ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

Sarcasm Detection on Twitter using Feature-Based Approach

Khan Shehla Kulsum, Nagesh Kamble

M.E. Student, Department of C.S.E., Shreeyash College of Engineering& Technology, Aurangabad, India

Asst. Professor, Department of C.S.E., Shreeyash College of Engineering& Technology, Aurangabad, India

ABSTRACT: A lot of research is being conducted in the field of Sentiment Analysis of literal statistics that exist on the web. Sarcasm can be defined as a sophisticated form of representation that is used to express mockery and irony widely used in social networks and micro blogging websites. Sarcasm might be used for different purposes, such as criticism or ridicule where, the speakers express their message in an implicit or hidden way. Automatic detection of sarcastic utterance is a challenging task because of the ambiguity that is inherent in sarcasm often makes it difficult even for human-beings to decide whether the content is sarcastic or not. It has impact on opinion mining and sentiment analysis but, is typically ignored in social media analysis as it is considered challenging task to handle. Existing methodologies for automatic sarcasm recognition depends mostly on lexical, verbal and linguistic indicators. Recognition of sarcasm can contribute to improved performance of these systems. In this paper, we propose a novel approach for identification of sarcastic utterance present in tweets. Our approach has two stages namely feature extraction and sarcasm classification. The feature extraction stage focuses on four sets of features based on sentiments, punctuation, syntax and pattern which cover the various forms of sarcasm namely wit, whimper and evasion. Next we make use of Machine Learning algorithms to classify tweets depending on whether utterance is sarcastic and non-sarcastic. We evaluate the significance of each of these feature sets and assess its added importance in the classification. Our proposed approach achieves an accuracy of 90% by using all the features together for sarcasm recognition.

KEYWORDS: Sarcasm detection; Twitter; Sentiment analysis; NLP; Feature extraction

I. INTRODUCTION

As stated above, the recognition of sarcasm aids in improving opinion mining and sentiment analysistasks when implemented on micro blogging websites such as Facebook, Twitter, etc. Sentiment Analysis refers to the technique of identification of attitude, feeling, opinion and emotion conveyed by online users towards any specific object such as person, occasion, issue, product, society, services etc. Sentiment analysis, opinion mining and NLP applications depend on words that have emotional content to detect the polarity whether positive or negative. But, it could be misleading from the look of the text. Sarcasm acts as an interfering-factor that flips the polarity of the given message into its opposite because sarcastic message usually expresses a negative sentiment using simply positive words or even exaggerated positive words. Therefore identification of contents that contain sarcasm is significant for the development and improvement of sentiment-analysis, opinion mining as well as many NLP applications, but sarcasm detection is conceptually and technically challenging task.

Sarcasm can be referred to mocking, contemptuous, or making use of ironic language with the intention to convey scorn or insult. According to Macmillan English dictionary, sarcasm is "the activity of saying or writing the opposite of what you mean, or of speaking in a way intended to make someone else feel stupid or show them that you are angry". Sarcasm is often conveyed verbally by the people by making use of heavy tone or using certain gestural signs like rolling of the eyes. These clues are not present when conveying sarcasm in written text, thus making it hard to recognize. On social media, these gestural signs and voice tones are often expressed by sarcasm-related emoticons such as :P, :/, ;-), etc.as well as by making use of use of exclamation marks, question marks and capital letter words. People make use of sarcasm in their day-to-day life to make jokes to be witty. Sarcasm is also used to criticize and condemn or make comments about persons, events, ideas, etc. We could broadly categorise the sarcasm into following three

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

categories namely whimper, wit, and evasion. When person is really annoyed, he usessarcasm as whimper where he refers to a negative situation using positive expressions or vice versa. Person uses sarcasm as wit by exaggerating, using special forms of speeches and tones unlike those when he usually talks. To avoid giving clear reply sarcasm is used as evasion, where people use complicated sentences, uncommon expression, etc.Sarcasm occurs commonly in content generated by the internet users such as blogs, forums and micro-posts, and is intrinsically difficult to analyse, not only for a machine but even for a human. Thus, presence of sarcasm can lead to shift in evaluative valence in either ways that is; shift from a plainly positive to anticipated negative sense, or the other way round. The state of the art methods for sentiment-analysis and opinion-mining are likely to have lower performances when scrutinizing information collected from social networks. Maynard and Greenwood [7] illustrated how the performance of sentiment analysis can be enhanced by identification of sarcasm within the context. Presence of sarcasm may hurt opinion mining systems since its obvious meaning is opposite from the actual anticipated meaning, therefore averaging on the sentiment would not be precise. A major challenge in opinion mining and sentiment analysis system is to automatically recognition the sarcasm present in the utterance. Identifying sarcastic utterance in textual context can improve automatic sentiment analysis of data collected from social networks and microblogging websites. Recognizing sarcasm is significant for natural language processing to evade misinterpreting sarcastic utterance as literal so that we can greatly enhance the performance of many NLP applications such as review summarization, brand monitoring systems, dialogue systems and ranking systems, to name a few. The successful identification of sarcastic utterance in the reviews can enhance the personalization of content ranking and recommendation systems.

