Multi-Modal Sarcasm Detection in Social Networks: A Comparative Review

Conference Paper · April 2022

DOI: 10.1109/ICCMC53470.2022.9753981

2 authors:

Poulami Dutta
Techno India
2 PUBLICATIONS **8** CITATIONS
SEE PROFILE

Chandan Kumar Bhattacharyya
Sister Nivedita University
24 PUBLICATIONS **87** CITATIONS
SEE PROFILE

Proceedings of the Sixth International Conference on Computing Methodologies and Communication (ICCMC 2022)
DVD Part Number: CFP22K25-DVD; ISBN: 978-1-6654-1027-4

Multi-Modal Sarcasm Detection in Social Networks: A Comparative Review

Poulami Dutta
Dept. of CSE
Techno Main Salt Lake
Kolkata, India
poulami.dutta@gmail.com

Chandan Kumar Bhattacharyya
Dept. of CSE
Sister Nivedita University
Kolkata, India
ckb.tig@gmail.com

***Abstract*—Sentiment Analysis (SA) has become an extremely sought after area of research especially post COVID-19 when people used to spent a lot of time on the social media to interact with each other. This interaction was done through posts having both textual and visual cues and also by participating in online discussions forums. Some of the inherent challenges encountered in the process of SA include discernment of sarcasm, irony, humor, negation, multi-polarity or Aspect-Level Sentiment Analysis (ASA) etc. Researchers are now gradually shifting their focus to the identification and detection of sarcasm and how it can empower SA. Sarcasm expresses a person's downside feelings by using positive words in an implicit way. It also has an overall impact on increasing the efficiency of the SA models. Eliciting sarcastic statements is tough for humans as well as for machines without the knowledge of the context or background in which it is expressed, body language and/or facial expression of the speaker and his voice modulation. This review paper studies some of the approaches used for sarcasm detection and also guides researchers in exploring the different modalities of data for developing applications like a virtual chat-bot or assistant, depression analysis, stress management system at workplace etc.**

Keywords*—Sentiment Analysis, Sarcasm Detection, Negation Detection, Multi-Modal, Social Networks.

I. Introduction

The computational study of judgements, viewpoints, feelings etc. on social media platforms like Twitter, Facebook, Instagram etc. is known as Sentiment Analysis (SA). It classifies the output as "positive", "negative" or "neutral/indifferent". Due to the growth of e-services such as e-commerce, e-tourism, e-business etc. and difficulties in the manual handling of the enormous volume of different types of data generated online, SA has become one of the most important research topics. [1].

The most challenging issue in the process of SA is sarcasm and negation detection whose existence is mostly ignored by researchers due to its inherent complexities. Sarcasm, a subset of irony, is characterized as an ironic or satirical wit that is intended to insult, mock, ridicule or amuse. It is the utterance of a positive statement with a negative intent that creates confusion in the overall meaning of the sentence making it hard to detect.

Verma et al. [2] has considered sarcasm to be a major roadblock while rectifying the polarity of a sentence. For e.g. the sentence: "My phone has a startling battery life of 1 hour." would be classified as positive by some and negative by few others. Training the model for sarcasm detection gets increasingly difficult with such ambiguous sentences [1]. Thus, to classify sarcasm, knowledge about the context or background is also necessary as observed by the authors in [2]. For example, "It feels great to waste some precious hours in traffic jam on my way to work." is a remark where a positive word "great' is used to express the person's downside feelings while "caught up in the traffic on the way to work" which in this case is the context.

Detecting emotion from the comments and reviews of social media websites is a contemporary research topic [3]. One of the applications of emotion detection is sarcasm detection. An emoji or an emoticon that is used to convey a person's mood or feelings is a pictorial representation (a facial expression) that uses punctuation marks, numbers, letters or imports from the GIF library. The same when used in images in contrasting situations make the sentence sarcastic. Sarcasm detection (SD) is a dual classification problem with binary target labels i.e. sarcastic and non-sarcastic.

We summarize our contributions as follows:

- We emphasize the need for multi-modality in sarcasm detection to achieve better accuracy and improved performance.
- The context in which the comment was made or the facial expression used can serve as an important cue for detecting sarcasm and this too has been one of our major findings.
- We establish the need for cross-lingual and multi-lingual data sets for extracting the overall sentiments in facebook posts, tweets etc.
- This paper also explores and compares the different models available in literature and proposes applications that can be developed from multi-modal data.

The rest of the paper is organized as follows: Classification of Sarcasm and Feature Set Analysis is given in Section II and Section III respectively. Section IV lists the Steps of a Sarcasm Detection Model and Section V details the techniques for detecting sarcasm. Section VI and VII deal with the issues and challenges respectively. Literature Review is presented in Section VIII along with the comparative study in Table II.

