An Overview of Multimodal Sentiment Analysis Research: Opportunities and Difficulties

Mohammad Aman Ullah, Md. Monirul Islam
Dept. of Computer Science and Engineering
International Islamic University Chittagong
Chittagong-4203, Bangladesh
ullah047@yahoo.com, monirliton@yahoo.com

Norhidayah Binti Azman, Zulkifly Mohd Zaki
Faculty of Science & Technology
University Sains Islam Malaysia (USIM)
Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia
dayah@usim.edu.my, zulkifly@usim.edu.my

Abstract-**The scatter form of multimedia data such as text, image, audio, and video posted regularly in the social media may contai n useful information for the organizations. But, this information s hould be derived with the use of some form of analysis known as Multimodal Sentiment Analysis (MSA). But, there is a lack of pro per analytic tools for such analysis. This paper presents a thorou gh overview of more than fifty most recent MSA research articles to find the gaps in terms of tasks, approaches theories and applic ations used till date. There seems to be no single approach, theor y, and tool which can support MSA. The study showed that each and every mode presents different difficulties which have not bee n fully solved yet, such as feature points of a face, voice clarity in audio, video summarization and so on, and are great research op portunities for the future researchers. Also, this research recomm ends a list of existing and upcoming difficulties and opportunities of MSA research.**

Keywords- Multimodal; Sentiments; Opportunities; Difficulties; Review

I. INTRODUCTION

Social networks such as Facebook or Twitter, become one form of shelter for the users, as they get rid of being expressing their thoughts to their virtual friends. Also, the different thought expressed by the users regarding organizations, political parties or leaders, products, etc. allows the different beneficiary to strengthen themselves. But the fact is, there are many users, who differ in their expressions, languages, and emotions. Some may express in English and some may in Bangla etc. Also, their emotions are expressed in different modes such as images, audio, and video etc., which are called Multimodal sentiments (MS). MS became the challenge for the researchers and equally sophisticated for a machine to understand. One of the studies that supports MS problems is a Multimodal sentiment analysis (MSA), which is the analysis of emotions, attitude, and opinion from audiovisual format [1].

To date, due to the proper development of tools, the analysis was confined mostly on texts, except a few on visual contents. Research shows that, the visual expression of emotions contains more information than the text. However, in the case of MSA, data from each modality should be processed separately to get sentiments. Then, the results from each modality should be integrated to get final polarity or affective states. Also, Studies show that, research on each and every modality poses different difficulties solely or as a whole. In this study, we have uncovered some of those difficulties for the future researchers to bring out the opportunities of research in this field. Moreover, we have also contributed the following through this study:

1. This work presents an overview of possible difficulties in an extension of existing research in the field.
2. This work presents an overview of possible extensions of recent research in the field.
3. This paper gives the new researchers a quick idea about different existing tools of MSA research.
4. This work presents different MSA tasks, approaches and applications in a single picture to help grasp the field easily.

The remainder of this paper is organized as follows: Section 2 provides a review of related works, Section 3 includes a brief Methodology of this research, Section 4 presents some ideas about the current and future opportunities in MSA Research, Section 5 presents some ideas about the current and future difficulties in MSA Research, Section 6 provides findings for future researchers, and finally Section 7 outlines the conclusions of the study.

II. RELATED WORK

With the advent of time, many works being done on the Multimodal sentiment analysis, such as Resource building, Task identification, Approaches development and searching for applications. Equally, many reviews being done with MSA research. Some of these are presented in this section as a literature review. Vohra et al. [2] do a survey on sentiment analysis concepts, its application, and possible difficulties by emphasizing only on textual SA, ignoring MSA issues. Vidula et al. [3] conducted a survey on Multimodal SA data set, methods to prepare the data set, SA techniques and upcoming difficulties and opportunities in this field to a few extent. An analysis on Multimodal sentiment data and models being reviewed by Fulse et al. [4], where, they discussed the difficulties and opportunities of integrating the results of different modes in an emotion reorganization system. Ravi et al. [5] surveyed the task approaches and applications of SA

978-1-5090-6004-7/17/$31.00 ©2017 IEEE

and discussed in details about the opportunities and difficulties of SA research, but ignored MSA research.

