International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 3165–3174

A Review on Sentiment Analysis from Multimodal Data

Rushali Patil[1], Sushama Shirke[2]

[1] *Army Institute of Technolog, Pune*

[2] *Army Institute of Technolog, Pune*
[1]rpatil@aitpune.edu.in, [2]sshirke@aitpune.edu.in

Abstract

A sentiment analysis is becoming a popular research area. Sentiments can be expressed in various forms like text, image, audio, video etc. Opinion of large population is anticipated by aggregating sentiments of individual and used in numerous applications. Traditional sentiment analysis system focuses on single modality to infer user's perception about subject. Such type of sentiment analysis has its own limitation and fails to employ other modality's expressiveness.

With the advent of technologies, the conventional system evolved into multimodal sentiment analysis which integrates multifarious data (i.e. text, audio, and visual) available over internet. Multimodality refers to the availability of more than one modality or medium. Every mode of data has unique features and helps users to express their emotions, opinions or attitude about the entity. Incorporating such features from multiple content enhance the effectiveness of sentiment analysis process. To find a constructive fusion mechanism to integrate these features is the challenging aspect of sentiment analysis. In this survey we have defined different modalities of sentiments, characteristics and fusion techniques of multimodal data. This paper gives an overview of different approaches for and applications of multimodal system.

Keywords: *Multi-modal data, Sentiment Analysis, Opinion Mining.*

1. Introduction

Nowadays social media has become very convenient and easy mode for knowledge sharing in various forms like text, images, audio, video, etc. People are freely sharing their opinions with each other through different platforms such as Facebook, Twitter, YouTube, and Foursquare. This huge amount of available information is further analyzed to help e-commerce, political reviews, recommender system and etc. Airport service quality is examined by user generated content on social media using text based sentiment analysis (SA) [21]. SA aims to acquire people's attitude hidden in shared views or opinions. SA helps in improving teaching and learning process by analyzing student's feedback from textual comments [19, 20]. Nowadays twitter is gaining more attention from people and have potential to influence the traditional media. Twitter allows users to create and post short text messages in the form of tweets. These tweets are categorized into positive or negative score [22]. The accuracy of tweet's content plays an important role in critical incidents such as natural calamities or social issues. This can be achieved by incorporating machine learning techniques with sentiment analysis in order to minimize spreading of misinformation [23].

People are expressing themselves through verbal, facial expressions, modulating tone of voice or body gestures. All these means exhibit strong correlation to affective computing which influences the judgment about entity (product, service, aspects etc.) into consideration. A multimodal sentiment analysis leverages SA by incorporating expressions from different modality to infer user's intent behind it [27, 5]. It also works

on finding semantic relationship between user's content available in text, audio, video, etc. forms and sentiment intensity exhibited by each modality [5]. Ongoing research in multimodal sentiment analysis covers three major areas:

- Analysis of images and tagged content posted on social media [24]
- Analysis of audio and visual data [5]
- Analysis in human-human and human-machine interactions [32]

Recent advances in multimodal sentiment analysis have concentrated on multimodal content (text, audio, visual, blogs, etc.) posted on social media to recognize sentiments [6, 21]. The information available on social media present in structured or unstructured form and processing such data is challenging task. A multimodal sentiment analysis has found application in various fields like Marketing, Business analytics, Box office prediction, Recommender system, Marketing intelligence, FOREX rate prediction etc. [1].

2. Multimodal Sentiment Analysis Terminologies

2.1 Forms of Modality

Fig.1 shows various easily available sources of sentiments like audio, text, images, facial expressions, EEG signals, keystroke patterns, etc. Microblogging [6] allows user to exchange their views about specific topic in the form of text, emoticons, images, videos. When emoticons are used in conversation, it intensifies the user's affective state. Keystroke dynamics, text, Photoplethysmography (PPG) signal to measure heart rate variation are introduced in [27], which described the user interaction with machine through keyboard is user and his emotional state dependent. This work uses web videos posted on social media like Facebook and YouTube for harvesting emotions. These videos are processed to extract audio, visual and textual content from it [5]. Electroencephalogram EEG waves are incorporated along with comments in textual format to predict user's preference towards particular advertisement [9]. The signature of a subject and brain

