

TWEET SARCASM: TWITTER

Keerthikha.M.S¹,S.Divya Lakshmi²

¹Assistant Professor, ²Assistant Professor, BIST,BIHER,Bharath University

¹keerthikha@bharathuniv.ac.in, ²divya@bharathuniv.ac.in

Abstract: The www is developing at an alarming price. customers have commenced collaborating actively on internet with the aid of giving their reviews on products, services and blogs. observe and evaluation of such reviews is known as Opinion Mining. however now and again users choose being sarcastic. Sarcasm is a linguistic phenomenon in which humans nation the other ofwhat they inreality imply. Sarcasm Detection is a difficult undertaking, even for people. it's far part of Opinion Mining that's studied so that specific sentiments can be analyzed and worked upon. on this paper, the overall mechanism of sarcasm detection is visible, that's rooted at the facts from Twitter, a famous micro blogging service.

Key words: Tweets, sarcasm detection.

1. Introduction

Textual information may be divided into categories, facts and critiques. information are objective statements whereas reviews are subjective statements. information nation the activities that had been occurred in the global. reviews indicate the exceptional sentiments, perceptions, observations or views approximately those events. What others assume has always been an essential and thrilling statistics for most people in decision-making manner. Opinion Mining in any commercial enterprise or corporation can be idea of as when someone desires to buy a telephone appears for remarks and critiques

someone who just bought a smartphone remarks on it
Writes approximately their enjoy

A smartphone manufacturer receives feedback from client

improve their products adjust advertising and marketing techniqueswhen it comes to sentiments or emotions

nobody is involved about the subject of the text but makes a speciality of its effective or terrible expressions. people can without problems explicit their critiques on social media services along with reviews, blogs, social networking sites as they offer a massive quantity of precious information. Now-a-days automated identity of sentiments is completed that's useful for lots NLP systems like evaluation summarization systems (SMO), speak structures and public media evaluation structures. Majorly the present sentiment extraction structures are based totally on polarity identity (e.g., advantageous vs. poor critiques){1-6}, however there are many beneficial and comparatively unexplored sentiment sorts which include sarcasm, irony or humor. on this paper, the sentiment sarcasm has been explored and its detection has been executed on Twitter{7-9}, as a platform. With the current trend of tagging posts the usage of HASHTAGS, a few social media services like Twitter permit customers to add exceptional hashtags to articles/tweets. for this reason blogs are used as a large dataset for sentiment getting to know and identity. on this paper, unique Twitter tags are used as sentiment labels. distinct punctuations, phrases, patterns inside the textual content are found for detecting sarcasm.

Associated paintings:

Sarcasm and irony are properly-studied and emerging standards in linguistics, psychology and cognitive technological know-how[1]. but within the opinion mining literature, among a lot of these principles, automation of sarcasm detection is tested as a hard hassle and has been approached in only a few studies[2]. Sentiment analysis obligations consists of foremost steps- (1) seeking out exclusive expressions, and (2) figuring out the polarity (terrible, nice or impartial) of the expressed sentiment. these steps are commonly achieved to check whether a

sentence conveys nice that means or bad. but in this paper, sarcastic and non-sarcastic tweets are prominent to discover the polarity of a sentence.

It has been proposed that sentiment phrases or phrases may have unique senses for this reason phrase sense disambiguation can improve analysis of sentiments[3]. All cited work.

Identifies evaluative sentiment expressions and their polarity. But, it's been mentioned that during many cases, definitely a sentence can not be judged as sarcastic or non-sarcastic without the encompassing text or content material [10-15]. for instance, the sentence "wherein am I?" may be assumed as sarcastic handiest if it is recognised that it's miles said in a evaluate of a GPS device. additionally, in some instances, only analyzing few sentences together can display the presence of sarcasm.

Information:

To form an algorithm for detecting sarcasm, first we want to educate that set of rules and we require facts for that. category is a directed studying activity, this means that, for the classifier to understand the distinction between distinct sentences, a few sentences labeled as sarcastic and others classified as non-sarcastic are needed. it may be performed by the usage of a web corpus which includes various sarcastic sentences [16], as an instance reviews, remarks, posts and so on and labelling is accomplished. but that is very monotonous workout in case of big information set. another option is to utilize the Twitter API to membership tweets with the label #sarcasm or #sarcastic, those will be the sarcastic tweets, and others that do not have such label, turns into non-sarcastic tweets.

