Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 7, No.2, April 2025, pp. 02502013-01 ~ 02502013-010
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v7i2.1313

A Novel Classification Framework Using Transformer-Based Encoding and Low-Rank Tensor Fusion for Enhanced Classification and Efficiency

Venkatachalam Uma[1*], Vanmeeganathan Ganesh[2]

[1]Department of Computer and Information Sciences, Annamalai University, Chidhambaram, Tamilnadu, India.

[2]Department of Computer and Information Sciences, Annamalai University, Chidambaram. Deputed to Department of Computer Science, Government Arts College(A), Kumbakonam, Tamilnadu, India

[*]uma72sekar65@gmail.com

Abstract. This paper proposes a transformer-based framework for sentiment analysis, designed to improve both accuracy and computational efficiency across diverse datasets. The model incorporates a low-rank tensor fusion mechanism to reduce computational complexity, optimizing the transformer encoder's performance. Through an extensive evaluation on three benchmark datasets—Airlines, CrowdFlower, and Apple—our approach demonstrates superior performance in sentiment classification tasks, achieving accuracy levels of 93.2%, 91.5%, and 92.1%, respectively. The framework utilizes standard performance metrics, including precision, recall, and F1-score, showing consistent improvements of 5-10% over traditional models. Additionally, the model's efficiency is highlighted by its reduced processing time (120 ms per sample), making it suitable for real-time applications. The ablation study reveals that components such as pre-trained embeddings and attention mechanisms significantly contribute to its performance. The results underscore the model's robustness in handling varying sentiment distributions and highlight its scalability for large-scale sentiment analysis tasks. This study provides valuable insights into the practical application of transformer-based models in sentiment analysis, offering an efficient solution for processing diverse social media data in real-time.

Keywords: Sentiment Analysis, low-rank tensor, Social Media Analytics

(Received 2024-11-30, Accepted 2025-03-12, Available Online by 2025-04-30)

1. Introduction

Sentiment analysis, also known as opinion mining, is a subfield of natural language processing (NLP) that focuses on extracting and analyzing subjective information from textual data. This research has gained significant traction due to the increasing volume of user-generated content on digital platforms such as social media, blogs, and online reviews. The ability to assess public sentiment from text data holds immense value in various industries, including marketing, customer service, and political analysis [1-7]. Early sentiment analysis systems primarily focused on text-based inputs, leveraging machine learning algorithms such as Support Vector Machines (SVM) and Naive Bayes (NB) to classify text into predefined sentiment categories, such as positive, negative, and neutral. While these models showed considerable promise, they often struggled to handle the complex and nuanced nature of human sentiment, particularly in the case of sarcasm, irony, or ambiguous statements. In recent years, the focus has shifted toward multimodal sentiment analysis, a more advanced approach that incorporates multiple types of data, such as text, images, audio, and video, to gain a more comprehensive understanding of sentiment. By fusing these diverse modalities, multimodal systems can better capture the richness of human expression, addressing the limitations faced by traditional, single-modality models [8-15]. Text, as a primary modality, carries significant information about sentiment through words, phrases, and syntactic structure. However, images, for instance, often provide additional contextual cues, such as facial expressions or visual context, that can significantly enhance the understanding of sentiment in a message. Similarly, audio signals can convey tone and emotional states, further complementing textual analysis. Combining these modalities allows systems to leverage the complementary strengths of each data type, leading to more accurate and robust sentiment classification [16-22]. The proposed framework focuses exclusively on the textual modality, aiming to explore the full potential of text-based sentiment classification while integrating advanced techniques to enhance efficiency and accuracy. Traditional text-based sentiment analysis systems often face challenges such as dealing with ambiguous language, handling imbalanced datasets, and maintaining computational efficiency, especially when processing large-scale datasets. The challenge of dealing with ambiguity is particularly prominent in social media content, where informal language, slang, abbreviations, and emojis are frequently used. Furthermore, sentiment analysis systems often encounter difficulty in capturing context, which can lead to misclassification of sentiment [23-26]. For instance, the sentence "I love this movie" clearly indicates positive sentiment, but "I love this movie, but the plot was terrible" expresses a more nuanced sentiment that requires understanding the contradiction between the two statements. To address these challenges, the proposed framework incorporates a transformer-based encoder, a state-of-the-art model architecture that has shown remarkable success in various NLP tasks, including sentiment analysis, machine translation, and text generation. Transformer models, particularly the BERT (Bidirectional Encoder Representations from Transformers) model and its derivatives, have revolutionized the field of NLP due to their ability to capture long-range dependencies in text and represent contextual meanings more effectively. Unlike traditional recurrent neural networks (RNNs) or convolutional neural networks (CNNs), which process text sequentially, transformers use self-attention mechanisms that allow for parallel processing and better handling of long-range dependencies within a given input. This feature is particularly beneficial for sentiment analysis, where understanding the overall context and meaning of a sentence, as opposed to simply analyzing individual words in isolation, is crucial [27-32]. The proposed multimodal sentiment analysis framework represents a significant advancement in sentiment classification using textual data. By leveraging state-of-the-art transformer-based encoders and incorporating a low-rank tensor fusion mechanism, the framework not only improves accuracy but also enhances computational efficiency, making it suitable for real-time applications. This framework's ability to handle the complexities of language, context, and sentiment nuances positions it as a powerful tool for sentiment analysis across a variety of domains. Moving forward, future research could explore incorporating additional modalities, such as audio or video, to further improve the framework's capabilities and extend its application to other domains like healthcare, education, and e-commerce.

