www.ThePharmaJournal.com
The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.03
TPI 2019; 8(2): 856-859

www.thepharmajournal.com
Received: 05-12-2018
Accepted: 11-01-2019

Surendra Singh Chauhan
Assistant Professor, Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India

Correspondence Author;
Surendra Singh Chauhan
Assistant Professor, Computer Science & Engineering, Lingaya's Vidyapeeth, Faridabad, Haryana, India

Sarcasm detection algorithm: Unraveling ironic expressions with precision

Surendra Singh Chauhan

DOI: https://doi.org/10.22271/tpi.2019.v8.i2n.25418

Abstract
In sarcasm, the public expresses their unfavourable feelings by utilising positive words in the text. Humans find it extremely difficult to recognise. By doing so, we can do fascination with sarcasm detection in tweets and other social media text. In this research, we investigate a novel pattern-based methodology for sarcasm recognition and a behavioural modelling strategy for efficient sarcasm detection through the analysis of tweet content while simultaneously utilising user activity features gained from previous behaviours. The accuracy and efficacy are evaluated using a variety of classifiers, including Random Forest, Support Vector Machine (SVM), k Nearest Neighbors (kNN), and Maximum Entropy.

Keywords: Sarcasm detection algorithm, Unraveling, precision

1. Introduction
1.1 What is Sarcasm? (Some Definition)
One of the expressions for anything pleasant with a bad intention is sarcasm. There are numerous definitions of sarcasm that use different approaches [1]. Several definitions include: Sarcasm is a sophisticated sort of ironic that is frequently utilized in social networks and microblogging platforms, according to Mondher Bouazizi. Sarcasm is seen as humor but now with incredible intellect, according to Francesca's statement in the Cambridge Dictionary [2].

1.2 Sentiment analysis
Sentiments are an author's attitude, beliefs, ideas, or feelings toward a person, object, place, business, or other entity. Sentiment analysis looks for the presenter's or author's viewpoint on a specific subject or the manuscript's overall relative polarity. The perspective could be the observer's assessment or judgment, their emotional condition, or their moving message emanating from behind. Our behaviour is greatly influenced by our opinions. Our opinions and truth-seeking insights are dependent on how other people see the world [3].
The wheel of emotions shown above illustrates the numerous human sentiments and indicates which mood group each one falls under. One of the emotions people use to communicate their opinions about people, places, products, or any circumstance on social media is sarcasm. Finding sarcasm in any speech or text will be useful for retrieving the proper meaning of sentences, user reviews, and any speech or text posted on social media that can persuade people to do a particular action [4].

1.3 What is Sarcasm detection?
Sarcasm detection refers to the process of identifying sarcasm in written or spoken text, including instructions, blogs, opinions, and reviews. In recent years, there has been an increasing focus on studying sarcasm analysis within opinion mining and sentiment analysis, although this research has primarily concentrated on English language data. The ability to detect sarcasm also has the potential to reveal insights into a person's psychology, mood, and even their health condition, which can aid in human-machine interaction and potentially assist in the early diagnosis of brain damage. The constant monitoring of sarcastic remarks or comments may also play a role in preventing suicide attempts among young individuals. (Rankin *et al*., 2010; McDonald, Skye *et al*.). (Joshi *et al*.) [5].

