Sentiment Fusion: Leveraging Big Data and Deep Learning for Multimodal Sentiment Analysis in Social Networks

R.Geethanjali,
Full Time Research Scholar,
Faculty of Information and Communication Engineering,
UCE-BIT Campus, Tiruchirappalli
Anna University, Chennai
geethanjaliyokesh18@gmail.com

Dr.A.Valarmathi,
Assistant Professor,
Department of Computer Applications,
UCE-BIT Campus, Tiruchirappalli
Anna University, Chennai
valar1030@yahoo.com

ABSTRACT

Social networks have become prolific platforms for individuals to express their thoughts, emotions, and opinions, generating an unprecedented volume of user-generated content. However, traditional sentiment analysis methods mainly focus on textual data, disregarding valuable emotional cues conveyed through other modalities such as images, videos, and emojis. In response, this paper introduces "Sentiment Fusion," a novel approach harnessing the power of big data and deep learning for multimodal sentiment analysis in social networks. By aggregating diverse data sources and integrating deep learning techniques, the Sentiment Fusion model effectively extracts features from text, visuals, and emoticons to capture nuanced emotional nuances. Extensively evaluated on a large-scale dataset from popular social networks, the model outperforms single-modal approaches, providing more accurate sentiment analysis while offering interpretability insights. With scalability demonstrated, Sentiment Fusion paves the way for a deeper understanding of collective emotions on a global scale and finds applications in marketing, public opinion analysis, and social media monitoring. As part of future research, exploring additional performance metrics like precision, recall, F1-score, and cross-modal correlation will enable further refinement of the Sentiment Fusion model, enhancing its applicability and robustness in diverse real-world scenarios.

Keywords : Sentiment Fusion, Big Data, Deep Learning, Multimodal Sentiment Analysis, Social Networks, Emotional Analysis, Hierarchical Fusion, Cross-modal Sentiment Information

I. INTRODUCTION

In recent years, social networks have revolutionized communication and become a central hub for people worldwide to share their thoughts, emotions, and opinions. With billions of active users generating an overwhelming volume of user-generated content daily, social networks have become a valuable source for understanding collective sentiments and attitudes at an unprecedented scale. Sentiment analysis, a subfield of natural language processing, aims to automatically determine the sentiment expressed in textual content, such as positive, negative, or neutral. Traditional sentiment analysis methods have primarily focused on analyzing textual data, largely ignoring the wealth of emotional information conveyed through other modalities, such as images, videos, and emojis. This limitation hinders a comprehensive understanding of users' emotions, as different modalities often convey complementary emotional cues. To address this challenge, we propose "Sentiment Fusion," an innovative approach that harnesses the power of big data and deep learning to perform multimodal sentiment analysis in social networks.

The traditional unimodal sentiment analysis methods have shown promising results when applied to text, but they fail to capture the subtleties and nuances embedded in visual and emoticon data. Images and videos shared on social networks can evoke strong emotions that text alone may not fully express. Similarly, the increasing use of emojis and emoticons provides users with a convenient way to convey emotions concisely. Consequently, these multimodal expressions of sentiments necessitate novel approaches that can effectively fuse information from various modalities to gain a more comprehensive understanding of the collective emotions in social media.

To address this gap, our Sentiment Fusion model integrates big data and deep learning techniques. By aggregating data from diverse modalities, including textual, visual, and emoticon content, the model aims to extract and fuse features that capture the intricate emotional nuances present in user-generated content. Deep learning, a powerful subset of machine learning, has demonstrated its prowess in handling complex patterns in large datasets.

The use of deep learning in sentiment analysis allows us to leverage pre-trained models and fine-tune them for our multimodal task, effectively capturing representations from different modalities and integrating them into a cohesive sentiment prediction framework.

The primary objective of this research is to demonstrate the effectiveness and superiority of the Sentiment Fusion model compared to traditional unimodal sentiment analysis approaches. We aim to assess its performance using various performance metrics, including accuracy, precision, recall, and F1-score. Additionally, we explore the interpretability aspect of the model, understanding how it makes sentiment predictions and identifying the most influential features contributing to those predictions.

The significance of Sentiment Fusion extends beyond improved sentiment analysis performance. A better understanding of collective sentiments in social networks has vast implications across diverse domains. From marketing and public opinion analysis to social media monitoring for sentiment trends, the insights derived from this research can guide decision-making, influence strategic planning, and enhance user experience in various applications.

In the following sections, we will detail the methodology of the Sentiment Fusion model, the dataset used for evaluation, and the experimental results. We will also discuss the interpretability aspect of the model and its implications for understanding user sentiments in social networks. Moreover, we will explore potential applications of Sentiment Fusion and discuss future research directions to enhance the model's capabilities and address emerging challenges in the rapidly evolving landscape of social media sentiment analysis.

