IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-1, Jan 2019 E-ISSN: 2347-2693

A Survey of Computational Approaches and Challenges in Multimodal Sentiment Analysis

Mahesh G. Huddar [1*], Sanjeev S. Sannakki [2], Vijay S. Rajpurohit [3]

[1]Dept. of Computer Science and Engineering, Hirasugar Institute of Technology, Nidasoshi, Belgaum, India
[2,3]Dept. of Computer Science and Engineering, Gogte Institute of Technology, Belgaum, Karnataka, India

[*]*Corresponding Author:* mailtomgh1@gmail.com, *Tel.: +917411043272*

Available online at: www.ijcseonline.org

Received: 02/Jan/2019, Accepted: 18/Jan/2019, Published: 31/Jan/2019

***Abstract* -** Most of the recent work in sentiment analysis is carried out on textual data. The text based sentiment analysis mainly relies on construction of word dictionaries, using machine learning techniques that learn and extract opinion from large text corpora. Text based sentiment analysis has numerous applications such as customer satisfaction analysis about a brand or product perception, to gauge voting intentions etc. With the rapid growth of social media, users post humongous volume of data in various modalities such as text, image, audio, and video. These multimodal data streams bring new opportunities for going beyond text based sentiment analysis and improving possible results. Since sentiment can be extracted from facial and vocal expressions, prosody and body posture, multimodal sentiment analysis offers new avenues in sentiment analysis. In multimodal sentiment analysis, sentiment is extracted from transcribed content, visual and vocal features. This survey defines sentiment, sentiment analysis, states problems and challenges in multimodal sentiment analysis and finally reviews some of the recent computational approaches used multimodal sentiment analysis.

***Keywords* -** Sentiment Analysis, Computational Approaches, Multimodal Sentiment Analysis, Challenges, Machine Learning

I. INTRODUCTION

Automatic sentiment analysis is to uncover one's opinion, position or attitude towards a topic, person or entity [1]. Extracting people's opinion towards a certain topic, entity or person has many applications. For example, Political parties are interested in knowing the opinion of voters to gauge the voting intensities [2]. Companies use sentiment analysis to understand people opinion about their product and service [3]. Initially only text based sentiment analysis was done but today we can witness huge growth in the multimodal data availability in the World Wide Web in the form of audio visual content [4]. Affect, feeling, emotion, sentiment and opinion are often used interchangeably in the literature [5]. In [6] authors discussed the definition of sentiment, emotion, and emotional disposition from the philosophical perspective. In [6] they define sentiment as "love or hate opinion that can be expressed in many forms such as, the affection you may have for your hamster, your devotion to your country, your dislike for the banking establishment, and your great fondness for the most recent electronic gadget". Definitions of emotion, opinion, sentiment and difference between them were discussed in context with sentiment analysis and emotion detection from text in [5]. In [5] they, differentiate sentiment from emotion based on their duration, sentiment as a long term while emotion as a short term. Effective sentiment analysis system is one which captures sentiment holder and entity in addition to the correct polarity. "Emotion recognition is the automatic discovery of an episodic emotional reaction, often of a single person; unlike opinions, emotions are short-term" [5]. In [7] authors define an opinion as a quadruple consisting of entity, sentiment holder, sentiment and claim. According to [7], sentiment analysis is a process of detecting sentiment of a claim a sentiment holder had on a particular entity. They formalize the problem of sentiment analysis as a problem of constructing a tuple consisting of entity, aspect of entity, time, sentiment holder and polarity such as positive or negative. General approach to multimodal sentiment analysis is shown in figure 1. In this survey we focus on computational approaches used in multimodal sentiment analysis and associated challenges. In the remainder of this survey, existing text-based, audio-visual and multimodal computational approaches for sentiment analysis are discussed in section 2. Section 3 reviews challenges in sentiment analysis from multimodalities. Finally, applications of multimodal sentiment analysis are discussed in section 4 and conclude the survey in section 5.

II. Sentiment Analysis: Computational Approaches

A. *Text: Sentiment analysis from text data*

This section provides an overview of supervised and unsupervised methods in sentiment analysis and future directions and limitations in the field of sentiment Natural Language Processing (NLP). Supervised method of sentiment analysis aims in the development of predictive models for automatic sentiment analysis using annotated datasets.

