Proceedings of the 7th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2025.AST27216

Hybrid Sentiment Analysis Model for Movie Reviews

Mingze Hua

Northwestern Polytechnical University, Xi'an, China
15189083108@163.com

Abstract. With the rapid development of natural language processing technology, BERT and CNN (Convolutional Neural Network) have achieved remarkable results in their respective fields. With the bidirectional Transformer encoder structure, BERT (Bidirectional Encoder Representations from Transformers) can effectively capture text context information and perform well in semantic understanding tasks. On the other hand, CNN is good at extracting local features through convolutional layers, which has obvious advantages when dealing with data with grid structure. However, traditional sentiment analysis methods have shortcomings in processing complex semantics and capturing local semantics, and the direct application of BERT also faces problems such as high computational cost and weak local feature capture ability. In this paper, a hybrid sentiment analysis model based on BERT and Convolutional Neural Network (CNN) is proposed to solve the above problems. By combining the advantages of BERT and CNN, the system can effectively capture the sentiment tendency of the text. Experimental results show that the accuracy of the model on IMDb movie review dataset is 91.5%, and the accuracy of the model on ChnSentiCorp dataset is 89.4%, which is significantly better than the baseline model.

Keywords: BERT, CNN, Sentiment Analysis, Movie Review

1. Introduction

In today's digital era, the film industry is booming, and the audience's evaluation and feedback data for the film are exploding. These movie reviews contain rich emotional information, and it is of great significance to conduct accurate sentiment analysis on them [1]. It can not only help audiences quickly select movies that meet their preferences, but also provide valuable market insights for film producers and distributors to formulate more accurate marketing strategies [2].

Traditional methods based on dictionaries and statistical learning have obvious limitations when dealing with complex texts such as movie reviews. For example, when dealing with polysemy, it is difficult for the dictionary-based method to accurately determine the sentiment orientation of a word according to the specific context. However, traditional machine learning methods extract statistical features such as word frequency and TF-IDF (Term Frequency-Inverse Document Frequency) for sentiment analysis, which are limited in modeling complex semantics and cannot fully understand the context dependencies in the text [3].

In recent years, significant breakthroughs have been made in pre-trained language models, especially BERT (Bidirectional Encoder Representations from Transformers). BERT significantly

© 2026 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.AST27216

improves the performance of sentiment analysis through deep semantic modeling [4]. However, there are still some problems in the direct application of BERT, such as high computational cost, which will bring large resource consumption when processing large-scale movie review data. At the same time, its ability to capture local semantics is insufficient, and it is difficult to accurately grasp the emotional cues expressed by local word combinations in the text.

To solve the above problems, this paper proposes a BERT-CNN (Convolutional Neural Network) hybrid model. In this model, BERT's powerful bidirectional semantic encoding ability is used to extract global semantic features, and the convolutional layer of CNN is used to capture local emotional cues, so as to form a hierarchical feature expression, which can capture the sentiment tendency in the text more comprehensively and accurately.

2. Related work

2.1. Evolution of sentiment analysis methods

In the early stage, lexicon-based sentiment analysis methods rely on manually constructed sentiment dictionaries, such as Senti WordNet [5]. This method is simple and direct, but due to the dynamic changes of language, new words and new semantics are constantly emerging, and the manually constructed dictionary is difficult to update in time, which makes it unable to adapt to the complex and changing language environment, and it is difficult to accurately judge the sentiment orientation of new words or semantic evolution.

Traditional machine learning methods use feature engineering to extract statistical features such as word frequency and TF-IDF to conduct sentiment analysis [6]. Although there is some progress compared with the dictionary-based method, there are still many shortcomings when dealing with complex semantics. For example, for some sentences with complex semantic structures, it is difficult to deeply understand the true meaning of the sentence by only relying on statistical features, which affects the accuracy of sentiment analysis.