Medhat et al. [6] conducted a detailed survey only on SA algorithms, didn't consider MSA. They have categorized the algorithms on the basis of their use in the data set and in the domain. They also summarize the approaches and data set on the basis of their applications. Marjan [7] conducted the review, where, she discussed different fusion techniques and their effect and outcome of emotion recognition system. Also, she discussed the performance and robustness issues of MSA in details along with future challenges and opportunities. A study conducted on SA concepts, existing research on SA and Future research questions in SA by Apple et al. [8]. They also compared different algorithms of SA. Finally, they have recommended some of the opportunities and difficulties of SA. But, they totally ignore the issue of MSA. Osimo [9] have focused on finding the research gap along with research challenges in SA for future researchers in this field. They also summarize current research, that being conducted in this field along with short term and long term research issues by ignoring MSA.

Most of the researchers on the above discussion ignored MSA issues. Though some of them noticed the issues, but they were limited in nature. In this study, we have derived the difficulties in existing research, present a useful recommendation on the opportunities for future research.

III. METHODOLOGY

In this paper, we have carefully selected and reviewed above fifty papers, as some of them are out of date, we have intentionally excluded them. Finally, we have focused on thirty three papers. We have brought together all the MSA task, approaches and applications in a single pictorial presentation as in Fig. 1 in the next section. Then, we have presented possible opportunities and difficulties poses by recent research papers in the next two sections. Finally, we have presented findings of our study, which in turns are the recommendations for the future researchers.

IV. MSA TASKS, APPROACHES, AND APPLICATIONS

From Fig. 1, it is clear that, there are no unique methods, models, and theories to deal with MSA. It is also seen that, there are many problems or tasks such as Sentiment Classification, Subjectivity Classification, and so on that need to be addressed carefully. It is also clear from fig. 1, in the case of MSA, each and every application type could be explored differently, such as some of them may want to classify subjectivity, and others may want to classify sentiments. In the case of approaches, some of them are used for classifications, and some of them are used for resource building, etc. It is also shown that, there are many application areas, where MSA could be applied such as Student feedback analysis, market prediction, etc. Above discussion indicates that, there are huge opportunities and difficulties in MSA research. So, in this paper, we have tried to address some of them through a detailed review on some of the recent works.