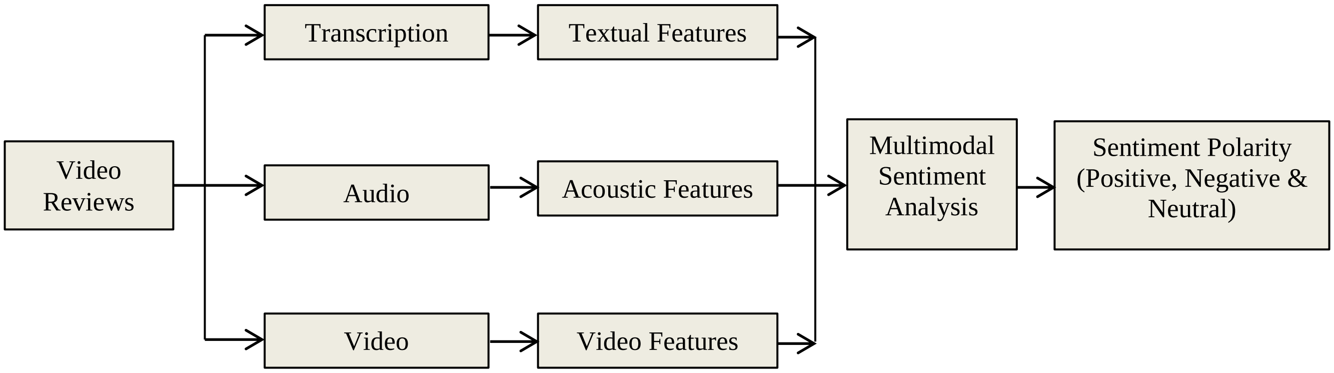

Figure 1 General approach to multimodal sentiment analysis

In this method feature vector is built for each annotated text entry, which is then used to train machine learning algorithms and finally validate the learning against annotated reference text. In the early days text is classified as either subjective or objective using supervised sentiment analysis [8]. One of the earliest works carried out using supervised approach of analysis of sentiment is prediction of stock market performance [9], leading to the development of an election outcome prediction [10] and movie box-office performance prediction [11]. In early days, to solve the problem of sentiment analysis probabilistic machine learning models such as support vector machines and Naïve Bayes [12] were used. These approaches use bag-of-words and lexicon based dictionaries models to calculate feature vectors of the reviews [13]. Most supervised sentiment analysis techniques are domain specific, such as social web, movie reviews or T.V. news. Arguably the most preferred and widely researched domain of sentiment analysis is the classification of polarity of tweets (Twitter short public messages of 140 characters in length).Since 2013 SemEval competition [14] is organized, in which sentiment analysis developers submit their tools for evaluation. Each tool is independently evaluated on the same dataset and the one which achieves best results is awarded as best tool of the year. In 2013 [15] a tool developed using SVM and a wide variety of features won best tool of the year award an improved model with fever features but more lexicon achieves best results in 2014 [16]. In 2015 a combination of SVM with other models such as maximum entropy and stochastic gradient descent optimization won the best performing model of the year [17]. In 2016, a tool called SwissCheese [18]; Convolutional Neural Network was used to train large dataset of tweets with emoticons achieved the best results. In 2017 SemEval competition, compared 2016 two changes were made Arabic tweets are introduced for all sub tasks and deep learning was used in developing sentiment analysis tool [19]. Bootstrap [20] parametric ensemble framework, is an alternative benchmark for Twitter sentiment analysis compared to other supervised machine learning methods. The main difficulty in the usage of supervised sentiment analysis methods is the need of huge amount of labeled datasets to train the machine learning algorithms (classifiers). The domain adaptation of supervised classifiers is difficult since some domains are too formal or produce lengthier textual data than other domains. For example, movie reviews are more formal and lengthier which allows us to create of tree banks [21], but the same methodology cannot be applied to social media data as the reviews are of short in nature (tweets are of 140 character in length). Due to unavailability of labeled data, unsupervised sentiment analysis methodology is used in certain applications. Usually expert knowledge is used to construct lexicon, where either words or phrases or both annotated with their associated sentiment. General Inquirer is one the most widely used reference lexica for sentiment analysis, which classifies the text based on corpus of positive and negative terms. To improve the recall of unsupervised sentiment analysis, usually General Inquirer is combined with other lexica [22]. Linguistic Inquiry and Word Count (LIWC) [23] software is used to count positive and negative terms in textual data. Deep learning approaches are the most recent and promising techniques in sentiment analysis and emotion detection [24]. Deep learning approaches compute word embedding's from large scale datasets that are relevant for sentiment analysis and emotion detection [25]. Deep Recurrent Neural Networks method was applied in [26] for subjectivity detection.