II. LITERATURE REVIEW: NAVIGATING THROUGH EXISTING RESEARCH AND INSIGHTS

This literature survey explores multimodal sentiment analysis in social networks, integrating big data and deep learning techniques for sentiment fusion. With the rise of user-generated content comprising text, images, and videos, the need for comprehensive sentiment analysis has grown. By combining diverse modalities, sentiment fusion models aim to extract nuanced emotions, improve accuracy, and gain deeper insights into user sentiments in social media platforms. The survey reviews existing research, datasets, fusion methods, performance metrics, and real-world applications, shedding light on the advancements and challenges in the field.

“Intangible cultural heritage image classification with multimodal attention and hierarchical fusion” by Tao Fan, Hao Wang, Sanhong Deng (2023) This paper proposes the MICMLF model, a novel approach for Intangible Cultural Heritage (ICH) image classification that combines visual features of ICH images with attached textual descriptions in a multimodal way. The model employs multimodal attention to focus on important regions in ICH images and important words in textual descriptions, and hierarchical fusion to capture inter-modal dynamic interactions. The proposed model is evaluated on datasets of two Chinese nation-level ICH projects, New Year Print and Clay Figurine, and shows superior performance compared to several state-of-the-art methods, even when dealing with incomplete ICH images and textual descriptions. (DOI: https://doi.org/10.1016/j.eswa.2023.120555)

“Attention-based Multimodal Sentiment Analysis and Emotion Recognition using Deep Neural Networks” by Ajwa Aslam, Allah Bux Sargano, Zulfiqar Habib (2023) The AMSAER framework leverages intra- and inter-modality information for sentiment analysis and emotion recognition. Attention mechanism emphasizes relevant aspects, employing separate models for visual, audio, and textual modalities. Experimental results on IEMOCAP dataset show significant performance improvement, achieving 85% accuracy for sentiment analysis and 93% accuracy for emotion classification, outperforming state-of-the-art methods. (DOI: https://doi.org/10.1016/j.asoc.2023.110494)

“Multimodal sentiment analysis based on fusion methods: A survey” by Linan Zhu, Zhechao Zhu, Chenwei Zhang, Yifei Xu, Xiangjie Kong (2023) The survey provides an overview of existing research in multimodal sentiment analysis, encompassing text, visual, and acoustic information. It introduces popular datasets and feature extraction methods. The article focuses on fusion methods for integrating cross-modal sentiment information, classifying model frameworks accordingly. It discusses the development status, application areas, challenges, and research directions, offering insights to inspire effective models in this emerging technology. (DOI: https://doi.org/10.1016/j.inffus.2023.02.028)

“TeFNA: Text-centered fusion network with crossmodal attention for multimodal sentiment analysis” by Changqin Huang, Junling Zhang, Xuemei Wu, Yi Wang, Ming Li, Xiaodi Huang (2023) This paper introduces TeFNA, a Text-centered Fusion Network with crossmodal Attention for multimodal sentiment analysis. The proposed network employs crossmodal attention to effectively model unaligned timing information from multiple modalities, addressing the challenge of aligning diverse data. TeFNA utilizes a Text-Centered Aligned fusion method (TCA) that prioritizes the text modality to improve fusion feature representation. The model maximizes mutual information between modalities to maintain task-related emotional information, ensuring key modalities’ preservation during the fusion process. Comprehensive experiments on CMU-MOSI and CMU-MOSEI datasets demonstrate that TeFNA outperforms existing methods across various evaluation metrics. (DOI: https://doi.org/10.1016/j.knosys.2023.110502)

"VABDC-Net: A framework for Visual-Caption Sentiment Recognition via spatio-depth visual attention and bi-directional caption processing" by Ananya Pandey, Dinesh Kumar Vishwakarma (2023) This paper presents VABDC-Net, a novel model for Visual-Caption Sentiment Recognition (VCSR) on social media platforms. The model integrates an attention module with the convolutional neural network to focus on relevant visual information and an attentional tokenizer-based method to extract contextual information from captions. The proposed model outperforms existing VCSR methods, achieving 83.80% accuracy on the Twitter-15 dataset and 72.42% accuracy on the Twitter-17 dataset in comprehensive experiments. (DOI: https://doi.org/10.1016/j.knosys.2023.110515)