The emergence of deep learning methods has brought new ideas for sentiment analysis, among which RNN, LSTM, and other models are widely used [7]. These models can capture the sequence dependencies in the text and solve the problem of processing context information to a certain extent. However, they still have limitations when dealing with long-distance dependency problems, and cannot well capture the semantic association between distant words in the text.

2.2. Advantages of BERT models

BERT was put forward in 2018 by the Google team as a training language model. Its core innovation lies in a two-way Transformer Encoder structure, which enables global modeling of text context information. Compared with traditional one-way language models (such as LSTM), BERT dynamically generates word vectors by combining the left and right contexts, which effectively solves the semantic ambiguity of words such as "bank." The model adopted a dual-task pre-training strategy of Masked Language Model (MLM) and Next Sentence Prediction (NSP), which showed significant advantages in semantic understanding tasks [6], and became a landmark technology in the field of natural language processing. However, BERT has limitations in long text processing: the computational complexity of its self-attention mechanism increases squarely with the length of the text, and its ability to capture local semantic patterns (such as continuous word order relationships) is weak, which provides an improvement direction for subsequent research such as Longformer and Reformer and other long text optimization models. BERT, through a two-way Transformer encoder,s

DOI: 10.54254/2755-2721/2025.AST27216

capture context semantics. The dynamic word embedding technology effectively resolves the problem of polysemy [8]. However, when BERT is directly applied to long texts, the computational cost is high, and the explicit modeling of local semantics is lacking.

2.3. Local feature extraction of CNN

CNN is a kind of deep learning model, especially suitable for processing mesh-structure data, such as images and texts. CNNs have achieved great success in the field of computer vision, but can also be applied to natural language processing tasks. A typical CNN contains convolutional layers and pooling layers.

Convolutional layers extract local features in the input data by applying a set of learnable filters, also known as convolutional kernels or feature detectors. Each filter slides over the input data and computes the dot product of the local regions, producing a Feature Map. The convolution operation can capture information about the spatial hierarchy in the data.

The pooling layer is used to reduce the spatial dimension of the feature map and reduce the number of parameters while preserving the most important features. Common pooling operations are Max Pooling and Average Pooling. CNN by convolution kernels extraction in text n - "gramm characteristics, suitable for capturing local semantic model [9]. It combines CNN and BERT, retaining global semantics while at the same time enhancing the expression of partial emotional signals.

3. Method design

3.1. Data preprocessing

The IMDb (English) and ChnSentiCorp (Chinese) datasets, which contain 50,000 movie reviews, were used in this study. These data cover different genres and different years of movie reviews, which are widely representative. The data is cleaned to remove invalid characters, such as special symbols, garbled codes, etc. At the same time, the stop words list is used to remove common words that have no substantial meaning for sentiment analysis. For Chinese text, the Jieba word segmentation tool is used for word segmentation, and the continuous Chinese text is segmented into single words for subsequent feature extraction and model processing. The sentiment labels (positive/negative) are converted into 0/1 numerical form, where 0 represents negative sentiment and 1 represents positive sentiment, so that the data can adapt to the input requirements of the model [10]. Finally, the data set was divided into a training set, a validation set, and a test set according to the ratio of 8:1:1. In the process of partitioning, the distribution of each category on different subsets was balanced, and the situation that a category was excessively concentrated in a subset was avoided to ensure the effectiveness of model training and evaluation.

3.2. Model architecture design

The pre-trained BERT model is used to transform the input text into context-sensitive word vectors. BERT can fully capture the text of the global semantic information, for the feelings of subsequent analysis provide a rich semantic representation. Local sentiment features in the text are captured by a 3-kernel a convolution operation. For example, "the film was not a bit wonderful." In such sentences, convolution layer can effectively capture the "don't" and the combination of the negative word and the adjective "good", and more accurately grasp the local emotion clues. AdaptiveMaxPool1d is used to retain the key features output by the convolutional layer, and reduce

DOI: 10.54254/2755-2721/2025.AST27216

the feature dimension to reduce the amount of calculation, while avoiding excessive loss of information. The fully connected layer (nn.Linear) is used to map the pooled features to the sentiment category, and the probability of the text belonging to the positive or negative sentiment is output [11].