B. Audio: Sentiment analysis from speech

Analysis of emotional and affective cues from speech data has a comparably long tradition [27]. However, sentiment analysis, exclusively from spoken data is a relatively new field of study. In [28] authors showed that prosody (accent) related features contain information on sentiment. Recent work on speech based sentiment and emotion analysis [27] [29] have focused on identifying several acoustic features such as pitch, utterance level [30], bandwidth, and duration. Results in [31] show that, speaker-dependent approach gives good results compared to speaker-independent approach. They used voice quality, prosodic and Mel Frequency Cepstral Coefficients (MFCC) as speech features, Gaussian mixture model (GMM) as a classifier and achieved an accuracy of about 98%. However, as many applications deals with different speakers (users), the speaker-dependent approach may not be suitable. For speaker-independent applications, authors have selected spectral, prosodic and voice features using Sequential Floating Forward Selection (SFFS) algorithm [32] and two-step classification approach on the Berlin Database of Emotional Speech (BDES) [33] and achieved an accuracy of 81% [34]. In [35] authors used 377 different features based on intensity, pitch, pause length, MFCC, voiced segment characteristics and Barkspectral bands and identified sadness and anger with an accuracy of 76.67% and 93.30% respectively.

C. Image: Visual sentiment analysis

Popular social networking websites, such as Flickr, Instagram contain large amounts of visual information in the form of images. Many of these images are manually annotated by the user. Visual sentiment analysis is a process of detecting and extracting sentiment and emotion expressed by means of facial expressions or body gestures. Along with object or entity, actions and locations visual content does contain cues about sentiment or emotion. For example, an image showing a hot coffee cup, beautiful flower or delicious cake is likely expressing the positive polarity of the publisher of the image. Sentiment analysis and emotion detection using computer vision techniques is a comparably new area of research. Automatic sentiment and emotional retrieval systems form images was first proposed by Colombo et al. [36]. In their system, they were able to extract emotional semantics such as joy, action, uneasiness and relaxation using high level representation of images. In [37] Jindal et al. progressively trained their framework built with Convolutional Neural Networks (CNN) using manually labeled Flicker dataset. Their proposed CNN based framework achieves better results compared to other frameworks. In [38] Jyoti et al. proposed a transfer learning based novel framework to predict emotion and sentiment. They use very deep convolutional neural network to find hyper-parameters and initialize their network to prevent from over fitting. They demonstrated their model using Twitter image dataset. Regulated Matrix Factorization Approach [39] was used to understand the sentiment from massive collection of images using image features and contextual information such as comments and annotated tags. In [40]Yu et al. used deep learning based convolutional neural network (CNN) for extracting sentiment from Chinese social media (SinaWeibo) using joint textual and visual features. For textual sentiment analysis they trained CNN using word2vec features and deep CNN for visual sentiment analysis. They used decision level fusion to get the final results. They demonstrated their framework using dataset collected from Chinese social media SinaWeibo. In [41] Wang et al. argued that neither visual features nor contextual information associated with images are by themselves insufficient for accurate sentiment labeling. They use supervised and unsupervised method to machine learning approaches to predict sentiment from images and associated tags. Luo et al. [42] use Progressively Trained and Domain Transferred Deep CNN Networks for image sentiment analysis. They collected half million images (machine labeled and noisy images) from Flicker. They employ progressive approach to train deep CNN network for such noisy data. They demonstrated their approach using manually labeled twitter images. Yang et al. [43] proposed a novel learning model based on both images and comments posted by their friends on images. In their experiment they demonstrated on Flicker dataset that the use of friend's comments, visual features and emoticons can significantly improve the accuracy. In [44] Frome et al. proposed a deep visual-semantic embedding model. They used both labeled images and semantic information gained from tag associated with images to train their model. They demonstrated their model using ImageNet dataset and proved that semantic information improves the accuracy by 18%.