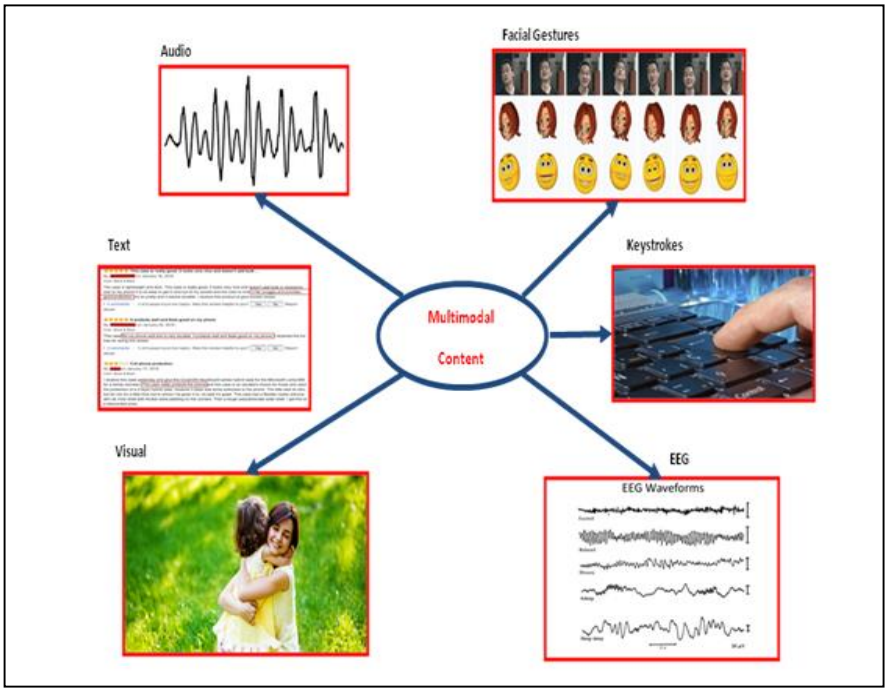

Fig.1: Multimodal data

signals (EEG) are correlated to each other and unique [7], hence can be used to authenticate and verify user's identity. Then audio and visual modalities can be used for automatic recognition of emotion [13]. Efficient multimodal framework is based on a fundamental unimodal sentiment analysis which employs only one modality like text, image, audio, etc. in analysis. Multimodal sentiment analysis is based on fusion of variety of information available from different sources. It combines data received from different modes for analysis process. The cost and complexity in analysis process varies as many features from different modalities are integrated to gain valuable insights [28]. This is due to different modalities have different features which are described below.

2.2 Characteristics of Multimodal data

- Data Representation: Every modality has unique data format and can be captured at different rates.
- Time: Time required for processing different media streams varies.
- Dependency: The modalities are correlated to each other or completely independent.
- Confidence Level: The confidence levels of different modalities vary in analysis process i.e. human crying can be easily recognized by audio data (has higher confidence) than video data.
- Cost: The initial data capturing and then processing incurs specific costs, which ultimately influences the fusion of different modalities

2.3 Information Fusion Techniques

Above characteristics influence the types of fusion carried out for analysis. Following Table 1 lists different fusion techniques along with its description [3].

Table 1. Information Fusion Techniques

Information Fusion Techniques	
Name	**Description**
Feature level/early fusion	It forms a common vector by fusing the features extracted from various modalities like audio, textual, image etc., and these combined features are used for analysis
Decision-level or late fusion	Each feature vector is processed by different classifier and classifier score is found as an output. This score is further concatenated with the score of other modality.
Hybrid fusion	It combines both feature-level and decision-level fusion methods
Model-level fusion	Model is built as per research needs and given problem space. This technique is based on observing data from various modalities to find the correlation between them.
Rule-based fusion methods	This technique applies various rules for fusing different modalities. e.g. Statistical methods, Majority voting and application specific custom-defined rules.
Classification-based fusion methods	It employs different classification algorithms for the multimodal content classification into previously defined classes

Estimation-based fusion methods	It is used to estimate moving object's state information. e.g. kalman filter, extended kalman filter

2.3 Multimodal Sentiment Analysis Framework

Fig. 2 shows general architecture for multimodal sentiment analysis. It consists of data collection, preprocessing, features extraction & vector representation, multimodal vector, classification and sentiment modules. Initial step in multimodal SA starts with data collection module. This data consists of text, audio, visual, etc from dataset or social media. In preprocessing phase, multimodal SA prepares data gathered from data collection module for further processing. Each modality has unique features and needs to be processed with separate attention. During feature extraction and vector representation phase, features are extracted from various modalities independently and represented in the form of vectors. Later these features vectors are amalgamated into multimodal feature vector to recognize sentiment of user content. In classification phase, data captured and processed through various phases is categorized into positive, negative or neutral sentiments with the help of multimodal feature vector. Last module in this architecture visualizes and presents user's sentiments based on above analysis.

