International Journal of Current Engineering and Technology E-ISSN 2277 – 4106, P-ISSN 2347 – 5161

Available at http://inpressco.com/category/ijcet

Research Article

Sarcasm detection on twitter using machine learning techniques

Tejal Patil Prof. Dhanashri S. kulkarni

Department Of Computer Engineering Dr.D.Y.Patil Institute of Technology Pimpri,Pune 41118,India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Sarcasm could be a subtle kind of irony wide utilized in social networks and small blogging websites. it's sometimes accustomed convey implicit info at intervals the message someone transmits. satire can be used for various functions, like criticism or mockery. Twitter became one in all the most important internet destinations for folks to precise their opinions, share their thoughts and report period of time events.

Keywords: *Twitter, sentiment analysis, sarcasm detection, machine learning.*

Introduction

There square measure completely different trends gap within the era of sentiment analysis, that analyze 'attitude and opinion folks in social media, that together with social sites like Facebook, Twitter, blogs, etc. the most aim of sentiment analysis is to spot the polarity (positive, negative or neutral) in a very given text. wit could be a special kind of sentiment that have the flexibility to flip the polarity of the given text. wit is outlined as 'the use of irony to mock or convey contempt'. wit could be a subtle type of sentiment specificion wherever speaker express their opinions opposite of what they mean. wit could be a distinction between positive sentiment word and a negative state of affairs.To notice mordant tweet. though it doesn't would like associate already-built user content as within the work of Rajadesingan et al. our approach considers the various kinds of wit anddetect the mordant tweets notwithstanding their house owners or their temporal context, with a exactness that reaches ninety one.1%. wit as a type of irony that's supposed to specific contempt. Since most of the main target on wit is to reinforce and the prevailing automatic sentiment analysis systems, we tend to conjointly use the 2 terms synonymously.

Literature Survey

In [1-3] Sarcasm is a sophisticated form of irony widely used in social networks and microblogging web sites. It is commonly used to convey implicit facts inside the message someone transmits. Sarcasm might be used for exceptional purposes, such as criticism or mockery. However, it is difficult even for human beings to recognize. Therefore, recognizing sarcastic statements can be very useful to improve computerized sentiment analysis of records gathered from microblogging web sites or social networks. Sentiment Analysis refers back to the identification and aggregation of attitudes and critiques expressed with the aid of Internet users closer to a specific topic. Automatic sarcasm detection is the challenge of predicting sarcasm in text. This is a important step to sentiment evaluation, considering prevalence and challenges of sarcasm in sentiment-bearing text. Beginning with an approach that used speech-based features , automatic sarcasim detection has witnesse dgreatint from these sentiment evaluation community. Sarcasm is described as witty language used to convey insults or scorn. It is applied for remarks that obviously mean the other individuals need to state, made preserving in mind the end intention to offend someone or to reprimand some thing hilariously. While speaking, it is very easy to distinguish sarcasm making use of pitch of voice, gesture, facial features etc. But in textual information, it's miles tough to detect sarcasm due to lack of defined factors. Sentimental evaluation is used to know someone's opinion, attitude toward specific event, agency etc. Sarcasm is one type of person's sentiment but used for taunting, insulting, to make a laugh of someone. Various algorithms are proposed to stumble on sarcasm based on one-of-a-kind features, domains and type of sarcasm. We used a Hadoop based framework that applied live tweets, system it and use hybrid set of rules which identifies sarcastic sentiment efficiently. Hybrid approach consider lexical and hyperbole function to improve performance of gadget by means of increasing accuracy, precision, F-score.

From [4-7], In current times, sarcasm evaluation has been one of the toughest demanding situations in Natural Language Processing (NLP). The property of sarcasm that makes it tough to research and detect is the gap between its literal and intended meaning.

