Multimedia Tools and Applications (2019) 78:24103–24119
https://doi.org/10.1007/s11042-019-7390-1

Sentiment analysis of multimodal twitter data

Akshi Kumar[1] · Geetanjali Garg[1]

Received: 8 October 2018 / Revised: 28 January 2019 / Accepted: 19 February 2019 /
Published online: 12 March 2019

Abstract
Text-driven sentiment analysis has been widely studied in the past decade, on both random and benchmark textual Twitter datasets. Few pertinent studies have also reported visual analysis of images to predict sentiment, but much of the work has analyzed a single modality data, that is either text or image or GIF video. More recently, as the images, memes and GIFs dominate the social feeds; typographic/infographic visual content has become a non-trivial element of social media. This multimodal text combines both text and image defining a novel visual language which needs to be analyzed as it has the potential to modify, confirm or grade the polarity of the sentiment. We propose a multimodal sentiment analysis model to determine the sentiment polarity and score for any incoming tweet, i.e., textual, image or info-graphic and typographic. Image sentiment scoring is done using SentiBank and SentiStrength scoring for Regions with convolution neural network (R-CNN). Text sentiment scoring is done using a novel context-aware hybrid (lexicon and machine learning) technique. Multimodal sentiment scoring is done by separating text from image using an optical character recognizer and then aggregating the independently processed image and text sentiment scores. High performance accuracy of 91.32% is observed for the random multimodal tweet dataset used to evaluate the proposed model. The research further demonstrates that combining both textual and image features outperforms separate models that rely exclusively on either images or text analysis.

Keywords Multimodal text · Twitter · Sentiment · Context-aware · Optical character recognition

1 Introduction

The current affordable and ubiquitous generation of Web provides substantial amount of opinionated social big data which facilitates decision making. Sentiment Analysis has gained

✉ Akshi Kumar
akshikumar@dce.ac.in

Geetanjali Garg
geetanjali.garg@dtu.ac.in

[1] Department of Computer Science & Engineering, Delhi Technological University, Delhi, India

importance as quick as a wink for tracking the mood/view of people by analyzing this unstructured, multimodal, informal, high-dimensional and noisy social data. It helps in gaining insights for a particular subject, topic, event or a matter within various domains such as market, business or government intelligence [30]. Formally, sentiment analysis is defined as the computational study of people's opinions, attitudes and emotions towards an entity [27]. It is a specialized type of Natural Language Processing (NLP) problem that relies on the analysis of the huge amount of user-generated online web content produced daily in social networks, blogs, e-commerce sites, and other user-editable web forums. Pertinent literature within the area report the use of various techniques such as machine learning, lexicon-based, hybrid techniques and concept-based techniques (contextual and ontology based) etc. to mine the opinion [31]. Out of these, the use of machine learning techniques to mine opinion especially on Twitter text data has been demonstrated the most [13, 20, 22, 39, 40]. Recently, visual communication using images to express views, opinions, feelings, emotions and sentiments has increased tremendously on social platforms like Flickr, Instagram, Twitter, Tumblr, etc. [11, 19, 22, 23].

Images are particularly powerful as they have cognition associated and visual experiences convey sentiments and emotions better. Consequently visual sentiment analysis has been of interest to researchers and it has been observed that deep learning techniques have outperformed the conventional machine learning techniques in analyzing the visual sentiment. Multimodal capabilities offered by popular social networking websites such as Facebook, Twitter, and Tumblr have further enabled mix of text and images in a variety of ways for better social engagement. The ascendant use of info-graphics, typographic-images, memes and GIFs in social feeds is a testimony to this. Visual content is interesting, engaging and effective. Visual content has both typographic as well infographic content. Typography deals with arranging size, style and weight of the right typoface to provide a visually pleasing format of the text and helps in holding reader's attention. Infography deals with graphic representation of the data to make it easier to perceive.

As discussed, text-driven sentiment analysis has been widely studied [3, 14, 27, 28] and few pertinent studies which report visual sentiment analysis of images are available in literature [5, 7, 19, 49, 54–56]. But, much of the reported work has analyzed a single modality data whereas multiple modalities of text and image remain unexplored. Moreover, human expressions are extremely complicated as statements, images and their mix can convey a wide range of emotions, and often require context to fully understand. Thus, the study to comprehend this text-image relationship is imperative as this combination can modify or enhance the semantics and consequently the sentiment. For example, consider the multimodal text given in the Fig. 1. Here the image of a growling lion depicts a negative, beastly behavior (negative sentiment polarity) which is *modified* by the textual content "*Go Hunt Your dream*" which is a motivational, positive statement (positive sentiment polarity).

Now, consider another example of multimodal tweet given in Fig. 2. Here the rainbow colored key- house depicts a happy home with positive vibes and the text "*Love has just checked in*" strengthens this happiness emotion and thus the positive polarity of the sentiment.

Motivated by this polarity shifts and depths due to content modality, we propose a model to analyze sentiments from multimodal Twitter data which will facilitate visual listening for social media analytics. The model analyzes the incoming tweet for its modality (text, image or image with text) and based on it forwards it to the respective processing module. The model has five modules:

- *Data Acquisition Module:* Input tweet to the model

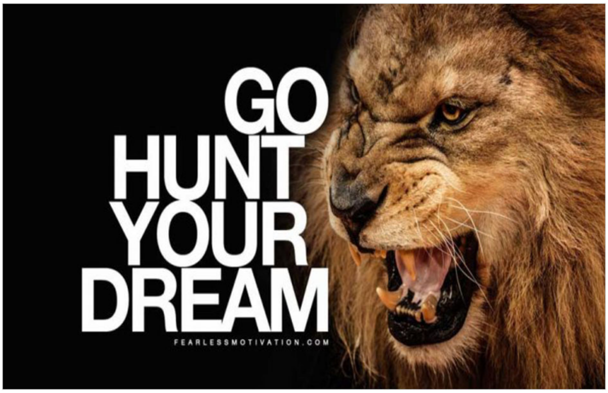


Fig. 1 Example Multimodal Text with sentiment modification

- *Image Module:* SentiBank and Regions with Convolution neural network (R-CNN) along with Senti-Strength based on the model given by Mandhyani et al. [37]
- *Text Module:* A hybrid of Lexicon and Machine Learning techniques (Gradient Boosting and SentiCircle [44])
- *Multimodal Module:* Image sent to Image Module for sentiment analysis and text is retrieved using Computer Vision API for Optical Character Recognition (CV/OCR) and then sent to the Text module for sentiment analysis;
- *Aggregation Module:* Resultant scores from Text and Image modules are combined to produce aggregate sentiment score for the multimodal tweet.

