Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD26888

Exploring CNN, LSTM and CNN-LSTM for Sentiment Analysis in Film Reviews

Mengyu Li

School of Computer Science, University of Birmingham, Birmingham, United Kingdom
mxl1099@student.bham.ac.uk

Abstract. With the widespread adoption of online platforms and user-generated content, the sentiment analysis of movie reviews has become a valuable task for academic research and industry applications. This paper systematically studies three main deep learning methods: convolutional neural network (CNN), long-term and short-term memory (LSTM) and its hybrid model CNN-LSTM. CNN captures local text features, while LSTM models long-term dependencies in the sequence. Hybrid models combine their advantages to provide higher accuracy and context understanding. Through the comparative analysis of data sets such as IMDb, Amazon and Twitter, the CNN-LSTM model has always shown excellent performance, with an accuracy rate of up to 98.09%. However, there are still challenges in dealing with irony and domain generalization. Future improvements may focus on integrating attention mechanisms and explainable artificial intelligence technology. This review provides valuable insights for the development of a robust, explainable and scalable sentiment analysis system for film reviews, and highlights potential directions for further enhancing model adaptability in real-word application.

Keywords: Sentiment analysis, Movie review, Convolutional neural network, Long-term and short-term memory, Hybrid CNN-LSTM model

1. Introduction

In recent year, with the rapid development of online platforms and social media, the way people share opinions has undergone profound changes, especially in the evaluation of entertainment products such as movies. Movie reviews are not only influencing audiences' choice on watching movie, but also significantly affecting the box office performance of movies, marketing strategies and the overall impression of the public. The significant amount of user-generated comment information makes efficient sentiment analysis techniques more important than ever before [1].

Sentiment analysis, also known as opinion mining, is a method that uses computational technology to classify text data and determine whether its sentiment tendency is positive, negative or neutral. However, given the complexity and subtleties inherent in human language, traditional analytical methods such as Naive Bayes and Support Vector Machine (SVM) have obvious limitations in capturing contextual meaning and semantic details [2]. Recently, deep learning method has become an important ingredient to solve these problems. The advantage of deep learning method

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.LD26888

is that it can automatically learn features directly from original text data without explicit feature engineering [3].

A vast amount of deep learning methods has already applied on sentiment analysis tasks. Reference Dholpuria et al. demonstrated that Convolutional Neural Networks (CNN) effectively capture local textual features and semantic representations, thus the classification accuracy of the CNN is better than traditional machine learning method [4]. In addition, Ali et al. compared multiple deep learning models, including CNN, Long Short-Term Memory Network (LSTM), and CNN-LSTM hybrid network, and found that the CNN-LSTM hybrid model performed better because it could simultaneously capture local and sequential patterns of text. The accuracy rate of this model on the IMDb movie review dataset has reached 89.2% [3].

Furthermore, Ullah et al. proposed a seven-layer deep neural network consisting of an embedding layer, a convolutional layer, a pooling layer, a dropout layer, and a fully connected layer for the sentiment classification of film reviews [1]. This model has a stronger generalization ability and robustness, and the accuracy rate achieves 92% on the binary sentiment analysis tasks. Additionally, Abinash emphasizes the potential of the hybrid natural inspiration method that combines particle swarm optimization (PSO) and flower pollination algorithm (FPA) with artificial neural network (ANN) in a variety of emotional classification, which has significantly improved the efficiency of feature selection and overall classification accuracy [5].

Based on previous research, this paper provides a current mainstream deep learning method especially on CNN, LSTM and their hybrid models-used for sentiment analysis of movies reviews. This paper will compare different methods' advantages and disadvantages as well as performance difference and analysis how different model designs influence sentiment classification through the review of existing research. Furthermore, this paper also points out the key challenges currently faced in this field, such as the difficulty of handling subtle language expressions, the trade-off between accuracy and interpretability, and the limited generalizability across domains. By integrating existing research results, this paper aims to provide valuable references and inspirations for researchers and practitioners dedicated to enhancing the sentiment analysis system for film reviews.

