INVESTIGATING THE IMPORTANCE OF HYPERBOLES TO DETECT SARCASM USING MACHINE LEARNING TECHNIQUES

Vithyatheri Govindan[1] and Vimala Balakrishnan[2]

[1,2] Faculty of Computer Science and Information Technology, University of Malaya

50603 Kuala Lumpur, Malaysia

Email: vithya13@hotmail.com[1], vimala.balakrishnan@um.edu.my[2] (corresponding author)

ABSTRACT

The present study aims to improve sarcasm detection mechanisms using multiple hyperboles such as interjection, intensifiers, capital letters, punctuation, and elongated words. A non-bias dataset consisting of the current pandemic related hashtags was used, namely #Chinesevirus and #Kungflu. Analysis and evaluation were performed with three distinguished machine learning algorithm that is Support Vector Machine, Random Forest and Random Forest with bagging classifiers. Each feature were analysed and the most significant hyperbole identifying sarcasm was assessed further by combining with other hyperboles. The experiments and analysis conducted using these hyperboles concluded that as a single or combined features, hyperboles enhance sarcasm especially in an unbiased dataset.

Keywords: Hyperbole, sarcasm, negative sentiment tweets, machine learning

1.0 INTRODUCTION

Sarcasm detection is an interesting challenge as it is subjective depending on the situation, people, language, or culture. Sarcasm is a combination of positive and negative words in a sentence which may be both funny and nasty [1, 2, 3] or used to criticize or praise [4, 5]. It is sometimes hard to determine if it is to mock or an attempt to convey a sharp, bitter, or cutting expression or remark [6, 7].

A vast majority of studies on sarcasm detection were based on Twitter attributes, particularly the content features such as text and hashtags. Hashtags (#) are often used as part of a text classification as they can convey the exact meaning of the intended sentences [8, 9] however at the same time user's self-declaration by explicitly using hashtags whilst commenting e.g. #sarcasm, does exploit or influence the annotation process [10]. Kunneman, *et al*., [11] and Ptacek, *et al*., [12] extracted data related to sarcasm hashtags (#sarcasm, #sarcastic, #not, #irony) for their research as this helps to identify the pattern in the text or comment whereas [13] did not even remove the sarcasm related hashtags during the annotation for the training set.

Text features are usually identified using lexical, hyperbole and pragmatic approaches, the three different approaches used for sarcasm detection using machine learning techniques [10, 14, 15]. In the lexical approach, words or phrases are analysed to determine their sentiment using word dictionaries [10, 16]. Capital letters, punctuation marks and elongated words are some of the text features extracted for hyperbole approach whereas emoticons and emojis are extracted for the pragmatic approach [2, 13]. Often these approaches are combined with machine learning techniques to perform sentiment analysis and detect sarcasm [1, 17].

Carston & Wearing in [18] and Burgers, *et al*., [19] identified hyperbole as a form of irony and considered it an important feature for sarcasm detection [10]. A various combination of text features is used to perform sarcasm detection, for example, in [14] study is to analyse sarcastic sentiments in real-time tweets with hyperbole and lexical approach scoring 0.87 and 0.82 F-score respectively, whereas [20] who adopted [14] approach for detecting sarcasm in Japanese review on computer games achieved an F-score of 63%. On the other hand, [2] combined all three approaches which resulted in an F-score of 63%. Studies such as [2], [14] and [20] focused on a single hyperbole feature. Sarcasm detection results can be improved by extensive feature engineering [2, 14] and a combination of features [20]. Therefore, the authors aim to investigate the effect of predominant hyperbole features that contributes to sarcasm detection.

The main contributions of this study are:

- The authors proposed a hyperbole-feature based sarcasm detection approach namely interjection, intensifiers, capital letters, punctuation marks and elongated words. The experiment shows that any single hyperbole feature can contribute significantly to sarcasm detection without the interference of other approaches.
- Sarcasm is often associated with negativity, e.g. combination pattern of positive sentiment in a negative situation or vice versa, hence in this study, the authors deliberately used negative sentiments tweets to conduct the analysis and classification for sarcasm detection.
- Current pandemic, COVID-19, has impacted and taken many lives. Hence this dataset provides invaluable insight into the sentiment of the people worldwide. The annotated dataset is used for feature extraction and sarcasm detection using several well-known machine learning algorithms in a non-sarcasm dataset.

It is to note that the terms sarcasm and irony will be used synonymously in this research as sarcasm is a form of verbal irony with contradicting meaning [4]. The remainder of the paper is structured as follows: Section 2 presents our research objectives; section 3 presents the background studies in sarcasm detection using various machine learning techniques and text feature approaches as well as a brief introduction to sarcasm/irony. The proposed model is discussed in Section 4. Results and discussion are presented in Section 5 with the practical implication of this study in Section 6. The conclusion of our research is presented in Section 7, which is the final section in this study.

Hyperbole features are often a strong indication that sarcasm exists. Strong words and texts often express the sentiment of the user's intention. The research objectives for this study are:

- To identify the hyperbole features in negative sentiment tweets, typically in identifying sarcasm.
- To develop a sarcasm detection model based on the hyperbole features identified using machine learning algorithms.
- To evaluate the single approach used in this study with other studies that used a combination of lexical, pragmatic and hyperbole.

2.0 BACKGROUND

2.1 Sarcasm/Irony

Sarcasm is a bitter attitude expressed that can contradict the actual meaning such as positive word may term as negative context or vice versa; 'Fantastic weather' when it rains. Sarcasm can be easily identified from the change of tone, facial expression, and body gestures. However, in text, this is a huge challenge faced by many researchers, specifically those focusing on sarcasm detection. Capital letters, heavy punctuation such as exclamation marks, exaggeration, interjections, usage of laughter expressions and emoticons can indirectly be associated with sarcasm in the text [7, 21]. Examples of the different hyperbole features will be elaborated in the next section.

Knowing the context, situation or topic addressed is vital in sarcasm detection as speakers may implicitly convey their comments [2]. Sarcasm may be inter-related with irony and satire [22] as there are not many distinct differences between the three. An irony is a form of sarcasm [23, 24] in an aggressive way [25] whereas satire is closely related to exaggeration [7]. The texts below show several examples of sarcasm:

Thanks China!
Fuck! See you in 2022 everyone Here s another on the way.
Are you CRAZY OR DRUNK ?!?

Many researchers have focused on different types of sarcasm detection techniques, further elaborated in the following section.

2.2 Hyperbole, Sentiment Analysis, Sarcasm Detection and Machine Learning

Sarcasm detection is a rather challenging task as it can be conveyed in many different forms. Features used for sarcasm detection are mainly extracted from the text to identify the pattern for text classification [10, 13, 26]. In Twitter, content features are the text and hashtag attributes in a tweet [27]. The three main text feature classification related to sarcasm detection are lexical, hyperbole and pragmatic.

2.2.1 Hyperbole

Hyperbole is a form of an extreme exaggeration or overstatement in a speech that could be absurd or funny. Hyperbole and sarcasm/irony are inter-related and often arise as a topic for discourse discussion between researchers. Hyperbole stresses on the intended meaning of words or statements to highlight expectations and reality. Hyperbole is a key marker for detecting sarcasm [3, 18, 19].

The common hyperbole features identified in the dataset used for this study are intensifier, interjection, capital letters, punctuation marks, repeated words, and elongated words. Intensifiers are used to strengthen the meaning of an expression or to show emphasis, commonly used in expressing compliments and apology [28]. Interjections in hyperbole are used to express strong emotions and are commonly used at the beginning of a sentence [29]. Table 1 shows the different types of hyperbole along with sample examples extracted from the Twitter dataset used in this study.