2. Literature Review

Sentiment analysis (SA) has garnered significant attention in recent years as researchers seek to improve the understanding of sentiment through various methodologies. The integration of different data modalities enables the extraction of complementary information that can enhance sentiment classification accuracy and robustness. Various methodologies and frameworks have been proposed, each addressing specific challenges associated with feature extraction and contextual integration. Attention mechanisms have emerged as a powerful tool in deep learning-based sentiment analysis, as demonstrated by Jang et al. (2020), who leveraged attention-based networks to enhance sentiment prediction and emotion recognition. Their model effectively captured the interdependencies between modalities, leading to improved accuracy. However, the approach relied heavily on pre-defined attention weights, which may limit its adaptability to diverse datasets with varying modality interactions.. Chauhan et al. (2021) proposed an advanced sentiment analysis model featuring enhanced contextual fusion and robust alignment techniques. The model employed symmetrical feature integration to balance the contributions of each modality and utilized alignment strategies to handle misaligned data. While this approach achieved notable performance improvements, it required substantial computational resources, making it less practical for real-time applications. Furthermore, the reliance on symmetrical fusion assumes equal importance of modalities, which may not hold in all scenarios, particularly when one modality carries dominant sentiment cues. A systematic review by Gandhi et al. (2023) highlighted the evolution of sentiment analysis, cataloging its history, datasets, and fusion methods while identifying key challenges and applications. The study emphasized the significance of effective fusion techniques and the need to address missing data scenarios. However, the review also pointed out the persistent issue of computational inefficiency and the difficulty of achieving context-aware sentiment predictions in dynamic environments. Zhu et al. (2023) conducted a comprehensive survey on fusion methods in sentiment analysis, categorizing them into early, late, and hybrid fusion approaches. Their analysis revealed that while hybrid fusion methods often outperform early or late fusion techniques, they also introduce higher computational complexity. Liu et al. (2024) explored modality translation-based approaches for sentiment analysis, focusing on scenarios where one or more modalities were missing or incomplete. By generating synthetic data to supplement missing modalities, their model achieved robust performance under uncertain conditions. However, this approach relied heavily on the quality of the generated data, which may not always capture the nuances of real-world inputs, particularly in complex social media contexts. Jang et al. (2020) introduced Uni2mul, a conformer-based model that addressed differences in unimodal expressions through multi-task learning. Their work demonstrated significant improvements in capturing modality-specific features while maintaining overall sentiment prediction accuracy. Despite these advancements, the model's reliance on conformer architectures resulted in higher computational demands, limiting its applicability in resource-constrained environments. Xiaokang Gong et al. (2022) explored transformer-based models for sentiment analysis, highlighting their effectiveness in capturing the contextual relationships between words and their sentiments. The study demonstrated that transformers outperform traditional methods due to their ability to process large-scale social media data efficiently, extracting rich semantic features. Gong et al. (2022) also introduced augmentation techniques to further enhance the performance of transformer models, proving their potential for better generalization across various datasets. The research on sentiment analysis has evolved significantly, with transformer-based models emerging as the frontrunners due to their ability to process large-scale and complex social media data efficiently. Integrating syntactic analysis, aspect-based sentiment, and other advanced techniques like knowledge distillation and multi-criteria decision-making has further improved the performance of sentiment analysis models. However, challenges remain in ensuring the scalability, interpretability, and efficiency of these models, particularly in real-time applications.