Detecting sarcastic sentiment within the domain of social media which includes Facebook, Twitter, on line blogs, reviews, etc. Has come to be an vital project as they influence every enterprise organization. In this article, a hyperbolic feature primarily based sarcasm detector for Twitter records is proposed. The hyperbolic functions include intensifiers and interjections of the textual content. The performance of the proposed device is analyzed the usage of numerous standard device getting to know strategies namely, Naive Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), and AdaBoost. The gadget attains an accuracy (%) of 75.12, 80.27, 80.67, 80.79, and 80.07 using NB, DT, SVM, RF, and AdaBoost respectively.

Sarcasm is considered one of the most hard trouble in sentiment evaluation. In our remark on Indonesian social media, for certain topics, people have a tendency to criticize something the use of sarcasm. Here, we proposed two additional functions to detect sarcasm after a not unusual sentiment evaluation is carried out. The capabilities are the negativity information and the wide variety of interjection words. We also employed translated SentiWordNet inside the sentiment classification. All the classifications were performed with device getting to know algorithms. The experimental results showed that the additional capabilities are quite effective within the sarcasm detection.

Sentiment Analysis has come to be a vast studies count for its in all likelihood in tapping into the sizable amount of opinions generated by way of the people. Sentiment evaluation offers with the computational behavior of opinion, sentiment in the textual content. People on occasion uses sarcastic text to specific their opinion within the textual content .Sarcasm is a type of communique act wherein the people write the contradictory of what they suggest in reality. The intrinsically vague nature of sarcasm every so often makes it difficult to understand. Recognizing sarcasm can sell many sentiment analysis applications. Automatic detecting sarcasm is an approach for predicting sarcasm in textual content. In this paper we've tried to speak of the beyond work that has been accomplished for detecting sarcasm in the textual content. This paper talk of methods, features, datasets, and issues related to sarcasm detection. Performance values associated with the beyond work also has been discussed. Various tables that present specific size of beyond paintings like dataset used, capabilities, techniques, overall performance values has additionally been discussed. Sarcasm is a sophisticated shape of sentiment expression where speaker explicit their evaluations opposite of what they mean. Sarcasm detection and Emotion detection from social netrunning web sites has been a superb area of study. With the growth of e-services consisting of e-commerce, e-tourism and ebusiness, the groups are very eager on exploiting emotion and sarcasm evaluation for his or her marketing strategies as a way to evaluate the public attitudes in the direction of their brand. Thus green emotion and sarcasm modeling machine may be a good way to the above hassle.

Proposed Methodology

A method to identify the emotional pattern and the word pattern in Twitter data to determine the changes in public opinion over the time. to classify tweets based on the sentiment polarity of the users towards specific topics.

A. Architecture

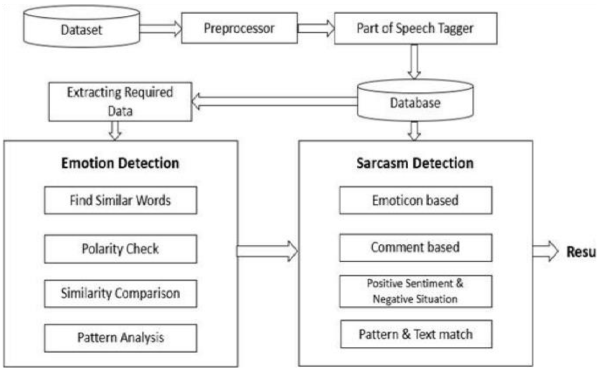

Fig.1. Architecture of Proposed System

Features of Proposed System To identify emotional pattern. To identify their emotions

B. Algorithms Support Vector Machine (SVM) Support Vector Machine is belongs to supervised machine learning formula which used for each classification or regression challenges. However, it's principally utilized in classification issues. during this formula, we tend to plot every knowledge item as a degree in n-dimensional area (where n is range of options you have) with the worth of every feature being the worth of a selected coordinate. Then, we tend to perform classification by finding the hyperplane that differentiate the 2 categories all right (look at the below snapshot)

