International Journal of Recent Development in Computer Technology & Software Applications
Vol. 4, Issue 2 – 2020
ISSN: 2581-6276

Sarcasm Detection in Hindi Tweets: A Review

Ms. Surbhi Sharma[1], Dr. Nisheeth Joshi[2]
[1]*Research Scholar, Department of CSE, Banasthali Vidhyapeeth, Niwai.*
[2]*Associate Professor, Department of CSE, Banasthali Vidhyapeeth, Niwai.*

Abstract

Sentiment analysis is a method to identify people's opinion, attitude, sentiment, and emotion in the direction of any precise goal which include individuals, events, topics, products, organizations, services, etc. Sarcasm is a special type of sentiment that comprise of words which might be opposite in that means to what is in reality being said (specially in a sense of insult, wit, irritation, humor). Sarcasm detection in Indian language Hindi is a challenging task in Natural Language Processing (NLP) because of the richness of morphology, Hindi being a fourth popular language within the world stay unexplored in sarcasm detection. Nowadays, posting Hindi sarcastic message on social media like Twitter, Facebook, WhatsApp, etc. has become a new trend to avoid direct negativity. Detecting those indirect negativities i.e., sarcasm within the social media Hindi textual content has come to be a vital undertaking as they influence each business organization. The belongings of sarcasm that makes it hard to analyze and detect is the gap among its literal and meant meaning. Therefore, an automated machine is needed for sarcasm detection in textual records which could be capable of identifying real sentiment of a given text in the presence of sarcasm. In the absence of sufficient resources, processing the NLP tasks which include POS tagging, sentiment analysis, text mining, sarcasm detection, etc., becomes hard for researchers. Here, we proposed a review for sarcasm detection in Hindi tweets.

Sarcasm detection methods inside the textual content may be labeled as rule-based, pattern-primarily based, system learning-based totally and context-primarily based.

Keywords: Sarcasm, Sentiment, Hindi Tweets, NLP.

Introduction

Social networking websites have become a popular platform for users to express their feelings and opinions on various topics, such as events, or products. Social media channels have become a popular platform to discuss ideas and to interact with people worldwide area. Twitter is also important social media network for people to express their feelings, opinions,

and thoughts. Users post more than 340 million tweets and 1.6 billion search queries every day [1] [2]. With 490 million speakers [1] across the world, Hindi stands fourth in recognition after Mandarin, Spanish, and English [2]. In social media such as Twitter, Face-book, WhatsApp, etc., maximum of the Indians now chooses Hindi for communication, and this generates large volumes of statistics. The guide process of mining the emotions from these massive data is a tedious process for individuals as well as organizations.

Therefore, an automated system is needed to perceive the sentiment mechanically from Hindi Text. Sentiment analysis is a task which identifies the orientation of a text closer to a specific goal which include products, individuals, organizations, etc. With the presence of sarcasm, the prediction of sentiment in text frequently goes wrong within the evaluation. Sarcasm frequently conveys terrible meaning the use of advantageous or intensified fine words.

For example, “I love waiting for all time for the doctor”. In the first look, the sentence conveys advantageous sentiment; but, it's miles sarcastic. Due to this, most of the existing sentiment analyzers fail to detect real sentiments.

Recently, many sarcasm detectors have been developed through researchers for textual content scripted in English [3–9]. But there is most effective one mentioned work to be had for detection of sarcasm in Hindi scripted textual content[10].The current work[10] does not recall the herbal Hindi tweets for the experiment.

Their education and trying out set consist of Hindi tweets translated from English scripted tweets.

In this paper, we proposed a framework for sarcasm detection in natural Hindi tweets using online Hindi information because the context.

