data handling. Overall, the study highlights the potential of sentiment analysis and the opportunities for improvement through continued research and innovation.

Keywords: Sentiment analysis, deep learning, LSTM, natural language processing.

1. Introduction

As social media and other online forums have grown in popularity, more people are using these platforms to express their feelings. To facilitate the understanding of emotions conveyed through media such as text and video, emotion analysis comes into being. Sentiment analysis (SA) is the technique of deriving information and subjective emotions from text resources through the application of specific procedures. Natural language processing (NLP), computational linguistics, and text mining are some of these techniques. Additionally, the emotional polarity—positive, negative, or neutral—is evaluated. Artificial intelligence and computer science sentiment analysis can assist the political, entertainment, and financial sectors in their analysis, forecasting, and management. There are many applications for sentiment analysis in society, politics, and business. It is undoubtedly useful for examining client comments, understanding public sentiment, formulating corresponding strategies and measures, and predicting election results and public support.

In earlier studies, sentiment analysis mainly focuses on the analysis of text, with an emphasis on determining the emotional polarity of the words, phrases and sentences. This process heavily relies on explicit emotive words (such as "hate" and "like"). Using machine learning algorithms and techniques is common in sentiment analysis, particularly for tasks involving binary classification

and the identification of positive or negative emotions. Supervised, unsupervised, and semi-supervised learning techniques are further classifications for machine learning algorithms [1]. Several machine learning techniques, including Naive Bayes, Support Vector Machines (SVM), K-Means, and Maximum Entropy, are utilized, depending on the training dataset. Nevertheless, previous research had its drawbacks, such as limited data sources and insufficient diversity and scale. The scope of text languages and fields is also narrow, preventing specialized analysis. In addition, it is difficult to deal with special context and implicit emotions in the early stage. But as the technology evolved, researchers began to use more complex and diverse methods for sentiment analysis, and the technique gradually improved.

Deep learning models like long short-term memory (LSTM), recurrent neural networks (RNN), and convolutional neural networks (CNN) have been more and more popular in sentiment analysis in recent years. Because sentiment analysis has significant implications for both supervised and unsupervised learning, deep learning is employed to research it [2]. The three deep learning techniques that were previously discussed are all well-developed and important deep learning models. In order to maximize the processing capability of the model, they are all hierarchical in structure and feature backpropagation and gradient descent methods. The main distinction between both is that RNNS is primarily used to process sequence data, whereas CNNS is used to analyze picture data. A unique type of RNN called an LSTM was created especially to address long-term reliance issues in lengthy sequences. Three different models offer distinct approaches to sentiment analysis, which are discussed in detail in subsequent sections. In addition, sentiment analysis is using more sophisticated Transformer-based models, like BERT and GPT, more frequently. It has been discovered that these models can detect personality from text and greatly increase analysis accuracy [3].

This article begins by introducing the background and significance of sentiment analysis in Section 1. In the following section, it will also list the deep learning methods and large models used in sentiment analysis, and compare and analyze the advantages and disadvantages of different methods through the conclusions drawn from the experiment in Section 3. In the last part, the existing research results of sentiment analysis are summarized and the future research direction in this field is suggested.

2. Methods

2.1. Convolutional Neural Network

Deep learning models called CNN are made to handle data with lattice topologies. CNNs consist of multiple convolutional layers, which are effective in extracting local features in NLP tasks by applying convolution operations through linear filters [4].

In previous studies, CNNs have been widely used to extract local features, especially when there is a large amount of space or content to be extracted. For instance, Liao et al. [4] processed and analyzed texts using single channel convolutional neural networks, effectively extracting features before beginning the classification of sentences. This is a basic and simple CNN model. Johnson and Zhang [5] proposed a new model called seq-CNN based on traditional CNN, which can effectively capture the features in sequence data. To comprehend the information in the text, one must be able to recognize local patterns in the sequence. In their investigation, Kalchbrenne et al. [5] also employed a novel CNN technique known as DCNN to enhance the modeling capacity of abstract or complicated data by delving deeper into the network structure. Cheng et al. [6] combined a Multi-Channel CNN with a Bidirectional GRU to enhance feature extraction from text. The Multi-Channel CNN offers several advantages: each convolutional layer operates independently, and the channels perform parallel operations. Additionally, after the pooling process, a fusion

operation takes place, where the features between channels are weighted and concatenated, enabling classification through fully connected layers. This approach differs from traditional CNN models. Various other CNN methods are also employed in sentiment analysis, including charCNN, Ada-CNN, and more [5]. These advanced models demonstrate the versatility of CNN architectures in handling different types of data and tasks.