Table 1: Example of common hyperbole features and texts

Hyperbole features	Markers/Identifiers (if any)	Text
Intensifiers	*absolutely*, *simply*	I am sad because your birth was also *completely* avoidable
Interjection	*oops, duh*, *oh*	Uh, let's see. I am guessing if they are getting tested for the that they are being PRESCRIBED the drugs by a licensed physician *DUH*!
Capital Letters		*WHAT DOES VOTING HAVE TO DO WITH PPL DYING AND OUT OF MONEY? SOMEBODY PLEASE TELL ME THAT S STILL A DEMOCRAT! WTF DOES THIS HAVE TO DO WITH THE FAKE NEWS HUSTLERS MAYBE Y ALL CAN ANSWER WHY IT'S IN NANCY'S BILL?*
Punctuation Marks	*!!*, *??*	Democrats are losing*!!!* They trying to negotiate with Americans lives on the line*!!!* Bad people*!!*
Elongated Words		*Ohhhhh Hellllll F ing Nooooo*! The House Spending Bill Would Allow 500 000 Visa Workers China and India To Keep Their U S Jobs During This Pandemic H 1B Lobby House Wuhan Bill Renews Work Permits for Foreign Workers
Repeated Words		*RESPIRATORS RESPIRATORS*?! I havent heard of anyone dying because a respirator was not available Have you?

2.2.2 Sentiment Analysis

Sentiment analysis uses artificial intelligence to identify, analyse and categorize opinions expressed in a piece of text to determine the writer's attitude, excitements, emotions and views towards a topic, product, posts, services, etc., is positive, negative, or neutral. Sentiment analysis is tsed to gauge human expression based on the text that does not have intonation or facial expression [6, 30, 48].

Sentiment analysis helps analyse different entities i.e., words, text, or phrases [7] by detecting their polarity. Riloff, *et al*., [22] focus on one sentiment that is positive sentiment tweets for their bootstrapping algorithm whereas [24] analyses how sarcasm impacts the polarity of tweets. One of the major challenges and essential subtask in sentiment analysis is sarcasm detection [31, 32].

2.2.3 Sarcasm Detection and Machine Learning

A recent study by [44] using Bidirectional Encoder Representations from Transformers (BERT) and Graph Convolutional Networks (GCN) to detect sarcasm content from news article. The model achieved an F-score of 89.6% whereas [45] compiled the dataset characteristics using Hybrid Neural Network framework by introducing Long Short-Term Memory (LSTM) module to encode the word context in every sentence to improve the sarcasm detection task. Similarly [47] applied LSTM in a context-based approach to detect sarcasm for English and Indonesian language. The model achieved an accuracy of 79%.

Kumar, *et al*., [33] developed a multi-head attention-based model for sarcasm detection. The hyperbole features consist of the total number of punctuation marks and interjections words in a tweet. Linguistic Inquiry and Word Count (LIWC) dictionary was used for extracting the semantic features and negative/positive situation/expression were extracted for the sentiment feature part. The model achieved a 77.48% F-score. Sonawane & Kohle [34] adapted a term co-occurrence-based feature selection typically for interjection and punctuation marks to detect sarcasm. The term co-occurrence is based on the frequency relating to positive and negative sentiments.

Jain, *et al*., [35] applied frequency occurrence for punctuation marks to indicate sarcasm exists in a real-time tweet. Lexical and hyperbole approach were combined to detect sarcasm in the English and Hindi mash-up language. Ren, *et al*., [1] developed a two-level memory network model for sarcasm detection. First-level memory is used for sentiment semantics and second-level memory is to capture the contract between sentiment and situation. Their study focused on the positive sentiment and negative situation mainly. Farha & Magdy [36] conducted a comprehensive study of different approaches for Arabic sentiment analysis. They reassessed the dataset to identify the presence of sarcasm which only existed around 16% in their dataset. Frenda, *et al*., [46] proposed a transformer-based system combined with linguistics features to understand the impact of offensive and hateful context in Italian tweets about controversial social issues e.g., politics, immigration, etc. The objective of their study is to analyse the emotion and linguistic markers such as quotation marks, ellipses, intensifiers, syntactic and semantic markers in sarcasm expression, investigate and classify the tweets, detect the impact of these features and finally detect sarcasm.

Kumar & Harish [26] proposed a novel approach using a content-based feature selection method for sarcasm classification based on Amazon product reviews. They used the Chi-square, Information Gain (IG) and Mutual Information (MI) to select relevant features, which were then clustered using k-means. SVM and RF were used for sarcasm detection with SVM and RF yielding 79.6% and 77.2%, respectively. On the other hand, [10] detected sarcasm by combining several approaches including lexicons, positive sentiment in a negative situation, emoticon, and hyperboles with results to be the best when these approaches were used in tandem.

Mukherjee & Bala [37] detected sarcasm based on the authorial style of the tweet users such as content words (i.e., shops, cats, happy), function words (i.e., the, and, she), part of speech tags, part of speech n-grams etc. Approximately 2,000 tweets were used for sarcasm detection with results indicating a 65% accuracy for Naïve Bayes. Meanwhile, [38] investigated personality-based features to exploit sentiment for sarcasm detection. Twitter data with hashtag #sarcasm was used along with Convolutional Neural Network (CNN) that automatically learns from a sarcasm corpus. Two-stage classifications were conducted, one with CNN and then followed by SVM for the final classification. The two-stage classification (CNN-SVM) outperformed CNN for all the datasets used by achieving an F-score of 87% for the balanced dataset, 92.32% for the imbalanced dataset and 93.30% for the training dataset.

Bharti, *et al*., [39] used a Hadoop-based framework to crawl real-time tweets amassing to approximately 1.45 million tweets with keywords such as #sarcasm, #sarcastic, sarcasm, sarcastic, happy, enjoy and sad to name a few. They used MapReduce for sentiment classification and interjection word start (IWS), parsing based lexical generation algorithm (PBLGA) and POS tagging for sarcasm detection with results indicating a high F-score of 97%.

A study by [13] used a pattern-based approach which resulted in an accuracy of 83.1% and a precision of 91.1%. They proposed four sets of features that covers different types of sarcasm identified through their manual annotation, namely, sentiment-related features, punctuation-related features, pattern features, syntactic and semantic features i.e., uncommon words and their meaning to form phrases or sentences. Sarcasm was categorized into three different types, that is, 'sarcasm as whimper', 'sarcasm as wit' and 'sarcasm as evasion'.

Signhaniya, *et al*., [2] used Neural Network (NN) for classification focusing on intensifiers, word features, sentiment, topic models, emoticons, and contextual features. They used a supervised machine learning approach that includes n-grams and features such as punctuation marks and emoticons, followed by NN semantic and syntactic words. Evaluation results showed NN to have an F-score of 47% for a large dataset and 63% for a small dataset. The testing dataset used in this research is highly imbalanced with a ratio of 1:15 (sarcastic versus non-sarcastic) which may have been contributed significantly to the low results attributed in their research.

The work of [24] showed how sarcasm detection can improve sentiment analysis of tweets. The authors extracted individual words by tokenizing the hashtags (e.g., #greatstart becomes #great and #start) using a Viterbi-like algorithm, which was then used for the sentiment analysis. Their results showed a precision of 91% for sarcasm detection. However, the tokenization technique may be inaccurate when splits are done incorrectly, for example, greats and tart (#greatstart) instead of great and start. Another limitation is that unknown entities such as people, location and organization that are used in the hashtag explicitly may affect the sentiment.

Riloff, *et al.*, [22] in his study, identified one type of sarcasm that is the contrast between positive sentiment like love, enjoy, etc with a negative situation that is unpleasant such as visiting a dentist or taking an exam. A bootstrapping algorithm that used a single seed word ‘love’ was used to detect the positive sentiments and negative situations in the sarcasm tweets. This learning process was applied 175,000 tweets consisting of both #sarcasm and non-sarcasm tweets. The bootstrapping algorithm automatically iterates to build the learned list of sentiments and situations which was then used to detect sarcasm in new tweets. A combined bootstrapping method with SVM was conducted which achieved a recall of 44% and f-score of 51% which is a slight improvement when the algorithm was run separately.

Table 2 shows a summary of the different approaches and machine learning techniques applied in sarcasm detection studies.

Table 2: Approaches and techniques applied by past researchers

Reference	Category				Techniques				
	Lexical	Pragmatic	Hyperbole	Others	SVM	RF	NB	NN	Others
Hyperbole-based Sarcasm Detection (HbSD)			/		/	/			/
[22]	/			/	/				
[24]	/			/					/
[2]	/	/	/		/			/	
[13]		/	/		/				
[38]				/	/			/	
[26]	/			/	/	/			/
[39]	/		/						/
[10]	/	/	/						
[37]	/			/			/		
[35]	/		/						/
[33]	/		/		/				/
[34]	/		/						/
[44]				/				/	/
[45]				/				/	
[46]	/		/					/	
[47]	/			/				/	/

One common pattern that can be identified includes the combination of various approaches (i.e., lexical, pragmatic, hyperbole, others) and different machine learning techniques, with the most common being SVM. Furthermore, most of the datasets retrieved from the multiple platforms consist or contain sarcasm related hashtags which help to identify a pattern for sarcasm.

