

# A Review on Sarcasm Detection Based on Machine Learning

Vaishnavi Vhora<sup>1</sup>, Vijaya Kamble<sup>2</sup>

<sup>1</sup>M.Tech Scholar, Department of Computer Science & Engineering, Guru Nanak Institute of Engineering and Technology, Nagpur, Maharashtra, India.

<sup>2</sup>Assistant Professor, Department of Computer Science & Engineering, Guru Nanak Institute of Engineering and Technology, Nagpur, Maharashtra, India.

## ABSTRACT

### Article Info

Volume 7, Issue 2

Page Number: 52-57

### Publication Issue :

March-April-2021

### Article History

Accepted : 01 March 2021

Published : 05 March 2021

Sarcasm is a subtle form of irony, which can be widely used social networks such as twitter. It is usually used to transmit hidden information, a message sent by people. Due to a different purposes Sarcasm can be used like criticism and ridicule. But even this is difficult for a person to recognize. The sarcastic reorganization system is very helpful for the improvement of automatic sentiment analysis collected from different social networks and microblogging sites. Sentiment analysis refers to internet users of a particular community, expressed attitudes and opinions of identification and aggregation. To detecting sarcasm we propose a pattern-based approach using Twitter data. We proposes four sets of features that include a lot of specific sarcasm. We use them to classify tweets as sarcastic and non-sarcastic. We also study each of the proposed feature sets and evaluate its additional cost classifications.

Keywords : Sarcasm detection, Twitter, Sentiment analysis, Machine learning

## I. INTRODUCTION

Today the twitter has been a very biggest network, by using peoples share their opinions and thoughts. In previous years, twitter content has grown again and is a typical example of big data. Twitter has been an official site which contain active 288 million users and sent 500 million tweets are daily. In this data, many companies and organizations interested in political events, popular products, or films towards the people opinion research purposes [1].

However, due to the limitations of the unofficial language and characters used by Twitter (that is, 140 characters per tweet), it is very difficult to understand the opinions of users and conduct such an

analysis. In addition, the presence of sarcasm is even more difficult: sarcastic when a person says that they are not what they mean [2].

Oxford dictionary express sarcasm as "the use of sarcasm to Express or convey contempt". Free Dictionary also defines sarcasm in the form of irony intended to express contempt. As sarcastic the main focus is on automated sentiment analysis of existing systems for improvement and improved, we also use two terms-synonyms that are [3]. Sarcasm detection is very difficult in real life.

As a rule, people use sarcasm in everyday life, not only jokes and humor, but also criticism or comments, ideas, types and effects.

Therefore, social networks are usually widely used, in particular microblogging sites such as Twitter. Thus, the modern approach to sentiment analysis and opinion analysis usually performs lower indicators when analyzing collected data, such sites. Maynard and Greenwood [4] show that the effectiveness of sarcastic analysis can be significantly improved when sarcasm is detected in sarcastic statements. Therefore, there is a need for an effective way to detect sarcasm. Identifying sarcasm helps with the task of analyzing mood when it is performed on microblogging sites such as Twitter. Mood analysis and opinion mining rely on emotional words in a text to detect its polarity (that is, whether it relates to "positively" or "negatively" in its thread). However, the appearance of the text can lead to confusion. A typical example is when there is sarcasm in the text. Some sarcastic texts are very common. "All your products are incredibly amazing !!! The some peoples are explain what he said, he actually doesn't mean. Although some are shows they are sarcastic, so its need to detect the sarcastic messages are automatically [5] [6]. So the aim of this paper is to propose a system to automatically detect a sarcastic tweet. Mainly the sarcasm is used in social network. To detect the sarcastic messages we define automatic system, this shows how the message or information is used. In short to detect the message is sarcastic or not. In this paper we also study different features.

## II. RELATED WORK

In recent years, attention paid to analyzing twitter's mood by researchers and a number of current documents have been applied to classifying tweets. Sriram [7] the author classifies tweets into a predefined set of General classes, including events, opinions, transactions and private messages, non-contextual features used, such as the presence of slang, phrases about temporary events, opinion by word, and information about Twitter users. The author of the article [8] [9] proposed a method for identifying emotional and verbal patterns in Twitter data. However, most of the work was done to classify

tweets based on the polarity of user sentiment towards the topic of specific, focusing on the content of the tweet. Various functions have been proposed. They include presence of diagrams [10], frequency and non-text features such as emoticons [11] [12]. The author [13], defines framework that learns to classify the words and that words emotions of context.