1. काले धन पे पेनल्टी 200% से घटा के 10% कर दी? काला धन वालों के सामने मोदी जी ने घुटने टेक दिए?- @ArvindKejriwal
2. दो दिन बाद शाहरुख खान अपना 51वां जन्मदिन मनाने वाले हैं, लेकिन उनकी हीरोइन की उम्र लगातार कम होती जा रही है
3. @Rajrrsingh #सुना_है! #iphone7 टिम कुक के टकले पे रख के चार्ज किया जायेगा!
4. आज सुबह मुझे सवच्छता भारत अभियान सड़क पर बिखरा हुआ मिला! #swachbharat #Hindi #clean #mock #sarcasm
5. #JioOffer का आधा से ज्यादा डेटा तो लोग सिर्फ ट्विटर पे अरविन्द केजरीवाल को ट्रोल करने में इस्तेमाल करते है.

Figure 1.A sample of Hindi sarcastic tweets

A sample of herbal Hindi sarcastic tweets is shown in Fig. 1

Twitter is a social media platform where users submit their perspectives of ordinary life. Many businesses and agencies have been inquisitive about this information for the cause of studying the opinion of people regards the political events, popular merchandise or Movies. When a specific product is launched, humans begin tweeting, writing critiques, posting comments, etc. On social media consisting of twitter. People flip to social media community

to read the comments, and opinions from other customers approximately a product before they decide whether or not to purchase or not. If the consumer review is right for the particular merchandise then the customers are purchase the product in any other case not. Organizations are also relying upon on these sites to know the reaction of customers for their merchandise and use the consumer comments to enhance their products [3]. Sentiment evaluation is the opinion of the user for the specific things. Sentiment analysis is the extraction of feeling from any communication (verbal/nonverbal).Two approaches to specific sentiment analysis.

1. **Explicit sentiments:** Direct expression of the opinion about the subject shows the presence of express sentiment.
2. **Implicit sentiments:** Whenever any sentence implies an opinion then such sentence shows the Presence of implicit sentiment (Indirect expression).

Sentiment evaluation and opinion mining depends on emotional phrases in a text to check its polarity (i.e., whether or not it deals positively or negatively with its theme) [4]. Sarcasm is a kind of sentiment wherein people explicit their bad emotions using fine word inside the text. The example of this is “I love the pain of breakup”. The love is the superb words however it explicit the poor feeling, along with breakup in this instance. It is normally used to transfer implicit data within the message someone transmits. It is difficult even for humans to recognize. Used Pattern based approach for detecting sarcasm on twitter. The definition of sarcasm is the hobby of saying or writing the alternative of what you mean, or of speakme in a way supposed to make someone else experience silly or show them that you are angry.

Literature review

In [3], authors display the interest in sarcasm delectation in the tweeter. For capturing real time tweets they use the Hadoop base framework, and tactics that tweets they used the different six algorithms such as parsing based lexicon technology set of rules (PBLGA), tweets contradicting with generic facts (TCUF), interjection word start (IWS), positive sentiment with antonym pair (PSWAP), Tweets contradicting with time-dependent facts (TCTDF), Likes dislikes contradiction (LDC), these set of rules are used identifies sarcastic sentiment effectively. This approach is greater appropriate for real time streaming tweets.

In [4], authors use the computational machine it's miles use for harnesses context incongruity as a foundation for sarcasm detection. Sarcasm classifier makes use of four sorts of functions: lexical, pragmatic, specific incongruity, and implicit incongruity features. They evaluate device on two text forms: tweets and dialogue discussion board posts. For improvement of performance of tweet makes use of the rule of thumb base set of rules, and to improve the performance for discussion board posts, uses the novel method to apply elicitor posts for sarcasm detection. This system additionally introduces error evaluation, the device future work

(a) function of numbers for sarcasm, and (b) conditions with subjective sentiment.

In [5], authors used the machine learning method to sarcasm detection on Twitter in two languages English and Czech. First paintings is sarcasm detection on Czech language. They used the 2 classifier Maximum Entropy (MaxEnt) and Support Vector Machine (SVM) with special combinations of features on each the Czech and English datasets. Also use the special preprocessing method together with Tokenizing, POS- tagging, no stemming and Removing prevent words, its use for finding the issue of Czech language.

