RESEARCH ARTICLE   OPEN ACCESS

Survey on Detection of Sarcasm in Sentiment Analysis

Dr.D.V.Nagarjana Devi*, K.VijayBhaskar**,Y.Pavan***
(Assistant Professor, IIIT-RGUKT Nuzvid
(Assistant professor,IIIT RGUKT Srikakulam,
*** (Student, IIIT RGUKT Srikakulam,*
Corresponding Author: Dr.D.V.Nagarjana Devi

ABSTRACT
Sarcasm detection is one of the important challenge in many areas like sentiment analysis, NLP etc. In sentiment analysis, sarcasm affects the polarity detection process. Sarcasm detection plays a crucial role. It is the unconventional way of conveying a message, which conflicts the context. The research is going on sarcasm detection by different approaches like FeatureBased approach, Rule-Based approach, learning approaches like machine learning and Deep learning, Dictionary-based approach, Hybrid approach, etc... In this paper we have compiled past work done in sarcasm detection and discussed different approaches, tasks, models used on different databases to detect sarcasm and also described how the sarcasm effect the sentiment, challenges and issues in sarcasm detection. Finally, we compared the result of the work in sarcasm detection of different researches done and concluded the better models and approaches which give more accuracy in sarcasm detection.
Keywords: Convolutional Neural Network (CNN), Hash-tag Tokenization (HTT), Sarcasm, SentimentAnalysis, Support Vector Machine (SVM)

I. INTRODUCTION

The present world categorized as a datadriven world. Due to advancements in networking, social platform and micro-blogging, sites like Facebook and Twitter introduced. These playing vital roles. This social media created an open platform to all for communication. In this, present era users are generating vast data in the social media. Sarcasm occurs frequently in user-generated content on the social platform

Most typically, Sarcasm expressed by a positive sentiment used with the negative situation, for instance, if we use any positive verbs like "love, happy, good" subsequently in negative circumstances such as "stuck in traffic, neglected or rejected".

Inherently it is hard to analyze for not only a machine but also a human. At present, sarcasm regularly used to mean a statement when people "state something as opposed to the real world or something in opposition to their real assumptions to be cunning or to arrive at an important resolution", as defined in the BBC sarcasm web page referenced beforehand. (Bousfield, 2007)

[3]Detection of sarcasm in the textual data is a key task of Natural Language Processing. But we can get interpretation between normal and sarcastic statements and can classify easily. In NLP, untruthful estimations can affect the accuracy in getting polarity. For example, a sentence like "So thrilled to go to college on Sunday" could classified as a high positive sentence. Regardless, it is a positive sentence with a negative situation specifies as the negative sentiment. The usage of sarcasm over every social media, e-commerce, and microblogging. Therefore, it is essential to separate noise data from the original data while dealing with NLP related tasks.

[2] Sarcastic phrases are irrelevant and contain emotions and sentiments in the data. Some literature surveys (Bosco et al., 2013; Ghosh et al., 2015; Far´ıas et al., 2016; Nozzaet al., 2017; Kannangara, 2018) include some features of feeling to recognize sarcasm using a state-of - the-art sentiment lexicon.

Subsequent sections in the paper, discuss the Effect of sarcasm in sentiment analysis, dataset mostly used for sarcasm, different approaches used in sarcasm detection and their result and finally challenges are facing in the sarcasm. People face many challenges in the detection of sarcasm, which discussed in detailed in the section VI.

II. SARCASM EFFECT ON SENTIMENT ANALYSIS

Sentiment Analysis is a methodology, which identifies feelings, attitude, opinion and emotion passed by online customers towards a specific thing, instance, products, occasions, issues, organizations, etc…

NLP applications such as sentiment analysis also known as opinion mining, mainly focus the emotional content in the text to identify polarity whether positive or negative. However, sarcasm interprets the sentiment ofthe text.

It acts like an interfering factor that changes the polarity because sarcastic sentences usually give the opposite meaning. Therefore, recognizing sarcasm in the content is significant to develop and improve the accuracy in sentiment analysis. Sentiment Analysis in addition to various NLP application mainly concentrate on text. So, it import to filter sarcasm in the text.

In general, sarcasm will affect the polarity of the opinion expressed. For instance, the reality i.e., "I love the movie and left the theater in middle" ravels us the negative sentiment. According to statement, "I love the movie" express positive opinion but and "I left the theater during the interval" express negative, finally the statement is sarcastic. In above situation, the changes in positive sentiment to negative sentiment, clarifies the contradictory nature of the opinion.