2.2. Recurrent neural network

Neural networks that are optimized for natural language processing are known as sequential data neural networks, or RNNs. Text, audio, and time series sequences can all be handled sequentially. Unlike other neural networks, RNNs have loop connections and allow this information to be retained in the network. Kurniasari and Setyanto [7] carried out sentiment classification by using RNN, which is different from feedforward neural networks in that its maximum sequence has the same time step. A hidden state vector (h_t) was introduced to encapsulate and summarize all the information of the previous time step.

$$h_t = \sigma(w^H h_{t-1} + w^x x_t) \quad (1)$$

where x is the time step's input vector and w is the weight matrix. Subtle shifts in textual sentiment can be recorded by analyzing time series and comprehending the text's context. The study found that RNN improved the accuracy compared to machine learning, with an accuracy of 91.98%, and paid attention to the overfitting of the model in the test.

RNNs are also characterized by their ability to handle longer input sequences, so that text of different lengths can be processed and analyzed more accurately. Following the RNN layer's processing of the sequence data, the fully connected layer organizes the RNN's output results into emotion categories before sending the results to the output layer. However, the traditional RNN model has problems such as gradient disappearance and gradient explosion, so many variant models of RNN have been proposed to solve these problems and optimize.

An RNN variant called the Gated Recurrent Unit (GRU) was created to solve the problem of disappearing gradients, which conventional RNNs frequently run into when processing lengthy sequences. To address this issue, GRU implements a gating mechanism that enhances training efficiency and model performance. Ali et al. [8] constructed a Bi-GRU sentiment analysis model, which captures and understands contextual information to a greater extent. This model can detect subtleties in expression, forecast the sentiment polarity of particular words, and model the aspect-based sentiment more precisely. By extracting the feature vector from the output layer and utilizing the label information in each segmentation, the two GRU layers enhance the model's performance.

Like GRU, LSTM is a special sort of RNN designed to solve the problems with gradient disappearance and explosion that come with traditional RNNs. The LSTM, which is more sophisticated in structure than the GRU, contains three gates: input, forget, and output. This makes it extremely good at handling complex long-term dependencies. LSTM plays a very important role in emotion analysis, and its cellular state and gating mechanism can remember and forget this information, helping models better understand complex emotions, such as sarcasm and puns. Compared with other sequence learning methods, the relative insensitivity of gap length is its advantage[9]. Based on bidirectional CNN and RNN, a sentiment analysis model known as ABCDM was presented by Basiri et al. [5] in order to extract context from the past and future. Two methods for identifying and categorizing the sentiment polarity of lengthy comments and brief tweets are convolution and pooling. The output of the LSTM and GRU branches was fed into the attention layer to enhance the interpretability of the semantics. When compared to some RNN and CNN models already in use, the sentiment classification accuracy of ABCDM was likewise higher.

2.3. Transformer

Google unveiled Bidirectional Encoder Representations from Transformers (BERT) in 2018. BERT is a deep learning model for natural language processing that takes into account context and the word itself to produce more accurate results. BERT makes use of the Transformer encoder architecture, which is stacked with several encoder layers and has a self-attention mechanism. The classification of emotions in text can be performed by fine-tuning BERT such as SMALL BERT. Although many variants of BERT are newly developed, many of them are already being applied to sentiment analysis and other tasks [10]. In Souza and Filho's research [10], Two basic models, m-BETR and BERTimbau were used and pre-trained and fine-tuned. By adjusting the weights of BERT to adapt to different sentiment analysis tasks, it was finally found that BERTimbau is the best variant in Brazilian Portuguese text sentiment analysis of the two BERT variants after training and fine-tuning. BERT's Transformer can help her understand emotional information in long and difficult sentences. Compared with CNN and RNN, BERT can better deal with complex emotions. In the pre-training stage, BERT uses the multi-task learning method to better understand the relationship between sentences and efficiently classify emotions.

GPT is also a widely used variant of Transformer that uses monomial autoregressive patterns for sentiment analysis. Zhan et al. [11] used GPT-3 for text processing. After preprocessing the data, the GPT-3 model performed sentiment analysis through fine-tuning technology, and was trained and evaluated. The research results reflected the great application potential and high accuracy of GPT in this task and there was still room for optimization and improvement.