3.3. Training strategy

Optimizer: Choose the AdamW optimizer and set the learning rate to 2e-5. The AdamW optimizer combines the advantages of the Adam algorithm and improves the weight decay, which can adaptively adjust the learning rate during the training process to improve the convergence speed and training effect of the model.

Loss function: The cross-entropy loss function is used, which can effectively measure the difference between the model prediction results and the true label, and the parameters of the model are optimized by minimizing the loss function.

Early stop mechanism: During the training process, the accuracy of the validation set is monitored. When the accuracy of the validation set does not improve within a certain number of rounds, the training is stopped and the current optimal model is saved to prevent the model from overfitting and ensure that the model has good generalization ability on unknown data.

4. Experiment and result analysis

4.1. Experimental setup

4.1.1. Dataset

IMDb: Contains 25,000 training sets and 25,000 test sets. This data set is a commonly used English data set in the field of movie review sentiment analysis, which is widely recognized and representative.

ChnSentiCorp: The Chinese sentiment dataset covers reviews from books, movies and other fields, and the movie reviews section provides rich Chinese data resources for this experiment.

4.1.2. Parameters are configured as follows

BERT model: BERT-base-Uncased is selected as the pre-training model, which has good performance and wide application in English text processing.

The maximum sequence length is set to 256, considering the computing resources of the model and the length of the text, this length can cover most of the content of the movie review, while avoiding the problem of high computational cost caused by too long sequences.

Batch size: Set to 16 to strike a balance between training efficiency and memory usage, and ensure that the model can make full use of data for parameter updates during training.

Training rounds: Three rounds of training are carried out. Through experimental observation, three rounds of training can not only make the model fully learn the data features, but also avoid the overfitting problem that may be caused by too many rounds of training.

DOI: 10.54254/2755-2721/2025.AST27216

4.2. Results and discussion

4.2.1. Comparison of model performance

The experiments compare the training accuracy, validation accuracy and test accuracy of the BERT-CNN hybrid model on different datasets, and the results are shown in Table 1.

Table 1. Comparison table of model performance

Datasets	Model combination	Training accuracy	Validation accuracy	Test accuracy
IMDb	BERT+CNN	0.910	0.915	0.908
ChnSentiCorp	BERT+CNN	0.874	0.894	0.882

Table 1 shows that on the IMDb dataset, the training accuracy of BERT-CNN model is 0.910, the validation accuracy is 0.915, and the test accuracy is 0.908. On the ChnSentiCorp dataset, the training accuracy is 0.874, the validation accuracy is 0.894, and the test accuracy is 0.882. This shows that the model achieves high accuracy on both English and Chinese movie review datasets, and has good sentiment analysis ability.

4.2.2. Result analysis

The English data set performs better: the accuracy on IMDb data set is higher than that on ChnSentiCorp data set. This may be because BERT is more fully pre-trained on English corpus, and has stronger semantic understanding and feature capture ability for English text. On the other hand, the process of Chinese word segmentation may introduce a small amount of noise, which affects the model's accurate judgment of Chinese text sentiment.

The effectiveness of CNN: Compared with the pure BERT model (assuming a validation accuracy of 89.4%), the performance of the BERT-CNN model after adding CNN is improved by 2.1%. This fully proves the importance of CNN in capturing local sentiment features. By combining the local feature extraction ability of CNN, the BERT-CNN model can capture text sentiment information more comprehensively, thereby improving the overall performance.

During the training process, the training loss decreases with each round, and the validation accuracy reaches its peak in the second round. This shows that the model can converge quickly and learn the features of the data in a short training round. At the same time, the early stopping mechanism effectively prevents overfitting, and the accuracy of the best model on the validation set reaches 91.5%, which ensures the generalization ability of the model.

4.3. Classification reports

On the IMDb test set, the classification of the model is reported as Table 2.