Fig.2: Multimodal Sentiment Analysis Framework

3. Multimodal Sentiment Analysis Approaches

3.1 Traditional

Huang [6] had applied 'multimodal joint sentiment topic model (MJST)'. This technique was especially for analyzing sentiment and topic from microblogs using latent dirichlet allocation (LDA). It also incorporated emoticons and microbloggers personality, as user expressed their hidden feeling through emoticons and user's personality influences his messages. Zadeh et al., [8] introduced Multimodal Opinion-Level Sentiment Intensity (MOSI) dataset, Multimodal Dictionary to represent correlation between facial gestures and spoken words. The interaction pattern is categorized as neutral, emphasize, positive and negative. The Multimodal Dictionary consists of two entries namely words and gestures. If word and gesture co-occurred then entry is made in dictionary. The nu-SVR6 is used for training a prediction model and a fivefold cross-validation methodology for testing.

Amir Zadeh et al. [9] presents a novel multimodal framework which incorporates EEG signals from user while watching videos and textual data in the form of comments. EEG signals are pre-processed with the help of Independent component analysis (ICA) and rating prediction is performed using random forest regression (RFR). Sentiment score from textual comments is computed by employing naïve bayes classification (NBC). Fusion takes place at decision level. The system is tested by 3 different regression models where RFR out-performs. RajkumarSaini et al. [7] has introduced a novel 'multimodal user identification with verification scheme'. This work exploited the two correlated biometric features, i.e. EEG- the brain signals and the signature. Feature extraction process retrieves PHOH-Pyramid Histogram of Orientation Gradients for signature data and gamma frequency band for EEG signals. Then sequential HMM classifier is applied to both the feature vectors. Ultimately identification and verification is accomplished in three cases: signature data, EEG signals and fusing signature with EEG data. This work first employs early fusion to concatenate brain signal and signature features into a common feature vector. In late fusion, feature vectors extracted from signature and brain signals are evaluated separately by respective classifier and final score is compared with the threshold value. The verification process compare X user against the highest probability of either PHOG or EEG. The EEG model exceeded PHOG in the user verification process.

Poria [5] has proposed novel multimodal sentiment analysis model which incorporates three modalities namely audio, visual and textual as an input. It has employed YouTube dataset, SenticNet and EmoSenticNet. Using Luxand FSDK 1.7 facial characteristic points (FCP) are extracted which are used to construct visual features. The OpenEAR open source software is used to have audio features and concept level sentiment analysis for textual features. These features are fused using feature level and decision level techniques. The performance is evaluated using classifiers like SVM, ANN and ELM where ELM gives best results over ANN with respect to accuracy and training time parameters.

2.2 Deep learning

Yu et al. [11] used convolutional neural-networks with 'Deep learning methods' for analysing the sentiment from textual and visual data. Chinese microblogsSinaWeibo is considered as a datasets. Textual feature extraction process segments textual data into Chinese phrases and Chinese character. Using word2vec tool two word vector vocabularies are obtained which are further processed using DNN to get DNN_text i.e. deep convoluted feature vector. The DNN model is employed to extract deep visual feature vector DNN_image. If text includes text and images then late fusion is applied otherwise only textual feature vector is utilized in sentiment analysis task. This proposed model is compared with baseline cross media bag of words model (CBM) and evaluated using two class (positive and negative) and three class (positive, negative and neutral) evaluations. The DNN model for text and images gives good performance.