Table III include the study of standard data sets. Finally, we conclude in Section IX and present the future scope in Section X.

II. Classification of Sarcasm

Researchers in [4] have classified sarcasm into five types, based on its feature and structure, given in Table I.

A. Sarcasm as an Incongruity of Sentiments

The incongruity between the textual and the situational sentiment is represented in this class of sarcasm as follows:

1) Contrast Between Negative Situation and Positive Sentiment

Conflicting opinions are expressed in a single sentence where disagreeable sentiments follow the pleasant ones, for example, *"Absolutely amazing when my bus is running late"* express pleasant sentiments in the context of an unpleasant scenario.

2) Contrast Between Positive Situation and Negative Sentiment

The sarcastic text here includes negative sentiments followed by a positive situation. The phrase *"How I hate Team Australia, they have once again won the T20 world cup!!"* conveys negative sentiment in a positive situation.

3) Contrasting Implications

This is a class of sarcasm where contrasting connotations appear in the same narrative. For example, *"I love receiving spams!"* expresses two contrasting connotations.

4) Variance of the Present with the Past

Disparity between historic and current comments of an user with respect to the same context is cited below. For example, *User 1: "Why travel between 'Thanksgiving' and 'Christmas'! Bad cold weather here in December!" User 2: "We love the kind words, thanks a lot!" User 1: "Wow!! I guess i should have said that i actually meant sarcasm"*. This example clearly shows that User 1 is actually expressing sarcasm which User 2 fails to understand.

5) Truthfulness Negation

This is the contradiction of textual contents with universal truths or facts. For example, *"This is one bickering couple. Opposite poles do not attract at all"*. We know as a matter of fact that "opposite poles attract" but this example shows otherwise.

6) Extrication of Temporal Facts

Here the contradiction is about the facts related to the specified event. For example, *"It was FUN spending my birthday with you and your boyfriend"* conveys sarcasm since the person does not seem quite happy doing so.

B. Sarcasm as a Means of Expressing Emotion

Sarcasm is also used to convey the different mental states the user is in depending on his/her mood. These are as follows:

1) Banter/Quip/Humour

Capital letters, question marks, punctuations, exaggerations etc. are used in sarcasm to express extreme happiness. For example, *"What a FANTASTIC weather!! I simply LOVE the rains!!"* expresses a person's happiness because of the rainy weather though his irritation is apparent.

2) Whine/Whimper

The angry mood of a person is captured in this type of sarcasm. For example, *"I am so glad my mom woke me up early morning for vacuuming my room!!#sarcasm"* clearly indicates a person's displeasure on being woken up early morning.

3) Equivocation

In this class of sarcasm a person circumvents giving a clear answer by using uncommon words or expressions. For e.g. *I:"Alice, you need to work hard!!"* and *Alice: "Ohh Yaah!! I am crystal clear about what needs to be done"* indicates Alice's way of keeping away from getting the work done.

4) Uncontrolled/Unrestrained

Hyperbolic text is used to depict the aggressive nature of a person. For example, *ME: "Hey Bob, I want you to work overtime today."* and *Bob: "Oh, don't worry!! I'll be there!! Would you like me to shine your shoes today??"*

C. Sarcasm as a Form of Written Expression

Sarcasm is expressed in writing by using a certain style through text as described below:

1) Prosodic Disparity

This is used to emphasize certain parts of the text. To convey exactly the opposite of what is meant, certain letters punctuation marks are repeated. For example, *"sooooo", "wooowww", "!!!!" etc.* Sometimes users capitalize certain words to indicate the magnitude, for example, *"Just WOOWWW!!" or "What an AMAZING weather!!"* or *"GR8888!!*.

2) Structural Disparity

In some sarcastic sentences, the first few words of the sentence express the opinion of the user while the contextual description is given in the later part. For example, *"I love it when my friends ignore me"*. convey the user's view in an annoying context.

3) Lexical Analysis

The hashtag, which is a strong indicator of sarcasm forms the basis here. (e.g. #sarcasm, #irony etc.) For example, *"I just love to party alone. #sarcasm."*

D. Sarcasm as a Function of Expertise

1) Language Competency

It is expected that the user would have a good grasp over the language. For example, *"He is a good person. #sarcasm."* explicitly indicates sarcasm. But the sentence, *"He is as good as evil."* expresses sarcasm even without being labelled with a '#'.