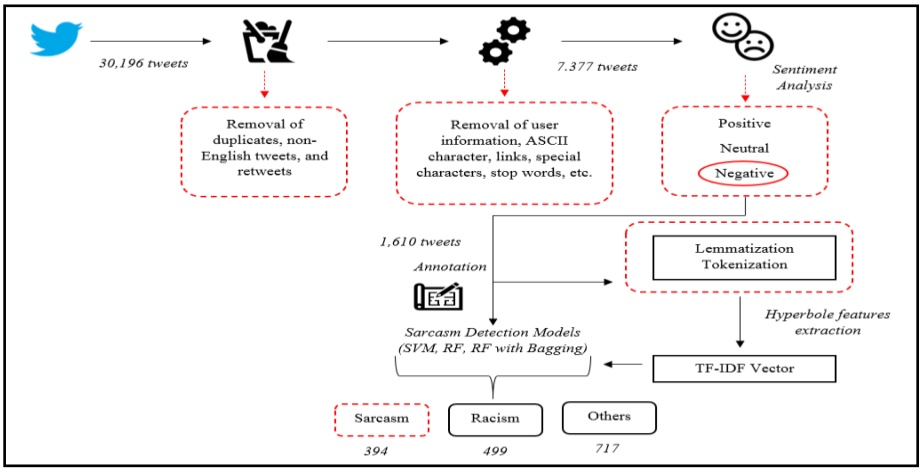
Past studies focus on a combination of approaches. This may impact the results achieved such that one cannot determine which approach is predominant in detecting sarcasm. In this study, hyperbole was particularly selected for sarcasm detection as hyperbole is a key marker and an important feature based on theoretical studies in the past. Hence the experiment for an empirical study to determine if hyperbole alone can significantly be used to identify sarcasm pattern in a text.

3.0 METHODOLOGY

This section will cover the different phases for the completion of this study with an overview of the proposed hyperbole-based sarcasm detection model comprising data acquisition and pre-processing, sentiment analysis, hyperbole feature extraction, sarcasm detection and the evaluation phase.

3.1 Hyperbole-based Sarcasm Detection Model

Fig 1 depicts the overall sarcasm detection model for the proposed hyperbole-based sarcasm detection model.



Note: TF-IDF = Term frequency – inverse document frequency

Fig 1: Sarcasm detection model

3.2 Data Acquisition and Pre-Processing

A total of 30,196 tweets were collected between 23rd March 2020 and 1st April 2020 using Streaming Twitter API. Tweets related to hashtags such as #ChineseVirus and #Kungflu were crawled during the coronavirus pandemic. These hashtags were used as they were the initial names referred to before coronavirus and COVID-19 were introduced. These hashtags were picked specifically as many countries were impacted due to the virus and their citizens being affected. These hashtags can provide valuable information to detect the sentiment among the people and specifically tweets users as countries like Italy went into total lockdown due to the rise of death among their citizens.

The dataset was cleaned and further pre-processed as they contain noise. Several pre-processing tasks were conducted as can be seen in Table 3 below.

Table 3: Pre-processing task

No	Task	Example	Removal
1	Non-English	*Kya, Maan Lia fake hai bhi* (What, is it worth considering)	/
2	URL and User mentions	https://t.co/LUai6iYgRH, @Test123	/

3	Retweets and duplicates	RT @xxxxxxx	/
4	Emoticons and emojis	😊, :P, :(, etc	/
5	Non-ACSII and alphanumeric	\xe2\x81\xbe	/
6	Punctuation marks	%, $, ^, &, *, +, etc.	/
7	Empty spaces		/
8	Hashtags	#chinesevirus	/
9	Lowercase/Uppercase		
10	Empty rows		/
11	Tweets less than three words	Good news	/

A Python script was prepared and ran for the pre-processing tasks namely NumPy and Pandas to store and manipulate the dataset. Tweets remained in their original form for capitalization as this will be used for hyperbole feature extraction. Punctuation marks such as exclamation mark and question marks remained as well. The hyperbole features will be explained in Section 3.5.1. A total of 7,377 tweets remained after completing all pre-processing tasks.

3.3 Sentiment Analysis

The Python library, TextBlob, was used to conduct the sentiment analysis in this study. TextBlob is a simple built-in API in Python library that offers several basic natural language processing features such as sentiment analysis, language detection, etc. The sentiment polarity ranges between -1 and 1 with < 0 being negative, 0 being neutral and > 0 being positive sentiment. Below are some example of tweets and their polarity scores. Majority of the tweets were neutral with only 26% of negative tweets detected from the overall pre-processed dataset.

Good Morning everyone *0.7 (positive)*
Is that a fake smile I see? *-0.1 (negative)*
Impact Of Coronavirus On Sports *0 (neutral)*

3.4 Dataset and Data Annotation

Dataset used for sarcasm detection in most studies are from sarcasm related hashtags, e.g. #sarcasm, #irony, #not, or manually annotated dataset. In a classification task, an annotated dataset is required for machine learning techniques to evaluate the accuracy. Manual annotation is time-consuming and costly, as annotators are required to read, analyse, and understand the content and classify it based on the predefined list of labels. However, studies such as [12] and [13] did not remove the sarcasm hashtags and these were used for annotation instead and labelled accordingly for identifying sarcasm. Three human experts were contacted for annotation who are familiar with the topic to avoid being bias. The experts were required to label the tweets based on three classes:

- Sarcasm/Irony: Tweets that are deliberate with sharp, bitter, or cutting remark to mock or convey attempt.
- Racism: Tweets showing or feeling discrimination or prejudice against people of other races or believing that a race/country is superior to another.
- Others: Tweets that do not fall into either class above.

Racism was included as part of the labelling as this study is part of large on-going research for sarcasm and racism. Furthermore, the dataset used for this study is related to the current pandemic situation that originated from China which then highly relates and contributes to both labels. They were given a week to complete their annotation. A total of 394 (24.50%) of the annotated tweets were labelled sarcastic, 499 (31%) and the remaining 717 (44.5%) were labelled Others.

The dataset employed for this study is relatively small. However, based on the analysis and annotation, the required dataset is sufficient whereby the top five hyperbole features exist in the sarcasm labelled tweets. [20] achieved a precision of 0.79 and recall of 0.56 on 140 reviews to detect sarcastic sentences. [11] and [37] worked on a small dataset as well for sarcasm detection. [11] proposed method detected 309 accurately sarcastic tweets out of 353 labelled tweets whereas [37] divided their dataset into several sets ranging between 650 to 2000 tweets and achieving an F-score ranging between 0.71 to 0.76. Section 4.5.4 will elaborate on the different baseline models used for this study.

3.5 Sarcasm Detection

3.5.1 Feature Extraction

A combination of different hyperboles can be seen in tweets [29] and these features can be extracted and used for sarcasm detection. Punctuations and interjections are the most common feature extractions that were used by many researchers using the hyperbole approach [8, 12, 3]. Based on observation of the annotated tweets, the five common hyperbole features that were extracted are interjection, intensifier, capital letters, elongated words, and punctuation marks. Feature extraction tasks used in this study is a combination of Python script and manual validation.

A Python script for tokenization and lemmatization were prepared for extracting interjection, intensifier, capital letters and elongated words. First, in tokenization, the tweets are broken into a set of individual words. Following that, English stop words such as 'the', 'of', 'a', etc. were removed. The final tweet only consists of lemmatized words. Fig 2 shows the feature extraction process with an example of the pre-processed tweet.

Punctuations marks were removed during the lemmatization process. In this study, the main punctuation marks extracted are exclamation marks and question marks. This feature was extracted adopting [13]. The number of exclamation marks and questions marks was calculated as it can reflect the importance of the mark.

When a feature is found to be matching, a value of one is set against the tweet and zero if none is found. As the dataset is not large, manual verification was done to analyse each annotated tweet for the different hyperbole features. The reasoning behind the manual verification is due to capital letters used in the tweets for entities such as political parties (CCP – Chinese Communist Party, GOP – a nickname for the Republican Party in the United States, NGO - Non-Governmental Organization), World Health Organization (WHO) and countries (USA – United States of America, UK – the United Kingdom, etc.) with no other capital letters exist in the same tweet. This will impact the detection model as these entities are not sarcastic in nature.