2. Research methodology

Convolutional Neural Networks (CNNs) have gradually become an efficient solution in the task of sentiment analysis for film reviews. The main advantage of CNN is that it has a powerful ability to capture local textual feature and semantic representations, which significantly improves the accuracy of text classification [4]. Abarja et al. demonstrated the effectiveness of CNN models in predicting film ratings by analyzing historical features such as actors, directors, film genres, and production companies [6]. Their research emphasizes that compared with the traditional collaborative filtering method, the integration of CNN with these historical features provides a more objective and accurate movie rating prediction, while the traditional collaborative filtering method relies on post-release elements such as user reviews [6]. Yu further verified the effectiveness of CNN by constructing a sentiment analysis system using 25,000 film reviews [7]. Yu used the enhanced bag-of-words (BoW) model combined with TF-IDF vectorization to build a CNN model, with an average accuracy rate of 80.62% in emotional classification. It has confirmed the robustness of CNN in extracting emotional information from text data. In addition, Tammina et al. used deep CNN to conduct sentiment analysis on IMDb movie reviews [8]. The experimental results showed that the CNN model effectively extracted the emotional characteristics in the text, which performed better than traditional machine learning methods such as support vector machine (SVM) and simple Bayes, further

DOI: 10.54254/2755-2721/2025.LD26888

consolidating the applicability of CNN in complex text sentiment analysis tasks. Moreover, a comparative study by Ali et al. on the IMDb dataset reported CNN's performance as superior compared to other standalone models, achieving an accuracy of 87.7%, which further emphasizes CNN's advantages in processing and classifying movie review text sentiment.

Long Short-Term Memory (LSTM) is a variant of advanced Recurrent Neural Networks (RNNs), which excels in sentiment analysis tasks due to its excellent contextual information learning and memory ability. In recent years, in the research in the field of sentiment analysis of film reviews, the LSTM model has shown the characteristics of continuous superiority over traditional classifiers and is especially good at capturing subtle sentiment expressions implicit in reviews [9].

Qaisar et al. used LSTM to conduct sentiment classification on the IMDb movie review dataset, achieving an impressive accuracy of approximately 89.9% [10]. This research highlight that the advantages of LSTM in processing text sequence data, especially it effectively solves the problem of vanishing gradients in traditional RNN models. Meanwhile, the study also pointed out that LSTM networks can effectively learn the emotional details in film reviews, thereby significantly improving the classification accuracy.

Similarly, Purohit et al. compare the performance of LSTM and Random Forest classifiers by using 50000 movie reviews [11]. Their research finds that LSTM is better than Random Forest classifier in understanding complex language and sentiment detail in movie reviews. Although the computing requirements are higher, the LSTM network based on sequence processing can capture more detailed context information and achieve accuracy, precision, recall rate and F1 score comparable to random forests (all approximately 0.85), demonstrating the great potential of LSTM in deep sentiment analysis applications [11].

Dashtipour et al. further enhanced the superiority of LSTM by implementing a context-aware deep learning method for Persian film reviews in their study which found that the stacked bidirectional LSTM model significantly improved in accuracy (95.61%) compared with traditional methods such as Support Vector Machine (SVM) and logistic regression [9]. This result clearly shows the powerful performance and flexibility of LSTM networks in processing texts in different languages.

Hybrid deep learning models that combine convolutional neural networks (CNNs) and long short-term memory networks (LSTM) have become a very promising architecture for sentiment analysis, especially suitable for scenarios that require local feature extraction and context semantic understanding, such as movie review analysis.

Feisheng used the CNN-LSTM hybrid model to systematically review the applications of sentiment analysis, emphasizing their potential in user-generated content-rich environments, such as digital media and film reviews [12]. Feisheng believes that the CNN layer is particularly effective in detecting local n-gram patterns and emotional clues, while the LSTM layer retains the semantic dependence across sequences [12]. According to Feisheng, compared with the independent CNN or LSTM model, the integration of these two components leads to better performance [12].

Paul et al. designed a CNN-LSTM model to classify e-commerce product reviews into emotional categories [13]. Despite the differences in the field, their research results are highly applicable to the film review task. The model includes text preprocessing, marking and embedding, followed by the CNN layer for local feature extraction and the LSTM layer for context modelling. Hybrid models show a strong ability to distinguish between different categories of subtle emotions, which shows that the same architecture can effectively handle the complex emotional textures found in movie reviews.

DOI: 10.54254/2755-2721/2025.LD26888

Venkatesh et al. used GloVe word embedding to develop a hybrid CNN-LSTM framework specifically for the emotional analysis of football-related tweets [14]. Although they focus on sports content, the performance of the model - a test accuracy of more than 92% - proves its robustness in handling short, informal and emotional texts. This makes the model structure highly related to film reviews, which also have the characteristics of colloquial language, irony and implicit emotions. Venkatesh and others emphasized that the CNN layer helps to extract spatial language characteristics, while the LSTM layer encodes the sequence stream, which enhances the classification performance in the dynamic text environment.