In [6], authors have investigated characteristics of sarcasm on Twitter. They are concerned not just with figuring out whether tweets are sarcastic or no longer, however additionally keep in mind the polarity of the tweets. They additionally have compiled a number of guidelines which enhance the accuracy of sentiment analysis while sarcasm is known to be present. Researcher have advanced a hash tag tokenizes for GATE approach in order that sentiment and sarcasm located within hash tag can be detected more easily. Hash tag tokenization approach is very beneficial for detection of sarcasm and assessments the polarity of the tweet i.e. nice or bad.

In [7], authors are used methods consisting of lexical and pragmatic factors which can be used for differentiate between sarcasm from fantastic and negative sentiments expressed in Twitter messages. They additionally created corpus of sarcastic Twitter messages in which dedication of the sarcasm of every message has been made by using its author. Corpus is used to evaluate sarcastic utterances in Twitter to utterances that display positive or negative attitudes without sarcasm.

In [8], authors have developed an underestimation recognizer to decide sarcasm on Twitter consists of a high-quality sentiment contrasted with a poor scenario of sarcasm in tweets. They use novel bootstrapping algorithm that robotically learns lists of positive sentiment phrases and terrible scenario phrases from sarcastic tweets. They display that determine contrasting contexts using the terms learned thru bootstrapping.

Rule-based approaches try and identify sarcasm via specific evidences. These evidences are captured in phrases of regulations that rely on indicators of sarcasm. Focus on identifying whether a given simile (of the form '* as a *') is intended to be sarcastic. They use Google seek in order to determine how possibly a simile is. They present a 9-step approach where at each step rule; a simile is validated using the quantity of seek results. Strength of this method is they present an error analysis similar to multiple rules [9]. The hash tag sentiment is a key indicator of sarcasm. Hash tags are often used by tweet authors to spotlight sarcasm, and hence, if the sentiment expressed by using a hash tag does no longer agree with rest of the tweet, the tweet is expected as sarcastic. They use a hash tag tokenizer to cut up hashtags made of concatenated phrases [6].

In [11], we've studied the distinctive techniques.

Following are the method for sarcasm detection on twitter.

1) Feature extraction

This method are used for annotating the data, it contain three categories.

A) Sarcasm as wit: when used as a wit, sarcasm is used with the cause of being funny.

B) Sarcasm as whimper: whilst used as whimper, sarcasm is employed to show how annoyed or angry the person is.

C) Sarcasm as evasion: it refers to the state of affairs while the individual wishes to keep away from giving a clean answer, thus, makes use of sarcasm.

2) Sentiment-related Features

It extracts sentimental components of the tweet and counts them. Positive emotional content material (e.G. Love, happy, etc.) and negative emotional content material (e.G. Hate, sad, etc.).Calculate the ratio of emotional phrases.

$$P(t) = (\& \cdot PW + pw) - (\& \cdot NW + nw) / (\& \cdot PW + pw) + (\& \cdot NW + nw)\ 1$$

t=tweet, pw=high quality phrases, nw =bad phrases, PW=extraordinarily emotional high-quality words, NW= surprisingly emotional negative phrases, & =weight larger than 1.

3) Punctuation-Related Features

It displays behavioral factors including low tones, Facial gestures or exaggeration. These elements are translated into a sure use of punctuation or repetition of vowels while the message is written.

- Number of exclamation marks
- Number of query marks
- Number of dots
- Number of all-capital words
- Number of quotes

4) Syntactic and Semantic Features

It refers to the situation when the character desires to keep away from giving a clear answer, thus, makes use of sarcasm.