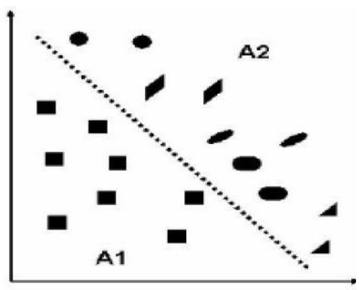

Fig.2. Devidetion of Labeled data and train data using hyperplane

The basic principle behind the operating of Support vector machines is easy produce a hyper plane that separates the dataset into categories. allow us to begin with a sample drawback. Suppose that for a given dataset, you have got to classify red triangles from blue circles. Your goal is to form a line that classifies the info into 2 categories, making a distinction between red triangles and blue circles. whereas one will theorize a transparent line that separates the 2 categories, there is several lines that may try this job. Therefore, there's not one line that you just will agree on which might perform this task. The principle of SVM depends on a linear separation in a very high dimension feature area wherever knowledge square measure mapped to think about the ultimate nonlinearity of the matter. to urge a decent level of generalization capability, the margin between the apparatus hyperplane and therefore the knowledge is maximized. A Support Vector Machine classifier is trained with matching score vectors. Hyper-plane may be a plane that linearly divides the n-dimensional knowledge points in 2 part. just in case of second, hyperplane is line, just in case of 3D it's plane.It is conjointly known as as n-dimensional line.

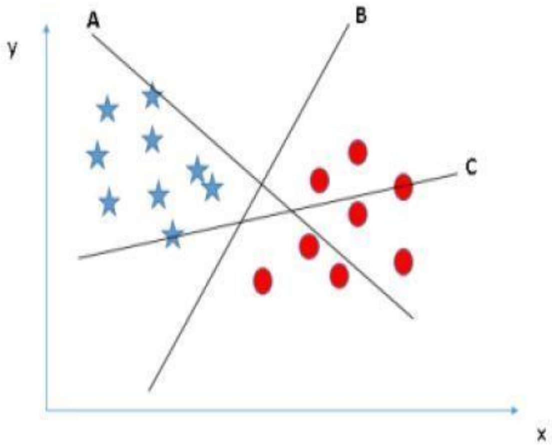

Fig.3. Ploting label data and train data

The goal of the help vector system algorithm is to discover ahyperplane in an N-dimensional space(N — the range of capabilities) that pretty classifies the statistics points.

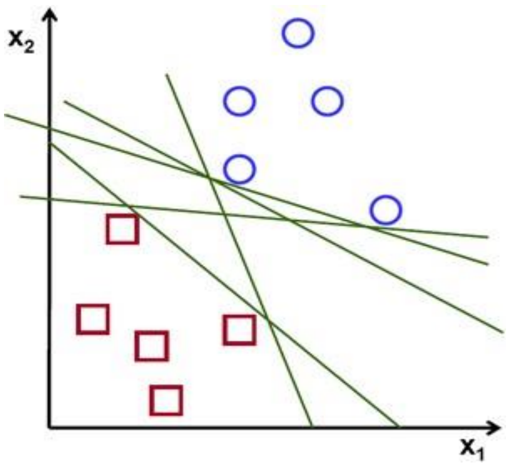

Fig.4. No. of dimensional space

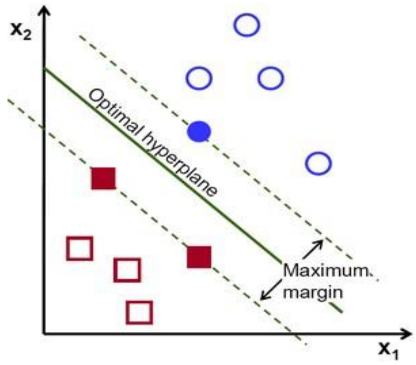

Fig.5. Optimal Hyperplane

To separate the two classes of facts points, there are many viable hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.E the most distance between statistics factors of both training. Maximizing the margin distance provides some reinforcement so that future statistics factors may be categorized with extra confidence.

