DOI: 10.54254/2755-2721/75/20240543

Balancing performance and efficiency: A CNN-LSTM hybrid model for sentiment analysis

Yizhen Zeng

China Agricultural University, Beijing, 100083, China

zengyizhen2021@163.com

Abstract. In the era of rapid e-commerce and social networks, the recognition of textual emotion has garnered escalating interest in recent years. This area of study proves pivotal for applications such as content recommendation and human-robot interaction. Prominent models such as CNN, LSTM, and Roberta have been used in natural language processing to decipher the subtleties of textual emotion. Despite the commendable accuracy achieved by Roberta, its extensive model parameters come at the cost of sluggish training speeds. To address this issue, we propose a novel hybrid approach that integrates Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. We employ this hybrid model to address the emotion classification problem using a dataset named "Emotion", and we compare its performance with that of individual models like Roberta, CNN, and LSTM. The new hybrid model exhibits a more compact architecture while maintaining a relatively high accuracy rate, boasting an impressive 91%. This hybrid architecture can provide a good compromise between efficiency and performance. This research contributes to the ongoing exploration of efficient and accurate models for textual emotion recognition, emphasizing the significance of balancing model complexity and training speed in practical applications.

Keywords: Emotion Classification, Machine Learning, Natural Language Processing, CNN, LSTM.

1. Introduction

Sentiment analysis on social network data is a burgeoning research area in natural language processing [1]. Even with such advances, there are still issues to be resolved, including as improving model reliability, cutting down on processing time, and adjusting methods for certain data kinds and domains [2]. In the field of natural language processing, prominent models such as Roberta, CNN, and LSTM have been widely employed to unravel the complexities of textual emotion [3-4].

Notably, researchers like K. L. Tan have combined RoBERTa with LSTM, achieving promising results on multiple sentiment classification datasets [5]. However, there is a notable absence of consideration for balancing model size and performance. There are researchers proposed a hybrid model, combining CNN and LSTM, lacks a comparative evaluation involving the Roberta model [6].

RoBERTa, a pre-trained language model in the field of natural language processing based on BERT, has exhibited excellent performance across many tasks, including text categorization, named entity identification, and question answering [7]. Despite the commendable accuracy achieved by Roberta, its extensive model parameters lead to slow training speeds. In response, this study introduces an

DOI: 10.54254/2755-2721/75/20240543

innovative hybrid approach integrating Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM). This hybrid model is compared with individual CNN, LSTM, and Roberta models. The study emphasizes the importance of balancing model complexity and training speed in practical applications, providing valuable insights for research in sentiment classification, machine learning, and natural language processing.

2. Literature Review

Several recent studies have explored innovative approaches to emotion classification, often integrating LSTM and CNN models to leverage their respective strengths. For instance, a hybrid model integrating CNN and Bi-LSTM architectures was proposed by Jang et al, augmented with an attention mechanism, aiming to enhance performance by capturing both local and global dependencies within the text [8].

Similarly, Umer et al. presented a CNN-LSTM hybrid model for emotion classification across diverse datasets, including hate speech, Twitter reviews, and women's e-commerce clothing data [9]. Their experiments revealed promising results, with the CNN-LSTM model achieving notably higher accuracy, particularly attaining 82% accuracy on the Twitter US dataset.

However, despite these advancements, a critical gap remains in the literature concerning the comparative evaluation of these hybrid models against well-performed architectures such as RoBERTa. Tan et al. suggested a novel hybrid sentiment analysis model that integrates LSTM with RoBERTa, showcasing encouraging results in emotion classification tasks [10]. Nevertheless, their study lacked a thorough investigation into the balance between model complexity and performance.

In addressing this gap, our study aims to contribute by conducting a comparative evaluation of a novel hybrid approach, integrating CNN and LSTM architectures, against RoBERTa and individual LSTM and CNN models. Through this analysis, we seek to provide insights into the trade-offs between model complexity, performance, and computational efficiency, thereby advancing research in emotion classification and natural language processing.

3. Methodology

3.1. Datasets Preprocessing

The dataset utilized in this study, named "Emotion," consists of 16,000 instances of English Twitter messages. Each instance is associated with a corresponding label denoting the emotional sentiment expressed in the message. "Emotion" is a diverse dataset curated from English Twitter messages, encompassing a range of emotions, including anger, fear, joy, love, sadness, and surprise.