D. Multimodal Sentiment Analysis

This is one of the emerging research area where sentiment from multimodal data is extracted using both verbal and non-verbal behaviors. There exists multiple standard datasets for multimodal sentiment analysis namely, YouTube [45], ICT-MMMO [46], CMU-MOSI dataset [47], MOUD [48], News Video [49]. Morency et al. [45] were the first to address multimodal sentiment analysis problem. They developed YouTube dataset which contains 47 videos including 27 male and 20 female speakers with their age ranging from 14 to 60 years. Videos were extracted using keywords such review, job, movie, product review etc. They manually annotated sentiment to each video namely positive, negative and neutral. They used openEAR [50] to extract Pitch and speech pause. Commercial facial expression analysis software was used to measure duration of smiling, nose position and looking away from camera. Transcribed speech was classified as either positive or negative using lexicons [51]. Trimodal (audio, video and transcribed text) features were classified using Hidden Markov Model classifier. Trimodal sentiment classification outperforms unimodal classifiers with an average F1 score of 0.55. Although the dataset was very small and the textual classification method was simple, with their approach authors demonstrates the future scope of multimodal sentiment analysis. Poria et al. in [52] extracted facial expressions such as distance between eyes, facial landmarks using open source software. OpenSMILE [53] was used to extract audio features. They used text2vec [54] and part of speech to extract textual features. The best modality was text, and with multimodal fusion they achieved an accuracy of 88.6% for two-class (negative vs. positive) classification. Recently, in his work Poria et al. [55] extended the previous work by using larger number of audio and textual features. Sentic computing paradigm [56] was used to extract textual features. Using audio and textual features, the authors were able to achieve F1 score up to 0.78 on three classes. Wöllmer et al. [46] proposed the model to extract sentiment from online user-generated movie reviews posted on YouTube. They developed ICT-MMMO dataset by collecting 370, 1 to 3 minutes videos from YouTube. Out of 370 videos, 228 were positive, 119 negative and 23 neutral reviews [46]. For speech analysis, it was transcribed both manually as well as using automatic speech recognition software. Further, authors performed cross-domain analysis using 102622 written reviews of 4901 movies. Facial features were extracted using open source tool and low level acoustic feature were extracted using OpenSMILE [53]. Linear SVM was trained using acoustic features and Bidirectional Long-Short-Term-Memory (BLSTM) recurrent neural network was trained using audiovisual features [57]. Using cross-corpus n-gram analysis authors were able to achieve average F score of 0.73. They achieved F1 score of only 0.66 without text. Pérez Rosas et al. in [58] analyzed MOUD “Multimodal Opinion Utterances Dataset” which contains product reviews collected from YouTube [48]. They selected 80 Spanish videos with 15 male and 65 female sparkers. From these videos manually they have selected 30 seconds form sentiment annotation. Each video was manually segmented into 30 second length segments, each covering a single topic and manually labeled as positive, negative and neutral polarity. Each video was manually transcribed to text. Words with frequency less than 10 were discarded while constructing bag-of-word representation. Audio features such as duration of pause, pitch, intensity and loudness were extracted. Like [45] commercial facial expression analysis software was used to extract smile duration, nose position, and looking away. Textual sentiment analysis achieved an accuracy of 65%. They used early fusion for multimodal analysis and achieved an accuracy of 75%. Their work proves that multimodal sentiment analysis language independent. McDuff et al. [59] demonstrate the use of facial expression analysis to assess preference of American voters. They collected 5 videos of 611 voter’s responses from a US presidential election debate. Using only facial expression authors were able to achieve an accuracy of 73%. Table 1 shows the summary of recent papers on multimodal sentiment analysis.

Table 1: Recent research papers on multimodal analysis

References	Language	Datasets	Modality	Feature Fusion / Decision fusion	Features
[45]	English	YouTube [45]	Text-Audio-Video	Decision Level Fusion	Linguistic lexicons, Acoustic features using OpenEAR [50], Facial features such as smile, nose position
[46]	English	ICT-MMMO [46]	Text-Audio-Video	Hybrid Fusion	Linguistic features such as Polarized words, Acoustic features pause, and voice intensity, facial expressions smile, gaze using OpenSMILE [53]
[52]	English	CMU-MOSI dataset [47]	Text-Audio-Video	Feature Level Fusion	Textual features using text2vec [54] and part of speech, Acoustic features, facial expressions using OpenSMILE [53]
[55]	English	CMU-MOSI dataset [47]	Text-Audio-Video	Feature Level Fusion	Linguistic features using Sentic computing paradigm [56], Acoustic features, facial expressions using OpenSMILE [53]
[58]	Spanish	MOUD [48]	Text-Audio-Video	Feature level fusion	Linguistic, Acoustic features such as pause, pitch, intensity and loudness and Visual features such as smile duration, nose position, and looking away
[59]	English	News Video [49]	Text-Audio-Video	Feature level fusion	Acoustic, lexical and visual features
[60]	English	YouTube [45]	Text-Audio-Video	Feature / Decision fusion	Linguistic, audio features extracted using openEAR [50] and Concept-gram and Sentic Net-based features
[61]	English	230 YouTube Videos	Text-Audio-Video	Feature / Decision fusion	Linguistic, audio features such as prosody, MFCC, Video features