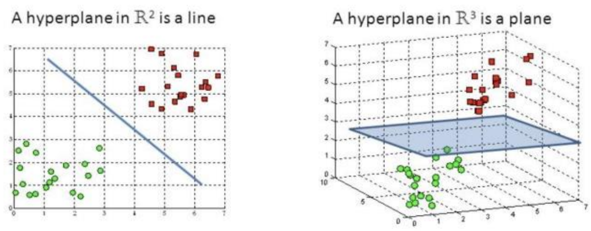

Fig.6. Statistics Factors of Margin

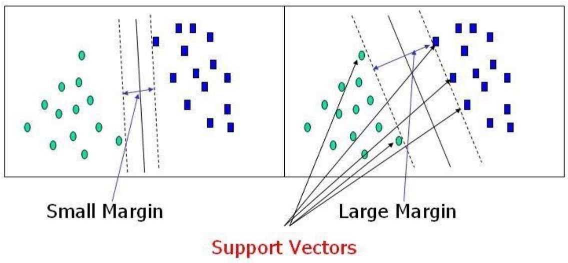

Fig.7. Difference Between Small and Large Margin

Hyperplanes are selection limitations that assist classify the records points. Data factors falling on either aspect of the hyperplane may be attributed to special instructions. Also, the dimension of the hyperplane relies upon upon the range of capabilities. If the range of input functions is 2, then the hyperplane is only a line. If the range of input features is 3, then the hyperplane will become a two-dimensional plane. It will become hard to assume while the wide variety of features exceeds 3. Support vectors are statistics factors which might be in the direction of the hyperplane and influence the position and orientation of the hyperplane. Using these help vectors, we maximize the margin of the classifier. Deleting the help vectors will trade the position of the hyperplane. These are the points that assist us build our SVM. In logistic regression, we take the output of the linear feature and squash the price inside the variety of [0,1] the use of the sigmoid characteristic. If the squashed price is more than a threshold cost(0.5) we assign it a label 1, else we assign it a label 0. In SVM, we take the output of the linear feature and if that output is greater than 1, we discover it with one class and if the output is - 1, we pick out is with any other class. Since the threshold values are modified to one and -1 in SVM, we reap this reinforcement range of values([-1,1]) which acts as margin.

Math Model:

"Algorithm: WSVE (S, D, k, y, V, μ, o)
Input: S = {(xi, Yi)}"=1, D, k, Y, V, μ, o
Output: h(:) begin
Select J such that E;e, D(j) < o, and it has minimum cardinality.
Set S* = {(xj, Y;)}J.
Set D* = Dj/EDj. h(:)
MWSV (S*, D*, k, y, v, μ). end
Output the hypothesis h(-)