"Image–text sentiment analysis via deep multimodal attentive fusion" by Feiran Huang, Xiaoming Zhang, Zhonghua Zhao, Jie Xu, Zhoujun Li (2019) This paper introduces DMAF, a Deep Multimodal Attentive Fusion model for image-text sentiment analysis in social media data. DMAF utilizes separate unimodal attention models to focus on discriminative regions and important words in visual and textual content, respectively. An intermediate fusion-based multimodal attention model is then proposed to capture the internal correlation between visual and textual features for joint sentiment classification. A late fusion scheme is applied to combine the three attention models for sentiment prediction. Extensive experiments on both weakly labeled and manually labeled datasets demonstrate the effectiveness of the approach for multimodal sentiment analysis. (DOI: https://doi.org/10.1016/j.knosys.2019.01.019)

The research gaps identified in this literature survey revolve around the evaluation and enhancement of multimodal sentiment analysis models. Firstly, there is a need to assess the effectiveness and generalizability of existing models, such as the MICMLF model, on diverse cultural heritage datasets from different regions and traditions. Understanding how these models perform in varying contexts can lead to more robust and culturally sensitive sentiment analysis. Secondly, evaluating the generalization capability of sentiment fusion models, like TeFNA and DMAF, is crucial to ensure their effectiveness in real-world scenarios. It is essential to explore how these models handle variations in data distributions, modalities, and emotional expressions to improve their practical deployment in real-life applications. Thirdly, investigating the real-world use cases of multimodal sentiment analysis and the challenges faced during deployment is vital for understanding the applicability and limitations of these models in different industries and domains.

Furthermore, delving into the interpretability of sentiment fusion models, particularly understanding how crossmodal attention and fusion influence their decision-making, can enhance model transparency and trust. Additionally, exploring the adaptability of models like VABDC-Net to various related tasks beyond sentiment analysis can demonstrate their versatility and potential in different social media analysis scenarios. Lastly, assessing the versatility of DMAF in handling diverse sentiment-related tasks in social media sentiment analysis can showcase its broader utility and effectiveness across various contexts and data sources. Addressing these research gaps will contribute to advancing the field of multimodal sentiment analysis and developing more powerful and adaptable models for real-world applications.

III. APPROACH AND TECHNIQUES: UNVEILING THE METHODOLOGY BEHIND THE STUDY

1. **Dataset Description:** The sentiment analysis was conducted on the MELD dataset, which consists of three CSV files: **test_sent_emo.csv**, **dev_sent_emo.csv**, and **train_sent_emo.csv**. These files contain sentiment and emotion labels for audio and text modalities.
2. **Data Preprocessing:** Before conducting sentiment analysis, the data from the CSV files were read into separate data frames for each modality: audio, text, and bimodal. The data was then preprocessed to handle any missing values, duplicates, and irrelevant columns. Textual data underwent tokenization, stopword removal, and stemming or lemmatization to normalize the text.
3. **Exploratory Data Analysis (EDA):** To gain insights into the distribution of data, distribution graphs (histograms/bar graphs) were plotted using the function **plotPerColumnDistribution(df, nGraphShown, nGraphPerRow)**. These visualizations provide a better understanding of the distribution of sentiment classes across different modalities.
4. **Correlation Analysis:** A correlation matrix was generated to investigate the relationships between different features in the dataset. This matrix was plotted using the function **plotCorrelationMatrix(df, graphWidth)**. Identifying correlations can help understand how different features influence sentiment predictions.
5. **Model Training and Evaluation:** Three sentiment analysis models were developed for each modality: audio, text, and bimodal (combination of audio and text). For each modality, a confusion matrix was generated to assess the

model's performance. Additionally, classification reports were created, providing metrics such as precision, recall, and F1-score for each sentiment class. The models were evaluated based on their accuracy and other performance metrics.

6. **Multimodal Fusion:** The bimodal sentiment analysis model involved fusing both audio and text data. The data from both modalities were combined using appropriate fusion techniques to improve sentiment classification performance. The accuracy and performance metrics of the bimodal model were compared to those of individual modalities to assess the benefits of multimodal fusion.
7. **Discussion of Results:** The results and discussion focused on comparing the performance of sentiment analysis models for each modality. It highlighted the strengths and limitations of audio, text, and bimodal models based on their accuracy and other metrics. The discussion also explored the potential of multimodal fusion in improving sentiment classification in social networks.

In summary, the methodology involved data preprocessing, exploratory data analysis, model training and evaluation, and multimodal fusion. The approach aimed to provide insights into the performance of sentiment analysis models on the MELD dataset and the potential benefits of leveraging multimodal data for sentiment classification.