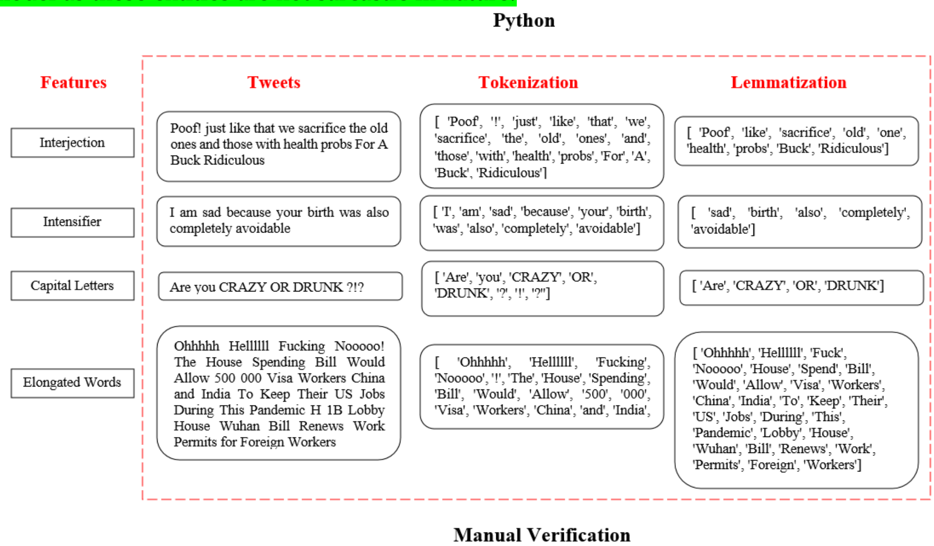


Fig 2: Hyperbole features extraction

3.5.2 Hyperbole-based Sarcasm Detection (HbSD)

As can be seen in Table 2 in section 2.2.3, one of the common machine learning algorithms applied for sarcasm detection was SVM based on past studies. In this study, three distinguished machine learning algorithms namely SVM, RF and RF with Bagging were used to detect sarcasm in the HbSD model.

SVM is a binary linear classifier and is a popular text classification that uses the dataset to find the best hyperplane by separating two classes with a maximum margin. RF is a learning technique that constructs multiple decision trees by splitting the dataset into smaller chunks and using them for creating trees or forest as this can reduce bias due to overfitting and class imbalance whereas RF with bagging in Python uses the bagging classifiers with the base estimator as the random forest classifier [32, 39, Jain et al., 2017].

The models were executed with 70-30 split for training and testing dataset in Python.

3.5.3 Evaluation Metrics

Considering the imbalanced dataset with Others as part of the annotated tweets, two baseline models were built. One model consists of all the three classes (Sarcasm, Racism and Others) and the second model consists of only Sarcasm and Racism class.

The standard evaluation metrics, namely precision, recall, F-score and accuracy were used to assess the effectiveness of the HbSD model. Precision refers to the number of correct classifications that match the human-annotated count whereas recall is the ability of the classifier to correctly classify the labelled sarcastic text against the total datasets. F-score is the harmonic mean of precision and recall. The formula for each standard metric is as below [20, 37]:

$$Precision = \frac{\sum positive\ data\ that\ match\ human\ classified\ count}{total\ positive\ data\ produced\ by\ tool}$$

$$Recall = \frac{total\ positive\ data\ that\ match\ human\ results}{total\ positive\ human\ classified\ data}$$

$$F - measure = \frac{2\ x\ precision\ x\ recall}{precision + recall}$$

The scores range between 0 and 1 whereby the higher score indicate a better classification.

3.5.4 Experiments

This study has two baseline models consisting of the three-class (BM^1) and two-class (BM^2) setup. BM^1 consists of Sarcasm, Racism and Others classified tweets and BM^2 consists of only Sarcasm and Racism tweets. The number of Sarcasm labelled tweets can be considered good in BM^2 with 45% labelled from a total of 893 tweets altogether. In real life, the Others labelled tweet will be common and will be in a higher percentage of a number. However, in this study, this was included as part of the BM^1 model (three-class setup) to draw the difference and impact in detecting sarcasm due to the high number of Others. Each hyperbole features were evaluated individually with both baseline models as well as with a combination of several features and all features.

In total, 28 models were evaluated for BM^1, BM^2 and BM^3. There are seven for BM^1 and 14 models with different combinations were evaluated for BM^2. A further test was conducted with a dataset retrieved from a past study [47] and was named as BM^3 with seven models evaluated using each hyperbole. The models are as below:

- Baseline: BM^1, BM^2 and BM^3
- Baseline + Interjection: BM^1 + $\text{HbSD}^{\text{Interjection}}$ BM^2 + $\text{HbSD}^{\text{Interjection}}$, BM^3 + $\text{HbSD}^{\text{Interjection}}$
- Baseline + Intensifier: BM^1 + $\text{HbSD}^{\text{Intensifier}}$, BM^2 + $\text{HbSD}^{\text{Intensifier}}$, BM^3 + $\text{HbSD}^{\text{Intensifier}}$
- Baseline + Capital Letter: BM^1 + $\text{HbSD}^{\text{CapitalLetter}}$, BM^2 + $\text{HbSD}^{\text{CapitalLetter}}$, BM^3 + $\text{HbSD}^{\text{CapitalLetter}}$
- Baseline + Elongated Word: BM^1 + $\text{HbSD}^{\text{ElongatedWord}}$, BM^2 + $\text{HbSD}^{\text{ElongatedWord}}$, BM^3 + $\text{HbSD}^{\text{ElongatedWord}}$
- Baseline + Punctuation: BM^1 + $\text{HbSD}^{\text{Punctuation}}$, BM^2 + $\text{HbSD}^{\text{Punctuation}}$, BM^3 + $\text{HbSD}^{\text{Punctuation}}$
- Baseline + All (represents all five hyperbole): BM^1 + HbSD^{All}, BM^2 + HbSD^{All}, BM^3 + HbSD^{All}
- Baseline + Interjection + Capital Letter: BM^2 + $\text{HbSD}^{\text{Interjection}}$ + $\text{HbSD}^{\text{CapitalLetter}}$
- Baseline + Interjection + Punctuation: BM^2 + $\text{HbSD}^{\text{Interjection}}$ + $\text{HbSD}^{\text{Punctuation}}$
- Baseline + Interjection + Elongated Word: BM^2 + $\text{HbSD}^{\text{Interjection}}$ + $\text{HbSD}^{\text{ElongatedWord}}$
- Baseline + Intensifier + Capital Letter: BM^2 + $\text{HbSD}^{\text{Intensifier}}$ + $\text{HbSD}^{\text{CapitalLetter}}$
- Baseline + Intensifier + Elongated Word: BM^2 + $\text{HbSD}^{\text{Intensifier}}$ + $\text{HbSD}^{\text{ElongatedWord}}$
- Baseline + Intensifier + Punctuation: BM^2 + $\text{HbSD}^{\text{Intensifier}}$ + $\text{HbSD}^{\text{Punctuation}}$
- Baseline + Intensifier + Interjection: BM^2 + $\text{HbSD}^{\text{Interjection}}$ + $\text{HbSD}^{\text{Intensifier}}$

4.0 RESULTS AND DISCUSSION

In this section, sentiment analysis, results achieved by the three machine learning algorithms for baseline and results with the different hyperbole features will be discussed in detail.

This study focuses only on negative sentiment tweets. A total of 1,610 negative tweets were annotated. Even though, the number of negative sentiment tweets is the least compared to positive (34.80%) and neutral (39.47%), contributing only 25.73% of the total pre-processed tweets, negative sentiment is often explicitly associated with negative feelings, emotion (sad, angry) or behaviour (aggressive) expressed but these are sometimes combined with positive words to convey them [11, 14, 20, 24]. Negative sentiment was selected to analyse how they contribute to sarcasm detection which is evaluated in baseline models.