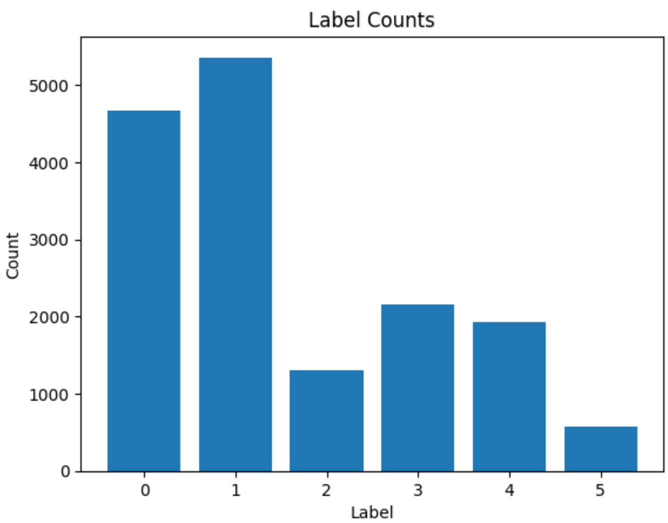

Figure 1. The Emotion dataset’s distribution

DOI: 10.54254/2755-2721/75/20240543

The datasets was chosen due to its relevance in capturing real-world emotional expressions within a social media context. The inclusion of multiple emotions allows for a nuanced exploration of sentiment analysis, providing a robust foundation for evaluating the proposed hybrid model's performance across diverse emotional states.

The text data underwent tokenization using the RoBERTa tokenizer initialized from the 'roberta-base' pre-trained model. Unlike traditional tokenizers, transformers like RoBERTa utilize a fixed vocabulary predefined during pre-training [11]. Therefore, the concept of the number of unique words is not directly applicable in this context. Sequences were then truncated or padded to a maximum length of 128 tokens to align with the model's input requirements. The categorical labels representing different emotional sentiments were encoded using the scikit-learn LabelEncoder, mapping them to numerical values for model training. This step ensures compatibility with machine learning algorithms that require numerical inputs.

3.2. Hybrid CNN-LSTM Sentiment Analysis Model

The model architecture comprises the following components, and the architecture is shown in Figure 2:

- Input Layer: Accepts sequences of a maximum length of 128 tokens.
- CNN layer: Utilizes four parallel Conv1D layers with a kernel size of 5 and ReLU activation.
- LSTM layer: Employs Bidirectional LSTM layers with a specified number of units, returning sequences at the first two layers and the final state at the last layer.
- Concatenation Layer: Merges the outputs from the CNN and LSTM components.
- Dense Layers: Two fully connected dense layers with ReLU activation.
- Output Layer: Produces the final predictions using a softmax activation function.

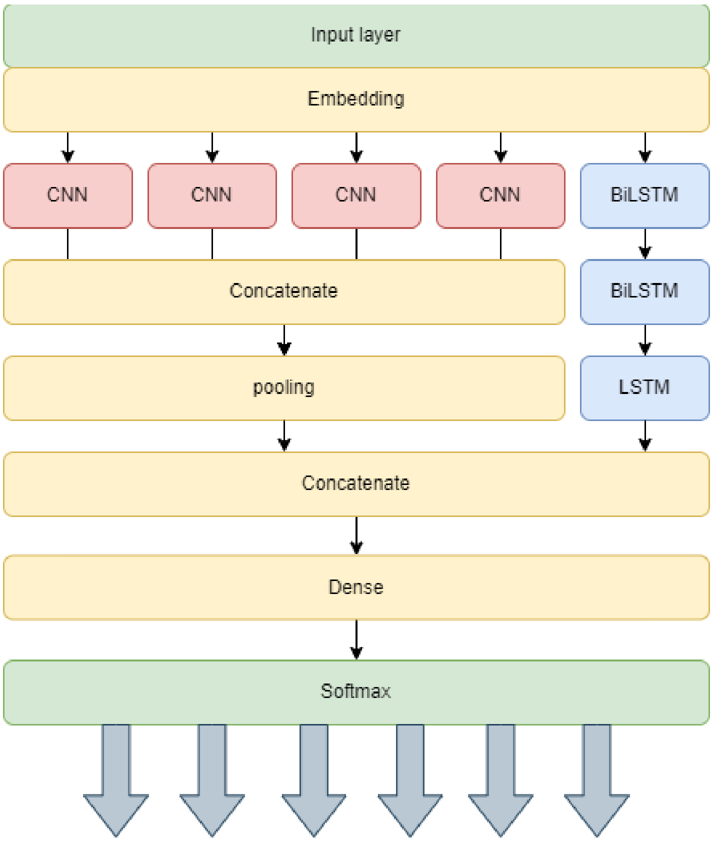

Figure 2. Hybrid CNN-LSTM Sentiment Analysis Model

DOI: 10.54254/2755-2721/75/20240543

CNN is widely employed for extracting local features, effectively capturing local structural information in input sequential data through convolutional operations. Simultaneously, for processing sequential data, especially in natural language text, Bidirectional Recurrent Neural Networks (Bi-RNNs) offer another powerful tool. As indicated by Alex Graves et al. in their 2005 study, the core principle of Bi-RNNs is to present each training sequence forward and backward to two separate recurrent neural networks that are both connected to the same output layer [12].