Result and Discussions

Fig.8. Welcome Page

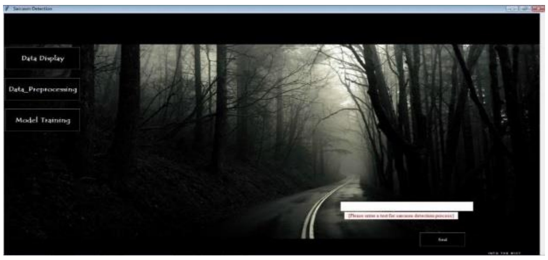

Fig.9. Welcome to prediction window

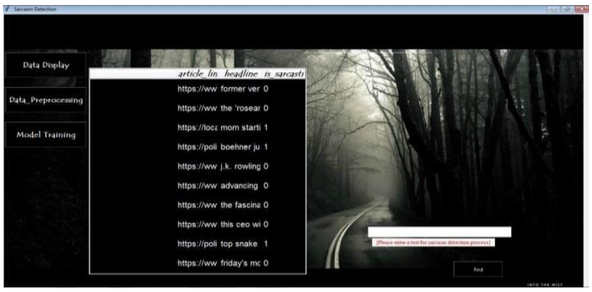

Fig.10. Display data for pre-processing

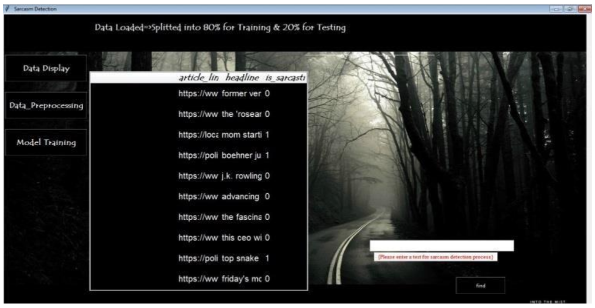

Fig.11. After pre-processing devide the data into 80% for training and 20% for testing

Fig.12. Model Training

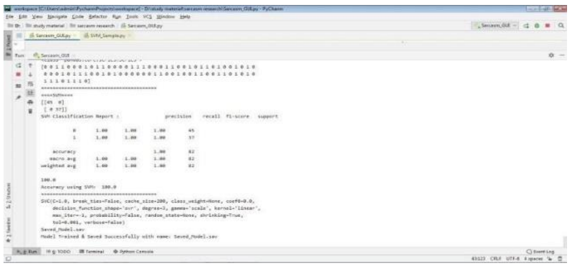

Fig.13. After Model Training shows Accuracy

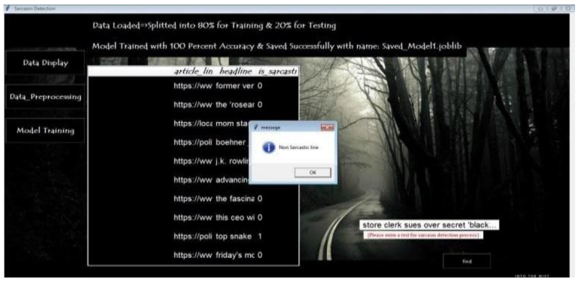

Fig.14. Prediction of Sarcasm Detection

Conclusions

Our paper uses combined approach of vario strategies like feeling detection, use of emoticons, patterns, etc. identifies the social website comment is sarcastic or not. thus it's needed to use combined approach that take completely different strategies and determine the comment is sarcastic or not. The witticism identification model could be a novel approach supported feeling model. The witticism identification model uses completely different algorithms, libraries And strategies in feeling detection section and its result's used for witticism detection.The planned technique makes use of the various elements of the tweet. Our approach makes use of Part-of-Speech tags to extract patterns characterizing the amount of witticism of tweets.

References

[1]. Automatically Identifying Fake News in Popular Twitter Threads by Cody Buntain Available: http://ieee xplore.ieee.org /abstract/ document/7100738/

[2]. S. Homoceanu, M. Loster, C. Lo_, and W.-T. Balke, ``Will I like it? Providing product overviews based on opinion excerpts,'' in Proc. IEEE CEC, Sep. 2011, pp. 26_33.

[3]. U. R. Hodeghatta, ``Sentiment analysis of Hollywood movies on Twitter,'' in Proc. IEEE/ACM ASONAM, Aug. 2013, pp. 1401_1404.

[4]. R. L. J. Brown, ``The pragmatics of verbal irony,'' in Language Use and the Uses of Language. Washington, DC, USA: Georgetown University Press, 1980, pp. 111_127.

[5]. S. Attardo, ``Irony as relevant inappropriateness,'' in Irony in Language and Thought. New York, NY, USA: Psychology Press, Jun. 2007, pp. 135_ 174.

[6]. R. W. Gibbs, Jr., and J. O'Brien, ``Psychological aspects of irony under- standing,'' J. Pragmatics, vol. 16, no. 6, pp. 523_530, Dec. 1991.