A work presented by Tzirakis et al., integrated a convolutional neural network (CNN) and long short term memory (LSTM) for audio and visual content to recognize emotions. Initially audio features are extracted using a CNN and visual features are obtained by applying deep residual network of 50 layers. Each modality network is pre-trained using two layers LSTM with 256 cells to capture temporal structure of both modalities. While training multimodal network two layers of LSTM are excluded and only 1280-dimensional audio features and 2048-dimensional visual features are taken into consideration. These features are concatenated into 3328-dimensional feature vectors and further processed using two layers LSTM with 256 cells. In comparisons to other models proposed system outperform in valence and arousal parameters [13]. Rakhee Sharma [30] has implemented multimodal sentiment analysis using deep neural network architecture. It uses text and image modalities for experiments. The experiments are performed on Large Movie Review dataset, Twitter image dataset and MOSI dataset. The text and image modalities are tested separately and compared in terms of accuracy.

TatoAnge [28] has employed semi-supervised deep multimodal architecture with user data to extract polarity of sentiments. This architecture extracts quiescent representation of a user from multimodal data using LSTM – Long Short-Term Memory, LSTM Auto Encoder, CNN- Convolutional Neural Network and DNN- Deep Neural Network. The model is experimented on Seempad multimodal dataset which includes text data with emotions and EEG. There are 5 branches: first one is LSTM accompanied by CNN which takes sequential text data, The second branch consists of LSTM Auto Encoder followed by CNN takes latent text data as input and extracts hidden structure of labeled or unlabeled data. The remaining 3 branches comprised of DNN for emotions, EEG data and other form of data. Each layer captures hidden features of each modality and combined into user latent vector to discriminate the polarity or predicting user's behavior. A new deep learning based triple layer sentiment analysis system [29] incorporates facial, verbal and vocal modalities. Each of these modalities is handled separately to acquire facial, verbal features and vocal characteristics of user and then integrated to accomplish multimodal analysis.

Fuhai Chen [14] introduced WS-MDL: a weakly supervised multimodal deep learning approach to predict sentiments of microblogs. The microblogs are rich in multimodal data like text, images and emoticons. In this work CNNs are trained using cheaply available noisy emoticons. Then probabilistic graphical modal is applied to remove noisy labels and to learn the relevancy between different modalities for inferring the confidence of label noise. Akshi Kumar [15] has proposed deep learning based sentiment analysis model for real time multimodal content. This model deals with text, image and info-graphic (i.e. mixture of text and image) modalities. It consists of four modules: 1. Discretization, 2. Text analytics, 3. Image analytics, 4. Decision module. The info-graphic modality is processed in discretization module using Google lens to separate the text content from the image and then forwarded to text analytics and image analytics module respectively. The sentiments from text data is determined using CNN augmented with SentiCircle whereas SVM with LBP features decides sentiments from images. The final sentiment polarity is predicted in decision module using Boolean OR operation. The proposed model for info-graphic modality outperforms text and image modality in terms of accuracy. A HDF-Hierarchical Deep Fusion [16] model employs text, image and social links for sentiment analysis. This modal aims to find correlation among these modalities using hierarchical LSTM and exploits social link with the help of weighted relation network.

2.3 Ensemble

The ensemble based multimodal sentiment analysis is proposed by Araque et al. [10]. This methodology divided the proposal into six tasks. First task has built a deep-learning based classifier for sentiment analysis using linear machine learning algorithm and a word

embeddings model. This is treated as an initial classifier for comparing successive results. Next task proposed two ensemble based techniques to alloy a baseline classifier with other surface classifiers. The surface and deep features are combined using two models to integrate information received from various sources in third task. In the fourth step taxonomy is introduced to classify proposed and the different existing models from the literature. Next experiments are conducted to test and compare the performance of these models with the deep learning based classifier. Finally, a statistical analysis of proposed models in comparisons with original baseline assures improved F1-Score.

SoujanyaPoria et al. [12] introduced ensemble based multimodal framework for extracting user opinion and emotions from video content. It employs YouTube videos for sentiment analysis and USC IEMOCAP database for multimodal emotion analysis. Visual feature is constructed by combining features obtained by CLM-Z and GAVAM. 6373 audio features consist of low level descriptor (LLD) are retrieved by openSMILE. To extract feature from textual content a convolutional neural network (CNN) is used. Then these features are classified by SVM classifier. It performs fusion at feature level and decision level. Multiple kernel learning (MKL) utilized in audio, visual and textual features shows significant improvement in fusion process.