III. CHALLENGES

Today sentiment analysis is domain dependent as it depends on data from where training data came. Designing a sentiment analysis system which is domain independent is an open issue and that needs to be addressed, for example, adapting a model which was trained using product reviews for analyzing micro blog posts. Other important challenges of automatic sentiment analysis include how to handle sarcasm and irony and ambiguous situations. "A number of methods have been proposed to identify sarcasm from textual data" in [1]. Along with textual review, facial and vocal expressions of Multimodal sentiment analysis can help in identifying the sarcastic reviews. Another challenge in multimodal sentiment analysis is "efficiently exploring intra-modality dynamics of a specific modality (unimodal interaction).Intra-modality dynamics are particularly challenging for the language analysis since multimodal sentiment analysis is performed on spoken language. A spoken opinion such as I think it was alright . . . Hmmm . . . let me think . . . yeah . . . no . . . ok yeah" almost never happens in written text. This volatile nature of spoken opinions, where proper language structure is often ignored, complicates sentiment analysis. Visual and acoustic modalities also contain their own intra-modality dynamics which are expressed through both space and time" [62]. There is also the issue of annotating sentiment to private data recorded in the laboratory, which limits the tedious task of annotating sentiment to a small group people who are authorized to access the data. As a result, we are not only bounded by the amount of data we can record in the laboratory but also by a limited ability to label large amount of data. The primary source of multimodal data is social web which contains large amount of multimedia. Social web is a rich resource of multimodal data. The problem is that the quality and the context of the recorded material may vary, and the data is limited to certain demographics that are more represented on the Internet. However, since the data is public, it can be easily labeled through crowd-sourcing.

IV. APPLICATIONS OF SENTIMENT ANALYSIS

Recently Sentiment analysis is commercially used to summarize customer opinions or attitude towards a product or a brand. Using automatic sentiment analysis we are opinions at low cost. Earlier companies or organizations conduct surveys or create focus groups manually to assess customer opinion, which was much slower, tedious and expensive task. With the emergence social media user post their opinion at large scale in the form of spoken reviews or video review on YouTube. Hence large amount of data is available free of cost, which makes sentiment analysis a low-cost endeavor. Multimedia analytics is a new domain where multimodal sentiment analysis is heavily used. The current work on visual sentiment analysis including SentiBank [63] found its way into much multimedia content analysis works [64] that are not directly linked to sentiment analysis. Following are the some of the applications of multimodal sentiment analysis.

A. Box office prediction

With the rapid growth of social media, many online reviews on movies available in both textual and video format. This provides an opportunity to predict the performance of movie box office. In [69] authors used fuzzy clustering and support vector machines on inverse document frequency as features for sentiment prediction. In [70] authors used Weka's K-Means clustering tools on twitter, YouTube and IMDB movie database to predict movie box office prediction.

B. Stock Market prediction

Despite the growth of computational world, volatile nature of stock market makes stock market prediction one of the difficult tasks. Sentiment analysis can be used to make successful strategies. In [66] authors have used artificial neural network (ANN) for NASDAQ stock market index prediction. They used back propagation for training ANN's.

C. Business Analytics

Today many organizations are using sentiment analysis for decision support system and business improvement. In [67] authors proposed novel approach to business analytics in contemporary organizations.

D. TV Program and News summarization

The work of Ellis et al. [68] built a system using multimodal sentiment analysis which automatically analyze the broadcast video news and create the summarization of TV programs. They used Amazon Mechanical Turk annotated 929 sentence length videos for summarizing TV program.

E. Recommender systems

Many application users are given recommendations depending on their past experience. For example, in retail industry if users have searched for a particular product, they will be given recommendations in future endeavours. In [65] authors have proposed a hybrid approach which accurately provides recommendations.

F. Politically trend prediction

Rising trends of social media increases the possibility of predicting outcome of electoral poll. In [71] proposed a novel approach to predict election results in France, Italy and USA. Politically and socially pervasive content can be identified using multimodal sentiment analysis.

V. CONCLUSION

In this paper opinion, sentiment, emotion, opinion mining and sentiment analysis and emotion detection were defined. We presented an overview of the concept of sentiment analysis from textual data, speech and visual data. Challenges and applications are discussed. Recent work from number of researcher proves the importance and future scope of multimodal sentiment analysis. Our review on computation methods in multimodal sentiment analysis shows how multimodal (trimodal i.e., textual, audio and visual data analysis) approach outperforms unimodal approaches.

REFERENCES

[1] L. Z. Bing Liu, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*, Springer US, 2013, pp. 415-463.