The novelty of the research is twofold. Firstly, the model is able to analyze sentiment for any incoming tweet, i.e., textual, visual or info-graphic and typographic. Secondly, it proposes a hybrid technique (Machine learning and lexicon) for context-aware textual sentiment analysis. The robustness of the individual modules is evaluated using benchmark datasets (STS-Gold for Text module [43]; Flickr 8k for Image module [17]) and the proposed model is evaluated

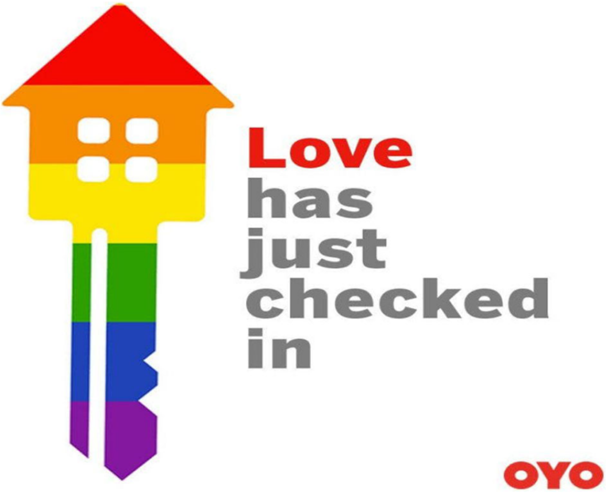


Fig. 2 Example Multimodal Text with sentiment strengthening

using random multimodal tweets collected on the recent topic related LGBT verdict of Indian Penal Court (IPC) section 377 in India (#section377). The results have been evaluated using accuracy as the performance metric.

The rest of the paper is structured as follows: Section 2 briefly discusses the pertinent work within the domain of textual, image and multimodal sentiment analysis followed by Section 3, which illustrates the proposed multimodal sentiment analysis model and its working details. Section 4 explicates the results and provides an analysis of the same. The final section, Section 5, concludes the study and expounds upon the scope of future work.

2 Related work

The term 'Sentiment Analysis' was initially witnessed in the published work [9] in 2003 and since then both primary [11, 13, 19–23, 25, 27–29, 31, 32, 39, 40] and secondary studies [24, 26, 30] have been reported across pertinent literature. Twitter, currently the most famous micro- blog connects people across the globe and has high level of user involvement. It has gradually emerged as a huge source of sentiment-rich data. Kumar and Jaiswal [22], have compared two social media (Twitter and Tumbler) for sentiment analysis and analyzed the performance using soft computing techniques. Also, literature is well-equipped with studies pertaining to sentiment analysis using machine learning paradigms on specifically textual user generated online content on social media [40] showed that emoticons could be used to collect a labeled dataset for sentiment analysis. Golder and Macy [14] investigated temporal patterns in emotion using tweets, and Bollen et al. [3] investigated the impact of collective mood states on the stock market.

Explosive growth of using images to express opinions in Microblog makes only text-based sentiment analysis an obsolete technique to understand the sentiments of users. Soleymani et al. [47] discusses the current work, challenges and opportunities in the field of multimodal sentiment analysis in detailed manner. There are many images in social networks that have similar emotional but different usual contents and opinion mining on these images is a challenging task. Besides the text based sentiment analysis, the image based sentiment analysis becomes important. The research in this field fall under three areas which are: aesthetics [8, 18, 38], emotion detection [33–36, 48, 51, 52, 54–56] and sentiment ontology [5]. Low level features of image are used in Emotion detection to detect the emotion in an image. J. Yanulevskaya et al. [51, 52] proposed a system to categorize emotion using low-level features. Machajdik and Hanbury [36] represented emotional content of an image by extracting low level features. Lu et al. [35] explored the relationship between emotions and shapes. Zhao et al. [54–56] proposed a principles-of-art-based emotion features (PAEF) which attained good performance in affective image classification and affective image retrieval. However, low-level image features are limited in the large-scale image sentiment detection. Visual Sentiment Ontology (VSO) was proposed by Borth et al. [5] which is based on such visual concepts which are strongly related to sentiments. They used high-level Adjective Noun Pairs (ANPs) which are strongly relevant with sentiment instead of the low-level features. Then they proposed SentiBank, which is used to detect the presence of ANPs in an image. Experiments are done on image tweets which demonstrate significant improvement in accuracy. Yang et al. [50] proposed an image discovery framework considering textual, visual and social features. They have proposed the Visual-Social-Textual Rank (VSTRank) algorithm to calculate the importance score for each image to identify images that have emotional content. Siersdorfer et al. [46] analyzed the relation between the sentiment of the metadata content to images and their visual content in social environment. For sentiment analysis of images authors' have extracted numerical values based on their textual metadata. Authors have used Support Vector Machine (SVM) to build machine learning model for sentiment analysis of images. It is concluded that the visual features of images can support in predicting the polarity of sentiments. Girshick et al. [12] have built a model "R-CNN: Regions with CNN features". Color histogram and SIFT-based features of images are used to train a sentiment polarity classifier. Gajarla and Gupta [11] used deep learning to predict the emotion depicted by an image. They have classified the Image categories as Love, Happiness, Violence, Fear and Sadness. They have collected data set from Flickr API. They have used three techniques which are SVM on high level features of VGG-ImageNet, fine-tuning on pre-trained models like RESNET, Places205 VGG16 and VGG-Image Net. Kumar et al. [23] proposed a visual sentiment framework using a convolutional neural network. They have used Flickr images for training the model and Twitter images for testing the proposed model.