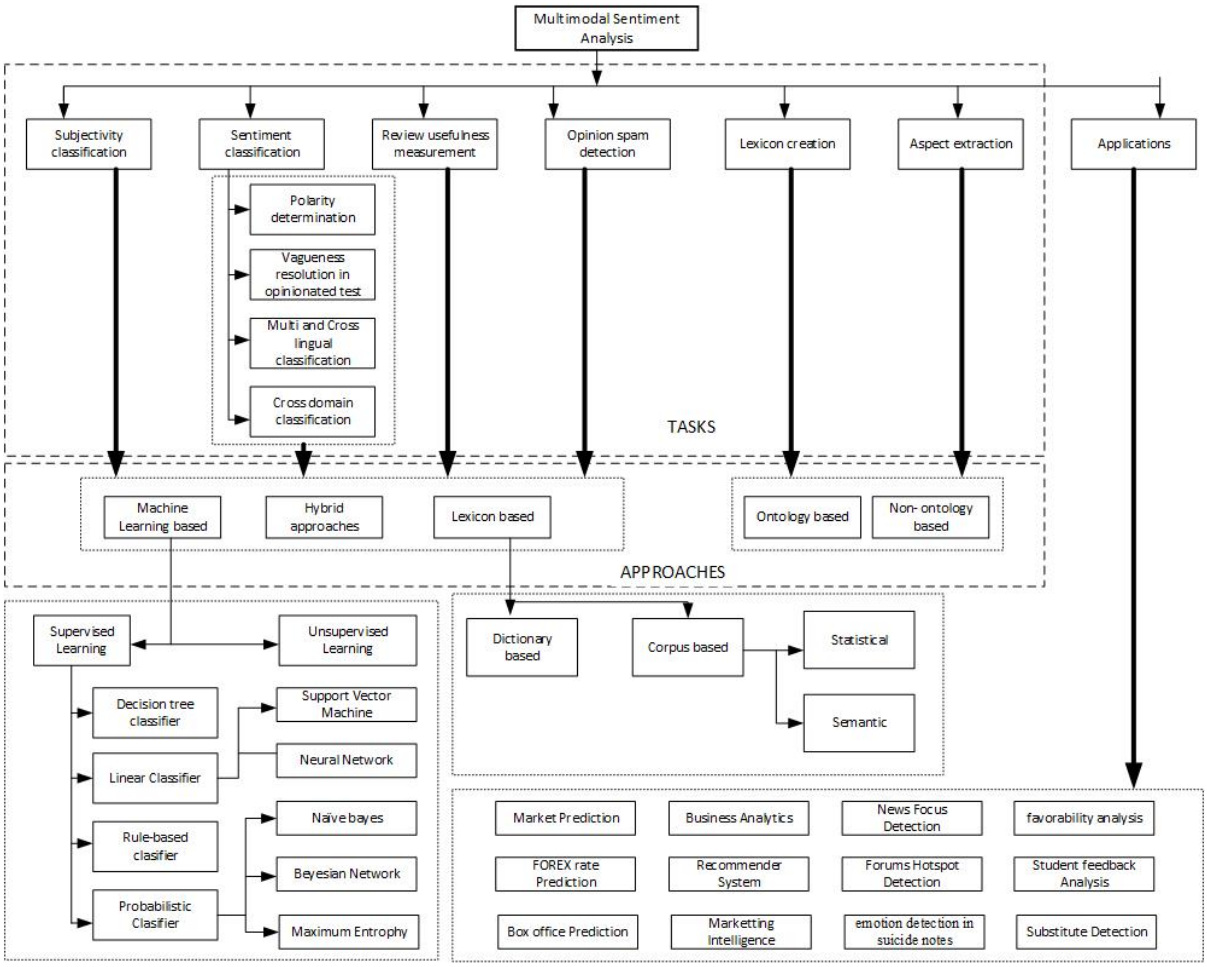

Figure 1. Multimodal Sentiment Analysis Tasks, Approaches, and applications

V. MSA RESEARCH OPPORTUNITIES

Facial activity-based approach on images and videos were applied by Joho et al. [10] for tracking motion vectors on a human face. This research creates a new opportunity to the researchers to explore more key points from the face as it only traces twelve key points using more techniques. For classification of sentiments using linear and Facial Part Actions model from facial expression was applied by Ahn [11]. The opportunity create by their research are- use of nonlinear model in the same domain, and identifying continuous changing emotions such as iris movement etc. more accurately. Dumoulin et al. [12] used hierarchical approach for affect recognition from the human face, where, they automatically detect emotions from the human face in the movies. But, the use of temporal smoothing for the improvement of emotion detection from movies will be the future research opportunity. Dupplaw et al. [13] used Testbed Platform framework for extracting sentiments of text and image data. But, the result of text and image sentiment analysis could be improved using Future Predictor Applications (FPA) and Media Content Analysis (MCA), which would be the upcoming opportunity. Chen et al. [14] used Weighted SVM and Proportion-SVM for sentiment concept classification from image and text data and achieved performance improvement from state-of-art research by 50%. They create the opportunity to apply the similar approach to many domains as they only applied to the political domain.

Similar modality data were explored by Baecchi et al. [15] for aspect extraction using some neural network models such as Skip-gram and De-noising Auto-encoders. As these models were applied to twitter data, it could also be applied to other social media data. Leeman-Munk [16] used deep learning-based and topology-based models on same modality data to find the effect of multimode over single mode data in student assessment and got better performance. As per them, the future opportunity is to explore more modality for creating personalized scaffolding. Kernel-based information fusion method (KCFA) and Extreme Learning Machine (ELM) were used by Poria et al.[17] to classify sentiments from the text, audio, and video. Their research outperformed the state-of-art system by 22.90%. However, future opportunities posed by them are the improvement of sentiment classification using more cognitive adapted fusion engine, and some unsupervised, and semi-supervised learning algorithms. Siddiquie et al.[18] also explored the same modality to measure "Review" usefulness by the use of Simple Linear Iterative Clustering (SLIC) Algorithm, SVM, and Convolution Neural Networks (CNNs) and found, politically attached videos presents more strongly negative viewer comments than non-persuasive

videos. Exploration of intents and aesthetics would be the possible future opportunities.