TABLE I: Classification of Sarcasm

Sl. No.	Classification of Sarcasm	Subclasses of Sarcasm
T1	Sarcasm as an Incongruity of Sentiments	i) Contrast Between Negative Situation and Positive Sentiment ii) Contrast Between Positive Situation and Negative Sentiment iii) Contrasting Implications iv) Variance of the Present with the Past v) Truthfulness Negation vi) Extrication of Temporal Facts
T2	Sarcasm as a Means of Expressing Emotion	i) Banter/Quip/Humour ii) Whine/Whimper iii) Equivocation iv) Uncontrolled/Unrestrained
T3	Sarcasm as a Form of Written Expression	i) Prosodic Disparity ii) Structural Disparity iii) Lexical Analysis
T4	Sarcasm as a Function of Expertise	i) Language Competency ii) Environment Prowess
T5	Behaviour-based Sarcasm	I) Prediction of Likes and Dislikes

2) Environment Prowess

Users are known to better express sarcasm when surrounded by a familiar environment. For example, *"I love it when i am woken up at 4 AM by my neighbor's kid who starts crying"* resonates the sentiment.

E. Behavior-based Sarcasm

User's behavior or point of view is considered here.

1) Prediction of Likes and Dislikes

Manifestation of sarcasm towards various products, services, events, movies etc. is done by using the "like" or "dislike" options. It may not be explicitly indicated but subtly expressed.

III. Feature Set Analysis

The occurrence of various types of sarcasm (in textual forms) is the basis of feature set analysis which serves the purpose of sarcasm detection in social networks.

A. Lexical Feature

This feature set includes textual properties like unigram, bigram, n-grams, #hashtags etc. which play an important role in sarcasm detection.

B. Pragmatic Feature

Considered to be one of the most powerful features for detecting sarcasm in textual data, this feature experiments with symbolic texts like emoticons, emojis, smileys, replies etc. to detect sarcasm.

C. Hyperbole Feature

The use of exaggeration as a parts of speech is known as a "hyperbole" viz. intensifiers, exclamations, quotations, punctuations etc.

D. Pattern-based Feature

High frequency words and their frequency of appearance in the textual content is used in this approach.

E. Syntactic Feature

This approach uses morphosyntactic features which contains a combination of morphology and syntax.

F. Contextual Feature

Information beyond textual input that is used for detecting sarcasm is referred to as the context or information or background knowledge. Explanations or supplementary data or information is used for this purpose.

G. Metaphoric Feature

Information about the author's feelings or emotions is conveyed through the extreme usage of positive or negative nouns, adjectives, proverbs, pseudonyms etc.

IV. Steps of Sarcasm Detection Model

The authors of the research article [2] have illustrated the following steps for detecting sarcasm also shown in Fig. 1:

A. DATA ASSEMBLAGE

A fundamental step is *Data Acquisition*. The two modes are, API (Application Programming Interface) and Accessible Data sets like MUStARD (Multi-modal Sarcasm Detection data set), SARC (Self annotated reddit Corpus), SemEval (Semantic Evaluation Data set) etc.

B. TOKENIZATION

The next step is Data Pre-Processing. For Natural Language Processing (NLP) tasks, this step includes stop words removal, lemmatization of data tokens, stemming, and tokenization.

C. FEATURES WITHDRAWAL

Here, various features are extracted from a data set to prepare the model. Different methods for feature extraction include the BoW (Bag of Words), Doc2Vec, Term Frequency - Inverse Document Frequency (TF-IDF), word2vec and GloVe.

D. FEATURE SELECTION

The most appropriate set is selected to enhance the performance of the classification model. Some of the popular techniques used are Chi-square and Mutual Information (MI).