In past few years, social media sites such as Twitter have gained enormous popularity and acceptance. Twitter allows users to post and read short text messages, called "tweets". Twitter has become one of the leading web destinations for users to show their attitude and opinion for a particular event or product, to share their feelings or to report real-time events, etc. The information generated by Twitter is important to many corporations, enterprises as well as government organizations for the purpose of learning the sentiments of people towards particular products [1], reviews about movies [2] or political events [3]. Twitter allows its users to make use of the informal language, slangs etc. and there is restriction on length of tweet in terms of characters which are limited to 140, therefore determining the feelings of users and performing analysis on it is quite tough. In addition, the existence of sarcasm makes the job even more difficult. On Twitter people often explicitly mark their tweets with the hashtag '#sarcasm'to avoid misunderstanding their sarcastic message. Most commonly, sarcasm is expressed by positive sentimentthat is used in contrastwith the negative circumstance,for example a positive emotion, such as "love" or "happy", followed by a negative situation such as taking exams or being ignored. Due to restriction on tweet length, user often make use of symbolic text and slangs in their tweets such as @USER, smiles, emoticons, interjections and punctuation mark. It has been observed that, there exists certain feature in the text such as lexical, hyperbole and pragmatic, that plays important role in sarcasm detection. There is high tendency of hyperbolic text to be classified as sarcastic as well as most tweets that start with the interjection have a higher probability to be categorised as sarcastic. In addition, the existence of intensifier in a text also raises the probability of sarcasm. Combination of features such as intensifier, interjection (e.g. wow, yay, yeah), quotes and punctuation-mark (e.g. ' ', ‘‘ ’’ , !!!!!), in a given text is called hyperbole. In a given text adjective and adverb act as an intensifier. The lexical features such as uni-gram, bi-gram, and n-gram are valuable to detect sarcasm. Another, feature that is valuable to identify sarcasm is pragmatics such as smiles and emoticons ;-), :P, ☺,etc. In addition to this, certain sarcastic patterns may also be present in the text.

The process of sarcasm detection using classification techniques makes use of the following steps:

Step1: Generation of Dataset and performing data Pre-processing

Today, the enormous amount of annotated corpora is available for Sentiment Analysis but for sarcasm detection no gold standard dataset is available, which is the biggest challenge for sarcasm detection. Also, Data obtained from online platform such as Twitter, Facebook, etc. are unstructured and does not follow grammar rules. So, Data Pre-processing is required to remove the noise present in the data set. Noise could be a user defined label, spelling mistakes, slang words, URLs, etc.

Step2: Feature extraction for Sarcasm Detection

Part of Speech (POS): POS is used to define the sematic meaning of word appearing in document i.e. adjective, noun, verb, etc. The features related tosarcastic patterns,sentiments, etc. are extracted from these POS-tagged tweets

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

Step3: Sarcasm Classification Techniques

The classification techniques for sentiment detection can be roughly divided into lexicon based approach, machine-learning (ML) approach, and hybrid approach. MLbased approach makes use of different ML algorithms like SVM, Naive Bayes, Decision-tree, etc. based on lexical features and, is further divided into supervised machine learning methods that uses large labelled data set and unsupervised Machine learning techniques, which are used where it is hard to find labelled corpora to train the classifier.

Step4: Supervised Machine Learning Techniques

Supervised learning methods can be used when sufficient labelled training corpora is available. Sarcasm detection problem can be formed as: Given training set T = {t1, t2, t3, ...,tn} with each tweet (ti) assigned one of the class label sarcastic or non-sarcastic. Next, Model M relates feature set of tweet to class labels. After that, given new tweet "t", model M is used to predict class label for it.

In this paper, we attempt to show an efficient approach to recognize sarcastic tweets. The proposed approach takes into consideration various forms of sarcasm and identifies the sarcasm present within the tweets irrespective of their user's previous knowledge or their temporal perspective, with an accuracy of 90%. The key contributions of the proposed approach include the following:

1) Our system provides an efficient way to recognize sarcastic tweets
2) We determine the importance of the different feature sets used and there combine effect to determine sarcasm.
3) We also make an attempt to understand how to use this knowledge to enhance the correctness of sentiment analysis, opinion mining and NLP applications.