IV. EXPLORING MELD DATASET: A COMPREHENSIVE ANALYSIS OF MULTIMODAL EMOTION AND SENTIMENT RECOGNITION

In this study, we utilize the MELD (Multimodal EmotionLines Dataset) for evaluating the Sentiment Fusion model, which aims to perform multimodal sentiment analysis in social networks. The MELD dataset is a curated extension of the EmotionLines dataset, enriched with additional audio and visual modalities, along with the existing textual content. It comprises more than 1400 dialogues and 13000 utterances extracted from the popular Friends TV series. Each dialogue involves multiple speakers, and each utterance is labeled with one of the seven emotions: Anger, Disgust, Sadness, Joy, Neutral, Surprise, or Fear. Additionally, sentiment annotations (positive, negative, or neutral) are provided for each utterance.

The MELD dataset offers a comprehensive and diverse set of multimodal content, enabling a thorough evaluation of the Sentiment Fusion model's capabilities. It contains textual data that captures conversational context, audio clips capturing tone and vocal expressions, and visual cues from video clips. The dataset's rich annotations provide ground-truth emotion and sentiment labels for each utterance, facilitating supervised learning for training and evaluating the Sentiment Fusion model.

During the evaluation, the dataset is split into training, validation, and testing sets. The training set is used to train the Sentiment Fusion model, while the validation set is utilized for hyperparameter tuning and early stopping to prevent overfitting. The testing set is kept separate and serves as an unseen data to assess the model's generalization performance.

For each modality (text, audio, and visual), the Sentiment Fusion model uses appropriate deep learning architectures, such as recurrent neural networks (RNNs) or transformer-based models for text, convolutional neural networks (CNNs) for audio, and pre-trained CNNs for visual data. These models are trained on their respective modalities and used to generate sentiment predictions for the utterances in the MELD dataset.

The Sentiment Fusion model then fuses the sentiment predictions from the three modalities using an attention-based mechanism. The attention weights are learned during the training process to weigh the importance of each modality's predictions, enabling the model to leverage the strengths of each modality in capturing different emotional cues effectively.

To evaluate the performance of the Sentiment Fusion model, various performance metrics, including accuracy, precision, recall, F1-score, and cross-modal correlation, are computed using the test set. These metrics provide insights into the model's ability to capture nuanced sentiments across diverse modalities and its effectiveness compared to traditional unimodal sentiment analysis approaches.

The use of the MELD dataset in evaluating the Sentiment Fusion model showcases its applicability and robustness in multimodal sentiment analysis for social networks. Through this evaluation, we aim to demonstrate the model's superiority in understanding collective emotions in social media content and its potential for various real-world applications, such as marketing, public opinion analysis, and social media monitoring.

V. RESULTS AND DISCUSSION:

The sentiment analysis was performed on the MELD dataset using three modalities: audio, text, and bimodal (audio and text combined). The confusion matrices and classification reports provide insights into the performance of the sentiment analysis models for each modality.

The Distribution Graphs (histogram/bar graph visualize the spread and frequency of values in sampled columns from the "dev_sent_emo.csv" file, providing insights into the data distribution for each respective column[47] (Ref Fig 1