[2] M. Prem, G. Wojciech and R. D. Lawrence, "Sentiment analysis of blogs by combining lexical knowledge with text classification," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, France, 2009.

[3] B. J. Jansen, M. Zhang, K. Sobel and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," *Journal of the Association for Information Science and Technology,* vol. 60, no. 11, p. 2169–2188, 2009.

[4] E. Cambria, B. Schuller and Y. Xia, "New Avenues in Opinion Mining and Sentiment Analysis," *IEEE Intelligent Systems,* vol. 28, no. 2, pp. 15 - 21, 2013.

[5] M. Munezero, C. S. Montero, E. Sutinen and J. Pajunen, "Are They Different? Affect, Feeling, Emotion, Sentiment, and Opinion Detection in Text," *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING,* vol. 5, no. 2, pp. 101-111, 2014.

[6] J. Deonna, C. Tappolet and F. Teroni, "Emotion, philosophical issues about," *WIREs Cogn Sci* , pp. 1-15, 2015.

[7] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *COLING '04 Proceedings of the 20th international conference on Computational Linguistics*, PA, USA, 2004.

[8] J. M. Wiebet, R. F. Bruce and T. P. O'Harat, "Development and Use of a Gold-Standard Data Set for Subjectivity Classifications," in *Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, 1999.

[9] J. Bollen, H. Mao and X.-J. Zeng, "Twitter mood predicts the stock market," *Journal of Computer Science,* vol. 2, no. 1, pp. 1-8, 2011.

[10] A. Tumasjan, T. O. Sprenger, P. G. Sandner and I. M. Welpe, "Predicting elections with Twitter: what 140 characters reveal about political sentiment," in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, George Washington University, 2010.

[11] S. Asur and B. A. Huberman, "Predicting the Future With Social Media," *Web Intelligence and. Intelligent Agent Technology (WI-IAT),* vol. 1, no. 6, pp. 492-299, 2010.

[12] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2002.

[13] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval,* vol. 2, no. 1-2, pp. 1-135, 2008.

[14] P. Nakov, T. Zesch, D. Cer and D. Jurgens, "International Workshop on Semantic Evaluation (SemEval 2013),," in *Association for Computational Linguistics.*, Atlanta, Georgia, 2013.

[15] S. M. Mohammad, S. Kiritchenko and X. Zhu, "NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets," in *In Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)*, Atlanta, USA, 2013.

[16] Y. Miura, S. Sakaki, K. Hattori and T. Ohkuma, "TeamX: A Sentiment Analyzer with Enhanced Lexicon Mapping and Weighting Scheme for Unbalanced Data," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, Ireland, 2014.

[17] M. Hagen, M. Potthast and B. Stein, "Webis: An Ensemble for Twitter Sentiment Detection," in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Denver, Colorado, 2015.

[18] J. Deriu, M. Gonzenbach, F. Uzdilli, A. Lucchi, V. D. Luca and M. Jaggi, "SwissCheese at SemEval-2016 Task 4: Sentiment Classification Using an Ensemble of Convolutional Neural Networks with Distant Supervision," in *Proceedings of the 10th International Workshop on Semantic Evaluation*, San Diego, CA, USA, 2016.

[19] S. Rosenthal, N. Farra and P. Nakov, "SemEval-2017: Sentiment Analysis in Twitter," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, Vancouver, Canada, 2017.

[20] A. Abbasi, A. Hassan and M. Dhar, "Benchmarking Twitter Sentiment Analysis Tools," *LREC,* p. 823–829, 2014.

[21] R. Socher, A. Perelygin, J. Wu, J. Chuang, M. C. D, A. Ng and C. Potts, "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2013.

[22] T. Wilson, J. Wiebe and P. Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," in *In: HLT/EMNLP*, Vancouver, BC, Canada, 2005.

[23] J. W. Pennebaker, R. L. Boyd, K. Jordan and K. Blackburn, "The Development and Psychometric Properties of LIWC2015," *Development Manual, University of Texas at Austin,* 2016.

[24] Q. T. Ain, M. Ali, A. Riaz, A. Noureen, M. Kamran, B. Hayat and A. Rehman, "Sentiment Analysis Using Deep Learning Techniques: A Review," in *(IJACSA) International Journal of Advanced Computer Science and Applications*, 2017.

[25] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu and B. Qin, "Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, Maryland, 2014.

[26] O. Irsoy and C. Cardie, "Opinion mining with deep recurrent neural networks," *EMNLP,* p. 720–728, 2014.