The remainder of our paper is organised as follows: Section II refer to some state of the research work done in the area of sarcasm recognition.Section III explainsthe proposed system for sarcasm recognition. Section IV we present the evaluation procedure and the results we obtained from our system and in Section V we provide the conclusion.

II. RELATED WORK

Sarcasm is studied well in the area of psychologists, behavioural scientists and linguists. However automatic recognition of sarcasm is challenging task in case of text mining and has been addressed by few researchers. Sriram et al. [8] suggested a method to enhance filtering of information in short text classification specifically on Twitter. Their approach made use of non-context features including presence of sentiment-words, slangs, phrases related to time event and the user information that exists in the text and used this to classify the tweets that fall into predefined set of generic-classes namely opinions, events, private messages, etc. Akcora et al. [9] proposed a system capable of identifying breaking point in the public opinion by making use of the word-pattern as well as emotive-pattern present in the textthat is used to decide how the public opinion varies over the time. They implemented an emotion corpus to identify emotions in tweets. Their method allows expanding opinion representation from binary options to multiple dimensions by providing additional granularity in classification. They propose combining set and vector space models to detect the public opinion and discover changes over time.Tepperman et al. [10] proposed an approach to determine the sarcastic utterance by studying the presence of expression "yeah right!". Their approach makes use of contextual clues along with prosodic and spectral hints to recognize the presence of sarcasm that is present in spoken dialogues. Davidov et al. [5] in his paper "Semi-supervised recognition of sarcastic sentences in Twitter and Amazon" proposed a semi-supervised algorithm that is capable of identifying whether sarcasm is present in the context or not. They experiment with semi supervised approach for sarcasm identification by using two very different datasets and created a gold standard sample in which each sentence was tagged by three annotators. Their approach detects presence of sarcasm by recognizing the sarcastic pattern present in the text. Maynard et al. [6] studied the effect of sarcasm on the polarity of tweets and developed set of rules that can be used to determine the polarity of tweet. They showed that performance of sentiment analysis can be highly improved by identifying the presence of sarcastic expression within the statements. They contemplate the effect of sentiment and sarcasm contained in hashtags, and developed a hashtag tokenizer for GATE. Not only they were concerned with detecting sarcasm within the tweet, but also considered a range of the sarcastic modifier on the meaning of the tweet and on the polarity of the sentiment expressed. Riloff et al. [11] recommended a system to determine a specific type of sarcasm used commonly on Twitter, where a positive emotion is

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

used to express a negative situation. On Twitter it has been noted that sarcastic tweetsusually contains a positive sentiment, such as “love”, “happy” or “enjoy”, that is followed by an expression which describes an negative situation or state such as “being ignored” or “taking exams”. Their system takes a group of sarcastic tweets and uses the seed word “love”and learns phrasal lexicons of positive-sentiments and negative-situations. The recognition of presence of sarcastic expressionwas done by using a bootstrapping algorithm. Identification ofcontrasting contexts by making use of the phrases that were learned through bootstrapping algorithm yields better recall for sarcasm recognition. Rajadesingan et al. [7] presented a behavioral modelling approach for identifying sarcasm present in tweets. They proposed an approachto determine presence of sarcasm that makes use of the behavioral traits that are intrinsic to users conveying sarcasm. They identified various forms of sarcasm present in Twitter, and showed that the historical information of the user collected from his previous tweets plays an important role in detecting sarcasm. It identifies behavioral traits expressed by user by collecting their past tweets and make use of behavioral and psychological theories to aid in the construction of a behavioral model that detect sarcasm. This modelling framework is known as SCUBA which determines the probability of the user being a sarcastic person by examining the user’s past tweets. This framework extracts features that identifies the di□erent forms of sarcasm and utilizes these features along with some labelled data in a supervised learning framework to decide sarcastic tweets. Bharti et al [12] proposed the following approaches for detecting sarcasm on Twitter: The first algorithm PBLGA (A parsing based lexical generation algorithm) identifies sarcasm as contrast between positive-sentiment and negative-situation and vice a versa by generating the lexicon to detect presence of sarcastic utterance in tweets. They used the learned lists of sentiment and situation phrases to identify sarcasm present in new tweets by recognizing those contexts that contain a positive-sentiment in close vicinity to a negative-situation phrase. The second algorithm IWS (Interjection_word_start) determines a hyperbole feature in tweets that starts with interjection word where, hyperbole is a grouping of interjection and intensifier. O. Tsur et al. [4] suggested an approach to identify sarcastic sentences present in online product reviews. A semi-supervised algorithm called SASI was proposed for identification of sarcastic utterance in product reviews. It involves two main stages, the first one is concerned with semi-supervised approach that performs pattern acquisition which is followed by sarcasm classification. During their work they found some strong features that can be used to classify sarcastic utterances. They extracted syntactic features along with pattern based features that were used to construct feature-vectors which were later utilized to build the model capable of assigning scores to the unlabelled examples.