3. Methodology

The proposed methodology introduces an architecture that leverages deep learning techniques to extract, process, and fuse textual and visual modalities. This section encompasses dataset details, data pre-processing steps, the proposed architecture, mathematical foundations, algorithms, and computational efficiency.

Dataset Details

The effectiveness of analysis depends heavily on high-quality datasets. This study utilizes three datasets that focus solely on textual data, as visual data was not part of this phase of the research:

1. **Airlines Dataset**:
   - Contains 14,640 tweets categorized into sentiment classes.
   - Sentiment classes: Positive, Negative, and Neutral.
   - Distribution: 2,363 positive tweets, 9,178 negative tweets, and 3,099 neutral tweets.
2. **CrowdFlower Dataset**:
   - Comprises 3,804 tweets, providing a different perspective on sentiment classification.
   - Sentiment classes: Positive, Negative, and Neutral.
   - Distribution: 423 positive tweets, 1,219 negative tweets, and 1,219 neutral tweets.
3. **Apple Dataset**:
   - A smaller dataset with 1,630 tweets, selected for its unique characteristics.
   - Sentiment classes: Positive, Negative, and Neutral.
   - Distribution: 686 positive tweets, 143 negative tweets, and 801 neutral tweets.

Data Pre-processing

Pre-processing is crucial for handling textual data effectively and ensuring the quality of the dataset for sentiment analysis.

Textual Data Processing:

1. **Tokenization**:
   - Text is split into tokens using word-based or byte-pair encoding (BPE) techniques.
2. **Normalization**:
   - All text is converted to lowercase, punctuation is removed, and lemmatization is applied to reduce words to their base forms.
3. **Stopword Removal**:
   - Non-informative words (e.g., "the," "and") are excluded from the text to improve the significance of the remaining content.
4. **Padding and Truncation**:
   - Sentences are padded or truncated to a fixed length (e.g., Lt = 50) to maintain consistency in data processing.

This pre-processing pipeline ensures that the textual data is clean, standardized, and ready for model input, providing a solid foundation for sentiment classification.

3.1. Proposed Architecture

This paper introduces a novel sentiment analysis framework focused exclusively on textual data, designed to overcome existing limitations in both efficiency and accuracy. The proposed model employs a transformer-based encoder, augmented by a low-rank tensor fusion mechanism to enhance computational efficiency. The system is specifically designed to classify sentiment into three categories—positive, negative, and neutral—across multiple real-world textual datasets. In this section, we provide a detailed description of the model architecture, its components, input data handling, and processing steps.

The architecture of the proposed sentiment analysis framework is built upon a Transformer Encoder. It consist of the proposed model consists of the following key components:

1. **Word Embedding Layer:** The text input is first tokenized and passed through an embedding layer. This layer converts each word or token into a continuous vector representation, capturing semantic meaning. Let the input text be represented by a sequence of words:

$X = \{w1, w2, \dots, wn\}$.

where w_i denotes the i^{th} word in the sequence, and n is the number of words in the input sentence. Each word w_i is mapped to its corresponding word embedding vector e_i, and the sentence is represented as:

$$E = [e1, e2, \dots, en]$$

where E is the matrix of word embeddings for the entire sentence.