In Table 5, the baseline model with three-class setup achieved an accuracy of 60.25% for the RF with Bagging algorithm which did not include any hyperbole features. The three-class setup tweets were later tested with each hyperbole features individually as well as with all hyperbole features. An improvement can be seen for SVM when interjection was introduced to the baseline model with an improvement of 1.45%. This is similar to studies by [10], [11] and [14] where words associated to interjections such as ah, oh, hmmm has a high tendency of being sarcastic.

When all features were introduced to BM^1, i.e., BM^1 + HbSD^{All}, the F-score improved for all the three algorithms compared to baseline model with 0.72, 0.71 and 0.70 for SVM, RF and RF with Bagging respectively. It can be concluded that hyperbole is an important marker for sarcasm detection [19] given that it is easy to interpret the meaning when there presents an interjection, intensifier, punctuation, etc. in a text [10, 11]. The F-score results were compared for all features as the accuracy achieved may be influenced by the imbalanced class in this baseline model.

Table 5: Results of three-class setup baseline model and/with different hyperbole features

Model	Combination	Algorithm											
		SVM				RF				RF + Bagging			
		A	P	R	F	A	P	R	F	A	P	R	F
Baseline Model^1 (BM^1)	Sarcasm + Racism + Others	59.83	0.62	0.75	0.68	60.04	0.61	0.81	0.69	60.25	0.60	0.85	0.70
HbSD^{Interjection}	BM^1 + HbSD^{Interjection}	61.28	0.59	0.82	0.68	57.56	0.56	0.78	0.65	59.00	0.56	0.85	0.67
HbSD^{Intensifier}	BM^1 + HbSD^{Intensifier}	59.21	0.57	0.80	0.66	58.80	0.56	0.82	0.67	57.35	0.54	0.88	0.67
HbSD^{CapitalLetter}	BM^1 + HbSD^{CapitalLetter}	58.18	0.58	0.78	0.66	59.83	0.59	0.80	0.68	58.18	0.56	0.84	0.67
HbSD^{Punctuation}	BM^1 + HbSD^{Punctuation}	57.97	0.55	0.78	0.65	56.94	0.55	0.83	0.66	56.73	0.54	0.88	0.67
HbSD^{ElongatedWord}	BM^1 + HbSD^{ElongatedWord}	55.90	0.51	0.83	0.63	58.80	0.54	0.88	0.67	56.52	0.51	0.92	0.66
HbSD^{All*}	BM^1 + HbSD^{All}	63.56	0.64	0.83	0.72	63.56	0.61	0.85	0.71	60.25	0.58	0.86	0.70

Note: HbSD^{All*} = BM^1 + HbSD^{Interjection} + HbSD^{Intensifier} + HbSD^{CapitalLetter} + HbSD^{Punctuation} + HbSD^{ElongatedWord}
A=Accuracy, P= Precision, R=Recall, F= F-score

Table 6 consists of a two-class setup (BM^2) results. Based on the accuracy achieved by individual features against the baseline model, interjection is the most common hyperbole feature used in a sarcastic tweet with an accuracy of 81.72% for RF algorithm. Past studies in sarcasm detection focusing on hyperbole mainly looked at interjection [14, 20, 34, 40] as one of the features. In this study, the result achieved by BM^2 + HbSD^{Interjection} as can be seen in Fig 3 is evident that interjection is indeed the main hyperbole feature for sarcasm detection and the reason why many researchers used it as part of their hyperbole approach.

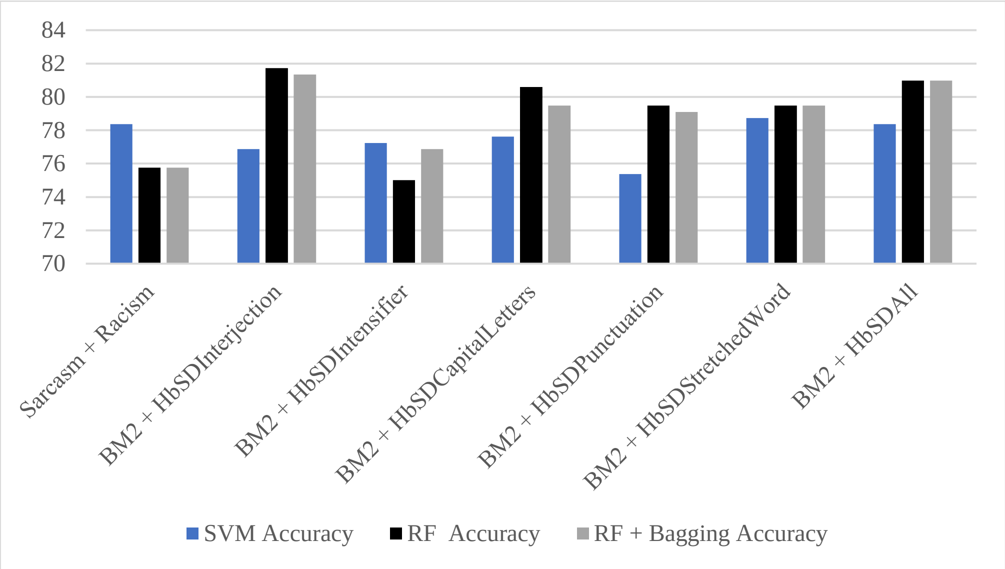


Fig 3: Accuracy for individual features for SVM, RF and RF with Bagging

An example of interjection (Hey) tweet that was part of the evaluation is as below:

Example 1: "*Hey stupid cunt you and shitcanned the relief bill because it did not have emission standards for airlines*".

As the interjection feature had a high accuracy among all single features that were evaluated, an additional experiment combining the other features with interjection was also conducted.

Table 6: Results of two-class setup baseline model and/with different hyperbole features

Model	Combination	Algorithm											
		SVM				RF				RF + Bagging			
		A	P	R	F	A	P	R	F	A	P	R	F
Baseline Model^2 (BM^2)	Sarcasm + Racism	78.36	0.83	0.79	0.81	75.75	0.83	0.73	0.78	75.75	0.84	0.72	0.78
$HbSD^{Interjection}$	BM^2 + $HbSD^I$	76.87	0.79	0.82	0.8	**81.72**	0.86	0.81	0.84	81.34	0.85	0.82	0.84
$HbSD^{Intensifier}$	BM2 + $HbSD^{If}$	77.24	0.82	0.77	0.79	75.00	0.78	0.77	0.77	76.87	0.79	0.79	0.79
$HbSD^{CapitalLetter}$	BM^2 + $HbSD^{CL}$	77.61	0.83	0.79	0.8	80.60	0.87	0.80	0.83	79.48	0.87	0.78	0.82
$HbSD^{Punctuation}$	BM^2 + $HbSD^P$	75.37	0.80	0.74	0.77	79.48	0.85	0.76	0.80	79.10	0.83	0.78	0.80
$HbSD^{ElongatedWord}$	BM^2 + $HbSD^{EW}$	78.73	0.77	0.84	0.81	79.48	0.79	0.84	0.81	79.48	0.78	0.85	0.81
BM^2 + $HbSD^{Interjecton}$ + $HbSD^{CapitalLetter}$	BM^2 + $HbSD^{ICL}$	78.73	0.80	0.80	0.80	78.73	0.80	0.81	0.80	81.34	0.83	0.83	0.83
BM^2 + $HbSD^{Interjecton}$ + $HbSD^{Punctuation}$	BM^2 + $HbSD^{IP}$	72.39	0.71	0.81	0.75	76.12	0.74	0.85	0.79	73.13	0.70	0.87	0.77
BM^2 + $HbSD^{Interjecton}$ + $HbSD^{ElongatedWord}$	BM^2 + $HbSD^{IEW}$	77.24	0.79	0.81	0.80	80.60	0.83	0.82	0.83	**82.84**	0.84	0.86	**0.85**
BM^2 + $HbSD^{Intensifier}$ + $HbSD^{CapitalLetter}$	BM^2 + $HbSD^{ICL}$	77.99	0.77	0.86	0.81	77.24	0.77	0.84	0.80	75.00	0.74	0.84	0.79
BM^2 + $HbSD^{Intensifier}$ + $HbSD^{ElongatedWord}$	BM^2 + $HbSD^{IfEW}$	77.61	0.84	0.76	0.80	77.99	0.85	0.76	0.80	77.24	0.83	0.76	0.80
BM^2 + $HbSD^{Intensifier}$ + $HbSD^{Punctuation}$	BM^2 + $HbSD^{IfP}$	75.00	0.80	0.74	0.77	75.37	0.79	0.77	0.78	73.88	0.78	0.74	0.76