3. Experimental results and analysis

3.1. Datasets

The models used in each study (CNN, LSTM and CNN-LSTM) were mainly based on open-source datasets for experiments. The most used data set is the IMDb large-scale movie review data set, which contains 50,000 highly polarized movie reviews, which are divided into training and test subsets on average. For instance, Purohit et al. used the IMDb datasets for LSTM-based sentiment classification, after applying text preprocessing steps such as tokenization, stop-word removal, and lemmatization [11]. Similarly, Paul et al. adopts e-commerce product reviews like the style of movie reviews [13]. The datasets have also undergone preprocessing operations such as tokenization, stop-word removal and lemmatization. These datasets show significant differences in text length, lexical complexity and domain terminology, which directly affect the generalization and accuracy of the model.

3.2. Evaluation metrics

The models use the standard metric of classification tasks to evaluate these models. These indicators include accuracy, which measures the proportion of correct predictions; precision, which reflects how many positive results of predictions are positive; recall, indicating the ability of the model to recognize all relevant instances; and f1 score, which is the average of the reconciliation of accuracy and recall rates. In all studies, accuracy and f1 scores are the most frequently reported indicators, providing a balanced evaluation of correctness and robustness.

3.3. Result analysis

To evaluate different deep learning models' effectiveness in sentiment analysis of movie and product reviews. This paper conducts a comparative analysis of three mainstream architectures - CNN, LSTM, and CNN-LSTM

Firstly, the CNN model is designed by Tammina uses Word2Vec word vectors as input and introduces Dropout (0.2) and Adam optimizer. On the IMDB and Amazon data sets, CNN achieved an accuracy rate of 74% on IMDB and 68% on Amazon respectively, which is significantly better than the traditional model: random forest (55.8% / 44.3%) and simple Bayes (54.3% / 46 .3%). The results show that CNN has obvious advantages in capturing local spatial characteristics of text.

Secondly, the LSTM model was applied to 10,000 preprocessed IMDB comments (half positive and half negative) and trained with the Adam optimizer, ultimately achieving an accuracy rate of 89.9% [10]. This indicates that LSTM performs well in handling long dependencies in text sequences.

DOI: 10.54254/2755-2721/2025.LD26888

Finally, Paul et al. adopted the CNN-LSTM hybrid model to analyze e-commerce reviews and achieved the best performance [13]. The accuracy rate of the model is as high as 98.09%, surpassing that of BiGRU (97.97%) and separate LSTM (97.82%). In addition, its F1 score also reached 98%, showing a good balance between accuracy and recall. CNN is responsible for extracting local features, and LSTM is responsible for processing chronological dependence, which makes the model particularly stable and accurate in classification tasks.

in summary, the hybrid CNN-LSTM model shows the most robust and accurate performance, followed by standalone LSTM and CNN. These results emphasize the significant of combining spatial and sequential modeling for sentiment classification tasks.

4. Discussion

Despite CNN, LSTM, and hybrid CNN-LSTM model perform well in the sentiment analysis of film reviews. However, there are still many challenges. One of the key problems is that it is difficult to deal with complex emotions such as sarcasm, irony, and cultural reference, which is easy to lead to misjudgment by classifiers. Furthermore, most models rely heavily on stander datasets such as IMDb, limiting their generalizability across language and domains. Although the hybrid CNN-LSTM shows the best performance on the rate of accuracy, but it has high model complexity and training time, making real-time deployment difficult. Additionally, interpretability is an issue that all deep models need to consider - especially in hybrid architecture, where the combination of space and time learning blurs the decision-making logic. Future research can consider introducing attention mechanisms or interpretable AI frameworks to improve transparency and enhance the adaptability of models in real situations with the help of diversified, multilingual and informal text data.

5. Conclusion

This paper conducts a comprehensive study on the application of three main deep learning architectures, convolutional neural network (CNN), long-term and short-term memory network (LSTM) and CNN-LSTM hybrid model, in emotional analysis of film reviews. Through the analysis and comparison of mechanisms and empirical performance of different academic sources, this study clearly understands their advantages, limitations and application potential.

The CNN model is good at capturing the local and spatial features of the text, which is especially useful for detecting n-gram-level emotional patterns. On the other hand, the LSTM model can model long-term dependencies and sequential information, making them more suitable for understanding the context and emotional evolution in longer comments. The CNN-LSTM hybrid model combines local feature extraction and time dynamics and performs well in most empirical studies. Specifically, several studies reviewed in this article show that CNN-LSTM is consistent.