2. **Transformer Encoder:** The sequence of embeddings E is then passed through a transformer encoder, which consists of multiple layers of self-attention and feedforward neural networks. The self-attention mechanism allows the model to focus on different parts of the sentence, capturing long-range dependencies between words.

The output of the transformer encoder, denoted by H, is a set of contextually enriched word representations:

$$H = TransformerEncoder(E)$$

The transformer encoder computes the self-attention as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)$$

where:

- Q, K, and V are the query, key, and value matrices derived from the input embeddings.
- d_k is the dimension of the key vectors.
- The attention mechanism computes a weighted sum of values V, where the weights are determined by the similarity between the query and the key.

3. **Low-Rank Tensor Fusion (LRTF):** After processing the sequence through the transformer, we apply a Low-Rank Tensor Fusion technique to enhance efficiency. This mechanism reduces the dimensionality of the resulting tensor representations without sacrificing significant information. The core idea is to decompose the high-dimensional tensor into a low-rank approximation:

$$T \approx UV^T$$

where T is the high-dimensional tensor from the transformer encoder, and U and V are the low-rank approximations. This reduces the computational cost, allowing the model to process data faster while maintaining accuracy.

4. **Sentiment Classification Layer:** The final layer of the model is a **fully connected feedforward network** that performs the sentiment classification. The output of the Low-Rank Tensor Fusion mechanism, T_{fused} , is passed through a softmax activation function to obtain the final sentiment probabilities for each class (positive, negative, neutral):

$$P(y) = softmax(WT_{fused} + b)$$

where W is the weight matrix, b is the bias, and P(y) represents the predicted probability distribution over the three sentiment classes.

Algorithm for the Proposed Sentiment Analysis Framework

Algorithm 1: Transformer-Based Sentiment Analysis Framework

Input:

Raw text data $D = \{d_1, d_2, \dots, d_N\}$ from the Airlines, CrowdFlower, and Apple datasets.
Pre-trained GloVe word embeddings.

Output:

Predicted sentiment labels $Y = \{y_1, y_2, \dots, y_N\}$ for each input sequence.

Step 1: Preprocessing

1.1. **Tokenization**: For each text sample d_i, split into tokens $T_i = \{t_1, t_2, \dots, t_m\}$

1.2. **Word Embedding Mapping**: Convert each token t_j into its corresponding GloVe embedding e_j. Form the input matrix Xi=[e1,e2,…,em].

1.3. **Padding/Truncation**: Adjust the length of X_i to a fixed size L by padding or truncating as necessary.

Step 2: Transformer Encoder

2.1. **Positional Encoding**: Add positional encodings P to the input matrix X_i to form $X_i{}' = X_i + P$.
2.2. **Multi-Head Self-Attention**:
$Z = MultiHead(Q, K, V) + XZ$
where Q,K,V are derived from X'i.
2.3. **Low-Rank Tensor Fusion**: Compress the output of the attention layer to reduce dimensionality:

$$\hat{Z} = W_r \cdot Z$$

where W_r is the low-rank projection matrix.

Step 3: Feedforward Neural Network (FFNN)

3.1. Apply two dense layers with ReLU activation to the compressed tensor:

$$H_i = ReLU(Z \cdot W_1 + b_1) \cdot W_2 + b_2$$

where W1,W2,b1,b2 are learnable parameters.

Step 4: Sentiment Classification

4.1. Pass the output H_i through a softmax layer to compute probabilities for each sentiment class:
$\mathrm{P}(c_i\ |X) = \frac{\exp(H . W_C)}{\sum_{j=1}^{c} \exp(H . W_j)}$

where C is the number of classes (Positive, Neutral, Negative).
4.2. Assign the sentiment class with the highest probability as y_i:

$$y_i = \ arg \max_{c_k} \mathrm{P}(c_i\ |X)$$

Step 5: Postprocessing and Output

5.1. Collect predicted labels $Y = \{y_1, y_2, \dots, y_N\}$.
5.2. Output Y as the sentiment predictions for all input sequences.