Mahesh G. Huddar [31] performed experiment on 2199 opinion utterances from MOSI dataset in which 1751 utterances are used for training and 458 for testing phase. In 4.2 seconds first video is segmented into average length of utterance, then features are extracted from MOSI dataset and normalized with min-max transformation technique. Audio, visual features are selected using feature selection method and assembled to form multimodal vector. Proposed ensemble model is implemented using support vector machine, Decision trees, Random forest classifier, KNN and Logistic regression and compared with voting and averaging baseline approaches.

2.4 Fuzzy logic based

The term fuzzy implies uncertainty and fuzzy logic mean reasoning like a human. Fuzzy logic is capable of certain or definite response (output) from given imprecise input. A fuzzy based approach [17] for multimodal sentiment analysis employs fuzzy logic to represent the polarity of sentiment from training text data. This fuzzy model is integrated with the information collected from the SenticNet and the General Inquirer vocabulary resources to infer polarity of other domain. A Convolution Fuzzy Sentiment Classifier [26] uses deep convolution neural network and fuzzy logic for multimodal sentiment analysis. Deep learning is applied to text, image and visual data individually to extract features of each modality. Later fuzzy classifier predicts the intensity of sentiment from four dimensional emotion space.

2.4 Quantum inspired

The existing sentiment analysis approaches concentrate on applying text based methodologies to form vector representation of documents. These methodologies are good in acquiring syntactic and semantic details from document but unsuccessful to retrieve sentiment content. YazhouZhang[25] introduces quantum inspired model (QSR) which focuses on incorporating sentiment and semantic information from documents. Experiments are performed on two publicly available Twitter dataset namely Obama-McCain Debate (OMD) and the Sentiment140. This work first extracts adjectives and adverbs as representative for sentiment phrases. Secondly, in quantum vector space single terms and sentiment phrases are modelled as a projector and finally enclosed into density matrix. The QSR model outperforms sentiment analysis conventional approaches.

A Quantum-inspired Multimodal Sentiment Analysis (QMSA) [18] framework comprises of two models namely, a QMR- Quantum-inspired Multimodal Representation and a QIMF- Multimodal decision Fusion strategy inspired by Quantum Interference. This

framework exploits text as well as image modalities. The first model (QMR) focuses on extracting precise semantic information from content features and on modeling correlation between various modalities in the form of density matrix. The QIMF model decides sentiment category based on double-slit experiment. The experiments are performed on two datasets built from Getty Images and Flickr photo sharing platform. The proposed framework gives best performance.

4. Challenges

Nowadays, a multimodal sentiment analysis is gaining more attention over traditional sentiment analysis due to vastly available multimodal data and becoming very popular research topic. It is observed that though lots of advancements are happening in this area, there are many issues and challenges need to be tackled to improve its performance. Some of the challenges in multimodal sentiment analysis are listed below:

- Sentiment Intensity Extraction: The existing computational strategies focus on extracting syntactic and semantic features from the user data. But there is lack of techniques to capture the polarity of sentiments expressed in various forms like expression, emoticons, gesture, etc. In reality sentiments plays important role in sentiment analysis process. [25]
- Multimodal features fusion: Each modality possesses a unique characteristic which helps in describing that specific modality. Hence same feature extraction techniques can't be applicable to all modalities (i.e. text, audio, visual. etc.). In multimodal sentiment analysis feature vectors are constructed for every modality separately and integrated to classify sentiments. To gain useful knowledge from different resources fusing all these feature vectors into single global vector is cumbersome task [5] [16].
- Domain Independent Sentiment Polarity Prediction: The existing MSA aims to predict sentiment polarity based on domain specific knowledge. It means MSA system trained using plain text content from document is unable to infer sentiments from microblogs. The Domain independent MSA makes system adaptable to the newly introduced domain [17].
- Semantic Gap: The semantic gap between low level data resource and high level semantic information is open an issue to researchers. An image is visual modality represented in the form of pixel. At low level these pixel has negligible meaning but when constructed into visual vector conveys useful information to user. A MSA should focus on minimizing this semantic gap [18].
- Cross-Modal Relationship: A large volume of data is available in the form of textual, visual, audio etc content for sentiment analysis. A user shares his opinion/view/sentiments simultaneously in different modality like textual feedback, facial expression, and gesture or voice modulation. When user is trying to express through various modalities, he aims to convey the same polarity of sentiment. The cross-modal relationship exploits similarities between different modalities for better sentiment prediction [5] [16].