Nevertheless, the current model still faces many challenges, such as the inability to effectively identify irony, antony and cultural meanings, the high complexity of hybrid model training, the strong dependence on pre-labelled data such as IMDb, and the widespread lack of interpretability in deep architecture, which limit the promotion and deployment of the model in a real complex environment.

Future research should focus on the introduction of attention mechanisms and interpretable AI frameworks to improve the transparency and credibility of the model; at the same time, expand the diversity of training data, such as the introduction of informal texts, multilingual content and

DOI: 10.54254/2755-2721/2025.LD26888

specific field comments, so as to improve the generalization ability of the model. Migration learning and domain adaptation strategies are also worth further exploring.

In short, although CNN, LSTM and CNN-LSTM methods have greatly promoted the development of emotional analysis, further research is still needed to make it more explainable, generalized and applicable to actual scenarios.

References

[1] Ullah, K., Rashad, A., Khan, M., Ghadi, Y., Aljuaid, H., Nawaz, Z. (2022). A Deep Neural Network-Based Approach for Sentiment Analysis of Movie Reviews, Complexity, 5217491, 9 pages. https: //doi.org/10.1155/2022/5217491

[2] Rahman, A. and Hossen, M. S. (2019). Sentiment Analysis on Movie Review Data Using Machine Learning Approach. 2019 International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, pp. 1-4, doi: 10.1109/ICBSLP47725.2019.201470.

[3] Ali, N. M., Abd El Hamid, M. M., & Youssif, A. (2019). Sentiment analysis for movies reviews dataset using deep learning models. International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol, 9.

[4] Dholpuria, T., Rana, Y. K., and Agrawal, C. (2018). A Sentiment analysis approach through deep learning for a movie review, 2018 8th International Conference on Communication Systems and Network Technologies (CSNT), Bhopal, India, pp. 173-181, doi: 10.1109/CSNT.2018.8820260.

[5] Abinash, T. (2025). Fuzzy rule based multi class sentiment analysis using hybrid nature inspired machine learning technique, PREPRINT (Version 1) available at Research Square [https: //doi.org/10.21203/rs.3.rs-6859329/v1

[6] Abarja, R. A., & Wibowo, A. (2020). Movie rating prediction using convolutional neural network based on historical values. Int. J, 8, 2156-2164.

[7] Yu, D. (2021, December). Intelligent Analysis System of Movie Reviews Using Deep Learning and Convolutional Neural Networks. In 2021 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS) (pp. 617-621). IEEE.

[8] Tammina, S., & Annareddy, S. (2020, January). Sentiment analysis on customer reviews using convolutional neural network. In 2020 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-6). IEEE.

[9] Dashtipour, K., Gogate, M., Adeel, A., Larijani, H., & Hussain, A. (2021). Sentiment Analysis of Persian Movie Reviews Using Deep Learning. Entropy, 23(5), 596. https: //doi.org/10.3390/e23050596

[10] Qaisar, S. M. (2020). Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory, " 2020 2nd International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, pp. 1-4, doi: 10.1109/ICCIS49240.2020.9257657.

[11] Purohit, S., Rajput, A., Vats, S., Mudgal, R., Shah, O. A., and Verma, A. (2024). Comparative Analysis of LSTM and Random Forest Algorithms for Sentiment Classification in Movie Reviews, 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, pp. 1053-1057, doi: 10.1109/ICAAIC60222.2024.10575246.

[12] Feisheng, L. (2024). Systematic Review of Sentiment Analysis: Insights Through CNN-LSTM Networks, 2024 5th International Conference on Industrial Engineering and Artificial Intelligence (IEAI), Bangkok, Thailand, pp. 102-109, doi: 10.1109/IEAI62569.2024.00026.

[13] Paul, P., Acharya, S., Misra, B., Majumder, S., Dey, N. and Pise, P. (2024). Sentiment Analysis for E-Commerce Product Reviews Using CNN-LSTM, 2024 First International Conference for Women in Computing (InCoWoCo), Pune, India, pp. 1-7, doi: 10.1109/InCoWoCo64194.2024.10863425.

[14] Venkatesh, S. U. Hegde, Z. A. S and N. Y, (2021). Hybrid CNN-LSTM Model with GloVe Word Vector for Sentiment Analysis on Football Specific Tweets, 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, pp. 1-8, doi: 10.1109/ICAECT49130.2021.9392516.