Successful results in multimodal sentiment analysis have been achieved using non-negative matrix factorization [49] and latent correlations [19]. Chen et al. [7] investigated the image posting behavior of social media users and found in their study that two thirds of the participants added an image to their tweets to enhance the emotion of the text. You and Luo [53] have analyzed the online sentiment changes of Twitter users considering both the textual and visual content. They also show the correlation between textual content and visual content. Poria et al. [41] gave a novel methodology for multimodal sentiment analysis, which consists in harvesting sentiments from Web videos by demonstrating a model that uses audio, visual and textual modalities as sources of information. Katsurai and Satoh [19], exploit correlations among multiple views: visual and textual views and a sentiment view the correlations being constructed using SentiWordNet. Dataset for this correlation is collected from Flickr and Instagram images. A novel Unsupervised Sentiment Analysis (USEA) framework for social media images has been proposed in a work [49] which uses relations between visual content and relevant contextual information. Authors found out both textual information and visual content for sentiment based image clustering in a non-negative matrix factorization framework. Cai and Xia [6] have used convolution neural networks (CNN) for sentiment analysis of multimodal tweets which consist of text and image. Two individual CNN architectures are used for learning textual features and visual features, which are combined as input of CNN architecture for exploiting the internal relation between text and image. Hare et al. [16] developed a system to analyze streams of image data. They explored trends in visual artefacts in the image tweets.

To the best of our knowledge, till date no work has been done on multimodal text tweets where the text is embedded in the image as a typographic or infographic visual content. This research proffers a model-based solution to this open problem of research within the sentiment analysis domain. The proposed multimodal sentiment analysis model is illustrated in the following section.

3 The proposed model for multimodal sentiment analysis

Multimedia involves multiple modalities of text, audio, images, videos, drawings etc. Twitter has transformed from a text based micro-blogging (very micro) platform to a visual one gradually. Recent studies also claim that the tweets with image links get 2x the engagement rate of those without [7, 19].

The proposed model takes into account both the modalities, text and image independently and their combination to analyze the sentiment in tweets. The incoming tweet is firstly examined for its modality type, that is, whether it is an image, a text or a multimodal text (image + text = typographic or info-graphic). The further processing is done on the basis of the identified modality type. For an image only tweet, the image module is implemented which uses an existing model of SentiBank [4] and R-CNN to determine the sentiment polarity and sentiment score of the image. For a text only tweet, the text module after pre-processing employs a machine learning based ensemble method (Gradient boosting) to classify the tweet in to one of the three polarity categories, namely, positive, negative or neutral. Post this; a lexicon based approach (SentiCircles), which captures contextual semantics, is used to determine the sentiment polarity and strength of the tweet. This polarity and strength obtained separately from the machine learning and lexicon based techniques is combined to do a scoring of sentiment which has the range [−3, 3]. Subsequently, for an image with text, the text is detected and extracted using a computer vision API and recognized using optical character

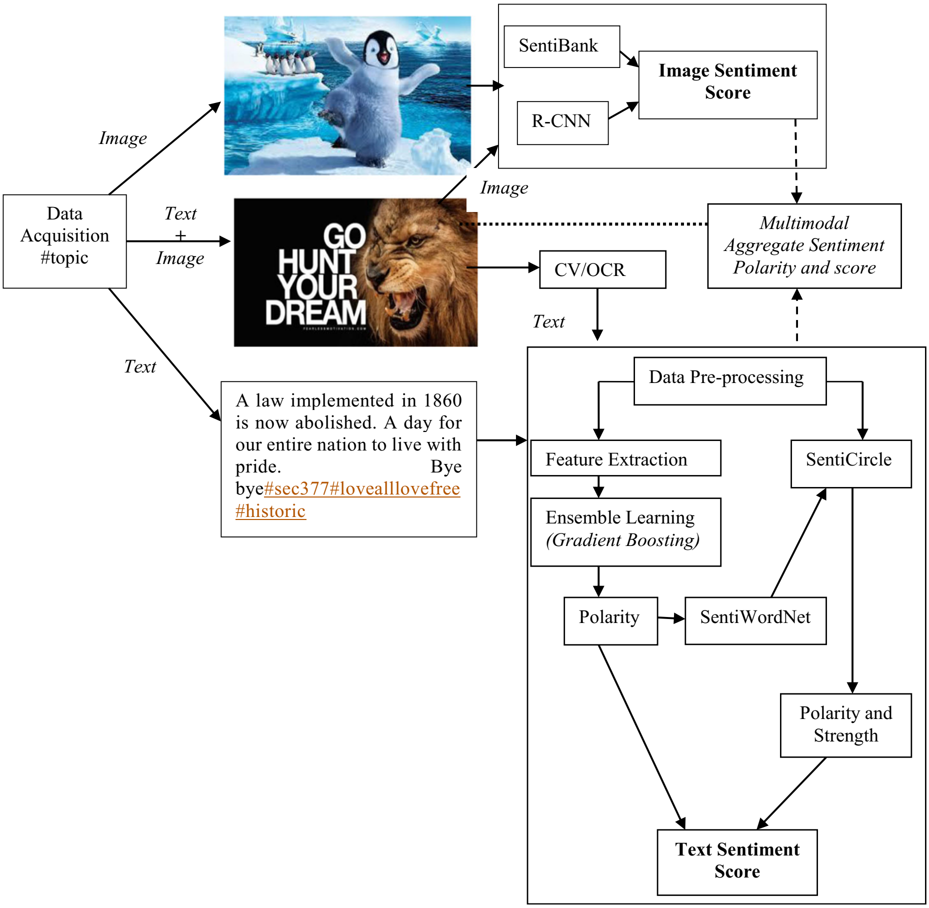


Fig. 3 Systematic flow of the proposed model

recognition approach. The recognized text is then processed in the text module whereas the image is sent to the image module for processing. The resultant scores from both these modules are combined to give an aggregate sentiment polarity and score. Figure 3 illustrates the systematic flow of the proposed model.

The following sub-sections expound the details:

3.1 Data acquisition

To evaluate the system using the aforesaid classification techniques tweets pertaining to a topic (#topic) are extracted from the publically available Twitter datasets using its API. 8000 multimodal tweets were collected on the recent topic related LGBT verdict of Indian Penal Court (IPC) section 377 in India using hashtag #section377.