[27] S. G. Koolagudi and K. S. Rao, "Emotion recognition from speech: a review," *International Journal of Speech Technology,* vol. 15, no. 2, pp. 99-117, 2012.

[28] R. M. Ver´onica P´erez-Rosas, "Sentiment Analysis of Online Spoken Reviews," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012.

[29] T. Johnstone, "Emotional speech elicited using computer games," in *Proceedings., Fourth International Conference on Spoken Language*, Philadelphia, PA, USA, 2002.

[30] C. L. Shao-Hsien, "Joint processing of audio-visual information for the recognition of emotional expressions in human-computer interaction," Doctoral Dissertation Joint processing of audio-visual information for

the recognition of emotional expressions in human-computer interaction, USA, 2000.

[31] E. Navas, I. Hernaez and I. Luengo, "An objective and subjective study of the role of semantics and prosodic features in building corpora for emotional TTS," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 14, no. 4, pp. 1117-1127, 2006.

[32] P. Pudil, F. Ferri and J. Novovicova, "Floating search methods for feature selection with nonmonotonic criterion functions," in *12th IAPR International Conference on Pattern Recognition*, 1994.

[33] G.N. Peerzade, R.R. Deshmukh, S.D. Waghmare, "*A Review: Speech Emotion Recognition*", International Journal of Computer Sciences and Engineering, Vol.6, Issue.3, pp.400-402, 2018.

[34] H. Atassi and A. Esposito, "A Speaker Independent Approach to the Classification of Emotional Vocal Expressions," in *20th IEEE International Conference on Tools with Artificial Intelligence*, 2008.

[35] G. Caridakis, G. Castellano, L. Kessous, A. Raouzaiou, L. Malatesta, S. Asteriadis and K. Karpouzis, "Multimodal emotion recognition from expressive faces, body gestures and speech," in *Artificial Intelligence and Innovations*, 2007.

[36] C. Colombo, A. D. Bimbo and P. Pala, "Semantics in visual information retrieval," *IEEE MultiMedia,* vol. 8, no. 6, pp. 38 - 53, 1999.

[37] S. Jindal and S. Singh, "Image Sentiment Analysis using Deep Convolutional Neural Networks with Domain Specific Fine Tuning," in *International Conference onInformation Processing (ICIP),* Pune, India, 2015.

[38] J. Islam and Y. Zhang, "Visual Sentiment Analysis for Social Images Using Transfer Learning Approach," in *2016 IEEE International Conferences on Social Computing and Networking,* Atlanta, GA, USA, 2016.

[39] Y. Wang, Y. Hu, S. Kambhampati and B. Li, "Inferring Sentiment from Web Images with Joint Inference on Visual and Social Cues: A Regulated Matrix Factorization Approach," in *Proceedings of the Ninth International AAAI Conference on Web and Social*, 2015.

[40] Y. Yu, H. Lin, J. Meng and Z. Zhao, "Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks," *MDPI journals on algorithms,* vol. 9, no. 41, pp. 1-11, 2016.

[41] Y. Wang and B. Li, "Sentiment Analysis for Social Media Images," in *IEEE International Conference onData Mining Workshop (ICDMW)*, Atlantic City, NJ, USA, 2015.

[42] Q. You, J. Luo, H. Jin and J. Yang, "Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks by Quanzeng," *Association for the Advancement of Artificial*, pp. 1-9, 2015.

[43] Y. Yang, J. Jia, S. Zhang, B. Wu, Q. Chen, J. Li, C. Xing and J. Tang, "How do your friends on social media disclose your emotions?," in *Twenty-Eighth AAAI Conference on Artificial Intelligence*, Canada, 2014.

[44] A. Frome, G. Corrado, J. Shlens, S. Bengio, J. Dean and T. Mikolov, "Devise: A deep visual-semantic embedding model," in *Advances in neural information processing systems*, Lake Tahoe, 2013.

[45] L.-P. Morency, R. Mihalcea and P. Doshi, "Towards multimodal sentiment analysis: harvesting opinions from the web," in *ACM International Conference on Multimodal Interfaces (ICMI)*, Alicante, Spain, 2011.

[46] M. Wöllmer, F. Weninger, T. Knaup and B. Schuller, "YouTube movie reviews: ssentiment analysis in an audio-visual context," *IEEE Intelligent Systems,* vol. 28, no. 3, pp. 46-52, 2013.

[47] *Ketan Sarvakar, Urvashi K Kuchara, "Sentiment Analysis of movie reviews: A new feature-based sentiment classification*", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.3, pp.8-12, 2018

[48] V. P´erez-Rosas, R. Mihalcea and L.-P. Morency, "Utterance-Level Multimodal Sentiment Analysis," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, 2013.