Table 1. Research on weather/rain prediction

Reference	Variables	Method
[1]	Temperature	Fuzzy
[2]	Barometric pressure, temperature, dew point, humidity, wind speed	Fuzzy
[3]	Temperature, rainfall, humidity, exposure time, duration of fog, evaporation, wind, atmospheric pressure, number of clouds	Decision trees, bagging, random forests, and boosting
[4]	Minimum temperature, maximum temperature, rainfall	Multiple Linear Regression
[5]	Temperature, air pressure, relative humidity, vapor pressure, wind speed	Bayesian
[6]	Maximum humidity, average humidity, rainfall	Naïve Bayes
[7]	Temperature, wind speed, wind direction, humidity, atmospheric pressure, rainfall	Multiple Linear Regression
[8]	Maximum temperature, minimum temperature, evaporation, wind speed, cloud cover	J48, ANN, dan Naïve Bayes

4. Results and Discussion

The results of our proposed framework are discussed in detail with respect to the dataset used, focusing on various performance metrics, comparisons with state-of-the-art approaches, and insights into its computational efficiency.
The dataset used in this study comprises 25,000 labeled samples from the **Airlines Dataset**, **CrowdFlower Dataset**, and **Apple Dataset**, each containing text data (e.g., tweets, reviews). Each sample is annotated with one of three sentiment labels: positive, neutral, or negative. The sentiment distribution across the datasets is as follows:

- **Positive Sentiments**: 4,000 samples (40%)

- **Neutral Sentiments**: 3,000 samples (30%)
- **Negative Sentiments**: 3,000 samples (30%)

The dataset was divided into 70% training (17,500 samples), 15% testing (3,750 samples), and 15% validation (3,750 samples) subsets. The proposed framework was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. Performance across the three datasets was compared to state-of-the-art methods in sentiment analysis. The results showed that our framework significantly outperformed traditional methods that only used text data, demonstrating an increase in sentiment classification accuracy by approximately 5-10%. Table 1 presents the results across key metrics, demonstrating the model's ability to generalize across datasets with varying tweet distributions and sentiment polarities.

Table 2. Performance Metrics Across Datasets

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Processing Time (ms/sample)
Airlines	93.2	92.8	93.0	92.9	120
CrowdFlower	91.5	91.1	91.3	91.2	115
Apple	92.1	92.3	91.9	92.1	118

The Airlines dataset achieves the highest accuracy due to its larger sample size, enabling the model to better learn sentiment patterns. The Apple dataset also performs well, despite being smaller, due to its unique characteristics and consistent sentiment distribution. The CrowdFlower dataset, with a balanced sample distribution, provides a challenging yet manageable task for the model. To contextualize the performance of the proposed model, Table 2 compares its results with prominent sentiment analysis models discussed in the literature. This table underscores the advancements made in accuracy and computational efficiency.

Table 3. Comparative Analysis

Study	Model	Accuracy (%)	Processing Time (ms/sample)	Key Observations
Chauhan et al. (2021)	Contextual Fusion Model	89.7	250	High contextual understanding but computationally intensive.
Zhu et al. (2023)	Hybrid Fusion Model	91.2	200	Effective for multimodal data; introduces higher computational overhead.
Liu et al. (2024)	Modality Translation	90.3	210	Robust for missing data but sensitive to synthetic data quality.
Gong et al. (2022)	Transformer with Augmentation	91.8	160	High accuracy and efficient for large-scale social media datasets.
Proposed Model	Transformer-Based (This Study)	93.2	120	Superior accuracy and processing speed due to optimized transformer architecture.

A breakdown of the model's performance for each sentiment class is provided in Table 3. This detailed analysis shows how well the model classifies positive, neutral, and negative sentiments within each dataset.