BM^2 + $HbSD^{Interjection}$ + $HbSD^{Intensifier}$	BM^2 + $HbSD^{IIf}$	79.48	0.83	0.80	0.82	80.97	0.81	0.88	0.84	82.46	0.83	0.88	0.85
BM^2 + $HbSD^{All*}$	BM^2 + $HbSD^{All}$	78.36	0.83	0.77	0.82	80.97	0.83	0.83	0.83	80.97	0.81	0.86	0.84

Note: $HbSD^{All*}$ = BM^2 + $HbSD^{Interjection}$ + $HbSD^{Intensifier}$ + $HbSD^{CapitalLetter}$ + $HbSD^{Punctuation}$ + $HbSD^{ElongatedWord}$
A=Accuracy, P=Precision, R=Recall, F=F-score

Capital letters and punctuation marks like exclamation and questions marks are used as special forms to indicate the difference in voice tones in social media [13]. Example 2 below is an example of the different hyperbole features such as interjection, capital letters, exclamation mark and intensifier showing the different intonation that is visible in a text.

Example 2: "*Damn Straight!! Never underestimate the FILTHY ROTTEN to waste an opportunity of A Health PANDEMIC CRISIS to USE for their agenda to CREATE AN initiative!*"

Based on the results in Table 6, interjection with capital letters combination did not give the best results. However, elongated word combined with interjection had the highest accuracy and F-score of 82.84% and 0.85 respectively compared to RF with Bagging technique, whereby interjection and intensifier features fell short slightly with a difference of 0.38 in accuracy. Elongated words can express strong emotions in a text to express the strength of a word such as 'yeahhhhhh', 'hmmmm', 'looooove' [27, 41, 42] and these emotions have a greater impact [43] and can indicate if the user is deliberate or exaggerating [7, 21].

As seen in Table 2 of Section 2 above, the SVM algorithm is the most common machine learning algorithm used for sarcasm detection. However, in this study, among the three algorithms used, the RF model was consistent for single hyperbole features as well as the combination of features. The F-score achieved in this study ranges between 0.78 – 0.84 for RF algorithm which is higher than the proposed model by [26] and [39]. [39] used intensifier and interjection in text whereas [26] used mutual information and clustering for sarcasm detection. The reason RF can perform better is due to the flexibility and the robustness of the algorithm to create any number of decision trees and then select the best option to get the prediction [39] as well as resolve overfitting issue. Fig 4 shows the overall performance achieved by the RF algorithm for BM^2.

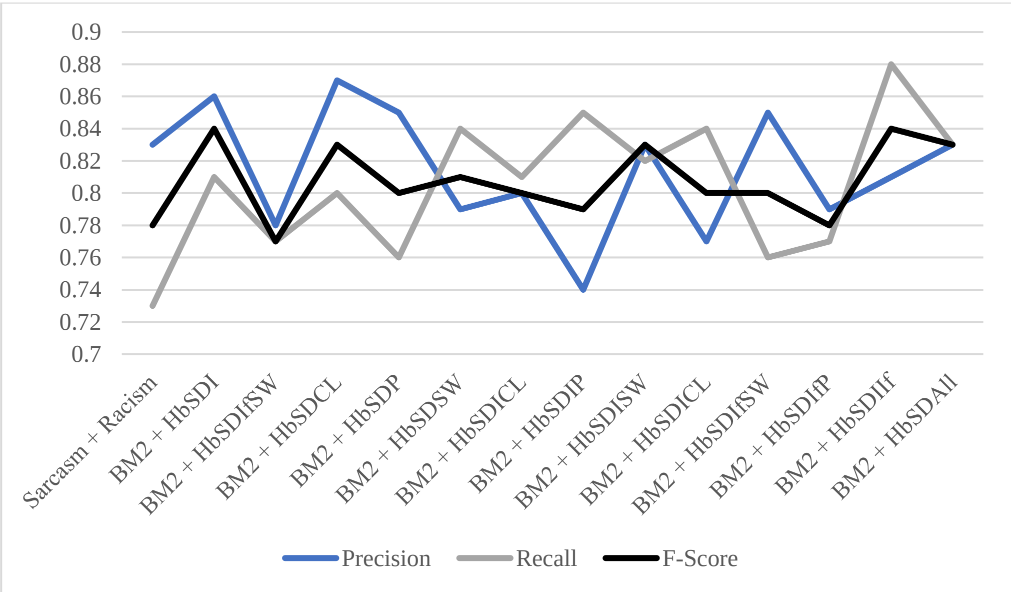


Fig 4: Random Forest performance

We retrieved the dataset from [47] and implemented our hyperbole approach on 4,554 tweets and our model achieved an impressive accuracy of 89.46% and 88.95% for the interjection and elongated words. The model with all hyperboles performed well with an accuracy of 89.31%. It can be noticed that our HbSD can be implemented on any dataset regardless of any approach used in the past. Hyperbole existence enhances the sarcasm detection model.

Full results can be seen in Table 7 below. The significant hyperbole identified in this open dataset are interjection and elongated words which confirms previous findings on the most significant hyperbole.

Table 7: Results of two-class setup baseline model and/with different hyperbole features on open dataset retrieved from [47]

Model	Combination	Algorithm											
		SVM				RF				RF + Bagging			
		A	P	R	F	A	P	R	F	A	P	R	F
Baseline Model[3] (BM[3])	Sarcasm + Others	88.29	0.86	0.59	0.70	88.21	0.92	0.54	0.68	87.34	**0.94**	0.48	0.64
$HbSD^{Interjection}$	BM^3 + $HbSD^I$	**89.46**	0.90	0.60	0.71	87.99	0.88	0.53	0.66	87.19	0.87	0.47	0.62
$HbSD^{Intensifier}$	BM^3 + $HbSD^{If}$	89.24	0.89	0.60	0.71	88.51	0.89	0.54	0.67	87.74	0.88	0.49	0.64
$HbSD^{CapitalLetter}$	BM^3 + $HbSD^{CL}$	88.58	0.89	0.56	0.69	87.92	0.88	0.50	0.65	87.04	0.87	0.46	0.62
$HbSD^{Punctuation}$	BM^3 + $HbSD^P$	87.99	0.88	0.56	0.68	87.92	0.88	0.53	0.66	86.90	0.87	0.47	0.62
$HbSD^{ElongatedWord}$	BM^3 + $HbSD^{EW}$	88.29	0.88	0.58	0.67	**88.95**	0.89	0.57	0.68	87.99	0.88	0.49	0.63
BM^2 + $HbSD^{All*}$	BM^3 + $HbSD^{All}$	88.95	0.89	0.61	0.71	88.14	0.88	0.55	0.67	87.77	0.88	0.51	0.65

Note: $HbSD^{All*}$ = BM^3 + $HbSD^{Interjection}$ + $HbSD^{Intensifier}$ + $HbSD^{CapitalLetter}$ + $HbSD^{Punctuation}$ + $HbSD^{ElongatedWord}$
A=Accuracy, P= Precision, R=Recall, F= F-score

5.0 PRACTICAL IMPLICATIONS

This study aims to identify the presence of hyperbole features and how they improve sarcasm detection as they are deemed as key markers in identifying sarcastic tweets. Firstly, based on the dataset used in this study, the top five hyperbole features that were predominant were identified. Based on the experiments conducted on negative sentiment tweets, they contribute significantly to sarcasm detection as seen in Table 6. For example, interjection scored the highest accuracy in both two-class (accuracy = 81.72%) and three-class (accuracy = 61.28%) setup as a single hyperbole feature as well as in the open dataset that was retrieved from past study with an accuracy of 89.46%. Capital letters, punctuation marks and elongated words only fell short slightly with an F-score of 0.83, 0.80 and 0.81 respectively compared to interjection in two-class setup with an F-score of 0.85. Nonetheless, hyperbolic tweets are a key marker for identifying sarcasm [29, 33].