III.   DATASETS

A dataset is a collection of information. Most regularly, a dataset relates to the substances of a single database or data matrix. There are two kinds of data sets i.e, Unbalanced and Balanced dataset. In unbalanced the data points for different classes are unequal. However, equality of data points in classes maintained by the balanced dataset. It is very hard to deal with an unequal dataset. So unbalanced dataset converted as a balanced dataset by using some data preprocessing techniques.

In the recent years, due to advancement in technology social media sites increased for the people communication. Twitter has gained large popularity. The users are communicating by post and short instant messages, called "tweets". Twitter turned as main platform for users to express their frame of mind, feelings and opinion on particular events or products, to share their opinion and to react on real-time issues, etc...

The data generated by Twitter is imperative to numerous partnership companies in addition to government associations to learn sentiment of the people on specific products, films or political events, etc.

As it was a data-driven world, there are multiple types of data. Mostly the data for sentiment analysis collected from social media feeds like posts, hashtags, etc. Reviews of movies, restaurants, and products sold online. According to datasets, there are different sentiment analyses commonly Tweet Sentiment Analysis: Sentiment Analysis of reviews. Sentiment Analysis of Multi-Domain Dataset Sentiment Analysis word-level dataset etc...Datasets

for sentiment analysis often come in the form of user-generated posts on social media (Facebook, Twitter, etc.) and customer reviews.
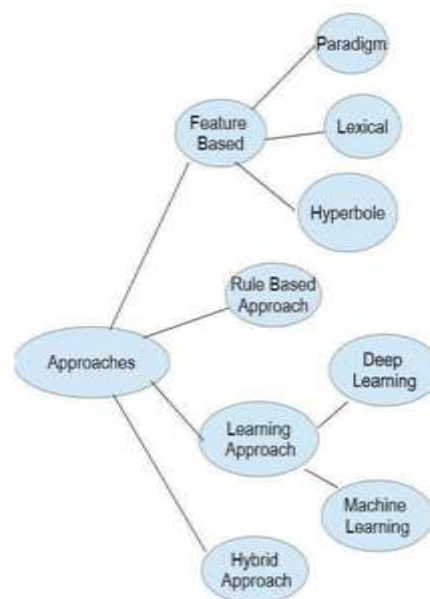
IV.   APPROACHES

Fig1: Approaches

4.1 Feature Based Approach:

In Feature Based Approach, Sarcasm classified by extracting features of text. The features extraction mainly concentrates on four parameters, i.e. features based on feelings, sentence structure, punctuations, and patterns that cover the different forms of sarcasm. There are mainly three types of feature-based classification in this approach, they are:

4.1.1 Lexical feature based classification

Textual characteristics like unigram, bigram, ngrams, etc... Considered as lexical features of the text. The sarcasm defined by these characteristics. (Kreuz et al.) First introduced the concept and noted that lexical characteristics play a vital role in detecting irony and sarcasm in the text.

4.2.2. Pragmatic feature based classification

To shorten the message and duration constraints, symbols and figurative text in tweets are common. These symbolic and figurative texts (such as smiles, emoticons, responses, @user, etc.) called pragmatic features. Pragmatic features are most powerful in Identifying sarcasm in message. Many researchers used this features and got better results in their work to detect sarcasm.

4.1.3 Hyperbole feature based classification:

Set of features like intensifier, interjection (e.g. wow, yay, yeah), quotes and punctuation mark (e.g.",""",!!!!!) are called hyperbole . Hyperbole is an important feature that recurrently used in text data for sarcasm detection. In hyperbolic text common features like intensification, interjection, quotes, punctuation, etc…identified.

In [9] they divided the features used in their models into five features: Tweet Features; Author Features;

Audience Features; Response Features and Environment Features. By comparing the performance of features five feature combinations isconsidered with tweet, they are:

(TWEET FEATURES+RESPONSE); (TWEET + AUDIENCE FEATURES); (TWEET + AUTHOR FEATURES).