III. PROPOSED APPROACH

Given some tweets, the objective is to classify them as sarcastic tweets or non-sarcastic tweets. To do this, we extract a certain features relating to sentiments, syntax, pattern and punctuation. These features are extracted to cover different forms of sarcasm namely wit, whimper, and evasion by making use of various components present in the tweet. We extract these features from the given set of tweets, which we refer to as training set. Next we provide this training set to ML (machine-learning) algorithms to carry out the classification. We first use the training set to train the model to detect the sarcastic and non-sarcastic tweets. Once the model is trained then we use the test data to determine the performance of our proposed approach.

We used Twitter’s streaming API to collect both sarcastic and non-sarcastic tweets. The sarcastic tweets were collected by querying the API for those tweets that contain the hashtag “#sarcasm”. We collected non-sarcastic tweets that deals with different subject matter with some emotional content. These tweets were filtered out to remove very short tweets, non-English tweets, etc. Finally were prepared the following datasets as discussed below:
- **Training set:** We use a dataset to train the classifier that contains six thousand tweets out of which half are sarcastic tweets, and the remaining half are non-sarcastic tweets.
- **Test Set**: this set contains one thousand tweets out of which half are sarcastic tweets, and half are non-sarcastic tweets which have been manually checked and classified. The test set is used to estimate the performances of ourproposed system.

System model
The System-model that is used to determine the sarcastic utterance present in the tweet is shown below. It gives the details of extracting the features from the give dataset and then performing the classification. As mentioned above the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

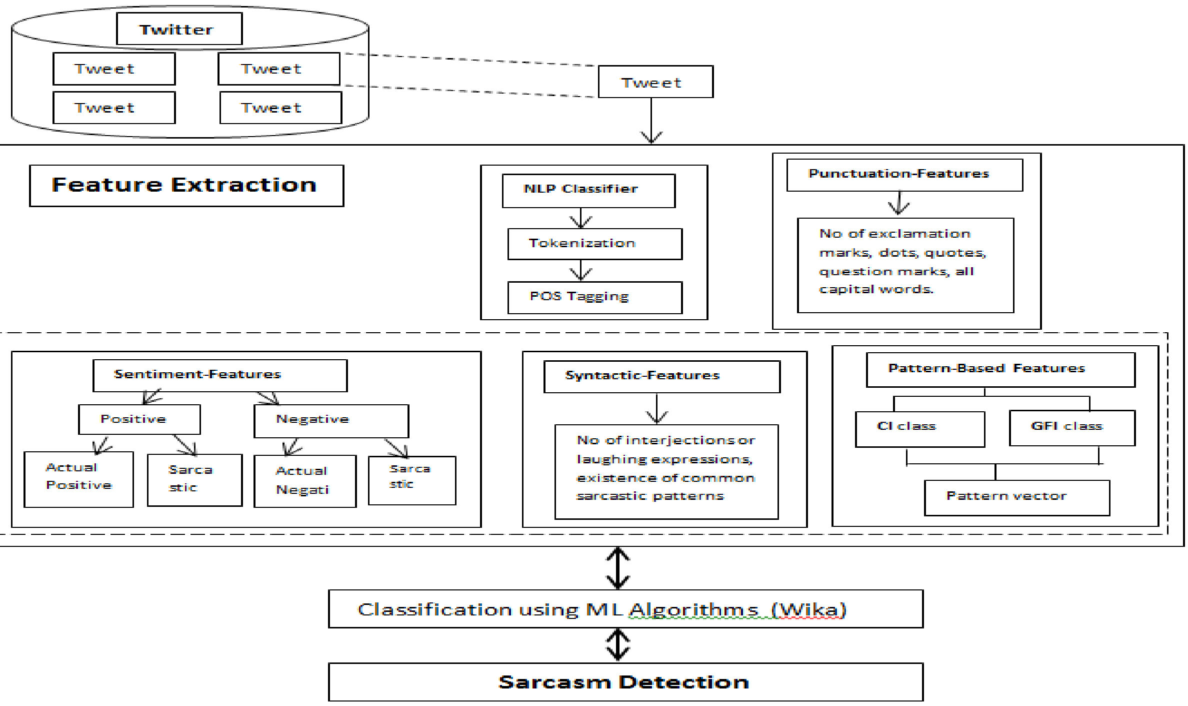
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

dataset consists of tweets collected by using streaming API provided by Twitter. Sarcastic tweets were collected by querying the API for tweets that contained hashtag #sarcasm. Non-sarcastic tweets were collected on various issues that have some emotional contents. The tweets may contain simple text as well as references to URL's, other Twitter users (using @<user>) or a content-tag called hashtags, for example #ihateyou, etc. Therefore it is important to first clean such text data before we could extract the features from it. For sarcastic tweets, we discard all tweets that contain https-address referring to a photo that contains sarcasm. We also remove the tweets written in languages other than English as well as those tweets whose length is less than 3 words and the duplicate tweets. The NPL classifier is used to perform tokenization and part of speech tagging. Features relating to punctuation, sentiment and common sarcastic pattern are extracted. Finally, the classifier is used to perform the classification on tweets to determine whether the tweet is sarcastic or non-sarcastic.