3.2 Image Sentiment Analysis

The image only content is processed using the SentiBank, Regions with Convolution neural network (R-CNN) and SentiStrength to obtain a sentiment score within the range [−2, 2]. This analytic model to determine sentiment in images was given by Mandhyani et al. [37] in the year 2017.

- *SentiBank:* It a large-scale visual sentiment ontology which includes 1200 semantic concepts and corresponding automatic classifiers. Each concept is defined as an Adjective Noun Pair (ANP), where adjective depicts the emotion for a specific object/scene described by a noun [4].
- *R-CNN:* R-CNN is one of the popular and efficient object detection models. R-CNN uses selective search for reducing the number of bounding boxes that are fed to the classifier. Selective search uses local cues like texture, intensity, color etc. to generate all the possible locations of the object. After this, these boxes are fed to CNN based classifier. In R-CNN the convolution neural network is forced to focus on a single region at a time because that way interference is minimized as it is expected that only a single object of interest will dominate in a given region. The regions are fed to a CNN for object detection.
- *SentiStrength:* SentiStrength is a sentiment analysis program which estimates the strength of positive and negative sentiment in short web texts, even for informal language. Words are classified and rated based on positive and negative strength, i.e., −1 (not negative) to −5 (extremely negative) and 1 (not positive) to 5 (extremely positive).

The pseudo-code to compute the image sentiment score is given in Fig. 4:

```
1.  Nouns from 1200 ANPs are separated
2.  Objects from 200 RCNN classes are taken
3.  Compute the distance between the noun and objects of 200 RCNN classes by using
    Radial basis function(RBF)
4.  New ANP weights are multiplied with SentiStrength of adjectives for each ANP.
5.  Sum-up all ANPs to get the sentiment score within the range [-2,2]
```

Fig. 4 Pseudo-code for image sentiment scoring

3.3 Textual sentiment analysis

The textual sentiment analysis is a multi-step process which consists of the following:

- *Data Pre-processing:* Data pre-processing is done for cleaning and transforming the data for relevant feature extraction. The HTML entities in the tweet are decoded (Eg. &amp is changed to &), URLs were removed, expressions corresponding to retweet (RT) at the beginning of the tweet are removed, contractions present in the tweet are replaced by their extended words (Eg, "I'll" is replaced with 'I will), punctuations present including hast-tag '#'etc., are removed. Further, three or more repetitive occurrences of a character are replaced with a single character. For example, *'happppy'* is changed to *'hapy'*. Terms in the tweet which contains only digits are removed. Extra spaces in the tweet are removed and finally, all the characters of the tweet were changed to lowercase. Additionally, all non-ASCII-English characters were removed, to keep the domain of the data specific to the English language. Part of Speech tagging is also done to extract common structural patterns such as verb, adverb, adjective and noun.
- *Feature Extraction:* This step identifies the characteristics of the datasets that are specifically useful in detecting sentiments. The classical bag-of-features framework is utilized. We form a list of tweet words from the tweets in the corpus, which are tagged as a noun, verb, adjective, adverb or pronoun using the Part-of-speech (POS) tagger provided by NLTK [2]. Now the frequency distribution of each tweet word in this list is obtained and the top 5000 most common words are considered. These words constitute the bag-of-words which are to be used as feature words to find the unigrams. Next, we form a feature vector corresponding to each tweet. The features used are:
  - Unigrams: presence/absence of feature words
  - Part-of-Speech(POS)features: count of nouns, verbs, adjectives, adverbs, interjections and pronouns
  - Negation: count of occurrences of negation word 'not'
  - Count of Emoticon features: Various combinations of punctuation marks have been mapped into six classes of emoticons:-Smiley(:),:-), (:), laugh(: D, xD), love(<3,:*), wink(;),;-D), frown(: -(,:(), and cry(:'() and their count is taken as feature
  - Count of elongated words(e.g. yummmy)
  - Count of capitalized words
  - Length of message.
- *Ensemble Learning:* An iterative learning model, gradient boosting is then used to train the textual sentiment analysis module. The gradient boosting is meta-model which consists of multiple weak models whose output is added together to get an overall prediction. The evaluated polarity is input the SentiWordNet [10] to obtain the respective sentiment score. SentiWordNet contains sentiment scores for all WordNet entries.
- *SentiCircle:* Each cleaned tweet is firstly tokenized, and each token is POS tagged using NLTK. Each token is lemmatized using WordNet Lemmatizer [15] and then stemmed to its root form using Porter Stemmer [42]. Based on the POS tag assigned to each token, it is scored using SentiWordNet. SentiWorNet offers a fixed, context-independent, word-sentiment orientations and strengths. SentiCircles considers the contextual co-occurrence patterns to capture conceptual information and update strength and polarity in sentiment lexicons accordingly.
  - Scoring from SentiWordNet

    - If the POS tag matches one of the tags in SentiWordNet for that term, then all positive and negative scores for that word corresponding to that tag are weighted average inversely according to their sense number separately. Else all positive and negative scores corresponding to that word are averaged.
    - If positive and negative scores are unequal, then higher of them is returned with appropriate sign, else machine learning output polarity is considered in deciding. If polarity is positive, then positive score is returned and if the polarity is negative, then negative score is returned. For neutral output of machine learning, positive score is returned.