[49] J. G. Ellis, B. Jou and S.-F. Chang, "Why We Watch the News: A Dataset for Exploring Sentiment in Broadcast Video News," in *ICMI'14*, Istanbul, Turkey, 2014.

[50] F. Eyben, M. Wöllmer and B. Schuller, "OpenEAR — Introducing the munich open-source emotion and affect recognition toolkit," in *3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 2009.

[51] J. Wiebe, T. Wilson and C. Cardie, "Annotating Expressions of Opinions and Emotions in Language," *Language Resources and Evaluation,* vol. 39, no. 2, p. 165–210, 2015.

[52] 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. vol. 174, pp. pp. 50-59.

[53] F. Eyben, M. Wöllmer and B. Schuller, "openSMILE: The Munich versatile and fast open-source audio feature extractor," in *ACM International Conference on Multimedia (MM)*, 2010.

[54] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems*, 2013.

[55] 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, 20176.

[56] Gagandeep Kaur, Kamaldeep Kaur, "*Sentiment Detection from Punjabi Text using Support Vector Machine*", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.6, pp.39-46, 2017

[57] M. Wöllmer, M. Kaiser, F. Eyben, B. Schuller and G. Rigoll, "LSTM-modeling of continuous emotions in an audiovisual affect recognition framework," *Image and Vision Computing,* vol. 31, no. 2, p. 53–163, 2013.

[58] V. P. Rosas, R. Mihalcea and L.-P. Morency, "Multimodal sentiment analysis of Spanish online videos," *IEEE Intelligent Systems,* vol. 28, no. 3, pp. 38-45, 2013.

[59] D. McDuff, R. Kaliouby, E. Kodra and R. Picard, "Measuring voter's candidate preference based on affective responses to election debates," in *Conference on Affective Computing and Intelligent Interaction (ACII),* US, 2013.

[60] S. Poria, E. Cambria, R. Bajpai and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion,* vol. 37, pp. 98-125, 2017.

[61] B. Siddiquie, D. Chisholm and A. Divakaran, "Exploiting multimodal affect and semantics to identify politically persuasive web videos," in *ACM on International Conference on Multimodal Interaction*, Seattle, Washington, 2015.

[62] A. Zadehy, M. Chen, S. Poria, E. Cambria and L.-P. Morency, "Tensor Fusion Network for Multimodal Sentiment Analysis," *arXiv*, 2017.

[63] D. Borth, R. Ji, T. Chen, T. Breuel and S.-F. Chang, "Large-scale Visual Sentiment Ontology and detectors using adjective noun pairs," in *ACM International Conference on Multimedia*, Barcelona, Spain, 2013.

[64] A. Khosla, A. D. Sarma and R. Hamid, "What makes an image popular?," in *International Conference on World Wide Web (WWW)*, 2014.

[65] X.-L. Zheng, C.-C. Chen, J.-L. Hung, W. He, F.-X. Hong and Z. Lin, "A Hybrid Trust-Based Recommender System for Online Communities of Practice," *IEEE Transactions on Learning Technologies,* vol. 8, no. 4, pp. 345 - 356, 2015.

[66] A. HedayatiMoghaddam, M. HedayatiMoghaddam and MortezaEsfandyari, "Stock market index prediction using artificial

neural network," *Journal of Economics, Finance and Administrative Science,* vol. 21, no. 41, pp. 89-93, Dec 2016.

[67] Shrija Madhu, "An approach to analyze suicidal tendency in blogs and tweets using Sentiment Analysis", International Journal of Scientific Research in Computer Science and Engineering, Vol.6, Issue.4, pp.34-36, 2018

[68] J. Ellis, B. Jou and S. Chang, "Why we watch the news: a dataset for exploring," in *ACM International Conference on Multimodal*, 2014.

[69] P. Nagamma, H. R. Pruthvi and K. K. Nisha, "An improved sentiment analysis of online movie reviews based on clustering for box-office prediction," in *International Conference on Computing, Communication & Automation (ICCCA)*, Noida, India, 2015.

[70] K. R. Apala, M. Jose and S. Motnam, "Prediction of movies box office performance using social media," in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Niagara Falls, ON, Canada, 2013.

[71] L. Curini, A. Ceron and S. M. Iacus, "To what extent sentiment analysis of Twitter is able to forecast electoral results? Evidence from France, Italy and the United States," in *7th ECPR General Conference*, Bordeaux, 2013.