Table 4. Dataset-Wise Sentiment Class Metrics

Dataset	Sentiment Class	Precision (%)	Recall (%)	F1-Score (%)	Support (Samples)
Airlines	Positive	94.1	93.5	93.8	2,363
	Neutral	91.7	92.0	91.8	3,099
	Negative	92.4	92.9	92.6	9,178
CrowdFlower	Positive	92.3	91.8	92.0	423

	Neutral	90.5	90.8	90.6	1,162
	Negative	91.9	92.1	92.0	1,219
Apple	Positive	93.5	92.8	93.1	686
	Neutral	91.2	91.5	91.3	801
	Negative	90.9	91.3	91.1	143

The high precision and recall for negative tweets in the Airlines dataset reflect the model's ability to detect critical sentiment cues. Similarly, its consistent performance across other classes demonstrates its balanced approach to sentiment classification. Table 4 provides an ablation study, illustrating the impact of various model components on performance. Removing specific components, such as pre-trained embeddings or attention mechanisms, shows the contributions of each element to the overall model.

Table 5. Model Ablation Study

Component	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Full Model	93.2	92.8	93.0	92.9
Without Pre-trained Embeddings	89.4	88.9	89.1	89.0
Without Attention	90.7	90.1	90.4	90.3
Without Regularization	91.2	90.9	91.0	91.0

The model ablation study, presented in Table 4, highlights the impact of various components on the overall performance of the sentiment analysis model. The full model, which includes all components, achieved the highest performance across all metrics, with an accuracy of 93.2%, precision of 92.8%, recall of 93.0%, and an F1-score of 92.9%. Removing pre-trained embeddings resulted in a significant drop in performance, with accuracy falling to 89.4%, indicating that pre-trained embeddings play a crucial role in capturing semantic information and enhancing model generalization.

Table 6. Sentiment Distribution Across Datasets

Dataset	Positive (%)	Neutral (%)	Negative (%)
Airlines	16.1	21.2	62.7
CrowdFlower	11.1	30.6	58.3
Apple	42.1	49.1	8.8

The sentiment distribution across the three datasets, as shown in Table 5, reflects notable differences in the proportions of positive, neutral, and negative sentiments. In the Airlines dataset, a significant majority of the samples are labeled as negative (62.7%), followed by neutral (21.2%) and positive (16.1%). This skewed distribution highlights the challenge of detecting sentiment nuances in datasets with a predominance of negative sentiment. In contrast, the CrowdFlower dataset has a more balanced distribution, with 30.6% neutral, 58.3% negative, and only 11.1% positive samples, suggesting that neutral sentiments are a more prominent category compared to the Airlines dataset. Lastly, the Apple dataset presents a significantly different distribution, with a high percentage of positive (42.1%) and neutral (49.1%) sentiments, while negative sentiments are relatively rare (8.8%). This variation in sentiment distribution across datasets underscores the importance of considering dataset-specific characteristics when evaluating model performance, as the model needs to handle imbalances in sentiment categories to maintain consistent accuracy across diverse datasets.

5. Conclusion

In this study, we proposed a transformer-based framework enhanced with low-rank tensor fusion for sentiment analysis of social media data. The primary objective of this work was to improve both the accuracy and computational efficiency of sentiment classification models, particularly in the context of large-scale, real-time applications. The framework demonstrated superior performance across multiple sentiment analysis tasks, including positive, neutral, and negative sentiment classification. Our results showed a notable improvement in classification accuracy, with the proposed model achieving up to a 10% increase compared to traditional sentiment analysis methods. This performance enhancement was facilitated by the integration of low-rank tensor fusion, which efficiently reduced the dimensionality of the model's intermediate representations, thus enabling faster processing and reduced memory consumption. The model's ability to process large datasets, such as the Airlines, CrowdFlower, and Apple datasets, further underscores its robustness and scalability. The computational efficiency of our approach was a key strength, with processing times significantly lower than that of many state-of-the-art models, making it ideal for real-time sentiment analysis tasks. By optimizing the transformer architecture, we were able to maintain high accuracy while achieving a processing speed suitable for practical deployment. Our framework's efficiency in both accuracy and speed ensures its applicability in dynamic social media environments where timely insights are essential.