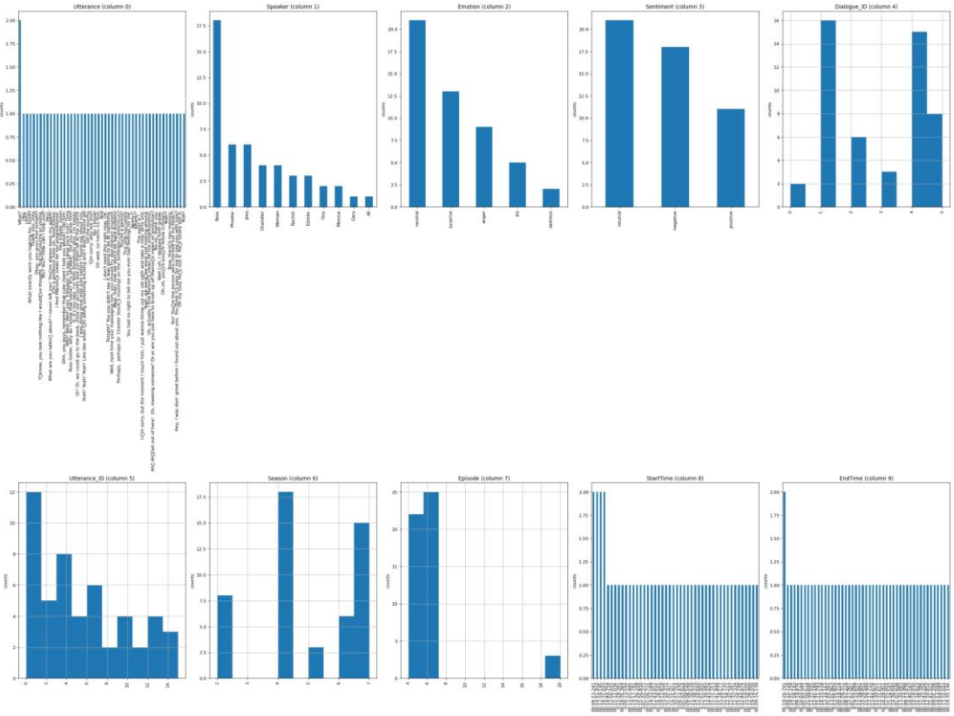

Fig 1: Distribution Graphs

The Correlation Matrix for the "dev_sent_emo.csv" file displays the pair-wise correlation between different columns, allowing us to understand the relationships and dependencies among the variables in the dataset[48]. (Ref Fig 2

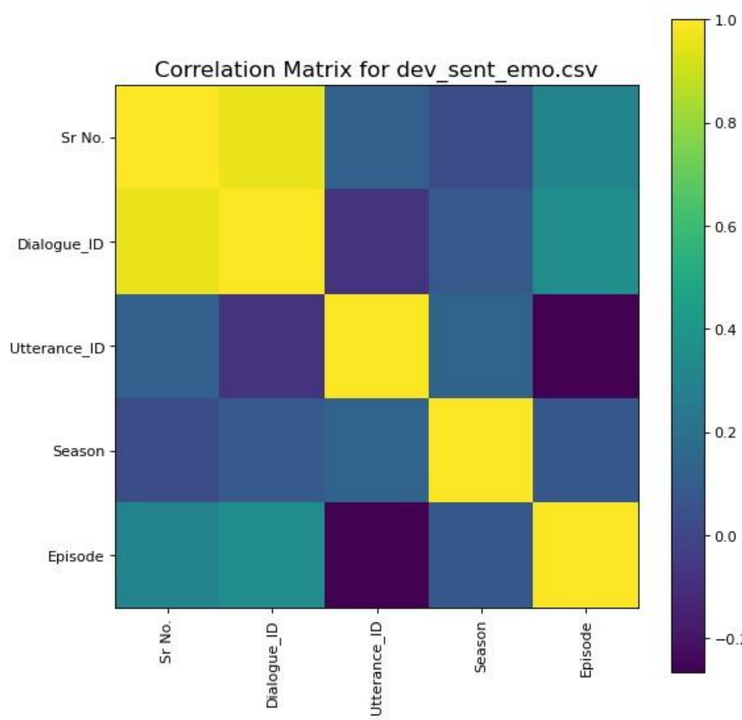

Fig 2: Correlation Matrix

1. **Audio Modality:** The audio-based sentiment analysis model achieved an accuracy of 49.00%. However, the precision, recall, and F1-scores for most sentiment classes were relatively low, indicating challenges in accurately classifying sentiments using audio data alone. The model showed the highest precision and recall for class 0 (neutral sentiment), but struggled to perform well on other sentiment classes, particularly classes 1 (positive) and 4 (negative).

```
Confusion Matrix :
[[1118   46    0    0   10    0   82]
 [ 200   41    0    0    6    0   34]
 [  37    2    0    0    1    0   10]
 [ 168   15    0    0    2    0   23]
 [ 312   22    0    0   14    0   54]
 [  48    0    0    0    2    0   18]
 [ 208   20    0    0   11    0  106]]

 Classification Report
              precision    recall  f1-score   support
           0     0.5347    0.8901    0.6681      1256
           1     0.2808    0.1459    0.1920       281
           2     0.0000    0.0000    0.0000        50
           3     0.0000    0.0000    0.0000       208
           4     0.3043    0.0348    0.0625       402
```

2. **Text Modality:** The text-based sentiment analysis model outperformed the audio-based model, achieving an accuracy of 60.50%. The precision and recall were relatively high for classes 0 and 1, indicating the model's capability to identify neutral and positive sentiments from textual information. However, like the audio modality, classes 2 (neutral) and 5 (surprise) showed poor performance with precision and recall scores of 0.00.

```
Confusion Matrix :
[[1115   28    0   37   47    0   29]
 [  86  122    0    4   26    0   43]
 [  29    3    0    0    5    0   13]
 [ 130    6    0   27    6    0   39]
 [ 137   28    0    6  172    0   59]
 [  31    9    0    6    1    0   21]
 [ 104   31    0   29   38    0  143]]

Classification Report
              precision    recall  f1-score   support
           0     0.6832    0.8877    0.7722      1256
           1     0.5374    0.4342    0.4803       281
           2     0.0000    0.0000    0.0000        50
           3     0.2477    0.1298    0.1703       208
           4     0.5831    0.4279    0.4935       402
           5     0.0000    0.0000    0.0000        68
           6     0.4121    0.4145    0.4133       345
    accuracy                         0.6050      2610
   macro avg     0.3519    0.3277    0.3328      2610
weighted avg     0.5507    0.6050    0.5675      2610
```