Twitter Streaming API:

We can easily integrate the Java application with the Twitter service using Twitter4J. This specification provides an API for representing a stream of data in web applications, as well as programmatically reading and writing it. Twitter offers a set of streaming APIs that gives its developers access to the global stream of Twitter data. The Twitter streaming API includes the following interfaces: ReadableStream, WritableStream, ByteStream, etc. Several streaming endpoints such as Public streams, User streams, Site streams, are offered by Twitter, each of which can be customized to certain use cases.

NLP Classifier

The NLP stands for Natural Language Processing and the NLP classifier get linguistic-annotations for a text and is a tool that takes in the data-items and classifies them into one of the k-classes. In natural language processing (NLP) we perform various tasks such as tokenization, part of speech tagging (PoS), lemmatization, etc. We used Stranford Classifier (Stanford CoreNLP) to implement the various NLP(Natural Language Processing) tasks.

Tokenization

Given some text, tokenization is the process of breaking it down into pieces which are called tokens. Tokenizationusually takes place at the word level. A token is defined as a sequence of characters present in the given

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

document which are grouped together to form a suitable semantic unit and can later be used during processing. Tokens can be a sequence of alphabets, or it can be combination of alphabets and numerals. The lexical analysis use tokenization to break the stream of text into words, phrases, etc. called tokens. The output of tokenization is a list of tokens which becomes input for further analysis. Various heuristics that a tokenizer uses includes: (i) the whitespace characters (space or line break) are used to separate the tokens, (ii) tokens can be a contiguous strings of alphabetic characters or numbers as well as alphanumeric characters, (iii) Punctuations and whitespaces may or may not be included in token.

Part-Of-Speech (POS) Tagging

Part-Of-Speech tagger reads the text and then assigns parts of speech to each token based on its definition as well as its context depending on its association with adjacent and related words in the text. POS tagger assigns one of the following part of speech noun, verb, adverb, conjunction, interjection, adjective, etc. Many applications in computational-science require fine-grained part of speech tagging, e.g. nouns can be distinguished as the singular, plural, or possessive forms.POS tagger uses notation NN for singular common nouns, NP for singular proper nouns and NNS for plural common nouns. The POS tagger uses algorithms that fall broadly into one of the two categories viz. rule-based and stochastic.

Feature Extraction

We extract four sets of features that are related to punctuation, sentiment, syntax, and pattern which are discussed below:

• **Sentiment-Based Features:** A common form of sarcasm usually used on social-media is whimper in which person makes use of a positive sentiment to describe a negative situation. Sarcasm is also expressed by making use of contradictory sentiments which can be recognized when a positive sentiment is used express a negative situation for example, “I love being ignored all the time”. To identify such sarcasm we study any kind of contradiction between word’s sentiments and other components in the given tweet. From each tweet we extract sentiment related components and count them. We used SentiStrength to create two lists of words which we call pw (positive words) and nw (negative words). The SentiStrength database contains a list of emotional-words, where the positive-words such as “happy”, “love”, etc. have score ranging from 1 to 5 and the negative-words such as “sad”, “hate”, etc. have the scores ranging from -1 to -5, where score 1 to 5 means ranging from almost positive to extremely positive and similarly score -1 to -5 means ranging from almost negative to extremely negative. In generalnouns have lower emotional content compared to verbs, adverbs and adjectives.Therefore positive-word and negative-words that have verbs (VB), adverbs (RB) and adjectives (JJ) POS tags associated with them are counted once more and we create two additional features denoted by PW and NW, where PW represents no. of highly emotional positive-words and NW represents number of highly emotional negative-words.Then we compute the ratio of emotional-words ρ(t), which is computed as follows:

$$\rho(t) = \frac{(\text{delta}\cdot PW + pw) - (\text{delta}\cdot NW + nw)}{(\text{delta}\cdot PW + pw) + (\text{delta}\cdot NW + nw)}$$

where, t stands for the given tweet, delta denotes the weight greater than 1 which is given to the highly emotional-words, pw, nw denoted the number of positive-words and number of negative words respectively and finally PW, NW denotes highly emotional positive-words and highly emotional negative-words respectively. We refer these features as sentiment features which we use during the classification.