- Use of uncommon words
- Number of unusual words
- Existence of common sarcastic expressions
- Number of interjections
- Number of laughing expressions

5) Pattern-Related Features

Pattern is described as an order series of phrases. Divide words into two classes: a first one called CI containing words of which the content material is important and a 2d one known as GFI containing the phrases of which the grammatical function is greater essential.

6) A behavioral modelling approaches

In this technique content material to observe how to broaden a systematic approach for effective sarcasm detection through not simplest reading the content of the tweets however through also exploiting the behavioral traits of customers derived from them beyond activities [10].

Sarcasm Detection Method

Sarcasm detection methods within the textual content can be classified as rule-based totally, pattern-primarily based, device learning-based totally definitely and context-primarily based.

1 Rule primarily based Approach

Rule-based method is the most basic method used for sarcasm detection in the textual content. Rule-based methods is a try and pick out sarcasm through precise evidence. In this method evidence is captured in the form of rules that depend on indicators of sarcasm. pick out sarcasm in similes the use of Google searches so as to determine how probable a simile is.

In this approach, we in particular cognizance on hyperbolic and syntactic capabilities of the textual content. Interjections, intensifiers and punctuation symbols are the most common hyperbole features used within the textual content to infer sarcastic messages. The severe adjective and extreme adverb act as intensifiers for the textual content. some examples of intensifiers are very well enjoyed, extraordinary weather, so beautiful, etc.

The Rule primarily based method iseasy to enforce and frequently attains suitable accuracy for textual content classification. Three rule-primarily based classification techniques are proposed, one each for English, Hindi and Telugu.

It affords a 9-step technique wherein at every step/rule, a simile is demonstrated as non-sarcastic using the wide variety of search results. Let consider an example of movie with negative tag, we elaborate the tree branches with different characteristics by applying the rule based approach.

International Journal of Recent Development in Computer Technology & Software Applications
Vol. 4, Issue 2 – 2020
ISSN: 2581-6276

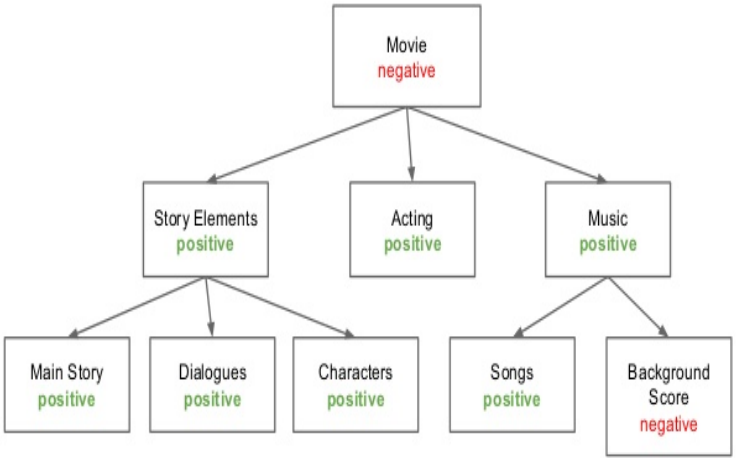

Figure 2.Example of Rule based Approach

[12] advise that hashtag sentiment is a key indicator of sarcasm. Hashtags are regularly utilized by tweet authors to spotlight sarcasm, and hence, if the sentiment expressed by way of a hashtag does no longer agree with rest of the tweet, the tweet is anticipated as sarcastic. They use a hashtag tokenizer to cut up hashtags made from concatenated words.

[6] present rule-primarily based classifiers. The first makes use of a parse–based totally lexicon era set of rules that creates parse bushes of sentences and identifies scenario terms that bear sentiment. If a negative phrase takes place in an effective sentence, then the sentence is expected as sarcastic. the second set of rules aims to seize hyperbolic sarcasm (i.e., by the use of interjections (such as '(wow)' and intensifiers (together with 'absolutely') that arise together.