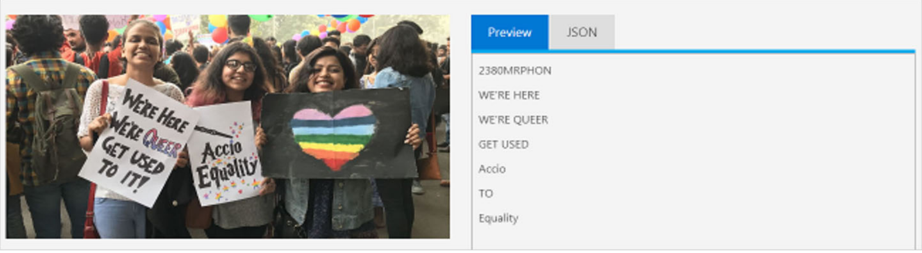
  - Negation handling terms that are preceded by any of the negative words listed in General Inquirer[1] under the NOTLW category1, have the sign of their score reversed. For example, in the tweet "*Uber Premier is not amazing!*", the term "*amazing*" is preceded by a negation. Therefore, instead of using its original sentiment score (0.75 in the SentiWordNet lexicon for example, this score negated (−0.75)
  - Term-Context vector is then created, that is for each word, and a vector of words that appear in context of the given word is formed. If the given word appears in any other tweet and that tweet matches the current tweet in having at-least one common user, topic etc., then all the words in that tweet is considered as part of the context-vector of the given word.
  - After forming term-context vector for each term in the tweet corpus, corresponding values of TDOC, θ, x and y are determined for each of the context terms in the context-vector of the term [44]
  - For each term, its sentiment polarity and strength is calculated by finding geometric median of all its context-terms using Weiszfeld's algorithm [1].
  - For each tweet, its sentiment polarity and strength is calculated by finding geometric median of all its terms using Weiszfeld's algorithm.

- Once the polarity from the ensemble learning algorithm and polarity & strength from SentiCircles is determined, these values are combined to determine the text sentiment score which has the range [−3, 3].

3.4 Multimodal text: text in image

For typographic or infographic multimodal text, we have used the Computer Vision API to extract text using OCR from the image. This process of text retrieval from image comprises of three sub-components viz. text detection, text extraction and text recognition. Text extraction is a crucial step in improving the accuracy and quality of the concluding recognition output. It aims at segmenting text from background that is to isolate text pixels from those of background. An effective text extraction method facilitates the use of commercial OCR without any amendments. For this we have used Open CV version3.4.2**,** EAST text detector proposed by Zhou et al. [57] which is a deep learning model, based on a novel architecture and training pattern. It is capable of running at near real-time at 13 FPS on 720p images and obtains state-of-the-art text detection accuracy. The extracted text is passed through an OCR for recognition.

[1] www.wjh.harvard.edu/~inquirer



2380MRPHON
WE'RE HERE
WE'RE QUEER
GET USED
Accio
TO
Equality

Fig. 5 Sample text extraction using the Computer Vision API

OCR is conversion of images of typed, handwritten or printed text into machine-encoded text. OCR first preprocess images by techniques like de-skewing, line & word detection, layout analysis, character segmentation etc. to improve recognition accuracy. The character recognition is generally done in two passes. The output of the first pass is transferred to second pass which is a kind of adaptive recognition. Second pass of recognition uses letter shapes recognized with high confidence on the first pass to recognize better the remaining letters on the second pass. OCR sometimes uses a post processing step which makes use of dictionary to improve upon accuracy. Figure 5 shows the sample text extraction using the API.

The text thus recognized is then sent to textual sentiment analysis module for determining the text sentiment score whereas the image is sent to the visual sentiment analysis module for finding the image sentiment score. These individual scores are combined to produce the aggregate sentiment score for the multimodal tweet. Figure 6 depicts the concept of multimodal sentiment analysis using optical character recognition.

4 Results and analysis

To investigate the robustness of the proposed model, the individual text and image modules are validated using benchmark datasets and the model is evaluated for random multimodal tweets. The empirical analysis is thus broadly divided into three parts; (i) Image sentiment analysis on benchmark Flickr 8 k dataset, (ii) Text sentiment analysis on benchmark STS-Gold dataset, and (iii) Multimodal text (text + image) sentiment analysis using randomly collected tweets on the selected topic.

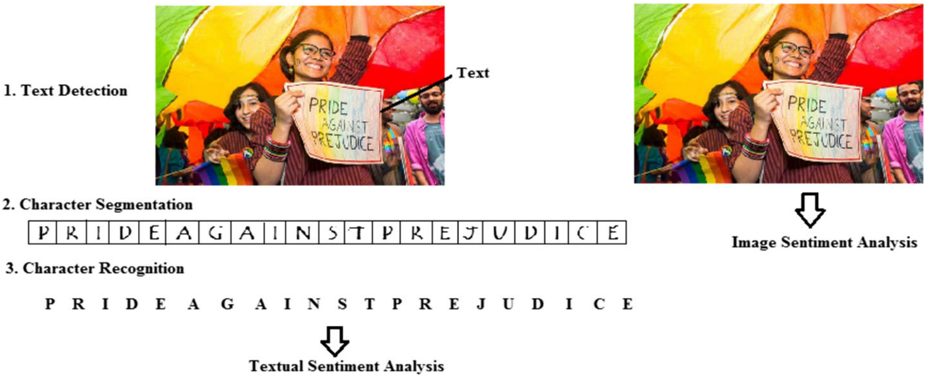

1. Text Detection
PRIDE AGAINST PREJUDICE
Text
PRIDE AGAINST PREJUDICE
2. Character Segmentation
P R I D E A G A I N S T P R E J U D I C E
Image Sentiment Analysis
3. Character Recognition
P R I D E A G A I N S T P R E J U D I C E
Textual Sentiment Analysis

Fig. 6 OCR and multimodal sentiment analysis

Table 1 Performance accuracy of image sentiment analysis techniques

Technique	Accuracy
SentiBank	71%
SentiBank+RCNN	76.04%

4.1 Image sentiment analysis

Image Sentiment is determined using a hybrid of SentiBank and RCNN. Flickr 8 k, a publically available dataset which comprises of images from flicker website is used to train and test the performance of RCNN for object detection. SentiBank consists of 1200 trained visual concept detectors providing a mid-level representation of sentiment. The results were evaluated initially by only using SentiBank technique and then using a combinational technique of SentiBank with RCNN. Table 1 depicts the performance accuracy.

Figure 7 shows the results graphically.

4.2 Text Sentiment Analysis

For textual analysis step we used STS-Gold, a standard dataset for Twitter sentiment analysis created by Saif et al. [45]. It contains a total of 2206 tweets, out of which 1402 are negative, 632 are positive and 77 are neutral. The performance of the text sentiment analysis module is evaluated using three approaches, the lexicon-only (SentiWordNet) technique, machine learning approach (Gaussian Naïve Bayesian, Decision Tree, Random Forest and Gradient Boosting) and hybrid approach. Table 2 depicts the accuracy results.

