

LEVERAGING MACHINE LEARNING TOOLS AND TECHNIQUES IN THE DETECTION OF PARODY ON TWITTER AND OTHER SOCIAL NETWORKING SITES

Saatvik Wadhwa

ABSTRACT

Parody is an inconspicuous type of contradiction, which can be broadly utilized informal communities like Twitter. It is normally used to send stowed away data, a message sent by individuals. Because of alternate purposes, I can utilize sarcasm like analysis and criticism. In any case, even this is hard for an individual to perceive. The snide redesign framework is extremely useful for improving programmed feeling investigation gathered from various interpersonal organizations and microblogging locales. Feeling examination alludes to web clients of a specific local area, communicated perspectives and assessments of ID and conglomeration. To distinguishing mockery, we propose an example-based methodology utilizing Twitter information. We present four arrangements of elements that incorporate a ton of explicit mockery. We use them to group tweets as mocking and non-snide. We likewise concentrate on every one of the proposed include sets and assess its extra expense groupings.

I. INTRODUCTION

Today, Twitter has been an extremely greatest organization by utilizing people groups to share their perspectives and considerations. In earlier years, twitter content has developed again and is a run of the mill illustration of enormous information. Twitter has been an authority site containing dynamic 288 million clients and sent 500 million tweets daily. In this information, many organizations and associations are inspired by political occasions, well-known items, or movies for individuals' assessment research purposes [1].

Notwithstanding, because of the restrictions of the informal language and characters utilized by Twitter (that is, 140 characters for each tweet), it is truly challenging to comprehend the assessments of clients and lead such an investigation. What's more, the presence of mockery is considerably more troublesome: wry when an individual says that they are not what they mean [2].

Oxford word reference expresses mockery as "the utilization of mockery to Express or passes on disdain". Free Dictionary likewise characterizes mockery as incongruity expected to communicate hatred. As mocking the fundamental spotlight is on mechanized opinion examination of existing frameworks for development and improvement, we likewise utilize two terms-equivalents [3]. Mockery discovery is extremely challenging, in actuality.

Generally speaking, individuals use mockery in day-to-day existence, jokes and humour, yet additionally analysis or remarks, thoughts, types and impacts.

In this manner, interpersonal organizations are typically broadly utilized, specifically microblogging destinations like Twitter. Consequently, the advanced way to deal with feeling investigation and assessment examination, as a rule, performs lower markers when dissecting gathered information, such destinations. Maynard

and Greenwood [4] show that the adequacy of mocking examination can be altogether further developed when mockery is distinguished in snide explanations. In this way, there is a requirement for a viable way of identifying mockery.

Distinguishing mockery assists with the assignment of examining disposition when it is performed on microblogging destinations like Twitter. Disposition examination and assessment mining depend on enthusiastic words in a text to recognize its extremity (that is, regardless of whether it identifies with "emphatically" or "contrarily" in its string). Nonetheless, the presence of the text can prompt disarray. A normal model is when there is mockery in the text. Some snide readers are extremely normal. "Every one of your items is inconceivably stunning!!!. A few people groups are

clarified what he said; he doesn't mean. Although some show they are mocking, its need to recognize the wry messages is consequently [5] [6].

So, the point of this paper is to propose a framework to distinguish a snide tweet naturally. Mostly the mockery is utilized in the interpersonal organization. We characterize the programmed framework to identify the disparaging messages; this shows how the message or data is used. In short, to determine the letter is disparaging or not. In this paper, we likewise concentrate on various components.

II. PROPOSED SYSTEM

Given a bunch of tweets, we will probably rank everyone as per if it is snide. Fig 1 shows the block diagram of the framework.

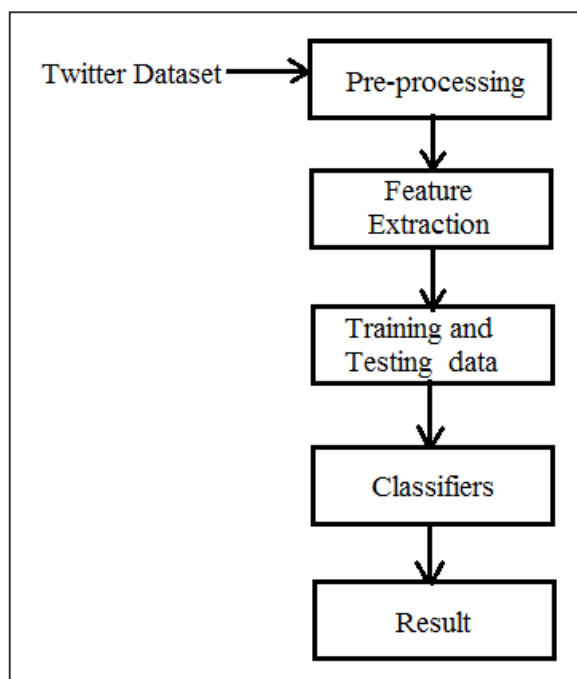


Fig 1: Block Diagram of Suggested framework

When the calculation is utilized on information, the machine learns based on kind of information, similar to we give input tweets, and their yield is either sure, negative or nonpartisan. In this way, when the machine learns itself, there isn't an issue

for which language input is given; the main matter is their result.