[7]. H. Grice, ``Further notes on logic and conversation,'' in Pragmatics: Syntax and Semantics. San Diego, CA, USA: Academic, 1978, pp. 113_127.

[8]. O. Tsur, D. Davidov, and A. Rappoport, ``ICWSM_A great catchy name: Semi-supervised recognition of sarcastic sentences in online prod- uct reviews,'' in Proc. AAAI Conf. Weblogs Soc. Media, May 2010,pp. 162_169.

[9]. Bharti, S. K., Vachha, B., Pradhan, R. K., Babu, K. S., & Jena, S. K. "Sarcastic sentiment detection in tweets streamed in real time: A big data approach." Digital Communications and Networks, 2(3), pp.108-121

[10]. Pang, Bo., Lee, L., and Vaithyanathan, S.. 2002. Thumb's up? Sentiment Classification using machine learning Techniques. Proceeding of the 7th Conference on Empirical Methods in Natural Language Processing (EMNLP-02). Stroudsburg, PA,USA.

[11]. Kim S. B., Rim H. C., Yook D. S. and Lim H. S., "Effective Methods for Improving Naive Bayes Text Classifiers", LNAI 2417, 2002, pp. 414-423

[12] Berger A. L., Pietra V. J. D., and Della S. A. P., "A Maximum Entropy Approach to Natural Language Processing", ACL

[13] Joachims, T. 1998. "Text Categorization with Support Vector Machine: Learning with Many Relevant Features". European Centre for Modern Languages.

[14] A. Rajadesingan, R. Zafarani, and H. Liu, "Sarcasm detection on twitter: A behavioral modeling approach," in Proceedings of the 8th Association for Computing Machinery International Conference on Web Search and Data Mining (WSDM), Shanghai, China, pp. 97–106, ACM, February 2-6, 2015.

[15] M. Bouazizi and T. Ohtsuki, "A pattern-based approach for sarcasm detection on twitter," IEEE Access, vol. 4, pp. 5477– 5488, September 2016.

[16] R. K. Gupta and Y. Yang, "Crystalnest at semeval-2017 task 4: Using sarcasm detection for enhancing sentiment classification and quantification," in Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval), Vancouver, Canada, pp. 626–633, ACL, August 3–4, 2017.

[17] D. Davidov, O. Tsur, and A. Rappoport, "Semi-supervised recognition of sarcastic sentences in twitter and amazon," in Proceedings of the 14th Conference on Computational Natural Language Learning (CONLL), Uppsala, Sweden, pp. 107–116, ACL, July 15-16, 2010.

[18] O. Tsur, D. Davidov, and A. Rappoport, "Icwsm-a great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews," in Proceedings of the 4th International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media (ICWSM), Washington, DC, USA, pp. 162–169, AAAI, May 23-26, 2010.

[19] S. Lukin and M. Walker, "Really? well. apparently bootstrapping improves the performance of sarcasm and nastiness classifiers for online dialogue," in Proceedings of the Workshop on Language in Social Media (LASM), Atlanta, Georgia, pp. 30–40, ACL, June 13,2013.

[20] E. Riloff, A. Qadir, P. Surve, L. D. Silva, N. Gilbert, and R. Huang, "Sarcasm as contrast between a positive sentiment and negative situation," in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Seattle, Washington, USA, pp. 704–714, ACL, October 18-21, 2013.

[21] S. K. Bharti, K. S. Babu, and S. K. J. Jena, "Parsing-based sarcasm sentiment recognition in twitter data," in Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Paris, France, pp. 1373–1380, IEEE, August 25–28, 2015.

[22] S. Amir, B. C. Wallace, H. Lyu, P. Carvalho, and M. J. Silva,"Modelling context with user embeddings for sarcasm detection in social media," in Proceedings of the 20th Special Interest Group on Natural Language Learning Conference on Computational Natural Language Learning (CoNLL), Berlin, Germany, pp. 167–177, 2016.