5. Conclusion

In this paper we have discussed different resources of multimodal content and various fusion techniques to understand user's intention or attitude towards product or service. The user data is available in different forms namely plain text, image, audio, video, microblogs, social links etc. A general framework along with

challenges is introduced for multimodal sentiment analysis. A numerous computational approaches are employed for multimodal sentiment analysis like traditional, deep learning, ensemble, fuzzy and quantum-inspired. All these techniques have proved multimodal sentiment analysis gives better sentiment prediction than unimodal.

6. References

6.1. Journal Article

[1] Kumar Ravi, Vadlamani Ravi, "*A survey on opinion mining and sentiment analysis: Tasks, approaches and applications"*, Knowledge-Based Systems Vol 89, November 2015, pp. 14-46. .
[2] WalaaMedhat, Ahmed Hassan, HodaKorashy, "*Sentiment analysis algorithms and applications: A survey*", Ain Shams Engineering Journal (2014) 5, pp. 1093-1113
[3] SoujanyaPoria, Erik Cambria, Rajiv Bajpai, Amir Hussain, *"A review of affective computing: From unimodal analysis to multimodal fusion"*, Information Fusion 37 (2017) pp. 98–125.
[4] Mohammad Soleymani, David Garciab, Brendan Jouc,1, BjörnSchullere,Shih-Fu Changc, MajaPanticf, "*A survey of multimodal sentiment analysis*", Image and Vision Computing 65(2017) pp. 3–14.
[5] SoujanyaPoria, Erik Cambria, Newton Howard, Guang-Bin Huang, Amir Hussain, "*Fusing Audio, Visual and Textual Clues for Sentiment Analysis from Multimodal Content*", Neurocomputing174 (2016) pp. 50–59.
[6] Faliang Huang, Shichao Zhang, Jilian Zhang, Ge Yu, "*Multimodal learning for topic sentiment analysis in microblogging*", Neurocomputing 253 (2017) pp. 144–153.
[7] RajkumarSaini , BarjinderKaur, Priyanka Singh, Pradeep Kumar, ParthaPratim Roy, Balasubramanian Raman, Dinesh Singh, *Don't just sign use brain too: "A novel multimodal approach for user identification and verification*", Information Sciences 430–431 (2018) pp. 163–178.
[8] Amir Zadeh, Rowan Zellers, Eli Pincus, Louis-Philippe Morency, "*Multimodal Sentiment Intensity Analysis in Videos: Facial Gestures and Verbal Messages*", IEEE Intelligent Systems (2016) pp. 82-88.
[9] HimaanshuGaubaa, Pradeep Kumara, ParthaPratimRoya, Priyanka Singh, Debi ProsadDograb, BalasubramanianRamana, "*Prediction of advertisement preference by fusing EEG response and sentiment analysis*", Neural Networks 92 (2017) pp. 77–88.
[10]Oscar Araque, Ignacio Corcuera-Platas, J.Fernando Sánchez-Rada, Carlos A. Iglesias, "*Enhancing deep learning sentiment analysis with ensemble techniques in social applications*", Expert Systems With Applications 77 (2017) pp. 236–246.
[11]Yuhai Yu, Hongfei Lin, JianaMeng and Zhehuan Zhao, "*Visual and Textual Sentiment Analysis of a Microblog Using Deep Convolutional Neural Networks"*
[12]SoujanyaPoria, HaiyunPeng, Amir Hussain, Newton Howard, Erik Cambria, "*Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis*", Neurocomputing 261 (2017) pp. 217–230.
[13]PanagiotisTzirakis, George Trigeorgis, Mihalis A. Nicolaou, Bjorn W. Schuller, and StefanosZafeiriou , "*End-to-End Multimodal Emotion Recognition Using Deep Neural Networks*", IEEE Journal Of Selected Topics In Signal Processing, Vol. 11, No. 8, December 2017.
[14]Fuhai Chen, RongrongJi, Jinsong Su, Donglin Cao, YueGao, *"Predicting Microblog Sentiments via Weakly Supervised Multimodal Deep Learning*", IEEE Transactions On Multimedia (2018), Vol. 20, No. 4, pp. 997-1007.
[15]Akshi Kumar, KathiravanSrinivasan, Cheng Wen-Huang, Albert Y. Zomaya, "*Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data*", Information Processing and Management (2020) 102141, Vol. 57, No.1, pp. 1-25.
[16]JieXu, Feiran Huang, Xiaoming Zhang, Senzhang Wang, Chaozhuo Li, Zhoujun Li, Yueying He, "*Sentiment analysis of social images via hierarchical deep fusion of content and links*", Applied Soft Computing Journal 80 (2019) pp. 387–399.
[17]Mauro Dragoni, GiulioPetrucci, "*A fuzzy-based strategy for multi-domain sentiment analysis*", International Journal of Approximate Reasoning 93 (2018) 59–73.
[18]Yazhou Zhang, Dawei Song, Peng Zhang, Panpan Wang, Jingfei Li et al., "*A Quantum-Inspired Multimodal Sentiment Analysis Framework*", Theoretical Computer Science Vol 752, Dec 2018, 21-40
[19]Sujata Rani, Parteek Kumar, *"A Sentiment Analysis System to Improve Teaching and Learning*", IEEE Computer (2017), Vol. 50 , No. 5, pp. 36 – 43.
[20]Samuel Cunningham-Nelson, MahsaBaktashmotlagh, Wageeh Boles, "*Visualizing Student Opinion Through Text Analysis*", IEEE Transactions On Education(2019), Vol. 62, No. 4, pp. 305 – 311.
[21]Luis Martin-Domingo, Juan Carlos Martín, Glen Mandsberg, "*Social media as a resource for sentiment analysis of Airport Service Quality (ASQ)*", Journal of Air Transport Management 78 (2019) pp. 106–115.