We disengage the distributed tweet from the dataset and separate data about past tweets (every client 80 tweets). The models are as per the following:

- 1) This paper is going along... #not
- 2) Finding out your companions survives tweets is the best inclination. #sarcasm

Component Extraction

Then, the framework includes extraction is completed for the information. Four elements are separated as per the following:

- 1) The elements identified with opinion:

Assessment mining or feeling investigation is the interaction by which an individual decides the feelings communicated in their composition. An extremely normal kind of paradox, generally utilized in both typical discussions, this sort of battle qualifies as "whine", the administration of the informal organization "Twitter". (e.g., "I love being overlooked all the time").

- 2) The components identified with Punctuation:

For recognizing any mockery, the element identified with feelings are sufficiently not, and all parts of tweets are not utilized. As referenced previously, mockery is a refined type of discourse: in addition to the fact that it plays a word or which means it changes over these viewpoints into accentuation or rehashed utilization of vowels when a message is composed, like a low-tone facial signal. After recognizing these angles, we extricate a bunch of qualified properties. The qualities for each tweet is ascertained:

- Number of every single capital word
- Number of dabs
- Number of interjection marks
- Number of question marks
- Number of statements

- 3) The elements identified with syntactic and Semantic:

Notwithstanding the capacities related to accentuation, some typical statements are generally utilized in a wry setting. Partner these articulations

in accentuation to decide whether what is being said is wry or not. Likewise, in different cases, individuals, when in doubt, use complex sentences or use surprising words to conceal the audience/peruser and make an unambiguous response. This is normal when mockery is utilized as an "avoidance", and the individual intends to hide their actual sentiments and feelings with sarcasm. Accordingly, we stress the accompanying attributes that mirror these perspectives:

- Use of exceptional words
- Number of contributions
- Number of unique words
- Existence of normal wry articulations
- Number of giggling articulations

E.g.: "You are unquestionably amusing - _-"

- 4) The component identified with the design

The determination example of the past subsection and qualified "general amusing articulation" is extremely normal and surprisingly. Be that as it may, their number is little, they are not special, and our preparation and test seals do exclude them to a great extent. For this situation, we will go further and concentrate on one more arrangement of components.

In this methodology, the words are characterized by high-recurrence words and content words that depend on their information, the recurrence of the recurrence, and deciding the example as high-recurrence words and openings in the arranged succession context-oriented words.

Preparing and Testing Data

The preparation set is what we are getting ready for, and reliability with our model mostly compares to the boundaries. At the same time, the test information is utilized uniquely for model execution evaluation. Preparing information yield accessible on the model while the testing information is inconspicuous information for which should finish forecasts. The K-overlap cross-approval is utilized multiple times for designing and testing datasets. To answer this issue, we use K-overlap Cross-Validation to separate the

information into folds and guarantee that each crease is utilized as a testing set sooner or later. Train and test the removed data, and play out an SVM, KNN, and Random timberland calculation for the expectation of snide or not and ascertain the exactness of every measure. This technique will prepare around 70% of the given informational index, and the leftover 30% will be utilized for testing purposes.

Order

We ran the order utilizing the classifiers like Support Vector Machine (SVM), k Nearest Neighbours (KNN), and Random woodland. The outcome segment presents the presentation of classifiers on the dataset.

III. CONCLUSION

This work proposes a framework that recognizes mockery in English just as on Hindi tweets on Twitter. Sarcasm is extremely reliant and profoundly logical; in this way, opinion and other context-oriented signs assist with recognizing the mockery text. The framework utilizes snide tweets, 9,104 tweets containing #sarcasm, and #not a dataset. The framework uses the SVM, KNN, and Random woods classifier. The methodology has shown great outcomes. All examples of wry do not cover in we extricated designs. So in future, we can consolidate neural networks, Genetic Algorithm and Pattern-based methods for more precision.

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 Microblogs Using 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 preprint arXiv: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.