Poria et al. [19] used decision- and feature-level fusion methods to merge affective information from text, audio, and video by applying Naïve Bayes, SVM, ELM, and Neural Networks algorithms and improved performance of state of the art research by 20%. The performance could be further improved by the investigation of gaze and smile-based facial expression, 3D face recognition, also applying more techniques. Pereira et al. [20] applied content-based multimodal sentiment analysis on text, image, audio, and video data on TV videos and got reasonable result and poses the possibilities of applying the same method in the exploration of general videos. Liu [21] used Multimodal Deep Belief Networks (MDBN) model for solving link prediction with the promising result, which can be further extended by applying the same model to other domains. For subjectivity classification from the text, image, and video, Maynard et al. [22]applied Active Shape Models, Active Appearance Models, Constrained Local Model, and Semantic and Rule-based approaches. Opportunities exist in their work are-an in-depth use of discourse analysis and sarcasm detection. Rhoet al. [23] used Thayer's model, and TWC model to identify emotions from images, audio, and video. But using Gaussian mixture model (GMM) and SVM may achieve more accurate results.

VI. MSA RESEARCH DIFFICULTIES

As per Vidula et al. [3], difficulties in MSA research are Creating multimodal data set, detecting hidden emotions from multimodal data, and Fusing multimode sentiments results. Microblog text, images, and visual content classification were being done by Zhao et al. [24]using SVM. Their Features extracted are high-level semantic image and low-level visual features and TF-IDF text features. The difficulties still needs to deal with are shortness between visual content and text, incompatibility and diversity of data in a microblog. Wang et al. [25] dealt with the difficulties of the semantic gap between low-level visual sentiments and high-level image features in the social images with the use of unsupervised SA model. But, SA for geo-location, link information, and user history are some of the difficulties that still need to be addressed. Maynard et al. [26]handled the SA difficulties such as noisy ungrammatical text, use of swear words, sarcasm, face model alignment inaccuracies from text and image with the use of rule-based approach and Locality Sensitive Hashing (LSH) respectively. More difficulties need to address are Coreference resolution, short utterances, and documents with implicit knowledge, identifying implicit and multi-dimensional tokens in images. As per Ji et al. [27], the upcoming difficulties in Visual Sentiment prediction are construction and deep understanding of visual and GIF ontology.

Marjan [7] conducted a survey and found the following upcoming difficulties that need to be addressed carefully for MSA, such as Identifying images with Hand occlusion, noise, low resolutions and movement of objects ,development of affect recognizer for recognition of multimodal human affective behavior, Fusing different modality features with respect to varying dynamic structure, time scale as well as metric levels, Extracting linguistic and paralinguistic features reliably from audio channel, building context model with different fields such as one's ID, speech etc. As per Saif et al. [28], difficulties in MSA research are Colloquial language, Short texts, Platform-specific elements, and Real-time Big Data. Panda et al.[29] Proposed a model to deal with the difficulties associated with multimodal music emotion recognition. Some other difficulties need to address in their work are- increasing size of the dataset incorporating standard, using semantic features in the stronger emotional state of song lyrics, melodic, MIDI and lyrical features of audio. Cambria et al. [30] proposed an approach to work with the continuous interpretation of multimodal conceptual and affective information in some domains. But, this work poses some other difficulties such as discrete interpretation of multimodal conceptual and affective information and including the different domain of multi-modality in the study of useful information.

The difficulties as per Langlet et al. [31] in dealing with human-agent interaction sentiments are- managing speech disfluencies and distinguishing between implicit and explicit sentiments of likes and dislikes. According to Zadeh et al.[32], existing and the upcoming difficulties in MSA research are the creation of the dataset. They have created aspect level dataset associating sentiment intensity and subjectivity annotations. However, still there exist difficulties in creating the dataset of multimodality in the document and sentence level. In a study Schuller et al. [33] found that, the key difficulties related to MSA research is to ethically collect, annotated and exploit the affective and behavioral corpus. From an investigation by Fulse et al.[4] on different input modes, their effect on each other, and fusion techniques and got multiple modalities as a whole presents the better results than single mode. As per them, MSA research difficulties include- dealing with Cultural effects, linguistic variation, diverse contexts, and moving from single modal to multimodal itself. A study was done by Yadav et al. [1]in the audiovisual format to explore sentiments and got highly remarkable improvement in performance than in text. Some other difficulties in similar research are- incorporating more feature points in analyzing sentiments from the face, exploration of larger videos, the fusion of speech features with facial emotions, and summarization of videos.

