International Journal of Research Publication and Reviews, Vol (5), Issue (9), September (2024), Page – 3354-3357

WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Review on Sentiment Analysis with Multi-Modal Deep Learning Techniques

Ankit Kumar[1], Nitesh Gupta[2], Anurag Shrivastava[3]

[1]MTech Scholar, CSE, Department, NIIST, Bhopal
[2]Associate Professor, CSE, Department, NIIST, Bhopal
[3]Associate Professor, CSE, Department, NIIST, Bhopal
[1]Ank14itk@gmail.com, [2]9.nitesh@gmail.com, [3]anurag.shri08@gmail.com

Abstract:

Sentiment analysis has become an essential tool in understanding user opinions and emotions across various platforms. Traditional approaches primarily rely on textual data, often overlooking the rich information embedded in other modalities such as audio and visual data. This review explores recent advancements in sentiment analysis using multi-modal deep learning techniques. This paper examines the integration of text, visual, and auditory information to better understand sentiment. Key deep learning models, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformer-based architectures, are discussed in the context of multi-modal inputs. Techniques such as feature fusion, cross-modal attention, and alignment are highlighted. The review also addresses challenges like data alignment across modalities, handling missing data, and computational complexity.

Keywords— Sentiment Analysis, Hybrid model, CNN, RNN, Deep Learning

I. Introduction

Sentiment analysis is a crucial task in natural language processing (NLP) that involves determining the sentiment expressed in a piece of text. It has wide-ranging applications, from analyzing customer reviews and social media interactions to gauging public opinion and market trends. Traditional sentiment analysis methods primarily focus on textual data, using techniques such as machine learning classifiers, lexicon-based approaches, and more recently, deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). While these methods have achieved significant success, they often fall short in fully capturing the richness of human emotions, which are not only conveyed through text but also through audio and visual cues [1]. Incorporating multi-modal data—text, audio, and visual inputs—into sentiment analysis can provide a more holistic understanding of sentiment. For instance, the tone of voice and facial expressions can significantly alter the perceived sentiment of spoken or written words. This multi-modal approach can enhance the accuracy of sentiment analysis models, making them more robust and reliable. This research aims to develop a multi-modal deep learning framework for sentiment analysis, leveraging the strengths of LSTM networks for processing sequential text and audio data, and the VGG16 network for extracting features from images. By integrating these modalities, the proposed model seeks to capture a more comprehensive representation of sentiment. This paper presents the design, implementation, and evaluation of this multi-modal framework, demonstrating its effectiveness in improving sentiment analysis accuracy. Our experiments indicate that the multi-modal approach significantly outperforms traditional single-modal methods, highlighting the potential of multi-modal deep learning techniques in sentiment analysis.

Figure 1. Sentiment Analysis

II. BACKGROUND AND RELATED WORK

This work [1] the rating of movie in twitter is taken to review a movie by using opinion mining This paper proposed a hybrid methods using SVM and PSO to classify the user opinions as positive, negative for the movie review dataset which could be used for better decisions. Authors [2] found that PSO affect the accuracy of SVM after the hybridization of SVM-PSO. The best accuracy level that gives in this study is 77% and has been achieved by SVM-PSO after data cleansing. On the other hand, the accuracy level of SVM-PSO still can be improved using enhancements of SVM that might be using another combination or variation of SVM with other optimization method. Authors [3] perform sentiment analysis from the point of view of the consumer review summarization model for capitalists. Author's outlined several research concerns and possible solutions for the challenges that occur when performing sentiment analysis for raw online reviews. Using the hybrid feature extraction method proposed in this work, the input pre-processed reviews can be transformed into meaningful feature vectors, allowing efficient, reliable, and robust sentiment analysis to take place. Authors [4] results show that sentiment analysis is an effective technique for classifying movie reviews. This analysis focused primarily on English-language movie reviews, and the models may not perform as effectively when applied to other languages due to linguistic variations and cultural differences. This study introduces a sentiment analysis approach using advanced deep learning models: Extra-Long Neural Network (XLNet), Long Short-Term Memory (LSTM), and Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM). Authors [5] Hybrid deep sentiment analysis learning models that combine long short-term memory (LSTM) networks, convolutional neural networks (CNN), and support vector machines (SVM) are built and tested on eight textual tweets and review datasets of different domains. %e hybrid models are compared against three single models, SVM, LSTM, and CNN. Both reliability and computation time were considered in the evaluation of each technique. %e hybrid models increased the accuracy for sentiment analysis compared with single models on all types of datasets, especially the combination of deep learning models with SVM. %e reliability of the latter was significantly higher. Authors [6] examine primary taxonomy and newly released multimodal fusion architectures. Recent developments in MSA architectures are divided into ten categories, namely early fusion, late fusion, hybrid fusion, model-level fusion, tensor fusion, hierarchical fusion, bi-modal fusion, attention-based fusion, quantum-based fusion and word-level fusion. A comparison of several architectural evolutions in terms of MSA fusion categories and their relative strengths and limitations are presented. Finally, a number of interdisciplinary applications and future research directions are proposed. Authors [7] review the multimodal sentiment analysis by combining several deep learning text and image processing models. These fusion techniques are RoBERTa with EfficientNet b3, RoBERTa with ResNet50, and BERT with MobileNetV2. This work focuses on improving sentiment analysis through the combination of text and image data. The performance of each fusion model is carefully analyzed using accuracy, confusion matrices, and ROC curves. The fusion techniques implemented in this study outperformed the previous benchmark models. Notably, the EfficientNet-b3 and RoBERTa combination achieves the highest accuracy (75%) and F1 score (74.9%).