DOI: 10.54254/2755-2721/2025.AST27216

Table 2. Model classification report

plaintext	precision	recall	f1-score	support
negative	0.91	0.93	0.92	12500
positive	0.92	0.90	0.91	12500
accuracy	-	-	0.91	25000
macro avg	0.91	0.91	0.91	25000
weighted avg	0.91	0.91	0.91	25000

From the classification report, it can be seen that the model classifies positive and negative emotions in a more balanced way. When predicting negative sentiment, the precision is 0.91, the recall is 0.93, and the F1 score is 0.92. When predicting positive sentiment, the precision is 0.92, the recall is 0.90, and the F1-score is 0.91. The overall accuracy reaches 0.91, and the F1 score also reaches 0.91, indicating that the model has good classification performance on the IMDb test set, which can meet the needs of practical applications.

5. Conclusion

The BERT-CNN model proposed in this paper significantly improves the accuracy of movie review sentiment analysis by combining the global semantic capture ability of BERT and the local feature extraction advantages of CNN. Experimental results on IMDb and ChnSentiCorp datasets show that the proposed model performs well in both English and Chinese movie review sentiment analysis tasks, which is superior to traditional models and some single models.

Future work can be further explored from the following directions: the introduction of unstructured data such as emojis and images. In actual movie review scenarios, emoticons and related images can also convey rich emotional information in addition to textual information. The fusion of these multimodal data with text data can capture the user's emotional tendency more comprehensively, and improve the accuracy and reliability of sentiment analysis. Fine-tuning is performed for specific fields, such as medical treatment and finance. Texts in different fields have different language styles and semantic characteristics. Through the fine-tuning of domain-specific data, the model can be better adapted to the sentiment analysis task in the domain, and the application effect of the model in the professional field can be improved. Through knowledge distillation (such as DistilBERT) and other technologies, it can reduce the computational cost and improve the deployment efficiency. With the increase in model size and complexity, the computational cost and deployment difficulty also increase accordingly. Knowledge distillation technology can be used to transfer the knowledge of large and complex models to small models without losing too much accuracy, thereby reducing the computational resource requirements of the model and facilitating more efficient deployment and application in practical scenarios.

References

[1] Zhang, A., & Hou, D. (2024). A review of Chinese film reviews in 2023. Yangtze River Literature Review, (02), 60-70.

[2] Ullah, R., Alam, A. M., & Zeb, A. (2021). How emotions in online reviews affect movie sales: Evidence from Hollywood. Journal of Retailing and Consumer Services, 85, 104304.

[3] Zhang, J. (2020). Research on the aesthetic education value transformation of film general education course under the OBE concept: Taking the Film Review Center of Jinzhong Institute of Information Science as an example.

DOI: 10.54254/2755-2721/2025.AST27216

Journal of Journalism Research Guide, 16(04), 104-108.
[4] Du, L. (2024). Film review in the era of artificial intelligence: From instruction engineering to thought simulation. Contemporary Film, (08), 45-50.
[5] Lu, W. (2024). Exploring the evaluation criteria of contemporary film criticism: From aesthetic to social significance. Theatre House, (21), 156-158.
[6] Liu, Y., Yang, K., & Dong, X. (2024). LSTM sentiment analysis based on movie review text. Journal of Changchun University of Technology, 45(03), 233-240. (in Chinese)
[7] Lin, Y. (2024). Chinese movie review sentiment classification method based on topic features and BERT model [Master's thesis, Jiangxi University of Finance and Economics]. (in Chinese)
[8] Wu, S. W. (2024). Research on film review sentiment analysis method based on deep learning [Master's thesis, Northwest Normal University]. (in Chinese)
[9] Li, S. (2024). Research and application of movie recommendation algorithm based on online reviews [Master's thesis, Anhui Jianzhu University].
[10] Chen, Y. (2024). Research on hybrid recommendation algorithm based on text sentiment analysis [Master's thesis, Wuhan Institute of Technology].
[11] Liu, C. (2023). Research on sentiment analysis of film reviews based on deep learning [Master's thesis, Changchun University].