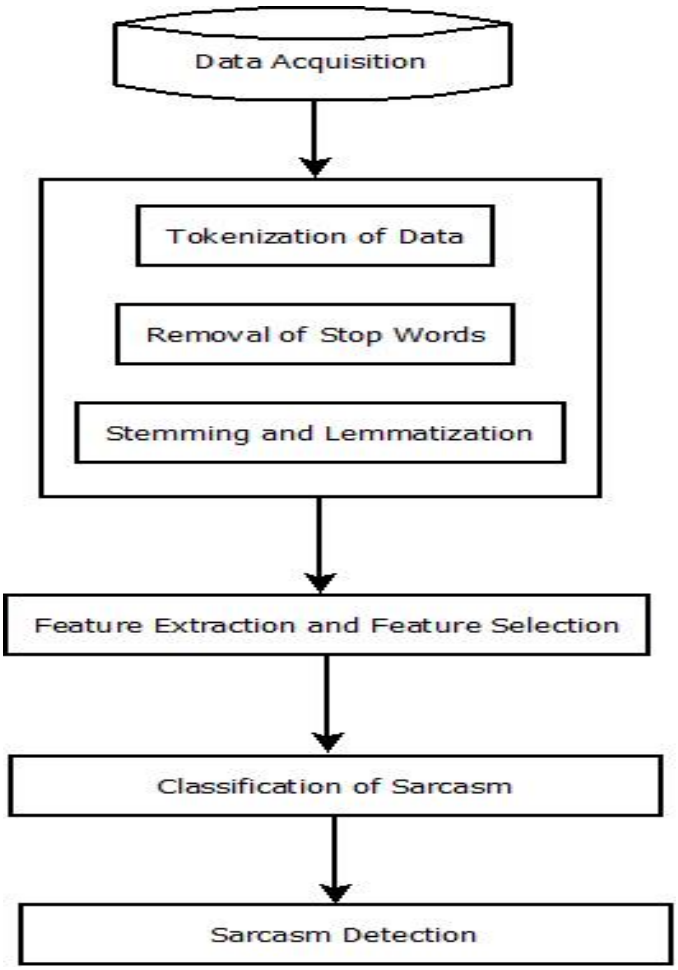

Fig. 1: Flow Chart for Sarcasm Detection Model

E. CLASSIFICATION TECHNIQUES

Sarcasm Detection (SD) is a binary classification problem done using Machine Learning (ML), Deep Learning (DL) and hybrid techniques.

F. EVALUATION METRICS

Precision 'P', F-score 'F', Recall 'R' and Accuracy 'A' are the evaluation metrics. Precision, given in Eq.1 is defined as the number of True Positives 'a' divided by the total number of True Positives plus the number of False Positives 'c'. 'R', expressed mathematically in Eq.2 is defined as the number of true positives that were recalled or found i.e. how many of the correct hits were found. Here, False Negatives are represented by 'd'. 'F' is a measure of a model's accuracy wrt a data set given in Eq.3. 'A' is defined mathematically in Eq.4 where 'b' denotes the True Negatives.

$$P = \frac{a}{(a+c)} \tag{1}$$

$$R = \frac{a}{(a+d)} \tag{2}$$

$$F = \frac{(2 \times P \times R)}{(P+R)} \tag{3}$$

$$A = \frac{a+b}{(a+b+c+d)} \tag{4}$$

V. TECHNIQUES FOR DETECTING SARCASM

There are five (5) approaches for detecting sarcasm as described in this section.

A. Rule-based Approach

This approach uses a 'hashtag'. It states that if a hashtag '#' is found in a tweet, it will be marked as sarcastic and everything else in that tweet will not matter. However, if the '#' is removed during pre-processing, then ambiguity arises in classifying the text as sarcastic or non-sarcastic.

B. Statistical Approach

Here, pattern-based features are identified by the classifiers. These features take up values based on exact match, partial match and no match at all. Pragmatic features like emoticons and emojis are also used in this approach.

C. Lexical Approach

This approach uses explicit terms/words like joyful, upset, anxious etc. to detect sarcasm. Four (4) features viz. overtness, acceptability, exaggeration and comparison are used here for training purposes using machine learning models. It also includes the dictionary-based and corpus-based methods.

D. Machine-Learning-based Approach

These approaches assess sarcasm based on features extracted from various data sets. Some research articles have demonstrated the influence of parts of speech e.g. adjectives and adverbs on classification of product reviews with an aim to detect sarcasm.

E. Deep-Learning-based Approach

The most popular of all approaches uses similarity between words as features to detect sarcasm. The features are then augmented to the most congruent and incongruent word-pairs. A combination of a few algorithms namely Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Deep Neural Networks (DNN) are used for achieving best results in this approach for detecting sarcasm.

In addition, there exists a hybrid approach that combines one or two of the above approaches into one system, accounting for more accuracy. It also combines an emoticon-based and a comment-based approach to detect sarcasm.

VI. ISSUES IN SARCASM DETECTION

There are issues that need to be addressed as cited by Chaudhari et al. [4] (2017) and have been described below.

A. Issues with Data Annotation

Sarcasm is explicitly revealed in sentences labelled with hashtags '#' but the same is arduous for cryptic and dubious data sets. For example, *"How I love bland food!!#"* clearly expresses sarcasm because of the usage of the '#' symbol. On removing the '#' symbol, the text *"How I love bland food!!"* no longer has any sarcastic interpretation. Here comes the necessity of contextual (background) feature for effectively detecting sarcasm.

B. Issues with Sentiment as a feature

Several approaches use sentiment as a feature for the sarcasm classifier used. Sarcastic sentences sometimes deceive the classifiers and thereby compromise on the accuracy of classification. For example, *"It's not like I wanted to have breakfast, but anyways. #sarcasm"* observes that the use of sarcasm does not flip the polarity of the sentence.

C. Issues with Classification Techniques

Some researchers use smaller data sets whereas some use larger data sets. For accuracy of results with respect to classification, it is necessary that the data set is balanced so that precision can be achieved.