[14] present rule-based totally classifiers that search for a tremendous verb and a negative state of affairs word ina sentence. The set of negative situation terms are extracted the usage of a well-structured, iterative algorithm that starts with a bootstrapped set of high-quality verbs and iteratively expands both the sets (namely, advantageous verbs and negative state of affairs terms). They experiment with special configurations of regulations along with proscribing the order of the verb and scenario phrase.

2 Pattern Based Approach

Sarcasm is a kind of sentiment where public expresses their negative feelings using fantastic word inside the text [4]. It is normally used to carry implicit statistics within the message a person transmits. Sarcasm may be used for extraordinary functions including criticism. It is very tough for people to understand. Recognizing sarcastic statements may be very useful to improve computerized sentiment evaluation of facts collected from specific web sites or social networks. Sarcasm is when someone says something distinct from what he means. Pattern based approach is used for detecting sarcasm on twitter.

International Journal of Recent Development in Computer Technology & Software Applications
Vol. 4, Issue 2 – 2020
ISSN: 2581-6276

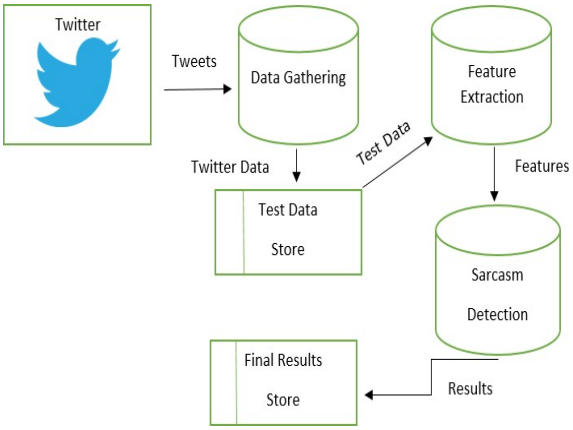

Figure 3.block diagram of sarcasm detection on twitter

3 System Learning Algorithms

Machine learning-based technique is the most not unusual technique used for classification. The performance of the system mastering classifiers often relies upon on dataset and characteristic set quality. In this thesis, lexical, syntactic, hyperbole, sentiment capabilities are utilized in various system getting to know algorithms. The classifiers evaluated are Naive Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Random Forest (RF) and AdaBoost. Among these classifiers, NB outperformed different classifiers because of independence of text capabilities. In text classification (particularly when schooling set is small), NB performs higher than different classifiers.

Most paintings in statistical sarcasm detection relies on different varieties of Support Vector Machines (SVM) [14]. [14] use SVM and Logistic Regression, with the χ 2 test used to discover discriminating features. [14] examine rule-based strategies with a SVM-based classifier. Diagram 4 explain the procedure for machine learning approach for sarcasm detection.