VII. FINDINGS

From the above study, we suggest the following opportunities exist for future researchers, such as:

1. Application of more machine learning algorithms in the existing work, and existing approaches to other domains.
2. Exploring the existing datasets with different models and techniques.
3. Changing the way of preprocessing.
4. Looking for sentiments in the images using more feature points
5. Search for appropriate tools, models and techniques to deal with multilingual datasets

6. Search for appropriate methods and tools to derive sentiments from audiovisual data
7. Find the appropriate tools for sarcasm problems in MSA
8. Extending the existing works with more modality
9. Look for more fusion techniques

From the above study, we suggest the following difficulties need to handle by future researchers, such as:

1. Creating Dataset other than English as well as Multimodal dataset.
2. Creating the dataset of different domain (movie, product, etc.) and integrating it.
3. Co reference resolution
4. Fusing analysis results from multiple modalities.
5. Detecting hidden emotions and Sarcasm.
6. Dealing with a short text, noisy and low resolution images and videos, and the implicit and the explicit meaning of different modalities.
7. Visual Ontology construction
8. Collecting and analyzing data ethically.
9. Applying different existing algorithms in current research.
10. Exploring more feature points of the face, larger videos, and video summarization.

VIII. CONCLUSIONS AND FUTURE WORK

With the passage of time, Sentiment analysis research has gone a far away in the development of required tools, algorithms, and techniques, but it is still a long way in case of such development for Multimodal sentiment analysis. This creates both the opportunities and the difficulties for the researchers of this field. So, this paper summarizes some of those opportunities such as exploring the existing datasets with different approaches, change the way of preprocessing combinations, etc. and difficulties such as coreference resolution, video summarization, and exploration of more feature points etc. from the most recent articles and provide necessary guidelines to the future researchers. This research also provides a single pictorial representation of MSA tasks and approaches. Thus, it opens huge possibilities for the future researchers of this field. In the future, we would like to conduct the study to identify and summarize many other opportunities and difficulties of MSA research.

REFERENCES

[1] S. K. Yadav, M. Bhushan, and S. Gupta, "Multimodal sentiment analysis: Sentiment analysis using audiovisual format," in Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on, 2015, pp. 1415-1419.

[2] M. S. Vohra and J. Teraiya, "Applications and challenges for sentiment analysis: A survey," in International Journal of Engineering Research and Technology, 2013.

[3] V. D. Bhat, V. S. Deshpande, and R. Sugandhi, "A Multimodal Sentiment Analysis Scheme to Detect Hidden Sentiments", 2014.

[4] S. Fulse, R. Sugandhi, and A. Mahajan, "A Survey on Multimodal Sentiment Analysis," in International Journal of Engineering Research and Technology, 2014.

[5] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," Knowledge-Based Systems, vol. 89, pp. 14-46, 2015.

[6] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, pp. 1093-1113, 2014.

[7] S. Marjan, " A Survey for Multimodal Sentiment Analysis Methods", In Int.J.Computer Technology & Applications, vol. 5, pp. 1470-1476, 2014.

[8] O. Appel, F. Chiclana, and J. Carter, "Main Concepts, State of the Art and Future Research Questions in Sentiment Analysis," Acta Polytechnica Hungarica, vol. 12, pp. 87-108, 2015.

[9] D. Osimo and F. Mureddu, "Research challenge on opinion mining and sentiment analysis," Universite de Paris-Sud, Laboratoire LIMSI-CNRS, Bâtiment, vol. 508, 2012.

[10] H. Joho, et al., "Looking at the viewer: analysing facial activity to detect personal highlights of multimedia contents", in Multimed Tools Appl © Springer Science+Business Media, vol. 51, pp. 505–523, 2011.

[11] J. Ahn, et al., "Conveying Real-Time Ambivalent Feelings through Asymmetric Facial Expressions", in Springer-Verlag Berlin Heidelberg, pp. 122–133, 2012.

[12] J. Dumoulin, et al., "Affect Recognition in a Realistic Movie Dataset Using aHierarchical Approach", in ASM'15, ACM, pp. 15-20, 2015.