• **Syntactic Features**

When people use sarcasm, they often make use of some common expression. To determine whether utterance is sarcastic or not,we can associate these commonly used expressions with the punctuations present in the text. Sometimes people use a specific form of sarcasm called evasion in which they make use of uncommon words and complicated

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

sentences so that it becomes ambiguous to the listener. Toextract the syntactic features, we identify use of uncommon words, presence of sarcastic words, no of interjections, and no. of laughing-expressions. We extract syntax-patterns of ranging between 3 to 6 words that occurs at least 10 or more times.

- **Punctuation-Related Features**

The behavioral aspects such as exaggeration, low tones, facial gestures (e.g. rolling of eyes), etc. are translated in use of punctuations or repetition of vowels in the written text. To detect these aspects, we extract features relating to punctuation. To do this we calculate the number of exclamation marks, question marks, number of dots, number of quotes, the number of all-capital words, and the number of emoticons. We set the counter for punctuation related features which is used to determine whether sarcasm is present or not. The use of exclamation by the user is not significant on its own and does not give clearing understanding whether the user is communicating sarcasm or some other emotion but when we use it with other features then this feature adds value for classification.

- **Pattern-Related Features**

Given a tweet, we divide its words into two classes which we refer to as class CI and class GFI. The word included in first class are the ones in which content is more important, whereas the words that are included in second class are the ones where the grammatical function is important. The words belonging to first class are lemmatized; and others are replaced by the expressions such as CD, FW, LS, MD NN, PRP, SYM,UH, WDT. Next, for each tweet we create the vector of words and extract patterns from the given set of tweets. The patternsthat appears two or more times are stored as sarcastic-patterns. Next for any new tweet, we compare the pattern present in the tweet with our sarcastic patterns; if a match is found then we increment its counter. These pattern features are used during classification of tweet into sarcastic or non-sarcastic.

Classification using Weka

The classification is performed by using "Weka" that has a variety of classifiers. Weka is a data mining/machine learning tool developed by University of Waikato(New Zealand). Weka is open-source software that is issued under the **GNU (General Public License)**. Weka is an assortment of machine learning algorithms for data mining tasks. Weka contains tools for data pre-processing, classification, clustering, regression, association rules, and visualization.

IV. Experimental Results

This section describes the results of the proposed approach. We perform cross-validation and train-and-test on the "Random Forest" classifier using different datasets. Finally, we make comparisons between the results obtained from our system and the proposed approach in the base paper. We make use of features that we extracted from the tweets and use them for classification. We evaluate the system using following KPIs (Key Performance Indicators): Accuracy, Precision and Recall.

For classification we make use of "Random Forests" classifier that works by building a multitude of decision-trees at the time of training the classifier and output the class, which is the mode of the classes (when using classification) or mean prediction (when using regression) of the individual trees. Random decision forests overcome the difficulty of decision tree's problem of over fitting its training set. We set the following parameter of the classifier: number of features=20, number of folds=10, Max Depth=10. We obtain an overall accuracy of 92% during cross-validation and 90% during train-and-test.

4.1 Performance of each Feature-set

We use Random Forest Classifierto check the performances of classification for each set of features individually.

4.1.1 Performance during Cross Validation:

We use that training set comprising of 6000 tweets half of which are sarcastic and half are non-sarcastic for cross-validation. Figure below shows the performances of classification during cross-validation for sentiment based features, punctuation based features, syntax based features, and pattern based features respectively.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