HashtagsIn Twitter:

A message may be of approximately 140 characters .except the regular textual content, a twitter message can contain references to other users (@<person>), hashtags (#hashtag) and URLs . as an example, : "@persona check out @personB for super ideas :) http://xxxxxx.com #satisfied #hour"[4]. So for building the corpus of sarcastic (S), poor (N) and positive(P) tweets, the annotations that tweeters assign to their tweets the use of hashtags are used. Twitter API is used to accumulate tweets that encompass hashtags of

sarcasm (#sarcastic, #sarcasm), direct high quality sentiment (e.g., #satisfied, #joy, #fortunate,

#great, #exciting), and direct negative sentiment (e.g., #unhappiness, #indignant, #frustrated, #horrific, #fail) [5]. also, automated filtering is carried out to cast off prices, spam, duplicates, tweets written in languages other than English. The advantage of the use of Twitter API is that we are able to have enough samples to meet our requirement. each day people write tweets, use sarcasm, that may be without difficulty accumulated, clubbed and saved in a database. however there may be a disadvantage in accumulating information from Twitter, that is, the facts is little noisy! humans additionally use the #sarcasm hashtag to reveal that the tweet is sarcastic, however a Human can't absolutely bet or assume that the tweet is sarcastic with out the label #sarcasm . So for this we want to pre-process the facts i.e cleaning up the data.

For doing this, all the tweets which include Non-ASCII characters, link to other tweets and non-sarcastic behaviour, are eliminated. After that each one the hashtags and all occurrences of the word sarcasm or sarcastic are eliminated from the closing tweets. And still if the tweet is atleast three phrases long, it is delivered to the dataset [6]. The above is done to put off all the noise.

2. Characteristic Engineering:

exceptional steps had been followed for doing feature extractions.

N-grams:

It is able to be bigrams and unigrams. these are organization of single word (example: seriously, awesome, high-quality, and so forth.) and double phrases (example: surely awful, wonderful splendid, very good, and so on.). To extract them from the last textual content, every tweet is surpassed through tokenization, stemming, uncapitalization after which every and every n-gram is delivered to a binary function dictionary [7]. Tokenization is a challenge of cutting via the individual sequence into bits and portions, known as tokens. Eg- "this is the antique record", w(sa) = "that", "is", "the", "vintage", "record" [8]. Stemming is utilized in linguistic morphology and retrieving information to describe the method of lowering inflected phrases to their word stem, base form or root. Eg: phrases fishing, fisher, fished has the equal root word 'Fish' [9].

Sentiments:

Sarcastic tweets are determined to be greater poor than non-sarcastic tweets. additionally, there may be a huge style of sentiments in tweets which are sarcastic. It starts off evolved with a pretty fine sentiment and ends with a bad sentiment (example: I experience getting slapped #sarcasm)[10]. So for this a sentence is broken into elements and sentiment analyzers are used on each component one after the other. Many studies work has been done on these analyzers. there may be an analyzer made called, SentiWordNet dictionary. It gives a fantastic and a poor sentiment price to each single phrase of the English language. with the aid of looking for the phrases within the dictionary, sentiment value may be given to every single part of the tweet. every other deployment of this sentiment analysis can use the python library tool TextBlob which contains a built-in sentiment ranking feature.

Sample Extraction:

For automated extraction of styles, the definitions about styles provided by using Davidov and Rappoport, are used. words are classified into content material words (CWs) and high-frequency words (HFWs). A word having more(much less) corpus frequency or occurrence than FH (FC) is stated to be a HFW (CW)[11]. All single punctuation characters or their consecutive sequences are considered as HFWs. A sample is stated to be an directed series of excessive frequency phrases and some slots for (CWs) content phrases.

Type algorithm:

The investigation at the applicability of pragmatic and lexical features in system mastering is completed to classify distinctive effective, negative and sarcastic Tweets . two widespread classifiers which might be commonly used in sentiment class are: logistic regression (LogR) and guide vector machine with sequential minimum optimization (SMO). In gadget mastering, help vector machines are supervised studying fashions having related gaining knowledge of algorithms (SMO) that may analyze facts and recognize patterns[12]. while, Logistic Regression is a regression wherein binary reaction variable is associated to a set of

explanatory variables, that can be discrete or continuous[13].

3. Evaluation

the primary purpose of evaluating is to learn how well the framework helps in identification and differentiation of sentiments defined through tags and to check if the framework may be effectively used to discover sentiments in new untagged sentences[14]. It can be finished using pass-validation approach. pass-validation, is a technique for assessing the effects of a statistical evaluation. it's miles especially used in settings where the aim is prediction, and one desires to estimate how as it should be a predictive version will perform in exercise[15]. The metric used to give an explanation for the move-validation is the F-rating. it's far a beneficial metric whilst there are greater samples from one precise class than from the other . as an example, if we've 10 times extra non-sarcastic messages or tweets than sarcastic then this kind of metric is considered. On the grounds that precision is wanted so some other metric is required. Precision defines the variety of rightly diagnosed sarcastic tweets upon the whole quantity of tweets that are sarcastic, while don't forget offers the quantity of sarcastic tweets rightly diagnosed upon the full wide variety of sarcastic tweets inside the go validation set. both don't forget and precision are higher ratings or ranking techniques to quantify the high-quality of an irony classifier. The F-rating provides the harmonic imply of precision and keep in mind[16].