[13] D. Dupplaw, et al., "Living Knowledge: A Platform and Testbed for Fact and Opinion Extraction from Multimodal Data", in Springer-Verlag Berlin Heidelberg, pp. 100–115, 2012.

[14] T. Chen, et al., "Object-Based Visual Sentiment ConceptAnalysis and Application", in MM'14, ACM, 2014, pp. 367-376.

[15] C. Baecchi, T. Uricchio, M. Bertini, and A. Del Bimbo, "A multimodal feature learning approach for sentiment analysis of social network multimedia," Multimedia Tools and Applications, vol. 75, pp. 2507-2525, 2016.

[16] S. Leeman-Munk, "Two Modes Are Better Than One:A Multimodal Assessment FrameworkIntegrating Student Writing and Drawing", in C. Conati et al. (Eds.): AIED 2015, LNAI 9112, Springer International Publishing Switzerland, pp. 205-215, 2015.

[17] S. Poria, E. Cambria, N. Howard, G.-B. Huang, and A. Hussain, "Fusing audio, visual and textual clues for sentiment analysis from multimodal content," Neurocomputing, vol. 174, pp. 50-59, 2016.

[18] B. Siddiquie, D. Chisholm, and A. Divakaran, "Exploiting Multimodal Affect and Semantics to Identify Politically Persuasive Web Videos," in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, 2015, pp. 203-210.

[19] S. Poria, A. Hussain, and E. Cambria, "Beyond text based sentiment analysis: Towards multi-modal systems," University of Stirling, Stirling FK9 4LA, UK, Tech. Rep, 2013.

[20] H. R. Pereira, et al., "Multimodal Sentiment Analysis for Automatic Estimationof Polarity Tension of TV News in TVNewscasts Videos", Webmedia '15, ACM , 2015. pp. 157-160.

[21] F. Liu, et al., "Multimodal Learning Based Approaches for Link Prediction in Social Networks", Springer International Publishing Switzerland, pp. 123–133, 2015.

[22] D. Maynard, D. Dupplaw, and J. Hare, "Multimodal sentiment analysis of social media," 2013.

[23] S. Rho, S.Yeo, "Bridging the semantic gap in multimedia emotion/moodrecognition for ubiquitous computing environment", Springer Journal of Supercomputing, vol. 65. pp. 274–286, 2013.

[24] S. Zhao, H. Yao, S. Zhao, X. Jiang, and X. Jiang, "Multi-modal microblog classification via multi-task learning," Multimedia Tools and Applications, pp. 1-18, 2014.

[25] Y. Wang, S. Wang, J. Tang, H. Liu, and B. Li, "Unsupervised sentiment analysis for social media images," in Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, 2015, pp. 2378-2379.

[26] D. Maynard and J. Hare, "Entity-Based Opinion Mining from Text and Multimedia," in Advances in Social Media Analysis, ed: Springer, 2015, pp. 65-86.

[27] R. Ji, D. Cao, Y. Zhou, and F. Chen, "Survey of visual sentiment prediction for social media analysis," Frontiers of Computer Science, pp. 1-10,2016.

[28] H. Saif, F. J. Ortega, M. Fernández, and I. Cantador, "Sentiment Analysis in Social Streams," 2016.

[29] R. Panda, et al., " Multi-Modal Music Emotion Recognition:A New Dataset, Methodology and Comparative Analysis", in 10th International Symposium on Computer Music Multidisciplinary Research – CMMR'2013, 2013.

[30] E. Cambria, N. Howard, J. Hsu, and A. Hussain, "Sentic blending: Scalable multimodal fusion for the continuous interpretation of semantics and sentics," in Computational Intelligence for Human-like Intelligence (CIHLI), 2013 IEEE Symposium on, 2013, pp. 108-117.

[31] C. Langlet and C. Clavel, "Adapting sentiment analysis to face-to-face human-agent interactions: from the detection to the evaluation issues," in Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on, 2015, pp. 14-20.

[32] A. Zadeh, R. Zellers, E. Pincus, and L.-P. Morency, "MOSI: Multimodal Corpus of Sentiment Intensity and Subjectivity Analysis in Online Opinion Videos," arXiv preprint arXiv:1606.06259, 2016.

[33] B. Schuller, J.-G. Ganascia, and L. Devillers, "Multimodal Sentiment Analysis in the Wild: Ethical considerations on Data Collection, Annotation, and Exploitation."