VII. Challenges in Sarcasm Detection

The most difficult area of SA is Sarcasm Detection which faces certain challenges that are highlighted below:

A. Difficulty in Sarcasm Detection from Text

The detection of sarcasm from only text is difficult as compared to image and speech and requires in depth study.Facial expression, body language, voice modulation etc. can be useful while identifying sarcasm.

B. Positive Words to convey Negative Sentiments

Sarcastic sentences use positive words to express a negative opinion or sentiment. A BoW approach might not be sufficient to analyse the sentiment from such sentences and may require additional features.

C. Short Text Usage

Sarcasm detection from short and noisy text is challenging due to the difficulties in feature extraction and analysis. Thus hashtags are used. Clarity of sarcastic sentences are difficult without these hashtags.

D. World Knowledge Integration

An example like *"Do not judge the book by its cover"* or *"All that glitters is not gold"*is actually an universally accepted sarcastic sentence provided the user is aware of the context.

E. Hyperbole

Hyperbole is the use of dramatization in a sentence that helps in sarcasm detection. It is an NLP (Natural Language Processing) problem that is still a sought after area of research.

VIII. Literature Review

This section gives an overview of some of the previous approaches used to detect sarcasm from multi-modal data like text, image, video etc. The review is presented in Table II which illustrates the performance of some of the most popular models (primarily based on the number of citations) in this domain. We have studied various methods for sarcasm detection including Machine Learning approaches like SVM, lexicon-based, decision trees, Naive Bayes classifier, logical regression; Deep Learning approaches like CNN, RNN, TF-IDF, Bi-LSTM etc. and Hybrid approaches that combine the best of both worlds using both rule-based and lexicon-based concepts. Some of the standard data sets used in the reviewed literature has been presented in Table III.

The analysis of multi-modal inputs and the sentiment/s attached using lexicon-based, machine-learning-based and hybrid approaches has been explored in [1]. A combination of different modalities yield better results in SA, as claimed by Mehta et al. [1]. [2], [4], [5], [6] and [7] discuss recent trends in the field of sarcasm detection along with the different techniques and challenges. A comparative study of the pattern-based approach; SVM (Support Vector Machine) and Voting Classifier approach which is a statistical approach; C-Net (Contextual Network) approach etc. has also been reviewed. The hyperbole approach exploiting the NLP technique or exaggerations used in text along with the hashtags #sarcasm, #not, #mockery, etc. help detect sarcastic tweets or posts without any dilemma. Psycho-linguistic i.e. correlation between language and psychology is an area where existing works on sarcasm detection are being done as pointed out by the authors in [7].

Researchers in [3] and [8] have proposed methods that group posts based on emotions and sentiments and detects the sarcastic ones from them. These emotions help in better understanding of the posts when compared to approaches which consider only the polarity. The use of the emotion and sentiment identifier using the "SentiWordNet" data set help find the emotion attached. It is the sentences with negativity in its second half that captures sarcasm, as studied here. [9] and [10] explore sarcasm detection as a field that is used to improve SA by using multi-modalities. The accuracy of the classifier used is potentially increased when visual modalities are blended with text. The authors have also observed the large dependence of sarcasm on the context in which it is spoken.

Castro et al. [11] have developed models giving due weightage to all three modalities of text, speech and image. A new dataset, "MUStARD" has been curated using audio-visual utterances compiled from popular TV shows. Having observed cross-modal incongruities the authors have stressed the need for multi-modal approaches. [12] has tried to look into the human process of detecting sarcasm on social media where users communicating with each other could be strangers and unaware of the background information. In [13], the authors use a hierarchical fusion model applying a bi-LSTM that extracts the image features, followed by the attribute features and then use the latter to learn the text features.

Aspect-level sentiment classification by means of transfer learning has been proposed by the "TransCapsule" model, [14]. It transfers knowledge from the document-level into the aspect-level with respect to each sentence using routing algorithms. By using a CNN, facial expression can be recognized, which achieves an accuracy of 96% [15]. The eXnet (Expression Net) library along with data augmentation techniques is used for attaining improved performance on real-time devices. [16] predicts the performance of students in order to bring about changes in learning outcomes. The aim is to overhaul the teaching-learning and evaluation/assessment paradigm in educational research. The most inquisitive and tough part in Natural Language Processing (NLP) is automatic text summarization which involves producing a precise and comprehensive summary of text from product titles, e-books etc. a method known as topic modeling as illustrated in [17]. [18] proposes an incongruity-aware attention network (IAWN) that detects sarcasm by using a word-level inconsistency between modalities and a scoring mechanism.

[19] and [20] claim to have curated a Hindi-English code-mixed data set "MaSaC" which is a first of its kind data set for multi-modal sarcasm detection and humor classification in dialogues. MSH-COMICS (**M**ulti-modal **S**arcasm Detection and **H**umor Classification in **CO**de-**MI**xed **C**onversation**S**) is an attention-rich architecture proposed for classifying speech. It studies the sentiment irregularities within and across modalities using graph convolutional network (GCN).