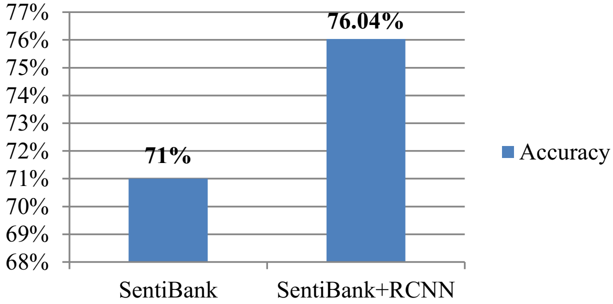


Fig. 7 Accuracy of image sentiment analysis techniques

Table 2 Performance accuracy of text sentiment analysis techniques

Approach	Technique	Accuracy
Lexicon Based	SentiWordNet (SWN)	59.43%
Machine Learning Based	Gradient Boosting (GB)	78.01%
	Decision Tree (DT)	71.35
	Random Forest (RF)	74%
	Gaussian Naïve Bayesian (NB)	64.45%
Hybrid	SentiWordNet +Gradient Boosting (SWN + GB)	82.21%

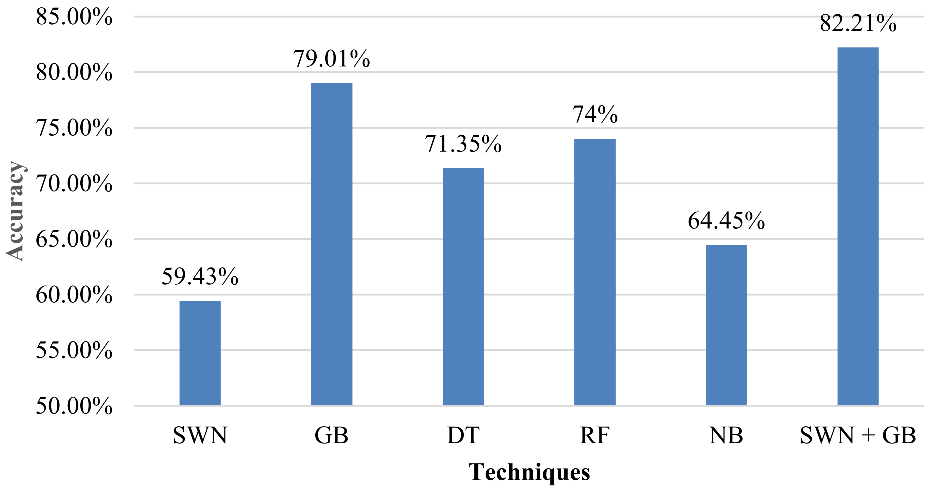


Fig. 8 Accuracy of text sentiment analysis techniques

Figure 8 shows these results graphically.

4.3 Multimodal Sentiment Analysis

Table 3 depicts the generic characteristics of text and image tweets:

8000 random multimodal tweets on the recent topic related LGBT verdict of Indian Penal Court (IPC) section 377 in India (#section377) were extracted. The distribution of modalities in these tweets is shown in the Fig. 9.

Table 3 Twitter text and image generic characteristics

Text Characteristics	Tweet length	• 280 characters • This limit is Not for Japanese, Korean, Chinese tweets
	Retweet	• No limit on number of times a tweet can be retweeted. • Only recent 100 people who retweeted a tweet will be shown.
	Tweet Limit	1000 tweets/day/Person
	Number of Language supported	40
Image Characteristics	GIF support	• Animated GIFs cannot be included in tweet with multiple images • We can send only 1 GIF per tweet • GIF in full is attached • Photos and GIFs attachment do not count towards character limit • Supports looping GIF
	Image limit	• We can select up to 4 photos to tweet at once • Up to 25 stickers can be attached to a photo
	Format of image tweet	• It accepts GIF,JPEG,PNG • Does not accept Bmp, TIFF
	Size of image tweet	• 5 Mb photo • 5 Mb GIF on mobile • Up to 15 Mb on web

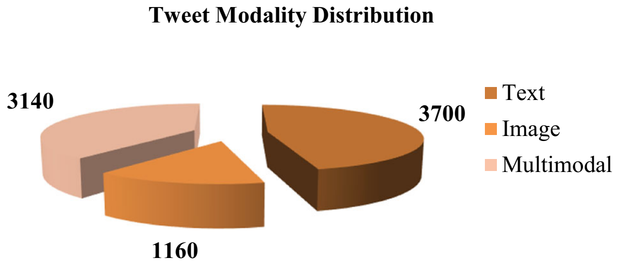


Fig. 9 Distribution of tweet modality types

The performance of the proposed model is evaluated for these multimodal tweets and the accuracy results for the same are shown in Table 4:

Figure 10 shows these results graphically.

5 Conclusion and future scope

Images are more expressive than text and at the same time text embedded or represented as an image further defines this power of expressiveness. This research proposed a model for sentiment analysis to capture this expressiveness for text in an image, both typographic or infographic, as sentiment polarity and strength. The multimodal sentiment analysis model for Twitter offered novelty in two ways. Firstly, it was able to handle multi-modalities in tweets, that is, text and image individually and text in an image to analyze the sentiment and secondly, the textual sentiment analysis was based on a hybrid of context-aware lexicon and ensemble learning. The model will serve as a visual listening tool for enhanced social media monitoring and analytics. The performance results were motivating and improve the generic sentiment

Table 4 Performance accuracy of proposed model

Module	Accuracy
Image module	77.63%
Text module	84.62%
Multimodal (proposed)	91.32%