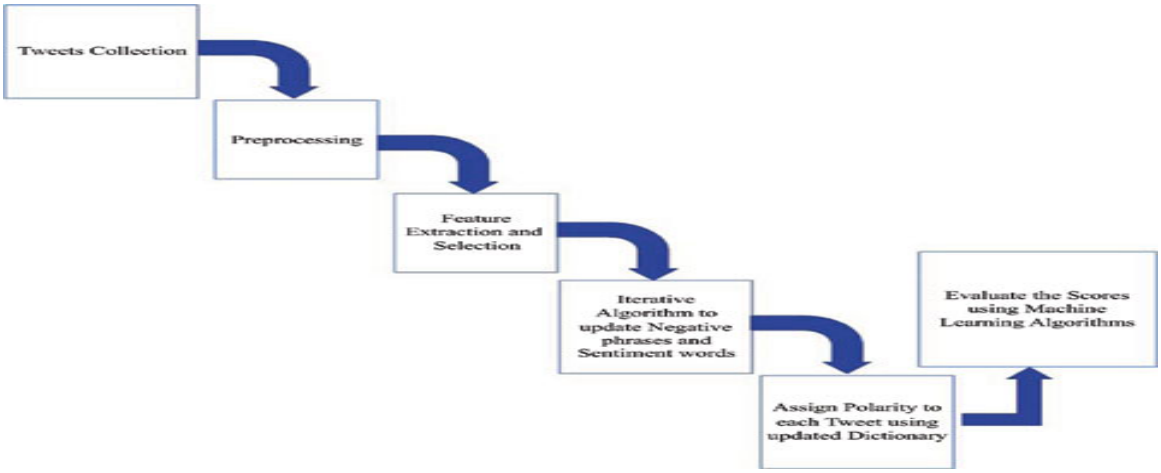

Figure 4.System(Machine Learning) Approach for sarcasm detection

4 Context based Approach/Feature Set

Context-based method is the most vital technique for text classification. Sarcasm can be detected by thinking about lexical, pragmatic, hyperbolic or different such capabilities of the textual content. Some functions also can be developed using sure styles such as unigram, bigram, trigram, etc. There may be features based totally on verbal or gestural clues such as emoticons, onomatopoeic expressions in laughter, nice interjections, citation marks, use of punctuation which can assist in detecting sarcasm. But all these features are not enough to identify sarcasm in text until the context of the text is known. The machine, in addition to human, should be aware about the context of the textual content and relate it to widespread international knowledge to be able to identify sarcasm more accurately. In this approach, we specifically cognizance on situation, topical, temporal, and ancient context of the textual content.

Most procedures use bag-of-phrases as capabilities. We attention on functions related to the text to be categorized. Contextual features (i.e., capabilities that use statistics beyond the textual content to be classified) are specified using context-based approach.

[15] introduce functions associated with ambiguity, unexpectedness, emotional scenario, etc. Ambiguity features cover structural, morpho- syntactic, semantic ambiguity, while unexpectedness capabilities degree semantic relatedness. [16] use a fixed of pa1erns, in particular fantastic verbs and negative state of affairs phrases, as features for a classifier (in addition to a rule-based classifier).

[17] explore pass-gram and man or woman n-gram-primarily based functions. [3] encompass seven sets of capabilities which include maximum/minimum/gap of intensity of adjectives and adverbs, max/min/average range of synonyms and synsets for words within the target text, etc. [4] use similar features for irony detection. [18] contain ellipsis, hyperbole and imbalance in their set of capabilities. [4] use functions corresponding to the linguistic theory of incongruity. The functions are categorized into units: implicit and express incongruity-based functions. In addition, in addition they use complex gaze-based totally features based on saliency graphs which connect words in a sentence with edges representing saccades between the phrases.

Conclusion

Sarcasm detection research has grown significantly within the past few years, necessitating a look-again at the overall picture that these character works have led to. This paper describes the framework for sarcasm detection in Hindi tweets and also describes the different techniques use for sarcasm detection we determined thatrule-based totally procedures seize evidence of sarcasm inside the shape of regulations along with that the sentiment of the hashtag does no longer suit the sentiment of the relaxation of the tweet. Statistical processes use features like sentiment changes, unique semi-supervised patterns, etc. To include context,

additional capabilities particular to the author, the communication and the topic were explored inside the past. An underlying theme of those beyond techniques (either in phrases of policies or functions) is attempting to pick out the 'irony' and 'hurtful nature' that is at the supply of sarcasm.

We have studied the different kind of methods for sarcasm detection; we additionally studied the pattern- based technique for sarcasm detection. In this paper, the methods are used to come across sarcasm or as nicely as check the behavioral approach of the user, the approach make used one-of-a-kind element of the tweet, and also by the use of of Part-of-Speech tags to extract styles characterizing the level of sarcasm of tweets. By the usage of #sarcasm collect all the sarcastic tweets. In this way we talk the specific method which includes Feature extraction, Sentiment-associated Features, Punctuation-Related Features, Syntactic and Semantic Features, Pattern-Related Features, behavioral modelling technique for detection of sarcasm inside the tweet. By the use of unique algorithm or classifier consisting of Random Forest, Support Vector Machine (SVM), okay Nearest Neighbors (k-NN) and Maximum Entropy, test the accuracy and performance. In future scope these approaches will show proper results.