Sarcasm has been used in everyday conversation in a very long time. Therefore, sarcasm in terms of psychological [14] and neurobiological [15] is the subject of deep research. However, it has been studied as a language behavior that characterizes a person. In [16], Burfoot and Baldwin introduced a set of attributes, including profanity and slang use, and they are certified "semantic action" and also used SVM classifier to classify satirical articles. In [17], studying the contextual components that are used to convey sarcastic ironic words, suggest that sarcasm requires the presence of four people. Tepperman [18], the proposed approach will find automatically sarcasm in conversational dialogue.

In [19] proposed to classify this text as political, humorous, sarcastic and sarcastic. In paper [20] formulate the task of detecting sarcasm as a task of removing ambiguity in the sense that the word can have a literal meaning or irony, and therefore the irony of the word. Kings et al. [21], this suggests that rather than trying to determine whether a tweet is sarcastic, it makes sense to consider the context: but most snippets aim to simply classify the typeset as sarcastic and non-sarcastic.

Davidov [22] and Tsur [23], a semi-controlled algorithm for sarcasm identification is proposed. They experimented with two sets of data: one is amazon and the other is Twitter. The results were interesting and their approach depended on the frequency of word detection, but the training sets were related to the topic and this approach treated what are called contextual words, regardless of grammatical features. It also does not distinguish between sentimental and unconscious words.

Samples that do not take into account words of emotional content or cancel them low in order to be able to reduce the probability approximation.

In [24], Twitter users relied on the hashtags they were employing to identify the irony of the tweet. Riloff et al. [25], when a positive mood contrasts with a negative situation, a method of detecting a certain type of irony is proposed. They use the single first word "love" and a series of sarcastic tweets to automatically detect expressions that indicate a positive mood or phrase quoting a negative situation. In [26] introduced behavioral modeling to detect irony on Twitter. They identified a variety of forms of irony and its manifestations on Twitter, showing the importance of historical information gathered from past tweets and identifying irony. It has proven to be very effective, but this approach is less effective in the absence of previous knowledge of the user. It is to determine the tweet from the collected data relying on the best feature extraction. In the real-time stream of tweets posted by random users, the approach is difficult to implement, and the knowledge base size is very fast. The author in [27], used machine learning to identify and detect sarcastic tweets, to investigate the impact of vocabulary and practical factors on the performance, and to classify positive and negative tweets.

### III. PROPOSED SYSTEM

This given a set of tweets, our goal is to rank each one according to whether it is sarcastic or not. Fig 1, shows the block diagram of the system.

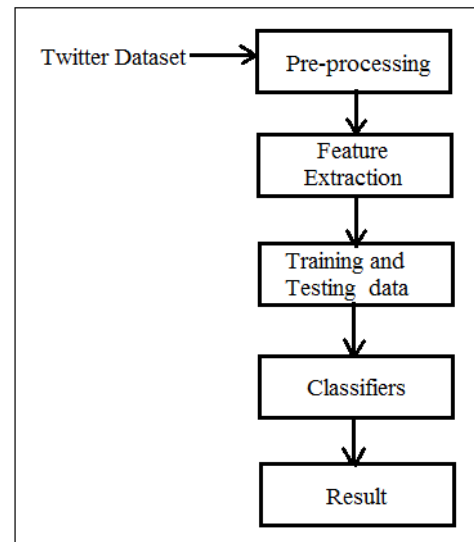


Fig. 1. Block diagram of the system

When the algorithm is used on data, the machine learns on the basis of type of data, like we provide input tweets, and there output is either positive, negative or neutral. So, when the machine learns itself, there is not any issue, for which language input is provided, the only matter is their output.

From the dataset, we extract the published tweet, and then extract information about past tweets (the each user 80 tweets). The examples are as follows:

- 1) This paper is coming along... #not
- 2) Finding out your friends lives through tweets is the greatest feeling. #sarcasm

#### Feature Extraction

Next, the system feature extraction is carried out for the data. Four features are extracted are as follows:

- 1) The features related to sentiment:

Opinion mining or sentiment analysis is the process by which an individual determines the emotions expressed in his or her writing. A very common type of irony, widely used in both normal conversations, this type of combat qualifies as "whimper", the management of the social network "twitter". (e.g., "I love being ignored all the time").