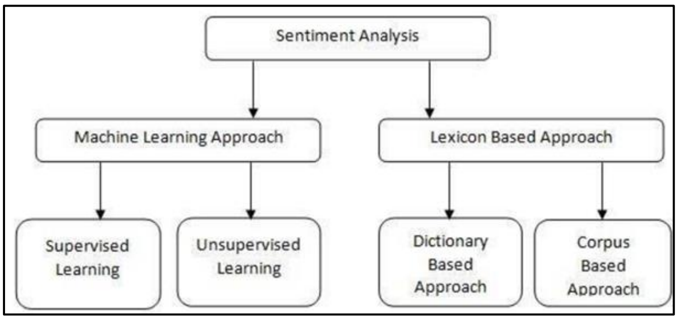


Fig 1: Techniques of Sentiment Analysis

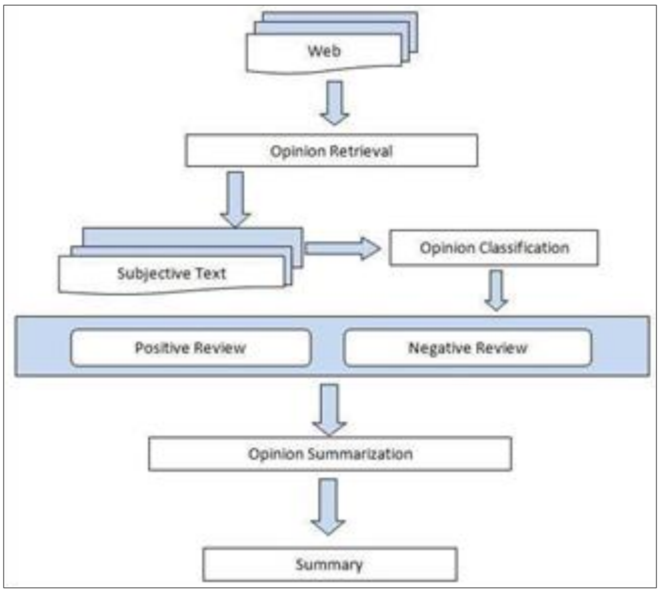


Fig 2: Sentimental Analysis Architecture

Review of literature

This work serves as the foundation for all subsequent research on the analysis and detection of sarcasm. In sentiment analysis, a relationship between positive and negative sentiments is established based on viewpoint. Furthermore, shifts in various sentiments, including happiness, sadness, and rage, are connected to opinion-based sentiment analysis. Opinions influence the sentiment analysis taken into consideration in this work. In this work, the celebrity names were used to test the score. e. With her score, Maria Sharapova was first.

The earlier research on sentiment analysis and opinion mining is discussed, along with the papers that addressed the categorization techniques for sarcasm detection [6].

Opinion mining of online product reviews was done in three steps by Hue and Liu (2004): First, features that customers have commented on are eliminated; second, opinion sentences are found in every review and are either accepted or rejected; and third, a demonstration of the outcome is made. They gathered product reviews from websites like CNN.com and Amazon.com for a variety of products that are sold online, including digital cameras, MP3 players, DVD players, and cell phones. Their primary method is the extraction and aggregation of opinion words, and features are chosen based solely on the opinion words themselves. When compared to opinion sentence extraction for MP3 (93 percent) and DVD (73 percent), they improved efficiency. Five products have an overall accuracy of between 64 and 84 percent [6].

(4) God bole *et al*. (2007) suggested categorization in a Word Net-sourced lexicon. They created distinct lexicons for every subject. Thus, the lexicon used in politics and 25 health are completely different. They created a graph model to extend polarities to additional words based on an initial lexicon [7].

For instance, all synonyms for the word "good" are marked as positive, and all antonyms of the word are marked as negative, if the word "good" is marked as positive. After that, a fresh iteration is finished for the following level by employing antonyms of antonyms and synonyms of synonyms, and so forth. The polarity score is affected by the distance. using the formula 1/cd, in which d is the distance from the node and c > 1. Every word in the system has defined polarities as a result of this kind of formulation. The polarity scores of each text can be found by dividing the total number of words by the sum of the polarity scores of those words, once the scores for each word have been obtained. The sentimental experiment involved creating a new graph for every case and computing the page rank independently for the positive and negative synonym sets. It has also been observed that using positive language greatly improves effectiveness. This implies that categorizing terms that are negative is more difficult [8].