III. FINDINGS OF THE SURVEY

The review of sentiment analysis using multi-modal deep learning techniques reveals several key insights across various aspects of model development, data handling, and challenges in the field:

Enhanced Sentiment Understanding through Multi-Modal Integration: Traditional text-based sentiment analysis struggles to capture the full spectrum of sentiment expressed across different content types. By integrating multiple modalities such as text, images, and audio, multi-modal models can capture richer context and nuanced emotions. The fusion of these modalities provides a deeper understanding of sentiment, especially in cases where text alone may be ambiguous or lacking in emotional cues. Deep Learning Models for Multi-Modal Sentiment Analysis: Various deep learning models have been applied to multi-modal sentiment analysis. Convolutional Neural Networks (CNNs) are primarily used for image and video data, Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks for textual and sequential data, and Transformer-based models for both text and multi-modal tasks. Feature fusion, cross-modal attention mechanisms, and late fusion strategies have proven effective in combining information from multiple modalities. Feature Fusion and Cross-Modal Attention: Feature fusion techniques, which combine text, image, and audio features, have shown superior performance in multi-modal sentiment analysis. Cross-modal attention, where one modality helps guide the interpretation of another, has also emerged as a powerful method for aligning data from different sources and enhancing the overall sentiment prediction.

Challenges in Multi-Modal Sentiment Analysis:

Data Alignment: Synchronizing and aligning multi-modal data remains a significant challenge, especially when data from different modalities (e.g., text and video) are not temporally aligned.

Missing Modality: In real-world applications, it is common to encounter missing or incomplete modalities, which can lead to degraded performance in multi-modal models.

Computational Overhead: The computational cost of training deep learning models with multi-modal inputs is substantially higher compared to uni-modal approaches, making scalability a concern.

Table 3.1: Important Aspects of Multi-Modal Deep Learning in Sentiment Analysis

Aspect	Description	Models/Techniques	Challenges
Data Modalities	Integration of text, images, audio, and video for deeper sentiment analysis	Feature Fusion, Cross-Modal Attention	Data alignment, missing modalities
Deep Learning Models	Use of CNNs for images, RNNs/LSTMs for text, Transformers for multi-modal tasks	CNNs, RNNs, LSTMs, Transformers	High computational cost, data heterogeneity
Feature Fusion Techniques	Combining features from multiple modalities to enhance sentiment understanding	Early Fusion, Late Fusion, Cross-Modal Attention	Ensuring effective modality interaction, avoiding overfitting

CONCLUSION

This review highlights the advancements and challenges in sentiment analysis using multi-modal deep learning techniques. By integrating diverse data modalities such as text, images, and audio, multi-modal approaches significantly enhance the accuracy and depth of sentiment detection compared to traditional text-based methods. Deep learning models like CNNs, RNNs, LSTMs, and Transformer-based architectures have been successfully employed to fuse information from multiple sources, with techniques like feature fusion and cross-modal attention playing a pivotal role.

However, challenges remain, including data alignment across modalities, handling incomplete data, and the high computational costs of multi-modal models. Despite these hurdles, multi-modal sentiment analysis holds great promise for improving sentiment detection in real-world applications, especially as future research focuses on model interpretability, the inclusion of new data modalities, and the development of real-time systems. The potential for more nuanced and accurate sentiment analysis in fields like marketing, social media, and human-computer interaction makes this area of research highly valuable and worth further exploration.

References

[1] K.Uma maheswari, Ph.D et al "Opinion Mining using Hybrid Methods" International Journal of Computer Applications (0975 – 8887) International Conference on Innovations in Computing Techniques (ICICT 2015)

[2] Abd. Samad Hasan Basaria et al "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization" 1877-7058 © 2013 The Authors. Published by Elsevier Ltd.