4.2.**Rule Based Approach:**

Rule-based methods uses particular evidence to recognize sarcasm. This evidence recorded in aspects of laws based on sarcasm indices. It has two sub-elements: the first focused on semantics; the second focused on statistics. Polarity in sarcastic phrases, parsing and many other are widely used semantic based strategy. Reyes et al. introduced the statistical method. The author is attempting to introduce statistical formula features such as tags, oddness, fashion, mental situations, and voting behavior. (Veale and Hao, 2010) concentrate on the identification of the sarcastic purpose of a specified simile. To determine how probable a simile is, they use Google search. They introduced a 9-step strategy, by using the amount of query outcomes; a simile is validated at each step / rule.

In [1] they removed the personal tokens from the hashtag by developing an algorithm certain conditions. Firstly, they created list of tokens, which matches against the Linux dictionary and transformed into a GATE gazetteer. In order to fit their gazetteers with current object styles like Locations, Organizations, etc. and the popular slang word dictionary used in their Twitter normalization tool ,they digitally revised the slang dictionary after some initial testing to extract most singlecharacter "words" like "h" (but retaining "real" terms like "i", "s" etc.) and few records are treated as non-words.

In [8] Ellen et al. (2013) introduced a technique for identifying sarcasm in tweets by comparing positive sentiments about a negative scenario. This technique uses a bootstrapping algorithm to know sentences related to positive sentiments and negative circumstances, which keep repeating between two stages. The first stage is to learn sentences of negative situations after positive sentiment, where "love" used as a root word. In second stage, they discover structure of positive sentiment expressions that happen close to expressions of negative situations. The resulted list of positive sentiment expressions and negative situations after multiple iteration warehoused. The warehoused data used to recognize sarcasm in tweets by identifying the positive sentiment close to expression of a negative situation according to the context. This technique is supposed to have the previous structure in many sarcastic tweets:

[VERBPHRASE] + [SITUATION PHRASE]

After multiple iteration processes sarcasm in the tweet recognize by identifying the positive feeling close to a expression of a negative situation in context, by using obtained list of negative situations and positive expressions.

4.3 **Learning Approaches**:

4.3.1 Deep Learning Approach:

Deep learning is a class of machine learning work based on the concept of artificial neural network. As deep learning-based models gain success for NLP issues and other classification tasks, few learning-based methods have identified for automatic detection of sarcasms as well.

In [2] they created a model depending on a pre-trained convolution neural network to sarcasm detection by extracting characteristics of feeling, emotion, and personality. These features, along with the base features of the network, enable the proposed models to perform the better calculation on datasets or better perfomance3.2.

4.3.2 Machine Learning:
4.3.2.1 Supervised Learning:

It is a type of machine learning algorithm, which can build a function from a labeled data. Supervised learning can build a good model because of the existence of labeled input value. This can happen because the anticipated outcomes already given in the training data, which need to be processed by the model. Supervised learning also provides the basis for a different algorithm with the same principle as Naïve Bayes, decision tree, logistic regression, etc.

In [3] various supervised learning algorithms used for sarcastic identification, such as Support Vector Machine (SVM), Logistic Regression, Naive Bayes and Decision Tree. In each classifier, if one of the four classifiers says it to be a sarcastic tweet, that tweet further regarded to be sarcastic without checking the outcomes of another classifier. The amount of precision been improved with this method.

4.3.2.2 Semi-supervised learning:

It is a kind of machine learning algorithm, which works as a combination of supervised learning and unsupervised learning. Semi-supervised learning deals with a big quantity of unlabeled information or small quantity of labeled information to create a model. Semi- supervised learning can referred as either Transductive learning or Inductive learning.

In [7] a novel Semi-supervised algorithm used to detect the sarcasm on Amazon customer feedback. The model mainly uses two algorithms: Semi -supervised algorithm to identify sarcastic patterns as features, which assist the classifier and classification algorithm to classify each phrase into sarcastic or nonsarcastic.

They analyzed their model on a big compilation of Amazon customer feedback for distinct product demonstrate excellent outcomes.

4.4 Hybrid Approach:

Combination of two or more different methodologies to create a new and better model known as hybrid methodology. The Hybrid approach includes a mixture of different approaches for better performance and accuracy in the model.

In [6] they used Support Vector Machine (SVM)to rank sarcastic tweets depending on their suggested characteristics and normal Ngrams. Their proposed classifier is a two-SVM ensemble with two distinct sets of features on twitter information and achieved 80% accuracy.