In 2007, Esuli and Sebastiani introduced a very intriguing plan that used the page rank algorithm to ascertain the polarities of terms. They made use of extended Word for this. Create a graph using Net such that each synonym set's polarity is determined by the polarity of its constituent parts. The main hypothesis is that there won't be any significant differences and that the level of negativity in each synonym set will be comparable.This will result in a graph of relationships between various synonym sets, which will impart to its neighbors the polarity properties of each node. This model has an intriguing feature that can be used in other situations involving word semantics [9].

K. Cai and associates. (2008) provided an explanation of sentiment analysis that combined an opinion-based strategy with a classification method. The opinion classification element divided the fragments into positive, negative, and neutral groups based on the comparative sentiment that each term in the fragment expressed. By using word support metrics, the sentiment subject recognition module finds the significance and simplifies the process beyond all sentiment groups [11].

Karamibekr and Ghorbani (2012) first investigated the differences between sentiment analysis of products and social issues mathematically. Subsequently, based on a few findings, they put forth a plan to regard the verb as the most important expression when expressing thoughts about societal issues. Results from experiments and statistics verify that allowing verbs is not only necessary and definite, but also enhances the sentiment analysis concert. They gathered their information from CNN Answers, Yahoo, and Procon.org. Opinion directories and opinion structure are the deciding factors when choosing features. Results for social issues and car models, based on the verb-oriented method, are 65 percent and 62 point five percent, respectively [12].
M. Eirinaki *et al*. (2012) put forth a plan for an opinion search engine. The two opinion mining algorithms were integrated in the suggested methodology. The outlooks are derived from features, and their positions are heavily reliant on the features as a stand-in for the object as a whole. Locals seem to detest a specific object that has multiple characteristics related to the outcome. 26 Their initial experimental evaluation on multiple patron review data sets revealed that their conclusions attained a very high degree of accuracy [13].
K Xu *et al*. (2011) presented a novel graphical model that uses user reviews to extract and visualize the comparative relationships between products. The interdependencies between relationships were taken into consideration in this work to support endeavours in the detection of potential risks. Additionally, fresh goods and advertising plans were created. Extensive experiments on a large dataset of Amazon customer reviews demonstrated that the suggested method outperformed other conventional methods in extracting comparative relations. Additionally, this technique can be used to manage business enterprise risk by analyzing rich user-generated data [14].
Pankaj Gupta *et al*. (2016) found that a number of fields based on sentiment analysis have not yet been thoroughly investigated and that it is critical to apply accurate knowledge in order to enhance earlier methods. It is appropriate to use text summarization to extract only the information that is helpful to users from the vast amount of gathered textual information. Furthermore, an intelligent model for sentiment analysis and data extraction could be made using the 27 machine learning techniques [15].
This study performed a survey on text summarization and reviews analysis. The advantages and disadvantages of the current approaches were examined through this survey [10].
M. Bouazizi and T. Ohtsuki, *et al*. (2019) investigated the use of multiclass classification in the categorization of Twitter users' online posts. Through this research, the benefits and drawbacks of this approach were examined. In this study, a novel model was put forth to represent the various sentiments and show how it helped to comprehend the relationships between sentiments. With the help of this model, multi-class classification's difficulties were resolved and its accuracy increased.

Methodology Used

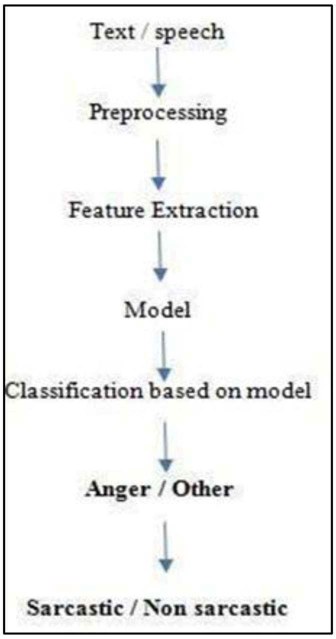
Extraction of micro blogs data and its preprocessing
Various clients use Twitter to express their opinions on a range of topics by posting information in various formats. The Twitter data sample is applied to two categories: the predictive and the analytical. Twitter API is used to generally collect tweeter data. twitter app appearance stands. studying interface. Software developers can now access and interact with publicly available Twitter data thanks to Twitter Tweets. Developers can use one of the public libraries available in several programming languages, or they can write their own scripts.