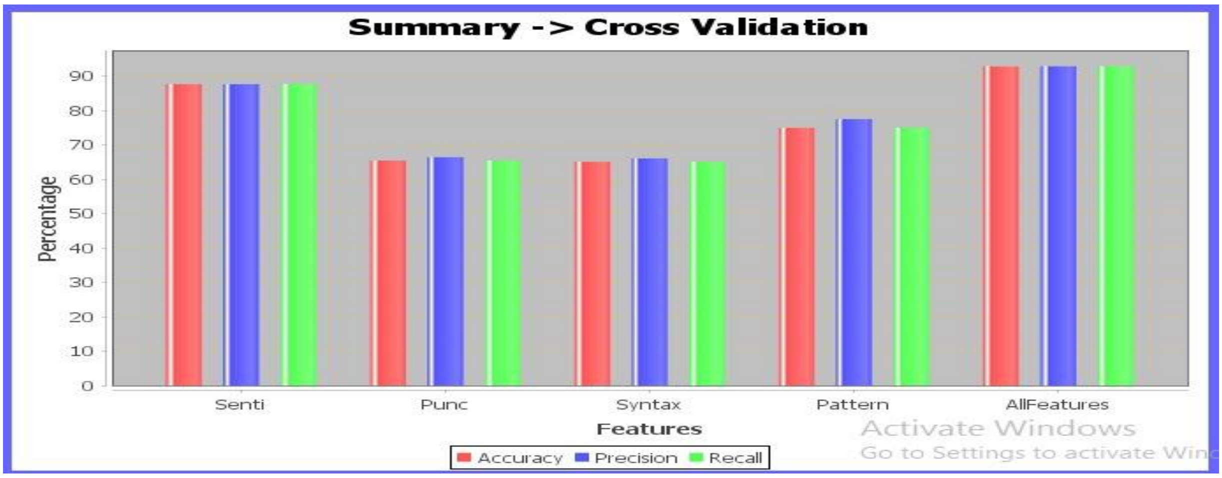


Fig. 2: Accuracy, Precision and Recall of classification during "Cross Validation" for each set of features.

Accuracy for sentiment based features is 87.65%, whereas, the accuracies of punctuation and syntax based features are 65.42% &65.07% respectively. For pattern based feature we obtain an accuracy of 74.94% during cross-validations. Table below show the accuracy, precision, recall and f-score for each of these feature-sets

Table 1: Performance of different features sets during "Cross validation"

Feature	Accuracy	Precision	Recall	F-score
Sentiment	87.65 %	87.66 %	87.65 %	87.64 %
Punctuation	65.42 %	66.42 %	65.42 %	64.89 %
Syntax	65.07 %	66.05 %	65.07 %	64.53 %
Pattern	74.94 %	77.49 %	74.94 %	74.35 %

From above analysis we observe that the performances of the sentiment and pattern based features is high during cross-validation whereas, the syntax-based featuresshows a low accuracy. We also observe that the features relating to punctuation and syntaxare not very useful(if used alone) for classifying tweets as sarcastic and non-sarcastic. Moreover owing to the noise present in thetweets and the informal language used by the users on Twitter, the performance of Part-of-Speech tagger is poor compared to when it is applied to a formal text. The Part-of-Speech tagger is not very effective in detecting interjections, as it classifies them as nouns in most cases. But, as we can see the precision given by Punctuation feature set exceeded 65% which shows the importance of these features in predicting sarcastic contents. Even though punctuation and syntax features perform poorly when used on its own, these features have higher added-value when used together with other features.

4.1.2. Performance during Train and Test

We use two datasets one for training the classifier and the second for testing it. The training-set consist of 6000 tweets out of which 3000 tweets are sarcastic and 3000 tweets are non-sarcastic whereas, the test-set consist of 1000 tweets (half of which are sarcastic and half are non-sarcastic). The performance of classification on feature-sets relating to sentiment, punctuation, syntax, and pattern is shown below.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

Fig. 3: Accuracy, Precision and Recall of classification during "Train and Test" for each set of features.

From the above figure we notice that the features-sets relating to sentiment and pattern gives higher accuracy during the classification of the test-set. The syntax-related features show low accuracy and f-score, due to low presence of syntax-patter and interjections in the tweets comprising the test-set.The performance of syntax and pattern based features can be further improved by obtaining more patterns by training the classifier on larger dataset. Table below shows the accuracy, precision, recall and f-score for each of these feature-sets.

Table 2: Performance of different features sets during "Train and Test"

Feature	Accuracy	Precision	Recall	F-score
Sentiment	84.87 %	85.02 %	84.87 %	84.85 %
Punctuation	62.02 %	62.67 %	62.02 %	61.53 %
Syntax	65.63 %	66.83 %	65.63 %	65.01 %
Pattern	71.24 %	73.2 %	71.24 %	70.62 %

4.2 Performance of the proposed system using all features together

When we use all the features relating to sentiment, punctuation, syntax and pattern together for classification, then the performance of classifier improves and we get an accuracy of around 92% during cross-validation and accuracy of around 90% when we use a separate test set. The performance of the proposed approach when wecombine all the features together is shown in Fig.2 and Fig.3. One of the important conclusions that we can draw from these experimental result is that, the accuracy is low when we use features (relating to sentiment, punctuation, syntax and pattern) separately for classification and accuracy is highly improved when all features are combined together.