IX. Conclusion

After going through the works of various authors, it has been observed that sarcasm is best detected with much more precision when combined with different modalities of data. Sarcasm, a subset of irony, is also detected using emojis, emoticons, numeric values, hashtags etc. A single approach might not be enough to identify a comment as sarcastic or non-sarcastic. It is hard to detect sarcasm without the listener having some kind of background knowledge or understanding of the facial expression or the body language of the speaker. This is also precisely the reason why more than text is necessary to properly detect sarcasm. Also, the hybrid approach that integrates deep learning with machine learning produces an improved performance for sarcasm detection using multi-modal data in social networks.

X. Future Work

For future work, modelling incongruities between facial expression and context, which are valuable cues for sarcasm detection can be done. An application can be developed by combining cognitive features with multi-modalities that will take multimedia data as input and produce the sarcasm attached to it as the output. A multi-modal data set in vernacular and non-English foreign languages can be curated to detect satire and classify humor for use in applications involving prediction of political outcomes, grievance redressal or stress management at workplace etc. This shall also find its relevance in NLP sub-domains like topic modelling.

References

[1] M. Mehta, K. Gupta, S. Tiwari, and Anamika, "A review on sentiment analysis of text, image and audio data," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 1660–1667.

[2] P. Verma, N. Shukla, and A. Shukla, "Techniques of sarcasm detection: A review," in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2021, pp. 968–972.

[3] S. Rendalkar and C. Chandankhede, "Sarcasm detection of online comments using emotion detection," in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2018, pp. 1244–1249.

[4] P. Chaudhari and C. Chandankhede, "Literature survey of sarcasm detection," in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2017, pp. 2041–2046.

[5] J. aboobaker and E. Ilavarasan, "A survey on sarcasm detection and challenges," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020, pp. 1234–1240.

[6] S. Raghav and E. Kumar, "Review of automatic sarcasm detection," in *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*. IEEE, 2017, pp. 1–6.

[7] S. G. Wicana, T. Y. İbisoglu, and U. Yavanoglu, "A review on sarcasm detection from machine-learning perspective," in *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*. IEEE, 2017, pp. 469–476.

[8] M. Adarsh and P. Ravikumar, "Sarcasm detection in text data to bring out genuine sentiments for sentimental analysis," in *2019 1st International Conference on Advances in Information Technology (ICAIT)*. IEEE, 2019, pp. 94–98.

[9] S. Sangwan, M. S. Akhtar, P. Behera, and A. Ekbal, "I didn't mean what i wrote! exploring multimodality for sarcasm detection," in *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[10] M. S. Razali, A. A. Halin, N. M. Norowi, and S. C. Doraisamy, "The importance of multimodality in sarcasm detection for sentiment analysis," in *2017 IEEE 15th Student Conference on Research and Development (SCOReD)*, 2017, pp. 56–60.

[11] S. Castro, D. Hazarika, V. Pérez-Rosas, R. Zimmermann, R. Mihalcea, and S. Poria, "Towards multimodal sarcasm detection (an _obviously_ perfect paper)," *arXiv preprint arXiv:1906.01815*, 2019.

[12] D. Das, "A multimodal approach to sarcasm detection on social media," 2019.

[13] Y. Cai, H. Cai, and X. Wan, "Multi-modal sarcasm detection in twitter with hierarchical fusion model," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 2506–2515.

[14] A. Sungheetha and R. Sharma, "Transcapsule model for sentiment classification," *Journal of Artificial Intelligence*, vol. 2, no. 03, pp. 163–169, 2020.

[15] K. Kottursamy, "A review on finding efficient approach to detect customer emotion analysis using deep learning analysis," *Journal of Trends in Computer Science and Smart Technology*, vol. 3, no. 2, pp. 95–113, 2021.

[16] A. Dhankhar, K. Solanki, S. Dalal *et al.*, "Predicting students performance using educational data mining and learning analytics: A systematic literature review," *Innovative Data Communication Technologies and Application*, pp. 127–140, 2021.

[17] A. D. Dhawale, S. B. Kulkarni, and V. M. Kumbhakarna, "A survey of distinctive prominence of automatic text summarization techniques using natural language processing," in *International Conference on Mobile Computing and Sustainable Informatics*. Springer, 2020, pp. 543–549.

[18] Y. Wu, Y. Zhao, X. Lu, B. Qin, Y. Wu, J. Sheng, and J. Li, "Modeling incongruity between modalities for multimodal sarcasm detection," *IEEE MultiMedia*, vol. 28, no. 2, pp. 86–95, 2021.