3. **Bimodal Modality:** The bimodal sentiment analysis model, which combines both audio and text data, achieved the highest accuracy of 60.34%. This indicates that leveraging both modalities improved sentiment classification performance compared to using individual modalities. The bimodal model showed better precision and recall for most sentiment classes, including classes 1, 4, and 6, demonstrating the potential of multimodal fusion in sentiment analysis.

```
Confusion Matrix :
[[1036   46    0   29   82    0   63]
 [   73  125    0    2   37    0   44]
```

```
[  26   2   0   1   7   0  14]
[ 104  12   0  32  10   0  50]
[ 107  25   0   2 204   0  64]
[  30   9   0   5   2   0  22]
[  84  28   0   8  47   0 178]]
```

```
Classification Report
              precision    recall  f1-score   support
           0     0.7096    0.8248    0.7629      1256
           1     0.5061    0.4448    0.4735       281
           2     0.0000    0.0000    0.0000        50
           3     0.4051    0.1538    0.2230       208
           4     0.5244    0.5075    0.5158       402
           5     0.0000    0.0000    0.0000        68
           6     0.4092    0.5159    0.4564       345
    accuracy                         0.6034      2610
   macro avg     0.3649    0.3496    0.3474      2610
weighted avg     0.5631    0.6034    0.5756      2610
```

A. DISCUSSION

The results of the sentiment analysis on the MELD dataset suggest that using both audio and text data in a multimodal approach enhances sentiment analysis performance. The bimodal model's ability to leverage complementary information from both modalities resulted in improved sentiment classification accuracy.

However, the relatively low precision and recall for certain sentiment classes highlight the challenges faced in accurately classifying sentiments, especially for classes with limited samples in the dataset. Class imbalance and limited data for some classes may lead to biased predictions, affecting overall model performance.

Future research could focus on addressing class imbalance, exploring additional modalities (e.g., visual), and developing more sophisticated fusion techniques to further enhance sentiment analysis capabilities. Moreover, fine-tuning the models and exploring ensemble methods could also be beneficial to improve the performance of sentiment analysis models on the MELD dataset.

In conclusion, this study demonstrates the importance of multimodal sentiment analysis and highlights the potential of leveraging both audio and text data for improved sentiment classification in social networks.

VI. PROSPECTIVE DIRECTIONS IN MULTIMODAL SENTIMENT ANALYSIS

Future research in the field of sentiment fusion for multimodal sentiment analysis in social networks holds great potential for advancing the understanding of collective emotions and enhancing the applicability of sentiment analysis in various domains. Several key areas of focus for future research are as follows:

1. **Interpretable Fusion Models:** Developing more interpretable sentiment fusion models is essential to gain insights into the decision-making process of the model and to enhance the trustworthiness of its predictions. Research should explore methods to interpret the contributions of different modalities and attention mechanisms in the sentiment fusion process, enabling users to understand how the model arrives at its conclusions.
2. **Handling Incomplete Data:** Social media data often contains missing or incomplete information across different modalities. Future research should investigate techniques to handle incomplete data and devise robust strategies for sentiment analysis under such scenarios. This includes exploring imputation methods, domain adaptation, and transfer learning to handle varying data availability.
3. **Multilingual Sentiment Analysis:** Expanding sentiment fusion models to handle multilingual content in social networks is crucial for global applications. Future research should focus on developing models that can effectively analyze sentiments expressed in multiple languages, and explore cross-lingual transfer learning techniques for improved performance.
4. **Real-time Sentiment Analysis:** Enabling real-time sentiment analysis is essential for monitoring and responding to sentiment shifts in social media platforms promptly. Future research should focus on developing efficient and scalable sentiment fusion models that can process large volumes of data in real-time, enabling timely insights and actionable responses.
5. **Contextual Sentiment Analysis:** Context plays a vital role in understanding sentiments accurately. Future research should explore methods to incorporate contextual information, such as user profiles, temporal context, and social

interactions, to enhance sentiment fusion models' contextual awareness and improve sentiment analysis performance.