To advantage an insight into what the algorithm has obtained, different feature coefficients inside the skilled SVM are noticed to look the maximum vital ones. consistent with a survey, for n-grams, the features that are critical to classify them into sarcastic tweets are how, what, best vicinity, a blast, shocking , simply eat, and so on. The maximum applicable n-grams for the type of non-sarcastic tweets are feeling extremely good, spend extra, praying, too humorous, be smart, goodnight, and so forth[17]. ultimately, it become discovered by using the classifier that sarcastic tweets are greater about expressing emotions and emotions, either positive or terrible, than non-sarcastic ones.

4. Conclusion

Sarcasm detection is a really captivating issue. It evaluates diverse feature sorts for sentiment extraction including sentiments, phrases, styles and n-grams, confirming that every characteristic type contributes to the sentiment type framework. As we have seen that it's miles viable to do sarcasm detection using NLP gear, one short and smooth way to improve this detector is to apply a spell corrector at the side of, for the tweets. this would assist in minimizing the order of dimensions of the dictionary for the n-gram features and could enhance the sentiment analysis operation as well. In the future, these methods can be carried out for automated clustering of sentiment types and sentiment dependency regulations and may be improved to hit upon some different non-literal shape of sentiments like humor.

References

- [1] Udayakumar R., Kaliyamurthie K.P., Khanaa, Thooyamani K.P., Data mining a boon: Predictive system for university topper women in academia, *World Applied Sciences Journal*, v-29, i-14, pp-86-90, 2014.
- [2] Kaliyamurthie K.P., Parameswari D., Udayakumar R., QOS aware privacy preserving location monitoring in wireless sensor network, *Indian Journal of Science and Technology*, v-6, i-SUPPL5, pp-4648-4652, 2013.
- [3] BrinthaRajakumari S., Nalini C., An efficient cost model for data storage with horizontal layout in the cloud, *Indian Journal of Science and Technology*, v-7, i-, pp-45-46, 2014.
- [4] BrinthaRajakumari S., Nalini C., An efficient data mining dataset preparation using aggregation in relational database, *Indian Journal of Science and Technology*, v-7, i-, pp-44-46, 2014.
- [5] Khanna V., Mohanta K., Saravanan T., Recovery of link quality degradation in wireless mesh networks, *Indian Journal of Science and Technology*, v-6, i-SUPPL.6, pp-4837-4843, 2013.
- [6] Khanaa V., Thooyamani K.P., Udayakumar R., A secure and efficient authentication system for distributed wireless sensor network, *World Applied Sciences Journal*, v-29, i-14, pp-304-308, 2014.
- [7] Udayakumar R., Khanaa V., Saravanan T., Saritha G., Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction, *Middle - East Journal of Scientific Research*, v-16, i-12, pp-1781-1785, 2013.
- [8] Khanaa V., Mohanta K., Saravanan. T., Performance analysis of FTTH using GEAPON in direct and external modulation, *Indian Journal of Science and Technology*, v-6, i-SUPPL.6, pp-4848-4852, 2013.
- [9] Kaliyamurthie K.P., Udayakumar R., Parameswari D., Mugunthan S.N., Highly secured online voting system over network, *Indian Journal of Science and Technology*, v-6, i-SUPPL.6, pp-4831-4836, 2013.
- [10] Thooyamani K.P., Khanaa V., Udayakumar R., Efficiently measuring denial of service attacks using appropriate metrics, *Middle - East Journal of Scientific Research*, v-20, i-12, pp-2464-2470, 2014.
- [11] R.Kalaiprasath, R.Elankavi, Dr.R.Udayakumar, Cloud Information Accountability (Cia) Framework Ensuring Accountability Of Data In Cloud And Security In End To End Process In Cloud Terminology, *International Journal Of Civil Engineering And Technology (Ijciety)* Volume 8, Issue 4, Pp. 376–385, April 2017.
- [12] R.Elankavi, R.Kalaiprasath, Dr.R.Udayakumar, A fast clustering algorithm for high-dimensional data, *International Journal Of Civil Engineering And Technology (Ijciety)*, Volume 8, Issue 5, Pp. 1220–1227, May 2017.
- [13] R. Kalaiprasath, R. Elankavi and Dr. R. Udayakumar. Cloud. Security and Compliance - A Semantic Approach in End to End Security, *International Journal Of Mechanical Engineering And Technology (Ijmet)*, Volume 8, Issue 5, pp-987-994, May 2017.
- [14] Thooyamani K.P., Khanaa V., Udayakumar R., Virtual instrumentation based process of agriculture by automation, *Middle - East Journal of Scientific Research*, v-20, i-12, pp-2604-2612, 2014.
- [15] Udayakumar R., Thooyamani K.P., Khanaa, Random projection based data perturbation using geometric transformation, *World Applied Sciences Journal*, v-29, i-14, pp-19-24, 2014.
- [16] Udayakumar R., Thooyamani K.P., Khanaa, Deploying site-to-site VPN connectivity: MPLS Vs IPSec, *World Applied Sciences Journal*, v-29, i-14, pp-6-10, 2014.