[3] Gagandeep Kaur1,2* and Amit Sharma3 "A deep learning‑based model using hybrid feature extraction approach for consumer sentiment analysis"

[4] Mian Muhammad Danyal1, Opinion Mining on Movie Reviews Based on Deep Learning Models DOI: 10.32604/jai.2023.045617 2023,

[5] Cach N. Dang et al "Hybrid Deep Learning Models for Sentiment Analysis" Hindawi 2021

[6] Ankita Gandhi et. al. "Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions" https://doi.org/10.1016/j.inffus.2022.09.025, www.elsevier.com/locate/inffus, 2023

[7] Muhaimin Bin Habib1 et al "Multimodal Sentiment Analysis using Deep Learning Fusion Techniques and Transformers" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 15, No. 6, 2024

[6] Lei Zhang and Bing Liu : Aspect and Entity Extraction for Opinion Mining. Springer-Verlag Berlin Heidelberg 2014., Studies in Big Data book series, Vol 1, pp. 1-40, Jul. 2014.

[7] Zhen Hai, Kuiyu Chang, Gao Cong : One Seed to Find Them All: Mining Opinion Features via Association. ACM CIKM'12., LNCS 6608, pp. 255-264, Nov. 2012

[8] Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang : Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance. ieee transactions on knowledge and data engineering, Volume 26, No. 3 pp. 623-634, 2014.

[9] Hui Song, Yan Yan, Xiaoqiang Liu : A Grammatical Dependency Improved CRF Learning Approach for Integrated Product Extraction. IEEE International Conference on Computer Science and Network Technology, pp. 1787-139, 2012.

[10] Luole Qi and Li Chen : Comparison of Model-Based Learning Methods for Feature-Level Opinion Mining. IEEE International Conferences on Web Intelligence and Intelligent Agent Technology, pp. 265-273, 2011.

[11] Arjun Mukherjee and Bing Liu: Aspect Extraction through Semi- Supervised Modeling. In: Association for Computational Linguistics., vol. 26, no. 3, pp. 339-348, Jul. 2012.

[12] Liviu, P.Dinu and Iulia Iuga.: The Naive Bayes Classifier in Opinion Mining:In Search of the Best Feature Set. Springer-Verlag Berlin Heidelberg, 2012.

[13] Xiuzhen Zhang., Yun Zhou.: Holistic Approaches to Identifying the Sentiment of Blogs Using Opinion Words. In: Springer-Verlag Berlin Heidelberg, 5–28, 2011.

[14] M Taysir Hassan A. Soliman., Mostafa A. Elmasry., Abdel Rahman Hedar, M. M. Doss.: Utilizing Support Vector Machines in Mining Online Customer Reviews. ICCTA (2012).

[15] Ye Jin Kwon., Young Bom Park.: A Study on Automatic Analysis of Social NetworkServices Using Opinion Mining. In: Springer-Verlag Berlin Heidelberg, 240–248, 2011.

[16] Anuj Sharma., Shubhamoy Dey: An Artificial Neural Network Based approach for Sentiment Analysis of Opinionated Text. In: ACM, 2012.

[17] Yulan He. : A Bayesian Modeling Approach to Multi-Dimensional Sentiment Distributions Prediction. In: ACM, Aug. 2012.

[18] Danushka Bollegala, David Weir and John Carroll : Cross-Domain Sentiment Classification using a Sentiment Sensitive Thesaurus. ieee transactions on knowledge and data engineering, pp. 1-14, 2012.

[19] Andrius Mudinas., Dell Zhang., Mark Levene. : Combining Lexicon and Learning based Approaches for Concept-Level Sentiment Analysis. In: ACM, Aug. 2012.

[20] Vamshi Krishna. B, Dr. Ajeet Kumar Pandey, Dr. Siva Kumar A. P "Topic Model Based Opinion Mining and Sentiment Analysis" 2018 International Conference on Computer Communication and Informatics (ICCCI -2018), Jan. 04 – 06, 2018, Coimbatore, INDIA

[21] Rita Sleiman, Kim-Phuc Tran "Natural Language Processing for Fashion Trends Detection" Proc. of the International Conference on Electrical, Computer and Energy Technologies (ICECET 2022)
20-22 June 2022, Prague-Czech Republic

[22] 1d.sai tvaritha, 2nithya shree j, 3saakshi ns 4surya prakash s, 5siyona ratheesh, 6shimil shijo "a review on sentiment analysis applications and approaches" 2022 JETIR June 2022, Volume 9, Issue 6 www.jetir.org (ISSN-2349-5162)

[23] Pansy Nandwani1 · Rupali Verma1 "A review on sentiment analysis and emotion detection from text" https://doi.org/10.1007/s13278-021-00776-6

[24] Hoong-Cheng Soong, Norazira Binti A Jalil, Ramesh Kumar Ayyasamy, Rehan Akbar "The Essential of Sentiment Analysis and Opinion Mining in Social Media" 978-1-5386-8546-4/19/$31.00 ©2019 IEEE