2) The features related to Punctuation:

For detecting any type of sarcasm the feature related to sentiments are not enough and they all components of tweets are not used. As mentioned before, sarcasm is a sophisticated form of speech: not only does it play a word or meaning, it converts these aspects into punctuation or repeated use of vowels when a message is written, such as a low-tone facial gesture. After detecting these aspects, we extract a set of qualified properties. The values for each tweet is calculates:

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

3) The features related to syntactic and Semantic:

In addition to the functions associated with punctuation, some common expressions are usually used in a sarcastic context. Associate these expressions in punctuation to determine if what is being said sarcastic or not. In addition, in other cases, people as a rule, complex sentences or use unusual words to hide the listener / reader to make a clear answer. This is common when sarcasm is used as an "evasion", and the person aims to hide their true feelings and opinions with sarcasm. Thus, we emphasize the following characteristics that reflect these aspects:

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

Eg: “You are incredibly funny -\_- ”

4) The feature related to pattern

The selection pattern of the previous subsection and qualified “general ironic expression “is very common and even in conversation. However, their number is

small, they are not unique, and our training and test seals largely do not include them. In this case, we will go further and extract another set of features.

In this approach, the words are classified according to two categories: high-frequency words and content words that are based on its data, the frequency of the frequency, and determining the sample as high-frequency words and slots in the ordered sequence of contextual words.

Training and Testing Data

The training set is what we are preparing, and consistent with our model mainly corresponds to the parameters, while the test data is used only for model performance assessment. Training data output available on the model while the testing data is unseen data for which predictions must be done. The K-fold cross-validation is used 10 times for training and testing datasets. To provide solution to this problem we use K-fold Cross Validation that dividing the data into folds and ensuring that each fold is used as a testing set at some point. Train and test the extracted data, and perform an SVM, KNN, and Random forest algorithm for prediction of sarcastic or not and calculate the accuracy of each algorithm. This method will train about 70%of the given data set, and the remaining 30% will be used for testing purposes.

Classification

We ran the classification using the classifiers such as Support Vector Machine (SVM), k Nearest Neighbors (KNN), and Random forest. The result section presents the performance of classifiers on the dataset.

IV. CONCLUSION

In this work, we propose a system which detects sarcasm on English as well as on Hindi tweets in Twitter. Sarcasm is very dependent and highly contextual; therefore, sentiment and other contextual clues to help detect the sarcasm text. The system uses sarcastic tweets, 9,104 tweets containing #sarcasm,

and #not dataset. The system uses the SVM, KNN, and Random forest classifier. The approach has shown good results. The all patterns of sarcastic not cover in we extracted patterns. So in future Neural Network, Genetic Algorithm and Pattern-based approach can be combined for more accuracy.