References

1. D. Chaffey, Global Social Media Research Summary 2016. URL ⟨http://www.smart insights.com/Social-media-marketing/social-media-strategy/new-global-social-media-research/⟩.
2. W. Tan, M.B. Blake, I. saleh, S. Dustdar, Social-network- sourced big data analytics, InternetComput.17 (5) (2014) 62-69.
3. S.K. Bharti B. Vachha, R.K. Pradhan, K.S. Babu, S.K. Jena "Sarcastic sentiment detection in tweets Streamed in real time: a big data approach", Elsevier 12 July 2016.
4. Aditya Joshi, Vinita Sharma, Pushpak Bhattacharyya "Harnessing Context Incongruity for Sarcasm Detection" Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Short Papers), pages 757–762, Beijing, China, July 26-31, 2015. C 2015 Association for Computational Linguisti.
5. Toma Ptacek Ivan Habernal and Jun Hong "Sarcasm Detection on Czech and English Twitter", Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 214– 223, Dublin, Ireland, August 23-292014.
6. D. Maynard, M. A. Greenwood. 2014. "Who cares about sarcastic tweets? Investigating the Impact of sarcasm on sentiment analysis", In Proceedings of the LREC 2014 May26-31.
7. R. Gonzalez-Ibanez, S. Muresan, and N. Wacholder. 2011. "Identifying Sarcasm in Twitter: A Closer Look". In Proceedings of the 49th Annual Meeting of Association for Computational Linguistics.

8. E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang, “Sarcasm as contrast between a positive Sentiment and negative situation”, in Proc. Con Empirical Methods Natural Lang. Process, Oct.2014, pp.704_714.
9. Tony Veale and Yanfen Hao. 2010. “Detecting Ironic Intent in Creative Comparisons”, In ECAI, Vol. 215.765-770.
10. Rajadesingan, R. Zafarani, and H. Liu, ``Sarcasm detection on Twitter A behavioral modeling Approach”, in Proc. 18th ACM Int. Conf. Web Search Data Mining, Feb. 2015, pp.79_106.
11. M. Bouazizi, T. Ohtsuki, “Pattern-Based Approach for Sarcasm Detection on Twitter” VOLUME 4, 10.1109/ACCESS.2016.2594194.
12. DianaMaynardandMarkAGreenwood.2014.Whocaresaboutsarcastictweets?investigatingtheimpactofsarcasm on sentiment analysis. In *Proceedings of LREC*.
13. Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra DeSilva, Nathan Gilbert, and Ruihong Huang. 2014. Sarcasmas Contrast between a Positive Sentiment and Negative Situation. In *EMNLP*.704-714.
14. Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Semi-supervised recognition of sarcastic sentences in twi1er andamazon. In *Proceedings of the Fourteenth Conferenceon Computational Natural Language Learning*. Association for Computational Linguistics, 107-116.
15. Antonio Reyes, Paolo Rosso, and Davide Buscaldi. 2012. From humor recognition to irony detection: fte figurative language of social media. *Data & Knowledge Engineering* 74 (2012), 1-12.
16. Nikita Desaiand Anand kumar D Dave. 2016. Sarcasm Detection in Hindi sentences using Support Vector machine. *International Journal* 4, 7 (2016).
17. Antonio Reyes, Paolo Rosso, and Tony Veale. 2013. A multidimensional approach for detecting irony in twi1er. *Language Resources and Evaluation* 47, 1 (2013), 239-268.
18. Konstantin Buschmeier, Philipp Cimiano, and Roman Klinger. 2014. An impact analysis of features in a classification approach to irony detection in product reviews. *ACL 2014* (2014), 42.