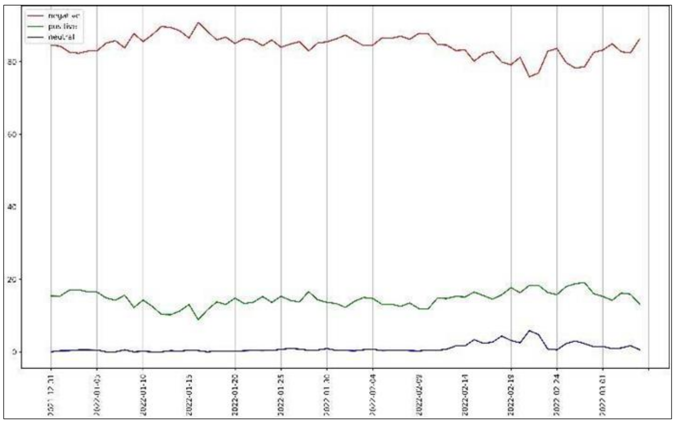
N-grams are used in this work to extract features. Below is a brief explanation of this methodology.
• N-grams: Natural Language Processing (NLP) and text mining applications widely use text N-grams. These are essentially a group of words that fit within a specific window. If variable X is the sentence's word count, then the number of n-grams for a given sentence K would be. One word advances in the n-gram computation.

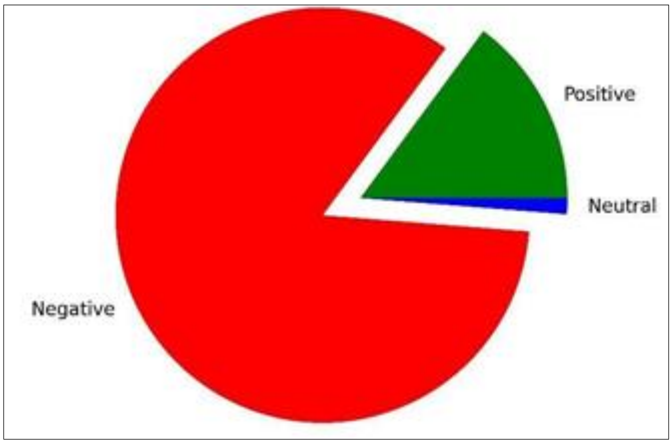
$$Ngrams_K = X - (N - 1)$$

There are many different tasks that can be completed with ngrams. For instance, n-grams are used to create bigram and trigram models in addition to unigram models when creating language models. Web scale n-gram models have been developed by Google and Microsoft. There are numerous uses for these models. These tasks involve spelling correction, word breaking, and text summarization. Feature development for supervised machine learning models, including SVM, MaxEnt models, Naive Bayes, and others, is another application for n-grams. Tokens e are meant to be used, not just for unigrams, but also for bigrams in the feature space [16].

Cope of work & results
More features may be extracted in future work to produce better outcomes. Additionally, the user profile-based weight factor can be used to enable bipolar classification of datasets by the system. Fake reviews can be eliminated from the extracted dataset by adding more functionality to this system. In the wheel of sentiments, sarcasm is categorized as a critical expression under the anger category. We proposed a model that will classify data based on its expression and appropriately classify into sub categories of in order to accurately detect sarcasm in the given speech or text fury. We will annotate each subcategory of anger sentiment to the existing model.