## V. REFERENCES

- [1]. Yi Tay, Mondher Bouazizi And Tomoaki Otsuki (Ohtsuki), "A Pattern-Based Approach For Sarcasm Detection On Twitter" in August 24, 2016
- [2]. Yi Tay†, Luu Anh Tuan, Siu Cheung Huiφ, JianSuδ, "Reasoning with Sarcasm by Reading In-between"arXiv:1805.02856v1 [cs.CL] 8 May 2018
- [3]. Aditya Joshi, Pushpak Bhattacharyya, Mark J. Carman.2018, "Sarcasm Detection Using Incongruity within Target Text", In: Investigations in Computational Sarcasm",Cognitive Systems Monographs, vol 37. Springer
- [4]. . Aditya Joshi, Pushpak Bhatta charyya,Mark J. Carman.2018, "Sarcasm Detection Using Contextual Incongruity. . . In: Investigations in Computational Sarcasm", Cognitive Systems Monographs, vol 37. Springer
- [5]. Shubhadeep Mukherjee, Dr. Pradip Kumar Bala, "Sarcasm Detection in MicroblogsUsing Naïve Bayes and Fuzzy Clustering" In Proceedings of Technology in Society,2017 pages 19-27
- [6]. Nishant Nikhil, Muktabh Mayank Srivastava, "Binarizer at SemEval-2018 Task 3:Parsing dependency and deep learning for irony detection"arXiv preprintarXiv:1805.01112
- [7]. B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in twitter to improve information filtering," in Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Jul. 2010.
- [8]. C. G. Akcora, M. A. Bayir, M. Demirbas, and H. Ferhatosmanoglu, "Identifying breakpoints in public opinion," in Proc. 1st Workshop Soc. Media Anal., Jul. 2010.
- [9]. M. W. Berry, Ed., Survey of Text Mining: Clustering, Classification, and Retrieval. New York, NY, USA: Springer-Verlag, 2004.
- [10]. B. Pang, L. Lillian, and V. Shivakumar, "Thumbs up: Sentiment classification using machine learning techniques," in Proc. ACL Conf. Empirical Methods Natural Lang. Process., vol. 10. Jul. 2002
- [11]. M. Boia, B. Faltings, C.-C. Musat, and P. Pu, "A Is worth a thousand words: How people attach sentiment to emoticons and words in tweets," in Proc. Int. Conf. Soc. Comput., Sep. 2013.
- [12]. K. Manuel, K. V.Indukuri, and P. R. Krishna, "Analyzing internet slang for sentiment mining, "in Proc. 2nd Vaagdevi Int. Conf. Inform. Technol. Real World Problems, Dec. 2010.
- [13]. L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, "Adaptive recursive neural network for target-dependent Twitter sentiment classification," in Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics, vol. 2. Jun. 2014,
- [14]. F. Stringfellow, Jr., The Meaning of Irony: A Psychoanalytic Investigation. New York, NY, USA: State Univ. New York, 1994.
- [15]. S. G. Shamay-Tsoory, R. Tomer, and J. Aharon-Peretz, "The neuroanatomical basis of understanding sarcasm and its relationship to social cognition," Neuropsychology, vol. 19, , May 2005.
- [16]. C. Burfoot and T. Baldwin, "Automatic satire detection: Are you having a laugh?" in Proc. ACL-IJCNLP, Aug. 2009.
- [17]. J. D. Campbell and A. N. Katz, "Are there necessary conditions for inducing a sense of

sarcastic irony?" Discourse Process. vol. 49, no. 6, Aug. 2012.

- [18]. J. Tepperman, D. Traum, and S. S. Narayanan, "Yeah right!: Sarcasm recognition for spoken dialogue systems," in Proc. InterSpeech, Sep. 2006
- [19]. T. Veale and Y. Hao, "Detecting ironic intent in creative comparisons," in Proc. ECAI, Aug. 2010,
- [20]. D. Ghosh, W. Guo, and S. Muresan, "Sarcastic or not: Word embeddings to predict the literal or sarcastic meaning of words," in Proc. EMNLP, Sep. 2015
- [21]. Z. Wang, Z. Wu, R. Wang, and Y. Ren, "Twitter sarcasm detection exploiting a context-based model," in Proc. Web Inf. Syst. Eng. (WISE), Nov. 2015
- [22]. O. Tsur, D. Davidov, and A. Rappoport, "ICWSM-A great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews," in Proc. AAAI Conf. Weblogs Soc. Media, May 2010,
- [23]. D. Davidov, O. Tsur, and A. Rappoport, "Semi-supervised recognition of sarcastic sentences in Twitter and Amazon," in Proc. 14th Conf. Comput. Natural Lang. Learn., Jul. 2010
- [24]. D. Maynard and M. A. Greenwood, "Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis," in Proc. 9th Int. Conf. Lang. Resour. Eval., May 2014
- [25]. 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. Conf. Empirical Methods Natural Lang. Process., Oct. 2013
- [26]. A. 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
- [27]. S. Muresan, R. Gonzalez-Ibanez, D. Ghosh, and N. Wacholder, "Identification of nonliteral

language in social media: A case study on sarcasm," J. Assoc. Inf. Sci. Technol., Jan. 2016.

**Cite this article as :**

Vaishnavi Vhora, Vijaya Kamble, "A Review on Sarcasm Detection Based on Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 2, pp. 52-57, March-April 2021. Journal URL : <https://ijsrcseit.com/CSEIT217221>