Secondly, this study focuses on a single approach that is hyperbole-based sarcasm detection model as opposed to other studies that combine lexical and/or pragmatic into their sarcasm model. It is evident that as a single approach, the hyperbole-based sarcasm detection model did improve the sarcasm mechanism compared to studies such as [33] and [35] scoring 77.48% and 69.45% respectively for F-score. Both studies combined lexical and hyperbole approaches. Bouazizi & Ohtsuki [13] combined lexical, hyperbole and pragmatic approach and achieved an F-score of 81.3% with Random Forest classifier.

Finally, the hyperbole-based sarcasm detection model evaluated in this study can be considered as a stable model as it reached an accuracy of 75% in most cases regardless of a single hyperbole feature or combined hyperbole features. The results achieved indicate that hyperbole features, if identified accurately, can improve the sarcasm detection mechanism. A lexical and pragmatic approach may enhance the performance for sarcasm detection, but they act as a supplementary approach to improve performance.

6.0 CONCLUSION, LIMITATION AND FUTURE WORK

The Hyperbole-based Sarcasm Detection (HbSD) model that manipulates the use of various hyperboles was proposed in this study. Experiment results demonstrate a shred of strong evidence that sarcasm can be detected on any dataset (unbiased), as opposed to past studies that predominantly used sarcasm related hashtags to retrieve data for researchers to be able to extract or identify sarcasm patterns [7, 13, 37]. Hyperbole features exist and sarcasm pattern can be extracted or identified regardless of the dataset as can be seen results shown on Table 7 above.

This study does have some limitations. The dataset retrieved in this study is imbalanced, hence the use of two-class setup with only Sarcasm and Racism class. Furthermore, the hyperbole features in this study were semi-manually extracted using tokenization and lemmatization. In future, a dictionary-based approach can be embedded for interjection and intensifier words which may help to reduce the manual work and time. This dictionary can be used for other languages; however, this may be time-consuming due to the nature of collecting and collaborating the words from a reliable resource and verifying them to be hyperbole.

This study is a small part of a larger research that is on-going. The authors would like to include the network features such as followers_count, friends_count and favorite_count that are available on tweets to analyse the user personalities and popularity for sarcasm detection. To the best of our knowledge, research analysing these attributes has not to be investigated for sarcasm detection and the authors would like to include this as part of the bigger picture for sarcasm detection.

ACKNOWLEDGEMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] Ren, L., Xu, B., Lin, H., Liu, X., & Yang, L. (2020). Sarcasm Detection with Sentiment Semantics Enhanced Multi-level Memory Network. *Neurocomputing*, 401, 320-326. DOI: https://doi.org/10.1016/j.neucom.2020.03.081

[2] Signhaniya, A., Shenoy, G., & Kondekar, R. (2015). Sarcasm Detection in Social Media. *Semantic Scholar*.

[3] Whalen, J. M., Pexman, P. M., Gill, A. J., & Nowson, S. (2013). Verbal Irony use in personal blogs. *Behaviour & Information Technology*, 32(6), 560-569. DOI: https://doi.org/10.1080/0144929X.2011.630418

[4] Zhu, N., & Wang, Z. (2020). The paradox of sarcasm: Theory of mind and sarcasm use in adults. *Personality and Individual Differences*, 163, 110035. DOI: https://doi.org/10.1016/j.paid.2020.110035

[5] Filik, R., Turcan, A., Thompson, D., Harvey, N., Davies, H., & Turner, A. (2016). Sarcasm and emoticons: Comprehension and emotional impact. *The Quarterly Journal of Experimental Psychology*, 69(11), 2130-2146. DOI: https://doi.org/10.1080/17470218.2015.1106566

[6] Vashishtha, S., & Susan, S. (2019). Fuzzy rule based unsupervised sentiment analysis from social media posts. Expert Systems With Applications, 138. DOI: https://doi.org/10.1016/j.eswa.2019.112834

[7] Sindhu, C., Vadivu, G., & Rao, M. V. (2018). A comprehensive study on sarcasm detection techniques in sentiment analysis. International Journal of Pure and Applied Mathematics, 118(22), 433-442.

[8] Al-Ghadhban, D., Alnkhilan, E., Tatwany, L., & Alrazgan, M. (2017). Arabic Sarcasm Detection in Twitter, In Proceedings of 2017 International Conference on Engineering & MIS (ICEMIS), May 8-10, 2017, Monastir, Tunisia.

[9] Ramteke, J., Godhia, D., Shah, S., & Shaikh, A. (2016). Election Results Prediction Using Twitter sentiment Analysis. In Proceedings of 2016 International Conference on Inventive Computation Technologies (ICICT), August, 26-27, 2016, Coimbatore, India. DOI: 10.1109/INVENTIVE.2016.7823280

[10] Vijayalaksmi, N., & Senthilrajan, A. (2017). A hybrid approach for Sarcasm Detection of Social Media Data. International Journal of Scientific and Research Publications, 7(5), 327-336.

[11] Kunneman, F., Liebrecht, C., van Mulken, M., & van den Bosch, A. (2015). Signalling sarcasm: From hyperbole to hashtag. *Information Processing and Management*, 51, 500-509. DOI: https://doi.org/10.1016/j.ipm.2014.07.006

[12] Ptacek, T., Habernal, I., & Hong, J. (2014). Sarcasm Detection on Czech and English Twitter. *In Proceedings of the 25th International Conference of Computational Linguistics (COLING 2014)*, August 23-29, 2014, Dublin, Ireland.

[13] Bouazizi, M., & Ohtsuki, T. O. (2016). A Pattern-Based Approach for Sarcasm Detection on Twitter. *IEEE Access*, 4, 5477-5488. DOI: https://doi.org/10.1109/ACCESS.2016.2594194

[14] Bharti, S. K., Vachha, B., Pradhan, R. K., Babu, K. S., & Jena, S. K. (2016). Sarcastic sentiment detection in tweets streamed in real-time: a big data approach. *Digital Communications and Networks*, 2, 108-121. DOI: 10.1016/j.dcan.2016.06.002

[15] Haripriya, V, Patil, P., G. (2017). A Survey of Sarcasm Detection in Social Media. *International Journal for Research in Applied Science & Engineering Technology*, 5, 1748-1753.

[16] Kolchyna, O., Souza, T. T. P., Treleaven, P., Aste, T. (2016). *Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination*. Handbook of Sentiment Analysis in Finance, G. Mitra and X. Yu, Eds., Albury Books, 2016, pp. 1-30.

[17] Mukhtar, N., Khan, M. A., & Chiragh, N. (2018). Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains. *Telematic and Informatics*, 35, 2173-2183. DOI: https://doi.org/10.1016/j.tele.2018.08.003

[18] Carston, R., & Wearing, C. (2015). Hyperbolic language and its relation to metaphor and irony. *Journal of Pragmatics*, 79, 79-92. DOI: https://doi.org/10.1016/j.pragma.2015.01.011

[19] Burgers, C., Renardel de Lavalette, K. Y., & Steen, G. J. (2018). Metaphor, hyperbole, and irony: Uses in isolation and combination in written discourse. *Journal of Pragmatics*, 127, 71-83. DOI: https://doi.org/10.1016/j.pragma.2018.01.009

[20] Suzuki, S., Orihara, R., Sei, Y., Tahara, Y., & Ohsuga, A. (2017). Sarcasm Detection Method to Improve Review Analysis. *In Proceedings of the 9th International Conference of Agents and Artificial Intelligence (ICAART 2017)*, February 24-26, 2017, Gaia, Portugal. DOI: 10.5220/0006192805190526.

[21] Liu P., Chen W., Ou G., Wang T., Yang D., Lei K. (2014). Sarcasm Detection in Social Media Based on Imbalanced Classification. In: Li F., Li G., Hwang S., Yao B., Zhang Z. (eds) Web-Age Information Management. WAIM 2014. Lecture Notes in Computer Science, vol 8485. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-08010-9_49

[22] Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N., & Huang, R. (2013). Sarcasm as Contrast between a Positive Sentiment and Negative Situation. *In Proceeding of the 2013 Conference on Empirical Methods in Natural Language Processing*, October 18-21, 2013, Seattle, Washington, USA.

[23] Joshi, A., Bhattacharyya, P. & Carman, M. J. (2017). Automatic Sarcasm Detection: A Survey. *ACM Computing Surveys*, 50(5), 1-22.