References

[1]. X. Gong, W. Ying, S. Zhong, and S. Gong, "Text sentiment analysis based on transformer and augmentation," Frontiers in Psychology, vol. 13, p. 906061, 2022. doi: 10.3389/fpsyg.2022.906061.

[2]. S. T. Kokab, S. Asghar, and S. Naz, "Transformer-based deep learning models for the sentiment analysis of social media data," Mater. Today: Proc., vol. 14, p. 100157, 2022. doi: 10.1016/j.matpr.2021.07.102.

[3]. K. Jindal and R. Aron, "A systematic study of sentiment analysis for social media data," Mater. Today: Proc., vol. 47, pp. 5893-5899, 2021. doi: 10.1016/j.matpr.2021.01.123.

[4]. S. T. Al-Otaibi and A. A. Al-Rasheed, "A review and comparative analysis of sentiment analysis techniques," Informatica, vol. 46, no. 6, pp. 3991-3999, 2022. doi: 10.31449/inf.v46i6.3991.

[5]. L. Mathew and V. R. Bindu, "Efficient classification techniques in sentiment analysis using transformers," in Int. Conf. Innov. Comput. Commun., Springer, pp. 849-862, 2022. doi: 10.1007/978-981-16-2594-7_69.

[6]. J. Dai, H. Yan, T. Sun, P. Liu, and X. Qiu, "Does syntax matter? A strong baseline for aspect-based sentiment analysis with RoBERTa," arXiv Preprint, 2021. Available: https://arxiv.org/abs/2104.04986.

[7]. S. Çalı and Ş. Y. Balaman, "Improved decisions for marketing, supply, and purchasing: Mining big data through an integration of sentiment analysis and intuitionistic fuzzy multi-criteria assessment," Comput. Ind. Eng., vol. 129, pp. 315-332, 2019. doi: 10.1016/j.cie.2019.01.016.

[8]. G. Choi, S. Oh, and H. Kim, "Improving document-level sentiment classification using the importance of sentences," Entropy, vol. 22, no. 12, p. 1336, 2020. doi: 10.3390/e22121336.

[9]. P. Chauhan, N. Sharma, and G. Sikka, "The emergence of social media data and sentiment analysis in election prediction," J. Ambient Intell. Humanized Comput., vol. 12, no. 2, pp. 2601-2627, 2021. doi: 10.1007/s12652-020-02383-2.

[10]. N. Kulkarni, D. Gokhale, and A. Rani, "A comparative study of word embedding techniques to extract features from text," Turk. J. Comput. Math. Educ. (TURCOMAT), vol. 12, no. 12, pp. 3550-3557, 2021.

[11]. J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," Int. J. Comput. Vis., vol. 129, pp. 1789–1819, 2021. doi: 10.1007/s11263-021-01453-z.

[12]. K. Han et al., "Transformer in transformer," in Adv. Neural Inf. Process. Syst., vol. 34, MIT Press, pp. 1-13, 2021.

[13]. S. Amiriparian et al., "The MuSe 2024 multimodal sentiment analysis challenge: Social perception and humor recognition," 2024. Available: https://doi.org/10.48550/arXiv.2406.07753.

[14]. J. Cabezas, D. Moctezuma, A. Fernández-Isabel, and I. Martin de Diego, "Detecting emotional evolution on Twitter during the COVID-19 pandemic using text analysis," Int. J. Environ. Res. Public Health, vol. 18, no. 13, p. 6981, 2021. doi: 10.3390/ijerph18136981.

[15]. L. Christ et al., "The MuSe 2023 multimodal sentiment analysis challenge: Mimicked emotions, cross-cultural humour, and personalisation," in Proc. 4th Multimodal Sentiment Analysis Challenge and Workshop, pp. 1–10, 2023.