Such a structure allows the model to capture information from both directions when dealing with sequential data, aiding in a more comprehensive understanding of patterns within the sequence. Taking sentiment analysis as an example, this model architecture can synthesize local features and dependencies within the sequence, thereby better capturing subtle nuances in emotional expressions.

By combining CNN and Bidirectional LSTM, the resulting model structure not only efficiently extracts local features but also captures global information within the sequence, exhibiting outstanding performance in tasks such as sentiment analysis. This integrated architecture provides the model with enhanced representational capabilities, enabling a more accurate understanding and classification of sentiment tendencies within textual data.

3.3. Evaluate

3.3.1. Model Parameters. Before assessing the performance of different models, we initially introduce the key parameters for each model. For the CNN, LSTM, CNN-LSTM, and RoBERTa models, we set the following crucial parameters. It is noteworthy that we employed the pre-trained RoBERTa model for training:

- CNN Model Parameters:embedding_dim, num_filters
- LSTM Model Parameters: embedding_dim, lstm_units
- CNN-LSTM Model Parameters: embedding_dim, num_filters, lstm_units
- RoBERTa Model Parameters: pre-trained_model, learning_rate

3.3.2. Model Performance Evaluation. Comprehensive assessments of each model's performance on the emotion dataset were conducted using a classification report. The following provides mathematical interpretations and key evaluation metrics:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

These evaluation metrics provide us with a thorough understanding of the performance of each model in sentiment analysis tasks, assisting us in comparing their strengths and weaknesses. Through detailed performance assessments, we can acquire a full understanding of each model's performance in a various aspects, establishing the basis for future discussion of the results.

DOI: 10.54254/2755-2721/75/20240543

4. Results and Discussion

We commence our analysis by scrutinizing the performance of CNN, LSTM, CNN-LSTM, and RoBERTa models on the emotion datasets. In Table 1, we present a comprehensive overview of their performance metrics alongside model sizes.

Table 1. Performance Metrics and Model Size Comparison between CNN, LSTM, CNN-LSTM, and RoBERTa models.

	CNN-LSTM	CNN	LSTM	RoBERTa
Accuracy	0.91	0.87	0.90	0.93
Macro Precision	0.87	0.79	0.88	0.89
Macro Recall	0.88	0.79	0.87	0.90
Macro F1-Score	0.87	0.79	0.97	0.89
Weighted Precision	0.91	0.88	0.90	0.93
Weighted Recall	0.91	0.87	0.90	0.93
Weighted F1-Score	0.91	0.87	0.90	0.93
Size(MB)	5.97	0.22	5.84	10.38

Considering the imbalanced nature of the emotion dataset, we place greater emphasis on weighted averages to comprehensively assess the performance of each model across different categories.

With an accuracy of 0.93, Table 1 demonstrates that RoBERTa outperforms all other models. CNN-LSTM comes in second place with an accuracy of 0.91.The performance of CNN-LSTM exceeds individual CNN and LSTM. Notably, the CNN-LSTM model demonstrates a smaller model size compared to RoBERTa, which could be advantageous in resource-constrained environments.

Table 2. Detailed Classification Reports

	precision	Recall	f1-score	Support
CNN-LSTM				
0	0.93	0.95	0.94	946
1	0.96	0.91	0.94	1021
2	0.80	0.89	0.84	296
3	0.89	0.91	0.90	427
4	0.88	0.85	0.86	397
5	0.76	0.74	0.75	113
accuracy			0.91	3200
macro avg	0.87	0.88	0.87	3200
weighted avg	0.91	0.91	0.91	3200
CNN				
0	0.93	0.93	0.93	946
1	0.96	0.90	0.93	1021
2	0.64	0.80	0.71	296
3	0.84	0.89	0.87	427
4	0.82	0.82	0.82	397
5	0.56	0.42	0.48	113
accuracy			0.87	3200
macro avg	0.79	0.79	0.79	3200

DOI: 10.54254/2755-2721/75/20240543

Table 2. (continued).