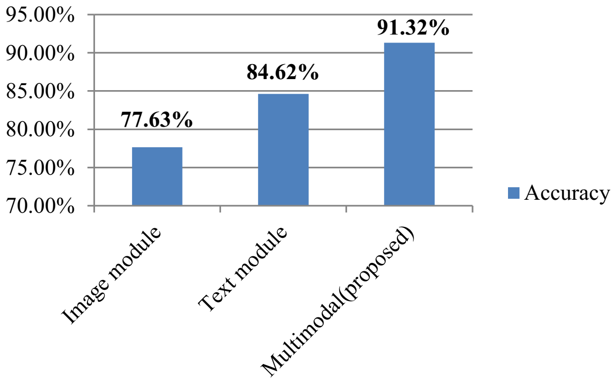


Fig. 10 Accuracy of proposed model

analysis task. The primary limitation of the model is that the text recognition is restricted by the capability of the Computer Vision API. Moreover, as social media is an informal way of communication, multilinguity (code-mix and code-switch languages, for example a mix of English and Hindi, a native Indian language) is widely seen, but such content (text or text in image) could not be processed. Also, the OCR has a limited capability for handwritten text recognition and suffers from reduced accuracy with 'lack of contrast' images where the text color and the background color are almost similar. In this research, only the text and image modality type had been considered whereas other modalities such as animated GIFs and memes define an open problem within the research domain. Also, the use of word embeddings and deep learning for context-aware sentiment analysis of text can be explored.

References

1. Aftab, K., Hartley, R. and Trumpf, J., 2015. Generalized weiszfeld algorithms for lq optimization
2. Bird S, Loper E (2004) NLTK: the natural language toolkit. In: Proceedings of the ACL 2004 on Interactive poster and demonstration sessions (p. 31). Association for Computational Linguistics
3. Bollen J, Mao H, Zeng X-J (2011) Twitter mood predicts the stock market. J Comput Sci 2(2011):1–8
4. Borth D, Chen T, Ji R, Chang SF (2013) Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In: Proceedings of the 21st ACM international conference on Multimedia (pp. 459–460). ACM
5. Borth D, Ji R, Chen T, Breuel T, Chang S-F (2013) Large-scale visual sentiment ontology and detectors using adjective noun pairs. In: Proceedings of the 21st ACM international conference on Multimedia, pp. 223–232. ACM
6. Cai G, Xia B (2015) Convolutional neural networks for multimedia sentiment analysis. In Natural Language Processing and Chinese Computing (pp. 159–167). Springer, Cham
7. Chen T, Salah Eldeen HM, He X, Kan MY, Lu D (2017) VELDA: Relating an Image Tweet's Text and Images. In AAAI 2015 Jan 25 (pp. 30–36)
8. Datta R, Joshi D, Li J, Wang JZ (2006) Studying aesthetics in photographic images using a computational approach. In: European Conference on Computer Vision (pp. 288–301). Springer, Berlin
9. Dave K, Lawrence S, Pennock DM (2003) Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. Proceedings of the 12th international conference on World Wide Web. ACM. 519–528
10. Esuli A, Sebastiani F (2007) SentiWordNet: a high-coverage lexical resource for opinion mining. Evaluation. 17:1–26
11. Gajarla V, Gupta A (2015) Emotion detection and sentiment analysis of images. Georgia Institute of Technology, Atlanta
12. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580–587
13. Go A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford 1(12)
14. Golder SA, Macy MW (2011) Diurnal and seasonal mood vary with work, sleep, and day length across diverse cultures. Science 333(6051):1878–1881
15. Hao T, Rusanov A, Boland MR, Weng C (2014) Clustering clinical trials with similar eligibility criteria features. J Biomed Inform 52:112–120
16. Hare JS, Samangooei S, Dupplaw DP, Lewis PH (2013) Twitter's visual pulse. In: Proceedings of the 3rd ACM conference on International conference on multimedia retrieval (pp. 297–298). ACM
17. Hodosh M, Young P, Hockenmaier J (2013) Framing image description as a ranking task: Data, models and evaluation metrics. J Artif Intell Res 47:853–899
18. Jia J, Wu S, Wang X, Hu P, Cai L, Tang J (2012) Can we understand van gogh's mood?: learning to infer affects from images in social networks. In: Proceedings of the 20th ACM international conference on Multimedia. ACM, pp. 857–860
19. Katsurai M, Satoh SI (2016) Image sentiment analysis using latent correlations among visual, textual, and sentiment views. In: Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on (pp. 2837–2841). IEEE