Table 3: Performance of all features sets used together during Cross-validation and Train-and-Test

All Features	Accuracy	Precision	Recall
Cross-validation	92.81 %	92.86 %	92.81 %
Train and Test	90.78 %	90.78 %	90.78 %

As can be seen from the above table, the performance of classifier on test set is lower than that during cross-validation. The accuracy and precision achieved during cross-validation isaround 92% and that on test set we obtain an accuracy of 90%.But from the above observations, it is quite clear that if we combine the feature sets relating to sentiment, pattern, punctuation and syntax together, then the performanceis far better compare to the performance when each feature set used on its own. Finally, we make comparison of our proposed approach for recognizing the presence of sarcasm in tweet with the state of art method proposed by M. Bouazizi et al [13] is shown in the table below

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 7, July 2017

Table 4: Performance of proposed system compared to the one used as Base line

All Features	Accuracy	Precision
M. Bouazizi and T. Ohtsuki (Cross-validation)	90.1 %	91.3 %
Proposed approach (Cross-validation)	92.81%	92.86 %
M. Bouazizi and T. Ohtsuki (test-set)	83.1 %	91.1 %
Proposed approach (test-set)	90.78%	90.78%

As can be observed from the above results our proposed approach achieves higher accuracy during cross-validation as well as on test set.

V. Conclusion and Future Work

In this paper, we proposed an efficient approach that is capable of automatically recognizing presence of sarcasm in the given tweet. For a given the set of tweets, the proposed system makes use of part-of-speech tagging to extract features relating to sentiments, syntax, punctuation and patterns by taking into account different forms of sarcasm and make use of different components of the tweets. Then we employ machine learning algorithm to perform the classification. As the dataset is relatively small in size, therefore all possible sarcastic patterns are not covered but, the efficiency can be enhances by training it on larger training set. Theproposed system can enhance the performance of opinion mining and sentiment analysis systems due to its ability to detect the sarcasm in the written text that is generally overlooked in these systems.Our system has shown decent results, but the extracted patterns do not include all sarcastic patterns and the result can further be enhanced by using a larger training set. Our proposed approach achieves an accuracy of 92% during cross-validation. We also study the significance of the proposed feature sets and evaluate their added value in the classification process. In upcoming work, we would like to determine how we can use the output of our proposed system to enhance the applications related to NLP as well as sentiment analysis system so that their performance can further be improved.

References

[1] S. Homoceanu, M. Loster, C. Lofi, and W.T. Balke. 201. ''Will I like it? Providing product overviews based on opinion excerpts,'' in Proceedings of IEEE CEC, pp. 26–33.
[2] U. R. Hodeghatta. 2013. ''Sentiment analysis of Hollywood movies on Twitter,'' in Proceedings of IEEE/ACM ASONAM, pp. 1401–1404.
[3] J. M. Soler, F. Cuartero, and M. Roblizo. 2012. ''Twitter as a tool for predicting elections results,'' in Proceedings of IEEE/ACM ASONAM, pp. 1194–1200.
[4] O. Tsur, D. Davidov, and A. Rappoport. 2010. ''ICWSM—A great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews,'' in Proceedings of AAAI Conference Weblogs and Social Media, pp. 162–169.
[5] D. Davidov, O. Tsur, and A. Rappoport. , 2010. ''Semi-supervised recognition of sarcastic sentences in Twitter and Amazon,'' in Proceedings of the 14th Conf. Computational Natural Language Learning, pp. 107–116.
[6] D. Maynard and M. A. Greenwood. 2014. ''Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis,'' in the Proceedings of the 9th Int. Conf. Language Resource Evaluation, pp. 4238–4243.
[7] A. Rajadesingan, R. Zafarani, and H. Liu. 2015. ''Sarcasm detection on Twitter: A behavioral modeling approach,'' in Proceedings of the 18th ACM Int. Conf. on Web Search Data Mining, pp. 79–106.
[8] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas. , 2010. ''Short text classification in twitter to improve information filtering,'' in the Proceedings of the 33rd Int. ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 841–842.
[9] C. G. Akcora, M. A. Bayir, M. Demirbas, and H. Ferhatosmanoglu. 2010. ''Identifying breakpoints in public opinion,'' in Proceedings of the 1st Workshop on Social Media Analytics, pp. 62–66.
[10] J. Tepperman, D. Traum, and S. S. Narayanan. 2006. ''Yeah right: Sarcasm recognition for spoken dialogue systems,'' in the Proceedings of the InterSpeech, pp. 1838–1841.
[11] E. Riloff, A. Qadir, P. Surve, L. De Silva, N. Gilbert, and R. Huang. 2013. ''Sarcasm as contrast between a positive sentiment and negative situation,'' in the Proceedings of the Conf. Empirical Methods Natural Language Processing, pp. 704–714.
[12] S. K. Bharti, K. S. Babu, and S. K. Jena. 2015. ''Parsing-based sarcasm sentiment recognition in Twitter data,'' in Proc. IEEE/ACM ASONAM, Aug. 2015, pp. 1373–1380.
[13] M. Bouazizi, T. Ohtsuki, 2016, "A pattern based approach for Sarcasm Detection on Twitter," IEEEAccess.