Conclusion

Sentiment analysis is a "suitcase" subject that draws from a wide range of academic fields, including computer science and social sciences like ethics, philosophy, and psychology. The sentiment analysis techniques that have been put forth thus far involve different stages. The dataset's redundant and missing values are eliminated during the pre-processing phase. An association was created between the attribute and the target set by the feature extraction technique. Enforcing the classification method—which can divide data into discrete classes like positive, negative, and neutral—is the final step in the classification process. Although the sentiments of the Twitter data are assessed using the hybrid classification method of the prior method, accuracy and precision can still be enhanced. This study develops a logistic regression, support vector machine, and random forest voting classification technique. The suggested model is put into practice using Python, and the accuracy, precision, and recall of the outcomes are examined. Customer reviews, chatbots, online forums, and even the diagnosis of brain injuries based on a person's writing ability will all benefit from this.

References

1. Wiebe JM, Bruce RF, O'Hara TP. Development and use of a Gold-standard Data Set for Subjectivity Classification. In: Proceeding of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics. USA; 1999. p. 246-253.
2. Pelleg D, Moore A. X-means: Extending Kmeans with Efficient Estimation of the Number of Clusters. In: Proc. of the 17th Int. Conference on Machine Learning. San Francisco, USA; 2000. p. 727-734.
3. Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment Classification using Machine Learning Techniques. In: Conference on Empirical Methods in Natural Language Processing. USA; 2002. p. 79-86.
4. Riloff E, Wiebe J. Learning Extraction Patterns for Subjective Expressions. In: Conference on Empirical Methods in Natural Language Processing. Japan; 2003. p. 105-112.
5. Wilson T, Wiebe J. Annotating opinions in the world Press. In: 4th SIG dial Workshop on Discourse and Dialogue. Sapporo, Japan; 2003. p. 13-22.
6. Dave K, Lawrence S, Pennock DM. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In: Proceedings of WWW. 2003. p. 519–528.
7. Hu, Liu. Mining and Summarizing Customer Reviews. In: International Conference on Knowledge Discovery and Data Mining. Seattle, USA; 2004. p. 168-177.
8. Pang B, Lee L. A sentimental education: Sentimental analysis using Subjectivity Summarization based on Minimum cuts. In: Proceeding of the 42nd Annual Meeting on Association for Computational Linguistics. USA; 2004. p. 271-278.
9. Kim SM, Hovy E. Determining the Sentiment of Opinions. In: Proceedings of the 20th International Conference on Computational Linguistics. USA; 2004. p. 1367-1373.
10. Popescu AM, Etzioni O. Extracting Product Features and Opinions from Reviews. In: Conference on Human Language Technology and Empirical Methods in Natural Language Processing. British Columbia; 2005. p. 339-346.
11. Wilson T, Wiebe J, Hoffmann P. Recognizing Contextual Polarity in Phrase-level sentiment analysis. In: Proceedings of the conference on human language technology and empirical methods in natural language processing. USA; 2005. p. 347-354.
12. Chau M, Xu J. Mining communities and their Relationships in Blogs: A study of online hate groups. International Journal of Human – Computer Studies. 2007;65(1):57-70.
13. Godbole N, Srinivasaiah M, Skiena S. Large-Scale Sentiment Analysis for News and Blogs. In: International Conference on Weblogs and social Media. USA; 2007. p. 21-24.
14. Esuli A, Sebastiani F. PageRanking WordNet Synsets: An Application to Opinion Mining. In: 45th Annual Meeting-Association for Computational linguistics. Prague, Czech Republic; 2007. Vol.45, p. 424-431.
15. Pang B, Lee L. Opinion Mining and Sentimental Analysis. Foundations and Trends in Information Retrieval. 2008;2(1-2):1-135.
16. Kaushik P, Yadav R. Reliability design protocol and block chain locating technique for mobile agent. Journal of Advances in Science and Technology (JAST). 2017;14(1):136-141. doi: 10.29070/JAST.