[19] M. Bedi, S. Kumar, M. S. Akhtar, and T. Chakraborty, "Multi-modal sarcasm detection and humor classification in code-mixed conversations," *IEEE Transactions on Affective Computing*, 2021.

[20] B. Liang, C. Lou, X. Li, L. Gui, M. Yang, and R. Xu, "Multi-modal sarcasm detection with interactive in-modal and cross-modal graphs," in *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 4707–4715.

TABLE II: Review of Previous Work - Comparison Table

Sl. No.	Title of Paper	Model Used	Type of Input Data	Data Set	Results/ Findings	Limitations
1.	A Review on Sentiment Analysis of Text, Image and Audio Data, 2021 [1]	Lexicon-basesd (VADER), Machine-learning based and Hybrid Model	Text, Image and Audio Data	VGG-Imagenet, VGG-Places205, ResNet-50, SentiBank, Flickr	It was found that sentiment extraction from multimedia data using deep learning models achieve better results.	Modelling cross-modal relationships for developing multimedia applications can be explored along with the incongruities for achieving state-of-art performances.
2.	Techniques of Sarcasm Detection: A Review, 2021 [2]	Hybrid Model	Textual Data	Twitter, Reddit, SemEval, politics data set	sAtt Bi-LSTM ConvNet shows outstanding results with an accuracy of 97.8% and 93.71% using SemEval and Random Tweets dataset respectively.	Multi-modal sarcasm detection using sAtt Bi-LSTM ConvNet can be done. Datasets having multi-media inputs may be curated to achieve better results.
3.	A Review on finding efficient approach to detect customer emotion analysis using deep learning analysis, 2021 [14]	CNN	Image or image frome from videos	Dataset for expression recognition, ICML - 2013 division	It has developed a CNN based model, eXnet for feature extraction that outperforms baseline models with an accuracy of 96.12%. This model recognizes facial expression.	Datset needs to be scaled for utilization of the proposed model on a GPU.
4.	Multi-modal Sarcasm Detection and Humor Classification in Code-mixed Conversations, 2021 [15]	LSTM	Video-clips and utterances/dialog	Hindi-English code-mixed dataset MaSaC	It has proposed a state-of-the-art architecture, MSH-COMICS, for multi-modal contextual classification of sentences, using the MaSaC dataset, that achieves an accuracy of 87.3% and 82.2% for multi-modal sarcasm detection and humor classification respectively.	It does not explore multi-lingual datasets for sarcasm detection and humor classification.
5.	Multi-Modal Sarcasm Detection with Interactive In-Modal and Cross-Modal Graphs, 2021 [16]	Graph Convolution Network (GCN)	Tweets with text and images	English tweets containing picture and hastag (#)	It establishes the inconsistencies among sentiments within a single modality and across multiple modalities by studying the role of text and image. Additionally, it also learns sarcasm without external or background knowledge.	It does not consider text and image integrated with speech for in-modal and cross-modal sarcasm detection.
6.	Predicting students performance using educational data mining and learning analytics: A systematic literature review, 2021 [19]	Deep learning networks	Student performance data	-	It predicts the performance of students in order to bring about changes in learning outcomes. The aim is to overhaul the teaching-learning and evaluation/assessment paradigm in educational research.	It does not exploit the role of multi-modality in prediction of students' performance.
7.	Modeling incongruity between modalities for multimodal sarcasm detection, 2021 [20]	Neural networks	Video clips with utterances and the context	MUStARD	It has proposed an incongruity-aware attention network (IAWN) that detects sarcasm by using a word-level inconsistency between modalities and a scoring mechanism.	Contextual incongruity that serves as a vital clue for sarcasm detection has not been discussed here.
8.	A survey of distinctive prominence of automatic text summarization techniques using natural language processing, 2020 [18]	Machine Learning and graph-based approaches	Textual inputs	-	The authors propose different techniques and methods applied to different languages.	The study can be adapted by researchers to get a brief idea about the evolution of different techniques for text summarization .
9.	TransCapsule Model for Sentiment Classification, 2020 [13]	Neural Network	Corpus of textual data	Online reviews	This methodology classifies aspect-level sentiment using the document-level data using transfer learning. It outperforms the existing baselines models.	Eliciting sarcasm from text is not studied here. Corpus can be extended to include image and speech.
10.	A Survey on Sarcasm detection and challenges, 2020 [5]	Support Vector Machine (SVM), Lexicon-based, Decision Trees	Textual Data	Twitter, WordNet	It studies the general architecture of sarcasm detection, the classifiers viz. SVM, Naive Bayes, random forest, decision tree, lexicon-based etc. and highlights the issues and challenges.	Thea authors have not dicussed about sarcasm detection from speech by observing the tone, body language, context etc. using unambiguous datasets other than the English language.
11.	I didn't mean what I wrote! Exploring Multimodality for Sarcasm Detection, 2020 [7]	Recurrent Neural Network (RNN)	Textual and Visual/ Image	Silver-Standard Data set and Gold-Standard Data set from Instagram	The proposed method achieves an accuracy of 66.17%, 70.0% and 71.5% respectively using text, text+image and text+image+transcript modality, whose inclusion plays an important role in enhancing the performance.	Gold standard multi-modal dataset like MUStARD can be used on the proposed methodology to achieve better performance.
12.	Towards Multimodal Sarcasm Detection (An Obviously Perfect Paper), 2019 [11]	BERT, pool15 layer of an ImageNet	Audio-visual utterances annotated with sarcasm labels	MUStARD	It was found that use of multimodal information for sarcasm detection reduces error rate by 12.9% in terms of F-score.	Fusion techniques that capture diaparities among modalities and also utilizes relationships like getsures can be identified.
13.	Multi-Modal Sarcasm Detection in Twitter with Hierarchical Fusion Model, 2019 [12]	Bi-LSTM	Image, text and image attributes	English tweets containing picture and hastag (#)	This model achieves an accuracy and F-score of 83.44% and 80.18% respectively when experimented with text, image and image atributes as feature sets.	Inclusion of audio and common sense into sarcasm detection model can be investigated.
14.	A Multimodal approach to sarcasm detection on social media, 2019 [17]	Deep Learning Networks	Facebook posts containing text, images and videos	Satire and fake news posts collected from Bengali satire sites Monikontho and Earki	This work extracts the overall sentiment from the posts that comprises of multi-modal data. Images and reactions are suggestive of the sentiments conveyed.	Language and alphabet specific cues can be exploited from non-English datasets for improving the performance of sarcasm detection algorithms.