20. Kouloumpis E, Wilson T, Moore JD (2011) Twitter sentiment analysis: The good the bad and the omg! ICWSM 11(538–541):164
21. Kumar A, Dogra P, Dabas V (2015) Emotion analysis of Twitter using opinion mining. In: Contemporary Computing (IC3), 2015 Eighth International Conference on, IEEE, pp. 285–290
22. Kumar A, Jaiswal A (2017) Empirical study of twitter and tumblr for sentiment analysis using soft computing techniques. Proceedings of the World Congress on Engineering and Computer Science 1:1–5
23. Kumar A, Jaiswal A (2017) Image sentiment analysis using convolutional neural network. In: International Conference on Intelligent Systems Design and Applications (pp. 464–473). Springer, Cham
24. Kumar A, Jaiswal A (2019) Systematic literature review of sentiment analysis on Twitter using soft computing techniques. Concurrency and Computation: Practice and Experience: e5107
25. Kumar A, Jaiswal A, Garg S, Verma S, Kumar S (2019) Sentiment Analysis Using Cuckoo Search for Optimized Feature Selection on Kaggle Tweets. International Journal of Information Retrieval Research (IJIRR) 9(1):1–15
26. Kumar, Akshi, Renu Khorwal, and Shweta Chaudhary (2016) A survey on sentiment analysis using swarm intelligence." Indian Journal of Science and Technology 9, no. 39
27. Kumar A, Sebastian TM (2012) Sentiment analysis: A perspective on its past, present and future. International Journal of Intelligent Systems and Applications 4(10):1–4
28. Kumar A, Sebastian TM (2012) Sentiment analysis on Twitter. IJCSI International Journal of Computer Science Issues 9(3):372–378
29. Kumar A, Sebastian TM (2012) Machine learning assisted sentiment analysis. In: Proceedings of International Conference on Computer Science & Engineering (ICCSE’2012), pp. 123–130
30. Kumar A, Sharma A (2017) Systematic literature review on opinion mining of big data for government intelligence. Webology 14(2)
31. Kumar A, Sharma A (2018) Socio-sentic framework for sustainable agricultural governance. Sustainable Computing: Informatics and Systems
32. Kumar A, Sharma A (2019) Opinion mining of Saubhagya Yojna for Digital India. In: International Conference on Innovative Computing and Communications, pp. 375–386. Springer, Singapore
33. Li B, Feng S, Xiong W, Hu W (2012) Scaring or pleasing: exploit emotional impact of an image. In: Proceedings of the 20th ACM international conference on Multimedia. ACM, pp. 1365–1366
34. Li B, Xiong W, Hu W, Ding X (2012) Context-aware affective images classification based on bilayer sparse representation. In: Proceedings of the 20th ACM international conference on Multimedia. ACM, pp. 721–724
35. Lu X, Suryanarayan P, Adams Jr RB, Li J, Newman MG, Wang JZ (2012) On shape and the computability of emotions. In Proceedings of the 20th ACM international conference on Multimedia (pp. 229–238). ACM
36. Machajdik J, Hanbury A (2010) Affective image classification using features inspired by psychology and art theory. In: Proceedings of the 18th ACM international conference on Multimedia. ACM, pp. 83-92
37. Mandhyani J, Khatri L, Ludhrani V, Nagdev R, Sahu S (2017) Image Sentiment Analysis. International Journal of Engineering Science 4566
38. Marchesotti L, Perronnin F, Larlus D, Csurka G (2011) Assessing the aesthetic quality of photographs using generic image descriptors. In: Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, pp. 1784–1791
39. Neethu MS, Rajasree R (2013) Sentiment analysis in twitter using machine learning techniques. In: Computing, Communications and Networking Technologies (ICCCNT), 2013 Fourth International Conference on, (pp. 1–5). IEEE
40. Pak A, Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. In: LREC, vol. 10, pp. 1320–1326
41. Poria S, Cambria E, Howard N, Huang GB, Hussain A (2016) Fusing audio, visual and textual clues for sentiment analysis from multimodal content. Neurocomputing 174:50–59
42. Porter MF (1980) An algorithm for suffix stripping. Program. 14(3):130–137
43. Saif H, Fernandez M, He Y, Alani H (2013) Evaluation datasets for Twitter sentiment analysis: a survey and a new dataset, the STS-Gold
44. Saif H, Fernandez M, He Y, Alani H (2014) Senticircles for contextual and conceptual semantic sentiment analysis of twitter. In: European Semantic Web Conference (pp. 83–98). Springer, Cham
45. Saif H, He Y, Fernandez M, Alani H (2016) Contextual semantics for sentiment analysis of Twitter. Inf Process Manag 52(1):5–19
46. Siersdorfer S, Minack E, Deng F, Hare J (2010) Analyzing and predicting sentiment of images on the social web. In: Proceedings of the 18th ACM international conference on Multimedia. ACM, pp. 715–718
47. Soleymani M, Garcia D, Jou B, Schuller B, Chang SF, Pantic M (2017) A survey of multimodal sentiment analysis. Image Vis Comput 65:3–14
48. Vonikakis V, Winkler S (2012) Emotion-based sequence of family photos. In: Proceedings of the 20th ACM international conference on Multimedia. ACM, pp. 1371–1372

49. Wang Y, Wang S, Tang J, Liu H, Li B (2015) Unsupervised sentiment analysis for social media images. In: IJCAI. pp. 2378–2379
50. Yang Y, Cui P, Zhu W, Zhao HV, Shi Y, Yang S (2014) Emotionally representative image discovery for social events. In: Proceedings of International Conference on Multimedia Retrieval. ACM, p. 177
51. Yanulevskaya V, Uijlings J, Bruni E, Sartori A, Zamboni E, Bacci F, Melcher D, Sebe N (2012) In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. In: Proceedings of the 20th ACM international conference on Multimedia. ACM, pp. 349–358
52. Yanulevskaya V, van Gemert JC, Roth K, Herbold AK, Sebe N, Geusebroek JM (2008) Emotional valence categorization using holistic image features. In: ICIP. pp. 101–104
53. You Q, Luo J (2013) Towards social imagematics: sentiment analysis in social multimedia. In: Proceedings of the Thirteenth International Workshop on Multimedia Data Mining, ACM, p. 3
54. Zhao S, Gao Y, Jiang X, Yao H, Chua TS, Sun X (2014) Exploring principles-of-art features for image emotion recognition. In: Proceedings of the 22nd ACM international conference on Multimedia (pp. 47–56). ACM
55. Zhao S, Yao H, Wang F, Jiang X, Zhang W (2014) Emotion based image musicalization. In: Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on (pp. 1–6). IEEE
56. Zhao S, Yao H, Yang Y, Zhang Y (2014) Affective image retrieval via multi-graph learning. In: Proceedings of the 22nd ACM international conference on Multimedia (pp. 1025–1028). ACM
57. Zhou X, Yao C, Wen H, Wang Y, Zhou S, He W, Liang J (2017) EAST: an efficient and accurate scene text detector. In Proc. CVPR, pp. 2642–2651

Akshi Kumar is an Assistant Professor in the Department of Computer Science & Engineering at Delhi Technological University (formerly Delhi College of Engineering). She has received her doctorate degree in Computer Engineering from the University of Delhi. She has received her M.Tech (Master of Technology) and BE (Bachelor of Engineering) degrees in Computer Engineering. She has many publications to her credit in various International Journals with high impact factor and International Conferences with best paper awards. Her research interests include Intelligent Systems, Text Mining, Sentiment Analysis, Social media analytics and soft computing. She has authored books & co-authored book-chapters within her domain of interest. She is a member of IEEE, IEEE (WIE) and ACM and a life member of IACSIT, IAENG, ISTE and CSI.

Geetanjali Garg is an Assistant Professor and research scholar in the Department of Computer Science & Engineering at Delhi Technological University (formerly Delhi College of Engineering). She has received her M.Tech (Master of Technology) degree in Computer Technology & Applications from the University of Delhi. She has received her B.Tech (Bachelor of Technology) degree in Computer Science & Engineering. Her research interests include Social media mining, Sentiment analysis,Semantic Web and Intelligent systems.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.