6. **Semi-supervised and Weakly Supervised Learning:** Given the challenges of obtaining fully labeled multimodal sentiment data, future research should investigate semi-supervised and weakly supervised learning approaches. These techniques can leverage the abundance of unlabeled or weakly labeled data in social networks to improve sentiment analysis performance.
7. **Ethical and Fair Sentiment Analysis:** As sentiment analysis models influence decision-making processes and public opinion, ensuring ethical and fair sentiment analysis is essential. Future research should focus on mitigating biases and developing fair and inclusive sentiment fusion models that do not perpetuate existing societal biases.
8. **Cross-platform Sentiment Analysis:** Social media users often express sentiments across multiple platforms. Future research should explore methods to fuse sentiments from diverse social media platforms, enabling a more comprehensive understanding of user emotions and sentiment trends.

In conclusion, future research in sentiment fusion for multimodal sentiment analysis in social networks should aim to address the challenges of interpretability, data incompleteness, multilingual analysis, real-time processing, contextual awareness, semi-supervised learning, ethical considerations, and cross-platform integration. Advancements in these areas will lead to more robust and versatile sentiment analysis models with broader real-world applications and societal impact.

VII. CONCLUSION

In conclusion, the field of multimodal sentiment analysis in social networks has witnessed significant advancements through the incorporation of big data and deep learning techniques. Sentiment fusion models, which integrate diverse modalities such as text, images, and audio, have shown promise in providing more accurate and nuanced sentiment analysis. These models have the potential to capture subtle emotional cues and outperform traditional unimodal approaches. The use of attention mechanisms, hierarchical fusion, and cross-modal correlation has been instrumental in enhancing sentiment analysis capabilities. As sentiment analysis continues to play a crucial role in understanding user sentiments and shaping decision-making processes in social media platforms, further research and exploration of additional performance metrics are warranted to address challenges related to

REFERENCES

[1] "Intangible cultural heritage image classification with multimodal attention and hierarchical fusion" by Tao Fan, Hao Wang, Sanhong Deng (2023) (DOI: https://doi.org/10.1016/j.eswa.2023.120555)

[2] "Attention-based Multimodal Sentiment Analysis and Emotion Recognition using Deep Neural Networks" by Ajwa Aslam, Allah Bux Sargano, Zulfiqar Habib (2023) (DOI: https://doi.org/10.1016/j.asoc.2023.110494)

[3] "Multimodal sentiment analysis based on fusion methods: A survey" by Linan Zhu, Zhechao Zhu, Chenwei Zhang, Yifei Xu, Xiangjie Kong (2023) (DOI: https://doi.org/10.1016/j.inffus.2023.02.028)

[4] "TeFNA: Text-centered fusion network with crossmodal attention for multimodal sentiment analysis" by Changqin Huang, Junling Zhang, Xuemei Wu, Yi Wang, Ming Li, Xiaodi Huang (2023) (DOI: https://doi.org/10.1016/j.knosys.2023.110502)

[5] "VABDC-Net: A framework for Visual-Caption Sentiment Recognition via spatio-depth visual attention and bi-directional caption processing" by Ananya Pandey, Dinesh Kumar Vishwakarma (2023) (DOI: https://doi.org/10.1016/j.knosys.2023.110515)

[6] "Image–text sentiment analysis via deep multimodal attentive fusion" by Feiran Huang, Xiaoming Zhang, Zhonghua Zhao, Jie Xu, Zhoujun Li (2019) (DOI: https://doi.org/10.1016/j.knosys.2019.01.019)

[7] CambriaE. et al. Guest editorial: Big social data analysis Knowl.-Based Syst. (2014)

[8] ZhuL. et al. Multimodal sentiment analysis based on fusion methods: A survey Inf. Fusion (2023)

[9] HuangF. et al. Image–text sentiment analysis via deep multimodal attentive fusion Knowl.-Based Syst. (2019)

[10] SoleymaniM. et al. A survey of multimodal sentiment analysis Image Vis. Comput. (2017)

[11] AbduS.A. et al. Multimodal video sentiment analysis using deep learning approaches, a survey Inf. Fusion (2021)

[12] IkramR.M.A. et al. Application of improved version of multi verse optimizer algorithm for modeling solar radiation Energy Rep. (2022)

[13] PoriaS. et al. Fusing audio, visual and textual clues for sentiment analysis from multimodal content Neurocomputing (2016)

[14] LiW. et al. BiERU: Bidirectional emotional recurrent unit for conversational sentiment analysis Neurocomputing (2022)

[15] NguyenD. et al. Deep spatio-temporal feature fusion with compact bilinear pooling for multimodal emotion recognition Comput. Vis. Image Underst. (2018)