3. Experiment

3.1. Dataset overview

IMDb is an internet database that offers details about television shows and films. With millions of reviews, it's one of the biggest movie and TV databases worldwide. The 50,000 movie reviews in the dataset included in this analysis, which was obtained from Haque [12], are all classified as either favorable or negative. 20% of the training samples were utilized for validation after the dataset was split into training and test sets at a ratio of 7:3. Zero-padding is used to ensure that reviews are of the same length for shorter reviews to facilitate training.

3.2. Process of experiment

Haque[12] trained the CNN model architecture for 8 epochs with a batch size of 128 because the loss would not be decreased in the next training phase. Similar to CNN, the LSTM model is trained across five epochs with a batch size of 128. In order to avoid overfitting, the loss function is then minimized using the Adam optimizer and the Dropout approach. With more convolutional and pooling layers than LSTM, the LSTM-CNN model combines the two and employs the same batch processing for training across six epochs. Cen et al. [13] used the same dataset and sample size to perform sentiment analysis classification using an RNN model, whereas Domadula and Sayyaparaju [14] predicted the sentiment of IMDb reviews using the BERT model. The two output nodes of BERT represent the probability of two sentiment classes, and the model adjusts the weights through learning to improve the prediction accuracy. Finally, the probability distribution of possible emotion labels (0 or 1) is output and the emotion is predicted.

3.3. Experiment result

The IMDb dataset is utilized for sentiment prediction and binary classification using the aforementioned four models. A comparison of several metrics, including accuracy, recall, specificity, precision, and F-score, is used to assess these models' performance.

Table 1: Performance comparisons of respective methods.

Evaluation index	Accuracy	Recall	Specificity	Precision	F-Score
CNN [12]	0.90	0.95	0.84	0.87	0.91
RNN [13]	0.68	-	-	-	-
LSTM [12]	0.88	0.82	0.90	0.90	0.86
LSTM-CNN [12]	0.89	0.90	0.87	0.87	0.88
BERT [14]	0.90	0.92	-	0.88	0.88

The table illustrates that the CNN and BERT models achieve the highest accuracy at 0.90, with CNN excelling in Recall (0.95) and BERT demonstrating balanced performance across multiple metrics. LSTM and LSTM-CNN models also perform well, particularly in Specificity and Precision. The RNN model, however, has the lowest reported accuracy (0.68) with missing data for other evaluation metrics.

3.4. Discussion

The table shows that CNN outperforms LSTM, RNN, and LSTM-CNN in terms of performance because phrase syntax can have an impact on semantics and the assessment of emotional polarity. Compared with the other two, CNN still has strong performance in processing sequence data. Compared with other models, although the recall rate of BERT is not as good as CNN, it is also ranked above other models, and the accuracy is also at the top, which shows that BERT has successfully identified a lot of true positive instances and can accurately classify emotions. Therefore, CNN and BERT can be said to be better performing deep learning models for sentiment analysis.

4. Conclusion

4.1. Summary of experiments

The experiment may summarise the various benefits and drawbacks of various models in handling sentiment analysis jobs by utilising the same data set and comparing the performance metrics.

CNN is capable of processing the input text for sentiment analysis as a one-dimensional image and using convolutional filters to extract local dependencies between words and phrases. Methods such as CNN have shown very good performance in identifying and processing text information, and have excellent performance in judging whether the sentiment polarity is positive, negative or neutral. At the same time, using different learning rates for training in multiple periods can make the performance of CNN model tend to be stable or steadily improve. CNN does, however, have some restrictions. CNN is frequently employed for the extraction of local features; but, in comparison to other models, it lacks context awareness, which can lead to errors in the analysis and evaluation of overall sentiment in lengthy texts. Second, in sentiment analysis, we need to be very aware of overfitting and hyperparameter adjustments.

LSTM improves accuracy because it can capture the sequential dependence of sequences and analyze text through the entire context. LSTM's ability to retain information makes it more accurate in analyzing longer texts and more effective in determining the overall mood. Although the LSTM

is very powerful, the training time is slower than other models such as CNN. Maintaining cell state and gating mechanisms increases the burden on memory, which becomes difficult when dealing with long sequences of text and large models.