[24] Maynard, D., & Greenwood, M. A. (2014). Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis. *In Proceedings of 9th International Conference on Language Resources and Evaluation*, May 26 – 31, 2014, Reykjavik, Iceland.

[25] Wang, P. A. (2013). #Irony or #Sarcasm – A Quantitative and Qualitative Study Based on Twitter. *In Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC, 27*), November 22-24, 2013, Taipei, Taiwan.

[26] Kumar, H. M. K., & Harish, B. S. (2018). Sarcasm classification: A novel approach by using Content Based Feature Selection Method. *In Proceedings of the 8th International Conference on Advances in Computing and Communication (ICACC-2018)*, September 13-15, 2018, Kochi, India. DOI: 10.1016/j.procs.2018.10.409.

[27] Wijeratne, S., Sheth, A., Bhatt, S., Balasuriya, L., Al-Olimat, H., Gaur, M., … Thirunarayanan, K. (2017). Feature Engineering for Twitter-based Applications. DOI: 10.1201/9781315181080-14

[28] Khedkar, S., & Shinde, S. (2020). Deep Learning and Ensemble Approach for Praise or Complaint Classification. *In Proceedings of International Conference on Computational Intelligence and Data Science (ICCIDS 2019)*, September 6-7, 2019, Gurgaon, India.

[29] Norrick, N. R. (2009). Interjections as pragmatic markers. *Journal of Pragmatics*, 41, 866-891. DOI: https://doi.org/10.1016/j.pragma.2008.08.005

[30] Alqaryouti, O., Siyam, N., Monem, A. A., & Shaalan, K. (2019). Aspect-Based Sentiment Analysis Using Smart Government Review Data. *Applied Computing and Informatics*. DOI: https://doi.org/10.1016/j.aci.2019.11.003

[31] Dharmavarapu, B., Durga, Bayana, J. (2019). Sarcasm Detection in Twitter using Sentiment Analysis. *International Journal of Recent Technology and Engineering*, 8(1S4), 642-645.

[32] Sarsam, S. M., Al-Samarraie, H., Alzahrani, A. I., & Wright, B. (2020). Sarcasm detection using machine learning algorithms in Twitter: A systematic review. *International Journal of Market Research*, 62(5), 578-598. DOI: https://doi.org/10.1177/1470785320921779

[33] Kumar, A., Narapareddy, V. T., Srikanth, V. A., Malapati, A., & Neti, L. B. M. (2020). Sarcasm Detection Using Multi-Head Attention Based Bidirectional LSTM. IEEE Access, 8, 6388-6397. DOI: 10.1109/ACCESS.2019.2963630.

[34] Sonawane, S., S. & Kohle, S. R. (2020). TCSD: Term Co-occurrence Based Sarcasm Detection from Twitter Trends. *In Proceedings of International Conference on Computational Intelligence and Data Science (ICCIDS 2019)*, September, 6-7, 2019, Gurgaon, India. DOI: https://doi.org/10.1016/j.procs.2020.03.422

[35] Jain, D., Kumar, A., & Garg, G. (2020). Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN. *Applied Soft Computing Journal*, 91, 106198.

[36] Farha, I. A., & Magdy, W. (2021). A comparative study of effective approaches for Arabic sentiment analysis. *Information Processing and Management*, 58, 102438. DOI: https://doi.org/10.1016/j.ipm.2020.102438

[37] Mukherjee, S., & Bala, P. K. (2017). Sarcasm detection in microblogs using Naïve Bayes and Fuzzy clustering. *Technology in Society*, 48, 19-27. DOI: https://doi.org/10.1016/j.techsoc.2016.10.003

[38] Poria, S., Cambria, E., Hazarika, D., & Vij, P. (2016). A Deeper Look in Sarcastic Tweets Using Deep Convotional Neural Networks. *In Proceedings of the 26th International Conference on Computational Linguatics : Technical Papers (COLING 2016)*, December 11-17, 2016, Osaka, Japan.

[39] Bharti, S. K., Naidu, R., & Babu, K. S. (2017). Hyperbolic Feature-based Sarcasm Detection in Tweets: A Machine Learning Approach. *In Proceedings of the 14th IEEE India Council International Conference (INDICON)*, December 15-17, 2017, Roorkee, India.

[40] Rahayu, D. A. P., Kuntur, S., & Hayatin, N. (2018). Sarcasm Detection on Indonesian Twitter Feeds. *In Proceedings of the 5th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI 2018)*, October, 16-18, 2018, Malang, Indonesia.

[41] Nadali, S. (2016) *Sarcasm Detection Model Based on Tweets' Strength using Hashtags and Non-Hashtags Sentiment Analysis* [Doctoral Dissertation, Universiti Putra Malaysia]. http://psasir.upm.edu.my/id/eprint/69401/1/FSKTM%202016%2046%20-%20IR.pdf

[42] Karthika, K., Gayathridevi, M., Marikkannan, M., (2017). A Combinational Approach for Sarcasm Detection in Twitter. *International Journal of Science and Research*, 7(7), 1040-1045.

[43] Thompsons, D., & Filik, R. (2016). Sarcasm in Written Communication: Emoticons are Efficient Markers of Intention. *Journal of Computer-Mediated Communication*, 21, 105-120. DOI:10.1111/jcc4.12156.

[44] Mohan, A., Nair, A. M., Jayakumar, B., & Muraleedharan, S. (2023). Sarcasm Detection Using Bidirectional Encoder Representations from Transformers and Graph Convolutional Networks. *In the Proceedings of the International Conference on Machine Learning and Data Engineering*, September 7-8, 2022, Dehradun, India. https://doi.org/10.1016/j.procs.2022.12.405

[45] Misra, R., & Arora, P. (2023). Sarcasm detection using news headlines dataset. *AI Open*, 4, 13-18. https://doi.org/10.1016/j.aiopen.2023.01.001

[46] Frenda, S., Cignarella, A. T., Basile, V., Bosco, C., Patti, V., & Rosso, P. (2022). The unbearable hurtfulness of sarcasm. *Expert Systems With Applications*, 193,116398. https://doi.org/10.1016/j.eswa.2021.116398

[47] Khotijah, S., Tirtawangsa, J., & Suryani, A. A. (2020). Using LSTM for Context Based Approach of Sarcasm Detection in Twitter. *In the Proceedings of International Conference on Advances in Information Technology (IAIT2020)*, July 1-3, 2020, Bangkok, Thailand. https://doi.org/10.1145/3406601.3406624

[48] Muhammad, W., Mushtaq, M., Junejo, K. N., & Khan, M. Y. (2020). Sentiment analysis of products reviews in the absence of labelled data using supervised learning approaches. *Malaysian Journal of Computer Science*, 33(2). https://doi.org/10.22452/mjcs.vol33no2.3

BIOGRAPHY

Vithyatheri Govindan is currently pursuing her PhD in University of Malaya, Malaysia. She completed her Diploma in Computer Science (Information Technology) in year 2000 and her degree, Bachelor of Science (Computer), in 2002, both from University Technology of Malaysia. She graduated with a Master of Computer Science from University of Malaya in 2011. She has been working in the Information Technology industry since 2006, starting as an Integration Development Engineer in Xybase Sdn Bhd., Malaysia, continued as a consultant in Tieto Sdn Bhd., Malaysia since 2007 and currently works in Mercer, a software company based in Singapore since 2012. She received 4th place in the National Postgraduate Poster Competition 2019 held in University Science of Malaysia, has co-authored three papers (2012, 2019 and 2022) and one conference paper (2011). Her current research interest is in data analytics and sentiment analysis.

Dr. Vimala Balakrishnan is an Associate Professor affiliated with the Faculty of Computer Science and Information Technology, University of Malaya since 2010. She obtained her Ph.D. in the field of Ergonomics from Multimedia University. Dr. Balakrishnan's main research interests are in machine learning (data analytics) and sentiment analysis, particularly related to social media. Her research domains include healthcare, education and social issues such as cyberbullying and misinformation dissemination. She has published more than 80 articles in top indexed journals, 50 conference proceedings and seven book chapters, has four patents and eight copyrights. She also serves as an Associate Editor to the Malaysian Journal of Computer Science, and as an Associate Member for the Global Science and Technology Forum. She is also a fellow for the Leadership in Innovation program, a prestigious award by the Royal Academy of Engineering, UK, and a Fulbright Visiting Scholar 2018.