In [8] they used the hybrid approach and Feature Based Approach. By using bootstrapped

eaching methods, the obtained records of favorable feeling and negative behavior and statements proves that these records can used to acknowledge sarcastic tweets.
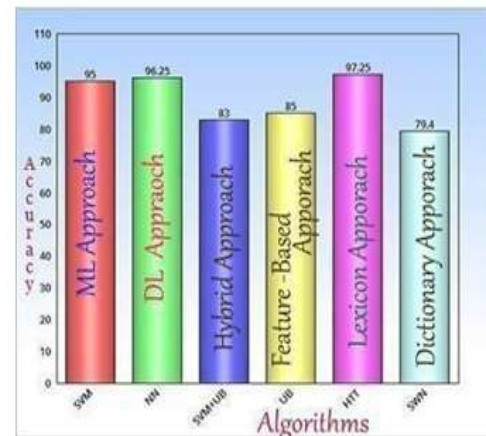
V. RESULT AND DISCUSSIONS

Fig 2: Result based on approaches

Among all theapproaches Lexicon, Deep Learning and Machine Learning Approaches have given maximum accuracy in identifying the sarcastic sentences.HTT model in the Lexical Approach, CNN model in the Deep Learning approach and SVM model in Machine Learning approach has given 97.25 %, 96.25% and 95% accuracy in identifying sarcasm in context.

Table1: Results comparison on different Approaches and models

Ref.no	Database	Approach	Models	Task	Result
[1]	Twitter	Lexicon **Rule based approach**	Hash-tag Tokenization	**Sarcasm classification and polarity detection**	97.25% ; 80%
[2]	Twitter	Deep Learning	CNN	Classification	Balanced -97.71% Imbalanced-94.80%
[3]	Twitter (Dutch tweets)	Machine Learning	Balanced Winnow algorithm	Sarcasm classification	85.00%
[4]	Twitter	Machine Learning	R.F; Gradient Boosting; Decision Tree; Adaptive Boost; L.R; N.B.	Classification and Regression	80.93% 80.60% 73.74% 78.09% 48.29% 75.18%
[5]	Twitter	Machine Learning	SVM; LR; NB; DT.	Classification and Regression	89%; 94%; 78%;94%
[6]	Twitter	Hybrid	SVM classifier with N-Gram	Feature Detection and classification	80.00%
[7]	Amazon	Semi-Supervised	KNN ;SASI	Classification	80.05%
[8]	Twitter	Hybrid(Feature based+ML)	SVM classifier with N-Gram Bootstrapped learning model	Classification and Feature Detection	69% precision 13% recall
[9]	Twitter	Feature based	Uni gram and bi gram; Binary Logistic Regression	Feature detection and classification	85.10%
[10]	Twitter (English &Czech)	Machine Learning	Max Entropy and SVM	Classification	English-94.7% Czech -58.2%
[11]	Twitter	Deep Learning	Deep Neural model	Feature Detection	Balanced-79.56 imbalanced-87.25%
[12]	Movie reviews	Machine learning	SVM	Classification and feature detection	83.00%
[13]	Hindi Tweets	Dictionary/Novel frame work(**feature based?**)	SENTIWORDNET	Mapping	79.40%
[14]	Amazon reviews	Machine learning	SVM and Neural Network	Classification	95%,88%

VI. CHALLENGES FACED IN SARACASM

In order to detect the sarcasm it is essential to comprehend the facts linked to an event. This allows us to detect the contradiction and helps to understand the sarcastic features in the context to detect the objective polarity.

Consider the instance, "I love the pain of breakup," it required high knowledge to detect sarcasm in the statement. In the example, "I love the pain" understands as a positive statement. However, "breakup" is describe the negative situation fallowed by positive sentiment gives the contradiction.

There are some other challenges like extraction of large amount of facts, anaphora while finding the sarcasm in the Statements. Writers prevent the extraction of characteristics and automatically learn the extraction of the features of sarcasm from the dataset using CNNs.The data sets used for the detection method present an issue owing to the unclear and unnecessary abundance of details described in the text. Extracting useful data from the huge amounts of crude data is quite difficult.

• Sarcasm can recognized with specific notation like tone of the speech, body language or facial expression in spoken declaration. However, there are no such clues that make it hard to detect for written text. It's very difficult for many Natural Language Processing systems

• Positive words used in sarcastic statements to express negative opinion, which make different sense.

• In some instances, domain knowledge in the issues and trends is essential to detect sarcasm, e.g. "Mind blowing achievement of the Indian team in Newzealand."