TABLE III: Some Popular Benchmark Data Sets used for Multi-Modal Sarcasm Detection

Sl. No.	Dataset	Brief Description	Entity	Paper Referred	Size	Link
1.	MUStARD (MUltimodal SARcasm Dataset), 2019	This dataset compiled from popular TV shows include audiovisual utterances annotated with sarcasm labels.	Video Clips with utterances and the context	[11], 2019	6421 videos	https://github.com/soujanyaporia/MUStARD
2.	MaSaC, 2021	It is a multimodal Code-Mixed corpus (Hindi-English) compiled from a popular hindi TV series, for detection of sarcasm as well as humour.	Videos with utterances	[15], 2021	15k utterances spread across 400 scenes from a total of 50 episodes	https://github.com/LCS2-IIITD/MSH-COMICS
3.	Twitter Dataset, 2021	This dataset is divided by into the training set, validation set and test set and preprocessed to separate words, emoticons, and hashtags.	Tweet, image and image attributes	[2], 2021 and [5], 2020	24k samples of tweets	https://github.com/sundeshgupta/FiLMing-Multimodal-Sarcasm-Detection
4.	Reddit, 2020	This dataset is generated by scraping comments using the sarcasm tag. This tag is used to indicate that the comment has an intended pun and need not be taken seriously.	Sarcastic comments	[2], 2021	1.3 million	https://www.kaggle.com/danofer/sarcasm
5.	SemEval, 2017	This is a dataset of tweets manually annotated for stance towards a given target, target of opinion (opinion towards), and sentiment (polarity).	Tweets	[2], 2021	70k tweets	https://github.com/cbaziotis/datastories-semeval2017-task4
6.	Sentiwordnet, 2018	SentiWordNet is a lexical resource for opinion mining.	Textual words	[3], 2018	More than 1,00,000 words	https://www.kaggle.com/nltkdata/sentiwordnet
7.	Tmblr, 2017	This dataset contains 100K animated GIFs and 120K sentences describing visual content of the GIFs.	GIFs with sentences	[8], 2017	100k animated GIFs and 120k sentences	https://github.com/yalesong/tumblr_chi2016
8.	Yahoo Flickr Sarcasm (YFS) dataset, 2019	It contains 443 images collected using keywords from the “sarcasm” class and 1403 images collected using keywords from the “non-sarcasm” class.	Images	[17], 2019	1846 images	Not provided by author
9.	Flickr, 2021	It has 31,000 images collected from Flickr, together with reference sentences provided by human annotators.	Images	[1], 2021	30k	https://www.kaggle.com/hsankesara/flickr-image-dataset
10.	Silver-Standard and Gold-Standard Dataset, 2020	The Instagram posts are classified as sarcastic and non-sarcastic based on the hashtags.	Instagram Posts	[7], 2020	20k, 1600	Not provided by the author

978-1-6654-1028-1/22/$31.00 ©2022 IEEE