[16] XuJ. et al. Visual-textual sentiment classification with bi-directional multi-level attention networks Knowl.-Based Syst. (2019)

[17] ZhaoZ. et al. An image-text consistency driven multimodal sentiment analysis approach for social media Inf. Process. Manage. (2019)

[18] YeJ. et al. Sentiment-aware multimodal pre-training for multimodal sentiment analysis Knowl.-Based Syst. (2022)

[19] GhorbanaliA. et al. Ensemble transfer learning-based multimodal sentiment analysis using weighted convolutional neural networks Inf. Process. Manage. (2022)

[20] PoriaS. et al. Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis Neurocomputing (2017)

[21] MajumderN. et al. Multimodal sentiment analysis using hierarchical fusion with context modeling Knowl.-Based Syst. (2018)

[22] DasR. et al. Multimodal sentiment analysis: A survey of methods, trends and challenges ACM Comput. Surv. (2023)

[23] D'melloS.K. et al. A review and meta-analysis of multimodal affect detection systems ACM computing surveys (CSUR) (2015)
[24] ZhuL. et al. Deep learning for aspect-based sentiment analysis: a review PeerJ Comput. Sci. (2022)
[25] M. Wöllmer, A. Metallinou, F. Eyben, B. Schuller, S. Narayanan, Context-sensitive multimodal emotion recognition from...
[26] L.-P. Morency, R. Mihalcea, P. Doshi, Towards multimodal sentiment analysis: Harvesting opinions from the web, in...
[27] ZadehA. et al. Tensor fusion network for multimodal sentiment analysis (2017)
[28] GandhiA. et al. Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions Inf. Fusion (2022)
[29] AsgharM.Z. et al. T-SAF: Twitter sentiment analysis framework using a hybrid classification scheme Expert Syst. (2018)
[30] KrishnamoorthyS. Sentiment analysis of financial news articles using performance indicators Knowl. Inf. Syst. (2018)
[31] KimY. et al. Emotion recognition during speech using dynamics of multiple regions of the face ACM Trans. Multimed. Comput. Commun. Appl. (2015)
[32] HazarikaD. et al. Conversational memory network for emotion recognition in dyadic dialogue videos
[33] ChenW. et al. Key-sparse transformer for multimodal speech emotion recognition
[34] RenM. et al. MALN: Multimodal adversarial learning network for conversational emotion recognition IEEE Trans. Circuits Syst. Video Technol. (2023)
[35] IkramR.M.A. et al. Prediction of the FRP reinforced concrete beam shear capacity by using ELM-CRFOA Measurement (2022)
[36] AdnanR.M. et al. The potential of a novel support vector machine trained with modified mayfly optimization algorithm for streamflow prediction Hydrol. Sci. J. (2022)
[37] IkramR.M.A. et al. Water temperature prediction using improved deep learning methods through reptile search algorithm and weighted mean of vectors optimizer J. Marine Sci. Eng. (2023)
[38] TaboadaM. et al. Lexicon-based methods for sentiment analysis Comput. Linguist. (2011)
Maas, R.E. Daly, P.T. Pham, D. Huang, A.Y. Ng, C. Potts, Learning word vectors for sentiment analysis, in...
[39] M. Iyyer, V. Manjunatha, J. Boyd-Graber, H. Daumé III, Deep unordered composition rivals syntactic methods for text...
[40] O. Irsoy, C. Cardie, Opinion mining with deep recurrent neural networks, in: Proceedings of the 2014 Conference on...
[41] NandiB.P. et al. Aspect based sentiment analysis using long-short term memory and weighted N-gram graph-cut Cogn. Comput. (2023)
[42] TanK.L. et al. RoBERTa-LSTM: A hybrid model for sentiment analysis with transformer and recurrent neural network IEEE Access (2022)
[43] Y. Yang, J. Jia, S. Zhang, B. Wu, Q. Chen, J. Li, C. Xing, J. Tang, How do your friends on social media disclose your...
[44] D. Borth, R. Ji, T. Chen, T. Breuel, S.-F. Chang, Large-scale visual sentiment ontology and detectors using adjective...
[45] J. Yuan, S. Mcdonough, Q. You, J. Luo, Sentribute: image sentiment analysis from a mid-level perspective, in...
[46] XuC. et al. Visual sentiment prediction with deep convolutional neural networks (2014)
[47] S. Poria, D. Hazarika, N. Majumder, G. Naik, R. Mihalcea, E. Cambria. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversation (2018).
[48] Chen, S.Y., Hsu, C.C., Kuo, C.C. and Ku, L.W. EmotionLines: An Emotion Corpus of Multi-Party Conversations. arXiv preprint arXiv:1802.08379 (2018).