• Sometimes, the hyperbole is used in sarcasm. Hyperbole implies exaggeration, i.e. using superlative degree words, e.g. "Outstanding Performance in work!"

VII. CONCLUSION/FUTURE WORK:

Sarcasm detection research has grown significantly in the past few years. It is necessary to give an overall picture of these individual works. In this paper, we compiled the past work and observed the result of different approaches. Our survey concludes that the CNN model in Deep Learning approach, Hash Tag Tokenization in Lexical Based approach and SVM and NN in the machine learning approach has given max accuracy in the sarcasm identification. We also hilted the issues in sarcasm i.e. Detection of contradiction (Feature extraction).

In sarcastic sentences, there are some positive sentences in it. The identification of sarcasm has grown significantly in the past few years, but detection of sentiment in sarcasm is not industrialized. The detection of sentiment in sarcasm results in high accuracy. Future work left in sarcasm is the detection of combined sentiment to sarcastic and nonsarcastic words, which results in max accuracy

REFERENCES

[1]. Who cares about sarcastic tweets? Investigating the impact of sarcasm on sentiment analysis Diana Maynard, Mark A. Greenwood University of Sheffield, Department of Computer Science ,Regent Court, 211 Portobello, Sheffield, S1 4 DP

[2]. NIHRIO at SemEval-2018 Task 3: A Simple and Accurate Neural Network Model for Irony Detection in TwitterThe perfect solution for detecting sarcasm in tweets #notChristine Liebrecht ,FlorianKunneman ,Antalvanden Bosch Centre for Language Studies Radboud University Nijmegen P.O.

[3]. Box 9103 NL6500 HD Nijmegen c.liebrecht@let.ru.nl

[4]. Sentiment Analysis for Sarcasm Detection on Streaming Short Text Data

[5]. AnukarshGPrasad;SanjanaS,SkandaMBhat,BSHa rishSriJayachamarajendra

[6]. College of Engineering Mysuru, Karnataka, India

[7]. Sarcasm Detection in Plain Text Using Machine Learning

[8]. D.V.Nagarjana Devi, Assistant Professor, IIIT-RGUKT Nuzvid, Dr.T.V.Rajanikanth and

[9]. Dr.V.V.S.S.S.Balaram, Professor, SNIST Hyderabad,TG.

[10]. Recognition of Sarcasm in Tweets Based on Concept Level Sentiment Analysis and Supervised Learning Approaches PiyorosTungthamthiti, KiyoakiShirai,MasnizahMohd Japan Advanced Institute of Science and Technology 1- 1, Asahidai, NomiCity, Ishikawa, Japan923-1292

[11]. Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews

[12]. Oren Tsur , Dmitry Davidov and Ari Rappoport Institute of Computer Science,The Hebrew University Jerusalem, Israel

[13]. Sarcasm as Contrast between a Positive Sentiment and Negative Situation

[14]. Ellen Riloff, AshequlQadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, Ruihong Huang School Of Computing University of Utah Salt Lake City, UT 84112

[15]. Contextualized Sarcasm Detection on Twitter David Bamman and Noah A.Smith School of Computer Science Carnegie Mellon University

[16]. A Study on Sarcasm Detection Algorithms

[17]. LohitaSivaprakasamandAartheJayaprakashSSNC ollegeofEngineeringInternationalJournalofEngine ering Technology Science and Research IJETSR www.ijetsr.com ISSN 2394 – 3386 Volume 4, Issue 9 September 2017

[18]. Tweet Sarcasm Detection Using Deep Neural Network

[19]. MeishanZhang and GuohongFu1 ,School of Computer Science and Technology, 2.YueZhang Singapore University of Technology and Design.

[20]. Sarcasm Detection in Hindi sentences using Support Vector machine

[21]. Prof. Nikita Desai and Anandkumar D. Dave IT Department, Dharmsinh Desai University Nadiad, India

[22]. Harnessing Online News for Sarcasm Detection in Hindi Tweets

[23]. Santosh Kumar Bharti ,KorraSathyaBabu , Sanjay Kumar JenaDepartment of Computer Science and Engineering, National Institute of Technology, Rourkela, Odisha, India.

[24]. SARCASM DETECTION IN ONLINE REVIEW TEXT

[25]. Srishti Sharma and ShampaChakraverty, Division

[26]. of ComputerEngineering,Netaji SubhasInstitute of Technology, India.