weighted avg	0.88	0.87	0.87	3200
LSTM				
0	0.95	0.95	0.95	946
1	0.94	0.92	0.93	1021
2	0.80	0.81	0.81	296
3	0.84	0.92	0.88	427
4	0.86	0.86	0.86	397
5	0.85	0.76	0.80	113
accuracy			0.90	3200
macro avg	0.88	0.87	0.87	3200
weighted avg	0.90	0.90	0.90	3200
RoBERTa				
0	0.98	0.95	0.97	946
1	0.96	0.94	0.95	1021
2	0.85	0.91	0.88	296
3	0.89	0.96	0.93	427
4	0.88	0.87	0.88	397
5	0.76	0.75	0.76	113
accuracy			0.93	3200
macro avg	0.89	0.90	0.89	3200
weighted avg	0.93	0.93	0.93	3200

Table 2 presents detailed classification reports for each model, delineating precision, recall, and F1-score for each sentiment category. Table 2 provides comprehensive classification reports for every model, including the F1-score, recall, and precision for every sentiment category. These reports offer a granular understanding of each model's performance across different sentiment categories, elucidating their strengths and weaknesses in sentiment classification tasks. Although the CNN-LSTM hybrid model has higher accuracy than CNN and LSTM in terms of overall accuracy, it cannot achieve the best performance in each category and needs to be improved.

Upon a holistic consideration of model size and performance, CNN-LSTM emerges as the preferred choice, showcasing a harmonious equilibrium. Despite RoBERTa's slight accuracy advantage, its substantially larger model size may exert notable pressure on resource consumption. Thus, the CNN-LSTM model stands out as a compelling option, demonstrating high-performance efficiency in sentiment analysis tasks.

5. Conclusion

In conclusion, the existence of large-scale models like BERT and RoBERTa has significantly advanced sentiment analysis in natural language processing. While these models have showcased remarkable accuracy, their extensive parameters result in challenges, especially in resource-constrained environments.

In response to this challenge, our study introduces a novel CNN-LSTM hybrid model for textual emotion recognition. By combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, our hybrid approach achieves a commendable accuracy rate of 91%, presenting a promising alternative to resource-intensive models like RoBERTa. The compact architecture of the CNN-LSTM hybrid model strikes a well-balanced compromise between efficiency and performance, addressing the trade-off between model complexity and training speed.

DOI: 10.54254/2755-2721/75/20240543

Our finding assists to the ongoing effort to develop efficient and accurate sentiment analysis methods. The comparative evaluation of the CNN-LSTM hybrid model against individual CNN, LSTM, and RoBERTa models highlights its efficacy in maintaining high accuracy while mitigating the challenges associated with extensive model parameters. This study underscores the importance of considering model size alongside performance in practical applications, providing useful insights for researchers as well as professionals in sentiment classification and machine learning.

As we move forward, further research can explore the generalizability of the proposed hybrid model across diverse datasets and domains. Additionally, investigating the interpretability of the CNN-LSTM model can enhance our understanding of its decision-making process. The ongoing pursuit of efficient and accurate models in textual emotion recognition remains vital for advancing applications in content recommendation, human-robot interaction, and other areas within the dynamic landscape of e-commerce and social networks.

References

[1] Liu, X., Shi, T., Zhou, G., Liu, M., Yin, Z., Yin, L., & Zheng, W. (2023). Emotion classification for short texts: an improved multi-label method. Humanities and Social Sciences Communications, 10, 1-9.

[2] Bucila, C., Caruana, R., & Niculescu-Mizil, A. (2006). Model compression. Knowledge Discovery and Data Mining.

[3] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. ArXiv, abs/1907.11692.

[4] Luan, Y., & Lin, S. (2019). Research on Text Classification Based on CNN and LSTM. 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 352-355.

[5] Tan, K.L., Lee, C.P., Anbananthen, K.S., & Lim, K.M. (2022). RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network. IEEE Access, 10, 21517-21525.

[6] Dang, C.N., García, M.N., & Prieta, F.D. (2021). Hybrid Deep Learning Models for Sentiment Analysis. Complex., 2021, 9986920:1-9986920:16.

[7] Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. North American Chapter of the Association for Computational Linguistics.

[8] Jang, B., Kim, M., Harerimana, G., Kang, S., & Kim, J.W. (2020). Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism. Applied Sciences.

[9] Umer, M.F., Ashraf, I., Mehmood, A., Kumari, S., Ullah, S., & Choi, G.S. (2020). Sentiment analysis of tweets using a unified convolutional neural network ‐ long short ‐ term memory network model. Computational Intelligence, 37, 409 - 434.

[10] Tan, K.L., Lee, C.P., Anbananthen, K.S., & Lim, K.M. (2022). RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network. IEEE Access, 10, 21517-21525.

[11] Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. Neural Information Processing Systems.

[12] Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM networks. Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005, 4, 2047-2052 vol. 4.