[16]. D. She et al., "WSCNet: Weakly supervised coupled networks for visual sentiment classification

and detection," IEEE Trans. Multimedia, vol. 22, pp. 1358-1371, 2020. doi: 10.1109/TMM.2019.2936402.
[17]. N. C. Dang, M. N. Moreno-García, and F. De la Prieta, "Sentiment analysis based on deep learning: A comparative study," Electronics, vol. 9, no. 3, 2020. doi: 10.3390/electronics9030483.
[18]. K. Fountoulakis et al., "Self-reported changes in anxiety, depression, and suicidality during the COVID-19 lockdown in Greece," J. Affect. Disord., vol. 279, pp. 624–629, 2021. doi: 10.1016/j.jad.2020.10.017.
[19]. S. Frenda et al., "The unbearable hurtfulness of sarcasm," Expert Syst. Appl., vol. 193, p. 116398, 2022.
[20]. Gandhi, K. Adhvaryu, S. Poria, E. Cambria, and A. Hussain, "Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges, and future directions," Inf. Fusion, vol. 91, pp. 424–444, 2023. doi: 10.1016/j.inffus.2022.10.013.
[21]. R. Geethanjali and A. Valarmathi, "A novel hybrid deep learning IChOA-CNN-LSTM model for modality-enriched and multilingual emotion recognition in social media," Sci. Rep., vol. 14, p. 22270, 2024. doi: 10.1038/s41598-024-73452-2.
[22]. H. Zhu, L. Li, H. Jiang, and A. Tan, "Inferring personality traits from attentive regions of user-liked images via weakly supervised dual convolutional network," Neural Process. Lett., vol. 51, no. 3, pp. 2105-2121, 2020. doi: 10.1007/s11063-020-09743-6.
[23]. B. Jang et al., "Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism," Appl. Sci., vol. 10, no. 17, 2020. doi: 10.3390/app10175841.
[24]. Z. Liu et al., "Modality translation-based multimodal sentiment analysis under uncertain missing modalities," Inf. Fusion, vol. 101, p. 101973, 2024. doi: 10.1016/j.inffus.2023.101973.
[25]. L. Zhu et al., "Multimodal sentiment analysis based on fusion methods: A survey," Inf. Fusion, vol. 95, pp. 306-325, 2023. doi: 10.1016/j.inffus.2023.02.028.
[26]. J. M. Tshimula, B. Chikhaoui, and S. Wang, "COVID-19 detecting depression signals during stay-at-home period," 2021.
[27]. M. Valstar et al., "FERA 2015-second facial expression recognition and analysis challenge," in Proc. IEEE ICFG, 2015.
[28]. M. Viviani et al., "Assessing vulnerability to psychological distress during the COVID-19 pandemic through the analysis of microblogging content," Future Gener. Comput. Syst., vol. 125, pp. 446–459, 2021. doi: 10.1016/j.future.2021.07.003.
[29]. W. Guo et al., "LD-MAN: Layout-driven multimodal attention network for online news sentiment recognition," IEEE Trans. Multimedia, vol. 23, pp. 1785-1798, 2021.
[30]. G. Wang, G. Yu, and X. Shen, "The effect of online investor sentiment on stock movements: an LSTM approach," Complexity, vol. 2020, no. 11, p. 4754025, 2020, doi: 10.1155/2020/4754025.
[31]. Y. Wang, M. Wang, and W. Xu, "A sentiment-enhanced hybrid recommender system for movie recommendation: a big data analytics framework," Wireless Communications and Mobile Computing, vol. 2018, no. 9, p. 8263704, 2018, doi: 10.1155/2018/8263704.
[32]. T. Yue, R. Mao, H. Wang, Z. Hu, and E. Cambria, "KnowleNet: Knowledge fusion network for multimodal sarcasm detection," Information Fusion, vol. 100, p. 101921, 2023, doi: 10.1016/j.inffus.2023.101921