[22]KashfiaSailunaz, RedaAlhajj, "*Emotion and sentiment analysis from Twitter text*", Journal of Computational Science 36 (2019) pp. 1-16.
[23]Gonzalo A. Ruz, Pablo A. Henríquez, Aldo Mascareño, "*Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers*", Future Generation Computer Systems 106 (2020) pp. 92–104.
[24]Ziyuan Zhao, Huiying Zhu, ZehaoXue, Zhao Liu, Jing Tian, Matthew Chin Heng Chua, Maofu Liu, "*An image-text consistency driven multimodal sentiment analysis approach for social media*", Information Processing and Management 56 (2019) Vol.56, No. 6, pp. 1-9.

[25]Yazhou Zhang, Dawei Song, Peng Zhang, Xiang Li, Panpan*Wang1"A quantum-inspired sentiment representation model for twitter sentiment analysis*", Applied Intelligence 49 (2019), 3093–3108.
[26]ItiChaturvedi, RanjanSatapathy, SandroCavallari, Erik Cambria, "*Fuzzy commonsense reasoning for multimodal sentiment analysis*", Pattern Recognition Letters 125 (2019) 264–270.

6.2. Conference Proceedings

[27]Anil Kumar K M, Kiran B R, Shreyas B R, Sylvester J Victor, "*A Multimodal Approach To Detect User's Emotion*", Procedia Computer Science 70 (2015) pp. 296 -303.
[28]TatoAnge, Nkambou Roger, DufresneAude, Frasson Claude, "*Semi-Supervised Multimodal Deep Learning Model for Polarity Detection in Arguments"*, IEEE International Joint Conference on Neural Networks (IJCNN) (2018),
[29]Vishal Shrivastava, VivekRichhariya, VivekRichhariya, VineetRichhariya, "*Puzzling Out Emotions: A Deep-Learning Approach to Multimodal Sentiment Analysis*", International Conference on Advanced Computation and Telecommunication (ICACAT) (2018).
[30]Rakhee Sharma, Le Ngoc Tan, Fatiha Sadat, "*Multimodal Sentiment Analysis Using Deep Learning*", 17th IEEE International Conference on Machine Learning and Applications(2018).
[31]Mahesh G. Huddar ;Sanjeev S. Sannakki ; Vijay S. Rajpurohit, "*An Ensemble Approach to Utterance Level Multimodal Sentiment Analysis*", International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), Dec 2018.
[32]AgnieszkaRozanska, Michal Podpora, "*Multimodal sentiment analysis applied to interaction between patients and humanoid robot Pepper*", IFAC PapersOnLine 52-27 (2019) pp. 411–414.