Because of its outstanding performance in terms of accuracy, precision, recall, and F1-score--all of which have produced extremely positive results in sentiment analysis--BERT is one of the fundamental models for sentiment analysis. However, there is a significant energy consumption and expensive computation and time expenditures, which compromise sustainability. Furthermore, biases in the training set may be amplified by BERT, leading to biased final results. We call this propagation of bias.

4.2. Future work

Sentiment analysis still has a lot to explore and need to be improved in the future. Developing models to accommodate many languages is one area of work, with a focus on low-resource languages to increase the applicability of sentiment analysis. At the moment, text information from forums and comments is the main focus of sentiment analysis. Future research could expand this by incorporating sentiment analysis for images, voice, and even video content. Additionally, sentiment analysis raises important privacy and ethical concerns. Future efforts should prioritize ensuring the security of input data and addressing ethical issues, particularly in guaranteeing that the output does not negatively impact individuals. Moreover, there is ample room for advancing personalized analysis, sentiment intensity analysis, and real-time sentiment analysis.

Overall, even though sentiment analysis has made significant strides in the field of natural language processing, more technological innovation and continuous experimental study can improve sentiment analysis's accuracy and usefulness.

References

- [1] Ahmad, M., Aftab, S., Muhammad, S. S., & Ahmad, S. (2017). Machine learning techniques for sentiment analysis: A review. *Int. J. Multidiscip. Sci. Eng.*, 8(3), 27.
- [2] Ain, Q. T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., & Rehman, A. (2017). Sentiment analysis using deep learning techniques: a review. *International Journal of Advanced Computer Science and Applications*, 8(6).
- [3] Tan, K. L., Lee, C. P., Anbananthen, K. S. M., & Lim, K. M. (2022). RoBERTa-LSTM: a hybrid model for sentiment analysis with transformer and recurrent neural network. *IEEE Access*, 10, 21517-21525.
- [4] Liao, S., Wang, J., Yu, R., Sato, K., & Cheng, Z. (2017). CNN for situations understanding based on sentiment analysis of twitter data. *Procedia computer science*, 111, 376-381.
- [5] Basiri, M. E., Nemati, S., Abdar, M., Cambria, E., & Acharya, U. R. (2021). ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis. *Future Generation Computer Systems*, 115, 279-294.
- [6] Cheng, Y., Sun, H., Chen, H., Li, M., Cai, Y., Cai, Z., & Huang, J. (2021). Sentiment analysis using multi-head attention capsules with multi-channel CNN and bidirectional GRU. *IEEE Access*, 9, 60383-60395.
- [7] Kurniasari, L., & Setyanto, A. (2020, February). Sentiment analysis using recurrent neural network. In *Journal of Physics: Conference Series* (Vol. 1471, No. 1, p. 012018). IOP Publishing.
- [8] Ali, W., Yang, Y., Qiu, X., Ke, Y., & Wang, Y. (2021). Aspect-level sentiment analysis based on bidirectional-GRU in SIoT. *IEEE Access*, 9, 69938-69950.
- [9] Murthy, G. S. N., Allu, S. R., Andhavarapu, B., Bagadi, M., & Belusonti, M. (2020). Text based sentiment analysis using LSTM. *Int. J. Eng. Res. Tech. Res.*, 9(05).
- [10] Souza, F. D., & Filho, J. B. D. O. E. S. (2022, March). BERT for sentiment analysis: pre-trained and fine-tuned alternatives. In *International Conference on Computational Processing of the Portuguese Language* (pp. 209-218). Cham: Springer International Publishing.
- [11] Zhan, T., Shi, C., Shi, Y., Li, H., & Lin, Y. (2024). Optimization Techniques for Sentiment Analysis Based on LLM (GPT-3). *arXiv preprint arXiv:2405.09770*.
- [12] Haque, M. R., Lima, S. A., & Mishu, S. Z. (2019, December). Performance analysis of different neural networks for sentiment analysis on IMDb movie reviews. In *2019 3rd International conference on electrical, computer & telecommunication engineering (ICECTE)* (pp. 161-164). IEEE.

- [13] Cen, P., Zhang, K., & Zheng, D. (2020). *Sentiment analysis using deep learning approach*. *J. Artif. Intell.*, 2(1), 17-27.
- [14] Domadula, P. S. S. V., & Sayyaparaju, S. S. (2023). *Sentiment analysis of IMDB movie reviews: a comparative study of Lexicon based approach and BERT Neural Network model*.